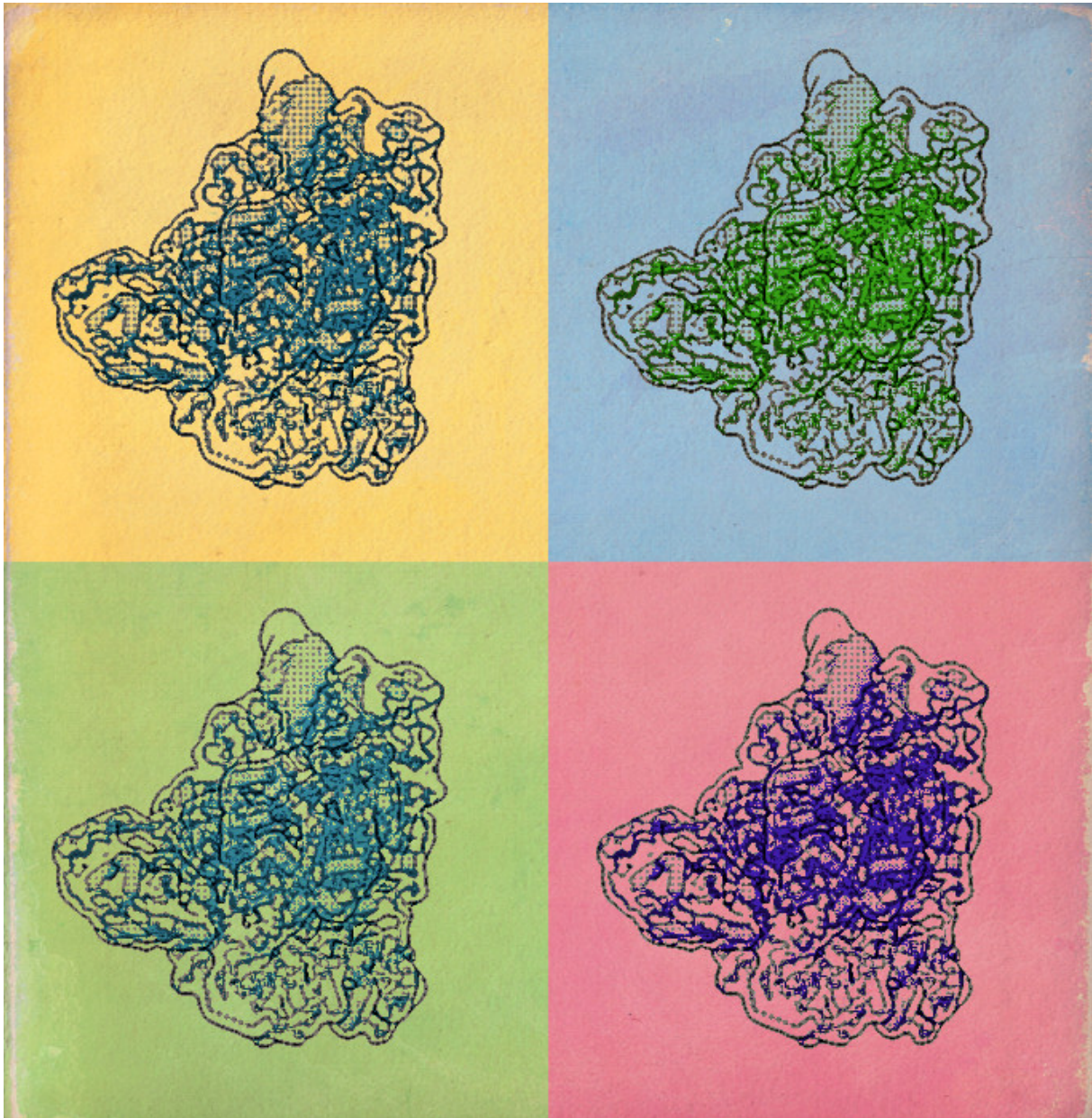


Structural Analysis of Bunyavirus L-protein During Replication and Transcription



**Doctoral Dissertation of:
Sigurður Rafn Þorkelsson
Department of Chemistry
Structural Cell Biology of Viruses
Submitted to the University of Hamburg 2023**

Reviewers:

Prof. Dr. Kay Grünewald

Prof. Dr. Wolfgang Maison

Date of Oral Defence: 11th of November 2023

Oral Defence committee:

Prof. Dr. Kay Grünewald

Prof. Dr. Daniel Wilson

Prof. Dr. Henning Tidow

Co-supervisor:

Dr. Emmanuelle Quemin

Thesis work was carried out at the Centre for Structural Systems Biology in Hamburg from the 1st of May 2019 to 31st of August 2022 in the Department of Structural Cell Biology of Viruses headed by Kay Grünewald and affiliated with the University of Hamburg – Department of Chemistry, MIN Faculty – and the Leibniz Institute of Virology.

List of publications:

1. Dominik Vogel*, **Sigurdur Rafn Thorkelsson***, Emmanuelle RJ Quemin, Kristina Meier, Tomas Kouba, Nadja Gogrefe, Carola Busch, Sophia Reindl, Stephan Günther, Stephen Cusack, Lau Grünewald, Maria Rosenthal. Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L-protein, *Nucleic Acids Research*, Volume 48, Issue 10, 04 June 2020, Pages 5749-5765, <https://doi.org/10.1093/nar/gkaa253>
2. Kristina Meier*, **Sigurdur Rafn Thorkelsson***, Emmanuelle RJ Quemin, Maria Rosenthal. Hantavirus Replication Cycle—An Updated Structural Virology Perspective. *Viruses* 13.8 (2021): 1561. <https://doi.org/10.3390/v13081561>
3. Tomas Kouba*, Dominik Vogel*, **Sigurdur Rafn Thorkelsson***, Emmanuelle RJ Quemin, Harry M Williams, Morlin Milewski, Carola Busch, Stephan Günther, Kay Grünewald, Maria Rosenthal, Stephen Cusack. Conformational changes in Lassa virus L-protein associated with promoter binding and RNA synthesis activity. *Nat Commun* 12, 7018 (2021). <https://doi.org/10.1038/s41467-021-27305-5>
4. Harry M. Williams*, **Sigurdur R. Thorkelsson***, Dominik Vogel, Morlin Milewski, Carola Busch, Stephen Cusack, Kay Grünewald, Emmanuelle R.J. Quemin, Maria Rosenthal. Structural insights into viral genome replication by the severe fever with thrombocytopenia syndrome virus L-protein. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkac1249>

*co-first authorship

Table of Contents

Abstract	6
Acknowledgements	8
1 Introduction	9
1.1 Bunyaviruses	9
1.2 Viral infection cycle	11
1.3 Severe fever with thrombocytopenia syndrome virus	12
1.4 Lassa virus	13
1.5 The L-protein	14
1.6. Principles of electron cryo microscopy single particle analysis	17
1.6.1 Sample preparation	18
1.6.2. EM grid support	20
1.6.3 High-end TEMs	21
1.6.4 Electron sources	23
1.6.5 Electron detectors	23
1.3.6 Image Processing	24
1.3.6.1 Motion correction	25
1.3.6.2 Contrast transfer function	26
1.3.6.3 3D Reconstruction	26
2 Scientific aim	29
3. Cumulative part of dissertation	29
3.1 Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L-protein	29
3.2 Conformational changes in Lassa virus L-protein associated with promoted binding and RNA synthesis activity	48
3.3 Structural insight into the viral genome replication of Severe fever with thrombocytopenia syndrome virus L-protein	67
4. Non-cumulative publications	87
4.1 Hantavirus Replication Cycle – An Updated Structural Virology Perspective	87
5. Own contribution to publications	104
6. Results and discussion	105
6.1 Overall L-protein structure	105
6.2 The RdRp core	106
6.3 RNA interaction	107
6.4 Cap binding domain	109

6.5 Endonuclease	110
6.6 Conclusion.....	111
6.7 Future prospects	113
7. Summary	118
8. References	119
9. Declaration on oath	128

Abstract

Bunyaviruses can be found throughout the globe and include multiple emerging strains that can cause severe disease in humans. They are grouped in a large and diverse order of segmented single-stranded negative sense RNA viruses. The viral genome typically contains three RNA segments: small, medium and large, which encode for the nucleoprotein, the glycoprotein precursor making the spike and the multifunctional L-protein, respectively.

The L-protein is essential for both replication and transcription of the bunyavirus genome. To perform these key functions, it has three important enzymatic domains: an RNA-dependent RNA polymerase (RdRp) domain, an endonuclease domain and a cap-binding domain (CBD).

To better understand bunyavirus infections, our structural analysis focused on L-proteins from phenuiviruses (Severe fever with thrombocytopenia syndrome virus) and arenaviruses (Lassa virus). Structures of these L proteins were solved at high resolution using cryo electron microscopy single-particle analysis and in multiple conformational states by mimicking replication and transcription *in vitro*. We observed precise interactions between the L-proteins and the conserved 5' and 3' ends of the viral genomic RNA including the stabilization of the RNA template during pre-initiation and the formation of the nascent RNA product during elongation. Notably, conformational shifts in the flexible endonuclease and CBD, in combination with the RdRp in particular, represent appealing therapeutic targets. Together with a comprehensive biochemical characterization, these findings allow for significant and novel insights into the mechanism of bunyavirus replication and transcription.

Zusammenfassung

Bunyaviren sind eine große und vielfältige Ordnung von segmentierten negativ Strang RNA-Viren. Sie sind auf der ganzen Welt zu finden, von denen mehrere Spezies beim Menschen schwere Krankheiten verursachen. Das Bunyavirus-Genom enthält in der Regel drei Segmente: ein kleines, ein mittleres und ein großes, die jeweils ein Nukleoprotein, ein Glykoprotein-Spike und das multifunktionale L-Protein kodieren.

Das L-Protein ist sowohl für die Replikation als auch für die Transkription des viralen Genoms unerlässlich. Um diese Funktionen zu erfüllen, hat es drei wesentliche enzymatische Funktionen: Eine RNA-abhängige RNA-Polymerase (RDRP)-Domäne, eine Endonuklease-Domäne und eine *cap-binding domain* (CBD). Die beiden letztgenannten Domänen werden zusammen verwendet, um sich an die mRNA des Wirts zu heften und deren 5'-Cap-Struktur abzuspalten, die beiden letzteren dienen zusammen dazu, sich an die mRNA des Wirts zu heften und deren 5'-Cap abzuspalten, das zur Initialisierung des eigenen Transkriptionsprodukts verwendet wird. Dieser Prozess wird "*cap snatching*" genannt.

Um den molekularen Mechanismus des viralen Lebenszyklus besser zu verstehen, wurden Strukturstudien der L-Proteine sowohl für das Virus des Schweren Fiebers mit Thrombozytopenie-Syndrom als auch für das Lassa-Virus durchgeführt. Die Proteine wurden mit Hilfe der Kryoelektronenmikroskopie in mehreren Konformationszuständen, darunter Apo, Prä-Initiation, Elongation und Transkription, in hoher Auflösung sichtbar gemacht. Diese Strukturen beschreiben die genauen Wechselwirkungen des Proteins mit den 5'- und 3'-Enden der viralen genomischen RNA, die Stabilisierung der genomischen RNA-Vorlage und des entstehenden Produkts sowie die dramatischen Veränderungen des Proteins in der flexiblen Endonuklease und CBD während der verschiedenen funktionellen Stadien. Der RDRP-Kern des Proteins weist eine hohe Ähnlichkeit zwischen den Spezies auf, wo hingegen bei den Endonuklease- und insbesondere bei den Cap-Bindungsdomänen-Motiven eine größere Divergenz besteht. Diese Erkenntnisse ermöglichen einen Einblick in den Mechanismus der viralen Replikation und die strukturelle Identifizierung konservierter Merkmale, die als therapeutische Ziele genutzt werden können.

Acknowledgements

My utmost gratitude goes to my supervisors, Emmanuelle Quemini and Kay Grunewald. I've had tremendous support that built a solid foundation for my research with increasing opportunity for me to take the lead as time went on. They motivated me to always push for excellence and allowed me to explore my own ideas all the while feeling like I had a safety net which I consider to be the essence of a good research environment.

The whole "Structural Cell Biology of Viruses" team has been incredibly supportive and helped me stay motivated. They have been my tribe during my time in Hamburg and hopefully in the future as well. I want to give a special mention to our climbing group. I hope you all stay healthy and happy.

I want to acknowledge my collaborators Maria Rosenthal, Dominik Vogel, Kristina Meyer, Harry Williams, Stephen Cusack and Tomas Kouba. Their knowledge and enthusiasm for bunyavirus L-proteins pushed me to squeeze every possible drop of information out of our data. To my parents for convincing me to seek an education and going down an academic path. My brother for our talks about the oddities of life. My partner, Lilja Harðardóttir, for helping me grow into the man I am today.

And a "thank you" to all the other people who have impacted my PhD journey.

1 Introduction

1.1 Bunyaviruses

Bunyavirales represents a large order of segmented negative sense RNA (snsRNA) viruses. Bunyaviruses are widely distributed around the world and can infect a wide range of organisms from arthropods to plants as well as vertebrates. They have a relatively small genome typically consisting of 3 RNA segments including a small segment, encoding for the nucleoprotein (N), a medium segment, encoding for the precursor of the glycoproteins Gn and Gc making the spike, and a large segment, encoding for the large (L) protein. Each of these genomic segments contains short untranslated regions (UTRs) with high sequence complementarity at the 5' and 3' ends (Kohl et al., 2004). These UTRs are highly conserved within a genus and play a specific role in viral genome replication and transcription (Lowen & Elliott, 2005).

The spherical bunyavirus virions (ca. 100 nm in diameter) consist of a lipid envelope, studded with the glycoproteins organized as spikes, and contain a copy of each genomic RNA segments, coated by multiple copies of N and packed together with L-proteins forming the viral ribonucleoprotein complex (RNP) (Fig.1) (Överby et al., 2008) .

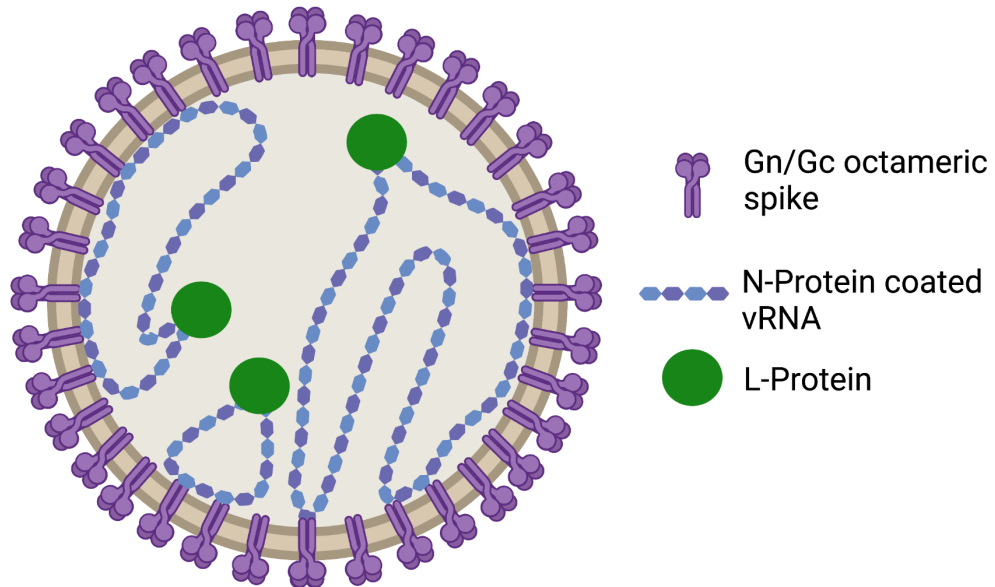


Figure 1. A cartoon representation of a typical bunyavirus virion. The negative sense genomic viral RNA segments are encapsidated by multiple copies of N protein (blue) and attached to a copy of the viral L-protein (green). The virus lipid envelope is decorated by glycoprotein spikes (purple) made of Gn and Gc. Created with Biorender.

Members of the *Bunyvirales* order are typically vector-borne and transmitted by arthropods such as mosquitoes or ticks (Wilson et al., 2022). Exceptions to this are found in the viral families *Hantaviridae* and *Arenaviridae*, which members are transmitted by rodents *via* contact with infected faeces, saliva and/or blood (Sarute & Ross, 2017)(Forbes et al., 2018). Some bunyaviruses can cause severe diseases in humans and livestock resulting in loss of lives and livelihood (M. H. Sun et al., 2022). Examples of these in humans are the nairovirus Crimean-Congo hemorrhagic fever virus (CCHFV) and the arenavirus Lassa virus (LASV). The severity and the extent of the disease symptoms vary between viral species and depend on individual cases. For CCHFV, symptoms can include fever with headache and malaise. Over the span of a few days, the disease progresses to muscle pain, vomiting and diarrhoea with severe cases resulting in death within 2 weeks due to multiple organ failure or circulatory shock (Mostafavi et al., 2014). The mortality rate for such viral infections ranges between 1% and 50% depending on the viral species and socio-economic environment. The molecular mechanisms underlying

the development of these diseases are still poorly understood and there is no effective antiviral strategies like treatment or diagnostic options.

1.2 The viral infection cycle

The Gc/Gn spikes mediate viral entry by binding to a specific receptor on the surface of the host cell. The cellular receptors specifically involved during bunyavirus infections are diverse between viral species and, with a majority still being uncharacterized, their entry pathway remains enigmatic (Hulwswit et al., 2021). Furthermore, some evidence suggest different entry routes such as apoptotic mimicry (*Arenaviridae*) (Amara & Mercer, 2015) or endocytosis (*Hantaviridae*, *Nairoviridae*, *Peribunyaviridae*) (Albornoz et al., 2016) as well as clathrin-dependent (e.g. La Crosse Virus) (Hollidge et al., 2012) and clathrin-independent entry (e.g. Rift Valley Fever Virus) (Harmon et al., 2012).

During the transport of virions in endosomes, there is dissociation of the viral glycoprotein from the cellular receptor. Subsequent maturation from early to late endosome where the pH becomes more acidic, triggers fusion between the viral and endosomal membranes, which is mediated by Gc/Gn and leads to the release of RNPs into the cytosol (Albornoz et al., 2016).

Viral genome replication and progeny assembly take place directly in the cytoplasm of the host cell (Walter & Barr, 2011). Both initial replication and transcription of the viral segmented negative sense RNA (snsRNA) genome rely on the L-protein packed in virions. To replicate the snsRNA genome, complementary RNA (cRNA) of the snsRNA must first be synthesized, the newly synthesized cRNA can then be used as a template to create new copies of the genome. Bunyavirus genomes tend to contain ambisense RNA segments, meaning that there can be oppositely oriented open reading frames expressed from both the genomic RNA

segments (gRNA) and their cRNA with the primary L, M and S being encoded on cRNA (Elliott, 2014).

Current model proposes that the newly replicated viral genome gRNA immediately gets packaged with N proteins to protect it from RNases while the conserved 5' and 3' terminal ends arrange as a panhandle bound to a copy of L-protein to form the RNP (Y. Sun et al., 2018). During transcription, a small segment of cellular capped mRNA is cleaved off by the L-protein endonuclease domain and used to prime transcription products (Olschewski et al., 2020). Viral mRNAs are then translated to proteins by the host cell machinery. Viral assembly happens in the Golgi apparatus, at the Endoplasmic Reticulum-Golgi intermediate compartments (Salanueva et al., 2003) or at the plasma membrane (Goldsmith et al., 1995). The assembly of progeny virions is mediated by interactions between the N-protein within an RNP and the C-terminal ectodomain of membrane-embedded Gn and Gc (Overby et al., 2007). Afterwards the virion is released in the extracellular milieu after fusion with the plasma membrane (Novoa et al., 2005).

1.3 Severe fever with thrombocytopenia syndrome virus

The *Phlebovirus* genus contains a few strains that are well-studied and known to infect humans, including Severe fever with thrombocytopenia syndrome virus or SFTSV in short (now known as Dabie bandavirus) which was first identified in 2009 (Yu et al., 2011). Phleboviruses have a typical bunyavirus virion architecture with three viral genome segments encoding for the L-protein, the glycoprotein precursor, and N (Casel et al., 2021). In addition, a non-structural protein expressed on opposite sense gRNA of the S segment was recently found to interfere with the host cell immune response (Wuerth et al., 2018). SFTSV is transmitted *via* tick bites with *Haemaphysalis Longhornis* considered to be the primary vector. SFTSV is now endemic in east Asia and has been detected in a variety of mammals including domestic and

wild animals (Chen et al., 2019). How the virus is maintained in nature is not entirely clear but it could be mediated by tick-to-tick transmission or through a tick-to-mammal cycle (Zhuang et al., 2018).

A SFTSV infection in human typically lasts between 7 to 14 days. The disease progression can be split into three main stages: Fever stage, Multiple organ failure stage and convalescence stage (Liu et al., 2014). The Fever stage starts after 5 days of incubation with symptoms including acute high fever and thrombocytopenia (Zhan et al., 2017). After ca. 5 days of Fever stage, in severe cases, starts the Multiple organ failure stage with decline in platelet numbers, disseminated intravascular coagulation which can lead death in about 5% of cases (Zhan et al., 2017). If the patient does not develop severe symptoms, the convalescence stage then takes over and the patient recovers over the span of 4-8 days (Liu et al., 2014).

1.4 Lassa virus

Members of the family *Arenaviridae* distinctly differ themselves from related families of other bunyaviruses in that they contain only two genome segments of viral RNA (Ferron et al., 2017). The large segment encodes for the essential L-protein and has an open reading frame on the negative sense strand encoding for a matrix (Z) protein. On the small segment, in addition to expressing the N protein, a truncated version of the glycoprotein precursor is expressed on the opposite strand (McLay et al., 2014). *Arenaviridae* virions have also been found to contain ribosomes from their host, giving the virion a “sandy” texture from which the name *Arenaviridae* is derived (Rowe et al., 1970). Arenaviruses do not rely on arthropod vectors but have natural reservoirs in small mammals and reptiles with their respective genera, *Mammarenavirus* and *Reptarenavirus* (Sarute & Ross, 2017). A zoonotic spread to humans can occur after exposure to aerosolized urine, faeces or saliva from an infected host. There are also reports of human-to-human transmission where the spread may have occurred

through exposure to blood containing high viral titer (Fisher-Hoch et al., 1995). Arenaviruses are classically divided into Old World and New World arenaviruses referencing their geographic location in the African and South American continents, respectively (Emonet et al., 2009). Although both Old World and New World arenaviruses are structurally similar, they differ in terms of disease symptoms and molecular mechanisms. One such difference is the receptor binding for cellular entry with Old World arenaviruses using α -dystroglycan (Cao et al., 1998) while New World arenaviruses use human transferrin 1 (McLay et al., 2014).

LASV is a mammarenavirus and a member of Old World arenaviruses. It can cause a viral haemorrhagic fever (VHF) when it infects humans. The primary natural reservoir for Lassa virus is the Natal multimammate rat (*Mastomys natalensis*) but it can also be found in other rodent species (Olayemi et al., 2016). The virus is endemic in west Africa and is estimated to infect hundred thousands of humans each year (Richmond, 2003). The mode of infection is not clear but is thought to be mainly *via* contact with infected animals (Richmond, 2003).

Lassa fever persists for 1-3 weeks with about 80% of cases being shorter with mild symptoms and the other 20% developing more severe symptoms including VHF (McCormick & Fisher-Hoch, 2002). Symptoms and case fatality rate vary between country which could be related to the prevalence of different viral strains (Garry, 2022).

1.5 The L-protein

The L-protein is essential for both replication and transcription of the viral genome utilizing three primary enzymatic modalities: a N-terminal endonuclease, an RNA-dependent RNA polymerase (RdRp) core and a C-terminal methyltransferase domain. The bunyavirus L-proteins have been shown to be proficient for RNA replication (Vogel et al., 2019) but the molecular mechanisms underlying the regulation of the different enzymatic functions,

recognition of genomic viral RNA and switch between *de novo* replication or primed transcription remain unknown.

Compared to all the bunyavirus genomic transcripts, the L-protein is the most highly conserved (Kapuscinski et al., 2021) and is considered to be a prime target for broad-acting antiviral drugs (Ter Horst et al., 2019). Typically, the encoded L-protein is around 250 kDa in size with exception for members of the *Nairoviridae* family having ~450 kDa L-protein (Ferron et al., 2017). Recently, it was also observed that a bunyavirus infecting fish, the Tilapia virus, has a ~150 kDa L-protein (Subramaniam et al., 2019).

A high resolution crystal structure and a medium resolution cryogenic electron microscopy (cryoEM) structure have been described of a truncated (1-1750 amino acids out of 2263) La Crosse orthobunyavirus (LACV) L-protein. The structures were solved in apo conformation and bound to its conserved 5' and 3' genomic ends, which had previously been shown to be essential for viral replication and transcription (Barr & Wertz, 2004). The L-protein has a right hand-like RdRp core with a peripheral N-terminal endonuclease and a putative methyltransferase C-terminal domain, not resolved in these structures (Gerlach et al., 2015). Overall, the architecture of the LACV L-protein resembles the heterotrimeric polymerase complex from Influenza virus, another ssRNA virus, in terms of domain organization. The overall arrangement of the conserved finger, palm and thumb subdomains and motifs called A, B, C and D is similar (Hussein et al., 2011) (Gerlach et al., 2015).

The 5' end of vRNA from LACV when interacting with the L-protein displays a hook-like conformation and involves residues primarily from the fingertips and finger domains. This 5' hook is formed by two base-pair interactions: G2-C10 and U3-G9 (Gerlach et al., 2015). The comparable 5' hook structure in the case of Influenza virus contains four base pairs. The 3' end interacts with multiple domains (thumb, thumb ring, core lobe, clamp, alpha ribbon 33) and is guided towards the active site of the RdRp core. In addition, the 5' end (nt 11-14) and 3' end (nt 10-13) form a short duplex stabilized by the RNA-binding lobe.

The L-proteins of bunyaviruses have a particular prime using a mechanism called prime-and-realign. After initial *de novo* product priming and elongation of 2- 3 nucleotides, the prime-and-realign mechanism causes a shift of the entire product backwards to the starting position of the template, thereby causing a short product overhang composed of repeats complementary to the initial template nucleotides (Vogel et al., 2019).

The L-protein has a separate mode for transcription. There are some heterologous non-viral sequences identified on the 5' termini of bunyavirus mRNA products (Ferron et al., 2017). Furthermore, based on the presence of a putative CBD on the C-terminal on multiple bunyavirus species, it has been proposed that bunyaviruses rely on a transcription process similar to Influenza virus (Olschewski et al., 2020). For this, the L-protein needs to cleave off a small part of cellular capped mRNA at the 5' ends through a process called cap-snatching (De Vlucht et al., 2018). Cap-snatching occurs for the trimeric Influenza polymerase complex with the CBD, located on the PB2 subunit, bound to the 5' end of a host mRNA. Subsequently and in conjunction with the endonuclease, located on the PA subunit, there is cleavage of 10-15 nucleotides downstream of the 5' end, thereby creating a product that can prime transcription of the viral genome (Walker & Fodor, 2019). When the CBD domain of a bunyavirus protein is expressed by itself, it shows low affinity to capped RNA (Gogrefe et al., 2019b; Rosenthal et al., 2017). In fact, crystal structures of the CBD domains of bunyavirus L-proteins bound to an m⁷GTP nucleotide are structurally similar to the influenza PB2 subunit. In particular, there are conserved aromatic residues in the CBD that mediate the interaction with the cap. However, given the low binding affinity, it can be expected that other parts of the bunyavirus L-protein are involved and/or a larger mRNA product increases affinity (Gogrefe et al., 2019b; Rosenthal et al., 2017).

While L-proteins contain conserved residues related to functional domains (*e.g.* endonuclease and RdRp) there is otherwise a large sequence variation in L-proteins between bunyavirus species (ter Horst et al., 2019). Given these variations in sequence, it could be

treacherous to infer that L-proteins from different species are structurally and functionally homologous to each other (Amroun et al., 2017). Therefore, studying the detailed structure of L-proteins from different families of Bunyaviruses is required to understand the function of these proteins, how they are related to infection and the viral life cycle, and whether a structure based antiviral could be used to target multiple species of L-proteins.

1.6. Principles of electron cryo microscopy single particle analysis

Electron microscopes were originally developed by Ernst Ruska in the 1930s. Since then, they have been a mainstay of scientific research of microscopic phenomena. The concept of imaging with electrons benefits from the low wavelength of the electrons (in the \sim pm range) and the fact that they can be focused using electromagnetic lenses. Electron microscopy imaging techniques can be split between scanning electron microscopy (SEM) and transmission electron microscopy (TEM). In SEM, the electron beam is used to probe the surface of a sample by detecting scattered and secondary electrons. Biological matter generally interacts weakly with electrons and to image it with SEM requires both drying and heavy metal coating to increase the signal to noise ratio at room temperature. When using TEM, the electron beam passes through the sample with electrons being scattered or transmitted as observed in the readout. The 2D image produced in TEM is called a micrograph where the background is bright with detected electrons and the sample casts a shadow onto the background by interacting with electrons and reducing electron signal.

One imaging modality of TEM is cryoEM, which relies on embedding a sample in a thin film of vitrified ice to image biological samples in a hydrated and close-to-native state. To this day, the SNR in cryoEM is so low that it requires multiple computational processing steps including motion correction, contrast transfer function (CTF) estimation and image averaging

to achieve high resolution structure determination. The sample preparation and computational methods used for cryoEM will be discussed in the following chapters.

The two methods of cryoEM that have been particularly insightful in structural biology are single particle analysis (SPA) and tomography. SPA relies on averaging thousands of 2D projection images of a homogenous sample to amplify the high resolution information coming from a specimen (Grassucci et al., 2007) while tomography takes multiple images of the same specimen at different tilt angles to allow for a 3D tomographic reconstruction (Koning et al., 2018).

Over the past decades, multiple advances in hardware, software, and sample preparation have improved the performance of SPA cryoEM drastically, leading to the so-called resolution revolution (Kühlbrandt, 2014) and allowed for biological material in near-native state to be visualized at atomic resolution (Nakane et al., 2020; Yip et al., 2020).

1.6.1 Sample preparation

When imaging organic material with TEM there is a challenge that the sample needs to enter a high vacuum environment in the microscope column, which prevents the imaging of hydrated specimen under standard conditions. Electron microscopists have come up with ways to embed or protect the sample to allow biological samples to be imaged.

Historically this was done by staining the sample with heavy metals (positive staining) with osmium tetroxide, dehydration and plastic embedding. These protocols allow the sample to be thinned by sectioning and survive in a high vacuum (Palade, 1952). Using TEM, these positively stained samples allowed for the description of cellular structures such as organelles (Gray, 1963) and even bacterial ribosomes (Lake, 1976) but some results could be unsound due to the fixation process. Analysis of the effects of these fixations showed that the sample is sensitive to changes in pH (Palade, 1952), interactions with heavy metals (Watson, 1958) and drying (Gusnard & Kirschner, 1977).

Negative staining allowed for the study of individual biological molecules such as virus particles or proteins (Brenner & Horne, 1959)(Ohi et al., 2004). Negative staining works by coating a sample with heavy metal salts such as uranyl acetate, ammonium molybdate or osmium tetroxide. These heavy metals stick to the surface of the molecule, which is then dried, resulting in an enhanced contrast in TEM. When the negatively stained samples are imaged by TEM, the beam mostly passes through the biological material but will not pass through the heavy metal stained surrounding it, resulting in improved contrast. The heavy metal staining also protects the sample from beam-induced damage allowing for practically unlimited exposure in a typical electron microscope (Gallagher et al., 2019).

Negative staining is a relatively fast method with sample preparation being on the minute scale and, when working with proteins, typically uses only 3-4 μL sample at 10-100 nM concentration making it a relatively time and material efficient screening method when assessing sample quality. The method also allows for 3D reconstruction of imaged samples but is limited to the grain size of the heavy metal which sets a high resolution limit of $\sim 15 \text{ \AA}$ or worse. It is also important to keep in mind that artefacts might arise due to the heavy metal staining and drying (Gallagher et al., 2019).

Since most biological molecules are natively immersed in aqueous solution, probing them in a hydrated context should minimize sample preparation artifacts. Under standard freezing conditions, liquid water will crystallize, which not only damages the biological specimens but also scatters the electron beam far more than amorphous water. For liquid water to be frozen in an amorphous state or “vitrified”, it has to be cooled very fast (approx. 10,000 K/s) to 100K or lower. For cryoEM SPA, a sample is vitrified by adding approximately 3 μL of the aqueous sample to a grid. Subsequently, most of the liquid is blotted away with filter paper resulting in a thin sample (5-10 μm thickness), which is then plunged into an efficient liquid heat conductor such as ethane cooled to liquid nitrogen temperature (Dubochet et al.,

1982). By vitrifying biological specimen, one is both able to image the sample using TEM and in a near-native state (Adrian et al., 1984).

1.6.2. EM grid support

Typical cryoEM sample support are referred to as “grids” and have been in use for decades (Stempak & Ward, 1964). Grids are made of a metal support with a circular shape of 3.05 mm in diameter and a pattern of squares with a thin carbon (20 to 40 nm in thickness) layer on top. To make the carbon film hydrophilic, and reduce a doming effect of the sample, the grids are exposed to plasma by glow discharging. For modern cryoEM grids from Quantifoil for example, the carbon film has regularly sized holes (0.6 – 4 μm diameter) with even spacing. These holes allow for the vitrified sample to be imaged with only the buffer solution contributing to the background.

Ideally, when applied to the grid and vitrified, individual particles contained in the sample should be randomly oriented in a single layer of vitreous and free-standing ice inside the holes of the thin support film. There, the individual particles can be imaged without background coming from the sample support. Usually for cryoEM SPA, copper and gold grids are used. When it comes to the holey film, the most common material is carbon for biological material but other types including gold and graphene oxide have proven beneficial for certain samples. Carbon film can be glow discharged and made hydrophobic so that applied samples can spread evenly onto the film before blotting the excess of liquid and plunge-freezing. While carbon tends to crinkle as it shrinks at a different rate than the metal support during vitrification, having both film and base support made of gold eliminates such problem, thereby limiting sample movement during imaging (Russo & Passmore, 2016). In practice, for biological samples that tend to adhere to the carbon films and are therefore not found in the free-standing ice inside film holes, the so-called gold-on-gold grids can be used.

1.6.3 High-end TEMs

The electron microscopes themselves have undergone substantial improvements including better electromagnetic lenses and sample stage stability. For all electron microscopes, it is optimal that they are operated under a vacuum as high as possible to minimize electron scattering during imaging. Other requirements include high voltage to emit electrons and to reach the acceleration needed to penetrate biological material. Electrons travel from the electron source, through an accelerator stack, bringing the voltage of the electrons to 200-300 keV. The electron beam then goes through a series of electromagnetic lenses, apertures, the specimen, and then an electron filter (if present) before hitting the detector.

The electron lenses can be categorized into three types: condenser lenses which are responsible for beam intensity and formation, objective lenses which are for imaging and projector lenses which control magnification (Fig.2) (Raimondi & Grinzato, 2021). The condenser and objective lenses can be followed by apertures in the microscope column, which physically block scattered electrons.

Another important factor is the stability of the stage that is constantly cooled to liquid nitrogen temperature. During image acquisition, a drift of a few Å or more can cause significant blurring of high-resolution signal. On modern high-end microscopes, one can expect that the stage moves less than 1 Å over the course of image acquisition. This motion can be corrected for in conjunction with sample motion caused by the electron beam (see chapter 1.3.6.1).

The microscope components are never perfectly aligned, which causes spherical and chromatic aberrations and reduces the final image quality. Even though these imperfections can be partially corrected post-acquisition by some recent image processing software, they are never fully recovered.

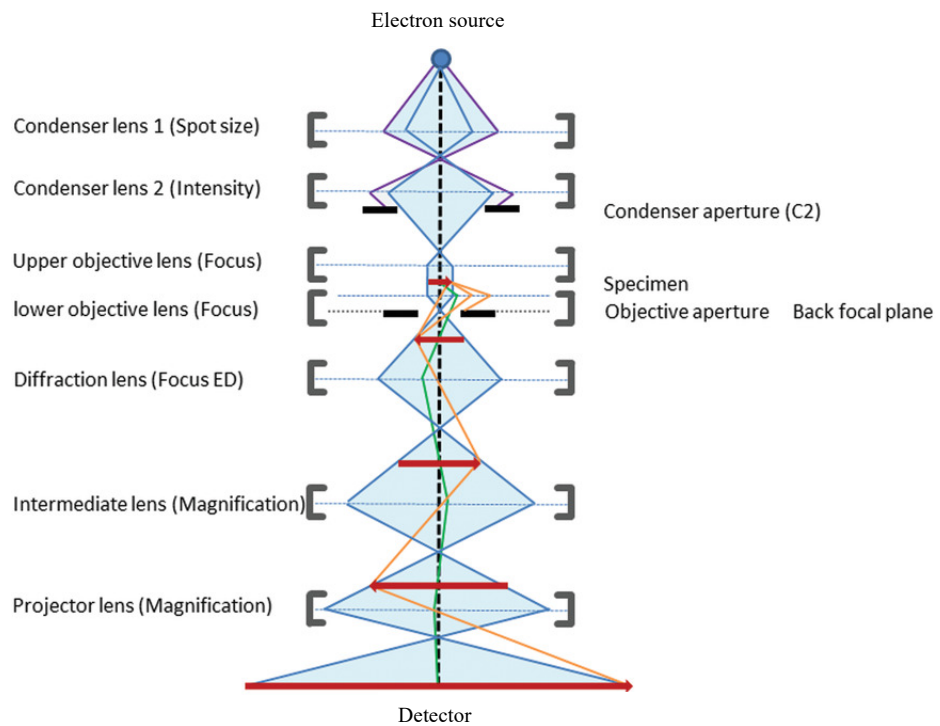


Figure 2. A schematic overview of a transmission electron microscope. An electron source is represented as a blue circle. The microscope's optical axis is represented as black dashed line; lenses are represented as black brackets and column centre are blue dotted lines; physical apertures are indicated by openings between black lines; the back focal plane is represented by a black dotted line. The specimen and its projected intermediates are represented as red arrows; the electron beam is represented as blue lines with filtered/excluded electrons as purple lines and scattered electrons as yellow and green lines. Adapted from Franken et al., 2020.

1.6.4 Electron sources

A primary component of an electron microscope is the electron source. Electron sources emit electrons from a narrow point. The electrons are then accelerated by a so-called accelerator stack which produces an electric field that is perpendicular to an electron emitted from the source. The accelerator stack builds up the electron's voltage, typically in the range of 120-300 keV which allows the electron to penetrate a thin vitrified biological specimen. The ideal electron source would provide a coherent beam with constant brightness.

Current electron sources can be split into either Thermionic sources or Field Emission Guns (FEGs):

Thermionic sources are commonly made of either a tungsten wire or a lanthanum hexaboride (LaB_6) crystal. These materials can be heated to high temperatures to make electrons spill out. The higher the temperature, the higher the emission with the tradeoff of shortening the lifespan of the source. Although these sources are acceptable for negative-staining TEM, they are not coherent or bright enough for high resolution cryoEM.

FEGs were introduced in the early 2000s and were a significant boost to the capabilities of cryoEM (Hewat & Neumann, 2002). FEGs have a sharper electron extraction point compared to thermionic sources and instead of electron emission happening due to electron clashes, they are instead drawn out by an electric field (Franken et al., 2020). Many high-end microscopes are equipped with a FEG which operates at high temperatures (1700-1800K), optimized for beam brightness. A FEG can be operated at room temperature and optimized for energy spread which is called a cold FEG (CFEG) (Nakane et al., 2020). Whether CFEGs outperform high temperature FEGs when it comes to cryoEM SPA is not yet clear.

1.6.5 Electron detectors

The recording of electron microscopy data has historically relied on using charge-coupled devices (CCD) or photographic films. When using a CCD camera for cryoEM, the energy deposited by an electron gets first converted into a photon which travels through a phosphor layer before being detected by the detector (Faruqi & Andrews, 1997). This process allows for multiple scattering events of the photon to happen before detection. It hence limits the spatial accuracy of detection and the signal of a given pixel of the CCD detector can be affected by charging of adjacent pixels, altogether reducing the signal-to-noise ratio (SNR) (Faruqi & Andrews, 1997). Direct electron detectors subvert these problems by containing a single thin layer which serves both for direct detection of electrons and sending the readout, thereby greatly improving spatial detection accuracy and SNR (Faruqi & Henderson, 2007). This leads to improved SNR in collected micrographs while also requiring less electron exposure for the sample.

Given the optimal dose rates (*e.g.* 15 electrons per pixel per second for Gatan K3 camera) and the rapid readout of direct electron detectors (second 1500 frames per second for Gatan K3 camera), one can reliably detect individual electrons on a single pixel level. Since multi-frame movies are collected where individual incident electrons can be detected or “counted”, it allows for correction of the specimen motion after exposure due to the beam warping the carbon film, the metal support bars of the grid and the vitrified sample (Henderson, 1992). By aligning the frames based on specimen movements and creating an average, the resulting motion-corrected micrograph contains higher resolution information than a single frame micrograph without motion correction (Brilot et al., 2012).

1.3.6 Image Processing

Even though raw micrographs can describe details, such as the overall structure of a cell, but with image processing one can enhance data quality and visualize molecules in much higher

detail. The first described three-dimensional reconstruction based on electron microscopy micrographs was published in 1968. D.J. De Rosier and A. Klug used 2D projection images and corresponding diffraction patterns to reconstruct the tail of phage T4 with helical symmetry. The proposed method was similar to 3D reconstruction in the field of X-ray crystallography (De Rosier & Klug, 1968).

Since then, multiple advances in image processing have helped cryoEM become a key method in structural biology. The main hurdles that had to be overcome in the field were how to compensate for the sample damage and motion caused by the electron beam during data acquisition and the low SNR of vitrified biological specimen.

1.3.6.1 Motion correction

Since exposure to the electron beam induces motion in the area of the specimen being imaged, this contributes to the loss of high resolution information on top of beam-induced damage to the particles. Furthermore, stage movement during image acquisition causes additional blurring causing further reduction in high resolution signal. To rescue this precious signal, multi frame movies collected during exposure can be motion corrected, bringing back a part of the blurred signal. This involves finding the shift between each frame of the movie, which can be described as a two-dimensional vector. This vector represents the difference between signal peak positions in a cross correlation map between frames (Li et al., 2013)(Grant & Grigorieff, 2015). To further compensate for anisotropy caused by beam-induced motion, each individual frame can be divided into a matrix (typically 5x5) where the shifts between frames are individually described within each section of the matrix. The resulting information is then combined in a smoothed polynomial function which is used to correct each individual pixel leading to even better reconstructions (Zheng et al., 2017).

There is radiation damage caused to a biological sample during exposure and each individual frame can also be dose-weighted. This assumes that the first frame of a movie

contains almost immaculate signal, while in the subsequent frames, the cumulative exposure of the sample requires the application of an increasing high-pass filter. The end result of dose weighting is that the pristine high resolution signal from earlier frames is preserved and undiluted by much lower SNR in latter frames (Li et al., 2013).

Instead of correcting motion on a whole micrograph scale, one can also do the correction on a single particle level. This polishing step requires tracking the movement of a particle through the exposure period, which can be done by searching for cross correlation of signal between frames (Brilot et al., 2012).

1.3.6.2 Contrast transfer function

All cryoEM micrographs are altered by slight corruptions of the signal transmitted in the projection chamber between the objective lens and the read out on the detector. The aberrations affecting the recorded micrographs are described mathematically as a contrast transfer function (CTF). Although high-end TEMs are quite stable, the alignment is never perfect and result in slight astigmatism and coma aberrations which leads to the loss of high-resolution frequencies in the CTF.

Many parameters of the CTF are known or estimated beforehand like the acceleration voltage and spherical aberration, directly from the microscope settings. Then, there are some approximations for the defocus values used during data collection, amplitude contrast and astigmatism. The latter three can be more accurately assessed while fitting the CTF. Once these parameters are set, software such as CTFFIND can be applied to correct for aberrations in the CTF during preprocessing of the data (Mindell & Grigorieff, 2003).

1.3.6.3 3D Reconstruction

Early image processing and 3D reconstructions from cryoEM images required extensive expertise in the matter. Recently, cryoEM image processing software like Relion (Scheres, 2012b) and cryoSPARC (Punjani et al., 2017) have made the prospect of achieving high

resolution 3D reconstructions much faster, requiring less human input and therefore less prone to bias.

One of the first problems that has to be overcome when reconstructing a protein of interest in cryoEM is how to pick the particles on collected micrographs. There are multiple approaches to particle picking. The first, and still used, method is picking by eye where a person has to view each micrograph and highlight individual particles manually. Manual picking has the upside of being fairly low in picking false-positives (*e.g.* ice contamination) but has the downside of being magnitudes slower than computational alternatives and also subject to individual biases. Manual picking also generally requires image contrast perceivable by the human eye requiring higher defocus values during acquisition, with higher defocus resulting in an unwanted reduction in high resolution signal.

When needing to pick vast amounts of particles, it is usually necessary to use computer automated processes. A simple version of computer automated particle picking is having the computer search for outstanding peaks in signal on a micrograph often referred to as “blob picking”. Optimized blob picking is orders of magnitude faster than its manual counterpart but is unable to distinguish particles of interest from contaminants and, if the sample is less than perfect, tends to give much higher ratio of false-positives. The second automatic method requires one to have a reference of what the particle of interest looks like, from prior 2D or 3D analysis. Such a reference can then be used as input for a template-based search by the computer. Template-based particle picking is much faster than manual picking but the method requires prior knowledge, is subject to template bias and also has a tendency to include more false-positives.

The third picking method is based on machine learning algorithms (Tegunov & Cramer, 2019)(Wagner et al., 2019)(Bepler et al., 2020). Particle picking machine learning agents are initially trained by input coordinates, which can come from either manual picking or template

picking, requiring minimal user input. The agent is then further trained by the user via positively or negatively rating the iterative particle picking cycles. These algorithms are able to match manual picking in accuracy and other automated computer particle picking algorithms in speed but the method is still subject to biases during training and other user input.

Following particle picking, the typical 3D reconstruction algorithm samples the 2D particle projection images against a provided 3D reference volume in Fourier space searching for cross-correlating signal. Modern algorithms improve the initial 3D reference iteratively by applying signal from collected cryoEM 2D particle images. Since the posterior of the 2D image in 3D space is unknown the algorithm universally samples the 2D image against the 3D template at an angular interval. The angular interval is automatically decided between each iteration and will continue until either the reference or the accuracy of angular assignment stop improving (Scheres, 2012a). With a Bayesian approach to automatic image refinement a particle is not assigned a single posterior in the refinement but instead given a likelihood integer for each sampled posterior. This can be done between multiple references which allows for the input particle images to be classified among the different references provided and iteratively refined to improve both the accuracy of the reference and the classification (Scheres, 2012b). This classification approach is an incredible tool for biological samples since one can resolve both conformational heterogeneity of a protein but also heterogeneity of different proteins in a sample or varying complex assemblies.

With bias affecting all current particle picking methods and 3D reconstructions, the cryoEM field came up with a set of practices to avoid the possibility of reconstructing a given template from noise since the SNR is so low. This bias problem can be avoided in 2D classification by using reference-free algorithms (Scheres et al., 2005) and in 3D by low pass filtering templates for particle picking and further reconstruction (Henderson, 2013). The

resolution of a reconstruction is then estimated using Fourier shell correlation cut-off = 0.143 of reconstructed half-maps to minimize noise contribution to the final reconstruction (Scheres & Chen, 2012).

All of these advancements in the cryoEM workflow have allowed for high resolution reconstruction of multiple biological molecules which had previously not been possible by other methods. The high resolution volumes allows for *de novo* models to be built which can give detailed insight into the mechanism of these complex molecular machines.

2 Scientific aim

Current information on bunyavirus protein function is fragmentary. Here, we focus on the L-protein, which is essential for both replication and transcription of the viral genome. In collaboration with the groups of Maria Rosenthal (Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany) and Stephen Cusack (European Molecular Biology Laboratory, Grenoble, France), we aim to resolve the structure of multiple L-proteins in different functional states and in parallel, to characterize the protein biochemically in order to elucidate the molecular function of these proteins. Having detailed insight into the molecular mechanism and regulation of bunyavirus replication and transcription, can help understand the diseases they cause and facilitate therapeutic development.

3. Cumulative part of dissertation

3.1 Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L-protein

In this work, the SFTSV L-protein was analysed using biochemical and structural approaches. We show that the L-protein is a functional polymerase capable of both RNA replication *de novo* and primed transcription. The high-resolution cryoEM structure of the apo SFTSV L-protein showed that the overall structure is remarkably similar to available structures of La Crosse Virus L-protein and, the more distantly related, Influenza polymerase complex considering their sequence divergence. This indicates that an antiviral drug could be effective against a variety of L-proteins.

Original publication and supplement at <https://doi.org/10.1093/nar/gkaa253>

Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L protein

Dominik Vogel^{1,†}, Sigurdur Rafn Thorkelsson^{2,†}, Emmanuelle R. J. Quemin², Kristina Meier¹, Tomas Kouba³, Nadja Gogrefe¹, Carola Busch¹, Sophia Reindl¹, Stephan Günther^{1,4}, Stephen Cusack³, Kay Grünewald² and Maria Rosenthal^{1,*}

¹Department of Virology, Bernhard-Nocht-Institute for Tropical Medicine, Hamburg, Hamburg 20359, Germany, ²Centre for Structural Systems Biology, Heinrich-Pette-Institute, Leibniz Institute for Experimental Virology, University of Hamburg, Hamburg, Germany, ³European Molecular Biology Laboratory, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France and ⁴German Center for Infection Research (DZIF), Partner site Hamburg – Lübeck – Borstel – Riems, Hamburg 20359, Germany

Received February 27, 2020; Revised March 31, 2020; Editorial Decision April 01, 2020; Accepted April 03, 2020

ABSTRACT

The *Bunyavirales* order contains several emerging viruses with high epidemic potential, including Severe fever with thrombocytopenia syndrome virus (SFTSV). The lack of medical countermeasures, such as vaccines and antivirals, is a limiting factor for the containment of any virus outbreak. To develop such antivirals a profound understanding of the viral replication process is essential. The L protein of bunyaviruses is a multi-functional and multi-domain protein performing both virus transcription and genome replication and, therefore, is an ideal drug target. We established expression and purification procedures for the full-length L protein of SFTSV. By combining single-particle electron cryo-microscopy and X-ray crystallography, we obtained 3D models covering ~70% of the SFTSV L protein in the apo-conformation including the polymerase core region, the endonuclease and the cap-binding domain. We compared this first L structure of the *Phenuiviridae* family to the structures of La Crosse peribunyavirus L protein and influenza orthomyxovirus polymerase. Together with a comprehensive biochemical characterization of the distinct functions of SFTSV L protein, this work provides a solid framework for future structural and functional studies of L protein–RNA interactions and the development of antiviral strategies against this group of emerging human pathogens.

INTRODUCTION

Severe fever with thrombocytopenia syndrome virus (SFTSV) is prevalent in East Asia and closely related to the Heartland virus that has been isolated in the US (1). Very recently an SFTSV-like virus has also been detected in bats from Germany (2). SFTSV belongs to the *Bunyavirales* order, established in 2018 to accommodate formerly separated bunya- and arenaviruses (3). Bunyaviruses are a diverse group of viruses with a segmented single-stranded RNA genome of negative polarity. They are distributed worldwide causing zoonoses with several outbreaks annually, mainly occurring in Low-to-Middle-Income countries and primarily affecting poor populations with restricted access to health care. Therefore, several bunyaviruses are listed on the WHO R&D Blueprint (4), a global strategy to enhance preparedness to future epidemics urging for the development of medical countermeasures, such as vaccines and antivirals, which are largely lacking. To develop antiviral strategies against emerging viruses, including SFTSV and other bunyaviruses a profound understanding of the viral life cycle is essential, especially of the similarities and differences within this virus group. The key player in bunyavirus transcription and genome replication is the large (L) protein with a size between 250 and 450 kDa, which contains multiple domains and functions including the viral RNA-dependent RNA polymerase (RdRp). For influenza virus, another segmented negative-sense RNA virus (sNSV), small molecules targeting different functions of the polymerase complex have been clinically approved (5,6). However, in contrast to influenza viruses, whose polymerase has been extensively investigated, structural

*To whom correspondence should be addressed. Tel: +49 40 42818 942; Fax: +49 40 42818 931; Email: rosenthal@bnitm.de

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

© The Author(s) 2020. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

and functional information about bunyavirus L proteins are scarce.

During the processes of genome replication and transcription, bunyavirus L protein synthesizes three distinct RNA species: (i) antigenomic complementary RNA (cRNA), (ii) genomic viral RNA (vRNA) and (iii) capped, mostly non-polyadenylated viral mRNA. Whereas genome replication is believed to be initiated *de novo* by the L protein, viral transcription is dependent on short, capped RNA primers derived from cellular RNAs by a mechanism called cap-snatching (7), which involves cap binding and endonuclease functions. The part of bunyavirus L protein that has been most characterized, both structurally and functionally, is the ~20 kDa endonuclease domain (7–9), which is generally located at the N terminus. Its metal-dependent RNA cleavage activity is required for the cap-snatching mechanism to steal 5' cap structures from cellular RNA and attach these fragments to the viral mRNA in order to be processed by the cellular translation machinery at the ribosome. For Rift Valley Fever virus (RVFV), a bunyavirus of the *Phenuiviridae* family, the C-terminal region of the L protein has been shown to contain a cap-binding domain (CBD), the second activity needed for the cap-snatching mechanism. This CBD is structurally similar to that of influenza virus but with distinct differences in its mode of cap-recognition (10,11). The third important activity is the RdRp, which is needed for both transcription and genome replication. To date, only for one bunyavirus L protein, that of La Crosse virus (LACV, *Peribunyaviridae* family), is more extensive high-resolution structural information available. The LACV structures with viral RNA contain ~77% of the L protein sequence including the endonuclease, RdRp domain and important protein–RNA interaction sites but are entirely missing the C-terminal domain (12). Even though further structural insights into distinct functional states of bunyavirus L protein are still lacking, these structures were a milestone on the path to a complete structural understanding of bunyavirus transcription and replication. In terms of functional data, biochemical characterization of Lassa and Machupo viruses (*Arenaviridae* family) (13–16) as well as LACV (12) L proteins have been published. Specific to each bunyavirus family, the 3' and 5' termini of the genomic RNA segments are highly conserved and almost fully complementary in sequence and are denoted as the conserved promoter regions. For influenza virus and LACV, the RNA 5' end forms a hook-like secondary structure that is bound in a specific pocket adjacent to the RNA synthesis active site of the L protein (12,17,18). The presence of a hook-conformation and its importance for robust RdRp activity have also been proposed for arenaviruses (13,16). Furthermore, for arenaviruses, it was shown that the L protein initiates genome synthesis *de novo* by a prime-and-realign mechanism (13) similar to the initiation of influenza virus vRNA synthesis from cRNA and Hantaan virus genome replication (19–22).

In order to deepen our understanding of bunyavirus genome replication and transcription, we aimed to structurally and functionally characterize the full-length L protein of a virus of the *Phenuiviridae* family. For this virus family, no high-resolution structural data on the L protein core polymerase domain is currently available. We estab-

lished expression and purification of the full-length SFTSV L protein in insect cells using a baculovirus expression system with a yield of up to 11 mg/l of expression culture. We present structural data on the RdRp core and adjacent L protein domains of SFTSV bunyavirus (*Phenuiviridae* family) obtained by electron cryo-microscopy (cryo-EM) and X-ray crystallography. Although we were unable to build the N-terminal endonuclease domain *de novo*, we could fit the recently published SFTSV endonuclease crystal structure (23) into the cryo-EM density. This allowed us to propose an integrated model for the L protein, which was verified using small-angle X-ray scattering (SAXS) data. Although the C-terminal region of the L protein was not resolved in the cryo-EM map, we solved the structure of the SFTSV CBD in complex with a cap-analogue by X-ray crystallography with a resolution of 1.35 Å and characterized how it interacts with cap-structures using biochemical and biophysical assays. We further used biochemical assays to characterize the different L protein activities, such as promoter binding, endonuclease and polymerase activities. In summary, we provide novel structural and functional information on SFTSV L protein that can serve as a basis for an improved understanding of bunyavirus transcription and genome replication processes and aid drug development.

MATERIALS AND METHODS

Cloning, expression and purification of SFTSV full-length L protein

The L gene of SFTSV strain AH12 (GenBank accession no. HQ116417) with a C-terminal StrepII-tag was chemically synthesized (Centic Biotech, Germany) and the plasmid-integrated gene was amplified via PCR. In the same step, primers were used for the introduction of mutations to the gene in the case of the catalytically inactive L proteins (D112A and D1126A). Amplified genes were cloned into an altered pFastBacHT B vector using the In-Fusion HD EcoDry Cloning Kit (Clontech). After transformation of DH10EMBacY *Escherichia coli* cells (kindly provided by Imre Berger), which contain a bacmid as well as a plasmid coding for a topoisomerase, with the pFastBac-plasmids, recombinant bacmids were isolated and transfected into Sf21 insect cells for recombinant baculovirus production. Hi5 insect cells were used for the expression of the StrepII-tagged L proteins. The harvested cells were resuspended in buffer A (50 mM HEPES(NaOH) pH 7.0, 1 M NaCl, 10% (w/w) glycerol and 2 mM dithiothreitol), supplemented with 0.05% (v/v) Tween20 and protease inhibitors (Roche, cOmplete mini), lysed by sonication and centrifuged two times at 20 000 × g for 30 min at 4°C. Soluble protein was loaded on Strep-TactinXT beads (IBA) and eluted with 50 mM Biotin (Applichem) in buffer B (50 mM HEPES(NaOH) pH 7.0, 500 mM NaCl, 10% (w/w) glycerol and 2 mM dithiothreitol). L protein-containing fractions were pooled and diluted with an equal volume of buffer C (20 mM HEPES(NaOH) pH 7.0) before loading onto a heparin column (HiTrap Heparin HP, GE Healthcare). Proteins were eluted with buffer A and concentrated using centrifugal filter units (Amicon Ultra, 30 kDa MWCO). The proteins were further purified by size-exclusion chromatography (Superdex 200, GE Healthcare) in buffer B and ei-

ther used for biochemical assays or SAXS experiments. Purified L proteins were concentrated as described above, flash frozen in liquid nitrogen and stored at -80°C .

Cloning, expression and purification of SFTSV cap-binding domain

The L gene region corresponding to amino acid residues 1695–1810 of SFTSV strain AH12 (GenBank accession no. HQ116417) was cloned into a pOPINF vector (24) using the NEBuilder HiFi DNA Assembly Cloning Kit (New England BioLabs). The protein, which based on sequence alignments corresponds to the CBD of RVFV (11) and therefore referred to as SFTSV CBD, was expressed as a wild-type version or mutated version (with single amino acid exchanges F1705A, Q1707A or Y1719A) in *E. coli* strain BL21 Gold (DE3) (Novagen) at 17°C overnight using TB medium and 0.5 mM isopropyl- β -D-thiogalactopyranoside for induction. After pelleting, the cells were resuspended in 50 mM Na-phosphate pH 7.5, 100 mM NaCl, 10 mM imidazole, Complete protease inhibitor EDTA-free (Roche), 0.4% (v/v) Triton X-100 and 0.025% (w/v) lysozyme and subsequently disrupted by sonication. The protein was purified from the soluble fraction after centrifugation by Ni affinity chromatography. Washing buffers contained 50 mM imidazole and 1 M NaCl or 50 mM imidazole and 100 mM NaCl. Elution buffer contained 500 mM imidazole and the eluted protein was immediately diluted with 20 mM Na-phosphate pH 6.5 followed by passing through an anion exchange chromatography column (HiTrap Q FF, GE Healthcare) and subsequent cation exchange chromatography (HiTrap SP FF, GE Healthcare, loading buffer: 50 mM Na-phosphate pH 6.5, 50 mM NaCl, 10% (w/v) glycerol, elution with salt gradient up to 1 M NaCl). The next step was the removal of the N-terminal His-tag by a GST-tagged 3C protease at 4°C overnight. After addition of up to 3 mM $m^7\text{GTP}$, the protein was concentrated for a final size exclusion chromatography (Superdex 200, 50 mM Na-phosphate, pH 6.5, 150 mM NaCl, 10% (w/v) glycerol). Purified proteins were concentrated using centrifugal devices with addition of up to 4 mM $m^7\text{GTP}$ (final concentration), flash frozen in liquid nitrogen, and stored at -80°C . For thermal shift assays and isothermal titration calorimetry, the purification procedure was done as described above but without any addition of $m^7\text{GTP}$.

Isothermal titration calorimetry

The affinity of SFTSV CBD to $m^7\text{GTP}$ and GTP was measured by isothermal titration calorimetry (ITC) using a MicroCal PEAQ-ITC instrument (Malvern Panalytical). Proteins were dialyzed overnight at 4°C against 50 mM Na-phosphate pH 6.5, 150 mM NaCl, 10% (w/v) glycerol. The ligand $m^7\text{GTP}$ was dissolved and GTP was diluted in the exact same dialysis buffer. Titrations were done with 150 μM CBD in the cell and 5.0–6.5 mM $m^7\text{GTP}$ or GTP in the syringe at 25°C with 19 injections of 2 μl (first injection 0.5 μl). Spacing between injections was constant with 150 s for all measurements. Data were analyzed and fitted with the respective PEAQ ITC evaluation software (Malvern) applying a single site binding model and fixing the stoichiometry value to 1.

Thermal stability assay

Thermal stability of SFTSV CBD was measured by thermofluor assay (25). The assay contained a final concentration of 8 μM of CBD protein, 20 mM Na-phosphate pH 6.5, 100 mM NaCl, SYPRO-Orange (final dilution 1:1000) and either no additive or between 1 and 10 mM of $m^7\text{GTP}$, $m^7\text{GpppG}$, GTP or ATP. Thermal stability of influenza A virus PB2 CBD and RVFV CBD was assessed at a final protein concentration of 10 and 8 μM , respectively, in the same setup as described for SFTSV CBD.

Electrophoretic mobility shift assay

RNAs (Supplementary Table S1) were chemically synthesized (Biomers) and labeled with T4 polynucleotide kinase (Thermo Fisher) and $[\gamma\text{-}^{32}\text{P}\text{-ATP}$ (Hartman Analytic), or ScriptCap $m^7\text{G}$ Capping System (CellScript) and $[\alpha\text{-}^{32}\text{P}\text{-GTP}$. Labeled RNA substrates were subsequently separated from unincorporated $[\gamma\text{-}^{32}\text{P}$ using Microspin G25 columns (GE Healthcare) or ethanol precipitation. RNA was heated for 3 min at 95°C and cooled down on ice to give single-stranded RNA (ssRNA). Reactions containing 0–14 pmol L protein and 2 pmol ^{32}P -labelled ssRNA were set up in 10 μl binding buffer (50 mM HEPES(NaOH) pH 7.0, 150 mM NaCl, 5 mM MgCl_2 , 2 mM dithiothreitol, 10% glycerol, 0.5 $\mu\text{g}/\mu\text{l}$ Poly(C) RNA (Sigma), 0.5 $\mu\text{g}/\mu\text{l}$ bovine serum albumin and 0.5 U/ μl RNasin (Promega)). After incubation on ice for 30 min, the products were separated by native gel electrophoresis using 6% polyacrylamide Tris-glycine gels and Tris-glycine buffer on ice. Signals were visualized by phosphor screen autoradiography using a Typhoon scanner (GE Healthcare) and quantified using ImageJ software (26) when necessary.

Endonuclease assay

Poly-A RNA 27mer and 40mer were chemically synthesized (Biomers) and ^{32}P -labeled with T4 polynucleotide kinase (Thermo Fisher) or ScriptCap $m^7\text{G}$ Capping System (CellScript). RNA substrates were subsequently purified with a Microspin G25 column (GE Healthcare). Reactions containing 2.5 pmol protein and 3 pmol ^{32}P -labeled RNA were carried out in a volume of 10 μl with 0.5 U/ μl RNasin (Promega), 100 mM HEPES(NaOH) pH 7.0, 100 mM NaCl, 50 mM KCl, 1 mM dithiothreitol, and 0.1 $\mu\text{g}/\mu\text{l}$ bovine serum albumin, divalent cations as indicated in the figure legends and incubated at 37°C for 30 min. The reaction was stopped by adding an equivalent volume of RNA loading buffer (98% formamide, 18 mM EDTA, 0.025 mM SDS, xylene cyanol and bromophenol blue) and heating the samples at 95°C for 3 min. Products were separated by electrophoresis on denaturing 7 M urea, 20% polyacrylamide Tris-borate-EDTA gels in 0.5-fold Tris-borate buffer. Signals were visualized by phosphor screen autoradiography using a Typhoon scanner (GE Healthcare).

Polymerase assay

If not indicated otherwise, 5 pmol L protein were pre-incubated for 15 min with 6 pmol of 5' promoter ssRNA (Supplementary Table S1) in assay buffer (100 mM

HEPES(NaOH) pH 7.0, 100 mM NaCl, 50 mM KCl, 5 mM MgCl₂, 0.5 U/μl RNasin (Promega), 2 mM dithiothreitol, and 0.1 μg/μl bovine serum albumin). Subsequently, 6 pmol of 3' promoter ssRNA (Supplementary Table S1) and NTPs (0.8 mM UTP/ATP/CTP and 0.5 mM GTP supplemented with 5 μCi [α]³²P-GTP) were added to a final reaction volume of 10 μl. After 1 h at 30°C the reaction was stopped by adding an equivalent volume of RNA loading buffer (98% formamide, 18 mM EDTA, 0.025 mM SDS, xylene cyanol and bromophenol blue) and heating the sample for 5 min at 98°C. Products were separated by electrophoresis on denaturing 7 M urea, 20% polyacrylamide Tris–borate–EDTA gels in 0.5-fold Tris–borate buffer. Signals were visualized by phosphor screen autoradiography using a Typhoon scanner (GE Healthcare). For primer extension assays, 2.5 nmol RNA oligonucleotides were ³²P-labeled with 6.6 pmol [α]³²P-GTP (10 μCi) and T4 polynucleotide kinase (Thermo Fisher) in a final volume of 20 μl standard reaction buffer. Reactions were stopped by heating to 95°C for 15 min. Primer extension assays were carried out as described for the standard polymerase reaction except 125 pmol of ³²P-labeled primer was used for each reaction instead of [α]³²P-GTP.

Double-stranded RNA separation assay

RNA oligonucleotides (Supplementary Table S1) were chemically synthesized (Biomers) and 3' labeled with T4 RNA ligase (Thermo Fisher) and pCp-Cy3 (Jena Bioscience). Labeled RNA substrates were subsequently separated from excess pCp-Cy3 by ethanol precipitation. RNA was mixed in 1:1 ratio in 100 mM HEPES(NaOH) pH 7.0, 100 mM NaCl, 50 mM KCl, 5 mM MgCl₂, 0.5 U/μl RNasin (Promega), 2 mM dithiothreitol, and 0.1 μg/μl bovine serum albumin and incubated at 30°C for 30 min before an equal volume of RNA loading buffer (98% formamide, 18 mM EDTA, 0.025 mM SDS) was added. The samples were heated at 95°C for 3 min before separation on denaturing 7 M urea, 20% polyacrylamide Tris–borate–EDTA gels and 0.5-fold Tris–borate buffer.

Crystallisation and crystallography

SFTSV CBD crystals grew within 14 days after mixing of one part SFTSV CBD at 18.6 mg/ml protein concentration supplemented with 4 mM m⁷GTP in 50 mM Na-phosphate pH 6.5, 150 mM NaCl, 10% (w/v) glycerol with two parts of 7% (v/v) 2-Butanol, 150 mM MES pH 6.0 and 27.4% PEG 4000 in a sitting drop vapor diffusion setup at 20°C. Crystals were flash frozen in liquid nitrogen without cryo protectants and datasets were obtained at beamline P14 of PETRA III at Deutsches Elektronen Synchrotron (DESY), Hamburg, Germany. Datasets were processed with XDS (27) and the SFTSV CBD structure was solved by molecular replacement using Phaser (28) and the RVFV CBD (PDB: 6QHG) monomer without the β-hairpin and loops as an input model. The structure was refined by iterative cycles of manual model building in Coot (29) and computational optimization with PHENIX (30). Visualization of structural data was done using either the PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC or UCSF Chimera (31).

Electron cryo-microscopy

Aliquots of 3 μl of SFTSV L protein diluted to 0.6 μM in buffer (20 mM HEPES pH 7.0, 500 mM NaCl, 20 mM MgCl₂) were applied to glow-discharged Quantifoil R 2/1 Au G200F4 grids, immediately blotted for 2 s using an FEI Vitrobot Mk IV (4°C, 100% humidity, blotting force –10) and plunge frozen in liquid ethane/propane cooled to liquid nitrogen temperature. The grids were loaded into a 300-keV Titan Krios transmission electron cryo-microscope (Thermo Scientific) equipped with a K3 direct electron detector and a GIF BioQuantum energy filter (Gatan). A total of 2626 movie frames were collected using the EPU software (Thermo Scientific) at a nominal magnification of ×105 000 with a pixel size of 0.87 Å and a defocus range of –1 to –3 μm. Each movie fractionated into 60 frames, was collected for a total exposure time of 3 s with a flux of 15 electrons per physical pixel per second and a total exposure of 59.45 e/Å² with a 20-eV slit for the GIF (Supplementary Table S2).

Cryo-EM data processing and model building

All movie frames were aligned using MotionCor2 (32) and particles were picked automatically using Warp (33). The motion-corrected micrographs were imported in Relion 3.0 (34) and used for contrast transfer function (CTF) parameter calculation with gctf (35). The particles were extracted using coordinates from Warp (~780 000) and subjected to two iterative rounds of reference-free 2D classification. Selected 2D classes representing the different projection views (575 157 particles) were initially classified into six 3D classes with image alignment using an *ab initio* volume created in cisTEM (36) and low-pass filtered to 40 Å, as a reference. The 3D classification with regularization parameter $T = 4$ was performed with 7.5° angular sampling for the first 25 iterations followed by 10 additional iterations using 3.7° angular sampling. One major class was identified as the most populated (223 604 particles) and better defined class, and was selected for auto-refinement and post-processing in Relion resulting in a 3D electron density map (SFTSV map A) with an estimated resolution of 3.8 Å using gold standard Fourier Shell Correlation (0.143 cut-off). The globally refined volume was then used as a template to sort particles again into 10 new 3D classes. This time a distinct class was visible containing additional volume for the C-terminal region (73 357 particles). This class submitted to auto-refine and post-processing and yielded a 3D EM map of 4.3 Å estimated resolution. The model was built *de novo* into SFTSV map A by the Map-to-model program included in Phenix (37) using the core domain of LACV L protein structure (PDB: 5AMQ) as a starting model. The structure was refined by iterative cycles of manual model building in Coot (29) and computational optimization with PHENIX (30).

Small angle X-ray scattering

Small angle X-ray scattering (SAXS) of SFTSV L protein was performed with an in-line size exclusion chromatography on a Superdex 200 Increase 5/150 GL column (GE Healthcare) with a buffer containing 50 mM HEPES(NaOH) pH 7, 500 mM NaCl, 5% (w/v) glycerol

and 2 mM dithiothreitol. Data was collected at the SAXS beamline P12 of PETRA III storage ring of the DESY, Hamburg, Germany (38) using a PILATUS 2M pixel detector at 3.0 m sample distance and 10 keV energy ($\lambda = 1.24 \text{ \AA}$), a momentum transfer range of $0.01 \text{ \AA}^{-1} < s < 0.45 \text{ \AA}^{-1}$ was covered ($s = 4\pi \sin \theta/\lambda$, where 2θ is the scattering angle). Data were analyzed using the ATSAS 2.8 package (39). The SEC-SAXS data were analyzed with CHROMIXS and the forward scattering $I(0)$ and the radius of gyration R_g were extracted from the Guinier approximation calculated with the AutoRG function within PRIMUS (40). GNOM (41) provided the pair distribution function $P(r)$ of the particle, the maximum size D_{\max} and the Porod volume. *Ab initio* reconstructions were generated with DAMMIF (42). Forty independent DAMMIF runs were compared and clustered into five main classes using DAMCLUST (43,44). Models within each class were superimposed by SUPCOMB (45) and averaged using DAMAVER (46). The structures were visualized using UCSF Chimera (31).

Integrated modelling

The SAXS envelopes of SFTSV L protein were visualized using molmap within UCSF Chimera (31) at 15 Å resolution and the SFTSV L protein 3D EM map A was low-pass filtered to 5 Å to smoothen the volume. After filtering, the flexible C-terminal region became visible in the 3D EM map A, which was subsequently used to guide the initial fitting into the SAXS envelope, and further refined with the 'fit in map' function of Chimera. The atomic model for SFTSV apo-L was fit into the 3D EM map A.

RESULTS AND DISCUSSION

Structure determination of SFTSV L protein by cryo-EM

We used a full-length construct of SFTSV L gene (strain AH12, GenBank accession HQ116417) for baculovirus-driven protein expression in insect cells. The L protein was either expressed as wild-type protein or as single-site mutants D112A (endonuclease inactive) or D1126A (RdRp inactive). The protein was purified to homogeneity, with a final yield of up to 11 mg pure SFTSV L protein per litre of cell culture, and used for cryo-EM experiments. Based on map A, derived from the single particle 3D reconstruction and with a resolution of 3.8 Å (Supplementary Figure S1, Supplementary Table S2), the apo-SFTSV L protein (apo-L) structure was determined for ~1100 residues mainly belonging to the RdRp core region (Figure 1). Density for the endonuclease domain was visible but the resolution was insufficient for *de novo* model building indicating mobility of this domain in the apo-conformation of the L protein. The apo-L structure determined by cryo-EM also lacks the C-terminal 500 residues. This domain seems to be very flexible, similar to the case of apo-LACV L, where the C-terminal 500 residues could also not be resolved using cryo-EM (12). Repeated classification resulted in an additional EM map that displayed additional electron density for the C-terminal domain. However, the overall resolution was not high enough to confidently build additional parts of the protein compared to map A. In summary, we present a model for a substantial part of the SFTSV L protein in

the apo-configuration and the cryo-EM map indicates the location and size of the C-terminal region.

SFTSV L protein is similar to the polymerase proteins of other segmented negative strand RNA viruses

SFTSV apo-L is structurally related to LACV L and the heterotrimeric influenza virus polymerase complex, although the similarity to LACV is higher (Supplementary Figures S2–S4). This is consistent with the phylogenetic relation between these viruses, which all belong to the sNSV group, and the previously described structural similarity between influenza virus polymerase and LACV L protein (12). Although the cryo-EM maps of SFTSV apo-L did not contain density sufficiently resolved to build the endonuclease domain *de novo*, it was possible to unambiguously place the SFTSV endonuclease structure recently obtained by X-ray crystallography (23) (residues 1–214 of PDB 6NTV) into the map by rigid body fitting (Figure 1C). The endonuclease domain (residues 1–214) is connected to an influenza virus PA-C like domain via an extended linker (residues 215–286) which wraps around the fingers and palm domains (Figure 1, Supplementary Alignment File). This linker is slightly shorter (~15 residues) than in LACV L and has a more extended conformation (Supplementary Figures S2–S5). Similar to LACV, the PA-C like domain (residues 287–758, with residues 342–358 and 371–445 missing interpretable density) is divided into two lobes, an α -helical 'core lobe' buttressing the palm and thumb domains of the L protein as well as an equivalent to the LACV vRNA binding lobe (vRBL) with a central β -sheet and at least two α -helices packing against both sides of the sheet (Figure 1 and Supplementary Figures S2–S4). A substantial part of this vRBL domain, which constitutes a 3' vRNA binding site in LACV L protein, is missing from the SFTSV structure as it lacks defined density in the map. According to LACV L protein and influenza virus polymerase complex, we expect these unresolved regions to represent the 'clamp', involved in 3' vRNA binding, and the 'arch', making contacts to the 5' vRNA (Supplementary Figures S2–S4). The arch is also disordered in the LACV L structure. The fingers, palm and thumb domains which contain most of the conserved RdRp active site motifs, are downstream of the PA-C like domain (Figure 1D, Supplementary Alignment File). The fingers domain (residues 759–948 and 995–1083) is again highly similar to the LACV counterpart (Supplementary Figures S3 and S4). However, the ribbon insertion (residues 853–870) extending from the fingers domain, is structurally different compared to the LACV α -ribbon and the influenza virus β -ribbon. In SFTSV L, the ribbon insertion is ~40 residues shorter than in LACV L and structurally disordered (Figure 1B, Supplementary Figures S2–S5). The corresponding β -ribbon of influenza virus, which is involved in binding of the RNA duplex region of the promoter and contains the nuclear localization sequence, is ~16 residues longer compared to SFTSV (Supplementary Figures S2–S4). The next specific feature of SFTSV L is the fingertips insertion (residues 913–920), a loop which is mostly conserved among bunyaviruses (47) and contains parts of the RdRp active site motif F (Figure 1D, Supplementary Alignment File). This loop has a different conformation in

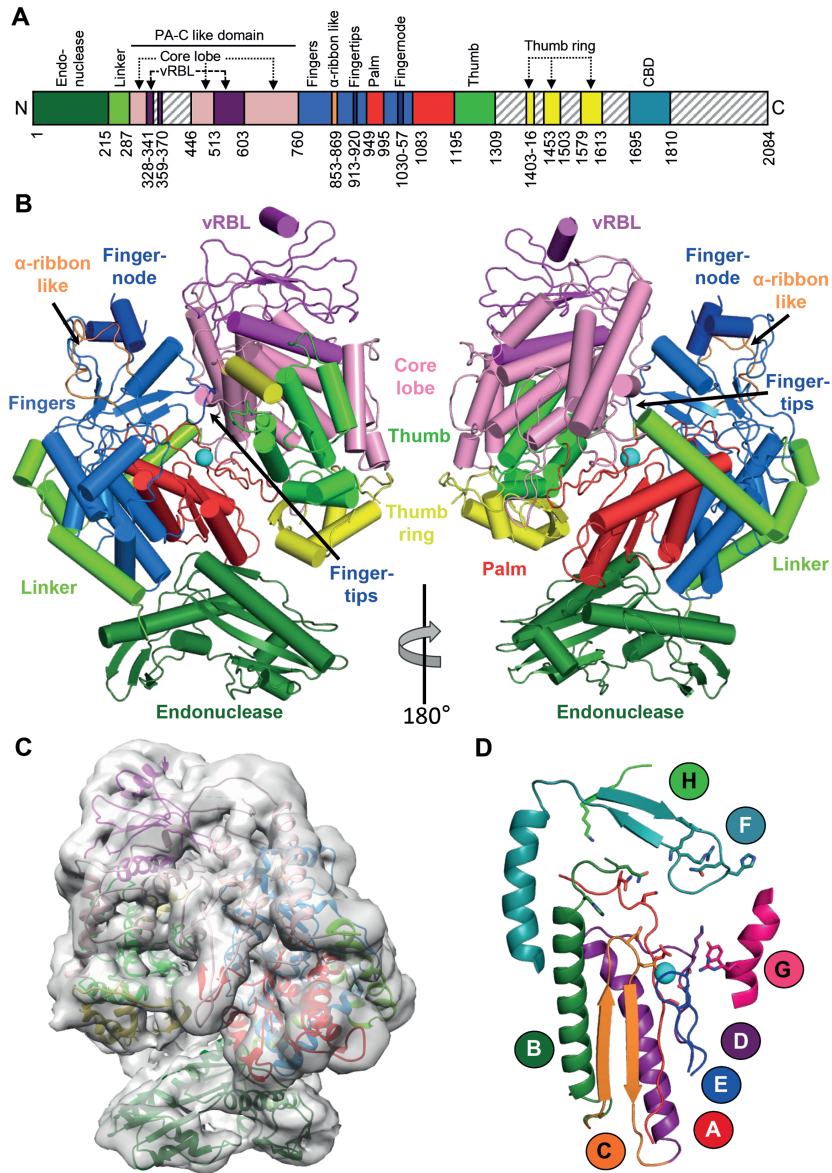


Figure 1. Apo structure of the SFTSV L protein. (A) Schematic linear representation of the domain structure of the SFTSV L protein. (B) Illustrated representation of two views of the apo-L cryo-EM structure as a ribbon diagram (PDB 6Y6K). Domains are colored as in (A). The magnesium ion in the active site is shown as a sphere. A more detailed view and comparison with the LACV L (PDB 5AMQ) and the influenza virus polymerase complex (PDB 6QCX) are shown in Supplementary Figures S2–S4. (C) Superposition of the 5 Å low-pass filtered cryo-EM map with the SFTSV apo-L structure model. (D) Representation of the polymerase core of the SFTSV L with the conserved motifs A–H as a ribbon diagram. The divalent magnesium ion is shown as a sphere.

SFTSV compared to LACV, which is probably related to a conformational change induced by the 5' RNA binding that binds as a stem-loop structure (the so-called 'hook') to the LACV L protein (47). A second insertion into the fingers domain is the so-called fingernode (residues 1030–1057), which is composed of two α -helices and involved in 5' vRNA binding in the LACV and influenza virus structures. A small part of what is presumably the connecting loop between the α -helices is absent in the SFTSV apo-L structure (residues 1039–1045). The fingernode is structurally very similar to the LACV fingernode but different from its counterpart in influenza virus which has a β -hairpin insertion (Supplementary Figures S2–S4). The fingers domain is followed by the palm domain of the RdRp (residues 949–994 and 1084–1194), which is structurally highly similar to LACV and influenza virus (Figure 1, Supplementary Figures S2–S4). The palm domain contains the RdRp active site motifs A and C that are involved in the coordination of the divalent metal ions and the catalysis. Additionally, the active site motifs D and E are part of the palm domain (Figure 1D). In contrast to the LACV palm domain, SFTSV L does not possess the so-called California insertion, whose function is unknown (Supplementary Figures S2–S5). The thumb domain of the RdRp (residues 1195–1308) is composed of a helical bundle and in contact with the previously mentioned PA-C like domain and the vRBL (Figure 1). Residues 1309–1402, which by analogy to the LACV and influenza virus structures presumably compose the priming loop and the bridge domain, are not visible in the SFTSV model (Figure 1, Supplementary Figures S2–S4). We cannot conclude on the length of the priming loop as the sequence identity to LACV is very low for this region (Supplementary Alignment File) and connected parts of the bridge are missing as well (Supplementary Figure S3). The so-called thumb ring is also only partly visible in the apo-L structure of SFTSV (visible parts include residues 1403–1416, 1452–1502, 1579–1587 and 1596–1612). Further regions missing from the cryo-EM structure are the lid and C-terminal domain including the CBD (Figure 1, Supplementary Figures S2–S4, Supplementary Alignment File). Overall, the SFTSV L protein has a very similar architecture as LACV and influenza virus polymerase proteins with some particular differences whose functional relevance will be interesting to determine. Regions missing from the SFTSV L and partly also from the LACV L structure indicate a high degree of flexibility of the C-terminal regions, especially the thumb ring, bridge, and C-terminal domain. Flexibility in the vRBL domain is most likely due to the absence of viral RNA to stabilize the structure but further structural data are needed to clarify this.

3' and 5' promoter RNA binding to SFTSV L protein

An important feature of all sNSV polymerase proteins is the ability to bind to the conserved RNA promoter ends, the almost complementary 3' and 5' termini of the genome segments. For LACV and influenza virus polymerases, distinct 3' and 5' RNA binding sites have been described (12,17). In electrophoretic mobility shift assays, we were able to detect interaction of SFTSV L protein with the conserved termini of all three genomic RNA segments (Figures 2A and

B, Supplementary Figure S6). To avoid RNA degradation during the experiment we used an endonuclease active site mutant (D112A) of the L protein. The migration behaviour of the protein–RNA complex in the native gel was different depending on whether 3' or 5' termini were bound (compare Figure 2A, right and left panel) indicating that 3' and 5' RNA binding induces distinct conformational states of the L protein. The migration behaviour of the complexes is consistent with previous findings on the LASV L protein (13). Interestingly, for the S segment, two different sequences of the conserved termini are published which differ from each other by an insertion/deletion of an A at position 9, counting from the 5' end (Figure 2B). We found that both versions bind to the L protein with comparable affinities (Supplementary Figure S6).

In LACV L protein, the 3' RNA was bound in a narrow cleft leading away from the polymerase active site and formed by the PA-C like domain, the thumb and thumb ring with the clamp of the vRBL serving as a lid over the cleft (Supplementary Figure S7A). The 5' vRNA was found to occupy a separate binding pocket in a hook-like conformation both in LACV as well as in influenza virus (12,17,18). By comparing with LACV L protein bound to the 3' and 5' promoter RNA, we suggest that the RNA binding sites of SFTSV L protein are in the equivalent locations. The 3' RNA is likely bound in a positively charged cleft formed by the thumb and the PA-C like domain, especially the vRBL (Figure 2C, Supplementary Figure S8A). The vRBL domain, however, would have to undergo a slight conformational change, as, in the current conformation, a loop between two β -strands (indicated in Figure 2C by a dashed circle) would block parts of the cleft. The clamp, which is missing clear density in the SFTSV cryo-EM map, indicating flexibility in the apo-L, could close the 3' RNA-binding cleft on the top as observed for LACV (12). Although the clamp is missing we observed potential RNA interacting residues whose location and number in this area would correspond to LACV 3' RNA binding site (compare Figure 2C and Supplementary Figure S7A, Supplementary Alignment File).

The 5' hook RNA binding site in LACV is formed by the PA-C like and the fingers domains. We identified potential 5' RNA-binding residues in SFTSV that closely match in location and number with the residues detected for 5' hook binding in LACV (Figure 2D, Supplementary Figure S7B, Supplementary Alignment File). Similar to LACV we observe clustering of positively charged residues in this pocket that, in terms of size, allows to accommodate the 5' RNA of LACV (Figure 2D, Supplementary Figures S7B and S8B). However, the exact residues interacting with the 5' RNA cannot be predicted, as the 3D structure of the SFTSV hook remains unclear. In an attempt to characterize the possible conformations of the SFTSV 5' RNA, we used the RNA secondary structure prediction program Mfold (48). This resulted in several possible hook structures but without consistency between the S, M and L segment promoters. Moreover, whereas in influenza virus and LACV the 5' hook structure forms within the first 10 nucleotides, this seems rather unlikely for SFTSV 5' termini judging from the predictions (Supplementary Figure S9). It may be that the RNA hook structure is stabilized by base-specific

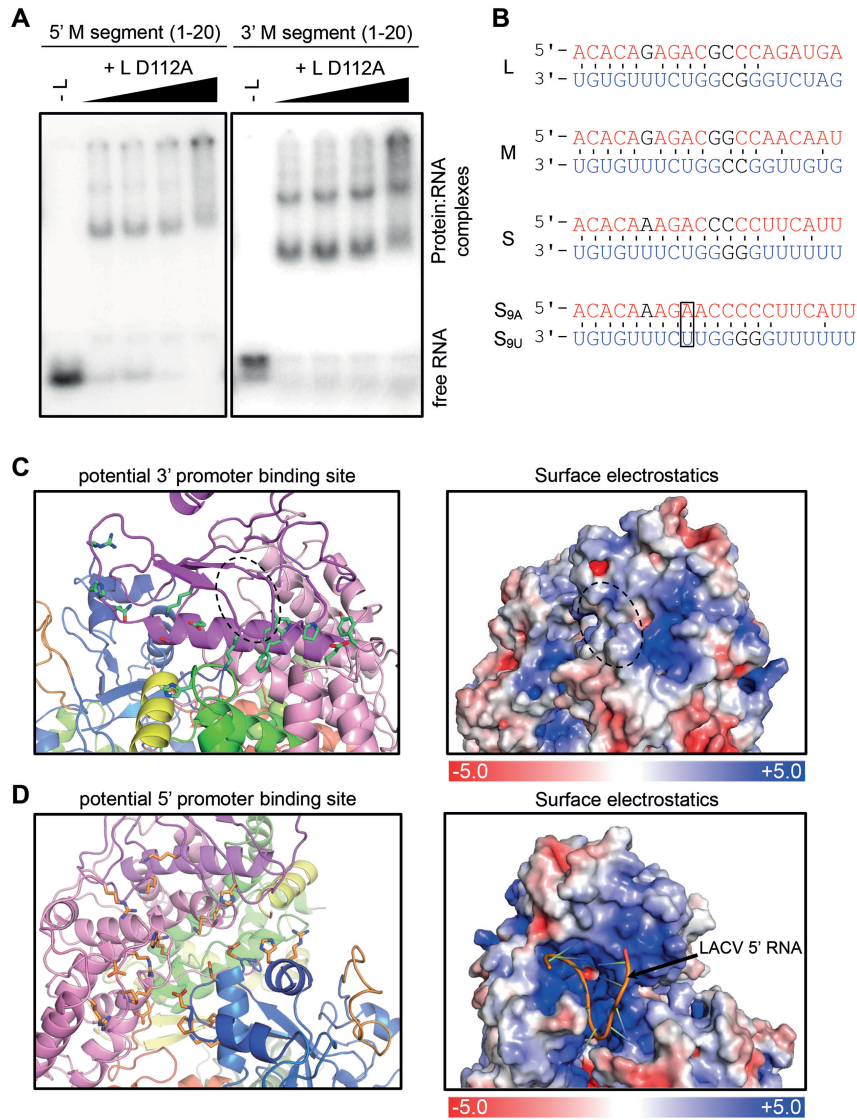


Figure 2. Interaction of the SFTSV L protein with its promoter RNAs. (A) Binding of SFTSV L protein to the 5' and 3' promoter ends (20 nt) of the M segment (Supplementary Table S1) was determined by an electrophoretic mobility shift assay. Increasing amounts of L protein (0–1.4 μ M) were incubated with 0.2 μ M of the indicated RNA (Supplementary Table S1). The protein–RNA complex was separated from the free RNA by native PAGE and visualized by phosphor screen autoradiography using a Typhoon scanner (GE Healthcare). (B) Conserved 3' (red) and 5' (blue) terminal sequences of the L, M and S segments. Watson-Crick base pairing is indicated by black lines. Bases which differ between but are conserved within the segments are shown in black. Additional bases at position 9, found in some sequences (5' S_{9A}, 3' S_{9U}), are marked with a frame. Potential 3' (C) and 5' (D) promoter RNA binding sites are shown as cartoon representation (left). Domains are colored according to Figure 1. Residues potentially involved in protein–RNA interaction are shown as green (C) or orange (D) sticks. See also Supplementary Alignment File. Right panels display surface electrostatics of the potential 3' and 5' RNA binding sites generated by APBS within PyMol. LACV 5' vRNA is shown as a cartoon within the SFTSV potential 5' vRNA binding site. For comparison, analogous figures for LACV vRNA binding sites are presented in Supplementary Figure S7.

protein–RNA interactions rather than base-pairing. However, the sequences reported for SFTSV genome segment termini vary significantly between the L and M segments on the one hand and the S segment on the other hand (Supplementary Figure S10). Therefore, additional structural information is necessary to make reliable conclusions about the putative 5′ hook structure and how it binds to the L protein. In summary, we observe potential 5′ and 3′ RNA binding sites in the SFTSV apo-L protein analogous to RNA binding sites reported for LACV L and also influenza virus polymerase.

The SFTSV L protein is an active polymerase

To characterize the enzymatic properties of SFTSV L protein, we performed *in vitro* RNA synthesis based on the assay conditions established for LASV (13) using a highly purified SFTSV L protein (Figure 3A). To avoid RNA degradation during the enzymatic reactions we used an endonuclease active site mutant (D112A) of the L protein. SFTSV L protein produces a ~35 nt RNA product independent of any primers, which was visualized by autoradiography based on incorporated radiolabelled [α]³²P-GTP after denaturing PAGE (Figure 3B). An L protein carrying a mutation in the RdRp active site motif C (D1126A), which served as a negative control to demonstrate the specificity of the assay, was inactive. The SFTSV L protein RdRp was only active when both the 3′ template vRNA and the 5′ vRNA were present in the reaction (Figure 3B). The activating role of the 5′ end is well known for influenza polymerase, and has also been described for arenaviruses, although the proposed fold of the arenavirus 5′ hook has yet to be confirmed by a structure (13,16–18).

Notably, for the sequences of the conserved genome segment termini, particularly the S segment, different sequences have been published. For the S segment these sequences differ in the insertion/deletion of an A at position 9, counted from the 5′ terminus, or a U in the case of the 3′ terminus (49–51). The S segment RNA including the A at position 9 (denoted as 9A) resulted in much stronger polymerase activity of the L protein compared to the S-segment promoter lacking the 9A, even though the affinity to the L protein seems to be comparable between the two S segment promoters (Supplementary Figures S6 and S11). The reduced polymerase activity with the S segment terminus lacking the 9A is consistent with results from Brennan *et al.* (2015) reporting that it was only possible to establish a reverse genetics system based on SFTSV S segment if the A at position 9 of the 5′ end as well as a corresponding U at the complementary 3′ end were present (51). Even though, we cannot conclude on the role of this 9A residue from our apo-L structure, this should be noted for future studies.

As already described for other polymerases (52,53), the enzymatic activity was also dependent on divalent metal ions. SFTSV RdRp displayed strong activity in the presence of magnesium ions with an activity plateau reached at Mg²⁺ concentrations of >2 mM (Figure 3C), which is similar to what has been reported for LASV RdRp and manganese ion concentrations (13). The presence of 5 mM Mg²⁺, nucleotides and both 3′ and 5′ promoter RNA was defined as the standard reaction conditions for *de novo* replication

by SFTSV L protein and resulted in a single and strong product band after 60 min incubation at 30°C (Figure 3B). In the presence of manganese ions, the RdRp activity was also detectable. However, with higher Mn²⁺ concentrations (>1 mM) the product band got more diffuse resulting either from digestion by the endonuclease (stimulated by the high concentration of Mn²⁺) or from less accurate replication initiation by the RdRp. For arenaviruses, the RdRp activity was greatly enhanced when the 5′ end had a single nucleotide G-overhang compared to the complementary 3′ promoter template strand (13,16), originating from a prime-and-realign mechanism for genome replication. The reason for this enhancing effect was speculated to result from either improved promoter binding or effects on the secondary structure of the 5′ hook. However, such an enhancing effect was not observed for SFTSV L protein (Supplementary Figure S12A).

The product detected in our *in vitro* primer-independent genome replication reactions was larger than expected. This has previously been observed for LASV in *in vitro* polymerase assays (13). It was argued that this could be due to missing termination signals as the assay contains only physically separated short promoter strands of 20 nt rather than the continuous RNA genome comprising all necessary *cis*-acting signals (13). Another possibility is that the template and the completely complementary product form very stable RNA duplexes, which are not separated even by the denaturing PAGE conditions used. We tested this hypothesis using perfectly complementary 20 nt RNAs that were fluorescently labelled instead of radioactively labelled, and performed denaturing PAGE at either 20°C (as usually done) or higher temperatures (60°C). Indeed, we were unable to separate the RNA duplex at electrophoresis temperatures of 20°C even though samples were heated to 95°C and supplemented with denaturing loading buffer prior to denaturing PAGE, whereas at 60°C we detected the RNA at the expected size (Supplementary Figure S12B). Therefore, we can explain the large size of our products in the assays by the difficulty to separate perfectly complementary RNA, a product that cannot be avoided when investigating viral genome replication. However, this does not compromise the specificity of our assay.

In summary, we established a polymerase assay for the SFTSV L protein in which the L protein synthesizes a specific product with the minimal components of Mg²⁺, nucleotides, conserved 3′ template and 5′ promoter strand.

L protein of SFTSV contains an active endonuclease

We tested the SFTSV full-length L protein for endonuclease activity using a ribonuclease assay with a radiolabeled 40mer ssRNA substrate. Substrate degradation in the presence of different divalent metal ions was detected by denaturing PAGE and autoradiography. We found that the endonuclease in SFTSV L protein was active in presence of either manganese or magnesium ions, but not calcium ions (Figure 3D). Residual activity was also detected when zinc, nickel and cobalt ions were added to the reaction (Supplementary Figure S12C). These results do not entirely match with reports on the isolated endonuclease domains of SFTSV and closely related Toscana virus (TOSV), for

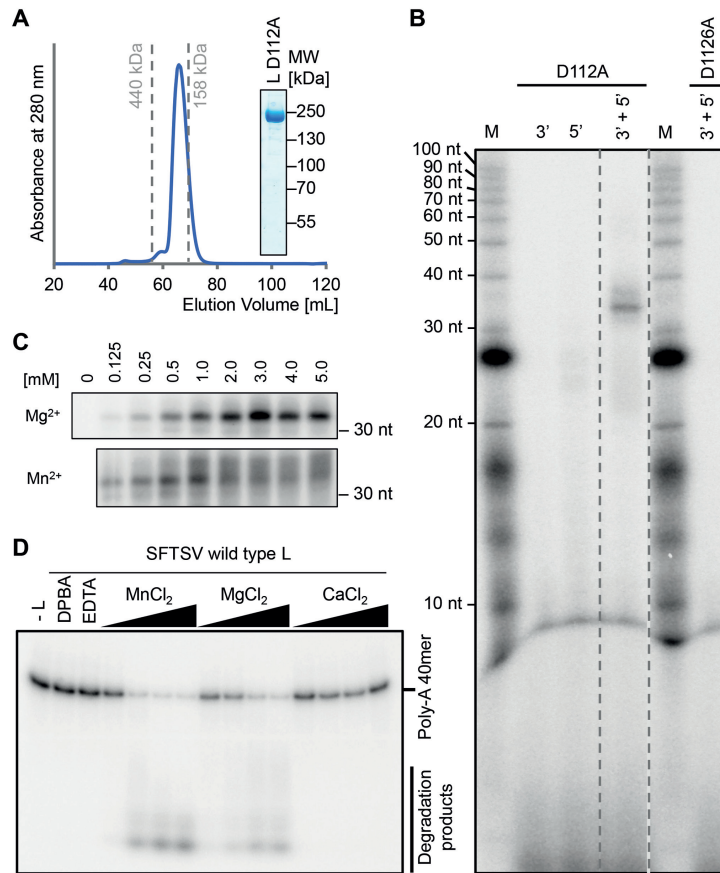


Figure 3. In vitro enzymatic activities of the SFTSV L protein. (A) Size exclusion chromatography and Coomassie stained SDS-PAGE analysis of purified SFTSV L (D112A) protein display the high purity and monodispersity of the L protein. Elution volumes of standard proteins (with sizes of 440 and 158 kDa) for column calibration are indicated. (B) The SFTSV L protein mediates RNA synthesis in an *in vitro* polymerase assay. SFTSV L (D112A) or an RdRp catalytically inactive mutant (D1126A) were incubated with the conserved 5' or/and 3' terminal 20 nt of the M segment (5' M: HO-ACAC AGAGACGGCCAACAAU-OH, 3' M: HO-UGUGUUUCUGGCCGGUUGUG-OH, Supplementary Table S1) in the presence of NTPs supplemented with [α]³²P-GTP for 60 min at 30°C. RNA products were separated by denaturing gel electrophoresis and visualized by autoradiography. (C) The RdRp activity of 500 nM SFTSV L (D112A) protein was analyzed in the presence of the indicated concentrations of MgCl₂ or MnCl₂. (D) 250 nM of wild-type L was incubated with ~0.3 μ M of radioactively labeled PolyA40 RNA substrate (Supplementary Table S1) in the presence of 5, 10, 25 and 50 μ M of the indicated Me(II) at 37°C for 30 min. Reactions without L protein, EDTA or the known endonuclease-specific inhibitor DPBA were used as negative controls. Reaction products were separated on a denaturing polyacrylamide gel and visualized by autoradiography.

which the endonuclease was inactive in the presence of magnesium ions (23,54). As a negative control, we expressed a full-length L protein with a mutation in the endonuclease active site (D112A). Contrary to previous findings on the isolated endonuclease domains of SFTSV and TOSV with mutations of the active site (23,54), the full-length D112A L protein mutant showed some residual RNA degradation activity in the endonuclease assay. We hence added either

EDTA or the known endonuclease-specific inhibitor DPBA (8) to our negative controls (Figure 3D).

In conclusion, we demonstrate that the endonuclease activity of the full-length SFTSV L protein is dependent on divalent manganese or magnesium cations. As in our polymerase and protein-RNA interaction assays we also observed some degradation of the polymerase products and promoter RNA, we conclude that the endonuclease cleaves

both viral and non-viral RNA. This suggests that the endonuclease activity is not sequence-specific and underlines the need for activity regulation on the one hand and the need for protection of viral RNA by binding to the L protein or viral nucleoprotein on the other hand.

Investigation of the mechanism of genome replication initiation

Based mainly on sequencing data from a number of sNSV, a prime-and-realign mechanism has been proposed for *de novo* initiation of genome replication (19–21) and possibly transcription (55–57). There is substantial evidence that the LASV L protein uses a prime-and-realign mechanism during replication initiation (13). LASV L protein initiates genome replication at position +2 of the template strand, produces a dinucleotide primer and realigns the 5' end of this primer to positions -1 and +1 of the template resulting in a single nucleotide overhang of the product relative to the complementary 3' template (13). To characterize the mechanism of genome replication initiation of the SFTSV L protein, we performed primer extension polymerase assays in the presence of different radioactively labelled RNA primers and compared the product size with the size of the *de novo* product (from polymerase reaction with radiolabelled [α] 32 P-GTP) (Figure 4A). In Figure 4C, we provide an overview of all possible products for each of the primers used. In all reactions, we observe only one single product band, indicating that the SFTSV L protein employs only one replication initiation site on the template (Figure 4A). We observe the same product size as the *de novo* product when using ACA or AC primers in the reaction. If CA and CAC primers are provided, the resulting products are ~1 nt smaller than the *de novo* product (Figures 4A and B). Therefore, there are three possible scenarios for genome replication initiation: (i) replication is initiated terminally at position +1 of the template strand, (ii) replication is initiated internally at position +3 and the nascent ACA primer is subsequently re-aligned to position -2/-1/+1 of the template strand or (iii) replication is initiated internally at position +3 and the nascent AC or ACA primer is re-aligned to the terminal position +1/+2/+3 of the template (Figure 4C). For further clarification, we used a longer primer (ACACAAA) for the reaction, which is complementary to the first seven nucleotides of the template strand and should only support terminal initiation (Figure 4C). This primer was incorporated in a product resulting in exactly the same size as the *de novo* product (Figures 4A and B) showing that we can exclude realignment of the primer to position -2/-1/+1. Although it is not possible to reliably discriminate between terminal initiation or priming and subsequent realignment of the primer to the terminal position of the template (Figure 4D), terminal initiation seems more likely as there is only one defined product band detected. In case of priming internally and realigning to the terminal position one would expect to also see a minor product band, two nucleotides shorter than the main product, resulting from missing realignment. At least, this has been observed for LASV L protein, which applies a prime-and-realign mechanism for genome replication initiation (13). We conclude that SFTSV initiates genome replication on a vRNA template either ter-

minally and without applying a prime-and-realign mechanism similar to influenza virus cRNA synthesis (22) or by priming at position +3 and subsequent realignment to the terminal position +1 (Figure 4D). In any case, SFTSV L protein does not seem to produce a single or di-nucleotide overhang of the 5' promoter end compared to the 3' end. A scenario proposed for hantavirus genome replication is that after internal priming and subsequent realignment, a single-nucleotide overhang is removed by the endonuclease resulting in a monophosphorylated 5' terminus (21). However, this mechanism is rather unlikely to occur during SFTSV replication initiation as we used an endonuclease inactive mutant of the L protein in our assays. Even though this mutant showed residual endonuclease activity, full cleavage would be necessary to produce the single product band we observed.

SFTSV L protein contains an active cap-binding domain

In analogy to influenza virus polymerase, the C-terminal region of the bunyavirus L protein has been suggested to contain the CBD that is needed for the cap-snatching mechanism employed by sNSV for transcription priming (7). Recently, the CBD of closely related RVFV has been determined and the residues interacting with a co-crystallized cap-analogue m⁷GTP have been proven to be essential for virus transcription in a cell-based minireplicon system (11). As the C-terminal domain of SFTSV apo-L could not be resolved from the cryo-EM data, we expressed only the putative CBD (residues 1695–1810) in *E. coli*. The purified protein crystallized as a monomer in complex with an m⁷GTP cap-analogue and the crystals diffracted to 1.35 Å resolution (Supplementary Table S3). The SFTSV CBD is structurally very similar to RVFV CBD with a 7-stranded mixed β -sheet, a β -hairpin at the periphery of the domain and a long α -helix packed against the β -sheet (Figure 5A). The key residues responsible for the interaction with the m⁷GTP are functionally conserved among phenuiviruses (Figures 5B, Supplementary Alignment File). The m⁷GTP is stacked between the two aromatic side chains of F1703 and Y1719 extending from the first β -strand and the β -hairpin, respectively. Further interactions are observed between Q1707 from the hinge between the first β -strand and the β -hairpin, L1772 from the end of the long α -helix and the carbonyl group of D1771 (Supplementary Table S4, Supplementary Figure S13). We characterized the interaction of m⁷GTP and the protein by isothermal titration calorimetry (ITC) and thermal stability assays (Figures 5C and D). ITC data demonstrate specific interaction of SFTSV CBD with m⁷GTP cap-analogue in contrast to extremely weak interaction with unmethylated GTP (Figure 5D). Analysis of ITC data provided a dissociation constant of $K_D \sim 138 \mu\text{M}$ for m⁷GTP binding to SFTSV CBD which is 5-fold lower than the K_D observed for the RVFV domain but still quite high compared to influenza virus PB2 CBD ($K_D \sim 1.5 \mu\text{M}$) or cellular cap-binding proteins ($K_D \sim 10\text{--}13 \text{ nM}$) (7,11,58–60). Thermal stability assays revealed a larger shift in the melting temperature (T_m) for SFTSV CBD upon addition of m⁷GTP compared to GTP or ATP (Figure 5C). Consistent with the higher affinity of SFTSV CBD for m⁷GTP determined by ITC, the shift in T_m

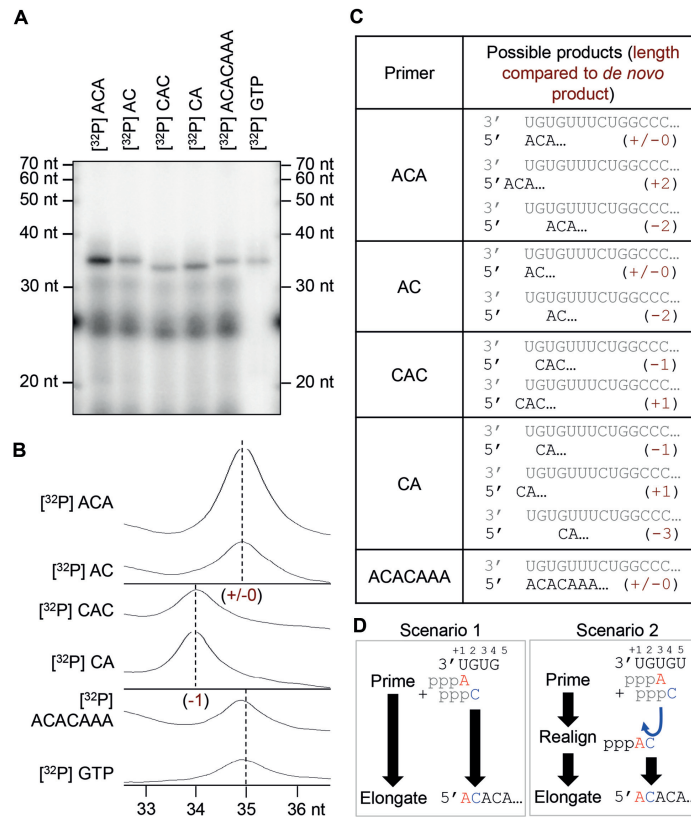


Figure 4. Initiation of replication. (A) RNA products synthesized by the SFTSV L (D112A) protein in the presence of the conserved 5' and 3' terminal 20 nt of the M segment (5' M: HO-ACACAGAGACGGCCAACA AU-OH, 3' M: HO-UGUGUUUCUGGCCGGUUGUG-OH, Supplementary Table S1) and radioactively labeled primers listed in (C). The experiment was performed as described for the standard polymerase assay (see Materials and Methods). Where indicated, radioactively labeled primers were used instead of [α - 32 P]-GTP. (B) The intensity profiles of the gel lanes from (A) were analyzed using ImageJ software (26) and illustrate the size differences between the product bands shown in (A). Sizes were determined by linear regression using the RNA marker (lower X-axis). (C) The table summarizes the RNA oligonucleotides used as primers in this experiment and the possible products with and without realignment. In brackets the length of the product RNA, relative to the *de novo* product is given. (D) Schematic representation of the possible priming scenarios. Scenario 1 depicts terminal initiation: ATP primes the reaction by binding to the first nucleotide (position +1) of the template and is further elongated without realignment, resulting in a product with the same length as the *de novo* reaction product (+/-0). Scenario 2 depicts internal initiation and realignment: the reaction is primed internally by binding of the first ATP to the position +3 and the addition of a C to form a di-nucleotide primer followed by dissociation of the AC dinucleotide and its realignment to position +1 and +2. This realigned AC dinucleotide is then elongated, resulting in the same product as the terminal initiation (+/-0).

was also higher compared to RVFV CBD (compare Figure 5C and Supplementary Figure S14A): +8°C for SFTSV versus +4.5°C for RVFV in the presence of 10 mM m^7 GTP. We observe this difference between SFTSV and RVFV in both assays even though the number of interactions between the protein and the m^7 GTP ligand is only slightly higher in SFTSV compared to RVFV (compare Supplementary Table S4 with data from Gogrefe *et al.*) (11). To provide additional evidence for the essential role of residues interacting with m^7 GTP, we expressed and purified individual F1703A, Y1719A and Q1707A mutants of the CBD and tested them

in the thermal stability assay (Supplementary Figure S14B). As expected, the mutated CBDs were not significantly thermally stabilized in the presence of m^7 GTP and the T_m of the mutated proteins was ~3–4°C lower than the T_m of the wild-type CBD indicating overall lower stability of the domain upon single mutation.

Although the cap-binding ability of the isolated domain seems to be clearly present we were unable to establish cap-dependent transcription assays for the full-length SFTSV L protein. In the presence of a 16 nt primer, capped or uncapped, designed to hybridize with the three nucleotides at

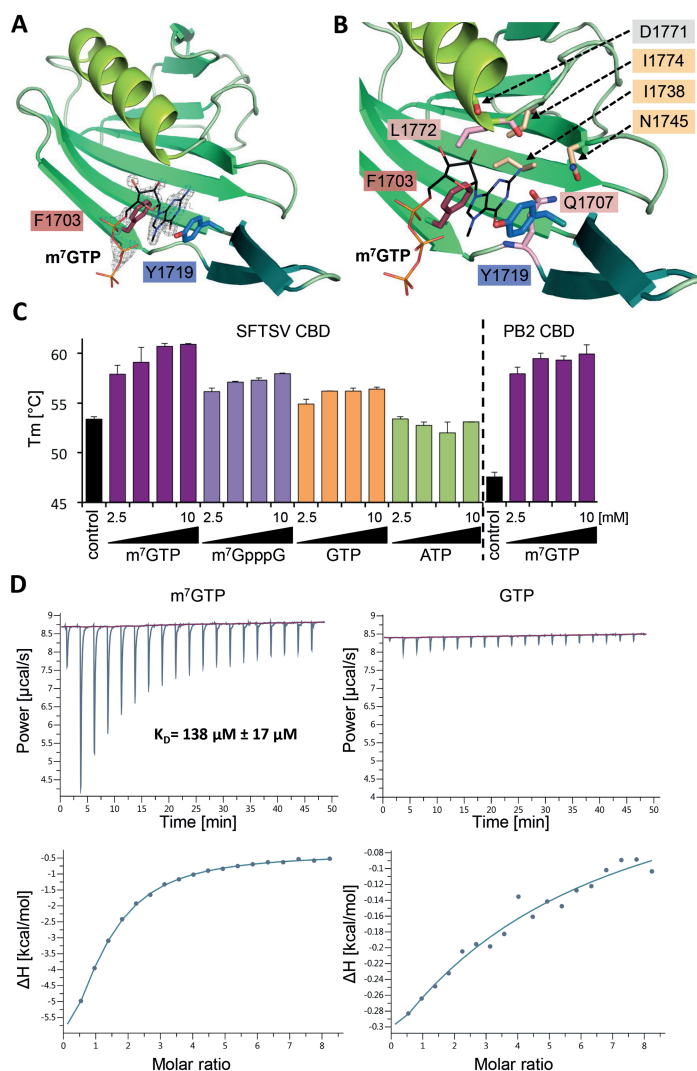


Figure 5. Structure of SFTSV CBD and m⁷GTP binding. (A) The figure shows SFTSV CBD crystal structure in complex with an m⁷GTP. SFTSV CBD is presented as a ribbon diagram with the side chains of the two aromatic residues (F1703, Y1719) involved in stacking interaction with the m⁷GTP ligand shown as sticks. m⁷GTP is presented as lines and the surrounding electron density (2I_oI_c-omit map at 2.0σ) as gray mesh. (B) In a close-up of the m⁷GTP binding site, the protein is shown as ribbon diagram and side chains involved in m⁷GTP binding as well as the carbonyl oxygen of residue D1771 are presented as sticks. The m⁷GTP is shown as lines. The residues I1774, I1738, N1745, are involved in stabilizing the binding site cavity, D1771 (carbonyl oxygen), Q1707 and L1772 directly interact with m⁷GTP. A detailed list of interactions between the CBD and the m⁷GTP ligand is given in Supplementary Table S4 and a ligand plot in Supplementary Figure S13A. (C) Thermal stability of SFTSV CBD and influenza A virus PB2 CBD was tested in the presence and absence of different concentrations (2.5, 5.0, 7.5 and 10 mM) of m⁷GTP, m⁷GpppG, GTP and ATP. The melting temperatures (T_m) are presented as mean and standard deviations of three independent measurements (n = 3). (D) The affinity of SFTSV CBD for m⁷GTP and GTP was measured by isothermal titration calorimetry at 25°C. A representative titration curve is shown. Titrations were done three times with 150 μM SFTSV CBD in the cell and 5.0–6.5 mM m⁷GTP or GTP in the syringe. The upper panel shows the raw data, the lower panel the integrated data fitted to a single-site binding model with the stoichiometry fixed to 1. The dissociation constant K_D is given as a mean and standard deviation of three independent measurements for m⁷GTP.

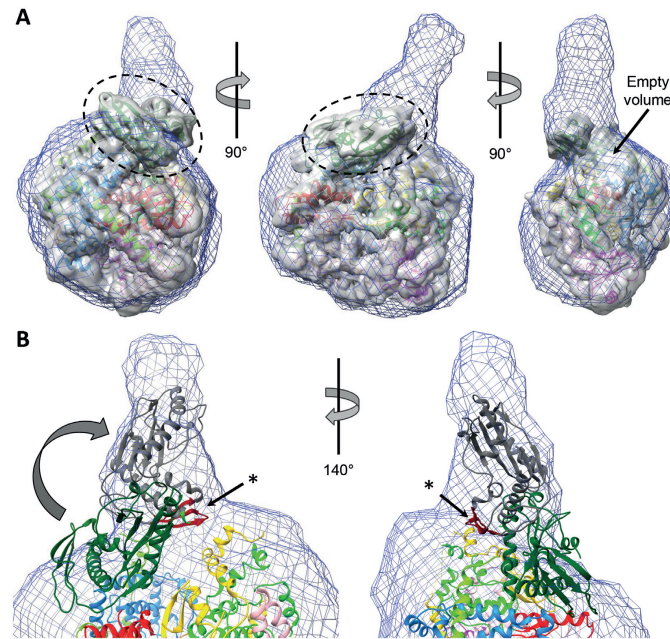


Figure 6. Integrative modelling of SFTSV apo-L protein. (A) A superposition of the SAXS envelope of cluster 1 (blue mesh), a 5Å low-pass filtered cryo-EM mapA (grey surface) and the structure model (ribbon diagram, colored according to Figure 1) is presented in three different orientations. The endonuclease domain, which is sticking out of the SAXS envelope, is marked by a dashed circle. An empty volume is also indicated. (B) Close-up of the endonuclease domain in the superposition of SAXS envelope and structure model. Potential movement of the endonuclease domain is indicated by an arrow and an alternative conformation depicted as grey ribbon diagram. An asterisk indicates the β -sheet (colored in red), for which a role in protein-protein interactions has been proposed.

the 3' end of the template strand, the L protein synthesized a product which was about ~12–16 nt larger than the *de novo* product (Supplementary Figure S15A). This result indicates that the primer was incorporated into the final product, but independent of the need for a 5' cap.

Binding of a capped primer to the CBD of the L protein could lead to endonuclease cleavage products of specific length depending on the distance between the CBD and the endonuclease active site. Therefore, we tested for cap-dependent endonuclease activity using poly-A RNA with either cap0 (m⁷GTP), cap1 (m⁷GpppNm) or no cap at the 5' end but did not detect any specific cleavage product (Supplementary Figure S15B). In summary, we provided evidence that residues 1695–1810 form a functional CBD within SFTSV L protein that is structurally similar to RVFV and influenza virus CBD, but were unable to demonstrate any cap-dependent polymerase activity of the full-length L protein. It remains unclear what activates the cap-binding function of the full-length L protein. Interaction with host factors or the viral nucleoprotein might be necessary for the L protein to switch to transcription mode (7). This is also consistent with the comparably low affinity detected for m⁷GTP binding to the CBD *in vitro*. Similar results have been reported for RVFV CBD (11). Further

studies are required to fully elucidate how bunyavirus cap snatching and cap-dependent transcription works.

Integrative modeling

We used the pure, monodisperse and monomeric full-length SFTSV L protein to perform SAXS experiments and obtain a low-resolution structure of the L protein in solution (Figure 3A, Supplementary Figure S16A). Three representative SAXS models were obtained by clustering analysis of forty *ab initio* dummy atom models and averaging of the structures within each of the three biggest clusters. All three models feature a compact core domain with a hollow center that is decorated with at least one protruding sub-domain (Supplementary Figure S16B), as observed for LASV L protein previously (13). However, the three SAXS models differ slightly in the size and location of a second protrusion from the core domain. We used these SAXS models for integrative modelling of an SFTSV L protein containing the described incomplete SFTSV cryo-EM structure (Figure 6A, Supplementary Figure S17). The cryo-EM map overall fits the SAXS envelopes, to the exception of the volume of the endonuclease domain (Figure 6A, dashed circle) indicating mobility of this domain relative to the polymerase core in

solution. As mentioned above, the endonuclease domain structure was added to the model by rigid-body fitting of the recently published crystal structure of the isolated domain (23). As our model of the L protein core region and the endonuclease crystal structure overlap within 14 residues, we were able to connect these two structures and include the endonuclease in our final model. These overlapping 14 residues form a helix, which has been demonstrated to be very flexible in its position relative to the endonuclease core, suggesting a role in regulation of the endonuclease activity by controlling access to the active site (23). The structural data presented here suggest classification of this flexible helix as the endonuclease-polymerase linker region (Figures 1A and B). In the SFTSV apo-L structure, this linker is in an extended conformation (Supplementary Figure S4A). However, it is conceivable that the linker can also be present in a more collapsed conformation as observed in LACV and influenza virus polymerase proteins, depending on the functional state of the L protein, which would be compatible with the hypothesized function in endonuclease activity control. Indeed, in the integrated model, it is conceivable that the endonuclease domain can rotate relative to the polymerase core and thereby fill the larger protrusion of the SAXS envelope (Figure 6B). In that state, the part of the endonuclease denoted as additional β -sheet in phenuivirus endonucleases, which was predicted to play a role in protein-protein interactions (23), could make contacts to the polymerase core or unresolved C-terminal region (Figure 6B).

Focussing on the C-terminal region of the L protein, in both the cryo-EM map as well as the SAXS envelopes, we observe low-resolution density volume into which it was not possible to build a structural model *de novo* (Figure 6A, indicated empty volume). This volume likely contains the C-terminal region especially missing parts of the thumb ring, bridge and lid. Notably, in the crystal structure of the arenavirus L protein C terminus, the long linker connection of the CBD to the L protein probably enables high mobility of this domain (10). The second protrusion of the SAXS envelope, which is less pronounced and not visible in all structures, might correspond to the CBD. However, this remains purely speculative based on the data we have and we therefore did not include the CBD in the model.

Combining high- and low-resolution structural data from cryo-EM and SAXS, we showed that the conformation of the polymerase is globally similar in solution and in cryo-EM. Furthermore, it is conceivable that the endonuclease is able to rotate with respect to the polymerase core.

CONCLUSIONS

Here we provide a comprehensive characterization of the SFTSV L protein structure and function using a combination of cryo-EM, X-ray crystallography, SAXS and biochemical assays. The structure of SFTSV L protein in the apo conformation closely resembles the LACV L protein and influenza virus polymerase complex structures and, by analogy, allows prediction of the RNA binding sites in SFTSV L protein. Notable differences between these three viral polymerases include the length and conformation of the ribbon-like insertion, the position of the endonuclease domain relative to the polymerase core, and the conforma-

tion of the endonuclease linker region. In particular, the ribbon-like insertion has been speculated to be involved in L protein-nucleoprotein interactions, an interface that must be highly specific for each virus (12,18). Integrating structural data from SAXS and cryo-EM experiments, reveals the likely flexibility of the endonuclease domain position relative to the polymerase core in solution. These observations are consistent with recently published analyses (23) and might explain the different positioning of the endonuclease relative to the polymerase core between LACV and SFTSV L structures. The C-terminal part of the bunyavirus L protein seems to be highly dynamic and was not well-resolved in the cryo-EM map. This has been previously observed for influenza virus PB2 (61) and LACV L (12). Mechanisms for stabilization of this region by viral RNA or other factors have to be defined in future studies of functionally relevant stages, i.e. initiation, elongation and termination of transcription. Additionally, the expression and purification procedures as well as the biochemical assays established here will foster further structural and functional studies on bunyavirus L proteins. We demonstrated that the L protein of SFTSV binds to both 3' and 5' promoter RNA *in vitro* inducing distinct conformational stages as concluded from electrophoretic mobility shift experiments. Furthermore, we show that SFTSV L likely initiates genome replication on vRNA *de novo* without applying a prime-and-realign mechanism and that cap-dependent transcription requires an unknown switch. Altogether, the structural and functional data presented here on the L protein of SFTSV advance our understanding of this complex and multi-domain protein that is essential for viral replication. We also provide significant insights into the commonalities and differences between sNSV polymerase proteins, which will be particularly important for the development of broad-acting antivirals.

DATA AVAILABILITY

Structural data are available from the PDB database (accession numbers 6Y6K and 6XYA). EM data have been deposited with the EMDB (accession number EMD-10706). SAXS data are available from the SASBDB database (accession number SASDHQ8). All remaining relevant data are included in the manuscript and its supplementary material.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The synchrotron MX data were collected at beamline P14 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We would like to thank Thomas Schneider and Isabel Bento for the assistance in using the beamline and Isabel Bento additionally for helpful advice. The synchrotron SAXS data were collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We would like to thank Tobias Graewert, Haydyn Mertens and Cy Jeffries for the assistance in using the beamline. We also thank

the team of the EMBL Hamburg Sample Preparation and Crystallization (SPC) facility for support of ITC measurements. The authors thank Imre Berger for providing the DH10EMBAcY *E. coli* and the team of the Eukaryotic Expression Facility (EEF) at EMBL Grenoble for support and advice.

FUNDING

Leibniz Association, Leibniz competition programme [K72/2017]; Wilhelm und Maria Kirmser-Stiftung; Part of this work was performed at the Cryo-EM Facility at CSSB, supported by the UHH and DFG [INST 152/772-11152/774-11152/775-11152/776-11152/777-1 FUGG]; Individual fellowship from the Alexander von Humboldt foundation (to E.Q.); T.K. holds a fellowship from the EMBL Interdisciplinary Postdocs (EI3POD) initiative co-funded by Marie Skłodowska-Curie [664726]; iNEXT [653706] funded by the Horizon 2020 programme of the European Union. Funding for open access charge: Leibniz Association.

Conflict of interest statement. The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

REFERENCES

- McMullan, L.K., Folk, S.M., Kelly, A.J., MacNeil, A., Goldsmith, C.S., Metcalfe, M.G., Batten, B.C., Albarino, C.G., Zaki, S.R., Rollin, P.E. *et al.* (2012) A new phlebovirus associated with severe febrile illness in Missouri. *N. Engl. J. Med.*, **367**, 834–841.
- Kohl, C., Brinkmann, A., Radonic, A., Dabrowski, P.W., Nitsche, A., Muhldorfer, K., Wibbelt, G. and Kurth, A. (2020) Zwiesel bat banyangvirus, a potentially zoonotic Huaiyangshan banyangvirus (Formerly known as SFTS)-like banyangvirus in Northern bats from Germany. *Sci. Rep.*, **10**, 1370.
- Maes, P., Alkhovsky, S.V., Bao, Y., Beer, M., Birkhead, M., Briese, T., Buchmeier, M.J., Calisher, C.H., Charrel, R.N., Choi, I.R. *et al.* (2018) Taxonomy of the family Arenaviridae and the order Bunyavirales: update 2018. *Arch. Virol.*, **163**, 2295–2310.
- Mehand, M.S., Al-Shorbaji, F., Millett, P. and Murgue, B. (2018) The WHO R&D Blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts. *Antiviral Res.*, **159**, 63–67.
- Finberg, R.W., Lanno, R., Anderson, D., Fleischhack, L.R., van Duijnhoven, W., Kaufman, R.S., Kosoglou, T., Vingerhoets, J. and Leopold, L. (2019) Phase 2b study of Pimodivir (JNJ-63623872) as monotherapy or in combination with oseltamivir for treatment of acute uncomplicated seasonal influenza A: TOPAZ trial. *J. Infect. Dis.*, **219**, 1026–1034.
- Noshi, T., Kitano, M., Taniguchi, K., Yamamoto, A., Omoto, S., Baba, K., Hashimoto, T., Ishida, K., Kushima, Y., Hattori, K. *et al.* (2018) In vitro characterization of baloxavir acid, a first-in-class cap-dependent endonuclease inhibitor of the influenza virus polymerase PA subunit. *Antiviral Res.*, **160**, 109–117.
- Olschewski, S., Cusack, S. and Rosenthal, M. (2020) The Cap-Snatching mechanism of bunyaviruses. *Trends Microbiol.*, **28**, 293–303.
- Reguera, J., Gerlach, P., Rosenthal, M., Gaudon, S., Coscia, F., Gunther, S. and Cusack, S. (2016) Comparative structural and functional analysis of bunyavirus and arenavirus Cap-Snatching endonucleases. *PLoS Pathog.*, **12**, e1005636.
- Holm, T., Kopicki, J.D., Busch, C., Olschewski, S., Rosenthal, M., Uetrecht, C., Gunther, S. and Reindl, S. (2018) Biochemical and structural studies reveal differences and commonalities among cap-snatching endonucleases from segmented negative-strand RNA viruses. *J. Biol. Chem.*, **293**, 19686–19698.
- Rosenthal, M., Gogrefe, N., Vogel, D., Reguera, J., Rauschenberger, B., Cusack, S., Gunther, S. and Reindl, S. (2017) Structural insights into reptarenavirus cap-snatching machinery. *PLoS Pathog.*, **13**, e1006400.
- Gogrefe, N., Reindl, S., Gunther, S. and Rosenthal, M. (2019) Structure of a functional cap-binding domain in Rift Valley fever virus L protein. *PLoS Pathog.*, **15**, e1007829.
- Gerlach, P., Malet, H., Cusack, S. and Reguera, J. (2015) Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell*, **161**, 1267–1279.
- Vogel, D., Rosenthal, M., Gogrefe, N., Reindl, S. and Gunther, S. (2019) Biochemical characterization of the Lassa virus L protein. *J. Biol. Chem.*, **294**, 8088–8100.
- Kranzusch, P.J., Schenk, A.D., Rahmeh, A.A., Radoshitzky, S.R., Bavari, S., Walz, T. and Whelan, S.P. (2010) Assembly of a functional Machupo virus polymerase complex. *Proc. Natl. Acad. Sci. USA*, **107**, 20069–20074.
- Kranzusch, P.J. and Whelan, S.P. (2011) Arenavirus Z protein controls viral RNA synthesis by locking a polymerase-promoter complex. *Proc. Natl. Acad. Sci. USA*, **108**, 19743–19748.
- Pyle, J.D. and Whelan, S.P.J. (2019) RNA ligands activate the Machupo virus polymerase and guide promoter usage. *Proc. Natl. Acad. Sci. USA*, **116**, 10518–10524.
- Pflug, A., Guilligay, D., Reich, S. and Cusack, S. (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*, **516**, 355–360.
- Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, J., Crepin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W. *et al.* (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature*, **516**, 361–366.
- 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. *J. Virol.*, **66**, 1370–1376.
- Polyak, S.J., Zheng, S. and Harnish, D.G. (1995) 5' termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. *J. Virol.*, **69**, 3211–3215.
- Garcin, D., Lezzi, M., Dobbs, M., Elliott, R.M., Schmaljohn, C., Kang, C.Y. and Kolakofsky, D. (1995) The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-extend mechanism for the initiation of RNA synthesis. *J. Virol.*, **69**, 5754–5762.
- 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. *J. Virol.*, **80**, 2337–2348.
- Wang, W., Shin, W.J., Zhang, B., Choi, Y., Yoo, J.S., Zimmerman, M.I., Frederick, T.E., Bowman, G.R., Gross, M.L., Leung, D.W. *et al.* (2020) The Cap-Snatching SFTSV endonuclease domain is an antiviral target. *Cell Rep.*, **30**, 153–163.
- Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I. and Owens, R.J. (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.*, **35**, e45.
- Cummings, M.D., Farnum, M.A. and Nelen, M.I. (2006) Universal screening methods and applications of ThermoFluor. *J. Biomol. Screen.*, **11**, 854–863.
- Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T. and Eliceiri, K.W. (2017) ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, **18**, 529.
- Kabsch, W. (2010) Xds. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 125–132.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 486–501.
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 213–221.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.











32. Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y. and Agard, D.A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.
33. Tegunov, D. and Cramer, P. (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods*, **16**, 1146–1152.
34. Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E. and Scheres, S.H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*, **7**, e42166.
35. Zhang, K. (2016) Gctf: Real-time CTF determination and correction. *J. Struct. Biol.*, **193**, 1–12.
36. Grant, T., Rohou, A. and Grigorieff, N. (2018) cisTEM, user-friendly software for single-particle image processing. *Elife*, **7**, e35383.
37. Lieschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J. et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, **75**, 861–877.
38. Blanchet, C.E., Spilotros, A., Schwemmer, F., Graewert, M.A., Kikhney, A., Jeffries, C.M., Franke, D., Mark, D., Zengerle, R., Cipriani, F. et al. (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.*, **48**, 431–443.
39. Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. et al. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.*, **50**, 1212–1225.
40. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277–1282.
41. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.*, **25**, 495–503.
42. Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.*, **42**, 342–346.
43. Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorka, C., Mertens, H.D., Konarev, P.V. and Svergun, D.I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **45**, 342–350.
44. Oide, M., Sekiguchi, Y., Fukuda, A., Okajima, K., Oroguchi, T. and Nakasako, M. (2018) Classification of ab initio models of proteins restored from small-angle X-ray scattering. *J. Synchrotron Radiat.*, **25**, 1379–1388.
45. Kozin, M. and Svergun, D. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.*, **34**, 33–41.
46. Volkov, V. and Svergun, D. (2003) Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.*, **36**, 860–864.
47. Amroun, A., Priet, S., de Lamballerie, X. and Querat, G. (2017) Bunyaviridae RdRps: structure, motifs, and RNA synthesis machinery. *Crit. Rev. Microbiol.*, **43**, 753–778.
48. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
49. Yun, S.M., Park, S.J., Park, S.W., Choi, W., Jeong, H.W., Choi, Y.K. and Lee, W.J. (2017) Molecular genomic characterization of tick- and human-derived severe fever with thrombocytopenia syndrome virus isolates from South Korea. *PLoS Negl. Trop. Dis.*, **11**, e0005893.
50. Yu, X.J., Liang, M.F., Zhang, S.Y., Liu, Y., Li, J.D., Sun, Y.L., Zhang, L., Zhang, Q.F., Popov, V.L., Li, C. et al. (2011) Fever with thrombocytopenia associated with a novel bunyavirus in China. *N. Engl. J. Med.*, **364**, 1523–1532.
51. Brennan, B., Li, P., Zhang, S., Li, A., Liang, M., Li, D. and Elliott, R.M. (2015) Reverse genetics system for severe fever with thrombocytopenia syndrome virus. *J. Virol.*, **89**, 3026–3037.
52. Zhong, W., Uss, A.S., Ferrari, E., Lau, J.Y. and Hong, Z. (2000) De novo initiation of RNA synthesis by hepatitis C virus nonstructural protein 5B polymerase. *J. Virol.*, **74**, 2017–2022.
53. Arnold, J.J., Gohara, D.W. and Cameron, C.E. (2004) Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mn²⁺. *Biochemistry*, **43**, 5138–5148.
54. Jones, R., Lessoued, S., Meier, K., Devignot, S., Barata-Garcia, S., Mate, M., Bragagnolo, G., Weber, F., Rosenthal, M. and Reguera, J. (2019) Structure and function of the Toscana virus cap-snatching endonuclease. *Nucleic Acids Res.*, **47**, 10914–10930.
55. Liu, X., Jin, J., Qiu, P., Gao, F., Lin, W., Xie, G., He, S., Liu, S., Du, Z. and Wu, Z. (2018) Rice stripe tenuivirus has a greater tendency to use the Prime-and-Realignment mechanism in transcription of genomic than in transcription of antigenomic template RNAs. *J. Virol.*, **92**, e01414–17.
56. Te Velthuis, A.J.W. and Oymans, J. (2017) Initiation, elongation and realignment during influenza virus mRNA synthesis. *J. Virol.*, **92**, e01775–17.
57. Jin, H. and Elliott, R.M. (1993) Non-viral sequences at the 5' ends of Dugbe nairovirus S mRNAs. *J. Gen. Virol.*, **74**, 2293–2297.
58. Mazza, C., Ohno, M., Segref, A., Mattaj, I.W. and Cusack, S. (2001) Crystal structure of the human nuclear cap binding complex. *Mol. Cell*, **8**, 383–396.
59. Niedzwiecka, A., Marcotrigiano, J., Stepinski, J., Jankowska-Anyszka, M., Wyslouch-Cieszyńska, A., Dadlez, M., Gingras, A.C., Mak, P., Darzynkiewicz, E., Sonenberg, N. et al. (2002) Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol.*, **319**, 615–635.
60. Byrn, R.A., Jones, S.M., Bennett, H.B., Bral, C., Clark, M.P., Jacobs, M.D., Kwong, A.D., Ledebor, M.W., Leeman, J.R., McNeil, C.F. et al. (2015) Preclinical activity of VX-787, a first-in-class, orally bioavailable inhibitor of the influenza virus polymerase PB2 subunit. *Antimicrob. Agents Chemother.*, **59**, 1569–1582.
61. Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., Pflug, A., Hengrung, N., El Omari, K., Baudin, F. et al. (2016) Influenza polymerase can adopt an alternative configuration involving a radical repacking of PB2 domains. *Mol. Cell*, **61**, 125–137.

3.2 Conformational changes in Lassa virus L-protein associated with promoted binding and RNA synthesis activity

Building on what had been learned from SFTSV L-protein, we aimed to visualize the L-protein from LASV in multiple different states of replication including apo, pre-initiation and RNA elongation. This was achieved by stalling the L-protein using a non-hydrolyzable nucleotide analog allowing for detailed insights into the L-protein interactions with viral RNA as well as conformational changes and important residues associated with elongation.

Original publication and supplement at doi.org/10.1038/s41467-021-27305-5

Conformational changes in Lassa virus L protein associated with promoter binding and RNA synthesis activity

Tomas Kouba ^{1,4}, Dominik Vogel ^{2,4}, Sigurdur R. Thorkelsson ^{3,4}, Emmanuelle R. J. Quemin ³, Harry M. Williams ², Morlin Milewski ², Carola Busch², Stephan Günther ², Kay Grünewald ³, Maria Rosenthal ^{2,5}✉ & Stephen Cusack ^{1,5}✉

Lassa virus is endemic in West Africa and can cause severe hemorrhagic fever. The viral L protein transcribes and replicates the RNA genome via its RNA-dependent RNA polymerase activity. Here, we present nine cryo-EM structures of the L protein in the apo-, promoter-bound pre-initiation and active RNA synthesis states. We characterize distinct binding pockets for the conserved 3' and 5' promoter RNAs and show how full-promoter binding induces a distinct pre-initiation conformation. In the apo- and early elongation states, the endonuclease is inhibited by two distinct L protein peptides, whereas in the pre-initiation state it is uninhibited. In the early elongation state, a template-product duplex is bound in the active site cavity together with an incoming non-hydrolysable nucleotide and the full C-terminal region of the L protein, including the putative cap-binding domain, is well-ordered. These data advance our mechanistic understanding of how this flexible and multifunctional molecular machine is activated.

¹European Molecular Biology Laboratory, Grenoble, France. ²Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. ³Centre for Structural Systems Biology, Leibniz Institute for Experimental Virology, University of Hamburg, Hamburg, Germany. ⁴These authors contributed equally: Tomas Kouba, Dominik Vogel, Sigurdur R. Thorkelsson. ⁵These authors jointly supervised this work: Maria Rosenthal, Stephen Cusack. ✉email: rosenthal@bnitm.de; cusack@embl.fr

Lassa virus (LASV) is a segmented, negative-strand RNA virus belonging to the family of *Arenaviridae* within the *Bunyavirales* order. It is a rodent-borne virus, endemic to West Africa and the causative agent of Lassa haemorrhagic fever, a febrile illness with increasing case numbers and a case fatality rate among hospitalised patients of ~15% in Nigeria in 2018¹. Recent studies applying computational modelling predict a total number of ~900,000 human infections per year across West Africa². The large (L) protein of LASV is a multi-domain molecular machine that binds the conserved 3' and 5' ends (the 'promoter') of each of the two viral RNA (vRNA) genome segments (denoted L and S) and plays a central role in the viral life cycle, which is entirely cytoplasmic. The L protein contains an RNA-dependent RNA polymerase (RdRp) activity and catalyses both viral transcription and genome replication. Each vRNA segment is a template for the synthesis of two different types of RNA products: capped viral transcripts (of L and NP genes) as well as an unmodified full-length complementary RNA (cRNA) copy, which is an intermediate of viral genome replication. The cRNA is a template for the second stage of replication, the synthesis of vRNA genome copies, as well as the production of further mRNA (of GPC and Z genes) by transcription. Transcription is initiated using a capped primer derived from host mRNA by a yet to be elucidated 'cap-snatching' mechanism involving the intrinsic endonuclease (EN) of the L protein and possibly its cap-binding domain (CBD)³. This results in viral mRNAs that have 1–7 host-derived nucleotides at the 5' end^{4–6}. Viral genome replication is initiated by a prime-and-realign mechanism resulting in an extra G nucleotide at the 5' end of the vRNA and cRNA⁷. The first structural studies of the complete arenavirus L protein were conducted on Machupo virus (MACV), which is related to LASV but belongs to the group of New World arenaviruses found in South America. Negative stain electron microscopy studies at low resolution revealed a donut-like molecule with accessory appendages⁸. In 2020, the first models of MACV and LASV L proteins were proposed based on cryo-electron microscopy (cryo-EM) with overall resolutions of ~3.6 and 3.9 Å, respectively⁹. These structures revealed that arenavirus L proteins are structurally similar to the polymerases of the

related La Crosse and influenza viruses^{10–14}. However, the reported arenavirus structures are incomplete and do not show the L protein in an active conformation⁹.

Here, we present nine cryo-EM structures that provide insights into the conformational rearrangements of LASV L protein that occur upon its activation into a functional RNA synthesis elongation state. This comprehensive structural study is complemented by biochemical data from *in vitro* assays with purified L protein and selected mutants as well as functional data in cells using the LASV mini-replicon system¹⁵. The results presented enhance our mechanistic understanding of the multifunctional LASV L protein and will guide targeted drug development approaches in the future.

Results

Overview of structures obtained. Nine cryo-EM structures of LASV L protein have been determined in the apo-state, bound to the 3' end of the genomic vRNA alone, bound to the full promoter, comprising the highly complementary 3' and 5' ends of the vRNA, or in a stalled, early elongation state (Table 1).

From grids of the apo-state, two different 3D classes were separated. In the APO-ENDO structure, at 3.35 Å resolution (Fig. 1a), the N-terminal EN is clearly resolved, packing against the polymerase core and with a peptide (residues 1092–1104) from the central region of the L protein bound in its active site cavity, presumably inhibiting its activity. In the second class, denoted APO-RIBBON (Fig. 1a), at 3.73 Å resolution, the EN is not resolved, but residues 822–1110 of the L protein form an extended structure including an α -bundle (α -ribbon, 843–884, with a third helix 907–925 packing against it) that is not visible in the APO-ENDO structure. Masked refinement of the common regions of the two apo-structures yielded a map of the APO-CORE (Fig. 1a) with an improved resolution of 3.14 Å that allowed a more accurate model to be built.

Upon incubation of L protein with nucleotides (nts) 1–16 of the vRNA 3' end alone (structures denoted 3END-CORE, 2.70 Å, 3END-ENDO, 3.04 Å) (Fig. 1b), nucleotides 1–6 from the 3' end bind specifically in a buried groove under the pyramid, a prominent feature in the N-terminal region of the L protein

Table 1 Overview of the different LASV L protein structures.

Identifier	Resolution [Å]	RNA ligands	PDB accession	EMDB accession	Comment
APO-CORE	3.14	-	7OCH	12807	High-resolution apo-polymerase core, best defined Zn ²⁺ coordination site
APO-ENDO	3.35	-	7OE3	12860	EN domain complete and bound to inhibitory peptide 1092–1104, well-defined interaction with polymerase core
APO-RIBBON	3.73	-	7OE7	12953	α -ribbon visible
3END-CORE	2.70	3' 1–16	7OEA	12862	Highest resolution core, 3' RNA bound to the secondary binding site, EN bound to inhibitory peptide 1092–1104
3END-ENDO	3.04	3' 1–16	7OEB	12863	EN domain complete and bound to inhibitory peptide 1092–1104, 3' RNA bound to the secondary binding site
PRE-INITIATION	3.34	5' 0–19 3' 1–19	7OJL	12955	dsRNA promoter bound, α -bundle and pendant visible
MID-LINK	3.50	5' 10–19 3' 1–19	7OJJ	12861	C terminus visible (low resolution), EN not potentially autoinhibited
DISTAL-PROMOTER	3.89	5' 10–19 3' 1–19	7OJK	12954	Distal dsRNA promoter bound, α -bundle and pendant visible
ELONGATION	2.92	5' 0–19 3' 1–19 C8 primer, UMPNPP	7OJN	12956	C terminus buildable with high confidence, RNA duplex in active site, UMPNPP, EN autoinhibited by peptide 173–190

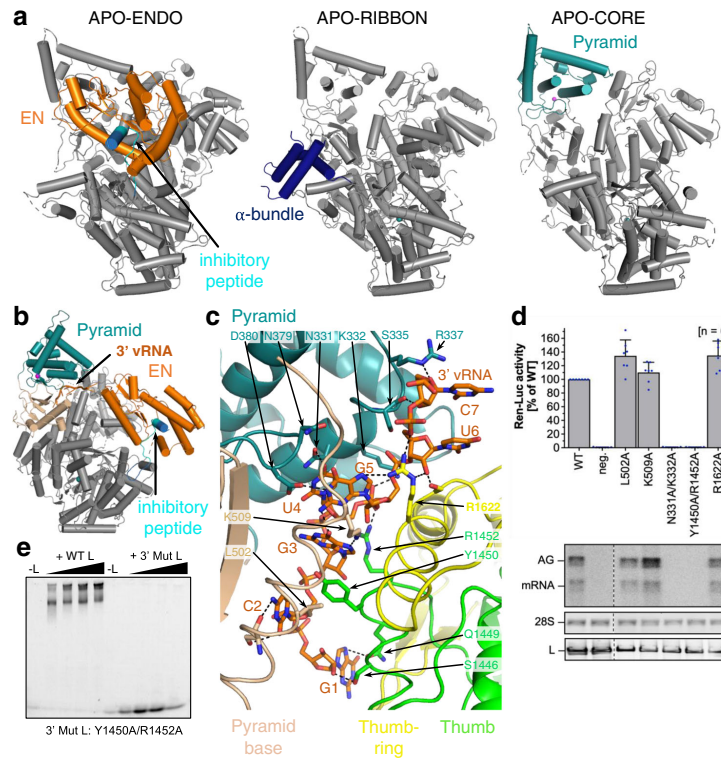


Fig. 1 L protein in the apo-state and with 3' viral RNA bound in the secondary binding site. **a** Ribbon diagram presentations of the structures APO-ENDO, APO-RIBBON and APO-CORE. Each of the respective experimental maps resolves distinct regions of the L protein better than the others and those regions are shown in colour and indicated by name in the respective structures. **b** Overall structure of L protein 3END-CORE as a ribbon diagram with the 3' vRNA bound below the pyramid domain. Pyramid (teal), pyramid base (wheat), EN domain (orange) and the inhibitory peptide (cyan) are highlighted. **c** Close-up of the secondary 3' vRNA binding site with the 3' vRNA nucleotides 1–7 (orange), pyramid domain (teal), pyramid base (wheat) as well as thumb (green) and thumb-ring (yellow) domains. Important amino acids in the RNA-protein interface are shown as sticks with respective labels. Hydrogen bonds are indicated by dotted lines. For selected regions, secondary structure depiction was disabled to enhance visibility. **d** LASV mini-replicon data for L proteins with mutations in the secondary 3' RNA binding site presenting luciferase reporter activity (in standardised relative light units relative to the wild-type L protein (WT)). Data were presented as mean values \pm SD of at least six biological replicates ($n = 6$), although for most mutants seven biological replicates were included. All biological replicates are shown as blue dots (top panel). Middle panels present Northern blotting results with signals for antigenomic viral RNA (AG), viral mRNA (mRNA) and 28 S ribosomal RNA (28 S) as a loading control, and the bottom panel shows Western blot detection of FLAG-tagged L proteins (L) to demonstrate general expressibility of the mutants. Source data are provided in a Source Data file. **e** Electrophoretic mobility shift assay of wild-type L protein (WT L) and mutant Y1450A/R1452A (Mut L) with 10 nt 3' viral RNA. L protein concentrations ranging from 0–1 μ M and 0.2 μ M of fluorescently labelled 3' vRNA (Supplementary Table 1) were used (see methods).

(Fig. 1c). This site corresponds to the secondary 3' end-binding site previously described for influenza virus, La Crosse virus (LACV) and MACV polymerase proteins^{9,13,16–18}. In these structures, the EN remains in the inhibited conformation as observed in the APO-ENDO structure (compare Fig. 1a and b). When the full vRNA promoter is bound (5' end nts 0–19, including an additional G0 according to the product expected from the prime-and-realign initiation mechanism, 3' end nts 1–19), a pre-initiation complex (PRE-INITIATION) is observed at 3.34 Å resolution (Fig. 2a, Supplementary Fig. 1). This structure reveals that the LASV vRNA promoter is organised similarly to those of influenza virus¹⁰ and LACV^{13,14} in comprising a single-stranded 5' end folded as a hook, a distal duplex region and a single-stranded 3' end, only partially visible, directed towards the

RNA synthesis active site (Fig. 2c). Overall, the protein conformation of the PRE-INITIATION structure resembles that of the APO-RIBBON, with an additional partially ordered insertion domain, previously denoted the pendant⁹, packing against the 3' strand of the promoter (see below) (Fig. 2a). Two further structures were obtained from a sample in which the L protein was incubated with a truncated promoter (5' nts 10–19, 3' nts 1–19), which lacks nucleotides 0–9 of the 5' end. One 3D class from this sample, obtained by focussed refinement on the promoter-bound region, (DISTAL-PROMOTER, 3.89 Å resolution), closely resembles the full promoter (PRE-INITIATION) structure, but additionally reveals a new position of the EN, without inhibitory peptide bound (Supplementary Fig. 1). The second 3D class from the same grid (MID-LINK, 3.50 Å

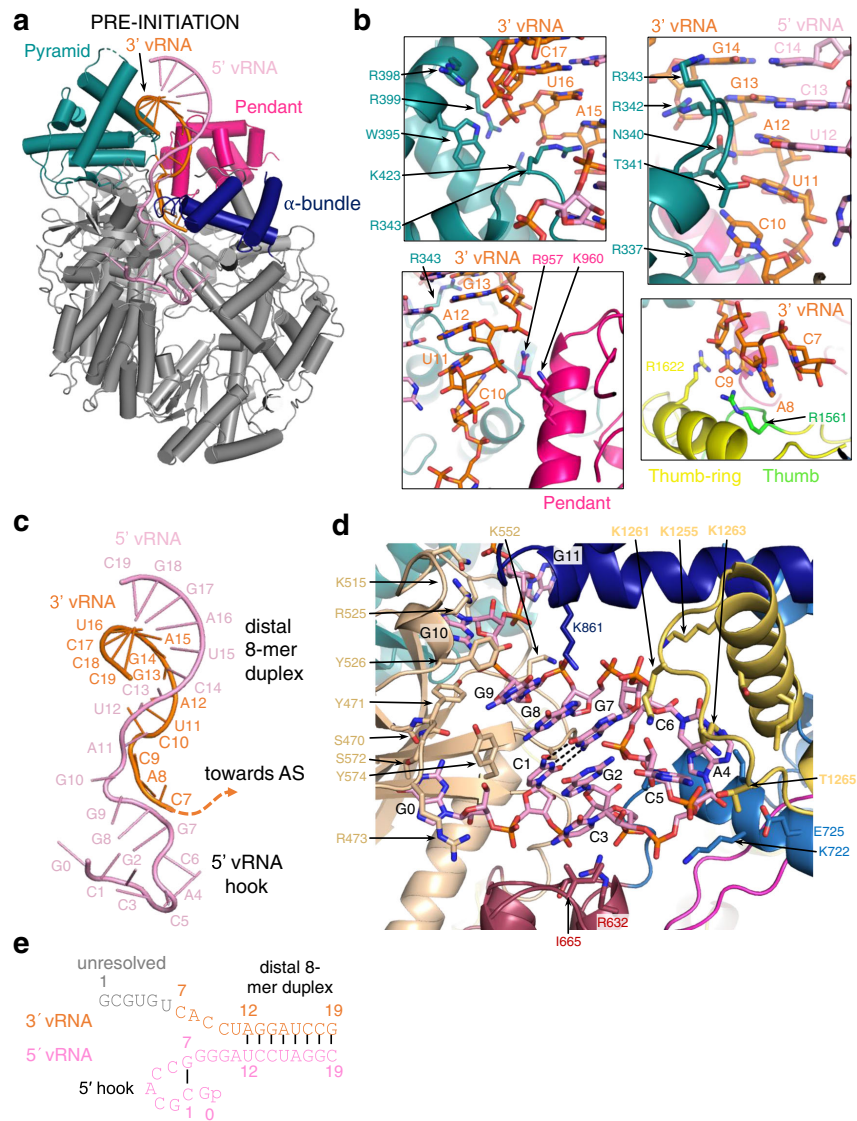


Fig. 2 L protein in the pre-initiation state. **a** Ribbon diagram of the PRE-INITIATION structure with pendant domain (pink), α -bundle (dark blue), pyramid (teal), 3' vRNA nts 7–19 (orange) and 5' vRNA nts 0–19 (pink) highlighted in colour and indicated by name. **b** Interactions of the L protein pyramid (teal), pendant (pink) thumb-ring (yellow) and thumb (green) domains with the 3' vRNA are shown. Important amino acid side chains and the RNA nucleotides of 3' and 5' vRNA are shown as sticks with respective labels. **c** Viral RNA observed in this structure with a 5' vRNA hook structure composed of 5' vRNA nts 0–9 and a distal duplex region involving 5' vRNA nts 12–19 and 3' vRNA nts 12–19. The 3' vRNA nts 1–11 are directed towards the RdRp active site (towards AS) but not resolved. **d** Close-up of the 5' vRNA hook binding site involving the fingers domain (blue), finger node (light yellow), pyramid base (wheat) and helical region (raspberry). Residues important for the RNA-protein interface and nucleotides are shown as sticks and are labelled. **e** Schematic presentation of the promoter RNA (3' vRNA in orange, 5' vRNA in pink) in the PRE-INITIATION structure. Nucleotides 1–6 of the 3' vRNA, which are not resolved, are coloured in grey. Distal duplex and 5' hook regions are labelled.

resolution) (Supplementary Fig. 2), obtained by focussed refinement of the other end of the polymerase, shows density for the C-terminal region of the L protein beyond residue 1834. This allows tentative modelling of domains that resemble the mid-link and 627-domains of the influenza virus polymerase PB2 subunit. At very low resolution, an envelope of the putative cap-binding

(CBD-like) domain is observed. A final structure (ELONGATION, 2.92 Å resolution) captures an early elongation state initiated with an uncapped primer and stalled after incorporation of four nucleotides by an incoming non-hydrolysable UTP analogue, UMPNPP (Figs. 3, 4). In this structure, the promoter duplex is disrupted due to translocation of the template and a

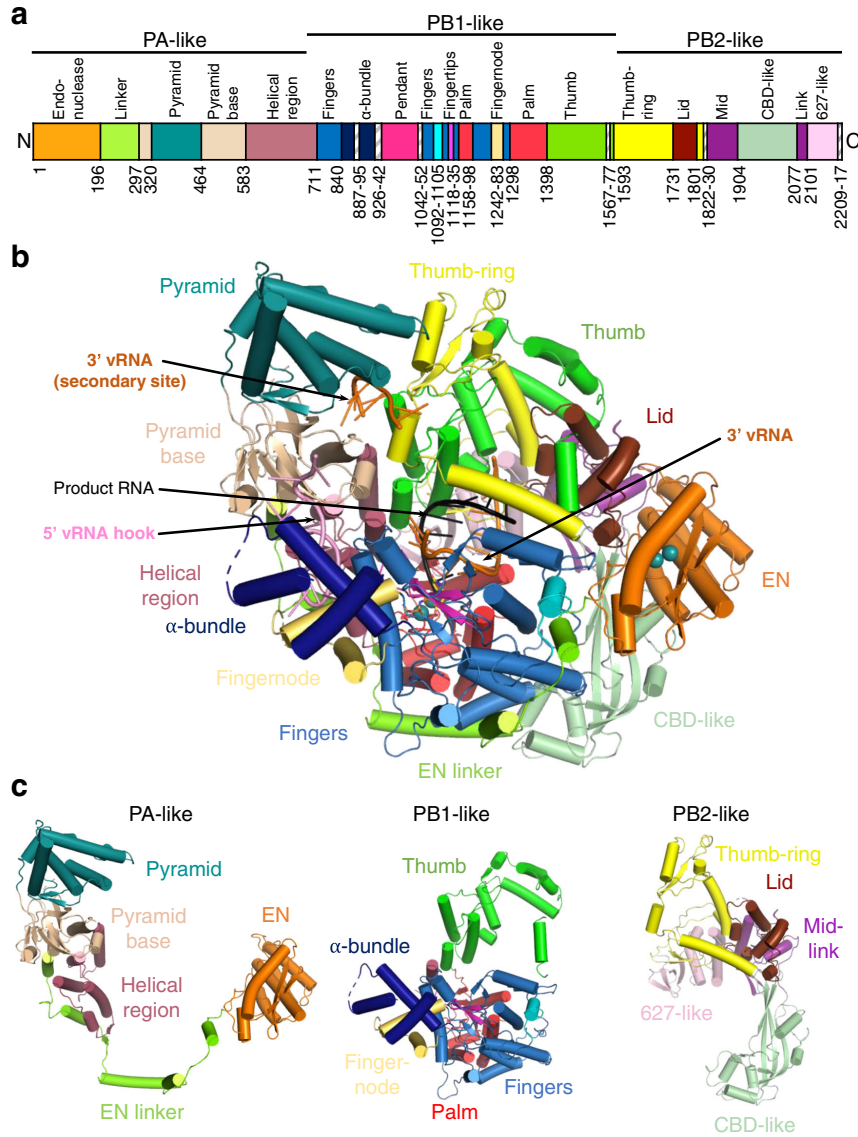
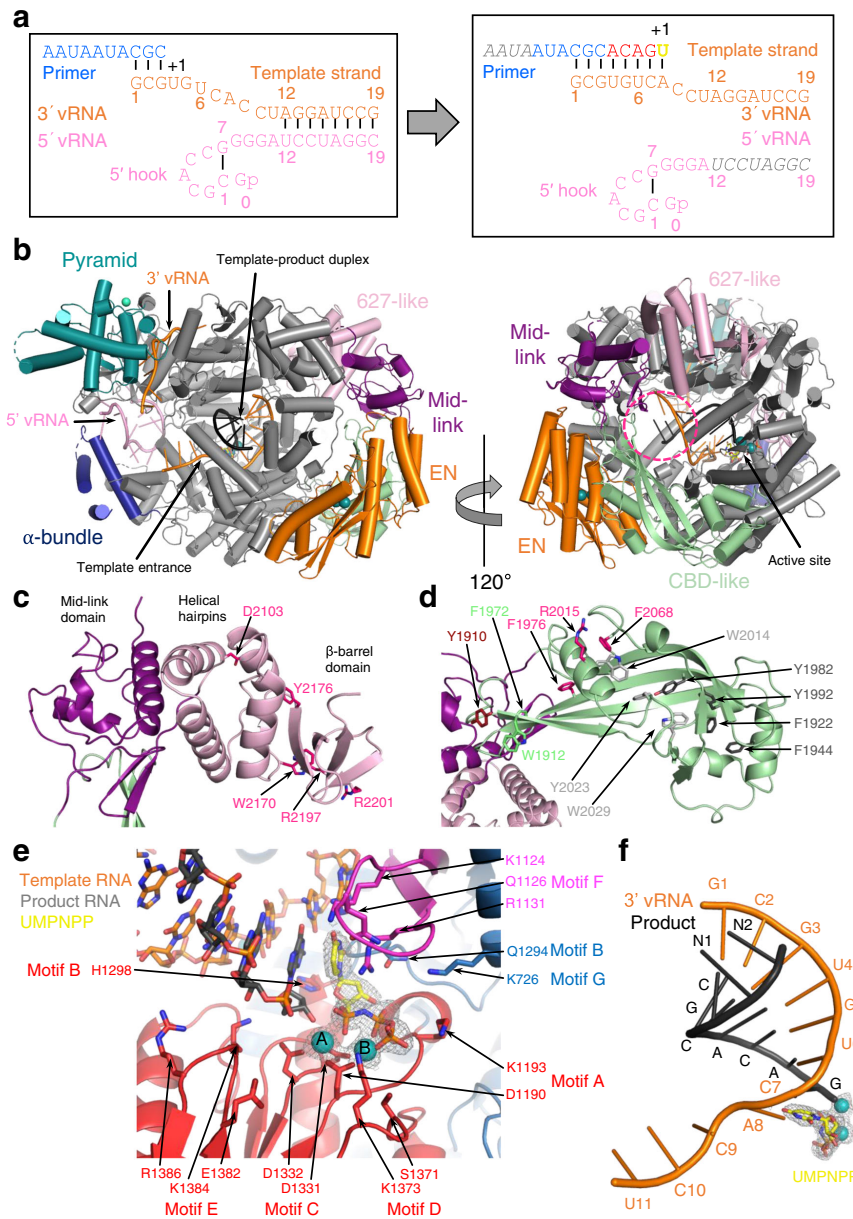


Fig. 3 Overview of the L protein structure. **a** Schematic linear presentation of the L protein domain structure. **b** Complete ELONGATION structure of the L protein presented as a ribbon diagram in front view. Domains are coloured according to (a) and labelled. 3' vRNA is coloured in orange, 5' vRNA in pink and product RNA in black. See also Supplementary Movie 3 for a 3D impression of the L protein and its domains. **c** Separate presentation of the PA-like, PB1-like and PB2-like regions of the L protein in the ELONGATION structure.



duplex of eight base pairs occupies the active site cavity (Fig. 4a). The complete C-terminal domain is well-resolved, and, due to rotation of the CBD-like domain, is now in a bent configuration, rather than the extended configuration seen in the MID-LINK structure. The C-terminal domain, together with the EN, now in a third distinct position, forms a ring around the putative product exit channel (Fig. 4b).

The different structures obtained (Table 1) reveal that LASV L, similar to influenza virus polymerase¹⁹, has a number of domains flexibly linked to the polymerase core, allowing multiple conformations of the protein. None of the here reported individual structures are complete (although the elongation structure model comprises ~90% of the residues, lacking mainly the pendant domain), but the integration of all the information leads to a coherent picture of the

Fig. 4 Elongation state of the L protein. **a** Schematic presentation of the primed reaction was carried out to obtain the ELONGATION structure with the L protein stalled in an early elongation state. Nucleotides that are not visible or not clearly assignable from the experimental map are shown in grey italics. **b** ELONGATION structure of the L protein presented as a ribbon diagram in two views as indicated. EN, pyramid, α -bundle, mid-link and 627-like domains are coloured. 3' vRNA is shown in orange, 5' RNA in pink and product RNA in black. A dashed circle (hot pink) indicates the putative product exit. **c** Close-up on the mid-link and 627-like domains with the respective structural features labelled and the side chains of amino acids shown to be selectively important for viral transcription by Lehmann et al. 2014²⁴ shown as pink sticks. **d** Close-up on the CBD-like domain with side chains of amino acids that have been tested in the LASV mini-replicon system shown as sticks (pink—selective role in viral transcription; light grey—no significant reduction of L protein activity upon mutation shown by Lehmann et al. 2014²⁴; green—no or weak effect on L protein function upon mutation; dark red—potential selective role in viral transcription; dark grey—general defect of L protein activity upon mutation). Corresponding mini-replicon data are presented in Supplementary Fig. 8. **e** Close-up of the polymerase active site with the template RNA (orange), the product RNA (dark grey), the non-hydrolysable UTP (UMPNNP, yellow) and catalytic manganese ions (teal, A and B) involving the palm (red), fingers (blue), fingertips (magenta) and thumb (green) domains of the L protein. Important side chains are shown as sticks and conserved RdRp active site motifs (A–G) are labelled. The map around the UMPNNP and the manganese ions is shown as a grey mesh. **f** Template-product duplex of the polymerase active site is shown as a ribbon diagram with the product in black, the 3' template in orange, the non-hydrolysable UTP (UMPNNP) in yellow and the catalytic manganese ions as teal spheres. The map around the ions and the UMPNNP is shown as a grey mesh.

overall LASV L protein structure and the significant conformational changes that occur upon promoter binding and the subsequent transition into the active elongation state.

Overview of LASV L protein structure. As previously established, MACV and LASV L proteins have an overall architecture similar to previously determined orthomyxovirus and bunyavirus polymerases⁹. With reference to the heterotrimeric influenza virus polymerase, the LASV L protein can be conveniently divided into PA-like (1–687), PB1-like (688–1592) and PB2-like (1593–2217) regions (Fig. 3a, c and Supplementary Fig. 3).

The PA-like region has an N-terminal EN (1–195), whose structure and properties have previously been studied^{20,21}. Whereas previous crystal structures of the isolated LASV EN are not resolved beyond residue ~173^{20,21}, the full L protein structures show that the domain comprises an additional helix ending at 190. The EN is followed by an extended linker (196–296) that wraps around the polymerase core (Fig. 3 and Supplementary Fig. 3). This connects to the pyramid base (297–319, 464–582) into which is inserted a prominent feature denoted the pyramid domain (320–463) (Figs. 1a, 3 and Supplementary Fig. 3). The pyramid domain is specific to Old World arenaviruses and results from residue insertions compared to New World arenaviruses (e.g. MACV) that considerably lengthen the two helices spanning 386–435 (Supplementary Data 1), giving it a characteristic angular shape. At the beginning of the pyramid domain, there is a structural zinc-binding site with coordinating ligands H316, C321, H364, and C366 (Supplementary Fig. 4). Mutational studies using the LASV mini-replicon system showed a general but incomplete reduction in L protein activity upon single-site exchanges to alanine (Supplementary Fig. 4), which further emphasises the structural role of this zinc-binding site. Indeed, sequence comparisons show that this site is specific to the LASV strain Bantou 289 and closely related strains, but not conserved in other LASV lineages or other arenaviruses (Supplementary Data 1). Interestingly, the MACV L protein also contains a zinc-binding site but in a different location, at the pyramid base (Supplementary Fig. 4).

The PB1-like region (Figs. 3a, c, 4e and Supplementary Fig. 3) contains the canonical fingers, palm and thumb with associated conserved polymerase motifs A–F (motif G is 641–RY, motif H is K1237¹³). The catalytic triad of aspartates are D1190 (motif A), D1331 and D1332 (motif C). As previously noted⁹, the fingertips (motif F, 1117–1137) are well structured even in the absence of bound promoter (Supplementary Fig. 5), unlike in influenza virus and LACV polymerases¹³. The LASV L PB1-like region is considerably larger than influenza virus PB1 (882 residues compared to 756), mainly due to the so-called Lassa insertion

(830–1069). This includes two flexibly linked modules: (i) a three-helix α -bundle (840–925), which includes an α -ribbon and (ii) the compact pendant domain (943–1042), the latter being only partially visible in our structures (Fig. 3 and Supplementary Fig. 3). The internal connection (887–895) between the α -ribbon and third helix of the α -bundle is disordered as are the flexible linkers before (926–942) and after (1042–1052) the pendant domain (Fig. 3a and Supplementary Fig. 6). These modules were previously observed in the MACV L structure⁹, but in different positions and configurations (Supplementary Fig. 6). The PB1-like region has an additional insertion in the fingers, called the finger node (1242–1283), not present in influenza virus PB1, but very similar to the finger node of LACV and likewise involved in binding the 5' hook (see below) (Fig. 2d)¹³. In all LASV structures, the region 1567–1577 is disordered. Moreover, it is so far unclear whether any protein segment might serve as a priming loop.

The PB2-like region (1593–2217) (Figs. 3a, c, 5c, d and Supplementary Fig. 3) has a similar overall organisation to influenza virus, LACV and Severe fever with thrombocytopenia syndrome bunyavirus (SFTSV) polymerases, with a 'thumb-ring', associated with the core and surrounding the thumb domain. Into the thumb-ring, a 'helical lid' domain (1731–1800) is inserted, which in influenza virus polymerase forces strand-separation during RNA synthesis²². This is followed by a short flexible linker to an array of C-terminal domains (1830–2217), visible at lower resolution in the MID-LINK structure (Supplementary Fig. 2), but fully buildable in the well-ordered ELONGATION structure (Figs. 3a–c and 4b–d). This includes the influenza-like split 'mid-link' domain (1831–1903, 2077–2100), into which is inserted the putative cap-binding domain (CBD-like, 1904–2076), followed by a 627-like domain that comprises two helical hairpins (2101–2168) and a terminal, compact β -barrel domain (2169–2208) (Figs. 3a–c, 4c–d and Supplementary Figs. 2, 3). The mid-link and 627-like domains are juxtaposed in the same way in the MID-LINK and ELONGATION structures, suggesting that they are rigidly associated (Supplementary Fig. 2), although possessing some rotational freedom as a whole with respect to the thumb domain (Supplementary Fig. 7). In contrast, the CBD-like domain has considerable rotational freedom, with a difference of ~84° in its orientation with respect to the mid-link domain in the MID-LINK and ELONGATION structures, respectively (Supplementary Fig. 7). The C-terminal region was previously visualised in a dimeric form of MACV L protein (PDB:6KLH), but at an insufficient resolution to build a correct model⁹. The MACV cryo-EM density (EMD-0710) for the C-terminal region is fully compatible with the LASV C-terminal model and shows an extended configuration similar to that observed in the MID-LINK

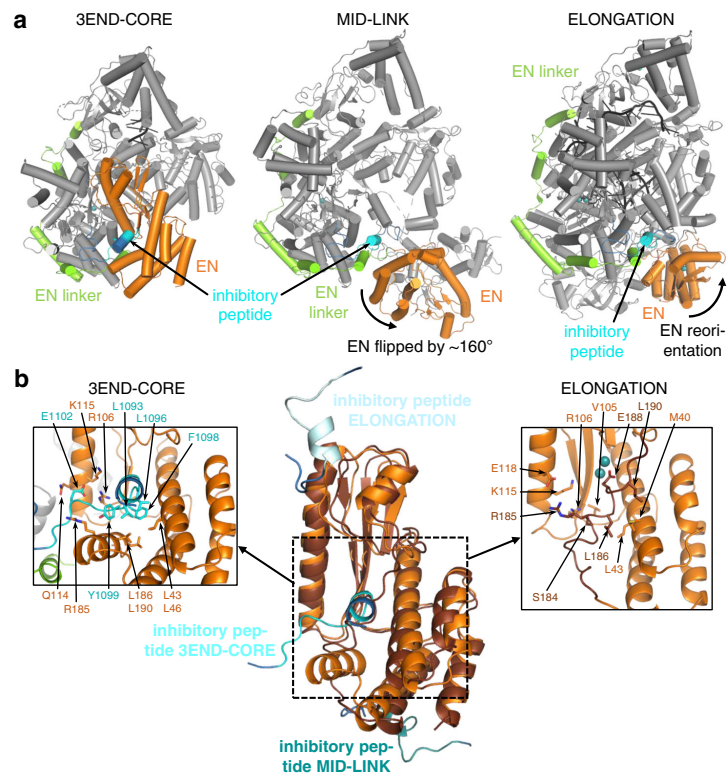


Fig. 5 Conformational flexibility of the endonuclease domain. **a** Overview of the three different conformations of the EN (orange) observed in the 3END-CORE, MID-LINK and ELONGATION structures. The EN linker (green) and the inhibitory peptide (cyan) are highlighted as well. **b** In the middle panel a superimposition of the EN domains of 3END-CORE (orange), corresponding also to the overall conformation of the EN domain in the MID-LINK structure, and ELONGATION (brown) with the respective positions of the inhibitory peptides in teal, cyan and light blue, respectively, is shown. Connections to the fingers domain are indicated in blue. The position of the inhibitory peptide of 3END-CORE is the same as in the APO-ENDO structure, similarly is the same position of the EN observed in both MID-LINK and DISTAL-PROMOTER structures. Right and left panels show close-ups of the autoinhibited EN active sites in the ELONGATION and 3END-CORE structures, respectively. Important residues of the protein-protein interactions are labelled and side chains are shown as sticks. For a focus on the inhibitory loop in the different conformations see Supplementary Fig. 11.

structure (Supplementary Fig. 7). The individual modules of the LASV C-terminal region have similar folds to that of the California Academy of Sciences reptaenavirus (CASV) L protein, despite low sequence homology (Supplementary Figs. 2 and 8)²³. However, the LASV CBD-like domain is considerably more elaborate, having ~ 170 residues compared to ~ 100 residues in CASV (Supplementary Fig. 8). The CASV CBD-like domain is minimalist, comprising a five-stranded mixed β -sheet with a transverse helical hairpin. In LASV, there are significant insertions in the $\beta 1$ - $\beta 2$ and $\beta 3$ - $\beta 4$ loops and in the loop of the helical hairpin, that fold together to extend the length of the domain (Fig. 4d and Supplementary Fig. 8). In the ELONGATION structure, the CBD-like domain is locked in position by interactions with the EN and core of the polymerase, with the helical hairpin insertion being particularly important for the latter interaction (Supplementary Fig. 9). In this conformation, the canonical cap-binding site between the C-terminal end of strand $\beta 1$ and the principal transverse helix, as observed in cap-bound CBDs of SFTSV, Rift-Valley fever virus (RVFV) or influenza virus (Supplementary Fig. 8), appears to be partly blocked. Moreover,

the CBD-like and 627-like helical-hairpin domains are particularly poorly conserved across arenavirus L proteins (as opposed to the mid-link and C-terminal β -barrel domain) (Supplementary Data 1 and Supplementary Fig. 8). Indeed, the most conserved region of the CBD-like domain is at the N-terminus of strand $\beta 1$ (1909-GYAW in LASV). Past and recent mutational analyses of the potential cap-binding aromatic residues of this domain using the LASV mini-replicon system did not identify any transcription-specific residues that might be responsible for cap-binding (Fig. 4d and Supplementary Fig. 8)²⁴. Neither have in vitro studies with isolated soluble domains of the CASV and LASV L protein C-terminal region detected any cap-binding activity²³, unlike for phenui- and orthomyxoviruses^{25–27}. It, therefore, remains an open question, whether in some, yet to be observed configuration of the L protein, a functional cap-binding site is formed.

Positional flexibility and regulation of the endonuclease. We observe three quite different locations of the EN, in each case

differently packed against the core of the L protein (Fig. 5). In the APO-ENDO and 3END-ENDO structures, the EN interacts with the core regions 1137–1142 and 1592–1604 (Supplementary Fig. 10). Mutation to alanine of F1592, which packs on P109 of the EN, resulted in a slight reduction in general L protein activity in the LASV mini-replicon system. A severe, general loss of 90% of activity was observed for L protein mutant P109G (Supplementary Fig. 10). The EN active site itself is exposed to the outside, but access is blocked by residues 1092–1105 from the PB1-like region, which we refer to as the ‘inhibitory peptide’ (Fig. 5a, b). The arena-conserved 1096-LCFYS motif (Supplementary Data 1) is intimately bound in the EN active site pocket (Fig. 5b and Supplementary Figs. 10, 11) and would thus prevent any substrate RNA binding there (e.g. superposition of LCMV EN structure with bound inhibitor shows overlap²⁸) (Supplementary Fig. 11). A similar interaction with the EN is observed in the apo-MACV structure (PDB:6KLD) (Supplementary Fig. 4).

In the MID-LINK and DISTAL-PROMOTER structures, the EN has flipped by ~160° around a hinge between G195–G199. The EN active site faces away from the rest of the polymerase and is exposed to the solvent, free of the inhibitory peptide (Fig. 5a, b). Instead, inhibitory peptide residues 1087–1099, as well as PB2-like segments 1759–1770 (lid), 1852–1860 and 1894–1896 (mid), 2077–2081 and 2089–2091 (link) pack against the back of the EN, stabilising it in its new location. For these two observed positions of the EN, the total buried surface area is comparable, 3005 Å² (APO-ENDO, autoinhibited) and 2721 Å² (MID-LINK, free), compatible with there being an equilibrium between the two states as observed in the two different apo-structures.

In the ELONGATION structure, the EN is stabilised in a third position (Fig. 5a, b) with its active site autoinhibited by a completely different mechanism involving the C-terminal region of the EN (173–190). This is redirected so that the 181–188 helix binds in and blocks the EN active site groove, with E188 coordinating, together with E51 and D89²¹, two cations in the active site (Fig. 5b and Supplementary Fig. 10). The inhibitory peptide remains at the same place with respect to the polymerase core, but due to the reorientation of the EN it packs against a different site on the EN, with, for instance, K1094 making a salt-bridge with E70 and Y1099 stacking against P81 (Fig. 5b and Supplementary Fig. 11). Diverse other regions of the L protein also contact the EN (Fig. 3b and Supplementary Fig. 9) and the total buried surface area of the EN in this location is 4060 Å².

To investigate the function of the inhibitory peptide, we performed a mutational analysis of residues 1092–1105 as well as interacting residues of the EN domain, as observed in the APO-ENDO and 3END-CORE structures (Supplementary Figs. 10, 11). We observed a severe general defect in L protein function for a number of mutants both in the EN domain (L43G/N, L46G/N, V105G, R106K, K115A, R185A, L186G, L190G/N) and in the ‘inhibitory peptide’ (L1093S, L1096A/N, C1097G, F1098A/S, Y1099A, E1102A) (Supplementary Fig. 11). For hantavirus L protein it was shown that an active EN can lead to RNA degradation and therefore lower protein expression levels²⁹. To exclude that the general defect of the mutants of the ‘inhibitory peptide’ interaction is caused by RNA degradation due to elevated activity of LASV EN, we combined the previous mutations with the EN inactivating mutation D89A³⁰, without observing any change in phenotype (Supplementary Fig. 12). Additionally, using *in vitro* polymerase assays only residual polymerase activity was detectable for L protein mutant E1102A and no activity for mutant Y1099A (Supplementary Fig. 13). We conclude that the ‘inhibitory peptide’ and other tested residues involved in the interaction play a general role in L protein activity but are not selectively important for transcription, this being consistent with the diversity of interactions we see for these residues when

comparing all observed conformations of the L protein (Fig. 5). Comparing the apo- and pre-initiation structures suggests that either promoter binding or mutations in the inhibitory peptide release the EN from autoinhibition. We tested this hypothesis by assaying purified L proteins with mutations in the inhibitory peptide for EN activity *in vitro*, using capped or uncapped RNA substrates and with either no promoter, 3′ end only, 5′ end only or both promoter RNAs present (Supplementary Fig. 14). The EN active site mutant E102A and the addition of the nuclease inhibitor DPBA served as negative controls. We found that, for the wild-type L protein, the only situation where weak EN activity is reproducibly detectable is when the 5′ end only or both promoter ends are bound, and the same is true for the L protein with mutations Q114A or E1102A, which are probably not sufficient to disrupt inhibitory peptide binding.

In summary, in both the apo- and early elongation states, the EN is autoinhibited, but by different mechanisms involving binding in the active site of either the ‘inhibitory peptide’ 1092–1105 or the C-terminal helix of the EN, respectively. Whilst 5′ end or full promoter binding partly activates the EN, consistent with the structure and presumed functional role in cap-dependent transcription priming of the pre-initiation state, its low intrinsic activity *in vitro* under any conditions tested by us and others^{20,21,31} suggests that the mechanism of EN activation may be more complex than expected from the currently available structural data.

3′ end binding in the secondary site. Incubation of LASV L protein with vRNA 3′ end nucleotides 1–16 yielded the currently highest resolution structure (3END-CORE, 2.70 Å). It features specific binding of nucleotides 1–7 in a tunnel under the pyramid (Fig. 1b, c). This site corresponds to the secondary 3′ end site previously observed for influenza virus polymerase^{16,17}, LACV L¹³ and MACV L⁹. The excellent cryo-EM density enables unambiguous base identification of nucleotides 2–5 (CGUG) of the 3′ end and placing of several water molecules in the protein–RNA interface (Supplementary Fig. 15). However, G1 and nucleotides 6–7 (UC) have poor density. In the MACV L structure, G1 is better defined perhaps due to its stabilisation by stacking on protein residue W534, which is substituted by L540 in LASV L (Supplementary Data 1). Nucleotides 3–5 of the 3′ end form a particularly compact arrangement with a direct interaction between G3 O6 and G5 N2, and U4 stacking underneath (Fig. 1c and Supplementary Fig. 15).

Several specific protein–RNA interactions are made with conserved arenavirus residues such as K332, D380, L502 and K509 from the pyramid and Y1450, R1452 and S1626 from the thumb and thumb-ring domains, thus involving the PA-, PB1 and PB2-like regions, as in influenza virus and LACV (Fig. 1c). Mutational analysis of the residues interacting with the 3′ end in the secondary binding site using the LASV mini-replicon system revealed a general defect in L protein function upon introduction of double mutations N331A/K332A and Y1450A/R1452A, whereas mutations L502A, K509A and R1622A did not interfere with L protein activity (Fig. 1d). Purified L protein mutant Y1450A/R1452A exhibited significantly reduced 3′ end binding ability compared to wild-type L protein (Fig. 1e). However, this mutant maintained polymerase activity in the presence of the 19 nt 3′ and 20 nt 5′ promoter RNAs and, in contrast to the wild-type L protein, showed polymerase activity with only the 19 nt 3′ promoter RNA present (Supplementary Fig. 16). This strongly suggests that in the wild-type L protein, in the absence of the 5′ end, the 3′ end is tightly sequestered in the secondary binding site and does not enter the active site (see discussion). In the presence of a 47 nt hairpin RNA containing the connected 3′ and 5′

promoter sequences of LASV, L-Y1450A/R1452A showed significantly reduced polymerase activity (Supplementary Fig. 16). This shows that 3' end binding in the secondary site is required for efficient RNA synthesis, either to sequester the template 3' end after passing through the active site and/or to prevent the unbound 3' end from forming double-stranded RNA with the template 5' end or the product RNA.

Full promoter binding. The PRE-INITIATION structure shows the full promoter (5' nts 0–19, 3' nts 1–19) bound to the LASV L protein and reveals several significant conformational changes that occur upon promoter binding (Fig. 2 and Supplementary Fig. 1). The LASV vRNA 5' and 3' ends are highly complementary over 19 nucleotides with only two mismatches at positions 6 and 8 (Supplementary Table 1). In addition, the 5' end carries an extra nucleotide (G0) arising from a prime-and-again mechanism during replication initiation^{4,5,32–34}. As expected, when bound to the L protein, the promoter does not adopt a fully double-stranded conformation but forms a structure resembling that observed for influenza virus and LACV (Fig. 2a, c, d, e and Supplementary Fig. 17). Nucleotides 12–19 from both strands form a distal 8-mer canonical A-form duplex, whereas nucleotides 1–11 of each end are single-stranded, which is consistent with previous mutational studies³⁵. Nucleotides 0–9 of the 5' end form a compact hook structure linked to the duplex region by nucleotides 10–11. The internal secondary structure of the hook differs from that of influenza virus and LACV polymerases by only having one canonical base-pair (C1–G7) upon which G2 and C3 are consecutively stacked on one side and G8 and G9 on the other side (Fig. 2d and Supplementary Fig. 17). The loop of the hook comprises nucleotides 4–6, with A4 stacking on C6. The hook makes extensive interactions with conserved residues from numerous different loops from both the PA-like and PB1-like regions of the L protein (e.g. residues 470–474, 518–526, 860–861 and 1255–1265) (Fig. 2d and Supplementary Movie 1). Several of these loops only become structured upon promoter binding. Of note, three aromatic residues are involved, Y471 (conserved in Old World arenaviruses only), Y526 (Y or F in all arenaviruses) and Y574 (conserved in all arenaviruses) (Supplementary Data 1). Y471 interacts with the phosphate of G10 and Y574 with the base of G9. Y526 extends and stabilises the central stacked backbone of the hook by packing on G9, with I665 playing a similar role on the other end by stacking against C3. G0 is base-specifically recognised by the carbonyl-oxygen of S470 and is sandwiched between P297 and R473, which could potentially interact with the terminal triphosphate (not present in the 5' RNA used for this structure) (Supplementary Fig. 17). Mutational analysis of the 5' binding site using the LASV mini-replicon system reveals that most mutations cause a severe general defect in L protein activity (Supplementary Fig. 1). Complete loss of function was observed for mutants Y474A, V514G/K515A, R525A/Y526A and K681A. An activity reduction to ~20% was observed for mutants R473A/T474A, Q551A/K552A, Y574A and K1263A/T1265A. These results indicate that 5' end binding is important for the general function of LASV L protein during both viral transcription and genome replication.

The 3' strand of the distal duplex specifically interacts with two regions of the L-protein pyramid (Figs. 2b and 6a). One is the 340-loop, which orders upon binding, with the arenavirus-conserved motif 340-NTRR making several major groove contacts with bases and phosphates of the 3' strand nucleotides 12–15 (and R337 with phosphate of C10). The second interacting region involves the Old World arenavirus-specific extended helices at the top of the pyramid, including W395, R399 (conserved in all Old World arenaviruses) and K423 contacting

the backbone phosphates of 3' strand nucleotides 16–17 (Fig. 2b and Supplementary Movie 1). The single-stranded 3' end nucleotides 11 to 7 are directed towards the polymerase active site, but 1–6 are not visible (Fig. 2a, c, e). Arenavirus-conserved R1561 interacts base-specifically with C9, which also stacks on chemically conserved R1622, both residues being from the thumb domain (Fig. 2b and Supplementary Movie 1).

Apart from the induced fit ordering of several promoter-interacting loops, there are also more global rearrangements. Firstly, distal duplex binding causes a major rotation of the entire pyramid by ~21.6°, enabling the summit helices to contact the 3' strand as described above (Fig. 6a). This rotation does not occur when just the 3' end is bound in the secondary site. Secondly, the α -bundle rotates slightly (4.5°, PRE-INITIATION versus APO-RIBBON) to allow interaction of the 860-region (860–861) with the 5' hook (Fig. 6a). Thirdly, the pendant domain (943–1040) becomes stably positioned by packing against the rotated pyramid domain and the thumb and thumb-ring (Fig. 6b). The pendant domain helix spanning residues 953–965, runs parallel to 3' nucleotides 9–12, but with only two direct interactions. R957 contacts the phosphate of A12 contributing to stabilisation of the 3' end interacting 340-loop and K960 is close to the phosphate of U11 (Fig. 2b). The pendant domain was first visualised in the MACV apo-L structure in a very different position that would superpose with the promoter duplex region as well as the α -bundle, as observed in the promoter-bound LASV L structure (Supplementary Fig. 6). Similarly, the α -bundle of MACV apo-L (with its different topology, see Supplementary Fig. 6) superposes with the pendant domain in LASV L, showing that these flexible and linked domains must rearrange from the apo-state observed in MACV L. Finally, the position of the pendant domain in the LASV PRE-INITIATION structure is incompatible, due to significant steric overlap, with the location of the inhibited EN in the APO-ENDO or 3END-ENDO structures (Fig. 6c). This important observation provides a plausible rationale for full promoter binding (as opposed to just the 3' end binding in the secondary site) inducing a flip of the EN to the alternative uninhibited location observed in the MID-LINK and DISTAL-PROMOTER structures (see above).

Elongation structure. To determine a structure of functionally active LASV polymerase in early elongation, we incubated promoter-bound L protein with a 10-mer uncapped primer, 5'-AAUAAUACGC-3' together with ATP, GTP, CTP and non-hydrolysable UMPNPP (denoted U_{PNPP}). Biochemical analysis (Supplementary Fig. 18) shows that various products are formed depending on whether (i) the 3' terminal triplet of the primer hybridises with the 3' end of the template (3'-GCGUGUCA...) giving 14-mer 5'-AAUAAUACGCACAGU $_{PNPP}$ or 18-mer: 5'-AAUAAUACGCACAG[U]GGAU $_{PNPP}$ products or (ii) just the 3' terminal nucleotide hybridises, giving 16-mer 5'-AAUAAUACGCACAGU $_{PNPP}$ or 20-mer 5'-AAUAAUACGCACAG[U]GGAU $_{PNPP}$ products (here, underline indicates primer hybridisation, italics indicate incorporated nucleotides, [U] misincorporation at an A in the template) (Supplementary Fig. 18). In each case, the longer product is formed if the A8 in the template is read-through by misincorporation. The most prominent products are the 14-mer and 18-mer. Upon plunge freezing on EM grids and performing 3D single-particle reconstruction, the sample gave a major class showing the stalled (i.e. pre-incorporation), early elongation state at 2.92 Å resolution. The high quality of the density allows unambiguous assignment of the template and product sequences. As expected, incoming UMPNPP is observed at the +1 position base-pairing with A8 of the template. There is good density for seven bases of the product

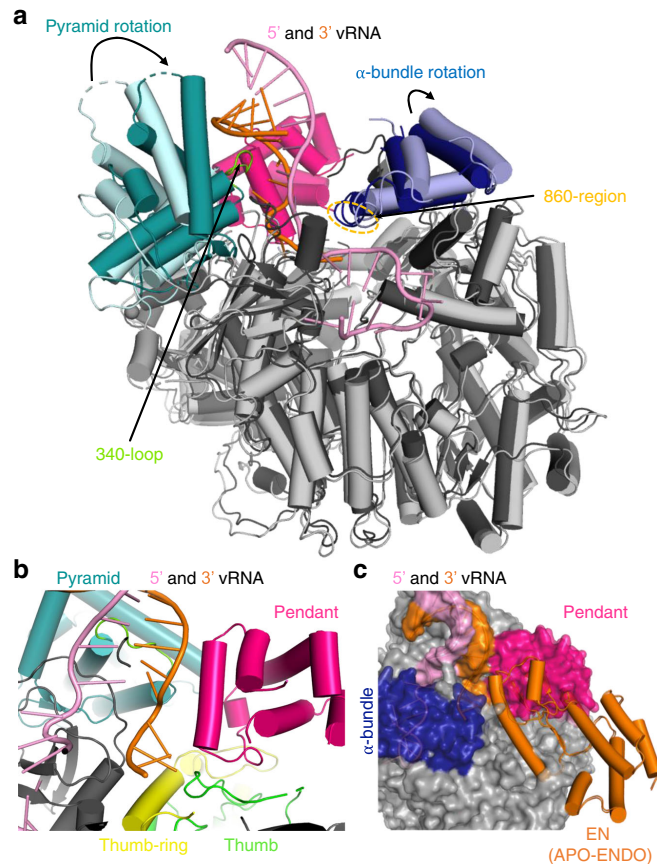


Fig. 6 Global rearrangements upon promoter binding. **a** PRE-INITIATION (dark grey, teal, dark blue, hot pink) and APO-RIBBON (light grey, light cyan, light blue) structures are superimposed. The pyramid and α -bundle rotations between apo- and promoter-bound structures are indicated as well as the 340-loop and the 860-region. Promoter RNA is shown in pink and orange. **b** Close-up of the interaction site between promoter RNA (pink and orange) and the pendant (hot pink), thumb-ring (yellow) and thumb (green) domains in the PRE-INITIATION structure. The pyramid domain is shown in teal. **c** Superposition of the PRE-INITIATION and APO-ENDO structures. PRE-INITIATION is presented as a transparent surface in grey with the pendant (hot pink) and α -bundle (blue) as well as the 5' (pink) and 3' (orange) vRNA highlighted in colour. The EN domain of the APO-ENDO structure is shown as an orange ribbon, which overlaps with the pendant domain volume of the PRE-INITIATION structure.

(5'-...N₁N₂CGCACAG), with the preceding two, which are unpaired with the template, having poorer density, prohibiting unambiguous identification (Supplementary Fig. 19). However, since the map density corresponding to N₂ looks more like an A, the product is most likely the 14-mer. Products that would contain mismatches due to read-through, maybe less stably bound to the polymerase. The active site cavity contains an 8-mer duplex from position +1 to -7 (+1 corresponding to the UMPNPP) (Fig. 4a, e, f), whereas as observed for related viral polymerases^{14,22} a 10-mer duplex (positions +1 to -9) is expected to fill the active site cavity before strand-separation. Nucleotides 1–11 of the template are visible, corresponding to positions +4 to -7. The duplex region of the promoter has melted due to translocation of the template but nucleotides 0–11 of the 5' end remain bound in the hook conformation as described

above. Consistent with the distal duplex being absent, the pyramid is not rotated and a second template 3' end (nucleotides 1–7) is actually bound in the secondary site (Fig. 4b). Whilst this is likely an artifact of performing the RNA synthesis reaction with an excess of template, it does confirm that secondary 3' end binding is compatible with elongation and consistent with the template docking in this site after exiting the active site cavity, as observed for influenza virus polymerase¹⁸. The pendant domain is not visible and the α -bundle only has weak density, probably due to the absence of the distal promoter duplex.

The configuration of the polymerase active site, as well as the binding of the incoming nucleotide and template, are canonical, involving conserved motifs A–D and the fingertips (motif F) (Fig. 4e and Supplementary Movie 2). The triphosphate and terminal 3' OH of the product strand are coordinated by two

manganese ions (A and B), held in place by D1190 (motif A) and 1331-DD (motif C). K1373 of motif D also contacts the γ -phosphate. Motif F residues R1131, positioned by Q1294 (motif B), and L1133 stack under the incoming nucleotide template bases at the +1 position, respectively, while K1124 (motif F) contacts the O4 of the incoming nucleotide (Supplementary Movie 2).

Compared to the unoccupied active site in the PRE-INITIATION structure, only minor adjustments to the active site loops occur, the most significant being the displacement of the central β -strands of the fingertips loop by about 2 Å to make room for the +1 base-pair to stack on R1131 and L1133 (Supplementary Fig. 20). Unlike in influenza virus polymerase²², motif B does not change conformation between the occupied and unoccupied states of the active site. However, to accommodate the growing template-product duplex, the helical lid (1731–1805) has to be displaced out of the active site cavity by about 8 Å (Supplementary Fig. 20). Coupled with the lid movement, the sharply kinked (requiring conserved G1595) pair of consecutive helices α 52– α 53 (1579–1611) also translates in the same direction, with helix α 53 forming one side of the active site cavity close to the distal part of the product strand. Interestingly, the first visible base of the product (position –9) is packed against T1583 from α 52, which could therefore play a role in strand-separation rather than the helical lid itself (Supplementary Fig. 20).

More generally, the conformation of the active elongating polymerase is stabilised by a number of new interactions between distant regions of the L protein sequence. For instance, in the new position of the helical lid, residues 1764–1766 contact the EN at F85 (close to the EN active site), contributing to the interactions which stabilise the EN in its third location (see above). Residues 811–820, disordered in all other structures, interact with EN linker 195–199, again only possible with the EN in its new location. Similarly, the inhibitory peptide residues 1087–1091 change conformation to allow simultaneous interaction with the EN (1089-TT with D129 and S82) and with the kink between α 52– α 53 (A1091 with T1591) (Fig. 5a and Supplementary Fig. 11). Most dramatic, is the stabilisation of the entire C-terminal domain, which, together with the exposed end of the palm, forms a ring with a ~30 Å diameter central pore, a putative product exit channel (Fig. 4b). The EN buttresses the proximal part of the CBD-like domain (e.g. Q32 with A1911) as well as the mid-link domain (e.g. A171 with K1895 and E34 with K1891), whereas the distal part of the CBD-like domain interacts with numerous loops from the polymerase core including EN linker residues 230–232 (e.g. H232-Q2045), fingers domain residues 793–799, 802–805 (e.g. V802-Y2030), 1215–1216 (e.g. D1216-K2062, K1215-E2053) and palm domain residues 1314–1318 (e.g. Y1314-Q2045) (Supplementary Fig. 9). The total buried surface area between the polymerase core and the CBD-like domain is 2884 Å². The extreme C-terminal 627-like domain (mainly the β -barrel and to a lesser extent, the helical hairpins) make interactions with multiple regions, notably 1715–1722 and 1812–1816 of the thumb-ring (e.g. F1715-Y2176/V2145, F1716-V2189, D1722-S2191/S2192, R1816-D2143, S1812/L1815-G2175), residues 691–694 of the helical region (e.g. M691-G2193) and the palm domain 1390–1392 loop (e.g. W1390-R2197) (Supplementary Fig. 9). The total buried surface area of the β -barrel domain is 1472 Å². Interestingly, W2170, R2197 and R2201, whose mutation leads to a transcription-specific defect²⁴, are intimately involved in the interface together with W1390 in the 1390-loop. Even though mutation of W1390 to alanine in a previous study did not impair overall L protein activity³⁶, from the structure we would not expect a small hydrophobic alanine residue to disturb the remaining contacts between these domains. Additionally, residues G1391 and D1392 were shown to be selectively

important for viral transcription³⁶, further emphasising the importance of this interaction site. Residue Y2176, also identified as being selectively important for viral transcription²⁴ interacts with the thumb-ring residue F1715 (Supplementary Fig. 9). These data suggest that the configuration of the C-terminal region and its interaction with the core, as observed in the ELONGATION structure is critical for transcription.

Discussion

Previous biochemical studies on LASV L⁷ and MACV L^{8,37} proteins have revealed certain features of promoter binding to arenavirus polymerases and the impact on RNA synthesis activity. In RNA binding experiments, it was shown that MACV L makes a tight complex with the 3' promoter strand with the identity of nucleotides 2–5 being particularly important⁸. This corresponds exactly with the binding specificity of the 3' end secondary site seen in our structural analysis. For both MACV and LASV, the most efficient *in vitro* RNA synthesis activity was observed using both 19-mer strands in 1:1 ratio as in the native promoter^{7,37}. For MACV L protein weak activity, which could be enhanced with a GpC primer, was also observed in presence of only the 3' strand⁸. The more specific requirements found for optimal unprimed RNA synthesis by LASV L were (i) the presence of the terminal non-templated G0 base on the 5' strand (i.e. 0–19); (ii) the two mismatches at positions 6 and 8 in the S segment promoter (only one in the L segment), rather than a perfect proximal duplex; (iii) a sufficiently long distal duplex region, preferably the full 19-mer⁷. For MACV, it was further shown that the G0 phosphates were not essential and that enhancement was achieved with 3' truncated 5' ends (e.g. maximal activity for 0–12 mer), although these experiments differed in that a GpC primer was systematically used³⁷. These observations are consistent with our structural analysis as well as the notion that the default mode of binding of the 3' end alone is in the secondary site and that full-promoter binding or a primer is required to dislodge it and permit RNA synthesis. More recently, it has been confirmed for both LASV and MACV L that *de novo* RNA synthesis is indeed enhanced by the presence of the 5' end, but, surprisingly, it was reported that cap-dependent transcription was inhibited by the 5' end⁹. This raises the question of the exact role of the 5' end binding in arenavirus L proteins.

In LACV and influenza virus, 5' end hook binding is required to order the fingertips loop into a functional configuration in the polymerase active site^{13,14}. This does not appear to be the case for arenavirus L proteins⁹ (Supplementary Fig. 5). In addition, for the influenza virus, it has been shown that 5' end binding stimulates EN activity, probably by favouring the transcription active configuration of the polymerase over the replicase conformation¹⁹. Evidence given above that 5' end (or full promoter) binding stimulates EN activity suggests that a similar conformational change mechanism may operate for arenavirus L proteins. Finally, whereas 5' end binding is required in orthomyxoviruses for poly(A) tail generation during transcription¹⁸, arenavirus L proteins terminate transcription by a very different mechanism without poly(A) tail synthesis^{38–41}. To investigate the functional consequences of 5' end hook binding further, we performed polymerase activity assays either (i) with the 3' promoter end alone, (ii) the 3' end together with the 5' (nts 0–19) end or (iii) the 3' end together with a truncated 5' (nts 0–12) end. In each case, the assays were performed in the presence or absence of 3 nt or 10 nt long uncapped primers (Supplementary Fig. 21). We used uncapped primers as we could not detect any difference in primed product formation between uncapped (hydroxylated or tri-phosphorylated 5') and cap0-capped primers for LASV L (Supplementary Fig. 22). Under the conditions of the reaction, no

products were formed by the 3' end alone unless a primer was present (Supplementary Fig. 21). Adding the promoter 5' end (nts 0–19) led to strong product formation even in the absence of primer. When the 5' end was truncated (nts 0–12), unprimed product formation was significantly reduced, whereas primed product formation was comparable to the corresponding conditions with the full 5' (nts 0–19) RNA. In the absence of the 5' end, structural and biochemical data show that the 3' end preferentially binds tightly in the secondary binding site. Bearing this in mind, we interpret our activity results to show that 5' end binding and/or the presence of a primer (capped or uncapped), that can hybridise with and stabilise the 3' end in the polymerase active site, stimulates RNA synthesis activity, presumably by promoting 3' end binding in the active site rather than the secondary site. Indeed, the major rotation of the pyramid domain induced by distal duplex binding shears the two sides of the 3' end-binding groove and prevents closure around the RNA, thus disfavours the secondary site 3' end binding when the full promoter is bound. Since the promoter duplex no longer exists during elongation, this mechanism does not prevent the template 3' end rebinding in the secondary site after going through the active site as observed for influenza polymerase¹⁸.

In conclusion, our structural and functional results support the hypothesis that full-promoter binding, including the 5' hook and distal duplex, induces the functionally ready pre-initiation state via the conformational changes described above, at the same time as releasing the EN from autoinhibition (Fig. 7). Given that our activity results are independent of whether the primer is capped or uncapped (Supplementary Fig. 22), we do not think they shed light on true cap-dependent transcription per se. It remains

unclear, whether the assays presented by Peng et al.⁹ truly reflect cap-dependent transcription, as the respective control (i.e. non-capped primer) was not included. Indeed, we have not been able to recapitulate true cap-dependent transcription with the LASV polymerase, due to its weak or non-existent endonuclease and cap-binding activities, emphasising that the mechanism of cap-snatching and cap-dependent transcription for arenavirus L protein remains enigmatic. This is even more remarkable considering the shortness of the capped primers used³, which it is difficult to imagine being bound in a canonical way by the CBD-like domain as well as reaching into the active site to hybridise with the template, a situation reminiscent of the Thogoto virus polymerase⁴².

The autoinhibited EN conformations appear in L protein configurations where the EN activity is not expected to be required and indeed could potentially be detrimental (Fig. 7). On the other hand, in the promoter-bound pre-initiation state, EN activity is required^{30,31}, presumably to generate capped primers. Correspondingly, in this state, we observe that the EN is not autoinhibited, although how it might act in collaboration with the putative CBD is unclear. Whilst these biochemical experiments support the structure-based hypothesis that 5' end only or full promoter binding activates the EN due to its displacement by the pendant and α -bundle domains, the observed EN activity is very weak (Supplementary Fig. 14), consistent with the barely detectable *in vitro* activity of the isolated EN domain^{20,21,31}. This suggests that some other L protein configuration or possibly a host factor may be required to fully activate the EN inside infected cells. Similarly, interaction with a currently unknown host cap-binding protein may be required to

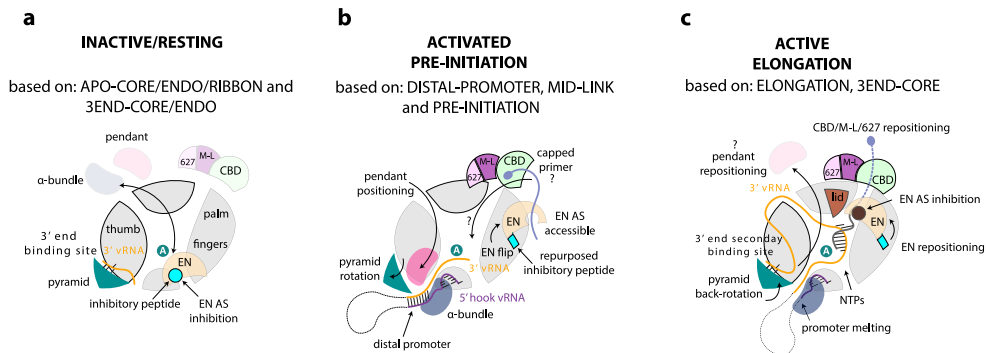


Fig. 7 Schematic diagram of conformational changes in Lassa virus L protein associated with promoter binding and RNA synthesis activity. **a** On the surface of the inactive/resting L protein core, there is mutually exclusive positioning (black double-arrow) of either the α -bundle and pendant (APO-RIBBON structure) or the EN domain (APO-ENDO, 3END-ENDO structures). When placed on the core, the inhibitory peptide (cyan) autoinhibits the EN domain by binding in its active site. In the absence of the 5' vRNA, the 3' vRNA binds preferentially, base-specifically, into a distinct secondary 3' RNA binding site between the pyramid and thumb domains (3END-ENDO/CORE structures). **b** Upon full promoter binding, major conformational changes occur (based on PRE-INITIATION, DISTAL-PROMOTER, MID-LINK structures). The 5' end nucleotides 0–9 are bound in a hook-like conformation in a specific pocket outside the active site. The distal promoter (formed by 3' and 5' vRNA nucleotides 12–19) is positioned by tight association with the α -bundle and pendant, with concomitant rotation of the pyramid domain. This forces release of the 3' vRNA from the secondary binding site, allowing it to be directed towards the RNA synthesis active site (marked by the white A in the teal circle). Positioning of the pendant domain next to the distal promoter displaces the EN domain, which relocates to the other end of the L protein core with the inhibitory peptide contacting its surface leaving the EN active site accessible. In this configuration, the EN is presumed to be in close vicinity to the CBD-like domain and could potentially cleave an incoming capped RNA to generate a transcription primer by 'cap-snatching'. How the capped primer associates with the CBD-like domain and how it is navigated towards the active site to initiate transcription remains elusive. **c** Upon transition to the elongation state, the distal promoter duplex melts and the pendant domain is released, which allows the pyramid to rotate back and re-establish the availability of the secondary 3' end binding site (ELONGATION structure). We presume the 3' vRNA template, after exiting the active site, wraps around the L protein core and rebinds to the secondary 3' end binding site, as described for influenza virus polymerase complex¹⁸. The EN repositions once again and together with the mid-link and CBD-like domains form a highly structured ring around the putative product exit channel. The EN active site is again autoinhibited, this time by the relocation of its C-terminal helix (181–188) to the active site.

present capped RNAs to the L protein. Identification of how arenavirus L proteins access host capped RNAs is a key requirement for further understanding of the mechanism of transcription.

Whilst this paper was under review, two papers were published describing the cryo-EM structures of the complex between the L protein and the multifunctional Z proteins of MACV and LASV⁴³ and of Junin virus⁴⁴. The Z protein is known to inhibit RNA synthesis^{7,45,46}. All L-Z structures show that the central, folded region of the Z protein, which includes the RING domain, binds to the external rim of the palm domain of L. In each case, the L protein is in an inactive state, with at most the 3' vRNA end bound in the secondary site. Only the polymerase core is visible, except in a dimeric form of the MACV L-Z complex, where the C-terminal domain, including the CBD-like domain, is ordered. Using modelling of the RNA synthesis active state of the L protein and hydrogen-deuterium exchange experiments, Xu et al. conclude⁴³ that the mechanism of RNA synthesis inhibition is allosteric, with Z protein binding conformationally restraining two key polymerase active site motifs. On the other hand, Kang et al. suggest⁴⁴ that Z protein binding to the Junin virus L protein 680-loop (690-loop in LASV L) locks the L protein in an inactive state with Z positioned to block the product exit channel. A key contact mediating the LASV L (strain G3278-SLE-2013) and Z interaction involves Z protein/W35 contacting L protein/M694-F1381-W1392 with similar interactions in MACV and Junin virus complexes (MACV: Z protein/W43 with L protein/F689-F1378-M1389; Junin virus: Z protein/W43 with L protein/F688-F1377-M1388). We note that the W1390 in our LASV strain Bantou 289 (W1392 in strain G3278-SLE-2013) is also central to the interaction between the 1390-loop and the extreme C-terminal 627-like domain (R2197) as observed in our early elongation state structure (see above and Supplementary Fig. 23). Indeed, superimposition of the active ELONGATION structure with the L-Z complex structure clearly reveals that the binding sites on the L protein core region (involving the 690-loop and the 1390-loop) of the 627-like domain and the Z protein, respectively, substantially overlap, showing that this binding is mutually exclusive (Supplementary Fig. 23). Furthermore, we see no significant change in conformation in the 690-loop between the two structures, and the Z protein, although not resolved in full length in the L-Z complex structures, does not appear to block the putative product exit channel (Supplementary Fig. 23). In addition, binding of Z protein or the 627-like domain of L to the palm domain might reduce the conformational flexibility of the active site motifs, without this necessarily causing inhibition of RNA synthesis. We, therefore, suggest an alternative explanation for the mechanism of inhibition of L by Z: binding of Z protein to L protein sterically prevents the establishment of the RNA synthesis active state of L, with its well-defined ring of peripheral domains surrounding the exit channel. These observations emphasise the importance of determining structures of active states of the L protein in order to understand functional mechanisms.

Methods

Expression and purification of LASV L protein. The L gene of LASV Bantou 289 (accession no. MK044799) contains a StrepII-tag at an internal position (after residue 407, 407strep) or a StrepII-His tandem tag at the C terminus (Cstrep) was cloned into a pFastBacHT B vector⁷. If indicated, mutations were introduced by mutagenic PCR before cloning. Using DH10EMBacY E. coli cells^{47,48}, recombinant baculoviruses were produced and subsequently used for protein expression in Hi5 insect cells⁷. The harvested Hi5 insect cells were resuspended in Buffer A (50 mM HEPES(NaOH) pH 7.0, 1 M NaCl, 10% (w/v) Glycerol and 2 mM dithiothreitol), supplemented with 0.05% (v/v) Tween20 and protease inhibitors (Roche, cOmplete mini), lysed by sonication and centrifuged two times (20,000 × g for 30 min at 4 °C). Soluble protein was loaded on Strep-TactinXT beads (IBA) and eluted with 50 mM Biotin (Applichem) in Buffer B (50 mM HEPES(NaOH) pH 7.0, 500 mM NaCl, 10% (w/v) Glycerol and 2 mM dithiothreitol). L protein-containing fractions

were pooled and diluted 1:1 with buffer C (20 mM HEPES(NaOH) pH 7.0) before loading on a heparin column (HiTrap Heparin HP, GE Healthcare). Proteins were eluted with Buffer A and concentrated using centrifugal filter units (Amicon Ultra, 30 kDa MWCO). The proteins were subsequently used for biochemical assays and structural studies. For endonuclease assays, the L proteins were further purified by size-exclusion chromatography (Superose 6, GE Healthcare) in buffer B. Pure L proteins were concentrated as described above, flash frozen and stored at -80 °C.

In vitro LASV L complex reconstitution for cryo-EM

APO-structures. The LASV L-Cstrep protein was first injected onto a Superdex 200 Increase 3.2/300 column (GE Healthcare) equilibrated in 40 mM HEPES pH 7.4 (4 °C), 500 mM NaCl, 10 mM MgCl₂ and 1 mM TCEP. About 50 µl fractions were collected and the protein was eluted at a 2 µM concentration. Protein was diluted to 0.7 µM and aliquots of 3 µl were applied to Quantifoil R1.2/1.3 Au 300 mesh grids, immediately blotted for 2 s and plunged into liquid ethane using an FEI Vitrobot IV (4 °C, 100% humidity).

PROMOTER-DUPLEX and MID-LINK structures. The LASV L-Cstrep protein was first injected onto a Superose 6 Increase 3.2/300 column (GE Healthcare) equilibrated at 4 °C in 40 mM HEPES pH 7.4, 400 mM NaCl, 10 mM MgCl₂ and 1 mM TCEP. About 50 µl fractions were collected and the protein was eluted at a 2 µM concentration. Protein was diluted to ~0.9 µM and mixed with a 1.3-fold molar excess of truncated promoter vRNAs (5' nts 10-19, 3' nts 1-19) (Supplementary Table 1) for 10 min at 4 °C. Aliquots of 3 µl were applied to Quantifoil R2/2 Au 300 mesh grids, immediately blotted for 2 s and plunge frozen into liquid ethane using an FEI Vitrobot IV (4 °C, 100% humidity).

3' END structures. The LASV L-Cstrep protein was first injected onto a Superose 6 Increase 3.2/300 column (GE Healthcare) equilibrated at 4 °C in 40 mM HEPES pH 7.4, 250 mM NaCl, 10 mM MgCl₂ and 1 mM TCEP. About 50 µl fractions were collected and the protein was eluted at a 2 µM concentration. Protein was diluted to ~1.8 µM and mixed with threefold molar excess of 3' (1-16) vRNA (Supplementary Table 1) for 10 min at 4 °C. Aliquots of 3 µl were applied to Quantifoil R2/2 Au 300 mesh grids, immediately blotted for 2 s and plunge frozen into liquid ethane using an FEI Vitrobot IV (4 °C, 100% humidity).

PRE-INITIATION structure. The LASV L-Cstrep protein with a concentration of 1.4 µM in assay buffer (100 mM HEPES(NaOH) pH 7.0, 50 mM NaCl, 50 mM KCl, 2 mM MnCl₂ and 2 mM dithiothreitol) was mixed sequentially with single-stranded 5' (0-19) vRNA and single-stranded 3' (1-19) vRNA in 1.2-fold and primer S1 in 7.1-fold molar excess (all RNAs are listed in Supplementary Table 1). After 45 min incubation on ice, the reaction was started by the addition of NTPs (0.25 mM GTP/ATP). After incubation at 30 °C for 2 h, 3 µl of the reaction was applied to glow-discharged Quantifoil R2/1 Au G200F4 grids, immediately blotted for 2 s using an FEI Vitrobot Mk IV (4 °C, 100% humidity, blotting force-10) and plunge frozen in liquid ethane/propane cooled to liquid nitrogen temperature.

ELONGATION structure. The LASV L-Cstrep protein with a concentration of 3 µM in assay buffer (100 mM HEPES(NaOH) pH 7.0, 50 mM NaCl, 50 mM KCl, 2 mM MnCl₂ and 2 mM dithiothreitol) was mixed sequentially with single-stranded 5' (0-19) vRNA and single-stranded 3' (1-19) vRNA in 1.7-fold and primer C8 in 3.3-fold molar excess (all RNAs are listed in Supplementary Table 1). After 45 min incubation on ice, the reaction was started by addition of NTPs (0.25 mM GTP/ATP/UMP/PPP and 0.125 mM CTP). After incubation at 30 °C for 2 h, 3 µl of the reaction was applied to glow-discharged Quantifoil R2/1 Au G200F4 grids, immediately blotted for 2 s using an FEI Vitrobot Mk IV (4 °C, 100% humidity, blotting force-10) and plunge frozen in liquid ethane/propane cooled to liquid nitrogen temperature.

Electron microscopy

APO-, DISTAL-PROMOTER, MID-LINK, PRE-INITIATION and ELONGATION structures. The grids were loaded into an FEI Tecnai Krios electron microscope at the Centre for Structural Systems Biology (CSSB) Cryo-EM facility, operated at an accelerating voltage of 300 kV and equipped with K3 direct electron counting camera (Gatan) positioned after a GIF BioQuantum energy filter (Gatan). Cryo-EM data were acquired using EPU software (FEI) at a nominal magnification of x105,000, with a pixel size of 0.85 or 0.87 Å per pixel. Movies of a total fluence of ~50 electrons per Å² were collected at ~1 e-/Å² per frame. A total number of 15,488 (APO-); 13,462 (DISTAL-PROMOTER and MID-LINK); 13,204 (PRE-INITIATION); 10,368 (ELONGATION) movies were acquired at a defocus range from -0.4 to -3.1 µm (Supplementary Table 2).

3' END- structures. The grids were loaded into an FEI Tecnai Krios electron microscope at European Synchrotron Radiation Facility (ESRF) beamline CM01⁴⁹, operated at an accelerating voltage of 300 kV and equipped with K2 Summit direct electron counting camera (Gatan) positioned after a GIF Quantum energy filter (Gatan). Cryo-EM data were acquired using EPU software (FEI) at a nominal magnification of x165,000, with a pixel size of 0.827 Å per pixel. Movies of a total fluence of ~50 electrons per Å² were collected at ~1 e-/Å² per frame. A total

number of 6616 movies were acquired at a defocus range from -0.3 to -2.8 μm (Supplementary Table 2).

Cryo-EM image processing. All movie frames were aligned and dose-weighted using the MotionCor2 programme (Supplementary Figs. 24, 26, 28, 30, 32). Thon rings from summed power spectra of every $4 e^{-}/\text{\AA}^2$ were used for contrast-transfer function parameter calculation with CTFIND 4.1⁵⁰. Particles were selected with WARP⁵¹. The further 2D and 3D cryo-EM image processing was performed in RELION 3.1⁵². First, particles were iteratively subjected to two rounds of 2D classification (Supplementary Figs. 24, 26, 28, 30, 32) at 2x binned pixel size. Particles in classes with poor structural features were removed.

3D analysis of the APO-structures. Two times binned particles (1491 k) were subjected to two rounds of 3D classifications with image alignment (Supplementary Fig. 25). The first round of 3D classification was restricted to ten classes and performed using a 60 \AA low-pass filtered initial model constructed from 16 most populated 2Ds class averages (Supplementary Fig. 25). Particles in classes with poor structural features were removed. The second classification (into ten classes) was done during two rounds of 25 iterations each, using the regularisation parameter $T = 4$. In the second round, local angular searches were performed at 3.5° to clearly separate structural species. Three major 3D species were identified: the bare CORE-like, the ENDO-like and RIBBON-like.

In the first branch of 3D classification, the focus was on the structure of the CORE of the LASV L protein. All identified species were pooled together (428.8 k particles) and CORE-focused global 3D refinement was performed. Another round of CORE-focused 3D classification was performed based on the global refinement with local angular searches at 0.9° to clearly separate structural species. The most defined class (67.4 k particles) was further CORE-focused 3D auto-refined and iteratively aberration-corrected. For Bayesian-polishing only the first 23 frames were used.

In the second branch of 3D classification, the focus was on the structure of the ENDO-like species. The ENDO-like species from CSSB DATA 1 and DATA 2 were pooled together (209.4 k particles) and globally 3D auto-refined. A global 3D classification was performed based on the global refinement with local angular searches at 0.9° . The most defined class (74 k particles) was further core-focused 3D auto-refined as described for the CORE-like species.

In the third branch of 3D classification, the focus was on the structure of the RIBBON-like species (35.4 k particles). The RIBBON structure was obtained in a similar way as the ENDO one.

3D analysis of the 3' END structures. Two times binned particles (654.8 k) were subjected to global 3D auto-refinement with 60 \AA low-pass filtered APO-ENDO structure as the initial model. In the first branch of 3D classification, the focus was on the structure of the 3' end binding site. A specific mask containing the 3' vRNA secondary binding site and pyramid domain area was created (yellow, Supplementary Fig. 27) and focused 3D classification without angular assignment was performed, using regularisation parameter $T = 4$. The most defined class (194 k particles) was globally 3D auto-refined, iteratively aberration-corrected and Bayesian-polished (only the first 23 frames were used). The resulting global refinement was then subjected to core-focused 3D classification, with angular assignment using regularisation parameter $T = 8$. The most defined class (84.5 k particles) was CORE-focused 3D auto-refined.

In the second branch of 3D classification, the focus was on the structure of the ENDO-like species. A specific mask containing the ENDO area was created (cyan, Supplementary Fig. 27) and ENDO-focused 3D classification without angular assignment was performed, using regularisation parameter $T = 4$. The most defined class (159 k particles) was globally 3D auto-refined, iteratively aberration-corrected and Bayesian-polished (only the first 26 frames were used). The resulting global refinement was then subjected to ENDO-focused (cyan mask, Supplementary Fig. 28) 3D classification, without angular assignment using regularisation parameter $T = 12$. The most defined class (40.2 k particles) was globally 3D auto-refined.

3D analysis of the DISTAL-PROMOTER and MID-LINK structures. Two times binned particles (2,081 k) were subjected to two rounds of three-dimensional classifications with image alignment (Supplementary Fig. 31). The first round of 3D classification was restricted to 12 classes and performed using 60 \AA low-pass filtered APO-ENDO structure as the initial model. Particles in classes with poor structural features were removed. The second classification (into ten classes) was done during two rounds of 25 iterations each, using the regularisation parameter $T = 4$. In the second round, local angular searches were performed at 3.5° to clearly separate structural species. Three major 3D species were identified: DISTAL-PROMOTER-like, the MID-LINK-like and ENDO-like. The ENDO-like particles were processed together with CSSB DATA 1 (Supplementary Fig. 25).

The DISTAL-PROMOTER-like and MID-LINK-like species were pooled (65 k particles) and globally 3D auto-refined. In the first branch of 3D classification, the focus was on the structure of the DISTAL-PROMOTER-like specie. A specific mask containing the DISTAL-PROMOTER area was created (pink, Supplementary Fig. 31) and DISTAL-PROMOTER-focused 3D classification without angular assignment was performed, using regularisation parameter $T = 4$. The most

defined class (23.7 k particles) was further DISTAL-PROMOTER-CORE-focused (purple mask, Supplementary Fig. 31) 3D auto-refined as described for the CORE-like specie.

In the second branch of 3D classification, the focus was on the structure of the MID-LINK-like specie. A specific mask containing the MID-LINK area was created (orange, Supplementary Fig. 31) and MID-LINK-focused 3D classification without angular assignment was performed, using regularisation parameter $T = 4$. The most defined class (40.7 k particles) was further MID-LINK-CORE-focused (yellow mask, Supplementary Fig. 31) 3D auto-refined as described for the CORE-like species.

3D analysis of the PRE-INITIATION structure. Particles (2,470 k) were subjected to two rounds of reference-free 2D classification. Particles in classes with secondary structure features were selected (1016 k particles) and used for an ab initio volume reconstruction and then 3D refined using the latter ab initio 60 \AA low-pass filtered volume reconstruction as the initial model. The particles were astigmatism-corrected with CTFrefine within RELION. The particles were then subjected to a 3D classification restricted to twelve classes using $T = 4$ and 7.5° sampling for 25 iterations and then 3.5° sampling for an additional ten iterations. Classes with comparable structural features were combined (319 k particles), 3D refined, aberration-corrected and Bayesian-polished then 3D refined again. The refined particles were then subjected to further 3D classification without image alignment and particles from the most defined class (119 k particles) were used for final 3D auto-refinement (Supplementary Figs. 28 and 29).

3D analysis of the ELONGATION structure. Particles (2452k) were binned four times and subjected to one round of reference-free 2D classification. Particles in classes with secondary structure features were selected (579 k particles) and subjected to 3D classification restricted to ten classes using $T = 4$ and 7.5° sampling for 25 iterations with the 60 \AA low-pass filtered DISTAL-PROMOTER volume as a reference. Classes with comparable structural features were combined (122 k particles) and 3D refined to 3.5 \AA resolution. The 20 \AA low-pass filtered refined volume was then used as a reference for the 3D classification of all extracted particles (2452k) restricted to eight classes using $T = 4$ and 7.5° sampling for 35 iterations. The most defined class (426 k particles) was selected for 3D refinement and then subjected to further 3D classification without image alignment restricted to six classes. The most defined class (79 k particles) was 3D refined, aberration-corrected and Bayesian-polished then finally 3D refined using SIDESPLITTER⁵³ (Supplementary Figs. 32 and 33).

Final steps. All final cryo-EM density maps were generated by the post-processing feature in RELION and sharpened or blurred into MTZ format using CCP-EM⁵⁴. The resolutions of the cryo-EM density maps were estimated at the 0.143 gold standard Fourier Shell Correlation (FSC) cut-off (Supplementary Table 2). A local resolution (Supplementary Table 2) was calculated using RELION and reference-based local amplitude scaling was performed by LocScale within CCP-EM⁵⁵.

Model building. The APO-CORE structure was constructed de novo with iterative rounds of model building with Coot⁵⁶ and real-space refinement with Phenix⁵⁷. Subsequent structures used this as a basis for the further model extension. Secondary structure prediction using JPRED⁵⁸ based on multiple sequence alignment of both New World and Old World arenaviruses (Supplementary Alignment file) was particularly helpful in guiding model building. Considerable care was taken to cross-check between structures for consistency of sequence assignment and to ascertain connectivity. This also enabled better completion of models in lower resolution maps by transfer of structural elements that could be more accurately modelled in a higher resolution map. A homology model based on the MACV pendant domain (PDB: 6KLD), rebuilt to correct for sequence misalignments (using the original map, EMD-0707), was initially used to help build the LASV pendant domain. A 3D structure prediction using AlphaFold⁵⁹ via the server <https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb> resulted in an improved model for the pendant domain, which was used for model building. Unexpectedly, despite apparent sequence homologies, the LASV α -bundle 827-VVFNK...IIDQY-925 has a completely different arrangement of helices (topologically impossible to align in 3D) than that of the equivalent region in MACV L 820-VVIPK...QVALA-917 (Supplementary Fig. 6), both structures being confirmed by good quality maps. This might partly explain why in the LASV L structure previously published⁹ (PDB:6KLC), based on a lower resolution 3.9 \AA map, the α -bundle is built in the reverse direction.

Buried surface areas were determined using the Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)⁶⁰. An overview of the Segment-based Manders' Overlap Coefficient (SMOC) scores⁶¹ for each of the structures is provided in Supplementary Fig. 34. Structure presentation was done using PyMOL (Schrodinger) and UCSF ChimeraX⁶².

Electrophoretic mobility shift assay. 3' RNA 1–10 nt (Supplementary Table 1) was chemically synthesised with a fluorophore at the 5' end (5' Cyanine3) (Biomers). Reactions containing 0–1 μM L protein and 0.2 μM labelled single-stranded RNA

were set up in binding buffer (100 mM HEPES(NaOH), pH 7.0, 100 mM NaCl, 50 mM KCl, 2 mM MnCl₂, 2 mM dithiothreitol, 0.1 µg/µL Poly(C) RNA (Sigma), 0.1 µg/µL bovine serum albumin and 0.5 U/µL RNasin (Promega)) and incubated on ice for 30 min. RNA-bound protein complexes were separated from unbound RNA by native gel electrophoresis at 4 °C, using 5% polyacrylamide Tris-glycine gels. Fluorescence signals were detected in the gel with the VILBER LOURMAT FUSION SL4 using the Starlight Module with an excitation wavelength of 523 and a 609 nm emission filter.

Endonuclease assay. An RNA 16-mer was chemically synthesised with either 5' Cap (TriLink BioTechnologies) or 5' Triphosphate (Chemgenes). For labelling pCp-Cy5 (Cytidine-5'-phosphate-3'-(6-aminoethyl)phosphate (Jena bioscience), was ligated to the 3' end of the 16 nt RNA using T4 RNA ligase (Thermo Scientific). The resulting labelled 17 nt RNA substrates (Supplementary Table 1) were separated from excess pCp-Cy5 by denaturing PAGE (7 M urea, 25% acrylamide 0.5-fold Tris-borate-EDTA). The clearly blue coloured product bands were excised from the gel. The gel pieces were grounded and the RNA was extracted two times with Tris-borate buffer. The pure labelled RNA was precipitated with 90% Ethanol from the supernatant after the addition of Ammonium acetate (2.5 M), washed two times with 90% Ethanol and dissolved in DEPC treated H₂O. Reactions containing 0.5 µM L protein were incubated, sequentially with 2.5 pmol of either single-stranded 5' promoter RNA, 3' promoter RNA (Supplementary Table 1) or both, on ice for 15 min in 5 µL assay buffer (100 mM HEPES(NaOH) pH 7.0, 100 mM NaCl, 50 mM KCl, 2 mM MnCl₂, 0.5 U/µL RNasin (Promega), 2 mM dithiothreitol, and 0.1 µg/µL bovine serum albumin). After the addition of ~0.3 µM labelled RNA the mix was incubated at 37 °C for 120 min. The reaction was stopped by adding an equivalent volume of RNA loading buffer (98% formamide, 18 mM EDTA, 0.025 mM SDS) and heating the samples at 95 °C for 5 min. Products were separated by denaturing PAGE on 7 M Urea, 25% polyacrylamide Tris-borate-EDTA (0.5-fold) gels and 0.5-fold Tris-borate buffer. Fluorescence signals were detected in the gel with the VILBER LOURMAT FUSION SL4 using the Starlight Module with an excitation wavelength of 624 nm and a 695 nm emission filter. Uncropped images are provided as Source Data.

Polymerase assay

Standard polymerase assay. If not indicated otherwise, 0.5 µM L protein was incubated sequentially with 1 µM of single-stranded 5' promoter RNA (Supplementary Table 1) and 1 µM single-stranded 3' promoter RNA (Supplementary Table 1) in assay buffer (100 mM HEPES(NaOH) pH 7.0, 50 mM NaCl, 50 mM KCl, 2 mM MnCl₂, 0.5 U/µL RNasin (Promega), 2 mM dithiothreitol) on ice for 15 min. The reaction was started by the addition of NTPs (0.25 mM UTP/ATP/CTP and 0.125 mM GTP supplemented with 166 nM, 5 µCi [α³²P-GTP) in a final reaction volume of 10 µL. After incubation at 30 °C for 2 h the reaction was stopped by adding an equivalent volume of RNA loading buffer (98% formamide, 18 mM EDTA, 0.025 mM SDS, xylene cyanol and bromophenol blue) and heating the sample at 95 °C for 5 min. Products were separated by native gel electrophoresis using 25% polyacrylamide 0.5-fold Tris-borate-EDTA gels and 0.5-fold Tris-borate running buffer. Signals were visualised by phosphor screen autoradiography using a Typhoon FLA-7000 phosphorimager (Fujifilm) and the respective FLA-7000 software. Uncropped images are provided as Source Data.

Primer-dependent polymerase assay. Primer GCG, C1, S1 and C8 (Supplementary Table 1) were chemically synthesised with 5'-hydroxy ends (Biomers), primer C8ppp (Supplementary Table 1) with 5' Triphosphate modification (Chemgenes). An N⁷-MeGppp (Cap0) was introduced at the 5' terminus of C8ppp using the ScriptCap m7G Capping System (CELLSCRIPT) with 1 nmol C8ppp oligo using the manufacturer's standard protocol. After the addition of Ammonium acetate (2.5 M) the capped RNA was precipitated with Ethanol (90%), washed two times with Ethanol (90%), dried and dissolved in DEPC treated H₂O. For primer-dependent reactions, 10 µM of the respective primer was added to LASV L bound to promoter RNA and the mix was again incubated on ice for 15 min. The reaction was started by the addition of NTPs (0.25 mM UTP/ATP/CTP and 0.125 mM GTP supplemented with 166 nM, 5 µCi [α³²P-GTP) in a final reaction volume of 10 µL.

LASV mini-replicon system. The experiments were performed in the context of the T7 RNA polymerase-based LASV mini-replicon system^{15,24,36}. L genes were amplified using mutagenic PCR from a pCITE2a-L template to either yield wild-type or mutated L gene expression cassettes. L gene PCR products were further gel-purified when additional nonspecific bands were visible in agarose gels and quantified spectrophotometrically. BSR-T7/5 cells stably expressing T7 RNA polymerase⁶⁵ were transfected per well of a 24-well plate with 250 ng of minigenome PCR product expressing Renilla luciferase (Ren-Luc), 250 ng of L gene PCR product, 250 ng of pCITE-NP expressing NP, and 10 ng of pCITE-FF-luc expressing firefly luciferase as an internal transfection control. At 24 h post transfection, either total cellular RNA was extracted for Northern blotting using an RNeasy Mini kit (Qiagen) or cells were lysed in 100 µL of passive lysis buffer (Promega) per well, and firefly luciferase and Ren-Luc activity were quantified using the dual-luciferase reporter assay system (Promega). Ren-Luc levels were corrected with the firefly luciferase levels (resulting in standardised relative light units [sRLU]) to compensate for differences in transfection efficiency or cell

density. Data were evaluated using Prism 7.0d and are always presented as the mean of a given amount (n) of biological replicates as well as the respective standard deviation (SD).

For Northern blot analysis, 750 ng of total cellular RNA was separated in a 1.5% agarose-formaldehyde gel and transferred onto a Roti[®]-Nylon plus membrane (pore size 0.45 µm, Carl Roth). After UV cross-linking and methylene blue staining to visualise 28 S rRNA the blots were hybridised with a ³²P-labelled riboprobe targeting the Ren-Luc gene. Transcripts of Ren-Luc genes and complementary replication intermediate RNA of the minigenome were visualised by autoradiography using a Typhoon FLA-7000 phosphorimager (Fujifilm) and the respective FLA-7000 software. Quantification of signals for antigenomic RNA and mRNA was done in ImageJ 2.0.0-rc-43/1.50e (Source file). To verify the expression of L protein mutants in BSR-T7/5 cells, the cells were transfected with 500 ng of PCR product expressing C-terminally 3xFLAG-tagged L protein mutants per well in a 24-well plate. Cells were additionally infected with modified vaccinia virus Ankara expressing T7 RNA polymerase (MVA-T7)⁶⁴ to boost the expression levels and thus facilitate detection by immunoblotting. After cell lysis and electrophoretic separation in a 3–8% Tris-acetate polyacrylamide gel, proteins were transferred to a nitrocellulose membrane (GE Healthcare). FLAG-tagged L protein mutants were detected using peroxidase-conjugated anti-FLAG M2 antibody (1:10,000) (A8592; Sigma-Aldrich) and bands were visualised by chemiluminescence using Super Signal West Femto substrate (Thermo Scientific) and a FUSION SL image acquisition system (Vilber Lourmat). All uncropped images are provided as Source Data.

Statistics and reproducibility. Unless stated otherwise in the figure legends all in vitro assays (electrophoretic mobility shift assay, polymerase assay, endonuclease assay), as well as Coomassie-stained SDS gels, have been repeated with similar results at least twice. The total number of micrographs included in each cryo-EM dataset is given in Supplementary Table 2, representative micrographs are shown in Supplementary Fig. 24, 26, 28, 30 and 32.

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

Data availability

Coordinates and structure factors or maps generated in this study have been deposited in the www.PDB or EMDB: Apo-structure of Lassa virus L protein (well-resolved polymerase core) [APO-CORE] EMD-12807, PDB ID 7OCH Apo-structure of Lassa virus L protein (well-resolved endonuclease) [APO-ENDO] EMD-12860, PDB ID 7OE3 Apo-structure of Lassa virus L protein (well-resolved α-ribbon) [APO-RIBBON] EMD-12953, PDB ID 7OE7 Lassa virus L protein bound to 3' promoter RNA (well-resolved polymerase core and 3' RNA secondary binding site) [3END-CORE] EMD-12862, PDB ID 7OEA Lassa virus L protein bound to 3' promoter RNA (well-resolved endonuclease) [3END-ENDO] EMD-12863, PDB ID 7OEB Lassa virus L protein in a pre-initiation conformation [PRE-INITIATION] EMD-12955, PDB ID 7OJL Lassa virus L protein with endonuclease and C-terminal domains in close proximity [MID-LINK] EMD-12861, PDB ID 7OJJ Lassa virus L protein bound to the distal promoter duplex [DISTAL-PROMOTER] EMD-12954, PDB ID 7OJK Lassa virus L protein in an elongation conformation [ELONGATION] EMD-12956, PDB ID 7OJN All other data supporting the findings of this study are available within the article, its supplementary information and source data files. Additional information, relevant data and unique biological materials will be available from the corresponding author upon reasonable request. Source data are provided with this paper.

Received: 14 June 2021; Accepted: 8 November 2021;

Published online: 02 December 2021

References

- Alkpede, G. O. et al. Caseload and case fatality of Lassa fever in Nigeria, 2001–2018: a specialist center's experience and its implications. *Front. Public Health* **7**, 170 (2019).
- Basinski, A. J. et al. Bridging the gap: using reservoir ecology and human serosurveys to estimate Lassa virus spillover in West Africa. *PLoS Comput. Biol.* **17**, e1008811 (2021).
- Olschewski, S., Cusack, S. & Rosenthal, M. The cap-snatching mechanism of Bunyaviruses. *Trends Microbiol.* **28**, 293–303 (2020).
- Raju, R. et al. Nontemplated bases at the 5' ends of Tacaribe virus mRNAs. *Virology* **174**, 53–59 (1990).
- Polyak, S. J., Zheng, S. & Harnish, D. G. 5' termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. *J. Virol.* **69**, 3211–3215 (1995).
- Meyer, B. J. & Southern, P. J. Concurrent sequence analysis of 5' and 3' RNA termini by intramolecular circularization reveals 5' nontemplated bases and 3'

- terminal heterogeneity for lymphocytic choriomeningitis virus mRNAs. *J. Virol.* **67**, 2621–2627 (1993).
7. Vogel, D., Rosenthal, M., Gogrefe, N., Reindl, S. & Gunther, S. Biochemical characterization of the Lassa virus L protein. *J. Biol. Chem.* **294**, 8088–8100 (2019).
 8. Kranzusch, P. J. et al. Assembly of a functional Machupo virus polymerase complex. *Proc. Natl Acad. Sci. USA* **107**, 20069–20074 (2010).
 9. Peng, R. et al. Structural insight into arenavirus replication machinery. *Nature* **579**, 615–619 (2020).
 10. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360 (2014).
 11. Reich, S. et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361–366 (2014).
 12. Hengrung, N. et al. Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* **527**, 114–117 (2015).
 13. 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).
 14. Arragain, B. et al. Pre-initiation and elongation structures of full-length Lassa virus polymerase reveal functionally important conformational changes. *Nat. Commun.* **11**, 3590 (2020).
 15. Hass, M., Goltz, U., Muller, S., Becker-Ziaja, B. & Gunther, S. Replicon system for Lassa virus. *J. Virol.* **78**, 13793–13803 (2004).
 16. Fan, H. et al. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature* **573**, 287–290 (2019).
 17. Peng, Q. et al. Structural insight into RNA synthesis by influenza D polymerase. *Nat. Microbiol.* **4**, 1750–1759 (2019).
 18. Wandzik, J. M. et al. A structure-based model for the complete transcription cycle of influenza polymerase. *Cell* **181**, 877–893 e821 (2020).
 19. Thierry, E. et al. Influenza polymerase can adopt an alternative configuration involving a radical repacking of PB2 domains. *Mol. Cell* **61**, 125–137 (2016).
 20. Wallat, G. D. et al. High-resolution structure of the N-terminal endonuclease domain of the lassa virus L polymerase in complex with magnesium ions. *PLoS ONE* **9**, e87577 (2014).
 21. Reguera, J. et al. Comparative structural and functional analysis of Bunyavirus and Arenavirus cap-snatching endonucleases. *PLoS Pathog.* **12**, e1005636 (2016).
 22. Kouba, T., Drcovca, P. & Cusack, S. Structural snapshots of actively transcribing influenza polymerase. *Nat. Struct. Mol. Biol.* **26**, 460–470 (2019).
 23. Rosenthal, M. et al. Structural insights into repletavirus cap-snatching machinery. *PLoS Pathog.* **13**, e1006400 (2017).
 24. Lehmann, M. et al. Role of the C terminus of Lassa virus L protein in viral mRNA synthesis. *J. Virol.* **88**, 8713–8717 (2014).
 25. Gogrefe, N., Reindl, S., Gunther, S. & Rosenthal, M. Structure of a functional cap-binding domain in Rift Valley fever virus L protein. *PLoS Pathog.* **15**, e1007829 (2019).
 26. Guilligay, D. et al. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* **15**, 500–506 (2008).
 27. Vogel, D. et al. Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L protein. *Nucleic Acids Res.* **48**, 5749–5765 (2020).
 28. Saez-Ayala, M. et al. Crystal structures of Lymphocytic choriomeningitis virus endonuclease domain complexed with diketo-acid ligands. *IUCr*. **5**, 223–235 (2018).
 29. Heinemann, P., Schmidt-Chanasit, J. & Gunther, S. The N terminus of Andes virus L protein suppresses mRNA and protein expression in mammalian cells. *J. Virol.* **87**, 6975–6985 (2013).
 30. Lelke, M., Brunotte, L., Busch, C. & Gunther, S. An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome. *J. Virol.* **84**, 1934–1944 (2010).
 31. Morin, B. et al. The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. *PLoS Pathog.* **6**, e1001038 (2010).
 32. Auperin, D. D., Romanowski, V., Galinski, M. & Bishop, D. H. Sequencing studies of pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. *J. Virol.* **52**, 897–904 (1984).
 33. Auperin, D. D. & McCormick, J. B. Nucleotide sequence of the Lassa virus (Josiah strain) S genome RNA and amino acid sequence comparison of the N and GPC proteins to other arenaviruses. *Virology* **168**, 421–425 (1989).
 34. Garcin, D. & Kolakofsky, D. Tacaribe arenavirus RNA synthesis in vitro is primer dependent and suggests an unusual model for the initiation of genome replication. *J. Virol.* **66**, 1370–1376 (1992).
 35. Hass, M. et al. Mutational analysis of the lassa virus promoter. *J. Virol.* **80**, 12414–12419 (2006).
 36. Hass, M., Lelke, M., Busch, C., Becker-Ziaja, B. & Gunther, S. Mutational evidence for a structural model of the Lassa virus RNA polymerase domain and identification of two residues, Gly1394 and Asp1395, that are critical for transcription but not replication of the genome. *J. Virol.* **82**, 10207–10217 (2008).
 37. Pyle, J. D. & Whelan, S. P. J. RNA ligands activate the Machupo virus polymerase and guide promoter usage. *Proc. Natl Acad. Sci. USA* **116**, 10518–10524 (2019).
 38. Iapalucci, S., Lopez, N. & Franze-Fernandez, M. T. The 3' end termini of the Tacaribe arenavirus subgenomic RNAs. *Virology* **182**, 269–278 (1991).
 39. Pinschewer, D. D., Perez, M. & de la Torre, J. C. Dual role of the lymphocytic choriomeningitis virus intergenic region in transcription termination and virus propagation. *J. Virol.* **79**, 4519–4526 (2005).
 40. Lopez, N. & Franze-Fernandez, M. T. A single stem-loop structure in Tacaribe arenavirus intergenic region is essential for transcription termination but is not required for a correct initiation of transcription and replication. *Virus Res.* **124**, 237–244 (2007).
 41. Bergeron, E. et al. Reverse genetics recovery of Lujo virus and role of virus RNA secondary structures in efficient virus growth. *J. Virol.* **86**, 10759–10765 (2012).
 42. Guilligay, D. et al. Comparative structural and functional analysis of orthomyxovirus polymerase cap-snatching domains. *PLoS ONE* **9**, e84973 (2014).
 43. Xu, X. et al. Cryo-EM structures of Lassa and Machupo virus polymerases complexed with cognate regulatory Z proteins identify targets for antivirals. *Nat. Microbiol.* **6**, 921–931 (2021).
 44. Kang, H. et al. Structural basis for recognition and regulation of arenavirus polymerase L by Z protein. *Nat. Commun.* **12**, 4134 (2021).
 45. Cornu, T. I. & de la Torre, J. C. RING finger Z protein of lymphocytic choriomeningitis virus (LCMV) inhibits transcription and RNA replication of an LCMV S-segment minigenome. *J. Virol.* **75**, 9415–9426 (2001).
 46. Kranzusch, P. J. & Whelan, S. P. Arenavirus Z protein controls viral RNA synthesis by locking a polymerase-promoter complex. *Proc. Natl Acad. Sci. USA* **108**, 19743–19748 (2011).
 47. Fitzgerald, D. J. et al. Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3**, 1021–1032 (2006).
 48. Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem. Sci.* **37**, 49–57 (2012).
 49. Kandiah, E. et al. CM01: a facility for cryo-electron microscopy at the European Synchrotron. *Acta Crystallogr. D. Struct. Biol.* **75**, 528–535 (2019).
 50. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
 51. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
 52. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCr*. **7**, 253–267 (2020).
 53. Ramlau, K., Palmer, C. M., Nakane, T. & Aylett, C. H. S. Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER. *J. Struct. Biol.* **211**, 107545 (2020).
 54. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. D. Struct. Biol.* **73**, 469–477 (2017).
 55. Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *Elife* <https://doi.org/10.7554/eLife.27131> (2017).
 56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
 57. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D. Struct. Biol.* **74**, 531–544 (2018).
 58. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
 59. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* <https://doi.org/10.1038/s41586-021-03819-2> (2021).
 60. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
 61. Joseph, A. P., Lagerstedt, I., Patwardhan, A., Topf, M. & Winn, M. Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. *J. Struct. Biol.* **199**, 12–26 (2017).
 62. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
 63. Buchholz, U. J., Finke, S. & Conzelmann, K. K. Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. *J. Virol.* **73**, 251–259 (1999).
 64. Sutter, G., Ohlmann, M. & Erfle, V. Non-replicating vaccinia vector efficiently expresses bacteriophage T7 RNA polymerase. *FEBS Lett.* **371**, 9–12 (1995).

Acknowledgements

We acknowledge the European Synchrotron Radiation Facility for the provision of beam time on CM01 and we would like to thank Daouda A.K. Traore for assistance. Furthermore, we want to thank Sophia Reindl and Nadja Hüttmann for advice and technical

support in the early phases of this project; Michael Hons, Wojtek Galej and Erika Pellegrini for access to the Glacios at EMBL Grenoble; Carolin Seuring and Cornelia Cazez for access to Cryo-EM facility at CSSB; Aymeric Peuch and Wolfgang Lugmayr for help with using the joint EMBL-IBS and the CSSB partition on the DESY computer cluster. We acknowledge funding of this project by the Leibniz Association's Leibniz competition programme (grant K72/2017 to S.G., K.G. and S.C.), the Federal Ministry of Education and Research of Germany (grant 01KI2019 to M.R.), the Wilhelm und Maria Kirmser-Stiftung, the Alexander von Humboldt foundation (FRA 1200789 HFST-P to E.R.J.Q.) as well as the EMBL Interdisciplinary Postdocs (EI3POD) initiative co-funded by Marie Skłodowska-Curie (grant 664726 to T.K.). Part of this work was performed at the Cryo-EM Facility at CSSB, supported by the UHH and DFG (grants INST 152/772-1 and 774-1 to K.G.).

Author contributions

T.K., D.V., E.R.J.Q., S.G., K.G., M.R. and S.C. conceived and supervised the project. C.B. and M.M. carried out cloning. D.V. and C.B. expressed and purified the proteins. T.K., D.V. and S.R.T. prepared the cryo-EM grids. T.K., S.R.T. and E.R.J.Q. collected and processed the cryo-EM data, D.V., H.M.W., M.R. and S.C. built and validated the models; D.V. and H.M.W. performed the in vitro experiments, M.M. performed the cell-based mini-genome experiments, T.K., D.V., H.M.W. and M.R. compiled the figures, M.R. and S.C. wrote the manuscript with input from all co-authors. The authors wish it to be known that M.R. and S.C., in addition to the joint supervision statement, have equally contributed to this manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27305-5>.

Correspondence and requests for materials should be addressed to Maria Rosenthal or Stephen Cusack.

Peer review information *Nature Communications* thanks Rachel Fearn and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer 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) 2021

3.3 Structural insight into the viral genome replication of Severe fever with thrombocytopenia syndrome virus L-protein

Previous findings on L-protein RNA replication described the pre-initiation and elongation states. To better understand the early and late elongation states of the L-protein, we solved the structure of the SFTSV L-protein in five different states of replication, filling in gaps in the RNA synthesis cycle. The high-resolution reconstructions of the L-protein allowed for a detailed description of the protein-RNA interactions at crucial replication stages starting from a pre-initiation state to late elongation. The structural and biochemical results elucidate the complex mechanisms of L-protein RNA replication.

Original publication and supplement at <https://doi.org/10.1093/nar/gkac1249>

using a Typhoon FLA-7000 phosphorimager (Fujifilm) operated with the FLA-7000 software.

Rift valley fever virus mini-replicon system

The experiments were performed using the RVFV mini-replicon system described previously (36). L genes were amplified using mutagenic PCR from a pCITE2a-L template to produce either wild-type or mutated L gene expression cassettes. PCR products were gel purified and quantified spectrophotometrically. All mutations were confirmed by sequencing of the PCR products. BSR-T7 cells were transfected per well of a 24 well plate with 250 ng of L gene PCR product, 500 ng of pCITE expressing NP, 750 ng of pCITE expressing the mini-genome RNA encoding Renilla luciferase (Ren-Luc), and 10 ng of pCITE expressing the firefly luciferase as an internal control, all under control of a T7 promoter. For transfection we used Lipofectamine 2000 according to the manufacturer's instructions. At 24 h post-transfection, either total cellular RNA was extracted for Northern blot analysis using an RNeasy Mini kit (Qiagen) or cells were lysed in 100 μ l of passive lysis buffer (Promega) per well, and firefly luciferase and Ren-Luc activity quantified using the dual-luciferase reporter assay system (Promega). Ren-Luc levels were corrected with the firefly luciferase levels (resulting in standardized relative light units [sRLU]) to compensate for differences in transfection efficiency or cell density. Data were evaluated in Prism and are always presented as the mean of a given amount (n) of biological replicates as well as the respective standard deviation (SD). To confirm general expressibility of the L protein mutants in BSR-T7 cells, the cells were transfected with 500 ng L gene PCR product expressing C-terminally 3xFLAG-tagged L protein mutants per well. Cells were additionally infected with Modified Vaccinia virus Ankara expressing a T7 RNA polymerase (MVA-T7) to boost the expression levels and, thereby facilitate detection by immunoblotting. At 24 h post-transfection, cells were lysed in 50 μ l passive lysis buffer (Promega) per well. After cell lysis and separation in a 3–8% Tris-acetate polyacrylamide gel (Invitrogen), proteins were transferred to a nitrocellulose membrane (Cy-tiva). FLAG-tagged L protein mutants were detected using horseradish peroxidase-conjugated anti-FLAG M2 antibody (1:9000) (A8592; Sigma-Aldrich) and bands were visualized by chemiluminescence using Clarity Max Western ECL Substrate (Bio-Rad) and a FUSION SL image acquisition system (Vilber Lourmat).

Northern blot analysis

For the northern blot analysis, 600 ng of total cellular RNA was separated in a 1.5% agarose-formaldehyde gel and transferred onto a Roti-Nylon plus membrane (pore size 0.45 μ m, Carl Roth). After UV cross-linking and methylene blue staining to visualize 28 S rRNA, the blots were hybridized with a 32P-labelled riboprobe targeting the Ren-Luc gene. Transcripts of the Ren-Luc gene and complementary replication intermediate RNA of the minigenome were visualized by autoradiography using a Typhoon FLA-7000 phosphorimager (Fujifilm) operated with the FLA-7000 software. Quantification of signals for antigenomic RNA and mRNA was done in Fiji.

Sample preparation for cryo-EM 5' HOOK structure

The SFTSV L protein at a concentration of 3 μ M in assay buffer (100 mM HEPES, 50 mM KCl, 2.5 mM MgCl₂, and 2 mM DTT, pH 7) was mixed with single-stranded L 5' (1–20) cRNA in a 2-fold molar excess. The mixture was incubated on ice for 30 min and then centrifuged at 15 000 \times g for 10 min at 4°C. Aliquots of 3 μ l were applied to Quantifoil R 2/1 Au G200F4 grids, immediately blotted for 2 s, and plunge frozen into liquid ethane/propane cooled to liquid nitrogen temperature using a FEI Vitrobot Mark IV (4°C, 100% humidity, blotting force –10).

Early-elongation, early-elongation-endo and late-elongation structures

The SFTSV L protein at a concentration of 3 μ M in assay buffer (100 mM HEPES, 50 mM KCl, 2.5 mM MgCl₂, and 2 mM DTT, pH 7) was mixed sequentially with single-stranded L 5' (1–20) cRNA and single-stranded L 3' (1–20) cRNA modified with 6 additional A bases, in a 2-fold molar excess. After 30 min of incubation on ice, the reaction was initiated by addition of NTPs (0.63 mM ATP/CTP/GTP/nhUTP). After incubation at 30°C for 2 h, the sample was centrifuged at 15 000 \times g for 10 min at 4°C. Aliquots of 3 μ l were applied to Quantifoil R 2/1 Au G200F4 grids, immediately blotted for 2 s, and plunge frozen into liquid ethane/propane cooled to liquid nitrogen temperature using a FEI Vitrobot Mark IV (4°C, 100% humidity, blotting force –10).

Resting structure

The SFTSV L protein at a concentration of 3 μ M in assay buffer (100 mM HEPES, 50 mM KCl, 2.5 mM MgCl₂, and 2 mM DTT, pH 7) was mixed with single-stranded L 5' (1–20) cRNA and single-stranded L 3' (1–20) cRNA in a 2-fold molar excess. After 30 min of incubation on ice, the reaction was supplemented with a short 2 nt primer (AC), which was at a 1:1 molar ratio with the RNA species. After incubation at 30°C for 2 h, the sample was then centrifuged at 15 000 \times g for 10 min at 4°C. Aliquots of 3 μ l were applied to Quantifoil R 2/1 Au G200F4 grids, immediately blotted for 2 s, and plunge frozen into liquid ethane/propane cooled to liquid nitrogen temperature using a FEI Vitrobot Mark IV (4°C, 100% humidity, blotting force –10).

Cryo-EM single-particle analysis

Data collection. Grids were loaded into a 300-kV Titan Krios transmission electron microscope (Thermo Scientific) equipped with a K3 direct electron detector and a post-column GIF BioQuantum energy filter (Gatan). Micrograph movies were typically collected using the EPU software (Thermo Scientific) at a nominal magnification of 105 000 \times with a pixel size of 0.85 Å using a defocus range of –0.8 to –2.6 μ m, unless stated otherwise. The sample was exposed with a dose of 15 electrons/physical pixel per second (estimated over vacuum) for 2.5 s and thus, a total dose of 50 electrons/Å². Recorded movie frames were fractionated into 50 frames. A summary of the key data collection

statistics for each cryo-EM structure is provided in Supplementary Table S2.

Image processing

All movie frames were imported in RELION 3.1 (37) or RELION 4.0 (38) and motion corrected using its own implementation of MotionCor2 (39). Particles were picked automatically using the pre-trained BoxNet in Warp (40) and the calculation of contrast transfer function (CTF) estimation and correction was done in RELION with CTFIND4 (41). Picked particles were extracted and processed in RELION 3.1/4.0 with visual inspection done with UCSF Chimera (42). First, particles were subjected to 2D classification at $4\times$ binned pixel size typically as an initial check of the quality of the dataset. Then, all particles were used for 3D classification in an iterative manner using a 50 Å low pass filtered map from apo SFTSV L protein (17). Incomplete, low resolution or 3D classes containing damaged particles were excluded from further data analyses. The remaining particles were re-extracted with the original pixel size value for further 3D classification and refinement as described for the 5' hook structure (Supplementary Figures S1 and S2); early-elongation, early-elongation-endo and late-elongation structures (Supplementary Figures S3 and S4); and, resting structure (Supplementary Figures S5 and S6). All cryo-EM maps were CTF refined and Bayesian polished before final refinement in relion except the resting structure which, after 3D classification, was imported as a particle stack in cryoSPARC (43) where final refinement was done with non-uniform refinement (44).

All final cryo-EM density maps were generated by the post-processing feature in RELION and sharpened or blurred into MTZ format using CCP-EM (45) except the early-elongation-endo structure where the final map was sharpened using DeepEMhancer (46) and the resting structure, which was sharpened using the local sharpening function of cryoSPARC. The resolutions of the cryo-EM density maps were estimated at the 0.143 gold standard Fourier Shell Correlation (FSC) cut-off. A local resolution was calculated using relion 3.1/4.0 except the resting structure which was calculated in cryoSPARC and displayed with Chimera visualization tools. A summary of the key processing statistics for each cryo-EM structure is provided in Supplementary Table S2.

Model building, refinement and validation

The published corrected apo structure of the SFTSV L protein (30) was used as the starting point for the RNA-bound structures published in this study which then underwent iterative rounds of model building in Coot (47) and real-space refinement in Phenix (48). The comprehensive validation tool in Phenix was used to assess model quality (49,50). Coordinates and EM maps are deposited in the EMDB and PDB databases as indicated in Table 1. Figures have been generated using CCP4mg (51), PyMOL (Schrödinger), and UCSF ChimeraX (52). A summary of the key refinement and validation statistics for each cryo-EM structure is provided in Supplementary Table S2.

RESULTS

Overview

We have determined five structures of the SFTSV L protein in several biologically relevant states ranging from pre-initiation to late-stage elongation. An overview of all structures including their key features is provided in Table 1. First, we incubated the L protein with nt 1–20 of the L segment 5' cRNA (5' hook, 3.4 Å) (Supplementary Table S1; Supplementary Figures S1 and S2). This initial structure enabled us to identify the pocket in which the L protein coordinates the 5' RNA terminus in a so-called hook-like conformation, which is known to be important for activation of the polymerase function (17).

To stall the L protein in an active genome replication mode, we incubated the L protein with nts 1–20 of the L segment 5' cRNA and a modified 3' cRNA template with 6 additional A's (Supplementary Table S1). These were incubated together in the presence of ATP, CTP and GTP spiked with non-hydrolysable UTP (nhUTP) (Supplementary Figures S3 and S4). With this, we observed three elongation-state structures: two in which the L protein is stalled at position G4 of the modified 3' cRNA template (EARLY-ELONGATION, 2.6 Å and EARLY-ELONGATION-ENDO, 3.6 Å) and another in which the L protein is stalled at position A21 of the modified 3' cRNA template (LATE-ELONGATION, 2.6 Å). While the binding of viral RNAs in the two early-elongation structures appear similar with the distal 5' cRNA region forming a duplex with the 3' cRNA template, there is a key difference in the overall conformation of the L protein. In the EARLY-ELONGATION-ENDO structure, the endonuclease domain adopts an entirely different 'raised' conformation by comparison to the endonuclease domain in the early-elongation structure, which is broadly similar to that seen in the apo and late-elongation structures. In the latter, the CBD can also be fitted and the putative cap-binding site appears blocked by the arginine blocking motif previously described (30–33). Therefore, in total, we obtained 3 structures of the L protein stalled in active genome replication mode: two stalled in the early stages of elongation, and a final structure stalled in a late stage of elongation.

To identify the structural basis for replication priming, we incubated the L protein with nts 1–20 of the L segment 5' cRNA and 3' cRNA termini and added a short 2 nt (AC) primer (Supplementary Table S1; Supplementary Figures S5 and S6). Processing of data collected from this grid led to a single structure (RESTING, 3.2 Å), in which the 5' RNA hook binds in the same way as that seen in our other structures. However, rather than formation of the distal duplex and the 3' cRNA template entering the active site, we observe the 3' cRNA terminus binding to an alternative region, denoted the 3' secondary binding site in analogy to similar binding sites identified in LASV (28) and LACV L proteins (13) and also the influenza virus polymerase complex (53). The distal duplex is not formed in this structure and we observe a weak signal in this region of the map that likely corresponds to an average of the RNA strands in multiple different conformations. Nevertheless, this structure demonstrates that both the 3' and 5' RNA termini can be

Table 1. Overview of cryo-EM structures. An overview of the cryo-EM structures is provided with the reported resolution (Å), any RNA ligands present, PDB and EMD accession codes, as well as supporting comments highlighting the key features of each structure

Name	Res (Å)	RNA Ligands	PDB Code	EMDB Code	Comment
5' Hook	3.4	L 5' (1–20)	8AS6	EMD-15607	Only 5' cRNA is bound to the L protein in the hook-like motif. The L protein core is clear and motif F is unstructured.
Early-elongation	2.6	L 5' (1–20) L 3' + 6A (1–26)	8AS7	EMD-15608	The 5' and 3' cRNA form a distal duplex structure with the 3' cRNA shown feeding into the L protein core. Short 3 nt product in L protein core. Stalling nhUTP clearly bound at +1 position.
Early-elongation-endo	3.6	L 5' (1–20) L 3' + 6A (1–26)	8ASB	EMD-15610	The 5' and 3' cRNA form a distal duplex structure with the 3' cRNA shown feeding into the L protein core. There is also a short 3 nt-long product in L protein core. However, in this structure, we see the endonuclease in a raised conformation unlike that seen elsewhere. Stalling nhUTP clearly bound at +1 position.
Late-elongation	2.6	L 5' (1–20) L 3' + 6A (1–26)	8ASD	EMD-15614	The distal duplex structure is disrupted and there is a long 10-bp product-template duplex in the L protein core. Stalling nhUTP clearly bound at +1 position. The CBD can also be fit with the putative cap-mRNA binding site occupied by the Arginine blocking motif.
Resting	3.2	L 5' (1–20) L 3' (1–20) Primer AC (1–2)	8ASG	EMD-15615	The 5' cRNA is bound as in the 5' HOOK structure. Excess of 3' cRNA is bound to the 3' secondary binding site. A small number of 3' cRNA nts can be fit in the L protein core but the region of map for the AC primer is not resolved.

bound simultaneously to the hook-binding and secondary binding sites, respectively. We propose that this might represent a resting conformation after finishing one round of genome replication and just before another pre-initiation complex is formed with the distal duplex stabilizing the 3' RNA and guiding it into the RdRp active site for initiation. However, in this structure, we can also confidently fit 2 nts of the 3' cRNA in the active site, which indicates that the observed map represents a mixture of resting and pre-initiation states, which would explain the relative lack of definition in some regions of the map, making it challenging to interpret.

The L protein during pre-initiation of viral genome replication

The 5' HOOK structure demonstrates that the SFTSV L protein undergoes several conformational changes as a result of binding to the terminal 20 nts of the 5' cRNA in comparison to the apo structure (PDB: 6L42, 6Y6K and 7ALP) (30–33). The most apparent changes being an opening up of the vRNA binding lobe (vRBL), the neighbouring fingers domain, and the PA-C like core lobe, allowing the 5' RNA to bind in a positively charged cleft (Figure 1A). This cleft is covered by a short 3-turn α -helix (M437–H447) and supported on the periphery by a short loop (H1038–D1046) (Figure 1A). Of note, both of these structural motifs were missing in the published apo structures, suggesting here that the 5' RNA has a role in stabilizing this region of the L protein.

Structure of the 5' cRNA hook

We find the 5' cRNA bound in each of our L protein structures and, depending on the structure, we can fit 11–17 nts into the corresponding map region. From this data, we

found that the first 10 nts of the SFTSV 5' RNA form the characteristic hook-like motif (Figure 1B). In addition, in our structures, the 5' RNA hook only appears to be stabilized by one canonical base-base interaction between C2 and G8. The SFTSV hook structure hence appears to be closer to that of LASV, which is also stabilized by one canonical base-base interaction (28), in comparison to that of LACV, which is stabilized by two canonical base-base interactions (14) (Supplementary Figure S7). In addition to the C2/G8 base-pairing interaction, amino acid sidechains from the fingers, finger node, and the core lobe contribute towards 5' RNA hook coordination.

For example, the K656 sidechain is shown to interact with the backbone phosphates of nts C2 and A3 (Figure 1B). The N765 sidechain also interacts with nt A3, with the sidechain amino group coordinating one of the A3 base nitrogens, while the carboxyl group interacts with the A3 ribose O2' hydroxyl. There are two sidechain interactions stabilizing nt C4: R1049 and H447. These coordinate the C4 ribose O2' hydroxyl and C4 base oxygen, respectively. The R695 sidechain also stacks against the C4 base providing additional stability. The A5 base is sandwiched between the R318 and P693 sidechains. Similarly, there are not many interactions visible stabilizing the G6 base, only the H447 sidechain coordinating the G6 ribose oxygen (Supplementary Figure S8A). In addition to this interaction, in our early-elongation structure, we noticed an interesting ring-shaped blob of density between the G6 base and the adjacent R1043 sidechain (Supplementary Figure S8B). After exploring several different options, including individual NTPs bound non-specifically at this position, we concluded that the best fit was a single HEPES molecule, which is the buffer used during the L protein/cRNA incubation step. Here, the HEPES ring is shown sandwiched in between the G6 base and the R1043 sidechain (Supplementary

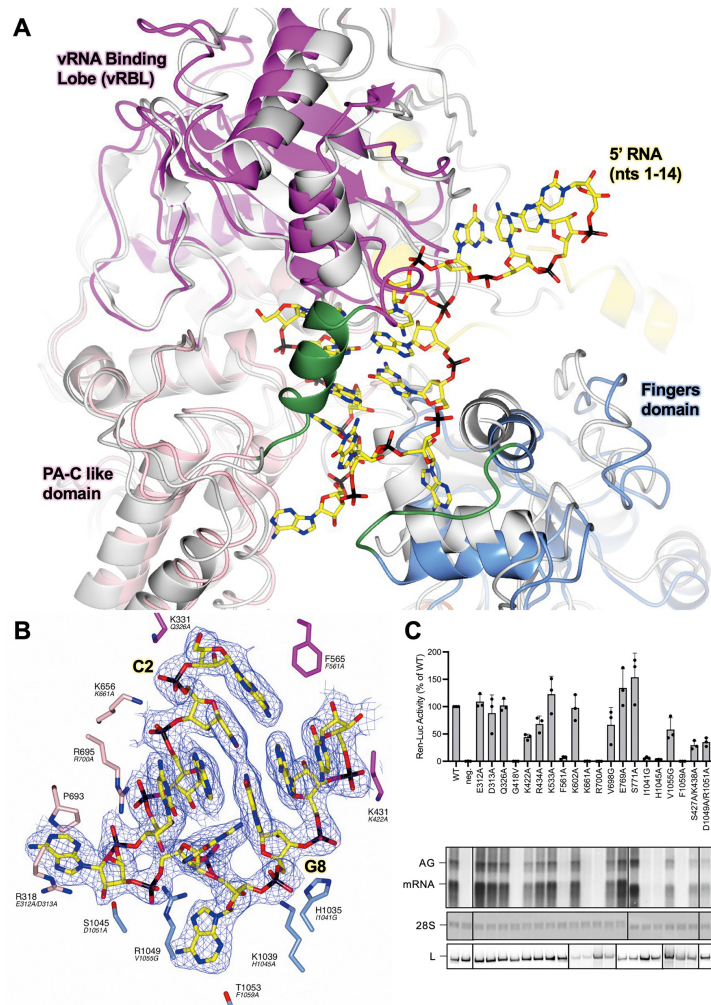


Figure 1. Conformational changes following the binding of the 5' cRNA hook. (A) The SFTSV L protein (5' HOOK) is shown with the published apo structure of the SFTSV L protein (PDB: 7ALP; backbone coloured grey) superposed by SSM in CCP4mg (51). A short 3-turn α -helix (M437–H447) and loop (H1038–D1046), which are both stabilized by the binding of the 5' cRNA nts, are shown in green. (B) A close-up of the SFTSV L protein hook-binding site is shown with the early-elongation map overlaid. Bound 5' cRNA is shown and coloured yellow with the exception of C2 and G8, which together form the only cognate base-base interaction and are labelled. Residues from the SFTSV L protein are labelled and coloured according to the assigned domain (pink for the PA-C like-domain, blue for the fingers, and magenta for the vRBL) with the corresponding RVFV L protein residue shown in black italicized text underneath. Not all hook-interacting residues are shown for clarity. (C) RVFV mini-replicon data for L protein with mutations in the proposed 5' hook binding site presenting luciferase reporter activity (in standardized relative light units relative to the wild-type L protein (WT)). Data were presented as mean values \pm SD of three biological replicates ($n = 3$). All biological replicates are shown as black dots (top panel). Middle panels present Northern blotting results with signals for antigenomic viral RNA (AG, equal to cRNA), viral mRNA (mRNA) and 28 S ribosomal RNA (28 S) as a loading control, and the bottom panel shows Western blot detection of FLAG-tagged L proteins (L) to demonstrate general expressibility of the mutants. Uncropped original blots/gels are provided in the Supplementary Data.

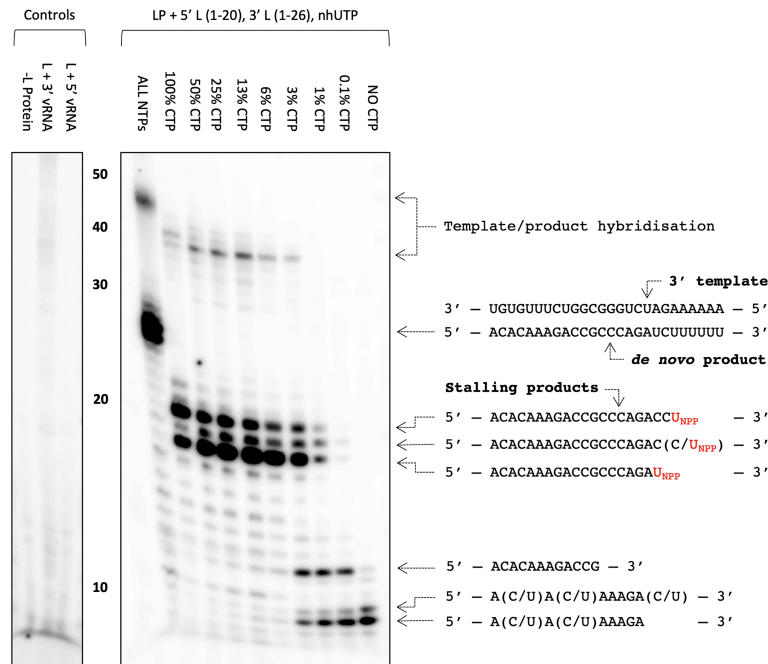


Figure 2. Stalling of the SFTSV L protein and misincorporation of CTP for UTP. The effect of removing CTP from the NTP mix in the presence of nhUTP (U_{NTP}) on the polymerase activity of purified SFTSV L protein (LP) was tested *in vitro*. The reactions were carried out with the L 5' cRNA (nts 1–20) and L 3' cRNA (nts 1–26) present under standard polymerase assay conditions (see Materials and Methods), with the exception of the NTPs used in the CTP titration conditions, which were diluted 1:10 resulting in a final concentration of 20 μ M ATP, UTP and 10 μ M GTP spiked with 166 nM, 5 μ Ci[α]³²P-GTP. The final concentration of CTP in the NTP mix varied from 20 to 0 μ M. Products were separated by denaturing gel electrophoresis and visualized by autoradiography. Uncropped original blots/gels are provided in the Supplementary Data.

the L protein could presumably also misincorporate nhUTP leading to the observed early stalling product.

To test this hypothesis, we selected the SFTSV S_{9U} segment vRNA as this genome end has a naturally occurring poly-G sequence. If the L protein can incorporate UTP in place of CTP, then in the absence of CTP, the L protein should be able to read-through the poly-G sequence (nts 11–15) by incorporating UTP. If not, then we should observe a stalling product around 11 nts. We additionally used a 16 nt primer (Primer B-OH, Supplementary Table S1) to increase any stalling product size by 13 nts, accounting for the 3 bp overlap with the template. In our *in vitro* polymerase assay setup, when we remove CTP from the NTP mix but retain UTP, we still see the full-length product (Supplementary Figure S10). When we then also titrate UTP out of the NTP mix, we see the emergence of a product at the approximate size of the expected primed stalling product (Supplementary Figure S10). Together, this strongly suggests that the L protein can also incorporate UTP in place of CTP supporting our hypothesis that the early-stalling structures we observed are likely the result of nhUTP being misincorporated in place of CTP. To conclude, we obtained 3 structures of the SFTSV L protein stalled in actively replicating

states: two structures stalled at position G4 of the 3' cRNA (EARLY-ELONGATION and EARLY-ELONGATION-ENDO), and one structure stalled at position U21 of the 3' cRNA (LATE-ELONGATION), which we refer to as early- and late-stage elongation states, respectively.

The L protein early elongation structures

The two early elongation structures, denoted EARLY-ELONGATION and EARLY-ELONGATION-ENDO, demonstrate that when both 5' and 3' 1–20 nts cRNAs are present, the SFTSV L protein undergoes further remodelling, mostly involving residues in the thumb and thumb ring domains. These structural changes allow the L protein to accommodate the 5'/3' distal duplex region, which is shown to curl around the vRBL, adopting a similar conformation to the full promoters observed for LACV (14) and LASV (28) (Figure 3A). However, there is a key difference in the number of canonical base-pair interactions between the 5' and 3' RNA strands. Unlike in other related bunyaviruses, the conserved 5' and 3' RNA termini of the SFTSV L segment are not particularly complementary with theoretically only 11 matches within

the first 20 nts, including only 2 exact matches (nts 13 and 14) in the distal duplex region (nts 11–18) (Supplementary Figure S11). However, in the EARLY-ELONGATION and EARLY-ELONGATION-ENDO structures, we observe 7 bp interactions in the distal duplex. This is possible because the 5' and 3' RNA strands are shifted by 1 nt in respect to each other, meaning that G11 of the 5' cRNA base-pairs with C12 of the 3' cRNA, C12 of the 5' cRNA base-pairs with G13 of the 3' cRNA and so on (Figure 3B). In LACV, the 5'/3' distal duplex is stabilized by 6 base-pair interactions (13), whereas, in LASV there are 8 base-pair interactions (28). This shift between the 3' and 5' RNA strands presumably enables the SFTSV L protein to overcome the general lack of complementarity in the 3'/5' duplex region.

In both the EARLY-ELONGATION and EARLY-ELONGATION-ENDO structures, we can reliably fit the 3' cRNA residues found in the space between the distal duplex and the template entry channel. Notably, the G11 base stacks against the K533 sidechain whilst also interacting directly with the H535 sidechain (Supplementary Figure S12). There is a sharp turn which allows the G10 base to sit in a pocket delineated by W1342–K1347 and L1399–S1400 (Supplementary Figure S12). While, the bases U9, C8, and U7 can be fit, this region is by comparison to the rest of the 3' cRNA less defined in both maps, which is most likely due to a low number of interactions with protein residues. The C8 base interacts with the S561 sidechain and the U7 backbone phosphate is coordinated by the K1401 sidechain (Supplementary Figure S12). The remaining template RNA nts 1–6 of the 3' cRNA, are overall much better defined and can be fit with high confidence in both structures. Bases U6 and U5 are found in the template entrance channel and G4 at the +1 position in the active site (Supplementary Figure S12). The only close interaction between the L protein and U6 and U5 is between the R871 sidechain and the U5 phosphate. In addition to the stalling nhUTP molecule, the G4 nt is coordinated by the Y923 and Q1080 carbonyls. There is only one close interaction between the remaining 3' cRNA and the L protein and this is between the G1081 carbonyl and the U3 ribose.

The stalling nhUTP contributes towards the coordination of a single Mg²⁺ ion, also found at the +1 position, with each of the nhUTP phospho-oxygens contributing to one coordinating bond (Figure 3C). The A986 carbonyl and sidechains of D985 and D1126 provide the three remaining coordination bonds required to complete the bipyramidal Mg²⁺ coordination shell. The volume for the 3 nt product strand in both early elongation structures is noisy but nevertheless good enough to fit the cognate bases for the template U1, G2 and U3 bases, which are A1, C2 and A3, respectively (Supplementary Figure S13). There are some interactions between the L protein and product RNA nts but these are limited. A3 is stabilized by interactions from the S1125 and D1126 sidechains, which interact with the ribose O2' and O3' hydroxyls. There is also a relatively long interaction between the A3 phosphate and the S1183 sidechain. G2 is sandwiched in between A1 and A3 while the G2 phosphate interacts with the R1197 sidechain of L. There are no close protein interactions stabilizing A1, which likely contributes to the poor map quality around this particular nt.

The L protein undergoes significant remodelling as elongation progresses

There are several noticeable changes to the L protein comparing pre-initiation to early-stage and then late-stage elongation. The most apparent being an expansion of the L protein core, which leads to an increase in the internal diameter of the L protein by ~6 Å (Supplementary Figure S14). This opening up allows the L protein core to accommodate not just the small 3 nt product we observe in the EARLY-ELONGATION and EARLY-ELONGATION-ENDO structures but also the 10-bp product-template duplex we find in the LATE-ELONGATION structure. Additionally, in comparison to the apo structure there is a rotation of the endonuclease domain by ~30° along the Z axis towards the L protein core (Figure 4A). This movement buries α4 of the endonuclease (residues 134–159, (24)) deeper within the L protein and, in doing so, partially opens the putative product exit channel in preparation for the later stages of elongation. It also provides space for the latter half of the thumb ring domain (roughly composed of L1509–K1577) to rotate away from the L protein core, backfilling the space left free by the movement of the endonuclease domain (Figure 4A and B). Finally, a beta-hairpin of the thumb domain (T1333–V1340) refolds and moves away from the L protein core (Figure 4B) unblocking the template exit channel.

State of RNA at late-stage elongation

While in our early-stage elongation structures the 5'/3' RNA distal duplex is maintained, in the later-stage LATE-ELONGATION structure, the distal duplex is disrupted completely. The SFTSV L protein is stalled through the incorporation of nhUTP at position A21 in the 3' RNA template and the L protein inner cavity contains a 10-bp product-template duplex (Figure 4C). For related polymerases, such as LASV and LACV L proteins, the product-template duplex in the RdRp active site cavity was found to be 8 bp and 9–10 bp, respectively (13,28). The late-elongation map is of sufficient quality and resolution (2.6 Å) to unambiguously identify each base in the growing strand. This led to the interesting finding that instead of stalling at the first A in the 3' cRNA sequence (position 19), the L protein instead stalls at position 21 on the template, which is the second A in the overall sequence and the first A in the 6A tail of the modified template used.

Product exit channel, template exit channel, and the 3' secondary binding site

At the base of the SFTSV L protein core, the product-template duplex splits. This seems to be enabled partly by thumb-ring residue H1573, the sidechain of which protrudes directly into the path of the growing duplex. Unfortunately, the bottom of the SFTSV L protein core does not overlay nicely with the same space in the RVFV L protein. It was not possible therefore to explore the impact of mutating the equivalent of H1573 in the RVFV mini-replicon system. Additionally, recombinant expression of a D112A/H1573A SFTSV L protein mutant did not yield sufficient amounts

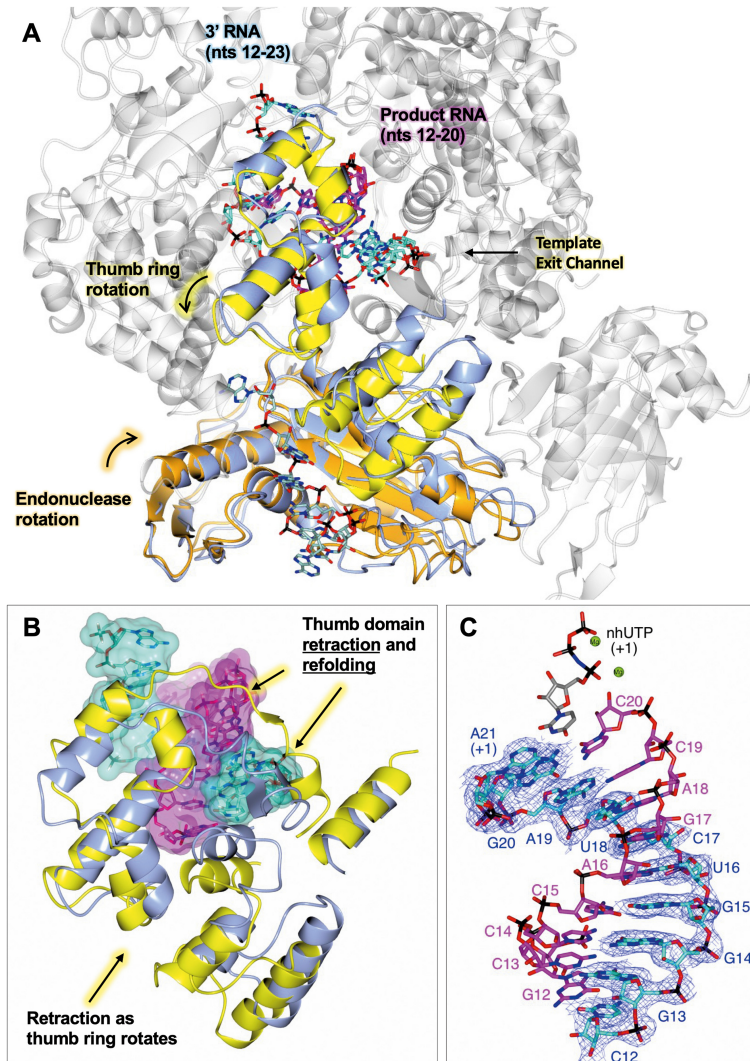


Figure 4. Conformational changes as the L protein proceeds from pre-initiation to late-stage elongation. (A) The SFTSV L protein (LATE-ELONGATION) is shown with the 5' HOOK structure of the SFTSV L protein superposed by SSM in CCP4mg (51). Parts in grey are very similar in both structures and only the LATE-ELONGATION structure is depicted. Domains of the LATE-ELONGATION structure that were shown to either rotate or remodel as the L protein moves from pre-initiation, i.e. hook binding, to late elongation are coloured (orange for the endonuclease and yellow for the thumb ring domain). The corresponding 5' HOOK structure domains are shown in blue. RNAs are shown as sticks with surface overlaid (50% transparency) and coloured either cyan (3' cRNA), magenta (progeny RNA), or pale green (endonuclease-bound RNA). Movements to the endonuclease and thumb ring domains precipitated by the progression to late elongation are indicated. (B) A close-up of the remodelling which leads to the opening of the template exit channel. As in A, the thumb domain and RNA of the LATE-ELONGATION structure is coloured yellow and cyan/magenta, respectively. The apo structure is shown in blue. (C) The 10-bp product-template duplex including the non-hydrolysable UTP (nhUTP) from the LATE-ELONGATION structure are shown as sticks with the respective map overlaid. The map is clipped to the 3' cRNA strand and contoured to 4 σ . Catalytically important magnesium ions (Mg) are shown as green spheres.

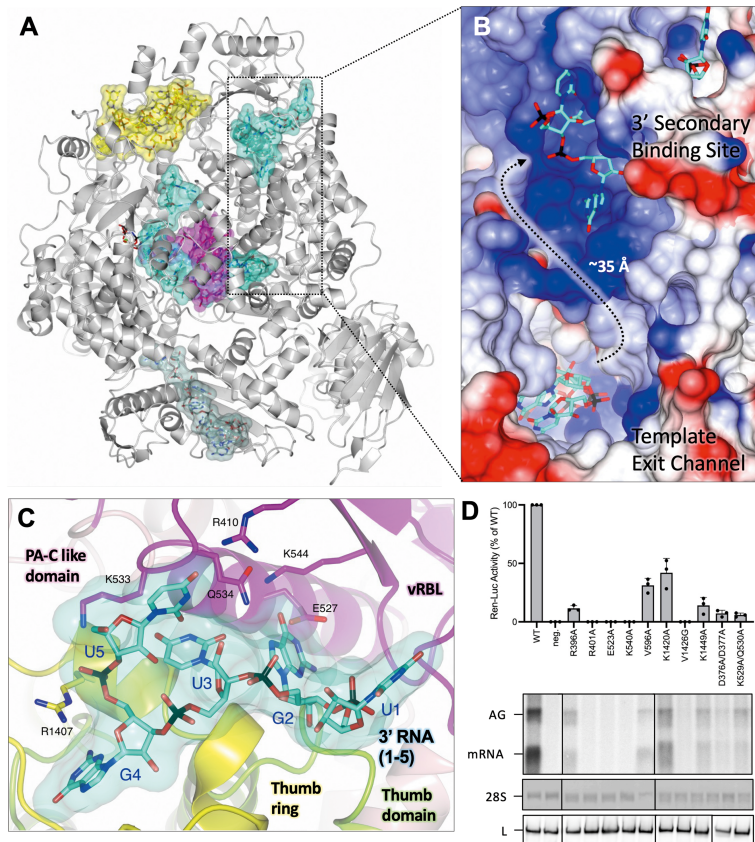


Figure 5. Interaction of 3' cRNA at the 3' secondary binding site. (A) The SFTSV L protein (LATE-ELONGATION) is shown with the protein depicted as ribbon in grey. RNAs are shown as sticks with surface overlaid (50% transparency) and coloured either yellow (5' cRNA), cyan (3' cRNA), magenta (progeny RNA), or pale green (endonuclease-bound RNA). (B) A close-up of the indicated area in A is shown as surface coloured by electrostatic potential demonstrating that the path connecting the template exit channel and 3' secondary binding site has a largely positive charge. (C) The 3' secondary binding site is shown with the L protein vRBL in magenta, the PA-C like domain in pink, thumb in green and thumb ring in yellow. The 3' cRNA nts bound to the protein domain, are shown as sticks and labelled accordingly. (D) RVFV mini-replicon data for L protein with mutations in the 3' secondary binding site presenting luciferase reporter activity (in standardized relative light units relative to the wild-type L protein (WT)). Data are presented as mean values \pm SD of three biological replicates ($n = 3$). All biological replicates are shown as black dots in the top panel. Middle panels present Northern blotting results with signals for antigenomic viral RNA (AG), viral mRNA (mRNA) and 28 S ribosomal RNA (28 S) as a loading control, and the bottom panel shows Western blot detection of FLAG-tagged L proteins (L) to demonstrate general expressibility of the mutants. Uncropped original blots/gels are provided in the Supplementary Data.

core, in the RESTING structure, the distal duplex is absent. The RESTING structure is reminiscent therefore of the LASV 3END-CORE (PDB: 7OEA) (28) structure in which excess 3' vRNA binds to the 3' secondary binding site instead of entering the L protein core. Although we clearly see 3' cRNA binding to the 3' secondary binding site, we also found that we could fit 4 nts of the template 3' cRNA in the L protein core, however, only 2 of these (U3 and G4) could be fit with confidence. The region of the map where

these 3' RNA nts are found is rather noisy but we think that similar to our early-elongation structures, the nts in the L protein core are likely the start of the 3' cRNA. Although we could not visualize initiation, given that the 3' cRNA repeatedly appears to reach at least 4 nts into the L protein core, and that we detect an RNA product corresponding exactly to the terminus of the 3' template without additional or missing nts in our biochemical assays (17), we think that for vRNA synthesis SFTSV primes internally and then re-

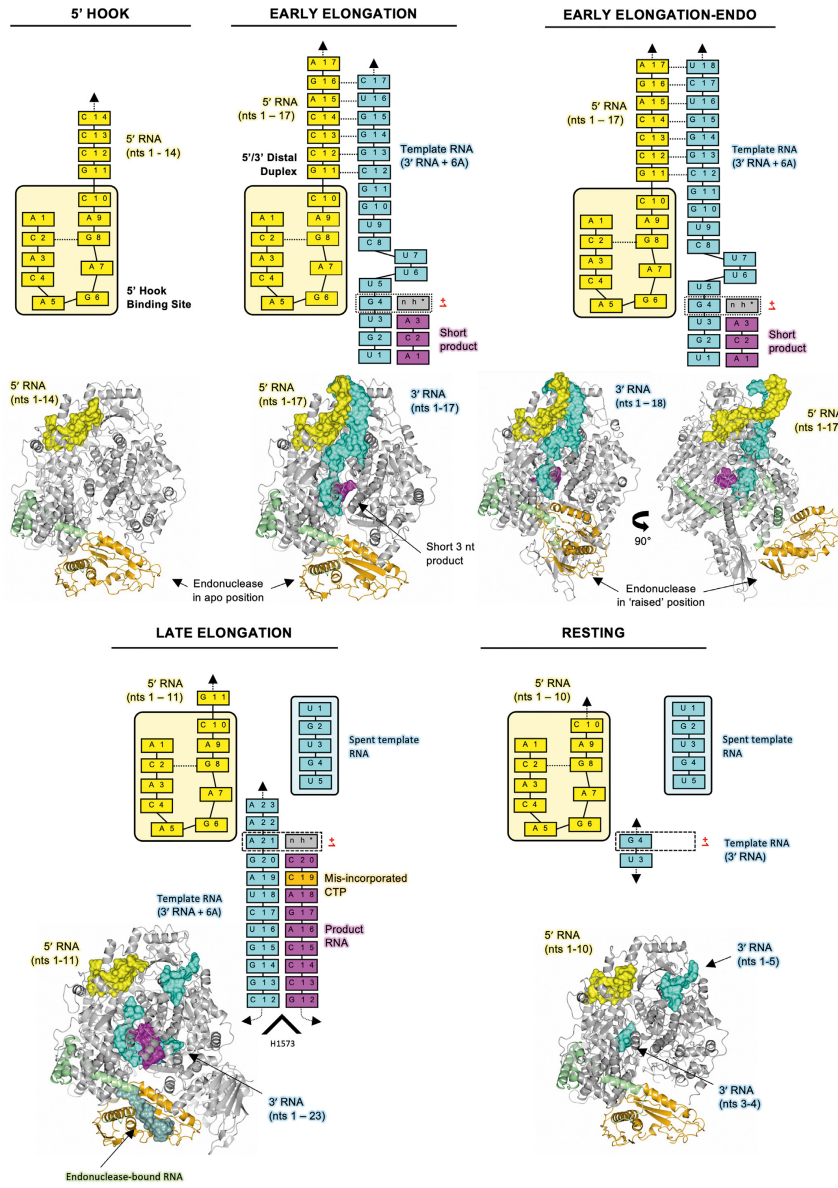


Figure 6. Overview of the SFTSV L protein at resting, pre-initiation, early-elongation, and late-elongation stages. A cartoon representation of each SFTSV L protein structure is shown in grey with the endonuclease and endonuclease-linker highlighted in orange and green, respectively. RNAs are shown as surface and coloured either yellow (5' cRNA), cyan (3' cRNA), magenta (progeny RNA), or pale green (endonuclease-bound RNA). In addition, the RNAs present in each structure are shown schematically and labelled.

low concentrations (≤ 0.5 mM) when compared to UTP (~ 1 mM) (67,68). The level of ATP is coupled to the metabolic status of the cell and is normally > 2 mM (67,68). The intracellular NTP concentrations are therefore even higher than those used in our *in vitro* polymerase assay (~ 0.2 mM) and may positively influence L protein fidelity. The mononegavirus respiratory syncytial virus (RSV) polymerase is capable of producing new RNA either *de novo* or via primer extension (69). However, in *in vitro* biochemical assays, it was found that the ratio of *de novo* vs primer extension RNA synthesis is influenced by NTP concentration reinforcing this idea that both the concentration of NTPs and the ratio of individual NTPs are important for RNA synthesis (69). Indeed, one suggested mechanism of action of the antiviral ribavirin is described to work by competitively inhibiting the cellular enzyme inosine 5'-monophosphate dehydrogenase, which ultimately leads to GTP depletion (70). In addition, at high concentrations (≥ 100 μ M) ribavirin has been shown to be incorporated as a GTP analogue into nascent viral RNA acting as a mutagen in following rounds of genome replication and transcription (71,72). Further factors known to influence the fidelity of polymerases are the availability and concentration of divalent metal ions with Mn^{2+} leading to higher rates of misincorporation compared to Mg^{2+} (73). Our results on the misincorporation of CTP for UTP and *vice versa* under our assay conditions emphasize that NTP concentrations and relative ratios are important to consider for *in vitro* drug testing against NSV RdRp targets. Although out of scope for this study, we think it would be particularly interesting to investigate whether bunyaviral infections have an influence on the intracellular NTP pools as significant changes in dNTP levels have recently been observed in bacteria in response to bacteriophage infection (74). Related to this topic, the low fidelity of the RdRp somewhat limits the confidence in the RNA product modelled in our late-elongation structure. In the corresponding *in vitro* reaction, we detect two products, one most likely with a misincorporated CTP instead of UTP at position 19 of the product as we used nhUTP ($> 95\%$ purity). Therefore, although we modelled the 20mer product with the described misincorporation, the cryo-EM data are most likely a mixture of both the 18mer and the 20mer. Interestingly, the mismatch does not result in a distortion of the RNA duplex, a scenario which has been described in the literature both for DNA and RNA suggesting that the rare tautomers giving rise to these mismatches are in fact relatively widespread in nucleic acids (75,76).

To conclude, the biochemical and structural data presented here has provided us with key insights into how genome replication is catalysed by the SFTSV L protein. We have shown that there are specific changes to the L protein that occur upon 5' RNA hook binding and that several L protein domains adapt further upon addition of 3' RNA to accommodate the 5'/3' RNA distal duplex. Our data show that as the L protein moves from pre-initiation to late-stage elongation, the L protein undergoes further remodelling with the inner cavity of the L protein opening up to accommodate the 10-bp product-template duplex. Similar to other sNSVs, we demonstrate that the SFTSV L protein has a functional 3' secondary binding site, which is occupied under both resting conditions and also dur-

ing the later stages of elongation. This structural data is accompanied by a comprehensive mutational analysis allowing us to identify key amino acid residues involved in binding RNA at the 5' RNA hook binding site and 3' secondary binding site. Altogether, this data provides an excellent foundation up on which further questions can be addressed including how the SFTSV L protein catalyses viral transcription.

DATA AVAILABILITY

Coordinates and structure factors or maps included in this paper have been deposited in the Worldwide Protein Data Bank (wwPDB) and the Electron Microscopy Data Bank (EMDB) with the following accession codes: SFTSV L protein bound to 5' promoter RNA [5' HOOK] EMD-15607 PDB-8AS6; SFTSV L protein early elongation structure [EARLY-ELONGATION] EMD-15608 PDB-8AS7; SFTSV L protein early elongation raised endonuclease structure [EARLY-ELONGATION-ENDO] EMD-15610 PDB-8ASB; SFTSV L protein late elongation structure [LATE-ELONGATION] EMD-15614 PDB-8ASD; SFTSV L protein resting structure [RESTING] EMD-15615 PDB-8ASG.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We want to thank Stephan Günther and Tomas Kouba for support and helpful discussions throughout the project. We thank Imre Berger for providing the DH10EMBaY *E. coli* and the team of the Eukaryotic Expression Facility (EEF) at European Molecular Biology Laboratory (EMBL) Grenoble for support and advice. We also acknowledge support by Carolin Seuring, Cornelia Cazey and Ulrike Laugks for access to the Cryo-EM multi-user facility at CSSB and providing time for sample preparation, screening, and data collection; Wolfgang Lugmayr for help and support in using the CSSB partition on the DESY computer cluster for cryo-EM data processing. We also want to thank Benoit Arragain for feedback on the manuscript.

FUNDING

Leibniz Association's Leibniz competition programme [K72/2017]; part of this work was performed at the Cryo-EM multi-user Facility at CSSB, headed by K.G. and supported by the UHH and DFG [INST 152/772-1, 774-1, 775-1, 776-1]; S.T. benefited from a travel grant from the Leibniz Institute for Virology; E.Q. was supported by an individual fellowship from the Alexander von Humboldt Foundation and a Klaus Tschira Boost Fund; M.R. received funding from the German Federal Ministry of Education and Research [01KI2019]. Funding for open access charge: Leibniz Association, institutional funding.

Conflict of interest statement. None declared.

- beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.
40. Tegunov, D. and Cramer, P. (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods*, **16**, 1146–1152.
 41. Mindell, J.A. and Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.*, **142**, 334–347.
 42. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
 43. Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, **14**, 290–296.
 44. Punjani, A., Zhang, H.W. and Fleet, D.J. (2020) Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods*, **17**, 1214–1221.
 45. Burnley, T., Palmer, C.M. and Winn, M. (2017) Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol.*, **73**, 469–477.
 46. Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S. and Vargas, J. (2021) DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.*, **4**, 874.
 47. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
 48. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A. and Adams, P.D. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.*, **74**, 531–544.
 49. Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D. and Urzhumtsev, A. (2018) New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.*, **74**, 814–840.
 50. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B. *et al.* (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293–315.
 51. McNicholas, S., Potterton, E., Wilson, K.S. and Noble, M.E. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 386–394.
 52. Pettersen, E.F., Goddard, T.D., Huang, C.R.C., Meng, E.E.C., Couch, G.S., Croll, T.I., Morris, J.H. and Ferrin, T.E. (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.*, **30**, 70–82.
 53. Wandzik, J.M., Kouba, T., Karuppusamy, M., Pflug, A., Drncova, P., Provaznik, J., Azevedo, N. and Cusack, S. (2020) A structure-Based model for the complete transcription cycle of influenza polymerase. *Cell*, **181**, 877–893.
 54. Campagnola, G., McDonald, S., Beaucourt, S., Vignuzzi, M. and Peersen, O.B. (2015) Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *J. Virol.*, **89**, 275–286.
 55. Fan, H., Walker, A.P., Carrique, L., Keown, J.R., Serna Martin, I., Karia, D., Sharps, J., Hengrung, N., Pardon, E., Steyaert, J. *et al.* (2019) Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature*, **573**, 287–290.
 56. Gerlach, P., Malet, H., Cusack, S. and Reguera, J. (2015) Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell*, **161**, 1267–1279.
 57. Peng, Q., Liu, Y.Q., Peng, R.C., Wang, M., Yang, W., Song, H., Chen, Y.H., Liu, S., Han, M., Zhang, X.Z. *et al.* (2019) Structural insight into RNA synthesis by influenza D polymerase. *Nat. Microbiol.*, **4**, 1750–1759.
 58. te Velthuis, A.J.W., Robb, N.C., Kapanidis, A.N. and Fodor, E. (2016) The role of the priming loop in influenza A virus RNA synthesis. *Nat. Microbiol.*, **1**, 16029.
 59. Kumar, G., Cuypers, M., Webby, R.R., Webb, T.R. and White, S.W. (2021) Structural insights into the substrate specificity of the endonuclease activity of the influenza virus cap-snatching mechanism. *Nucleic Acids Res.*, **49**, 1609–1618.
 60. 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. *J. Virol.*, **80**, 2337–2348.
 61. 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. *J. Virol.*, **66**, 1370–1376.
 62. Garcin, D., Lezzi, M., Dobbs, M., Elliott, R.M., Schmaljohn, C., Kang, C.Y. and Kolakofsky, D. (1995) The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. *J. Virol.*, **69**, 5754–5762.
 63. Polyak, S.J., Zheng, S. and Harnish, D.G. (1995) 5' Termini of Pichinde Arenavirus S-RNAs and Messenger-RNAs contain nontemplated nucleotides. *J. Virol.*, **69**, 3211–3215.
 64. Prehaud, C., Lopez, N., Blok, M.J., Oby, V. and Bouloy, M. (1997) Analysis of the 3' terminal sequence recognized by the Rift Valley fever virus transcription complex in its ambisense S segment. *Virology*, **227**, 189–197.
 65. Holm, T., Kopicik, J.D., Busch, C., Olschewski, S., Rosenthal, M., Uetrecht, C., Gunther, S. and Reindl, S. (2018) Biochemical and structural studies reveal differences and commonalities among cap-snatching endonucleases from segmented negative-strand RNA viruses. *J. Biol. Chem.*, **293**, 19686–19698.
 66. Kuang, W., Zhang, H., Cai, Y., Zhang, G., Deng, F., Li, H., Hu, Z., Guo, Y., Wang, M., Zhou, Y. *et al.* (2022) Insights into two-Metal-Ion catalytic mechanism of cap-Snatching endonuclease of Ebirur Lake virus in Bunyavirales. *J. Virol.*, **96**, e0208521.
 67. Kong, Z., Jia, S., Chabes, A.L., Appelblad, P., Lundmark, R., Moritz, T. and Chabes, A. (2018) Simultaneous determination of ribonucleoside and deoxyribonucleoside triphosphates in biological samples by hydrophilic interaction liquid chromatography coupled with tandem mass spectrometry. *Nucleic Acids Res.*, **46**, e66.
 68. Traut, T.W. (1994) Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.*, **140**, 1–22.
 69. Noton, S.L., DeFlube, L.R., Tremaglio, C.Z. and Fearn, R. (2012) The respiratory syncytial virus polymerase has multiple RNA synthesis activities at the promoter. *PLoS Pathog.*, **8**, e1002980.
 70. Streeter, D.G., Witkowski, J.T., Khare, G.P., Sidwell, R.W., Bauer, R.J., Robins, R.K. and Simon, L.N. (1973) Mechanism of action of 1-D-ribofuranosyl-1,2,4-triazole-3-carboxamide (Virazole), a new broad-spectrum antiviral agent. *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 1174–1178.
 71. Crotty, S., Maag, D., Arnold, J.J., Zhong, W., Lau, J.Y., Hong, Z., Andino, R. and Cameron, C.E. (2000) The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat. Med.*, **6**, 1375–1379.
 72. Nystrom, K., Waldenstrom, J., Tang, K.W. and Lagging, M. (2019) Ribavirin: pharmacology, multiple modes of action and possible future perspectives. *Future Virol.*, **14**, 153–160.
 73. Dominguez, O., Ruiz, J.F., Lain de Lera, T., Garcia-Diaz, M., Gonzalez, M.A., Kirchhoff, T., Martinez, A.C., Bernad, A. and Blanco, L. (2000) DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J.*, **19**, 1731–1742.
 74. Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., Yirmiya, E., Avraham, C., Brandis, A., Mehlman, T. *et al.* (2022) Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nat. Microbiol.*, **7**, 1200–1209.
 75. Kimsey, I.J., Petzold, K., Sathyamoorthy, B., Stein, Z.W. and Al-Hashimi, H.M. (2015) Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature*, **519**, 315–320.
 76. Bebenek, K., Pedersen, L.C. and Kunkel, T.A. (2011) Replication infidelity via a mismatch with Watson-Crick geometry. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 1862–1867.

4. Non-cumulative publications





4.1 Hantavirus Replication Cycle – An Updated Structural Virology Perspective

A literature review on recent findings on the replication cycle of bunyaviruses of the *Hantaviridae* family with a focus on structural biology.

Original publication at <https://doi.org/10.3390/v13081561>

Review

Hantavirus Replication Cycle—An Updated Structural Virology Perspective

Kristina Meier ^{1,†}, Sigurdur R. Thorkelsson ^{2,†}, Emmanuelle R. J. Quemin ^{2,*} and Maria Rosenthal ^{1,3,*}

¹ Department of Virology, Bernhard Nocht Institute for Tropical Medicine, 20359 Hamburg, Germany; kristina.meier@bniitm.de

² Centre for Structural Systems Biology, Leibniz Institute for Experimental Virology, University of Hamburg, 22607 Hamburg, Germany; sigurdur.thorkelsson@cssb-hamburg.de

³ Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, 22525 Hamburg, Germany

* Correspondence: emmanuelle.quemin@cssb-hamburg.de (E.R.J.Q.); rosenthal@bniitm.de (M.R.); Tel.: +49-40-899887-717 (E.R.J.Q.); +49-40-42818-930 (M.R.)

† These authors contributed equally to the work.

Abstract: Hantaviruses infect a wide range of hosts including insectivores and rodents and can also cause zoonotic infections in humans, which can lead to severe disease with possible fatal outcomes. Hantavirus outbreaks are usually linked to the population dynamics of the host animals and their habitats being in close proximity to humans, which is becoming increasingly important in a globalized world. Currently there is neither an approved vaccine nor a specific and effective antiviral treatment available for use in humans. Hantaviruses belong to the order *Bunyavirales* with a tri-segmented negative-sense RNA genome. They encode only five viral proteins and replicate and transcribe their genome in the cytoplasm of infected cells. However, many details of the viral amplification cycle are still unknown. In recent years, structural biology methods such as cryo-electron tomography, cryo-electron microscopy, and crystallography have contributed essentially to our understanding of virus entry by membrane fusion as well as genome encapsidation by the nucleoprotein. In this review, we provide an update on the hantavirus replication cycle with a special focus on structural virology aspects.

Keywords: hantaviruses; structural virology; cryo-electron tomography; cryo-electron microscopy; X-ray crystallography; viral fusion glycoproteins; viral replication; viral transcription; viral genome encapsidation; virion assembly



Citation: Meier, K.; Thorkelsson, S.R.; Quemin, E.R.J.; Rosenthal, M. Hantavirus Replication Cycle—An Updated Structural Virology Perspective. *Viruses* **2021**, *13*, 1561. <https://doi.org/10.3390/v13081561>

Academic Editors: Kumiko Yoshimatsu and Hiroaki Kariwa

Received: 18 June 2021
Accepted: 2 August 2021
Published: 6 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bunyavirales is a large viral order that includes many emerging viruses with high epidemic potential [1]. Notably, several bunyaviruses from the *Hantaviridae* family can infect humans, causing thrombocytopenia, capillary permeability (leading to vascular leakage), and immunopathology due to the activation of the innate and adaptive immune systems (reviewed in [2]). Whereas Old World hantaviruses, such as Hantaan virus (HTNV) and Puumala virus (PUUV), which are prevalent in Europe and Asia, can cause hemorrhagic fever with renal syndrome [3], the New World hantaviruses, such as Andes virus (ANDV) and Sin Nombre virus (SNV), which are found in the Americas, primarily cause hantavirus cardiopulmonary syndrome [4].

Humans are generally considered as dead-end hosts with the exception of some reports of person-to-person transmission for ANDV in Argentina [5], including recent cases of “super-spreaders” of the same viral strain [6] and the suspicion of transmission *via* the transfusion of platelets and blood products for PUUV in Finland [7]. Outbreaks of hantaviruses correlate with the population dynamics of their carriers, and each viral species seems to have a different and specific primary reservoir, which is usually either rodents or insectivore mammals such as moles, shrews, and bats [3,4]. Based on a rodent-hantavirus

codivergence hypothesis, it has been proposed that hantaviruses have been co-evolving for a long time with their natural hosts, which are chronically infected with high viremia but remain asymptomatic [8]. However, the mechanisms that support hantavirus replication in a given host and that limit the spill over or adaptation to new organisms remain poorly understood (reviewed in [9,10]). Human infections upon exposure to contaminated aerosolized secreta or excreta have been reported almost exclusively from rodents so far, with case fatality rates ranging from 0.2% for the PUUV endemic in Europe [3] to ~50% in the case of ANDV in South America [4]. The possibility that hantaviruses circulating in insectivores can cause human diseases might have been overlooked and requires closer attention [2].

The enveloped virions are spherical or pleomorphic, are decorated with spikes, and encase the tri-segmented single-stranded viral RNA (vRNA) genome of negative polarity (Figure 1) [11–14]. The small (S), medium (M), and large (L) genomic segments encode four structural proteins: the nucleoprotein N, the glycoproteins Gn and Gc (resulting from the maturation of the glycoprotein precursor GPC after co-translational cleavage by the cellular signal peptidase complex), and the large (L) protein [15]. In some hantaviruses, the S segment also encodes a nonstructural protein (NSs) [16]. Although for ANDV, the NSs might work as an immunosuppressor of the type I interferon induction pathway [17], its functions in other hantaviruses and potentially additional roles during infection are unknown [18]. The envelope glycoproteins Gn and Gc form the spike complex responsible for receptor-binding and Gc-mediated membrane fusion. Integrins have been identified as receptors *in vitro* although there is no information on the receptors and co-receptors used in the natural context of infection (reviewed in [19]). Similar to other bunyaviruses such as the La Crosse virus (LACV, *Peribunyaviridae*) [20], the Crimean–Congo Hemorrhagic Fever virus (*Nairoviridae*) [21–23], and the Uukuniemi virus (*Phenuiviridae*) [24], hantaviruses appear to rely on several pathways for entry including macropinocytosis and endocytosis that is either clathrin-, calveolin- or cholesterol-dependent [19]. Particles then travel through the endocytosis pathway, and the low pH inside endosomes triggers a conformational change in the Gc glycoproteins [25]. This leads to the insertion of the Gc fusion loop into the endosomal membrane, the fusion of the latter with the viral envelope, and the subsequent release of the virion content. Each vRNA segment is flanked by non-coding regions at the 5' and 3' termini, exhibiting complementary sequences that are predicted to form a so-called panhandle structure, essential for the viral transcription and genome replication conducted by the viral L protein in the host cell cytoplasm. Based on recent structural data on the L protein, the formation of the panhandle might also be attributed to the L protein binding to both RNA ends [26–31]. Following the synthesis of virion components, the glycoproteins, which are specific for each hantavirus, play a key role in virus assembly and maturation (reviewed in [32]). In particular, the cytosolic tail of Gn is likely interacting with N proteins of the ribonucleoprotein complex [19]. Finally, it has been proposed that the virions being assembled bud into the Golgi apparatus and are released by exocytosis (Old World hantaviruses) or are released directly at the plasma membrane (New World hantaviruses), but the details of virion egress are largely unknown.

In this review, we summarize our current understanding of the molecular and structural biology of the hantavirus replication cycle, focusing on recent insights from structural virology studies. We discuss somewhat contradictory hypotheses regarding hantavirus entry, replication, and assembly. Finally, we highlight open questions in the field, which are not only critical to increase our knowledge on hantavirus–host interactions but to also develop specific countermeasures against emerging bunyaviruses.

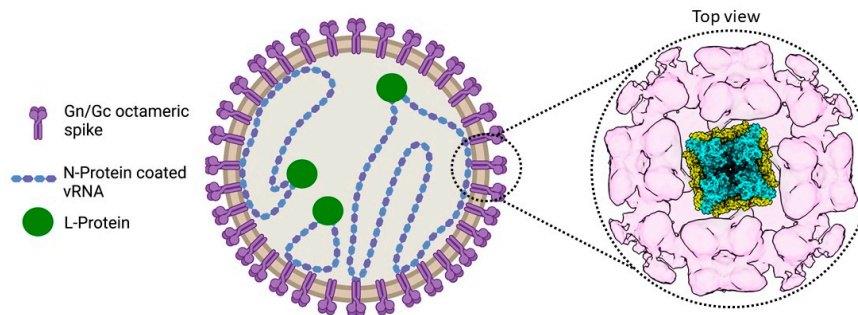


Figure 1. The hantavirus virion. A drawing of a hantavirus virion is presented (left) including a top view of the map of the TULV glycoprotein lattice (right), which was obtained by electron cryo-tomography and subvolume averaging (EMD-11236). A model of the prefusion tetrameric Gn/Gc spike complex (PDB: 6ZJM), presented as a cartoon, was fitted into the central volume [24].

2. Literature Review

2.1. Entry into the Host Cell

While bunyaviruses from the *Phenuiviridae* family have quite regular particles with $T = 12$ icosahedral quasi-symmetry and exhibit low plasticity in terms of size and shape [33–35], hantaviruses are pleomorphic with a diameter of 120–160 nm and vary from round to elongated [14,36]. The viral envelope is decorated with glycoproteins Gn and Gc in a grid-like pattern specific to hantaviruses [11–13,25]. However, this grid-like pattern can be also interrupted, resulting in bare patches of membrane on the virion surface [14]. Inside, the virion carries three genome segments (L, M, and S) which are encapsidated by the nucleoprotein N and are associated with the L protein (Figure 1) [15] (see assembly section).

Gc is a class II viral fusion protein with a three-domain architecture that is characterized by a β -sheet-rich secondary structure. Gn and Gc form the spike complex, which is by itself sufficient for host cell entry [13,37]. Consistent with biochemical data on the Tula virus (TULV), cryo-electron tomography analysis of HTNV, TULV, and PUUV have shown that the characteristic squared spike complexes are tetrameric and formed by the heterodimers of Gn and Gc (Figures 1 and 2a) [12,13,38]. In the assembled fusion protein, Gc forms a lattice around the Gn protomers, which together form the spike through Gn:Gn interactions, maintaining Gc in a metastable pre-fusion conformation [25,37,38]. The fusion loop of hantavirus Gc is significantly different from related class II fusion proteins in that the residues involved in membrane insertion are split between three loops compared to all being on a single fusion loop for flaviviruses and alphaviruses (Figure 2b) [25,37]. The dissociation of the Gn/Gc complexes at a low pH is thought to make the tripartite fusion loop accessible for membrane fusion, although this dissociation does not seem to trigger membrane fusion and seems to even be reversible [39]. While Gc is responsible for membrane fusion, the role of Gn in viral entry remains unclear. However, it was recently shown that Gn can be a target for neutralizing antibodies against HTNV, emphasizing its importance for the spike function [40]. The conformational shift between pre- and post-fusion of Gn/Gc on viral particles is triggered by acidification in the endosome following virion uptake [34].

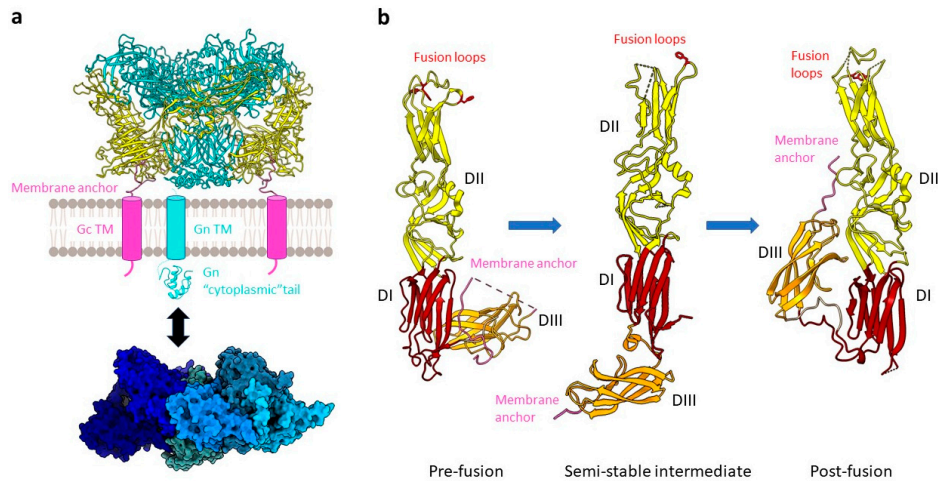


Figure 2. The hantavirus spike complex. (a) A pre-fusion spike complex on the virus surface composed of the Gn (teal) and Gc (yellow) proteins forming a tetrameric assembly (example of ANDV, PDB: 6ZJM) is presented as a cartoon. The viral envelope (grey membrane) as well as the transmembrane regions (TM) of Gn (cyan cylinder) and Gc (pink cylinder) are schematically shown. One copy of the Gn cytoplasmic tail (cyan, PDB: 2K9H) and a heptameric RNP-like assembly of the N protein (blue; PDB: 6I2N) are displayed as a cartoon and surface representation, respectively. An interaction between the latter is indicated (black arrow). (b) A comparison of pre-fusion ANDV Gc (PDB: 6Y5F), an intermediate state of HTNV stabilized by an antibody (antibody not shown) (PDB: 5LJY), and the ANDV Gc post-fusion conformation (PDB: 6Y6Q) is shown. Domains are colored as DI, red; DII, yellow; DIII, orange. Key residues W766, Y745, F900 for ANDV and F250 for HTNV of the tripartite fusion loop are highlighted in red with the side chains displayed as sticks.

Although the cellular receptors in rodents are still unknown, there is evidence that the human integrins used are (i) $\alpha V\beta 3$ in the case of SNV [41], HTNV, the Seoul virus, and PUUV [42,43]; (ii) $\alpha V\beta 1$ for the Sangassou virus [44]; and (iii) $\beta 1$ integrins for the Prospect Hill virus [41,42]. Apart from integrins, other cell surface proteins mediate virus entry *in vitro*, including the decay-accelerating factor CD55 and the complement receptor gC1qR/p32 [43–48]. So far, only protocadherin-1 has been shown to play a role in the entry of all New World hantaviruses (comprehensively reviewed by [19]). The hantaviruses ANDV and HTNV additionally depend on membrane cholesterol for host cell entry, most likely for endosomal escape and membrane fusion [49–51].

Despite the differences observed in the required macropinocytosis-related kinases ML-7 and ML-9 between HTNV, an Old World hantavirus, and ANDV, a New World hantavirus, both viruses enter human respiratory epithelial cells using a pathway that depends on sodium proton exchangers and actin, supporting that the entry process can involve micropinocytosis [52,53]. HTNV and PUUV also use clathrin-dependent endocytosis for host cell entry [53,54]. Additionally, it has been reported that HTNV, PUUV, and the Black Creek Canal virus (BCCV) tend to preferentially enter at the apical site of epithelial and endothelial cells [46,55]. However, more sound evidence is needed to uncover the intricacies of the hantavirus entry pathway(s). The current lack of relevant *in vitro* models is indeed a bottleneck to any study on hantavirus–host interactions in general, given their narrow host range and diversity [19,56].

After uptake, viral particles are transported to early or late endosomes [19], where the low pH of the compartment triggers large conformational changes in Gn/Gc and the Gn:Gc interface gets disrupted [40]. Similar to what has been observed for other class II fusion proteins [57], the transition of Gc from pre- to post-fusion is associated with rearrangements of the three domains in their relative orientation. First, a large relocation of

the β -sheet stack at the C-terminus of domain III occurs. Upon additional reorganization of domain II, the buried key non-polar residues at the tip of the domain, constituting the tripartite fusion loop, become exposed on the molecular surface (Figure 2b). In fact, upon acidification, there is formation of a carboxylate–carboxylic acid hydrogen bond structuring the membrane-binding region [25,37,58]. Altogether, these conformational changes result in the assembly of protein monomers into trimers, which are stabilized by a conserved N-terminal segment or “N-tail”, which is so far unique to hantavirus Gc, along with membrane fusion, virion uncoating, and the release of the viral genome. Postfusion structures of class II fusion proteins are commonly trimers [57], which is also the case for the postfusion structures of ANDV Gc [37] and Gn/Gc [25]. The hantavirus-specific N-terminal tail of Gc was found to stabilize the Gc trimer [37]. Whether the three protomers of the fusion-active trimer originate from the same tetrameric prefusion complex or assemble from different tetrameric complexes during virion entry is still unclear.

2.2. Viral Genome Replication and Transcription

Viral ribonucleoproteins (vRNPs) consisting of the genomic RNA encapsidated by N protein and associated with the L protein are the functional units of genome replication and transcription (Figure 3). Both of these functions are catalyzed by the L protein that harbors the RNA-dependent RNA polymerase (RdRp).

Upon the release of the vRNPs into the cytoplasm, genome replication is initiated *de novo* and proceeds *via* a positive-sense complementary intermediate RNA (antigenome or cRNA) which is, similarly to the vRNA, encapsidated by N proteins. Interestingly, although self-initiating polymerases like the L protein typically use purines as initiating nucleotides, the hantavirus genome commences with a uridine monophosphate, suggesting initiation *via* a priming and realigning mechanism with the subsequent cleavage of the resulting overhang [59,60].

Although the bunyavirus vRNA and cRNA are always associated with N proteins, this is generally not the case for viral mRNA. However, it was reported that the encapsidation of the LACV (*Peribunyaviridae*) S segment mRNA with the N protein can occur, albeit with much lower affinity than that of vRNA and cRNA. Additionally, encapsidation of the S segment mRNA by the N protein was shown to prevent its own translation, which may serve as a negative feedback expression control at high cellular concentrations of N [61].

The bunyavirus genome segments are flanked by non-coding regions (also called untranslated regions, UTRs) which are highly conserved in each bunyavirus family. Based on the sequence complementarity of these UTRs, it was hypothesized that they promote circularization of the genome segments into a so-called panhandle conformation [62,63]. However, more recent structural data on closely related bunya- and orthomyxoviruses suggests that both RNA ends can be bound by the L protein at specific sites within the protein [26–30]. Thus, the genome circularization observed in electron microscopy (EM) (reviewed in [64]) might be the result of the L protein:RNA interaction rather than of panhandle formation *via* base-pairing but could also rely on both mechanisms (Figure 3b).

Severson et al. conducted binding studies with either (i) full-length vRNA; (ii) vRNA deletion mutants lacking either the UTRs or the first twelve 5' terminal nucleotides; (iii) using oligonucleotides corresponding to the 3'; or (iv) 5' termini of the vRNA only [65]. These studies led them to postulate a cis-acting encapsidation signal within the noncoding region of the HTNV 5' terminal vRNA that would be recognized by the N protein. Upon specific recognition of this 5' terminal signal, genome encapsidation would then be driven by the specific interactions among N protein monomers as well as proposedly sequence-unspecific interactions between N proteins and the remaining RNA along the genome or antigenome, respectively. While for HTNV, the N protein was reported to bind single-stranded RNA with a higher affinity than double-stranded RNA, the PUUV N protein has been reported to preferentially bind to double-stranded rather than single-stranded vRNA [66]. Along this line, SNV N was found to specifically bind to the panhandle composed of the 3' and 5' vRNA in an artificial minipanhandle RNA but not the single-

stranded region of this RNA [67]. Whether the exact mechanism of RNA encapsidation by N is based on sequence-recognition and/or secondary structure-specific binding requires further investigation, including on the essential experimental controls, which were partially incomplete in the mentioned studies.

Crystal structures of the hantavirus N protein core revealed two lobes clamping a positively charged RNA binding groove with N- and C-terminal extensions linked to the respective lobe and mediating N protein oligomerization [68,69]. Recent cryo-EM studies showed that HTNV N forms a left-handed helical assembly, in which one N contacts 6 other protomers (Figure 3b). The N- and C-terminal arms of an N protomer interact with the N- and C-terminal lobes of the previous and successive protomers in the helix, respectively. The C-terminal extension can also rotate to contact the N-terminal arm of the same subunit, creating a positively charged groove in the helical assembly that is compatible with viral RNA binding [70]. Helix formation is likely coupled to RNA-binding since hantavirus N proteins were reported to be mostly trimeric [67] or hexameric [69] in solution. Indeed, Arragain et al. observed cellular RNA bound to the N protein after recombinant expression in the absence of viral RNA, suggesting that N protein:RNA interaction and the formation of the helical assembly may not depend on virus-specific RNA sequences [70]. It is still unclear how the L protein gains access to the encapsidated RNA, but interaction of the L protein and the N-terminus of the N protein seems to be necessary for viral RNA synthesis [71]. In the currently proposed model, this interaction with N brings the L protein in close proximity to the RNA, and the helical assembly of the N protomers would only be locally disrupted at the sites where the L protein reads the encapsidated RNA [70]. Of note, in addition to this compact helical conformation, a more flexible pearl-necklace-like conformation of the RNA-associated N protein has been reported [12,68], which would allow for more structural flexibility and may facilitate genome circularization.

Like other segmented negative-strand RNA viruses, hantaviruses employ a cap-snatching mechanism in which short, capped primers are cleaved off of host cell mRNAs and are subsequently used to prime viral mRNA synthesis [59]. The cap-snatching mechanism of the influenza virus (*Orthomyxoviridae*) has been extensively studied [72–74] and may occur analogously to that of bunyaviruses with the exception that orthomyxoviruses carry a heterotrimeric polymerase complex consisting of PA, PB1, and PB2 subunits. Additionally, orthomyxoviruses replicate and transcribe their genome within the cell nucleus, whereas bunyavirus genome replication and transcription occur in the cytoplasm [75,76].

The cap-snatching endonuclease of hantaviruses is located at the N-terminus of the L protein. It is classified as a His⁺ endonuclease, as it contains a catalytically important histidine residue upstream of the metal-coordinating PD-D/E-K active site motif. Overall, this domain is similar to the endonucleases of the influenza A virus PA subunit and the LACV L, but the hantavirus endonuclease has a significantly higher *in vitro* activity [77,78], which seems to limit its own recombinant expression in cells [79]. Sequestration of host cell mRNA has been postulated to occur at a distinct cap-binding site of the N protein, which would subsequently interact with the L protein for RNA cleavage and transcription initiation [80]. However, the structural characterization of the HTNV N protein did not yield evidence of a canonical cap-binding motif [69,70]. In contrast, recent studies provide structural and some functional evidence of a cap-binding domain in the C-terminus of arenavirus, phenuivirus, and peribunyavirus L proteins that are similar to the PB2 subunit of the influenza virus polymerase complex [27,30,81–83].

While the influenza polymerase complex has been shown to interact with the host cell polymerase II within the nucleus to gain access to mRNA caps [73,84], the specific cellular mRNA targets of bunyaviruses remain unclear (reviewed by [85]). Some studies suggest that hantaviruses snatch caps from non-sense RNA in processing bodies (P bodies), which are non-membranous cytoplasmic compartments specialized in RNA turnover, including the decapping machinery enzymes 1a, 1b, and 2 (DCP1a, DCP1b, DCP2) [86]. SNV has been reported to preferentially steal caps from mRNAs with premature stop codons [87],

and SNV N was reported to colocalize with DCP1a [86]. Concomitantly, the transcription of the Rift Valley Fever virus (RVFV, *Phenuiviridae*) was shown to be restricted by the P body-associated decapping machinery, suggesting a shared pool of mRNA targets. Notably, RVFV N protein localization partially overlaps with P body-resident proteins [88]. In contrast, a recent study found no significant colocalization of TULV N protein and P body markers but reported significant colocalization of TULV RNA and N protein with stress granule-resident T-cell restricted intracellular antigen 1 (TIA-1) as well as an increase in the number of stress granules in TULV-infected cells [89]. Stress granules, similar to P bodies, are non-membranous compartments but rather serve as storage sites for translation-stalled mRNAs under cellular stress more than as sites of RNA decay. Stress granules contain small ribosomal subunits as well as translation initiation factors such as the eukaryotic initiation factors 4E and 4G (eIF4E, eIF4G). While eIF4E is also a constituent of the P bodies, eIF4G is restricted to stress granules [90,91]. However, stress granules and P bodies share some similarities in protein composition and are able to interact with one another and exchange RNA as well as possibly protein components (Figure 3a) [92–94]. Thus, both may serve as a site for cap-snatching and more research is required to further elucidate the roles of specific RNA granules during bunyavirus infection.

After transcription of viral genes *via* cap-snatching (Figure 3a), translation follows. For HTNV, it was proposed that the N protein could substitute the function of the eukaryotic initiation factor 4F complex (eIF4F), resulting in the preferential translation of viral mRNA [95,96]. However, structural characterization of HTNV N showed similarity to the tumor suppressor protein Programmed Cell Death 4 (PDCD4) [69], which disturbs the formation of eIF4F [97,98], and HTNV N may therefore also impede cellular mRNA translation by structurally mimicking PDCD4.

Another aspect of cytoplasmic replication is the possibility of coupling transcription to translation, which has been suggested for the Bunyamwera virus (BUNV, *Peribunyaviridae*) and may serve to avoid premature transcription termination [99].

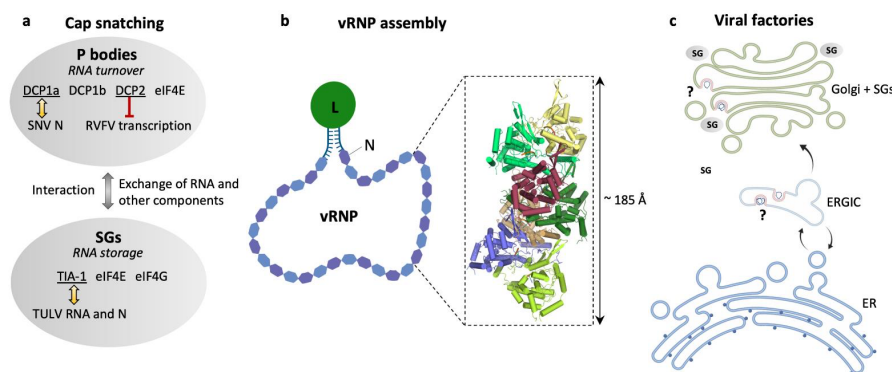


Figure 3. Hantavirus ribonucleoproteins and the location of viral genome replication and transcription. (a) Schematic representation of cellular RNA granules that have been reported as sites for hantavirus cap snatching: processing (P) bodies [86] and stress granules (SGs) [89]. Selected marker proteins for both granules are shown as well as eIF4E to emphasize its presence in both granules. Yellow arrows indicate reported colocalization of P body marker DCP1a and SNV N [86] or SGs marker TIA-1 and TULV RNA and N [89], and the red barred arrow indicates inhibition of RVFV transcription by P body-resident protein DCP2 [88]. (b) Schematic representation of a viral ribonucleoprotein (vRNP). The viral RNA is associated with N and L proteins in a panhandle-like conformation with the complementary genome ends forming a partially double-stranded region and the termini is most probably bound to the L protein. The close-up represents a heptameric assembly of N (PDB: 6I2N) into a helical RNP-like structure with tri-nucleotide RNAs (orange) bound to each N protomer. (c) Overview of possible locations of viral factories for genome transcription and replication in the endoplasmic reticulum (ER), endoplasmic reticulum–Golgi intermediate compartment (ERGIC) [100], and the Golgi apparatus associated with SGs [89].

Apart from the source of mRNA caps, the exact site of hantavirus genome replication and transcription has yet to be determined. Ramanathan et al. reported the formation of perinuclear structures by the HTNV N protein and their colocalization with markers of the endoplasmic reticulum–Golgi intermediate compartment (ERGIC) but only little colocalization with markers of only the endoplasmic reticulum (ER) or Golgi apparatus. Therefore, they suggested the formation of viral factories, virus-induced compartments, at the ERGIC and that the N protein would be quickly assembled into virions as soon as it reached the Golgi apparatus [100]. Davies et al. also reported perinuclear structures formed by the TULV N protein. These structures were of tubular nature and stained positive for Golgi markers and vRNA as well as for cRNA. Furthermore, they observed the recruitment of stress granules to these structures, suggesting the formation of viral factories within a structurally remodeled Golgi associated with stress granules (Figure 3c) [89].

2.3. Assembly and Egress of Viral Progeny

In the ER, the maturation of the glycoprotein precursor GPC by co-translational cleavage happens at the conserved pentapeptide motif WAASA [101]. The Gn and Gc glycoproteins then travel from the ER to the Golgi apparatus, and they oligomerize to form heterodimers [25,102–105]. Studies on PUUV revealed that Gn relies on Gc to be transported from the ER to the Golgi and that this process involves the C-terminal cytoplasmic tail of Gc [106]. In the full spike complex, the Gn/Gc dimer contains two glycan chains that are thought to originate from the ER, which stabilize the spike structure [25] and are essential for viral assembly and entry [103].

Although it is generally considered that Old World hantaviruses assemble in the Golgi apparatus like other bunyaviruses (e.g., the Uukuniemi Virus [107]), it has been proposed that the assembly of New World hantaviruses happens at the plasma membrane. This is based on the fact that the viral particles of the SNV and the BCCV were observed in the extracellular space close to the plasma membrane by conventional EM [55,108]. However, SNV virus-like particles were also found in the Golgi in one of these studies [108]. In addition, SNV glycoproteins could be detected at the plasma membrane at late infection time points [105]. However, SNV was detected in the perinuclear regions of the pulmonary endothelial cells in an EM pathogenesis investigation of tissue samples from patients [109] while analysis of a new hantavirus isolate in Vero cells reported several budding sites [110]. Recent work on the Old World HTNV and New World ANDV made use of high-pressure freezing and freeze-substitution to better preserve samples for EM and reported fragmentation and unstacking of the Golgi and herniation of the rough ER but no intracellular viral particles. At 7 and 9 days of infection, HTNV virions were observed extracellularly, close to projections of the plasma membrane, supposedly as a consequence of viral budding directly at the cell surface [111]. More systematic studies of both Old and New World hantaviruses, especially immunolabeling and 3D analysis of viral assembly and egress are still lacking to draw any conclusion on the assembly sites.

Gn/Gc octameric spikes serve a vital role in the budding of the virion. The lateral interactions between Gc proteins at the inter-spike interfaces have been proposed to be sufficient to induce the membrane curvature that would facilitate viral budding into the Golgi [13,38]. In fact, expression of the ANDV and PUUV Gn and Gc glycoproteins alone leads to the formation of virus-like particles released into the extracellular milieu [112]. In addition, there is significant reduction in virion production upon the disruption of the Gc:Gc inter-spike interfaces [25,38].

Gn and Gc are both type I integral transmembrane proteins with a C-terminal tail following the hydrophobic anchor domain [39]. The N protein was described to interact with both the cytoplasmic tail of Gc and Gn in PUUV [39]. Gn contains a 110 amino acid-long cytoplasmic tail, which interacts with N and might act as a substitute for the lack of matrix protein. This interaction is thought to be mediated by two zinc finger domains of the Gn cytoplasmic tail, which form a compact and unique fold [25,39,113,114]. Notably, the HTNV N protein also associates with RNA [67,70], forming ribonucleoprotein-like

particles when expressed alone in mammalian cells but does not lead to the production of virus-like particles. For the recruitment of N into virus-like particles, the co-expression of Gn and Gc is required [115]. The N protein is essential for the encapsidation of the viral RNA segments by wrapping around the vRNA in a helical manner with 3.6 subunits per twist, fitting the negatively charged nucleotides into a positively charged groove within the helical structure [70] (see Section 2.2). *Via* interaction with the cytoplasmic tail of Gn, it confers recruitment of the genomic RNA segments into virions, but the exact sequence of these assembly steps is not yet clear. Some evidence suggests that the vRNA, the N protein, and the glycoproteins accumulate at the Golgi [89,116] and that microtubules are necessary for the correct movement of the viral components to the assembly site [100]. After budding into the Golgi apparatus, the viral particle is transported to the plasma membrane where it is then released *via* exocytosis. The involvement of multivesicular bodies or recycling endosomes in mature particle egress has been proposed and still remains largely unknown [117]. The N protein of the BCCV co-localizes with actin in infected cells, but the link between such association and viral assembly or release remains to be determined [118].

Of note, although the N protein can be detected in infected cells as early as 4 h post infection [100], putative viral factories were not perceptible at 36 h post infection, at which point the N protein was mostly localized in the perinuclear puncta. Larger, tubular structures of N protein assemblies became apparent at 7 days post infection, although no further time points between 36 h and 7 days post infection have been tested [89]. This is in accordance with reports noting the slow growth of hantaviruses [14], and hence, the assembly of N protein and other components into viral factories and recruitment of cellular structures as described in Section 2.2 may be a rate-determining step in the hantavirus replication cycle.

In summary, whether hantavirus assembly occurs at a specific site or at multiple locations remains unclear [55,108–110] as does the role of the cytoskeleton in viral component trafficking.

3. Discussion

The *Bunyavirales* order encompasses a very diverse number of viruses [1]. Members of the *Hantaviridae* family exhibit a number of specific features regarding their host range, also including insectivores and rodents, their direct transmission from persistently infected reservoirs to humans without using invertebrates as vectors, and their limited ability of human-to-human transmission, to name a few. Hantaviruses mostly grow very slowly and produce low titers in cell culture [119,120]. Although this slow viral growth rate seems rather inefficient, it might just be the reason that viral replication is possible at all, by staying below the radar of host's antiviral defenses, but it makes working with these viruses more tedious. Additionally, there is a lack of relevant *in vitro* models [19,56], hampering studies on hantavirus–host interactions. Therefore, many details of the hantavirus replication cycle are still unknown.

First, the cellular receptor(s) used by hantaviruses to attach to cells as well as the interaction between the host receptor and the viral glycoprotein complex are not clear. Structural studies of viral particles by cryo-electron tomography and of isolated glycoproteins by X-ray crystallography greatly improved our mechanistic understanding of the low pH-induced conformational changes involved in membrane fusion [25,37]. Nevertheless, receptor-binding studies were unable to find a common determinant for hantavirus attachment and entry. Several candidates for receptors for attachment and entry have been described, but only protocadherin-1 could be shown to be relevant in all New World hantaviruses (comprehensively reviewed by [19]). Although hantaviruses are amplified in different cell lines, which cell types are initially infected in a natural infection has not been proven, which is in line with the lack of knowledge on the natural receptors used by hantaviruses. Second, a multitude of pathways have been proposed for the uptake of hantaviruses by the cell, including macropinocytosis and endocytosis that is either clathrin-, calveolin- or cholesterol-dependent [19]. Additional research is required to determine

which of them is the road that is mainly used and whether this depends on the cell type, the exact virus species, or both.

Next, the mechanisms of viral replication and transcription also remain poorly understood. Recent structural data show how the hantavirus nucleoproteins can assemble to RNP-like structures in the absence of viral RNA with an impressive complexity of interactions observed between the N protomers and with a positively charged putative RNA binding groove buried inside the filamentous structure [70]. However, there is no structural data on the full-length L protein of hantaviruses, which contains the viral RdRp. The hantavirus L protein is a difficult target for *in vitro* expression due to its size of more than 200 kDa and the fact that its endonuclease domain is highly active, presumably digesting all sorts of RNA within cells, leading to a strong cytopathic effect, hence limiting its own expression [79] and possibly also general virus growth. Recombinant expression, biochemical characterization, and structure determination of the isolated endonuclease domain confirmed this hypothesis, as only significantly less active mutants of the domain could be produced, which were shown to digest a broad spectrum of substrates *in vitro* [77,78].

There are several structures of the full-length bunyavirus L proteins that have been published, such as for arenaviruses [30,121], phenuiviruses [28,83], and peribunyaviruses [26,27,31]. However, it remains unclear if these structures would enable reliable 3D modelling of the complete hantavirus L protein due to the mentioned diversity among bunyaviruses, which is also reflected in the L protein sequences and overall domain architecture. Previous attempts to predict the phenuivirus L protein structure based on the published structures of single domains of related polymerases did not yield entirely reliable models [122] when compared to experimentally determined structures [28,82,83]. However, this was before the game-changing appearance of AlphaFold, a novel structure-prediction tool with great potential [123].

While the endonuclease has been identified in the N-terminal domain of the full-length L protein [77,78], the cap-binding domain was hypothesized to be located in the N protein [80]. The latter is quite unlikely, as neither a canonical cap-binding site in the structure nor a convincing cap-binding activity have been demonstrated for N [69,80], which was reviewed by [85]. The current lack of structural information on the hantavirus L protein hence makes it impossible to propose a conclusive model of hantavirus cap-snatching. Recently, the expansion of the number of potential proteins expressed by bunyaviruses was proposed by a mechanism denoted as “start-snatching”. This process is based on cap-snatching and results from the introduction of a host start codon upstream of the viral open reading frames [124] but needs to be experimentally tested for hantaviruses.

Additional open questions relate to the exact cellular localization of viral genome replication and transcription as well as their regulation. Several studies present contradictory results regarding the replication and transcription site with the involvement of stress granules, P bodies, and different subcellular locations [86,89,100]. Importantly, most of the current studies used the overexpression of the N protein [71,86,100], which likely results in artificial distributions of the protein and the induction of cellular stress. As pointed out above, studies of infected cells are limited by the slow growth rate of hantaviruses and the low expression levels of the viral proteins [100,119,120]. However, further research and, in particular, research on the development of more relevant infection models is needed to answer these questions and to improve our understanding of the key stages of hantavirus infection.

After viral genome replication and transcription, the newly synthesized proteins and viral genomes have to be assembled to form new virions. Whether the genome segments are specifically recruited, as seen in influenza viruses [125], or packaged by statistical likelihood, as proposed for RVFV [126,127], is currently unknown. The cytoplasmic tail of Gc seems to be important for the recruitment of viral genomes into budding virions *via* an interaction with N [115]. Interestingly, the cytoplasmic tail of Gc (comprising 110 residues) has a size that is very similar to that of the Z protein specific to arenaviruses [128], which is not expressed by members of other bunyavirus families. Similar to the zinc-binding

Z protein, the hantavirus Gc cytoplasmic tail was shown to contain a zinc-coordination site [113], and both have the same localization within the virion, lining the inner side of the viral envelope [128].

The cytoskeletal components important for transporting the subunits to the virion assembly site as well as the localization of the assembly site itself are also a matter of debate. The diverse scenarios propose assembly in the Golgi apparatus, the ERGIC, and the plasma membrane [89,100,105,109,110]. Similar to the studies of the N protein, investigation of the glycoprotein localization that is largely involved in the overexpression of the proteins and therefore potential artifacts have to be carefully considered. Whereas the sample preparation procedures for some of the earlier studies were also likely not optimal [55,108], a more recent study presented the same diversity of potential assembly sites [111], leaving it open as to if the assembly sites differ between hantaviruses or cell types or if they are simply dependent on the experimental setup (e.g., overexpression or sample preparation).

4. Conclusions

In summary, despite the enormous progress made in the field over the past years, many key aspects of the hantavirus replication cycle remain poorly understood. In particular, the mechanisms at play during genome replication, transcription, and particle assembly are unclear and are debated in the field. The biggest challenges seem to be the lack of suitable model systems and the slow viral growth rates in cell cultures, impeding detailed studies in the natural context of infection. We are confident that technological advances in the near future will aid in overcoming most of the noted barriers and will provide us with a comprehensive overview of hantavirus diversity and pathogenesis.

Author Contributions: E.R.J.Q. and M.R. prepared the outline; K.M. and S.R.T. made the figures; K.M., S.R.T., E.R.J.Q. and M.R. jointly wrote and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: Research in the MR group is funded by the German Research Foundation (grant number RO 5954/2-1), the German Ministry for Education and Research (grant number 01K12019) as well as the Leibniz Association (Leibniz Competition Programme, grant number K72-2017). We also want to acknowledge funding by the Klaus Tschira Boost Fund to ERJQ (KT 33).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Parts of the figures were created using BioRender.com in June 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Maes, P.; Alkhovsky, S.V.; Bao, Y.; Beer, M.; Birkhead, M.; Briese, T.; Buchmeier, M.J.; Calisher, C.H.; Charrel, R.N.; Choi, I.R.; et al. Taxonomy of the family Arenaviridae and the order Bunyavirales: Update 2018. *Arch. Virol.* **2018**, *163*, 2295–2310. [[CrossRef](#)] [[PubMed](#)]
2. Avsic-Zupanc, T.; Saksida, A.; Korva, M. Hantavirus infections. *Clin. Microbiol. Infect.* **2019**, *21S*, e6–e16. [[CrossRef](#)]
3. Vaheri, A.; Henttonen, H.; Voutilainen, L.; Mustonen, J.; Sironen, T.; Vapalahti, O. Hantavirus infections in Europe and their impact on public health. *Rev. Med. Virol.* **2013**, *23*, 35–49. [[CrossRef](#)]
4. Hjelle, B.; Torres-Perez, F. Hantaviruses in the Americas and their role as emerging pathogens. *Viruses* **2010**, *2*, 2559–2586. [[CrossRef](#)] [[PubMed](#)]
5. Padula, P.J.; Edelstein, A.; Miguel, S.D.L.; Lopez, N.M.; Rossi, C.M.; Rabinovich, R.D. Hantavirus Pulmonary Syndrome Outbreak in Argentina: Molecular Evidence for Person-to-Person Transmission of Andes Virus. *Virology* **1998**. [[CrossRef](#)]
6. Martinez, V.P.; Di Paola, N.; Alonso, D.O.; Perez-Sautu, U.; Bellomo, C.M.; Iglesias, A.A.; Coelho, R.M.; Lopez, B.; Periolo, N.; Larson, P.A.; et al. “Super-Spreaders” and Person-to-Person Transmission of Andes Virus in Argentina. *N. Engl. J. Med.* **2020**, *383*, 2230–2241. [[CrossRef](#)]
7. Sinisalo, M.; Vapalahti, O.; Ekblom-Kullberg, S.; Laine, O.; Makela, S.; Rintala, H.; Vaheri, A. Headache and low platelets in a patient with acute leukemia. *J. Clin. Virol.* **2010**, *48*, 159–161. [[CrossRef](#)]

8. Ramsden, C.; Melo, F.L.; Figueiredo, L.M.; Holmes, E.C.; Zanotto, P.M.; Consortium, V. High rates of molecular evolution in hantaviruses. *Mol. Biol. Evol.* **2008**, *25*, 1488–1492. [[CrossRef](#)] [[PubMed](#)]
9. Ermonval, M.; Baychelier, F.; Tordo, N. What Do We Know about How Hantaviruses Interact with Their Different Hosts? *Viruses* **2016**, *8*, 223. [[CrossRef](#)] [[PubMed](#)]
10. Forbes, K.M.; Sironen, T.; Plyusnin, A. Hantavirus maintenance and transmission in reservoir host populations. *Curr. Opin. Virol.* **2018**, *28*, 1–6. [[CrossRef](#)] [[PubMed](#)]
11. Li, S.; Rissanen, I.; Zeltina, A.; Hepojoki, J.; Raghvani, J.; Harlos, K.; Pybus, O.G.; Huiskonen, J.T.; Bowden, T.A. A Molecular-Level Account of the Antigenic Hantaviral Surface. *Cell Rep.* **2016**, *15*, 959–967. [[CrossRef](#)]
12. Battisti, A.J.; Chu, Y.K.; Chipman, P.R.; Kaufmann, B.; Jonsson, C.B.; Rossmann, M.G. Structural studies of Hantaan virus. *J. Virol.* **2011**, *85*, 835–841. [[CrossRef](#)] [[PubMed](#)]
13. Huiskonen, J.T.; Hepojoki, J.; Laurinmaki, P.; Vaehri, A.; Lankinen, H.; Butcher, S.J.; Grunewald, K. Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses. *J. Virol.* **2010**, *84*, 4889–4897. [[CrossRef](#)]
14. Parvate, A.; Williams, E.P.; Taylor, M.K.; Chu, Y.K.; Lanman, J.; Saphire, E.O.; Jonsson, C.B. Diverse Morphology and Structural Features of Old and New World Hantaviruses. *Viruses* **2019**, *11*, 862. [[CrossRef](#)] [[PubMed](#)]
15. Muyangwa, M.; Martynova, E.V.; Khaiboullina, S.F.; Morzunov, S.P.; Rizvanov, A.A. Hantaviral Proteins: Structure, Functions, and Role in Hantavirus Infection. *Front. Microbiol.* **2015**, *6*, 1326. [[CrossRef](#)] [[PubMed](#)]
16. Vera-Otarola, J.; Solis, L.; Soto-Rifo, R.; Ricci, E.P.; Pino, K.; Tischler, N.D.; Ohlmann, T.; Darlix, J.L.; Lopez-Lastra, M. The Andes hantavirus NSs protein is expressed from the viral small mRNA by a leaky scanning mechanism. *J. Virol.* **2012**, *86*, 2176–2187. [[CrossRef](#)] [[PubMed](#)]
17. Vera-Otarola, J.; Solis, L.; Lowy, F.; Olguin, V.; Angulo, J.; Pino, K.; Tischler, N.D.; Otth, C.; Padula, P.; Lopez-Lastra, M. The Andes Orthohantavirus NSs Protein Antagonizes the Type I Interferon Response by Inhibiting MAVS Signaling. *J. Virol.* **2020**, *94*. [[CrossRef](#)]
18. Ronnberg, T.; Jaaskelainen, K.; Blot, G.; Parviainen, V.; Vaehri, A.; Renkonen, R.; Bouloy, M.; Plyusnin, A. Searching for cellular partners of hantaviral nonstructural protein NSs: Y2H screening of mouse cDNA library and analysis of cellular interactome. *PLoS ONE* **2012**, *7*, e34307. [[CrossRef](#)]
19. Mittler, E.; Dieterle, M.E.; Kleinfelder, L.M.; Slough, M.M.; Chandran, K.; Jangra, R.K. Hantavirus entry: Perspectives and recent advances. *Adv. Virus Res.* **2019**, *104*, 185–224.
20. Hollidge, B.S.; Nedelsky, N.B.; Salzano, M.V.; Fraser, J.W.; Gonzalez-Scarano, F.; Soldan, S.S. Orthobunyavirus entry into neurons and other mammalian cells occurs via clathrin-mediated endocytosis and requires trafficking into early endosomes. *J. Virol.* **2012**, *86*, 7988–8001. [[CrossRef](#)]
21. Simon, M.; Johansson, C.; Lundkvist, A.; Mirazimi, A. Microtubule-dependent and microtubule-independent steps in Crimean-Congo hemorrhagic fever virus replication cycle. *Virology* **2009**, *385*, 313–322. [[CrossRef](#)]
22. Simon, M.; Johansson, C.; Mirazimi, A. Crimean-Congo hemorrhagic fever virus entry and replication is clathrin-, pH- and cholesterol-dependent. *J. Gen. Virol.* **2009**, *90*, 210–215. [[CrossRef](#)] [[PubMed](#)]
23. Garrison, A.R.; Radoshitzky, S.R.; Kota, K.P.; Pegoraro, G.; Ruthel, G.; Kuhn, J.H.; Altamura, L.A.; Kwilas, S.A.; Bavari, S.; Hauke, V.; et al. Crimean-Congo hemorrhagic fever virus utilizes a clathrin- and early endosome-dependent entry pathway. *Virology* **2013**, *444*, 45–54. [[CrossRef](#)]
24. Lozach, P.Y.; Mancini, R.; Bitto, D.; Meier, R.; Oestereich, L.; Overby, A.K.; Pettersson, R.F.; Helenius, A. Entry of bunyaviruses into mammalian cells. *Cell Host Microbe* **2010**, *7*, 488–499. [[CrossRef](#)]
25. Serris, A.; Stass, R.; Bignon, E.A.; Muenz, N.A.; Manuguerra, J.C.; Jangra, R.K.; Li, S.; Chandran, K.; Tischler, N.D.; Huiskonen, J.T.; et al. The Hantavirus Surface Glycoprotein Lattice and Its Fusion Control Mechanism. *Cell* **2020**, *183*, 442–456.e16. [[CrossRef](#)] [[PubMed](#)]
26. Gerlach, P.; Malet, H.; Cusack, S.; Reguera, J. Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter. *Cell* **2015**, *161*, 1267–1279. [[CrossRef](#)]
27. Arragain, B.; Effantin, G.; Gerlach, P.; Reguera, J.; Schoehn, G.; Cusack, S.; Malet, H. Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes. *Nat. Commun.* **2020**, *11*, 3590. [[CrossRef](#)] [[PubMed](#)]
28. Wang, P.; Liu, L.; Liu, A.; Yan, L.; He, Y.; Shen, S.; Hu, M.; Guo, Y.; Liu, H.; Liu, C.; et al. Structure of severe fever with thrombocytopenia syndrome virus L protein elucidates the mechanisms of viral transcription initiation. *Nat. Microbiol.* **2020**, *5*, 864–871. [[CrossRef](#)]
29. Li, M.-L.; Ramirez, B.C.; Krug, R.M. RNA-dependent activation of primer RNA production by influenza virus polymerase: Different regions of the same protein subunit constitute the two required RNA-binding sites. *Embo J.* **1998**, *17*, 5844–5852. [[CrossRef](#)]
30. Kouba, T.; Vogel, D.; Thorkelsson, S.R.; Quemin, E.R.J.; Williams, H.M.; Milewski, M.; Busch, C.; Günther, S.; Grunewald, K.; Rosenthal, M.; et al. Conformational changes in Lassa virus L protein associated with promoter binding and RNA synthesis activity. *BioRxiv* **2021**. [[CrossRef](#)]
31. Arragain, B.; Trouilleton, Q.D.; Baudin, F.; Cusack, S.; Schoehn, G.; Malet, H. Structural snapshots of La Crosse virus polymerase reveal the mechanisms underlying Peribunyaviridae replication and transcription. *BioRxiv* **2021**. [[CrossRef](#)]

32. Cifuentes-Munoz, N.; Salazar-Quiroz, N.; Tischler, N.D. Hantavirus Gn and Gc envelope glycoproteins: Key structural units for virus cell entry and virus assembly. *Viruses* **2014**, *6*, 1801–1822. [[CrossRef](#)] [[PubMed](#)]
33. Freiberg, A.N.; Sherman, M.B.; Morais, M.C.; Holbrook, M.R.; Watowich, S.J. Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography. *J. Virol.* **2008**, *82*, 10341–10348. [[CrossRef](#)] [[PubMed](#)]
34. Overby, A.K.; Pettersson, R.F.; Grunewald, K.; Huiskonen, J.T. Insights into bunyavirus architecture from electron cryotomography of Uukuniemi virus. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2375–2379. [[CrossRef](#)]
35. Huiskonen, J.T.; Overby, A.K.; Weber, F.; Grunewald, K. Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: Evidence for GN-GC glycoprotein heterodimers. *J. Virol.* **2009**, *83*, 3762–3769. [[CrossRef](#)] [[PubMed](#)]
36. Martin, M.L.; Lindsey-Regnery, H.; Sasso, D.R.; McCormick, J.B.; Palmer, E. Distinction between Bunyaviridae genera by surface structure and comparison with Hantaan virus using negative stain electron microscopy. *Arch. Virol.* **1985**, *86*, 17–28. [[CrossRef](#)]
37. Guardado-Calvo, P.; Bignon, E.A.; Stettner, E.; Jeffers, S.A.; Perez-Vargas, J.; Pehau-Arnaudet, G.; Tortorici, M.A.; Jestin, J.L.; England, P.; Tischler, N.D.; et al. Mechanistic Insight into Bunyavirus-Induced Membrane Fusion from Structure-Function Analyses of the Hantavirus Envelope Glycoprotein Gc. *PLoS Pathog.* **2016**, *12*, e1005813. [[CrossRef](#)] [[PubMed](#)]
38. Bignon, E.A.; Albornoz, A.; Guardado-Calvo, P.; Rey, F.A.; Tischler, N.D. Molecular organization and dynamics of the fusion protein Gc at the hantavirus surface. *eLife* **2019**, *8*, e46028. [[CrossRef](#)]
39. Hepojoki, J.; Strandin, T.; Wang, H.; Vapalahti, O.; Vaheri, A.; Lankinen, H. Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein. *J. Gen. Virol.* **2010**, *91*, 2341–2350. [[CrossRef](#)]
40. Rissanen, I.; Stass, R.; Zeltina, A.; Li, S.; Hepojoki, J.; Harlos, K.; Gilbert, R.J.C.; Huiskonen, J.T.; Bowden, T.A.; Sundquist, W.I. Structural Transitions of the Conserved and Metastable Hantaviral Glycoprotein Envelope. *J. Virol.* **2017**, *91*. [[CrossRef](#)]
41. Gavrilovskaya, I.N.; Shepley, M.; Shaw, R.; Ginsberg, M.H.; Mackow, E.R. beta3 Integrins mediate the cellular entry of hantaviruses that cause respiratory failure. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7074–7079. [[CrossRef](#)]
42. Gavrilovskaya, I.N.; Brown, E.J.; Ginsberg, M.H.; Mackow, E.R. Cellular entry of hantaviruses which cause hemorrhagic fever with renal syndrome is mediated by beta3 integrins. *J. Virol.* **1999**, *73*, 3951–3959. [[CrossRef](#)]
43. Mou, D.L.; Wang, Y.P.; Huang, C.X.; Li, G.Y.; Pan, L.; Yang, W.S.; Bai, X.F. Cellular entry of Hantaan virus A9 strain: Specific interactions with beta3 integrins and a novel 70kDa protein. *Biochem. Biophys. Res. Commun.* **2006**, *339*, 611–617. [[CrossRef](#)] [[PubMed](#)]
44. Klempa, B.; Witkowski, P.T.; Popugaeva, E.; Auste, B.; Koivogui, L.; Fichet-Calvet, E.; Strecker, T.; Ter Meulen, J.; Kruger, D.H. Sangassou virus, the first hantavirus isolate from Africa, displays genetic and functional properties distinct from those of other murinae-associated hantaviruses. *J. Virol.* **2012**, *86*, 3819–3827. [[CrossRef](#)] [[PubMed](#)]
45. Jangra, R.K.; Herbert, A.S.; Li, R.; Jae, L.T.; Kleinfelder, L.M.; Slough, M.M.; Barker, S.L.; Guardado-Calvo, P.; Roman-Sosa, G.; Dieterle, M.E.; et al. Protocadherin-1 is essential for cell entry by New World hantaviruses. *Nature* **2018**, *563*, 559–563. [[CrossRef](#)]
46. Krautkramer, E.; Zeier, M. Hantavirus causing hemorrhagic fever with renal syndrome enters from the apical surface and requires decay-accelerating factor (DAF/CD55). *J. Virol.* **2008**, *82*, 4257–4264. [[CrossRef](#)] [[PubMed](#)]
47. Choi, Y.; Kwon, Y.C.; Kim, S.I.; Park, J.M.; Lee, K.H.; Ahn, B.Y. A hantavirus causing hemorrhagic fever with renal syndrome requires gC1qR/p32 for efficient cell binding and infection. *Virology* **2008**, *381*, 178–183. [[CrossRef](#)] [[PubMed](#)]
48. Kim, T.Y.; Choi, Y.; Cheong, H.S.; Choe, J. Identification of a cell surface 30 kDa protein as a candidate receptor for Hantaan virus. *J. Gen. Virol.* **2002**, *83*, 767–773. [[CrossRef](#)] [[PubMed](#)]
49. Kleinfelder, L.M.; Jangra, R.K.; Jae, L.T.; Herbert, A.S.; Mittler, E.; Stiles, K.M.; Wirchnianski, A.S.; Kielian, M.; Brummelkamp, T.R.; Dye, J.M.; et al. Haploid Genetic Screen Reveals a Profound and Direct Dependence on Cholesterol for Hantavirus Membrane Fusion. *mBio* **2015**, *6*, e00801. [[CrossRef](#)]
50. Cifuentes-Munoz, N.; Darlix, J.L.; Tischler, N.D. Development of a lentiviral vector system to study the role of the Andes virus glycoproteins. *Virus Res.* **2010**, *153*, 29–35. [[CrossRef](#)]
51. Chiang, C.F.; Flint, M.; Lin, J.S.; Spiropoulou, C.F. Endocytic Pathways Used by Andes Virus to Enter Primary Human Lung Endothelial Cells. *PLoS ONE* **2016**, *11*, e0164768. [[CrossRef](#)]
52. Torriani, G.; Mayor, J.; Zimmer, G.; Kunz, S.; Rothenberger, S.; Engler, O. Macropinocytosis contributes to hantavirus entry into human airway epithelial cells. *Virology* **2019**, *531*, 57–68. [[CrossRef](#)] [[PubMed](#)]
53. Bauherr, S.; Larsberg, F.; Petrich, A.; Sperber, H.S.; Klose-Grzelka, V.; Luckner, M.; Azab, W.; Schade, M.; Hofer, C.T.; Lehmann, M.J.; et al. Macropinocytosis and Clathrin-Dependent Endocytosis Play Pivotal Roles for the Infectious Entry of Puumala Virus. *J. Virol.* **2020**, *94*. [[CrossRef](#)]
54. Jin, M.; Park, J.; Lee, S.; Park, B.; Shin, J.; Song, K.J.; Ahn, T.I.; Hwang, S.Y.; Ahn, B.Y.; Ahn, K. Hantaan virus enters cells by clathrin-dependent receptor-mediated endocytosis. *Virology* **2002**, *294*, 60–69. [[CrossRef](#)] [[PubMed](#)]
55. Ravkov, E.V.; Nichol, S.T.; Compans, R.W. Polarized entry and release in epithelial cells of Black Creek Canal virus, a New World hantavirus. *J. Virol.* **1997**, *71*, 1147–1154. [[CrossRef](#)]
56. Eckerle, I.; Lenk, M.; Ulrich, R.G. More novel hantaviruses and diversifying reservoir hosts—time for development of reservoir-derived cell culture models? *Viruses* **2014**, *6*, 951–967. [[CrossRef](#)] [[PubMed](#)]
57. Modis, Y. Class II fusion proteins. *Adv. Exp. Med. Biol.* **2013**, *790*, 150–166. [[PubMed](#)]
58. Willensky, S.; Bar-Rogovsky, H.; Bignon, E.A.; Tischler, N.D.; Modis, Y.; Dessau, M. Crystal Structure of Glycoprotein C from a Hantavirus in the Post-fusion Conformation. *PLoS Pathog.* **2016**, *12*, e1005948. [[CrossRef](#)]

59. Garcin, D.; Lezzi, M.; Dobbs, M.; Elliott, R.M.; Schmaljohn, C.; Kang, C.Y.; Kolakofsky, D. The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. *J. Virol.* **1995**, *69*, 5754–5762. [[CrossRef](#)]
60. te Velthuis, A.J. Common and unique features of viral RNA-dependent polymerases. *Cell Mol. Life Sci.* **2014**, *71*, 4403–4420. [[CrossRef](#)] [[PubMed](#)]
61. Hacker, D.; Raju, R.; Kolakofsky, D. La Crosse Virus Nucleocapsid Protein Controls Its Own Synthesis in Mosquito Cells by Encapsidating Its mRNA. *J. Virol.* **1989**, *63*, 5166–5174. [[CrossRef](#)]
62. Obijeski, J.F.; Bishop, D.H.L.; Palmer, E.L.; Murphy, F.A. Segmented Genome and Nucleocapsid of La Crosse Virus. *J. Virol.* **1976**, *20*, 664–675. [[CrossRef](#)]
63. Hewlett, M.J.; Pettersson, R.F.; Baltimore, D. Circular Forms of Uukuniemi Virion RNA: An Electron Microscopic Study. *J. Virol.* **1977**, *21*. [[CrossRef](#)]
64. Hepojoki, J.; Strandin, T.; Lankinen, H.; Vaheri, A. Hantavirus structure–molecular interactions behind the scene. *J. Gen. Virol.* **2012**, *93*, 1631–1644. [[CrossRef](#)]
65. Severson, W.E.; Xu, X.; Jonsson, C.B. cis-Acting signals in encapsidation of Hantaan virus S-segment viral genomic RNA by its N protein. *J. Virol.* **2001**, *75*, 2646–2652. [[CrossRef](#)] [[PubMed](#)]
66. Gött, P.; Stohwasser, R.; Schnitzler, P.; Darai, G.; Bautz, E.K.F. RNA Binding of Recombinant Nucleocapsid Proteins of Hantaviruses. *Virology* **1993**, *194*. [[CrossRef](#)] [[PubMed](#)]
67. Mir, M.A.; Panganiban, A.T. Trimeric Hantavirus Nucleocapsid Protein Binds Specifically to the Viral RNA Panhandle. *J. Virol.* **2004**, *78*, 8281–8288. [[CrossRef](#)] [[PubMed](#)]
68. Guo, Y.; Wang, W.; Sun, Y.; Ma, C.; Wang, X.; Wang, X.; Liu, P.; Shen, S.; Li, B.; Lin, J.; et al. Crystal Structure of the Core Region of Hantavirus Nucleocapsid Protein Reveals the Mechanism for Ribonucleoprotein Complex Formation. *J. Virol.* **2016**, *90*, 1048–1061. [[CrossRef](#)]
69. Olal, D.; Daumke, O. Structure of the Hantavirus Nucleoprotein Provides Insights into the Mechanism of RNA Encapsidation. *Cell Rep.* **2016**, *14*, 2092–2099. [[CrossRef](#)]
70. Arragain, B.; Reguera, J.; Desfosses, A.; Gutsche, I.; Schoehn, G.; Malet, H. High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms. *eLife* **2019**, *8*. [[CrossRef](#)]
71. Cheng, E.; Wang, Z.; Mir, M.A. 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. *J. Virol.* **2014**, *88*, 8706–8712. [[CrossRef](#)] [[PubMed](#)]
72. Plotch, S.J.; Bouloy, M.; Ulmanen, I.; Krug, R.M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **1981**, *23*, 847–858. [[CrossRef](#)]
73. Engelhardt, O.G.; Smith, M.; Fodor, E. Association of the Influenza A Virus RNA-Dependent RNA Polymerase with Cellular RNA Polymerase II. *J. Virol.* **2005**, *79*, 5812–5818. [[CrossRef](#)] [[PubMed](#)]
74. Reich, S.; Guilligay, D.; Pflug, A.; Malet, H.; Berger, I.; Crepin, T.; Hart, D.; Lunardi, T.; Nanao, M.; Ruigrok, R.W.; et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **2014**, *516*, 361–366. [[CrossRef](#)] [[PubMed](#)]
75. Jackson, D.A.; Caton, A.J.; McCready, S.J.; Cook, P.R. Influenza virus RNA is synthesized at fixed sites in the nucleus. *Nature* **1982**, *296*, 366–368. [[CrossRef](#)] [[PubMed](#)]
76. Rossier, C.; Patterson, J.; Kolakofsky, D. La Crosse Virus Small Genome mRNA Is Made in the Cytoplasm. *J. Virol.* **1986**, *59*, 647–650. [[CrossRef](#)]
77. Fernández-García, Y.; Reguera, J.; Busch, C.; Witte, G.; Sánchez-Ramos, O.; Betzel, C.; Cusack, S.; Günther, S.; Reindl, S. Atomic Structure and Biochemical Characterization of an RNA Endonuclease in the N Terminus of Andes Virus L Protein. *PLoS Pathog.* **2016**, *12*, e1005635. [[CrossRef](#)] [[PubMed](#)]
78. Reguera, J.; Gerlach, P.; Rosenthal, M.; Gaudon, S.; Coscia, F.; Günther, S.; Cusack, S. Comparative Structural and Functional Analysis of Bunyavirus and Arenavirus Cap-Snatching Endonucleases. *PLoS Pathog.* **2016**, *12*, e1005636. [[CrossRef](#)]
79. Heinemann, P.; Schmidt-Chanasit, J.; Gunther, S. The N terminus of Andes virus L protein suppresses mRNA and protein expression in mammalian cells. *J. Virol.* **2013**, *87*, 6975–6985. [[CrossRef](#)]
80. Mir, M.A.; Sheema, S.; Haseeb, A.; Haque, A. Hantavirus Nucleocapsid Protein Has Distinct m7G Cap- and RNA-binding Sites. *J. Biol. Chem.* **2010**, *285*, 11357–11368. [[CrossRef](#)]
81. Rosenthal, M.; Gogrefe, N.; Vogel, D.; Reguera, J.; Rauschenberger, B.; Cusack, S.; Gunther, S.; Reindl, S. Structural insights into reptarenavirus cap-snatching machinery. *PLoS Pathog.* **2017**, *13*, e1006400. [[CrossRef](#)]
82. Gogrefe, N.; Reindl, S.; Gunther, S.; Rosenthal, M. Structure of a functional cap-binding domain in Rift Valley fever virus L protein. *PLoS Pathog.* **2019**, *15*, e1007829. [[CrossRef](#)] [[PubMed](#)]
83. Vogel, D.; Thorkelsson, S.R.; Quemin, E.R.J.; Meier, K.; Kouba, T.; Gogrefe, N.; Busch, C.; Reindl, S.; Gunther, S.; Cusack, S.; et al. Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L protein. *Nucleic Acids Res.* **2020**, *48*, 5749–5765. [[CrossRef](#)] [[PubMed](#)]
84. Lukarska, M.; Fournier, G.; Pflug, A.; Resa-Infante, P.; Reich, S.; Naffakh, N.; Cusack, S. Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* **2016**, *541*, 117–121. [[CrossRef](#)] [[PubMed](#)]
85. Olschewski, S.; Cusack, S.; Rosenthal, M. The Cap-Snatching Mechanism of Bunyaviruses. *Trends Microbiol.* **2020**, *28*, 293–303. [[CrossRef](#)]

86. Mir, M.A.; Duran, W.A.; Hjelle, B.L.; Ye, C.; Panganiban, A.T. Storage of cellular 5' mRNA caps in P bodies for viral cap-snatching. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 19294–19299. [[CrossRef](#)]
87. Cheng, E.; Mir, M.A. Signatures of host mRNA 5' terminus for efficient hantavirus cap snatching. *J. Virol.* **2012**, *86*, 10173–10185. [[CrossRef](#)]
88. Hopkins, K.C.; McLane, L.M.; Maqbool, T.; Panda, D.; Gordesky-Gold, B.; Cherry, S. A genome-wide RNAi screen reveals that mRNA decapping restricts bunyaviral replication by limiting the pools of Dcp2-accessible targets for cap-snatching. *Genes Dev.* **2013**, *27*, 1511–1525. [[CrossRef](#)] [[PubMed](#)]
89. Davies, K.A.; Chadwick, B.; Hewson, R.; Fontana, J.; Mankouri, J.; Barr, J.N. The RNA Replication Site of Tula Orthohantavirus Resides within a Remodelled Golgi Network. *Cells* **2020**, *9*, 1569. [[CrossRef](#)]
90. Kedersha, N.; Chen, S.; Gilks, N.; Li, W.; Miller, I.J.; Stahl, J.; Anderson, P. Evidence That Ternary Complex (eIF2-GTP-tRNA^{iMet})–Deficient Preinitiation Complexes Are Core Constituents of Mammalian Stress Granules. *Mol. Biol. Cell* **2002**, *13*, 195–210. [[CrossRef](#)]
91. Kedersha, N.; Stoeklin, G.; Ayodele, M.; Yacono, P.; Lykke-Andersen, J.; Fritzler, M.J.; Scheuner, D.; Kaufman, R.J.; Golan, D.E.; Anderson, P. Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.* **2005**, *169*, 871–884. [[CrossRef](#)]
92. Youn, J.Y.; Dyakov, B.J.A.; Zhang, J.; Knight, J.D.R.; Vernon, R.M.; Forman-Kay, J.D.; Gingras, A.C. Properties of Stress Granule and P-Body Proteomes. *Mol. Cell* **2019**, *76*, 286–294. [[CrossRef](#)] [[PubMed](#)]
93. Tian, S.; Curnutte, H.A.; Trcek, T. RNA Granules: A View from the RNA Perspective. *Molecules* **2020**, *25*, 3130. [[CrossRef](#)] [[PubMed](#)]
94. Thomas, M.G.; Loschi, M.; Desbats, M.A.; Boccaccio, G.L. RNA granules: The good, the bad and the ugly. *Cell Signal.* **2011**, *23*, 324–334. [[CrossRef](#)] [[PubMed](#)]
95. Hussein, I.T.M.; Mir, M.A. How hantaviruses modulate pathways for efficient replication? *Front. Biosci.* **2013**, *5*, 154–166.
96. Mir, M.A.; Panganiban, A.T. A protein that replaces the entire cellular eIF4F complex. *Embo J.* **2008**, *27*, 3129–3139. [[CrossRef](#)]
97. Suzuki, C.; Garces, R.G.; Edmonds, K.A.; Hiller, S.; Hyberts, S.G.; Marintchev, A.; Wagner, G. PDCD4 inhibits translation initiation by binding to eIF4A using both its MA3 domains. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3274–3279. [[CrossRef](#)]
98. Chang, J.H.; Cho, Y.H.; Sohn, S.Y.; Choi, J.M.; Kim, A.; Kim, Y.C.; Jang, S.K.; Cho, Y. Crystal structure of the eIF4A-PDCD4 complex. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3148–3153. [[CrossRef](#)]
99. Barr, J.N. Bunyavirus mRNA synthesis is coupled to translation to prevent premature transcription termination. *Rna* **2007**, *13*, 731–736. [[CrossRef](#)]
100. Ramanathan, H.N.; Chung, D.-H.; Plane, S.J.; Sztul, E.; Chu, Y.-k.; Guttieri, M.C.; McDowell, M.; Ali, G.; Jonsson, C.B. Dynein-Dependent Transport of the Hantaan Virus Nucleocapsid Protein to the Endoplasmic Reticulum-Golgi Intermediate Compartment. *J. Virol.* **2007**, *81*, 8634–8647. [[CrossRef](#)]
101. Lober, C.; Anheier, B.; Lindow, S.; Klenk, H.D.; Feldmann, H. The Hantaan virus glycoprotein precursor is cleaved at the conserved pentapeptide WAASA. *Virology* **2001**, *289*, 224–229. [[CrossRef](#)]
102. Petazzi, R.A.; Koikkarah, A.A.; Tischler, N.D.; Chiantia, S. Detection of Envelope Glycoprotein Assembly from Old-World Hantaviruses in the Golgi Apparatus of Living Cells. *J. Virol.* **2020**. [[CrossRef](#)]
103. Deyde, V.M.; Rizvanov, A.A.; Chase, J.; Otteson, E.W.; St Jeor, S.C. Interactions and trafficking of Andes and Sin Nombre Hantavirus glycoproteins G1 and G2. *Virology* **2005**, *331*, 307–315. [[CrossRef](#)] [[PubMed](#)]
104. Ruusala, A.; Persson, R.; Schmaljohn, C.S.; Pettersson, R.F. Coexpression of the Membrane Glycoproteins G1 and G2 of Hantaan Virus Is Required for Targeting to the Golgi Complex. *Virology* **1992**, *186*. [[CrossRef](#)]
105. Spiropoulou, C.F.; Goldsmith, C.S.; Shoemaker, T.R.; Peters, C.J.; Compans, R.W. Sin Nombre virus glycoprotein trafficking. *Virology* **2003**, *308*, 48–63. [[CrossRef](#)]
106. Sperber, H.S.; Welke, R.W.; Petazzi, R.A.; Bergmann, R.; Schade, M.; Shai, Y.; Chiantia, S.; Herrmann, A.; Schwarzer, R. Self-association and subcellular localization of Puumala hantavirus envelope proteins. *Sci. Rep.* **2019**, *9*, 707. [[CrossRef](#)]
107. Jääntti, J.; Hilden, P.; Rönkä, H.; Mäkiranta, V.; Keränen, S.; Kuismanen, E. Immunocytochemical Analysis of Uukuniemi Virus Budding Compartments: Role of the Intermediate Compartment and the Golgi Stack in Virus Maturation. *J. Virol.* **1997**, *71*, 1162–1172. [[CrossRef](#)]
108. Goldsmith, C.S.; Elliott, L.H.; Peters, C.J.; Zaki, S.R. Ultrastructural characteristics of Sin Nombre virus, causative agent of hantavirus pulmonary syndrome. *Arch. Virol.* **1995**, *140*, 2107–2122. [[CrossRef](#)] [[PubMed](#)]
109. Sheriff, Z.; Greer, P.W.; Coffield, L.M.; Goldsmith, C.S.; Nolte, K.B.; Foucar, K.; Feddersen, R.M.; Zumwalt, R.E.; Miller, G.L.; Khan, A.S.; et al. Hantavirus Pulmonary Syndrome Pathogenesis of an Emerging Infectious Disease. *Am. J. Pathol.* **1995**, *146*, 552–579.
110. Xu, F.; Yang, Z.; Wang, L.; Lee, Y.L.; Yang, C.C.; Xiao, S.Y.; Xiao, H.; Wen, L. Morphological characterization of hantavirus HV114 by electron microscopy. *Intervirology* **2007**, *50*, 166–172. [[CrossRef](#)]
111. Parvate, A.; Sengupta, R.; Williams, E.P.; Xue, Y.; Chu, Y.K.; Stahelin, R.V.; Jonsson, C.B. Cryofixation of Inactivated Hantavirus-Infected Cells as a Method for Obtaining High-Quality Ultrastructural Preservation for Electron Microscopic Studies. *Front. Cell Infect. Microbiol.* **2020**, *10*, 580339. [[CrossRef](#)] [[PubMed](#)]
112. Acuna, R.; Cifuentes-Munoz, N.; Marquez, C.L.; Bulling, M.; Klingstrom, J.; Mancini, R.; Lozach, P.Y.; Tischler, N.D. Hantavirus Gn and Gc glycoproteins self-assemble into virus-like particles. *J. Virol.* **2014**, *88*, 2344–2348. [[CrossRef](#)] [[PubMed](#)]

113. Estrada, D.F.; Boudreaux, D.M.; Zhong, D.; St Jeor, S.C.; De Guzman, R.N. The Hantavirus Glycoprotein G1 Tail Contains Dual CCHC-type Classical Zinc Fingers. *J. Biol. Chem.* **2009**, *284*, 8654–8660. [[CrossRef](#)]
114. Estrada, D.F.; Conner, M.; Jeor, S.C.; Guzman, R.N. The Structure of the Hantavirus Zinc Finger Domain is Conserved and Represents the Only Natively Folded Region of the Gn Cytoplasmic Tail. *Front. Microbiol.* **2011**, *2*, 251. [[CrossRef](#)]
115. Betenbaugh, M.; Yu, M.; Kuehl, K.; White, J.; Pennock, D.; Spik, K.; Schmaljohn, C. Nucleocapsid- and virus-like particles assemble in cells infected with recombinant baculoviruses or vaccinia viruses expressing the M and the S segments of Hantaan virus. *Virus Res.* **1995**, *38*, 111–124. [[CrossRef](#)]
116. Kariwa, H.; Tanabe, H.; Mizutani, T.; Kon, Y.; Lokugamage, K.; Lokugamage, N.; Iwasa, M.A.; Hagiya, T.; Araki, K.; Yoshimatsu, K.; et al. Synthesis of Seoul virus RNA and structural proteins in cultured cells. *Arch. Virol.* **2003**, *148*, 1671–1685. [[CrossRef](#)]
117. Rowe, R.K.; Suszko, J.W.; Pekosz, A. Roles for the recycling endosome, Rab8, and Rab11 in hantavirus release from epithelial cells. *Virology* **2008**, *382*, 239–249. [[CrossRef](#)] [[PubMed](#)]
118. Ravkov, E.V.; Nichol, S.T.; Peters, C.J.; Compans, R.W. Role of actin microfilaments in Black Creek Canal virus morphogenesis. *J. Virol.* **1998**, *72*, 2865–2870. [[CrossRef](#)]
119. Kukkonen, S.K.J.; Vaheri, A.; Plyusnin, A. Completion of the Tula hantavirus genome sequence: Properties of the L segment and heterogeneity found in the 3′ termini of S and L genome RNAs. *J. Gen. Virol.* **1998**. [[CrossRef](#)]
120. Meyer, B.J.; Schmaljohn, C. Accumulation of Terminally Deleted RNAs May Play a Role in Seoul Virus Persistence. *J. Virol.* **2000**, *74*, 1321–1331. [[CrossRef](#)]
121. Peng, R.; Xu, X.; Jing, J.; Wang, M.; Peng, Q.; Liu, S.; Wu, Y.; Bao, X.; Wang, P.; Qi, J.; et al. Structural insight into arenavirus replication machinery. *Nature* **2020**, *579*, 615–619. [[CrossRef](#)]
122. Gogovi, G.K.; Almsned, F.; Bracci, N.; Kehn-Hall, K.; Shehu, A.; Blaisten-Barojas, E. Modeling the Tertiary Structure of the Rift Valley Fever Virus L Protein. *Molecules* **2019**, *24*, 1768. [[CrossRef](#)] [[PubMed](#)]
123. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, 1–11. [[CrossRef](#)]
124. Ho, J.S.Y.; Angel, M.; Ma, Y.; Sloan, E.; Wang, G.; Martinez-Romero, C.; Alenquer, M.; Roudko, V.; Chung, L.; Zheng, S.; et al. Hybrid Gene Origination Creates Human-Virus Chimeric Proteins during Infection. *Cell* **2020**, *181*, 1502–1517.e23. [[CrossRef](#)]
125. Dadonaite, B.; Gilbertson, B.; Knight, M.L.; Trifkovic, S.; Rockman, S.; Laederach, A.; Brown, L.E.; Fodor, E.; Bauer, D.L.V. The structure of the influenza A virus genome. *Nat. Microbiol.* **2019**, *4*, 1781–1789. [[CrossRef](#)] [[PubMed](#)]
126. Wichgers Schreur, P.J.; Kortekaas, J. Single-Molecule FISH Reveals Non-selective Packaging of Rift Valley Fever Virus Genome Segments. *PLoS Pathog.* **2016**, *12*, e1005800. [[CrossRef](#)] [[PubMed](#)]
127. Bermudez-Mendez, E.; Katrukha, E.A.; Spruit, C.M.; Kortekaas, J.; Wichgers Schreur, P.J. Visualizing the ribonucleoprotein content of single bunyavirus virions reveals more efficient genome packaging in the arthropod host. *Commun. Biol.* **2021**, *4*, 345. [[CrossRef](#)] [[PubMed](#)]
128. Fehling, S.K.; Lennartz, F.; Strecker, T. Multifunctional nature of the arenavirus RING finger protein Z. *Viruses* **2012**, *4*, 2973–3011. [[CrossRef](#)]

5. Own contribution to publications

I, Sigurdur Thorkelsson, carried out the sample preparation for cryoEM, data collection and image processing as well as design of research experiments, execution of the experiments, evaluation and interpretation of the data with supervisors and collaborators, and finally, participated in the preparation of the publications with co-authors.

6. Results and discussion

6.1 Overall L-protein structure

Before I started my PhD, only bunyavirus L-protein fragments (Reguera et al., 2010) or truncated versions of the L-protein, omitting the flexible C-terminal domain (Gerlach et al., 2015), had been structurally characterized. Based on previous biochemical characterization, there was evidence for the presence of an endonuclease enzymatic domain in the L-protein from orthobunyaviruses (Reguera et al., 2010), arenaviruses (Morin et al., 2010) and hantaviruses (Reguera et al., 2016). However, the cap-binding domain and full-length protein had eluded structural characterization by X-ray crystallography or NMR due to its relatively large size, complex substrate interactions or flexibility. With the recent advances in cryoEM, including direct electron detectors, electron microscope automation, computing infrastructure and image processing software, the structure and function of the L-protein of bunyaviruses could be revisited. Using cryoEM SPA, we were now able to study L-proteins in full-length and functionally relevant stages to improve our understanding of the molecular mechanism of this Swiss army knife of nsRNA virus infections.

During my PhD, I took advantage of reactions based on non-hydrolyzable NTPs and modified genomic viral RNAs set-up and characterized in the lab of Maria Rosenthal that we optimized for cryoEM SPA. With this, we could produce multiple structures of L-proteins at high resolution (<4 Å) from both SFTSV (*Phenuiviridae*) (Vogel et al., 2020; Williams et al., 2022) and LASV (*Arenaviridae*) (Kouba et al., 2021) in a variety of functional states. Complementary to the structural analysis I performed, the lab of Maria Rosenthal has shown biochemically that the L-proteins from SFTSV and LASV are active RdRps capable of both *de novo* replication and primed transcription of modified genomic viral RNA *in vitro*. Interestingly, the L-protein could also prime transcription *in vitro* when incubated with RNA primers, independently of a 5' cap (Kouba et al., 2021; Vogel et al., 2020). In the case of SFTSV, the L-protein structure

could be resolved with an occupied CBD showing how the capped 5' mRNA end is sandwiched inside the domain and threaded towards the RDRP active site (Williams et al., 2023).

At the structural level, these proteins display a remarkable overall similarity compared with the currently available bunyavirus L-protein structures, namely that of La Crosse virus (*Peribunyaviridae*) (Arragain et al., 2020)(Arragain et al., 2022), or the Influenza virus polymerase complex (Kouba et al., 2019)(Fan et al., 2019)(Wandzik et al., 2021) (Fig.3). In the meantime, other studies of L-protein structures describing different conformations and interactions with the viral Z protein have been published (Kouba et al., 2021; Vogel et al., 2020)(P. Wang et al., 2021)(Arragain et al., 2020, 2022)(X. Wang et al., 2022)(Peng et al., 2020)(Kang et al., 2021)(Ma et al., 2021). Altogether with the results of my PhD work, the wealth of data now available allow for significant and detailed insight into the molecular mechanism of bunyavirus L-protein replication, transcription and interactions with some partners.

6.2 The RdRp core

RdRps have structural homology even with divergence in sequence (Venkataraman et al., 2018). The RdRp is the most highly conserved domain in the L-protein making it an opportune target for antivirals (ter Horst et al., 2019). The RdRp core of L-proteins have a signature cupped right hand architecture of RdRp with three subdomains termed palm, fingers and thumb (te Velthuis et al., 2021). In the palm subdomain of the L-protein, there are the conserved polymerase motifs A and B, which form the NTP entry tunnel and facilitate NTP selectivity with motif H supporting the structure of motif B. While motif C stabilizes metal ion binding in the polymerase core, motif D supports the recruitment of NTPs complementary to the template and motif E, serves as a hinge for the thumb domain movement and participates in primer 3' end positioning. In the polymerase fingers is motif F, which coordinates the binding of the 5'

end of the viral genomic RNA. Then there is motif G in the polymerase core lobe, which supports priming of incoming NTPs (Kouba et al., 2021; Vogel et al., 2020).

For bunyaviruses, the L-protein requires a viral RNA template which is recognized by conserved complementary sequences, specific to each viral family at the termini of the viral genome RNA segments S, M, and L (Ferron et al., 2017). A product is synthesized by the RdRp in a process that requires divalent metal ions such as magnesium or manganese (Kouba et al., 2021; Vogel et al., 2020).

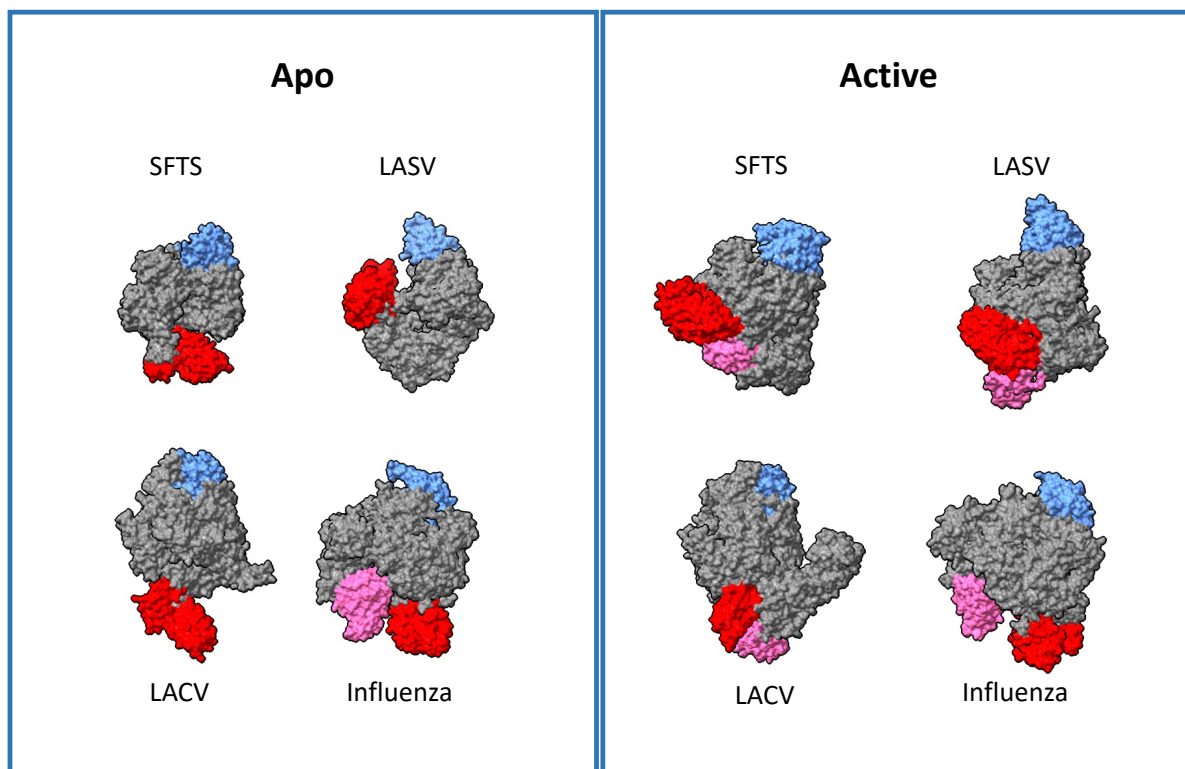


Figure 3. A comparison of L-proteins from SFTS, LASV, LACV and the trimeric Influenza polymerase complex in apo and active state with the endonuclease in red, cap-binding domain in pink and RNA-binding lobe coloured teal.

6.3 RNA interaction

Both LASV (arenavirus) and SFTSV (phenuivirus) L-proteins have an RNA binding domain (RBD) which function is to attract and guide viral gRNA and cRNA towards the RdRp via specific interaction with the 5' and 3' end sequences. The 5' end is guided towards a primary RNA binding site where the 5' forms a hook-like structure with a single base pair in both LASV

and SFTSV while the 3' is initially guided towards the active site for replication (Kouba et al., 2021; Williams et al., 2022). When the 3' end has been copied it exits the active site and is threaded into a secondary RNA binding site adjacent to the RBD while elongation continues (Kouba et al., 2021; Williams et al., 2022).

A noticeable difference in the architecture the RBDs of LASV and SFTSV is zinc-binding site in LASV and a conformational twist of the RBD in the LASV L-protein when bound to viral RNA while SFTSV RBD stays rigid between the apo and RNA-bound conformations. The RBD recognizes specific sequences at the 5' and 3' end of the viral genome. During pre-initiation, the 3' end is guided towards the active site of the RdRp while during elongation, it is threaded into a secondary binding site. In the meantime, the 5' end sits in a peripheral groove of the RdRp core forming a hook-like structure, stabilized by a single base pair for both LASV (Kouba et al., 2021; Williams et al., 2022). Some loops in the RBD that are disordered in the apo conformation, become ordered when bound to viral RNA like the 5' binding loop (SFTSV a.a. 412-436, LASV a.a. 507-529) (Kouba et al., 2021; Williams et al., 2022).

To replicate the nsRNA viral genome, a complementary RNA (cRNA) intermediate is first synthesized and can then be used as a template to synthesize copies of the nsRNA genome. For this, the RBD needs to be able to recognize the 5' and 3' ends of the cRNA for a successful replication. Notably, the cRNA intermediate be used for ambisense transcription of viral L-proteins. In the SFTSV genome, there is expression of N proteins from the cRNA of the S genome segment and in the LASV genome, it is the Z-protein and glycoprotein on the L and S segment, respectively (Ferron et al., 2017).

After replication of a viral genome segment, it gets covered with viral N proteins and attaches to an L-protein forming the RNP (Hopkins et al., 2022). These RNPs are then packed into a virion requiring one copy of each S, M and L segment to form a functional virion. Although the packaging of genome segments seems to be mostly a stochastic process, affected

primarily by the number of copies of each segment in infected cells (Bermúdez-Méndez et al., 2021). The L-protein affinity to the N protein covered genome segments and the rate of replication and transcription of each segment could affect the efficiency of packaging in virion being assembled.

6.4 Cap binding domain

Prior to this work, it had been proposed that the bunyaviruses had a CBD based on sequence homology to the Influenza virus RdRp complex. The CBD binds to 5' capped RNAs (Gogrefe et al., 2019a). These mRNAs are then cleaved 10-20 nucleotides upstream of the cap by the endonuclease domain, thereby creating a product for priming viral transcription (Olschewski et al., 2020). The structural and biochemical work presented here supports that members of the *Phenuiviridae* (Vogel et al., 2020) and *Arenaviridae* (Kouba et al., 2021) have a functional CBD to perform transcription priming *via* cap-snatching. Indeed, the CBD of L-proteins serves a key role in bunyavirus transcription. The C-termini of L-proteins in the apo conformation including the CBD are too flexible for high-resolution alignment in cryoEM SPA. Recently, a CBD domain (117 residues) of Rift valley fever virus (phenuivirus) was crystallized while bound to m7GTP (Gogrefe et al., 2019a). In the crystal structure, the CBD sandwiches the capped nucleotide with two aromatic residues (Gogrefe et al., 2019a) similar to Influenza virus (Guilligay et al., 2008). Another apo CBD was crystallized from California Academy of Sciences virus (arenavirus) without a capped nucleotide. This structure shows a CBD which does not contain the aromatic residue sandwich (Rosenthal et al., 2017). However, one can expect that an additional trigger, other than the presence of a capped nucleotide, is required for such a conformational change of the CBD into a functional state.

The CBD varies slightly in structure between SFTSV and LASV with the latter being particularly small and without aromatic residues in an orientation to sandwich capped RNA. However, the elongation structure of LASV containing a rigid CBD was solved also in absence

of capped RNA or other analogues, which leads me to expect that there is a rearrangement of the CBD in presence of mRNA substrate. Biochemical characterization of LASV L-protein transcription showed that priming can happen independent of the 5' cap indicating that there is a possibility that the L-protein can prime transcription *via* other routes than cap-snatching, at least *in vitro* (Kouba et al., 2021).

6.5 Endonuclease

The endonuclease domain is a key component in the cap-snatching function of bunyavirus L-proteins. In practice, cleavage by the endonuclease results in a truncated host mRNA product with a 5' methylated cap that can prime transcription for the L-protein. The endonuclease domain is quite flexible and in the apo conformation, it can have different positions relative to the RdRp core between the L-proteins from different segmented nsRNA viruses (Fig. 3). However, during active elongation, there is a shift of the endonuclease domain towards the polymerase core adjacent to a rigid CBD found in SFTSV (PDB:8ASB), LACV (PDBID:7ORL), LASV (PDB:7OJN) and influenza virus (PDB:6QCS). Interestingly, the LASV L-protein has clearly two different modes of endonuclease auto-inhibition. There is an endonuclease C-terminal α helix (a.a 173-190) during elongation and an inhibitory peptide (a.a 1083-1099) in apo conformation. However, the endonuclease of SFTSV L-protein has no distinguishable mechanisms of auto-inhibition. In the apo or when stalled during elongation, the endonuclease was mostly uninhibited with the exception of a rare conformation where the endonuclease is raised towards the RdRp core causing α helix (a.a. 211-233), thereby blocking its access to the active site. The relevance of this rare SFTSV auto inhibition needs further confirmation. This involves residues 181-188 in the elongation conformation and 1092-1105 in the apo state (Kouba et al., 2021). The endonuclease auto-inhibition explains why the LASV L-protein barely had any endonuclease activity unless inhibitory peptide residues were mutated (*e.g.* Y1099A) (Kouba et al., 2021).

In the structural studies of these proteins, there are point mutations in the highly conserved metal coordinating sites – H34K in LACV and D112A for SFTSV – that drastically reduce endonuclease activity and make the recombinant protein easier to over-express and less prone to aggregation in solution (Vogel et al., 2020). In our cryoEM maps of the SFTSV D112A L-protein mutant, however, there is weak density where the metal ion is normally coordinated. It could be that the metal ion is present in a part of the particles used for reconstruction and/or poorly coordinated.

It should be noted that the SFTS D112A mutant is expected to have improper metal coordination and, indeed, has reduced endonuclease activity in vitro (Vogel et al., 2020). This has the caveat that the mutation could affect the overall structure and conformational changes observed for the SFTS L-protein. I made attempts to reconstruct wild type (wt) SFTSV L-protein in the same conditions as the D112A mutant but particles would not align to high resolution and 3D reconstructions plateaued at ~ 10 Å resolution. The reason for this difference in sample behavior is not clear. The wt protein had noticeably higher aggregation tendency during sample preparation and the reconstruction seems to have the rough outline of a folded L-protein but it might be more dynamic in solution than the D112A mutant. Due to the limited yield when expressing and purifying SFTS wt L-protein this was not optimized further. Future improvements in the expression and purification of wt SFTS L-protein would allow cryoEM sample optimization using for example PEGylation or crosslinking.

6.6 Conclusion

The L-proteins are essential for both transcription and replication of the viral genome of bunyaviruses. They are highly flexible and dynamic multidomain proteins, which make them difficult targets for biochemical and structural analysis. One way for us to view them in active conformation by cryoEM single-particle analysis was to load them with the conserved 5' and

3' RNA ends of the viral genome. Furthermore, we could also stall the polymerisation reaction during elongation by incorporation of non-hydrolysable NTPs. Taking advantage of the comprehensive biochemical characterization done by collaborators in the group of Maria Rosenthal, I have been able to visualize the proteins in multiple states from apo, to mid and late replication as well as transcription initiation. My PhD work uncovered several structures of the L-proteins of the phenuivirus SFTSV and arenavirus LASV at high resolution. Altogether, our collective effort provides novel and significant insight on the complex molecular mechanisms of L-protein functions at the structural level, shedding light on different stages of bunyavirus replication and transcription. The fundamental knowledge gained lays the foundation for targeted drug design against the essential and structurally conserved motifs of L-proteins and provides a strong framework for elucidating the replication machinery of other bunyaviruses (Fig. 4).

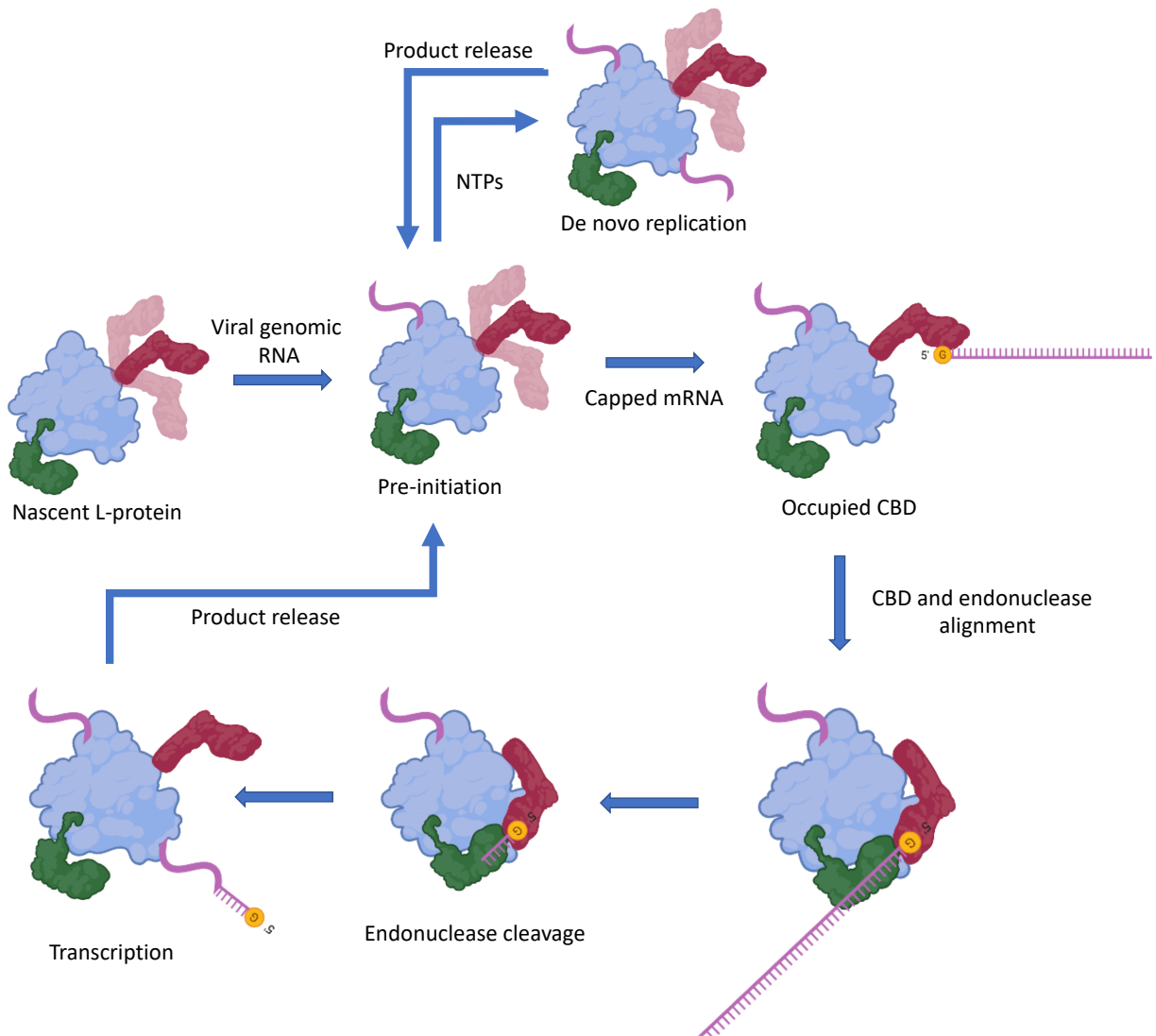


Figure 4. A cartoon flowchart of Bunyavirus L-protein states. RDRP is coloured blue, cap-binding domain in red, endonuclease in green and RNA in purple.

6.7 Future prospects

As is often the case in science, tackling one question brings forward multiple new questions. In particular, several points remain open on the L-proteins of SFTSV (*Phenuiviridae*) and LASV (*Arenaviridae*). In the available structures, including ours, there is no priming loop so we still ignore how these proteins prime replication *de novo* with the viral RNA templates. The process of priming *de novo* is a thermodynamically expensive process and is mediated by a priming loop in Influenza virus (Wandzik et al., 2021) and the peribunyavirus La Crosse virus (Arragain

et al., 2020). Our attempts at stalling the SFTSV L-protein during priming proved unsuccessful so far. This is probably due to the L-proteins propensity to incorporate mismatched base pairs during RNA synthesis. Using molecules other than non-hydrolyzable nucleotides, such as the antiviral ribavirin, might better stall the polymerase in early priming while still allowing for visualization of the priming mechanism. Time-resolved cryoEM approaches using microfluidics and spray-on-grid freezing devices (Dandey et al., 2020) might offer options for capturing the L-protein in a more native replication cycle rather than stalling the polymerase.

Of note, other viral L-proteins from members of the *Hantaviridae* were also included in our structural studies. For Sin Nombre Virus in particular, we had difficulties when it came to cryoEM sample preparation both in terms of protein aggregation during sample preparation and preferred orientation of the protein during 3D reconstruction. Optimization in terms of testing different buffer conditions, PEGylation, crosslinking, grid support and jet freezing could remedy this problem for SNV and are being further tested and optimized in the group of Maria Rosenthal.

It remains unclear why bunyavirus infections can be lethal to human hosts. An interesting avenue would be to better understand the effect of an infection on a host cell. One could use recently developed correlative approaches, electron cryo tomography and focused ion beam milling workflows (Quemin et al., 2020) to follow the progression of the cellular infection at the level of macromolecular complexes and how the host cell is remodelled during infection. It would also be interesting to track the location of different viral components during infection and to reconstruct distinct macromolecular structures (*e.g.* viral factories (Fontana et al., 2008)) at high resolution with sub-tomogram averaging (Bharat & Scheres, 2016) to understand on a molecular level what effects a bunyavirus infection has and how to combat it.

There is also limited knowledge of L-protein interactions with other proteins like the LASV L-protein in complex with the viral Z matrix protein (Xu et al., 2021). The interface of the Z protein would inhibit the rigid conformation observed during elongation and hence turn

off viral transcription. Further studies on the role and structure of L-proteins in the ribonucleoprotein complex formation are required to reveal how these essential functions are regulated during bunyavirus infections. Another interesting avenue to explore is whether the L-proteins interact with host proteins like shown for the Influenza virus RdRp complex, which recruits the human polymerase II (Krischuns et al., 2022). Such experiments could be performed *e.g.* by co-immunoprecipitation and native mass spec to find interesting candidates which could then be further characterized.

LASV is endemic in west Africa resulting in thousands of deaths each year and no effective treatment options available (Hansen et al., 2021). The recent L-protein structures are a good foundation for structure-based drug design. A strategy targeting multiple vital functions of this macromolecule (*e.g.* Cap-binding, RdRp, protein-protein interactions) at the same time could prove most effective. Indeed, there might be a good chance to have a drug cocktail that would be effective against multiple bunyaviruses and even more distantly related (segmented) nsRNA viruses.

Recently, a major breakthrough in the field of structural biology was made in protein structure prediction. Machine-learning algorithms like AlphaFold 2 (Jumper et al., 2021) and RoseTTAfold (Baek et al., 2021) have reached incredible accuracy when it comes to predicting tertiary and even quaternary structure predictions only using amino acid sequence as input from the user. These programs are even able to predict new folds (Jumper et al., 2021). They do however, struggle when it comes to flexible domains and if the prediction's confidence score is low, it tends to be a good predictor of a disordered domain. I do think that it is important to consider protein structures in general as an ensemble because proteins do not have a single rigid conformation like in crystals but can be highly flexible or display a variety of conformations based on their environments and functional regulations. There is still a lack of protein structure prediction when it comes to substrate or nucleotide interactions. One aspect to this limitation comes from the fact that there is still a relative low amount of experimental data to train

machine-learning algorithms for these tasks. The current state of protein prediction is an incredibly helpful resource for structural biology depending on the validity of the output. I'm convinced that in the future, the field will move more towards displaying proteins as flexible structures with gradual movements and considering the environmental context.

In this thesis, we focus on using cryoEM SPA to describe a highly dynamic and flexible protein in semi-rigid states. The cryoEM image processing process has a strong bias for particles that align to high-resolution alignments while filtering out lower resolution particles for example due to a rare conformation (low amount of particles), preferred orientation or a incompatible 3D reference, but these particles are potentially interesting states of the protein. Cross correlation based methods are not optimal at describing highly flexible proteins in cryoEM since they require large numbers of highly homogenous particles to reach high resolution but there are hopes with machine learning based algorithms such as cryoDRGN (Zhong et al., 2021) to describe continuous heterogeneity. This would allow for description of scarce conformations and flexible domains that tend to not contribute to high-resolution alignment. An issue with this approach is that it can be hard to verify whether the output of such algorithms is accurate, real and/or biologically relevant. If the SNR in cryoEM micrographs could be amplified with further improvements of direct electron detectors and improved phase plates (Schwartz et al., 2019) perhaps flexibility can start to be more reliably described, allowing for pseudo time-resolved reconstructions.

The field of virology has been on the spotlight in recent years in particular during the SARS-CoV2 pandemic. With this regained interest, the potential of fundamental research to boost our knowledge and hence, pandemic preparedness, among other antiviral strategies, has never been more apparent. The rapid evolution in structural virology in the cryoEM post-resolution revolution era, further amplified by machine learning and structure prediction, has opened up research avenues previously thought impossible. The work presented here has allowed for detailed insights into the molecular mechanisms of bunyavirus replication and

transcription, which lays a foundation for future developments against these deadly viruses and potential pandemic threats.

7. Summary

Bunyaviruses are emerging human pathogens with epidemic potential that remain poorly understood. These viruses rely on the so-called L-protein for both replication and transcription of the viral negative-stranded RNA genome. The structure we obtained by cryoEM single particle analysis are informed by biochemical assays from collaborators. In a collective effort to elucidate the replication mechanisms of bunyaviruses, we uncovered the structure and function of L-proteins of SFTSV and LASV in biologically relevant conformations, which will facilitate the development of antiviral strategies.

8. References

- Adrian, M., Dubochet, J., Lepault, J., & McDowell, A. W. (1984). Cryo-electron microscopy of viruses. *Nature*, *308*(5954), 32–36. <https://doi.org/10.1038/308032a0>
- Albornoz, A., Hoffmann, A., Lozach, P.-Y., & Tischler, N. (2016). Early Bunyavirus-Host Cell Interactions. *Viruses*, *8*(5), 143. <https://doi.org/10.3390/v8050143>
- Amara, A., & Mercer, J. (2015). Viral apoptotic mimicry. *Nature Reviews Microbiology*, *13*(8), 461–469. <https://doi.org/10.1038/nrmicro3469>
- Amroun, A., Priet, S., de Lamballerie, X., & Qu erat, G. (2017). Bunyaviridae RdRps: structure, motifs, and RNA synthesis machinery. *Critical Reviews in Microbiology*, *43*(6), 753–778. <https://doi.org/10.1080/1040841X.2017.1307805>
- Arragain, B., Durieux Trouillette, Q., Baudin, F., Provaznik, J., Azevedo, N., Cusack, S., Schoehn, G., & Malet, H. (2022). Structural snapshots of La Crosse virus polymerase reveal the mechanisms underlying Peribunyaviridae replication and transcription. *Nature Communications*, *13*(1). <https://doi.org/10.1038/s41467-022-28428-z>
- Arragain, B., Effantin, G., Gerlach, P., Reguera, J., Schoehn, G., Cusack, S., & Malet, H. (2020). Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes. *Nature Communications*, *11*(1), 3590. <https://doi.org/10.1038/s41467-020-17349-4>
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Mill an, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871–876. <https://doi.org/10.1126/science.abj8754>
- Barr, J. N., & Wertz, G. W. (2004). Bunyamwera bunyavirus RNA synthesis requires cooperation of 3'- and 5'-terminal sequences. *Journal of Virology*, *78*(3), 1129–1138. <https://doi.org/10.1128/jvi.78.3.1129-1138.2004>
- Bepler, T., Kelley, K., Noble, A. J., & Berger, B. (2020). Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nature Communications*, *11*(1), 5208. <https://doi.org/10.1038/s41467-020-18952-1>
- Berm dez-M endez, E., Katrukha, E. A., Spruit, C. M., Kortekaas, J., & Wichgers Schreur, P. J. (2021). Visualizing the ribonucleoprotein content of single bunyavirus virions reveals more efficient genome packaging in the arthropod host. *Communications Biology*, *4*(1), 345. <https://doi.org/10.1038/s42003-021-01821-y>
- Bharat, T. A. M., & Scheres, S. H. W. (2016). Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature Protocols*, *11*(11), 2054–2065. <https://doi.org/10.1038/nprot.2016.124>
- Brenner, S., & Horne, R. W. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochimica et Biophysica Acta*, *34*, 103–110. [https://doi.org/10.1016/0006-3002\(59\)90237-9](https://doi.org/10.1016/0006-3002(59)90237-9)
- Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R., & Grigorieff, N. (2012). Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology*, *177*(3), 630–637. <https://doi.org/10.1016/j.jsb.2012.02.003>
- Cao, W., Henry, M. D., Borrow, P., Yamada, H., Elder, J. H., Ravkov, E. V, Nichol, S. T., Compans, R. W., Campbell, K. P., & Oldstone, M. B. (1998). Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science (New York, N.Y.)*, *282*(5396), 2079–2081.

<https://doi.org/10.1126/science.282.5396.2079>

- Casel, M. A., Park, S. J., & Choi, Y. K. (2021). Severe fever with thrombocytopenia syndrome virus: emerging novel phlebovirus and their control strategy. *Experimental & Molecular Medicine*, 53(5), 713–722. <https://doi.org/10.1038/s12276-021-00610-1>
- Chen, C., Li, P., Li, K. F., Wang, H. L., Dai, Y. X., Cheng, X., & Yan, J. B. (2019). Animals as amplification hosts in the spread of severe fever with thrombocytopenia syndrome virus: A systematic review and meta-analysis. *International Journal of Infectious Diseases*, 79, 77–84. <https://doi.org/10.1016/j.ijid.2018.11.017>
- Dandey, V. P., Budell, W. C., Wei, H., Bobe, D., Maruthi, K., Kopylov, M., Eng, E. T., Kahn, P. A., Hinshaw, J. E., Kundu, N., Nimigeon, C. M., Fan, C., Sukomon, N., Darst, S. A., Saecker, R. M., Chen, J., Malone, B., Potter, C. S., & Carragher, B. (2020). Time-resolved cryo-EM using Spotiton. *Nature Methods*, 17(9), 897–900. <https://doi.org/10.1038/s41592-020-0925-6>
- De Rosier, D. J., & Klug, A. (1968). Reconstruction of three dimensional fiber structures from orthogonal projections. *Nature*, 217, 130–134.
- De Vlugt, C., Sikora, D., & Pelchat, M. (2018). Insight into Influenza: A Virus Cap-Snatching. *Viruses*, 10(11). <https://doi.org/10.3390/v10110641>
- Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A., & Homo, J.-C. (1982). Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy*, 128(3), 219–237. <https://doi.org/10.1111/j.1365-2818.1982.tb04625.x>
- Elliott, R. M. (2014). Orthobunyaviruses: recent genetic and structural insights. *Nature Reviews. Microbiology*, 12(10), 673–685. <https://doi.org/10.1038/nrmicro3332>
- Emonet, S. F., de la Torre, J. C., Domingo, E., & Sevilla, N. (2009). Arenavirus genetic diversity and its biological implications. *Infection, Genetics and Evolution*, 9(4), 417–429. <https://doi.org/10.1016/j.meegid.2009.03.005>
- Fan, H., Walker, A. P., Carrique, L., Keown, J. R., Serna Martin, I., Karia, D., Sharps, J., Hengrung, N., Pardon, E., Steyaert, J., Grimes, J. M., & Fodor, E. (2019). Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature*, 573(7773), 287–290. <https://doi.org/10.1038/s41586-019-1530-7>
- Faruqi, A. ., & Henderson, R. (2007). Electronic detectors for electron microscopy. *Current Opinion in Structural Biology*, 17(5), 549–555. <https://doi.org/10.1016/j.sbi.2007.08.014>
- Faruqi, A. R., & Andrews, H. N. (1997). Cooled CCD camera with tapered fibre optics for electron microscopy. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 392(1–3), 233–236. [https://doi.org/10.1016/S0168-9002\(97\)00217-9](https://doi.org/10.1016/S0168-9002(97)00217-9)
- Ferron, F., Weber, F., de la Torre, J. C., & Reguera, J. (2017). Transcription and replication mechanisms of Bunyaviridae and Arenaviridae L proteins. *Virus Research*, 234(January), 118–134. <https://doi.org/10.1016/j.virusres.2017.01.018>
- Fisher-Hoch, S. P., Tomori, O., Nasidi, A., Perez-Oronoz, G. I., Fakile, Y., Hutwagner, L., & McCormick, J. B. (1995). Review of cases of nosocomial Lassa fever in Nigeria: the high price of poor medical practice. *BMJ*, 311(7009), 857–859. <https://doi.org/10.1136/bmj.311.7009.857>
- Fontana, J., López-Montero, N., Elliott, R. M., Fernández, J. J., & Risco, C. (2008). The unique architecture of Bunyamwera virus factories around the Golgi complex. *Cellular Microbiology*, 10(10), 2012–2028. <https://doi.org/10.1111/j.1462-5822.2008.01184.x>
- Forbes, K. M., Sironen, T., & Plyusnin, A. (2018). Hantavirus maintenance and transmission in reservoir host populations. *Current Opinion in Virology*, 28, 1–6. <https://doi.org/10.1016/j.coviro.2017.09.003>

- Franken, L. E., Grünewald, K., Boekema, E. J., & Stuart, M. C. A. (2020). A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small*, *16*(14), 1906198. <https://doi.org/10.1002/sml.201906198>
- Gallagher, J. R., Kim, A. J., Gulati, N. M., & Harris, A. K. (2019). Negative-Stain Transmission Electron Microscopy of Molecular Complexes for Image Analysis by 2D Class Averaging. *Current Protocols in Microbiology*, *54*(1). <https://doi.org/10.1002/cpmc.90>
- Garry, R. F. (2022). Lassa fever — the road ahead. *Nature Reviews Microbiology*. <https://doi.org/10.1038/s41579-022-00789-8>
- Gerlach, P., Malet, H., Cusack, S., & Reguera, J. (2015). Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell*, *161*(6), 1267–1279. <https://doi.org/10.1016/j.cell.2015.05.006>
- Gogrefe, N., Reindl, S., Günther, S., & Rosenthal, M. (2019a). Structure of a functional cap-binding domain in Rift Valley fever virus L protein. *PLoS Pathogens*, *15*(5), e1007829. <https://doi.org/10.1371/journal.ppat.1007829>
- Gogrefe, N., Reindl, S., Günther, S., & Rosenthal, M. (2019b). Structure of a functional cap-binding domain in Rift Valley fever virus L protein. *PLOS Pathogens*, *15*(5), e1007829. <https://doi.org/10.1371/journal.ppat.1007829>
- Goldsmith, C. S., Elliott, L. H., Peters, C. J., & Zaki, S. R. (1995). Ultrastructural characteristics of Sin Nombre virus, causative agent of hantavirus pulmonary syndrome. *Archives of Virology*, *140*(12), 2107–2122. <https://doi.org/10.1007/BF01323234>
- Grant, T., & Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *ELife*, *4*. <https://doi.org/10.7554/eLife.06980>
- Grassucci, R. A., Taylor, D. J., & Frank, J. (2007). Preparation of macromolecular complexes for cryo-electron microscopy. *Nature Protocols*, *2*(12), 3239–3246. <https://doi.org/10.1038/nprot.2007.452>
- Gray, E. G. (1963). Electron microscopy of presynaptic organelles of the spinal cord. *Journal of Anatomy*, *97*(Pt 1), 101–106. <http://www.ncbi.nlm.nih.gov/pubmed/13949972>
- Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R. W. H., Ortin, J., Hart, D. J., & Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Structural & Molecular Biology*, *15*(5), 500–506. <https://doi.org/10.1038/nsmb.1421>
- Gusnard, D., & Kirschner, R. H. (1977). Cell and organelle shrinkage during preparation for scanning electron microscopy: effects of fixation, dehydration and critical point drying. *Journal of Microscopy*, *110*(1), 51–57. <https://doi.org/10.1111/j.1365-2818.1977.tb00012.x>
- Hansen, F., Jarvis, M. A., Feldmann, H., & Rosenke, K. (2021). Lassa Virus Treatment Options. *Microorganisms*, *9*(4), 772. <https://doi.org/10.3390/microorganisms9040772>
- Harmon, B., Schudel, B. R., Maar, D., Kozina, C., Ikegami, T., Tseng, C.-T. K., & Negrete, O. A. (2012). Rift Valley fever virus strain MP-12 enters mammalian host cells via caveola-mediated endocytosis. *Journal of Virology*, *86*(23), 12954–12970. <https://doi.org/10.1128/JVI.02242-12>
- Henderson, R. (1992). Image contrast in high-resolution electron microscopy of biological macromolecules: TMV in ice. *Ultramicroscopy*, *46*(1–4), 1–18. [https://doi.org/10.1016/0304-3991\(92\)90003-3](https://doi.org/10.1016/0304-3991(92)90003-3)
- Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, *110*(45), 18037–

18041. <https://doi.org/10.1073/pnas.1314449110>
- Hewat, E. A., & Neumann, E. (2002). Characterization of the performance of a 200-kV field emission gun for cryo-electron microscopy of biological molecules. *Journal of Structural Biology*, *139*(1), 60–64. [https://doi.org/10.1016/S1047-8477\(02\)00515-4](https://doi.org/10.1016/S1047-8477(02)00515-4)
- Hollidge, B. S., Nedelsky, N. B., Salzano, M.-V., Fraser, J. W., González-Scarano, F., & Soldan, S. S. (2012). Orthobunyavirus Entry into Neurons and Other Mammalian Cells Occurs via Clathrin-Mediated Endocytosis and Requires Trafficking into Early Endosomes. *Journal of Virology*, *86*(15), 7988–8001. <https://doi.org/10.1128/JVI.00140-12>
- Hopkins, F. R., Álvarez-Rodríguez, B., Heath, G. R., Panayi, K., Hover, S., Edwards, T. A., Barr, J. N., & Fontana, J. (2022). The Native Orthobunyavirus Ribonucleoprotein Possesses a Helical Architecture. *MBio*, *13*(4). <https://doi.org/10.1128/mbio.01405-22>
- Hulswit, R. J. G., Paesen, G. C., Bowden, T. A., & Shi, X. (2021). Recent Advances in Bunyavirus Glycoprotein Research: Precursor Processing, Receptor Binding and Structure. *Viruses*, *13*(2), 353. <https://doi.org/10.3390/v13020353>
- Hussein, I. T. M., Haseeb, A., Haque, A., & Mir, M. A. (2011). Recent advances in hantavirus molecular biology and disease. *Advances in Applied Microbiology*, *74*, 35–75. <https://doi.org/10.1016/B978-0-12-387022-3.00006-9>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kang, H., Cong, J., Wang, C., Ji, W., Xin, Y., Qian, Y., Li, X., Chen, Y., & Rao, Z. (2021). Structural basis for recognition and regulation of arenavirus polymerase L by Z protein. *Nature Communications*, *12*(1), 4134. <https://doi.org/10.1038/s41467-021-24458-1>
- Kapuscinski, M. L., Bergren, N. A., Russell, B. J., Lee, J. S., Borland, E. M., Hartman, D. A., King, D. C., Hughes, H. R., Burkhalter, K. L., Kading, R. C., & Stenglein, M. D. (2021). Genomic characterization of 99 viruses from the bunyavirus families Nairoviridae, Peribunyaviridae, and Phenuiviridae, including 35 previously unsequenced viruses. In *PLoS Pathogens* (Vol. 17, Issue 3). <https://doi.org/10.1371/journal.ppat.1009315>
- Kohl, A., Dunn, E. F., Lowen, A. C., & Elliott, R. M. (2004). Complementarity, sequence and structural elements within the 3' and 5' non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength. *Journal of General Virology*, *85*(11), 3269–3278. <https://doi.org/10.1099/vir.0.80407-0>
- Koning, R. I., Koster, A. J., & Sharp, T. H. (2018). Advances in cryo-electron tomography for biology and medicine. *Annals of Anatomy - Anatomischer Anzeiger*, *217*, 82–96. <https://doi.org/10.1016/j.aanat.2018.02.004>
- Kouba, T., Drncová, P., & Cusack, S. (2019). Structural snapshots of actively transcribing influenza polymerase. *Nature Structural & Molecular Biology*, *26*(6), 460–470. <https://doi.org/10.1038/s41594-019-0232-z>
- Kouba, T., Vogel, D., Thorkelsson, S. R., Quemin, E. R. J., Williams, H. M., Milewski, M., Busch, C., Günther, S., Grünewald, K., Rosenthal, M., & Cusack, S. (2021). Conformational changes in Lassa virus L protein associated with promoter binding and RNA synthesis activity. *Nature Communications*, *12*(1), 7018. <https://doi.org/10.1038/s41467-021-27305-5>
- Krischuns, T., Isel, C., Drncova, P., Lukarska, M., Pflug, A., Paisant, S., Navratil, V., Cusack, S., & Naffakh, N. (2022). Type B and type A influenza polymerases have evolved distinct binding interfaces to recruit the RNA polymerase II CTD. *PLoS Pathogens*, *18*(5),

- e1010328. <https://doi.org/10.1371/journal.ppat.1010328>
- Kühlbrandt, W. (2014). The Resolution Revolution. *Science*, 343(6178), 1443–1444. <https://doi.org/10.1126/science.1251652>
- Lake, J. A. (1976). Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. *Journal of Molecular Biology*, 105(1), 131–159. [https://doi.org/10.1016/0022-2836\(76\)90200-X](https://doi.org/10.1016/0022-2836(76)90200-X)
- Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., & Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, 10(6), 584–590. <https://doi.org/10.1038/nmeth.2472>
- Liu, Q., He, B., Huang, S.-Y., Wei, F., & Zhu, X.-Q. (2014). Severe fever with thrombocytopenia syndrome, an emerging tick-borne zoonosis. *The Lancet Infectious Diseases*, 14(8), 763–772. [https://doi.org/10.1016/S1473-3099\(14\)70718-2](https://doi.org/10.1016/S1473-3099(14)70718-2)
- Lowen, A. C., & Elliott, R. M. (2005). Mutational analyses of the nonconserved sequences in the Bunyamwera Orthobunyavirus S segment untranslated regions. *Journal of Virology*, 79(20), 12861–12870. <https://doi.org/10.1128/JVI.79.20.12861-12870.2005>
- Ma, J., Zhang, S., & Zhang, X. (2021). Structure of Machupo virus polymerase in complex with matrix protein Z. *Nature Communications*, 12(1), 6163. <https://doi.org/10.1038/s41467-021-26432-3>
- McCormick, J. B., & Fisher-Hoch, S. P. (2002). Lassa fever. *Current Topics in Microbiology and Immunology*, 262, 75–109. https://doi.org/10.1007/978-3-642-56029-3_4
- McLay, L., Liang, Y., & Ly, H. (2014). Comparative analysis of disease pathogenesis and molecular mechanisms of New World and Old World arenavirus infections. *Journal of General Virology*, 95(1), 1–15. <https://doi.org/10.1099/vir.0.057000-0>
- Mindell, J. A., & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology*, 142(3), 334–347. [https://doi.org/10.1016/s1047-8477\(03\)00069-8](https://doi.org/10.1016/s1047-8477(03)00069-8)
- Morin, B., Coutard, B., Lelke, M., Ferron, F., Kerber, R., Jamal, S., Frangeul, A., Baronti, C., Charrel, R., de Lamballerie, X., Vonnrhein, C., Lescar, J., Bricogne, G., Günther, S., & Canard, B. (2010). The N-Terminal Domain of the Arenavirus L Protein Is an RNA Endonuclease Essential in mRNA Transcription. *PLoS Pathogens*, 6(9), e1001038. <https://doi.org/10.1371/journal.ppat.1001038>
- Mostafavi, E., Pourhossein, B., & Chinikar, S. (2014). Clinical symptoms and laboratory findings supporting early diagnosis of Crimean-Congo hemorrhagic fever in Iran. *Journal of Medical Virology*, 86(7), 1188–1192. <https://doi.org/10.1002/jmv.23922>
- Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., Grigoras, I. T., Malinauskaitė, L., Malinauskas, T., Miehl, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., ... Scheres, S. H. W. (2020). Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832), 152–156. <https://doi.org/10.1038/s41586-020-2829-0>
- Novoa, R. R., Calderita, G., Cabezas, P., Elliott, R. M., & Risco, C. (2005). Key Golgi factors for structural and functional maturation of bunyamwera virus. *Journal of Virology*, 79(17), 10852–10863. <https://doi.org/10.1128/JVI.79.17.10852-10863.2005>
- Ohi, M., Li, Y., Cheng, Y., & Walz, T. (2004). Negative staining and image classification — powerful tools in modern electron microscopy. *Biological Procedures Online*, 6(1), 23–34. <https://doi.org/10.1251/bpo70>
- Olayemi, A., Cadar, D., Magassouba, N., Obadare, A., Kourouma, F., Oyeyiola, A., Fasogbon, S., Igbokwe, J., Rieger, T., Bockholt, S., Jérôme, H., Schmidt-Chanasit, J., Garigliany, M.,

- Lorenzen, S., Igbahenah, F., Fichet, J.-N., Ortsega, D., Omilabu, S., Günther, S., & Fichet-Calvet, E. (2016). New Hosts of The Lassa Virus. *Scientific Reports*, *6*(1), 25280. <https://doi.org/10.1038/srep25280>
- Olschewski, S., Cusack, S., & Rosenthal, M. (2020). The Cap-Snatching Mechanism of Bunyaviruses. *Trends in Microbiology*, *28*(4), 293–303. <https://doi.org/10.1016/j.tim.2019.12.006>
- Överby, A. K., Pettersson, R. F., Grünewald, K., & Huiskonen, J. T. (2008). Insights into bunyavirus architecture from electron cryotomography of Uukuniemi virus. *Proceedings of the National Academy of Sciences*, *105*(7), 2375–2379. <https://doi.org/10.1073/pnas.0708738105>
- Overby, A. K., Pettersson, R. F., & 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), 3198–3205. <https://doi.org/10.1128/JVI.02655-06>
- Palade, G. E. (1952). A STUDY OF FIXATION FOR ELECTRON MICROSCOPY. *Journal of Experimental Medicine*, *95*(3), 285–298. <https://doi.org/10.1084/jem.95.3.285>
- Peng, R., Xu, X., Jing, J., Wang, M., Peng, Q., Liu, S., Wu, Y., Bao, X., Wang, P., Qi, J., Gao, G. F., & Shi, Y. (2020). Structural insight into arenavirus replication machinery. *Nature*, *579*(7800), 615–619. <https://doi.org/10.1038/s41586-020-2114-2>
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290–296. <https://doi.org/10.1038/nmeth.4169>
- Quemin, E. R. J., Machala, E. A., Vollmer, B., Pražák, V., Vasishtan, D., Rosch, R., Grange, M., Franken, L. E., Baker, L. A., & Grünewald, K. (2020). Cellular Electron Cryo-Tomography to Study Virus-Host Interactions. *Annual Review of Virology*, *7*(1), 239–262. <https://doi.org/10.1146/annurev-virology-021920-115935>
- Raimondi, V., & Grinzato, A. (2021). A basic introduction to single particles cryo-electron microscopy. *AIMS Biophysics*, *9*(1), 5–20. <https://doi.org/10.3934/biophy.2022002>
- Reguera, J., Gerlach, P., Rosenthal, M., Gaudon, S., Coscia, F., Günther, S., & Cusack, S. (2016). Comparative Structural and Functional Analysis of Bunyavirus and Arenavirus Cap-Snatching Endonucleases. *PLOS Pathogens*, *12*(6), e1005636. <https://doi.org/10.1371/journal.ppat.1005636>
- Reguera, J., Weber, F., & 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), e1001101. <https://doi.org/10.1371/journal.ppat.1001101>
- Richmond, J. K. (2003). Lassa fever: epidemiology, clinical features, and social consequences. *BMJ*, *327*(7426), 1271–1275. <https://doi.org/10.1136/bmj.327.7426.1271>
- Rosenthal, M., Gogrefe, N., Vogel, D., Reguera, J., Rauschenberger, B., Cusack, S., Günther, S., & Reindl, S. (2017). Structural insights into reptarenavirus cap-snatching machinery. *PLOS Pathogens*, *13*(5), e1006400. <https://doi.org/10.1371/journal.ppat.1006400>
- Rowe, W. P., Murphy, F. A., Bergold, G. H., Casals, J., Hotchin, J., Johnson, K. M., Lehmann-Grube, F., Mims, C. A., Traub, E., & Webb, P. A. (1970). Arenoviruses: proposed name for a newly defined virus group. *Journal of Virology*, *5*(5), 651–652. <https://doi.org/10.1128/JVI.5.5.651-652.1970>
- Russo, C. J., & Passmore, L. A. (2016). Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens. *Journal of Structural Biology*, *193*(1), 33–44. <https://doi.org/10.1016/j.jsb.2015.11.006>

- Salanueva, I. J., Novoa, R. R., Cabezas, P., López-Iglesias, C., Carrascosa, J. L., Elliott, R. M., & Risco, C. (2003). Polymorphism and structural maturation of bunyamwera virus in Golgi and post-Golgi compartments. *Journal of Virology*, *77*(2), 1368–1381. <https://doi.org/10.1128/jvi.77.2.1368-1381.2003>
- Sarute, N., & Ross, S. R. (2017). New World Arenavirus Biology. *Annual Review of Virology*, *4*(1), 141–158. <https://doi.org/10.1146/annurev-virology-101416-042001>
- Scheres, S. H. W. (2012a). A Bayesian View on Cryo-EM Structure Determination. *Journal of Molecular Biology*, *415*(2), 406–418. <https://doi.org/10.1016/j.jmb.2011.11.010>
- Scheres, S. H. W. (2012b). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530. <https://doi.org/10.1016/j.jsb.2012.09.006>
- Scheres, S. H. W., & Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nature Methods*, *9*(9), 853–854. <https://doi.org/10.1038/nmeth.2115>
- Scheres, S. H. W., Valle, M., Nuñez, R., Sorzano, C. O. S., Marabini, R., Herman, G. T., & Carazo, J.-M. (2005). Maximum-likelihood Multi-reference Refinement for Electron Microscopy Images. *Journal of Molecular Biology*, *348*(1), 139–149. <https://doi.org/10.1016/j.jmb.2005.02.031>
- Schwartz, O., Axelrod, J. J., Campbell, S. L., Turnbaugh, C., Glaeser, R. M., & Müller, H. (2019). Laser phase plate for transmission electron microscopy. *Nature Methods*, *16*(10), 1016–1020. <https://doi.org/10.1038/s41592-019-0552-2>
- Stempak, J. G., & Ward, R. T. (1964). AN IMPROVED STAINING METHOD FOR ELECTRON MICROSCOPY. *The Journal of Cell Biology*, *22*(3), 697–701. <https://doi.org/10.1083/jcb.22.3.697>
- Subramaniam, K., Ferguson, H. W., Kabuusu, R., & Waltzek, T. B. (2019). Genome Sequence of Tilapia Lake Virus Associated with Syncytial Hepatitis of Tilapia in an Ecuadorian Aquaculture Facility. *Microbiology Resource Announcements*, *8*(18). <https://doi.org/10.1128/MRA.00084-19>
- Sun, M. H., Ji, Y. F., Li, G. H., Shao, J. W., Chen, R. X., Gong, H. Y., Chen, S. Y., & Chen, J. M. (2022). Highly adaptive Phenuiviridae with biomedical importance in multiple fields. *Journal of Medical Virology*, *November 2021*, 1–14. <https://doi.org/10.1002/jmv.27618>
- Sun, Y., Li, J., Gao, G. F., Tien, P., & Liu, W. (2018). Bunyavirales ribonucleoproteins: the viral replication and transcription machinery. *Critical Reviews in Microbiology*, *44*(5), 522–540. <https://doi.org/10.1080/1040841X.2018.1446901>
- te Velhuis, A. J. W., Grimes, J. M., & Fodor, E. (2021). Publisher Correction: Structural insights into RNA polymerases of negative-sense RNA viruses. *Nature Reviews Microbiology*, *19*(3), 220–220. <https://doi.org/10.1038/s41579-021-00524-9>
- Tegunov, D., & Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods*, *16*(11), 1146–1152. <https://doi.org/10.1038/s41592-019-0580-y>
- ter Horst, S., Conceição-Neto, N., Neyts, J., & Rocha-Pereira, J. (2019). Structural and functional similarities in bunyaviruses: Perspectives for pan-bunya antivirals. *Reviews in Medical Virology*, *29*(3), e2039. <https://doi.org/10.1002/rmv.2039>
- Ter Horst, S., Conceição-Neto, N., Neyts, J., & Rocha-Pereira, J. (2019). Structural and functional similarities in bunyaviruses: Perspectives for pan-bunya antivirals. *Reviews in Medical Virology*, *29*(3), e2039. <https://doi.org/10.1002/rmv.2039>
- Venkataraman, S., Prasad, B., & Selvarajan, R. (2018). RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses*, *10*(2), 76. <https://doi.org/10.3390/v10020076>

- Vogel, D., Rosenthal, M., Gogrefe, N., Reindl, S., & Günther, S. (2019). Biochemical characterization of the Lassa virus L protein. *Journal of Biological Chemistry*, 294(20), 8088–8100. <https://doi.org/10.1074/JBC.RA118.006973>
- Vogel, D., Thorkelsson, S. R., Quemin, E. R. J., Meier, K., Kouba, T., Gogrefe, N., Busch, C., Reindl, S., Günther, S., Cusack, S., Grünewald, K., & Rosenthal, M. (2020). Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L protein. *Nucleic Acids Research*, 48(10), 5749–5765. <https://doi.org/10.1093/nar/gkaa253>
- Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C., & Raunser, S. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, 2(1), 218. <https://doi.org/10.1038/s42003-019-0437-z>
- Walker, A. P., & Fodor, E. (2019). Interplay between Influenza Virus and the Host RNA Polymerase II Transcriptional Machinery. *Trends in Microbiology*, 27(5), 398–407. <https://doi.org/10.1016/J.TIM.2018.12.013>
- Walter, C. T., & Barr, J. N. (2011). Recent advances in the molecular and cellular biology of bunyaviruses. *Journal of General Virology*, 92(11), 2467–2484. <https://doi.org/10.1099/vir.0.035105-0>
- Wandzik, J. M., Kouba, T., & Cusack, S. (2021). Structure and Function of Influenza Polymerase. *Cold Spring Harbor Perspectives in Medicine*, 11(9), a038372. <https://doi.org/10.1101/cshperspect.a038372>
- Wang, P., Liu, L., Liu, A., Yan, L., He, Y., Shen, S., Hu, M., Guo, Y., Liu, H., Liu, C., Lu, Y., Wang, P., Deng, F., Rao, Z., & Lou, Z. (2021). Author Correction: Structure of severe fever with thrombocytopenia syndrome virus L protein elucidates the mechanisms of viral transcription initiation. *Nature Microbiology*, 6(5), 697–698. <https://doi.org/10.1038/s41564-021-00906-y>
- Wang, X., Hu, C., Ye, W., Wang, J., Dong, X., Xu, J., Li, X., Zhang, M., Lu, H., Zhang, F., Wu, W., Dai, S., Wang, H.-W., & Chen, Z. (2022). Structure of Rift Valley Fever Virus RNA-Dependent RNA Polymerase. *Journal of Virology*, 96(3), e0171321. <https://doi.org/10.1128/JVI.01713-21>
- Watson, M. L. (1958). Staining of Tissue Sections for Electron Microscopy with Heavy Metals. *The Journal of Biophysical and Biochemical Cytology*, 4(6), 727–730. <https://doi.org/10.1083/jcb.4.6.727>
- Williams, H. M., Thorkelsson, S. R., Vogel, D., Milewski, M., Busch, C., Cusack, S., Grünewald, K., Quemin, E. R. J., & Rosenthal, M. (2022). Structural insights into viral genome replication by the Severe fever with thrombocytopenia syndrome virus L protein. *BioRxiv*. <https://doi.org/https://doi.org/10.1101/2022.08.25.505333>
- Williams, H. M., Thorkelsson, S. R., Vogel, D., Milewski, M., Busch, C., Cusack, S., Grünewald, K., Quemin, E. R. J., & Rosenthal, M. (2023). Structural insights into viral genome replication by the severe fever with thrombocytopenia syndrome virus L protein. *Nucleic Acids Research*, 51(3), 1424–1442. <https://doi.org/10.1093/nar/gkac1249>
- Wilson, W. C., Mitzel, D., Cohnstaedt, L. W., Noronha, L., Drolet, B. S., & McVey, D. S. (2022). Bunyavirales. In *Veterinary Microbiology* (pp. 589–595). Wiley. <https://doi.org/10.1002/9781119650836.ch58>
- Wuerth, J. D., Habjan, M., Wulle, J., Superti-Furga, G., Pichlmair, A., & Weber, F. (2018). NSs Protein of Sandfly Fever Sicilian Phlebovirus Counteracts Interferon (IFN) Induction by Masking the DNA-Binding Domain of IFN Regulatory Factor 3. *Journal of Virology*,

- 92(23). <https://doi.org/10.1128/JVI.01202-18>
- Xu, X., Peng, R., Peng, Q., Wang, M., Xu, Y., Liu, S., Tian, X., Deng, H., Tong, Y., Hu, X., Zhong, J., Wang, P., Qi, J., Gao, G. F., & Shi, Y. (2021). Cryo-EM structures of Lassa and Machupo virus polymerases complexed with cognate regulatory Z proteins identify targets for antivirals. *Nature Microbiology*, *6*(7), 921–931. <https://doi.org/10.1038/s41564-021-00916-w>
- Yip, K. M., Fischer, N., Paknia, E., Chari, A., & Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature*, *587*(7832), 157–161. <https://doi.org/10.1038/s41586-020-2833-4>
- Yu, X.-J., Liang, M.-F., Zhang, S.-Y., Liu, Y., Li, J.-D., Sun, Y.-L., Zhang, L., Zhang, Q.-F., Popov, V. L., Li, C., Qu, J., Li, Q., Zhang, Y.-P., Hai, R., Wu, W., Wang, Q., Zhan, F.-X., Wang, X.-J., Kan, B., ... Li, D.-X. (2011). Fever with thrombocytopenia associated with a novel bunyavirus in China. *The New England Journal of Medicine*, *364*(16), 1523–1532. <https://doi.org/10.1056/NEJMoa1010095>
- Zhan, J., Wang, Q., Cheng, J., Hu, B., Li, J., Zhan, F., Song, Y., & Guo, D. (2017). Current status of severe fever with thrombocytopenia syndrome in China. *Virologica Sinica*, *32*(1), 51–62. <https://doi.org/10.1007/s12250-016-3931-1>
- Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, *14*(4), 331–332. <https://doi.org/10.1038/nmeth.4193>
- Zhong, E. D., Bepler, T., Berger, B., & Davis, J. H. (2021). CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, *18*(2), 176–185. <https://doi.org/10.1038/s41592-020-01049-4>
- Zhuang, L., Sun, Y., Cui, X.-M., Tang, F., Hu, J.-G., Wang, L.-Y., Cui, N., Yang, Z.-D., Huang, D.-D., Zhang, X.-A., Liu, W., & Cao, W.-C. (2018). Transmission of Severe Fever with Thrombocytopenia Syndrome Virus by *Haemaphysalis longicornis* Ticks, China. *Emerging Infectious Diseases*, *24*(5). <https://doi.org/10.3201/eid2405.151435>

9. Declaration on oath

I hereby declare on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids. The submitted written version corresponds to the version on the electronic storage medium. I hereby declare that I have not previously applied or pursued for a doctorate (Ph.D. studies).