# Structure and function of the La Crosse orthobunyavirus polymerase

Piotr Gerlach

UNIVERSITÉ DE GRENOBLE

EMBL

# THESIS / THÈSE

To obtain the title of / Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Discipline / Spécialité : Structural Biology – Nanobiology / Biologie Structurale et Nanobiologie

Arrêté ministériel : 7 août 2006

Presented by / Présentée par

**Piotr GERLACH**

Thesis supervisor / Thèse dirigée par **Dr. Stephen CUSACK**

Thesis prepared at / Thèse préparée au sein du
**European Molecular Biology Laboratory (EMBL)**
**Grenoble Outstation**

In / Dans **l'École Doctorale Chimie et Sciences du Vivant**

# Structure and function of the La Crosse orthobunyavirus polymerase

## La structure et la fonction de la polymérase d'orthobunyavirus La Crosse

Public defense / Thèse soutenue publiquement le **29/06/2015**
Jury members / Devant le jury composé de :

**Dr. Stephen CUSACK**
Directeur de thèse
**Dr. Bruno CANARD**
Rapporteur
**Dr. John BARR**
Rapporteur
**Prof. Daniel KOLAKOFSKY**
Examinateur
**Prof. Winfried WEISSENHORN**
Président

*Université Joseph Fourier / Université Pierre Mendès France /*
*Université Stendhal / Université de Savoie / Grenoble INP*

SIGILLUM UNIVERSITATIS GRATIANOPOLITANAE 1339

TO MY FAMILY

# SUMMARY

Viruses are not more than particles composed of lipids and/or proteins with genetic information – the viral RNA or DNA genome – embedded inside. In order to be efficient, once they enter the host cell they need to multiply this genetic information, package it into new viral particles and spread out from the cell. While in order to produce viral proteins viruses highjack cellular machinery, for replicating their genome most viruses use their own, specialized polymerases.

*Bunyaviridae* is the largest viral family of segmented negative-strand RNA viruses, comprising also *Arenaviridae* and *Orthomyxoviridae* families. Some bunyaviruses are causative agents of severe human diseases including heamorrhagic fevers, encephalitis and meningitis. Others infect a variety of plants and animals posing a significant economic threat to the crop cultivation and cattle breeding.

RNA-dependent RNA polymerases of segmented negative-strand RNA viruses are multifunctional machines, able to perform both *de novo* genome replication via positive-strand cRNA intermediate, and viral mRNA transcription using cap-snatched host-derived mRNA primers. Viral RNA genome of bunyaviruses, arenaviruses, and orthomyxoviruses is divided into three, two, and eight segments respectively. Each segment, coated by nucleoproteins and attached through its conserved 3′ and 5′ ends to the polymerase, constitutes an individual ribonucleoprotein particle – an autonomous RNA synthesis unit.

The scope of the PhD project described in this thesis was the structural and functional characterization of the La Crosse orthobunyavirus polymerase, also named the L protein. It was based on the hypothesis that all polymerases of segmented negative-strand RNA viruses share a similar domain organization and mode of action. During the 1$^{st}$ year attempts were made to confirm and characterize a putative C-terminal cap-binding domain. During the 2$^{nd}$ year project was extended to study 3′ and 5′ vRNA ends interactions with the full length and C-terminus truncated L protein. Facing difficulties to establish replication and transcription assays *in vitro*, vRNA binding studies and co-crystallizastion were continued during the 3$^{rd}$ year. This finally led to the main achievement of the thesis – the x-ray structure of La Crosse orthobunyavirus polymerase in complex with vRNA. Obtained structure is a breakthrough in the bunyavirus field. It reveals – unlike it was initially believed – conserved, sequence specific and separate binding sites for 3′ and 5′ vRNA ends located within the polymerase. The 5′ vRNA end binding allosterically structures one of the conserved catalytic motifs within the polymerase active site. The structure sheds also some new light on bunyaviral replication and transcription mechanisms. There exist two distinct product and template exit channels, suggesting that the nascent RNA strand is separated from the template and leaves the polymerase as the single-strand RNA. Close proximity of the template entry and exit channels explains how the polymerase can translocate along the genomic template with minimal disruption of the RNP.

In parallel to the La Crosse polymerase structure, structures of Influenza A and B heterotrimeric polymerases in complex with vRNA were also obtained in Stephen Cusack group. This gave a great opportunity to compare the domain organization and the nature of vRNA binding by viral polymerases belonging to *Bunyaviriadae* and *Orthomyxoviridae* families, and proved that despite minimal sequence homology the structural similarities are striking. This strongly suggests an evolutionary common ancestor, which can possibly be shared with non-segmented negative-strand RNA viruses as well.

## RÉSUMÉ

Les virus ne sont rien de plus que des particules composées de lipides et/ou de protéines qui encapsulent de l'information génétique composée d'ARN ou d'ADN. Au cours du cycle viral, les virus entrent dans la cellule hôte où ils dupliquent leur génome, puis forment de nouvelles particules virales qui ressortiront de la cellule pour se diffuser. Alors que pour produire leurs protéines virales les virus détournent la machinerie cellulaire, ils utilisent pour la plupart leur propre polymérase spécifique pour répliquer leur génome.

Les *Bunyaviridae* sont une grande famille des virus à ARN simple brin segmenté de polarité négative. Les *Arenaviridae* et les *Orthomyxoviridae* sont les deux autres familles de ce type. Certains bunyavirus provoquent des maladies humaines graves, comme des fièvres hémorragiques, des encéphalites et des méningites. D'autres infectent des plantes et animaux, posant une menace économique sérieuse en agronomie.

Les ARN polymérases ARN-dépendante de virus à ARN négatif segmenté sont des machineries multi-fonctionnelles, capables de répliquer le génome viral et de le transcrire en ARNs messagers. La réplication est effectuée *de novo*, en utilisant un intermédiaire d'ARN complémentaire de polarité positive, alors que la transcription est initiée par vol de coiffe d'ARN cellulaire. Chaque segment du génome viral est recouvert par des nucléoprotéines et fixé à la polymérase par ses extrémités 3' et 5' conservées. Le complexe ARN viral/nucléoprotéines/polymérase forme une ribonucléoprotéine, qui est l'unité fonctionnelle de la réplication/transcription.

L'objectif de mon projet de thèse était la caractérisation structurale et fonctionnelle de la polymérase du virus La Crosse, également nommée protéine L. Ce projet était basé sur l'hypothèse que toutes les polymérases de virus à ARN négatif segmenté pourraient partager une organisation et un mode d'action similaire. Lors de la première année de ma thèse, j'ai tenté de caractériser le domaine C-terminal, que nous supposions être responsable de la fixation de coiffe. Au cours de la deuxième année, j'ai étendu mes recherches sur l'étude de l'interaction entre les extrémités de l'ARN viral et la protéine L (protéine entière et construction tronquée en C-terminal). Confronté à des difficultés pour établir des tests de réplication et de transcription *in vitro*, j'ai poursuivi mes recherches en troisième année avec l'étude d'interactions et de co-cristallisation entre polymérase et ARN viral. Cela a finalement conduit au résultat principal de ma thèse - la détermination de la structure par cristallographie aux rayons X de la polymérase de virus de La Crosse en complexe avec les extrémités 3' et 5' de l'ARN viral. La structure obtenue constitue une percée dans le domaine de bunyavirus. Elle révèle – à la différence de ce qui avait été initialement proposé – que les extrémités 3' et 5' de l'ARN se lient dans deux sites séparés et conservés. La liaison de l'extrémité 5' de l'ARN viral stabilise de façon allostérique l'un des motifs catalytiques du site actif de la polymérase. La structure révèle l'existence de deux tunnels séparés pour l'ARN produit et l'ARN matrice de

sortir, ce qui suggère que le brin d'ARN naissant est séparé de la matrice et quitte la polymérase comme ARN simple brin. La proximité des tunnels d'entrée et de sortie de la matrice explique comment la polymérase peut se déplacer le long de l'ARN génomique avec une perturbation minimale de la ribonucléoprotéine.

En parallèle de la structure de la polymérase du virus La Crosse, les structures des polymérases hétérotrimériques de la grippe A et B en complexe avec l'ARN viral ont également été déterminées au sein du groupe du Dr. Stephen Cusack. La comparaison de l'organisation des polymérases des deux familles et de la nature de leur liaison avec l'ARN viral montre que, malgré une homologie de séquence minimale, des similitudes structurelles sont frappantes. Cela suggère fortement la présence d'un ancêtre commun.

## Table of Contents


**1. INTRODUCTION ........ 13**

**1.1 Bunyaviruses ........ 14**

1.1.1 Classification, distribution, and transmission ........ 14

1.1.2 Viral cycle ........ 17

1.1.3 Lytic vs. persistent infection ........ 19

1.1.4 Organization of the viral genome ........ 20

1.1.5 L protein – the polymerase ........ 23

1.1.6 N protein – the nucleoprotein ........ 25

1.1.7 NSs and NSm – the non-structural proteins ........ 28

1.1.8 Gn and Gc – the glycoproteins ........ 29

1.1.9 Role of genomic non-translated regions in RNA synthesis ........ 30

1.1.10 Cap-snatching initiates transcription ........ 31

1.1.11 Translation-dependent transcription ........ 32

1.1.12 Prime-and-realign model for replication initiation ........ 33

**1.2 RNA-dependent RNA Polymerases ........ 35**

1.2.1 Primordial polymerases – from RNA to protein ........ 35

1.2.2 Nucleotidyl transfer reaction – chemistry ........ 36

1.2.3 Polymerases – overview and classification ........ 36

1.2.4 Cellular RNA-dependent RNA polymerases ........ 37

1.2.5 Viral RNA-dependent RNA polymerases ........ 39

1.2.6 RdRp – catalytic motifs ........ 42

1.2.7 *De novo* initiation – different strategies ........ 44

1.2.8 Additional domains and multimerisation ........ 45

**1.3 Aim of the thesis ........ 49**

**2. METHODS ........ 51**

**2.1 Protein expression and purification ... 52**

2.1.1 Constructs for the bacteria expression ... 52

2.1.2 Constructs for the insect cell expression ... 54

2.1.3 Bacteria expression ... 56

2.1.4 Insect cells expression ... 56

2.1.5 L protein C-terminus constructs purification ... 57

2.1.6 L and $L_{1750}$ proteins purification ... 57

**2.2 RNA oligos ... 58**

2.2.1 Synthetic RNA oligos ... 58

2.2.2 T7 transcription of RNA oligos ... 60

**2.3 Functional characterization ... 64**

2.3.1 Limited proteolysis experiments ... 64

2.3.2 RNA-protein UV crosslinking ... 64

2.3.3 RNA-protein binding studies – radioactive EMSA assays ... 65

2.3.4 RNP assembly ... 65

2.3.5 Isothermal titration calorimetry ... 66

2.3.6 Microscale thermophoresis ... 67

**2.4 Structural characterization ... 68**

2.4.1 SAXS – data collection and analysis ... 68

2.4.2 X-ray crystallography – protein crystallization ... 69

2.4.3 X-ray crystallography – data collection ... 71

2.4.4 X-ray crystallography – data processing ... 72

2.4.5 $L_{1750}$-vRNA structure solution and model building ... 73

2.4.6 L-vRNA structure solution and model building ... 74

**3. RESULTS ... 75**

**3.1 LACV L protein C terminus – structural characterization.......................................... 76**

3.1.1 L6 and L6-derived constructs – design, expression, and purification ..................... 76

3.1.2 7AR and 8SR constructs – crystallisation trials ...................................................... 79

3.1.3 8SR construct – CD and 1D $^1$H-NMR analysis.......................................................... 80

3.1.4 3A and L7 and other C-terminal construct – expression and purification .............. 82

3.1.5 L7 C-terminal construct – limited proteolysis........................................................ 82

3.1.6 L7-derived 2$^{nd}$ set of C-terminal constructs.......................................................... 85

**3.2 LACV L protein C terminus – functional studies and binding assays .......................... 86**

3.2.1 L protein C terminus nuclease activity assays ....................................................... 86

3.2.2 8SR cap-binding assays ......................................................................................... 86

3.2.3 RNA binding assays ............................................................................................... 89

3.2.4 Protein-protein interactions .................................................................................. 89

**3.3 LACV full length L protein and $L_{1750}$ construct – RNA binding.................................. 90**

3.3.1 L protein and $L_{1750}$ – expression and purification .................................................. 90

3.3.2 Electrophoretic mobility shift assays ..................................................................... 90

3.3.3 Fluorescence anisotropy......................................................................................... 92

3.3.4 Isothermal titration calorimetry ............................................................................ 94

3.3.5 Microscale thermophoresis .................................................................................... 94

3.3.6 vRNA protection during $L_{1750}$ and L protein trypsination ....................................... 95

3.3.7 $L_{1750}$ – vRNA UV crosslinking .................................................................................. 97

**3.4 LACV full length L protein and $L_{1750}$ construct – structural characterization ............. 98**

3.4.1 $L_{1750}$ and L protein – SAXS analysis ....................................................................... 98

3.4.2 $L_{1750}$-vRNA co-crystallization experiments............................................................ 98

3.4.3 ATTACHED ARTICLE ................................................................................................ 105

3.4.4 L-vRNA crystallization and data collection........................................................... 150

3.4.5 L-vRNA structure solution and structural features.............................................. 151

**3.5 LACV full length L protein and $L_{1750}$ construct – functional studies......................... 154**

3.5.1 L and $L_{1750}$ – cap-binding assays ........ 154
3.5.2 Polymerase activity assays – transcription ........ 156
3.5.3 Polymerase activity assays – replication ........ 158
3.5.4 Ribonucleoprotein assembly ........ 159
3.5.5 L protein interactions with host translation factors ........ 163

**4. DISCUSSION ........ 165**

**5. ACKNOWLEDGEMENTS ........ 177**

**6. REFERENCES ........ 179**

# 1. INTRODUCTION

## 1.1 Bunyaviruses

*Bunyaviridae* family was defined in 1973 and at that point comprised a single *Bunyavirus* genus (Porterfield et al., 1975). Its name comes from the Bunyamwera virus (BUNV), isolated in 1943 during yellow fewer in Uganda (Smithburn et al., 1946). Since then more than 350 isolated viruses have been identified which makes *Bunyaviridae* the largest and the most diverse family of segmented negative-strand RNA viruses (sNSV), after *Arenaviridae* and *Orthomyxoviridae*. Apart from being the causative agents of serious zoonotic human diseases, bunyaviruses infect a wide range of crops and cattle, which makes them an important health and economic threat.

Bunyaviruses are enveloped viruses with the RNA genome (vRNA) divided into three segments and coding, depending on the genera, from four to six viral proteins. Depending on the size of these genomic segments, the size of viral structural proteins, and the consensus sequences of the 3’ and 5’ vRNA extremities bunyaviruses can be subdivided into five genera: *Orthobunyavirus*, *Hantavirus*, *Nairovirus*, *Phlebovirus* and *Tospovirus* (Fig. 1.1).

This chapter describes the current understanding of cellular and molecular biology of bunyaviruses putting the main emphasis on the genome organization and its transcription and replication mechanisms. Several comprehensive and broader reviews on bunyavirus biology are available (Elliott, 2014; Plyusnin and Elliott, 2011; Walter and Barr, 2011).

### 1.1.1 Classification, distribution, and transmission

Bunyaviruses are arthropod-borne viruses (with the exception of rodent-borne hantaviruses) and their worldwide distribution is limited to the natural habitat of their insect hosts, usually mosquitos. In most cases a particular bunyavirus can be transmitted by only one arthropod species. This high dependence on the insect vector greatly limits the ecological niche of each bunyavirus, leading on the other hand to the efficient viral emergence and re-emergence each time the reservoir vector adapts to a new environment.

**A**

| RNA segment | Genus | | | | |
|---|---|---|---|---|---|
| | *Orthobunyavirus* | *Hantavirus* | *Nairovirus* | *Phlebovirus* | *Tospovirus* |
| L | 6.9 | 6.5 | 12.2 | 6.4 | 8.9 |
| M | 4.5 | 3.6 | 4.9 | 3.5 | 4.8 |
| S | 1.0 | 1.7 | 1.7 | 1.7 | 2.9 |
| Total | 12.4 | 11.8 | 18.8 | 11.6 | 16.6 |

**B**

| Structural proteins | Genus | | | | |
|---|---|---|---|---|---|
| | *Orthobunyavirus* | *Hantavirus* | *Nairovirus* | *Phlebovirus* | *Tospovirus* |
| L | 250 | 250 | 450 | 250 | 330 |
| Gc | 110 | 55 | 75 | 65 | 75 |
| Gn | 35 | 70 | 35 | 55-70 | 45 |
| N | 25 | 50 | 50 | 30 | 30 |

**C**

| | |
|---|---|
| *Orthobunyavirus* | 3′ UCAUCACAUGA---------UCGUGUGAUGA 5′ |
| *Hantavirus* | 3′ AUCAUCAUCUG------------AUGAUGAU 5′ |
| *Nairovirus* | 3′ AGAGUUUCU-------------AGAAACUCU 5′ |
| *Phlebovirus* | 3′ UGUGUUUC---------------GAAACACA 5′ |
| *Tospovirus* | 3′ UCUCGUUA--------------CUAACGAGA 5′ |

**Figure 1.1** *Bunyaviridae* genera classification based on RNA segments length, viral protein size, and conserved sequence of genomic ends; adapted from (Plyusnin and Elliott, 2011). A) Genomic segments length (given in kb). B) Sizes of viral structural proteins (given in kDa). C) Consensus 3′ and 5′ vRNA terminal sequences

*Orthobunyavirus* genus is the largest one within *Bunyaviridae* family, and includes more than 170 named viruses, grouped into 48 species, and 18 serogroups. Initally named *Bunyavirus* it was renamed in 2005 in order to avoid confusion while referring to the entire bunyaviral family (Nichol et al., 2005). Orthobunyaviruses are mostly transmitted by mosquitos from the *Aedes* genus, with some being transmitted by *Haemaphysalis* ticks or *Culicoides* midges. A good example of newly emerged orthobunyavirus is the Schmallenberg virus (SBV) which causes a significant threat in western european farms since summer 2011, infecting cattle, sheep and goat herds (Hoffmann et al., 2012). On the other hand, viruses belonging to the California serogroup (e.g. La Crosse virus (LACV)) cause encephalitis and meningitis mainly in children and young adults (Haddow and Odoi, 2009).

Hantaviruses, unlike other bunyaviruses, are mainly rodent-borne viruses. The *Hantavirus* genus has been added to the *Bunyaviridae* family in 1987 (Schmaljohn and Dalrymple, 1983). Hantaviruses are distributed worldwide and can generally be divided into Old World hantaviruses, causing haemorrhagic fever with renal syndrome (e.g. Hantaan virus (HTNV) with 50% mortality rate), and New World hantaviruses, causing hantavirus pulmonary syndrome – e.g. Sin Nombre virus (SNV) and Andes virus (ANDV). Hantavirus reservoir hosts belong mainly to the families *Cricetidae* and *Muridae*, of the order *Rodentia*, but some viruses were also isolated from mammals belonging to the order *Soricomorpha*. Hantaviruses have the ability to adapt their viral cycle to either mammalian reservoir or final host – establishing persistence in rodents and causing disease in humans. There are numbers of studies addressing virus-host co-evolution of hantaviruses (Nemirov et al., 2004; Plyusnin and Morzunov, 2001).

Nairoviruses are transmitted principally by ixodid ticks and are restricted to Africa and Eurasia. The best studied representative of the genus, the Crimean Congo haemorrhagic fever virus (CCHFV), is a highly pathogenic, BSL-4 virus with 50% mortality rate in man. It has initially been described in 1944 during an epidemic in Crimea. Since then its geographic distribution appeared to range from western China through southern Europe, to south Africa, being probably the biggest known viral distribution for a nonavian visrus (Swanepoel et al., 1983). Nairobi sheep disease virus (NSDV) – another example of a highly pathogenic agent from *Nairovirus* genus – causes up to 90% mortality rate among domestic animals (Marczinke and Nichol, 2002).

Phleboviruses are emerging human pathogens with Rift Valley fever virus (RVFV) causing high mortality haemorrhagic fevers in both Africa and Middle East. In 1991 viruses related to the Uukuniemi virus (UUKV) were reclassified and added to the *Phlebovirus* genus (Francki et al., 1991; Palacios et al., 2013). Phleboviruses can be further classified based on the arthropod vectors that transmit them. Sandfly fever viruses (SFV), Uukuniemy virus (UUKV), and Rift Valley fever virus are transmitted by *Phlebotomus* sand flies, *Haemaphysalis* ticks or *Aedes* mosquitoes respectively. The type of the arthropod vector affects the biology of phleboviruses.

Tospoviruses are the only bunyaviruses infecting plants, threathening global agricultural productivity (Pappu et al., 2009; Turina et al., 2012). The *Tospovirus* genus is the newest one established, and initially contained only the tomato spotted wilt virus (TSWV). Thrips – plant-feeding arthropods from the family *Thripidae* – are the major vectors of tospoviruses. Recent introduction of one of them in Europe – the *Frankliniella occidentalis* thrip – allowed the tospovirus spread in the Mediterranean, leading to one of the most serious threats to vegetable crops in this region.

### 1.1.2 Viral cycle

Bunyaviruses enter the cell by means of the endocytosis, although detailed entry strategies may differ between the genera. Firstly the Gn/Gc glycoproteins have to be recognized by various cell surface receptors. Hantaviruses requires receptors from the integrin family (Gavrilovskaya et al., 1998), while phleboviruses and orthobunyaviruses entry is promoted by DC-SIGN (Lozach et al., 2011). In the next step, as shown for orthobunyaviruses, clathrin-mediated endocytosis takes place (Hollidge et al., 2012; Santos et al., 2008). Following internalization the endosome undergoes the acid-activated transition from its early to the late stage. At pH of around 5.5 the viral membrane fuses with the endosomal one which releases the viral genomes to the cytoplasm (Lozach et al., 2010). Each genomic segment is attached to the viral polymerase (the L protein), and coated by numerous nucleoproteins (the N proteins). Together they form ribonucleoproteins (RNP) – the functional RNA synthesis units.

Upon entry to the cell and release of the RNPs the primary transcription occurs. During this process negative sense viral genomes (vRNA) can be used directly as templates to produce the viral mRNA in a single RNA synthesis round. Production of the viral mRNA is initiated via the cap-snatching – the sNSV-specific mechanism which involves binding of host capped mRNAs to the RNPs, cleavage of these RNAs close to the 5' cap by a viral endonuclease intrinsic to the L protein, and use of the short capped fragments as primers for viral mRNA transcription (Morin et al., 2010; Plotch et al., 1981; Reguera et al., 2010; Reich et al., 2014). Transcription stops before the end of the viral template, resulting in a shorter 3' end which, in case of bunyaviruses, is not poly-adenylated (Bouloy et al., 1990; Collett, 1986; Hutchinson et al., 1996). The viral mRNA is then translated by the cellular machinery producing the viral proteins.

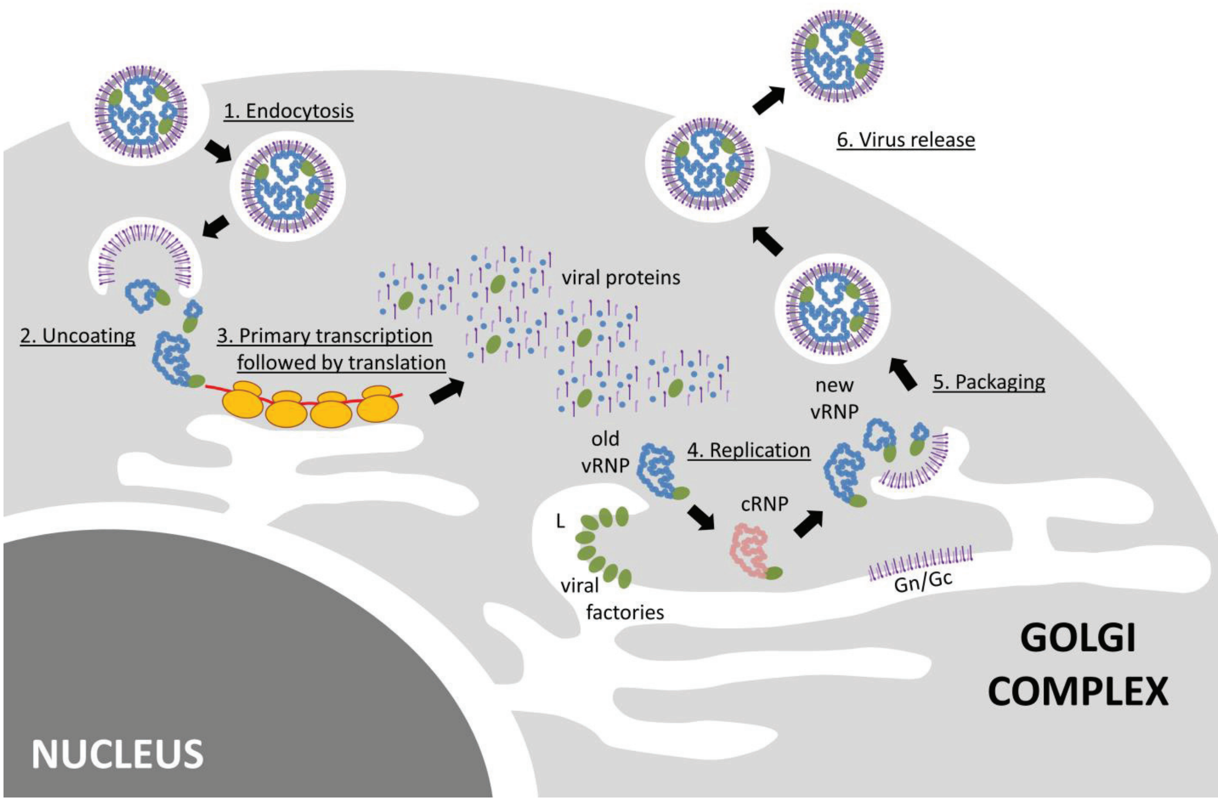



**Figure 1.2** Model of the bunyavirus infection cycle. After clathrin-mediated endocytosis (1), and pH-dependent fusion of the viral and endosomal membranes, the vRNPs are realeased to the host-cell cytoplasm (2), where they undergo primary transcription enabling translation of the viral proteins by the host machinery (3). Accumulation of the viral proteins (the L and N proteins) switches on replication of the vRNPs through the intermediate cRNPs (4). Viral replication is thought to occur within the viral factories localized in the Golgi complex. Gn/Gc glycoproteins embedded in the Golgi membranes enable packaging of the multiplied vRNPs (5). Newly formed virions are transported inside the vesicles to the periphery of the cell, and released (6).

In the following phase of infection, accumulation of the viral L and N proteins enables replication of the genomic segments to take place. Unlike one-step transcription, replication requires the synthesis of a positive sense, complementary RNA (cRNA) which is then used as the template to produce vRNA. Moreover, in contrast to transcription, replication initiates *de novo* and results in full copies of genomic segments. During both cRNA and vRNA synthesis the emerging nascent RNA strand gets coated by N proteins. A distinctive feature of *Bunyaviridae* is that the polymerase performs both transcription and replication in the cytoplasm, similarly to the *Arenaviridae*, and differently from the *Orthomyxoviridae*, whose RNPs are transcribed and replicated inside the nucleus.

Electron microscopy studies using cells infected with Bunyamwera virus (BUNV) revealed formation of unique structures within the Golgi complex – the viral factories. It is postulated viral replicating machineries are localized within these membrane compartments, which results in local accumulation of the newly produced RNPs, without the need of transferring them to the assembly and packaging sites (Fontana et al., 2008; Novoa et al., 2005). Nature of these factories is different in mammalian and mosquito cells (López-Montero and Risco, 2011)

The final phase of infection is the assembly of newly produced viral RNPs into virions. In mammalian cells they are internalized to the lumen of Golgi complex (Fontana et al., 2008). Afterwards they bud into secretory vesicles inside which they travel towards the plasma membrane to be released. It remains an open question how the RNPs representing three genomic segments are properly co-packaged into the same virion. Studies performed on RVFV suggest that elements responsible for coordinated packaging of viral segments lies within their non-coding ends (Terasaki et al., 2011). Successful generation of an artificial 4-segment RVFV proved a certain flexibility in vRNPs packaging (Wichgers Schreur et al., 2014).

### 1.1.3 Lytic vs. persistent infection

Insects are the natural reservoir vectors for most bunyaviruses (except hantaviruses), with mosquitos from the *Aedes* genus being the most frequent bunyavirus carriers. Humans become infected when bitten by the appropriate insect vector. There is almost no evidence for horizontal human-to-human transmission, with some exception like the Andes hantavirus (ANDV) (Padula et al., 1998), or CCHFV being nosocomially transmitted during surgical procedures (Mardani et al., 2009). There is also very few cases of vertical human-to-human transmission (Adam and Karsany, 2008). It is thus generally agreed that human infection is a dead-end event. Instead, bunyaviruses can be back-transmitted from human to the insect vector during blood-feeding by an arthropod. Within insects bunyaviruses can be spread both transovarially (vertical transmission) and venereally (horizontal transmission). The former way of inter-arthropod transmission allows bunyaviruses to survive winter seasons. Viral infectious cycles are adapted to the host type and vary in the outcome. Unlike lytic infection of the vertebrate host cells, in insect reservoir host vector cells bunyaviruses usually cause mild and persistent infection. This adaptation for persistence is most likely the effect of the long host-virus co-evolution. Studies comparing LACV infection in mammalian BHK cells and insect C6/36 cells revealed that viral replication but not transcription is down-regulated in the insect cells. It was suggested that expanding pool of nucleoprotein coats at some point its own mRNA leading to the nucleoprotein translation shutdown (Hacker et al., 1989; Rossier et al., 1988).

### 1.1.4 Organization of the viral genome

Single-stranded negative sense RNA viruses can have their genomic material either as a single RNA molecule, as in case of the order Mononegavirales (comprising the rabies virus (RSV), the vesicular stomatitis virus (VSV), or the *Filoviridae* ebolavirus), or divided into several segments – two, three, and eight for *Arenaviridae*, *Bunyaviridae*, and *Orthomyxoviridae* respectively. The genome segmentation provides several advantages. Since the polymerase mutation rate in negative-strand RNA viruses is very high (around $1.0 \times 10^{-4}$ mutations per base (Drake, 1993)), the longer the genome is, the higher number of mutations it accumulates. The genome segmentation permits the reassortment of different segments when the cell is co-infected by several viruses. This greatly boosts genetic diversity and ensures elimination of genomic segments with deleterious mutations from the pool (Pressing and Reanney, 1984).

Having several shorter genomic segments might by beneficial for packing into virions from the ergonomic point of view. However it would require a sophisticated packaging system, ensuring that each genomic segment is packaged into the same virus.

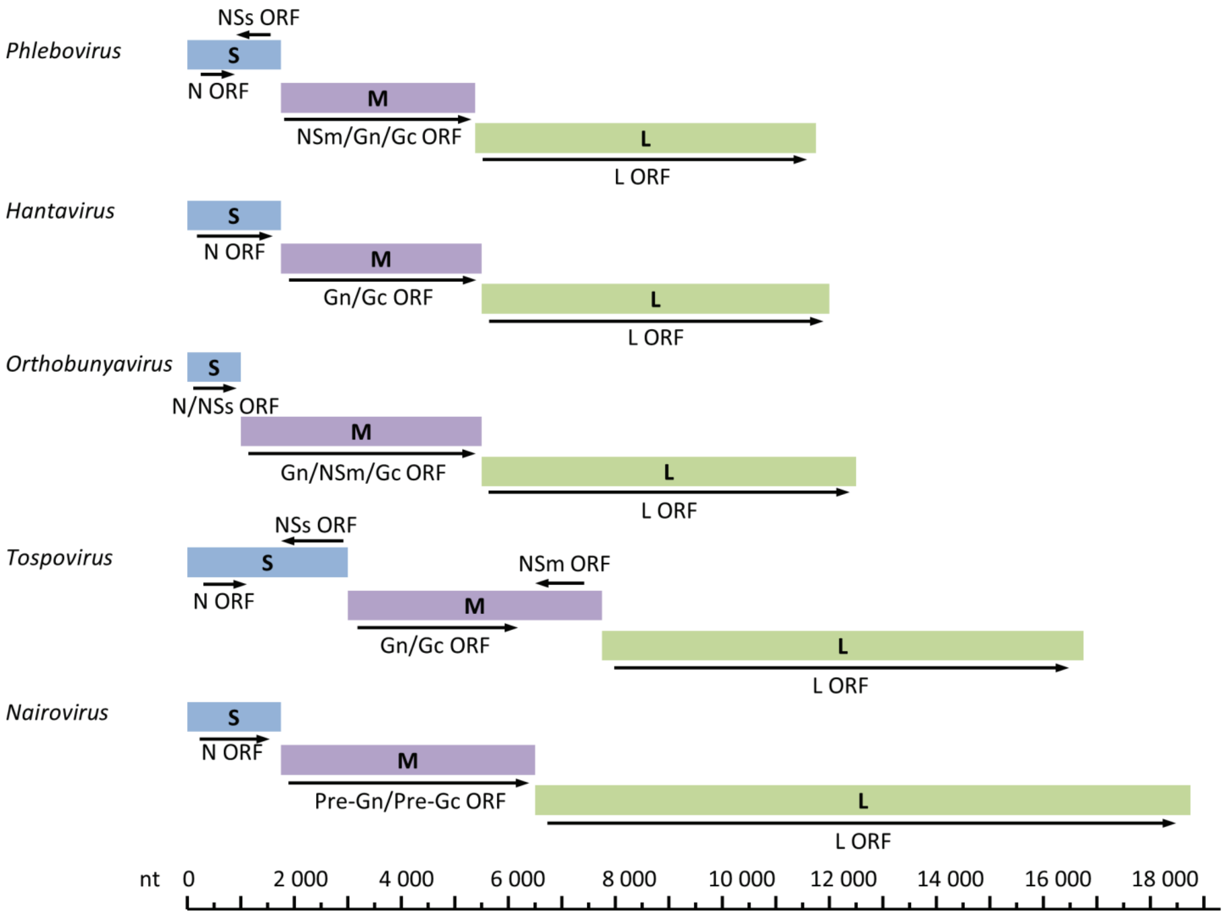



**Figure 1.3** Bunyavirus genomic segments. Open reading frames of the proteins (ORF) are marked together with the coding strategy (black arrows). Phleboviruses and Tospoviruses exhibit the ambisense coding strategy. Size of the segments is reflected by the scale (bottom). Adapted from (Walter and Barr, 2011).

Presence of the three *Orthobunyaviridae* genomic segments was revealed by electron microscopy studies (Obijeski et al., 1976; Pattnaik and Abraham, 1983). In order to highlight the difference in their size the genomic segments were called large (L), medium (M), and small (S). Their exact length varies and is genus-dependent (Fig 1.1.B, Fig. 1.3). The S segment codes for N protein and in some genera for NSs protein (Akashi and Bishop, 1983; Bouloy et al., 1984; Gentsch and Bishop, 1978; Gerbaud et al., 1987), the M segment codes for glycoproteins Gn and Gc and for NSm protein in some genera (Eshita and Bishop, 1984; Gentsch and Bishop, 1979; Grady et al., 1987; Lees et al., 1986; Pardigon et al., 1988), and the L segment codes for L protein – the polymerase (Fig. 1.3).

There is high level of similarity throughout the five *Bunyaviridae* genera concerning genetic organization of the segments and coding strategies they exhibit (Bishop et al., 1982; Clerx-van Haaster et al., 1982). In most of the cases viral proteins are coded in a negative-sense manner. However, some viral mRNAs are transcribed also from the positive-sense templates. This ambisense coding strategy is used by phleboviruses and tospoviruses to code for non-structural proteins NSs and NSm. An artificially generated RVFV phlebovirus, with swapped ORFs coding for N and NSs proteins, showed that in consequence the expression level of these proteins is altered (Brennan et al., 2014). In orthobunyaviruses the S segment codes both N and NSs protein and their ORFs overlap. The NSs protein is translated from the same viral mRNA as the N protein, but an alternative AUG start codon is used (Fuller et al., 1983). An exception from that rule was noticed in Brazoran virus which contains an enlarged S segment where the NSs ORF precedes that of the significantly larger N protein (Lanciotti et al., 2013). Successful approach to generate an artificial BUNV orthobunyavirus in which the NSs protein was coded in the antisense manner within the S segment proved certain plasticity of the BUNV genome (van Knippenberg and Elliott, 2015).

Sequencing studies revealed that bunyavirus genomic segments are linear and possess 3′ and 5′ non-translated ends (NTRs) (Haaster and Bishop, 1980; Obijeski et al., 1980). The NTRs length is genus and segment specific but the 5′ NTR is usually twice longer than the 3′ NTR. In case of LACV the 3′ NTRs are 60 nucleotides long in both M and L segments, and 80 nucleotide long in the S segment. The LACV 5′ NTRs are 125, 140, and 195 nucleotides long in L, M, and S segments respectively. A characteristic feature of bunyavirus genomic NTRs is their high conservation and extended complementarity between first 20-30 nucleotides of 3′ and 5′ ends. In *Orthubunyavirus* genus the sequence of initial 11 nucleotides is exactly the same for all three segments. Following nucleotides are also conserved but segment specific.

Despite linear nature of the bunyavirus genomic RNA segments, electron microscopy analysis of extracted virions revealed that genomic segments are circular (Pettersson and von Bonsdorff, 1975). It was shown that the circularity was not due to covalent closure (Dahlberg et al., 1977). Instead, the RNAse A protection assays performed on purified genomic RNA suggested formation of double-stranded structure (Patterson et al., 1983). The existence of

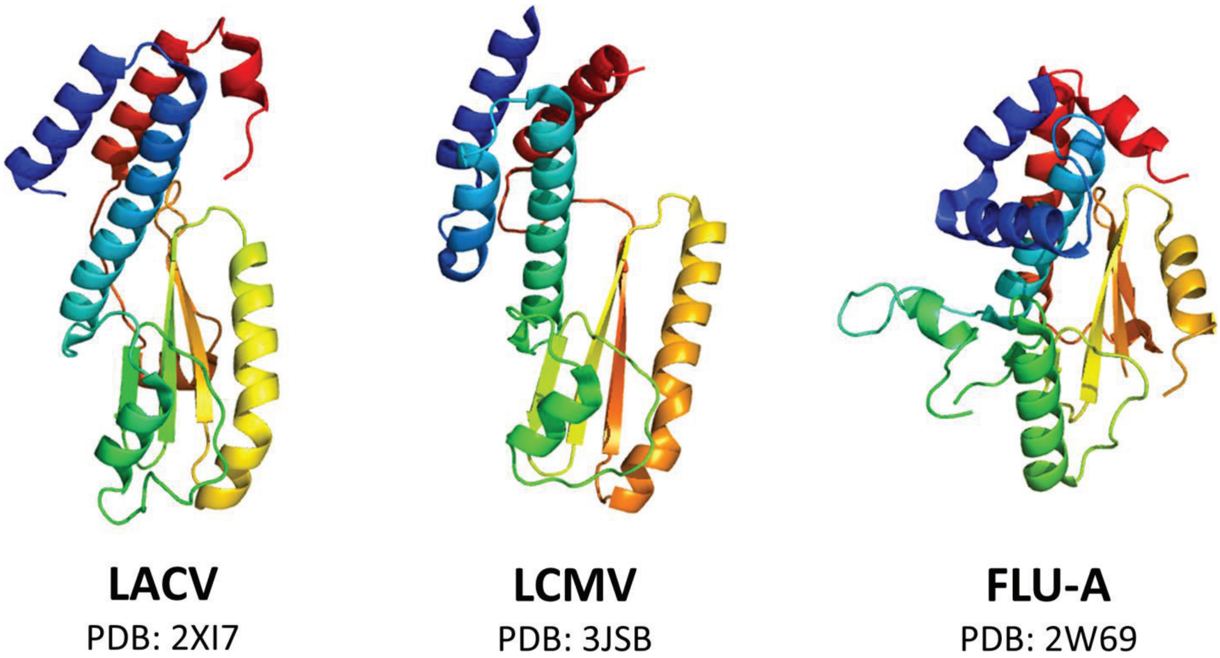



**Figure 1.4** Comparison of endonuclease domains from polymerases of viruses belonging to three sNSV families – LACV (Reguera et al., 2010), LCMV (Morin et al., 2010), and Influenza A (Dias et al., 2009). Proteins are rainbow-colored highlighting the N terminus (blue) and the C terminus (red).

such dsRNA structure was further confirmed *in vivo* by primer extension upon psoralen crosslinking, although it did not exclude that the very NTR ends remain single-stranded and primer extension is stopped by the duplex region distal from the ends (Raju and Kolakofsky, 1989).

### 1.1.5 L protein – the polymerase

The L segments codes for a single, large L protein which is the viral RNA dependent RNA polymerase (RdRp). Its size ranges from 250 kDa in *Orthobunyavirus*, *Hantavirus*, and *Phlebovirus*, to 330 in *Tospovirus* kDa, and 450 kDa in *Nairovirus*. The L proteins of bunyaviruses are multifunctional proteins able to interact with vRNA and cRNA promoters, to transcribe the viral mRNA through the cap-snatching mechanism, and to copy the viral genome by *de novo* initiated replication. Multiple sequence alignment of RdRps belonging to several groups of negative strand RNA viruses highlighted a region containing strongly conserved polymerase catalytic motifs F (preA), A, B, C, D, and E (Müller et al., 1994; Poch et al., 1990). Mutagenesis of conserved residues within these motifs abolishes RNA synthesis activity of the BUNV L protein (Jin and Elliott, 1992).

Apart from the central catalytic RdRp domain, the bunyavirus L proteins contain the cap-snatching endonuclease domains located at the N terminus (Patterson et al., 1984). The structure of LACV endonuclease (Reguera et al., 2010) exhibits similar fold to the endonucleases from Lymphocytic choriomeningitis (LCMV) or Lassa arenaviruses (Morin et al., 2010; Wallat et al., 2014), and the Influenza A PA subunit endonuclease (Dias et al., 2009) (Fig. 1.4). Mutagenesis of the conserved catalytic endonuclease residues results in a lack of viral transcription in the LACV minireplicon system (Klemm et al., 2013). Multiple sequence alignements indicate that all bunyavirus endonucleases are related to the type II endonucleases with a common active site motif PD…(D/E)XK.

The L proteins belonging to the genera *Nairovirus* and *Tospovirus* are significantly bigger than the ones from remaining bunyaviruses. Apart from catalytic polymerase motifs and endonuclease domain sequence alignments reveal additional modules. Studies with CCHFV nairovirus resulted in a crystal structure of one of these modules – a cysteine de-ubiquitinating motif (DUB) assigned to the ovarian tumour (OTU) superfamily (Capodagli et al., 2011; Frias-Staheli et al., 2007; James et al., 2011). It is proposed that the role of such domain in CCHFV L protein is to remove ubiquitin and ubiquitin-like modifications from host cell proteins, which in turn suppress the host-cell antiviral response by inhibiting the RIG I activation by ubiquitination (van Kasteren et al., 2012).

There are several conserved features suggesting that all sNSV polymerases share a common overall architecture. The monomeric *Bunyaviridae* and *Arenaviridae* polymerases and the heterotrimeric *Orthomyxoviridae* Influenza virus polymerase are similar in size. The Influenza virus polymerase PA, PB1, and PB2 subunits interact in the head-to-tail manner. The localization of the catalytic RdRp motifs is always central and the endonuclease domain is positioned at the N terminus of the polymerase. Altogether these features strongly suggest that all sNSV polymerases share common, modular architecture (Fig. 1.5). However it remains an open question whether bunyavirus L polymerase possess within its C terminus a cap-binding domain, homologous to the Influenza PB2 subunit cap-binding domain (Guilligay et al., 2008).

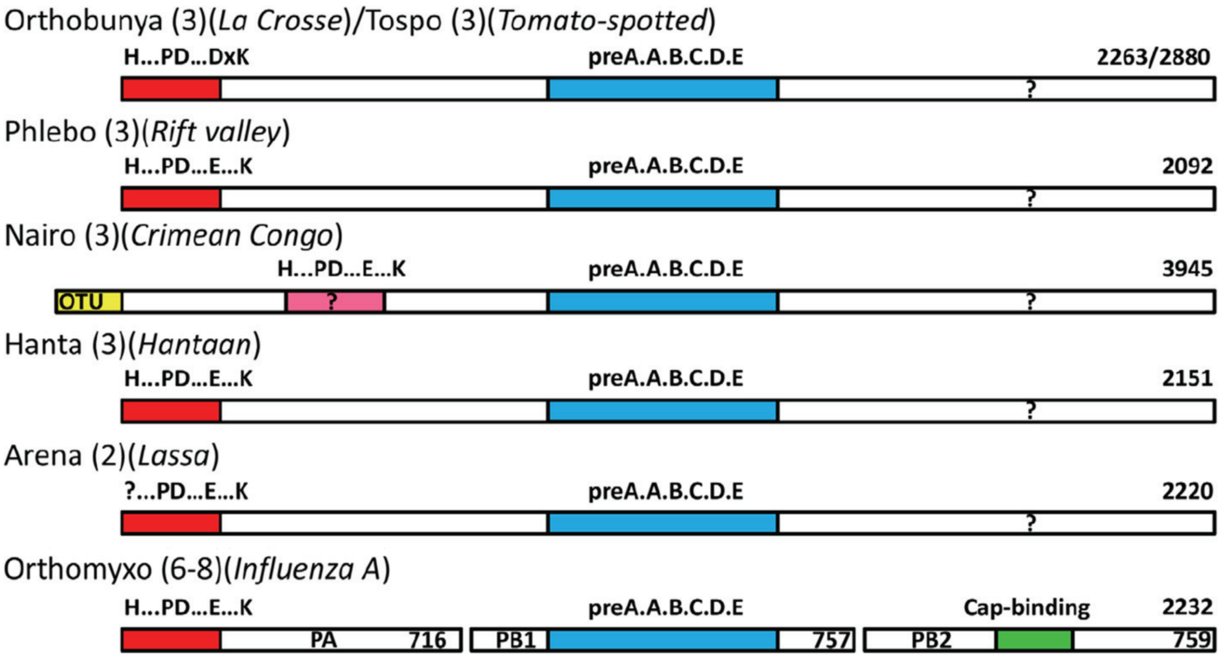



**Figure 1.5** Schematic diagram of the polymerase architecture of negative strand segmented viruses The polymerase proteins of different genera of negative strand segmented RNA viruses are represented as bars (not drawn to scale). The number of genome segments and an example species, the length of whose polymerase in amino acids is indicated at the right end, are indicated in brackets. The red bar at the left indicates the endonuclease domain with the particular sequence signature. The blue bar represents the conserved RdRp motifs in the central region. Three orthomyxovirus polymerase subunits PA, PB1 and PB2 are represented co-linearly. The green bar in PB2 represents the cap-binding domain, which is possibly located in the C-terminal region of the L-proteins of other sNSV (question mark). Figure and the caption to the figure are adapted from (Reguera et al., 2010).

### 1.1.6 N protein – the nucleoprotein

Bunyavirus nucleoprotein is the most abundantly produced viral protein during infection. It is critical for the genomic RNA replication since it coats both negative sense vRNA and positive sense cRNA, but not the viral mRNA. As revealed for BUNV the nucleoprotein preferentially binds single-stranded RNA and reveals higher affinity for the putative 5′ RNA encapsidation signal (Osborne and Elliott, 2000).

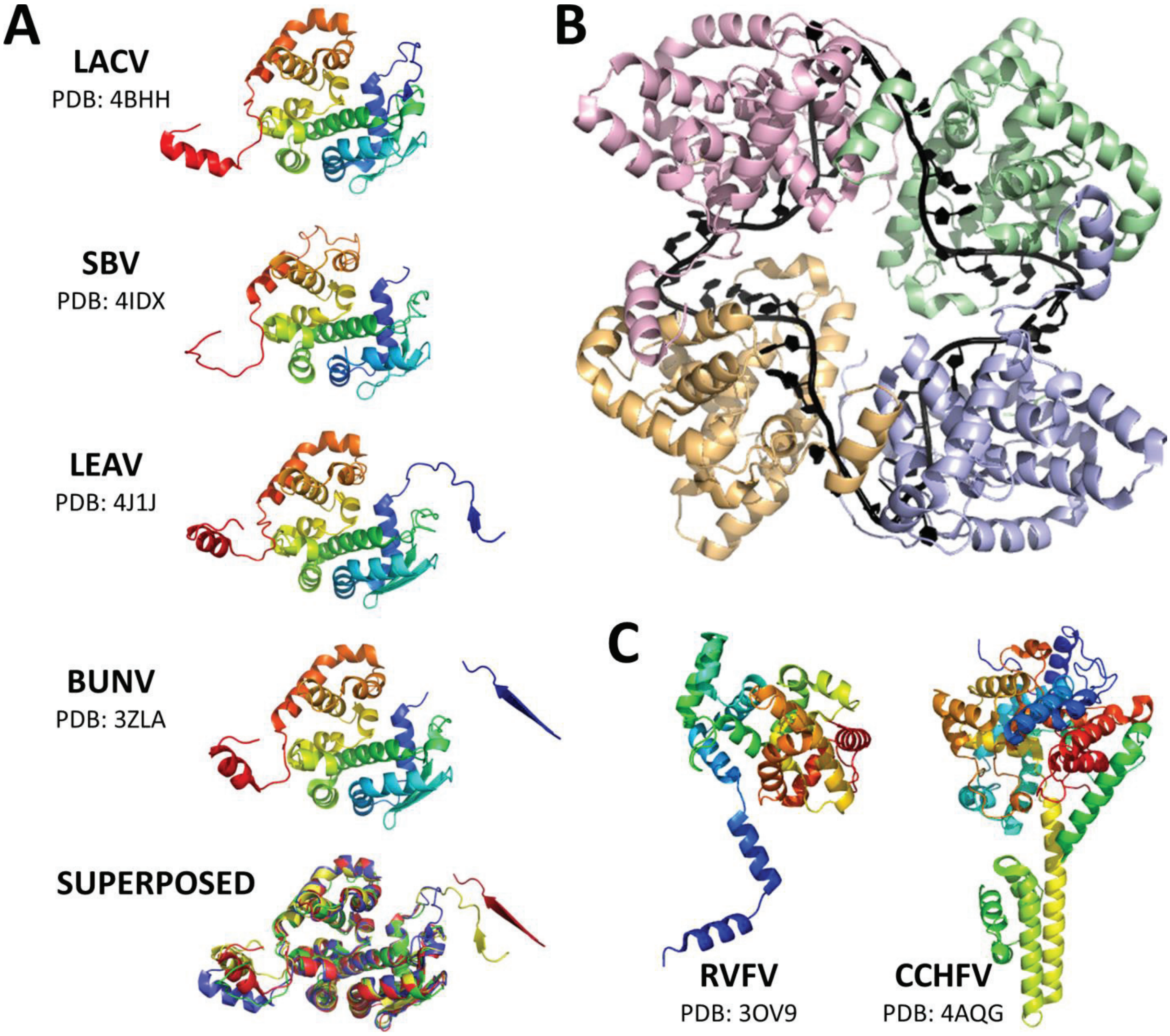



**Figure 1.6** Bunyavirus nucleoproteins. A) Structures of nucleoproteins belonging to *Orthobunyavirus* LACV (Reguera et al., 2013), SBV (Dong et al., 2013a), LEAV (Niu et al., 2013), and BUNV (Ariza et al., 2013). Proteins are rainbow-colored highlighting the N terminus (blue) and the C terminus (red). Bottom figure represents all four structures superposed – LACV N in blue, SBV N in green, LEAV N in yellow, and BUNV N in red. B) Tetrameric LACV N in complex with RNA. Single nucleoprotein sequesters 11 nucleotides. C) Structures of nucleoproteins belonging to *Phlebovirus* RVFV (Ferron et al., 2011) and *Nairovirus* CCHFV (Wang et al., 2012) Proteins are rainbow-colored highlighting the N terminus (blue) and the C terminus (red).

Atomic structures of nucleoproteins belonging to phlebovirus RVFV (Ferron et al., 2011; Raymond et al., 2010, 2012), nairovirus CCHFV (Carter et al., 2012; Guo et al., 2012; Wang et al., 2012), and more recently orthobunyavirus BUNV (Ariza et al., 2013; Li et al., 2013), LACV (Reguera et al., 2013), Schmallenberg virus (Ariza et al., 2013; Dong et al., 2013a, 2013b), and Leanyer virus (Niu et al., 2013) are available (Fig. 1.6). As reviewed in (Reguera et al., 2014; Ruigrok et al., 2011), unlike non-segmented NSV nucleoproteins, whose structures share the same fold, comparison of *Bunyaviridae* nucleoproteins with other available sNSV nucleoprotein structures – those from *Orthomyxoviridae* Influenza A (Chenavas et al., 2013; Ye et al., 2006), Influenza B (Ng et al., 2012), and infectious salmon anemia virus (Zheng et al., 2013), and those from *Arenaviridae* Lassa virus (Brunotte et al., 2011; Qi et al., 2010) and Junin virus (Zhang et al., 2013) – clearly shows that there is no common fold of the sNSV nucleoprotein. Instead, given different protein sizes, various oligomerization modes through flexible terminal arms, and different natures of the RNA-binding, it seems plausible that various sNSV families and even various genera belonging to the same family have developed their unique types of nucleoprotein-RNA interaction, resulting in different RNP architecture (Reguera et al., 2014; Ruigrok et al., 2011).

The RNA sequence specificity required for the BUNV N protein binding (Osborne and Elliott, 2000) contrasts with the unspecific nature of other bunyavirus N protein RNA binding revealed by X-ray crystallography. The *Orthobunyavirus* LACV and the *Phlebovirus* RVFV N proteins sequester 11 or 7 bases respectively in a sequence unspecific manner (Raymond et al., 2012; Reguera et al., 2010). It is proposed that nucleoproteins form an “armor” covering the genomic RNA. The assembly of nucleoproteins is stabilized by head-to-tail interactions mediated by the N and C terminal mobile arms which link the nucleoproteins together into long and flexible RNPs.

Although the major role of the sNSV nucleoproteins is the encapsidation of genomic RNA and thus formation of the RNPs, there are reports suggesting additional functions of nucleoproteins. Structure of the *Nairovirus* CCHFV N protein revealed a distinct metal-dependent binding site which turned out to have a DNA-specific endonuclease activity (Guo et al., 2012). Another example of the nucleoprotein with a nuclease activity comes from the

*Arenaviridae* family. The C-terminal domain of the Lassa virus N protein contains a 3'-5' RNA exoribonuclease activity, which is thought to suppress the interferon induction. Moreover, the same N protein has a deep cavity able to accommodate a nucleotide and possibly a m7GpppN cap structure, suggesting its role in the cap-snatching mediated transcription. Similar cap-binding function, although not supported by structural information, was proposed for the 50 kDa *Hantavirus* HANV nucleoprotein. It is thought that during translation-dependent transcription the HANV N protein replaces entire eIF4F translation initiation complex, and uses the pool of the host mRNAs localized in P bodies, thus protecting them from decapping and degradation (Mir and Panganiban, 2008; Mir et al., 2008, 2010).

Bunyavirus nucleoproteins are also targets of the host-cell innate immune response. It was shown that the interferon-induced MxA protein can inhibit viruses belonging to various *Bunyaviridae* genera, possibly by affecting the N protein expression level and by direct interaction with it, which significantly reduces the pool of free N protein essential for replication (Andersson et al., 2004; Bridgen et al., 2004; Frese et al., 1996; Haller et al., 2007). Studies on the *Orthobunyavirus* LACV and BUNV revealed that MxA sequesters the N protein and co-localizes with it in cytoplasmic inclusions, thus repressing the viral replication (Kochs et al., 2002; Reichelt et al., 2004). Moreover, studies on the *Phlebovirus* RVFV showed that MxA targets as well the primary transcription at the early infection stage (Habjan et al., 2009).

### 1.1.7 NSs and NSm – the non-structural proteins

The main role of the non-structural protein NSs is to affect the host-cell innate antiviral immunity response, and it is achieved at the interferon (IFN) transcription level. In BUNV-infected mammalian cells NSs prevents the phosphorylation of serine-2 within the host-cell RNA Pol II CTD heptapeptide repeats which results in a non-specific down regulation of the host-cell mRNA (Thomas et al., 2004). Studies with LACV-infected cells suggest that NSs induces degradation of the RNA Pol II subunit RPBI (Verbruggen et al., 2011). Detailed studies using RVFV-infected cells revealed that NSs interacts with E3 ubiquitin ligase FBXO3 and recruits it to destroy the general TFIIH-p62 transcription factor by enzymatic removal of p62 (Kainulainen et al., 2014).

An interesting observation is that all nairoviruses and majority of the hantaviruses do not possess NSs protein. It is supposed that they evade the host-cell innate immunity using a different strategy. Genomes of hantaviruses end with the 5'-monophosphate instead of the 5'-triphosphate (Garcin et al., 1995). This is a side effect of the prime-and-realign mechanism (described in more detail below) and the benefit it gives is that hantavirus genomes avoid detection by the RIG-I RNA helicase, an interferon-inducing factor (Habjan et al., 2008).

Less is known about the NSm protein coded within the M segment. In *Tospovirus* it is coded in the ambisense manner and translated from its own mRNA, whereas in *Orthobunyavirus* and *Phlebovirus* it is produced together with both Gn and Gc glycoproteins as a polyprotein precursor. It is supposed that the BUNV NSm protein is required for virus assembly (Shi et al., 2006) and as a membrane protein localizes within the Golgi-related virus-induced elongated structures (Fontana et al., 2008).

Apart from already mentioned *Nairovirus* and *Hantavirus* there exist examples of viruses from other *Bunyaviridae* genera which lost the capacity to express both NSs and NSm proteins. This is the case for several orthobunyaviruses from the California serogroup or the recently described Herbert-like orthobunyavirus (Marklewitz et al., 2013).

### 1.1.8 Gn and Gc – the glycoproteins

Gn and Gc are membrane glycoproteins expressed as a polyprotein precursor, which depending on the genera contains as well the NSm protein. The polyprotein precursor co-transcriptionally inserted into the ER membrane. Upon proteolytic processing Gn and Gc proteins form a heterodimer which is later on targeted to the Golgi complex and retained there due to the Gn Golgi-retention signal (Andersson and Pettersson, 1998; Gerrard and Nichol, 2002; Matsuoka et al., 1994; Shi et al., 2004; Snippe et al., 2007).

Both Gn and Gc proteins bear the N-terminal glycosylation modification and are rich in evolutionally conserved cysteine residues, that form intermolecular and intramolecular disulphide bonds. The Gn-Gc heterodimer is essential for the virus assembly, and remaining anchored to the viral envelope it allows attachment to the target cells. It was shown that the Gn interacts with both L and N protein which would facilitate the packaging (Piper et al., 2011). The Gc on the other hand contains an N-terminal ectodomain essential to mediate cell fusion (Shi et al., 2009).

### 1.1.9 Role of genomic non-translated regions in RNA synthesis

The 3′ and 5′ non-translated regions (NTRs) are functionally important for many aspects of viral cycle. They comprise for instance signals for the RNP specific encapsidation (Osborne and Elliott, 2000), but also for genomic segments packaging into new virions (Kohl et al., 2006; Terasaki et al., 2011). Genomic NTRs are also important for RNA synthesis (Flick et al., 2004; Gauliard et al., 2006). Moreover, the 5′ end seems to be more critical for RNA synthesis since when shortened to 67 nucleotides it cannot rescue the virus, while 22-nucleotide short 3′ end still can (Lowen and Elliott, 2005). Similarly for the RVFV initial 13 nucleotides of the 3′ NTR end were sufficient for RNA synthesis (Prehaud et al., 1997). On the other hand artificial BUNV genomic segments containing only 25-nucleotide long 3′ and 5′ NTRs flanking the reference gene proved to be efficiently replicated and transcribed (Barr and Wertz, 2005). Detailed mutagenesis studies targeting the initial, highly conserved 11 nucleotides of the BUNV genomes proved that any modification within this region, even a double mutations maintaining the complementarity, greatly reduce RNA synthesis (Kohl et al., 2004). Similar mutagenesis studies leading to the same conclusion has been done for the UUKV (Flick et al., 2002). Mutagenesis studies on the BUNV NTRs showed as well that while shortening the 3′ NTR by a single terminal nucleotide does not affect RNA synthesis, it does so in case of the 5′ end (Kohl et al., 2004).

The difference between segments in promoting RNA replication was reported to be M > L > S. It was proposed that it is related to the NTR sequences after the initial 11 nucleotides (Barr et al., 2003). Shuffling of non-coding and coding regions between BUNV segments suggested a certain interplay between 3′ and 5′ ends (Barr and Wertz, 2004; Barr et al., 2003; Kohl et al., 2004; Lowen et al., 2005). Detailed investigation of these two aspects led to the conclusion that a short region of about 5 nucleotides, and following the initial 11 nucleotides, can be freely mutated but needs to form dsRNA in order to efficiently promote replication (Barr and Wertz, 2004; Kohl et al., 2004). The difference in sequence within this so called distal duplex modulated the level of RNA replication, suggesting that this might be a mechanism to control RNA synthesis level from different genomic segments.

Extensive nucleotide complementarity between the orthobunyavirus 3′ and 5′ NTRs ends is broken by a single mismatch resulting in G:U Wobble base pair at position 9. It was proposed that the difference between vRNA and cRNA at this position in the 3′ but not 5′ NTR end is the key factor promoting transcription and replication or only replication in case of vRNA and cRNA respectively (Barr and Wertz, 2005). Distal NTR regions of both 3′ and 5′ NTRs, but outside the initial 11 nucleotides, are also involved in the transcription-replication switch, but their contribution is less evident than it is in case of 3′ U9 from the wobble pair (Barr et al., 2005).

### 1.1.10 Cap-snatching initiates transcription

After entering the cell bunyaviruses stay in the cytoplasm when they perform the primary transcription and in a later infection phase – the replication. As described above 5′ end of the viral mRNA contains host-derived capped sequence (Bishop et al., 1983; Collett, 1986). At its 3′ end the viral mRNA is shorter relative to the template and may fold into stem-loop structures stabilizing the transcript in absence of the 3′polyadenylation (Abraham and Pattnaik, 1983; Hacker et al., 1990; Hutchinson et al., 1996). In order to produce such viral mRNA with nonviral 5′ ends, bunyavirus utilize the cap-snatching mechanism – using L protein endonuclease domain and putative cap-binding domain they steal capped RNA oligos from cellular mRNA (Patterson et al., 1984; Reguera et al., 2010).

It remains unclear whether bunyavirus polymerase exhibit any sequence preference for the host mRNA used for cap-snatching. There are reports suggesting that snatched primers able to form few base pairs with terminal nucleotides of the 3′ template end are selected more favourably. Studies on HANV mRNA revealed that majority of the host-derived 5′ sequences contained a G residue at position -1 (Garcin et al., 1995). Genome-wide RNAi screening in Drosophila cells identified Dcp2 as an antiviral factor, and revealed that RVFV performs cap-snatching on host mRNAs targeted for decapping and degradation. Sequencing of the viral mRNA showed no sequence preference for the host mRNA used in cap-snatching by RFVF (Hopkins et al., 2013). In contrast, studies of LACV mRNA 5′ ends in infected mosquito cell cultures (C6/36) and embryos revealed that initial sequence variability decreases during infection, leading to several predominant sequences present in the viral mRNAs in the later infection stages. Certain host-derived 5′ mRNA sequences were found to be preferentially used depending on the developmental stage of the infected insect (Dobie et al., 1997).

#### 1.1.11 Translation-dependent transcription

Another peculiarity about bunyavirus transcription is the need for ongoing translation. Translation ensures that the viral transcription will not finish prematurely. Since cellular mRNA appears always as RNP in the cytoplasm, it is possible that when performing cap-snatching bunyavirus directly hijack the entire translation machinery.

Initially, during characterization of Akabane virus RNA species, it was noticed that the cycloheximidine and puromycin – drugs inhibiting translation – decrease the level of RNA synthesis in both early and late phases of infection (Pattnaik and Abraham, 1983). This was further confirmed and small genome transcripts, products of premature transcription termination, were observed upon treating the cells with the same drugs (Patterson and Kolakofsky, 1984). As proposed later, the shutdown in viral transcription results in a decrease of available nucleoproteins which cause a subsequent decrease in the viral replication yields (Raju and Kolakofsky, 1986). Studies performed in vitro showed that purified virions need reticulocyte lysate for the transcription of complete mRNA. This activity can be blocked by the same set of translation-inhibiting drugs (Bellocq et al., 1987).

Studies on Germiston virus transcription revealed that scanning of the nascent mRNA by the small 40S subunit is sufficient to avoid premature termination (Vialat and Bouloy, 1992). This model was later confirmed by the introduction of additional translational stop codons into

BUNV genomic RNA (Barr, 2007). It is suggested that the role of the scanning ribosome is to grab the emerging viral mRNA, and to avoid RNA-RNA interaction between mRNA and the template. Another obvious advantage of this model is the efficient directing of the emerging viral mRNA for protein translation.

It cannot be excluded that additional mechanisms, other than RNA-RNA interactions, might cause the premature transcription termination. One possible mechanism could involve coating of the emerging mRNA by nucleoproteins like it happens in insect cells during persistent infection (Hacker et al., 1989). Interestingly, comparison of viral infections in both mammalian (BHK) and insect cells (C6/36) led to the intriguing discovery that translation requirement for complete viral mRNA synthesis is cell type dependent concerning only mammalian and not insect cells (Raju et al., 1989).

### 1.1.12 Prime-and-realign model for replication initiation

In contrast to primer-dependent transcription initiation, bunyaviral polymerases initiate replication de novo. Given that the conserved 3′ end of the template contains short, trinucleotide repeats (Fig. 1.1.C), the prime-and-realign strategy is proposed to facilitate the initiation step. In this model RNA synthesis is initiated internally – from the second terminal repeat – to generate a short primer, which is then shifted and positions on the first terminal repeat. Upon this initial release and re-annealing of the primer, polymerase proceeds to the elongation phase of RNA synthesis.

Studies of HANV genomic sequence revealed that the terminal U residue instead of the 5′ triphosphate has a single phosphate group. This suggests that replication starts with another nucleotide – possibly a GTP which could be placed opposite to C at position 3+ in the template. Following addition of few nucleotides, the short primer is realigned at the template terminus, resulting in an overhanging GTP which is removed later on (Garcin et al., 1995). Analysis of the 3′ end sequence requirements for RVFV efficient RNA synthesis revealed that the terminal repeat GU can be omitted without affecting synthesis efficiency (Prehaud et al., 1997). Similar observations upon removal of a terminal nucleotide from the BUNV 3′ genomic end were made

(Kohl et al., 2004). This supports the prime-and-realign model and explains the importance of short terminal repeats in the template 3’ end. One or two nucleotides absent at the very 3’ template end do not affect internal initiation. Great advantage of such mechanism is that it ensures the preservation of genomic ends.

## 1.2 RNA-dependent RNA Polymerases

### 1.2.1 Primordial polymerases – from RNA to protein

It is very plausible that one of the first enzymatic reactions that appeared during early evolution of life was the genome replication ability of polymerases. It actually defined the frames for the evolution itself as we understand it today – the mutation-driven change in genomic sequences. Assuming that it was one of the core enzymatic activities one can presume that its chemistry is simple and universal. This hypothesis – proposed originally by Tom Steitz (Steitz, 1998) can be easily proven by high level of conservation in the polymerase reaction chemistry across different domains of life – prokaryotic, eukaryotic and viral, no matter if it concerns DNA or RNA polymerases.

Before protein enzymes appeared the polymerization reaction – according to the RNA world hypothesis – could only by carried by the RNA itself, either replicating other RNA molecules or having the self-replication ability. Such pioneer ribozymes catalyzing RNA-dependent RNA polymerization have gone extinct, or are still waiting to be discovered. However, successful attempts were pursued to artificially generate, using in vitro selection pressure, an RNA molecule able to catalyze repetitive formation of a 3’/5’-phosphodiester bond – the class I ligase ribozyme (Ekland and Bartel, 1996; Ekland et al., 1995). Structural studies of the class I ligase demonstrated compact and folded nature of this ribozyme and proved that catalytic magnesium ions are coordinated at the active site via backbone phosphates (Koldobskaya et al., 2011; Robertson and Scott, 2007; Shechner et al., 2009).

Primordial ribozymes possibly, similar to the artificial class I ligase, had to face an important obstacle – their requirements for replication were fundamentally incompatible. On one hand it is favorable to perform such NTP condensation reaction within small compartments defined by fatty acids, which ensures high local NTP concentration. On the other hand chemistry of this reaction requires high concentration of divalent metal ions like Mg2+, which in turn disintegrate fatty acid vesicles causing their precipitation. This could be overcome – as elegantly proposed in the screen of Mg2+ chelators – by adding citric acid to the equation (Adamala and Szostak, 2013). Two carboxyl groups of the citric acid establish a coordination network for magnesium ions preventing fatty acids from precipitation, but more importantly contributing to stabilization of the incoming NTP. It is not improbable that at the early days of life chelators like the citric acid were replaced by short acidic peptides – ancestors of the current polymerases (Milner-White and Russell, 2011).

### 1.2.2 Nucleotidyl transfer reaction – chemistry

All protein polymerases known to date share the same 'two-metal-ion' mechanism of nucleotide addition (Fig. 1.7.A). In most cases those metal ions are Mg2+. The first one – $Mg^{2+}$ A – is always coordinated within the active site by catalytically important aspartic acids. The second one – $Mg^{2+}$ B – is imported together with an incoming NTP. The role of the magnesium ion A is to decrease the affinity of the ribose 3'OH for its hydrogen atom within the growing RNA. This facilitates the $3'O^{-}$ attack on the α-phosphate of an incoming NTP. Upon this nucleophile attack a pentacovalent transition state is formed, stabilized by both magnesium ions. Once the phosphodiester bound is formed, magnesium ion B remains bound to the released pyrophosphate group which leaves the polymerase catalytic site (Steitz, 1998). Polymerases build the nascent RNA/DNA strand in the RNA/DNA template dependent manner. The chemistry of the nucleotide addition defines the 5'-3' directionality of the RNA/DNA chain extension. The template strand is 'read' in the reverse, 3'-5' direction, and the incoming NTPs/dNTPs are stabilized via Watson-Crick nucleotide base pairing – hydrogen bonds formed between guanine and cytosine or adenine and uracil/thymine.

### 1.2.3 Polymerases – overview and classification

The first polymerase discovered and purified was the DNA polymerase I belonging to Escherichia coli (Bessman et al., 1958). For its identification and characterization of magnesium ion and DNA template-dependent DNA synthesis Arthur Kornberg and Severe Ochoa obtained the Nobel prize in medicine in 1959. 30 years later the first crystal structure of a polymerase has been solved – the C-terminal part of DNA polymaresa I known as the Klenow fragment (Ollis et al., 1985). The structure revealed a general architecture which later on appeared to be shared between various polymerases. It can be described as the 'right-hand' fold composed of three subdomains – fingers, palm, and thumb. While fingers and thumb domain are mostly α-helical, the palm subdomain comprises a β-sheet, similar to the RNA-recognition motif fold (RRM). The palm domain contains majority of the catalytic residues, thus defining the active site of the polymerase. Based on this features and structural characteristics of palm subdomains polymerases can be divided into two major groups – the DNA/RNA polymerase superfamily and the nuclotidyltransferase superfamily. The first group comprises DNA polymerase I, lesion bypass DNA polymerase, reverse transcriptase, T7 RNA polymerase, RNA-dependent RNA polymerase, and dsRNA phage RNA-dependent RNA polymerase.

### 1.2.4 Cellular RNA-dependent RNA polymerases

Cellular RNA-dependent RNA polymerases (cRdRp) play a role in RNA interference mechanism. They are supposed to amplify the RNAi signal producing dsRNA – substrate for Dicer – in the ssRNA template-dependent manner (Maida and Masutomi, 2011). Until now cRdRps have been found in fungi (QDE-1 in *Neurospora crassa* (Cogoni and Macino, 1999; Makeyev and Bamford, 2002), nematodes (EGO-1 in *Caenorhabditis elegans* (Maniar and Fire, 2011), insects (D-elp1 in *Drosophila melanogaster* (Lipardi and Paterson, 2009)), yeast (Rdp-1 in Schizosaccharomyces pombe (Sugiyama et al., 2005)), and various plant species. Structural analysis of the fungi QDE-1 revealed striking similarity to the catalytic core of the DNA-dependent RNA polymerases responsible for transcription (Salgado et al., 2006). It remains an open question whether analogical cellular RdRps exist in vertebrates.

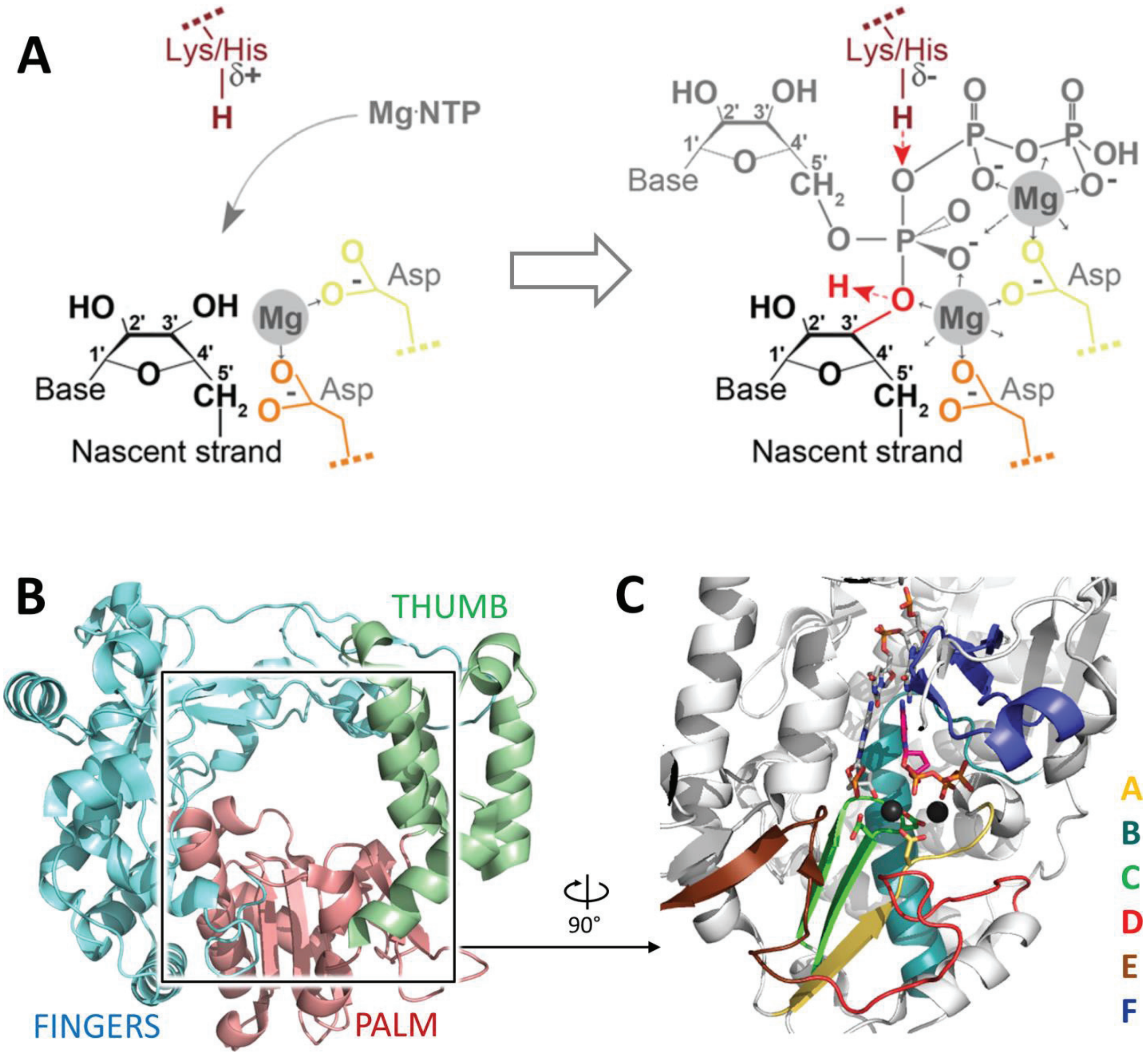



**Figure 1.7** A) Chemistry of the nucleotidyl transfer reaction, adapted from (te Velthuis, 2014). B) Poliovirus 3Dpol structure; fingers, palm, and thumb domains are highlighted in blue, red, and green respectively; PDB: 3OLB (Gong and Peersen, 2010); C) the same structure coloured in light grey with the RdRp catalytic motifs highlighted - F (dark blue), A (yellow), B (cyan), C (green), D (red), and E (brown); nucleotides of the template and product strands are marked as light grey sticks; the incoming NTP is marked as pink sticks; two catalytic metal ions are marked as grey spheres.

### 1.2.5 Viral RNA-dependent RNA polymerases

Viral RNA-dependent RNA polymerases (RdRp) are essential during the viral cycle since they can both transcribe and replicate the viral genome, which can be a double or single stranded RNA molecule (dsRNA or ssRNA respectively). In case of ssRNA genome, the RNA can be either a positive or a negative strand. Positive strand RNA genome is in the same time the viral mRNA which can directly be used for translation of the viral proteins encoded in it (e.g. *Flaviridae)*. In case of the negative strand RNA genome a transcription step is required in order to produce the viral mRNA before the translation can occur. Unlike majority of positive strand RNA viruses which genome is a single, long RNA molecule, the genome of negative strand RNA viruses can be either non-segmented (Mononegavirales, e.g Ebola virus), or divided into several segments. Based on the number of those segments (which could be regarded as viral chromosomes) the group of segmented negative-strand RNA viruses (sNSV) is further divided into *Arenaviridae* (e.g. Lassa virus), *Bunyaviridae* (e.g. Schmallenberg virus), and *Orthomyxoviridae* (e.g. Influenza A, B, and C virus), bearing two, three, or eight genomic segments respectively.

What all these different viruses share are the specialized RNA-dependent RNA polymerases using specific strategies in order to transcribe and replicate the viral genome. The only exception from this rule is the Hepatitis D virus (HDV) – a small RNA virus which does not have its own RdRp (Abbas and Afzal, 2013; Lai, 2005). RNA-dependent RNA polymerases are the template-dependent nucleotide transferases. Despite very little overall sequence conservation, they are similar in length (400-700 aminoacids) and share the conserved 3D architecture (Černý et al., 2014; Ferrer-Orta et al., 2006; Ng et al., 2008; te Velthuis, 2014). RdRps all adopt the same, right-hand like architecture which comprise three main subdomains – fingers, palm, and thumb (Fig. 1.7.B). Unlike other polymerases presenting more open conformation, RdRps adopt a closed-hand conformation. This is achieved due to the fingertips – an RdRp specific loops emerging from the fingers subdomain and connecting it with the thumb subdomain.

Up to date there is around 100 non-redundant, viral RdRp (EC=2.7.7.48) structures obtained by X-ray crystallography and deposited in the PDB database. This search has been performed for polypeptide chains longer than 300 residues. Polymerase structures with sequence similarity higher than 90% are represented by one hit. The list comprises La Crosse bunyavirus polymerase structures obtained during this PhD thesis. Statistics presented in Figure 1.8 highlight the number of RdRp structures from different viral source, protein sequence length, obtained resolution, and proportion of the structures containing protein alone or in complex with RNA. Table 1.1 gathers representative examples of those viral RdRp structures.

| STRUCTURES OF VIRAL RNA-DEPENDENT RNA POLYMERASES | | | |
|---|---|---|---|
| **Virus** | **Highlights** | **PDB codes** | **References** |
| Phage ϕ6 | $Mn^{2+}$ ion role in transition from de novo initiation to elongation | 4A8F – 4 A8Y | (Wright et al., 2012) |
| | Impact of $Mn^{2+}$ ion on the active site | 2JL9, 2JLF, 2JLG | (Poranen et al., 2008) |
| | $Ca^{2+}$ inhibits the RdRp and allows to capture it at certain state | 1UVI – 1UVM | (Salgado et al., 2004) |
| | Mechanism of initiation | 1HHS, 1HHT, 1HI0, 1HI1, 1HI8 | (Butcher et al., 2001) |
| Hepatitis C virus | Detailed study of initiation and initiation-replication transition | 4WTA – 4WTM | (Appleby et al., 2015) |
| | Elongation state; Δ-βhairpin construct | 4E76, 4E78, 4E7A | (Mosley et al., 2012) |
| | Resistance to non-nucleoside inhibitors | 3GSZ | (Rydberg et al., 2009) |
| | Non-nucleoside inhibitors profiling | 3HKW, 3HKY | (Nyanguile et al., 2010) |
| | Replication in strains J6 and JFH1 | 2XWH, 2XXD | (Schmitt et al., 2011) |
| | Highly potent irreversible inhibitors | 3TYQ | (Chen et al., 2012) |
| | Evidence for de novo initiation | 1NB4, 1NB6, 1NB7 | (O'Farrell et al., 2003) |
| | Complex with nucleotides and m.ions | 1GX5, 1GX6 | (Bressanelli et al., 2002) |
| | HCV NS5B structure at 2.8 Å | 1CSJ | (Bressanelli et al., 1999) |
| | HCV NS5B structure at 2.5 Å | 1QUV | (Ago et al., 1999) |
| | HCV NS5B structure at 1.9 Å | 1C2P | (Lesburg et al., 1999) |
| Poliovirus | Motif B loop mediates translocation | 4NLO – 4NLY | (Sholders and Peersen, 2014) |
| | Engineered RNA make crystal contacts | 4K4S – 4K4W | (Gong et al., 2013) |
| | Structural basis for active site closure | 3OL6 – 3OL9, 3Ola, 3OLB | (Gong and Peersen, 2010) |
| | 3CD – precursor of 3C(pro) and 3D(pol) | 2IJD, 2IJF | (Marcotte et al., 2007) |
| | Proteolysis-dep. Activation of 3D(pol) | 1RAJ, 1RA6, 1RA7, 1TQL | (Thompson and Peersen, 2004) |
| | Original 3D(pol) structure | 1RDR | (Hansen et al., 1997) |
| Foot-and-mouth disease virus | Mutant polymerases with reduced sensitivity to ribavirin | 3KLV, 3KMQ, 3KMS, 3KNA, 3KOA | (Ferrer-Orta et al., 2010) |
| | In crystallo RNA synthesis captured | 2E9R, 2E9T, 2E9Z, 2EC0 | (Ferrer-Orta et al., 2007) |
| | Apo / RNA-bound FMDV 3D structure | 1U09, 1WNE | (Ferrer-Orta et al., 2004) |
| Simian rotavirus | Apo / RNA-bound cage-like, 4-channel VP1 structure | 2R7Q – 2R7X | (Lu et al., 2008) |

| | | | |
|---|---|---|---|
| Norwalk virus | RNA-polymerase complex with substrate analog – putative inhibitor | 3H5X, 3H5Y | (Zamyatkin et al., 2009) |
| | RNA-polymerase complex in closed conformation | 3BSN, 3BSO | (Zamyatkin et al., 2008) |
| | Original norovirus RdRp structure | 1SH0, 1SH2, 1SH3 | (Ng et al., 2004) |
| Mammalian orthoreovirus | RNA synthesis in a cage | 1N1H, 1N35, 1N38, 1MUK | (Tao et al., 2002) |
| Human rhinovirus | Common cold antiviral therapy target | 1XR5, 1XR6, 1XR7 | (Love et al., 2004) |
| | Engineered RNA make crystal contacts | 4K50 | (Gong et al., 2013) |
| Japanese encephalitis virus | Motif F ordering upon GTP binding | 4HDG, 4HDH, 4MTP | (Surana et al., 2014) |
| | JEV NS5 MTase RdRp interface | 4K6M | (Lu and Gong, 2013) |
| Infectious pancreatic necrosis virus | Covalent RNA-polymerase complex formation through N-terminal serine | 2YI8, 2YI9, 2YIA, 2YIB | (Graham et al., 2011) |
| Infectious bursal disease virus | Birnavirus VP1 polymerase shows re-ordered catalytic motifs: C – A – B | 2PGG | (Pan et al., 2007) |
| Bovine viral diarrhea virus | Role of BVDV pol N-terminal domain | 2CJQ | (Choi et al., 2006) |
| | Role of GTP in *de novo* initiation | 1S48, 1S49, 1S4F | (Choi et al., 2004) |
| Dengue virus | DENV-3 polymerase/inhibitor structure | 3VWS, 4HHJ | (Noble et al., 2013) |
| | NS5 MTase requirements for cap analogs and capped RNA | 2P3L – 2P3Q, 2P40, 2P41 | (Egloff et al., 2007) |
| | DENV pol activity and atomic structure | 2J7U, 2J7W | (Yap et al., 2007) |
| La Crosse virus | First bunyavirus pol structure (3/4 of L protein); in complex with 3'/5' vRNA | 5AMQ, 5AMR | (Gerlach et al., 2015) |
| Influenza B virus | Structural insight into cap-snatching and RNA synthesis by influenza RdRp | 4WRT, 4WSA | (Reich et al., 2014) |
| Influenza A virus | Entire bat FluA heterotrimeric RdRp in complex with 3'/5' vRNA promoter | 4WSB | (Pflug et al., 2014) |
| Coxsackievirus | Engineered RNA make crystal contacts | 4K4X – 4K4Z | (Gong et al., 2013) |
| | Mutations in palm domain affetc fidelity of viral RdRps | 4WFX – 4WFZ | (Campagnola et al., 2015) |
| | VPg peptide – primer for initiation | 3CDU, 3CDW | (Gruez et al., 2008) |
| Rabbit Hemorrhagic fever virus | First calcivirus RdRp structure | 1KHV, 1KHW | (Ng et al., 2002) |
| Tomato mosaic virus | Structural basis of viral superfamily 1 helicases | 3VKW | (Nishikiori et al., 2012) |
| Phage Qβ | EF-Tu, EF-Ts, and ribosomal S1 are parts of Qbeta replicase | 4Q7J | (Takeshita et al., 2014) |
| | Elongation complexes with increasing length of the product | 3AVT – 3AVY | (Takeshita and Tomita, 2012) |
| | Qbeta replicase composed of both viral and host proteins | 3MMP | (Kidmose et al., 2010) |
| Infectious bursal disease virus | Activation mechanism of the IBDV VP1 polymerase | 2R70, 2R72, 2PUS, 2QJ1 | (Garriga et al., 2007) |
| West Nile virus | Active and inactive WNV NS5(pol) | 2HCN, 2HCS, 2HFZ | (Malet et al., 2007) |
| Sapporo virus | First structure of sapovirus 3D(pol) | 2CKW | (Fullerton et al., 2007) |

**Table 1.1** List of available viral RdRp X-ray structures

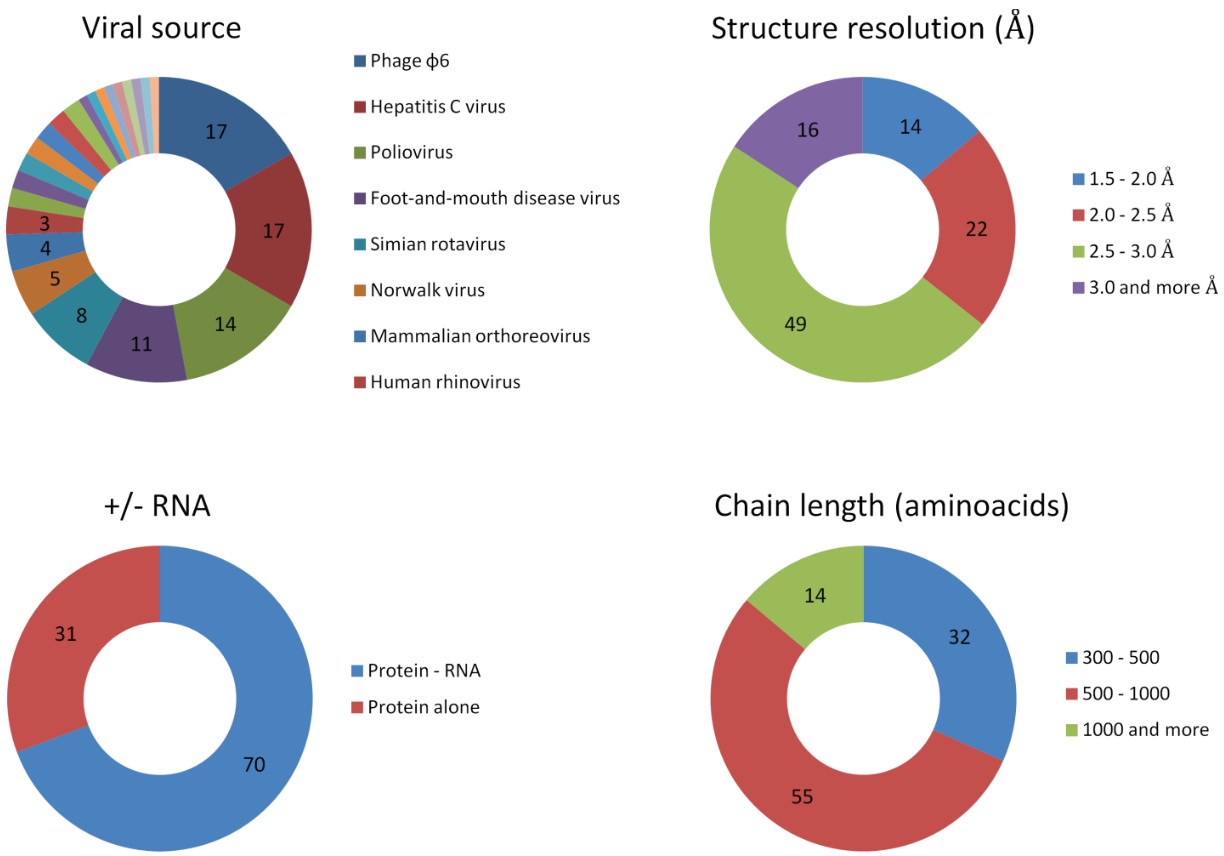



**Figure 1.8** Statistics concerning viral RdRp structures available in the PDB database. Search has been performed asking for the EC=2.7.7.48 viral proteins longer than 300 residues. Polymerase structures with sequence similarity higher than 90% are represented by one hit

### 1.2.6 RdRp – catalytic motifs

The central catalytic site of RdRp, unlike other regions of the polymerase, contains highly conserved elements also called the homomorphs – the catalytic motifs responsible for the chemistry of nucleotidyl transfer reaction (Fig. 1.7.C). Depending on the study up to seven motifs can be specified – G, F1-3 (pre-A), A, B, C, D, and E (Bruenn, 2003). Viruses belonging to *Birnaviridae* family (e.g. IBDV - infectious bursal disease virus, or IPNV – infectious pancreatic necrosis virus) are exceptional in having catalytic motif C preceding motif A (Gorbalenya et al., 2002). Those catalytic motifs adopt conserved secondary structures, specific for each of them. The role of the motifs is to define the shape of the internal polymerase cavity. They are directly responsible for the proper selection of an incoming NTP, and they orchestrate the RNA synthesis, thus defining the polymerase fidelity and processivity. As shown by mutagenesis studies of several viral RdRps, point mutations within conserved palm domain residues can affect the fidelity and the reaction rate, resulting as well in the change of viral pathogenesis and fitness (Arnold et al., 2005; Campagnola et al., 2015; Pfeiffer and Kirkegaard, 2005).

Motif A lies in the linker connecting one of the β-strands from the palm subdomain with an α-helix from the fingers subdomain. It contains a conserved, magnesium-binding aspartic acid which together with aspartic acids from motif C co-ordinates both divalent metal ions involved in catalysis. Studies of the poliovirus 3d polymerase suggests that only when the right incoming NTP, complementary to the template, is accommodated in the catalytic site, the proper coordination of both magnesium ions can be established. This in turn results in ordering of motif A β-strand, which can be aligned with the motif C β-strands, thus forming an RRM fold (Gong and Peersen, 2010). This elegant, structuring-driven mechanism, ensures that the catalytically competent state of the active site – often referred to as the 'closed complex' – can only be reached when the right NTP has entered the active site.

Motif B is composed of a flexible loop followed by a long helix. It contains conserved asparagine forming hydrogen bond with the ribose 2'OH group of the incoming NTP, thus discriminating rNTPs over dNTPs (Gohara et al., 2000). The long helix of motif B builds the bottom of the active site cavity. It allows correct positioning of the incoming template, and provides a stabilizing platform for the template-product duplex. Finally, polymerase state-dependent steric clashes between the template and the motif B, observed in Polio 3d(pol) structures (Sholders and Peersen, 2014), suggest that motif B mediates the polymerase translocation, contributing to the Brownian ratchet mechanism. Flexible nature of the motif B loop can be a drug target leading to mispositioning or the RNA template (Garriga et al., 2013)

Motif C is located in the β-turn of the palm domain. It is composed of the GDD triplet in case of dsRNA and positive strand ssRNA viruses, the GDNQ in case of non-segmented NSV viruses, or the SDD triplet in case of segmented NSV viruses. In all those cases both aspartic acids participate in coordination of the two magnesium ions used by polymerase to catalyze the nucleotidyl transfer reaction. The amino-terminal aspartate is critical since apart from positioning the α-phosphate of the incoming NTP via the magnesium ion, it also reduces the pKa of the 3'OH from the nascent strand. This facilitates the ribose attack on the incoming NTP α-phosphate and results in phosphodiester bond formation.

Motif D contains positively charged residues that contribute in generation of the basic environment within the NTP entry channel and are directly involved in positioning of incoming NTPs. Some studies propose that a one of its conserved protonated lysines, positioned close to the β-phosphate of an incoming NTP, may play a role of a general acid and protonate the PPi group upon release from the incorporated NTP (Castro et al., 2007, 2009; Yang et al., 2012). Other studies revealed that mutation of conserved residues from motif D both affects the fidelity of the polymerase and lowers the reaction rate (Liu et al., 2013; Shen et al., 2012). In fact the conformational change of motif D mediated by the interaction with the β-phosphate is believed to be one of the checkpoint of the right incoming NTP selection.

Motif E is located within a β-turn-β (similar to motif C). Its putative role is to position oligonucleotide primer or priming NTP. It does so by stabilization the NTP triphosphate like it was shown for *Flaviviridae* RdRp (Bressanelli et al., 2002; Choi and Rossmann, 2009)

Motif F (in some studies called also preA) is located within the fingertips in an elongated β-loop-β structure. It contributes to the NTP entry channel formation and its conserved, positively charged residues are able to neutralize phosphate groups of the incoming NTPs. One of the main observations unraveled by LACV L-vRNA structure is the allosteric ordering of the motif F, driven by 5' vRNA binding (Gerlach et al., 2015). Similar structuring has been observed in JEV NS5 polymerase upon GTP binding (Surana et al., 2014). This suggests that differently mediated motif F structuring might be a common feature of all RdRps.

### 1.2.7 *De novo* initiation – different strategies

Most of the RNA (and DNA) polymerases when they commence to build the nascent nucleic acid strand do not actually start from scratch. Instead, they extend an existing RNA (or DNA) strand which serves a role of the primer. Viral RdRps are the exception to that rule. They are indeed able to initiate the RNA synthesis *de novo* in absence of the primer. This very initiation step is critical and difficult to establish. It requires both priming and incoming NTPs properly positioned in the active site with all its elements in the exact orientation in order to allow the first nucloetidyl transfer reaction to happen. After this rate limiting step which results in the formation of a dinucleotide primer the polymerase can carry on with the RNA synthesis through the standard strand elongation.

Various RdRps have established different approaches to facilitate the *de novo* initiation. *Flaviviridae* for instance have a specific loop emerging from the thumb subdomain - β-hairpin in HVC (Hong et al., 2001), β-thumb in BVDV (Choi et al., 2004) – that occludes the active site cavity. The very tip of this structural element contains an aromatic residue – a tyrosine or a histidine – which by stacking interactions between its aromatic ring and the base of the priming NTP helps to hold this it in position.

Similar approach – utilizing the C-terminal domain to facilitate initiaton – is used by the dsRNA bacteriophage ϕ6 (Butcher et al., 2001). The template 3'end CTP (position T1) is initially locked in the pocket. A conserved aromatic amino acid (tyrosine or tryptophan) is a platform, providing stacking interaction to stabilize the incoming GTP, opposite to the T2 template-CTP. In the following step the template ratchets back, making the T1 template-CTP accessible for another incoming GTP. Following the phosphodiester bond formation, template-2xGTP duplex ratchets down displacing the C-terminal domain, and the elongation phase starts.

### 1.2.8 Additional domains and multimerisation

Transcription and replication of viral RNA genomes often require additional enzymatic activities beyond the ordinary nucleotide extension. In order to acquire them many viral polymerases possess extra domains flanking the RdRp domain. Segmented negative strand viruses from *Bunya*- and *Arenaviridae* families are very good example. Their single chain, 250 kDa or bigger, polymerases (also named L proteins) apart from having the RdRp in the middle, possess the N-terminal cap-snatching endonucleases providing capped RNA primers for transcription, derived from the host mRNA (Morin et al., 2010; Reguera et al., 2010). As discussed in 'Results' of this thesis – apart from the endonuclease domain, bunyaviral L protein contains as well domains responsible for genomic RNA ends recognition, and others contributing to RNA channels formation and stabilizing the RdRp domain itself (Gerlach et al., 2015).

Slightly different approach – for which Orthomyxoviridae Influenza virus is the best example – is to assemble the polymerase complex from several protein subunits. Segmented genome of Influenza virus codes for three such subunits – PA, PB1, and PB2 – each approximately 80 kDa in size. All subunits are transported inside the host cell nucleus (PA-PB1 as dimer separately from the PB2) where they assemble into heterotrimeric functional polymerase. Structural studies have shown the identity of the PA N-terminal cap-snatching endonuclease domain (Dias et al., 2009), PA-C domain which function was initially unclear (He et al., 2008), but also domains belonging to the PB2 subunit – cap-binding domain (Guilligay 2010), 627 domain, and NLS domain (Tarendeau et al., 2008). What for some time remained unclear was the organization of the PB1 subunit – predicted to contain the RdRp domain, and the overall architecture of entire heterotrimer. Structures of Influenza A and B complete heterotrimeric polymerases in complex with their vRNA promoters allowed to answer these questions (Pflug et al., 2014; Reich et al., 2014) (Fig. 1.9). Firstly it became clear that all three subunits are greatly intertwined, with the PB1 subunit containing indeed the classical RdRp domain, although with some sNSV specific structural elements. All three subunints participate in vRNA 3' and 5' ends binding with great contribution from the PA-C domain. PA-N endonuclease and PA-C domain are separated by the long flexible linker allowing to position them on the opposite sides of the PB1 RdRp. Tightly bound PA-C domain and PB1 subunit, together with the PB2 N-terminal region, constitute the rigid core of the Influenza polymerase. Other domains – PA-N endonuclease and PB2 C-terminal cap-binding, mid, 627, and NLS domains are flexibly attached and can be oriented differently according to the polymerase state (Reich et al., 2014)(Cusack et al., unpublished data).

Other good examples of multidomain polymerases are those of the Vesicular stomatis virsus (VSV) or Ebola virus (EBOV) from the order Mononegavirales, or the positive RNA strand flaviviruses like Dengue virus (DENV) and Hepatitis C virus (HCV). Unlike segmented NSV these viruses are able to perform all the post-transcriptional RNA 5'-end modifications necessary for

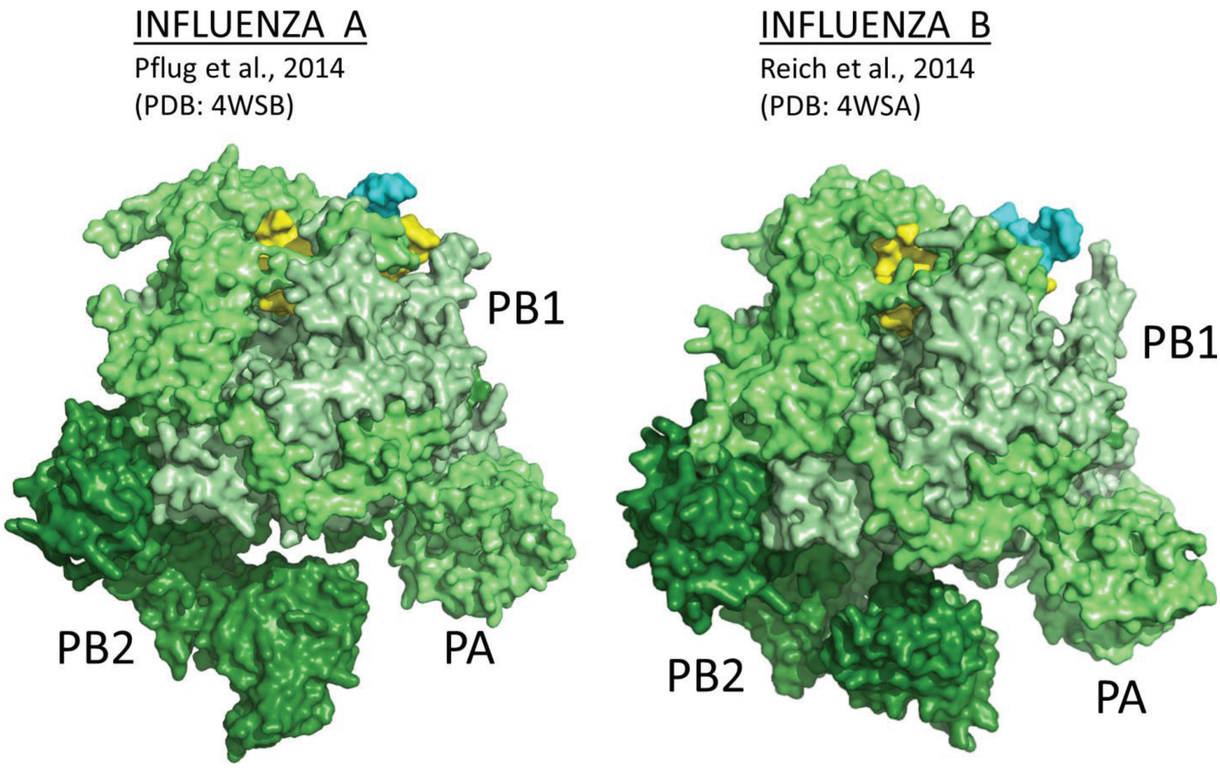



**Figure 1.9** Comparison of the heterotrimeric polymerases from Influenza A and Influenza B virus (Pflug et al., 2014; Reich et al., 2014). PA, PB1, and PB2 subunits are coloured with different shades of green. Genomic 3' and 5' RNA is coloured in cyan and yellow respectively.

the cap formation. Their polymerases contain a 2'-O-methyltransferase (MTase), a guanine-N7-MTase, and a polyribonucleotidyltransferase (PRNTase) or a guanylyltransferase (GTase) in case of Mononegavirales (Li et al., 2005; Ogino et al., 2010) or flaviviruses (Egloff et al., 2002; Issur et al., 2009) respectively.

Another example of complex polymerase is the HIV-1 reverse transcriptase, an RNA-dependent DNA polymerase (RdDp). In this case polymerase is a heterodimer composed of viral p66 subunit, containing RNase H and RdDp domains, and its processed version p51 bearing only the RdDp domain (Hostomsky et al., 1992; Wang et al., 1994). In consequence the HIV-1 RT complex contains two RdDp domains of which only the one belonging to the p66 is active. Moreover the heterodimer can bind randomly the viral RNA/DNA duplex and reorient itself based on what enzymatic activity is needed (Abbondanzieri et al., 2008; Liu et al., 2008).

The HIV-1 RT with its dual RdDp complex is not a unique case among polymerases of RNA viruses. The oligomerisation of polymerases has been observed as well for RdRps of Poliovirus (PV), FMDV, HCV, Norovirus (NV), and IBDV. The role of such assembly is not clear but some evidences suggest that it might facilitate template binding, increase polymerase activity, or basically stabiles the RdRp (Lyle et al., 2002; Spagnolo et al., 2010). Studies on Influenza virus revealed a possibility of dimer formation or oligomerization of the RdRps (Chang et al., 2015; Jorba et al., 2008). It was proposed that binding of an inactive secondary RdRp to the cRNP can stimulate vRNA synthesis – which suggests a cooperative functioning of both polymerases during replication (York et al., 2013). No evidence for such cooperation between two polymerases was found in case of Influenza virus transcription (Jorba et al., 2009).

Another strategy to form a functional polymerase – instead of using multisubunit or multidomain enzyme of viral origin – is to hijack host factors. The best studied example of such strategy adopted by an RNA virus is the replicase of bacteriophage Qβ (Kidmose et al., 2010; Takeshita et al., 2014). The replicase complex is composed of the viral β-subunit being the catalytic RdRp, two bacterial translation elongation factors – EF-Tu and EF-Ts, and the third bacterial protein – ribosomal protein S1. Translation factors contribute to the general architecture of the replicase. On one hand they stabilize the β-subunit, playing the role of chaperons, on the other they contribute in defining RNA tunnels leading in and out from the catalytic site. EF-Tu is a ubiquitous GTPase whose role is to bind aa-tRNA and deliver it to the translating ribosome. By chance the highly structured genome of bacteriophage Qβ possesses a tRNA-like structure at its 3' end which is recognized by the EF-Tu, and after de novo initiation exits through the EF-Tu-formed template exit tunnel. Other, yet not structurally proven examples, exist for viral RdRps binding host translation elongation factors. Studies revealed that an eukaryotic counterpart of the EF-Tu, eEF1A, participates in formation of the RdRp complexes belonging to bovine viral diarrhea virus (Johnson et al., 2001), polio virus (Harris et al., 1994), and tobacco mosaic virus (Yamaji et al., 2006)

## 1.3 Aim of the thesis

The PhD thesis project aimed to study the structure and function of a bunyavirus L protein alone and in complex with vRNA in order to better understand mechanisms that drive sNSV polymerases and to provide knowledge that could be extrapolated to other polymerases of this group of viruses, comprising *Arenaviridae*, *Bunyaviridae*, and *Orthomyxoviridae*. As a model in our studies we used the L protein from La Crosse virus belonging to the *Orthobunyavirus* genus. The project was designed based on the initial assumption that *Bunyaviridae* LACV and *Orthomyxoviridae* Influenza virus polymerases share similar domain organization. This hypothesis was strengthen by: (1) the similar size of the monomeric LACV L protein and the Influenza polymerase heterotrimer, (2) the linearity of the PA, PB1, and PB2 Influenza polymerase subunits, mediated by the head-to-tail interactions, (3) the same position of the central, catalytic RdRp motifs and the N-terminal endonuclease domains of both LACV L protein and Influenza PA subunit, whose structures were previously obtained in the Cusack group (Dias et al., 2009; Reguera et al., 2010). Based on this hypothesis, the identification and structural studies of the Influenza PB2 subunit cap-binding domain performed in the Cusack group (Guilligay et al., 2008) indicated that a homologous domain could be located within the C terminus of the LACV L protein.

The goal of the PhD thesis project was to:

1. Identify and characterize both structurally and functionally a putative cap-binding domain of the LACV L protein.
2. Study the nature of the vRNA binding by the full length LACV L protein or its truncated forms.
3. Establish the *in vitro* assays for studying replication and transcription activity of the LACV L protein.
4. Characterize structurally the LACV L protein or its truncated forms either as an apo form or in complex with the vRNA.

# 2. METHODS

## 2.1 Protein expression and purification

### 2.1.1 Constructs for the bacteria expression

The La Crosse virus strain LACV/mosquito/1978 L protein sequence was used in all construct preparations (GenBank code: EF485038.1, UniProt code: A5HC98). The synthetic, codon-optimized constructs (DNA 2.0) bearing N-terminal His-tag followed by TEV protease cleavage site (MGHHHHHH$_{\text{6xHis-tag}}$DYDIPTTENLYFQ$_{\text{TEV}}$G) were used. The initial constructs already available in the lab at the beginning of the PhD project were designed based on the ESPRIT approach (Yumerefendi et al., 2010) performed on the last 800 residues of the L protein, followed by limited proteolysis with papain. They were later re-designed according to secondary structure prediction softwares: Prof prediction, GOR4 (Sen et al., 2005), and Porter (Pollastri and McLysaght, 2005). DNA oligos (Invitrogen) used for cloning are listed in the Tabale 2.1 below.

All the L protein C-terminal constructs were cloned via NdeI and NotI into the pESPRIT, a pET-derived high copy number plasmids. Generation of new constructs with modified boundaries was achieved in a two-step PCR. During the first one the NotI restriction site was introduced by the overhang of the reversed primer. During the second step the nested forward primer allowed to reintroduce the NdeI restriction site followed by the His-tag and TEV cleavage site. The PCR reactions were carried in 50 µL volumes using either Pyrobest polymerase (TaKaRa) or PFU polymerase (produced in house). The reaction mixtures and temperature cycle are presented below:

| Reagent | Amount | | Temp. | Time | Cycles |
|---|---|---|---|---|---|
| Polymerase | 0.02 U / µL | | 95°C | 1 min | 1 |
| DNA template | 10 ng / µL | | 95°C | 30 sec | 35 |
| F/R primer | 2 µM | | 55°C | 30 sec | |
| dNTPs | 0.25-1 mM | | 72°C | 1 min | |
| Pol. buffer | 1X | | 72°C | 5 min | 1 |
| $H_2O$ | up to 50 µL | | 4°C | - | - |

PCR products were than purified using PCR Purification Kit (Qiagen), and cleaved with restriction enzymes (1 hour at 37°C). Next they were resolved on agarose gel, purified with Gel Extraction Kit (Qiagen), and ligated with the vector in 4:1 molar ration using 2X Ligation Kit (TaKaRa). Following 2-3 hours of incubation, ligation mixtures were directly used to transform *Escherichia coli* TOP10. Amplified DNA plasmids were extracted using Miniprep Kit (Qiagen). Those with bearing proper insertions were sent for sequencing (Eurofins MWG Operon).

| PCR PRIMERS | |
|---|---|
| NdeI-His-TEV-F (2nd PCR) | GCGGATCCATATGGGCCACCATCATCACCACCATGATTATGATAT TCCAACTACCGAGAATTTGTATTTTCAGGGG |
| | |
| L-LKVA--180-F | GAATTTGTATTTTCAGGGGCTGAAGGTCGCACATGG |
| L-MPNY--900-F | GAATTTGTATTTTCAGGGGATGCCGAATTACACCGATT |
| L-LPGQ-1310-F | GAATTTGTATTTTCAGGGGCTGCCGGGTCAGAGCA |
| L-ALEK-1536-F | GAATTTGTATTTTCAGGGGGCTCTGGAAAAGGAACCGG |
| L-LKEM-1685-F | GAATTTGTATTTTCAGGGGCTGAAGGAGATGACGCGC |
| L-VFIL-1708-F | GAATTTGTATTTTCAGGGGGTCTTCATCCTCCCCGCG |
| L-SGGH-1746-F | GAATTTGTATTTTCAGGGGTCCGGTGGCCATAAGGC |
| L-ATSE-1757-F | GAATTTGTATTTTCAGGGGGCGACCAGCGAGCAGAAC |
| L-SEQN-1759-F | GAATTTGTATTTTCAGGGGAGCGAGCAGAACATCGC |
| L-IAFE-1763-F | GAATTTGTATTTTCAGGGGATCGCTTTTGAGTGCTTC |
| L-RIYN-1960-F | GAATTTGTATTTTCAGGGGCGTATCTACAATGAGATTACG |
| L-YRNA-2110-F | GAATTTGTATTTTCAGGGGTACCGTAATGCGATTAAACTG |
| | |
| L-RNAM--900-R | AGCTCGCGGCCGCTCATTACATTGCGTTACGGAGCAT |
| L-YNML-1310-R | AGCTCGCGGCCGCTCATTACAGCATATTGTAAGTCAGG |
| L-GGHK-1750-R | AGCTCGCGGCCGCTCATTACTTATGGCCACCGGACA |
| L-TDFI-1819-R | AGCTCGCGGCCGCTCATTAGATGAAATCGGTACGGTTG |
| L-LMKS-2039-R | AGCTCGCGGCCGCTCATTAGCTCTTCATCAGCTCGG |
| L-DLLN-2044-R | AGCTCGCGGCCGCTCATTAATTCAACAGATCCTGGCTC |
| L-DNIR-2051-R | AGCTCGCGGCCGCTCATTAGCGAATGTTGTCGTAATTC |
| L-EFLS-2078-R | AGCTCGCGGCCGCTCATTAGCTCAGGAACTCCAGAC |
| L-DDPM-2082-R | AGCTCGCGGCCGCTCATTACATCGGATCGTCGCTCA |
| L-NFTE-2086-R | AGCTCGCGGCCGCTCATTACTCGGTGAAATTCATCGG |
| L-EAIH-2091-R | AGCTCGCGGCCGCTCATTAGTGGATTGCCTCGCCCTC |
| L-HSTP-2094-R | AGCTCGCGGCCGCTCATTACGGGGTGCTGTGGATTG |
| L-PIFN-2097-R | AGCTCGCGGCCGCTCATTAATTGAAAATCGGGGTGCTG |
| L-NIYY-2100-R | AGCTCGCGGCCGCTCATTAATAGTAAATATTGAAAATCGGG |
| L-IYYS-2101-R | AGCTCGCGGCCGCTCATTAAGAATAGTAAATATTGAAAATCG |
| L-LLIE-2119-R | AGCTCGCGGCCGCTCATTACTCGATCAGCAGTTTAATC |
| L-EEAF-2129-R | AGCTCGCGGCCGCTCATTAGAAGGCCTCCTCAAAGAT |
| L-FTFS-2132-R | AGCTCGCGGCCGCTCATTAGCTAAACGTGAAGGCCTC |

| | |
|---|---|
| L-IKLL-2155-R | AGCTCGCGGCCGCTCATTACAGCAGCTTGATGAGGC |
| L-STVI-2164-R | AGCTCGCGGCCGCTCATTAGATCACGGTAGACCACTC |
| L-VIDK-2166-R | AGCTCGCGGCCGCTCATTATTTGTCGATCACGGTAGAC |
| L-DHMY-2181-R | AGCTCGCGGCCGCTCATTAGTACATGTGATCCATTCCG |
| L-SFDV-2186-R | AGCTCGCGGCCGCTCATTACACGTCGAAGCTGTGGT |
| L-PKCF-2190-R | AGCTCGCGGCCGCTCATTAAAAACATTTCGGCACGTCG |
| L-IALI-2234-R | AGCTCGCGGCCGCTCATTAAATCAGAGCGATACATTTCTTC |
| L-EFTK-2249-R | AGCTCGCGGCCGCTCATTATTTCGTAAACTCAGAGAAGT |
| L-KLMK-2252-R | AGCTCGCGGCCGCTCATTATTTCATCAGTTTCGTAAACTC |

**Table 2.1** DNA oligos used for cloning of the L protein C-terminal constructs. NdeI and NotI restriction sites are highlighted in red in forward (F) and reversed (R) primers respectively. Sequences corresponding to the newly established constructs are highlighted in blue. The nested region for the second PCR is highlighted in grey.

### 2.1.2 Constructs for the insect cell expression

The same LACV L protein sequence and N-terminal tag as described above has been used for the L full length and $L_{1750}$ (ΔCterm construct covering residues 1-1750) constructs. $L_{1750}$ construct was designed based on limited proteolysis with trypsin performed on the L full length protein. Its C-terminal end corresponds to the boundaries of soluble domains derived from ESPRIT approach (Yumerefendi et al., 2010). Both L full length and $L_{1750}$ constructs were cloned via NdeI and NotI (New England Biolabs) into the pFastBac vector.

Based on in vivo minireplicon studies on LACV transcription and replication, performed by the Cusack group in collaboration with Dr. Friedmann Weber (Klemm et al., 2013) a single H34K mutation has been introduced to inactivate the endonuclease in L and $L_{1750}$ constructs. The mutated L constructs were supposed to remain replication but not transcription activity, and should not degrade RNA oligos used in various binding and activity *in vitro* assays.

Based on $L_{1750}$ structure in complex with 3' and 5' vRNA (Gerlach et al., 2015), full length L and $L_{1750}$ constructs with major deletions were designed. The 'clamp' deletion removes residues 355-400, introducing two glycines instead. The 'α-ribbon' deletion removes residues 848-905, introducing one glycine instead. The 'finger node' deletion removes residues 1105-1135, introducing one glycine instead. The 'priming loop' deletion removes residues 1408-1424, introducing two glycines instead. Detailed list of DNA oligos used for mutagenesis and deletions is given in the Table 2.2 below.

The 25 µL PCR reactions were performed with Phusion High-Fidelity polymerase (New England Biolabs). Primers introducing mutation and temperature cycles were designed based on the QuickChange site-directed mutagenesis protocol (Stratagene). The reaction mixtures and temperature cycle are presented below:

| Reagent | Amount |
|---|---|
| Polymerase | 0.04 U / µL |
| DNA template | 50-100 ng |
| F/R primer | 0.5 µM |
| dNTPs | 0.2 mM |
| Pol. buffer | 1X |
| DMSO | 3% |
| $H_2O$ | up to 25 µL |

| Temp. | Time | Cycles |
|---|---|---|
| 98°C | 30 sec | 1 |
| 98°C | 10 sec | |
| 45°C | 30 sec | 20 |
| 72°C | 10 min | |
| 72°C | 10 min | 1 |
| 4°C | - | - |
| | | |

PCR products were than treated with DpnI restriction enzyme (1 hour at 37°C) used directly to transform the *Escherichia coli* TOP10 cells.

| SITE DIRECTED MUTAGENESIS PCR PRIMERS | |
|---|---|
| L-H34K-F | CTGCTGATGGCCCGTAAGGACTATTTCGGACGC |
| L-H34K-R | GCGTCCGAAATAGTCCTTACGGGCCATCAGCAG |
| L-ΔClamp-F | GCAAAATGATGGATATTGGTGGTGGTATCAACAATGACCTGATCG |
| L-ΔClamp-R | CGATCAGGTCATTGTTGATACCACCACCAATATCCATCATTTTGC |
| L-ΔαRibbon-F | GCAAGTCTTGTCTGAAGATTGGTGGTTACATTTCCACCAAGGTC |
| L-ΔαRibbon-R | GACCTTGGTGGAAATGTAACCACCAATCTTCAGACAAGACTTGC |
| L-ΔFNode-F | CATGGACAAGGAGCTGATCGGTACCGTGCTGATTAAACGTAAC |
| L-ΔFNode-R | GTTACGTTTAATCAGCACGGTACCGATCAGCTCCTTGTCCATG |
| L-ΔPLoop-F | GCTACTTGGTTCTGGATGCGGGTGGTAGCATTCTGACGCCACG |
| L-ΔPLoop-R | CGTGGCGTCAGAATGCTACCACCCGCATCCAGAACCAAGTAGC |

**Table 2.2** DNA oligos used for site directed mutagenesis and deletions. Nucleotides introducing H34K mutation are highlighted in red. Nucleotides replacing deleted region with glycines are highlighted in blue.

### 2.1.3 Bacteria expression

L protein C-terminal constructs were expressed in *Escherichia coli* BL21 strain transformed with pESPRIT plasmid, a pET-derived high copy number plasmids, carrying appropriate inserts. Bacteria were grown in suspension in LB + kanamycin media at 37°C to reach the $OD_{600}$ = 0.6 – 0.8. Expression was induced with 1 mM IPTG, and carried over night at 20°C.

### 2.1.4 Insect cells expression

The Bac-to-Bac insect cell expression system (Invitrogen) was used for expression. Briefly, the pFastBac donor plasmid and *Escherichia coli* TOP10 were used to clone and amplify desired vectors. They were than transformed into DH10Bac competent cells containing the bacmid and the helper plasmid. Cells with successfully performed transposition were selected on LB agar plates containing kanamycin, gentamycin, tetracycline, Bluo-gal, and IPTG. Freshly isolated recombinant bacmids were transfected into $0.5x10^6$/mL of adherent Sf21 cells (*Spodoptera frugiperda*) in Sf-900 II SFM medium (Life Technologies), using the X-tremeGENE HP transfection agent (Roche), and 3 ml 6-well plates. $V_0$ baculoviruses were than amplified in bigger volumes of suspension Sf21 cells, collected, and stored at 4°C. Proteins were expressed High Five insect cells grown in Express Five SFM medium (Life Technologies). After reaching desirable volume at $0.5x10^6$/mL concentration, cells were infected with 0.5% (v/v) of the baculovirus and incubated for additional 3-4 days. Cells were harvested by centrifugation (10 min, 4°C, 5000 G) and were either used directly for purification, or stored at -80°C.

Selenomethionine derivative of $L_{1750}$ was expressed in High Five insect cells grown in ESF 921 medium (Expression Systems). 8 hours post infection medium was replaced by ESF 921 methionine-free medium (Expression Systems) and after additional 16 hours (24 hours post infection) selenomethionine (ACROS Organics) was added (100 mg/L).

### 2.1.5 L protein C-terminus constructs purification

Bacterial pellets were re-suspended in lysis buffer (20 mM Tris-HCl pH 7, 150 mM NaCl, 10 mM βMe, 20 mM imidazole) supplemented with protease inhibitors (ROCHE, complete mini, EDTA-free) and lysed by sonication. Cell debris was spun off (30 min, 4 °C, 35000 G) and clarified supernatant was loaded onto immobilized nickel ion affinity chromatography (IMAC) column and washed with the lysis buffer supplemented with 1 M NaCl. Proteins were eluted using the lysis buffer supplemented with 300 mM imidazole. Appropriate fractions were merged and dialyzed overnight against the lysis buffer but without the imidazole. During this step TEV protease (home source) was added to the protein solution to remove the N-terminal His-tag. Following dialysis sample was loaded on the 2$^{nd}$ Ni column to separate the cleaved protein from the the His-tagged TEV protease. Finally samples were subjected to size exclusion chromatography (GE, Superdex S75).

### 2.1.6 L and $L_{1750}$ proteins purification

Pellets of High Five cells were re-suspended in buffer A (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, pH 8) supplemented with protease inhibitors (ROCHE, complete mini, EDTA-free) and lysed by mild sonication (3 min on ice in 10 sec pulses). Cell debris was spun off (30 min, 4 °C, 35000 G) and clarified supernatant was filtered with 0.22 µm cut off in order to enhance the flow through chelating sepharose (GE Healthcare) in upcoming chromatography steps. After filtration the protein was salt out of solution by adding ammonium sulfate to the supernatant (0.5 g/ml) and incubating 1 hour at 4 °C. The precipitated protein was collected by centrifugation (30 min, 4 °C, 70000 G) and re-suspended in the same volume of buffer A. Following final centrifugation step (30 min, 4 °C, 70000 G) the protein solution was loaded onto immobilized nickel ion affinity chromatography (IMAC) column and washed with buffer A supplemented with 1 M NaCl. Polymerase was eluted using buffer A with 300 mM imidazole. Appropriate fractions were merged and dialyzed overnight against buffer B (50 mM Tris-HCl, 250 mM NaCl, 0.5 mM TCEP, pH 8). Next the polymerase solution was loaded on a heparin column (GE, HiTrap Heparin HP, 2 x 5ml) and eluted by a gradient of buffer B supplemented with 1 M NaCl. This step allowed further purification and concentration of the polymerase in the same time. At this stage RNA-free-sample was either supplemented with 10% glycerol, aliqouted, flash-frozen in liquid nitrogen and stored at -80 °C, or was directly subjected to size exclusion chromatography (GE, Superdex S200) in buffer C (20 mM Tris-HCl, 150 mM NaCl, 2 mM TCEP, pH 8).

In case of L full length protein purified for co-crystallization with vRNA, a modification of that procedure was used. Namely an excess of the appropriate RNA oligo (or pre-annealed oligos) was added to the polymerase sample eluted from the heparin column. Mild, stepwise dialysis against buffer B with decreasing concentration of NaCl (from 600 mM to 150 mM) allowed polymerase-vRNA complex formation, preventing protein precipitation from low salt solution.

## 2.2 RNA oligos

### 2.2.1 Synthetic RNA oligos

All synthetic RNA used for co-crystallization experiments, activity assays and radiolabelling were initially acquired from Sigma-Aldrich and later on from GE Healthcare Dharmacon. Fluorescently labelled RNA oligos were initially acquired from Sigma-Aldrich and later on from IBA. RNA oligos with 5'ppp modification used for capping with VCE were acquired from ChemGenes. Table below lists all the RNA oligos used:

| | |
|---|---|
| 3'1-56 WT | GGUUUGAUUCAAAACUUUAAUAUUCAAACGUUAUCUAUACUUGGUAGUACACUACU |
| 3'1-25 WT | UAUCUAUACUUGGUAGUACACUACU |
| 3'1-22 WT | CUAUACUUGGUAGUACACUACU |
| 3'1-20 WT | AUACUUGGUAGUACACUACU |
| 3'1-18 WT | ACUUGGUAGUACACUACU |
| 3'1-17 WT | CUUGGUAGUACACUACU |
| 3'1-16 WT | UUGGUAGUACACUACU |
| 3'1-16 cRNA | UUGGUAGCACACUACU |
| 3'1-15 WT | UGGUAGUACACUACU |

| | |
|---|---|
| 3'1-14 WT | GGUAGUACACUACU |
| 3'1-12 WT | UAGUACACUACU |
| 3'1-10 WT | GUACACUACU |
| 3'1-8 WT | ACACUACU |
| 3'9-17 WT | CUUGGUAGU |
| 3'9-16 WT | UUGGUAGU |
| 3'9-15 WT | UGGUAGU |
| | |
| 3'1-11WT-12-20mut | UAUGAACCAAGUACACUACU |
| 3'1-17 mut | CUUGGUAAAGUUCUACU |
| 3'1-16 mut | UUGGUAAAGUUCUACU |
| 3'1-15 mut | UGGUAAAGUUCUACU |
| 3'1-17 FLUmut | CUUGGUGCUACAGCUAC |
| 3'1-16 FLUmut | UUGGUGCUACAGCUAC |
| 3'1-15 FLUmut | UGGUGCUACAGCUAC |
| | |
| 3'5A(6-14WT) | GGUAGUACAAAAAA |
| 3'(1-5WT)4G(10-14WT) | GGUAGGGGGCUACU |
| 3'(1-9WT)5A | AAAAAUACACUACU |
| 3'3A(1-14WT)3A | AAAGGUAGUACACUACUAAA |
| 3'5G(6-9WT)5C | CCCCCUACAGGGGG |
| | |
| 5'p1-59 WT | pAGUAGUGUGCUACCAAGUAUAAAAUAAUGUUUGCAAAAUAAAUUUACGCUGUUGUUAGU |
| 5'1-25 WT | AGUAGUGUGCUACCAAGUAUAAAAU |
| 5'ppp1-25 WT | pppAGUAGUGUGCUACCAAGUAUAAAAU |
| 5'p1-17 WT | pAGUAGUGUGCUACCAAG |
| 5'p1-16 WT | pAGUAGUGUGCUACCAA |
| 5'p1-15 WT | pAGUAGUGUGCUACCA |
| 5'p1-14 WT | pAGUAGUGUGCUACC |
| 5'p1-13 WT | pAGUAGUGUGCUAC |
| 5'p1-12 WT | pAGUAGUGUGCUA |
| 5'p1-11 WT | pAGUAGUGUGCU |
| 5'p1-11 cRNA | pAGUAGUGUACU |
| 5'p1-10 WT | pAGUAGUGUGC |
| 5'p1-10 cRNA | pAGUAGUGUAC |
| 5'p1-8 WT | pAGUAGUGU |
| 5'9-25 WT | GCUACCAAGUAUAAAAU |
| 5'9-22 WT | GCUACCAAGUAUAA |
| 5'9-20 WT | GCUACCAAGUAU |
| 5'9-18 WT | GCUACCAAGU |
| 5'9-17 WT | GCUACCAAG |
| 5'9-16 WT | GCUACCAA |
| 5'9-16 cRNA | ACUACCAA |
| 5'9-15 WT | GCUACCA |
| 5'9-14 WT | GCUACC |
| 5'8-16 WT | UGCUACCAA |
| 5'7-16 WT | GUGCUACCAA |
| 5'6-16 WT | UGUGCUACCAA |
| 5'12-16 WT | ACCAA |
| 5'11-16 WT | UACCAA |
| 5'10-16 WT | CUACCAA |
| | |
| 5'p1-16 mut | pGACCGUGGGUCACCAA |

| | |
|---|---|
| 5'p1-11 mut | pGACAGUGUGUC |
| 5'p1-10 mut | pGACAGUGUGU |
| 5'p1-17 BPmut | pACGAGUGUCGUACCAAG |
| 5'p1-16 BPmut | pACGAGUGUCGUACCAA |
| 5'p1-15 BPmut | pACGAGUGUCGUACCA |
| | |
| 3'1-16 WT[Br] | U(U5Br)GG(U5Br)AG(U5Br)ACAC(U5Br)ACU |
| 5'9-16 WT[Br] | GC(U5Br)ACCAA |
| | |
| 3'1-25 WT[Cy5] | [Cy5]-UAUCUAUACUUGGUAGUACACUACU |
| 3'1-25 WT[DY647] | [DY647]-UAUCUAUACUUGGUAG(U4S)ACAC(U4S)ACU |
| 3'1-25 WT[FAM] | [FAM-EX-5]-UAUCUAUACUUGGUAGUACACUACU |
| 3'1-25 A6C[FAM] | [FAM-EX-5]-UAUCUAUACUUGGUAGUACCCUACU |
| 3'1-25 A8C[FAM] | [FAM-EX-5]-UAUCUAUACUUGGUAGUCCACUACU |
| | |
| 5'ppp1-25 WT[Cy5] | pppAG(U4S)AGUGUGCUACCAAGUAUAAAAU-[Cy5] |
| 5'p1-25 WT[FAM] | pAGUAGUGUGCUACCAAGUAUAAAAU-[FAM-EX-5] |
| 5'p1-25 cRNA[FAM] | pAGUAGUGUACUACCAAGUAUAAAAU-[FAM-EX-5] |
| 5'9-25 WT[FAM] | GCUACCAAGUAUAAAAU-[FAM-EX-5] |
| 5'8-25 WT[FAM] | UGCUACCAAGUAUAAAAU-[FAM-EX-5] |
| 5'7-25 WT[FAM] | GUGCUACCAAGUAUAAAAU-[FAM-EX-5] |
| 5'6-25 WT[FAM] | UGUGCUACCAAGUAUAAAAU-[FAM-EX-5] |
| | |
| α-globin ppp mRNA | ppp(Am)CACUUCUGGUCCAGUCCGA |
| α-globin mRNA | ACACUUCUGGUCCAGUCCGA |

### 2.2.2 T7 transcription of RNA oligos

T7 in vitro transcription was used to produce radioactively labelled RNA oligos for electrophoretic mobility shift assays (EMSA). Non labelled long RNA oligos for RNP assembly were also T7-transcribed. 2 µM DNA oligos were preheated and annealed in order to form double stranded T7 promoter with long overhang being the reaction template, depicted below.

MECHANISM:
5' **TAATACGACTCACTATAGGG**XXXXXXXXXXXXXXXXXXXXXX 3'
3' **ATTATGCTGAGTGATATCCC**YYYYYYYYYYYYYYYYYYYYYY 5'

OLIGOS TO ORDER:
5' **TAATACGACTCACTATAGGG** 3'
5' YYYYYYYYYYYYYYYYYYYYYY**CCCTATAGTGAGTCGTATTA** 3'

RNA PRODUCT:
5' **GGG**XXXXXXXXXXXXXXXXXXXXXX 3'

Reactions were carried in 10 µl or 10 ml volume for radioactively labelled or non-labelled RNA respectively. DNA templates were mixed with 4 mM ATP/UTP/CTP/GTP (Jena Bioscience) and, if needed, [α32P]-GTP (PerkinElmer) in the TDMST buffer containing 40 mM Tris pH 8.0, 5 mM DTT, 1 mM spermidine, 0.01% Triton X-100, 30 mM $MgCl_2$, and 1% PEG8000. To start the transcription 50 µg/ml of the home source T7 polymerase was added. Reactions were incubated in 37°C for 3-5 hours. RNA products were than resolved on big urea TBE 15-20% polyacrylamide gels. Silica gels (Merck) were used to visualize RNA bands by 254 nm UV light shadowing. Bands were then cut from the gel and incubated overnight in the RNA elution buffer containing 0.5 M ammonium acetate pH 5.0, 1 mM EDTA pH 8.0, and 0.1% SDS.

DNA oligos comprising various TATA promoters and templates used in T7 transcription reactions are listed in the table below:

| | |
|---|---|
| T7prom+AGT | TAATACGACTCACTATAAGT |
| T7prom+GTG | TAATACGACTCACTATAGTG |
| T7prom+GCT | TAATACGACTCACTATAGCT |
| T7prom+GCC | TAATACGACTCACTATAGCC |
| T7prom+GAG | TAATACGACTCACTATAGAG |
| T7prom+GGA | TAATACGACTCACTATAGGA |
| T7prom+GGG | TAATACGACTCACTATAGGG |
| T7prom+GGT | TAATACGACTCACTATAGGT |
| T7prom+CGT | TAATACGACTCACTATACGT |
| T7prom+TGT | TAATACGACTCACTATATGT |
| T7prom+AAT | TAATACGACTCACTATAAAT |
| T7prom+ACT | TAATACGACTCACTATAACT |
| T7prom+ATT | TAATACGACTCACTATAATT |
| T7prom+AGA | TAATACGACTCACTATAAGA |
| T7prom+AGG | TAATACGACTCACTATAAGG |
| T7prom+AGC | TAATACGACTCACTATAAGC |
| | |
| mRNA50/T1 | CATATCGTGTCATATCGTGTCATATCGTGTCTTAACGTGTCAATTCATCCTATAGTGAGTCGTATTA |
| mRNA40/T1 | CATATCGTGTCATATCGTGTCTTAACGTGTCAATTCATCCTATAGTGAGTCGTATTA |
| mRNA30/T1 | CATATCGTGTCTTAACGTGTCAATTCATCCTATAGTGAGTCGTATTA |
| mRNA20/T1 | CTTAACGTGTCAATTCATCCTATAGTGAGTCGTATTA |
| mRNA10/T1 | CAATTCATCCTATAGTGAGTCGTATTA |
| | |
| 3'1-30(GG) | AGTAGTGTACTACCAAGTATAGATAACGTTCCTATAGTGAGTCGTATTA |
| 3'1-25(GG) | AGTAGTGTACTACCAAGTATAGATACCTATAGTGAGTCGTATTA |
| 3'20wt(GG) | AGTAGTGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'15wt5mut(GG) | AGTAGTGTACTACCATTCTCCCTATAGTGAGTCGTATTA |
| 3'10wt10mut(GG) | AGTAGTGTACCCTTCTTCTCCCTATAGTGAGTCGTATTA |
| 3'5wt15mut(GG) | AGTAGCTCCTCCTTCTTCTCCCTATAGTGAGTCGTATTA |
| | |
| 5'1-30 | ACATTATTTTATACTTGGTAGCACACTACTTATAGTGAGTCGTATTA |

| | |
|---|---|
| 5'1-25 | ATTTTATACTTGGTAGCACACTACTTATAGTGAGTCGTATTA |
| 5'20wt | ATACTTGGTAGCACACTACTTATAGTGAGTCGTATTA |
| 5'15wt5mut(GA) | CCTTCTGGTAGCACACTACTTATAGTGAGTCGTATTA |
| 5'10wt10mut(GA) | CCTTCCTTCCGCACACTACTTATAGTGAGTCGTATTA |
| 5'5wt15mut(GA) | CCTTCCTTCCTTCTCCTACTTATAGTGAGTCGTATTA |
| | |
| PH:3'20(7G)5'1-20 | AGTAGTGTACTACCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGT GAGTCGTATTA |
| PH:3'20(7G)5'5-20 | AGTAGTGTACTACCAAGTATCCCCCCCATACTTGGTAGCACACTATAGTGAGT CGTATTA |
| PH:3'20(7G)5'7-20 | AGTAGTGTACTACCAAGTATCCCCCCCATACTTGGTAGCACTATAGTGAGTCG TATTA |
| PH:3'20(7G)5'9-20 | AGTAGTGTACTACCAAGTATCCCCCCCATACTTGGTAGCTATAGTGAGTCGTA TTA |
| PH:3'20(7G)5'12-20 | AGTAGTGTACTACCAAGTATCCCCCCCATACTTGGCTATAGTGAGTCGTATTA |
| PH:3'20(7G)5'15-20 | AGTAGTGTACTACCAAGTATCCCCCCCATACTCTATAGTGAGTCGTATTA |
| PH:3'1-20(7G)5'20 | AGTAGTGTACTACCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGT GAGTCGTATTA |
| PH:3'3-20(7G)5'20 | TAGTGTACTACCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGTGA GTCGTATTA |
| PH:3'5-20(7G)5'20 | GTGTACTACCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGTGAGT CGTATTA |
| PH:3'7-20(7G)5'20 | GTACTACCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGTGAGTCG TATTA |
| PH:3'9-20(7G)5'20 | ACTACCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGTGAGTCGTA TTA |
| PH:3'11-20(7G)5'20 | TACCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGTGAGTCGTATT A |
| PH:3'13-20(7G)5'20 | CCAAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGTGAGTCGTATTA |
| PH:3'15-20(7G)5'20 | AAGTATCCCCCCCATACTTGGTAGCACACTACTTATAGTGAGTCGTATTA |
| PH:3'20(7G)5'9 | AGTAGTGTACTACCAAGTATCCCCCCCATACTTGGCTATAGTGAGTCGTATTA |
| PH:3'20(7G)5'6 | AGTAGTGTACTACCAAGTATCCCCCCCATACTCTATAGTGAGTCGTATTA |
| | |
| 3'20:U4A | AGTTGTGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:U4G | AGTCGTGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:U4C | AGTGGTGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:C5A | AGTATTGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:C5G | AGTACTGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:C5U | AGTAATGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:A6G | AGTAGCGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:A6C | AGTAGGGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:A6U | AGTAGAGTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:C7A | AGTAGTTTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:C7G | AGTAGTCTACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:C7U | AGTAGTATACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:A8G | AGTAGTGCACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:A8C | AGTAGTGGACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:A8U | AGTAGTGAACTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:U9A | AGTAGTGTTCTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:U9G | AGTAGTGTCCTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:U9C | AGTAGTGTGCTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:G10A | AGTAGTGTATTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:G10C | AGTAGTGTAGTACCAAGTATCCTATAGTGAGTCGTATTA |
| 3'20:G10U | AGTAGTGTAATACCAAGTATCCTATAGTGAGTCGTATTA |

| | |
|---|---|
| 5'20:A1G | ATACTTGGTAGCACACTACCTATAGTGAGTCGTATTA |
| 5'20:A1C | ATACTTGGTAGCACACTACGTATAGTGAGTCGTATTA |
| 5'20:A1U | ATACTTGGTAGCACACTACATATAGTGAGTCGTATTA |
| 5'20:G2A | ATACTTGGTAGCACACTATTTATAGTGAGTCGTATTA |
| 5'20:G2C | ATACTTGGTAGCACACTAGTTATAGTGAGTCGTATTA |
| 5'20:G2U | ATACTTGGTAGCACACTAATTATAGTGAGTCGTATTA |
| 5'20:U3A | ATACTTGGTAGCACACTTCTTATAGTGAGTCGTATTA |
| 5'20:U3G | ATACTTGGTAGCACACTCCTTATAGTGAGTCGTATTA |
| 5'20:U3C | ATACTTGGTAGCACACTGCTTATAGTGAGTCGTATTA |
| 5'20:A4G | ATACTTGGTAGCACACCACTTATAGTGAGTCGTATTA |
| 5'20:A4C | ATACTTGGTAGCACACGACTTATAGTGAGTCGTATTA |
| 5'20:A4U | ATACTTGGTAGCACACAACTTATAGTGAGTCGTATTA |
| 5'20:G5A | ATACTTGGTAGCACATTACTTATAGTGAGTCGTATTA |
| 5'20:G5C | ATACTTGGTAGCACAGTACTTATAGTGAGTCGTATTA |
| 5'20:G5U | ATACTTGGTAGCACAATACTTATAGTGAGTCGTATTA |
| 5'20:U6A | ATACTTGGTAGCACTCTACTTATAGTGAGTCGTATTA |
| 5'20:U6G | ATACTTGGTAGCACCCTACTTATAGTGAGTCGTATTA |
| 5'20:U6C | ATACTTGGTAGCACGCTACTTATAGTGAGTCGTATTA |
| 5'20:G7A | ATACTTGGTAGCATACTACTTATAGTGAGTCGTATTA |
| 5'20:G7C | ATACTTGGTAGCAGACTACTTATAGTGAGTCGTATTA |
| 5'20:G7U | ATACTTGGTAGCAAACTACTTATAGTGAGTCGTATTA |
| 5'20:U8A | ATACTTGGTAGCTCACTACTTATAGTGAGTCGTATTA |
| 5'20:U8G | ATACTTGGTAGCCCACTACTTATAGTGAGTCGTATTA |
| 5'20:U8C | ATACTTGGTAGCGCACTACTTATAGTGAGTCGTATTA |
| 5'20:G9A | ATACTTGGTAGTACACTACTTATAGTGAGTCGTATTA |
| 5'20:G9C | ATACTTGGTAGGACACTACTTATAGTGAGTCGTATTA |
| 5'20:G9U | ATACTTGGTAGAACACTACTTATAGTGAGTCGTATTA |
| 5'20:C10A | ATACTTGGTATCACACTACTTATAGTGAGTCGTATTA |
| 5'20:C10G | ATACTTGGTACCACACTACTTATAGTGAGTCGTATTA |
| 5'20:C10U | ATACTTGGTAACACACTACTTATAGTGAGTCGTATTA |
| | |
| 3'52wt | AGTAGTGTACTACCAAGTATAGATAACGTTTGAATATTAAAGTTTTGAATCCT ATAGTGAGTCGTATTA |
| 3'53wt | AGTAGTGTACTACCAAGTATAGATAACGTTTGAATATTAAAGTTTTGAATCCC TATAGTGAGTCGTATTA |
| 3'54wt | AGTAGTGTACTACCAAGTATAGATAACGTTTGAATATTAAAGTTTTGAATCAC CTATAGTGAGTCGTATTA |
| 3'55wt | AGTAGTGTACTACCAAGTATAGATAACGTTTGAATATTAAAGTTTTGAATCAA CCTATAGTGAGTCGTATTA |
| 3'56wt | AGTAGTGTACTACCAAGTATAGATAACGTTTGAATATTAAAGTTTTGAATCAA ACCTATAGTGAGTCGTATTA |
| | |
| 5'55wt | ACAACAGCGTAAATTTATTTTGCAAACATTATTTTATACTTGGTAGCACACTA CTTATAGTGAGTCGTATTA |
| 5'56wt | AACAACAGCGTAAATTTATTTTGCAAACATTATTTTATACTTGGTAGCACACT ACTTATAGTGAGTCGTATTA |
| 5'57wt | TAACAACAGCGTAAATTTATTTTGCAAACATTATTTTATACTTGGTAGCACAC TACTTATAGTGAGTCGTATTA |
| 5'58wt | CTAACAACAGCGTAAATTTATTTTGCAAACATTATTTTATACTTGGTAGCACA CTACTTATAGTGAGTCGTATTA |
| 5'59wt | ACTAACAACAGCGTAAATTTATTTTGCAAACATTATTTTATACTTGGTAGCAC ACTACTTATAGTGAGTCGTATTA |

## 2.3 Functional characterization

### 2.3.1 Limited proteolysis experiments

Papain limited proteolysis of the L7 C-terminal construct was carried for 1-2 hours at room temperature with 1:500 (w/w) papain-L7 ratio and stopped by addition of the SDS-PAGE loading buffer and sample boiling, and visualized on SDS-PAGE.

Trypsin protection assays were performed on both $L_{1750}$ and full length L proteins using 25-nucleotide long 3'vRNA or 5'vRNA ends. 1 mg/ml of protein sample was incubated with the vRNA for 1-2 hours prior to trypsination. 1:1 protein-RNA ratio or a slight excess of RNA to ensure saturation were used. Apo protein was used as a reference. Trypsin limited proteolysis was carried for 1 hour at room temperature with 1:1000 (w/w) trypsin-L protein ratio and stopped by addition of the SDS-PAGE loading buffer and sample boiling. Products of $L_{1750}$ or L tripsination in presence or absence of the vRNAs were visualized on gradient SDS-PAGE (Bolt 4-12% Bis-Tris Plus Gel, Life Technologies). The cleavage products were identified and analysed by various techniques, including western-blot, ESI-TOF-MS, MALDI-TOF-MS, MALDI-TOF-MS with N-termini acetylation, and Edman degradation.

### 2.3.2 RNA-protein UV crosslinking

Prior to UV crosslinking 1µM of $L_{1750}$ or full length L protein was saturated with 2 µM of 3' or 5'vRNA bearing 2 or 1 photo-activable 4-thio-uracil (4SU) respectively (Kramer et al., 2011). In some cases excess of polyU RNA was used to minimize non-specific L protein – vRNA interactions. Both vRNAs were additionally labeled with a fluorophore on their non-interacting ends for tracking protein-RNA crosslinks. 5'vRNA was labelled with Cy5 and 3'vRNA with DY647 – an analog of Cy5. Following incubation vRNA – protein mixtures were loaded inside inverted caps of the Eppendorf tubes, placed on ice, and exposed to 365 nm UV light. Distance from the UV lamp was approximately 1 cm. Samples were crosslinked for 5-10 min (1min on / 1min off).

Following crosslinking samples were tripsinated as described above. Cleavage products were visualized on gradient SDS-PAGE (Bolt 4-12% Bis-Tris Plus Gel, Life Technologies). Prior to Coomasie staining gels were scanned for fluorescence signal using Typhoon. Images resulted from gel staining and fluorescence scanning were superposed in order to identify cleavage products crosslinked to vRNA. Selected bands were analysed by MALDI-TOF-MS, MALDI-TOF-MS with N-termini acetylation.

### 2.3.3 RNA-protein binding studies – radioactive EMSA assays

For each RNA-binding experiment 10 µM of $L_{1750}$ in 10 µl buffer (100 mM Tris, 100 mM NaCl, 0.5 mM DPBA, 5 mM βME, pH 8) was mixed with radioactive RNA and 1 µl of non-specific polyU RNA (Sigma). In order to maintain similar amount of radioactive RNA in each mixture the volumes added were scaled based on ImageQuant and Geiger counter measurements performed prior to the experiment. Mixtures were incubated at RT for several hours and resolved on native TG gels (top and bottom part of the gel – 10 and 20% polyacrylamide respectively). Radioactive signal from shifted bands was recorded with Typhoon and quantified with ImageQuant. In each case the amount of bound RNA was measured in reference to the control RNA of known strong affinity to the protein.

A modification of the EMSA method was used with the synthetic, fluorescently-labelled RNA oligos. RNA-protein binding analysis was analysed by gel EMSA as described above. Fluorescence signal was scanned with Typhoon.

### 2.3.4 RNP assembly

Trials were made in order to assemble in vitro the L or $L_{1750}$ protein together with single-stranded vRNA and nucleoproteins (NP). Buffer for the assembly contained 20 mM Tris ph 8.0, 150 mM NaCl, and 0.5 mM TCEP. The following vRNA, either T7 in vitro transcribed or synthetic, were used: 3’52nt, 3’53nt, 3’56nt, 5’55nt, 5’56nt, and 5’59nt. Nucleoprotein samples were provided by Dr. Juan Reguera and were purified using thiocyanate as described previously (Reguera et al., 2013). Prior to the assembly N proteins were subjected to the size exclusion chromatography. In case of full length L protein the vRNA was added directly to the high salt heparin elution protein samples (see L purification procedure) and mixtures were dialysed against the assembly buffer. In both cases – L and $L_{1750}$ – either the polymerase was pre-incubated with vRNA before N protein addition, or all three components were mixed together from the start. In this later scenario the assembly buffer was implemented with 1M NaCl and the cocktail was gradually dialysed down to 150 mM NaCl. This approach was similar to the L full length purification procedure, and was performed to force proper order of assembly, with vRNA binding first to the polymerase due to the affinity higher than vRNA-N protein affinity. Components for assembly assay were mixed at 1:1:4 ratio of polymerase, vRNA, and NP, and at low µM concentrations. Knowing the length of vRNA 3' and 5' ends specifically bound by the polymerase to be ~8 and ~11 nucleotides respectively (Gerlach et al., 2015), as well as the length of nucleotides accommodated by single NP (Reguera et al., 2013), the properly formed RNPs should contain four NP per a single polymerase-vRNA complex. Assemblies were carried at 4°C with long incubation times up to 48 hours. To verify whether complexes were formed samples were subjected to the size exclusion chromatography and collected fractions resolved on gradient SDS-PAGE (Bolt 4-12% Bis-Tris Plus Gel, Life Technologies)

### 2.3.5 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a quantitative technique based on heat consumption or heat release generated upon titration of a ligand into sample solution. Each time the change in heat is encountered the system must apply a certain amount of µcal/sec in order to restore the temperature. ITC allows accurate determination of the binding affinities ($K_a$) between analysed molecules in solution. It also provides information concerning enthalpy changes ($\Delta H$) and binding stoichiometry (n). Set of these experimentally measured parameters allows calculation of changes in Gibbs energy ($\Delta G$) and entropy ($\Delta S$) using the following formula:

$$\Delta G = -RT\ln Ka = \Delta H - T\Delta S$$

All measurements were performed at 25°C, using the ITC200 Microcalorimeter (MicroCal Inc.). $L_{1750}$ protein sample at 0.015 – 0.03 mM was titrated with 3′-OH or 5′-ppp 25 nucleotide long vRNA at approximately 10 x higher concentration (RNA oligos with triphosphate modification were obtained from Chemgenes). Prior to a single titration experiment 200 – 250 µl of protein sample in buffer containing 20 mM Tris ph 8.0, 150 mM NaCl, and 5 mM βME was loaded inside the sample cell. In order to overcome protein precipitation during the experiment, buffer containing 300 mM NaCl was also used. Syringe was filled with 40 – 50 µl of titrant vRNA. In order to minimize the buffer-related effects both sample and ligand solutions were dialysed against the same buffer. Moreover a blank measurement was performed in which vRNA was titrated into the sample buffer alone. Results of such control titration were later on subtracted form experimental data. During single experiment a total of 26 injections, 1.5 µl each, were made. Injections lasted 3 sec each, and were separated by 180 sec spacing. Reference power was set up to 5 µcal/sec. Obtained binding isotherms were analysed by Levenberg–Marquardt nonlinear regression using the ITC-adapted Origin Software (MicroCal Inc.). In each case a simple one binding-site fitting strategy was adopted.

### 2.3.6 Microscale thermophoresis

Microscale thermophoresis (MST) is a technique allowing to measure binding affinities (Kd) between analysed molecules in solution. As a principle it uses microscopic temperature gradients which trigger movements of molecules. Experiments are easy to perform and they consume very little material. For a single run 16 two-fold dilutions of the ligand (small molecule, RNA) are mixed with fluorescently-labelled molecule (protein) usually at nM concentration, in the final volume of 15-20 µl. Following incubation the mixtures are loaded into separate glass capillaries within which the measurement is performed. Infra-red (IR) laser heats sequentially the capillaries, thus increasing locally sample temperature by few Celsius degrees.

MST experiments were performed using the Monolith NT.115 Green/Red device (NanoTemper Technologies GmbH). $L_{1750}$ protein was fluorescently labelled using amine reactive MO-L002 Monolith™ Protein Labeling Kit GREEN-NHS. 50 nM of labelled $L_{1750}$ protein was than incubated with 25 nucleotide long 5′ vRNA end bearing either mono- or triphosphate modification at the very 5′ end. 16 dilutions of vRNA ranging from 40 µM to 1 nM were used. IR irradiation lasted 50 sec per one capillary, and the 15% of IR laser power was used. Experimental points were analysed and the binding curve fitted using NanoTemper software.

## 2.4 Structural characterization

### 2.4.1 SAXS – data collection and analysis

Small angle X-ray scattering (SAXS) is a technique that allows low resolution structural analysis of a macromolecular sample in solution. The intensity of the X-ray scattered during an experiment depends on the scattering angle. The signal generated by differently oriented molecules is averaged and becomes proportional to the scattering from a single molecule. Since different shapes of molecules generate characteristic scattering curves, SAXS can usually provide helpful information concerning the overall shape and dimensions of the analyzed object (Guinier plot, Rg – radius of gyration). SAXS 3D models can be used to fit in more accurate atomic structures.

Both $L_{1750}$ and L were subjected to the SAXS analysis, either as apo or vRNA bound proteins. Samples were prepared in buffer containing 20 mM Tris pH 8.0, 150 mM NaCl, 5 mM βME, and 2.5% glycerol. In one set of apo protein samples a buffer with stabilising additives was tested – 500 mM NaCl and 200 mM arginine. For each sample a series of four two fold dilutions were tested starting from the following concetrations: L-apo 1.83 mg/ml, L-3′ 1.3 mg/ml, L-5′ 1.7 mg/ml, L-additives 0.5 mg/ml, $L_{1750}$-apo 7.2 mg/ml, $L_{1750}$-3′ 25.0 mg/ml, $L_{1750}$-5′ 16.4 mg/ml, $L_{1750}$-additives 3.5 mg/ml. Results obtained for different concentrations were used to correct for oligomerization or other inter particle effects.

Data were collected at the ESRF BioSAXS beamline BM29 having wide energy range (7.0 – 15.0 keV) and equipped with a PILATUS 1M detector (Pernot et al., 2013). 35-40 µl samples were loaded into 2 mm quartz capillaries, and ISPyB BioSAXS interface (De Maria Antolinos et al., 2015) and BsxCuBE software were used to conduct the experiment and automatically process the images. Background buffer scattering was subtracted from each sample measurement using EMBL-developed PRIMUS program (Konarev et al., 2003). Ab-inition models were generated using programs DAMMIF (Franke and Svergun, 2009).

### 2.4.2 X-ray crystallography – protein crystallization

$L_{1750}$ protein in buffer C was adjusted to a concentration of 10 mg/ml and mixed in 3:4 ratio with vRNA. The vRNA used was an equimolar mixture of nucleotides 1-16 from the M-segment vRNA 3′ end (3′-UCAUCACAUGAUGGUU-5′) and nucleotides 9-16 from M-segment vRNA 5′ end (5′-GCUACCAA-3′)(GE Dharmacon), preheated to 90°C and annealed by slow cooling down. Sitting drops were prepared by mixing equal volumes (100 nl each) of the sample and reservoir solution using a Cartesian PixSys 4200 crystallization robot (Genomic Solutions). Crystallization trials were performed by vapor diffusion at 20°C. The brush-like needle clusters emerged after 2-3 days from several conditions containing 15% (w/v) of PEG 4000 or 20% (w/v) of PEG 3350 as precipitant and 0.2 M of various sulfate or citrate salts (Crystal Screen Lite and PEG/Ion from Hampton Research). After an additive screen and several rounds of pH and precipitant/salt concentration optimization the most convenient conditions were narrowed down to: 0.1 M Hepes, 0.2 M sodium citrate tribasic dihydrate, 0.3 M ammonium sulfate, 15% PEG 3350, pH 7. Despite decreasing crystallization temperature to 4 °C to reduce nucleation and growth speed crystals maintained tendency to form multicrystal clusters. Finally the microseeding procedure allowed to obtain the rod-shaped single crystals. In brief the best single-like looking clusters were transferred to the fresh drop of mother liquor, crushed with microseeding tools, and used to prepare a series of seeds dilutions, which were then added directly to the new crystallization drops (1 µl protein-RNA sample + 1 µl of mother liquor + 0.25 µl of seeds dilution).

Selenomethionine derivative crystals were grown in the same conditions as native ones, using initially native seeds and then switching to selenomethionine seeds if necessary. Heavy metal derivatives were prepared by soaking native crystals in mother liquor supplemented with either 2.5-5 mM $K_2PtCl_4$ (1-2 hours soaking) or with 5-10 mM $Ta_6Br_{12}$ tantalum cluster (Jena Bioscience)(1-2 days soaking). The tantalum cluster soaked crystals became visibly green. Prior to cryo-cooling the crystals were soaked in the mother liquor supplemented with 20-30% glycerol but without the heavy metal. This back-soaking step was performed in order to decrease the amount of heavy metal ions which did not bound the crystallized protein, but which would increase the X-ray absorption and radiation damage.

In order to obtain crystals containing missing 5′ end, the initial $L_{1750}$-vRNA crystals were soaked with nucleotides 1-8, 1-10, 1-11, and 1-12 from M-segment vRNA 5′ end (5′-pAGUAGUGUGCUA-3′ and shorter)(GE Dharmacon).

$L_{1750}$-vRNA crystals (both native and selenomethionine derivative) started to grow overnight at 4 °C, but they required at least 3 days to reach their maximum size. Prior to the X-ray diffraction experiment they were cryo-cooled to minimize the radiation damage caused by the free radicals and positive ionization. All crystals were cryo-protected in mother liquor supplemented with a cryo-protectant. While 20-30% glycerol proved to be the most efficient, in some cases ethylene glycol or PEG 400 were also tested. After few seconds of soaking in the cryo-protectant solution, crystals were quickly plunged and flash-frozen in liquid nitrogen at approximately -196 °C.

In order to obtain L-vRNA crystals the RNA oligos used for co-crystallization were added during the L full length purification procedure (see above). Based on previous experience with $L_{1750}$-vRNA co-crystallization the following lengths of partially double-stranded 3′ vRNA were tested: 3′1-15/5′9-15, 3′1-16/5′9-16, and 3′1-17/5′9-16 (GE Dharmacon). Following L-vRNA complex purification on the size exclusion chromatography, samples were concentrated to 5 mg/ml and supplemented with fresh vRNA to ensure that all L proteins are saturated with vRNA. Initial, precipitate-like hits obtained in presence of pre-annealed, partially double-stranded 3′1-16/5′9-16 vRNA in buffer containing 100 mM Tris buffer pH 8.0, 100 mM NaCl, and 8% PEG 4000, were manually reproduced and optimized in hanging drops. During soaking with the glycerol cryo-protectant a dehydration approach was used in order to rigidify the crystals and increase their diffraction capacities. It was achieved by the stepwise exchange of the cryo-protectant solution with increasing concentration of glycerol from 5 to 30%. Crystals were incubated 0.5-1 hour at each glycerol concentration point, and flash-frozen at the end.

### 2.4.3 X-ray crystallography – data collection

All data collection experiments were performed at the European Synchrotron Radiation Facility (ESRF) on ID29 and ID23-1 beamlines both equipped with Pilatus 6M-F detectors. MxCuBE software was used to conduct the diffraction experiments (Gabadinho et al., 2010). Following the collection of 2-4 initial images 90° from each other, the strategy was designed based on the EDNA (**E**nhanced automate**D** collectio**N** of dat**A**) diffraction plan (Incardona et al., 2009). In some cases additional, manual investigation of the frames in the Mosflm software was needed (Leslie, 2006). Data collection strategy is planned based on initial autoindexing performed by EDNA which estimates the crystal orientation, its space group, and probable dimension of the unit cell. To design the data collection set of parameters need to be specified. Listed below are example parameters for a standard data collection at ID29: oscillation start (defines the starting angle at which the crystal is held by the goniometer arm); oscillation range = 0.1 (a single frame will be collected every 0.1°); number of images = 2000 (together with oscillation defines the angle of rotation during entire data collection, here 2000 x 0.1° = 200°); exposure time = 0.03 sec (time used to expose a single frame – in this case entire experiment last 2000 x 0.03 sec = 60 sec); transmission given in % (allows to set up beam attenuators according to the current machine power [mA]). In special cases a mesh scan was performed covering entire crystal in order to define the position resulting in the highest diffraction (Bowler et al., 2010). The helical data collection was frequently used in case of long bar-like crystals. It allows to define two points at the crystal edges between which the data collection will be done. While collection proceeds, the goniometer apart from rotating the crystal is sliding it from one end to the other. This minimizes the radiation damage by spreading it throughout the whole crystal, which usually results in better data quality. It is particularly helpful during anomalous data collection. Another useful device available at some ESRF beamlines is the MiniKappa goniometer (Brockhauser et al., 2013). After performing initial characteristics of the crystal MiniKappa can align the crystal along its longest axe, which results in recording the Bijvoet mates (a reflection and the Friedel pair of its symmetry equivalent) on the same image.

### 2.4.4 X-ray crystallography – data processing

Data collection experiment results in a set of diffraction images. They need to be processed in order to obtain a list of Miller indices (hkl) with their corresponding intensities. To achieve this data was integrated and scaled with XDS and XSCALE (Kabsch, 2010). XDS is a command line program composed of several subprograms which are run in a row unless defined otherwise (XYCORR, INIT, COLSPOT, IDXREF, INTEGRATE, CORRECT). The first three prepare diffraction images date for actual analysis – they correct image axes, calculate the background, correct for beam deviations, and finally localize the spots. The IDXREF performs an indexing step – it calculates crystal orientation, matrix and symmetry of its lattice and refines parameters defined in the XDS.INP input file. If IDXREF succeeds in indexing majority of the spots according to the space group chosen, the XDS pipeline can proceed. The next step is the spot integration performed by INTEGRATE. Unlike Mosflm who implements a two-dimensional integration, XDS performs a three-dimensional integration. Tight spacing of frames results in many spots spread not only in two dimensions on a single image, but also in the third dimension which results in the same spot appearing on several images. XDS INTEGRATE takes this into account and sums up all the recorded signal belonging to the same spot into a single intensity. Finally XDS runs CORRECT in order to introduce correction factors to the intensities, determined space group if remained unknown so far, and refine the unit cell parameters. Final statistics inform about the completeness and quality of the data. Integrated intensities are saved in the XDS_ASCII.HKL file

XDS generated intensities are not free from some systematic measurement errors – deviation in beam intensity, different exposed crystal volume due to rotation, radiation damage decreasing the signal in the later part of the data, and detector-related errors. All those errors need to be corrected by scaling the data. This was done by XSCALE programme.

### 2.4.5 $L_{1750}$-vRNA structure solution and model building

$L_{1750}$-vRNA complexes crystallize in P212121 space group with one complex per ASU (unit cell dimensions: 102.5, 141.0, and 165.5 Å). The first native crystals diffracted to 3.1 Å, but it turned out the resolution can be improved to 2.8 Å when native data was collected from crystals soaked with 2 or 5 mM tantalum clusters. In order to obtain experimental phases anomalous datasets from platinum, tantalum, and selenomethionine derivatives were collected. Due to radiation damage and metal-specific limitation none of those datasets had anomalous signal that would be strong enough in high resolution shells to allow structure solution. Two of the best selenomethionine datasets scaled together gave anomalous signal at 4.0-4.5 Å with overall resolution at 3.5 Å. Exceptionally good 5 mM tantalum cluster derivatives after scaling four datasets, measured at the absorption peak, gave strong anomalous signal in the 4.5-5.0 Å resolution range, with overall resolution at 3.5 Å and 13x multiplicity. This dataset was used to localize four sites for tantalum clusters in SAD experiment at autoSHARP (Vonrhein et al., 2007). These initial substructure was subjected to SHARP together with previously collected selenomethionine, platinum, and native datasets and refined using multiple isomorphous replacement with anomalous signal approach (MIRAS). The map obtained was of very good quality and – as it turned out once the model was built – only less than 10% of the protein was not visible in the density.

Initially the automatic fitting of the endonuclease domain (PDB 2XI5) did not work. Its position became clear only after fitting of FluB PB1 subunit (PDB 4WSB) and it had to be further split within the hinge region in order to properly place both lobes in the electron density. Methionine positions were located using the anomalous differences from selenomethionine crystals. The model was iteratively built with Coot (Emsley and Cowtan, 2004) and refined with REFMAC (Murshudov et al., 1997). Autobuilded models done in Buccaneer (Cowtan, 2006) were used to extend the model in previously unbuilt regions and to verify the ones already built.

Exceptionally good quality of the RNA electron density allowed unambiguous building of co-crystallized vRNA and revealed its critical role in crystallization. Apart from the single-stranded overhang buried in the protein cavity, the solvent-exposed short helix is on its blunt end interacting with the neighbouring protein molecule, thus greatly contributing to the crystal packing.

Native $L_{1750}$-vRNA crystals soaked with missing 5′ ends diffracted to the following resolutions: 5′p1-8 – 2.8 Å, 5′p1-10 – 2.8 Å, and 5′p1-11 – 3.1 Å. $L_{1750}$-vRNA crystal soaked with 5′p1-12 showed no RNA density in the appropriate binding site, but since it diffracted to 2.8 Å the dataset was still useful in building and refining previous models. Structures from all those crystals were solved by subjecting processed data to REFMAC rigid body refinement and refining one the available models against them.

PyMOL was used to draw all the structure figures (DeLano 2002).

### 2.4.6 L-vRNA structure solution and model building

L-vRNA complexes crystallize in C2 space group with two complexes per ASU (unit cell dimensions: 371.5, 145.2, 232.8 Å; 90°, 116°, 90°). Initially crystals diffracted to 8-9 Å. The resolution could be further improved applying the stepwise cryo-cooling procedure (described above), and using 100% ID29 transmission combined with helical data collection. The best dataset obtained reached 4.4 Å resolution. Structure was solved with MOLREP (Vagin and Teplyakov, 2010) using $L_{1750}$-vRNA structure as a model.

# 3. RESULTS

## 3.1 LACV L protein C terminus – structural characterization

### 3.1.1 L6 and L6-derived constructs – design, expression, and purification

Prior to the PhD project described in this thesis The ESPRIT method (Yumerefendi et al., 2010) performed by Dr. Juan Reguera from the Cusack group on the last 800 residues of the LACV L protein revealed candidates for potentially soluble constructs (Fig. 3.1.A). The longest one called 3A was expressed in *E. coli* and treated with papain by Dr. Juan Reguera in order to define the edges of the structured domain (Fig. 3.1.B). The following L6 construct served as a starting point to establish a high yield purification protocol (Fig. 3.1.C). Initially, a significant

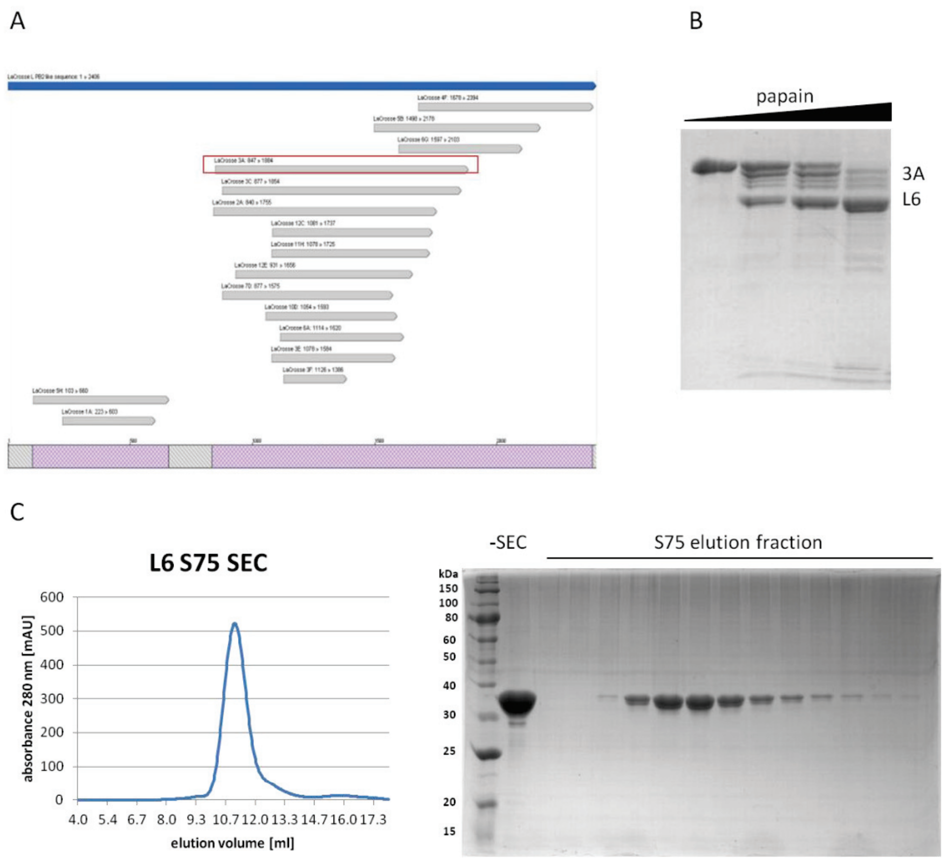
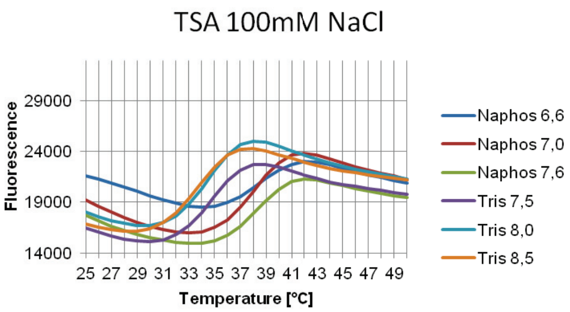
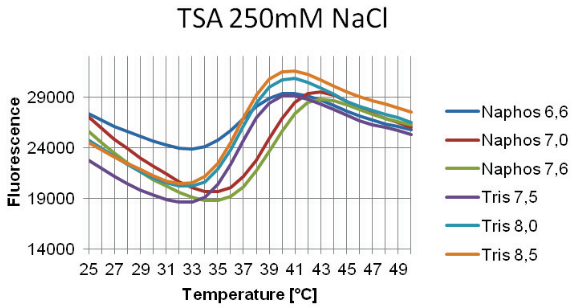



**Figure 3.1** A) ESPRIT technique-derived soluble constructs from La Crosse polymerase C-terminal 800 residues-long region (performed by Juan Reguera); B) L6 – a papain resistant fragment derived from 3A (performed by Juan Reguera); C) S75 size exclusion chromatography of purified L6.

Melting temperature

| | Naphos 6,6 | Naphos 7,0 | Naphos 7,6 | Tris 7,5 | Tris 8,0 | Tris 8,5 |
|---|---|---|---|---|---|---|
| **100mM** | 38,5 | 38 | 38,5 | 34,5 | 34 | 33,5 |
| **250mM** | 37 | 39 | 39 | 36,5 | 36,5 | 36,5 |
| **500mM** | 40,5 | 40,5 | 40 | 39 | 39 | 38,5 |



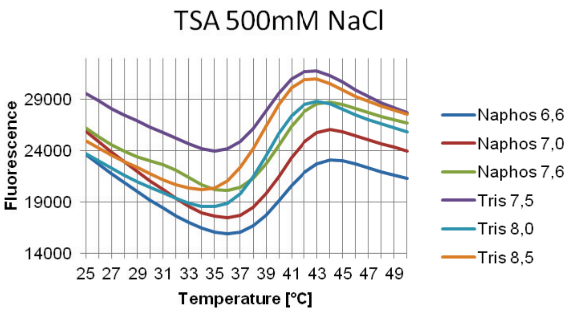




**Figure 3.2** Thermal stability assays performed on L6 showing the impact of different phosphate and Tris buffer, pH, and salt concentration on the melting point of the protein.

Amount of protein was precipitating, especially during final concentration steps after size exclusion chromatography. Thermal stability assays (TSA) were performed in order to define the melting temperature of the construct and to screen for the optimal buffer and salt concentration. The TSA results showed that the L6 construct has a melting temperature below 45°C which decreases the chances for obtaining a diffracting crystal (Dupeux et al., 2011), and that higher salt concentration and phosphate buffer tend to improve the stability of the protein (Fig. 3.2). It is worth noting that the L6 gives a well-defined unfolding curve meaning that the protein is at least partially structured.

Based on these results and following three different 2nd structure prediction software outputs and the multiple alignment of *Orthobunyavirus* L proteins, a new set of constructs were designed in order to refine the domain boundaries (Fig. 3.3). Among those constructs four were not soluble, one was not susceptible for TEV cleavage, two were precipitating heavily during purification, and the last two – 7AR and 8SR – gave a high yield of expression and could be finally concentrated up to 25 mg/ml. Several purification trials testing different buffer composition led us to the optimization of the 8SR purification protocol. Final buffer contained:

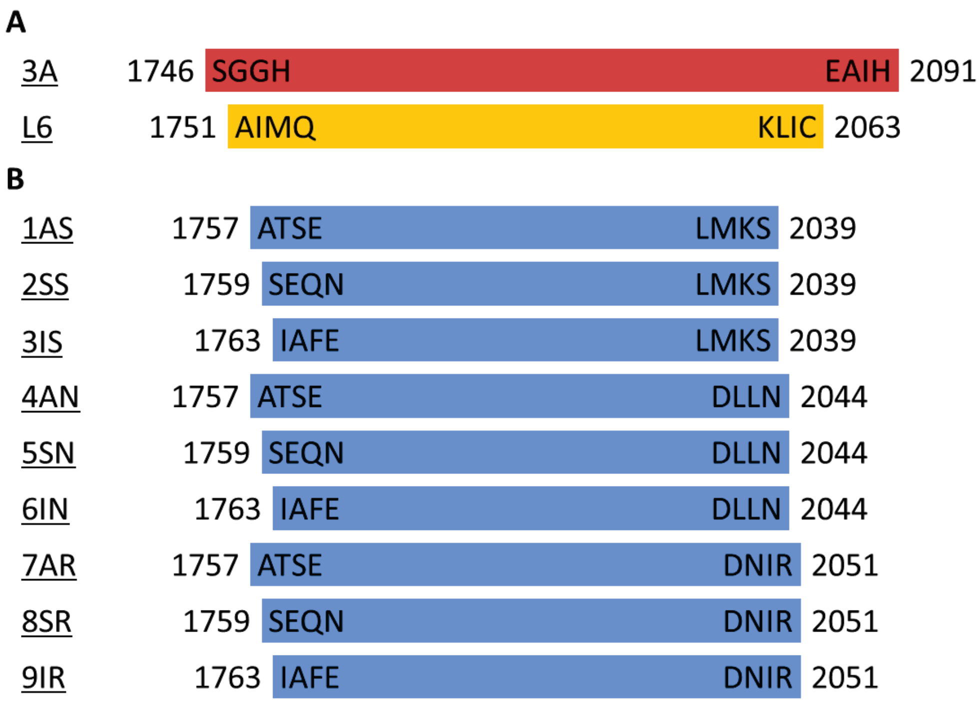



**Figure 3.3** A) Initial constructs after ESPRIT technique and papain digestion (Juan Reguera); B) Constructs designed based on <Prof prediction>, <GOR 4> and <Porter> softwares.

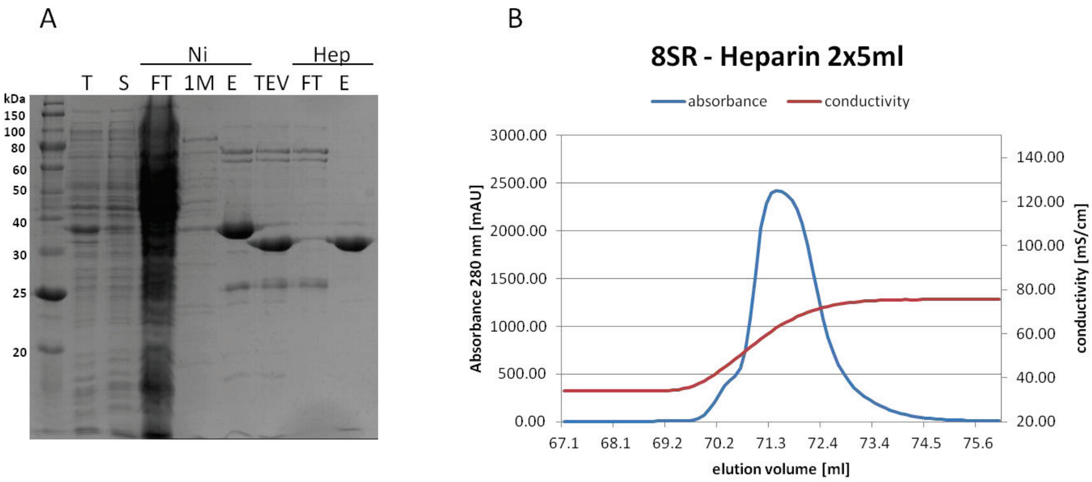



**Figure 3.4** A) SDS-PAGE showing improved purification protocol for 8SR: T – total fraction after sonication, S – soluble fraction, Ni – nickel column, FT – flow through, 1M – 1 M NaCl washing step, E – elution, TEV – sample after dialysis with TEV, Hep – Heparin column; B) 8SR profile after 2x5ml Heparin column chromatography against NaCl gradient.

20 mM Tris pH 7.0, 0.4 M NaCl, 5% glycerol, 20 mM imidazole, and 10 mM βME. The substitution of the 2$^{nd}$ Ni column step by the Heparin affinity chromatography resulted in more efficient purification and increased the final amount of protein sample. The NaCl gradient used during Heparin chromatography ranged from 0,4 M to 1 M with an elution peak at around 0,66 M concentration (Fig. 3.4).

### 3.1.2 7AR and 8SR constructs – crystallisation trials

Both 7AR and 8SR samples were used to set up crystallisation trials at the HTX facility. The crystal hits appeared in conditions containing sulphate anions. The most promising crystals grew after 1 week in 4°C from 8SR sample mixed with 100 mM tri-sodium citrate dihydrate pH 5,6 and 1 M magnesium sulfate heptahydrate (Fig. 3.5.A). They were flash-cooled in the same buffer supplemented with 20% glycerol but showed no diffraction pattern. These initial 8SR crystals could not be further improved by an additive screen or seeding. The remaining crystal hits were circular-shaped packs of clusters (Fig. 3.5.B). They tend to appear in 1/3 of additive screen drops but they also gave no diffraction pattern. To check whether they were protein crystals they were boiled in Laemmli buffer and analysed on the SDS-PAGE (Fig. 3.5.C). As mentioned above it was not possible to reproduce the initial 8SR crystals either in the sitting

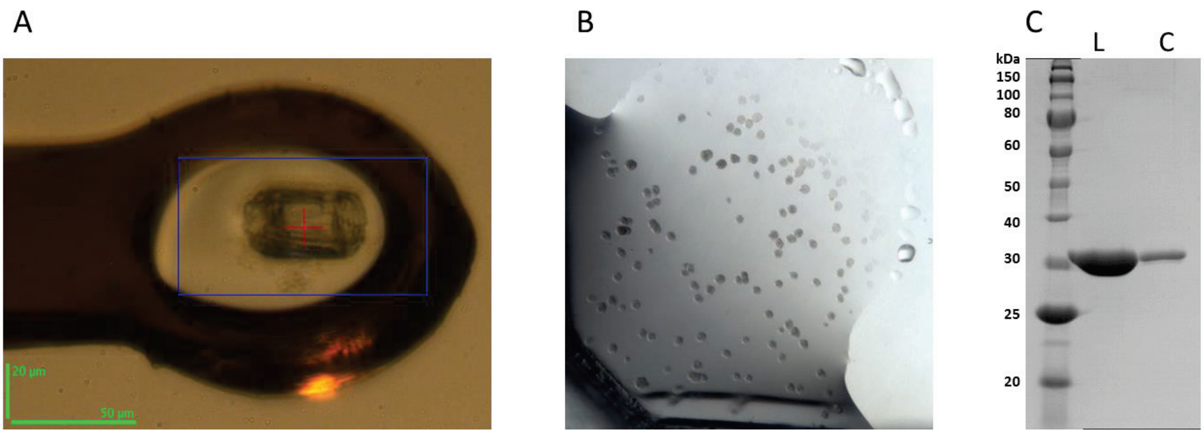



**Figure 3.5** A) 8SR initial crystal hit; B) Multiple 8SR circle-shaped packs of clusters obtained from additive screens; C) SDS-PAGE gel of 8SR pure liquid sample (L) and a 'circular crystal' boiled in the Laemmli buffer (C).

drop or hanging drop vapour diffusion experiments performed in the 24-well plates. However, the shower of tiny 8SR crystals of similar three dimensional shape appeared in the MPCS microfluidic experiment (Gerdts et al., 2008). Initially this experiment was supposed to screen for a proper magnesium sulphate concentration, where the range from 0,5 M to 1,25 M was tested. Unfortunately those crystals also did not diffract.

### 3.1.3 8SR construct – CD and 1D $^1$H-NMR analysis

Taking all previous results into account, a possible scenario has raised, saying that 8SR – being a derivative of the papain resistant and at least partially structured L6 construct – is either too flexible to be a promising candidate for crystal screening or it contains only part of the domain or even lies in between two domains. Circular dichroism spectroscopy and 1D $^1$H-NMR have been used to further investigate the conformational state of this construct. For both techniques the concentration of chloride anions had to be decreased as much as possible. This resulted in a very low protein solubility that lowered the concentration of the analyzed samples. In case of the CD experiment the remaining chloride anions at 200 mM concentration masked the far UV spectrum which is crucial to define with high accuracy the secondary structure of the protein and to calculate the α-helix/β-strand ratio. Nevertheless, the CD spectrum showed a peak characteristic for α-helical structures, and it was possible to conclude that ~20% of the protein is α-helical, ~40% is composed of random coils, and the rest of the protein is non-structured (Fig. 3.6.A).

In spite of the partial precipitation of the 8SR sample during 1D $^1$H-NMR experiment, the data obtained further confirmed the CD analysis conclusions. By using 1D $^1$H-SOFAST experiment it was possible to observe that detectable $^1$H resonances of the sample are mostly located in the random coil regions (Fig. 3.6.B). Moreover, the $^1$H Het$^{noe}$-SOFAST experiment showed that the protein detectable by $^1$H-NMR had a relatively low proton density and behaved rather like an unfolded protein with only few elements of regular folding (λ=0,46) (Fig. 3.6.C).

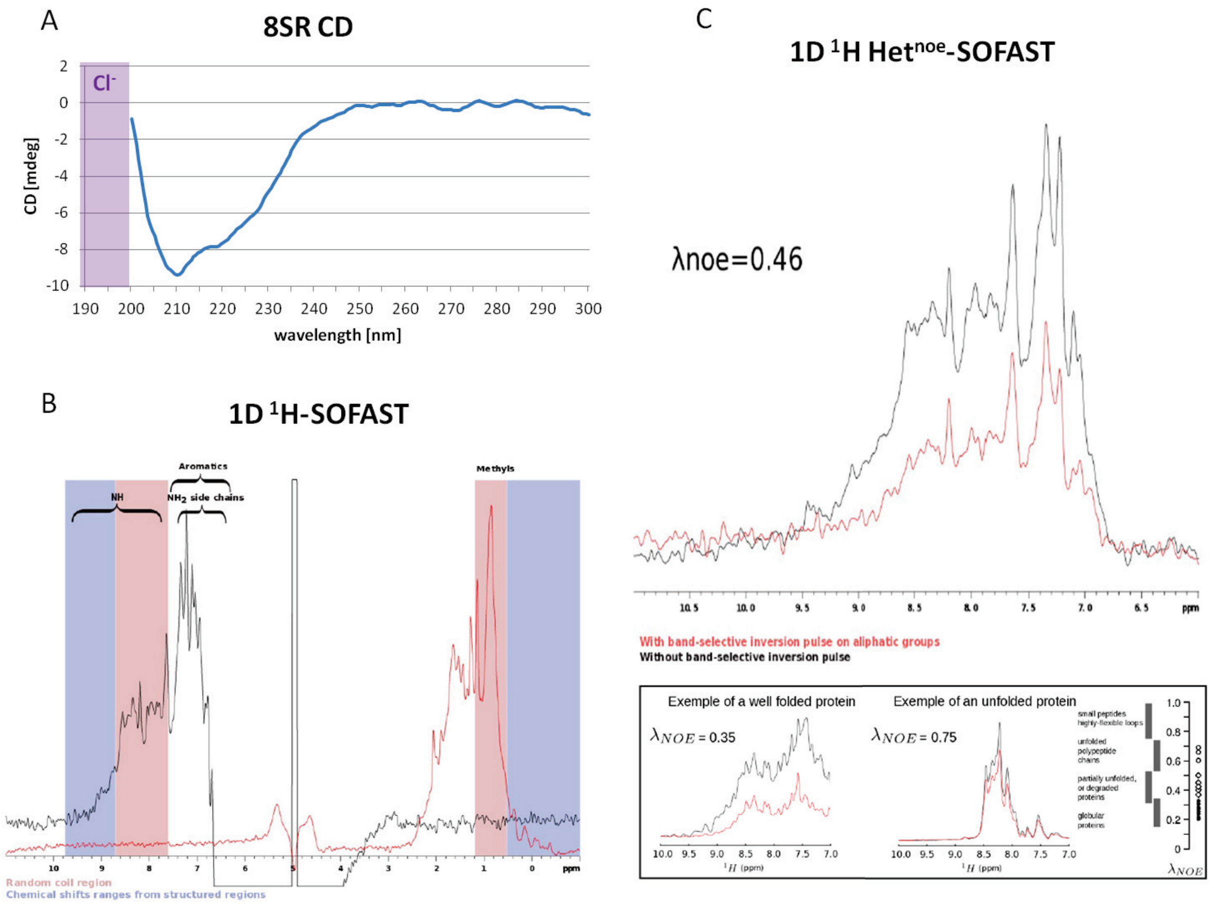


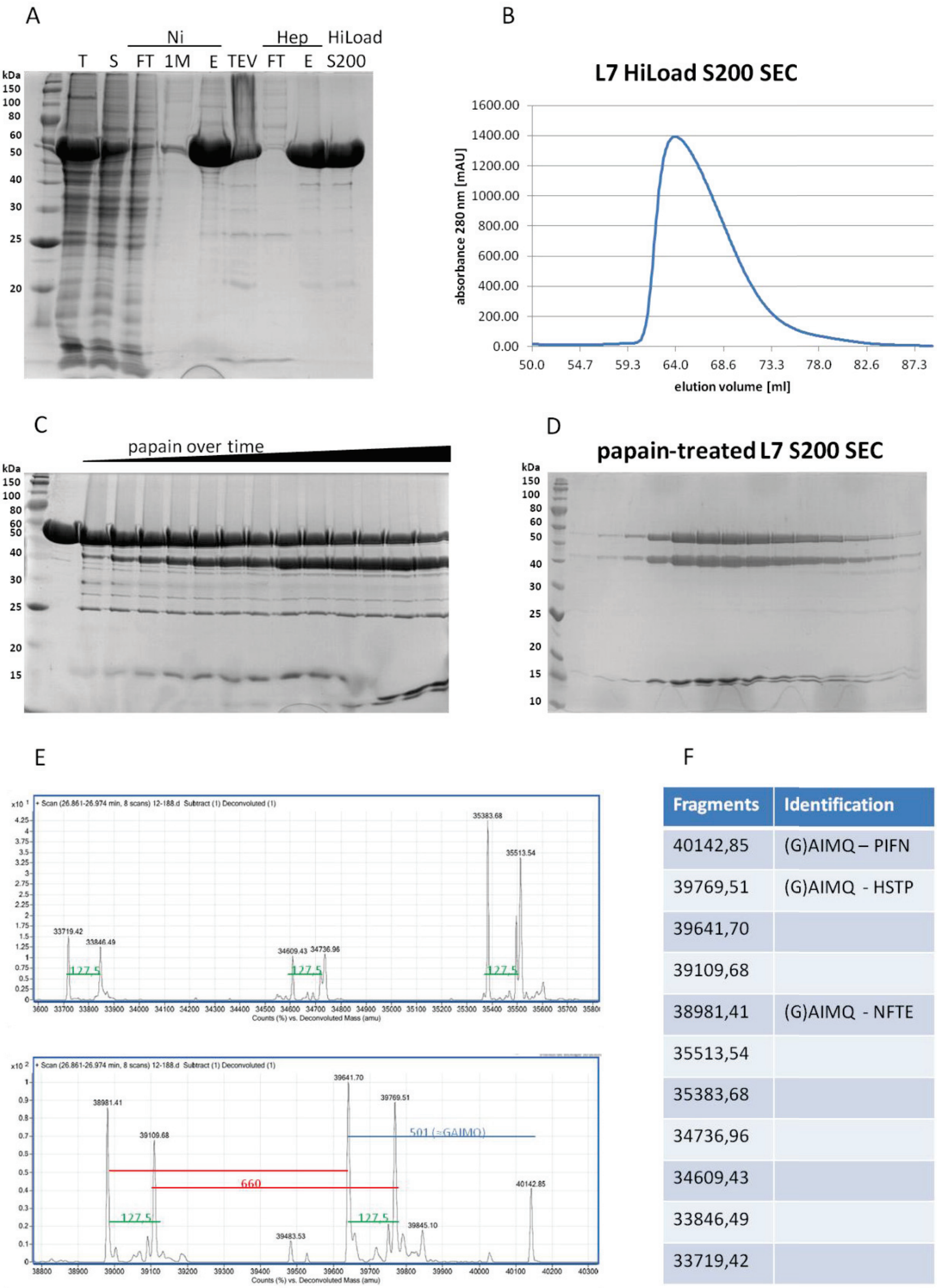

**Figure 3.6** A) 8SR circular dichroism spectra, the signal in far UV region highlighted in violet is masked by chloride anions; B) 1D 1H-SOFAST – distribution of NH and $CH_3$ hydrogens between random coils (red) and structured regions (blue); C) 1H Het$^{noe}$-SOFAST – ratio (λ) between NH – $CH_3$ NOE with irradiated $CH_3$ (NOE 'off') and the regular NH – $CH_3$ NOE signal (NOE 'on').

### 3.1.4 3A and L7 and other C-terminal construct – expression and purification

The *E. coli* expression and purification trials were performed for 3A and L7 constructs – the first one being the initial ESPRIT-derived construct and later one sharing with the L6 the same N terminus and going downstream untill the C terminus of the L protein (aa 1751-2263). In case of the 3A it occurred that the C-terminal biotin tag of around 20 aa, added for ESPRIT technique purposes (Yumerefendi et al., 2010) has a huge impact on the protein solubility. 3A construct bearing only the original L protein sequence was completely insoluble. This could suggest that there is a structured region downstream to the C-terminus of 3A, which when removed impairs protein solubility and capability of crystallization. In order to verify this hypothesis the L7 – the longest C-terminal construct analyzed so far – was expressed and purified from the Hi5 insect cells. Purification protocol already established for the 8SR was implemented with one modification – the lysis buffer contained additional 200 mM NDSB – a non-ionic detergent used to stabilize proteins (Fig. 3.7.A-B).

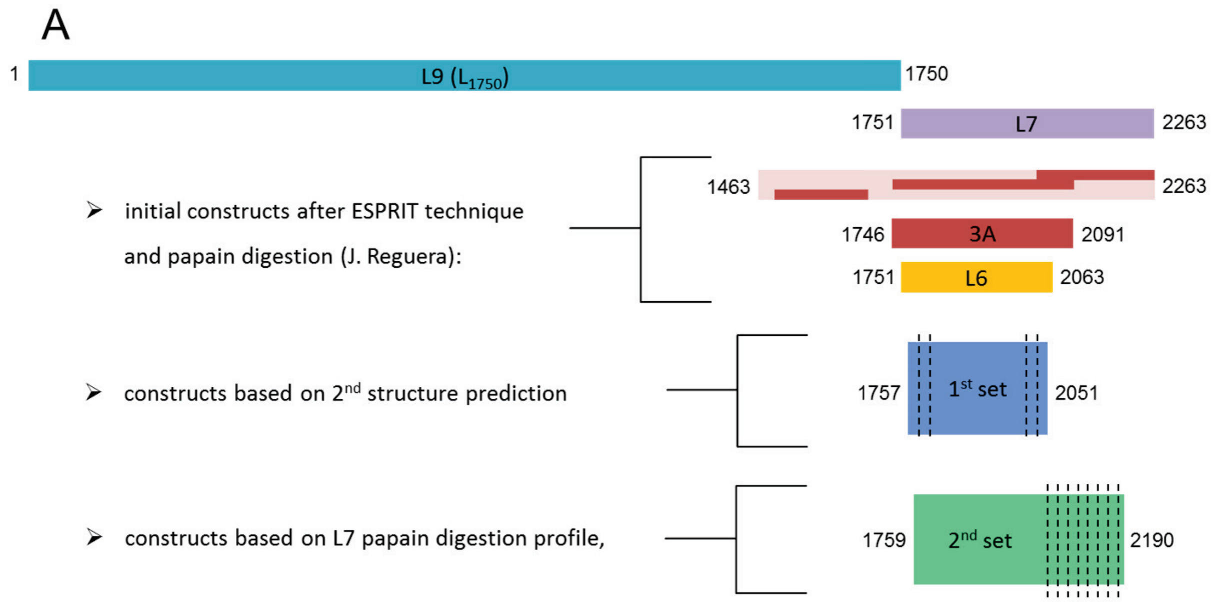
### 3.1.5 L7 C-terminal construct – limited proteolysis

The L7 construct was further treated with papain in the limited proteolysis experiment in order to define autonomous well-structured domains. The L7-papain ratio was 500:1 (w/w). The digestion was held for 2 hours in room temperature and reaction was stopped after each 10 min by adding Laemmli loading buffer and boiling the sample in 100°C. The new papain-resistant fragments resolved on SDS-PAGE suggested the presence of two main species of around 40 kDa and 15 kDa (Fig. 3.7.C). In the next step papain-treated L7 sample was loaded onto S200 size exclusion chromatography column. Surprisingly the uncleaved L7 co-eluted together with 40 kDa and a smaller 15 kDa digestion products (Fig. 3.7.D). This suggests a presence of an internal loop, connecting these two fragments. Apparently cleaving-off the loop does not lead to the disassembly of the two subunits.




| Fragments | Identification |
|---|---|
| 40142,85 | (G)AIMQ – PIFN |
| 39769,51 | (G)AIMQ - HSTP |
| 39641,70 | |
| 39109,68 | |
| 38981,41 | (G)AIMQ - NFTE |
| 35513,54 | |
| 35383,68 | |
| 34736,96 | |
| 34609,43 | |
| 33846,49 | |
| 33719,42 | |

**Figure 3.7** A) SDS-PAGE showing purification protocol for L7. B) L7 profile after HiLoad S200 size exclusion chromatography. C) time course of L7 limited proteolysis with 500:1 (w/w) papain, each line represents 10 min interval. D) papain-treated L7 profile after S200 size exclusion chromatography. E) ESI-MS results for L7-derived fragments; F) list of detected masses and identifications.




B

| | MW (kDa) | sequence | cross-validation |
|---|---|---|---|
| TRYPSIN | 79.246 | (1555) R \| LLMRD – LINSK \| F (2237) | Lprot specific, Cterm WB |
| | 49.428 | (1536) R \| ALEKE – HMAIR \| T (1958) | Lprot specific, Cterm WB |
| | 37.218 | (1536) R \| ALEKE – WQTSR \| H (1854) | Lprot specific, Cterm WB |
| | 33.517 | (1959) R \| TRIYN – SEFTK \| L (2249) | Lprot specific, Cterm WB |
| | 24.976 | (1536) R \| ALEKE – SGGHK \| A (1750) | MALDI (Acetyl-Nterm), L9 specific |
| | 16.289 | (1959) R \| TRIYN – IYYSK \| R (2112) | Lprot specific |
| | | | |
| PAPAIN | 40.142 | (1751) [G]AIMQH – TPIFN \| I (2097) | Edman degr. (60% results) |
| | 39.769 | (1751) [G]AIMQH – IHSTP \| I (2094) | Edman degr. (60% results) |
| | 39.641 | (1755) Q \| HNATS – TPIFN \| I (2097) | Edman degr. (40% results) |
| | 17.460 | (2101) Y \| SKRGE – SEFTK \| L (2249) | Edman degr. |
| | 17.833 | (2101) Y \| SKRGE – TKLMK \| L (2252) | Edman degr. |

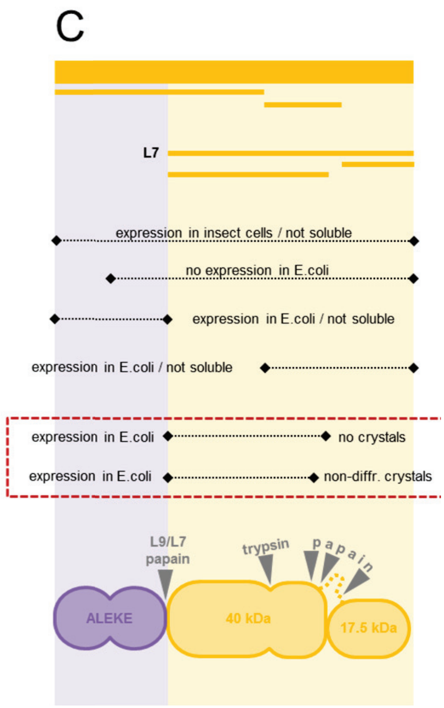



**Figure 3.8** A) Overview of the LACV L protein C-terminal constructs. B) Summary of the C-terminal protein fragments derived from trypsin or papain limited proteolysis. C) Overview of constructs designed based on limited proteolysis and tested for expression, solubility and crystallization, and a cartoon proposing the possible domain architecture of the C terminus

In order to define the exact sequence and mass of the 40 kDa fragment different analytical techniques were used. Trypsin digestion followed by MALDI-TOF analysis showed sequence coverage for this fragment from Ala-1751 to Ser-2101. Molecular mass calculated for such a construct equals 40,5 kDa. N-terminal sequencing has revealed that 60% of the analysed 40 kDa fragment starts from the Ala-1751, the N terminus of the L7 construct. Another 40% start from the His-1755. In parallel papain-treated L7 sample was analyzed by ESI-MS in denaturing conditions in order to define the mass of all species that appeared after digestion. ESI-MS analysis did not provide an unequivocal answer suggesting that there are numerous papain cleavage sites ~160 aa upstream from the L7 C terminus. The largest mass detected was 40,14 kDa which would match with the construct starting at the Ala-1751 and ending at the Asn-2097 (Fig. 3.7.E-F and Fig 3.8.B). This might confirm the 3A disrupted C-terminus hypothesis and explain the decrease of its solubility by the lack of a structured region situated downstream.

### 3.1.6 L7-derived 2$^{nd}$ set of C-terminal constructs

Based on data derived from the L7 papain limited proteolysis and followed by 2$^{nd}$ structure prediction and multiple sequence alignment a new set of constructs was designed. They start from the Ser-1759, like in case of the 8SR construct, and they differ in C termini with the shortest ending at the Ser-2078 and the longest going up to the Phe-2190 (Fig. 3.8.A). Majority of these constructs could not be expressed or purified from *E. coli*.

In general whenever the C-terminal construct could be purified and remained soluble, it was subjected for crystallization trials. The most promising were the 3A and the L7 constructs, which could be concentrated up to 5 mg/ml and remained soluble. Nevertheless, apart from initial crystallization hits for the 8SR construct, none of the following C-terminal constructs could be crystallized.

## 3.2 LACV L protein C terminus – functional studies and binding assays

### 3.2.1 L protein C terminus nuclease activity assays

The 3A-RNA interaction experiments with the set of different RNAs previously performed by Dr. Juan Reguera have led to the observation that this region of the L protein may contain a nuclease activity. In order to confirm this hypothesis and to define the substrate specificity of this putative nuclease as well as its metal ion dependence a set of RNA and DNA digestions has been performed (Fig. 3.9.A-C) Reactions were held in room temperature for 1-2 hours. For each construct the last purification step buffers were used. Reactions were stopped by adding the loading buffer containing 8M urea. Samples were analyzed by 15% urea PAGE. Provided results show no divalent metal ion dependence (Fig. 3.9.B). The DNA substrates remained uncleaved. The RNA substrates, both single- and double-stranded revealed a weak digestion pattern which most likely can be explained by the contaminations that remain in the sample after purification procedure (Fig. 3.9.A). There is no significant difference between the L7, the 3A and the 8SR construct RNA digestion ability (Fig. 3.9.C). Presented results lead to the conclusion that there is no nuclease activity within the LACV L protein C terminus.

### 3.2.2 8SR cap-binding assays

The 7mGTP sepharose resin was used to test the cap-binding activity of the LACV L protein C terminus. 0,25 mg of the 8SR in 1 ml sample buffer was loaded on the 0,1 mL of the resin. The washing buffer contained 20 mM Tris pH 7,0 and 200 mM NaCl. After three rounds of washing the elution was performed by implementing the washing buffer with 1 mM 7mGTP. Collected fractions were analyzed by SDS-PAGE. No 7mGTP-binding activity was detected (Fig. 3.10.A).

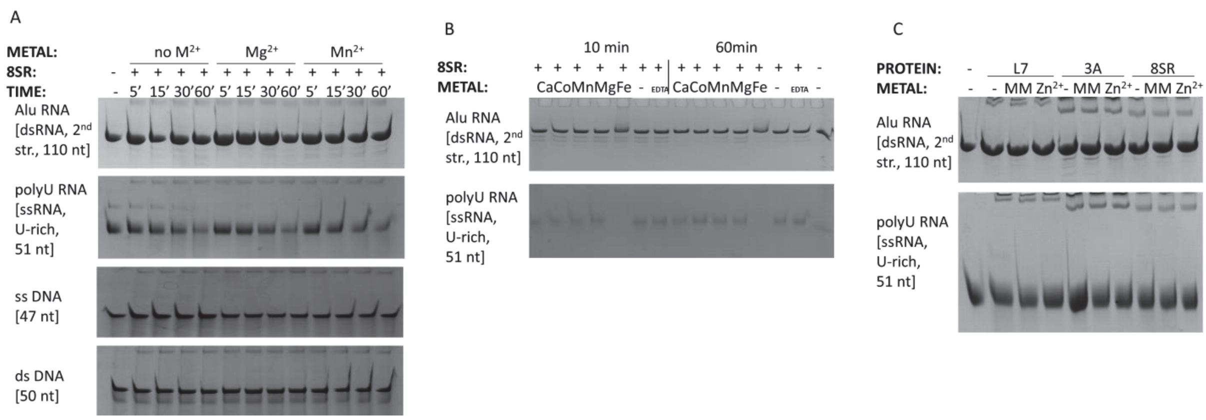



**Figure 3.9** A) 8SR nuclease assays – substrate specificity: 1 hour digestion of Alu dsRNA, polyU ssRNA, ssDNA and dsDNA; B) 8SR nuclease assays – divalent metal ion dependance: 1 hour digestion of Alu dsRNA, polyU ssRNA C) comparison of nuclease activity of L7, 3A and 8SR

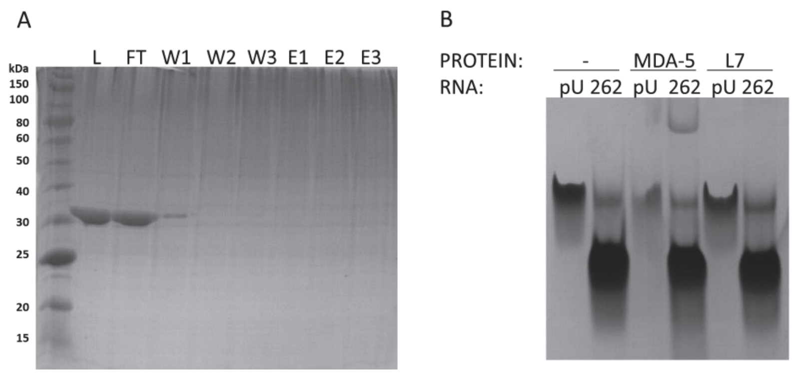



**Figure 3.10** A) 8SR construct binding to the 7mGTP sepharose resin. L – sample loaded on the resin; FT – flow through of the protein unbound to the resin; W1-W3 – washing with 200 mM NaCl; E1-E3 – elution with 1 mM 7mGTP. B) L7 construct RNA binding by electrophoretic mobility shift in native PAGE.

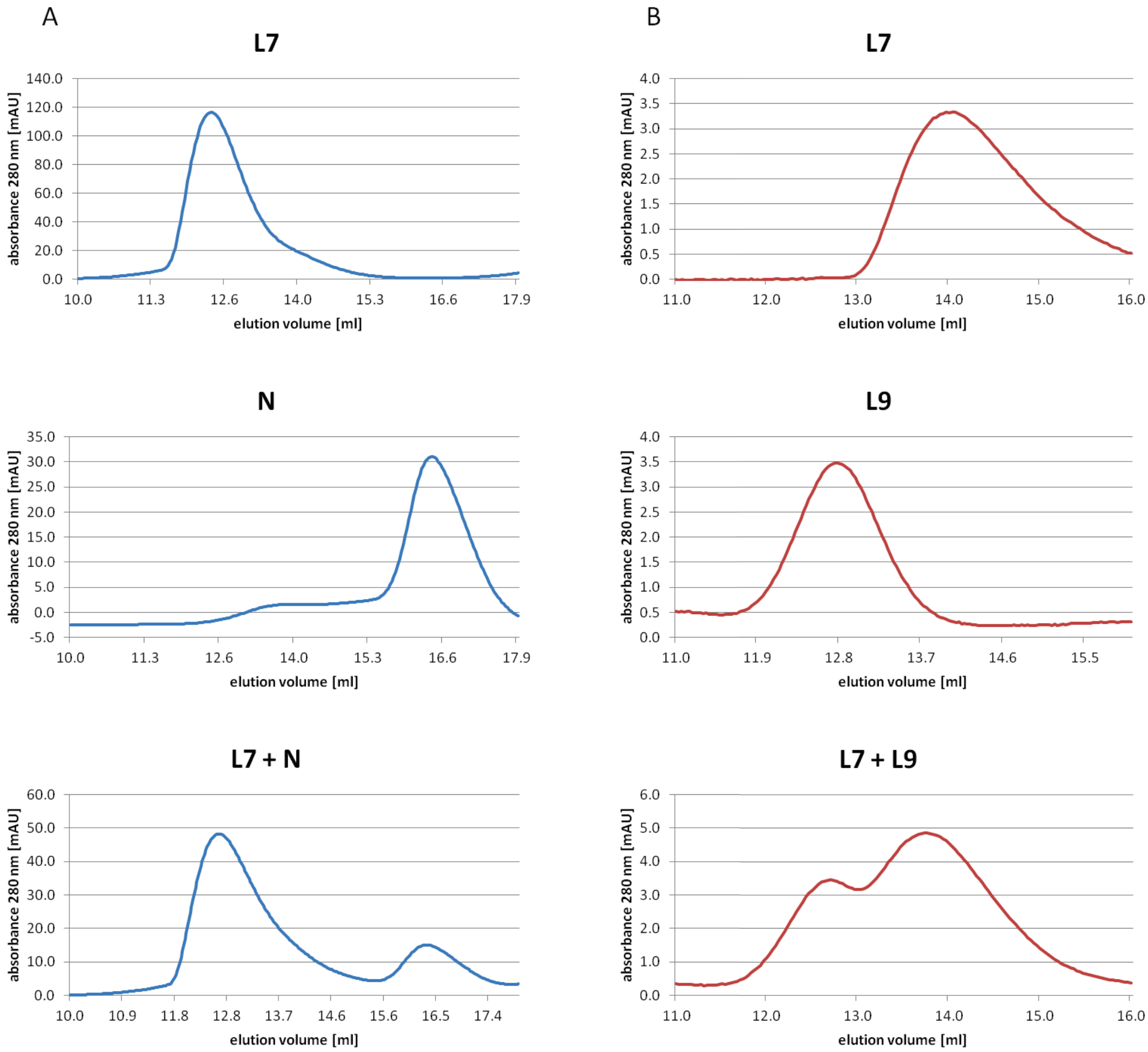



**Figure 3.11** A) L7-N interaction analyzed by S200 SEC; B) L7-L9 interaction analyzed by S200 SEC

### 3.2.3 RNA binding assays

In order to test whether the C-terminal region of L protein has any RNA-binding activity, L7 construct was mixed either with polyU ssRNA or with dsRNA ‘262’ containing two hairpin structures. MDA-5 protein was used as a positive control (both ‘262’ RNA and MDA-5 samples were kindly provided by Dr. Emiko Uchikawa). Incubation was held on ice for 1 hour. After that samples were analyzed by 8% native PAGE and stained with methylene blue. No RNA-binding activity has been noticed for L7 construct (Fig. 3.10.B).

### 3.2.4 Protein-protein interactions

In parallel the possibility of protein-protein interaction between L7 and nucleoprotein (N) as well as between L7 and L9 (aa 1-1750) has been tested. Appropriate pairs of proteins have been mixed together and left for 1 hour incubation on ice. Samples were than analyzed by S200 Sepharose size exclusion chromatography. The results obtained indicate that either in case of L7-N (Fig. 3.11.A) or L7-L9 (Fig. 3.11.B) the proteins migrate separately. This suggests that for these pairs of proteins there are no interactions detectable by size exclusion chromatography.

## 3.3 LACV full length L protein and $L_{1750}$ construct – RNA binding

### 3.3.1 L protein and $L_{1750}$ – expression and purification

The $L_{1750}$ construct (also called L9) comprises residues 1-1750 of the full length L protein and lacks C-terminal 513 residues. It was designed following trypsin limited proteolysis of the full length L protein performed by Dr. Juan Reguera. Its C-terminal end was also confirmed by the ESPRIT analysis. Both $L_{1750}$ and L protein constructs were expressed in Hi5 insect cells using baculovirus expression system (a detailed description of the purification protocols can be found in 'Methods'). Following sonication, the proteins were purified from the insect cell lysate using nickel affinity column, heparin column, and size exclusion column as the final step. Later on the purification procedure was extended by addition of the ammonium sulfate precipitation step performed before the nickel column. This allowed to remove cellular nucleotides and nucleic acids at the beginning of the purification. Unlike the $L_{1750}$, L protein had tendency to precipitate during final purification steps and could not be concentrated to the satisfactory concentrations, required for crystallization experiments. To overcome this L protein sample eluted from the heparin column by the buffer containing ~1M NaCl was supplemented with the vRNA and after dialysis against the buffer containing 150 mM NaCl purification was continued. Addition of the vRNA led to L-vRNA complex formation which stabilized the protein. Figures 3.12.A and 3.12.B present the examples of purified full length L protein and $L_{1750}$ respectively.

### 3.3.2 Electrophoretic mobility shift assays

The 3' and 5' ends of bunyavirus genomic RNA segments are highly conserved and exhibit extended complementarity to each other. In order to study the nature of interaction between vRNA ends and the full length L protein or the $L_{1750}$, various RNA-protein binding experiments and competition assays were done.

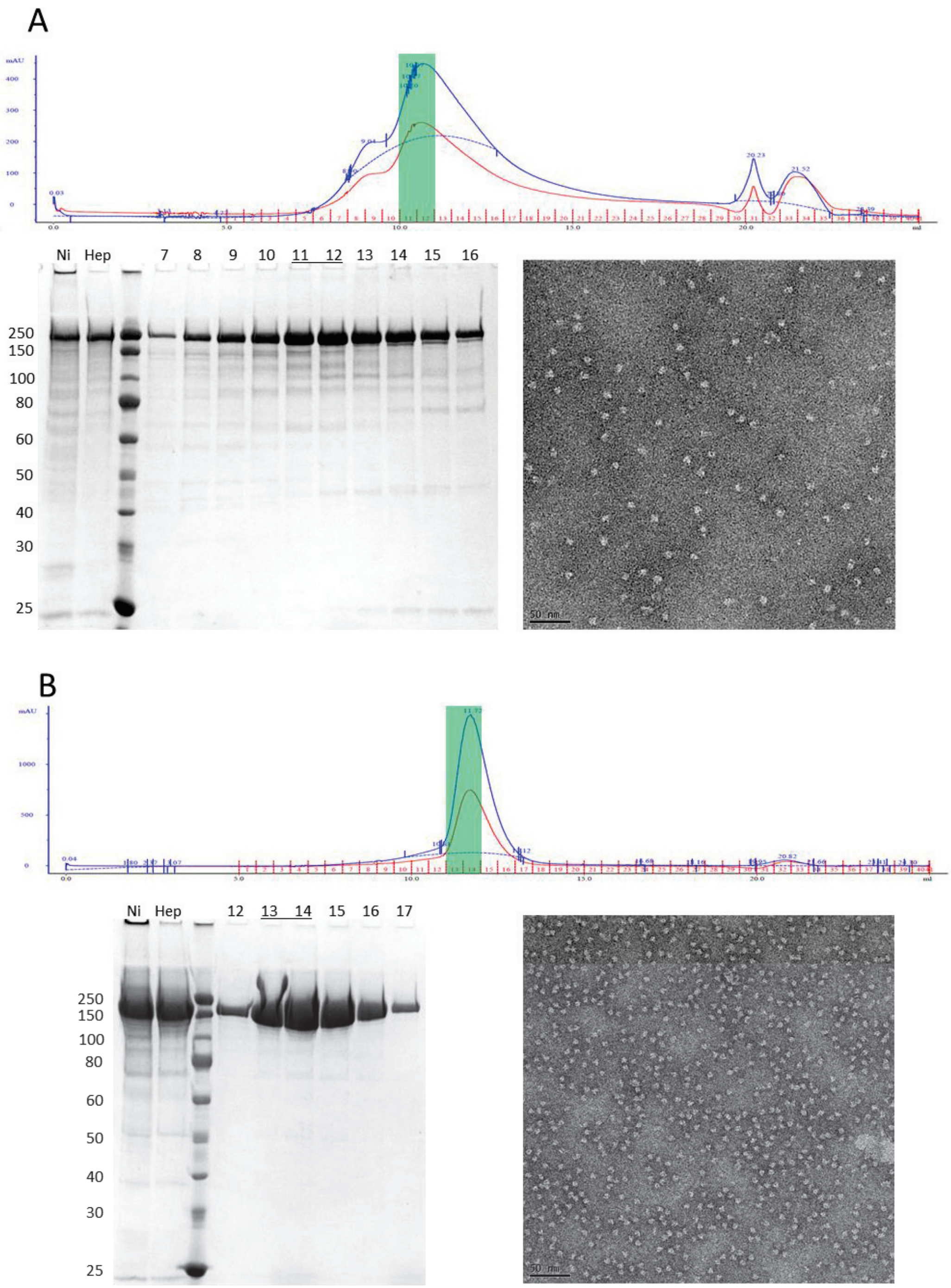



**Figure 3.12** L full length and $L_{1750}$ purified proteins are presented in panel A and B respectively; upper part of each panel represents the size exclusion profile, marked in green are fractions used for negative stain EM analysis; bottom left part of each panel represents SDS-PAGE; bottom right part of each panel represents the micrograph from negative stain EM

In electrophoretic mobility shift assays (EMSA) both radioactively and fluorescently labelled RNA were used. Labelled vRNA at low concentrations were mixed with the excess of the protein. An excess of the non-labelled poly-uracil RNA (polyU) was added in the assays to minimize unspecific binding of the vRNA by the protein. In the competition assays the non-labelled vRNA was used. Upon several hours of incubation mixtures were resolved in the native PAGE.

EMSA assay with 0.1 µM of the fluorescently labelled 3′ vRNA and 2 µM of the $L_{1750}$ revealed that all vRNA binds to the protein. Since it could not be competed by the polyU RNA, it suggested strong and specific interaction. The binding was partially competed by the same but non-labelled 3′ vRNA. Trials to compete the labelled 3′ vRNA by the non-labelled 5′ vRNA resulted in dsRNA formation. In native PAGE this dsRNA was migrating slower than ssRNA, as expected, but could not be bound efficiently by the polymerase (Fig. 3.13.A). Reciprocal experiments performed with the labelled 5′ vRNA revealed similar effect (data not shown). These results suggest that the $L_{1750}$ efficiently binds the 3′ and 5′ vRNA ends as single-stranded and not double-stranded molecules.

### 3.3.3 Fluorescence anisotropy

Fluorescence anisotropy technique was used to determine precise affinity values of the 3′ and 5′ vRNA ends binding by the $L_{1750}$ protein. As in case of the EMSA experiments, fluorescently labelled single-stranded 3′ and 5′ vRNA were used either alone or in presence of competing, non-labelled RNA. Labelled RNAs at fixed concentration were titrated with the $L_{1750}$ protein. Experimental data was used to calculate the dissociation constants (Kd). Results are presented in Figure 3.13.B, and in the supplementary material of the attached publication (Gerlach et al., 2015).

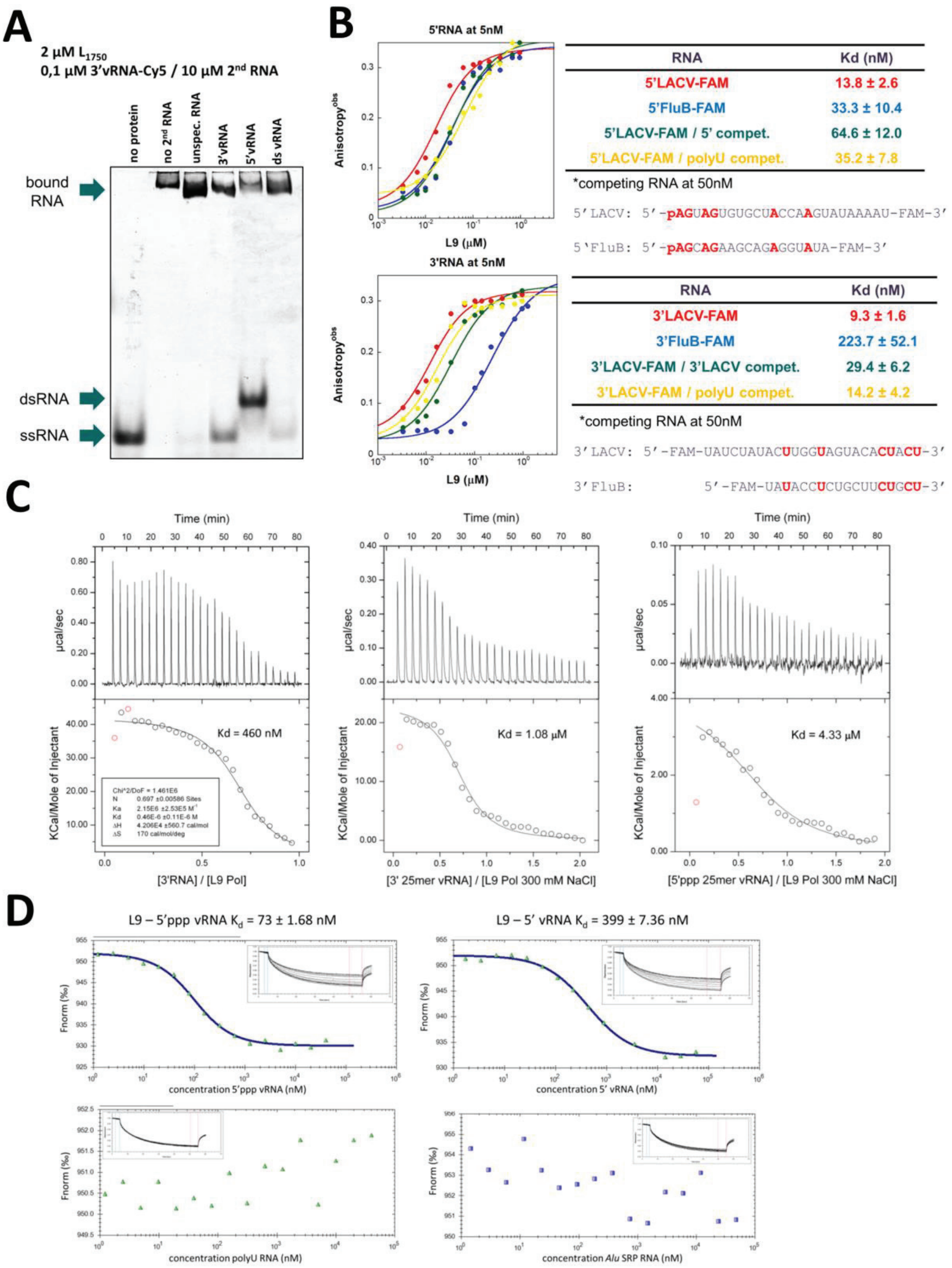



**Figure 3.13** $L_{1750}$-vRNA binding studies: A) EMSA assay in the native PAGE using fluorescently labelled 3′ vRNA. B) Fluorescence anisotropy with fluorescently labelled 3′ and 5′ vRNA; top panel represents results obtained for the 5′ vRNA, bottom panel represents results obtained for the 3′ vRNA; Influenza B 3′ and 5′ end genomic sequences were used as control RNA. C) ITC experiments with 25 nucleotide-long 3′ and 5′ppp vRNA; results presented in two right panels were obtained using 300 mM NaCl buffer. D) MST experiments with 25 nucleotide-long 5′ppp and 5′p vRNA; non-specific polyU and *Alu* SRP RNA were used as negative controls.

Both single-stranded 3′ and 5′ vRNA exhibit very strong, nanomolar affinity to the $L_{1750}$ protein. When competition assays with polyU RNA or non-labelled 3′ and 5′ vRNA were performed, both RNA were mixed together prior to $L_{1750}$ titration. It revealed that, at the initial phase of the titration, both 3′ and 5′ vRNA compete more for the $L_{1750}$ than the non-specific polyU RNA, resulting in increased Kd values. The Influenza B vRNA sequences have the same nucleotides as LACV vRNA in some positions (Fig. 3.13.B – nucleotides highlighted in red). As expected, the Influenza B 3′ vRNA exhibits much weaker affinity to the $L_{1750}$ than its LACV homolog. Interestingly, this is not the case for the Influenza B 5′ vRNA, which is bound by $L_{1750}$ with similar affinity than the LACV 5′ vRNA.

### 3.3.4 Isothermal titration calorimetry

Isothermal titration calorimetry was also used to determine affinities between 3′ or 5′ vRNA and $L_{1750}$. Due to ITC requirements protein concentration was at 30 µM, and the concentration of titrated vRNA was 10 times higher. At these conditions, and performing the experiment at room temperature, protein tends to precipitate, biasing the results. Nevertheless, measurement performed with the 3′ vRNA revealed significant affinity, with Kd=460 nM. To overcome protein precipitation the NaCl concentration in the protein buffer was increased to 300 mM. Higher salt prevents protein precipitation but it disrupts RNA-protein interactions mediated mostly by salt-bridges. The results obtained at these conditions do not reveal exact affinities but allow to compare the binding of various vRNA by the $L_{1750}$. Data obtained revealed that the 5′ppp vRNA binds weaker to the $L_{1750}$ than the 3′ vRNA (Fig. 3.13.C).

### 3.3.5 Microscale thermophoresis

The $L_{1750}$-vRNA binding was also analysed by the microscale thermophoresis. This technique requires much lower concentration of the protein. However, a significant drawback is the fluorophore labelling of one of the interacting molecules, which might affect the binding.

During the experiments fluorescently labelled $L_{1750}$ at 50 nM concentration was incubated with the vRNA at 16 different concentrations, ranging from 40 µM to 1 nM. In ths case both 5′ppp and 5′p vRNA were tested. Calculated Kd were 73 nM and 399 nM for 5′ppp and 5′p vRNA respectively (Fig. 3.13.D). These values are drastically different from the ones obtained by ITC, but they resemble the Kd range obtained by fluorescence anisotropy experiments. Difference between binding affinities of the 5′ppp and 5′p vRNAs to the $L_{1750}$ suggests the importance of the triphosphate group in efficient recognition of the 5′ genomic end.

### 3.3.6 vRNA protection during $L_{1750}$ and L protein trypsination

To further characterize the effect of the vRNA-$L_{1750}$ binding on the protein conformation comparative limited tripsination assays were performed with eiher apo form of the $L_{1750}$, or in complex with the 3′ or 5′ vRNA end. Following $L_{1750}$-vRNA incubation the trypsin was added in the 1:1000 trypsin – $L_{1750}$ ratio (w/w). Reactions were carried at room temperature and were stopped every 20 minutes by addition of the SDS-PAGE loading buffer. Upon SDS-PAGE analysis it was noticed that the interaction with vRNAs induces changes in the trypsin digestion pattern. Moreover, these changes were 3′ and 5′ vRNA specific (Fig. 3.14.A). Similar effects were observed when the full length protein was used. L protein trypsination products were further analysed by western-blot with antibodies recognizing the N-terminal His-tag, the N-terminal endonuclease or the C terminus of the protein (Fig. 3.14.C). Mass spectrometry analysis and Edman degradation allowed to assign vRNA-mediated differences in the trypsination pattern to the N-terminal region of the protein.

Tryspin digestion fragments are summarized in the supplementary material of the attached publication (Gerlach et al., 2015). Upon interaction with the 3′ vRNA the N-terminal 1-368 trypsin digest observed in the apo polymerase profile is extended to 1-430. On the other hand upon adding the 5′ vRNA trypsin cleavage at K430 disappears and the 431-862 fragment is extended to 377-862. These results together strongly suggest that the region in the polymerase between residues K368 and K430 changes its conformation or at least becomes inaccessible for the trypsin upon 3′ or 5′ vRNA binding (Fig. 3.14.A).

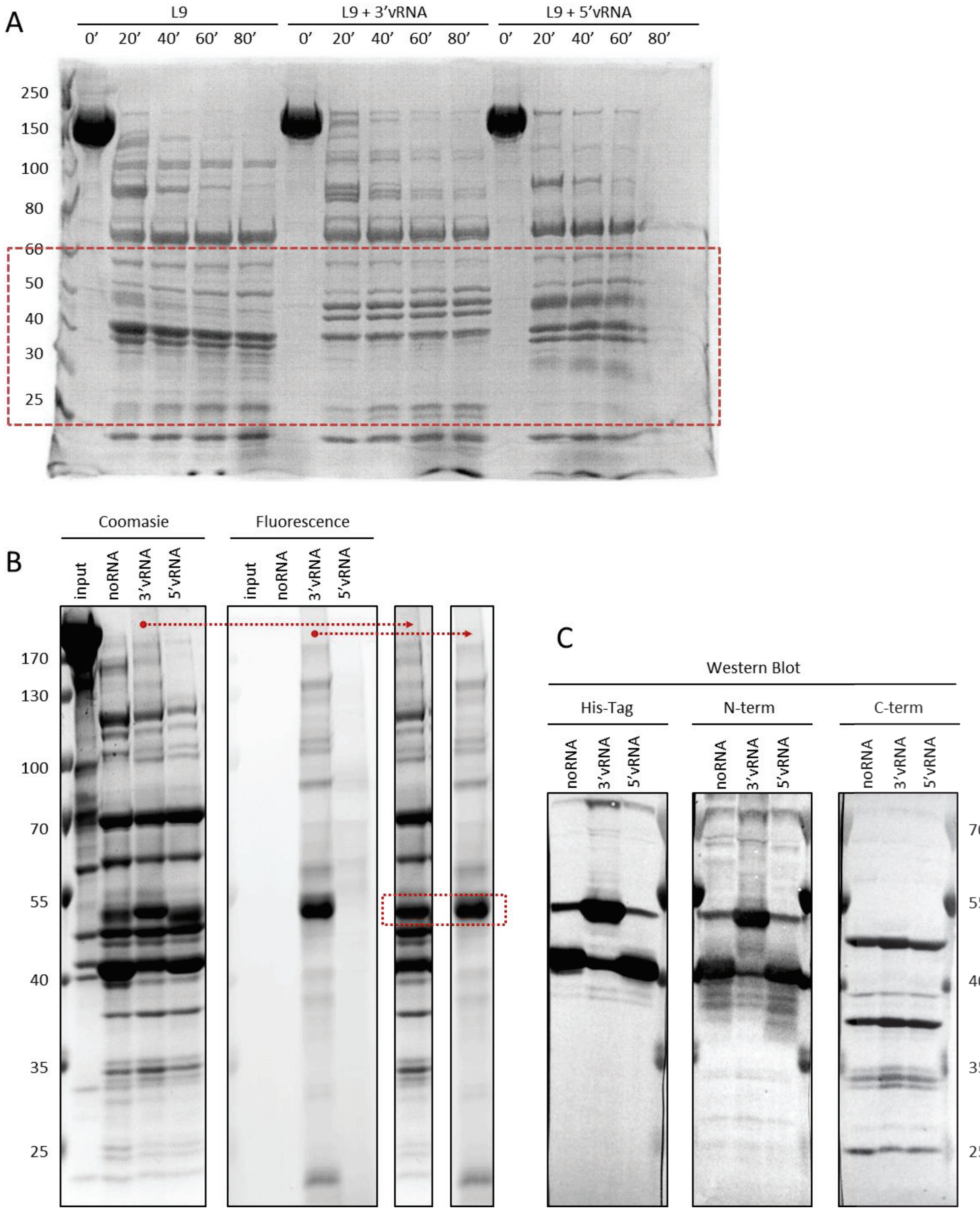



**Figure 3.14** A) Trypsin limited proteolysis of the apo-$L_{1750}$, $L_{1750}$-3'vRNA, and $L_{1750}$-5'RNA (1:1000 w/w trypsin-$L_{1750}$ ratio). B) $L_{1750}$-vRNA 4SU crosslinking. RNA-fluorescence has been scanned with Typhoon before the Coomasie staining. Selected gel lines are duplicated next to each other for easier comparison of $L_{1750}$-cRNA crosslinks. C) Western-blot of the full length L protein after trypsin digestion. Various patterns exhibited by anti-HisTag, anti-Nterm and anti-Cterm antibodies.

### 3.3.7 $L_{1750}$ – vRNA UV crosslinking

The vRNA protection assay, which allowed to identify that the N-terminal region of the L protein interacts with the vRNA, was further modified in order to narrow down which part of the N-terminal region is involved. To do so RNA-protein UV crosslinking step was performed prior to the trypsination. Specially designed, synthetic 3' or 5'vRNA were bearing two or one modified, photoactivable 4-thio-uracil (4SU) respectively (Kramer et al., 2011). Both vRNAs were additionally labelled with a fluorophore on their non-interacting ends – DY647 in case of 3'vRNA and Cy5 in case of 5'vRNA. Fluorescence allowed to track the protein-RNA crosslinks. Like described previously, following trypsination samples were subjected to SDS-PAGE. A single 3'vRNA-protein crosslink was noticed. It migrated like the N-terminal trypsin digestion fragments (Fig. 3.14.B). In contrast no 5'vRNA-protein crosslink was observed.

The 3'vRNA-$L_{1750}$ protein crosslink was cut out from the gel and subjected to mass spectrometry analysis. No specific mass, different from the reference sample, could be detected. This is most likely due to very low efficiency of RNA-protein crosslinking resulting in low amount of the RNA-protein crosslinks, below mass spectrometry detection capacities.

## 3.4 LACV full length L protein and $L_{1750}$ construct – structural characterization

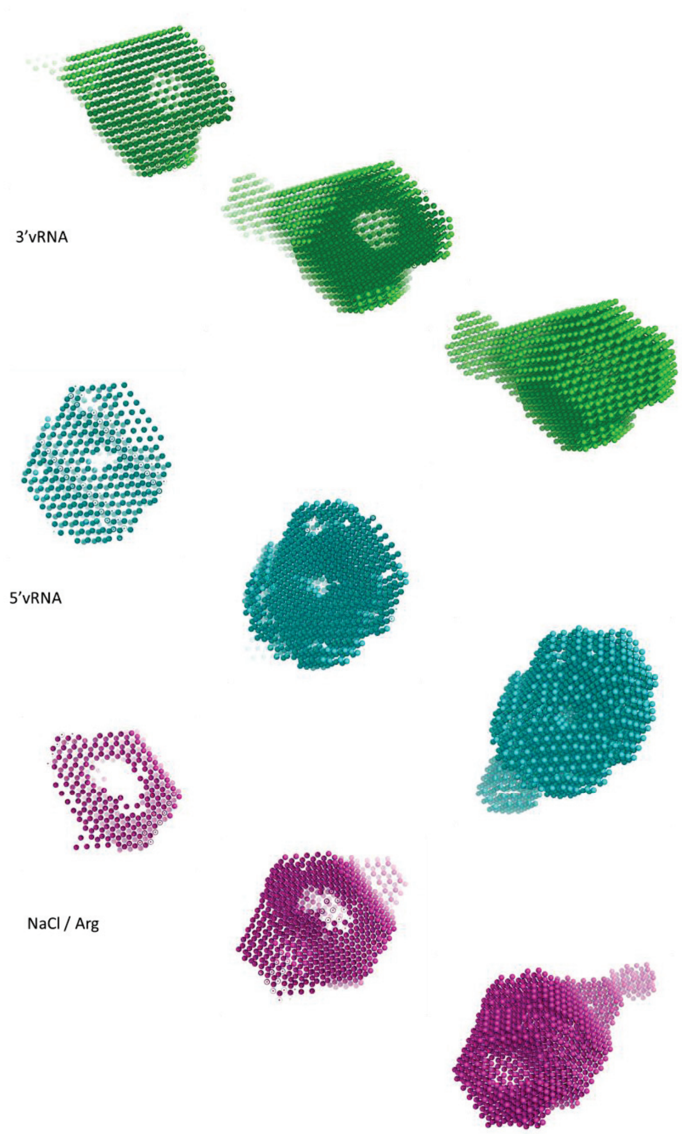
### 3.4.1 $L_{1750}$ and L protein – SAXS analysis

SAXS measurements have been performed on both the full length L protein and the $L_{1750}$. Different concentrations of samples were analysed. We noticed that radius of gyration (Rg) raises together with sample concentration suggesting oligomerization or other non-specific intermolecular interactions. These interactions can be diminished upon addition of the 3′ or 5′ 25-nucleotide long vRNA. The 5′ vRNA contained 5′ppp modification. Rg for such samples are 4.1 nm and 5.5 nm for $L_{1750}$ and L protein respectively. Similar effect was obtained in absence of vRNA but upon addition of stabilizing agents to the buffer, namely 500 mM NaCl and 200 mM arginine.

Because of the data quality reasons Guinier plots were fitted only for the $L_{1750}$ sample. They suggest the presence of cavities or a hole inside the protein which can be seen in the ab initio models. Results of the SAXS analysis are summarized in Figure 3.15.

### 3.4.2 $L_{1750}$-vRNA co-crystallization experiments

Both the full length L and the $L_{1750}$ protein were subjected to many crystallization trials. Co-crystallization trials were performed with either separate 3′ or 5′ vRNA ends, or both ends mixed together. Initially all these trials were fruitless leading to the protein precipitation in majority of crystallization conditions. Following detailed studies of $L_{1750}$ and L-vRNA binding, described in more details above and in the attached publication (Gerlach et al., 2015), new set of RNA oligos, representing partially double-stranded 3′ vRNA end, was used in co-crystallization experiments with more stable $L_{1750}$ protein. Detailed composition of tested RNA is presented in Fig. 3.16.




| | Rg | Dmax | Volume |
|---|---|---|---|
| **L** | 7.5 | 24.65 | 999 |
| **L_ADD** | 5.076 | 15.86 | 487 |
| **L_3RNA** | 5.54 | 19.39 | 666 |
| **L_5RNA** | 5.61 | 19.56 | 653 |
| | | | |
| **L9** | 4.27 | 14.94 | 353 |
| **L9_ADD** | 4.188 | 13.04 | 340 |
| **L9_3RNA** | 4.128 | 14.45 | 341 |
| **L9_5RNA** | 4.103 | 14.36 | 333 |

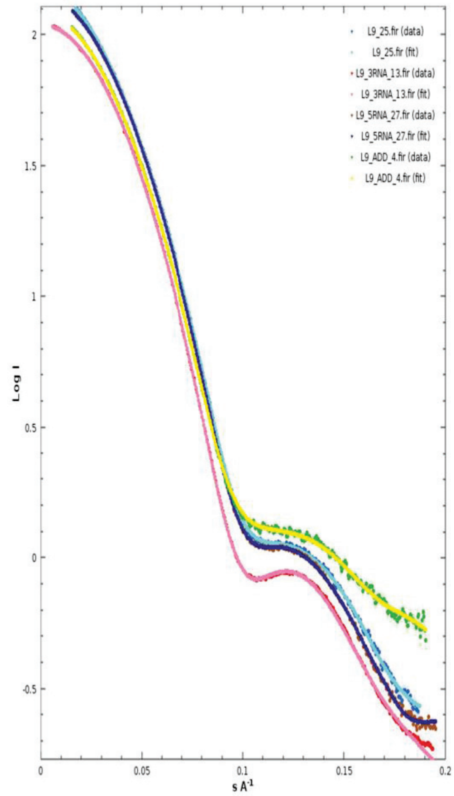



**Figure 3.15** SAXS measurements of the full length L protein and the $L_{1750}$ construct (L9). Table in the top right part of the figure summarizes differences between L protein and $L_{1750}$ Rg values; Chart in the bottom right represents Guinier plot fitting of the $L_{1750}$ sample derived data; Left part of the figure represents slices through the ab initio models showing overall shape and cavities of three different $L_{1750}$ samples.

The $L_{1750}$-vRNA crystals appeared only in presence of the partially double-stranded RNA composed of the oligo representing nucleotides 1-16 from 3′ vRNA pre-annealed to the oligo representing nucleotides 9-16 from 5′ vRNA. This combination of RNAs allowed to have an eight-nucleotide overhang of conserved 3′ vRNA terminal nucleotides, and a blunt dsRNA end on the opposite site. Other combinations bearing one nucleotide overhang instead of a blunt end were also giving crystals but of lower quality (Fig. 3.16). The $L_{1750}$-vRNA crystals in the form of brush-like clusters of thin needles grew in the narrow range of conditions containing 0,2 M sulfate or citrate salts, and 20% PEG 3350 or 15% PEG 4000 (Fig. 3.17.A). Additive screen was performed to optimize these initial crystals (Fig. 3.17.B). It revealed that 10 mM $Mn^{2+}$ but not $Mg^{2+}$ affects the crystal growth. It allowed also, by using both citrate and sulfate salts in the mother liquor, to separate the clusters of needles into single thin needles. Screening of the citrate and sulfate concentrations allowed to find the optimal mother liquor composition: 0.1 M Hepes, 0.2 M sodium citrate tribasic dihydrate, 0.3 M ammonium sulfate, 15% PEG 3350, pH 7 (Fig. 3.17.C). The final bar-like single crystals were obtained using microseeding procedure, which allowed to control the nucleation and greatly improved quality of the crystals. Figure 3.18 presents example of a single $L_{1750}$-vRNA crystal and the diffraction pattern it generated.

As described in the attached publication (Gerlach et al., 2015) it was also possible to soak the $L_{1750}$ – 3′ vRNA crystals with the missing 5′ vRNA end (8, 10, and 11 nucleotide-long) which resulted in another structure revealing two, separate binding sites for the 3′ and 5′ vRNA ends. However, due to the co-crystallization strategy involving partially double-stranded 3′ vRNA, it was impossible to track the continuity between 5′ vRNA hook and the short 5′ vRNA strand involved in formation of the partial 3′ vRNA duplex. Also, due to the conformation in which $L_{1750}$ crystallized, 5′ vRNA end longer than 11 nucleotides could not be soaked into the crystal most likely because of the steric clashes that would appear between nucleotide 12 of the 5′ vRNA and neighbouring protein residues.

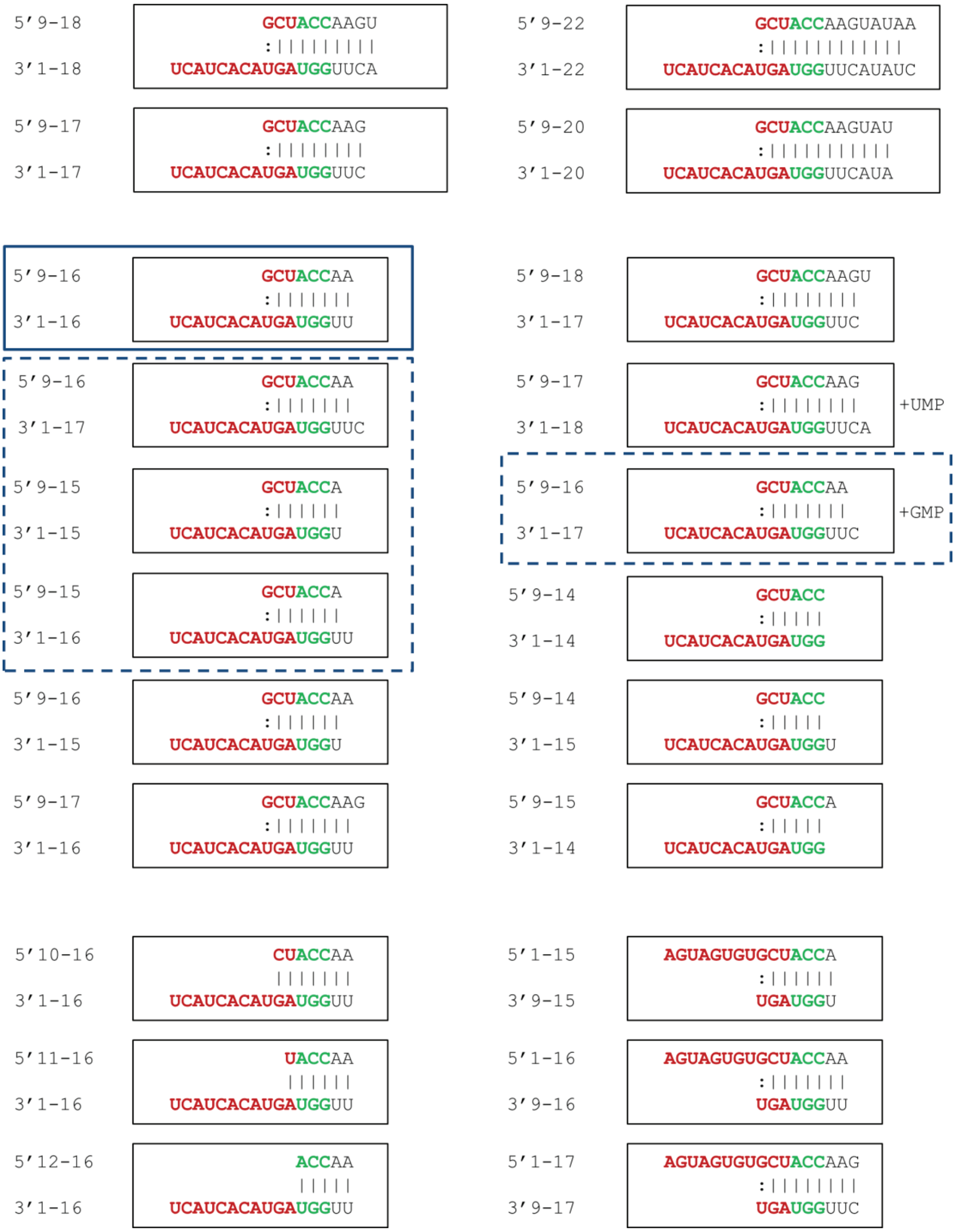



**Figure 3.16** Partially double-stranded RNA used in co-crystallization experiments with $L_{1750}$. Initial, conserved 11 nucleotides are marked in red, following duplex region (Barr and Wertz, 2004) is marked in green. RNA optimal for crystallization is marked with blue rectangle. RNAs that also allowed co-crystallization, but resulting in lower quality crystals, are marked with blue dashed rectangles.

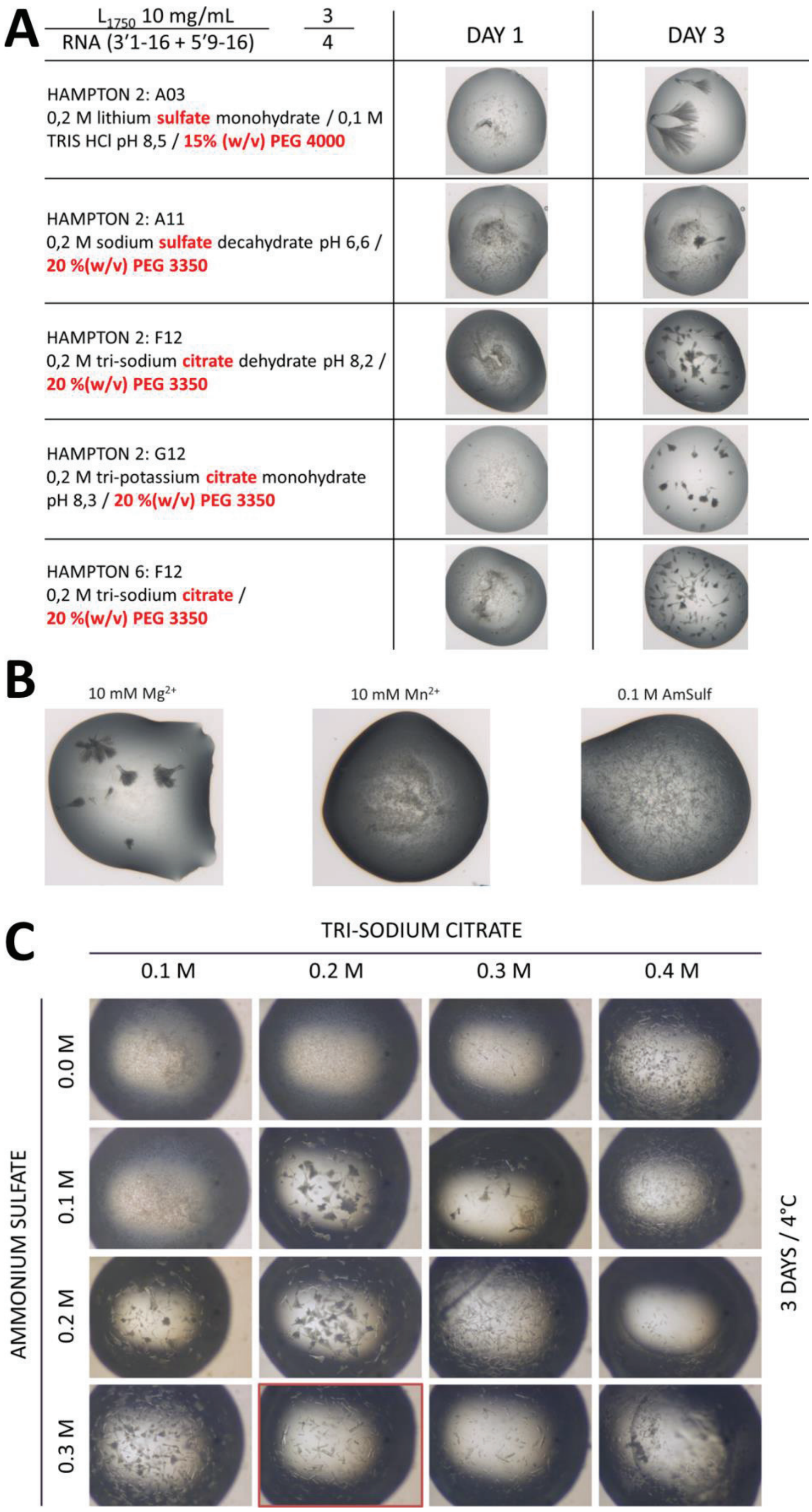



**Figure 3.17** Summary of the $L_{1750}$-vRNA crystallization: A – initial crystals in conditions containing sulfate or citrate salts and PEG 3350 or PEG 4000; B – additive screen; C – citrate and sulfate concentration optimization; red rectangle highlights the best combination.

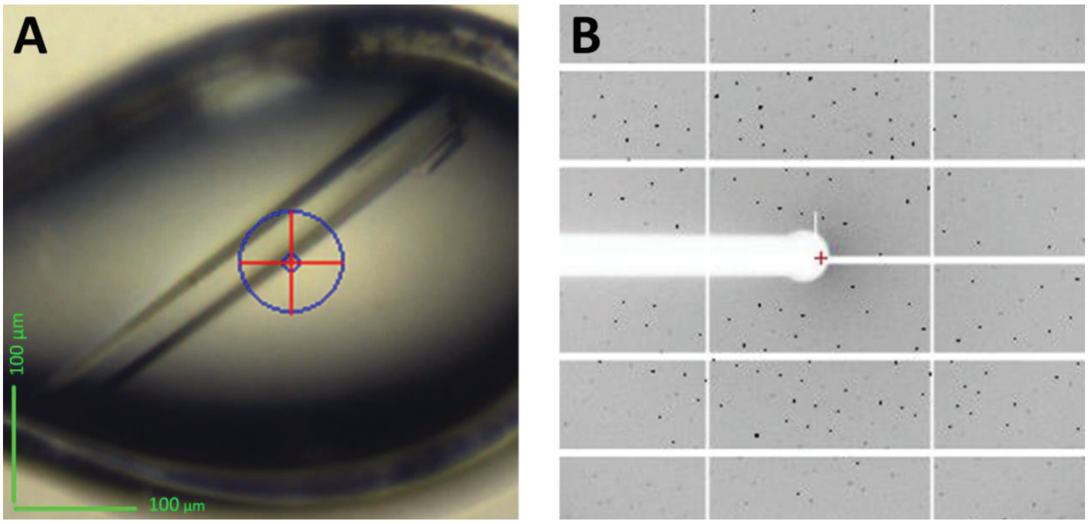



**Figure 3.18** $L_{1750}$-vRNA crystals: A – a single bar-like crystal inside the cryo-loop during diffraction experiment at the beamline; B – example of the diffraction pattern

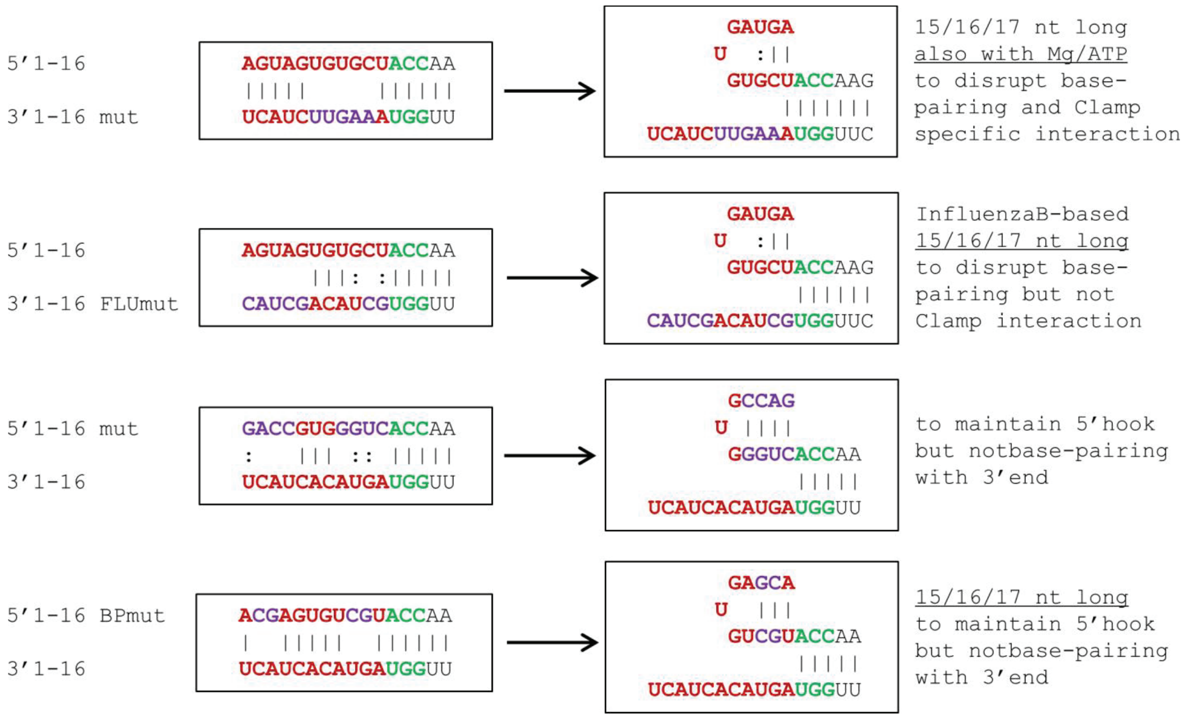



**Figure 3.19** Modified double-stranded RNA used in co-crystallization experiments with $L_{1750}$. In the two top RNA sets the 3′ vRNA end is mutated to disrupt base-pairing with the 5′ vRNA end. In the two bottom RNA sets 5′ vRNA is mutated in order to maintain its secondary hook structure and to disrupt base-pairing with the 3′ vRNA. Mutated nucleotides are marked in violet.

Various sets of modified vRNA ends were used in order to examine whether it is possible to co-crystallize the $L_{1750}$ with both 5′ and 3′ RNA in a different conformation, which would allow accommodation of both vRNA ends during the crystallization process. The main obstacle is that the wild type vRNA ends are fully complementary to each other resulting in formation of the double-stranded RNA, which cannot interact with the polymerase and has the propensity to crystallize on its own (data not shown). To avoid this scenario vRNA ends were mutated following different strategies in order to diminish the double strand formation (Fig. 3.19). Firstly, the 3′ vRNA was mutated to impair the sequence-specific binding within its binding site, and thus to facilitate its positioning inside the template entry tunnel. Secondly, the 5′ vRNA was mutated within its hook base-pairing region in order to maintain the 5′ vRNA hook structure, but to reduce its propensity for annealing with the 3′ vRNA end. The 3′ and 5′ vRNA ends were either mixed together before being added to the $L_{1750}$ sample, or they were added sequentially, with the 5′ vRNA added first so it could form stable hook structure and induce freely all the conformational changes within the L protein. We assumed that this could affect structural protein elements around the template entry tunnel allowing the 3′ vRNA entry into this tunnel. Until now the co-crystallization experiments with these modified RNAs were not successful.

### 3.4.3 ATTACHED ARTICLE

**Structural insights into bunyavirus replication and its regulation by the vRNA promoter**

Piotr Gerlach1,2†, Hélène Malet1,2†, Stephen Cusack1,2* and Juan Reguera1,2*

[1]European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, CS90181, 38042 Grenoble Cedex 9, France.

[2]Unit of Virus Host-Cell Interactions (UMI 3265), Univ. Grenoble Alpes-EMBL-CNRS, 71 Avenue des Martyrs, CS90181, 38042 Grenoble Cedex 9, France.

[†]Joint first author

*Corresponding authors.

Tel. (33)476207238

Email: cusack@embl.fr, jreguera@embl.fr

First version submitted to Cell – 6th February

Accepted, minor corrections required – 27th February 2015

Revised version submitted to Cell – 13th March 2015

Revised version accepted – 20th March 2015

Final acceptance – 7th April 2015

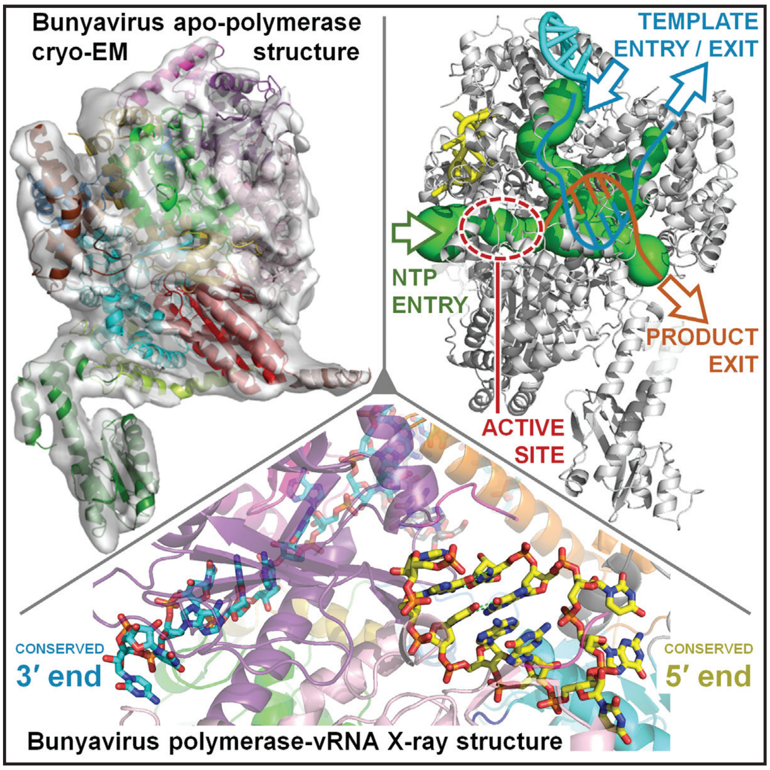
Published online – 21st May 2015

Article

Cell

# Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter

## Graphical Abstract




## Authors

Piotr Gerlach, Hélène Malet, Stephen Cusack, Juan Reguera

## Correspondence

cusack@embl.fr (S.C.),
jreguera@embl.fr (J.R.)

## In Brief

The structure of the monomeric bunyavirus polymerase reveals that divergent segmented negative-strand RNA virus polymerases have a common overall architecture, explains how viral RNA binding allosterically regulates polymerase activity, and suggests a replication model that could apply to all related RNA viruses.

## Highlights

- Bunyavirus polymerase binds its complementary 3′ and 5′ vRNA ends in distinct sites
- 5′ vRNA binding allosterically structures and activates the polymerase catalytic site
- Distinct template/product exit tunnels explain RNA synthesis in a circularized RNP
- Monomeric bunyavirus and trimeric influenza polymerases are structurally similar

## Accession Numbers

5AMR
5AMQ
EMD-2930

Gerlach et al., 2015, Cell *161*, 1–13
June 4, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.cell.2015.05.006

CellPress

Please cite this article in press as: Gerlach et al., Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter, Cell (2015), http://dx.doi.org/10.1016/j.cell.2015.05.006



Cell

# Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter

Piotr Gerlach,[1,2,3] Hélène Malet,[1,2,3] Stephen Cusack,[1,2,*] and Juan Reguera[1,2,*]
[1]European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, CS90181, 38042 Grenoble Cedex 9, France
[2]Unit of Virus Host-Cell Interactions (UMI 3265), University Grenoble Alpes-EMBL-CNRS, 71 Avenue des Martyrs, CS90181, 38042 Grenoble Cedex 9, France
[3]Co-first author
*Correspondence: cusack@embl.fr (S.C.), jreguera@embl.fr (J.R.)
http://dx.doi.org/10.1016/j.cell.2015.05.006


## SUMMARY

**Segmented negative-strand RNA virus (sNSV) polymerases transcribe and replicate the viral RNA (vRNA) within a ribonucleoprotein particle (RNP). We present cryo-EM and X-ray structures of, respectively, apo- and vRNA bound La Crosse orthobunyavirus (LACV) polymerase that give atomic-resolution insight into how such RNPs perform RNA synthesis. The complementary 3′ and 5′ vRNA extremities are sequence specifically bound in separate sites on the polymerase. The 5′ end binds as a stem-loop, allosterically structuring functionally important polymerase active site loops. Identification of distinct template and product exit tunnels allows proposal of a detailed model for template-directed replication with minimal disruption to the circularised RNP. The similar overall architecture and vRNA binding of monomeric LACV to heterotrimeric influenza polymerase, despite high sequence divergence, suggests that all sNSV polymerases have a common evolutionary origin and mechanism of RNA synthesis. These results will aid development of replication inhibitors of diverse, serious human pathogenic viruses.**

## INTRODUCTION

*Arenaviridae*, *Bunyaviridae*, and *Orthomyxoviridae*, the principal families of segmented negative single-stranded RNA viruses (sNSV), each include serious human pathogens such as Lassa fever, Crimean-Congo haemorrhagic fever, and influenza viruses, respectively. *Orthomyxoviruses* have six to eight genome segments, whereas *Bunyaviridae* (reviewed in Elliott, 2014) have three, and *Arenaviridae* two. For each segment, transcription, generating capped viral mRNAs, and replication, generating full-length genome or antigenome copies (vRNA and cRNA, respectively), are performed by the same virally encoded RNA-dependent RNA polymerase (RdRp). For arena- and bunyaviruses, which replicate in the cytoplasm, the RdRp is the single-chain L protein, whereas for orthomyxoviruses, which replicate in the nucleus, it is a hetero-trimeric complex, formed by the PA, PB1, and PB2 subunits (Fodor, 2013; Guu et al., 2012; Morin et al., 2013; te Velthuis, 2014). vRNA genome segments are always packaged by multiple copies of the viral nucleoprotein (NP) together with one copy of the RdRp into filamentous ribonucleoprotein particles (RNPs), which are the functional replication and transcription units (Reguera et al., 2014).

sNSV polymerases have two unique features. First, they perform transcription by the "cap-snatching" mechanism, whereby short 5′ capped RNA fragments are cleaved from host cell mRNA by an endonuclease intrinsic to the RdRp and then used to prime synthesis of viral mRNAs (Morin et al., 2010; Plotch et al., 1981; Reguera et al., 2010; Reich et al., 2014). Second, they recognize each genome segment via their highly conserved, quasi-complementary 3′ and 5′ extremities (over a length of 13–19 nucleotides), known as the promoter (Barr and Wertz, 2004). Correlated with this, sNSV RNPs are generally circularized, which is thought to occur by base pairing between the genome ends (forming a double-stranded "panhandle") and/or the simultaneous binding of both ends to the polymerase (Reguera et al., 2014). vRNA promoter binding to influenza polymerase was visualized recently for the first time in a co-crystal structure. This revealed that each vRNA extremity binds sequence specifically as a single strand to distinct sites on the polymerase but then come together to form a short duplex of about four base pairs (Pflug et al., 2014). Furthermore, this mode of promoter binding is required for activation of diverse influenza polymerase functions (Fodor, 2013). For bunyaviruses, exact self-complementarity of the genome ends extends for 15–19 nts (except for only one G-U mismatch in the case of orthobunyaviruses [Barr and Wertz, 2004; Kohl et al., 2004]), potentially allowing formation of a much more stable panhandle than for influenza vRNA. However, the exact nature of the vRNA-vRNA and vRNA-L interactions that circularize bunyavirus RNPs are not known. Cross linking suggests that the vRNA ends within bunyavirus RNPs are base paired at least partially (Raju and Kolakofsky, 1989) and a distal duplex region is essential for RNA synthesis by bunyavirus (Barr and Wertz, 2004; Kohl et al., 2004) and arenavirus (Kranzusch et al., 2010) polymerases. The absence of significant sequence similarity, outside of the cap-snatching endonuclease (Reguera et al., 2010) and the conserved RdRp motifs (Müller et al., 1994), between *Arenaviridae* and *Bunyaviridae* L

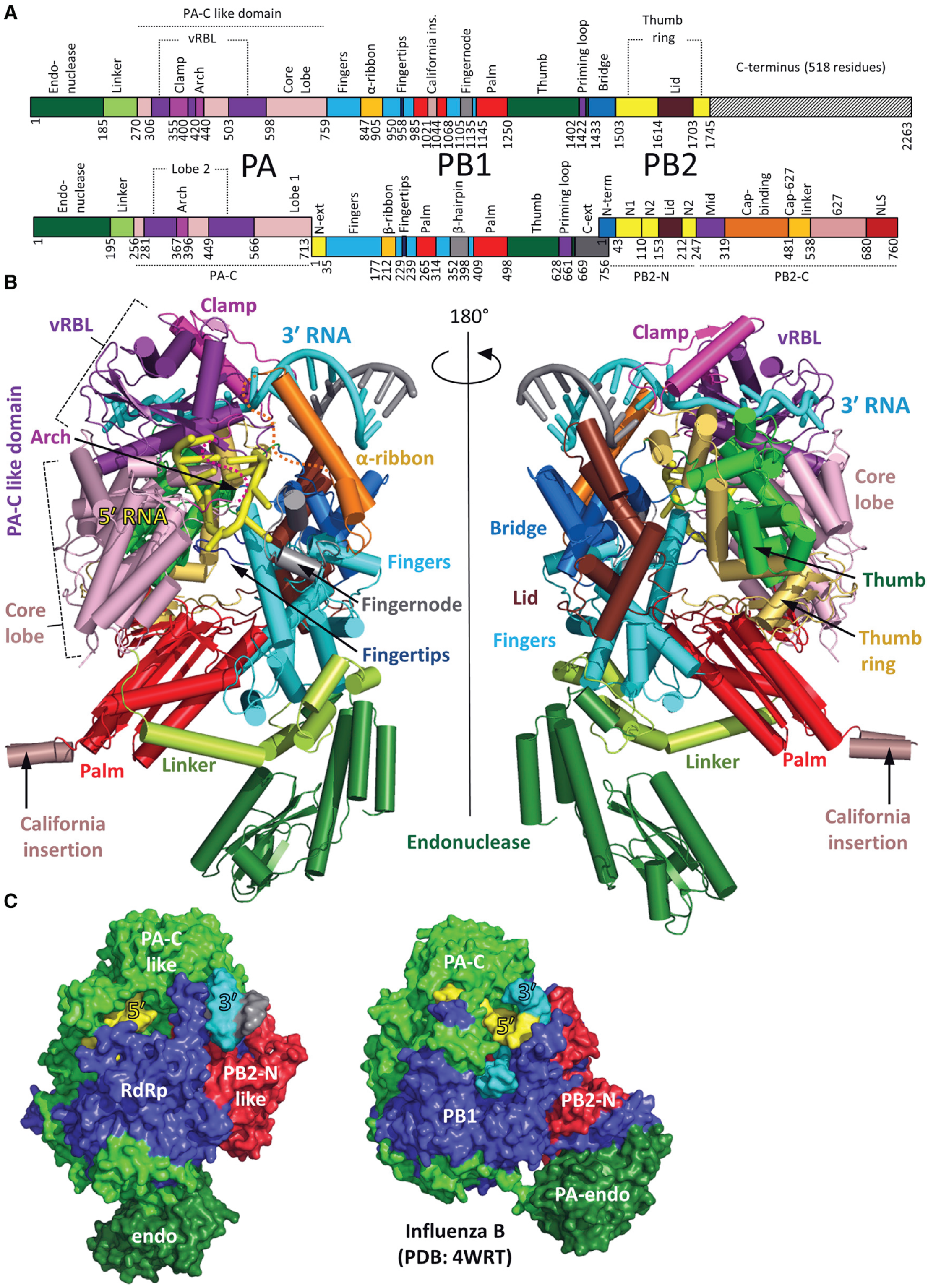



**Figure 1. Overall Structure of LACV Polymerase**

(A) Schematic representation of the domain structure of the monomeric LACV polymerase (top) aligned to that of heterotrimeric (PA-PB1-PB2) influenza polymerase (bottom). Structurally or functionally equivalent domains are similarly colored. A notable difference with the influenza polymerase is the clamp (magenta), involved in 3′ vRNA end binding, which is inserted into the LACV PA-C like domain. The LACV α-ribbon (orange) is structurally equivalent to the influenza β-ribbon

*(legend continued on next page)*

proteins and *Orthomyxoviridae* heterotrimeric polymerases also poses the question as to whether all sNSV have a structurally and evolutionary conserved architecture to match their functional similarity.

To answer this question and those related to promoter binding, we determined the crystal structure of 77% of the L protein from La Crosse orthobunyavirus (LACV) in complex with just the 3′ or both 3′ and 5′ conserved genomic RNA ends and the cryo-EM structure of the apo-form. LACV is a potentially serious but rare, mosquito-transmitted human pathogen that causes 50–100 cases of encephalitis per year in the USA (http://www.cdc.gov/lac/) (Elliott, 2014; Haddow and Odoi, 2009). The structure reveals high similarity, but also interesting differences, to the equivalent influenza complex (Pflug et al., 2014; Reich et al., 2014). In particular, it shows the structural basis of the specific recognition of the vRNA 3′ end, the allosteric regulation mediated by vRNA 5′ end binding, and the likely path of the template into the polymerase active site cavity and out again. These findings, combined with those gained from the influenza polymerase structures, provide new insight into the common mechanism of action, the conserved features, and the diversity among sNSV polymerases.

## RESULTS

### Structure Determination of the LACV $L_{1750}$ Protein

$L_{1750}$, a construct comprising residues 1–1750 (out of 2263) of the LACV L protein, was expressed in insect cells and purified to homogeneity in milligram amounts (Figure S1A, Figures S2A and S2B). In vitro RNA-protein interaction experiments show that the separated single stranded LACV genomic extremities each bind with high affinity and specificity to the polymerase with dissociation constants of 13.8 $\pm$ 2.6 nM and 9.3 $\pm$ 1.6 nM for the 5′ and the 3′ ends, respectively, whereas the polymerase only has low affinity ($\sim$1.5 $\mu$M) for the pre-annealed double stranded panhandle (Figure S3A). However, as shown by mobility shift assays, the polymerase still binds with high affinity to partially double stranded 3′ vRNA provided the first eight nucleotides from the 3′ end are single stranded (Figure S3B). Similar conclusions were previously found for Machupo arenavirus polymerase (Kranzusch et al., 2010). Using these results, co-crystals of $L_{1750}$ diffracting up to 2.6 Å resolution were obtained with nucleotides 1–16 of the genomic 3′ vRNA (3′OH-UCAUCACAUGAUGGUU), to which was annealed a complementary 8-mer (5′OH-GCUACCAA), corresponding to nucleotides 9–16 from the 5′ vRNA extremity. The structure was solved by multiple isomorphous replacement with anomalous scattering (Figures S1B–S1D). Soaking the first 10 nucleotides of the 5′ vRNA (5′p-AGUAGUGUGC) into the crystals gave a new structure at 3.0 Å resolution that revealed the 5′ end in a distinct binding site. Crystallographic statistics are given in Table S1 and a sequence alignment of representative orthobunyavirus polymerases, annotated with the secondary structure, is shown in Data S1. The $L_{1750}$ model contains 1652 residues (94.4% complete) with several connecting loops missing, some of which become ordered upon 5′ vRNA binding (Figure 1, Data S1). The structure of apo-$L_{1750}$ was determined from cryo-EM images by single particle 3D reconstruction at 8.3 Å resolution and allows visualization of secondary structure elements in most of the protein (Figure 2, Figure S2).

### Overall Structure of LACV Polymerase

The RNA-bound and apo-LACV $L_{1750}$ structures display the same overall shape with a large globular central core and a flexible protrusion (Figures 1 and 2). The overall structural organization is strikingly similar to that of the influenza polymerase (Pflug et al., 2014) despite the complete lack of extended sequence homology (Figures 1A and 1C, Figure S4). In fact $L_{1750}$ corresponds precisely to PA, PB1, and PB2-N (residues 1–250 of PB2), confirming the linear, head-to-tail mapping of the influenza heterotrimeric polymerase onto the orthobunyavirus L protein, as previously proposed (Reguera et al., 2010). The central PB1-like RdRp region of $L_{1750}$ (residues 758–1433) contains the canonical fingers, fingertips, palm, and thumb domains with the conserved polymerase motifs exposed into the internal RNA synthesis chamber. It is buttressed on one side by the PA-C like region, which also has distinct pockets for the 3′ and 5′ vRNA extremities, and on the other by the PB2-N like region.

The previously described N-terminal endonuclease domain (residues 1–184) (Reguera et al., 2010) is solvent exposed and differently orientated compared to influenza polymerase (Figure 1C, Figure S4A). However, it is clearly flexibly linked to the central polymerase core, as revealed by its lack of density in one 3D class of the EM map (Figure 2B). An extended linker (residues 185–270), analogous to the influenza PA-linker, that packs on and stabilizes the fingers and palm domains of the RdRp (Figure 1B, Figures S4A and S4B), connects the endonuclease to the PA-C like domain (residues 271–759), which is divided into two lobes (Figures 1A and 1B). The larger "core-lobe" is $\alpha$-helical and buttresses the thumb and palm domains of the RdRp. The second lobe is mainly involved in vRNA promoter interactions and is therefore called the vRNA binding lobe (vRBL). It has a central $\beta$ sheet with, on one side, a structure denoted the "clamp" (residues 355–400) that binds the 3′ of the vRNA (see below) and which has no equivalent in influenza PA. On the opposite side is a long loop (residues 420–440), analogous to the influenza PA-arch, which binds the vRNA 5′ end (Figure 1B,

despite being inserted in a different loop of the fingers domain. The LACV palm domain has an insertion specific for the California serogroup of orthobunyaviruses (salmon). The LACV fingernode (gray) is functionally equivalent to the influenza $\beta$-hairpin. The PB1 C-ext/PB2-Nterm interface is replaced by the LACV bridge domain. The LACV thumb ring domain (yellow) is structurally homologous to the influenza PB2 N1 and N2 domains. $L_{1750}$ lacks the last 518 residues of the L protein currently of unknown structure (black stripes).

(B) Illustrated representation of two views of the crystal structure of $L_{1750}$ in complex with the 3′ (cyan) and 5′ vRNA (yellow). Protein domains are colored as in (A).

(C) Structural comparison between $L_{1750}$ and influenza (FluB2 structure, PDB: 4WRT) polymerases with equivalent PA-like, PB1-like and PB2-N like regions colored green, blue, and red, respectively. The 3′ (cyan) and 5′ vRNA (yellow) vRNAs are indicated. A more detailed structural comparison is in Figure S4.

See also Figures S1 and S4.

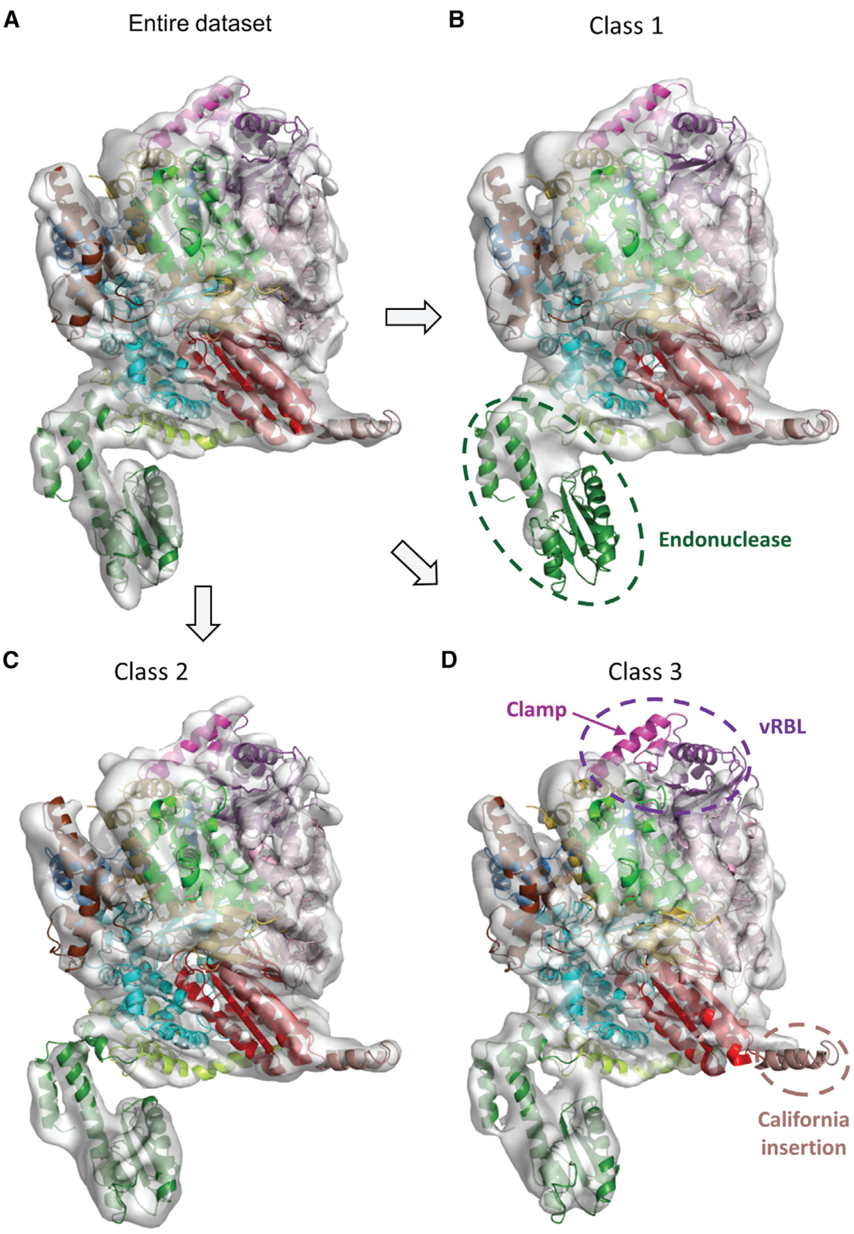



**Figure 2. Cryo-EM Reconstruction of Apo-$L_{1750}$**

(A) 3D reconstruction of the apo-$L_{1750}$ containing the entire dataset of cryo-EM imaged particles, determined at 8.3 Å resolution. The dataset can be separated into three distinct states:

(B) A 3D class displaying only partial density for the endonuclease (9.7 Å resolution).

(C) A 3D class displaying density for all regions of the polymerase (9.7 Å resolution).

(D) A 3D class lacking density for most of the vRBL domain and California insertion (9.3 Å resolution). Flexible regions are indicated with dotted lines. The domains are colored as in Figure 1.

See also Figure S2.

Figure S4). 3D classification of the apo $L_{1750}$ cryo-EM data allows generation of EM maps corresponding to two states of the vRBL. In the first, the same conformation is observed as in the crystal structure, but the arch is invisible, whereas in the second state, the most of the vRBL is invisible, suggesting its enhanced flexibility in the absence of bound vRNA (Figures 2C and 2D). Two specific insertions emerge from the fingers domain, the partially ordered "α-ribbon" (residues 847–905), structurally equivalent to the influenza PB1 β-ribbon (but emerging from a different fingers domain loop, Figure S4B) and the "fingernode" (residues 1105–1135), functionally equivalent to the PB1 β-hairpin. The fingernode folds into two α helices (α43–44) linked by a flexible loop, and together with the vRBL arch, it plays a central role in sequence specific binding of the vRNA 5′ end (see below). The palm domain appears unusually elongated, partly due to the insertion of a solvent exposed helical hairpin of unknown function that is specific for the California orthobunyavirus serogroup including LACV (Figure 1B, Data S1). The thumb domain is surrounded by the PA-C like domain core lobe and a set of helices (α58–62 and α67) and strands (β31–34), denoted the thumb ring which is structurally homologous to the influenza PB2 N1 and N2 domains. A loop (residues 1402 to 1422) at the C-terminal end of the thumb is likely deployed into the polymerase cavity but lacks electron density map, indicating mobility. It is analogous to the influenza PB1 putative priming loop but is significantly shorter. The thumb domain is followed by a helical bundle called the bridge (residues 1433–1503), which replaces the helical PB1-PB2 interface and closes the circular architecture of the polymerase around its internal cavity. The highly conserved connection between the bridge and the thumb ring (residues 1498–1506) partly defines the template entry channel (see below). Inserted in the thumb ring is the lid (residues 1614–1703) which as in influenza PB2, borders the exit channel.

## vRNA Promoter 3′ End Recognition

Nucleotides 1–8 from the vRNA 3′ end are bound in an extended, single-stranded configuration in a narrow cleft over which the clamp closes (Figure 3A). Diverse regions of the polymerase contribute 3′ end RNA binding loops, including both lobes of the PA-C like domain, the thumb (residues 1307–1315) and the thumb ring domain (residues 1513–1517). The protein-RNA interface buries a total of ~3460 $Å^2$ of surface area and includes >30 protein-RNA hydrogen bonds, indicating a high degree of sequence specificity (Figure S5A, Table S2). Nucleotides U1,

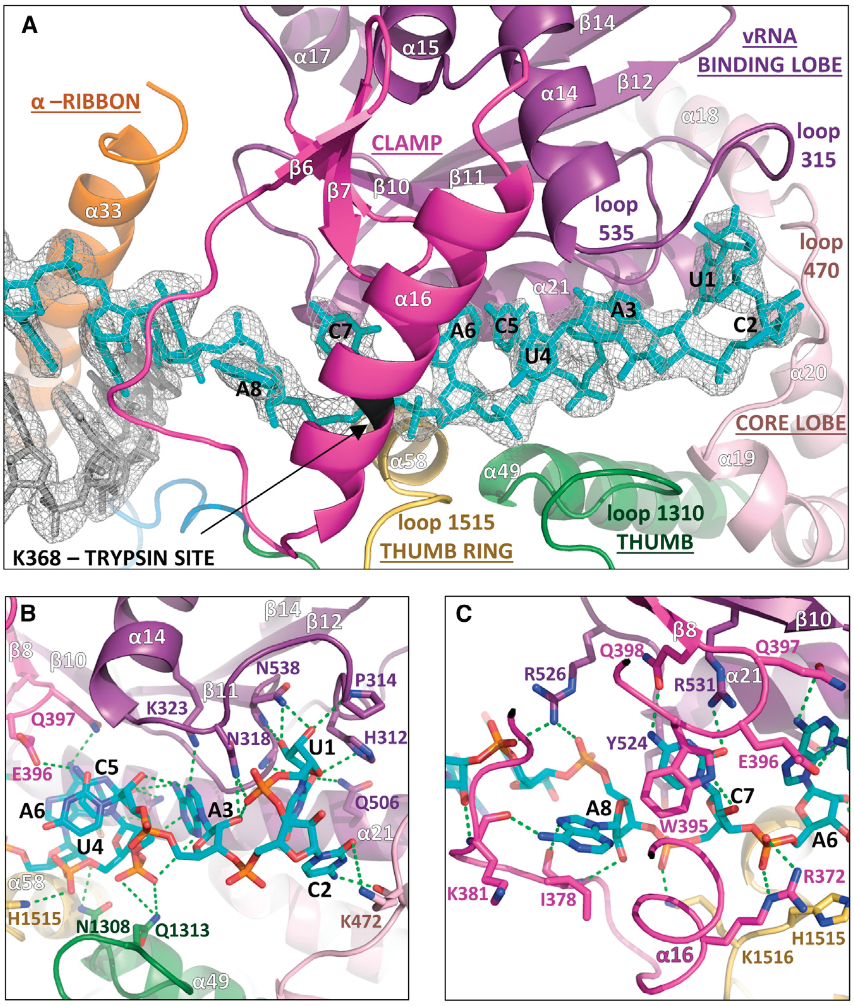



**Figure 3. 3′ vRNA End Binding to LACV Polymerase**

(A) Overview of the 3′ vRNA (cyan sticks) binding site showing the clamp (magenta) and other interacting loops colored as in Figure 1A. The distal short complementary strand is in gray sticks. The RNA electron density is from the final 2Fo-Fc map contoured at 1.5 σ. K368 on helix α16 is protected from trypsin cleavage upon 3′ end binding.

(B) Protein-RNA interactions of nucleotides 1–6 of the 3′ vRNA extremity. Hydrogen bonds are shown as green dotted lines.

(C) Protein-RNA interactions of the clamp with 3′ vRNA nucleotides 6–9.

See also Figures S3 and S5.

C2, and A3 are orientated into individual pockets by an extensive hydrogen bond network with residues from helices α14 and α21 and two loops from the vRBL (residues 312–316 and 535–539) and another loop from the core lobe (residues 469–473) (Figure 3B). The very 3′ end is completely sequestered by the stacking of His312 onto the U1 ribose and hydrogen bonds from the 2′ and 3′ OH to Pro314 and Asn538 (Figure 3B). Nucleotides U4, C5, and A6 are stacked on each other with their bases facing the protein and their phosphates interacting with His1515 from the thumb ring and Arg372 from the clamp (Figures 3B and 3C). Nucleotides C7 and A8 are again orientated into separate pockets mainly formed by clamp residues, with C7 stacking between conserved Trp395 and Tyr524 as well as making base specific hydrogen bonds to residues Gln398 and Arg531. A8 stacks on Ile 378 and its N6 makes a base-specific hydrogen bond with the backbone of Lys381 (Figure 3C). These structural observations are consistent with RNA binding experiments with all possible single substitutions in 3′ end nucleotides 1–11, which show that nucleotides 6–8 are the most critical for sequence specific binding (Figure S3C). 3′ end nucleotides 9-UGAUGGUU-16 form a duplex with the co-crystallized complementary oligonucleotide 5′OH-GCUACCAA (Figure 3A) and the 3′ end backbone is neutralised by basic residues Lys859, Lys862, Arg869, and Lys870 from helix α33 of the α-ribbon (Figure S5A).

## vRNA Promoter 5′ End Recognition and Induced Structural Changes

The 10 first nucleotides of the 5′ vRNA, soaked into the crystal, binds as a stem-loop to the polymerase in a similar configuration and location to the ten nucleotide 5′ hook of influenza vRNA promoter (Pflug et al., 2014) (Figure 4, Figure S6A). However, LACV 5′ hook is less compact with only two base pairs between G2-C10 and U3-G9, compared to four in influenza. Bases G5 and A4 are consecutively stacked on U3, whereas U6, G7, and U8 are splayed out in the loop region, compared with only one in influenza. The 5′ vRNA interaction with the polymerase buries a total of ~3030 Å$^2$ of surface area and includes 37 protein-RNA hydrogen bonds (Figure S5B, Table S2). Upon RNA binding both the arch and the fingernode are structurally reconfigured to promote protein-RNA interactions (Figures S6B and S6C). An array of conserved, mainly positively charged, residues stabilizes the 5′ vRNA backbone phosphates (A1-Lys423, G2-Lys302/Arg592, U3-His306, A4-Arg600/Thr642, G5-Lys643/Tyr677, U6-Arg292, G7-Lys768, U8-His760/His761) (Figure S5B). Nucleotide A1 stacks onto the planar backbone of Cys419 and Gly420 and consecutive base stacking of nucleotides 1–5 is interrupted by conserved arch residue Pro440, which stacks on base G5 forcing base U6 to flip out (Figures 4A and 4B). Highly specific, induced fit interactions are made from residues 1116–1123 of the fingernode loop to flipped out bases G7 and U8. The loop structurally reconfigures to allow G7 to stack on Tyr1120 and make base specific interactions with Gln1116, Asp1123, and Lys768 (α30). U8 stacks on Gln1116 and makes three base specific main-chain interactions with the peptide 1118–1120 (Figure 4C, Figure S6C). Adjacent, conserved His760 and His761 on α30, further stabilize the RNA loop conformation by binding the U8 phosphate and by stacking onto the G9 and G5 ribose moieties, respectively (Figures 4B and 4C).

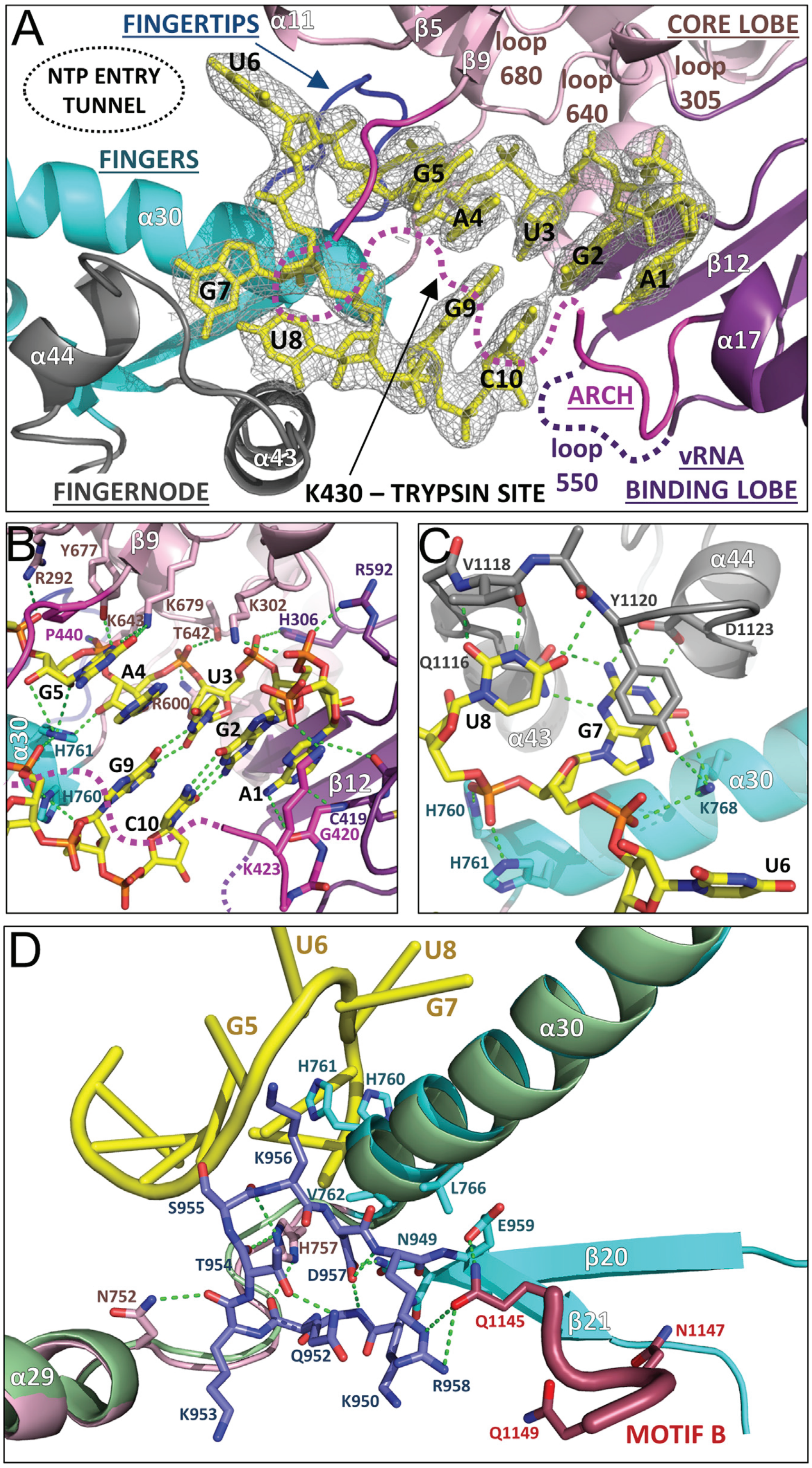



**Figure 4. 5′ vRNA End Binding and Induced Structural Changes**

(A) Overview of the 5′ vRNA stem-loop (yellow sticks) binding site with interacting loops colored as in Figure 1A. The RNA electron density is from the final 2Fo-Fc map contoured at 1.5 σ. K430 on the arch is protected from trypsin cleavage upon 5′ end binding.

(B) Protein-RNA interactions in the 5′ vRNA stem region with hydrogen bonds as green dotted lines.

(C) Protein-RNA interactions of the fingernode loop with 5′ vRNA loop bases G7 and U8.

(D) Superposition of the $L_{1750}$ −3′ vRNA structure without (light green ribbons) and with (colored as in Figure 1A) soaking of nucleotides 1–10 of the 5′ vRNA. Upon 5′ vRNA binding (yellow) the backbone interaction with His760 and His761 pulls helix α30 up allowing stabilization of an ordered configuration of the fingertips residues 949–958 (blue sticks). Multiple new contacts are formed, including hydrophobic interactions with α30 residues V762 and L766 and hydrogen bonds (dashed green lines) with residues from the linker region between PA-C like domain α29 and fingers domain α30, notably His757. Hydrogen bonds between Arg958 and Glu959 to Gln1145 stabilize polymerase active site motif B (dark red).

See also Figures S3, S5 and S6.

Biochemical evidence for the involvement of the clamp and arch in, respectively, 3′ and 5′ end binding comes from proteolysis experiments that show clamp residue Lys368 and arch residue Lys430 are protected from trypsination by 3′ and 5′ end binding, respectively (Figures 3A and 4A, Figure S3D). However, 5′ vRNA binding not only induces conformational changes in RNA binding loops but also structures elements of the polymerase active site (Figure 4D). Most significant is the complete ordering of the fingertips loop (residues 950–958), which contains motif F. The interactions of His760 and His761 pull helix α30 toward the 5′ vRNA, and the consequent displacements of Val762 and Leu766 make room for the fingertips to order into a structured active form through multiple contacts with residues from the PA-C like domain α29, fingers domain α30 and the linker joining them (Figure 4D). Conserved His757 plays a key role in stabilizing the fingertips β-turn through multiple hydrogen bonds to backbone carbonyl groups (Figure 4D). Fingertips ordering has a knock-on effect on stabilization of motif B notably through the interaction of conserved motif B Gln1145 with Arg958 and Glu959 (Figure 4D). This is the first observation of functionally important allosteric effects associated with 5′ vRNA binding and is only observed when 10–11 5′ end nucleotides are bound, but not eight (data not shown). Interestingly, soaking in 5′ cRNA nucleotides 1–10

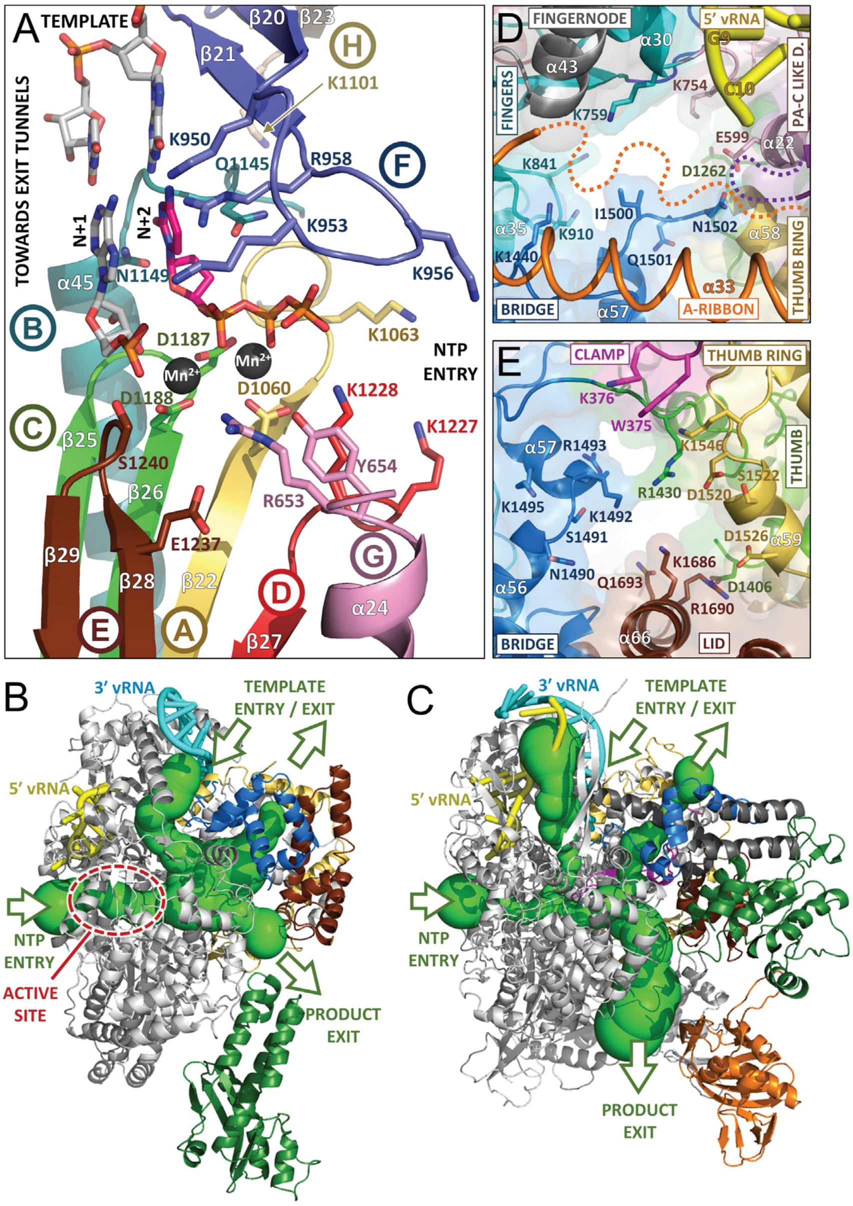



**Figure 5. The LACV Polymerase Active Site and Entrance and Exit Tunnels.**

(A) The arrangement of the conserved RdRp motifs in the LACV active site colored gold, light blue, green, red, brown, and blue for motifs A–F, respectively. Additional sNSV specific motifs G (from the PA-C like domain) and H are shown in pink and gray (see Figure S7). Superposition of the polio virus elongation complex structure (PDB: 3OLB, 3OL8) shows the positions of the catalytic divalent cations (black spheres), the priming nucleotide (N+1, gray) and incoming NTP (N+2, magenta) and template strand (light gray sticks).

(B) The LACV polymerase structure (gray cartoon) with the 5′ and 3′ vRNA in, respectively, yellow and cyan is shown with the tunnels (green) marked with arrows as template entry, NTP entry, product, and template exit, as calculated with MOLE 2.0 (Sehnal et al., 2013). The endonuclease, bridge, thumb-ring and lid are, respectively, in forest green, blue, gold, and brown.

(C) The same representation and orientation as (B) for the influenza A polymerase structure (PDB: 4WSB) with additionally the PB2 cap-binding domain in orange, the putative priming loop in magenta and the PB1 C-extension in dark gray.

(D) Diagram showing the conserved residues forming the template entrance in LACV polymerase which is partially occluded by the flexible α-ribbon (orange). Colors are as in Figure 1A.

(E) As (D) but showing the putative template exit channel in LACV polymerase.

See also Figure S7.

(which differ from vRNA only in the substitution G9 to A) shows that the wobble base pair becomes canonical U3-A9 with no discernible difference on the induced polymerase rearrangements compared to vRNA (data not shown). This is consistent with only the identity of 3′ position 9 leading to a significant difference in propensity for transcription between vRNA and cRNA (Barr and Wertz, 2005). In the 5′ vRNA end 1–11 soaked structure, nucleotide U11 is only partially ordered and cannot base pair with A1 without displacement of conserved Arg595 which hydrogen bonds to N1 of A1 (Figure S6D). Furthermore soaking 5′ RNAs longer than 11 nucleotides resulted in no binding in the crystal, suggesting that further rearrangements of the polymerase (notably the α-ribbon) and probably a sharp turn in the RNA (as observed in influenza 5′ end between nucleotides 10–11) may be required to allow binding of the complete 5′ end, but which are incompatible with the crystal packing (Figure S6D).

## The Active Site Cavity and Its Entrance and Exit Tunnels

The L protein internal active site chamber, where nucleotide addition occurs, is formed by the conserved polymerase motifs A-F and configured like other RNA polymerases, notably influenza polymerase(Pflug et al., 2014; te Velthuis, 2014) (Figure 5A). Motif F forms part of the fingertips, the flexible loop between fingers strands β20 and β21 that is only fully ordered when the 5′ vRNA is bound (see above). The other conserved polymerase motifs are all in the palm domain: motif A (1060-DMSKWS) between palm β22 and α41 with divalent cation binding D1060; motif B (1145-QGNFNYTSSY) between β24 and the long α45 with conserved N1149; the catalytic motif C (1186-SDD) in the turn between strands β25 and β26; motif D (1223-QANMKKTY) just before β27 and motif E (1236-KEFVSLFN) forming the tight loop between β28 and β29. Interestingly, structural alignment with influenza polymerase allows identification of two new active site motifs, denoted G and H, which appear to be conserved specifically in sNSV polymerases (Figure 5A, Figures S7A and

S7B). Motif G (653-**R**YMI in LACV, 658-**R**KLL in influenza PA) is in helix α24 in the core-lobe of the PA-C like domain (Figure S7A), the conserved arginine being positioned to interact with the priming NTP (Figure 5A). Motif H (1101-**K**ELIL in LACV and 347-**K**VARL in influenza PB1) forms a β strand (β23 in LACV) with the conserved lysine stabilizing the motif B backbone conformation by hydrogen bonding to multiple carbonyl-oxygens (Figure S7B).

Four positively charged solvent accessible tunnels, visualized using MOLE 2.0 (Sehnal et al., 2013) converge into a central inner cavity where the RdRp motifs mediate template directed RNA synthesis (Figure 5B, Figure S7C). The tunnels are delimited by residues conserved among Orthobunyavirus polymerases (Figure S7D). The configuration of the template entry channel, the NTP entry channel and the nascent strand exit channel is similar to that described for influenza polymerase (Figure 5C) (Pflug et al., 2014). The template channel entrance is defined by the vRBL, fingers and bridge and is partially obscured by the α-ribbon (Figure 5D), which together with several loops of the vRBL β sheet that are deployed toward the entrance but disordered in the structure, may modulate access. The NTP entry channel is lined by conserved basic residues R287, K673 (PA-C like), K956, R958 (fingertips), K1063 (motif A), K1227, and K1228 (motif D), some of which are only positioned correctly upon 5′ vRNA binding (Figure S7E). The product strand exit tunnel is surrounded by the lid domain and the thumb ring mainly by the extended joining linkers and by fingers and palm opposite side of the NTP entry channel (Figure S7F). In a LACV L there is a more obvious extra channel that we postulate is for the template to exit. In influenza polymerase the equivalent channel is present, but narrowed by the presence of the putative priming loop (Figure 5C), which is 15 residues longer than in LACV L. The putative template exit channel is defined by the thumb, thumb ring, bridge, and lid domains and lined by conserved basic residues R1430, K1492, R1493, K1686, and R1690 (Figure 5E). As discussed below, the arrangement of the tunnels in LACV L protein suggests an elegant strategy for RNA synthesis whereby the polymerase forces separation of the template and product strands and directs each down distinct exit channels on opposite sides of the molecule.

## DISCUSSION

### Initiation of RNA Synthesis Requires Significant Conformational Changes in Protein and vRNA

In the $L_{1750}$-vRNA complex, the 3′ end of the template is tightly and specifically bound but not accessing the entry tunnel, corresponding perhaps to the highly stable state that occurs in inactive RNPs, for example within virions. For active RNA synthesis, the 3′ end clearly has to be relocated into the polymerase active site (Figures 6A and 6B). Furthermore it has been shown that complementarity and presumably base pairing between at least 3′ and 5′ nucleotides 12 to 16 are required for RNA synthesis by orthobunyavirus polymerases (Barr and Wertz, 2004; Kohl et al., 2004). A pre-initiation configuration, with the 3′ end specifically bound on the outside but not entering the polymerase active site, was also visualized for influenza polymerase (Pflug et al., 2014; Reich et al., 2014). However, whereas in the influenza structures, both 3′ and 5′ extremities of the promoter simultaneously bind their separate single-stranded binding sites and form a distal duplex region (with 11–14 of 5′ base pairing with 10–13 of the 3′ end), for LACV polymerase in the crystallized conformation, this appears to be impossible. First, nucleotides 11 of the 3′ and 5′ ends are around 25 Å apart, too far to see how 3′ and 5′ nucleotides 12 to 16 could base pair (Figure 6A). Second, in the all $L_{1750}$ structures, bases 9–16 of the 3′ end are already base paired with the co-crystallized cRNA, corresponding in sequence to 5′ nucleotides 9–16. Thus when 5′ nucleotides 1–11 are soaked into the crystal, nucleotides 9–11 are present in two distinct locations greater than 20 Å apart, which obviously cannot happen physiologically (Figure 6A). Given the similar mode of 5′ end binding to both LACV and influenza polymerases (and that the 5′ end remains bound as observed during initiation; Reich et al., 2014) and the similar requirement of a short distal 3′–5′ duplex, we propose that an alternative configuration of the bound promoter likely exists more analogous to the influenza pre-initiation conformation (Figure 6B). Release and repositioning of the 3′ end could occur by swinging of the clamp, without necessarily letting go of the 3′ end RNA, with a concomitant reorientation of the α-ribbon into a position analogous to the β-ribbon in influenza polymerase (Reich et al., 2014), to stabilize duplex formation (Figure 6B, Figure S6D). The flexibility of the vRBL and α-ribbon as seen by 3D classification of the EM images, together with the lower local resolution of these regions in the cryo-EM map, shows that such movements are plausible (Figure 2, Figure S2F). In the case of influenza virus, extrapolation of the template from the observed duplex region, based on the poliovirus polymerase elongation complex model (Gong and Peersen, 2010), would result in the 3′ end overshooting the polymerase active site by three nucleotides as previously discussed (Reich et al., 2014). For LACV, similar modeling suggests that this overshoot is accentuated. Assuming that for LACV the duplex is from nucleotides 12–15 of both strands (Barr and Wertz, 2004; Kohl et al., 2004), based on the influenza/polio models, the template would overshoot the polymerase active site by 5 or 4 nucleotides, depending on whether the LACV 3′–5′ 12–12 base pair corresponds to influenza 3′-5′ 10–11 or 11–12 base pair, respectively (Figure 6B). In the case of cap-dependent transcription, this overhang could favor base pairing with the incoming capped primer. In the case of replication, this situation could be explained by (1) a different, less direct path of the single stranded template, so that the 3′ nucleotide 1 was placed directly in the polymerase active site (i.e., the modeling is misleading); (2) a mechanism of internal initiation followed by realignment, dependent on the triplet repeat at the beginning of the template (3′-UCAUCA), as has been described for some bunya- and arena-viruses (Guu et al., 2012) (i.e., internal initiation at position 4 followed by realignment of the AGU triplet); or (3) initiation starting at nucleotide 1 but duplex formation between 3′ and 5′ nucleotides 12 to 16 only occurring after 4–5 nucleotides have been synthesized and the template has translocated further into the active site cavity. This latter possibility would be consistent with the observed position of LACV 3′ end nucleotide 8 being close to that of the nucleotide 8 counting from the active site along the polio template (Figure 6B).

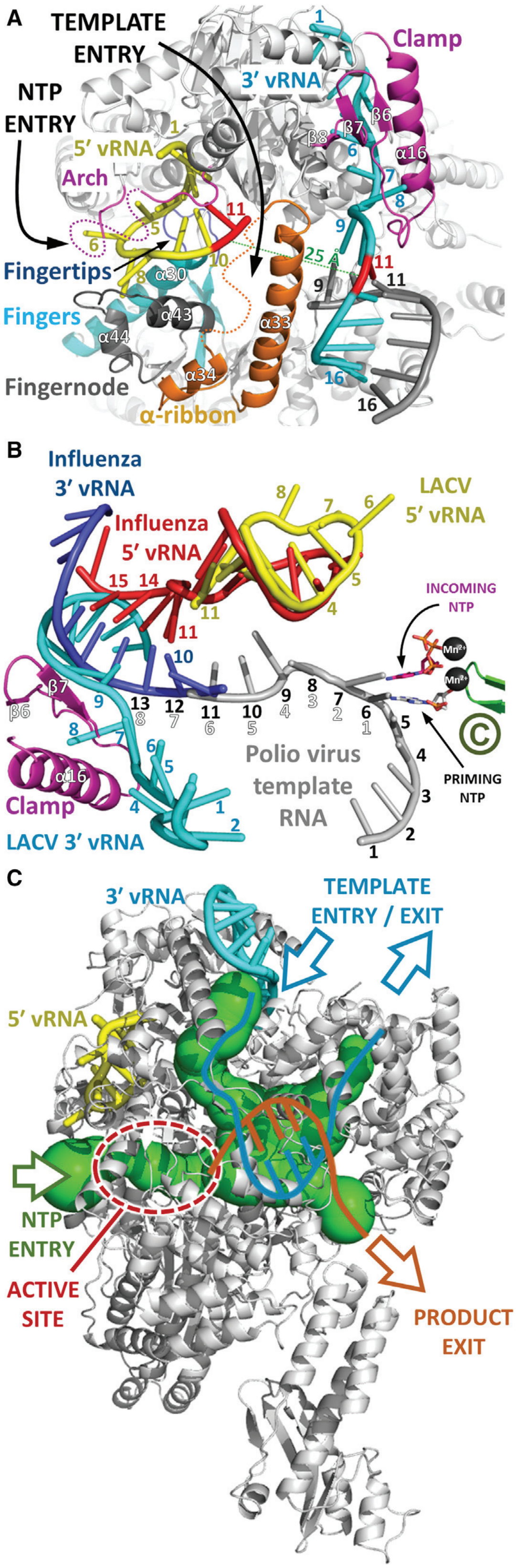



**Figure 6. Model of RNA Synthesis by LACV Polymerase**

(A) Illustrated representation of the LACV polymerase (gray) looking down the template entry channel showing the disposition of key structural elements (arch, clamp, α-ribbon, fingertips, fingernode) colored as in Figure 1A. The 5′ and 3′ vRNA extremities are, respectively, yellow and cyan tubes, except that nucleotide 11 in each case is in red highlighting their wide separation (>20 Å). The figure shows the impossibility of formation of a distal 5′ and 3′ duplex between nucleotides 12–15 of each strand, while maintaining the single-stranded ends bound as in the observed conformation.

(B) Model for the initiation conformation of LACV based on superposition with the influenza polymerase (PDB: 4WSB) and the poliovirus elongation complex (PDB: 3OLB, 3OL8) structures. The observed 5′ and 3′ vRNAs are, respectively, red and blue for influenza and yellow and cyan for LACV and numbered accordingly. The LACV clamp binding to the 3′ end is in magenta. The poliovirus template strand is in gray and the active site is indicated by motif C (green), the catalytic divalent cations (black) and the priming and incoming NTPs (gray and magenta, respectively). The influenza vRNA distal duplex starts with the 3′-5′ 10:11 base pair (labeled). The template nucleotide numbering in outline white counts back from the active site, assuming initiation at position 1. The template nucleotide numbering in black numbers counts along the LACV template assuming the first LACV 3′-5′ base pair 12:12 aligns with the influenza 10:11 base pair. This would allow for connectivity between the distal LACV duplex and the observed 5′ end hook binding but imply an overshoot of the active site by 5 nucleotides. This is discussed further in the text.

(C) Model of the elongation state showing trajectories of template RNA (cyan) and product RNA (orange) and NTPs through the polymerase tunnels (green). The observed positions of the 3′ and 5′ ends are shown as well as the position of the active site. After a short template-product duplex, which is accommodated in the interior cavity, each strand exits separately along different tunnels, the template back to the front of the polymerase where it can re-integrate into the RNP and the nascent strand to the rear where product processing occurs i.e., progeny cRNP assembly in the case of replication or mRNP assembly or translation coupling in the case of transcription.

## The Template Pathway through the Polymerase

Whatever mode of initiation, during elongation, a duplex formed by template and nascent strands starts to grow in the active site cavity as visualized in the structure of the poliovirus polymerase elongation complex (Gong and Peersen, 2010) (PDB: 3OL7) (Figure 6C). However, after a complete double-helical turn, the nascent and emerging template strands would appear to clash with the thumb and C-terminal lid domains, respectively, as described for the influenza case (Reich et al., 2014). To resolve this situation, we propose that the strands are forced to separate and each is extruded along a separate exit tunnel (Figure 6C). The nascent strand would exit to the rear of the polymerase and, by analogy to the influenza case, the cap-snatched transcription primer would probably enter by the same route (Reich et al., 2014). The template would exit to the front of polymerase through the template exit channel on the same side and not far from the template entrance (Figure 6C). The proximity of the entry and exit channels would allow reintegration of the vRNA genome template into the RNP with minimal disruption (Figure 7). Modeling shows that the template path through the entry channel, cavity, and the exit channel would accommodate around 20 nucleotides. Given that ∼11 nucleotides of the 5′ and ∼9 of the 3′ ends directly bind the polymerase in the pre-initiation state, and also a single LACV NP can accommodate 11 nucleotides (Reguera et al., 2013), no free RNA needs to be exposed, nor NPs removed or added, during template reading. The mobile clamp together with the α-ribbon

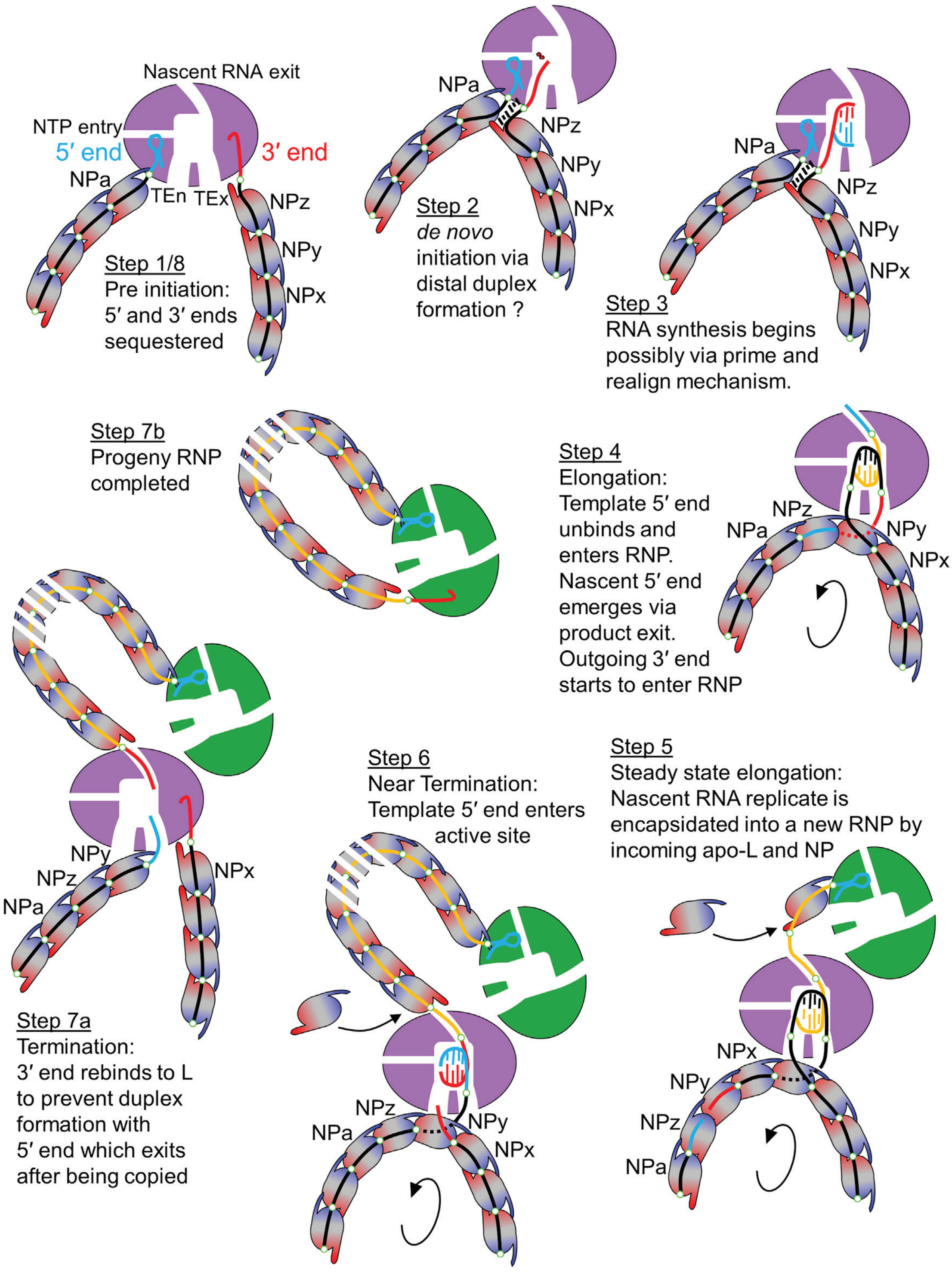



**Figure 7. Schematic Model of vRNA Replication**

An LACV RNP is schematically represented with the polymerase (purple or green), with template entrance (TEn), template exit (TEx), NTP entry and nascent RNA exit channels as marked, interacting with the viral RNA (black or yellow) and proximal NPs (ellipses colored with a blue-to-red gradient). The complementary 5′ and 3′ vRNA ends are, respectively, cyan and red. The NPs form a chain linked together by flexible NP-NP interactions involving the N-terminal arm (blue) and the C-terminal arm (red) and each NP sequesters 11 nucleotides RNA (Reguera et al., 2014; Reguera et al., 2013). Small circles mark consecutive 11 nucleotide segments of the vRNA. The polymerase itself can sequester around 20–22 template nucleotides.

*(legend continued on next page)*

and proximal NPs could mediate the RNA template translocation driven by the polymerase motor.

## CONCLUSION

Our structures of the LACV L protein in the apo-state and with one or both vRNA ends bound suggest that assembly of a functional initiation complex is a multistep process. The structures clearly show that the polymerase has highly specific and distinct sites for the single-stranded 3′ and 5′ vRNA ends, preventing them from forming an extended panhandle. The mode of 5′ end hook binding is similar to that observed for influenza polymerase, but here we directly observe the associated allosteric effects that are essential for structuring critical active site loops. On the other hand the 3′ end appears to be preferentially and tightly bound in a groove closed by a clamp on the side of the polymerase. Elucidation of the exact purpose of this binding site and the mechanism for 3′ end relocation into the template tunnel for the initiation of RNA synthesis are clear questions for future studies. Furthermore, the extended complementarity of the LACV 3′ and 5′ ends appears to be a major obstacle to reconstitute the LACV initiation complex in vitro, since incubation, even sequentially, of the complete 3′ and 5′ ends leads to the preferred formation of a long, stable duplex, which has low affinity for the polymerase. Probably for the same reason, we have not yet been able to demonstrate robust template directed RNA synthesis activity for either $L_{1750}$ or full-length polymerase (see Supplemental Experimental Procedures). It is therefore likely that in the case of bunya-, and probably, arenavirus polymerases (Kranzusch et al., 2010), to avoid stable base pairing of the highly complementary vRNA promoter, the free ends are prevented from ever meeting each other by the sequential mode of assembly of nascent RNPs starting at one end of a growing replicate (Figure 7). Thus it is plausible that, unlike the situation for recombinant influenza polymerase, which is fully active when reconstituted only with the vRNA promoter (Reich et al., 2014), NP-L and NP-RNA interactions may be required for bunyavirus polymerase activity.

The striking structural similarities between the single-chain LACV and heterotrimeric influenza polymerases strongly support the idea of an evolutionary common ancestor. Indeed, it now seems plausible that all sNSV polymerases (i.e., from arena-, bunya-, and orthomyxovirus families) have a similar architecture, despite very low overall sequence homology, and this is supported by structure-based identification of new common motifs (Figures S7A and S7B). However, this does not mean that these polymerases will not have idiosyncratic family and sub-family differences. For example, whereas the arch and fingernode that bind the 5′ hook have structural and functional homologs in influenza, the LACV 3′ end binding site and the clamp structure that pins it in place has no such equivalent. Similarly, it is reasonable to suppose that each polymerase is adapted to its cognate nucleoprotein, whose size, structure, mode of RNA binding, and number of nucleotides bound (e.g., 11 for orthobunyaviruses, 7 for phleboviruses), are very different for each sNSV family (Reguera et al., 2014). In this context, it is intriguing that LACV and influenza contain, respectively, an α- or β-ribbon, equivalently located extended and flexibly hinged structures that could both play a role in both RNA and NP interactions and could be adapted to the respective NP structures (Figure S6D) (Reich et al., 2014). Finally, it is highly significant that the $L_{1750}$ construct ends precisely at the same position as separates PB2-N and PB2-C in influenza polymerase (Figure 1). PB2-C, which includes the cap-binding domain and the C-terminal nuclear localization motif (not relevant for cytoplasmic LACV), has already been shown to be loosely associated with the rest of influenza polymerase (Reich et al., 2014). It remains open as to how much the C-terminal residues 1751–2263 of LACV L, missing in $L_{1750}$, are structurally homologous to PB2-C. In particular it is still unknown whether there is a cap-binding domain, for which there is no direct evidence yet, but, in the case of Lassa arenavirus a specific requirement for C-terminal residues for mRNA transcription has been established (Lehmann et al., 2014).

(1) In the inactive state, whether after vRNP assembly or in virions, both ends of the genomic RNA are sequestered into the specific 5′ and 3′ RNA binding sites of the polymerase, thus circularizing the RNP.

(2) For de novo RNA synthesis or cap-dependent transcription (not shown) the 3′ end is relocated into the polymerase active site for initiation, by an unknown mechanism. Distal 3′-5′ duplex formation may occur before or after initiation depending on whether initiation is internal (followed by prime and align) or at position 1 (see Figure 6B and main text). Duplex formation could bring the NPs at the 5′ and 3′ (NPa and NPz) closer enhancing the circularization of the NP scaffold but would need to be dissociated to proceed with elongation.

(3) With the 5′ end bound to the allosteric site for the activation of the RNA synthesis, a nascent cRNA begins to be synthesized.

(4) As elongation proceeds, the template dissociates from the proximal NP and is channelled into the active site. Because of the proximity of the entrance and exit channels the disruption of the RNA-NP assembly may only affect one NP. Early on, the 5′ end is detached from its specific binding site on the polymerase and enters the RNP by loading onto NPz. As incoming template is released from NPy on one side, the outgoing 3′ end is loaded on it from the other side. More generally, the RNA being pulled into the cavity by the polymerase motor detaches from the proximal NP which is pulled to the left thus pushing the NP-RNA array in the direction of the arrow. This model would imply that 5′ end binding is only required to activate initiation. This would be a difference from the influenza situation where the maintenance of 5′ end binding is required, at least during transcription, for self-polyadenylation to occur.

(5) Once the nascent c5′ end emerges from the exit channel it can recruit an incoming apo-polymerase as the first step in encapsidating the progeny cRNP with incoming apo-NPs. This may be facilitated by polymerase dimer formation (see main text).

(6) Approaching termination the template 5′ end would be copied and the template 3′ end (bound to NPy) would approach its starting point.

(7a) At termination the template 3′ end rebinds to its specific binding site on the polymerase to avoid base pairing with the emerging template 5′ end which subsequently rebinds to its polymerase binding site, thus completing the replication cycle.

(7b) Due to polymerase dimer formation, the nascent c3′ end, which emerges last from the product exit channel, can easily find and bind to specific 3′ binding site on the green polymerase, thus completing progeny cRNP formation. Without polymerase dimer formation being maintained throughout replication (or other mechanism for keeping the polymerases in close proximity), it is unclear how the c3′ could find and bind to the correct polymerase which may have diffused far away.

Based on the $L_{1750}$ structure, we propose a model for RNA replication in which there are clearly separated exit tunnels for the single-stranded template and product (Figure 6C, Figure 7). The proximity of the template entrance and exit, on one side of the polymerase, is compatible with processive template reading with minimal disruption of the RNP. Template RNA would progressively dissociate from proximal NP as it translocates through the entrance tunnel, into the polymerase internal chamber and out again, to be reincorporated into the RNA free NP that concomitantly translocates round the outside of the polymerase between the entrance and exit tunnels, held together by flexible NP-NP interactions (Reguera et al., 2013) (Figure 7). Meanwhile, the products exit to the other side of the polymerase, thus allowing spatial separation of template translocation and product processing. In the case of (anti)-genome replicates, product processing involves assembly into progeny RNPs, possibly first with an incoming apo-polymerase binding specifically the emerging, nascent 5′ end and subsequently progressive packaging by incoming NPs (Figure 7). We propose that robust polymerase dimerization is necessary to ensure efficient and correct circularisation of progeny RNP (Figure 7), and this is consistent with some observations concerning replication by influenza polymerase (Jorba et al., 2008; York et al., 2013), but other influenza replication models involve more complicated higher order polymerization (Chang et al., 2015; Jorba et al., 2009). In the case of transcription, as suggested by the structure of influenza polymerase, the cap-snatched primer would enter the internal cavity via the product exit channel and then, upon elongation, extrude out in the same direction, where it likely interacts with host translation factors (in the case of bunyaviruses, transcription is closely coupled to translation; Barr, 2007). As suggested for influenza polymerase (Reich et al., 2014), the main role of the PB2-C like region may be in these product processing mechanisms. We think these concepts are likely to be applicable to all sNSVs polymerase and possibly those of nsNSV as well, which also operate in an RNP context.

## EXPERIMENTAL PROCEDURES

### Protein Production, Crystallization, and Structure Determination

Residues 1–1750 ($L_{1750}$) of the polymerase (L protein) sequence of La Crosse virus (LACV) were expressed in insect cells from a synthetic gene inserted in a pFastBac vector. Purified protein at 5 mg/ml was crystallized with an equimolar mixture of nucleotides 1–16 from the 3′ and nucleotides 9–16 from the 5′ vRNA ends (Dharmacon). Crystals were improved by microseeding. Diffraction data were collected on beamlines ID23-1 or ID29 at the European Synchrotron Radiation Facility and integrated with XDS (Kabsch, 2010). The structure was solved by the multiple isomorphous replacement with anomalous signal method using selenium, platinum, and tantalum cluster derivatives with SHARP (Bricogne et al., 2003). LACV endonuclease (PDB: 2XI5) and influenza polymerase PA and PB1 subunits (PDB: 4WSB) were used as a guide to model building. Selenomethionine positions from the anomalous difference map helped align the sequence and autobuilding with BUCCANEER (Cowtan, 2006) was useful to extend the model, which was refined with REFMAC (Murshudov et al., 1997) and PHENIX (Afonine et al., 2012). 5′ end nucleotides 1–10 or 1–11 were soaked into pre-grown crystals to reveal the 5′ end binding site.

### Electron Microscopy

Cryo-EM grids were prepared by applying 4 μl of $L_{1750}$ at 0.2 mg/ml to a quantifoil grid, blotting excess solution and then freezing in liquid ethane. Cryo-EM images of apo-$L_{1750}$ were collected on a Krios microscope at 80 kV with a Falcon II direct detector (FEI) at magnification 138,129 times. FEI EPU automatisation software was used to collect 6,129 micrographs with a defocus between 0.5 and 2 μm, an exposure time of 0.5 s and a dose of 14e$^{-}$/Å$^{2}$. After contrast transfer function correction, 10,102 manually picked particles were used to derive initial class averages with IMAGIC (van Heel et al., 1996). These were then used to select ~180,000 particles for input to 3D reconstruction and refinement with RELION 1.3 (Scheres, 2012), using as initial model the $L_{1750}$-vRNA crystal structure filtered at 30 Å resolution, leading to a map at 8.3 Å resolution. The dataset was subsequently partitioned by 3D classification resulting into three structures which revealed the less well-defined, flexible regions.

### Polymerase-vRNA Binding Studies

For electrophoretic mobility shift assays, radioactively labeled RNAs were produced by in vitro transcription with T7 polymerase. For binding assays, 10 μM of $L_{1750}$ in 10 μl buffer was mixed with radiolabelled RNA and 1 μl of non-specific poly(U) RNA (Sigma). Mixtures were incubated at room temperature for several hours and resolved on native TG gels. Radioactive signal from shifted bands was recorded with a Typhoon and quantified with ImageQuant.

For fluorescence anisotropy measurements, 5 nM 25-nucleotide long RNA oligos corresponding to 3′ or 5′ vRNA (IBA), labeled with fluorescein on the appropriate non-interacting end, were titrated with $L_{1750}$ in order to obtain 10–15 protein concentration points ranging from 3 nM to 1 μM. Fluorescence and fluorescence anisotropy measurements used 495 nm excitation and 515 nm emission wavelengths. KaleidaGraph (Synergy Software) was used to evaluate the data and derive dissociation constants.

For proteolysis protection experiments, $L_{1750}$ was incubated at 1 mg/ml with 1:1 molar ratios of 25-nucleotide long 3′ or 5′ genomic ends and then digested for 1 hr at room temperature with trypsin (1:1,000 w/w). Products of digestion were analyzed by various techniques including: SDS-PAGE, western-blot, ESI-TOF-MS, MALDI-TOF-MS, MALDI-TOF-MS with N-terminal acetylation, and N-terminal sequencing of protein fragments by Edman degradation.

For more details see Supplemental Information.

## ACCESSION NUMBERS

The accession number for the 8.3 Å cryo-EM map of apo-$LACV_{1750}$ reported in this paper is EMDB: EMD-2930. The accession numbers for the $L_{1750}$-3′ end and L1750-3′/5′ ends are PDB: 5AMR and 5AMQ, respectively.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, two tables, one data file and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.05.006.

## AUTHOR CONTRIBUTIONS

P.G. and J.R. did protein production, biochemistry, RNA binding studies, and X-ray data collection and structural analysis. P.G. crystallized the $L_{1750}$-vRNA complex. J.R. identified the $L_{1750}$ construct and solved the crystal structure. H.M. did EM analysis. S.C. and J.R. co-directed the project and wrote the paper with input from P.G. and H.M.

## ACKNOWLEDGMENTS

We thank Wim Hagen and John Briggs for cryo-EM data collection at EMBL Heidelberg, Leandro Estrozi, Irina Gutsche, and Guy Schoehn for advice on EM image analysis; the staff of the EMBL Eukaryotic Expression and High Throughput Crystallisation facilities, and the EMBL-ESRF JSBG staff for access to ESRF beamlines. Darren Hart helped with construct design and together with Philippe Mas and Damien Maurin with the ESPRIT analysis. Mass-spectroscopy was performed by Joanna Kirkpatrick (EMBL Proteomics Core Facility) and Luca Signor (IBS), Edman degradation by Jean-Pierre Andrieu (IBS) and Stefan Reich (EMBL) helped with RNA binding studies. This work used the platforms of the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) within the Grenoble Partnership for Structural Biology (PSB), with support from FRISBI (ANR-10-INSB-05-02), GRAL (ANR-10-LABX-49-01) and, for the EM facility, the Rhône-Alpes Region and the Fondation Recherche Medicale (FRM). H.M. and J.R. were supported by EMBO long-term Fellowships for some of this work. S.C. acknowledges support by ANR grant ArenaBunya-L and ERC Advanced Grant V-RNA (322586).

Received: February 6, 2015
Revised: March 20, 2015
Accepted: April 7, 2015
Published: May 21, 2015

## REFERENCES

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Barr, J.N. (2007). Bunyavirus mRNA synthesis is coupled to translation to prevent premature transcription termination. RNA *13*, 731–736.

Barr, J.N., and Wertz, G.W. (2004). Bunyamwera bunyavirus RNA synthesis requires cooperation of 3′- and 5′-terminal sequences. J. Virol. *78*, 1129–1138.

Barr, J.N., and Wertz, G.W. (2005). Role of the conserved nucleotide mismatch within 3′- and 5′-terminal regions of Bunyamwera virus in signaling transcription. J. Virol. *79*, 3586–3594.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. *59*, 2023–2030.

Chang, S., Sun, D., Liang, H., Wang, J., Li, J., Guo, L., Wang, X., Guan, C., Boruah, B.M., Yuan, L., et al. (2015). Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 Å resolution. Mol. Cell *57*, 925–935.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Elliott, R.M. (2014). Orthobunyaviruses: recent genetic and structural insights. Nat. Rev. Microbiol. *12*, 673–685.

Fodor, E. (2013). The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. Acta Virol. *57*, 113–122.

Gong, P., and Peersen, O.B. (2010). Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. USA *107*, 22505–22510.

Guu, T.S., Zheng, W., and Tao, Y.J. (2012). Bunyavirus: structure and replication. Adv. Exp. Med. Biol. *726*, 245–266.

Haddow, A.D., and Odoi, A. (2009). The incidence risk, clustering, and clinical presentation of La Crosse virus infections in the eastern United States, 2003-2007. PLoS ONE *4*, e6145.

Jorba, N., Area, E., and Ortín, J. (2008). Oligomerization of the influenza virus polymerase complex in vivo. J. Gen. Virol. *89*, 520–524.

Jorba, N., Coloma, R., and Ortín, J. (2009). Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. PLoS Pathog. *5*, e1000462.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr. D Biol. Crystallogr. *66*, 133–144.

Kohl, A., Dunn, E.F., Lowen, A.C., and 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. J. Gen. Virol. *85*, 3269–3278.

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.

Lehmann, M., Pahlmann, M., Jérôme, H., Busch, C., Lelke, M., and Günther, S. (2014). Role of the C terminus of Lassa virus L protein in viral mRNA synthesis. J. Virol. *88*, 8713–8717.

Morin, B., Coutard, B., Lelke, M., Ferron, F., Kerber, R., Jamal, S., Frangeul, A., Baronti, C., Charrel, R., de Lamballerie, X., et al. (2010). The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. PLoS Pathog. *6*, e1001038.

Morin, B., Kranzusch, P.J., Rahmeh, A.A., and Whelan, S.P. (2013). The polymerase of negative-stranded RNA viruses. Curr Opin Virol *3*, 103–110.

Müller, R., Poch, O., Delarue, M., Bishop, D.H., and Bouloy, M. (1994). Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. J. Gen. Virol. *75*, 1345–1352.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

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.

Plotch, S.J., Bouloy, M., Ulmanen, I., and Krug, R.M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell *23*, 847–858.

Raju, R., and Kolakofsky, D. (1989). The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base paired. J. Virol. *63*, 122–128.

Reguera, J., Weber, F., and Cusack, S. (2010). Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. PLoS Pathog. *6*, e1001101.

Reguera, J., Malet, H., Weber, F., and Cusack, S. (2013). Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein. Proc. Natl. Acad. Sci. USA *110*, 7246–7251.

Reguera, J., Cusack, S., and Kolakofsky, D. (2014). Segmented negative strand RNA virus nucleoprotein structure. Curr Opin Virol *5*, 7–15.

Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W., and Cusack, S. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature *516*, 361–366.

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

Sehnal, D., Svobodova Varekova, R., Berka, K., Pravda, L., Navratilova, V., Banas, P., Ionescu, C.M., Otyepka, M., and Koca, J. (2013). MOLE 2.0: advanced approach for analysis of biomacromolecular channels. J. cheminform. *5*, 39.

te Velthuis, A.J. (2014). Common and unique features of viral RNA-dependent polymerases. Cell. Mol. Life Sci. *71*, 4403–4420.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17–24.

York, A., Hengrung, N., Vreede, F.T., Huiskonen, J.T., and Fodor, E. (2013). Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. Proc. Natl. Acad. Sci. USA *110*, E4238–E4245.

# Structural insights into bunyavirus replication and its regulation by the vRNA promoter

Piotr Gerlach[1,2†], Hélène Malet[1,2†], Stephen Cusack[1,2*] and Juan Reguera[1,2*]

[1]European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, CS90181, 38042 Grenoble Cedex 9, France.

[2]Unit of Virus Host-Cell Interactions (UMI 3265), Univ. Grenoble Alpes-EMBL-CNRS, 71 Avenue des Martyrs, CS90181, 38042 Grenoble Cedex 9, France.

[†]Joint first author

*Corresponding author.

Tel. (33)476207238

Email: cusack@embl.fr, jreguera@embl.fr

## Supplementary Material

- Supplementary Figure Legends
- Supplementary Table Legends
- Extended Experimental Procedures
- Supplementary References.

## Supplementary Figure Legends.

**Figure S1. Purification and X-ray structure solution. Related to Figure 1.**

A. SDS PAGE of purified $L_{1750}$ after a S200 gel filtration run.

B. Scheme for X-ray crystal structure solution by the MIRAS methods using tantalum cluster, platinum and selenomethionine derivatives.

C. X-ray energy scan of selenomethionine derivative at Se absorption edge.

D. X-ray energy scan of tantalum derivative at Ta absorption edge.

E. Experimentally phased map at 2.85 Å resolution contoured at 1 σ (cyan mesh) with superposed final model showing a number of selenomethionines (anomalous difference peaks contoured at 4.0 σ, orange mesh) and one tantalum cluster position (anomalous difference peaks contoured at 4.0 σ, green mesh).

**Figure S2. Electron microscopy of apo-$L_{1750}$. Related to Figure 2.**

A. Negative stain EM micrograph of purified $L_{1750}$. Scale bar is indicated.

B. Representative cryo-EM micrograph collected at 80kV on a Krios microscope using a Falcon II direct detector (defocus of the chosen micrograph: 1.4µm).

C. Comparison between *ab initio* 2D class averages (top) and reprojections of the 3D reconstruction (bottom), showing the reliability of the 3D reconstruction. Below each comparison, the corresponding view of the 3D reconstruction with the fitted pseudo-atomic model is displayed.

D. Fourier Shell Correlation (FSC) based on the gold standard FSC=0.143 criterion shows a resolution of 8.3 Å.

E. Separated helices are clearly visible inside the map, consistent with the 8.3 Å resolution identified by the FSC.

F. Local resolution identifies flexible parts of the polymerase. A cut-away view is displayed with the pseudo-atomic model on the left and the cryo-EM reconstruction coloured by resolution on the right. A bar indicates the colour code corresponding to the local resolution. The most flexible parts, which correspond to the vRBL domain and the α-ribbon, are highlighted with a dotted ellipse.

**Figure S3. Biophysical and biochemical analysis of vRNA binding. Related to Figures 3 and 4.**

A. Measurements of Kd by fluorescence anisotropy.

B. Binding of duplexes with different length 3′ overhangs.

C. Sequence analysis of 3′ end binding.

D. Protection from trypsin cleavage by 3′ or 5′ end vRNA binding.

A. 3′ and 5′ vRNA affinity to $L_{1750}$ measured by fluorescence anisotropy. 5 nM of 25-mer RNAs corresponding to 3′ (red) or 5′ (blue) vRNA and labelled with fluorescein were titrated with $L_{1750}$ over the protein concentration range from 3 nM to 1 μM (left panel). Comparison of ssRNA (red) and dsRNA (green) binding to $L_{1750}$ upon titration with a broader protein concentration range (right panel). The calculated affinity values are indicated in the offset tables.

B. Mobility shift assay of radiolabelled, artificial panhandle RNAs with different lengths of 3′ overhang bound to $L_{1750}$. The RNAs used are schematically represented showing the overhangs and colouring the polymerase interacting nucleotides in red, the distal duplex nucleotides in green, the up/downstream nucleotides in black and guanine linker in white. The radioactive signals belonging to shifted bands were recorded with a Typhoon and quantified with ImageQuant. Amounts of bound RNAs were normalised against the amount of bound 22 nt long 3′ ssRNA used as a reference. The graphic plots the average values and standard deviation from four independent experiments. Radiography for one MSA experiment is shown in the offset maintaining the same order of RNAs as in the graphic but shifting the ssRNA control to the right lane.

C. Sequence specificity of the 3′ vRNA end binding to $L_{1750}$. Mobility shift assay of the wild type radiolabelled 22 nt long 3′ vRNA (WT) (top-left, same as in panel B) and all possible 33 point mutants within the first 3′ 11 nucleotides. As in panel B amounts of bound RNAs were normalised against the amount of bound WT 3′ vRNA used as a reference. The graphic plots the average values and standard deviations from four independent experiments. The RNA variants tested are loaded in the gel lanes maintaining the same order as in the graphic.

D. Protection from trypsin cleavage by 3′ or 5′ end vRNA binding. Limited trypsination of $L_{1750}$ without RNA, with 3′ vRNA and with 5′ vRNA was visualized by SDS-PAGE. The cleavage products were identified by ESI-TOF mass spectrometry to determine their molecular mass, tryptic MALDI-TOF to identify the tryptic peptides belonging to each, and Edman degradation to sequence the N-term of each band. The resulting molecular weight, peptide boundaries and methods used to cross-validate the identification on each experiment are shown in the table. The results are schematically summarised on the bottom right of the panel showing the $L_{1750}$ construct as a coloured bar where the regions before and after the trypsin cleavage site are coloured in red and blue respectively, the RdRp region in green and the $L_{1750}$ C-term region in yellow. The differentially cleaved region is represented by the dashed red lines box. The protein fragments belonging to each region are highlighted with dashed lines squares on the SDS-PAGE.

**Figure S4. Structure comparison between LACV and influenza polymerase. Related to Figure 1.**

In each panel equivalent structural features are highlighted for LACV (left) and influenza A (PDB 4WSB) (right) polymerases after superposition. The root mean square structural similarity for Cα positions is 3.18 Å for PA, 3.24 Å for PB1 and 4.33 Å for the first third of PB2 (PB2-N).

A. PA like region.

B. PB1 like region.

C. PB2-N like region.

**Figure S5. Details of protein-RNA interactions. Related to Figures 3 and 4**.

**A**. Residues (colour coded as shown, according to Fig. 1A) interacting with the 3′ vRNA and short duplex region. Based on data from CONTACT (Table S3)

**B**. As **A** but for the 5′ vRNA.

**Figure S6. Comparison of 5′ vRNA binding in LACV and influenza and induced structural changes in the arch and fingernode. Related to Figure 4.**

**A**. Stick model of the LACV (left) and influenza (right) 5′ vRNA stem loops. The base-paired residues are in pink, the additional stacked residues in grey and the flipped out bases of the loop in light cyan. Hydrogen bonds are shown as green dashed lines. Nucleotide U11, which is partially ordered in the LACV structure, is shown in yellow.

**B**. The structure of $L_{1750}$ before (light green) and after (coloured as in Figure 1A) soaking with 10 nucleotides 5′ vRNA (yellow tube). 5′ vRNA binding prompts a reorientation of arch residues Val437 and Ser438 and induces a significant reconfiguration of the arch.

**C**. The interaction with 5′ vRNA loop bases G7 and U8 radically changes the configuration of the fingernode loop (light green before to grey after) allowing the stacking of Tyr1120 onto G7 which makes base-specific hydrogen bonds with Gln1116 and Asp1123 and base-specific recognition of U8 by main-chain interactions.

**D**. Influenza 5′ vRNA nucleotides 1-16 (red tube) and 3′ vRNA nucleotides 10-14 (blue tube) are superposed on the $L_{1750}$ 5′ vRNA soaked structure after superposition of bat influenza FluA structure (PDB 4WSB) and LACV polymerases. LACV polymerase is shown as a grey cartoon with 5′ vRNA nucleotides 1-10 in yellow. Partially ordered LACV U11 is highlighted in yellow sticks and corresponds roughly to A10 in influenza. The LACV α-ribbon (orange) would clash with a distal, base-paired region positioned as in influenza, but could rotate (arrow) to play a role similar to the PB1 β-ribbon (green) in binding the duplex region.

**Figure S7. Newly identified conserved motifs in all sNSV polymerases and details of the polymerase tunnels. Related to Figure 5.**

**A**. The conservation of motif G (**R**Yφφ, bold absolutely conserved, φ denoting a hydrophobic residue) is shown in an alignment (left) including two strains each from Orthomyxoviruses (Influenza A and B), Bunyaviruses (Ortho: LACV/Bunyamwera, Hanta: Hantaan/Imgin, Phlebo: Rift Valley fever/Uukuniemi, Nairo: Crimean-Congo haemorrhagic fever/Kupe, Tospo: Tomato spotted wilt/ Peanut bud necrosis) and Arenavirus (Lassa/Junin). The environment of motif G is shown for LACV (middle) and influenza (right) with the neighbouring motifs A, C, D (including conserved Lys1228/481) and E. Structurally similar elements are coloured according to Figure 1A.

**B.** The same for motif H (**K**xφxφ) showing for LACV (middle) and influenza (right) the interactions of the conserved lysine with the backbone of the motif B loop. Neighbouring LACV fingernode (grey) is highlighted (middle) as well as the analogous β-hairpin (grey) in influenza polymerase (right). In both cases neighbouring motifs A, B, and F are marked. Structurally similar elements are coloured according to Figure 1A.

**C**. Section through the $L_{1750}$ polymerase structure in van der Waals surface representation coloured according to surface electrostatic potential showing that the four conserved channels and central cavity are positively charged. The 3′ and 5′ RNAs are shown as blue and yellow ribbons respectively. Green arrows indicate the NTP and RNA traffic following the homology with influenza and other similar RNA polymerases.

**D**. As **C** but solvent accessible residues are coloured according to their degree of conservation amongst orthobunyaviruses (violet > 90%, 90% < pink > 60%, white < 60%).

**E**. The NTP entry channel is lined by conserved basic residues R287, K673 (PA-C like), K956, R958 (fingertips), K1063 (motif A), K1227, K1228 (motif D) some of which are only positioned correctly upon 5′ vRNA binding. Colouring of domains is as in Fig 1A.

**F**. View into the product exit channel showing contributing conserved residues from the fingers, bridge thumb ring and lid (coloured as in Fig.1A).

**Table S1. Crystallographic data collection and refinement statistics.**

**Table S2. Sequence alignment of representative orthobunyavirus polymerases with superposed secondary structure of LACV polymerase and domains annotated and coloured according to Figure 1A.**

**Table S3. Table of protein-vRNA hydrogen bonding contacts.**

## Extended Experimental Procedures

### Protein production and crystallisation

The L protein sequence of the La Crosse virus strain LACV/mosquito/1978 (GenBank code: EF485038.1, UniProt code: A5HC98) was used to design a synthetic, codon-optimized gene also coding for an N-terminal His-tag followed by Tobacco Etch Virus (TEV) protease cleavage site (MG$\underline{\text{HHHHHH}}_{\text{6xHis-tag}}$DYDIPTT$\underline{\text{ENLYFQ}}_{\text{TEV}}\underline{\text{G}}$) (DNA 2.0). L protein construct 1-1750 ($L_{1750}$) was found by limited trypsin digestion of purified full-length L, its C-terminus also corresponding to boundaries of soluble domains identified by ESPRIT ((Yumerefendi et al., 2010)(data not shown).

$L_{1750}$ was expressed in Hi5 insect cells using the Bac-to-Bac expression system with a pFastBac vector. Cells were collected by centrifugation, re-suspended in lysis buffer (50mM Tris-HCl, 500 mM NaCl, 20 mM imidazole 0.5 mM TCEP, pH 8) supplemented with protease inhibitors (Roche, complete mini, EDTA-free), and lysed by sonication. Cell lysate was clarified by centrifugation for 1 h at 35,000 g and further purified by precipitation with $NH_4SO_4$ at 0.5 g/ml lysate. The protein fraction was pelleted by 30 min centrifugation at 70,000 g and re-suspended in lysis buffer. After a final clarification step (30 min at 70,000 g), the polymerase was purified from the supernatant by sequential nickel ion affinity, heparin affinity and size exclusion chromatography (Fig. S1A).

$L_{1750}$ protein at 5 mg/ml concentration in 20 mM Tris-HCl at pH 8, 150 mM NaCl, 2 mM TCEP, was mixed in 3:4 ratio with an equimolar mixture of nucleotides 1-16 from the 3′ vRNA end (3′-UCAUCACAUGAUGGUU-5′) and nucleotides 9-16 from the 5′ vRNA end (5′-GCUACCAA-3′), which had been pre-annealed. All oligonucleotides were purchased from Dharmacon -GE Healthcare. Brush-like clusters of needles emerged after 2-3 days from several conditions containing 15% (w/v) of PEG 4000 or 20% (w/v) of PEG 3350 as precipitant and 0.2 M of various sulphate or citrate salts. After an additive screen and pH and precipitant/salt concentration optimization, the optimal conditions were found to be 0.1 M Hepes, 0.2 M sodium citrate tribasic dihydrate, 0.3 M ammonium sulphate, 15% PEG 3350, pH 7. Microseeding allowed production of rod-shaped single crystals suitable for data collection. Seleno-methionine derivative crystals were grown in the same conditions as native ones, using initially native seeds and then switching to seleno-methionine seeds. Heavy metal derivatives were prepared by soaking native crystals in mother liquor supplemented with either 2.5-5 mM $K_2PtCl_4$ for 1-2 hours, or with 5-10 mM $Ta_6Br_{12}$ tantalum cluster (Jena Bioscience) for 1-2 days which turned the crystals green. Crystals containing the 5′ genomic end were obtained by soaking $L_{1750}$-vRNA crystals with 5′ nucleotides 1-8, 1-10, 1-11 and 1-12 (5′-pAGUAGUGUGCUA-3′ and shorter).

**Data collection, structure solution and model building**

All datasets were collected on ID23-1 and ID29 beamlines at ESRF. Data were integrated and scaled with XDS (Kabsch, 2010).

$L_{1750}$-vRNA complexes crystallize in the $P2_12_12_1$ space group (a= 102.5, b=141.0, c=165.5 Å) with one complex per asymmetric unit and 60 % solvent and diffract up to 2.7 Å resolution. To obtain experimental phases optimised anomalous datasets from platinum, tantalum and selenomethionine derivatives were collected at 4.8, 4.2-4.4 and 3.5 Å resolution, respectively (Fig. S1B-D).

The structure was finally solved by first using a highly redundant, 5 Å resolution Ta dataset, measured at the absorption peak, from which four sites for tantalum clusters were localized in a SAD experiment using autoSHARP (Vonrhein et al., 2007). This substructure was then refined and improved using the multiple isomorphous replacement with anomalous signal (MIRAS) method with SHARP (Bricogne et al., 2003), using the same dataset at 4.2 Å resolution, another dataset measured at the Ta inflection point at 4.4 Å from the same crystal, SeMet derivative data measured at the Se absorption peak, 5 mM platinum soaked crystals at the Pt absorption peak and native data at 2.85 Å resolution. After several rounds of refinement and sites detection by anomalous difference maps, 16 tantalum cluster, 9 platinum sites and 53 selenium sites, were found, leading to an interpretable map at 2.85 Å resolution. LACV endonuclease (PDB 2XI5) and influenza polymerase PA and PB1 subunits (PDB 4WSB) could be fitted into the map as a guide to model building (Fig. S1E). The rest of $L_{1750}$-vRNA complex model was iteratively built with COOT (Emsley and Cowtan, 2004) and refined with REFMAC (Murshudov et al., 1997), using selenomethionine positions from the anomalous difference map to help align the sequence. Autobuilding with BUCCANNEER (Cowtan, 2006) was useful to extend the model. The excellent quality of the RNA electron density allowed unambiguous building of co-crystallized vRNA and revealed that the distal short duplex stacks with its exterior blunt end on a neighbouring protein molecule, thus explaining its requirement for crystallisation.

Native $L_{1750}$-vRNA crystals soaked with the 5′ genomic ends diffracted to the following resolutions: 5′p1-8 –3.0 Å, 5′p1-10 –3.0 Å, 5′p1-11 –3.1 Å and 5′p1-12–2.8 Å. All showed extra 5′ RNA density (Fig. 4A) except that with 5′p1-12.

Structure figures were drawn with Pymol (DeLano, 2002).

**Electron microscopy**

Purified $L_{1750}$ was initially checked for homogeneity by negative stain electron microscopy. 4 µl of 0.02 mg/ml of sample was applied to the clear side of carbon on a carbon-mica interface and stained with 1% (wt/vol) sodium silico-tungstate pH 7.5. Images were recorded under low-dose conditions with a JEOL 1200 EX II microscope at 100 kV using a nominal magnification of 40,000× (Fig. S2A). Homogeneous samples were subsequently used for cryo-EM analysis. For cryo-EM grids preparation, 4µl of polymerase at 0.2mg/ml was applied onto a glow-discharged and pumped quantifoil grid 400 mesh 2/1 (Quantifoil Micro Tools GmbH, Germany), excess solution was blotted during 2 s with a

Vitrobot (FEI) and the grid frozen in liquid ethane. Data collection was performed on a FEI Krios microscope operated at 80 kV. CryoEM micrographs were collected on a Falcon II electron direct detector at a magnification of 138,129× and with a binning factor of 2, giving a pixel size of 2.02 Å (Fig. S2B). FEI EPU automatization software was used to collect 6129 micrographs with a defocus between 0.5 and 2 µm, an exposure time of 0.5 s and a dose of $14e^{-}/Å^{2}$. The contrast transfer function (CTF) for each micrograph was determined with CTFFIND3 (Mindell and Grigorieff, 2003). The 4646 best micrographs, based on manual inspection and CTF quality, were kept for subsequent image processing. 10,102 particles were manually picked using EMAN boxer (Ludtke et al., 1999) and used for reference-free 2D MSA classification as implemented in IMAGIC (van Heel et al., 1996). The resulting class averages were used for automated particle picking with a procedure based on the Fast Projection Matching algorithm (Estrozi and Navaza, 2008). False positives from the automated picking procedure were discarded after 2D classification in IMAGIC and RELION 1.3 (Scheres, 2012) leading to a 178,983 particle dataset. Class averages of this clean dataset allowed identification of secondary structure elements of the polymerase (Fig. S2C). An initial 3D model was generated by filtering the $L_{1750}$-vRNA crystal structure at 30 Å resolution and refined as implemented in RELION 1.3 autorefine using the cleaned dataset, leading to a map at 8.3 Å resolution (Fig. 2A, Fig. S2D, E)(Scheres and Chen, 2012). Local resolution variations were estimated using ResMap (Fig. S2F), identifying regions with lower resolution corresponding to flexible parts of the polymerase (Kucukelbir et al., 2014). In a subsequent 3D classification run, an angular sampling of 3.75° was combined with local angular searches around the refined orientations and led to separation of the dataset into three different subsets, corresponding to three different states of the polymerase (Fig. 2). The first subset of 90960 particles led to a 9.7 Å resolution EM map which shows only partial density for the endonuclease, depicting its flexibility (Fig. 2B). A second subset containing 43832 particles led to a 9.7 Å resolution reconstruction with improved density in vRBL (Fig. 2C). The third subset containing 44191 particles yielded a map at 9.3 Å resolution which clearly shows that the vRBL is highly flexible (Fig. 2D). Reported resolutions are based on the gold-standard FSC = 0.143 criterion and are calculated without applying any mask. Prior to visualization, all density maps were sharpened by applying a negative B-factor of 900 $Å^2$. The crystallographic model of the endonuclease (residues 1-184) and the polymerase core domain (185-1745) were fitted as two rigid bodies into the EM map using Chimera Fit-in-map module (Pettersen et al., 2004).

**Polymerase-vRNA binding studies**

For electrophoretic mobility shift assays (EMSA) (Fig. S3B, D) radioactively labelled RNAs were produced by *in vitro* transcription with T7 polymerase (produced in house). For each 10 μl transcription 2 μM DNA oligos were preheated and annealed in order to form double stranded T7 promoter with long overhang being the reaction template. Reactions were carried with ATP/UTP/CTP (Jena Bioscience) and [$\alpha^{32}$P]-GTP (PerkinElmer). RNA products were than resolved and purified from urea TBE 15-20 % polyacrylamide gels. For each binding experiment 10 μM of $L_{1750}$ in 10 μl buffer (100 mM Tris, 100 mM NaCl, 0.5 mM DPBA (an endonuclease inhibitor), 5 mM β-mercaptoethanol, pH 8) was mixed with radiolabelled RNA and 1 μl of non-specific poly(U) RNA (Sigma). To maintain a similar amount of radioactive RNA in each mixture the volumes added were scaled based on ImageQuant and Geiger counter measurements performed prior to the experiment. Mixtures were incubated at room temperature for several hours and resolved on native TG gels (top and bottom parts of the gel with 10 and 20% polyacrylamide respectively). Radioactive signal from shifted bands was recorded with a Typhoon and quantified with ImageQuant. In each case the amount of bound RNA was measured with reference to the control RNA of known strong affinity to the protein.

For fluorescence anisotropy measurements (Fig. S3A), 5 nM 25-nucleotide long RNA oligos corresponding to 3′ or 5′ vRNA (IBA), labelled with fluorescein on the appropriate non-interacting end, in 150 µl buffer (100 mM Tris, 150 mM NaCl, pH 8) were titrated with $L_{1750}$ in order to obtain 10-15 protein concentration points ranging from 3 nM to 1 µM. Fluorescence and fluorescence anisotropy measurements were carried out in quartz cuvettes, using 495 nm excitation and 515 nm emission wavelengths. KaleidaGraph (Synergy Software) was used to evaluate the data and derive dissociation constants.

For proteolysis protection experiments (Fig. S3D), $L_{1750}$ was incubated at 1 mg/ml with 1:1 molar ratios of 25-nucleotide long 3′ or 5′ genomic ends and then digested for 1 hour at room temperature with trypsin (1:1000 w/w). Products of digestion were analyzed by various techniques including: SDS-PAGE, western-blot, ESI-TOF-MS, MALDI-TOF-MS, MALDI-TOF-MS with N-terminal acetylation, and N terminal sequencing of protein fragments by Edman degradation.

**Polymerase activity assays.**

$L_{1750}$ or full length L protein and synthetic RNA of various lengths corresponding to the 3' and 5' vRNA ends (GE Dharmacon) were used. RNAs were preheated to 90°C prior to the reaction and, depending on the approach, added either sequentially or simultaneously (as pre-annealed) to the protein sample. The replication assays were done in 10 µl aliquots mixing 0.5 µM of L-protein, 0.5 µM of 3' RNA with or without 0.5 µM 5' RNA, 0.5 mM UTP/ATP/GTP/CTP mixture (Jena Bioscience) and [α32P]-UTP (PerkinElmer), in 20 mM Tris, 150 mM NaCl, 5 mM MgCl2, pH 8, and incubated at 30°C for 2-4h. Reaction products were loaded on urea TBE 15-20 % polyacrylamide gels. The auto-radiographs showed very weak product bands with no consistent product with the template length as would be expected for a replication product. Products only few nucleotides longer than the 3' RNA template used in the reaction were frequently observed. In most cases the control reaction containing only the radioactive UTP led to the same output as reactions containing all four NTPs, suggesting non-templated nucleotide addition. The presence of regulatory 5' RNA end affected the output of the reactions, not making however the experiment less ambiguous (data not shown). The inconclusive polymerase activity suggests that other viral (e.g. NP) or cellular factors might be essential for its proper function.

## Supplementary References


Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta crystallographica Section D, Biological crystallography *59*, 2023-2030.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta crystallographica Section D, Biological crystallography *62*, 1002-1011.

DeLano, W.L. (2002). PyMOL Molecular Graphics System. available online at http://wwwpymolsourceforgenet.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography *60*, 2126-2132.

Estrozi, L.F., and Navaza, J. (2008). Ab initio high-resolution single-particle 3D reconstructions: the symmetry adapted functions way. Journal of structural biology *172*, 253-260.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta crystallographica Section D, Biological crystallography *66*, 133-144.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. Nature methods *11*, 63-65.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. Journal of structural biology *128*, 82-97.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of structural biology *142*, 334-347.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta crystallographica Section D, Biological crystallography *53*, 240-255.

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.

Reguera, J., Cusack, S., and Kolakofsky, D. (2014). Segmented negative strand RNA virus nucleoprotein structure. Current opinion in virology *5*, 7-15.

Reguera, J., Malet, H., Weber, F., and Cusack, S. (2013). Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein. Proceedings of the National Academy of Sciences of the United States of America *110*, 7246-7251.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. Journal of structural biology *180*, 519-530.

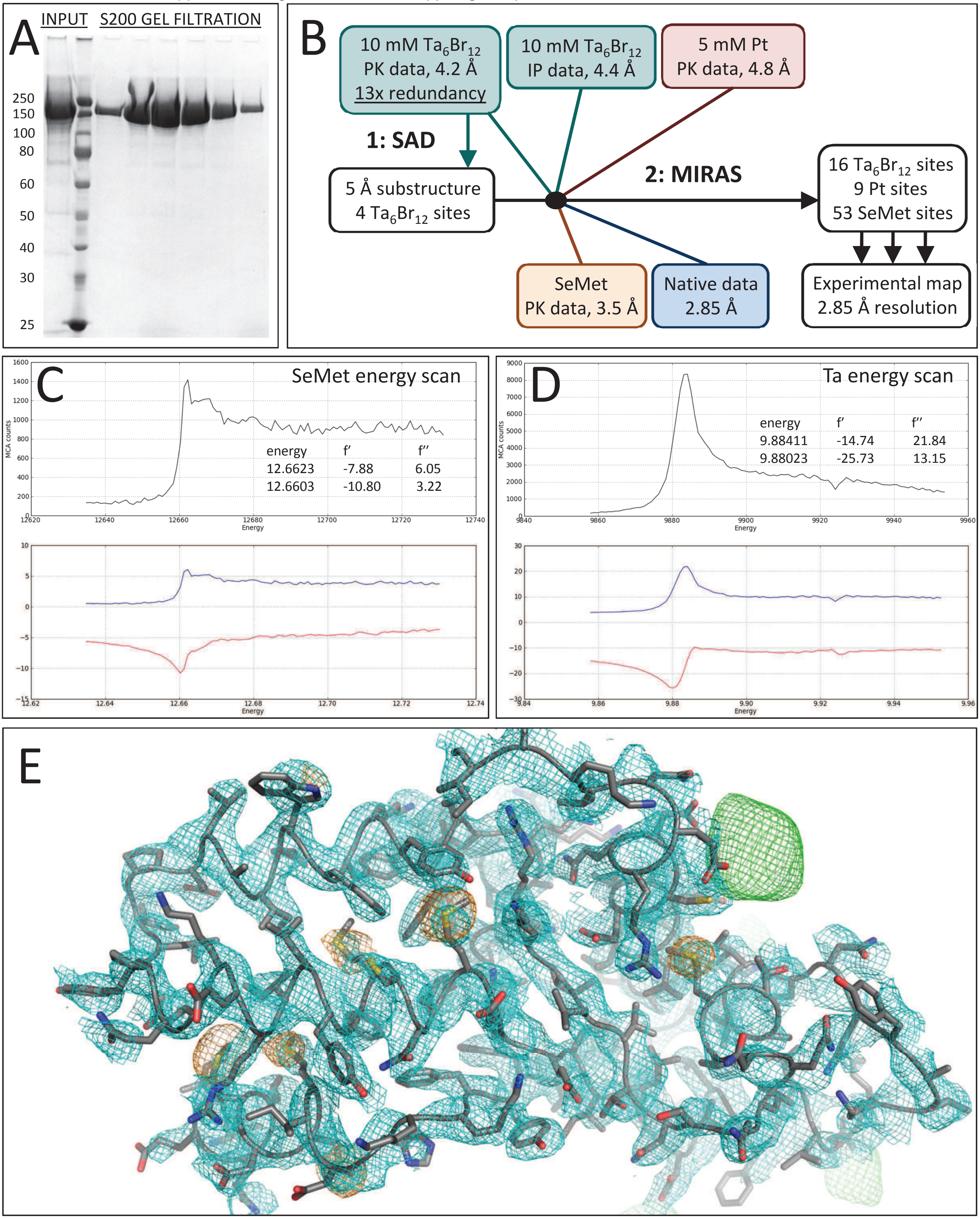
Scheres, S.H., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. Nature methods *9*, 853-854.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. Journal of structural biology *116*, 17-24.

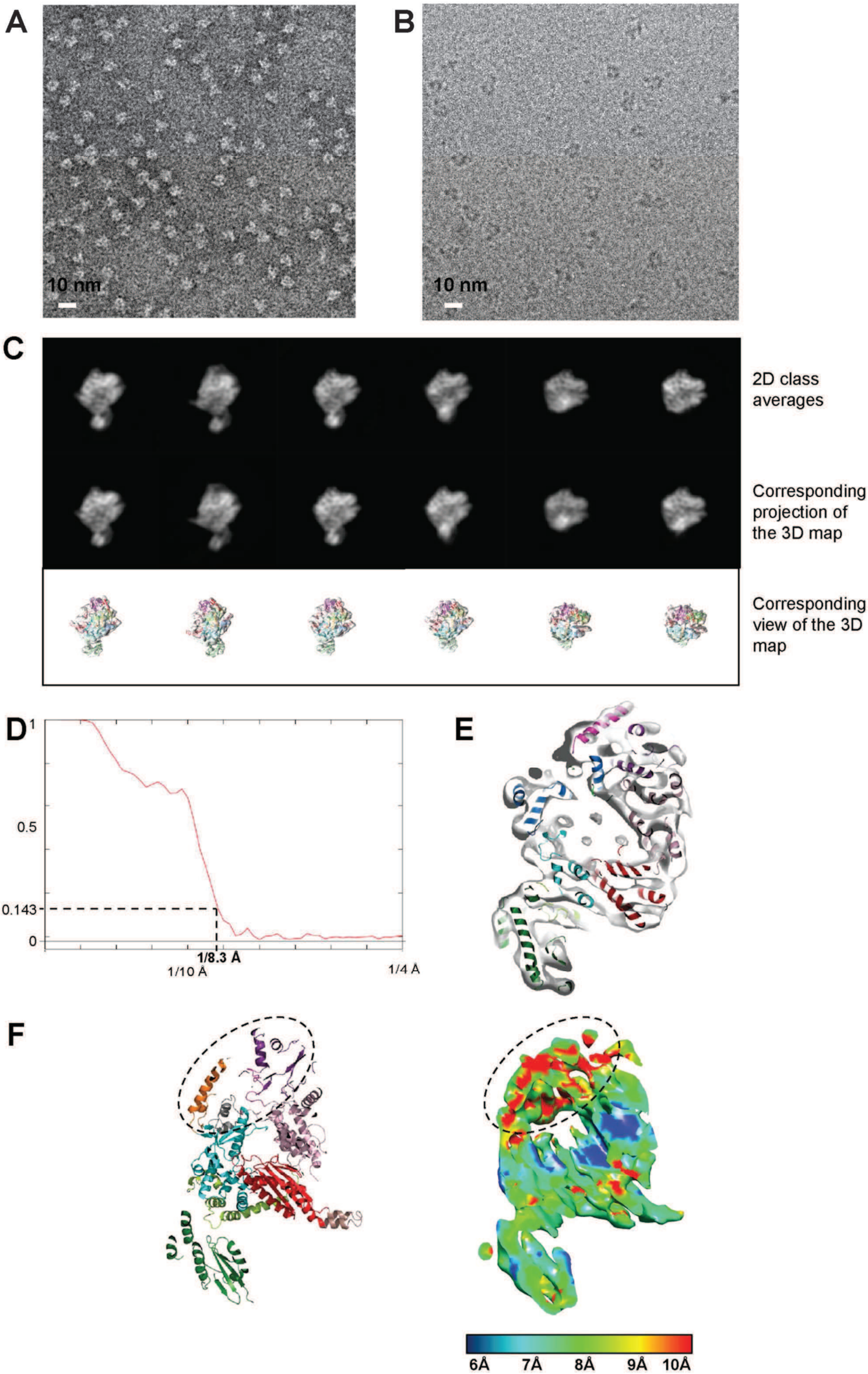
Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol Biol *364*, 215-230.

Yumerefendi, H., Tarendeau, F., Mas, P.J., and Hart, D.J. (2010). ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. Journal of structural biology *172*, 66-74.

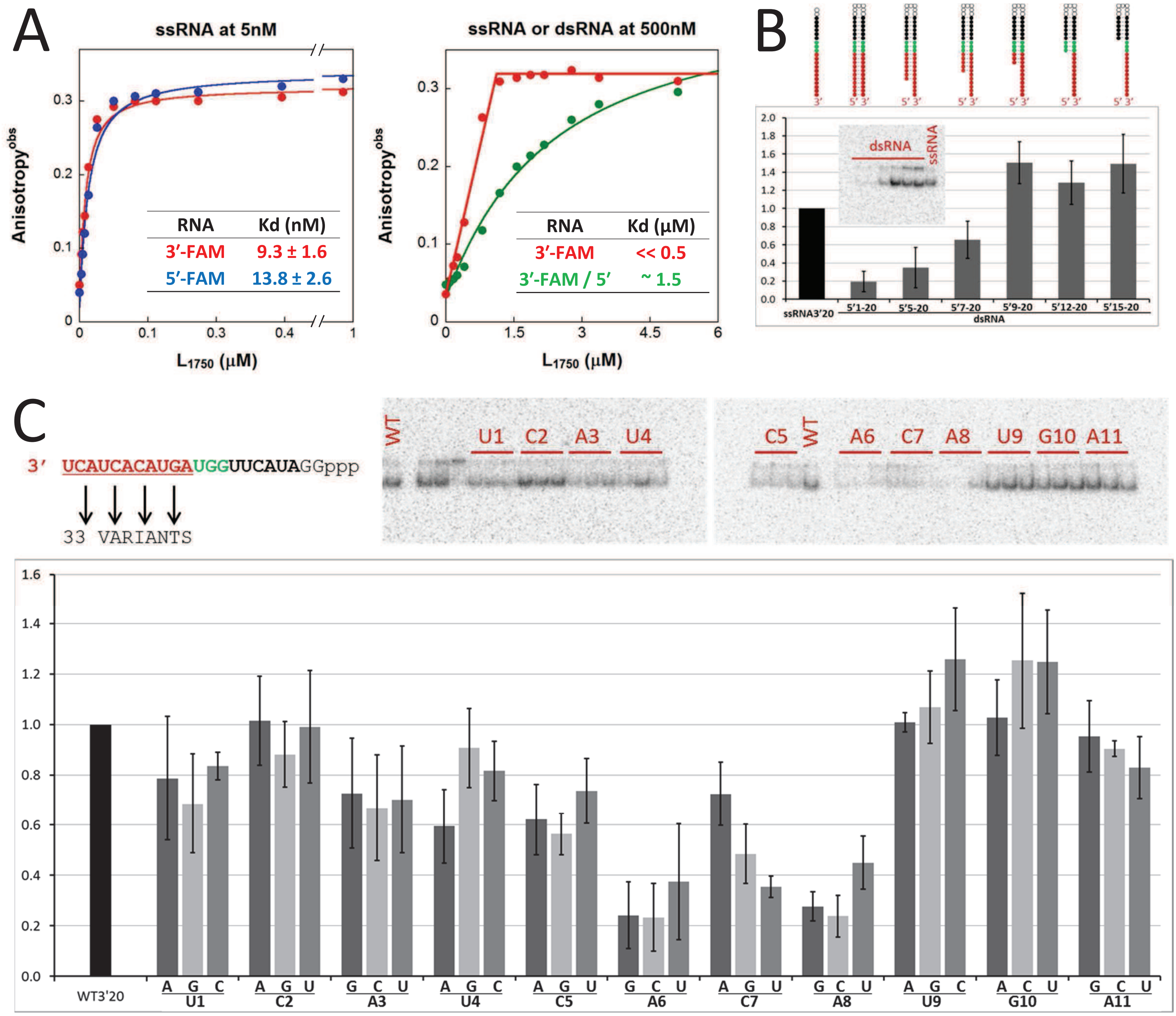
# Figure S1

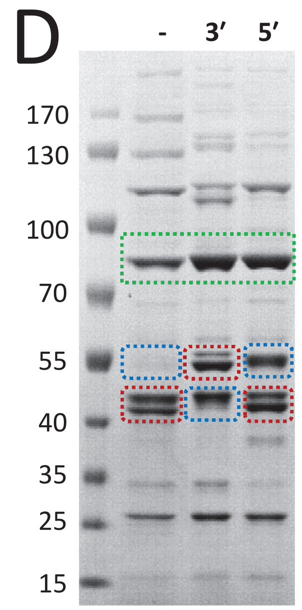

A
INPUT
S200 GEL FILTRATION
250
150
100
80
60
50
40
30
25
B
10 mM $Ta_6Br_{12}$
PK data, 4.2 Å
13x redundancy
10 mM $Ta_6Br_{12}$
IP data, 4.4 Å
5 mM Pt
PK data, 4.8 Å
1: SAD
5 Å substructure
4 $Ta_6Br_{12}$ sites
2: MIRAS
16 $Ta_6Br_{12}$ sites
9 Pt sites
53 SeMet sites
SeMet
PK data, 3.5 Å
Native data
2.85 Å
Experimental map
2.85 Å resolution
C
SeMet energy scan
MCA counts
Energy
energy f' f''
12.6623 -7.88 6.05
12.6603 -10.80 3.22
D
Ta energy scan
MCA counts
Energy
energy f' f''
9.88411 -14.74 21.84
9.88023 -25.73 13.15
E

# Figure S2

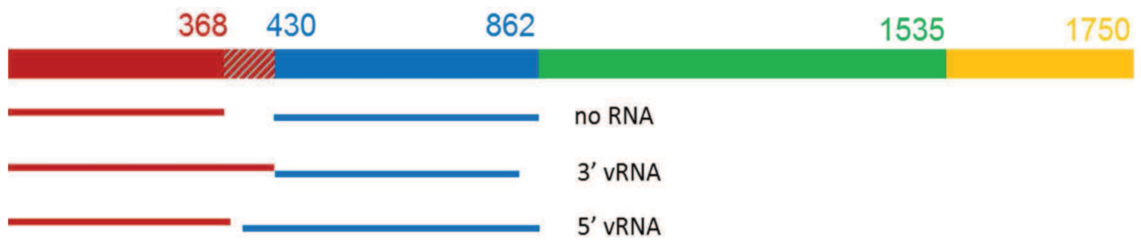

A
10 nm
B
10 nm
C
2D class averages
Corresponding projection of the 3D map
Corresponding view of the 3D map
D
1
0.5
0.143
0
1/8.3 Å
1/10 Å
1/4 Å
E
F
6Å
7Å
8Å
9Å
10Å

# Figure S3

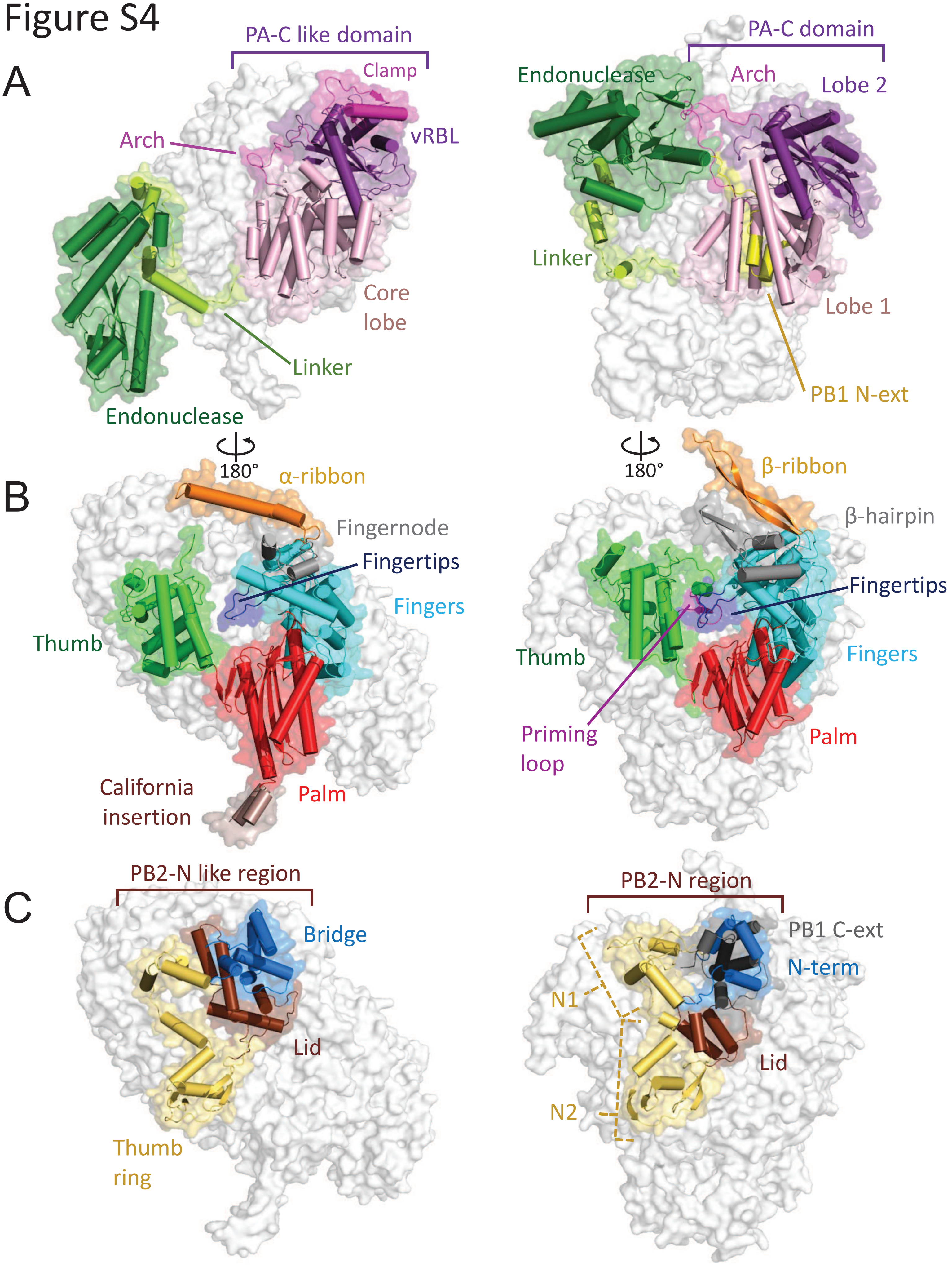


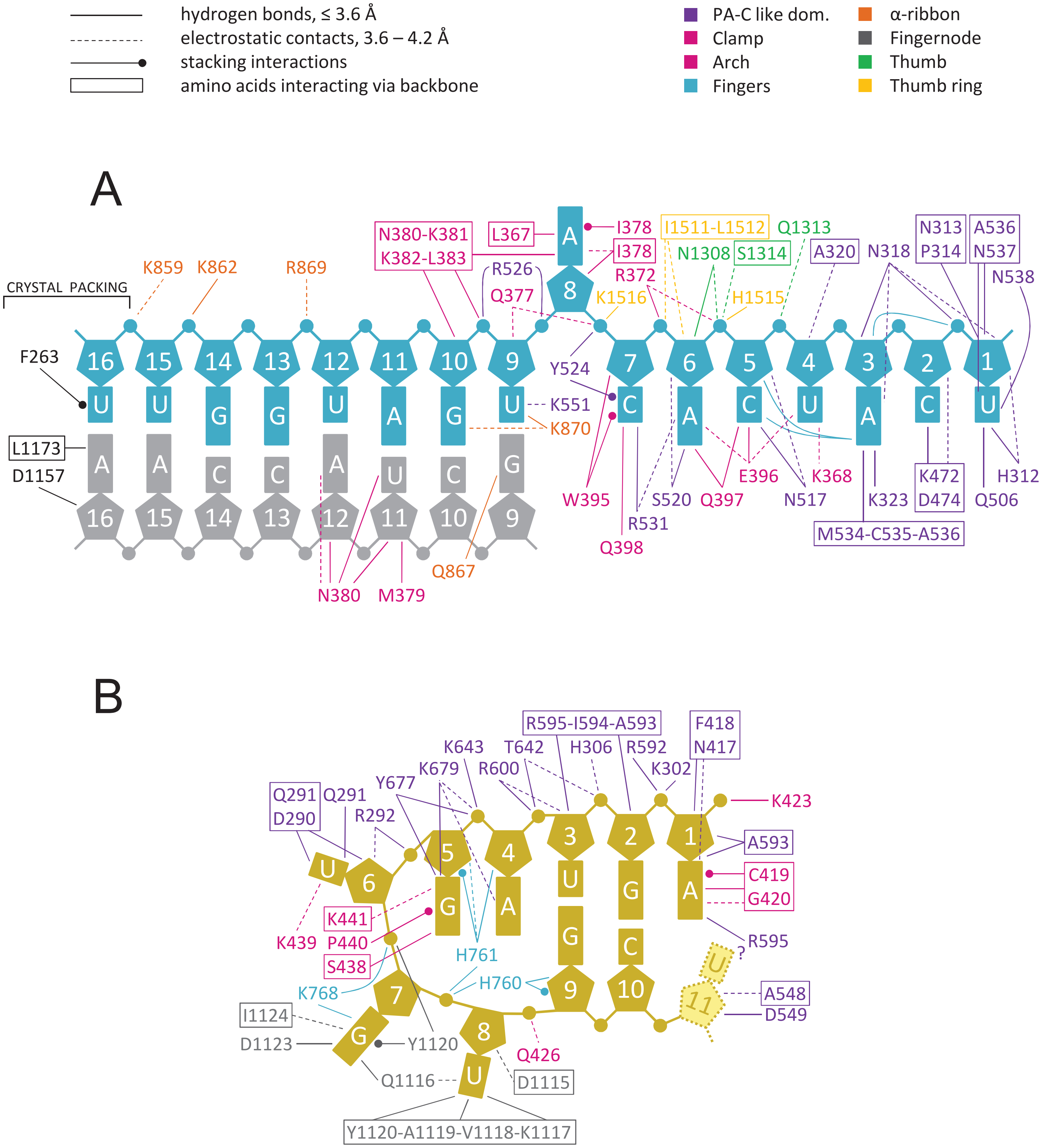

| | MW (kDa) | sequence | cross-validation |
|---|---|---|---|
| - | 50.125 | (431) K \| MLEDL – VKQKK \| I (862) | Edman degr., MALDI (Acetyl-Nterm) |
| | 43.270 | (1) GMDYQ – CMSLK \| S (368) | Edman degr., MALDI (Acetyl-Nterm), Nterm WB |
| | | | |
| 3' | ~ 47 | (431) K \| MLEDL – | Edman degr., MALDI (Acetyl-Nterm) |
| | 50.658 | (1) GMDYQ – QFKNK \| M (430) | Edman degr., MALDI (Acetyl-Nterm), Nterm WB |
| | | | |
| 5' | 56.582 | (377) K \| QIMNK – VKQKK \| I (862) | MALDI (Acetyl-Nterm) |
| | 43.713 | (1) GMDYQ – KSKAR \| S (372) | Edman degr., MALDI (Acetyl-Nterm), Nterm WB |



# Figure S4


A
PA-C like domain
Clamp
Arch
vRBL
Core lobe
Linker
Endonuclease
PA-C domain
Endonuclease
Arch
Lobe 2
Linker
Lobe 1
PB1 N-ext
B
180°
α-ribbon
Fingernode
Fingertips
Fingers
Thumb
California insertion
Palm
180°
β-ribbon
β-hairpin
Fingertips
Thumb
Fingers
Priming loop
Palm
C
PB2-N like region
Bridge
Lid
Thumb ring
PB2-N region
PB1 C-ext
N-term
N1
N2
Lid

# Figure S5



hydrogen bonds, ≤ 3.6 Å
electrostatic contacts, 3.6 – 4.2 Å
stacking interactions
amino acids interacting via backbone
PA-C like dom.
Clamp
Arch
Fingers
α-ribbon
Fingernode
Thumb
Thumb ring
A
CRYSTAL PACKING
F263
L1173
D1157
K859
K862
R869
N380-K381
K382-L383
L367
R526
Q377
I378
I378
R372
K1516
I1511-L1512
Q1313
N1308
S1314
H1515
A320
N318
N313
P314
A536
N537
N538
Y524
K551
K870
W395
R531
Q398
S520
Q397
E396
N517
K368
K323
M534-C535-A536
K472
D474
Q506
H312
Q867
N380
M379
B
R595-I594-A593
F418
N417
K643
T642
H306
R592
K679
R600
K302
Q291
D290
Q291
Y677
R292
K423
A593
C419
G420
K441
K439
P440
S438
H761
R595
?
K768
H760
A548
D549
I1124
D1123
Y1120
Q426
Q1116
D1115
Y1120-A1119-V1118-K1117

# Figure S6

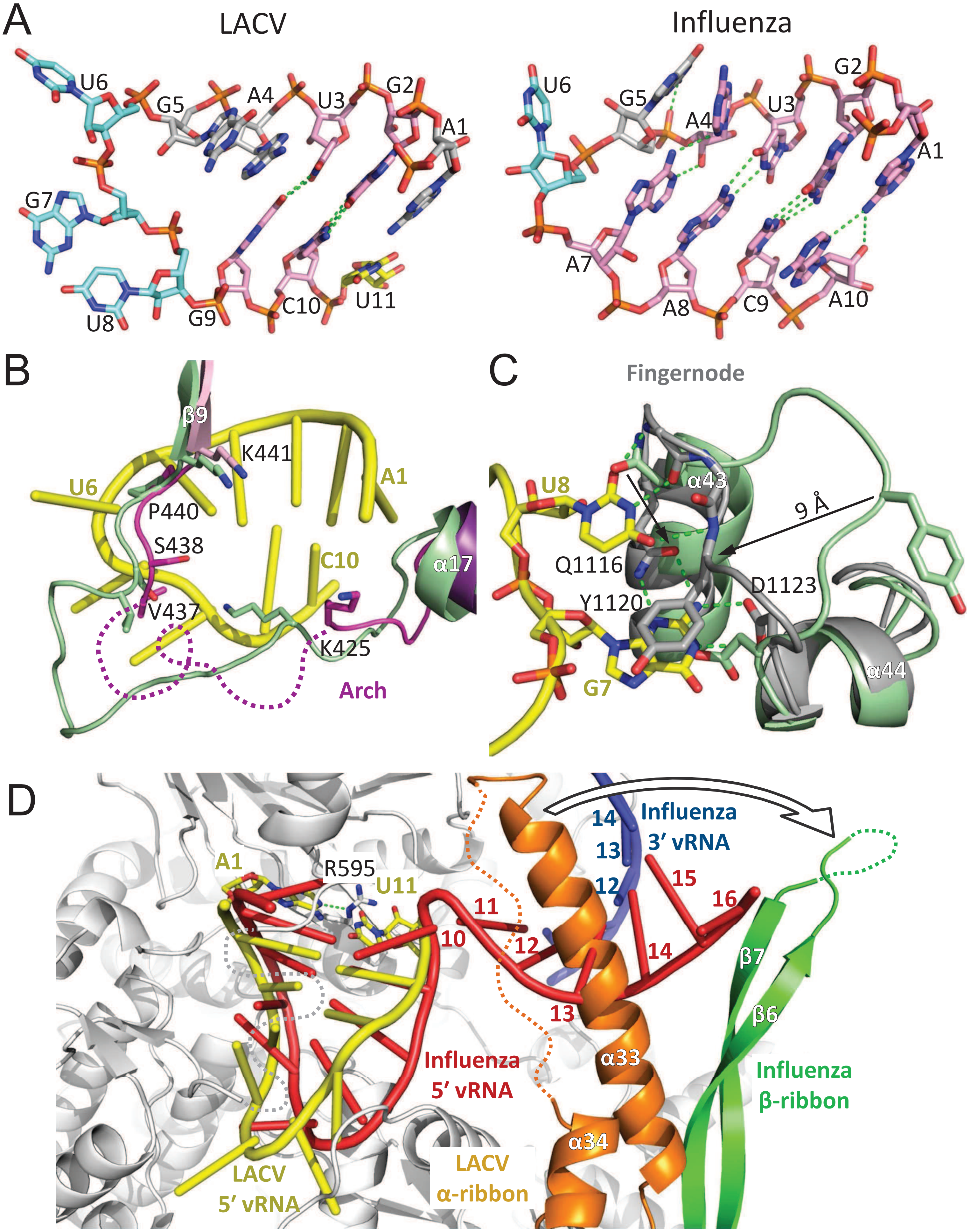
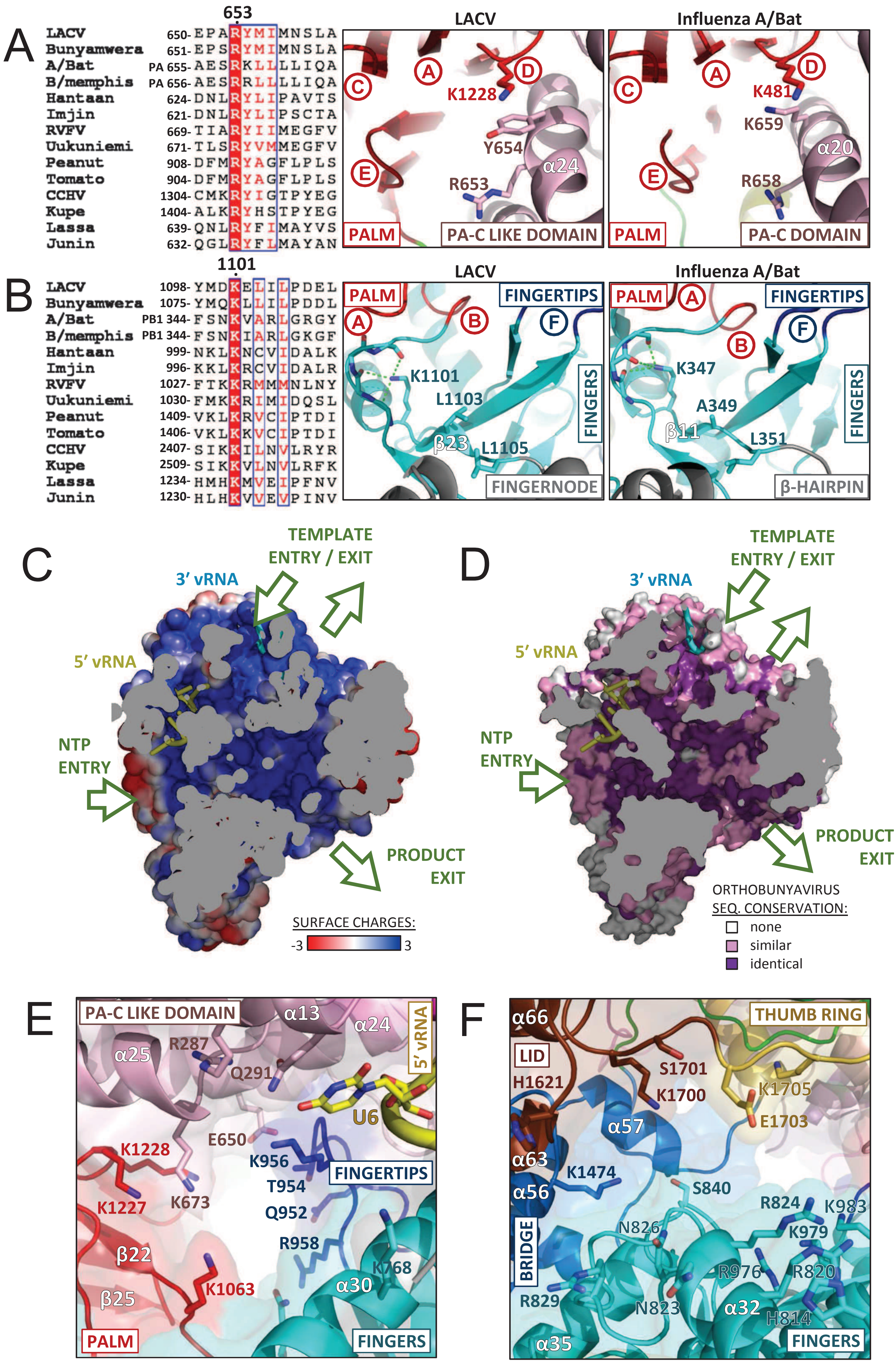

A
LACV
Influenza
U6
G5
A4
U3
G2
A1
G7
U8
G9
C10
U11
A7
A8
C9
A10
B
β9
K441
U6
A1
P440
S438
C10
α17
V437
K425
Arch
C
Fingernode
U8
α43
9 Å
Q1116
D1123
Y1120
G7
α44
D
Influenza
3′ vRNA
14
13
12
A1
R595
U11
15
16
11
10
12
14
13
β7
β6
Influenza
5′ vRNA
α33
Influenza
β-ribbon
α34
LACV
α-ribbon
LACV
5′ vRNA

Figure S7



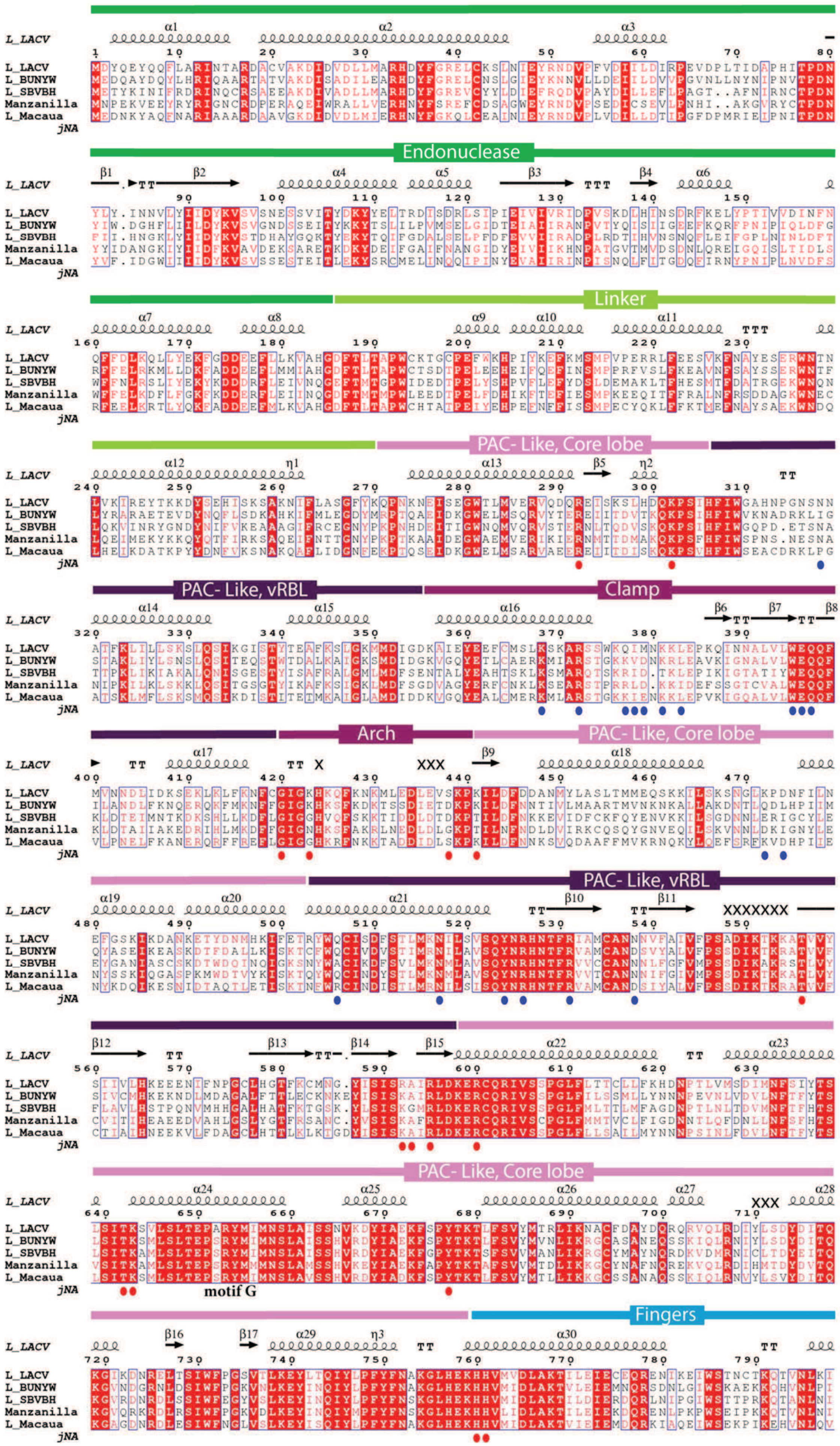
A
653
LACV 650- EPARYMIMNSLA
Bunyamwera 651- EPSRYMIMNSLA
A/Bat PA 655- AESRKLLLLIQA
B/memphis PA 656- AESRRLLLLIQA
Hantaan 624- DNLRYLIPAVTS
Imjin 621- DNLRYLIPSCTA
RVFV 669- TIARYIIMEGFV
Uukuniemi 671- TLSRYVMMEGFV
Peanut 908- DFMRYAGFLPLS
Tomato 904- DFMRYAGFLPLS
CCHV 1304- CMKRYIGTPYEG
Kupe 1404- ALKRYHSTPYEG
Lassa 639- QNLRYFIMAYVS
Junin 632- QGLRYFLMAYAN
LACV
C
A
D
K1228
Y654
E
R653
α24
PALM
PA-C LIKE DOMAIN
Influenza A/Bat
K481
K659
α20
R658
PA-C DOMAIN
B
1101
LACV 1098- YMDKELILPDEL
Bunyamwera 1075- YMQKLLILPDDL
A/Bat PB1 344- FSNKVARLGRGY
B/memphis PB1 344- FSNKIARLGKGF
Hantaan 999- NKLKNCVIDALK
Imjin 996- KKLKRCVIDALR
RVFV 1027- FTKKRMMMNLNY
Uukuniemi 1030- FMKKRIMIDQSL
Peanut 1409- VKLKRVCIPTDI
Tomato 1406- VKLKRVCIPTDI
CCHV 2407- SIKKILNVLRYR
Kupe 2509- SIKKVLNVLRFK
Lassa 1234- HMHKMVEIPFNV
Junin 1230- HLHKVVEVPINV
FINGERTIPS
F
K1101
L1103
β23
L1105
FINGERS
FINGERNODE
K347
A349
β11
L351
β-HAIRPIN
C
TEMPLATE ENTRY / EXIT
3′ vRNA
5′ vRNA
NTP ENTRY
PRODUCT EXIT
SURFACE CHARGES:
-3
3
D
ORTHOBUNYAVIRUS SEQ. CONSERVATION:
none
similar
identical
E
PA-C LIKE DOMAIN
α13
α24
5′ vRNA
α25
R287
Q291
U6
E650
K1228
K956
FINGERTIPS
T954
K1227
K673
Q952
R958
β22
K768
α30
K1063
β25
PALM
FINGERS
F
α66
THUMB RING
LID
S1701
H1621
K1705
K1700
α57
E1703
α63
K1474
α56
S840
R824
K983
N826
K979
BRIDGE
R976
R820
R829
N823
α32
H814
α35
FINGERS

# Table S1

| | Native_1 | SeMet_Pk | Ta6Br12_Pk | Ta6Br12_Ip | Pt_Pk | Native_2 | 5′ 10 nt soak |
|---|---|---|---|---|---|---|---|
| **Data collection** | | | | | | | |
| Space Group | $P2_12_12_1$ | $P\ 2_12_12_1$ | $P\ 2_12_12_1$ | $P\ 2_12_12_1$ | $P\ 2_12_12_1$ | $P\ 2_12_12_1$ | $P\ 2_12_12_1$ |
| Cell dimensions | | | | | | | |
| a (Å) | 102.09 | 102.39 | 102.25 | 103.00 | 104.93 | 102.0 | 103.00 |
| b (Å) | 140.90 | 141.07 | 140.95 | 141.00 | 141.25 | 140.7 | 141.10 |
| c (Å) | 162.54 | 163.09 | 161.97 | 163.80 | 163.49 | 162.4 | 165.50 |
| α ($^0$) | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.0 | 90.00 |
| β ($^0$) | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.0 | 90.00 |
| γ ($^0$) | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 | 90.0 | 90.00 |
| | | | | | | | |
| Resolution (Å) | 50 – 2.8 | 50 - 3.49 | 50-4.2 | 50-4.4 | 50 – 4.8 | 50-2.57 | 50-3.0 |
| (last shell) | (2.9-2.8) | (3.58-3.49) | (4.3-4.2) | (4.5-4.4) | (4.9-4.8) | (2.6-2.57) | (3.12-3.0) |
| Beamline (ESRF) | ID29 | ID29 | ID29 | ID29 | ID29 | ID23_1 | ID23_1 |
| Wavelength (Å) | 1.25425 | 0.97908 | 1.25438 | 1.25484 | 1.07156 | 0.99187 | 0.97924 |
| Completeness (%) | 99.1 (97.4) | 96.9 (95.8) | 99,8 (99,8) | 98.5 (99.3) | 93.2 (93.9) | 94.7 (97.4) | 99.5 (99.5) |
| Redundancy | 3.35 (3.28) | 6.94 (6.62) | 13.91 (14.34) | 6.6 (6.66) | 2.3 (2.01) | 3.55 (3.56) | 4.32 (4.52) |
| R-merge | 7.9 (66.5) | 25.4 (74.0) | 23 (90.2) | 28.1 (89.9) | 7.0 (57.4) | 8.0 (90.6) | 9.2 (82.1) |
| I/σI | 11.37 (1.93) | 7.48 (2.63) | 10.8 (3.60) | 6.36 (2.56) | 8.11 (2.05) | 11.2 (1.28) | 11.91 (1.55) |
| No. of sites | - | 53 | 16 | 13 | 10 | - | - |
| Phasing power acen | - | 7.547 | 3.407 | 3.076 | 1.712 | - | - |
| | | | | | | | |
| **Refinement** | | | | | | | |
| Resolution (Å) | | | | | | 48.65-2.57 | 49.2–3.0 |
| No. reflections | | | | | | 71088 | 46142 |
| $R_{work}/R_{free}$ | | | | | | 20.38/25.23 | 19.23/25.24 |
| | | | | | | (36.6/42.4) | (35.6/36.9) |
| **No. residues** | | | | | | | |
| Protein | | | | | | 1657 | 1666 |
| RNA | | | | | | 24 | 34 |
| **No. atoms** | | | | | | | |
| Protein | | | | | | 13571 | 13618 |
| RNA | | | | | | 500 | 716 |
| Water | | | | | | 92 | - |
| **B-factors (Å$^2$)** | | | | | | | |
| Overall | | | | | | 69.6 | 89.5 |
| Protein | | | | | | 67.2 | 90.7 |
| RNA | | | | | | 59.2 | 76.9 |
| Water | | | | | | 59.2 | - |
| **Ramachandran (%)** | | | | | | | |
| Outliers | | | | | | 0.1 | 0.3 |
| Favoured | | | | | | 96.19 | 95.14 |
| RMS deviations | | | | | | | |
| Bond lengths (Å) | | | | | | 0.003 | 0.007 |
| Bond angles ($^0$) | | | | | | 0.634 | 1.148 |

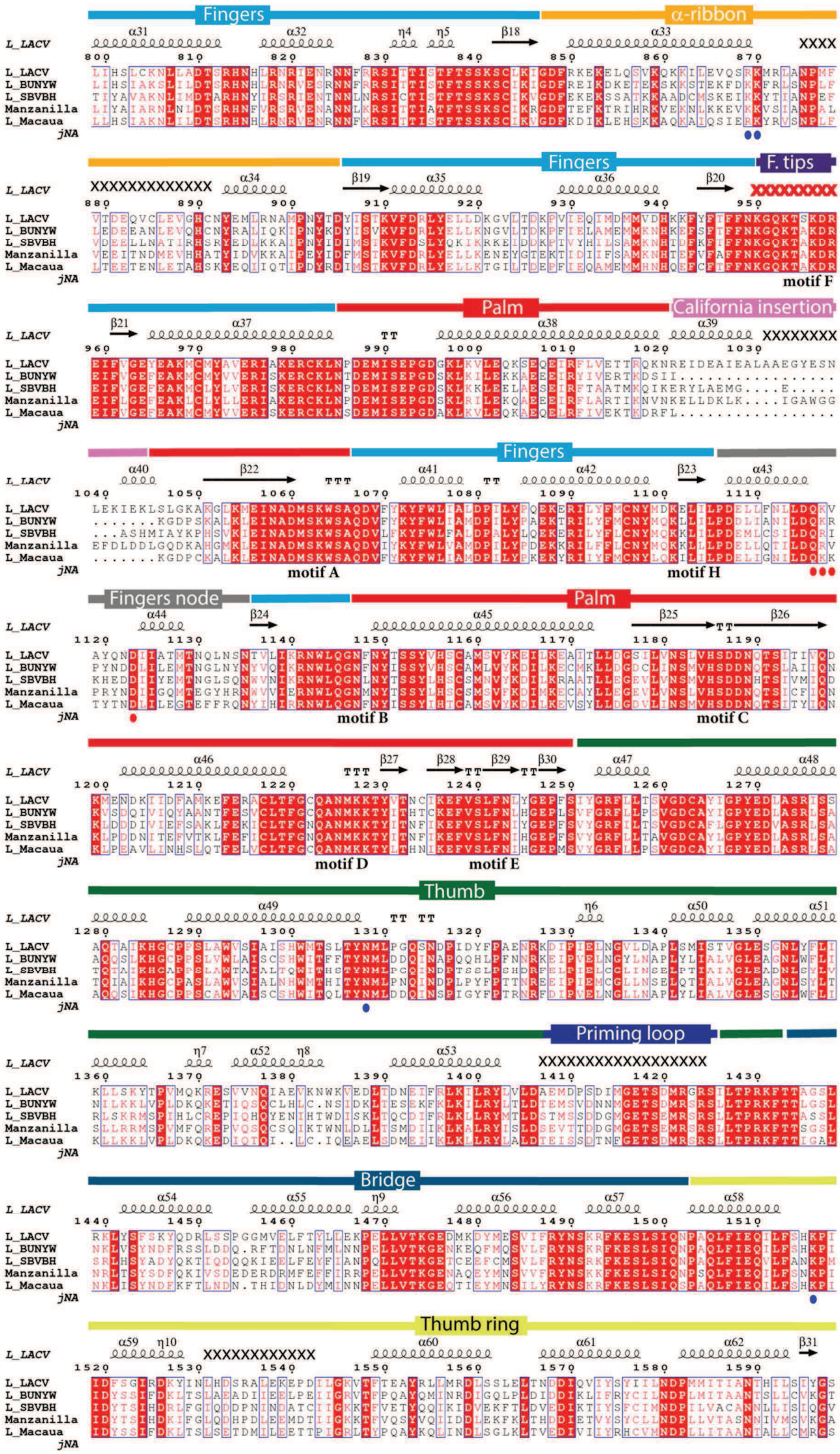
# Table S2

A

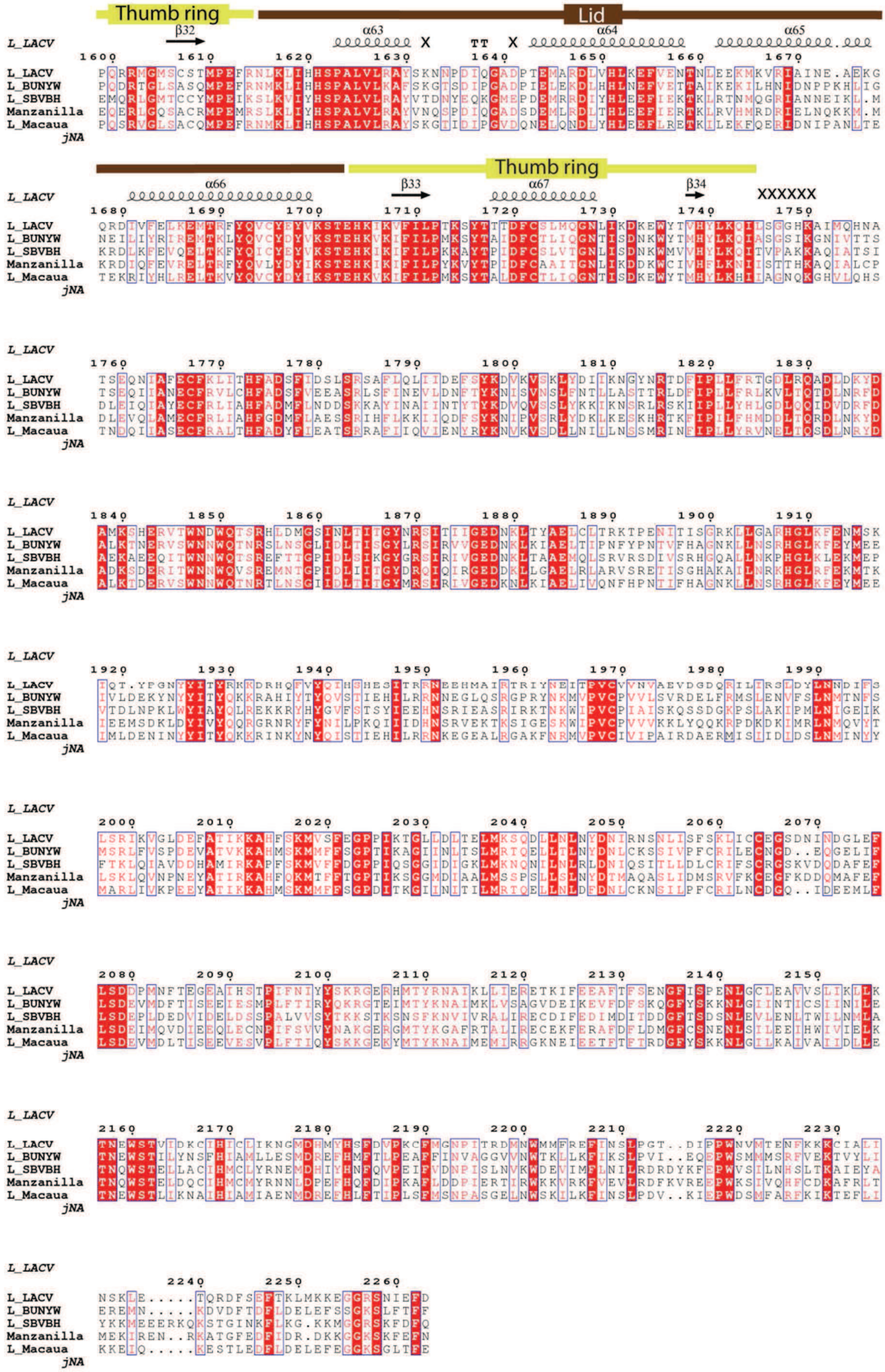


L_LACV
L_BUNYW
L_SBVBH
Manzanilla
L_Macaua
jNA
Endonuclease
Linker
PAC- Like, Core lobe
PAC- Like, vRBL
Clamp
Arch
motif G
Fingers

B


Fingers
α-ribbon
F. tips
Palm
California insertion
Fingers node
Thumb
Priming loop
Bridge
Thumb ring
L_LACV
L_BUNYW
L_SBVBH
Manzanilla
L_Macaua
jNA
motif F
motif A
motif H
motif B
motif C
motif D
motif E

C



# Table S3

## 3′ vRNA

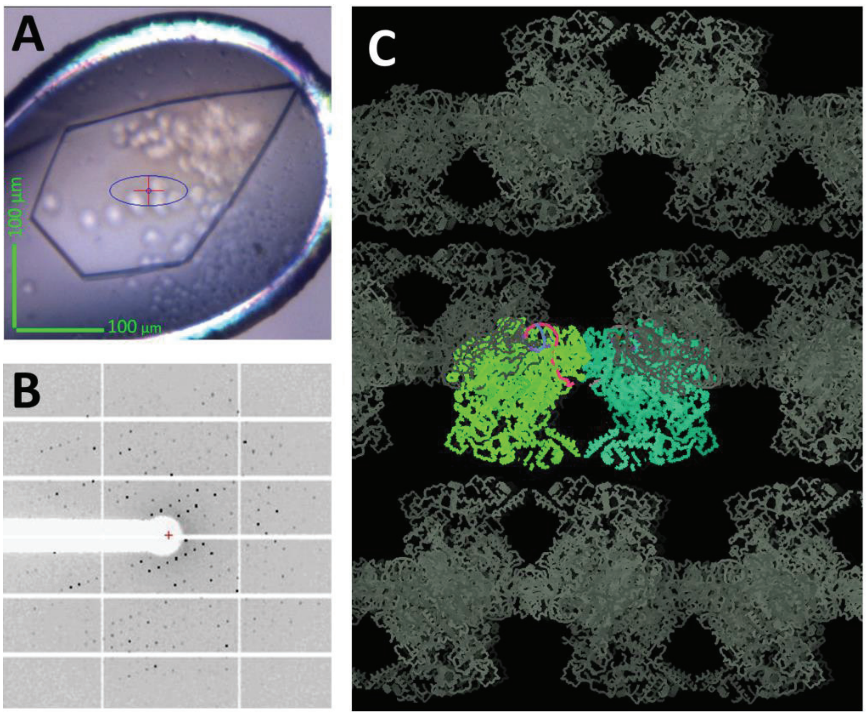
| RNA base | | Atom | | Residue | | Atom | | Distance (Å) | |
|---|---|---|---|---|---|---|---|---|---|
| G | 14B | OP2 | ... | Lys | 862A | NZ | ... | 2.59 | *** |
| G | 10B | O2' | ... | Lys | 381A | N | ... | 2.85 | *** |
| U | 9B | OP1 | ... | Leu | 383A | N | ... | 2.94 | *** |
| | | | ... | Arg | 526A | NH2 | ... | 2.88 | *** |
| U | 9B | O4 | ... | Lys | 870A | NZ | ... | 3.13 | *** |
| A | 8B | OP1 | ... | Arg | 526A | NH2 | ... | 3.06 | *** |
| A | 8B | O2' | ... | Ile | 378A | N | ... | 3.22 | *** |
| A | 8B | N7 | ... | Lys | 381A | O | ... | 3.42 | * |
| A | 8B | N6 | ... | Lys | 381A | O | ... | 2.79 | *** |
| C | 7B | OP1 | ... | Lys | 1516A | NZ | ... | 2.39 | *** |
| C | 7B | O2' | ... | Trp | 395A | NE1 | ... | 3.25 | *** |
| C | 7B | N4 | ... | Gln | 398A | NE2 | ... | 3.11 | *** |
| C | 7B | N3 | ... | Arg | 531A | NH1 | ... | 2.86 | *** |
| C | 7B | O2 | ... | Arg | 531A | NH2 | ... | 3.42 | * |
| | | | ... | Arg | 531A | NH1 | ... | 3.06 | *** |
| C | 7B | OP2 | ... | Tyr | 524A | OH | ... | 2.67 | *** |
| A | 6B | OP1 | ... | Arg | 372A | NE | ... | 3.07 | *** |
| | | | ... | Arg | 372A | NH2 | ... | 3.41 | * |
| A | 6B | OP2 | ... | Arg | 372A | NH2 | ... | 2.86 | *** |
| A | 6B | O5' | ... | Arg | 372A | NH2 | ... | 3.49 | * |
| A | 6B | O3' | ... | Asn | 1308A | ND2 | ... | 3.11 | *** |
| A | 6B | O2' | ... | Asn | 1308A | OD1 | ... | 3.52 | * |
| | | | ... | Asn | 1308A | ND2 | ... | 3.46 | * |
| A | 6B | N3 | ... | Ser | 520A | OG | ... | 3.51 | * |
| A | 6B | N6 | ... | Gln | 397A | OE1 | ... | 3.60 | * |
| C | 5B | OP1 | ... | His | 1515A | NE2 | ... | 3.35 | * |
| C | 5B | N4 | ... | Glu | 396A | OE1 | ... | 3.26 | *** |
| | | | ... | Gln | 397A | NE2 | ... | 3.17 | *** |
| C | 5B | N3 | ... | Gln | 397A | NE2 | ... | 3.58 | * |
| C | 5B | O2 | ... | Asn | 517A | ND2 | ... | 2.91 | *** |
| U | 4B | O4 | ... | Lys | 368A | NZ | ... | 3.49 | * |
| A | 3B | O2' | ... | Asn | 318A | ND2 | ... | 3.03 | *** |
| A | 3B | N7 | ... | Lys | 323A | NZ | ... | 3.03 | *** |
| A | 3B | N6 | ... | Cys | 535A | N | ... | 3.29 | *** |
| C | 2B | N4 | ... | Lys | 472A | O | ... | 2.83 | *** |
| C | 2B | N3 | ... | Lys | 472A | N | ... | 3.41 | * |
| | | | ... | Lys | 472A | O | ... | 3.42 | * |
| U | 1B | O3' | ... | Pro | 314A | N | ... | 3.45 | * |
| | | | ... | Asn | 538A | ND2 | ... | 2.71 | *** |
| | | | ... | Asn | 313A | O | ... | 3.18 | *** |
| U | 1B | O2' | ... | Asn | 318A | OD1 | ... | 3.53 | * |
| | | | ... | Asn | 538A | N | ... | 2.99 | *** |
| | | | ... | Asn | 538A | ND2 | ... | 2.85 | *** |
| | | | ... | Ala | 536A | O | ... | 2.75 | *** |
| U | 1B | N1 | ... | Ala | 536A | O | ... | 3.19 | *** |
| U | 1B | N3 | ... | Ala | 536A | O | ... | 3.57 | * |
| U | 1B | O2 | ... | Ala | 536A | O | ... | 3.55 | * |
| | | | ... | His | 312A | ND1 | ... | 3.56 | * |
| | | | ... | Gln | 506A | NE2 | ... | 3.03 | *** |
| U | 1B | OP2 | ... | Asn | 318A | ND2 | ... | 3.40 | * |

## 5′ vRNA

| RNA base | | Atom | | Residue | | Atom | | Distance (Å) | |
|---|---|---|---|---|---|---|---|---|---|
| A | 1C | OP2 | ... | Lys | 423A | NZ | ... | 3.20 | *** |
| A | 1C | O2' | ... | Ala | 593A | N | ... | 3.09 | *** |
| | | | ... | Ala | 593A | O | ... | 2.63 | *** |
| A | 1C | N3 | ... | Cys | 419A | O | ... | 3.51 | * |
| A | 1C | N1 | ... | Cys | 419A | O | ... | 3.21 | *** |
| | | | ... | Arg | 595A | NE | ... | 2.95 | *** |
| G | 2C | OP1 | ... | Arg | 592A | NH1 | ... | 2.69 | *** |
| G | 2C | O2' | ... | Arg | 595A | N | ... | 2.65 | *** |
| | | | ... | Arg | 595A | O | ... | 2.78 | *** |
| G | 2C | O4' | ... | Ala | 593A | O | ... | 3.14 | *** |
| G | 2C | OP2 | ... | Lys | 302A | NZ | ... | 3.34 | * |
| U | 3C | O3' | ... | Arg | 600A | NH2 | ... | 3.02 | *** |
| | | | ... | Thr | 642A | OG1 | ... | 3.16 | *** |
| U | 3C | O4' | ... | Arg | 595A | O | ... | 3.48 | * |
| A | 4C | OP1 | ... | Arg | 600A | NH2 | ... | 2.82 | *** |
| | | | ... | Thr | 642A | OG1 | ... | 3.00 | *** |
| | | | ... | Lys | 643A | N | ... | 2.96 | *** |
| A | 4C | OP2 | ... | Thr | 642A | OG1 | ... | 3.13 | *** |
| A | 4C | O2' | ... | His | 761A | ND1 | ... | 2.68 | *** |
| G | 5C | OP1 | ... | Lys | 643A | NZ | ... | 2.86 | *** |
| G | 5C | O5' | ... | Tyr | 677A | OH | ... | 3.14 | *** |
| G | 5C | O4' | ... | His | 761A | NE2 | ... | 3.49 | * |
| G | 5C | N7 | ... | Tyr | 677A | OH | ... | 2.94 | *** |
| | | | ... | Lys | 679A | NZ | ... | 2.83 | *** |
| G | 5C | N2 | ... | Ser | 438A | O | ... | 3.53 | * |
| G | 5C | OP2 | ... | Tyr | 677A | OH | ... | 2.72 | *** |
| U | 6C | O5' | ... | Arg | 292A | NH2 | ... | 3.37 | * |
| U | 6C | O4' | ... | Gln | 291A | O | ... | 3.23 | *** |
| U | 6C | N1 | ... | Gln | 291A | O | ... | 3.60 | * |
| U | 6C | OP2 | ... | Arg | 292A | NH1 | ... | 3.49 | * |
| | | | ... | Arg | 292A | NH2 | ... | 3.26 | *** |
| G | 7C | N7 | ... | Tyr | 1120A | OH | ... | 3.36 | * |
| | | | ... | Lys | 768A | NZ | ... | 3.16 | *** |
| G | 7C | N3 | ... | Gln | 1116A | NE2 | ... | 3.02 | *** |
| G | 7C | N2 | ... | Gln | 1116A | OE1 | ... | 3.10 | *** |
| | | | ... | Gln | 1116A | NE2 | ... | 3.44 | * |
| | | | ... | Asp | 1123A | OD2 | ... | 2.79 | *** |
| G | 7C | N1 | ... | Asp | 1123A | OD2 | ... | 3.46 | * |
| | | | ... | Asp | 1123A | OD1 | ... | 3.22 | *** |
| G | 7C | OP2 | ... | Tyr | 1120A | OH | ... | 2.92 | *** |
| U | 8C | OP1 | ... | His | 761A | NE2 | ... | 3.07 | *** |
| U | 8C | N3 | ... | Val | 1118A | O | ... | 2.55 | *** |
| U | 8C | O2 | ... | Lys | 1117A | N | ... | 3.01 | *** |
| | | | ... | Val | 1118A | N | ... | 3.00 | *** |
| | | | ... | Val | 1118A | O | ... | 3.26 | *** |
| U | 8C | O4 | ... | Val | 1118A | O | ... | 3.47 | * |
| | | | ... | Tyr | 1120A | N | ... | 3.21 | *** |
| U | 8C | OP2 | ... | His | 760A | NE2 | ... | 2.61 | *** |
| G | 9C | O4' | ... | His | 760A | NE2 | ... | 3.13 | *** |

### 3.4.4 L-vRNA crystallization and data collection

Similarly to the $L_{1750}$ construct the full length L protein was expressed in Hi5 insect cells using baculovirus expression system. As described in 'Methods' full length L protein was purified using the protocol established for $L_{1750}$, although one significant modification was introduced. When the L protein sample was being purified for crystallization experiments the vRNA was added to the concentrated sample eluted from the heparin column with the high salt buffer (~1M NaCl). Dialysis of such mixture against the final low salt buffer led to L-vRNA complex formation and allowed to avoid protein precipitation. After final size exclusion chromatography L-vRNA complexes were concentrated up to ~5 mg/ml. Samples were supplemented with additional vRNA in order to saturate the protein prior to co-crystallization experiments.




D

| | Lprot-vRNA |
|---|---|
| **Data collection** | |
| Space Group | *C2* |
| Cell dimensions | |
| a (Å) | 366.7 |
| b (Å) | 144.1 |
| c (Å) | 230.7 |
| α (⁰) | 90.00 |
| β (⁰) | 115.9 |
| γ (⁰) | 90.00 |
| Resolution (Å) | 50 – 4.4 |
| (last shell) | (4.45-4.40) |
| Beamline (ESRF) | ID29 |
| Wavelength (Å) | 0.97239 |
| Completeness (%) | 99.2 (99.6) |
| Redundancy | 3.85 (4.03) |
| R-meas (%) | 17.2 (147.6) |
| I/σI | 6.54 (1.16) |
| **Refinement** | |
| Resolution (Å) | 50 – 4.4 |
| No. reflections | 65120 |
| $R_{work}/R_{free}$ | 33.43/38.05 |
| **B-factors (Å²)** | |
| Overall | 203.92 |
| **RMS deviations** | |
| Bond lengths (Å) | 0.008 |
| Bond angles (⁰) | 1.050 |

**Figure 3.20** L-vRNA crystals: A – a single plate-like L-vRNA crystal inside the cryo-loop during diffraction experiment at the beamline; B – example of the diffraction pattern; C – crystal packing revealing two L-vRNA molecules in the asymmetric unit and importance of helical C terminus in formation of the crystal contact; D – crystallographic statistics

As previously for $L_{1750}$-vRNA crystals, the L-vRNA crystals grew only in case of 16-nucleotide long, partially double-stranded 3' vRNA. Initial hits – dense, round precipitates – were obtained in 100 mM Tris buffer pH 8.0, 100 mM NaCl, and 8% PEG 4000. When manually reproduced in hanging drops they yielded a fast-growing, big, and thin hexagonal plates (Fig. 3.20.A).

Initial L-vRNA crystals were soft and fragile and if successfully handled and cryo-protected with 20% glycerol showed weak diffraction of ~8 Å. In order to improve the resolution 30% glycerol cryo-protectant was used and crystals were soaked in a stepwise manner in cryo-solutions of increasing glycerol concentration. Cryo-protected L-vRNA crystals were then subjected to diffraction experiments, using both helical collection strategy and maximum transmission of the ID29 beamline. The best crystal diffracted to 4.4 Å (Fig. 3.20.B).

### 3.4.5 L-vRNA structure solution and structural features

L-vRNA crystallized in the C2 space group (366.7 Å, 144.1 Å, 230.7 Å, 90°, 115.9°, 90°) (Fig. 3.20.D). Structure was solved by molecular replacement using the $L_{1750}$-vRNA as a model. There are two molecules of the L-vRNA in the asymmetric unit weakly connected via protein-protein interactions between the PA-C like domains and the emerging C termini. The same type of RNA-protein stacking interactions involving the blunt end of the vRNA, as observed in the $L_{1750}$-vRNA crystals, participates in crystal packing. The C termini of the L proteins interact with themselves not only within the asymmetric unit but also with the C termini from the L proteins of the neighbouring crystal layer, thus forming major crystal contacts (Fig. 3.20.C). Low resolution of the structure does not allow to build in the residues from the C terminus. However it is good enough to judge the secondary structure and the connectivity within the emerged C terminus, although some of its fragments could not be connected. The portion of the C terminus which appears in the L-vRNA C2 structure is solely α-helical. The only exception is a small β-strand, forming of which structures the loop 707-718 from the central region of the L protein. In the $L_{1750}$-vRNA structure there is no electron density for this loop, which suggests that it gets stabilised upon interaction with the C terminal region.

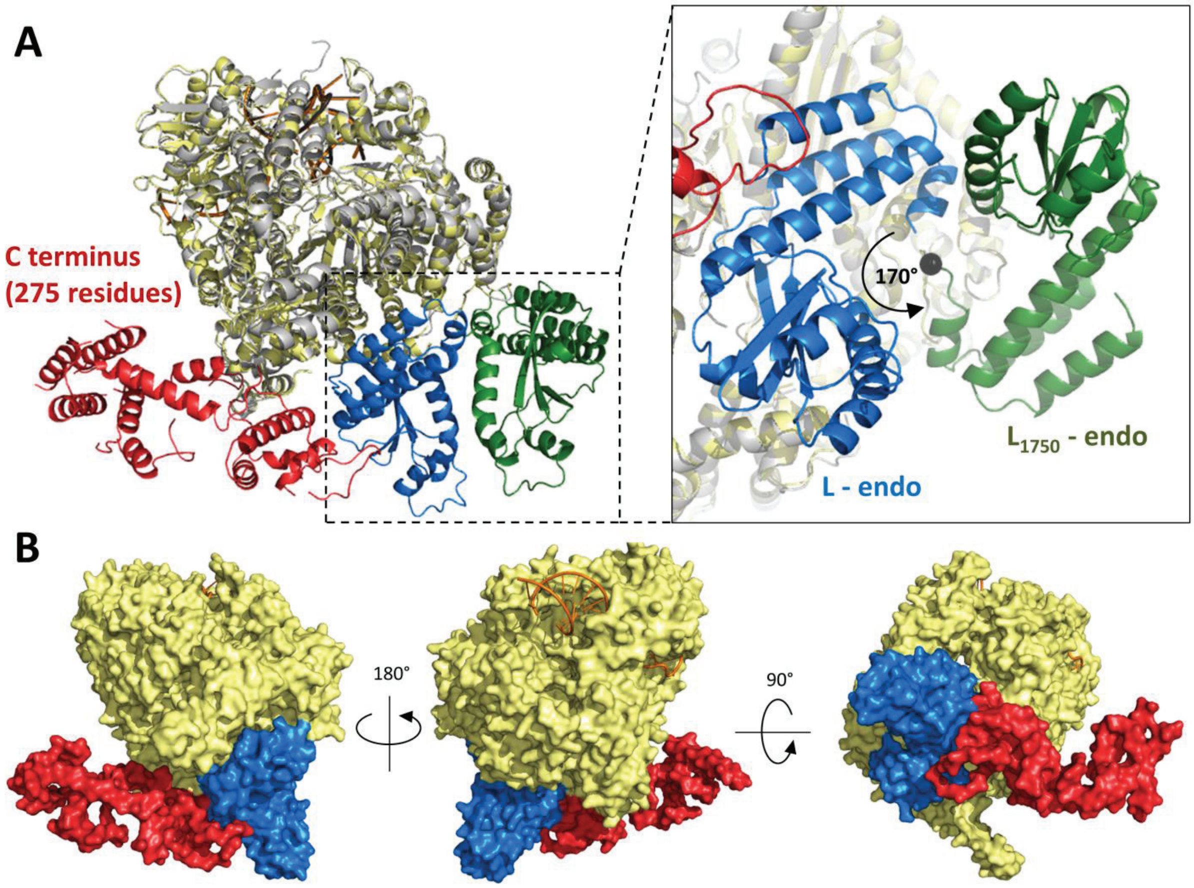



**Figure 3.21** L-vRNA structure: A – superposition of the $L_{1750}$-vRNA and L-vRNA structure; RNA strands are coloured black and orange in $L_{1750}$-vRNA and L-vRNA respectively; core unchanged part of the polymerase is coloured light grey and light yellow in $L_{1750}$ and L respectively; endonuclease is coloured green and blue in $L_{1750}$ and L respectively; α-helical C terminus is cloloured red; zoomed-in image highlights the difference in endonuclease position between $L_{1750}$ and L structures. B – overview of the L-vRNA structure; L protein is coloured as in A.

Out of 513 residues, missing in the $L_{1750}$ construct, the L protein C terminus α-helices allow to place around 275 alanines (Fig. 3.21.A-B). Although there is enough space, given the very high solvent content and the extensive free areas around the emerged C terminus, yet there is no electron density for the remaining ~240 residues (Fig. 3.20.C).

Given the electron microscopy data obtained by Dr. Hélène Mallet for both the $L_{1750}$ (Gerlach et al., 2015) and the full length L protein (not published), and the protein-protein interaction experiments with the $L_{1750}$ and the L7 constructs (see 3.2.4), we know that the C terminus does not maintain stable contacts with the body of the protein. It rather hangs flexibly, and is able to adopt multiple conformations. This flexibility is probably affecting the crystal packing, resulting in the lack of density for certain regions, high overall B factors, and poor diffraction. To stabilize the C terminus whether by boosting crystal contacts or by blocking it in a defined conformation, will be a critical point to improve the diffraction and to obtain the structure of this region.

As far as we can see from the current L-vRNA model, although some structural similarities can be noticed between the emerged C terminus and the Influenza virus PB2 mid domain, we do not see any structural similarity with the cap-binding domain. Whether bunyavirus L protein contains a cap-binding domain remains an open question. However, cap-binding experiments described below (see 3.5.1) suggest that this activity can be located at the very C terminus of the protein, not visible in the L-vRNA C2 structure.

Superposition of the L-vRNA and the $L_{1750}$-vRNA unravels another striking feature. The position of the endonuclease domain differs significantly between those two structures. In the L-vRNA structure it is rotated 170° comparing to the $L_{1750}$-vRNA structure. The pivotal point is located within the flexible linker between the endonuclease domain and the helical linker – namely the Gly185 (Fig. 3.21.A). Interestingly in the L-vRNA structure the endonuclease domain directly contacts the emerged C terminus, and more specifically its part corresponding to the mid domain of Influenza PB2 protein. In this position the endonuclease domain almost completely blocks the product exit tunnel. However its active site is oriented in the opposite way, so the emerging nascent RNA strand may not reach it. It is worth noticing that in both $L_{1750}$-vRNA and L-vRNA structures the position of the endonuclease domain is not forced by the crystal packing. Together with the cryo-EM data, this evidences that the endonuclease domain has certain freedom to move, and could be oriented differently at various polymerase functional states.

## 3.5 LACV full length L protein and $L_{1750}$ construct – functional studies

### 3.5.1 L and $L_{1750}$ – cap-binding assays

Following unsuccessful trials to identify the cap-binding activity within the C terminus of the L protein by means of 7mGTP-sepharose binding, EMSA assays were performed with radioactively labelled capped RNA. Following protein samples were tested: the 260 kDa full length L, the 200 kDa $L_{1750}$ (the L9 construct), and the L21 construct (~230 kDa ΔC-terminus). Synthetic RNA oligo bearing α-globin mRNA sequence and 5'ppp modification was radioactively labelled by vaccinia capping enzyme (VCE) within the cap structure. The polyU RNA was used in all binding assays in order to minimize the unspecific protein-RNA interactions. Experiments were performed without $Mg^{2+}$ to avoid RNA degradation by the N-terminal endonuclease. Knowing that L protein specifically binds 3' and 5' vRNA ends their impact on cap-binding activity was also tested. The cap-binding complex (CBC) was used as a positive control.

Results obtained indicate that the 5' vRNA enhances the cap-binding if compared with the apo protein. In contrast the 3' vRNA decreases the cap-binding activity. Despite absence of the Mg2+ ions in the buffer, degradation of the radiolabelled capped-RNA was observed. This might be due to residual amounts of magnesium ions which remained bound to the protein during purification procedure. Worth noticing is the significant increase in the capped-RNA cleavage in presence of the 3' vRNA (Fig. 3.22.A). Similar experiments using endonuclease inhibitor (DPBA) or the H34K mutant of the L protein – that inactivates the endonuclease – were performed (not shown). Another experiment comparing putative cap-binding activity of the full length L protein, the L21 construct and the $L_{1750}$ (the L9 construct) strongly suggests that the cap-binding region lies within the very C terminus of the polymerase or that at least its presence is required to activate the cap-binding. The LACV N protein showed no capped-RNA binding activity (Fig. 3.22.B).

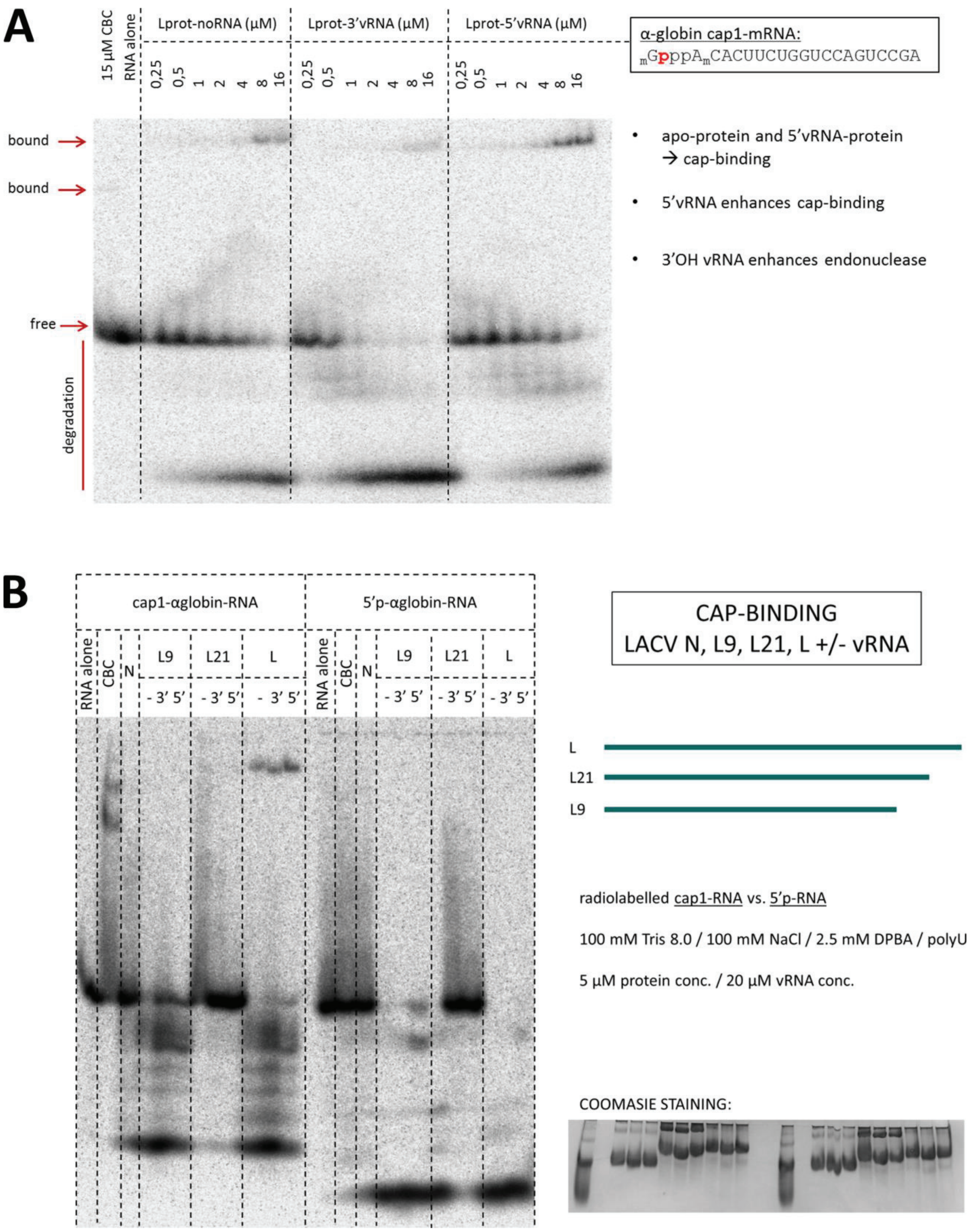



**Figure 3.22** L and $L_{1750}$ – cap-binding assays. A) EMSA assay with radioactively labelled capped RNA and L protein as apo form or in complex with 3′ or 5′ vRNA; polyU RNA was used to minimize unspecific RNA-protein interactions; samples in buffer without Mg2+ were incubated ON in RT and resolved on 8-20% PAGE (TG). B) similar EMSA experiment but including full length L protein, $L_{1750}$ (L9), and L21 construct; gel was stained with Coomasie (bottom right panel) in order to check whther all the proteins entered the gel. In both experiments (Fig. A and B) CBC was used as the positive control.

It has to be emphasized that cap-binding EMSA assays were performed many times and proved to vary depending on the batch of protein or incubation times used. Additional experiments are envisaged for the future in order to increase the binding signal and to reach more reproducibility. This will hopefully allow to confirm our conclusions about the putative cap-binding activity of the L protein.

### 3.5.2 Polymerase activity assays – transcription

In order to check whether recombinant LACV polymerase can perform RNA transcription in vitro a set of reactions have been designed (Fig. 3.23.A). Radioactively labelled capped α-globin mRNA, was used as a substrate for cap-snatching. The 25 nucleotide-long 3′ vRNA was used as the template for transcription in all the reactions. The polyU RNA was included in all the reactions in order to minimize unspecific protein-RNA interactions. The following negative controls reactions were performed – with EDTA, with DPBA (the endonuclease inhibitor), and with the L protein H34K endonuclease mutant. Additionally some reactions were implemented with the 25 nucleotide-long 5′ vRNA either added separately, or pre-mixed together with the 3′ vRNA. The influence of N protein on transcription was checked. It has been also tested whether the $L_{1750}$ (L9) maintains transcriptase activity.

The results obtained do not show strong evidences for the L protein transcription activity in vitro. Instead, two main activities observed are the endonuclease-mediated capped RNA degradation and the nucleotide extension of the capped RNA, the latter being template-independent and not random in terms of nucleotide addition. As observed in the control lines, the endonuclease activity can be inhibited in presence of EDTA or DPBA. Moreover, the L protein with H34K inactivated endonuclease does not degrade capped mRNA either. As observed previously during cap-binding EMSA assays, binding of the 3′ vRNA increases slightly the endonuclease activity. However, in contrast the cap-binding EMSA assays, the $L_{1750}$ (L9) does not exhibit any endonuclease activity. Presence of nucleoprotein had no significant effect on L protein activities.

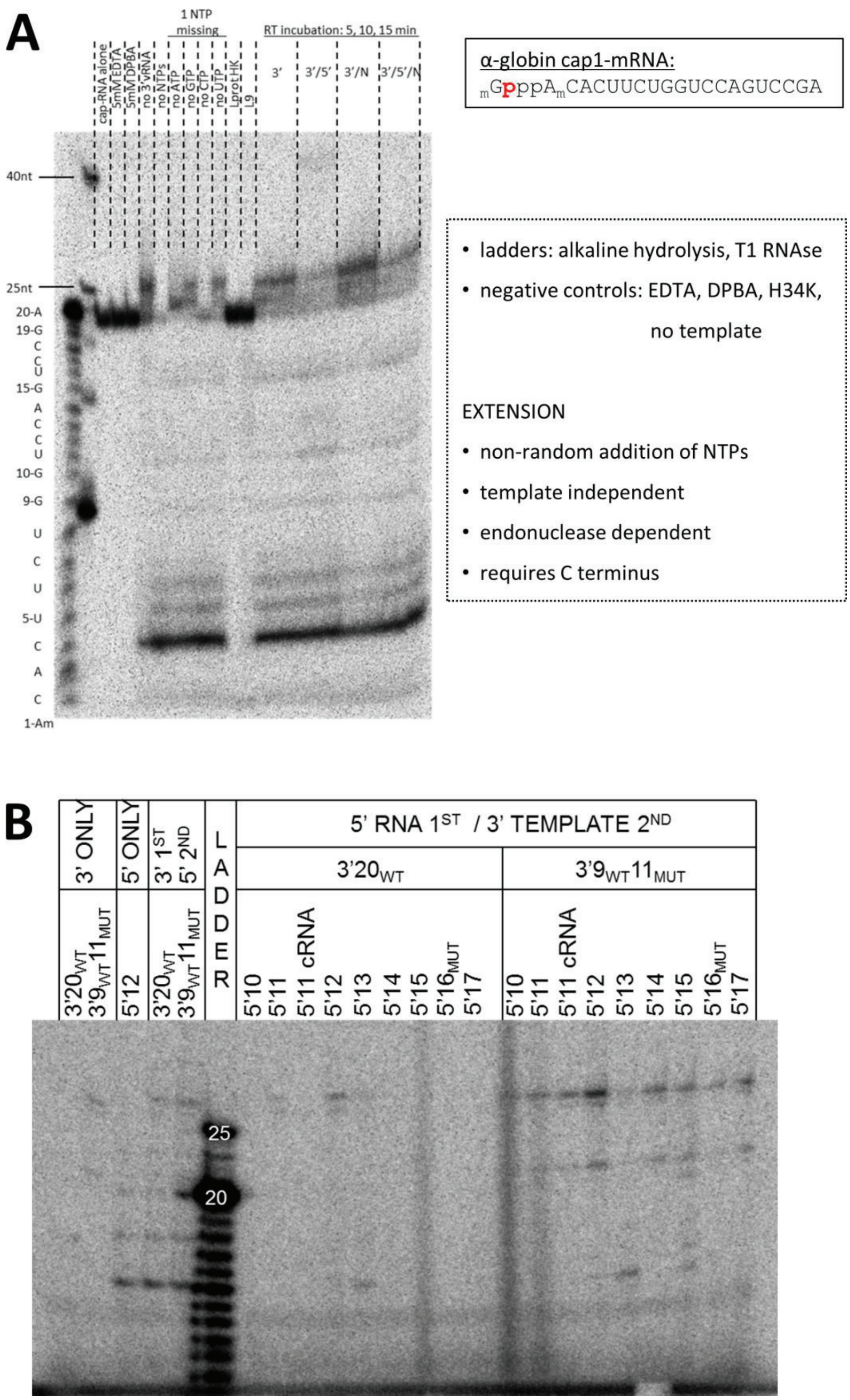



**Figure 3.23** Polymerase activity assays – transcription and replication. A) Transcritption assay using radiolabelled α-globin RNA; 5 µM L protein/$L_{1750}$/N protein; 10 µM vRNA, buffer contains 8 mM $Mg^{2+}$; two first lines – ladders generated from the input mRNA by alkaline hydrolysis and T1 RNAse digestion. B) Replication assay using $^{32}$P-UTP; $3'20_{WT}$ or $3'9_{WT}11_{MUT}$ were used as templates; 0.5 µM $L_{1750}$/vRNA, buffer contains 5 mM $Mg^{2+}$.

The transcription product of the expected size range – corresponding to the properly cap-snatched primer (~15 nucleotides) extended in a template-dependent manner (+25 nucleotide) into a final length of ~40 nucleotides – could only be weakly observed in reactions containing both 3′ and 5′ vRNA. This would suggest that 5′ vRNA apart of enhancing cap-binding is also required during transcription. However, this result does not reflect robust enzymatic activity of the L protein if compared with influenza virus polymerase (Reich et al., 2014).

The main product of L protein activity remains enigmatic. It appears to be a template-independent, non-random capped-RNA extension. Its synthesis is not modulated by addition of the vRNA. It affects capped-RNA but not 3′ vRNA (not shown). What is even more confusing is that it can be completely shut down by the single H34K mutation inactivating cap-snatching endonuclease. Transcription reactions with one out of four nucleotides omitted showed that the sequence being built is not random and it most likely starts with a CTP followed by an ATP (Fig. 3.23.A).

### 3.5.3 Polymerase activity assays – replication

Replication assays were performed with either L protein or $L_{1750}$ using 20 nucleotide-long 3′ vRNA as a template. Reactions were performed with radioactive $^{32}$P-UTP and other non-radioactive nucleotides which, in presence of polymerase activity, should lead to the radioactive labelling of the nascent RNA strand. The efficiency of the reactions was always low making it difficult to reproduce the results and evidencing the need of forward optimization of the experiments. The resulting products did not correspond to the expected 20 nucleotide-long RNAs. In contrast, the controls reactions where only the 32P-UTP was used revealed signal corresponding to the 20 nucleotide-long RNA. This suggests that in conditions tested, like in case of transcription assays, L protein reveals nucleotide extension rather than template-dependent polymerase activity.

Some of the replication experiments were testing the impact of the 5′ vRNA on the L protein replicase activity. The L protein H34K endonuclease mutant was used to exclude the

RNA degradation. Reactions were designed based on previous observations suggesting that the 5′ vRNA might be important for various activities of the L protein (e.g. cap-binding, transcription). Moreover, as seen in the $L_{1750}$-vRNA structure, 5′ vRNA proved to have an allosteric effect on the catalytic site of the L protein. A set of 5′ vRNA oligos of different lengths was used, ranging from 10 to 17 and including the 5′ cRNA. The RNA oligos representing the 3′ vRNA template and the 5′ vRNA were added sequentially to the L protein. To ensure that the 3′ and 5′ vRNA ends will not anneal, modified 3′ vRNA template was also used, having initial 9 nucleotides WT and following 11 nucleotides mutated (Fig. 3.23.B).

The results highlight the role of the 5′ vRNA, especially when it is incubated with the L protein prior to addition of the 3′ vRNA template. An RNA product, approximately 27 nucleotide-long, can be observed when the 12 nucleotide-long 5′ vRNA is used (but not others). The output of the reaction is different when the 5′ vRNA is added after the template 3′ vRNA, or comparing to reactions when only the template 3′ vRNA was used. Other RNA products, the nature of which is uneasy to explain, can also be observed. For instance in the reaction with the 12 nucleotide-long 5′ vRNA but omitting the template 3′ vRNA a ladder of 14, 17, and 20 nucleotide-long RNAs appears.

### 3.5.4 Ribonucleoprotein assembly

Attempts were made to assemble the $L_{1750}$ or L protein with vRNA and N proteins, in order to obtain samples for both replication and transcription activity assays, and co-crystallization experiments. Given the tendency of N proteins to form tetramers on RNA molecules, and knowing that one N protein sequesters 11 nucleotides (Reguera et al., 2013), various 3′ and 5′ vRNA were designed, in order to allow simultaneous binding of a single $L_{1750}$/L protein and four N proteins. vRNA were either synthetic or T7 in vitro transcribed. Following initial 1-2 hours incubation of the $L_{1750}$ protein with the 3′ or 5′ vRNA, the N proteins were added and the mixture was let for longer incubation over few days at 4°C. The final assembly mixtures contained 1:1:4 ratio of $L_{1750}$, vRNA, and N protein at low µM concentrations.

Following assembly samples were analysed by the size exclusion chromatography and collected fractions resolved on gradient SDS-PAGE.

Control samples revealed that in absence of vRNA, $L_{1750}$ and N protieins do not form complexes and elute separately as monomeric proteins. N proteins in presence of the vRNA form indeed tetramers and elute between elution volumes of monomeric $L_{1750}$ and N protein. In all four assembly samples – 3'53, 3'56, 5'56, and 5'59 – a new gel filtration peak appears strongly suggesting that an RNP was formed (Fig. 3.24). Analysis of elution fractions on gradient SDS-PAGE confirms that in presence of vRNA a subset of N proteins co-elutes with $L_{1750}$. It is mostly visible in case of both 3'53 and 3'56 vRNA. In all four cases however a significant portion of the tetrameric N proteins remains unbound to the $L_{1750}$. In this type of experiment it is impossible to judge whether such behaviour of the tetrameric N proteins is due to the weak interaction of the $L_{1750}$ with the rest of the RNP, which would lead to dynamic assembly and disassembly resulting in certain equilibrium. Another, more pragmatic, explanation would be that prior to the size exclusion all N protein tetramers are bound to the $L_{1750}$, but during gel filtration samples become greatly diluted which leads to the dis-assembly. Whatever the explanation, it is evident that the RNP assembly is mediated by the vRNA, since the $L_{1750}$ and N proteins do not form the complex in its absence.

Assembled RNPs – samples before gel filtration – were subjected to the co-crystallization experiments. No crystals were obtained until now.

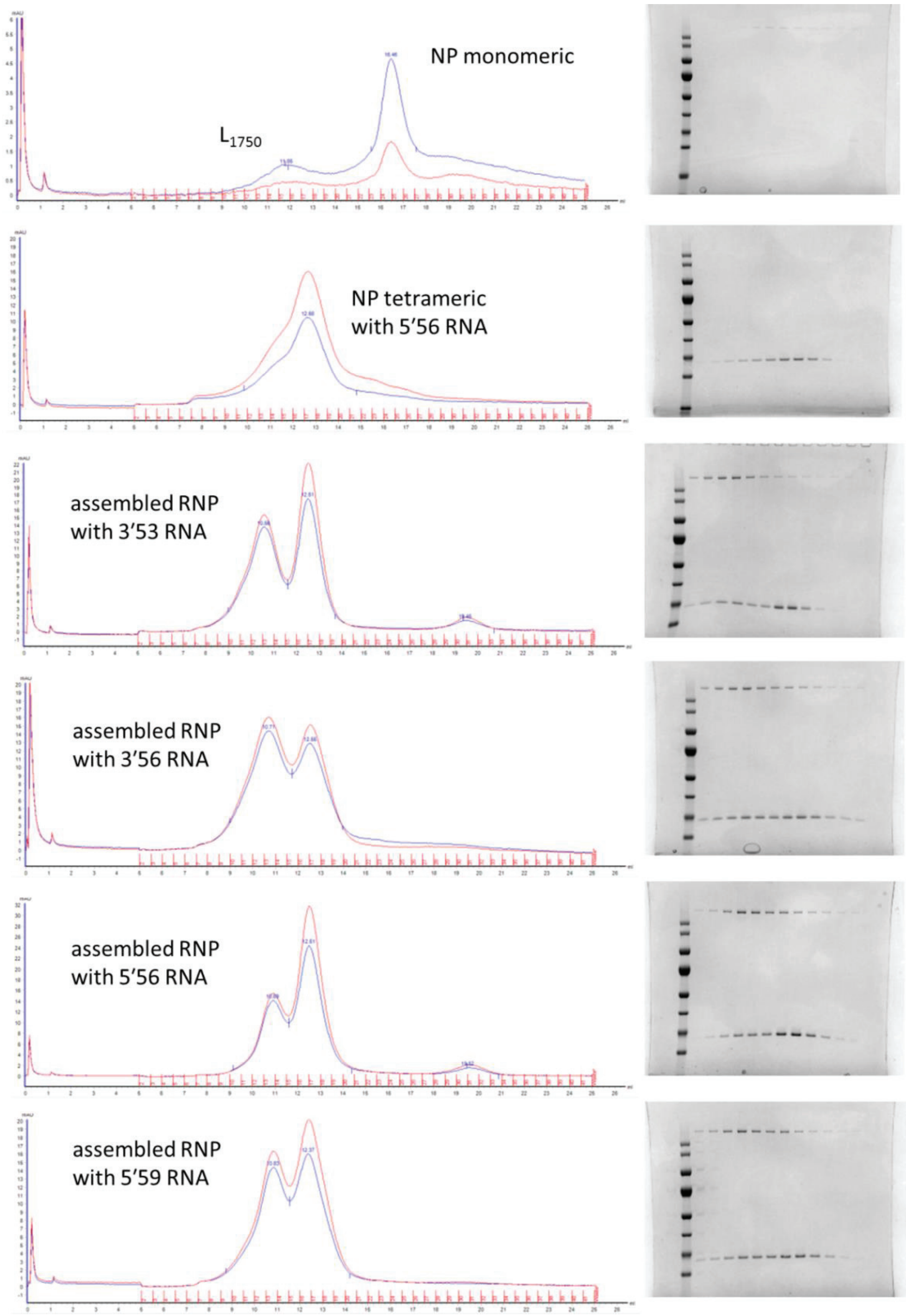



**Figure 3.24** RNP assembly of the L1750, N proteins and various 3' or 5' vRNA whose length should accommodate single L1750 and a tetramer of N proteins. RNP assembly was tested by gel filtration followed by SDS-PAGE. Two upper panels represent controls with either monomeric L1750 and N protein, or the N protein tetramer assembled on the RNA. All four RNP assembly experiments (bottom panels) reveal a distinct peaks representing the RNPs.

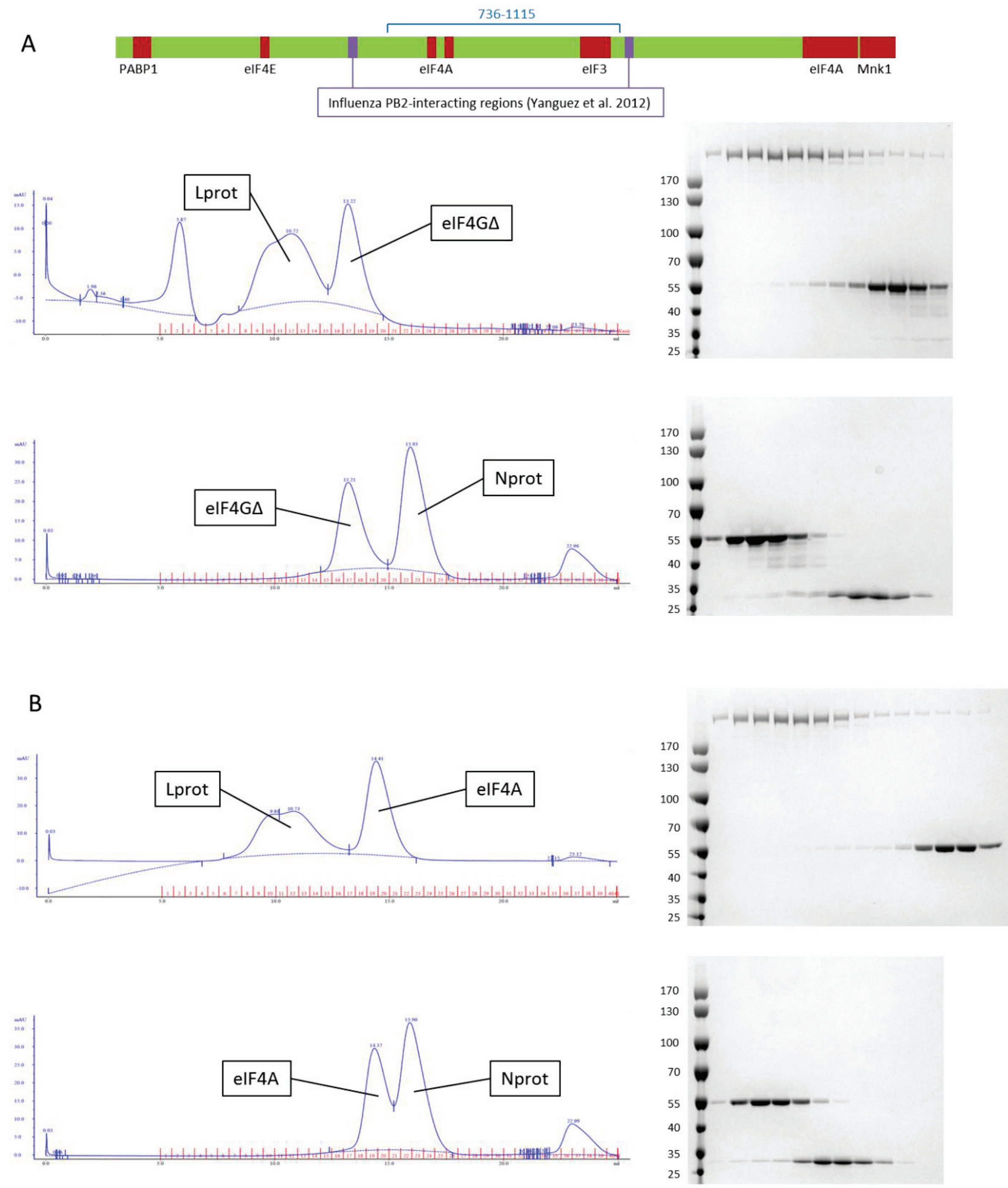



**Figure 3.25** Interaction with translation factors analysed by gel filtration followed by SDS-PAGE. A) No interactions were observed between L protein or N protein and the eIF4GΔ – an eIF4G construct lacking both N and C-terminal ends. B) No interactions were observed between L protein or N protein and the eIF4A helicase.

### 3.5.5 L protein interactions with host translation factors

In light of the studies showing that bunyavirus transcription requires ongoing translation (Barr, 2007; Bellocq and Kolakofsky, 1987), and proving that LACV infection affects the pool of some translation factors (Hodges and Connor, 2013), it was decided to study a potential direct interactions between the L protein or N protein and some of the subunits belonging to the eIF4F translation initiation complex.

The full length eIF4A helicase and the eIF4GΔ fragment – covering residues 736-1115, and lacking both N and C-terminal ends – were kindly provided by Dr. Boris Eliseev from Dr. Christiane Schaffitzel team. The eIF4G protein is a platform which binds various translation factors, including the eIf4A and the eIF4E, thus contributing to the eIF4F complex formation. The proteins were mixed in pair wise combinations and analyzed by gel filtration followed by SDS-PAGE (Fig. 3.25.AB). No interactions have been observed. However, it has to be emphasized that the eIF4GΔ fragment used in these assays does not comprise regions responsible for the interaction with the Influenza PB2 subunit (Yángüez et al., 2012).

# 4. DISCUSSION

Bunyavirus L protein is a multifunctional, 260 kDa single chain, RNA-dependent RNA polymerase able to both replicate and transcribe the negative-sense viral RNA genome. It initiates both processes bound to the pseudo-circularized, nucleoprotein-coated genomic segments via their conserved 3′ and 5′ extremities – the promoter. Lack of structural information of the bunyavirus polymerase, especially in complex with vRNA, makes the understanding of bunyaviral replication and transcription mechanics elusive and incomplete.

The aim of this PhD thesis project was to provide structural and functional insight into the mechanisms that drive sNSV polymerases. We hoped that studying the La Crosse virus L protein we will gain knowledge that could be extrapolated to other sNSV, comprising *Arenaviridae*, *Bunyaviridae*, and *Orthomyxoviridae*. The project was designed based on the initial assumption that *Bunyaviridae* LACV and *Orthomyxoviridae* Influenza virus polymerases share similar domain organization. In spite of the fact that, unlike the bunyavirus L protein, the Influenza virus polymerase is a heterotrimer composed of PA, PB1, and PB2 subunit, many features – the position of the endonuclease domains (Dias et al., 2009; Reguera et al., 2010), the position of the highly conserved RdRp motifs, the size of both polymerases, and the head-to-tail interactions between the Influenza polymerase subunits – were in favour for such hypothesis.

Focus of the 1st year of the thesis project was to confirm and characterize both structurally and functionally a putative cap-binding domain isolated from the LACV L protein. Based on the homologous Influenza virus cap-binding domain positioned in the polymerase PB2 subunit (Guilligay et al., 2008), we assumed that the LACV cap-binding domain is located within the C terminus of the L protein. Constructs were designed following the ESPRIT soluble domain screening within that region (Yumerefendi et al., 2010), followed by limited proteolysis and *in silico* prediction of the domain boundaries. Inspite of many protein constructs tested neither crystallization trials nor functional cap-binding assays were successful.

During the 2nd year the project has been broaden to use the LACV full length L protein and a more stable C-terminus truncated construct named the $L_{1750}$ and comprising 3/4 of the protein. Both constructs were subjected to numerous trials of crystallization and co-crystallization with vRNA, initially without any success. In parallel enzymatic activities and vRNA-binding studies were carried out *in vitro*. Replicase and transcriptase assays provided an unclear outcome suggesting that either the polymerase cannot synthesize RNA properly in experimental conditions tested, or it requires some additional factors. Cap-binding studies strongly suggested that the cap-binding activity may indeed be located within the C terminus of the protein, although without providing unambiguous evidence. Binding affinities of the genomic 3′ and 5′ vRNA ends to the L protein – analysed by various approaches including EMSA, ITC, and MST – revealed that the bunyavirus polymerase interacts strongly with both genomic ends as separate ssRNA and residually when they form dsRNA. Limited proteolysis of the 3′ or 5′ vRNA-protein complexes performed with trypsin revealed different patterns of cleavage, suggesting conformational rearrangement upon vRNA binding and/or direct masking of the trypsination sites by the vRNA molecules. The MS analysis and Edman degradation of the proteolytic fragments allowed to narrow down the region interacting with vRNA to the N-terminal part of the L protein (residues 368-430) and, surprisingly, to exclude interactions with the core RdRp domain.

During 3rd year of the project LACV L protein-vRNA binding studies were continued. They greatly helped to rationalize co-crystallization attempts, which finally resulted in the first bunyavirus polymerase crystals. The vRNA which made crystallization of the $L_{1750}$ possible is a partially double-stranded molecule with single-stranded part comprising 3′ genomic end. Optimized $L_{1750}$-vRNA crystals diffracted initially to 3.1 Å. Structure was solved with experimental phases from tantalum, platinum and selenomethionine derivative crystals. Resolution was further improved up to 2.6 Å.

Details of the $L_{1750}$-vRNA structure and conclusions that were made based on it are described in the publication attached to the ‘Results’ chapter of this thesis (Gerlach et al., 2015). Below major observations are recapitulated and some additional features – not included in the publication – are discussed in light of other results obtained during the PhD project.

To our knowledge the LACV $L_{1750}$-vRNA crystal structure is the first structure of a *Bunyaviridae* polymerase. It reveals for the first time detailed domain architecture of that multifunctional protein providing insight into many common RdRp elements and revealing new, specific structural features. Our structure covers ~3/4 of the full length L protein and represents the pre-initiation state of the polymerase. The solvent-exposed N-terminal endonuclease is followed by an elongated helical linker leading to a compact domain which forms one side of the molecule, and contributes to the vRNA binding. The central, right-handed RdRp domain contains both classical features – fingers, fingertips, palm, and thumb – as well as specific modules. The C-terminal domains close the circular architecture of the polymerase.

Apart from the structure of the LACV $L_{1750}$-vRNA complex the Cusack group obtained as well protein structures of the bat Influenza A and human Influenza B polymerase in complex with their viral promoters (Pflug et al., 2014; Reich et al., 2014). This provides unique chance to analyse and compare in detail the architecture of hitherto mysterious sNSV RNA dependent RNA polymerases. The major and straightforward conclusion is that bunyavirus monomeric and influenza virus heterotrimeric polymerases exhibit extremely high fold similarity. This fact greatly exceeds the initial hypothesis concerning domain organization similarity. In spite of almost none – except for some catalytically important motifs – sequence identity, the majority of secondary structures position is highly similar in both polymerases. This allows structural comparison of the LACV L protein N terminus, its core RdRp region, and the C-terminal region with the Influenza polymerase PA, PB1 and PB2 subunits. It also raises the question about the evolution of sNSV polymerases and suggests that *Bunyaviridae* (and possibly *Arenaviridae*) polymerases are evolutionary older and had to be split into 3 subunits in order to form the *Orthomyxoviridae* heterotrimeric polymerase.

Another major observation that can be made based on the $L_{1750}$-vRNA complex structure is the presence of two separate binding sites, specifically recognizing both 3' and 5' RNA genomic ends. Our *in vitro* $L_{1750}$-vRNA binding studies confirm that the L protein recognizes both genomic ends as single-stranded molecules via their highly conserved initial nucleotides. Trypsin cleavage sites within the L protein are protected upon incubation with vRNA which further confirms that indeed in solution both 3' and 5' RNA genomic ends bind to their specific sites observed in the crystal structure. The 3′ vRNA end is recognized by the $L_{1750}$ as a non-structured stretch of 8 nucleotides. Distant parts of the $L_{1750}$ protein are contributing to the 3′ vRNA binding groove formation, although the majority of interactions is mediated by the vRBL from the PA-C like domain. Noteworthy structural element is the clamp specifically interacting with nucleotides A6-C7-A8. The 5′ vRNA end forms a stem-loop containing 10-11 initial nucleotides. It is bound in the pocket formed by the fingernode emerging from the fingers domain and the vRBL with its elongated loop called the arch. Binding of the 5′ vRNA allosterically stabilizes the motif F – one of the conserved catalytic RdRp motifs.

Until now, given the extensive complementarity between bunyavirus RNA genomic ends – it was generally assumed that they interact with the L protein as dsRNA panhandle (Barr and Wertz, 2004; Kohl et al., 2004; Patterson et al., 1983; Raju and Kolakofsky, 1989). Our structure provides completely new image of bunyavirus polymerase-vRNA interactions and creates the need for revision of previously reported studies on bunyaviral promoters and polymerase activity. We propose that the terminal, conserved 11 nucleotides do not anneal during entire viral cycle. The only exception from that rule is when the L protein falls from the RNP creating the chance for the 3′ and 5′ ends to form a dsRNA which would be then recognized by the host-cell innate immune system receptors like RIG-I (Weber et al., 2013). The dsRNA formation observed in LACV vRNP and cRNP upon psoralen cross-linking (Raju and Kolakofsky, 1989) can be explained by the distal duplex formation, which was proven to be essential for the efficient RNA synthesis (Barr and Wertz, 2004). As discussed in more detail in the attached publication (Gerlach et al., 2015) it is possible that the 3′ end, while entering the polymerase active site through the template entry tunnel, forms such distal duplex with the 5′ end which terminal bases are bound by the L protein in the form of a stem loop. If this would be true, the terminal 3′-UCA repeat would overshoot the active site, and the replication would initiate at the second UCA repeat. This is consistent with the prime-and-realign mechanism, and would imply that the distal duplex formation stabilizes the initiation state of the polymerase. Interestingly, it was reported that the assembly of a functional *Arenaviridae* Machupo virus polymerase complex also requires single-stranded genomic RNA ends (Kranzusch et al., 2010).

Another intriguing feature of the bunyavirus polymerase is its capacity to switch between transcription and replication modes. It was reported that the G:U wobble pair at position 9 in the *Orthobunnyavirus* genomic RNA is critical in mediating this switch. Mutagenesis studies performed on the BUNV genomes revealed that while the 3′ vRNA 9-U promotes transcription, the 3′ cRNA 9-C promotes replication (Barr and Wertz, 2005). Our $L_{1750}$-vRNA structures with either 5′ vRNA or 5′ cRNA show no significant difference in the 5′ RNA binding and the way it allosterically regulates the active site. It remains to be structurally verified whether the 3′ vRNA and 3′ cRNA templates affect differently the polymerase residues within the template entry tunnel or the polymerase active site.

Analysis of the $L_{1750}$-vRNA structure revealed four tunnels leading into and from the inner polymerase cavity which likely serve the purposes of NTP entry, template entry, template exit, and product exit. Given the significant structural similarity among viral RdRp domains and the availability of the RdRp structures in complex with RNA, we were able to assign all four tunnels. Superposing the $L_{1750}$-vRNA structure with structures of the HCV or the Poliovirus polymerase in complex with the template-product RNA duplex (Gong and Peersen, 2010; Mosley et al., 2012) we noticed that the LACV polymerase product exit tunnel is too narrow to accommodate such RNA duplex. Poliovirus polymerase-RNA structure was particularly helpful since having long RNA duplex in the active site cavity – covering approximately a single complete double helix turn – it suggests that in LACV polymerase case such duplex undergoes separation, forced by the C-terminal region of the L protein, mostly the lid domain. Upon separation of both RNA strands the product exits, like in the Poliovirus polymerase case, through the product exit tunnel, and the template is directed towards another, distinct tunnel, which is located very close to the template entry tunnel. (Fig. 4.1.A). Comparing the distribution of the four tunnels from the LACV L protein with the four tunnels from the Qβ replicase-RNA structure (Takeshita and Tomita, 2012) ensured us that the fourth LACV L protein tunnel is for the template strand to exit the active site cavity (Fig. 4.1.B).

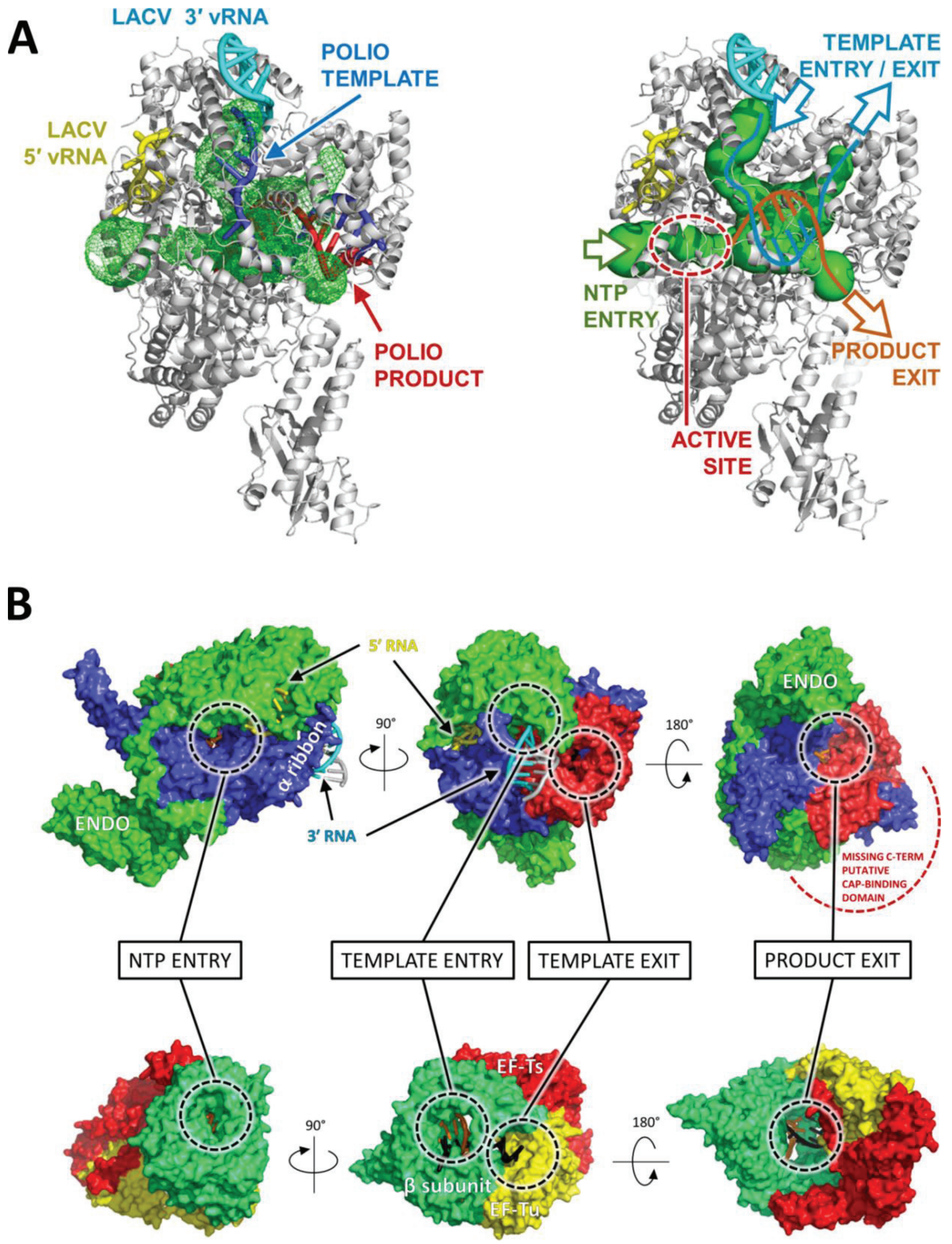



**Figure 4.1** Distribution of the four tunnels leading into and out from the internal cavity active site of the $L_{1750}$. A) Tunnels and internal cavity are represented as green mesh; LACV 3′ and 5′ vRNA are coloured in cyan and yellow respectively; Poliovirus polymerase was superposed onto the $L_{1750}$-vRNA structure (PDB: 3OLB)(Gong and Peersen, 2010); Poliovirus template and product RNA strands are coloured in blue and red respectively; Left panel proposes the RNA duplex separation inside the $L_{1750}$ and highlights RNA product and template separate exit tunnels. B) Comparison of the four tunnel distribution between the $L_{1750}$-vRNA structure and the Qβ replicase structure (PDB: 3AVY)(Takeshita and Tomita, 2012).

Not surprisingly all four tunnels are built by highly conserved and positively charged residues, able to accommodate the nucleic acids. Distribution of the tunnels, and especially both template and entry and exit tunnels, allowed us to propose the replication model, details of which are described in (Gerlach et al., 2015). In brief, the model proposes that, in the context of the RNP, the close proximity of the template tunnels allows immediate re-coating of the genomic RNA by nucleoproteins once it has passed through the polymerase active site and was used as a template. During a single replication cycle, when the L protein travels along the vRNP or cRNP templates, not a single nucleoprotein is lost or any additional needed. This is supported by the several findings from the crystal structures: (a) each nucleoprotein sequesters 11 nucleotides, (b) the L protein itself, accommodating both 3′ and 5′ genomic ends during the pre-initiation state, sequesters approximately 20 nucleotides, (c) modelling of the Poliovirus template RNA inside the template tunnels and the active site of the LACV L protein shows that the entire template RNA pathway can accommodate around 20 nucleotides. During RNA synthesis the template RNA would unwind from the nucleoprotein and would enter through the template entry tunnel. At the same time the template emerging from the template exit tunnel would bind to the very same RNA-free nucleoprotein. This is possible because the distance between the entry and the exit tunnels is not greater than the length of a single nucleoprotein. The L protein processing in such manner along the RNP would generate a shift of the nucleoprotein neckless in regards to the genomic RNA after each replication round. Proposed model explains how the RNP can preserve its structural integrity during transcription and replication, avoiding the N protein disassembly and minimally disrupting the RNP. This is consistent with previous observations in *Orthomixovirus* RNP studies that show consistently assembled RNPs during its replication (Arranz et al., 2012).

The replication model which we propose explains as well how the emerging nascent strand RNA is directly packaged into the cRNP or vRNP. It assumes that the second L protein binds to the processing L protein at the rear, product exit site. The first synthesized 5′ vRNA or cRNA is immediately bound as a hook structure in its binding pocket. The rest of the emerging nascent strand is coated by the N proteins while inactive L protein remains attached to the active one during entire replication cycle. At the replication termination the last synthesized 3′

end is immediately bound in the 3′ binding groove of the inactive L protein. Although dimerization between *Bunyaviridae* L protein has not been reported, there are studies suggesting that dimerization of the Influenza polymerase is required for replication (Arranz et al., 2012; York et al., 2013).

Initial hypothesis saying that all sNSV polymerases share the same domain architecture, confirmed strongly by comparison of the $L_{1750}$-vRNA structure with Influenza A and B polymerase structures (Pflug et al., 2014; Reich et al., 2014), justifies the assumption that the bunyavirus cap-binding domain is located within the C terminus of the L protein. The L protein-vRNA structure reveals low resolution electron density which allows to position approximately 250 out of 500 residues missing in the $L_{1750}$-vRNA structure, however with not enough resolution to see any side chains. There are no features that would resemble the Influenza cap-binding domain, or at least a β-strand like the one present in the cap-binding domain (Guilligay et al., 2008). The cap-binding EMSA assays performed with the radioactively labelled capped RNA and the full length L protein, the $L_{1750}$ construct, or the L21 230 kDa construct missing the very C terminus showed that the cap-binding activity may be indeed be located within the very C terminus of the L protein. However, it has to be emphasized that the EMSA cap-binding assays could not be robustly reproduced and many times showed inconsistent results. This can be explained by either a non-specific binding of the capped RNA, which occurred more via the backbone of the nucleic acid and not the via the cap moiety itself. However, it cannot be excluded that the cap-binding activity emerges only at certain conformation of the L protein, or requires an unknown host factor that would stabilize the active conformation of the cap-binding domain. Features revealed by the L protein-vRNA structure are in favour of such scenario. We observed that a region of the stabilized C terminus – which can be assigned as the homolog of the Influenza PB2 mid domain – interacts directly with the N-terminal endonuclease domain. Similarly, different part of the C terminus could either interact with other region of the L protein, the N protein, or with an unknown host factor. Although by gel filtration we did not observe such interactions between the isolated LACV L protein C terminus (the L7 construct) and the $L_{1750}$ or the N protein, it cannot be excluded that they would occur in the full length L protein context.

Orthobunyavirus L protein requires an ongoing translation to avoid premature transcription termination (Barr, 2007; Bellocq and Kolakofsky, 1987). As revealed by (Vialat and Bouloy, 1992) the smaller 40S ribosome subunit scanning along the viral mRNA is sufficient to ensure non-interrupted viral transcription. During host-cell translation the 40S subunit is recruited onto the mRNA by the eIF4F complex composed of the initiation factors – the scaffolding eIF4G, the eIF4A helicase, and the eIF4E cap-binding protein (Gross et al., 2003; Myasnikov et al., 2009). The ambiguities concerning bunyavirus L protein cap-binding domain, cytoplasmic location of the transcribing L proteins, and the requirement of the ongoing translation, open the possibility for a direct interaction between the L protein and the host translation initiation factors – like those forming the eIF4F complex. Interestingly, analysis of the host-cell proteins upon LACV infection demonstrate the effect on eIF2α phosphorylation and eIF4E dephosphorylation, resulting in global shut down of the host translation (Hodges and Connor, 2013). Studies on the *Orthomyxoviridae* Influenza virus translation revealed its independence from the cellular cap-binding factor eIF4E, proving in the same time direct interaction between the eIF4G and the influenza PB2 subunit, which suggests that this viral protein is able to replace cellular eIF4E factor (Yángüez et al., 2012). Similarly, it was reported that the *Arenaviridae* Junin virus translation is also independent from the cellular cap-binding factor eIF4E, and instead the N protein associates with the eIF4A and the eIF4G cellular factors (Linero et al., 2013). In light of these studies we decided to examine the possibility of the direct interaction between the LACV L protein and subunits of the cellular eIF4F complex. So far only the eIF4A and the middle region of the eIF4G were tested by gel filtration and showed no interaction. Following studies, including the eIF4E translation factor, are being performed.

Another solution to the cap-binding requirement comes from the studies on the *Hantavirus* HANV. It was proposed that the cap-binding activity resides within the HANV N protein which is thought to replace the entire eIF4F complex and actively participate in transcription both interacting with the L protein and providing the cap-snatched RNA oligos acquired from the pool of the host mRNA directed for decapping in P-bodies (Mir and Panganiban, 2008; Mir et al., 2008, 2010). In general it cannot be excluded that the N proteins from various sNSV viruses possess additional functions other than RNA-binding and genome encapsidation. A good example is the *Arenaviridae* Lassa virus N protein which has an exonuclease activity (Hastie et al., 2011; Qi et al., 2010). However, in case of the 50 kDa HANV N protein the lack of crystal structure makes it difficult to judge its functions with no ambiguity. In light of the LACV N protein crystal structure (Reguera et al., 2013) and our negative results from the EMSA capped-RNA binding studies it can be stated that the *Orthobunyavirus* N protein does not have the cap-binding activity. We also did not observe any N protein impact on transcription in our L protein activity assays. It remains to be tested whether RNP complexes assembled in vitro with the L protein, the N proteins, and approximately 55-nucleotide long 3' or 5' genomic RNA ends will exhibit replication or transcription activity.

# 5. ACKNOWLEDGEMENTS

*First of all I would like to say thank you to my PhD supervisor, Stephen Cusack, for giving me the opportunity to work on such a demanding, interesting, and rewarding project. For many years the structural architecture of the sNSV polymerases was a mystery. Being a PhD student in Stephen's group I had an enormous privilege to witness how this enigma is being solved. Firstly, the structures of both Influenza A and B virus polymerases were obtained, and few months later we got the first structure of the bunyavirus polymerase. It will not be an exaggeration to claim that these are real breakthroughs and will pave a way for new interesting studies in the sNSV field. I am truly grateful for having a chance to participate in this achievement. That was a great adventure!*

*I also want to say thank you to Juan Reguera who, together with Stephen, was responsible for the bunyavirus project. He led and supervised me during my entire stay in EMBL Grenoble. He taught me many things which are important in being a scientist, both from the practical, experimental design point of view, and from the theoretical one, which concerns asking the scientific questions and choosing right approaches in finding the answers. Being a stubborn student I did not make this task easy, but I have learned a lot, and I will be always grateful for that.*

*Many thanks to my PhD thesis defense jury members, Bruno Canard, Richard Elliott, Daniel Kolakofsky, and Winfried Weissenhorn for their time in revising and commenting this work. Thanks as well to my EMBL Thesis Advisory Committee, Matthias Hentze, Ramesh Pillai, and Winfried Weissenhorn for their help and fruitful discussions during my PhD project.*

*I would also like to thank all the people from the EMBL staff and various facilities without the help of whom all this work would not be possible – the EEF facility, the HTX crystallization platform, the ESRF beam line scientists, the MS platforms both in IBS Grenoble, and EMBL Heidelberg.*

*Great thanks go to the entire Cusack group. During these 3.5 years many members have left and many new ones came, but the atmosphere within the group was always great. Everybody was very friendly, willing to help, and open for scientific discussions. Wherever I will go next I will always keep the Cusack group spirit as a reference.*

*Finally I want to thank my family, my wife Agata, my parents and my younger sister. Without their love and support I would not be where I am today.*

P. Gerlach

*09.05.2015, Grenoble, France*

# 6. REFERENCES

Abbas, Z., and Afzal, R. (2013). Life cycle and pathogenesis of hepatitis D virus: A review. World J. Hepatol. *5*, 666–675.

Abbondanzieri, E.A., Bokinsky, G., Rausch, J.W., Zhang, J.X., Le Grice, S.F.J., and Zhuang, X. (2008). Dynamic binding orientations direct activity of HIV reverse transcriptase. Nature *453*, 184–189.

Abraham, G., and Pattnaik, A.K. (1983). Early RNA synthesis in Bunyamwera virus-infected cells. J. Gen. Virol. *64 (Pt 6)*, 1277–1290.

Adam, I., and Karsany, M.S. (2008). Case report: Rift Valley Fever with vertical transmission in a pregnant Sudanese woman. J. Med. Virol. *80*, 929.

Adamala, K., and Szostak, J.W. (2013). Nonenzymatic template-directed RNA synthesis inside model protocells. Science *342*, 1098–1100.

Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K., and Miyano, M. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Struct. Lond. Engl. 1993 *7*, 1417–1426.

Akashi, H., and Bishop, D.H. (1983). Comparison of the sequences and coding of La Crosse and snowshoe hare bunyavirus S RNA species. J. Virol. *45*, 1155–1158.

Andersson, A.M., and Pettersson, R.F. (1998). Targeting of a short peptide derived from the cytoplasmic tail of the G1 membrane glycoprotein of Uukuniemi virus (Bunyaviridae) to the Golgi complex. J. Virol. *72*, 9585–9596.

Andersson, I., Bladh, L., Mousavi-Jazi, M., Magnusson, K.-E., Lundkvist, A., Haller, O., and Mirazimi, A. (2004). Human MxA protein inhibits the replication of Crimean-Congo hemorrhagic fever virus. J. Virol. *78*, 4323–4329.

Appleby, T.C., Perry, J.K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D., Wetmore, D.R., McGrath, M.E., Ray, A.S., et al. (2015). Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. Science *347*, 771–775.

Ariza, A., Tanner, S.J., Walter, C.T., Dent, K.C., Shepherd, D.A., Wu, W., Matthews, S.V., Hiscox, J.A., Green, T.J., Luo, M., et al. (2013). Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. Nucleic Acids Res. *41*, 5912–5926.

Arnold, J.J., Vignuzzi, M., Stone, J.K., Andino, R., and Cameron, C.E. (2005). Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. J. Biol. Chem. *280*, 25706–25716.

Arranz, R., Coloma, R., Chichón, F.J., Conesa, J.J., Carrascosa, J.L., Valpuesta, J.M., Ortín, J., and Martín-Benito, J. (2012). The structure of native influenza virion ribonucleoproteins. Science *338*, 1634–1637.

Barr, J.N. (2007). Bunyavirus mRNA synthesis is coupled to translation to prevent premature transcription termination. RNA N. Y. N *13*, 731–736.

Barr, J.N., and Wertz, G.W. (2004). Bunyamwera bunyavirus RNA synthesis requires cooperation of 3'- and 5'-terminal sequences. J. Virol. *78*, 1129–1138.

Barr, J.N., and Wertz, G.W. (2005). Role of the conserved nucleotide mismatch within 3'- and 5'-terminal regions of Bunyamwera virus in signaling transcription. J. Virol. *79*, 3586–3594.

Barr, J.N., Elliott, R.M., Dunn, E.F., and Wertz, G.W. (2003). Segment-specific terminal sequences of Bunyamwera bunyavirus regulate genome replication. Virology *311*, 326–338.

Barr, J.N., Rodgers, J.W., and Wertz, G.W. (2005). The Bunyamwera virus mRNA transcription signal resides within both the 3' and the 5' terminal regions and allows ambisense transcription from a model RNA segment. J. Virol. *79*, 12602–12607.

Bellocq, C., and Kolakofsky, D. (1987). Translational requirement for La Crosse virus S-mRNA synthesis: a possible mechanism. J. Virol. *61*, 3960–3967.

Bellocq, C., Raju, R., Patterson, J., and Kolakofsky, D. (1987). Translational requirement of La Crosse virus S-mRNA synthesis: in vitro studies. J. Virol. *61*, 87–95.

Bessman, M.J., Lehman, I.R., Simms, E.S., and Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction. J. Biol. Chem. *233*, 171–177.

Bishop, D.H., Gay, M.E., and Matsuoko, Y. (1983). Nonviral heterogeneous sequences are present at the 5' ends of one species of snowshoe hare bunyavirus S complementary RNA. Nucleic Acids Res. *11*, 6409–6418.

Bishop, D.H.L., Gould, K.G., Akashi, H., and Haaster, C.M.C. (1982). The complete sequence and coding content of snowshoe hare bunyavirus small (S) viral RNA species. Nucleic Acids Res. *10*, 3703–3713.

Bouloy, M., Vialat, P., Girard, M., and Pardigon, N. (1984). A transcript from the S segment of the Germiston bunyavirus is uncapped and codes for the nucleoprotein and a nonstructural protein. J. Virol. *49*, 717–723.

Bouloy, M., Pardigon, N., Vialat, P., Gerbaud, S., and Girard, M. (1990). Characterization of the 5' and 3' ends of viral messenger RNAs isolated from BHK21 cells infected with Germiston virus (Bunyavirus). Virology *175*, 50–58.

Bowler, M.W., Guijarro, M., Petitdemange, S., Baker, I., Svensson, O., Burghammer, M., Mueller-Dieckmann, C., Gordon, E.J., Flot, D., McSweeney, S.M., et al. (2010). Diffraction cartography: applying microbeams to macromolecular crystallography sample evaluation and data collection. Acta Crystallogr. D Biol. Crystallogr. *66*, 855–864.

Brennan, B., Welch, S.R., and Elliott, R.M. (2014). The consequences of reconfiguring the ambisense S genome segment of Rift Valley fever virus on viral replication in mammalian and mosquito cells and for genome packaging. PLoS Pathog. *10*, e1003922.

Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De Francesco, R., and Rey, F.A. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proc. Natl. Acad. Sci. U. S. A. *96*, 13034–13039.

Bressanelli, S., Tomei, L., Rey, F.A., and De Francesco, R. (2002). Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. J. Virol. *76*, 3482–3492.

Bridgen, A., Dalrymple, D.A., Weber, F., and Elliott, R.M. (2004). Inhibition of Dugbe nairovirus replication by human MxA protein. Virus Res. *99*, 47–50.

Brockhauser, S., Ravelli, R.B.G., and McCarthy, A.A. (2013). The use of a mini-κ goniometer head in macromolecular crystallography diffraction experiments. Acta Crystallogr. D Biol. Crystallogr. *69*, 1241–1251.

Bruenn, J.A. (2003). A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. Nucleic Acids Res. *31*, 1821–1829.

Brunotte, L., Kerber, R., Shang, W., Hauer, F., Hass, M., Gabriel, M., Lelke, M., Busch, C., Stark, H., Svergun, D.I., et al. (2011). Structure of the Lassa virus nucleoprotein revealed by X-ray crystallography, small-angle X-ray scattering, and electron microscopy. J. Biol. Chem. *286*, 38748–38756.

Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., and Stuart, D.I. (2001). A mechanism for initiating RNA-dependent RNA polymerization. Nature *410*, 235–240.

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.

Capodagli, G.C., McKercher, M.A., Baker, E.A., Masters, E.M., Brunzelle, J.S., and Pegan, S.D. (2011). Structural analysis of a viral ovarian tumor domain protease from the Crimean-Congo hemorrhagic fever virus in complex with covalently bonded ubiquitin. J. Virol. *85*, 3621–3630.

Carter, S.D., Surtees, R., Walter, C.T., Ariza, A., Bergeron, É., Nichol, S.T., Hiscox, J.A., Edwards, T.A., and Barr, J.N. (2012). Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein. J. Virol. *86*, 10914–10923.

Castro, C., Smidansky, E., Maksimchuk, K.R., Arnold, J.J., Korneeva, V.S., Götte, M., Konigsberg, W., and Cameron, C.E. (2007). Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. Proc. Natl. Acad. Sci. U. S. A. *104*, 4267–4272.

Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A., Götte, M., Konigsberg, W., and Cameron, C.E. (2009). Nucleic acid polymerases use a general acid for nucleotidyl transfer. Nat. Struct. Mol. Biol. *16*, 212–218.

Černý, J., Černá Bolfíková, B., Valdés, J.J., Grubhoffer, L., and Růžek, D. (2014). Evolution of tertiary structure of viral RNA dependent polymerases. PloS One *9*, e96070.

Chang, S., Sun, D., Liang, H., Wang, J., Li, J., Guo, L., Wang, X., Guan, C., Boruah, B.M., Yuan, L., et al. (2015). Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 Å resolution. Mol. Cell *57*, 925–935.

Chen, K.X., Lesburg, C.A., Vibulbhan, B., Yang, W., Chan, T.-Y., Venkatraman, S., Velazquez, F., Zeng, Q., Bennett, F., Anilkumar, G.N., et al. (2012). A novel class of highly potent irreversible hepatitis C virus NS5B polymerase inhibitors. J. Med. Chem. *55*, 2089–2101.

Chenavas, S., Estrozi, L.F., Slama-Schwok, A., Delmas, B., Di Primo, C., Baudin, F., Li, X., Crépin, T., and Ruigrok, R.W.H. (2013). Monomeric nucleoprotein of influenza A virus. PLoS Pathog. *9*, e1003275.

Choi, K.H., and Rossmann, M.G. (2009). RNA-dependent RNA polymerases from Flaviviridae. Curr. Opin. Struct. Biol. *19*, 746–751.

Choi, K.H., Groarke, J.M., Young, D.C., Kuhn, R.J., Smith, J.L., Pevear, D.C., and Rossmann, M.G. (2004). The structure of the RNA-dependent RNA polymerase from bovine viral diarrhea virus establishes the role of GTP in de novo initiation. Proc. Natl. Acad. Sci. U. S. A. *101*, 4425–4430.

Choi, K.H., Gallei, A., Becher, P., and Rossmann, M.G. (2006). The structure of bovine viral diarrhea virus RNA-dependent RNA polymerase and its amino-terminal domain. Struct. Lond. Engl. 1993 *14*, 1107–1113.

Clerx-van Haaster, C.M., Akashi, H., Auperin, D.D., and Bishop, D.H. (1982). Nucleotide sequence analyses and predicted coding of bunyavirus genome RNA species. J. Virol. *41*, 119–128.

Cogoni, C., and Macino, G. (1999). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. Nature *399*, 166–169.

Collett, M.S. (1986). Messenger RNA of the M segment RNA of Rift Valley fever virus. Virology *151*, 151–156.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Dahlberg, J.E., Obijeski, J.F., and Korb, J. (1977). Electron microscopy of the segmented RNA genome ofLa Crosse virus: absence of circular molecules. J. Virol. *22*, 203–209.

Dias, A., Bouvier, D., Crépin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S., and Ruigrok, R.W.H. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature *458*, 914–918.

Dobie, D.K., Blair, C.D., Chandler, L.J., Rayms-Keller, A., McGaw, M.M., Wasieloski, L.P., and Beaty, B.J. (1997). Analysis of LaCrosse virus S mRNA 5’ termini in infected mosquito cells and Aedes triseriatus mosquitoes. J. Virol. *71*, 4395–4399.

Dong, H., Li, P., Elliott, R.M., and Dong, C. (2013a). Structure of Schmallenberg orthobunyavirus nucleoprotein suggests a novel mechanism of genome encapsidation. J. Virol. *87*, 5593–5601.

Dong, H., Li, P., Böttcher, B., Elliott, R.M., and Dong, C. (2013b). Crystal structure of Schmallenberg orthobunyavirus nucleoprotein-RNA complex reveals a novel RNA sequestration mechanism. RNA N. Y. N *19*, 1129–1136.

Drake, J.W. (1993). Rates of spontaneous mutation among RNA viruses. Proc. Natl. Acad. Sci. U. S. A. *90*, 4171–4175.

Dupeux, F., Röwer, M., Seroul, G., Blot, D., and Márquez, J.A. (2011). A thermal stability assay can help to estimate the crystallization likelihood of biological samples. Acta Crystallogr. D Biol. Crystallogr. *67*, 915–919.

Egloff, M.-P., Benarroch, D., Selisko, B., Romette, J.-L., and Canard, B. (2002). An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J. *21*, 2757–2768.

Egloff, M.-P., Decroly, E., Malet, H., Selisko, B., Benarroch, D., Ferron, F., and Canard, B. (2007). Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. J. Mol. Biol. *372*, 723–736.

Ekland, E.H., and Bartel, D.P. (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature *382*, 373–376.

Ekland, E.H., Szostak, J.W., and Bartel, D.P. (1995). Structurally complex and highly active RNA ligases derived from random RNA sequences. Science *269*, 364–370.

Elliott, R.M. (2014). Orthobunyaviruses: recent genetic and structural insights. Nat. Rev. Microbiol. *12*, 673–685.

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

Eshita, Y., and Bishop, D.H.L. (1984). The complete sequence of the M RNA of snowshoe hare bunyavirus reveals the presence of internal hydrophobic domains in the viral glycoprotein. Virology *137*, 227–240.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmís, C., Domingo, E., and Verdaguer, N. (2004). Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. J. Biol. Chem. *279*, 47212–47221.

Ferrer-Orta, C., Arias, A., Escarmís, C., and Verdaguer, N. (2006). A comparison of viral RNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. *16*, 27–34.

Ferrer-Orta, C., Arias, A., Pérez-Luque, R., Escarmís, C., Domingo, E., and Verdaguer, N. (2007). Sequential structures provide insights into the fidelity of RNA replication. Proc. Natl. Acad. Sci. U. S. A. *104*, 9463–9468.

Ferrer-Orta, C., Sierra, M., Agudo, R., de la Higuera, I., Arias, A., Pérez-Luque, R., Escarmís, C., Domingo, E., and Verdaguer, N. (2010). Structure of foot-and-mouth disease virus mutant polymerases with reduced sensitivity to ribavirin. J. Virol. *84*, 6188–6199.

Ferron, F., Li, Z., Danek, E.I., Luo, D., Wong, Y., Coutard, B., Lantez, V., Charrel, R., Canard, B., Walz, T., et al. (2011). The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. PLoS Pathog. *7*, e1002030.

Flick, K., Katz, A., Overby, A., Feldmann, H., Pettersson, R.F., and Flick, R. (2004). Functional analysis of the noncoding regions of the Uukuniemi virus (Bunyaviridae) RNA segments. J. Virol. *78*, 11726–11738.

Flick, R., Elgh, F., and Pettersson, R.F. (2002). Mutational analysis of the Uukuniemi virus (Bunyaviridae family) promoter reveals two elements of functional importance. J. Virol. *76*, 10849–10860.

Fontana, J., López-Montero, N., Elliott, R.M., Fernández, J.J., and Risco, C. (2008). The unique architecture of Bunyamwera virus factories around the Golgi complex. Cell. Microbiol. *10*, 2012–2028.

Francki, R.I.B., Fauquet, C.M., Knudson, D.L., and Brown, F. (1991). Classification and Nomenclature of Viruses (Vienna: Springer Vienna).

Franke, D., and Svergun, D.I. (2009). *DAMMIF* , a program for rapid *ab-initio* shape determination in small-angle scattering. J. Appl. Crystallogr. *42*, 342–346.

Frese, M., Kochs, G., Feldmann, H., Hertkorn, C., and Haller, O. (1996). Inhibition of bunyaviruses, phleboviruses, and hantaviruses by human MxA protein. J. Virol. *70*, 915–923.

Frias-Staheli, N., Giannakopoulos, N.V., Kikkert, M., Taylor, S.L., Bridgen, A., Paragas, J., Richt, J.A., Rowland, R.R., Schmaljohn, C.S., Lenschow, D.J., et al. (2007). Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. Cell Host Microbe *2*, 404–416.

Fuller, F., Bhown, A.S., and Bishop, D.H. (1983). Bunyavirus nucleoprotein, N, and a non-structural protein, NSS, are coded by overlapping reading frames in the S RNA. J. Gen. Virol. *64 (Pt 8)*, 1705–1714.

Fullerton, S.W.B., Blaschke, M., Coutard, B., Gebhardt, J., Gorbalenya, A., Canard, B., Tucker, P.A., and Rohayem, J. (2007). Structural and functional characterization of sapovirus RNA-dependent RNA polymerase. J. Virol. *81*, 1858–1871.

Gabadinho, J., Beteva, A., Guijarro, M., Rey-Bakaikoa, V., Spruce, D., Bowler, M.W., Brockhauser, S., Flot, D., Gordon, E.J., Hall, D.R., et al. (2010). MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. J. Synchrotron Radiat. *17*, 700–707.

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.

Garriga, D., Navarro, A., Querol-Audí, J., Abaitua, F., Rodríguez, J.F., and Verdaguer, N. (2007). Activation mechanism of a noncanonical RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. *104*, 20540–20545.

Garriga, D., Ferrer-Orta, C., Querol-Audí, J., Oliva, B., and Verdaguer, N. (2013). Role of motif B loop in allosteric regulation of RNA-dependent RNA polymerization activity. J. Mol. Biol. *425*, 2279–2287.

Gauliard, N., Billecocq, A., Flick, R., and Bouloy, M. (2006). Rift Valley fever virus noncoding regions of L, M and S segments regulate RNA synthesis. Virology *351*, 170–179.

Gavrilovskaya, I.N., Shepley, M., Shaw, R., Ginsberg, M.H., and Mackow, E.R. (1998). beta3 Integrins mediate the cellular entry of hantaviruses that cause respiratory failure. Proc. Natl. Acad. Sci. U. S. A. *95*, 7074–7079.

Gentsch, J.R., and Bishop, D.H. (1978). Small viral RNA segment of bunyaviruses codes for viral nucleocapsid protein. J. Virol. *28*, 417–419.

Gentsch, J.R., and Bishop, D.L. (1979). M viral RNA segment of bunyaviruses codes for two glycoproteins, G1 and G2. J. Virol. *30*, 767–770.

Gerbaud, S., Vialat, P., Pardigon, N., Wychowski, C., Girard, M., and Bouloy, M. (1987). The S segment of the Germiston virus RNA genome can code for three proteins. Virus Res. *8*, 1–13.

Gerdts, C.J., Elliott, M., Lovell, S., Mixon, M.B., Napuli, A.J., Staker, B.L., Nollert, P., and Stewart, L. (2008). The plug-based nanovolume Microcapillary Protein Crystallization System (MPCS). Acta Crystallogr. D Biol. Crystallogr. *64*, 1116–1122.

Gerlach, P., Malet, H., Cusack, S., and Reguera, J. (2015). Structural insights into bunyavirus replication and its regulation by the vRNA promoter. Cell *in press*.

Gerrard, S.R., and Nichol, S.T. (2002). Characterization of the Golgi retention motif of Rift Valley fever virus G(N) glycoprotein. J. Virol. *76*, 12200–12210.

Gohara, D.W., Crotty, S., Arnold, J.J., Yoder, J.D., Andino, R., and Cameron, C.E. (2000). Poliovirus RNA-dependent RNA polymerase (3Dpol): structural, biochemical, and biological analysis of conserved structural motifs A and B. J. Biol. Chem. *275*, 25523–25532.

Gong, P., and Peersen, O.B. (2010). Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. *107*, 22505–22510.

Gong, P., Kortus, M.G., Nix, J.C., Davis, R.E., and Peersen, O.B. (2013). Structures of coxsackievirus, rhinovirus, and poliovirus polymerase elongation complexes solved by engineering RNA mediated crystal contacts. PloS One *8*, e60272.

Gorbalenya, A.E., Pringle, F.M., Zeddam, J.-L., Luke, B.T., Cameron, C.E., Kalmakoff, J., Hanzlik, T.N., Gordon, K.H.J., and Ward, V.K. (2002). The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. J. Mol. Biol. *324*, 47–62.

Grady, L.J., Sanders, M.L., and Campbell, W.P. (1987). The sequence of the M RNA of an isolate of La Crosse virus. J. Gen. Virol. *68 ( Pt 12)*, 3057–3071.

Graham, S.C., Sarin, L.P., Bahar, M.W., Myers, R.A., Stuart, D.I., Bamford, D.H., and Grimes, J.M. (2011). The N-terminus of the RNA polymerase from infectious pancreatic necrosis virus is the determinant of genome attachment. PLoS Pathog. *7*, e1002085.

Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., McCarthy, J.E.G., and Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell *115*, 739–750.

Gruez, A., Selisko, B., Roberts, M., Bricogne, G., Bussetta, C., Jabafi, I., Coutard, B., De Palma, A.M., Neyts, J., and Canard, B. (2008). The crystal structure of coxsackievirus B3 RNA-dependent RNA polymerase in complex with its protein primer VPg confirms the existence of a second VPg binding site on Picornaviridae polymerases. J. Virol. *82*, 9577–9590.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W.H., Ortin, J., Hart, D.J., et al. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat. Struct. Mol. Biol. *15*, 500–506.

Guo, Y., Wang, W., Ji, W., Deng, M., Sun, Y., Zhou, H., Yang, C., Deng, F., Wang, H., Hu, Z., et al. (2012). Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. Proc. Natl. Acad. Sci. U. S. A. *109*, 5046–5051.

Haaster, C.C., and Bishop, D.H. (1980). Analyses of the 3’-terminal sequences of snowshoe hare and La Crosse Bunyaviruses. Virology *105*, 564–574.

Habjan, M., Andersson, I., Klingström, J., Schümann, M., Martin, A., Zimmermann, P., Wagner, V., Pichlmair, A., Schneider, U., Mühlberger, E., et al. (2008). Processing of genome 5’ termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction. PloS One *3*, e2032.

Habjan, M., Penski, N., Wagner, V., Spiegel, M., Overby, A.K., Kochs, G., Huiskonen, J.T., and Weber, F. (2009). Efficient production of Rift Valley fever virus-like particles: The antiviral protein MxA can inhibit primary transcription of bunyaviruses. Virology *385*, 400–408.

Hacker, D., Raju, R., and Kolakofsky, D. (1989). La Crosse virus nucleocapsid protein controls its own synthesis in mosquito cells by encapsidating its mRNA. J. Virol. *63*, 5166–5174.

Hacker, D., Rochat, S., and Kolakofsky, D. (1990). Anti-mRNAs in La Crosse bunyavirus-infected cells. J. Virol. *64*, 5051–5057.

Haddow, A.D., and Odoi, A. (2009). The incidence risk, clustering, and clinical presentation of La Crosse virus infections in the eastern United States, 2003-2007. PloS One *4*, e6145.

Haller, O., Staeheli, P., and Kochs, G. (2007). Interferon-induced Mx proteins in antiviral host defense. Biochimie *89*, 812–818.

Hansen, J.L., Long, A.M., and Schultz, S.C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. Struct. Lond. Engl. 1993 *5*, 1109–1122.

Harris, K.S., Xiang, W., Alexander, L., Lane, W.S., Paul, A.V., and Wimmer, E. (1994). Interaction of poliovirus polypeptide 3CDpro with the 5' and 3' termini of the poliovirus genome. Identification of viral and cellular cofactors needed for efficient binding. J. Biol. Chem. *269*, 27004–27014.

Hastie, K.M., Kimberlin, C.R., Zandonatti, M.A., MacRae, I.J., and Saphire, E.O. (2011). Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3' to 5' exonuclease activity essential for immune suppression. Proc. Natl. Acad. Sci. U. S. A. *108*, 2396–2401.

He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., et al. (2008). Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. Nature *454*, 1123–1126.

Hodges, E.N., and Connor, J.H. (2013). Translational control by negative-strand RNA viruses: methods for the study of a crucial virus/host interaction. Methods San Diego Calif *59*, 180–187.

Hoffmann, B., Scheuch, M., Höper, D., Jungblut, R., Holsteg, M., Schirrmeier, H., Eschbaumer, M., Goller, K.V., Wernike, K., Fischer, M., et al. (2012). Novel orthobunyavirus in Cattle, Europe, 2011. Emerg. Infect. Dis. *18*, 469–472.

Hollidge, B.S., Nedelsky, N.B., Salzano, M.-V., Fraser, J.W., González-Scarano, F., and Soldan, S.S. (2012). Orthobunyavirus entry into neurons and other mammalian cells occurs via clathrin-mediated endocytosis and requires trafficking into early endosomes. J. Virol. *86*, 7988–8001.

Hong, Z., Cameron, C.E., Walker, M.P., Castro, C., Yao, N.H., Lau, J.Y.N., and Zhong, W.D. (2001). A novel mechanism to ensure terminal initiation by hepatitis C virus NS5B polymerase. Virology *285*, 6–11.

Hopkins, K.C., McLane, L.M., Maqbool, T., Panda, D., Gordesky-Gold, B., and Cherry, S. (2013). 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. *27*, 1511–1525.

Hostomsky, Z., Hostomska, Z., Fu, T.B., and Taylor, J. (1992). Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. J. Virol. *66*, 3179–3182.

Hutchinson, K.L., Peters, C.J., and Nichol, S.T. (1996). Sin Nombre virus mRNA synthesis. Virology *224*, 139–149.

Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J. Synchrotron Radiat. *16*, 872–879.

Issur, M., Geiss, B.J., Bougie, I., Picard-Jean, F., Despins, S., Mayette, J., Hobdey, S.E., and Bisaillon, M. (2009). The flavivirus NS5 protein is a true RNA guanylyltransferase that catalyzes a two-step reaction to form the RNA cap structure. RNA N. Y. N *15*, 2340–2350.

James, T.W., Frias-Staheli, N., Bacik, J.-P., Levingston Macleod, J.M., Khajehpour, M., García-Sastre, A., and Mark, B.L. (2011). Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease. Proc. Natl. Acad. Sci. U. S. A. *108*, 2222–2227.

Jin, H., and Elliott, R.M. (1992). Mutagenesis of the L protein encoded by Bunyamwera virus and production of monospecific antibodies. J. Gen. Virol. *73 ( Pt 9)*, 2235–2244.

Johnson, C.M., Perez, D.R., French, R., Merrick, W.C., and Donis, R.O. (2001). The NS5A protein of bovine viral diarrhoea virus interacts with the alpha subunit of translation elongation factor-1. J. Gen. Virol. *82*, 2935–2943.

Jorba, N., Area, E., and Ortín, J. (2008). Oligomerization of the influenza virus polymerase complex in vivo. J. Gen. Virol. *89*, 520–524.

Jorba, N., Coloma, R., and Ortín, J. (2009). Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. PLoS Pathog. *5*, e1000462.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kainulainen, M., Habjan, M., Hubel, P., Busch, L., Lau, S., Colinge, J., Superti-Furga, G., Pichlmair, A., and Weber, F. (2014). Virulence factor NSs of rift valley fever virus recruits the F-box protein FBXO3 to degrade subunit p62 of general transcription factor TFIIH. J. Virol. *88*, 3464–3473.

Van Kasteren, P.B., Beugeling, C., Ninaber, D.K., Frias-Staheli, N., van Boheemen, S., García-Sastre, A., Snijder, E.J., and Kikkert, M. (2012). Arterivirus and nairovirus ovarian tumor domain-containing Deubiquitinases target activated RIG-I to control innate immune signaling. J. Virol. *86*, 773–785.

Kidmose, R.T., Vasiliev, N.N., Chetverin, A.B., Andersen, G.R., and Knudsen, C.R. (2010). Structure of the Qbeta replicase, an RNA-dependent RNA polymerase consisting of viral and host proteins. Proc. Natl. Acad. Sci. U. S. A. *107*, 10884–10889.

Klemm, C., Reguera, J., Cusack, S., Zielecki, F., Kochs, G., and Weber, F. (2013). Systems to establish bunyavirus genome replication in the absence of transcription. J. Virol. *87*, 8205–8212.

Van Knippenberg, I., and Elliott, R.M. (2015). Flexibility of bunyavirus genomes: creation of an orthobunyavirus with an ambisense s segment. J. Virol. *89*, 5525–5535.

Kochs, G., Janzen, C., Hohenberg, H., and Haller, O. (2002). Antivirally active MxA protein sequesters La Crosse virus nucleocapsid protein into perinuclear complexes. Proc. Natl. Acad. Sci. U. S. A. *99*, 3153–3158.

Kohl, A., Dunn, E.F., Lowen, A.C., and 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. J. Gen. Virol. *85*, 3269–3278.

Kohl, A., Lowen, A.C., Léonard, V.H.J., and Elliott, R.M. (2006). Genetic elements regulating packaging of the Bunyamwera orthobunyavirus genome. J. Gen. Virol. *87*, 177–187.

Koldobskaya, Y., Duguid, E.M., Shechner, D.M., Suslov, N.B., Ye, J., Sidhu, S.S., Bartel, D.P., Koide, S., Kossiakoff, A.A., and Piccirilli, J.A. (2011). A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. Nat. Struct. Mol. Biol. *18*, 100–106.

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.

Kramer, K., Hummel, P., Hsiao, H.-H., Luo, X., Wahl, M., and Urlaub, H. (2011). Mass-spectrometric analysis of proteins cross-linked to 4-thio-uracil- and 5-bromo-uracil-substituted RNA. Int. J. Mass Spectrom. *304*, 184–194.

Kranzusch, P.J., Schenk, A.D., Rahmeh, A.A., Radoshitzky, S.R., Bavari, S., Walz, T., and Whelan, S.P.J. (2010). Assembly of a functional Machupo virus polymerase complex. Proc. Natl. Acad. Sci. U. S. A. *107*, 20069–20074.

Lai, M.M.C. (2005). RNA Replication without RNA-Dependent RNA Polymerase: Surprises from Hepatitis Delta Virus. J. Virol. *79*, 7951–7958.

Lanciotti, R.S., Kosoy, O.I., Bosco-Lauth, A.M., Pohl, J., Stuchlik, O., Reed, M., and Lambert, A.J. (2013). Isolation of a novel orthobunyavirus (Brazoran virus) with a 1.7 kb S segment that encodes a unique nucleocapsid protein possessing two putative functional domains. Virology *444*, 55–63.

Lees, J.F., Pringle, C.R., and Elliott, R.M. (1986). Nucleotide sequence of the Bunyamwera virus M RNA segment: conservation of structural features in the Bunyavirus glycoprotein gene product. Virology *148*, 1–14.

Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F., and Weber, P.C. (1999). Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. Nat. Struct. Biol. *6*, 937–943.

Leslie, A.G.W. (2006). The integration of macromolecular diffraction data. Acta Crystallogr. D Biol. Crystallogr. *62*, 48–57.

Li, B., Wang, Q., Pan, X., Fernández de Castro, I., Sun, Y., Guo, Y., Tao, X., Risco, C., Sui, S.-F., and Lou, Z. (2013). Bunyamwera virus possesses a distinct nucleocapsid protein to facilitate genome encapsidation. Proc. Natl. Acad. Sci. U. S. A. *110*, 9048–9053.

Li, J., Fontaine-Rodriguez, E.C., and Whelan, S.P.J. (2005). Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity. J. Virol. *79*, 13373–13384.

Linero, F., Welnowska, E., Carrasco, L., and Scolaro, L. (2013). Participation of eIF4F complex in Junin virus infection: blockage of eIF4E does not impair virus replication. Cell. Microbiol. *15*, 1766–1782.

Lipardi, C., and Paterson, B.M. (2009). Identification of an RNA-dependent RNA polymerase in Drosophila involved in RNAi and transposon suppression. Proc. Natl. Acad. Sci. U. S. A. *106*, 15645–15650.

Liu, S., Abbondanzieri, E.A., Rausch, J.W., Le Grice, S.F.J., and Zhuang, X. (2008). Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. Science *322*, 1092–1097.

Liu, X., Yang, X., Lee, C.A., Moustafa, I.M., Smidansky, E.D., Lum, D., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2013). Vaccine-derived mutation in motif D of poliovirus RNA-dependent RNA polymerase lowers nucleotide incorporation fidelity. J. Biol. Chem. *288*, 32753–32765.

López-Montero, N., and Risco, C. (2011). Self-protection and survival of arbovirus-infected mosquito cells. Cell. Microbiol. *13*, 300–315.

Love, R.A., Maegley, K.A., Yu, X., Ferre, R.A., Lingardo, L.K., Diehl, W., Parge, H.E., Dragovich, P.S., and Fuhrman, S.A. (2004). The crystal structure of the RNA-dependent RNA polymerase from human rhinovirus: a dual function target for common cold antiviral therapy. Struct. Lond. Engl. 1993 *12*, 1533–1544.

Lowen, A.C., and Elliott, R.M. (2005). Mutational analyses of the nonconserved sequences in the Bunyamwera Orthobunyavirus S segment untranslated regions. J. Virol. *79*, 12861–12870.

Lowen, A.C., Boyd, A., Fazakerley, J.K., and Elliott, R.M. (2005). Attenuation of bunyavirus replication by rearrangement of viral coding and noncoding sequences. J. Virol. *79*, 6940–6946.

Lozach, P.-Y., Mancini, R., Bitto, D., Meier, R., Oestereich, L., Overby, A.K., Pettersson, R.F., and Helenius, A. (2010). Entry of bunyaviruses into mammalian cells. Cell Host Microbe *7*, 488–499.

Lozach, P.-Y., Kühbacher, A., Meier, R., Mancini, R., Bitto, D., Bouloy, M., and Helenius, A. (2011). DC-SIGN as a receptor for phleboviruses. Cell Host Microbe *10*, 75–88.

Lu, G., and Gong, P. (2013). Crystal Structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. PLoS Pathog. *9*, e1003549.

Lu, X., McDonald, S.M., Tortorici, M.A., Tao, Y.J., Vasquez-Del Carpio, R., Nibert, M.L., Patton, J.T., and Harrison, S.C. (2008). Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. Struct. Lond. Engl. 1993 *16*, 1678–1688.

Lyle, J.M., Bullitt, E., Bienz, K., and Kirkegaard, K. (2002). Visualization and functional analysis of RNA-dependent RNA polymerase lattices. Science *296*, 2218–2222.

Maida, Y., and Masutomi, K. (2011). RNA-dependent RNA polymerases in RNA silencing. Biol. Chem. *392*, 299–304.

Makeyev, E.V., and Bamford, D.H. (2002). Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. Mol. Cell *10*, 1417–1427.

Malet, H., Egloff, M.-P., Selisko, B., Butcher, R.E., Wright, P.J., Roberts, M., Gruez, A., Sulzenbacher, G., Vonrhein, C., Bricogne, G., et al. (2007). Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. J. Biol. Chem. *282*, 10678–10689.

Maniar, J.M., and Fire, A.Z. (2011). EGO-1, a C. elegans RdRP, Modulates Gene Expression via Production of mRNA-Templated Short Antisense RNAs. Curr. Biol. CB *21*, 449–459.

Marcotte, L.L., Wass, A.B., Gohara, D.W., Pathak, H.B., Arnold, J.J., Filman, D.J., Cameron, C.E., and Hogle, J.M. (2007). Crystal structure of poliovirus 3CD protein: virally encoded protease and precursor to the RNA-dependent RNA polymerase. J. Virol. *81*, 3583–3596.

Marczinke, B.I., and Nichol, S.T. (2002). Nairobi sheep disease virus, an important tick-borne pathogen of sheep and goats in Africa, is also present in Asia. Virology *303*, 146–151.

Mardani, M., Keshtkar-Jahromi, M., Ataie, B., and Adibi, P. (2009). Crimean-Congo hemorrhagic fever virus as a nosocomial pathogen in Iran. Am. J. Trop. Med. Hyg. *81*, 675–678.

De Maria Antolinos, A., Pernot, P., Brennich, M.E., Kieffer, J., Bowler, M.W., Delageniere, S., Ohlsson, S., Malbet Monaco, S., Ashton, A., Franke, D., et al. (2015). ISPyB for BioSAXS, the gateway to user autonomy in solution scattering experiments. Acta Crystallogr. D Biol. Crystallogr. *71*, 76–85.

Marklewitz, M., Zirkel, F., Rwego, I.B., Heidemann, H., Trippner, P., Kurth, A., Kallies, R., Briese, T., Lipkin, W.I., Drosten, C., et al. (2013). Discovery of a unique novel clade of mosquito-associated bunyaviruses. J. Virol. *87*, 12850–12865.

Matsuoka, Y., Chen, S.Y., and Compans, R.W. (1994). A signal for Golgi retention in the bunyavirus G1 glycoprotein. J. Biol. Chem. *269*, 22565–22573.

Milner-White, E.J., and Russell, M.J. (2011). Functional capabilities of the earliest peptides and the emergence of life. Genes *2*, 671–688.

Mir, M.A., and Panganiban, A.T. (2008). A protein that replaces the entire cellular eIF4F complex. EMBO J. *27*, 3129–3139.

Mir, M.A., Duran, W.A., Hjelle, B.L., Ye, C., and Panganiban, A.T. (2008). Storage of cellular 5’ mRNA caps in P bodies for viral cap-snatching. Proc. Natl. Acad. Sci. U. S. A. *105*, 19294–19299.

Mir, M.A., Sheema, S., Haseeb, A., and Haque, A. (2010). Hantavirus nucleocapsid protein has distinct m7G cap- and RNA-binding sites. J. Biol. Chem. *285*, 11357–11368.

Morin, B., Coutard, B., Lelke, M., Ferron, F., Kerber, R., Jamal, S., Frangeul, A., Baronti, C., Charrel, R., de Lamballerie, X., et al. (2010). The N-Terminal Domain of the Arenavirus L

Protein Is an RNA Endonuclease Essential in mRNA Transcription. PLoS Pathog *6*, e1001038.

Mosley, R.T., Edwards, T.E., Murakami, E., Lam, A.M., Grice, R.L., Du, J., Sofia, M.J., Furman, P.A., and Otto, M.J. (2012). Structure of hepatitis C virus polymerase in complex with primer-template RNA. J. Virol. *86*, 6503–6511.

Müller, R., Poch, O., Delarue, M., Bishop, D.H., and Bouloy, M. (1994). Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. J. Gen. Virol. *75 ( Pt 6)*, 1345–1352.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Myasnikov, A.G., Simonetti, A., Marzi, S., and Klaholz, B.P. (2009). Structure-function insights into prokaryotic and eukaryotic translation initiation. Curr. Opin. Struct. Biol. *19*, 300–309.

Nemirov, K., Vaheri, A., and Plyusnin, A. (2004). Hantaviruses: co-evolution with natural hosts. 201–228.

Ng, A.K.-L., Lam, M.K.-H., Zhang, H., Liu, J., Au, S.W.-N., Chan, P.K.-S., Wang, J., and Shaw, P.-C. (2012). Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. J. Virol. *86*, 6758–6767.

Ng, K.K.S., Cherney, M.M., Vazquez, A.L., Machin, A., Alonso, J.M.M., Parra, F., and James, M.N.G. (2002). Crystal structures of active and inactive conformations of a caliciviral RNA-dependent RNA polymerase. J. Biol. Chem. *277*, 1381–1387.

Ng, K.K.-S., Pendás-Franco, N., Rojo, J., Boga, J.A., Machín, A., Alonso, J.M.M., and Parra, F. (2004). Crystal structure of norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. J. Biol. Chem. *279*, 16638–16645.

Ng, K.K.S., Arnold, J.J., and Cameron, C.E. (2008). Structure-function relationships among RNA-dependent RNA polymerases. Curr. Top. Microbiol. Immunol. *320*, 137–156.

Nichol, S.T., Beaty, B.J., Elliott, R.M., Goldbach, R., Plyusnin, A., Schmaljohn, C.S., and Tesh, R.B. (2005). The Negative Sense Single Stranded RNA Viruses - Bunyaviruses. In Virus Taxonomy, (San Diego: Academic Press), pp. 695–716.

Nishikiori, M., Sugiyama, S., Xiang, H., Niiyama, M., Ishibashi, K., Inoue, T., Ishikawa, M., Matsumura, H., and Katoh, E. (2012). Crystal structure of the superfamily 1 helicase from Tomato mosaic virus. J. Virol. *86*, 7565–7576.

Niu, F., Shaw, N., Wang, Y.E., Jiao, L., Ding, W., Li, X., Zhu, P., Upur, H., Ouyang, S., Cheng, G., et al. (2013). Structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex reveals unique architecture for RNA encapsidation. Proc. Natl. Acad. Sci. U. S. A. *110*, 9054–9059.

Noble, C.G., Lim, S.P., Chen, Y.-L., Liew, C.W., Yap, L., Lescar, J., and Shi, P.-Y. (2013). Conformational flexibility of the Dengue virus RNA-dependent RNA polymerase revealed by a complex with an inhibitor. J. Virol. *87*, 5291–5295.

Novoa, R.R., Calderita, G., Cabezas, P., Elliott, R.M., and Risco, C. (2005). Key Golgi factors for structural and functional maturation of bunyamwera virus. J. Virol. *79*, 10852–10863.

Nyanguile, O., Devogelaere, B., Vijgen, L., Van den Broeck, W., Pauwels, F., Cummings, M.D., De Bondt, H.L., Vos, A.M., Berke, J.M., Lenz, O., et al. (2010). 1a/1b subtype profiling of nonnucleoside polymerase inhibitors of hepatitis C virus. J. Virol. *84*, 2923–2934.

Obijeski, J.F., Bishop, D.H., Palmer, E.L., and Murphy, F.A. (1976). Segmented genome and nucleocapsid of La Crosse virus. J. Virol. *20*, 664–675.

Obijeski, J.F., McCauley, J., and Shekei, J.J. (1980). Nucleotide sequences at the termini of La Crosse virus RNAs. Nucleic Acids Res. *8*, 2431–2438.

O'Farrell, D., Trowbridge, R., Rowlands, D., and Jäger, J. (2003). Substrate complexes of hepatitis C virus RNA polymerase (HC-J4): structural evidence for nucleotide import and de-novo initiation. J. Mol. Biol. *326*, 1025–1035.

Ogino, T., Yadav, S.P., and Banerjee, A.K. (2010). Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. *107*, 3463–3468.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., and Steitz, T.A. (1985). Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. Nature *313*, 762–766.

Osborne, J.C., and Elliott, R.M. (2000). RNA binding properties of bunyamwera virus nucleocapsid protein and selective binding to an element in the 5' terminus of the negative-sense S segment. J. Virol. *74*, 9946–9952.

Padula, P.J., Edelstein, A., Miguel, S.D., López, N.M., Rossi, C.M., and Rabinovich, R.D. (1998). Hantavirus pulmonary syndrome outbreak in Argentina: molecular evidence for person-to-person transmission of Andes virus. Virology *241*, 323–330.

Palacios, G., Savji, N., Travassos da Rosa, A., Guzman, H., Yu, X., Desai, A., Rosen, G.E., Hutchison, S., Lipkin, W.I., and Tesh, R. (2013). Characterization of the Uukuniemi Virus Group (Phlebovirus: Bunyaviridae): Evidence for Seven Distinct Species. J. Virol. *87*, 3187–3195.

Pan, J., Vakharia, V.N., and Tao, Y.J. (2007). The structure of a birnavirus polymerase reveals a distinct active site topology. Proc. Natl. Acad. Sci. U. S. A. *104*, 7385–7390.

Pappu, H.R., Jones, R. a. C., and Jain, R.K. (2009). Global status of tospovirus epidemics in diverse cropping systems: successes achieved and challenges ahead. Virus Res. *141*, 219–236.

Pardigon, N., Vialat, P., Gerbaud, S., Girard, M., and Bouloy, M. (1988). Nucleotide sequence of the M segment of Germiston virus: comparison of the M gene product of several bunyaviruses. Virus Res. *11*, 73–85.

Patterson, J.L., and Kolakofsky, D. (1984). Characterization of La Crosse virus small-genome transcripts. J. Virol. *49*, 680–685.

Patterson, J.L., Kolakofsky, D., Holloway, B.P., and Obijeski, J.F. (1983). Isolation of the ends of La Crosse virus small RNA as a double-stranded structure. J. Virol. *45*, 882–884.

Patterson, J.L., Holloway, B., and Kolakofsky, D. (1984). La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. J. Virol. *52*, 215–222.

Pattnaik, A.K., and Abraham, G. (1983). Identification of four complementary RNA species in Akabane virus-infected cells. J. Virol. *47*, 452–462.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J. Synchrotron Radiat. *20*, 660–664.

Pettersson, R.F., and von Bonsdorff, C.H. (1975). Ribonucleoproteins of Uukuniemi virus are circular. J. Virol. *15*, 386–392.

Pfeiffer, J.K., and Kirkegaard, K. (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. *1*, e11.

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.

Piper, M.E., Sorenson, D.R., and Gerrard, S.R. (2011). Efficient cellular release of Rift Valley fever virus requires genomic RNA. PloS One *6*, e18070.

Plotch, S.J., Bouloy, M., Ulmanen, I., and Krug, R.M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell *23*, 847–858.

Plyusnin, A., and Elliott, R.M. (2011). Bunyaviridae: Molecular and Cellular Biology (Horizon Scientific Press).

Plyusnin, A., and Morzunov, S.P. (2001). Virus evolution and genetic diversity of hantaviruses and their rodent hosts. Curr. Top. Microbiol. Immunol. *256*, 47–75.

Poch, O., Blumberg, B.M., Bougueleret, L., and Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. *71 ( Pt 5)*, 1153–1162.

Pollastri, G., and McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. Bioinforma. Oxf. Engl. *21*, 1719–1720.

Poranen, M.M., Salgado, P.S., Koivunen, M.R.L., Wright, S., Bamford, D.H., Stuart, D.I., and Grimes, J.M. (2008). Structural explanation for the role of Mn2+ in the activity of phi6 RNA-dependent RNA polymerase. Nucleic Acids Res. *36*, 6633–6644.

Porterfield, J.S., Casals, J., Chumakov, M.P., Gaidamovich, S.Y., Hannoun, C., Holmes, I.H., Horzinek, M.C., Mussgay, M., Oker-Blom, N., and Russell, P.K. (1975). Bunyaviruses and Bunyaviridae. Intervirology *6*, 13–24.

Prehaud, C., Lopez, N., Blok, M.J., Obry, 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.

Pressing, J., and Reanney, D.C. (1984). Divided genomes and intrinsic noise. J. Mol. Evol. *20*, 135–146.

Qi, X., Lan, S., Wang, W., Schelde, L.M., Dong, H., Wallat, G.D., Ly, H., Liang, Y., and Dong, C. (2010). Cap binding and immune evasion revealed by Lassa nucleoprotein structure. Nature *468*, 779–783.

Raju, R., and Kolakofsky, D. (1986). Inhibitors of protein synthesis inhibit both La Crosse virus S-mRNA and S genome syntheses in vivo. Virus Res. *5*, 1–9.

Raju, R., and Kolakofsky, D. (1989). The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base paired. J. Virol. *63*, 122–128.

Raju, R., Raju, L., and Kolakofsky, D. (1989). The translational requirement for complete La Crosse virus mRNA synthesis is cell-type dependent. J. Virol. *63*, 5159–5165.

Raymond, D.D., Piper, M.E., Gerrard, S.R., and Smith, J.L. (2010). Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. Proc. Natl. Acad. Sci. U. S. A. *107*, 11769–11774.

Raymond, D.D., Piper, M.E., Gerrard, S.R., Skiniotis, G., and Smith, J.L. (2012). Phleboviruses encapsidate their genomes by sequestering RNA bases. Proc. Natl. Acad. Sci. U. S. A. *109*, 19208–19213.

Reguera, J., Weber, F., and Cusack, S. (2010). Bunyaviridae RNA Polymerases (L-Protein) Have an N-Terminal, Influenza-Like Endonuclease Domain, Essential for Viral Cap-Dependent Transcription. PLoS Pathog. *6*.

Reguera, J., Malet, H., Weber, F., and Cusack, S. (2013). Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein. Proc. Natl. Acad. Sci. U. S. A. *110*, 7246–7251.

Reguera, J., Cusack, S., and Kolakofsky, D. (2014). Segmented negative strand RNA virus nucleoprotein structure. Curr. Opin. Virol. *5*, 7–15.

Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W.H., et al. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature *516*, 361–366.

Reichelt, M., Stertz, S., Krijnse-Locker, J., Haller, O., and Kochs, G. (2004). Missorting of LaCrosse Virus Nucleocapsid Protein by the Interferon-Induced MxA GTPase Involves Smooth ER Membranes. Traffic *5*, 772–784.

Robertson, M.P., and Scott, W.G. (2007). The structural basis of ribozyme-catalyzed RNA assembly. Science *315*, 1549–1553.

Rossier, C., Raju, R., and Kolakofsky, D. (1988). LaCrosse virus gene expression in mammalian and mosquito cells. Virology *165*, 539–548.

Ruigrok, R.W.H., Crépin, T., and Kolakofsky, D. (2011). Nucleoproteins and nucleocapsids of negative-strand RNA viruses. Curr. Opin. Microbiol. *14*, 504–510.

Rydberg, E.H., Cellucci, A., Bartholomew, L., Mattu, M., Barbato, G., Ludmerer, S.W., Graham, D.J., Altamura, S., Paonessa, G., De Francesco, R., et al. (2009). Structural basis for resistance of the genotype 2b hepatitis C virus NS5B polymerase to site A non-nucleoside inhibitors. J. Mol. Biol. *390*, 1048–1059.

Salgado, P.S., Makeyev, E.V., Butcher, S.J., Bamford, D.H., Stuart, D.I., and Grimes, J.M. (2004). The structural basis for RNA specificity and Ca2+ inhibition of an RNA-dependent RNA polymerase. Struct. Lond. Engl. 1993 *12*, 307–316.

Salgado, P.S., Koivunen, M.R.L., Makeyev, E.V., Bamford, D.H., Stuart, D.I., and Grimes, J.M. (2006). The structure of an RNAi polymerase links RNA silencing and transcription. PLoS Biol. *4*, e434.

Santos, R.I.M., Rodrigues, A.H., Silva, M.L., Mortara, R.A., Rossi, M.A., Jamur, M.C., Oliver, C., and Arruda, E. (2008). Oropouche virus entry into HeLa cells involves clathrin and requires endosomal acidification. Virus Res. *138*, 139–143.

Schmaljohn, C.S., and Dalrymple, J.M. (1983). Analysis of Hantaan virus RNA: evidence for a new genus of bunyaviridae. Virology *131*, 482–491.

Schmitt, M., Scrima, N., Radujkovic, D., Caillet-Saguy, C., Simister, P.C., Friebe, P., Wicht, O., Klein, R., Bartenschlager, R., Lohmann, V., et al. (2011). A comprehensive structure-function comparison of hepatitis C virus strain JFH1 and J6 polymerases reveals a key residue stimulating replication in cell culture across genotypes. J. Virol. *85*, 2565–2581.

Sen, T.Z., Jernigan, R.L., Garnier, J., and Kloczkowski, A. (2005). GOR V server for protein secondary structure prediction. Bioinforma. Oxf. Engl. *21*, 2787–2788.

Shechner, D.M., Grant, R.A., Bagby, S.C., Koldobskaya, Y., Piccirilli, J.A., and Bartel, D.P. (2009). Crystal structure of the catalytic core of an RNA-polymerase ribozyme. Science *326*, 1271–1275.

Shen, H., Sun, H., and Li, G. (2012). What is the role of motif D in the nucleotide incorporation catalyzed by the RNA-dependent RNA polymerase from poliovirus? PLoS Comput. Biol. *8*, e1002851.

Shi, X., Lappin, D.F., and Elliott, R.M. (2004). Mapping the Golgi targeting and retention signal of Bunyamwera virus glycoproteins. J. Virol. *78*, 10793–10802.

Shi, X., Kohl, A., Léonard, V.H.J., Li, P., McLees, A., and Elliott, R.M. (2006). Requirement of the N-terminal region of orthobunyavirus nonstructural protein NSm for virus assembly and morphogenesis. J. Virol. *80*, 8089–8099.

Shi, X., Goli, J., Clark, G., Brauburger, K., and Elliott, R.M. (2009). Functional analysis of the Bunyamwera orthobunyavirus Gc glycoprotein. J. Gen. Virol. *90*, 2483–2492.

Sholders, A.J., and Peersen, O.B. (2014). Distinct conformations of a putative translocation element in poliovirus polymerase. J. Mol. Biol. *426*, 1407–1419.

Smithburn, K.C., Haddow, A.J., and Mahaffy, A.F. (1946). A neurotropic virus isolated from Aedes mosquitoes caught in the Semliki forest. Am. J. Trop. Med. Hyg. *26*, 189–208.

Snippe, M., Smeenk, L., Goldbach, R., and Kormelink, R. (2007). The cytoplasmic domain of tomato spotted wilt virus Gn glycoprotein is required for Golgi localisation and interaction with Gc. Virology *363*, 272–279.

Spagnolo, J.F., Rossignol, E., Bullitt, E., and Kirkegaard, K. (2010). Enzymatic and nonenzymatic functions of viral RNA-dependent RNA polymerases within oligomeric arrays. RNA N. Y. N *16*, 382–393.

Steitz, T.A. (1998). A mechanism for all polymerases. Nature *391*, 231–232.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S.I.S. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. Proc. Natl. Acad. Sci. U. S. A. *102*, 152–157.

Surana, P., Satchidanandam, V., and Nair, D.T. (2014). RNA-dependent RNA polymerase of Japanese encephalitis virus binds the initiator nucleotide GTP to form a mechanistically important pre-initiation state. Nucleic Acids Res. *42*, 2758–2773.

Swanepoel, R., Struthers, J.K., Shepherd, A.J., McGillivray, G.M., Nel, M.J., and Jupp, P.G. (1983). Crimean-congo hemorrhagic fever in South Africa. Am. J. Trop. Med. Hyg. *32*, 1407–1415.

Takeshita, D., and Tomita, K. (2012). Molecular basis for RNA polymerization by Qβ replicase. Nat. Struct. Mol. Biol. *19*, 229–237.

Takeshita, D., Yamashita, S., and Tomita, K. (2014). Molecular insights into replication initiation by Qβ replicase using ribosomal protein S1. Nucleic Acids Res. *42*, 10809–10822.

Tao, Y., Farsetta, D.L., Nibert, M.L., and Harrison, S.C. (2002). RNA synthesis in a cage--structural studies of reovirus polymerase lambda3. Cell *111*, 733–745.

Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W.H., Cusack, S., and Hart, D.J. (2008). Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. PLoS Pathog. *4*, e1000136.

Terasaki, K., Murakami, S., Lokugamage, K.G., and Makino, S. (2011). Mechanism of tripartite RNA genome packaging in Rift Valley fever virus. Proc. Natl. Acad. Sci. U. S. A. *108*, 804–809.

Thomas, D., Blakqori, G., Wagner, V., Banholzer, M., Kessler, N., Elliott, R.M., Haller, O., and Weber, F. (2004). Inhibition of RNA polymerase II phosphorylation by a viral interferon antagonist. J. Biol. Chem. *279*, 31471–31477.

Thompson, A.A., and Peersen, O.B. (2004). Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. EMBO J. *23*, 3462–3471.

Turina, M., Tavella, L., and Ciuffo, M. (2012). Tospoviruses in the Mediterranean area. Adv. Virus Res. *84*, 403–437.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. *66*, 22–25.

Te Velthuis, A.J.W. (2014). Common and unique features of viral RNA-dependent polymerases. Cell. Mol. Life Sci. CMLS *71*, 4403–4420.

Verbruggen, P., Ruf, M., Blakqori, G., Överby, A.K., Heidemann, M., Eick, D., and Weber, F. (2011). Interferon antagonist NSs of La Crosse virus triggers a DNA damage response-like degradation of transcribing RNA polymerase II. J. Biol. Chem. *286*, 3681–3692.

Vialat, P., and Bouloy, M. (1992). Germiston virus transcriptase requires active 40S ribosomal subunits and utilizes capped cellular RNAs. J. Virol. *66*, 685–693.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol. Biol. Clifton NJ *364*, 215–230.

Wallat, G.D., Huang, Q., Wang, W., Dong, H., Ly, H., Liang, Y., and Dong, C. (2014). High-Resolution Structure of the N-Terminal Endonuclease Domain of the Lassa Virus L Polymerase in Complex with Magnesium Ions. PLoS ONE *9*, e87577.

Walter, C.T., and Barr, J.N. (2011). Recent advances in the molecular and cellular biology of bunyaviruses. J. Gen. Virol. *92*, 2467–2484.

Wang, J., Smerdon, S.J., Jäger, J., Kohlstaedt, L.A., Rice, P.A., Friedman, J.M., and Steitz, T.A. (1994). Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. Proc. Natl. Acad. Sci. U. S. A. *91*, 7242–7246.

Wang, Y., Dutta, S., Karlberg, H., Devignot, S., Weber, F., Hao, Q., Tan, Y.J., Mirazimi, A., and Kotaka, M. (2012). Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage. J. Virol. *86*, 12294–12303.

Weber, M., Gawanbacht, A., Habjan, M., Rang, A., Borner, C., Schmidt, A.M., Veitinger, S., Jacob, R., Devignot, S., Kochs, G., et al. (2013). Incoming RNA virus nucleocapsids containing a 5'-triphosphorylated genome activate RIG-I and antiviral signaling. Cell Host Microbe *13*, 336–346.

Wichgers Schreur, P.J., Oreshkova, N., Moormann, R.J.M., and Kortekaas, J. (2014). Creation of Rift Valley fever viruses with four-segmented genomes reveals flexibility in bunyavirus genome packaging. J. Virol. *88*, 10883–10893.

Wright, S., Poranen, M.M., Bamford, D.H., Stuart, D.I., and Grimes, J.M. (2012). Noncatalytic ions direct the RNA-dependent RNA polymerase of bacterial double-stranded RNA virus φ6 from de novo initiation to elongation. J. Virol. *86*, 2837–2849.

Yamaji, Y., Kobayashi, T., Hamada, K., Sakurai, K., Yoshii, A., Suzuki, M., Namba, S., and Hibi, T. (2006). In vivo interaction between Tobacco mosaic virus RNA-dependent RNA polymerase and host translation elongation factor 1A. Virology *347*, 100–108.

Yang, X., Smidansky, E.D., Maksimchuk, K.R., Lum, D., Welch, J.L., Arnold, J.J., Cameron, C.E., and Boehr, D.D. (2012). Motif D of viral RNA-dependent RNA polymerases determines efficiency and fidelity of nucleotide addition. Struct. Lond. Engl. 1993 *20*, 1519–1527.

Yángüez, E., Rodriguez, P., Goodfellow, I., and Nieto, A. (2012). Influenza virus polymerase confers independence of the cellular cap-binding factor eIF4E for viral mRNA translation. Virology *422*, 297–307.

Yap, T.L., Xu, T., Chen, Y.-L., Malet, H., Egloff, M.-P., Canard, B., Vasudevan, S.G., and Lescar, J. (2007). Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. J. Virol. *81*, 4753–4765.

Ye, Q., Krug, R.M., and Tao, Y.J. (2006). The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. Nature *444*, 1078–1082.

York, A., Hengrung, N., Vreede, F.T., Huiskonen, J.T., and Fodor, E. (2013). Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. Proc. Natl. Acad. Sci. U. S. A. *110*, E4238–E4245.

Yumerefendi, H., Tarendeau, F., Mas, P.J., and Hart, D.J. (2010). ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. J. Struct. Biol. *172*, 66–74.

Zamyatkin, D.F., Parra, F., Alonso, J.M.M., Harki, D.A., Peterson, B.R., Grochulski, P., and Ng, K.K.-S. (2008). Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. J. Biol. Chem. *283*, 7705–7712.

Zamyatkin, D.F., Parra, F., Machín, A., Grochulski, P., and Ng, K.K.-S. (2009). Binding of 2'-amino-2'-deoxycytidine-5'-triphosphate to norovirus polymerase induces rearrangement of the active site. J. Mol. Biol. *390*, 10–16.

Zhang, Y., Li, L., Liu, X., Dong, S., Wang, W., Huo, T., Guo, Y., Rao, Z., and Yang, C. (2013). Crystal structure of Junin virus nucleoprotein. J. Gen. Virol. *94*, 2175–2183.

Zheng, W., Olson, J., Vakharia, V., and Tao, Y.J. (2013). The crystal structure and RNA-binding of an orthomyxovirus nucleoprotein. PLoS Pathog. *9*, e1003624.