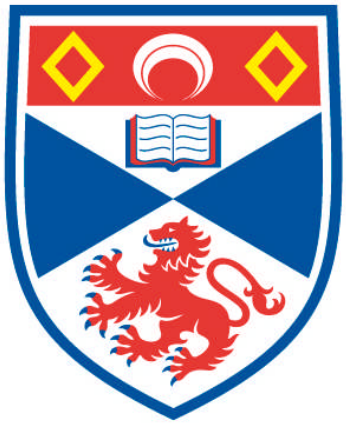
STRUCTURAL AND FUNCTIONAL STUDIES OF NOVEL MECHANISMS OF LASSA FEVER VIRUS NUCLEOPROTEIN IN IMMUNE SUPPRESSION, VIRAL RNA TRANSCRIPTION AND REPLICATION

Xiaoxuan Qi

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2012

Structural and functional studies of novel mechanisms of Lassa fever virus nucleoprotein in immune suppression, viral RNA transcription and replication.

Xiaoxuan Qi

A thesis submitted for the degree of Doctor of Philosophy

University of St Andrews
October 2012

Declaration

I, Xiaoxuan Qi, certify that this thesis, which is approximately 30,000 words in length has been written by me, is the record of work carried out by myself and that it has not been submitted in any previous application for a higher degree.

Date: 16/10/2012 Signature of the candidate:

I was admitted as an MRes student to University of St Andrews in November 2007 and as a candidate for the degree of Ph.D from July 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2012.

Date: 16/10/2012 Signature of the candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 16/10/2012 Signature of supervisor:

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis: Access to Printed copy and electronic publication of thesis through the University of St Andrews.

Date: 16/10/2012 Signature of the candidate:

Signature of supervisor:

Abstract:

Lassa fever virus is one of the most dangerous viruses of *arenaviridae* family, causing more than 500,000 infections per year in Africa. The fatality rate for hospitalized patients is as high as 20%. Due to the high fatality and lack of efficient licensed drugs and vaccines to treat and prevent, Lassa fever virus is classified as a Category A priority pathogen and biosafety level-4 agent by the Centers for Disease Control and Prevention of the USA. Cases were also found in the America and European countries, highlighting its potency to be a bioterrorism weapon.

Like other areanaviruses, Lassa virus has developed a unique interferon suppression mechanism to evade from the host immune system, in which Lassa nucleoprotein plays the key role. To understand the LASV nucleoprotein functions, we tried to determine the first arenaviral nucleoprotein structure, LASV nucleoprotein. The LASV nucleoprotein (NP) was overexpressed and purified. The NP protein was crystallized and the structure was determined to 1.80 Å resolution. The crystals belong to space group *P3*, with the unit cell parameters a = b = 177.16 Å, c = 56.49 Å, α = β = 90° and γ = 120°. The LASV NP structure contains two domains, which are not similar to any reported viral nucleoprotein structures. The N-terminal domain has a novel structure with a cavity, which we proposed for cap binding, and the C-terminus is a 3'-5' ribonuclease, which is responsible for suppressing interferon production. To characterize the possible interaction between NP and other arenaviral protein, we also overexpressed and purified LASV Z. Interesting, both NP and Z proteins have two forms and the purified NP protein and monomeric Z protein bind RNA.

It is surprising that only the oligomeric Z protein interacts with NP protein but the monomeric Z protein does not as determined by Isothermal Titration Calorimetry (ITC). Our studies have reported the first arenaviral nucleoprotein structure, revealed the novel mechanism for the cap binding and immune suppression, which set up a platform for development novel drugs and vaccines to treat deadly arenaviral infections.

Acknowledgements

I would like to express my sincere gratitude to Professor Dr. Changjiang Dong for his supervision, encouragement and guidance throughout my Ph.D study. Prof. Changjiang has not only been a patient teacher, but also a reasonable mentor. The research field he led me into has broadened my sight of the scientific world. Studying and living in a foreign country alone was not easy for me, his advice always brought me back on the right way when I got lost.

I would like to thank people who have given me technical support and friendly discussion, too numerous to name, but particular to Dr. Catherine Botting and her team for mass spectrometry analysis, Professor Yuying Liang's group at Emory University for the collaboration work of the NP protein in vivo assays, the staff working on Diamond light sources synchrotron IO2 and IO3 station and technician in the kitchen of BSRC building.

My gratitude also goes to my family and my friends. I would like to thank you all for sharing my happiness and sadness, my anger and my growth, the work may not be accomplished without your support and unconditional love.

This work was financially supported by a partial Scholarship from School of Chemistry, University of St Andrews, and this research project was funded by Wellcome Trust.

Abbreviations

I. Amino acids

Ala, A	Alanine	Leu,L	Leucine
Arg,R	Arginine	Lys,K	Lysine
Asn, N	Asparagine	Met,M	Methionine
Asp,D	Aspartate	Phe,F	Phenylalanine
Cys,C	Cystine	Pro,P	Proline
Gln,Q	Glutamine	Ser,S	Serine
Glu,E	Glutamate	Thr,T	Threonine
Gly,G	Glycine	Trp,W	Tryptophan
His,H	Histidine	Tyr,Y	Tyrosine
Ile,I	Isoleucine	Val,V	Valine

II. Chemicals

CHES	2-n-cyclohexylamino ethane sulfonic acid
DMSO	Dimethyl sulfoxide
DNAse	Deoxyribonuclease
DTT	Dithiothreitol
EB	Ethidum bromide
EDTA	Ethylenediaminetetraacetic acid
IPTG	Isopropyl thio-β-D-galactoside
LB	Luria Bertani
MES	2-(N-morpholino) ethanesulfonic acid

PBS	Phophate buffered saline
PEG	Polyethylene glycol
Se-Met	Selenomethionine

III. Arenavirus

GPC	Glycoprotein (GP1 & GP2)
GTOV	Guanarito virus
L	RNA dependent RNA polymerase L protein
LASV	Lassa Fever virus
LCMV	Lymphocytic choriomeningitis virus
JUNV	Junin virus
MACV	Machupo virus
NP	Nucleoprotein
SABV	Sabia virus
Z	Matrix Z protein

IV. Symbols

3’	3 prime end (DNA)	5’	5’ prime end (DNA)
a,b,c	Unit cell dimensions	λ	Wavelength
Å	Angstrom	μ	Micro (one millionth)
α,β,γ	Angles of unit cell	V_m	Matthews' coefficient
σ(*I*)	Error in intensity	V_{max}	Maximum velocity
ε	Extinction coefficient		

V. Technical and others

α-DG	α-dystroglycan
β-DG	β-dystroglycan
$2F_o$-F_c	Maximum likelihood/σ_A-weighted $2F_o$-F_c map
3-D	Three dimensions
AP-1	Activator protein
ATF2-c-Jun	Activating transcription factor 2 that interact with c-Jun
B factor	Temperature factor
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
CARDs	Caspase activation and recruitment domains
CBP	CREB binding protein
CCD	Charge coupled devices
CCP4	Collaborative Computing Project Number 4
CPSF30	Cleavage and polyadenylation specificity factor
CTD	C terminal domain
Da	Dalton
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
dTTP	Deoxythymidine triphosphate
EBOV	Ebola virus
EBV	Epstein-barr virus
ED	Effector domain
eIF2α	Eukaryotic translation initiation factor 2

eIf4E	Eukaryotic initiation factor 4E
EP	Eppendorf
ESPRIT	Expression of soluble protein by random incremental truncation
HA	Heamagglutinin
HCMV	Human cytomegalovirus
HF	Haemorrhagic fever
HSV-1	Herpes simplex virus 1
IFN	Interferon
IGR	Intergenic region
IRF3	Interferon-regulatory factor 3
IRF7	Interferon-regulatory factor 7
ISGF3	ISG factor 3
ISRE	IFN-stimulated response elements
ITC	Isothermal titration calorimetry
Jak PTKs	Janus protein tyrosine kinases
LATV	Latino virus
LGP2	Laboratory of genetics and physiology-2
M	Molar concentration
M1	Matrix 1 protein
M2	Matrix 2 protein
MACV	Machupo virus
MAD	Multiple wavelength anomalous dispersion
MALDI-TOF	Matrix-assisted laser desorption time-of-flight
MBP	Maltose binding protein
MDA5	Melanoma differentiation-associated gene 5
m^7G	Methylated guanosine at 7^{th} carbon

m^7GTP	Methylated guanosine-5'-triphosphate at 7^{th} carbon
MR	Molecular replacement
MS	Mass spectrometry
MW	Molecular weight
MxA	Myxovirus resistance gene A
NA	Neuraminidase
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NLS	Nuclear localization signals
NSP1	Non-structural protein 1
NS2	Non-structural protein 2
OAS	2'- 5' oligoadenylate synthetase
OD_{600}	Optical density at 600 nm
p300	E1A-associated protein
PA	Polymerase acidic protein
PAMPs	Pathogen-associated molecular patterns
PB1	Polymerase basic protein 1
PB1-F2	Polymerase basic protein 1-F2
PB2	Polymerase basic protein 2
PCR	Polymerase chain reactions
PDB	Protein Data Bank
Pfu	DNA proof reading polymerase from *Pyrococcus furiosus*
PICV	Pichinde virus
PKR	Protein kinase R
PML	Promyelocyte leukemia protein
PRRs	Pattern-recognition receptors

RBD	RNA-binding domain
RdRp	RNA dependent RNA polymerase
RIG-I	Retinoic acid inducible gene I
RLRs	Retinoic acid inducible gene I like receptors
r.m.s.d.	Root mean square deviation
RNP	Ribonucleoprotein
r.p.m.	Round per minute
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SKI-1/S1P	Subtilisin kexin isozyme-1/site-1 protease
ssDNA	Single stranded deoxyribonucleic acid
ssRNA	Single strand RNA
Stats	Signal transducers and activators of transcription
TCRV	Tacaribe virus
TEV	Tobacco etches virus
TfR1	Transferring receptor 1
TLRs	Toll-like receptors
UTP	Uridine-5’-triphosphate
WT	Wild type
WWAV	Whitewater Arroyo virus

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background information of Arenaviruse

1.1.1 Arenaviruse epidemiology and evolution

Arenaviridae viruses are a big group of enveloped and single-stranded RNA viruses, which have at least 22 members and can be divided into two subgroups the new world and the old world viruses according to the initial isolation places (Charrel et al., 2008). The old world group comprises seven viruses: Lymphocytic choriomeningitis virus (LCMV), Lassa virus (LASV), Mopeia virus, Mobala virus, Ippy virus, Morogoro virus and Kodoko virus, whereas the new world group has 15 viruses and is divided into three lineage: Clade A, B and C (Charrel, 2011). Lassa virus (LASV) is an old world virus that causes Lassa fever (LF) that is an endemic disease in West African countries like Nigeria, Guinea, Liberia, and Sierra Leone. Lassa fever virus infects 100,000 to 500,000 people every year and about 15% to 20% hospitalized patients die from the infection (Khan et al., 2008). It was first found in 1969 when two missionary nurses died in Nigeria, and the virus was named after the town where the first cases were found.

Lassa virus, along with other four new world arenaviruses Junin virus (JUNV), Machupo virus (MACV), Guanarito virus (GTOV) and Sabia virus (SABV) are the most well-known severe haemorrhagic fever (HF) causing arenaviruses. The cases also were reported in America and Europe (Rodrigo, 2011). The four new world arenaviruses that are all belong to the Clade B cause of Argentine, Bolivian, Venezuelan and Brazilian HF, respectively. New emerging arenaviruses that cause HF

like Chapare virus and Lujo virus are recently discovered and studied. These two viruses are found in Bolivia and Southern Africa, respectively.

Furthermore, the prototypic arenaviruse LCMV which was first isolated in 1933 is considered to be a neglected human pathogen and is an important model to study persistent viral infection. LCMV is the causation of central nervous system disease, congentital malformations and has recently been reported to cause deaths in organ-transplant patients (Fischer, 2006; Palacios, 2008). Due to the high mortality and the lack efficient treatments for later infection patients, LAVS, JUNV, MACV, GTOV and SABV are categorized as diseases catalog A pathogens and biosafety level-4 agents by the National Institutes of Allergy and Infection of the USA (Illick et al., 2008).

1.1.2 Transmission and Pathology

Most arenaviruses are transmitted by certain closely related rodent species, in which the viruses could have very high titer in rodent bodies and do not result in any disease. The viral haemorrhagic fever (VHF) these viruses caused is characterized into several degrees along with shock syndrome in the final stage. Both the clinical and the experimental data show that the vascular endothelium participates in the transmission pathway. Lassa virus is typically transmitted by aerosolized virus particles via a rodent of the genus Mastomys natalensis (Table 1.1). Humans can be infected by direct exposure with the host or with food that has been contaminated with feces or

urine from infected mice. Lassa virus may also be spread between humans through direct contact or contact with contaminated urine, blood, feces and other bodily fluids (Fisher-Hoch, 1987). The incubation period is 7-18 days. The typical symptoms include fever, retrosternal pain, cough, vomiting, diarrhea, facial swelling and mucosal bleeding.

The LASV viruses invade into dendritic cells, macrophages, hepatocytes and endothelial cells and replicate themselves by hijacking the cell machineries (Baize, et al 2009). Although the infection leads to little cellular damage, and is insufficient for organ failure like the hepatic failure, it is still causing patients to die as the anti-viral immune response is absent. The viral induced immunosuppression during infection is not well known, and the only antiviral drug, ribavirin, is most likely effective within the first week of infection. However, due to the LASV fever sympathies are very similar to other popular HF diseases in west African and lack efficient diagnostic methods, the LAVS infection patients most likely to miss the best time to be treated using ribavirin. Therefore, the ideal vaccine is supposed to based on a replication competent engineered and most likely to be preventive vaccine (Geisbert, et al 2005). Our research indicates the potential drug target by proclaiming the Lassa virus nucleoprotein structure and its role in the immune responses induced during LASV infection.

Table 1.1 Viral haemorrhagic fever caused by arenaviruses LASV, JUNV, MACV and filoviruses Ebola and Marburg virus (Table from Stefan Kunz, 2009).

Parameter	LASV	JUNV	MACV	Ebola	Marburg
Fever	yes	yes	yes	yes	yes
Hypotension	yes	yes	yes	yes	yes
Haemorrhages	weak[a]	occasional	occasional	occasional	occasional
Lymphopenia	yes	yes	yes	yes	yes
Hepatic lesions	yes	yes	yes	yes	yes
Vascular lesions	no	no	no	no	no
Thrombocytopenia	rare	yes	yes	yes	yes
DIC	no[b]	some[c]	some[c]	yes	yes
Infection of MP	yes	yes	yes	yes	yes
Infection of DC	yes	yes	yes	yes	yes

[a] Haemorrhages in LASV are limited to mucosal surfaces and blood loss is mild. [b] Fibrin deposits have been reported in rare cases. [c] In some cases biochemical evidence for DIC and detection of fibrin deposits.

1.1.3 Genome structure of Arenaviruse

Lassa virus like other arenaviruses is an enveloped virus with a bisegmented ambisense RNA genome that comprises a small RNA segment (S-RNA, 3.4 kb) and a large RNA segment (L-RNA, 7.2 kb). Each segment is reported to direct the synthesis of two viral genes in opposite orientation by ambisense coding strategy, separated by a non-coding intergenic region (IGR) with a hairpin structure. For Lassa virus, the S-RNA is always abundant and encodes the major structural elements – the internal nucleoprotein (NP, 64kDa) and the external precursor polypeptide (glycoprotein) GPC that is posttranslationally cleaved into two viral glycoproteins GPC1 (42 kDa) and GPC2 (38 kDa) by protease SKI-1/S1P. The L-RNA encodes a RNA-dependent RNA polymerase (L, 200 kDa) and a small RING finger protein (Z, 11 kDa) that binds zinc and acts as a matrix protein (Fig. 1.1).

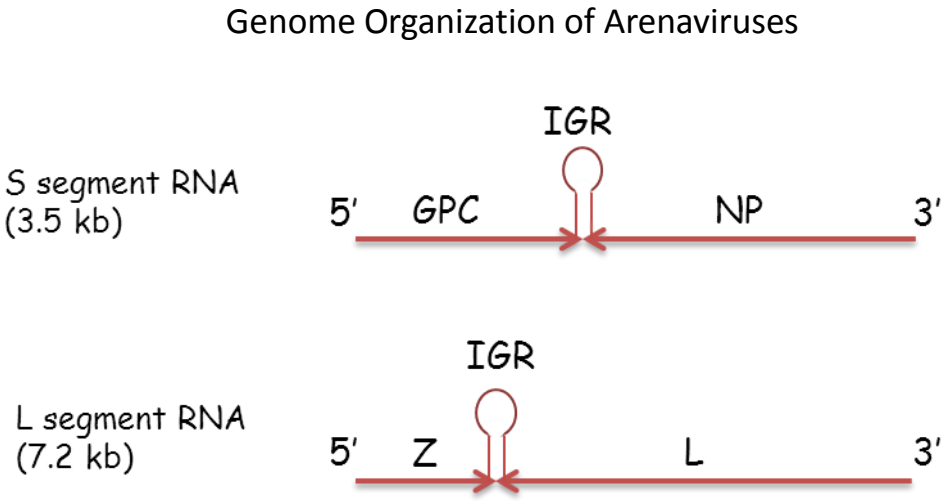


Figure 1.1 The genomic organization of arenaviruses. The two segments encode four viral proteins NP, GPC, L and Z.

There are 19 nucleotides at the 3' and 5' ends of both L and S genome segments are complementary to each other and highly conserved among arenaviruses. They are believed to be the binding sites of the viral polymerase. The nucleotide sequences of L and S IGR are different in most cases, but some IGRs on the same segment are also highly conserved. The IGR is the termination signal during the viral RNA transcription. The NP and L coding regions share the same transcription strategy by being transcribed into a genomic complementary mRNA from viral sense strand 3' end to 5'end. The whole process happens right after the RNP is delivered into the cytoplasm of the infected host cells, and is initiated by the promoter located at 3' end of each segment. There will be a cap structure at the 5' end of the transcribed NP mRNA. Once the transcription of NP mRNA finished, the viral polymerase will then move across IGR to replicate a full length copy of antigenome RNA which will later serves as the template for the synthesis of GPC and Z mRNA (Fig. 1.2) (Emonet, 2009).

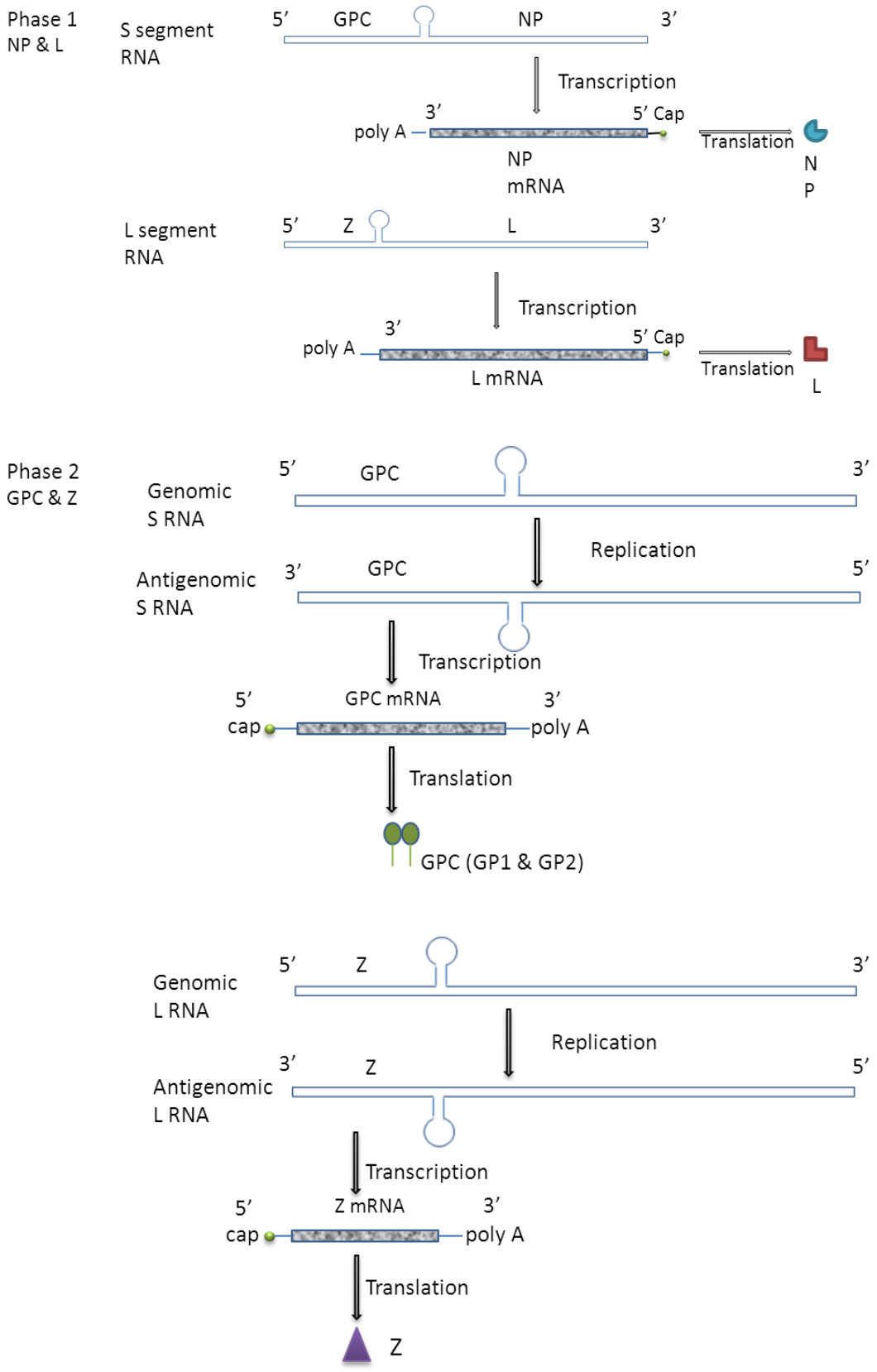


Figure 1.2 The replication and transcription pathways of Arenaviruses. The two segments of arenaviurse use the ambisense coding strategy to replicate the viral genome and transcribe the four proteins with the priority of NP and L protein, and then GPC and Z

proteins (Figure was derived from Emonet, 2009).

1.1.4 Receptors and the infection pathway

The viral infection pathway is initiated by binding the receptors at cell surface. The Old world group arenaviruses like LASV and LCMV use α-dystroglycan (α-DG) as a primary cellular receptor while most of the New World group arenaviuses like JUNV, MACV, GTOV and SABV use transferrin receptor (TfR1) as the cellular cargo receptor. This makes a significant difference between the two groups (Rojek, 2008). Dystroglycan (DG) is encoded by a single gene, and is processed into two products α-dystroglycan and β-dystroglycan. The α-dystroglycan is a highly conserved and ubiquitously expressed cell surface receptor for proteins of the extracellular matrix and is associated with β-DG, which is a transmembrane protein linked to the cytoskeleton. These receptors are key elements during virus infection.

LASV uses an unusual endocytotic pathway that is independent of clathrin, caveolin, dynamin and actin, while most of the new world enveloped viruses use clathrin coated pits for entry cell. The viruses are rapidly delivered to endosomes via vesicular trafficking once they entry the cell. The LASV was only mildly affected by dominant negative mutant of GTPase Rab5 which inhibits vesicular trafficking from the plasma membrane to the early endosome and was independent of Rab7 that is implicated in delivery from early to late endosomes.

1.1.5 Lassa NP

The NP protein interacts with the viral RNA forming RNP, which is vital for the genomic RNA protection, the RNA transcription and replication. In addition, the NP protein of Lassa fever virus also is reported to interact with Z protein in vivo, and interact with host cell factors regulating human immune responses (Eichler et al., 2004; Mahanty et al., 2003). There are also a serial of cell based functional assays designed for the anti-IFN activity of arenavirus NP protein. The studies in Martínez-Sobrido group showed that almost all the NP from Arenaviridae family (beside TCRV) can suppress the induction of type I interferon by inhibiting the translocation of interferon regulator factor 3 (IRF-3). The residue D389 from Lassa NP (corresponding to LMCV D382) is crucial for the anti-IFN activity, but there is no direct evidence to show why these amino acids are involved in this activity (Martínez-Sobrido, 2006; Martínez-Sobrido, 2007; Martínez-Sobrido, 2009). Therefore structural and biochemical studies on the NP and Z proteins are critical for us to understand the important biological processes and help us to find the novel targets for drug and vaccine development, which may help to treat and prevent the infections of lassa fever virus and other *Arenaviruses*.

1.1.6 Lassa GPC, Z and L protein

The pre-glycoprotein is cleaved by the cellular proprotein convertase subtilisin kexin isozyme-1/site-1 protease (SKI-1/S1P) in the endoplasmic reticulum to give two glycoproteins GP1 and GP2 as the final products, which are essential components of

the viral surface (Lenz, 2001). These proteins are essential for the virus to entry into host cells (Garbutt, 2004). The N-terminal GP1 is located at the top of the glycoprotein spike and is responsible for cell attachment that the GP1 binds to the α-dystroglycan receptor on the cell surface (Cao, 1998), while the C-terminal domain of GP2 is the transmembrane domain for fusing the viral particle into the cell. GP1 and GP2 attached to each other noncovalently. The infection requires GP1 dissociation from GP2 at low pH during entry of viral particles to endosome. The dissociation will expose fusogenic domains of GP2 and then trigger membrane penetration (Burns, 1993).

Z protein is the driving force for *Arenavirus* budding (Perez, 2003;Strecker, 2003), and research showed that Z protein alone could form virus-like particles in host cells (Eichler, 2004). Z protein is abundant in viral particles and contains a zinc-binding RING motif. It also contains two proline-rich late domains, PPPY and PTAP, at the C terminus of Z protein. The two domains are involved in the fission of virus particles from the plasma membrane (Strecker, 2003). Although it is only 11 kD, Z protein has been reported to have multiple functions, regulating virus replication (Eichler, 2004), involving in transcription (Garcin, 1993) and interacting with host cellular proteins such as oncoprotein promyelocyte leukemia protein (PML) (Borden, 1998), the ribosomal protein P0 and the eukaryotic initiation factor 4E (Borden, 1998; Kentsis, 2001; Volpon, 2010). However, how the Z protein interaction with the host proteins and what are the mechanisms are still elusive.

L protein is essential for viral replication and transcription. It mediates the synthesis of two viral RNA segments. The protein is split into three structural domains with the central domain harboring the RNA-dependent RNA polymerase (Fig. 1.3).

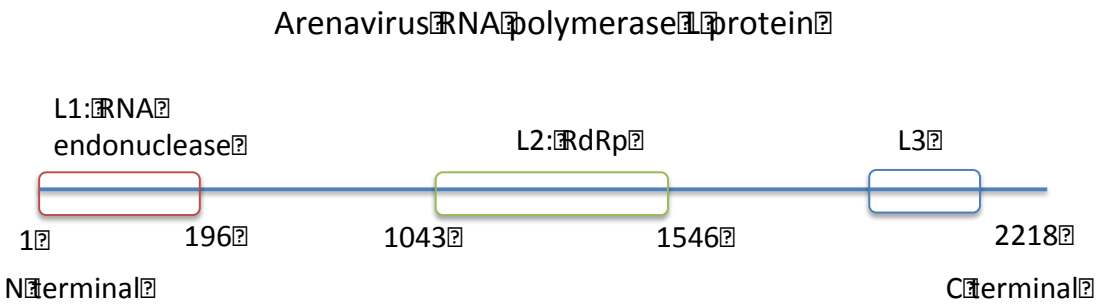


Figure 1.3 Three functional domains of Arenavirus RNA polyermase L protein. L1 domain is the endonuclease domain that may participate in the "cap-snatching" mechanism; L2 domain is the RNA dependent RNA polymerase; L3 is not currently well studied (data from Vieth, 2003 & pfam).

Its oligomerization may be crucial for function like other negative strand RNA viruses (Smallwood, 2002; Smallwood, 2004; Sanchez, 2005; Zamoto-Niikura, 2010). Z protein binds L protein at two binding sites: N terminal residues between 160 and 290 of the first domain and the RdRp (the second) domain, respectively. The first 200 amino acids at the N terminal of L1 domain hypothetically contains an endonuclease corresponding to the N terminus of influenza virus PA which is known to have the endonuclease activity (Lelke, 2010). The structure of N terminal domain of LCMV L protein consists of 1 to 196 residues confirming that the L1 domain if an RNA endonuclease (Morin, 2010). Currently, the function of the second domain and the C terminal of the third domain is unknown, the structure of lassa virus L protein has not been determined, but many biochemical studies have shown its participation in the

"cap snatching" mechanism during viral transcription (Fodor, 2002; Vieth, 2003; Hara, 2006) (see Appendix I.1 for the diagram of LASV virion).

1.2 Cap-snatching

Cap is an essential structure for mRNA transcription and translation, which chemical structure is shown in figure 1.4. The cap structure initiates the translation, protects the mRNA from exonucleases degradation, and it also regulates nuclear export along with the cap binding complex (CBC) for influenza viruses. In influenza A virus, once the mRNA is exported to cytoplasm, the translation factors eIF-4E and eIF-4G take place of the CBC, the translation initiates after ribosome recognize these factors. The cap that consists of a guanine nucleotide is connected with the mRNA via a 5' to 5' triphosphate bridge. The 7^{th} carbon on guanosine is methylated by methyl transferase, which leads the cap is abbreviated as m^7G.

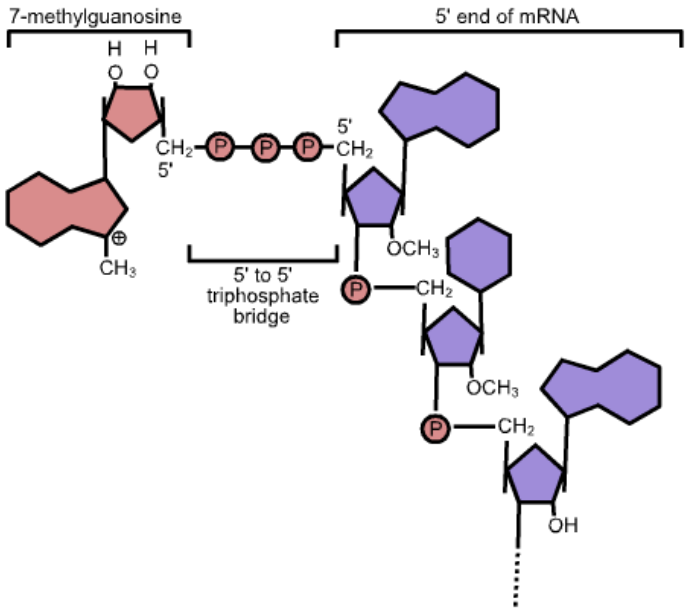


Figure 1.4 The 7-methylguanosine cap structure (colored in pink) that is kinked related to the nucleotides by the 5' to 5' triphosphate bond instead of the phosphodiester bond.

Like influenza viruses and bunyaviruses, arenaviruses can't synthesize mRNA cap themselves. Therefore, arenaviruses have to steal host mRNA caps for viral mRNA translation. Compared to influenza viruses and bunyaviruses cap-snatching, arenaviruses have two special features: firstly, the cap snatch occurs in the cytoplasm; secondly, the length of the cap template that the arenaviruses steal from host cells is normally 1 to 6 nt, but 8 to 16 nt for influenza viruses and bunyaviruses.

1.2.1 Influenza A virus and the cap-snatching mechanism

The most well studied cap-snatching mechanism in negative stranded RNA viruses is the initiation of viral mRNA transcription in influenza A virus (Bouloy et al. 1978; Beaton and Krug et al., 1981). These cap snatching viruses which include

arenaviruses, bunyaviruses,tenuiviruses and orthomyxoviruses are all from segmented negative stranded RNA virus. Unlike non- segmented negative stranded RNA viruses, these viruses are not able to synthesize the cap for its own viral mRNA, so they have to steal the cap either from the host nuclear pre-mRNA or the host cytoplasmic mRNA (Rey 2010). Influenza A viruses are segmented negative-stranded and ambisense RNA virus which belongs to the family of *Orthomyxoviridae* and is most well studied with this mechanism. Beside influenza A virus, other negative-stranded RNA viruses (like Germiston virus or TSWV) are found using the same transcription initiation strategy by using cytoplastic RNA cap as primers rather than nuclear RNA cap (Vialat and Bouloy,1992; Duijsings, 2001).

The genome of Influenza A viruses contains eight RNA segments. These segments encode 11 viral elements: hemagglutinnin(HA), neuraminidase (NA), matrix 1 protein (M1), matrix 2 protein (M2), nucleoprotein (NP), non-structural protein 1(NSP1), non-structural protein 2(NS2), polymerase acidic protein (PA), polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2) and polymerase basic protein 1-F2 (PB1-F2). The influenza A viral RNA polymerase complex which consist of three subunits PA, PB1and PB2 is responsible for replication and transcription of the viral RNA genome in the host nuclei. These three polymerases are the core factors of the virus, which locate at one end of the viral ribonucleoprotein (RNP) along with small amount of nuclear export protein (also called NS2). The viral RNAs are wrapped around the nucleoprotein (NP) (Samji, 2009).

After the virus enters the host cell, the RNP complex that is made up of NP, PA, PB1 and PB2 were released and then enter the host nuclei by using the nuclear localization signals (NLS) (Fig. 1.5). The viral negative sense strand RNAs were converted into positive sense strand RNA to serve as the template for viral RNA transcription. Although the full mechanism of the replication is unknown, many researches show that the RNA dependent RNA polymerase (RdRp) initiates RNA synthesis internally on the vrial RNA.

A mature mRNA is supposed to have both a 5' methylated cap and a 3' poly (A) tail. However, influenza viruses do not encode any protein for cap synthesis. The hypothesis comes up with the cap-snatching mechanism which shows the viral RNA must have stolen the cap from host cellular mRNA. Both structural and biochemical studies later confirmed this idea. In the nuclei, the PB1 subunit holds the viral RNA within its centre, while the PB2 subunit is found to bind the 5'methylguanosine cap of the host pre-mRNA, then the PA subunit as an endonuclease to cleave 10- 13 nucleotides from the cellular pre-mRNA that is used as the primers for viral mRNA synthesis. These capped, polyadenylated positive sense viral mRNA then are exported to the cytoplasm to translate into viral proteins.

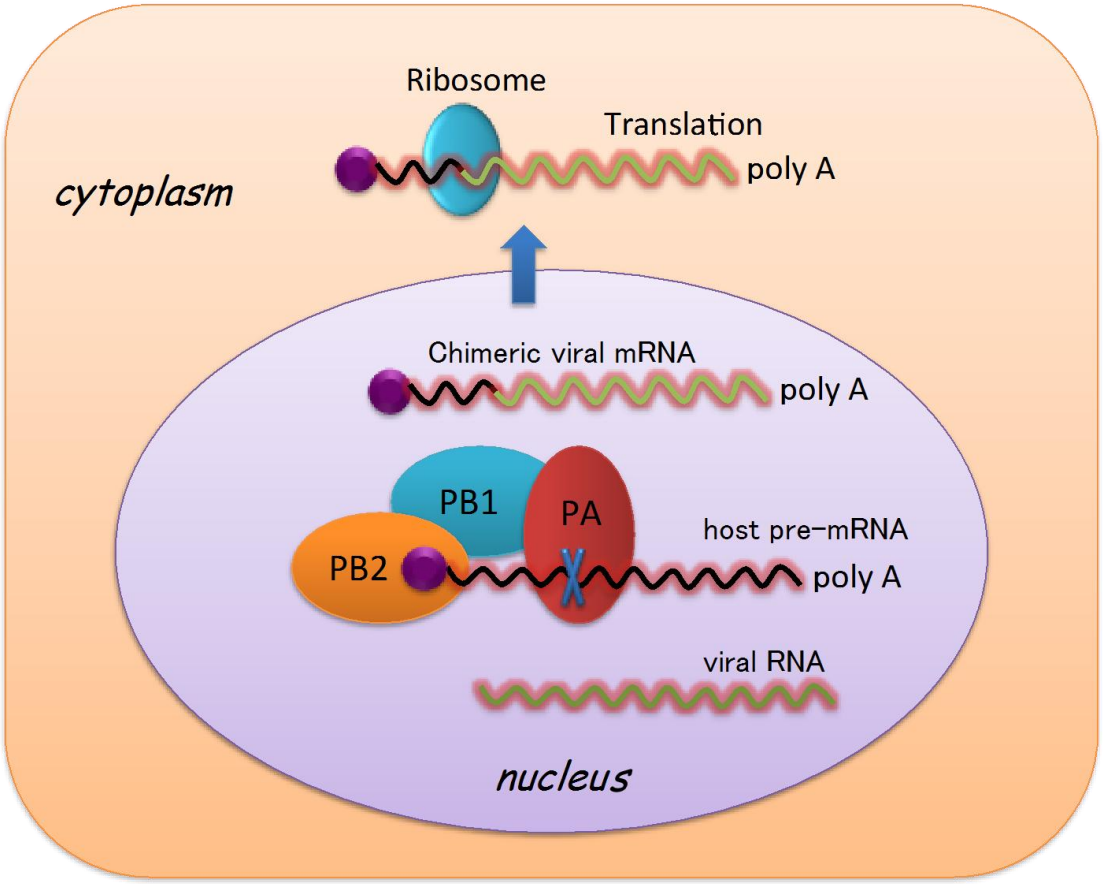


Figure 1.5 Cap snatching mechanism in Influenza A virus. PA, PB1 and PB2 subunits form the heterotrimeric polymerase complex that carries the cap snatching mechanism.

1.2.2 The polymerase subunits' structures reveal the cap-snatching mechanism

1.2.2.1 The cap binding site on PB2 subunit

Altough none of the full length structures of the complex subunits have been solved, the structural domains of half of the trimeric complex PA, PB1 and PB2 have been solved by either x-ray crystallography or NMR, which give us an overview of the polymerases' function.

The full length of PB2 subunit is insoluble, a new technology ESPRIT (Expression of

Soluble Protein by Random Incremental Truncation) provides soluble fragments of PB2 by screening almost 90, 000 fragments, which led to several domains structures (Yumerefendi, Tarendeau et al. 2010). The C-terminal domain of PB2 contains two independent domains: the domain that contains residue 627 and the nuclear localizaiton signal (NLS) domain (Fig. 1.6). The residue 627 is conserved as Glu and Lys in either avian viruses or human, the specific host range is unknown so far. Many studies refer that it's the key factor for PB2-NP interaction in host adaptation (Naffakh, 2000; Crescenzao-C, 2002; Labadie, 2007). A complex structure shows that the N-terminal domain of PB2 subunit (residues 1-37) tightly associated with the C terminal domain of PB1 subunit (residues 678-757) (sugiyama, 2009).

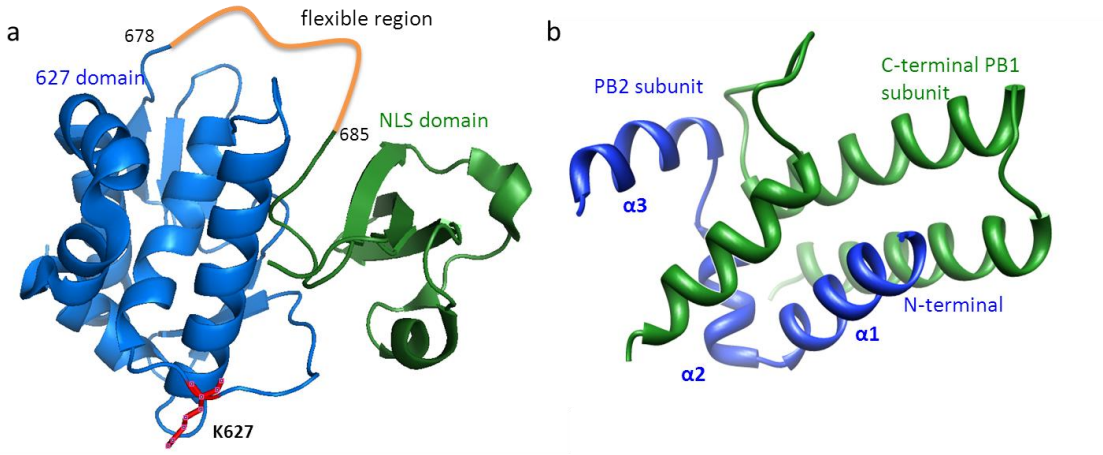


Figure 1.6 The C terminal domain and N terminal domain of PB2. a, the C terminal domain of PB2 consists of two independent domains: the 627 domain (colored in blue) that contains residue Lys 627 (colored in red) that is crutial for host adaptation and the nuclear localization signal (NLS) domain (colored in green) (PDB code: 2VY6); b, the N terminal domain of PB2 subunit (colored in blue) in complex with C terminal domain of PB1 subunit (colored in green) (PDB code: 2ZTT).

PB2 was considered to be the cap binding protein during the initiation of viral transcription since it was found that PB2 can recognize 5' cap (Ulmanen, 1981), the

crystal structure of the central domain of PB2 (amino acid 318 to 483) in complex with cap analog m^7GTP later confirmed that PB2 holds the cap binding site (Guiligay, 2008). The residues that anchor the m^7GTP with putative hydrogen bonds are K339, R355, E361, K376, N429 and H432. The aromatic sandwhich architechture of the typical cap binding proteins are accommodated by Tryptophan and Tryptophan for eIF4E (Marcotrigiano, 1997), Tyrosine and Tyrosine for cap binding complex (CBC) (Mazza, 2002), Tyrosine and Phenylalanine for vaccinia virus VP39 (Hodel, 1998). The positvely charged m^7G methylated base is recognized by PB2 aromatic residues Phe 404 and His 357 that hold the π-electrons in the way of electrostatic interactions, the histidine is not observed in any well characterized cap binding proteins before (Fig. 1.7b). These two residues along with residues Phe 325, Phe 330 and Phe 363 forms the novel hydrophobic region. The acidic residue E361 that is crucial for base specific recognition is also involved in triphosphate stabilization, the acidic cavity holds the interaction between the methylated m^7G and the acidic acid Glu/Asp by both salt bridges and hydrogen bonds (Guilligay, 2008; Fechter, 2005).

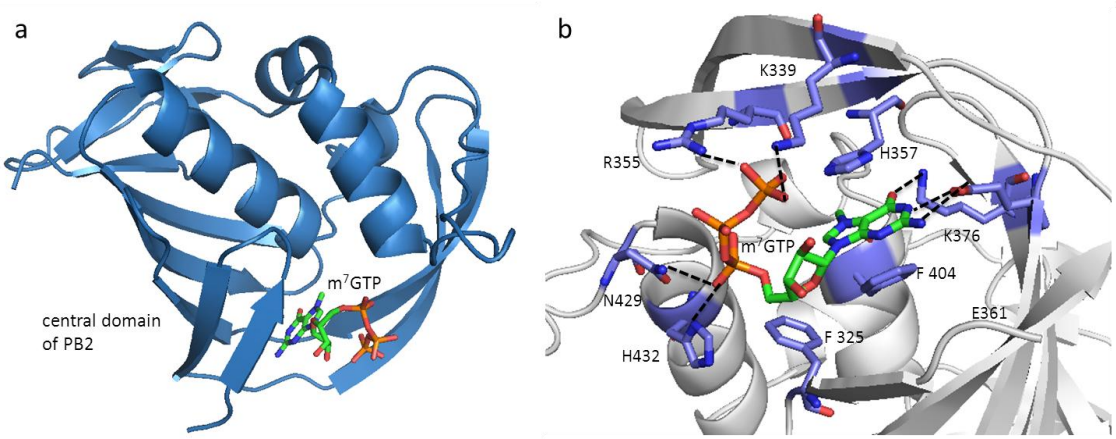


Figure 1.7 The central cap binding domain of PB2 in complex with cap analog m^7GTP. a, the cap binding domain comprises 12 β sheets and 4 α helices (PDB code: 2VQZ); b, the

amino acids that interact with m^7GTP.

1.2.2.2 The endonuclease subunit PA and the PB1 subunit

The structure of PB1 sunbunit has been poorly characterized, although it was previously considered to be the cap cleaving polymerase since it is predicted to have a classic polymerase fold in the center from its conserved motif (Poch, Sauvaget et al. 1989). The C terminal PB1 interacts with N terminal domain of PB2 as previously descirbed, a short peptide on the N-terminal of PB1 was found in a complex with the C-terminal domain of PA subunit, these interactions are essential for the 250 KDa heterotrimeric polymerase complex to function properly and could be the good targets for anti viral drug design (Obayashi, Yoshida, 2008) (Fig. 1.8).

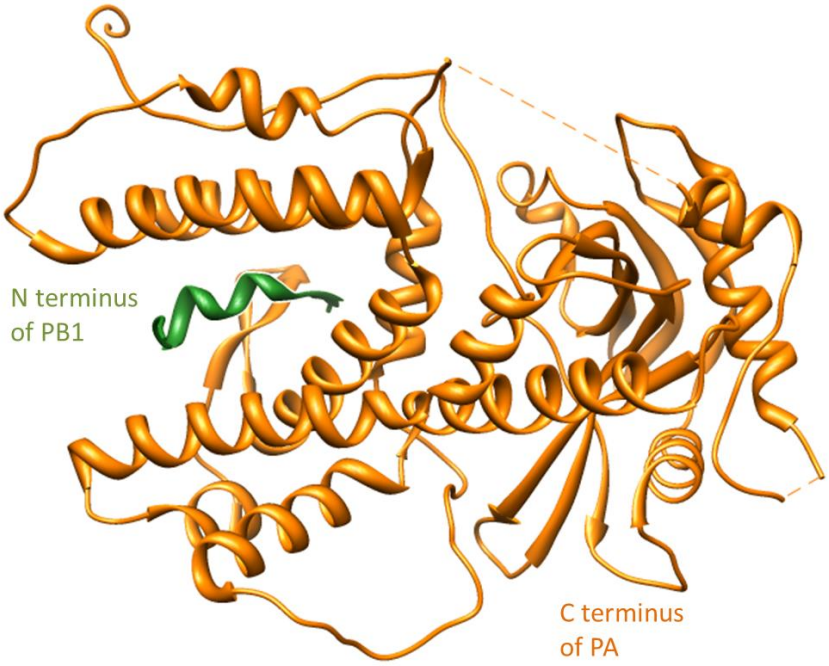


Figure 1.8 The complex structure of N terminus of PB1 (residues 1-81) (colored in light blue) and C terminus of PA (colored in orange) (PDB code: 3EBJ).

The endonuclease PA subunit consists of two domains: C-terminal domain and N

terminal domain. The N-terminal domain of PA subunit has an endonuclease active site which is similar to the nuclease PD-(D/E) XK family. It suggests that the PA subunit holds the function of cleaving host pre-mRNA during cap-snatching rather than PB1 subunit, and the activity is Mg^{2+} dependent (Fig. 1.9). Although the native crystal structure did not provide any evidence on how PA subunit cleaves capped mRNA, the three residues P107, D108, E119 out of the six residues (plus E80, L106 and H41) that coordinate the Mg^{2+} ion align very well with many endonucleases, which suggests that the (P)DX_N(D/E)XK motif of PA N-terminal subunit comprises the endonuclease activity (Yuan,2009; Dias, 2009).

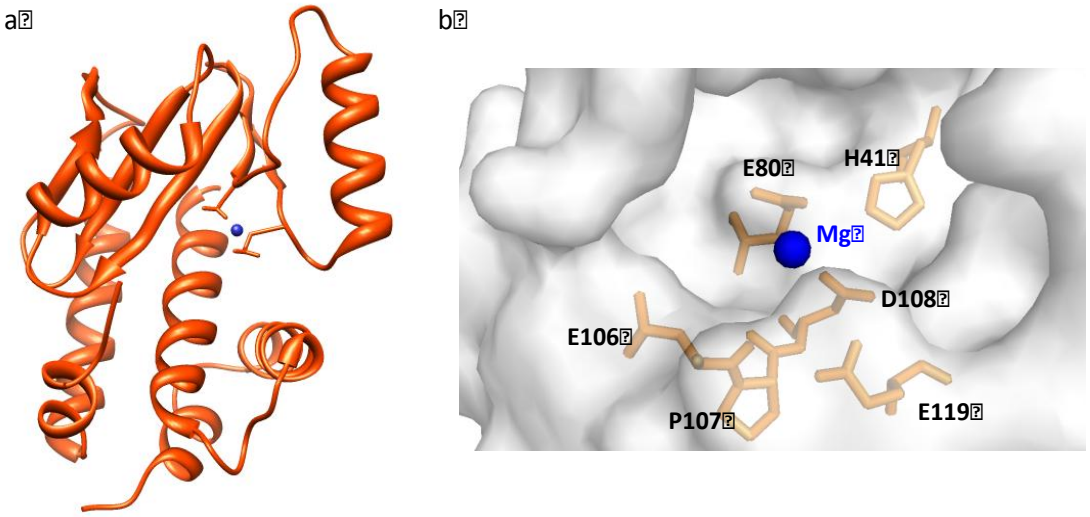


Figure 1.9 The structure of N-terminal domain of PA that reveals the endonuclease active site. Mg^{2+} is colored in blue. The residues around the active site areP107, D108, E80, E106, E119 and H41.

The similar polymerase endonuclease function exits in almost all known negative strand segements viruses like arenaviruses (2 segments), bunyaviruses (3 segments), tenuiviruses (4-6 segments) and orthomyxoviruses (6-8 segments influenza A virus)

by sequence ananlysis (Fig. 1.10). Recently, strucutural evidences of an endonuclease of the N terminal domain of L protein from both La crosse orthobunyavirus and Lymphocytic Choriomeningitis virus are present (Reguera, 2010; Morin, 2010).

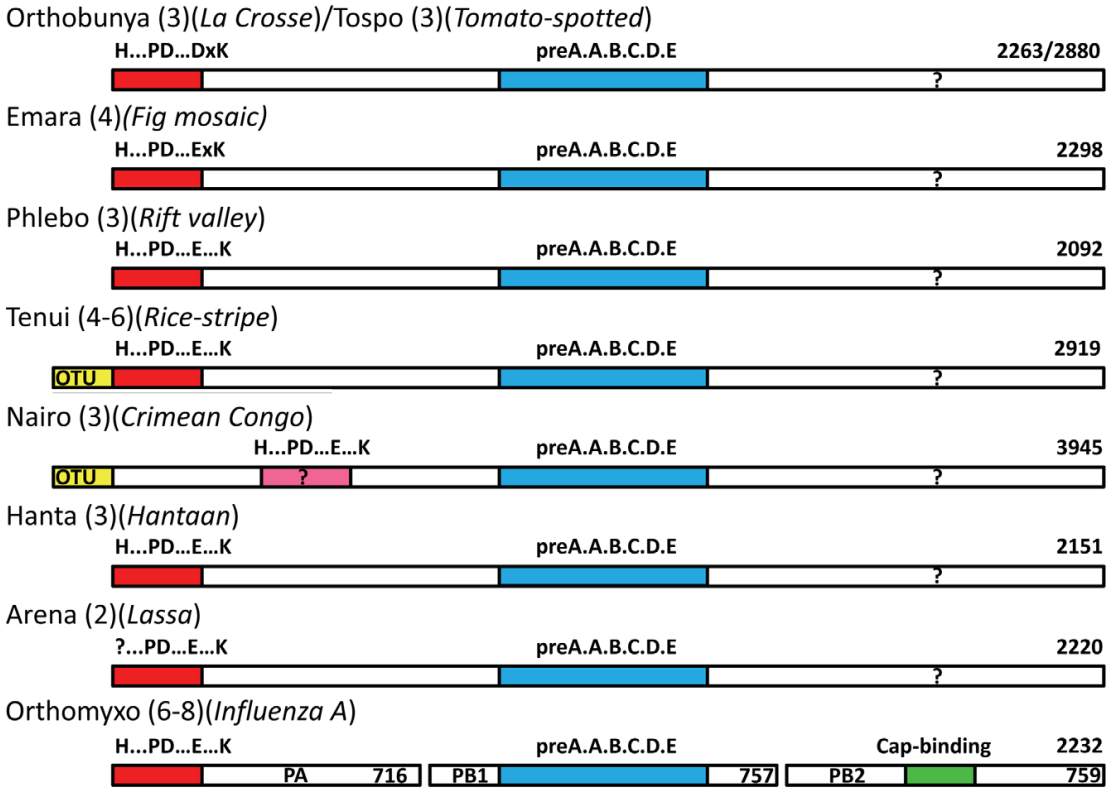


Figure 1.10 Sequence alignment of different negative strand segment viruses. The endonulease domains which are shown in red are all located at N terminus of L protein in each virus beside in influenza A virus it is located at the N termimus of PA subunit. The blue region is the conserved polymerase domain (figure from Reguera, 2010)

1.3 Interferons and immune response

Lassa virus nucleoprotein shows an immune suppression function by suppressing type I interferon induction. In this chapter, the role of interferons in immune response and nucleoproteins in both Old world and New World representative arenaviruses which hold the same function will be given as basic information.

1.3.1 The two immune systems and interferons classification

Immune system is a protecting system for an organism to detect and kill pathogen or tumor cells. Immune system comprises two types, innate immune system and adaptive immune system. Innate immune system acts rapidly upon infection in a non-specific manner while adaptive immune system provides specialized long lasting defences. Normally in higher mammals, the adaptive immune system is very well developed to generate specific antibodies and T cells to either destroy the infected cell or disable the virus function. But this process can take days or weeks for efficient development, so the organism needs to have initial sensors to detect viral infection at early stage. Interferons (IFN) play an important role in innate immune responses and influence some adaptive immune responses, the antiviral activity of IFN was first found in 1957 by Isaacs and Lindenmann (Isaacs, 1957). IFNs have three major biological functions: antiviral activity, antitumour activity and immunomodulatory effects. IFN is classified into two types, Type I IFN and Type II IFN. The discovery of type I, type II interferons and IFN-like cytokines along with their receptors gave us an idea of the innate and adaptive immune response. Type I IFN consists of 7 classes: IFN-α, IFN-β, IFN-δ, IFN-ε, IFN-κ, IFN-τ, IFN-ω. IFN-α has 13 subspecies whereas the other 6 Type I IFN have only one subspecies. Type II IFN has only IFN-γ. Besides these two types, four IFN-like cytokines have been found: limitin, interleukin-28A (IL-28A), IL-28B and IL-29 (Pestka, 2004). Recently, the IL-28A, IL-28B and IL-29 are also classified as Type III IFNλ_1， IFNλ_2 and IFNλ_3, respectively (Kotenko, 2003). All the proteins above can be found in humans except IFN-δ, IFN-τ and limitin (only

in mice). The three types of IFN I, II and III use different signal transduction pathways and their genes are encoded on human chromosome 9, 12 and 19, respectively (Pestka, 2004).

1.3.2 Type I IFN expression and signal transduction pathway

Various NP proteins from Arenaviridae family have been documented for their interferon inhibition function (Martines-Sobrido, 2006; Martines-Sobrido, 2007). Since almost all the NPs inhibit induction of Type I IFN with a notable except of Tacaribe virus. A brief introduction of Type I IFN induction process, signal transduction pathway will be given.

1.3.2.1 Induction of IFN expression

The host pattern-recognition receptors (PRRs) which are found in innate immune cells can recognize the pathogen-associated molecular patterns (PAMPs) like viral genomic single strand RNA, double strand RNA and DNA to serve as initial detection of viral infection. Toll-like receptors (TLRs) and retinoic acid inducible gene I (RIG-I) like receptors (RLRs) are the two main classes of PRRs which can activate the antiviral response including induction of Type I IFNs and Inflammatory cytokines. TLRs are transmembrane proteins that can detect virus on cell surface or within endosomes, whereas RLRs function in cytoplasm as viral sensors. There are three known members of the RLRs family: RIG-1, melanoma differentiation-associated gene 5 (MDA5) and laboratory of genetics and physiology-2 (LGP2). Both RIG-I and

MDA-5 contain a DExD/H box helicase domain that will recognize viral RNAs, the difference is that RIG-I preferentially recognizes 5’ end phosphorylated single strand RNA and double stand RNA while MDA-5 binds 5’non-phophorylated long dsRNA. Because of this difference, RIG-1 and MDA-5 are responsible for recognizing distinct types of viruses (McCartney, 2009).

Type I IFN-α and IFN-β that were identified as a prototype of many cytokine families. IFN- α and IFN-β had not been reported until 1978 (Berthold, 1978; Rubinstein, 1978; Pestka, 2007). Although the main function of type I IFNα/β is on generating an antiviral state, more evidences have shown that they are also involved in other actions like anti tumor activities and the control of cell growth. Expression of type I IFN is regulated at transcriptional level by transcription factors including NF-κB, ATF2-c-Jun, interferon-regulatory factor 3 (IRF3) and IRF7.

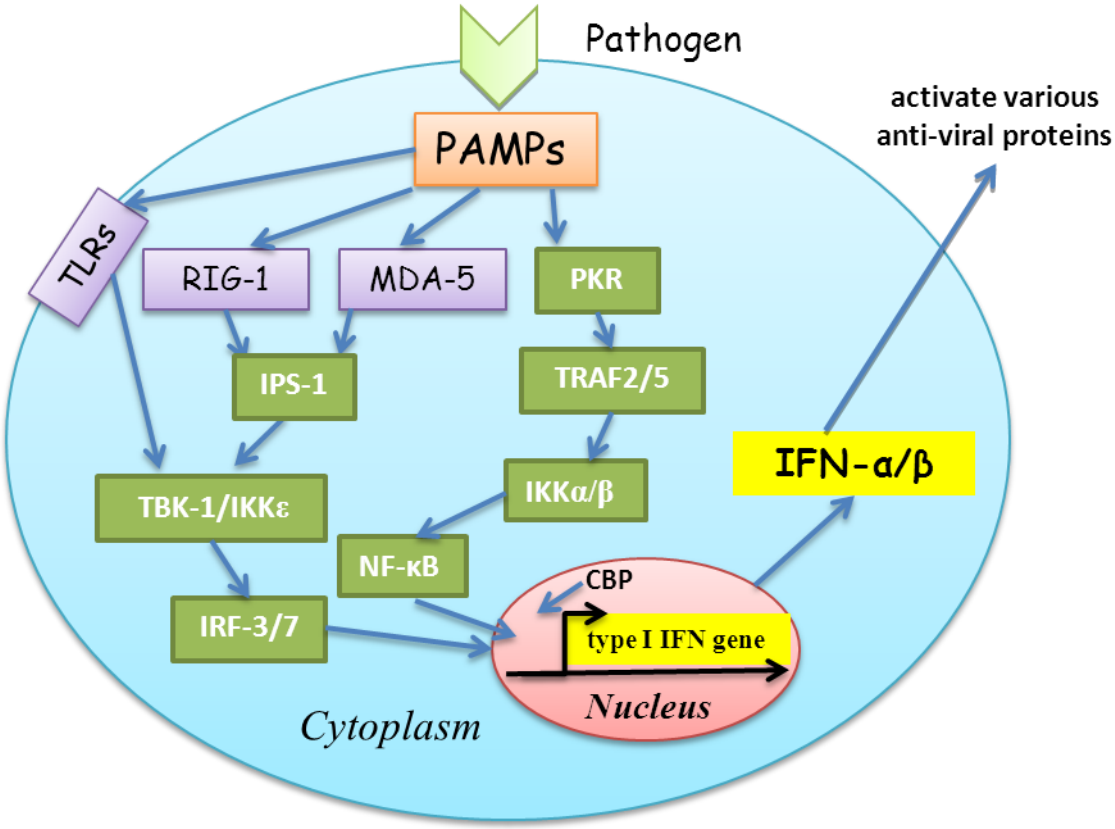


Figure 1.11 The induction pathway of type I IFN. Viruses infect cells and produce different PAMPs molecules. The cellular receptors can detect different PAMPs and trigger different pathways to produce IFNs, and IFNs can activate different interferon stimulation proteins, which can control the viral infections.

Nucleoprotein from arenavirus lymphocytic choriomeningitis virus (LCMV) has been documented for its inhibition of type I interferon production by blocking IRF-3 (Martinez-Sobrido, 2007). IRF-3 is located in cytoplasm in an inactive state, it will be activated by phosphorylation. The C-terminal domain of IRF-3 is phosphorylated by kinases IKK-ε and TBK-1to induce IRF-3 dimerization, then IRF-3 is translocated to nuclei, form a complex CREB-binding protein CBP/p300 which later initiates IFN-β transcription (Yoneyama, 1998). The two kinases are activated by RIG-1, MDA-5, or probably some Toll-like receptors (TLRs) via IPS-1. Unlike IRF-3, IRF-7 needs to be induced by IFN since its expression level is low in most cells. There are also

transcription factors ATF2-c-Jun (AP-1) and NF-κB that are activated by the PKR pathway, which are promoted as a more general stress response. All these factors together can up-regulate IFN-β mRNA synthesis. IRF-3 is thought to be responsible for initiating IFN-β induction rather than IFN-α, whereas IRF-7 can activate both IFN-α and IFN-β induction. Homodimer IRF-7 or heterodimer IRF-3 and IRF-7 play the major role for induction of IFNα/β rather than the homodimer IRF-3 alone, which means IRF-3 has little contribution without IRF-7 (Honda, 2005a; Honda 2005b) (Fig. 1.11).

1.3.2.2 Type I IFN signal transduction pathway

The IFN and IFN like molecules use the Jak-Stat signal transduction pathway. For type I IFN, the receptor consists of two subunits: IFN-αR1and IFN-αR2 that are associated with janus protein tyrosine kinases (Jak PTKs) Tyk2 and Jak1, respectively (Uze, 1990). The initiation of type I IFN signal transduction is the activation of the receptor associated Jak PTKs. These kinases will then phosphorylated their downstream substrates Stat1 and Stat2 that are the two members of the signal transducers and activators of transcription (Stats) family. Although Stat1 and Stat2 are the two main factors used in IFN-α and IFN-β signal pathway, Stat3, Stat4, Stat5 and Stat6 also have been reported in the same pathway under diverse conditions. Stat 3 and Stat5 are reported to be pre-associated with Stat1. The activation of Stat4 and Stat6 are specialized to certain cell types like endothelial cells or lymphoid origin cells (Fasler-Kan, 1998; Torpey, 2004).

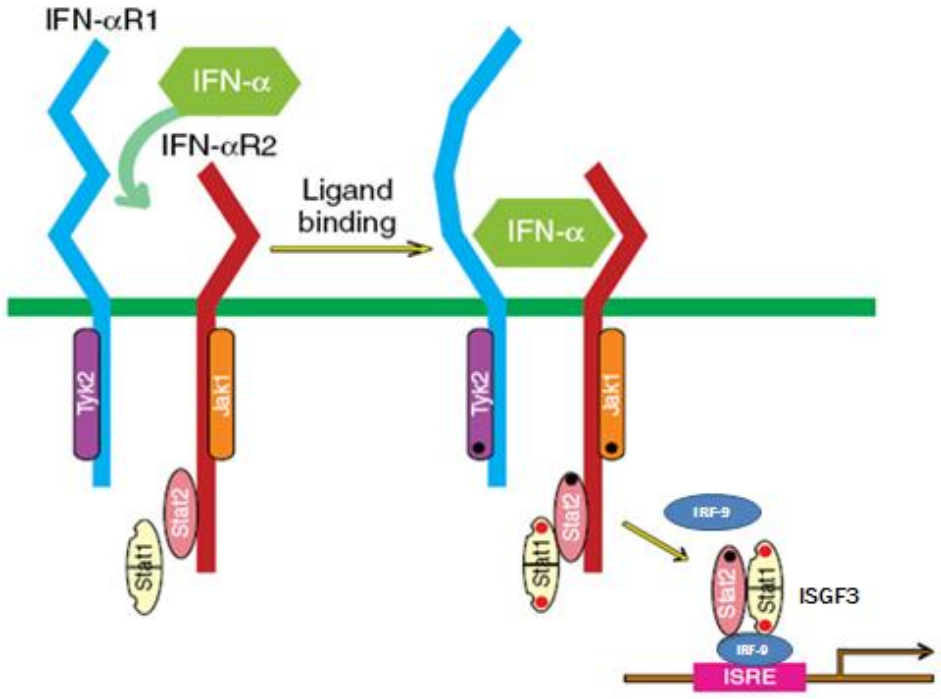


Figure 1.12 The complex of Type I IFN receptors (picture modified from Pestka, 2004).

Phosphorylated Stat1 and Stat2 dimerize and recruit a non-phosphorylated factor IRF-9 to form a heterotrimer ISG factor 3 (ISGF3) that is translocated into nuclei, then initiate the transcription of numerous ISGs by binding IFN-stimulated response elements (ISRE). These antiviral ISGs proteins can limit the virus replication in cells. In influenza viruses, the most well studied antiviral proteins are: dsRNA-activated protein kinase PKR, 2'-5'oligoadenylate synthetase (OAS), myxovirus resistance gene A (MxA), viperin and IFN-stimulated gene ISG15. Besides this classic ISGF3 complex inducing signal pathway, there are also several complex receptors that induced by type I IFN that will not be discussed further (Fig. 1.12).

1.3.3 Virus evasion from IFN system

Human cells have different receptors to detect viral infections, triggering different pathways to produce interferons, which are the first line for defending viral infections. However, different RNA viruses have developed different strategies to evade human immune responses, for examples, influenza A virus, human herpesvirus and arenavirus whose evasion pathways are described in the following paragraphs.

1.3.3.1 Influenza A virus

Type I IFN has been reported for its critical role in responding to the influenza A virus infection. Influenza A virus counteracts with the host defence system by primarily targeting either the IFN induction or signalling cascades. The non-structural protein 1(NSP1) of influenza A virus limits innate immune responses with several strategies: NS1 protein limits RIG-I mediated IFN-β promoter by inhibiting the activation and translocation of the transcription factors IRF-2, NF-κB and ATF2-c-Jun, which prevents PAMPs from being detected by the immune defense system; NS1 regulates host cell gene expression by binding cleavage and polyadenylation specificity factor (CPSF30) to prevent it from processing 3’ end of pre-mRNA into polyadenylated mature mRNA; influenza A virus replicates in nuclei, which prevent the RNA PAMP from being detected by sensors in cytoplasm; NS1 also inhibits antiviral proteins like 2’-5’ OAS/RNase L and TRIM25/RIG-I mediated IFN production.

1.3.3.2 Immune evasion for human herpesviruses

In herpes simplex virus 1 (HSV-1), viral proteins ICP34.5 and ICP0 may inhibit type I IFN induction by targeting cellular TBK1 and IRF3, respectively. Its ICP27 protein effects on IFN-α dependent signal pathway by inhibit STAT1 phosphorylation. The herpes Us11 protein interferes with the host defence system by blocking the antiviral protein OAS (Vandevenne, 2010). The Human cytomegalovirus (HCMV) evades the immune system by disrupting the IFN-β expression or the IFN-α signal transduction pathway. Its viral proteins pp65 and IE86 act as IFN-β antagnists by dephosphorylating IRF-3 and reducing NF- κB activation (Abate, 2004; Sanchez, 2002). The viral proteins of Epstein-barr virus (EBV) are also found to inhibit both IRF-7 and IRF-3 to prevent target gene expression or binding with responsive promoters (Bentz, 2010).

1.3.3.3 Arenaviruses

The nucleoprotein of prototypic arenavirus Lmphocytic Choriomeningitis Virus was the first viral protein in the *Arenaviridae* family found to inhibit the type I IFN expression by an early block in IRF-3 activation pathway (Martinez-Sobrado, 2006). The experiment result shows LCMV NP blocks the activation and nuclear translocation of Sendai virus induced transcription factor IRF-3 to interefere with IFN-β promoter expression rather than directly disrupt IFN singnalling. But the limited expression of IFN-β will later cause the low level of antiviral proteins ISGs expression. The NPs from both Old World and New Wold arenavirus except Tacaribe

virus (TCRV) are all found to have suppression on type I IFN-β by inhibiting IRF-3. The NPs tested for the inhibition function are from old world arenaviruses LASV and LCMV, and new wold arenaviruses Junin virus (JUNV), Whitewater Arroyo virus (WWAV), Pichinde virus (PICV), Machupo virus (MACV), TCRV, Latino virus (LATV) and plus NS1 from influenza virus (Martinez-Sobrado, 2007). But so far, there is no structural evidence of any NPs from Arenaviridae family to support this theory. Structure of NP from Lassa virus then becomes the first discovery of both IFN immune suppression and cap snatching mechanism in Arenavirus.

1.3.4 Structural bases for viral immune suppressions

Many viruses evade from the host immune system by antagonizing antiviral signalling pathways. Both Ebola virus (EBOV) and Influenza A virus are found to bind dsRNA, this binding will consequently antagonize the host PAMP sensors to recognize the viral intruder and suppress the innate immune response.

1.3.4.1 Ebola virus VP35

Ebola virus is a nonsegmented, enveloped, negative-strand RNA virus that causes severe hemorrhagic fever. VP35 is one of the seven proteins of Ebola viruses, whose C-terminal interferon inhibitory domain binds dsRNA to prevent host receptors RIG-I and MDA-5 to recognize the pathogen intrusion. VP 35 is also a multifunctional protein in mediating viral replication (Cárdenas et al., 2006) and nulceocapsid formation, blocking interferon regulatory factor 3 (IRF-3) phosphorylation and

inhibiting RNA dependent protein kinase R (PKR) activation (Schumann, 2009). The complex structure of viral protein VP35 and dsRNA that forms a symmetric dimer shows two distinct ways of molecule binding: one monomer (B) binds dsRNA backbone and caps the blunt terminus of the dsRNA, where the other monomer (A) is engaged in dimerization and binding only the phosphate backbone of dsRNA (Kimberlin, 2010; Leung, 2010) (Fig. 1.13).

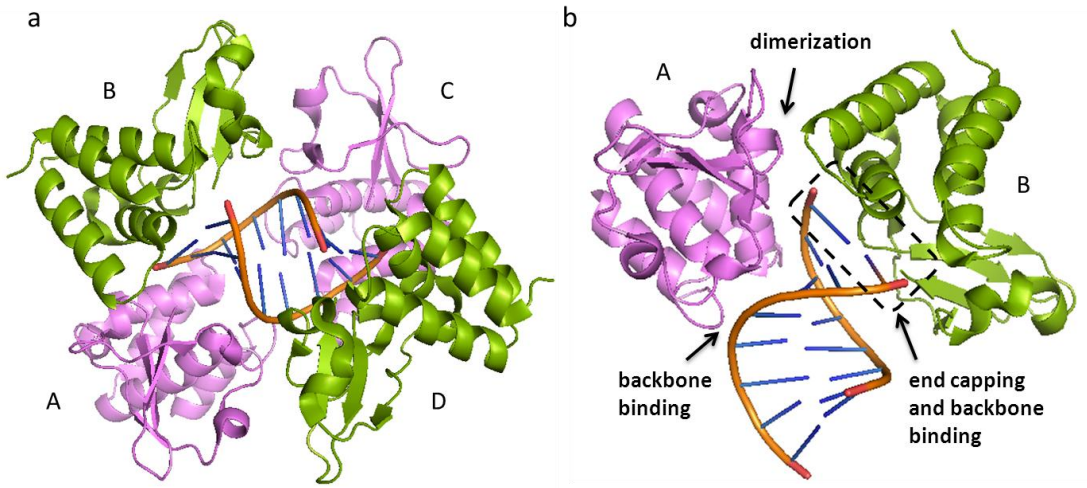


Figure 1.13 Complex of dsRNA and Ebola virus VP35. a, the complex of dsRNA and two dimers of VP35 with four molecules A, B, C, D; b, the picture shows a dimer VP35 with molecules A and B binds an 8-bp dsRNA oligomer. The pink monomer binds the phosphate backbone of the dsRNA; the green monomer caps the blunt end of dsRNA (PDB code: 3L26).

In molecule B, residues R305, K309, K319 and K339 are involved in dsRNA binding, residues R312 and R322 bind directly to the phospohdiester bond of dsRNA. In molecule A, residues R312, R322 and K309 do not contact with dsRNA, these residues instead form the dimerization interface between the two proteins. Another unique finding in this complex structure is that the central basic patch also

accommodates the 5’ end cap of the dsRNA by several conserved hydrophobic residues (Fig. 1.14).

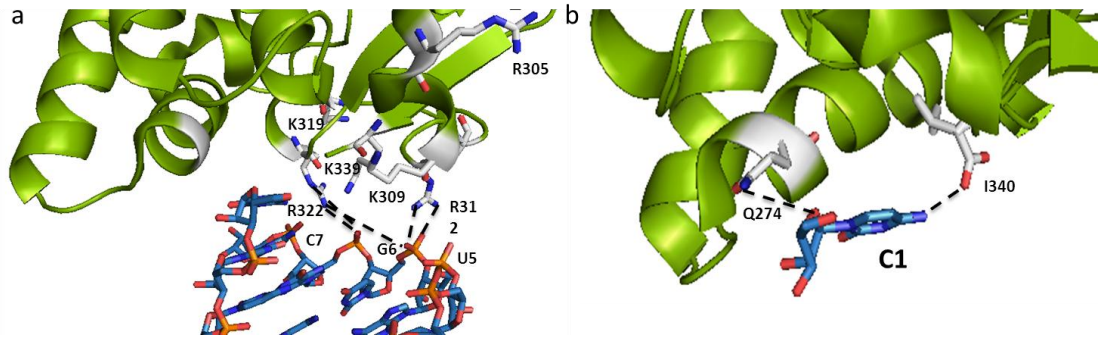


Figure 1.14 The residues of VP35 that are involved in phosphodiester bond binding and dsRNA 5’ terminal capping. a, residues R305, K309, R312, R322, K319 and K339 are key for binding the backbones of the dsRNA oligomer; b, residues Q274 and I340 are critical for binding 5’ base of dsRNA (PDB code: 3L26).

Vertebrates have developed various receptors to recognize viral intrusion, retinoic acid inducible gene I (RIG-I) is one of the innate immune patterns recognition receptors that induce type I interferon expression. RIG-I consists of a DECH helicase and C-terminal domain that bind dsRNA and N-terminal domain that has two caspase activation and recruitment domains (CARDs). Once dsRNA and ATP bind to C-terminal domain (CTD) and helicase domain, the structural conformation of RIG-I is altered to release CARDs on N-terminal domain for downstream signalling. The recently published complex structures of RIG-I and blunt end 5’ppp dsRNA reveal that C-terminal domain of RIG-I caps the terminal base of dsRNA, whereas the helicase and repressor domain binds the phosphate backbone of dsRNA (Wang, 2010; Kowalinski, 2011) (Fig. 1.15). RIG-I is a dsRNA dependent ATPase, lacking the CTD significantly reducing its ATPase activity (Cui, 2008). The C-terminal domain

of RIG-I caps the terminal base of blunt end 5'ppp dsRNA, the electrostatics map shows the capping site that consists of His847, Lys849, Lys861, Lys858 and Phe853 are positively charged. The helicase domain interact mainly with 3' strand, whereas the CTD stacks the 5' end of dsRNA.

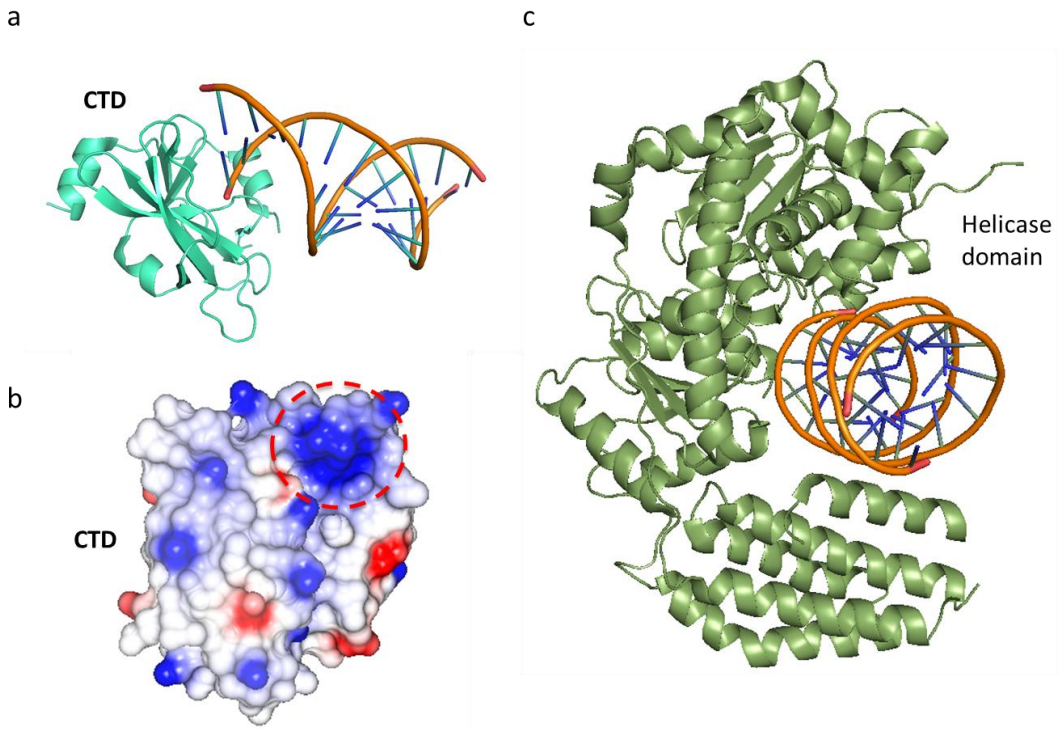


Figure 1.15 Structures of RIG-I in complex with blunt end 5'ppp dsRNA. a, complex structure of RIG-I CTD and 14mer blunt end 5'OH dsRNA (PDB code: 4A2X); b, electrostatics potential map of RIG-I CTD, the capping region is marked with red dash line, positively charged surface is colored in blue and negatively charged surface is colored in red; c, the complex of RIG-I helicase domain binding the backbone of 19mer dsRNA (PDB code: 4A36).

Combining with the in vitro and in vivo functional studies, EBOV VP35 inhibits IFN-β promoter activation by shielding the 5' base of the blunt end dsRNA and binding the backbone of dsRNA in a similar way as RIG-I. Viral protein VP35 binds with dsRNA preventing the activation of RIG-I, which contributes to the evasion of host innate immune response of Ebola virus.

1.3.4.2 NS1 protein of Influenza A virus

The non-structural 1 (NS1) protein in Influenza A virus has been reported to have multiple functions in both innate immune response and adaptive immune response. NS1 has two functional domains that are N terminal RNA-binding domain (RBD) and C terminal effector domain (ED). NS1 regulates the posttranscriptional steps by inhibiting host mRNA nuclear export, pre-mRNA splicing and activation of dsRNA activated protein kinase (PKR). The binding between NS1 RBD and dsRNA blocks the activation of PKR that is responsible for phosphorylation of the α subunit of the eukaryotic translation initiation factor 2 (elF2α). The virus consequently is shielded from the antiviral response by decreasing viral and cellular protein synthesis (Lu, 1995). The structure of NS1 RBD-dsRNA complex shows the dimeric six helical fold NS1 RBD recognizes the major groove of dsRNA with the special Arg38-Arg38 and Arg35-Arg46 pairs (Fig. 1.16).

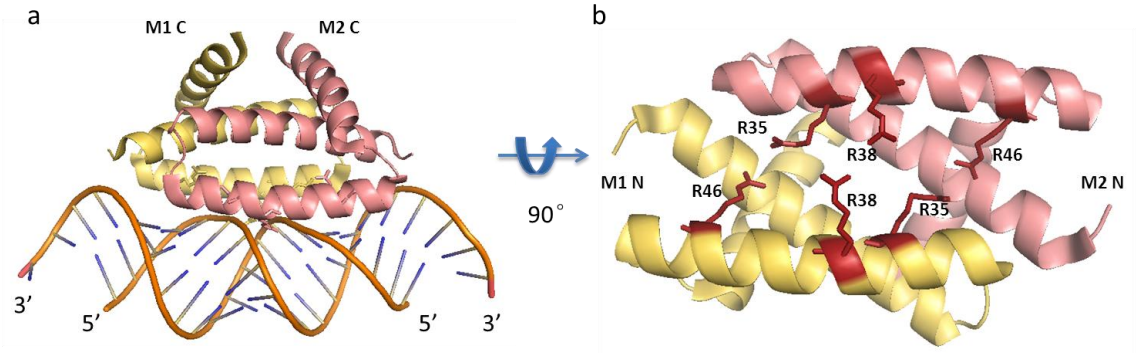


Figure 1.16 The complex of dimeric NS1 RNA-binding domain and dsRNA. a, the dimeric NS1 with monomer 1 colored in yellow and monomer 2 colored in pink; b, residues Arg 35, Arg38 and Arg46 that are crucial for protein and dsRNA binding are colored in red (PDB code: 2ZKO).

The residues on the interaction surface (Arg35, Arg38, Ser42 or Thr49) of NS1 RBD distinguish dsRNA and dsDNA by recognizing the sugar phosphate backbone and 2'-OH groups on the RNA strand either directly or via water bridges. The Isothermal Titration Calorimetry (ITC) assay shows that the R35AR46A double mutant and R38A mutant lose dsRNA binding function. These residues may also be critical for the formation of NS1 RBD dimerization (Cheng, 2009). Unlike VP35 and RIG-I, NS1 shields viral RNA only by binding the phosphate backbone, which suggests a different way for influenza A virus evading from host innate immune system.

1.4 Objective of this thesis

Arenaviral nucleoprotein encapsidates with viral genomic RNA forming ribonucleoprotein complex, playing essential roles in viral RNA replication, transcription and immune suppression. However, little is known about the molecular mechanism. The objective of this thesis is to elucidate the molecular mechanism by determining LASV NP protein structure and carrying out biochemical and cellular studies.

Chapter 2: Lassa virus NP protein over-expression and purification

To solve the structure of LASV NP, the NP protein was prepared. In this chapter, we are going to discuss how to make the expression construct by molecular cloning, and how to express, purify and identify LASV NP protein.

2.1 Preparation of Lassa NP expression construct

The entire NP gene of Lassa fever virus (Josiah strain) that is subcloned into pLOU3 vector (see Appendix I.2), which is a derivative plasmid from pMAL-C2X, containing a 6His-tag at the N-terminus of MBP and the Tev proteinase cleavage site between the MBP and target protein. The plasmid pCAGGS containing the LASV NP gene was provided by our collaborator Dr. Yuying Liang from Emory University, USA.

2.1.1 The NP gene was obtained from the original plasmid

The original plasmid pCAGGS that contains the NP gene was digested by restriction endonucleases EcoRI and XhoI (restriction enzymes from Promega/New England Biolabs). Briefly, the NP gene is cut from the plasmid in 60 µl digestion solution containing 50µl plasmid, 6µl Buffer H, 1µl EcoRI, 1µl XhoI, 1µl BSA and 1µl H_2O. The digestion was carried out at 37 ℃ for 1.5 hours and the NP gene was separated and purified using 1% agarose gel electrophoresis in TAE buffer. The pLOU3 plasmid was prepared using the same way as the NP gene by EcoRI and SalI endonucleases except the plasmid had an additional treatment with 1 µl Shrimp phosphatase at 37 °C for 30 min, which can reduce the plasmid self-ligation. The DNAs in the agarose gel were stained with ethidium bromide (EB) and visualized using a UV transilluminator

(UVP BioDoc-It system).

2.1.2 Gel extraction

Both the NP gene and the pLOU3 plasmid treated with the enzymes were isolated from the 1% agarose gel using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions. The gel slices containing proper DNA fragment were mixed with 3 volumes of Buffer QG and the mixture was incubated at 50°C till the gel was completely dissolved. Afterwards, 1 volume (equal to the weight of the gel) of isopropanol was added to the mixture and the sample was applied into the QIAquick Spin Column. After 1 minute centrifugation, the DNA was separated from the flow-through contaminants bound to the column. The column was washed once with 0.5ml of Buffer QG to remove remained agarose and twice with 0.8ml of buffer PE to remove contaminants. Finally the DNA was eluted using 50µl sterilized ddH_2O.

2.1.3 Ligation

The NP gene fragment was inserted into the pLOU3 plasmid by T4 DNA ligase (Promega) and the reaction was carried out in 20 µl solution containing 3µl NP fragment, 5µl pLOU3, 1µl T4 ligase, 2µl T4 ligase buffer (×10) and 9 µl H_2O at 16°C overnight.

2.1.4 Preparation of competent cells

DH5α/Tam1 E.coli cells were incubated from 1ml overnight culture in 100ml LB

(10g/l Bactotryptone, 10g/l NaCl, 5mM NaOH, 5g/l yeast extract) for 2 hours at 37°C in incubator (New Brunswick Scientific). The cultures then were chilled on ice for 30 mins before transfer into two prechilled 50ml falcon tubes. The cultures were pelleted at 2800 rpm for 5 mins at 4°C, the supernatant was discarded and the tubes were put up side down to drain away the water. The pellets were resuspended in 12.5ml 100mM $CaCl_2$, 2.5ml 40mM $MgSO_4$ and prechilled autoclaved glycerol was added to 10% final concentration (v/v). The cells were centrifuged at 2800 rpm for 5 mins at 4°C. The pellets were then resuspended in 2ml of ice cold 0.1M $CaCl_2$. The component cells were aliquoted into 0.2 to 0.5ml and flash frozen cells in liquid nitrogen and stored in ultra low temperature freezer (New Brunswick Scientific) at -80°C.

2.1.5 Transformation

The ligation products were transformed into E.coli strain DH5α. 50µl competent cells were thawed on ice, and 5µl of ligated plasmids were added into the competent cells. The mixture was incubated for 30 minutes on ice and heat shocked at 42°C for 45 seconds. Then cells were incubated for 2 minutes on ice, and 100µl LB medium was added and cells were incubated at 37°C for 1 hour at 200rpm. The cells were spread on an agar plate containing 50 ug/ml Ampicillin (Melford laboratories) and grown overnight at 37 °C.

2.1.6 Evaluation of NP expression plasmid

In order to check whether the correct NP gene was inserted into the pLOU3 plasmid, two single clones were picked from the transformation plate, cultured separately in 10ml LB medium in present of 100 ug/ml Ampicillin overnight at 37 °C. The plasmids were purified from the overnight culture using QIAprep miniprep Kit according to the manufacture's instruction. Briefly, the 10 ml of E.coli overnight culture was transferred into a 50 ml test tube and centrifuged at 4,000 rpm for 10 minutes at room temperature to harvest the cells. The cell pellet was resuspended in 200µl Buffer P1 and 250µl Buffer P2 was added to lysize the cells. After gently mixing, 350µl Buffer N3 was added for neutralization. The mixture was then spun down at 13,000rpm for 10 minutes. The supernatant was loaded into a QIAprep Spin Column. After one-minute centrifugation, the DNA was bound to the column. The column was then washed twice with 0.8 ml of buffer PE to remove contaminants and finally the plasmids were eluted into 50µl sterilized ddH_2O. 10 ul of the pLOU3-NP plasmids were digested with EcoRI and Hind III to cut the NP fragments from the plasmids and then the size of both the NP fragments were checked on a 1% agarose gel as describing above. The agarose gel electrophoresis showed that there is a band for NP gene corresponding to the right size around 1.7 Kb. The NP gene fragment is shown in figure3 as the lower band in the electrophoresis gel (Fig. 2.1). This result indicates that the NP gene was successfully inserted in the expression plasmid.

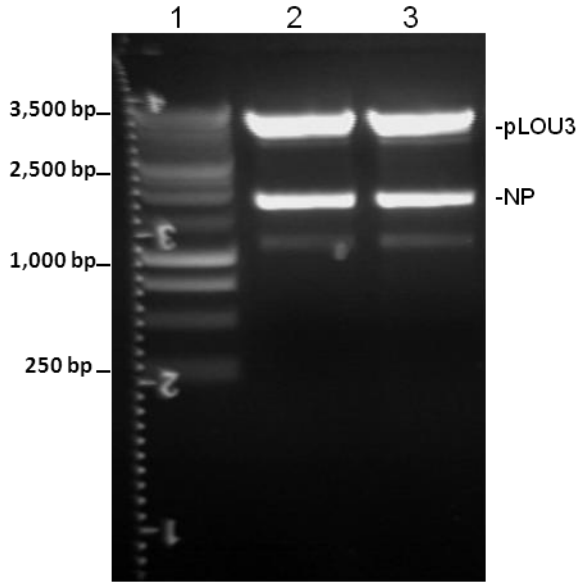


Figure 2.1 The electrophoresis gel of digested pLOU3-NP plasmid. Lane 1 is the 250 bp DNA ladder; lane 2 and lane 3 are the digested plasmids from two clones. The upper bands of the digested plasmids are the vector pLOU3, whereas the lower bands are the target NP gene fragments.

2.2 Protein expression

Once the correct plasmids were confirmed, the plasmids were transformed into expression strain Rosetta cells for protein expression trials as described above. Although other expression constructs like pLOU2 were also tried in *E.coli* expression system, only NP-pLOU3 expressed certain amount of the target NP protein. The single colony from Rosetta transformation plate was then picked and inoculated into 500 ml LB medium in present of 50 μg/ml Ampcilin and 34 μg/ml chloramphenicol (Duchefa Biochemic), and cultured overnight at 37 °C and 200 rpm. The overnight culture was inoculated into 10 liters of the LB medium containing 50 μg/ml Ampcillin and 34μg/ml chloramphenicol. The cell cultures were incubated at 37°C and 200 rpm in an incubator. The NP protein was induced by isopropyl-beta-d-thiogalactopyranoside (IPTG) to a final concentration of 0.03mM

when the OD_{600} of the cell culture reached to 0.6. The NP protein was induced overnight at 20°C.

2.3 Protein purification

2.3.1 Methods of purification

The cells were harvested by centrifugation at 8,000*g* for 10 min at 4 °C. The cell pellet collected from 10L culture was resuspended in 100 ml of amylose binding buffer (10% glycerol, 20mM Tris pH7.5, 0.2M NaCl and 1mM EDTA) with 2 protease inhibitor tablets (Roche), 1 uM DNase I (Sigma) and 1 mM Phenylmethylsulfonyl fluoride (PMSF) (Sigma). The cells were disrupted by a cell disruptor (Constant system Ltd) and the broken cell mixtures were centrifuged at 20,000 rpm for 30 min to remove the cell debris. The supernatant was applied into an amylose column and the column was washed using 10 column volumes of amylose binding buffer. The fusion protein was eluted with amylose elution buffer (10% glycerol, 20mM Tris pH7.5, 0.2 M NaCl, 1mM EDTA and 10 mM Maltose). The protein buffer was changed to Tev cleavage buffer (1/50 PBS, 50 mM Tris-HCl, 0.5 mM EDTA, 1 mM DTT, 0.3 M NaCl and 10% glycerol) using a desalting column (Hiprep 26/10, GE healthcare). Fractions containing HisMBP-NP was pooled and the fusion protein was cleaved with 1 ml His-taggedTEV protease at room temperature overnight. The cleaved sample was applied to the amylose resin again to remove the MBP. The flow-through sample was concentrated to 5 ml and the NP protein was purified by a gel filtration column (Hiload^{TM} 16/60 Superdex 200, GE healthcare) that

was equilibrated by the gel filtration buffer (20mM Tris pH 7.5, 0.3M NaCl and 10% glycerol).

The fractions containing purified NP protein was pooled and concentrated to 7 mg/ml and frozen at -80 °C. The protein's identity and integrity were checked by Electrospray ionization mass spectrometry (ESI-MS) and SDS-PAGE protein gel. The SDS-PAGEs were carried out using pre-cast NuPAGE 4-12% Bis-Tris gels and Powerease 500 according to the manufacturer's manual, with the protein molecular standards of Mark12 unstained standard (Invitrogen). The concentration of purified protein was determined using Bradford solution (Sigma Aldrich) and BSA as standard protein.

2.3.2 Evaluation of NP protein purification

The NP protein was successfully expressed as a MBP fusion protein, and the fusion protein was purified and finally the MBP fusion protein was cleaved off by Tev proteinase. As shown in figure 2.2, although there is still certain amount of NP protein in the flow-through, the eluted NP protein is enough for the next purification step, which shows that the expression of the NP protein was very successful. After the cleavage by Tev proteinase and going through the amylose column, almost all the MPB and uncleaved fusion protein were removed. After gel filtration, NP protein was purified (Fig. 2.3 & 2.4). The protein was finally concentrated to 7 mg/ml and the volume is around 1.5ml, which means a total amount of 10.5 mg NP protein was

obtained from the 10-liter culture.

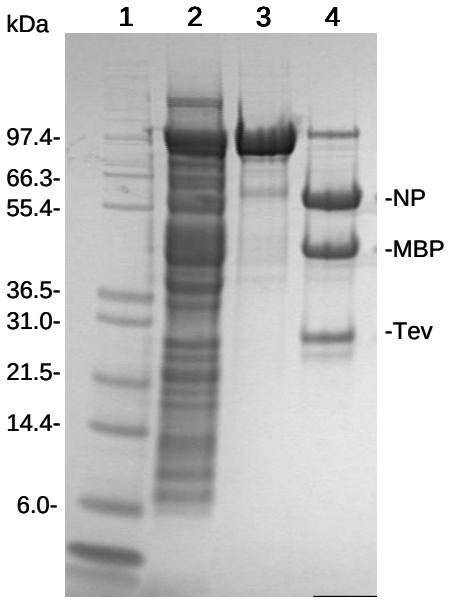


Figure 2.2 The LASV NP was purified by amylose beads. The expression level of NP and the first purification step after tev proteinase cleavage are shown by SDS-PAGE gel. Lane 1: Protein Marker 12; lane 2 flow-through proteins from amylose resin; lane 3 NP elution from amylose resin; lane 4 proteins after tev proteinase cleavage.

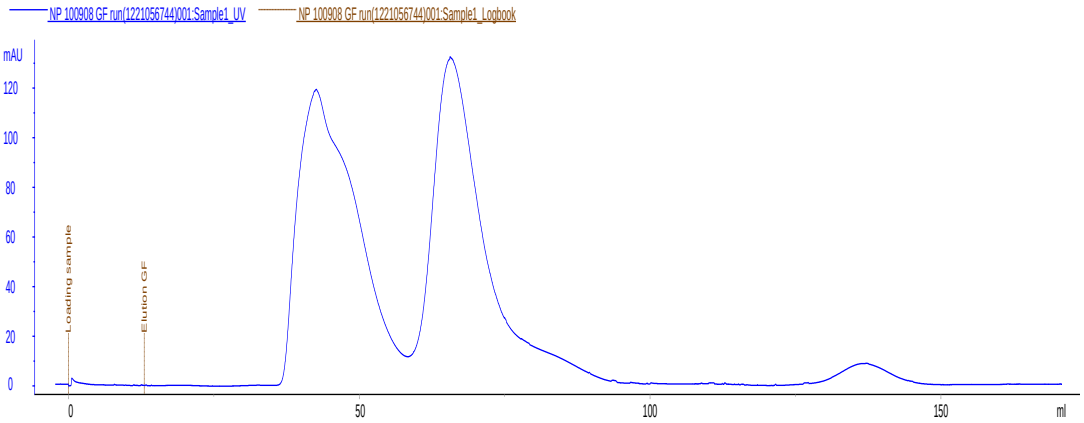


Figure 2.3 The LASV NP protein purification by gel filtration column. The gel filtration chromatography was carried out using a superdex 200 column, the first peak is NP protein and the second peak is MBP.

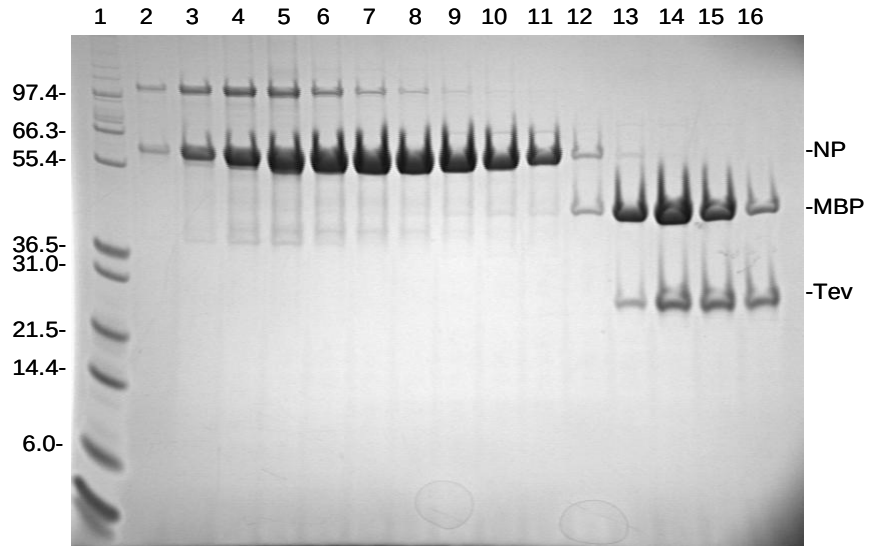


Figure 2.4 SDS-PAGE protein gel for samples from gel filtration. Lane 1 Protein Marker 12; lanes 2 to 11 are NP protein from peak 1; lanes 12 to 16 show the MBP and tev proteinase from peak 2; lanes 7 to 11 fractions were pooled and concentrated.

2.3.3 Identification of NP

To confirm the identity of the target NP protein, the band on SDS-PAGE after gel filtration was cut and sent for the electrospray ionization mass spectrometry (ESI-MS). The trypsin digestion with mass spectroscopy confirmed that the purified protein was the NP of Lassa fever virus. To determine the whole protein molecular weight by mass spectrum, the purified protein was submitted for analysis. The whole mass of native NP was determined to be 64,240 Da (as shown in Fig. 2.5), which matches calculation from the amino acid sequence.

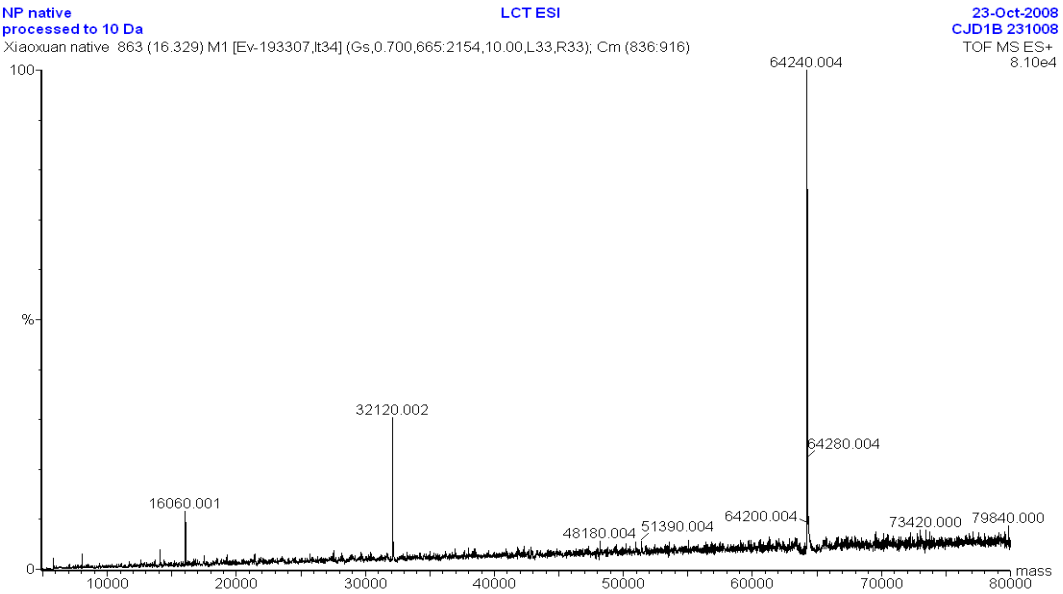


Figure 2.5 The Electrospray ionization mass spectrometry (ESI-MS) result of native NP protein.

2.3.4 Expression, purification and crystallization of Selenomethionine (SeMet) labeled protein

As there is no any similar structure available in the PDB database for structure determination using molecular replacement. We therefore decided to determine the structure by Selenomethionine (SeMet) labeled crystals using multi wavelength anomalous diffraction (MDA). The SeMet labeled protein was set to express in *E.coli* system, unfortunately the level of expression is very poor. The NP protein was cultured in SeMet medium which was prepared as follows (per litre): 1g NH_4Cl, 3g KH_2PO_4, 6g $Na_2HPO_4{\cdot}7H_2O$, 20g glucose 0.3g $MgSO_4$, 1g NaCl, 35ml glycerol, and 0.5g of each of the 19 common amino acid (excluding Met). The final pH was adjusted to around 7.4. The antibiotic Ampicillin was added to 50ug/ml, chloramphenicol was added to final concentration at 34ug/ml. When OD_{600} reaches 0.7, amino acids lysine, phenylanaline, and threonine were added at 100 mg/liter;

amino acids isoleucine, leucine and valine were added at 50 mg/liter; 0.02g thiamine and biotin were added to the culture. 50mg SeMet was added after 10 mimutes, the temperature was adjusted to 20°C when 0.03mM IPTG was added after another 10 minutes. The purification steps are the same as native protein. However, we were unable to crystallize the labeled protein. The phase problem was then solved by heavy atom samarium soaking during structure determination (see section 3.1.3).

2.3.5 Different multimerization of LASV NP

According to the slightly shift curve on the first peak in the gel filtration map (Fig. 2.3), the purified nucleoprotein was assumed to have two forms of oligomers. In order to check whether there are two forms of the NP protein, the purified NP protein was applied into a Superdex 200 10/300 (GE Healthcare Life Sciences) column using a buffer containing 20mM Tris pH 7.5, 0.3M NaCl and 10% glycerol. To calibrate the column, protein standards (Sigma) including five proteins thyroglobulin, y-globulin, ovalbumin, myoglobulin and vitb12 were run with the same buffer. The gel filtration result is shown in figure 2.6. According to the elution volumes of the protein standards, a standard curve was obtained with an equation y= -0.5764x + 19.336 (Fig. 2.7).

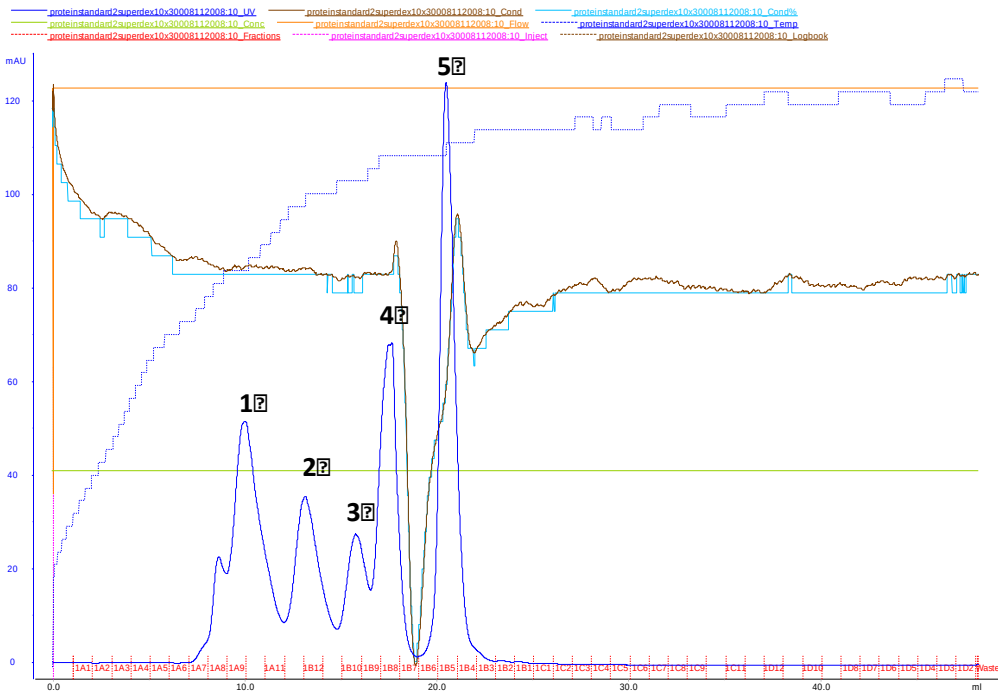


Figure 2.6 The gel filtration of protein standard from Superdex 200 10/300 column. The five protein standards are peak 1: thyroglobulin, peak 2: y-globulin, peak 3: ovalbumin, peak 4: myoglobulin, peak 5: vitb12.

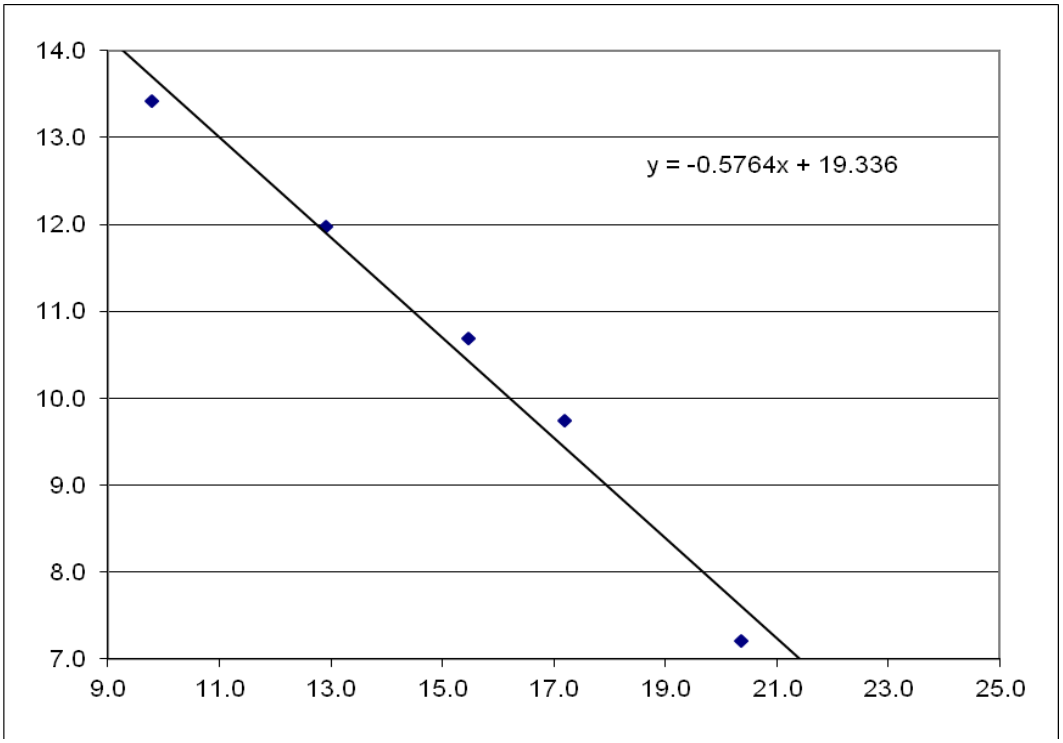


Figure 2.7 The equation for calculating NP molecular weight based on the elution volume (X) and the In(Molecular Weight) (Y) of the five standard protein.

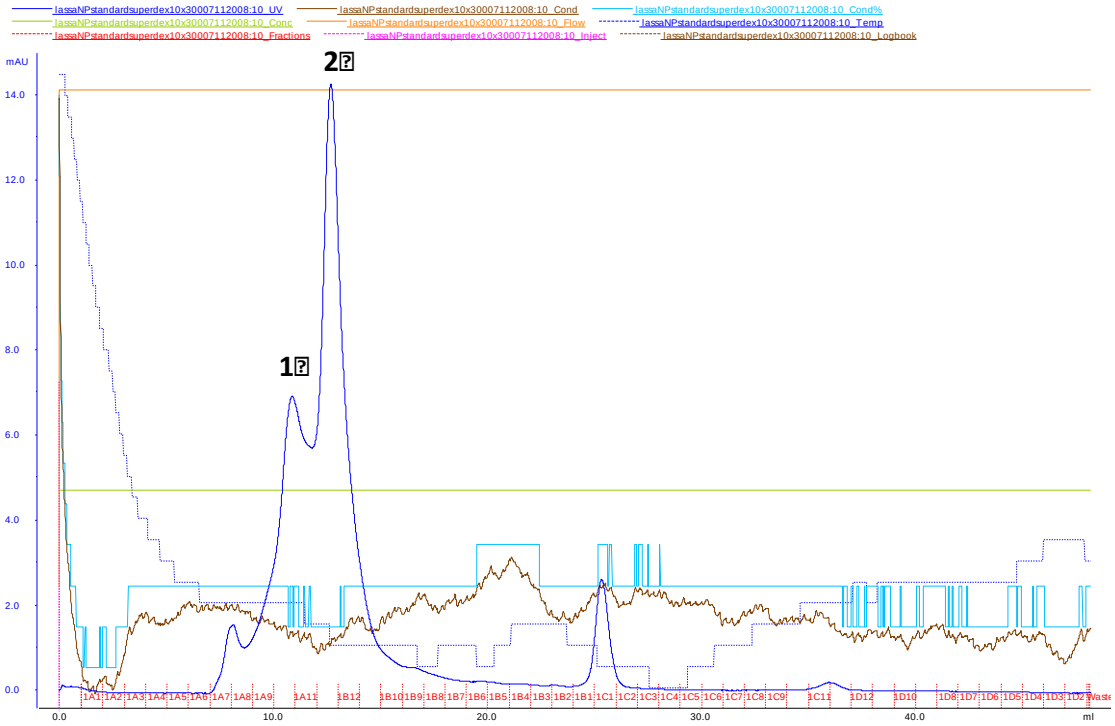


Figure 2.8 Gel filtration of NP protein by a Superdex 200 column. There are two main peaks, indicating NP protein has two forms, which are hexamer (peak 1) and trimer (peak 2).

Table 2.1 The statics table of the molecular weight of oligomer NP (protein x and protein y) that is calculated based on five standard protein thyroglobulin, y-globulin, ovalbumin, myoglobulin and vitb12.

standard	Peak	MW	ln(MW)	Ve	calculated MW	in kDa	error %
thyroglobulin	b	670000	13.4	9.8	1027658	1028	53
y-globulin	c	158000	12.0	12.9	196511	197	24
ovalbumin	d	44000	10.7	15.5	50520	51	15
myoglobulin	e	17000	9.7	17.2	20267	20	19
vitb12	d	1350	7.2	20.4	3780	4	180
unknown							
protein x	1			10.9	466606	467	
protein y	2			12.7	165334	165	

100 ul of the purified NP protein was then loaded into the column, and the gel filtration was carried out according to the manufacture's instruction (Fig. 2.8). There are two main peaks appeared. The elution volume for peak 1 is 10.9ml, which

indicates the molecular weight of the oligomer NP is around 467kDa. The mass of native NP protein is around 64kDa, which means there are almost 6 monomers in this oligomer (Table 2.1). Since the accuracy of this equation shifts with the elution volume, and based on the observation of other NP protein, this oligomer NP is hexamer. The elution volume of the second peak is 12.7 ml, indicating the molecular weight is around 165kD, which means there are 3 monomers in this oligomer, indicating NP oligomer is trimer. A recent published paper (Brunotte, 2011) also confirmed the existence of the symmetric trimer of Lassa virus nucleoprotein (AV strain) by presenting small-angle X-ray scattering (SAXS) and electron microscopy (EM) result.

Chapter3: Structure of Lassa nucleoprotein

In this chapter, we determine the molecular structure of LASV NP protein. Each of the functional domains is described in detail on the basis of corresponding amino acids. Both the exoribonuclease activity site and the potential cap binding site are revealed from this structure. The oligomerization and the interface between each monomer are also present in the following chapter.

3.1 Native NP crystal structure

3.1.1 Crystallization of native Lassa NP

Crystallization conditions for LASV NP were screened on 2 well MRC plates with the sitting drop vapor diffusion method using Cartesian Dispensing systems. The drops were built with 0.15 ul of the protein and 0.15 ul of the mother liquid using two different protein concentrations (7 mg/ml and 3.5 mg/ml), with 100 ul of mother liquid in each reservior. The Classic, JCSG+, Nextal, PEGs, and PEGs II (Hampton, Jena, Qiagen) kits were used in the initial screenings. The initial crystals were formed in 0.2 M LiCl, 20% PEG3350 at 20°C for a week. After optimization, hexagonal prism crystals were grown in 0.2-0.3 M LiCl and 12-16% PEG3350 at 20 °C for 5 days. The optimization buffer conditions are list below (Table 3.1).

Table 3.1 Optimization buffer conditions for crystallizing NP protein.

PEG3350 (v/m, %)	LiCl (M)			
	0.15	0.2	0.25	0.3
12	240ul PEG 75ul LiCl 685ul H_2O	240ul PEG 100ul LiCl 660ul H_2O	240ul PEG 125ul LiCl 635ul H_2O	240ul PEG 150ul LiCl 610ul H_2O
14	280ul PEG 75ul LiCl 645ul H_2O	280ul PEG 100ul LiCl 620ul H_2O	280ul PEG 125ul LiCl 595ul H_2O	280ul PEG 150ul LiCl 570ul H_2O
16	320ul PEG 75ul LiCl 605ul H_2O	320ul PEG 100ul LiCl 580ul H_2O	320ul PEG 125ul LiCl 555ul H_2O	320ul PEG 150ul LiCl 530ul H_2O
18	360ul PEG 75ul LiCl 565ul H_2O	360ul PEG 100ul LiCl 540ul H_2O	360ul PEG 125ul LiCl 515ul H_2O	360ul PEG 150ul LiCl 490ul H_2O
20	400ul PEG 75ul LiCl 525ul H_2O	400ul PEG 100ul LiCl 500ul H_2O	400ul PEG 125ul LiCl 475ul H_2O	400ul PEG 150ul LiCl 450ul H_2O
22	440ul PEG 75ul LiCl 485ul H_2O	440ul PEG 100ul LiCl 485ul H_2O	440ul PEG 125ul LiCl 485ul H_2O	440ul PEG 150ul LiCl 485ul H_2O
24	480ul PEG 75ul LiCl 445ul H_2O	480ul PEG 100ul LiCl 420ul H_2O	480ul PEG 125ul LiCl 395ul H_2O	480ul PEG 150ul LiCl 370ul H_2O

The NP protein crystals grew with dimension of 0.2 × 0.2 × 0.6 mm, which are suitable for X-ray diffraction. To confirm the crystal is from the whole NP protein, a crystal was picked up and washed three times in a cryoprotectant, and dissolved in

SDS-loading buffer for SDS-PAGE. The result showed there was a single band corresponding to the whole NP protein's molecular weight (Fig. 3.1)

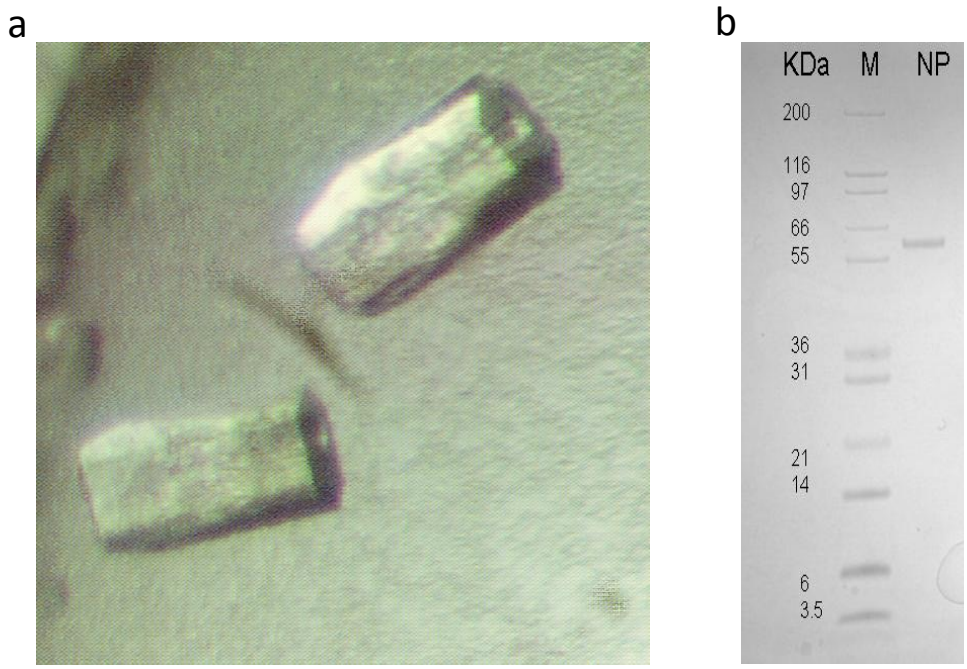


Figure 3.1 The crystals of native Lassa NP protein. a, Photograph of the NP crystals from Lassa fever virus; b, the SDS-PAGE gel of dissolved native NP crystal, indicating the crystals were from the whole NP protein.

3.1.2 In house X-ray data collection

All the crystals were cryo-protected with a cryoprotectant containing 15% to 20% glycerol, 0.2 M $LiCl_2$, 16% PEG3350. The crystal which is primitive hexagon were first tested using in house Rigaku MicroMax-007 HFM fine focus X-ray generator and CCD detector Saturn 944+ with Actor robot. One of the crystals was selected and the data were collected at oscillation angel 0.1° and 100 second exposure per frame with 60 mm of distance from the crystal to detector, using the Cu radiation with the wavelength of 1.5418Å. Total 1000 frames were recordered. The in house processed data indicates that the crystal belongs to space group P321. The unit cell parameters

are a = b = 176.35 Å, c = 56.40 Å, α = β = 90° and γ = 120°. The data processed to resolution 2.36 Å and the detail statistics are list in table 3.2.

Table 3.2 In house data collection statistics of the native NP protein crystal.

Data collection	Values
Wavelength (Å)	1.5418
Resolution (Å)	50-2.36 (2.40-2.36)
Space group	P321
Cell dimension (Å, °)	a = b = 176.35 c = 56.40 α = β = 90 γ = 120
No. unique reflections	41405 (1998)
Completeness	99.7 (97.1)
Redundancy	5.4 (3.9)
I/σ(I)	27.6 (7.2)
R_{merge}*	0.057 (0.166)

* $R_{merge}=\Sigma_{hkl}\Sigma_i|I_i(hkl)-(I(hkl))|/\ \Sigma_{hkl}\Sigma_i I_i(hkl)$, *h*, *k*, *l*-indices of independent reflections with the average intensity <*I*>, I_i-intensities of independent reflections.

The final data were collected at IO2 and IO3 of Diamond light sources UK with the crystals protected by cryoprotectant that contain 15% to 20% glycerol in the crystallization conditions. The high resolution data sets of native NP were collected to a resolution of 1.8 Angstrom (Table 3.3).

Table 3.3 Data collection statistics of native crystal at synchrotron:

Data collection	Values
Wavelength (Å)	0.979
Resolution (Å)	47.63-1.80 (1.81-1.795)
Space group	P321
Cell dimension (Å, °)	a = b = 177.16. c = 56.49
	α = β = 90 γ = 120
No. unique reflections	184629 (5543)
Completeness	99.7 (99.9)
Redundancy	5.6 (5.5)
I/σ(I)	16.4 (4.2)
R_{merge}	0.073 (0.393)

3.1.3 Phase determination

As no similar structures exist in PDB database for structure determination using molecular replacement, and we could not obtain selenomethionine labelled NP crystals, we have tried to soak different heavy atoms to the native crystals in order to find the phase for the LASV NP structure determination using in house x-ray machine. Only samarium soaking allowed us to collect good data to show good anomalous signals. We therefore decided to collect MAD datasets for the structure determination at Diamond I03 beamstation. The MAD data sets were collected at a wavelength of 1.8Å for peak data, 1.8Å for inflection data and 1.45Å for remote data from a single

crystal (Table 3.4).

Table 3.4 The MAD data collection for NP crystal at synchrotron:

Data collection	Multiwavelength experiment		
	peak	inflection	remote
Wavelength Å	1.83	1.84	1.45
Space group	*P*3	*P*3	*P*3
Resolution(highest shell) Å	57.86-2.5 (2.53-2.50)	57.94-2.5 (2.53-2.50)	58.00-2.5 (2.53-2.50)
Cell constants (Å) α=β=90°, γ=120°	a=b=176.75, c=56.35	a=b=177.01, c=56.45	a=b=177.20, c=56.54
Unique reflections	68131(2042)	68424(2029)	68691(2057)
Average redundancy	5.5(5.4)	5.5(5.4)	5.6(5.5)
I/σ	15.3 (4)	13.6 (2.6)	15.5 (3.3)
Completeness (%)	100 (100)	100 (100)	100 (100)
Anomonous Completeness (%)	97.1 (92.6)	97.1 (93.2)	99.1 (97.4)
(High resolution shell) R_{merge}	0.062 (0.342)	0.066 (0.541)	0.06 (0.433)

The original data were indexed and scaled by imosflm and ccp4i, respectively (Fig. 3.2 & 3.3)

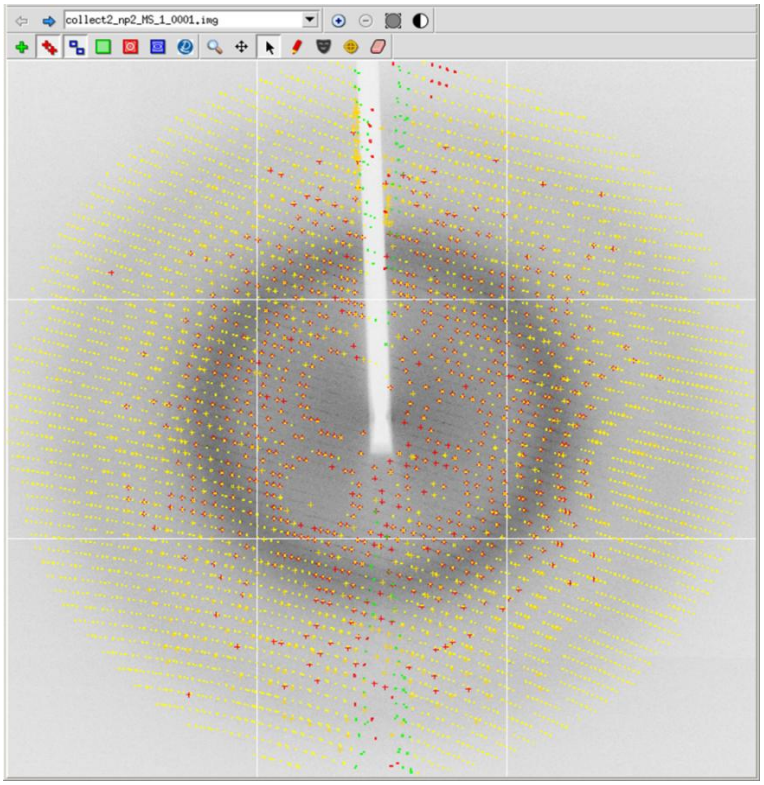


Figure 3.2 X-ray diffraction pattern of optimized crystal of native NP protein. The yellow spots are partially recorded reflection and the red spots are spatially overlapped reflection that will not be integrated.

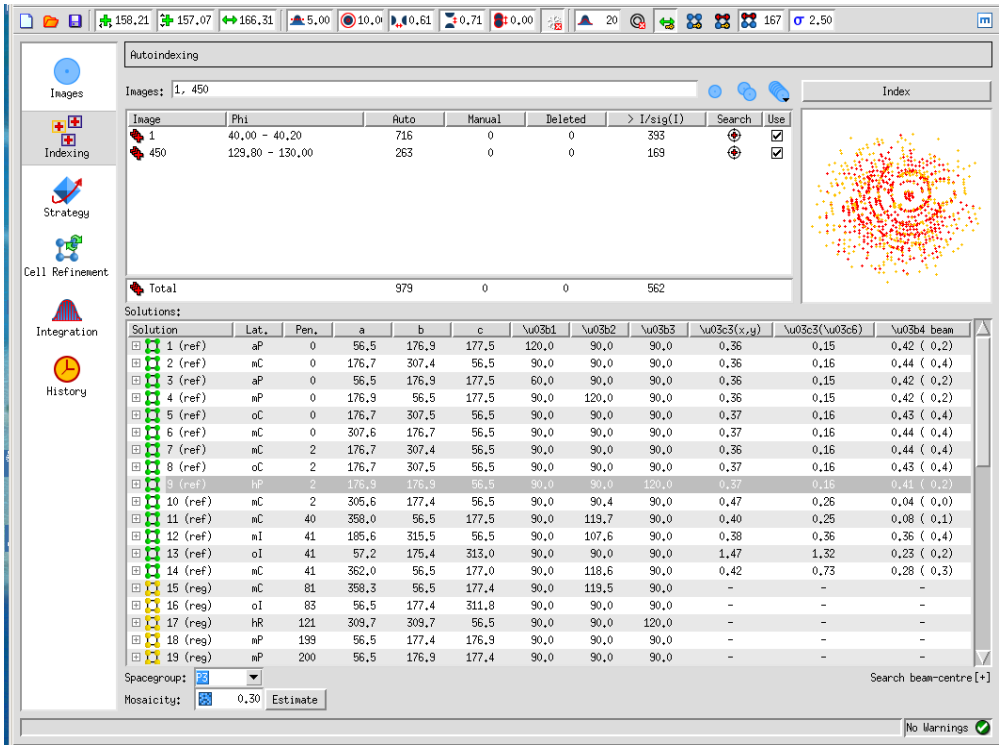


Figure 3.3 Indexing of native NP protein crystal data by imosflm. The p3 space group was selected after indexing.

After indexing, space group *P*3 was suggested during processing data from native NP protein by imosflm. The final data sets were collect from Diamond light sources UK, the data sets processed for multiwavelength anomalous dispersion (MAD) method with samarium derivative native NP are listed above (Table 3.4). The crystals were soaked in 100 mM samarium acetate in 0.2 M LiCl and 16% PEG3350 overnight and back soaked into 15% glycerol, 0.2M LiCl and 16% PEG3350 for 5 min. The crystals were frozen in liquid nitrogen and the data were collected in IO2 station of Diamond. The crystals were scanned by fluorescence scanner to determine the peak, inflection and remote wavelengths to collect complete MAD datasets. More than twenty crystals were screened and one of the best crystals was chosen to collect individual peak wavelength, inflection wavelength and remote wavelength datasets. The three datasets were indexed and integrated by imosflm and scaled by ccp4i SCALA separately.

3.1.4 Model building and refinement of Lassa NP crystal structure

Although the original structure was determined by SOLVE and RESOLVE with the space group of *P*321, the real space group *P*3 with three subunits in an asymmetric unit that was found during refinement will be used in this thesis. By using the multiple wavelength anomalous diffraction (MAD) data sets, the initial phases were obtained with samarium derivative data were processed by Phenix Autosolve, the initial electron density map was then assigned to build the model by Buccaneer and Coot.

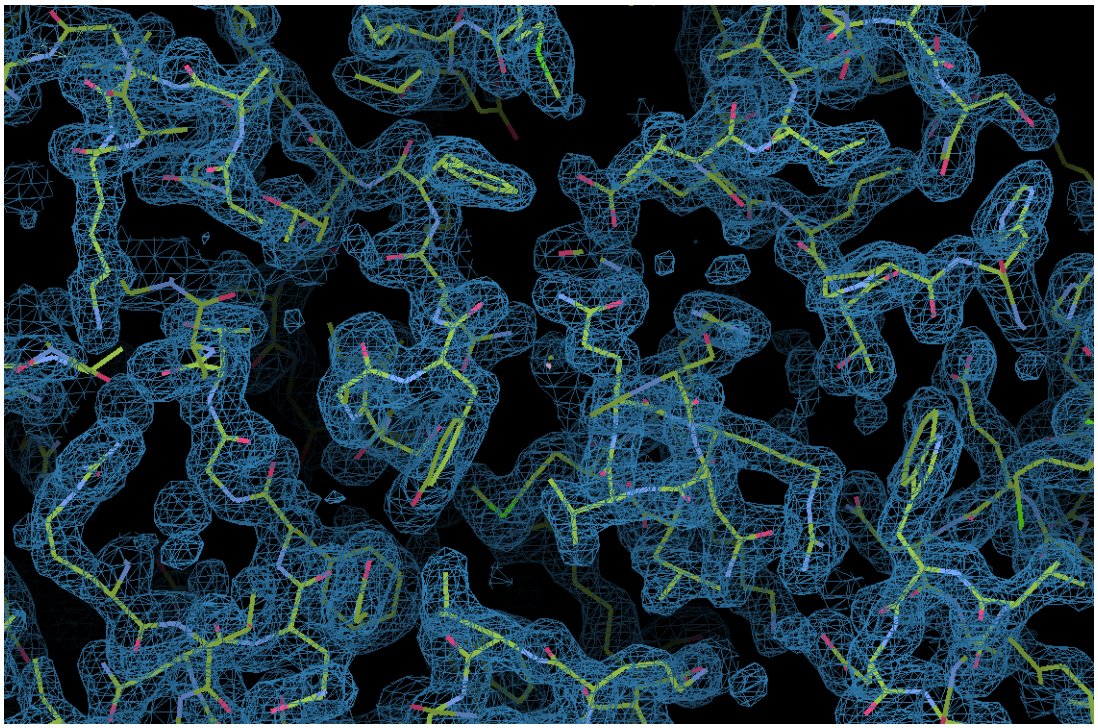


Figure 3.4 The 2*Fo-Fc* electron density map that contoured at 1.8σ is colored in blue. The crystal structural model that references to the amino acid sequence of LASV NP is built manually.

The high-resolution native NP structure was then determined using molecular replacement and the initial NP model built by buccaneer. The structure was built by Coot (Fig. 3.4), and the structure was refined and de-twinned by using REFMAC5. The water molecules were added into the structure by ARP/wARP. To improve the initial phases, the trial strucutures were then interactively processed between real space and reciprocal space until the R_{factor} and R_{free} reached 0.1781 and 0.2014, respectively.

During structure refinement, the crystal was found heavily twined with a twining fraction of 0.43, as we could not drop the Rfree and Rfactor from 0.37 and 0.32 for the 1.8 Å data, the structure then refined and de-twinned by using REFMAC5. The

final model was obtained with R_{factor} and R_{free} reach to 0.178 and 0.201, respectively (Table 3.5). At last, the structure models were evaluated by MolProbity.

Table 3.5 Refinement statistics of native NP crystal structure.

Refinement	Native NP
R_{factor}[a]	0.178
R_{free}[b]	0.201
Rmsd bonds (Å)/angels (°)	0.04/2.62
Residues in Ramachandran core (%)	96.83
PDB ID	3MWP

a $R_{factor} = \Sigma \, ||F_{obs} - F_{calc}|| \, / \, \Sigma|F_{obs}|$, F_{obs} is the observed structure factors, wherea F_{calc} is the calculation structure factors.
b R_{free} measures the degree to which the model predicts the diffraction data for the test set.

The final model has an excellent geometry with 97.15% of the total residues in preferred regions, 2.58% of the total residues in allowed regions and only 0.20% of total residues (1 residue) in outliers (Fig. 3.5). The electron densities for residues 1-6, 147-157, 339-363, 518-521, 562-569 were not well defined, indicating the residues in these regions are disorded. The final protomer model contains 514 residues from the total 569 residues. The model of LASV native NP protein was deposited into the PDB with the accession code 3MWP.

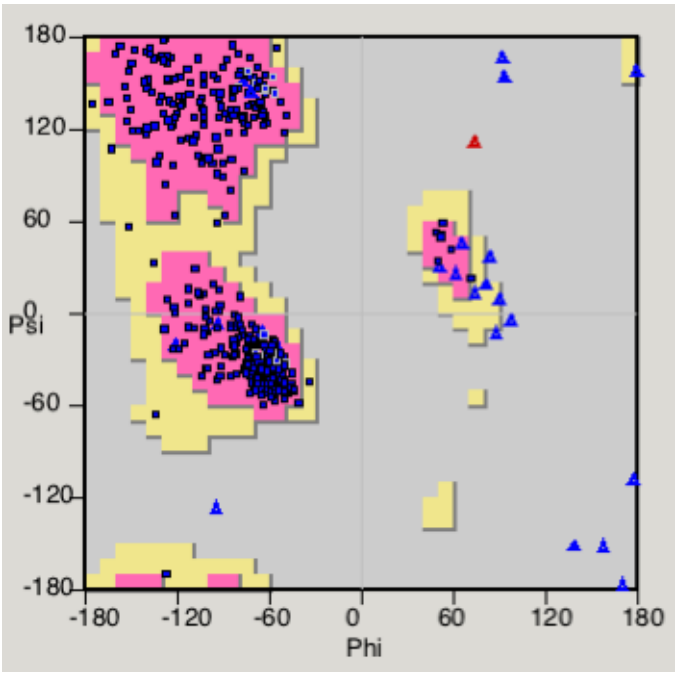


Figure 3.5 Ramachandran plot of the Lassa NP monomer structure. Produced with Ramachandran Plot program by Coot.

3.2 Overall structure of LASV NP

Like other viral NPs (Aurélie, 2006; MacLellan, 2007), the LASV NP structure contains two distinct domains which are the N-terminal domain and the C-terminal domains. However, the two domain structures are totally different from any known nucleoprotein structures of other negative-stranded viruses such as Influenza virus, Vesicular Stomatitis virus, Rabies virus, Respirary syncytical virus and Borna virus (Table 3.6). The NP protomer consists of 24 α-helix and 12 β-strands, with C-terminal domain forming the head, N-terminal domain forming the body and residues 339-363 disordered in between (Fig. 3.6). In the structure, we also have not found the RNA. To date, this is the first crystal structure of nucleoprotein in *Arenaviridae* family.

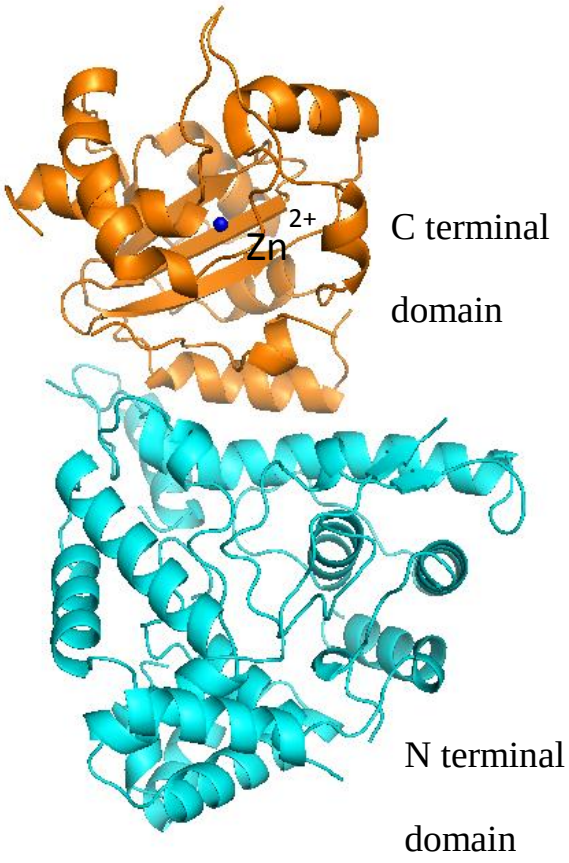


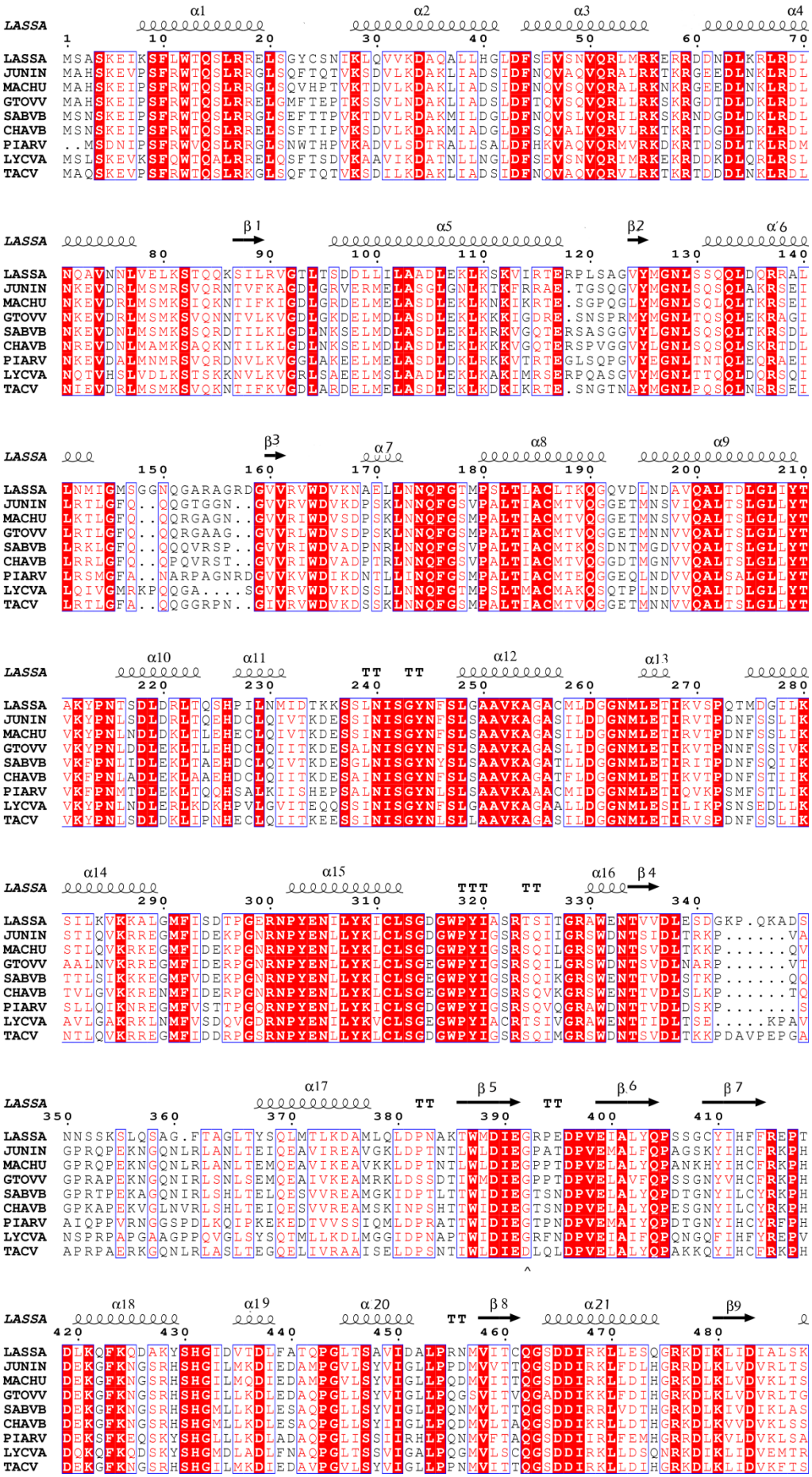
Figure 3.6 The monomer model of native NP crystal structure consists of C-terminal domain (residue 364 to 561) and N-terminal domain (residue 7 to 338) (PDB code: 3MWP).

Table 3.6 Structural comparison of LASV NP and other negative strand RNA viruses.

NP	DALI score*	RMS deviation (Å)	Number of aligned C-alpha atoms	Total residues	PDB code
Influenza virus	1.7	4.1	54	499	2iqh
Vesicular stomatitis virus	1.2	3.6	50	421	2gic
Rabies virus	0.8	2.9	42	450	2gtt
Borna virus	1.8	3.1	41	375	ln93
Respirary syncytical virus	0.8	5.6	63	391	2wj8

*The scores below 2 are not significant. The analysis was performed using the DaliLite server http://www.ebi.ac.uk/Tools/dalilite/index.html.

Amino acid sequence alignment of NPs from Arenaviridae family members, Lassa fever virus (LASV) (sequence access number P13699), Sabia virus (Q90038), Chapare virus (B2C4J1), Junin virus (P14239), Machupo (P26578), Guanarito (Q8AZ66), Pichinde virus (P03541), Lymphocytic choriomeningitis virus (P09992) and Tacaribe virus (P18140) are shown in figure 30, which shows that the NPs share very high similarity in amino acid sequence (almost 32% identity) (Fig. 3.7). That means that the LASV NP model can be served a good model for other nucleoprotein of arenaviruses, and the discoveries from the LASV NP can apply to other arenaviral nucleoprotein.


LASSA
LASSA MSASKEIKSFLWTQSLRREL SGYCSNIKLQVVKDAQALLHGLDFSFVSNVQRLMRKERRDDNDLKRLRDL
JUNIN MAHSKEVPSFRWTQSLRRGL SQFTQTVKSDVLKDAKLLIADSIDFNQVAQVQRAIRKTKRGEEDLNKLRDL
MACHU MAHSKEIPSFRWTQSLRRGL SQVHPTVKTDVLKDAKLLIADSIDFNQVSQVQRALRKNKRGEEDLNKLRDL
GTOVV MAHSKEIPSFRWTQSLRREL GMFTEPTKSSVLNDAKLLIADSLDFTQVSQVQRLIRKSKRGDTDLDKLRDL
SABVB MSNSKEIPSFRWTQSLRRGL SEFTTPVKTDVLRDAKMILDGLDFNQVSIVQRIIRKSKRNDGDLDKLRDL
CHAVB MSNSKEIPSFRWTQSLRREL SSFTIPVKSDVLKDAKMIADGLDFSQVALVQRVLRKTKRTDGDLDKLRDL
PIARV ..MSDNIPSFRWVQSLRRGL SNWTHPVKADVLSDTRALLSALDFHKVAQVQRMVRKDKRTDSDLTKLRDM
LYCVA MSLSKEVKSFQWTQALRREL QSFTSDVKAAVIKDATNLLNGLDFSEVSNVQRIMRKEKRDDKDLQRLRSL
TACV MAQSKEVPSFRWTQSLRKGL SQFTQTVKSDILKDAKLLIADSIDFNQVAQVQRVLRKTKRTDDDLNKLRDL
α1
α2
α3
α4
β1
α5
β2
α'6
LASSA NQAVNNLVELKSTQQKSILRVGTLTSDDLLILAADLEKLKSKVIRTERPLSAGVYMGNLSSQQLDQRRAL
JUNIN NKEVDRLMSMRSVQRNTVFKAGDLGRVERMELASGLGNLKTKFRRAE.TGSQGVYMGNLSQSQLAKRSEI
MACHU NKEVDRLMSMKSIQKNTIFKIGDLGRDELMELASDLEKLKNKIKRTE.SGPQGLYMGNLSQLQLTKRSEI
GTOVV NKEVDRLMSMKSVQNNTVLKVGDLGKDELMDLASDLEKLKKKIGDRE.SNSPRMYMGNLTQSQLEKRAGI
SABVB NKEVDNLMSMKSSQRDTILKLGDLNKSELMDLASDLEKLKRKVGQTERSASGGVYLGNLSQSQLTKRSDL
CHAVB NREVDNLMAMKSAQKNTILKLGDLNKSELMDLASDLEKLKKKVGQTERSPVGGVYLGNLSQSQLSKRTDL
PIARV NKEVDALMNMRSVQRDNVLKVGGLAKEELMELASDLDKLRKKVTRTEGLSQPGVYEGNLTNTQLEQRAEI
LYCVA NQTVHSLVDLKSTSKKNVLKVGRLSAEELMSLAADLEKLKAKIMRSERPQASGVYMGNLTTQQLDQRSQI
TACV NIEVDRLMSMKSVQKNTIFKVGDLARDELMELASDLEKLKDKIKRTE.SNGTNAYMGNLPQSQLNRRSEI
β3
α7
α8
α9
LASSA LNMIGMSGGNQGARAGRDGVVRVWDVKNAELLNNQFGTMPSLTLACLTKQGQVDLNDAVQALTDLGLIYT
JUNIN LRTLGFQ..QQGTGGN..GVVRVWDVKDPSKLNNQFGSVPALTIACMTVQGGETMNSVIQALTSLGLLYT
MACHU LKTLGFQ..QQRGAGN..GVVRIWDVSDPSKLNNQFGSMPALTIACMTVQGGETMNSVVQALTSLGLLYT
GTOVV LRTLGFQ..QQRGAAG..GVVRLWDVSDPSKLNNQFGSMPALTIACMTVQGGETMNNVVQALTSLGLLYT
SABVB LRKLGFQ..QQQVRSP..GVVRIWDVADPNRLNNQFGSVPALTIACMTKQSDNTMGDVVQALTSLGLLYT
CHAVB LRRIGFQ..QPQVRST..GVVRIWDVADPTRLNNQFGSVPALTIACMTVQGGDTMGNVVQALTSLGLLYT
PIARV LRSMGFA..NARPAGNRDGVVKVWDIKDNTLLINQFGSMPALTIACMTEQGGEQLNDVVQALSALGLLYT
LYCVA LQIVGMRKPQQGA....SGVVRVWDVKDSSLLNNQFGTMPSLTMACMAKQSQTPLNDVVQALTDLGLLYT
TACV LRTLGFA..QQGGRPN..GIVRVWDVKDSSKLNNQFGSMPALTIACMTVQGGETMNNVVQALTSLGLLYT
α10
α11
TT
TT
α12
α13
LASSA AKYPNTSDLDRLTQSHPILNMIDTKKSSINISGYNFSLGAAVKAGACMLDGGNMLETIKVSPQTMDGILK
JUNIN VKYPNLSDLDRLTQEHDCLQIVTKDESSINISGYNFSLSAAVKAGASILDGGNMLETIRVTPDNFSSLIK
MACHU VKYPNLNDLDKLTLEHECLQIVTKDESSINISGYNFSLSAAVKAGASILDGGNMLETIRVTPDNFSSLIK
GTOVV VKYPNLDDLEKLTLEHDCLQIITKDESAINISGYNFSLSAAVKAGASLIDGGNMLETIKVTPNNFSSIVK
SABVB VKFPNLIDLEKLTAEHDCLQIVTKDESGINISGYNYSLSAAVKAGATLLDGGNMLETIRITPDNFSQIIK
CHAVB VKFPNLADLEKLAAEHDCLQIITKDESAINISGYNFSLSAAVKAGATFLDGGNMLETIKVTPDNFSTIIK
PIARV VKFPNMTDLEKLTQQHSALKIISHEPSAINISGYNLSLSAAVKAAACMIDGGNMLETIQVKPSMFSTLIK
LYCVA VKYPNLNDLERLKDKHPVLGVITEQQSSINISGYNFSLGAAVKAGAALLDGGNMLESILIKPSNSEDLLK
TACV VKYPNLSDLDKLIPNHECLQIITKEESSINISGYNFSLSLAAVKAGASILDGGNMLETIRVSPDNFSSLIK
α14
α15
TTT
TT
α16
β4
LASSA SILKVKKALGMFISDTPGERNPYENILYKICLSGDGWPYIASRTSITGRAWENTVVDLESDGKP.QKADS
JUNIN STIQVKRREGMFIDEKPGNRNPYENLLYKLCLSGDGWPYIGSRSQIIGRSWDNTSIDLTRKP.......VA
MACHU STLQVKRKEGMFIDEKPGNRNPYENLLYKLCLSGDGWPYIGSRSQILGRSWDNTSVDLTKKP.......QV
GTOVV AALNVKRREGMFIDERPGNRNPYENLLYKLCLSGEGWPYIGSRSQILGRSWDNTSVDLNARP.......VT
SABVB TTLSIKKKEGMFVDEKPGNRNPYENLLYKICLSGEGWPYIGSRSQIKGRSWENTTVDLSTKP.......QQ
CHAVB TVLGVKKRENMFIDERPGNRNPYENLLYKLCLSGEGWPYIGSRSQVKGRSWENTTVDLSLKP.......TQ
PIARV SLLQIKNREGMFVSTTPGQRNPYENLLYKICLSGDGWPYIGSRSQVQGRAWDNTTVDLDSKP.......S
LYCVA AVLGAKRKLNMFVSDQVGDRNPYENILYKVCLSGEGWPYIACRTSIVGRAWENTTIDLTSE.....KPAV
TACV NTLQVKRREGMFIDDRPGSRNPYENLLYKLCLSGDGWPYIGSRSQIMGRSWDNTSVDLTKKPDAVPEPGA
α17
TT
β5
TT
β6
β7
LASSA NNSSKSLQSAG.FTAGLTYSQLMTLKDAMLQLDPNAKTWMDIEGRPEDPVEIALYQPSSGCYIHFFREPT
JUNIN GPRQPEKNGQNLRLANLTEIQEAVIREAVGKLDPTNTLWLDIEGPATDPVEMALFQPAGSKYIHCFRKPH
MACHU GPRQPEKNGQNLRLANLTEMQEAVIKEAVKKLDPTNTLWLDIEGPPTDPVEIALYQPANKHYIHCFRKPH
GTOVV GPRAPEKNGQNIRLSNLSEMQEAIVKEAMRKLDSSDTIWMDIEGPPTDPVELAVFQPSSGNYVHCFRKPH
SABVB GPRTPEKAGQNIRLSHLTELQESVVREAMGKIDPTLTTWIDIEGTSNDPVELALYQPDTGNYILCYRKPH
CHAVB GPKAPEKVGLNVRLSHLTEIQESVVREAMSKINPSHTTWIDIEGTSNDPVELALYQPESGNYILCYRKPH
PIARV AIQPPVRNGGSPDLKQIPKEKEDTVVSSIQMLDPRATTWIDIEGTPNDPVEMAIYQPDTGNYIHCYRFPH
LYCVA NSPRPAPGAAGPPQVGLSYSQTMLLKDLMGGIDPNAPTWIDIEGRFNDPVEIAIFQPQNGQFIHFYREPV
TACV APRPAERKGQNLRLASLTEGQELIVRAAISELDPSNTIWLDIEDLQLDPVELALYQPAKKQYIHCFRKPH
^
α18
α19
α 20
TT
β8
α21
β9
LASSA DLKQFKQDAKYSHGIDVTDLFATQPGLTSAVIDALPRNMVITCQGSDDIRKLLESQGRKDIKLIDIALSK
JUNIN DEKGFKNGSRHSHGILMKDIEDAMPGVLSYVIGLLPPDMVTTQGSDDIRKLFDLHGRRDLKLVDVRLTS
MACHU DEKGFKNGSRHSHGILMQDIEDAMPGVLSYVIGLLPQDMVITTQGSDDIRKLLDIHGRKDLKLVDVKLTS
GTOVV DEKGFKNGSRHSHGILLKDLEDAQPGLLSYVIGLLPQGSVITVQGADDIKKLFDIHGRKDLKLVDVRLTG
SABVB DEKGFKNGSRHSHGMLLKDLESAQPGLLSYVIGLLPQNMVLTTQGSDDIRRLVDTHGRKDLKIVDIKLAS
CHAVB DEKGFKNGSRHSHGMLLKDLESAQPGLLSYIIGLLPQDMVLTAQGSDDIKRLLDTHGRKDLKVVDVKLSS
PIARV DEKSFKEQSKYSHGLLLKDLADAQPGLLSSIIRHLPQNMVFTAQGSDDIRLFEMHGRRDLKVLDVKLSA
LYCVA DQKQFKQDSKYSHGMDLADLFNAQPGLTSSVIGALPQGMVLSCQGSDDIRKLLDSQNRKDIKLIDVEMTR
TACV DEKGFKNGSRHSHGILMKDIEDAVPGVLSYVIGLLPPNMVITTQGSDDIRKLLDIHGRKDLKLIDVKFTS

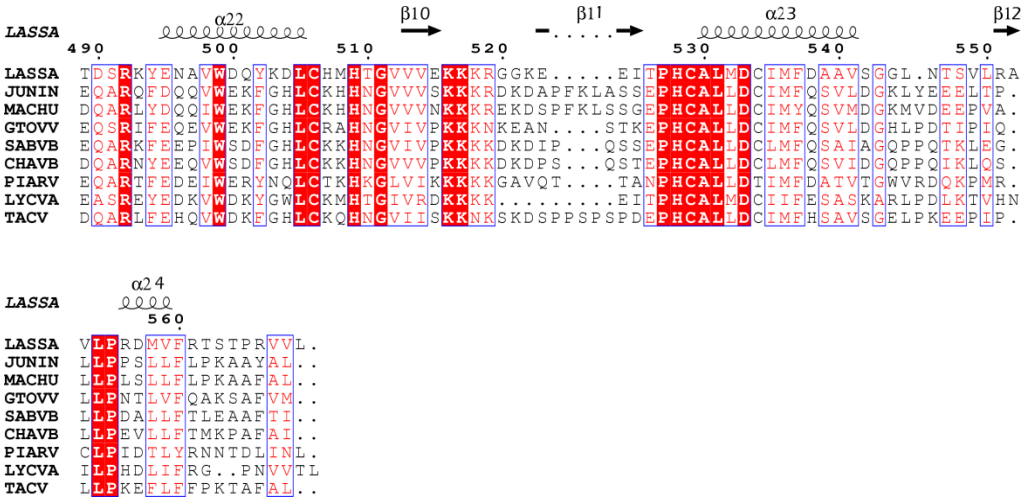


Figure 3.7 Sequence alignment of arenaviral NP proteins. The amino acid sequence of the NPs from the arenavirus family is highly conserved. LASSA, SABVB, CHAVB, JUNIN, MACHU, GTOVV, PIARV, LCMV and TACV represent Lassa fever virus (LASV) (sequence access number P13699), Sabia virus (Q90038), Chapare virus (B2C4J1), Junin virus (P14239), Machupo (P26578), Guanarito (Q8AZ66), Pichinde virus (P03541), Lymphocytic choriomeningitis virus (P09992) and Tacaribe virus (P18140), respectively. The secondary structures shown were based on the LASV NP structure. The conserved residues are shown in red. The residue D389 of TACV NP (corresponding to G392 of LASV NP) is indicated by ^.

3.3 Structural analysis of LASV NP

3.3.1 The structure of N-terminal domain of LASV NP

The N-terminal domain of LASV NP (residues 7 to 338) mainly consists of α-helices and coils. There are 16 α-helices and 4 β-strands within this structural model. The first visible residue of N-terminal domain is I7 which is located on α1-helix (colored in blue) at the front of the model picture shown below (Fig. 3.8a), the polypeptide chain then goes around the model from outside to inside for the next 15α-helices and 4 β-strands. The N-terminal domain forms a body-like structure. The bottom of the structure is made of helices α2, α8, α9, α10, α11 and α12, and the back made of two parallel helices α3, α4, helices α6, α7, and β strands β2 and β3, and the abdomen composed of helices α1, α5, α13, α14, α15, α16, α17 and anti-parallel β1 and β4

strands.

The N-terminal domain terminates at residues L338 that is a short loop right after β4. The N-terminal domain adopts a novel fold that could be the cap binding site for cellular mRNA within its center where the electrostatic surface potential map (Fig. 3.14 a) shows that there is a highly positively charged cavity. But the comparison of LASV NP and other well-known cap bingding proteins as vaccinia methyl transferase, human nuclear cap binding protein, mouse eIF4E and influenza virus cap binding protein PB2 from Dali server indicates that there is no structural similarity (Table 3.7). The following mutation study at binding site and functional study will be presented in the next chapter.

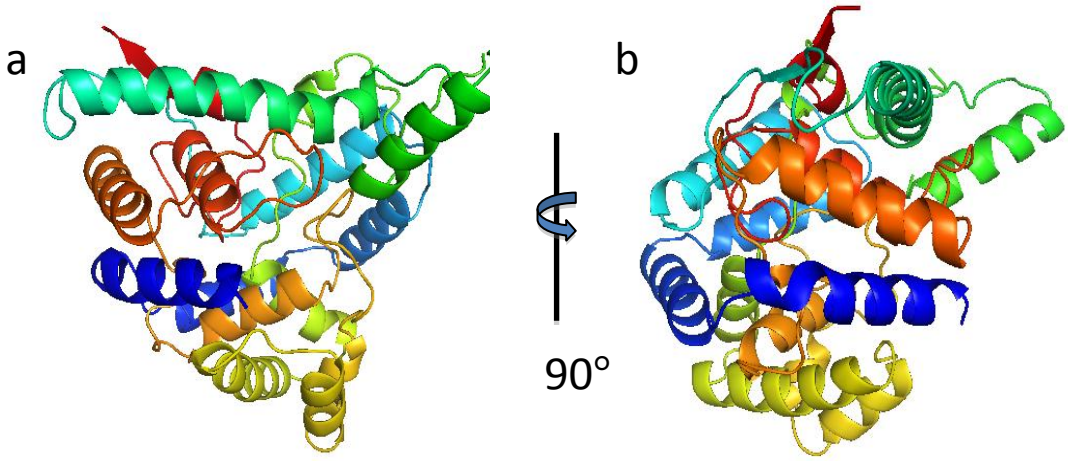


Figure 3.8 The structural model of N-terminal domain of LASV NP, whereas b is after 90° rotation to a along y-axis.

Table 3.7 Structural comparison of the N domain of LASV NP with other well-known cap-binding proteins using the DaliLite server.

	DALI score*	RMS deviation	Number of aligned C-alpha atoms	Total residues	PDB code
Vaccinia methyl transferase	No[#]			294	1av6
Human nuclear cap binding protein	0.7	4.0	57	621	1h2t
Mouse elF4E	0.8	3.9	62	181	1ej1
Influenza virus cap binding protein PB2	0.9	3.3	42	164	2vqz

*Score below 2 is not significant.

The Dali server failed to provide a score due to the dissimilarity of the two structures.

3.3.1.1 LASV NP structure in complex with dTTP

3.3.1.1.1 Crystallization of native NP and dTTP

As the N-terminal domain of the LASV NP has a highly positive charged cavity, we speculated that the cavity might be a cap binding site. To test this hypothesis, we tried to soak m^7GpppG, triphosphorylated, dephosphorylated and monophosphorylated ribonucleotides. The NP complexed with dTTP was obtained by co-crystallization of NP protein with 50mM dTTP in 0.2M $LiCl_2$ and 20% PEG3350 at 20°C.

3.3.1.1.2 Data collection

The data of the NP and dTTP complex structure were collected at IO3 of Diamond light sources UK with the crystals protected by cryoprotectant that contain 15% to 20% glycerol, 0.2 M $LiCl_2$, 16% PEG3350. The space group is also *P3*, the unit cell parameters are a = b = 176.15 Å, c = 56.47 Å, α = β = 90° and γ = 120°. The data processed to resolution 1.98 Å and the detail statistics are list in table 3.8.

Table 3.8 Data collection statistics of NP in complex with dTTP:

Data collection	Values
Wavelength (Å)	0.978
Resolution (Å)	29.75-1.98 (2.09-1.98)
Space group	P3
Cell dimension (Å, °)	a = b = 176.15 c = 56.47
	α = β = 90 γ = 120
No. unique reflections	134699 (19075)
Completeness	99.2 (95.8)
Redundancy	3.8 (3.6)
I/σ(I)	11.9 (2.6)
R_{merge}	0.074 (0.565)

3.3.1.1.3 The dTTP structure determination and refinement

The NP/dTTP compelx structure was determined by molecular replacement based on the native NP crystal structure. After the initial model was built, several cycles of

building and refinement were performed until the R_{factor} and R_{free} reached to 0.154 and 0.182 (Table 3.9). The F_o-F_c maps were calculated before the ligands dTTP was added into the structure.

Table 3.9 The refinement stastics of NP and dTTP complex.

Refinement	dTTP complex
R_{factor}	0.154
R_{free}	0.182
Rmsd bonds (Å)/angels (°)	0.02/1.61
Residues in Ramachandran core (%)	95.21
PDB ID	3MX2

The structure was refined and de-twinned by using REFMAC5, and the water molecules were added into the structure by ARP/wARP. At last, the structure model was evaluated by MolProbity. The coordinate and structure factor of LASV NP with dTTP were deposited into the PDB with the accession code 3MX2 (Fig. 3.9).

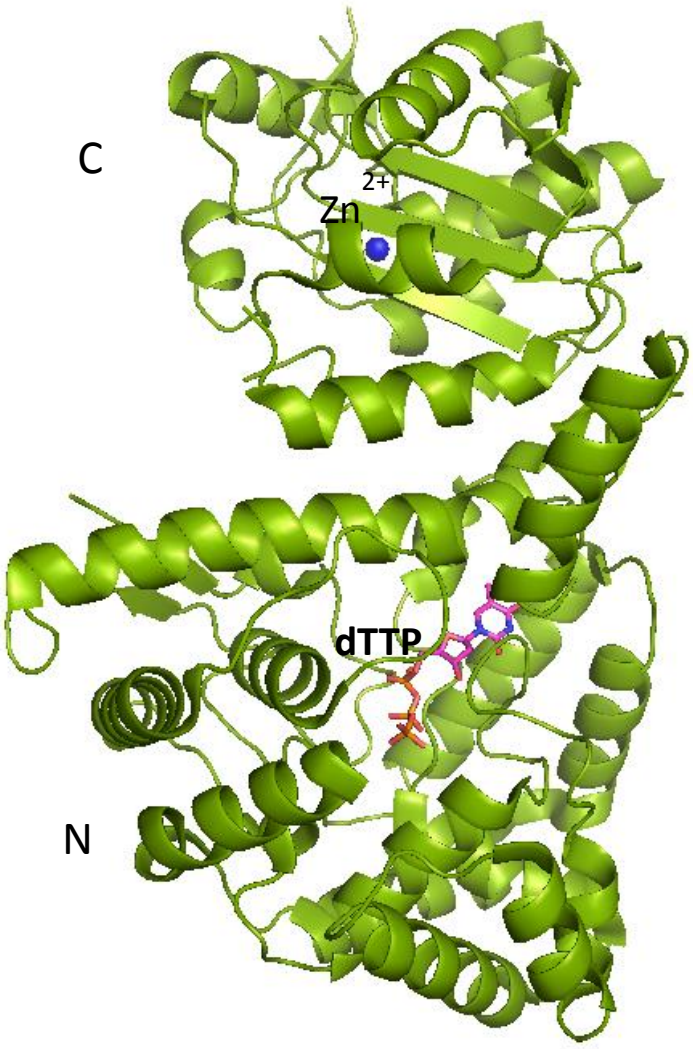


Figure 3.9 The crystal structure model of NP in complex with dTTP. Both the C terminal domain and the N terminal domain of NP are colored in green, whereas the ligand dTTP is colored in magenta and orange based on the element.

3.3.1.2 LASV NP structure in complex with UTP

3.3.1.2.1 Crystallization of native NP and UTP

UTP is an another cap analogue, which is more similar to cap structure with 2-OH presents in nucleotide. To check how UTP binds in the cavity, the NP complexed with UTP was obtained by co-crystallization of NP protein with 50mM UTP in 0.2M $LiCl_2$ and 20% PEG3350 at 20°C.

3.3.1.2.2 Data collection

The data of the NP and UTP complex structure were collected at IO3 of Diamond light sources UK with the crystals protected by cryoprotectant that contain 15% to 20% glycerol, 0.2 M $LiCl_2$, 16% PEG3350. The space group is also *P3*, the unit cell parameters are a = b = 177.11Å, c = 56.72Å, α = β = 90° and γ = 120°. The data processed to resolution 1.91 Å and the detail statistics are list in table 3.10.

Table 3.10 Data collection statistics of native NP in complex with UTP:

Data collection	Values
Wavelength (Å)	0.978
Resolution (Å)	88.56-1.91 (1.96-1.91)
Space group	P3
Cell dimension (Å, °)	a = b = 177.11 c = 56.72
	α = β = 90 γ = 120
No. unique reflections	79046 (5845)
Completeness	99.9 (99.9)
Redundancy	4.5 (4.2)
I/σ(I)	12.6 (3.3)
R_{merge}	0.075 (0.482)

3.3.1.2.3 The UTP complex structure determination and refinement

The NP/UTP complex structure was determined by molecular replacement based on the native NP crystal structure. After the initial model was built, the refinement and

model building cycles were performed until the R_{factor} and R_{free} reached to 0.186 and 0.225 (Table 3.11). The F_o-F_c maps were calculated before the ligands UTP was added into the structure.

Table 3.11 The structure refinement statistics of NP and UTP complex.

Refinement	UTP complex
R_{factor}	0.186
R_{free}	0.225
Rmsd bonds (Å)/angels (°)	0.02/1.78
Residues in Ramachandran core (%)	94.46
PDB ID	3MX5

The structure was refined and de-twinned by using REFMAC5, and the water molecules were added into the structure by ARP/wARP. At last, the structure model was evaluated by MolProbity. The model of LASV NP in complex with UTP was deposited into the PDB with the accession code 3MX5 (Fig. 3.10).

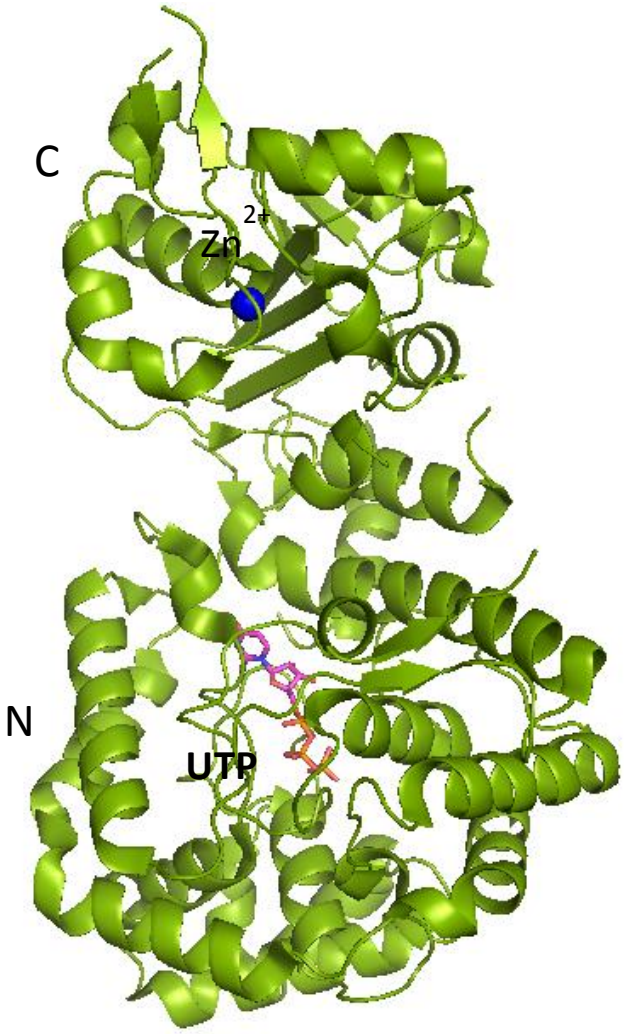


Figure 3.10 The crystal structure model of NP in complex with UTP. Both the C terminal domain and the N terminal domain of NP are colored in green, whereas the ligand UTP is colored in magenta and orange based on the element.

3.3.2 The structure of C-terminal domain of LASV NP

Since almost 25 residues are disordered between the N-terminal domain and C-terminal domain, the first residue that can be seen is alanine 364. The C-terminal domain (residues 364-561) of LASV NP forms the head-like structure that consists of one anti-parallel β sheet (β5, β6 and β7 strands) in the center which surrounded by helix α17, α 19, α 20, α 24 and strands β10, β11 and β12 at one side and helices α18,

α21, α22, and α23 at the other side, and one parallel β sheet (â8 and â9). The centered β sheet is composed of three anti-parallel β strands forming sandwich architecture. The strands β5, β6 and β7 of the first β sheet are connected to each other head to tail by loop only, and each of them is in opposite orientation to the adjacent strand. The β8 and β9 strands form the second β sheet. These two β strands are parallel to each other with a 30^{o} kink separated by α21 (Fig. 3.11).

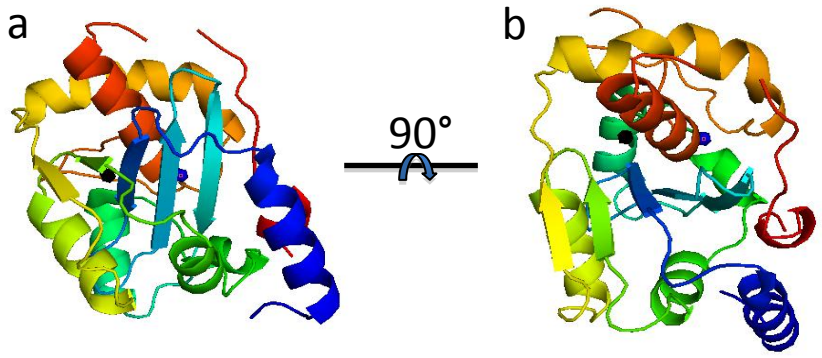


Figure 3.11 The structural model of C-terminal domain of LASV NP, whereas b is 90° rotated to a along x-axis. Manganese ion is colored in black sphere, while zinc ion is colored in blue sphere.

To search the similar structure of C-terminal domain in PDB, LASV NP C-terminal domain structure coordinate was put into Dali server (http://ekhidna.biocenter.helsinki.fi/dali_server), and it was very surprise that the C-terminal domain is very similar to several known 3'-5' exonucleases/exoribonucleases, which are from the DEDDH subfamily of the DEDD (DnaQ) superfamily (Fig. 3.12) as below.

No: Chain Z rmsd lali nres %id PDB Description

1: 1j54-A 10.6 2.4 126 174 13 PDB MOLECULE: **DNA POLYMERASE III, EPSILON CHAIN**;

2: 1j53-A 10.6 2.4 126 174 13 PDB MOLECULE: DNA POLYMERASE III, EPSILON CHAIN;

3: 2ido-A 10.3 2.4 126 172 13 PDB MOLECULE: DNA POLYMERASE III EPSILON SUBUNIT;

4: 2gui-A 10.2 2.4 125 176 14 PDB MOLECULE: DNA POLYMERASE III EPSILON SUBUNIT;

5: 2ido-C 10.2 2.4 127 175 13 PDB MOLECULE: DNA POLYMERASE III EPSILON SUBUNIT;

6: 1y97-B 8.6 2.9 125 211 12 PDB MOLECULE: **THREE PRIME REPAIR EXONUCLEASE 2**;

7: 2ioc-A 8.5 3.0 127 218 10 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

8: 1w0h-A 8.5 2.9 127 200 15 PDB MOLECULE: **3'-5' EXONUCLEASE ERI1**;

9: 2ioc-B 8.4 3.0 127 220 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

10: 2o4g-A 8.4 3.0 127 217 10 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

11: 3b6p-D 8.4 3.0 127 217 10 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

12: 2o4g-D 8.4 3.0 127 217 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

13: 3b6o-D 8.4 3.0 127 217 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

14: 2o4i-B 8.4 3.0 128 218 11 PDB MOLECULE: 5'-

15: 3b6o-A 8.4 3.0 127 217 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

16: 2o4g-C 8.4 3.0 126 217 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

17: 3b6p-A 8.4 3.0 127 217 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

18: 2oa8-A 8.3 3.0 127 220 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

19: 2o4i-A 8.3 3.0 127 218 11 PDB MOLECULE: 5'-

20: 2o4g-B 8.3 3.0 127 217 10 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

21: 3b6p-C 8.3 3.0 127 217 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

22: 3b6o-B 8.3 3.1 127 217 10 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

23: 3b6o-C 8.3 3.0 127 217 10 PDB MOLECULE: THREE PRIME REPAIR

EXONUCLEASE 1;

24: 3b6p-B 8.3 3.1 127 217 11 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

25: 2oa8-B 8.2 3.0 127 226 10 PDB MOLECULE: THREE PRIME REPAIR EXONUCLEASE 1;

26: 1zbu-B 8.0 2.9 125 304 15 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

27: 1zbu-A 7.9 3.0 127 225 15 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

28: 1zbh-C 7.9 3.1 129 225 15 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

29: 1zbh-D 7.8 12.4 157 289 15 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

30: 1zbh-A 7.8 10.6 149 289 14 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

31: 1zbh-B 7.8 3.0 128 225 15 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

32: 1zbu-C 7.8 3.0 127 225 15 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

33: 2f96-B 7.8 3.6 126 201 13 PDB MOLECULE: **RIBONUCLEASE T**;

34: 2p1j-A 7.7 2.4 111 164 13 PDB MOLECULE: **DNA POLYMERASE III POLC-TYPE**;

35: 2igi-A 7.7 3.0 126 180 13 PDB MOLECULE: **OLIGORIBONUCLEASE**;

36: 2igi-B 7.7 3.1 126 179 13 PDB MOLECULE: OLIGORIBONUCLEASE;

37: 1zbu-D 7.6 3.1 127 224 15 PDB MOLECULE: 3'-5' EXONUCLEASE ERI1;

38: 2is3-C 7.6 3.9 128 203 14 PDB MOLECULE: RIBONUCLEASE T;

39: 1fxx-A 7.5 2.6 129 459 13 PDB MOLECULE: **EXONUCLEASE I**;

40: 2p1j-B 7.3 2.9 113 171 12 PDB MOLECULE: DNA POLYMERASE III POLC-TYPE;

41: 1yta-D 7.3 3.3 128 180 13 PDB MOLECULE: OLIGORIBONUCLEASE;

42: 1yta-C 7.2 3.3 128 180 13 PDB MOLECULE: OLIGORIBONUCLEASE;

43: 1yta-B 7.2 3.1 127 180 13 PDB MOLECULE: OLIGORIBONUCLEASE;

44: 3cg7-B 7.0 3.0 129 295 8 PDB MOLECULE: **CELL DEATH-RELATED NUCLEASE 4**;

45: 3cg7-A 7.0 3.0 129 296 8 PDB MOLECULE: CELL DEATH-RELATED NUCLEASE 4;

46: 3cm6-A 7.0 3.1 129 294 8 PDB MOLECULE: CELL DEATH-RELATED NUCLEASE 4;

47: 3cm5-A 7.0 3.1 128 294 8 PDB MOLECULE: CELL DEATH-RELATED NUCLEASE 4;

48: 2is3-D 7.0 3.9 125 203 12 PDB MOLECULE: RIBONUCLEASE T;

49: 3cm5-B 6.9 3.1 130 294 8 PDB MOLECULE: CELL DEATH-RELATED NUCLEASE 4;

50: 3cm6-B 6.9 3.1 130 296 8 PDB MOLECULE: CELL DEATH-RELATED NUCLEASE 4;

Figure 3.12 A list of the proteins whose Z-score are higher than 2 in Dali database. In the first 50 molecules, DNA polymerase III (Epsilon chain), Three prime repair exonuclease 1(human TREX 1), 3'-5' exonuclease ERI1, Ribonuclease T, DNA polymerase III POLC-type, oligoribonuclease and cell death-related nuclease 4 all show very high similarity to LASV NP C-terminal domain.

Although the structural comparison from Dali server between influenza A PA N-terminal endonuclease domain and C terminal domain of LASV NP shows that there is no significant similarity (Table 3.12), the C-terminal domain of LASV NP is structural similar to *E.coli* DNA polymerase III epsilon subunit, RNase T and human TREX 1 (Fig. 3.12). All these proteins are 3'-5' exonucleases. The *E.coli* DNA polymerase III epsilon subunit is responsible for proofreading activity during DNA replication by enzymatically excising the incorrect nucleotide(s) from the newly synthesised DNA, and RNase T plays an important role in bacterial RNA maturation by trimming and modifying the 3' ends of rRNA and tRNA. TREX1 is a dominant 3' to 5' exonuclease in mammalian cells, which degrades both ssDNA and dsDNA that are released into cytoplasm during apoptosis. Natural mutations of the catalytic residues of human TREX1 lead to aberrant accumulation of non-degraded DNAs that are recognized by the cellular pattern-recognition receptor to trigger the IFN production, a persistently high level of which can lead to autoimmunue diseases, such as Aicardi-Goutieres Syndrome and chilblain lupus. We propose that the NP protein functionally mimics human TREX1 by removing the viral PAMP RNAs that otherwise activate the cellular RNA pattern-recognition receptors to trigger the IFN production. The superimposition of LASV NP C-terminal domain and human Trex1 protein (Fig. 3.13) shows that the centered β sheet described above is highly conserved. Particularly, the five catalytic residues from Lassa NP and the Trex1 are at identical positions, which further suggest that the LASV NP is a 3'-5' exonuclease. The human Trex1 protein has two Mn^{2+} cations in the active site whereas the C-terminal domain

structure contains not only one Manganese cation (black). The electrostatics potential map (Fig. 3.14 b) indicates that there is a highly negatively charged cavity, which is a typical DEDDH family enzyme. The C-terminal domain of LASV NP ends at residue R561, and the rest of residues are disordered.

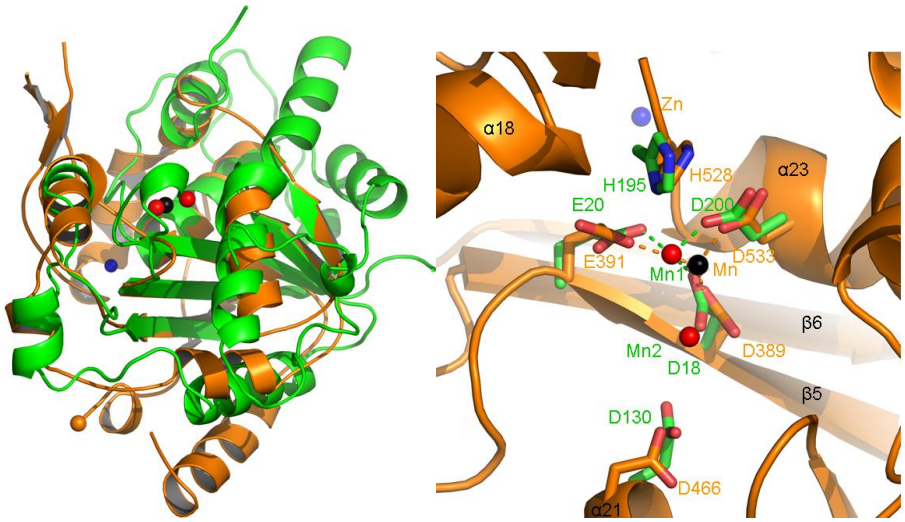


Figure 3.13 The superimposition of LASV NP C-terminal domain and human Trex1 protein. The C-terminal domain of LASV NP is colored in orange with Mn^{2+} colored in black, Zn^{2+} colored in blue; The human Trex1protein is colored in green with its two Mn^{2+} colored in red. The result shows that the two proteins are very similar.

Table 3.12 Structural comparison of the LASV NP and the influenza A virus PA N-terminal endonuclease domain.

	DALI score*	RMS deviation	Number of aligned C-alpha atoms	Total residues	PDB code
Influenza virus PA endonuclease	1.9	10.4	38	197	3ebj

* Scores below 2 are not significant.

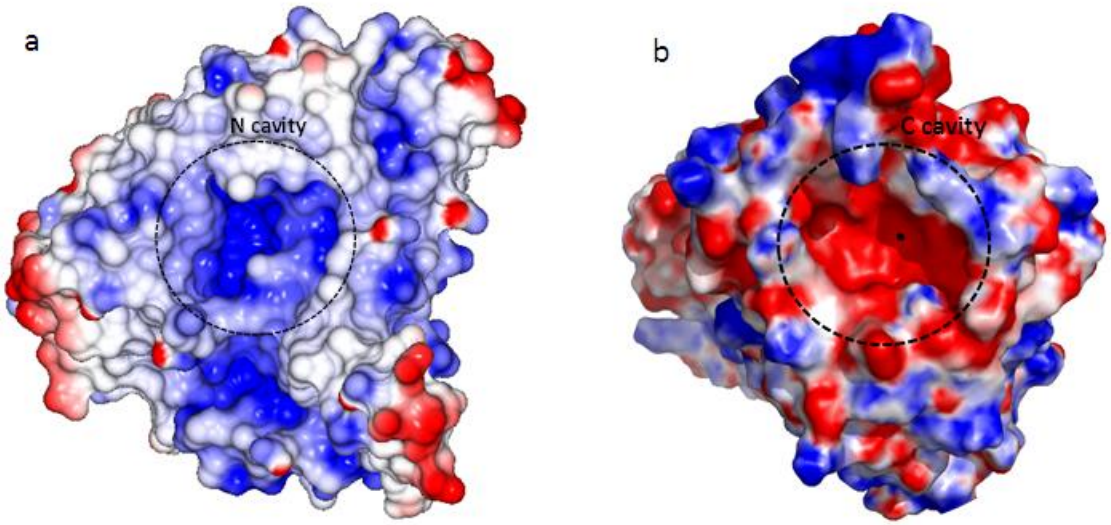


Figure 3.14 Electrostatic surface potential map for the two domain of LASV NP. a, the N-terminal domain with highly positively charged cavity colored in blue; b, the C-terminal domain with highly negatively charged cavity colored in red, and the black dot represents the Mn^{2+}.

3.3.2.1 LASV NP structure in complex with Mn^{2+}

3.3.2.1.1 Crystallization of native NP and Mn^{2+}

As described above that the C-terminal domain structure of the LASV NP is very similar to 3'-5' exonucleases, which are metal-dependent enzymes. Normally, magnesium and manganese are the most important metals for their activities. To check how Mg^{2+} or Mn^{2+} bind on the NP exoribonuclease, we tried to soak Mg^{2+} and Mn^{2+} into native NP crystal. Only NP complexed with manganese ion was obtained by soaking the crystals in 0.2 M $NaCl_2$, 20% PEG3350 and 15% glycerol containing 50mM $MnCl_2$. The presence of the manganese and the zinc ions was confirmed by fluorescence scanning at the Diamond light sources UK.

3.3.2.1.2 Data collection

The data of the NP and Mn^{2+} complex crystals were collected at IO2 and IO3 of Diamond light sources UK with the crystals protected by cryoprotectant that contain 15% to 20% glycerol, 0.2 M $LiCl_2$, 16% PEG3350. The space group is also *P3*, the unit cell parameters are a = b = 176.81 Å, c = 56.43 Å, α = β = 90° and γ = 120°. The data processed to resolution 1.99 Å and the detail statistics are list in table 3.13.

Table 3.13 Data collection statistics of NP and Mn^{2+} complex:

Data collection	Values
Wavelength (Å)	0.978
Resolution (Å)	56.43-1.99 (2.01-1.99)
Space group	P3
Cell dimension (Å, °)	a = b = 176.81. c = 56.43
	α = β = 90 γ = 120
No. unique reflections	134882 (4075)
Completeness	99.2 (99.2)
Redundancy	3.7 (3.3)
I/σ(I)	11.8 (2.9)
R_{merge}	0.069 (0.39)

3.3.2.1.3 The NP/Mn^{2+} complex structure determination and refinement

The NP/Mn^{2+} complex structure was determined by molecular replacement based on the native NP crystal structure. After the initial model was built, the refinement and

model building were carried out until the R_{factor} and R_{free} reach 0.179 and 0.214 (Table 3.14). The F_o-F_c maps were calculated before the ligands zinc and manganese were added into the structure.

Table 3.14 The refinement statistics of NP and Mn^{2+} complex.

Refinement	Mn^{2+} complex
R_{factor}	0.179
R_{free}	0.214
Rmsd bonds (Å)/angels (°)	0.02/1.62
Residues in Ramachandran core (%)	95.84
PDB ID	3MWT

The structure was refined and de-twinned by using REFMAC5, and the water molecules were added into the structure by ARP/wARP. At last, the structure model was evaluated by MolProbity. The model (Fig. 3.15 & 3.16) of LASV NP with Mn^{2+} was deposited into the PDB with the accession code 3MWT.

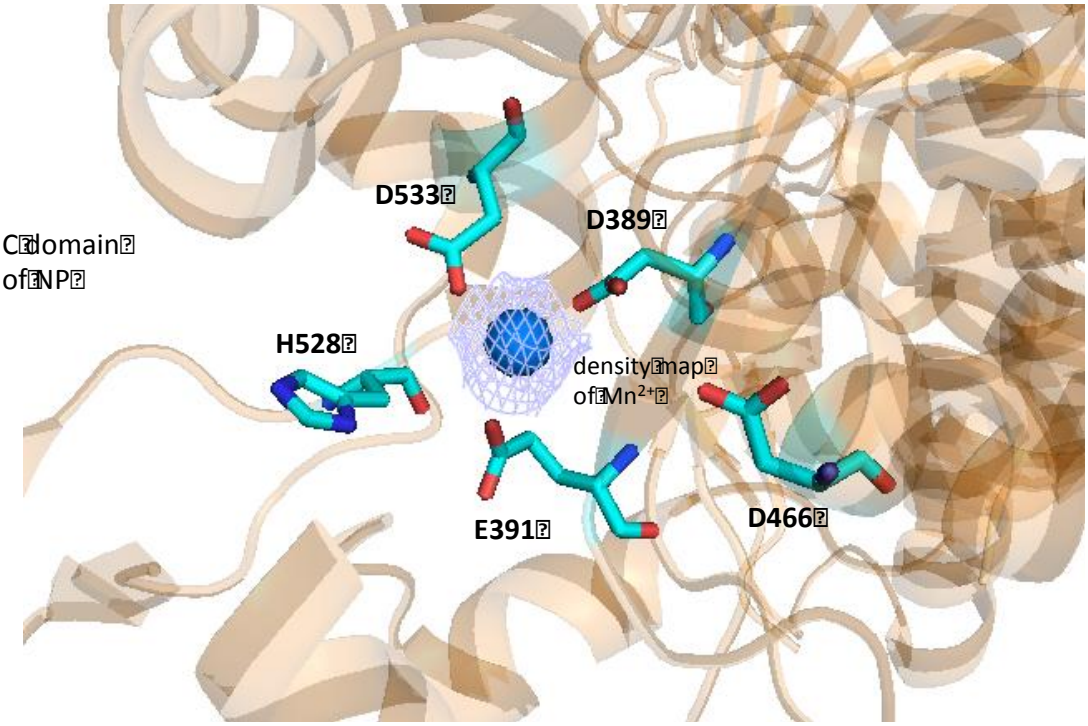


Figure 3.15 Electron density map for manganese ion that resides within the catalytic center of LASV NP protein. The corresponding residues D389, E391, D466, D533 and H528 are colored in cyan, manganese is colored in marine and the F_o-F_c map at 2.5 σ is colored in light blue.

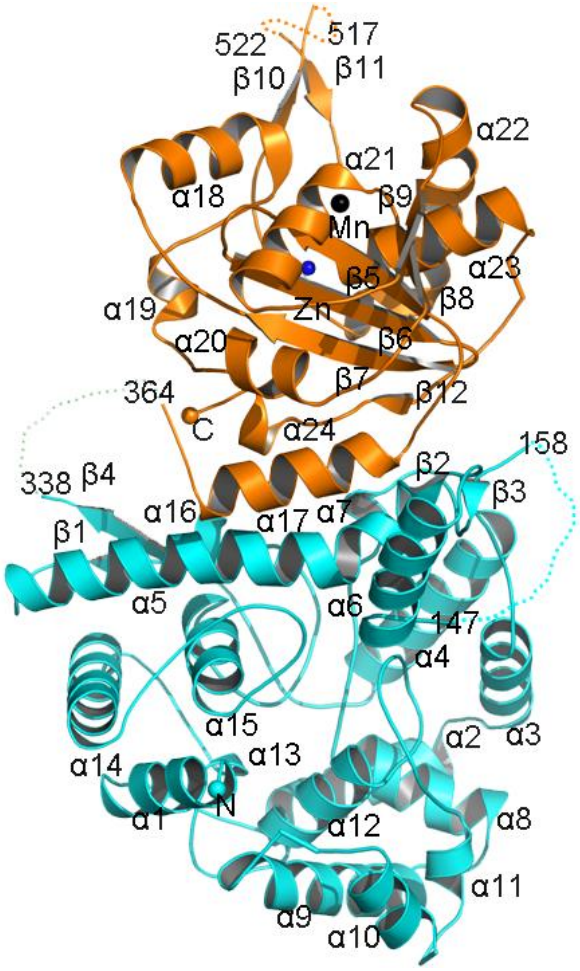


Figure 3.16 The crystal structure model of LASV NP protomer in complex with Mn^{2+}. The N-terminal domain is colored in cyan with the cyan sphere indicating the N terminus; the C-terminal domain is colored in orange with the orange sphere i.ndicating the C terminus. The black sphere shows Mn^{2+}, whereas the blue sphere shows Zn^{2+}. The dotted lines represent the disordered loops

3.6 Trimeric NP and its interfaces

There are two different trimeric architectures in crystal lattice that are asymmetric trimer and crystallographic three fold axis symmetric trimer (shown in Fig. 3.17). We assumed that the native trimeric NP protein forms the symmetric trimer in nature.

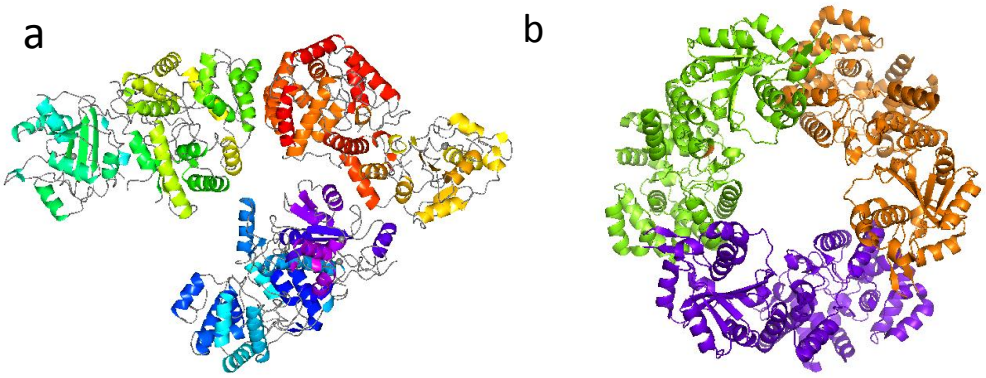


Figure 3.17 The asymmetric and symmetric trimeric structures of LASV NP. a, the asymmetric trimer with each protomer colored in yellow green, yellow orange and purple blue, respectively; b, the symmetric trimer with crystallographic three fold axis symmetry in crystal lattice.

There are two forms of interface between every dimerized monomer subunits of these two forms trimer architectures, which are tail to tail (N-terminus to N-terminus) for asymmetric trimer and head to tail (N-terminus to C-terminus) for symmetric trimer. The subnuits in symmetric trimer connect to adjacent monomer in a head to tail orientation that forms a ring structure (Fig. 3.18).

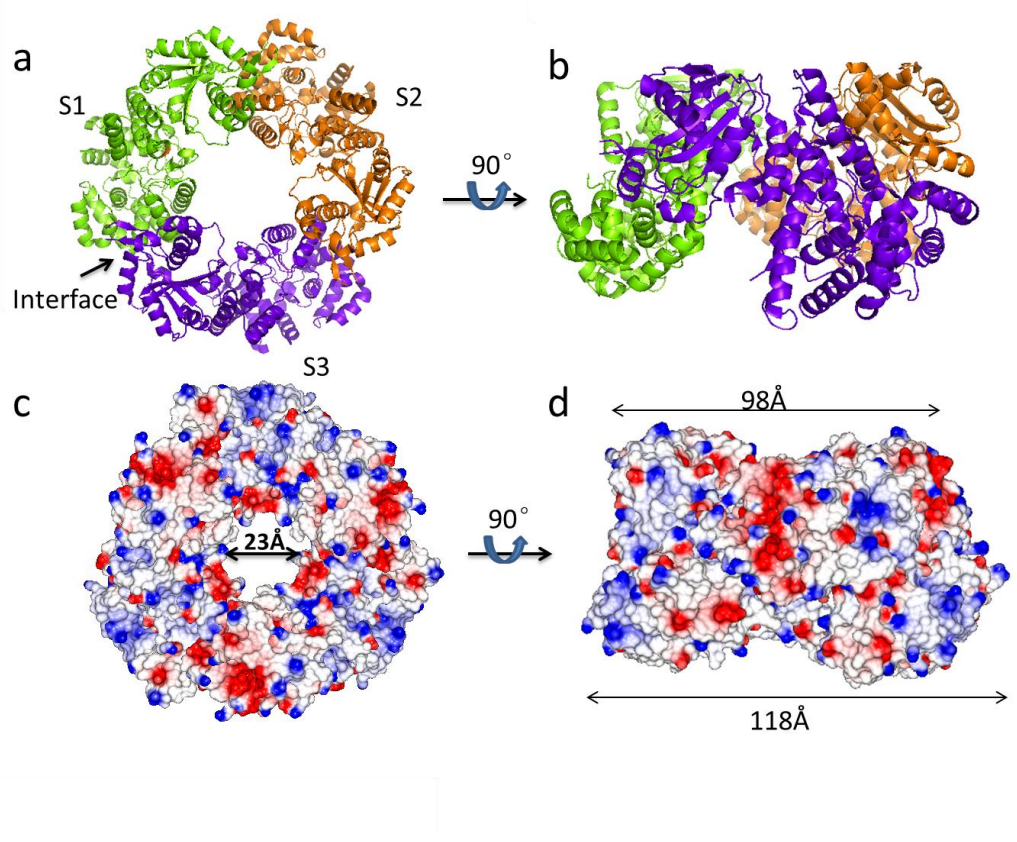


Figure 3.18 The symmetric trimer of LASV NP in solution. a, the ribbon diagram of LASV NP trimer, each of the subunits (labeled as S1,S2 and S3) colored in green, blue and orange, respectively; b, the ribbon diagram of symmetrice trimer after 90° rotation along x-axis; c, the electrostatic surface potential map of the symmetric trimer shows that the diameter of the center hole is around 23Å; d, 90° rotation of the electrostratic surface potential map shows that the diameter of the head of the symmetric trimer is around 98Å, whereas the bottom is around 118Å.

The interface area between the subunits is 455Å^2, representing 1.9% of the total surface area of a subunit which is 23,343Å^2. The central hole of the trimeric structure is 23Å in diameter, whereas the head ring is 98Å and the body ring is 118Å (Fig. 3.18). The pridicted residues that are involved in dimerization on the interfaces between each subunit of the symmetric trimer are list in two groups. They are R52, K56, K189, Q190 and N230 on N-terminal domain, and D434, D437, N496, D500, T510 and K522 on C-terminal domain according to the LIGPLOT result (Brunotte, 2011) (Fig. 3.19).

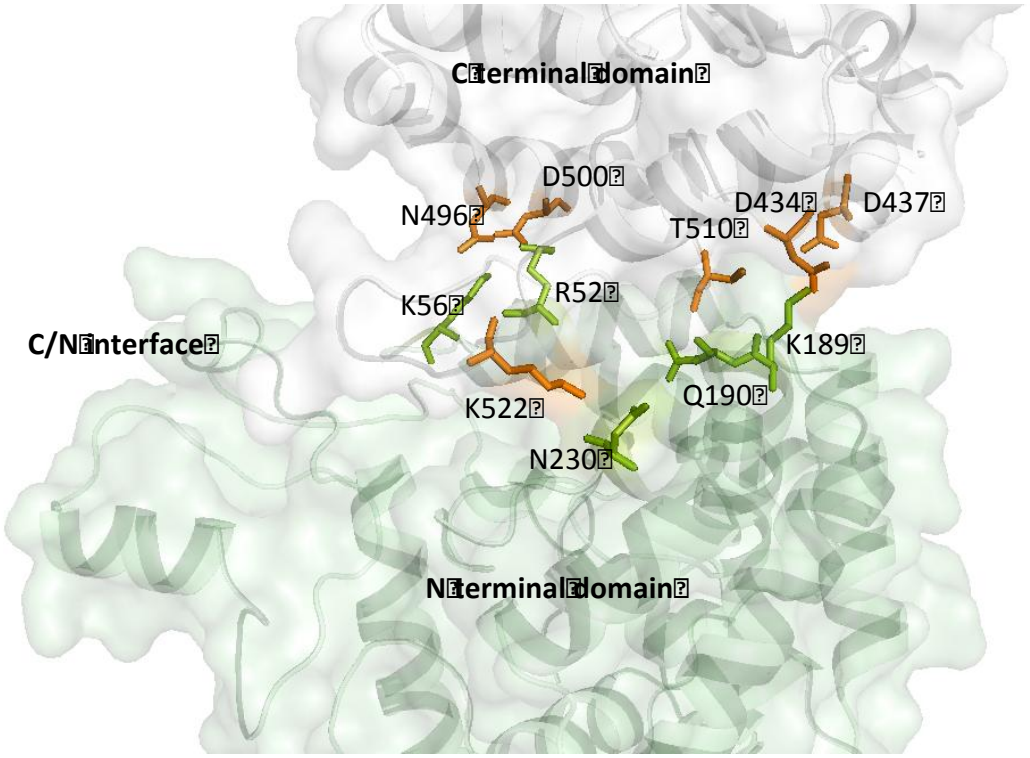


Figure 3.19 The interface between each subunit of the symmetric trimer. The residues involved in the symmetric trimer on N terminal domain (green) are colored in green, and the residues on C terminal domain (grey) are colored in orange.

Chapter 4: Functional study of LASV NP

Based on the structural information, several assays were reported in this chapter to characterize the functions of LASV NP. From the structure, the C-terminal domain of the NP is 3'-5' exoribonuclease, which may play crucial role in immune suppression, and the N-terminal domain can bind cap, which may be involved in cap-snatching. To test the functions, NP mutants were made for catalytic residues, zinc binding residues and the cap binding residues, and in vitro and in vivo assays were performed.

4.1 Purified LASV NP protein binds RNAs

LASV nucleoprotein plays an essential role in protecting viral genomic RNA and forming ribonucleoprotein (RNP) complex, many purified nucleoproteins of negative-strand RNA viruses (Rabies virus, Respiratory Syncytial virus and Vesicular Stomatitis virus) contain RNAs from the expression host (Albertini, 2006; MacLellan, 2007; Todd, 2006). We wondered whether the purified LASV NP contains RNAs from the E.coli expression cells. To test this, the LASV NP was loaded on a 2% agarose gel and the RNAs were detected as below.

4.1.1 The agarose gel

A 2 % agarose gel was prepared with TAE buffer. 5μl of RNA loading buffer (6x) was added to 15μl purified NP protein. The concentration of the NP protein is 5mg/ml. The gel was run for a short time to avoid RNA running off the gel as the RNA size is quite small. Briefly, 1 gram agarose powder was added into 100 ml 1× TAE buffer,

and the mixture was heated in a microwave oven until the solution became clear. After the solution was cooled for 10 minutes, 5µl ethidium bromide (10 mg/ml stock), a fluorescent dye, which can intercalate between bases of nucleic acids (DNA and RNA) was added. The electrophoresis was carried out under a voltage of 110v for 15 minutes in 1xTAE buffer. The RNAs were visualized using a UV transilluminator (UVP BioDot-IT system). The RNAs that are bound with NP protein are shown as a very strong band on the following gel picture (Fig. 4.1), even after so many purification steps the RNAs still stay.

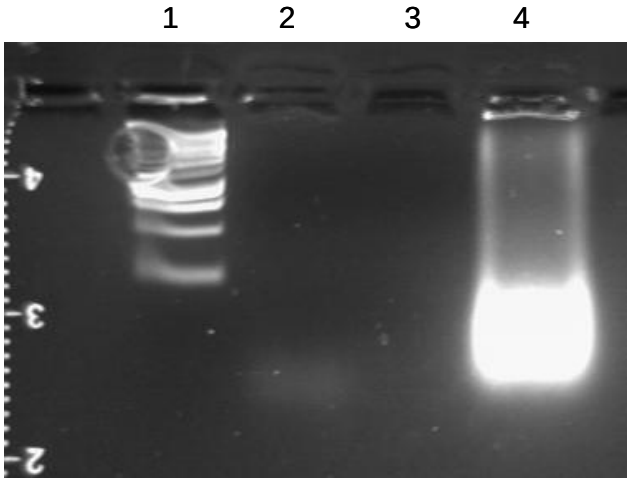


Figure 4.1 Electrophoresis gel of the bound RNAs in Z and NP proteins. Lane 1: DNA marker; lane 2: monomer Z protein from LASV; lane3: Z protein oligomer; lane4: NP protein.

4.1.2 Denature polyacrylamide gel electrophoresis of RNA

To determine the length of the bound RNAs in the LASV NP, we decided to use urea-denaturing polyacrylamide gel with a single strand RNA standard kit. 20 ml of 16% polyacrylamide gel was made using 2.0ml 10X TBE, 9.6g urea and 12ml 25% acrylamide solution (with gloves, neurotoxin). After adding 10 µl TEMED and 100 µl

10% fresh ammonium persulfate to the gel solution, The gel solution mix was swirled gently, and then quickly pour into the gel plate without any bubbles. The comb was put into the top of the gel leaving the gel to polymerize for about 1 hour in a Bio-RAD gel cassette.

Once the gel was polymerized, the comb was removed and the gel cassette was put into the reservoir. The gel was pre-run in 1× TBE for 35 minutes to remove some contaminants. The sample was prepared with 15µl of the purified NP protein and 5µl of the formamid loading buffer, and then heat at 90°C for 5 minutes. The electrophoresis was carried out for 1.5 hrs and the RNA gel was stained in 0.04% EB containing 1× TBE for 20 min. The RNAs were visualized using the UV transilluminator.

The gel showed that there are two main bands of the bound RNAs, corresponding to 120 bp and 100 bp respectively (Fig. 4.2). As the RNAs are from the host E.coli cells and they are nonspecific binding, it is not surprised that the lengths of the RNA are different. Although the NP is a 3'-5' exoribonuclease, this result indicates that the NP protein can protect the RNAs from the exonuclease.

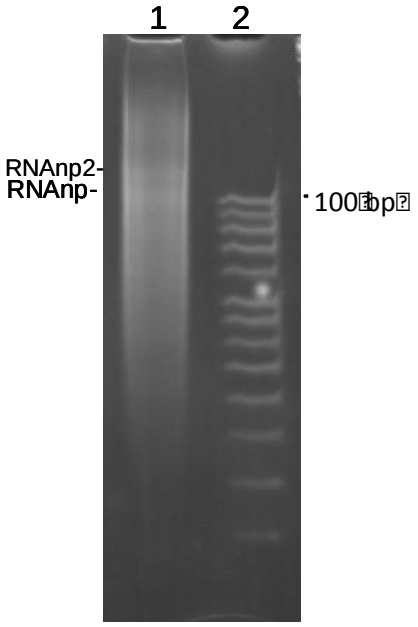


Figure 4.2 Denaturing eletrophoresis gel of NP-RNA complex. Lane 1: NP protein with RNA; lane 2: RNA marker.

4.2 Characterization of the 3'-5' exoribonuclease of LASV NP C-terminal domain

4.2.1 The LASV NP is a 3'-5' nuclease

The LASV NP structure and the NP in complex with Mn2+ have shown that the LASV NP is a 3'-5' nuclease, belonging to DEDDh subfamily with the catalytic residues D389, E391, D466, H528 and D533. Interestingly, Residue D389 is proven to be crucial for immune suppression (corresponding to D382 with LCMV/NP, Martínez-Sobrido, 2009).

In order to determine the residues which are key for the 3'-5' exoribonuclease function that leading to the suppression of type I interferon. In vitro and in vivo assays were performed. The catalytic residues D389, E391, G392, D466, H528 and D533

that reside around Mn^{2+} and are conserved in all arenavirus, which indicates that the catalytic residues are essential (Fig. 4.3).

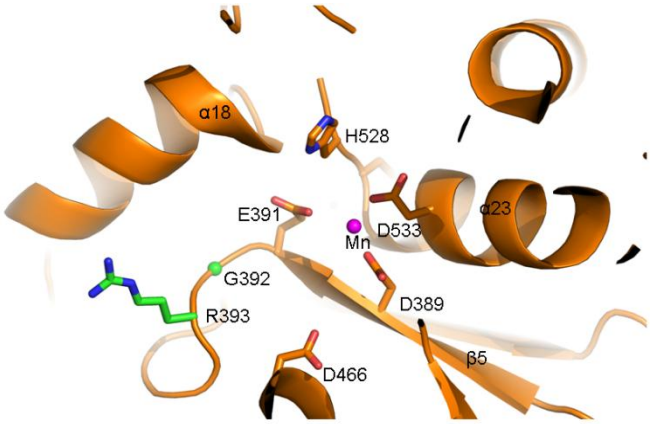


Figure 4.3 The nuclease active site of C-terminal domain of NP and the related residues. Mn^{2+} is colored in pink, residue G392 is colored in green, and the rest residues are D389, E391, D466, H528 and D533.

4.2.2 Generation of catalytic residue mutants of LASV NP

4.2.2.1 Mutants primers design and molecular cloning

The substitution of the catalytic residues D389A, E391A, G392A, D466A, H528A, C529A, D533A by alanine were performed according to the modified protocol from QuikChange™ (Liu, 2008).

To increase the PCR efficiency, the designed primers contain 22 to 25 bases of non-overlapping region at its 3' end and 19 to 21 bases of the overlapping region at its primer to primer complementary 5' end (Fig. 4.4) depending on each mutants, thus the melting temperature for the non-overlapping region is 9°C to 12°C higher than the overlapping region (see Appendix I.3 for the mutants primers). These newly designed

primers decrease the self-annealing of the complementary primers and help to bridge the strand break of the newly synthesized DNA from the first PCR cycle which will be used as template for next cycle of the PCR reaction. As the parental templates were methylated, which will be removed by *Dpn*I, and the newly synthesized DNA containing the mutations will be transformed into DH5α competent cells.

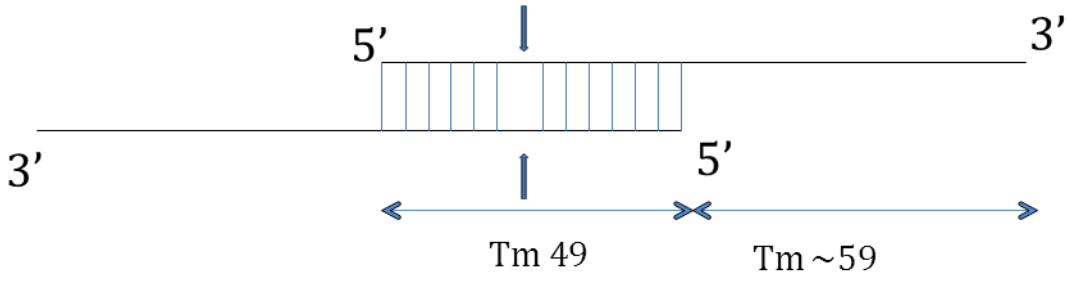


Figure 4.4 Primer designed for site directed mutants of LASV NP. The primers contain 3' end non-overlapping region with the melting temperature is around 59°C and 5' overlapping region with the melting temperature is around 49°C. The mutated site is located on the overlapping region.

The 50ul PCR water solution contains 2-10 ng template, 1uM forward primer, 1uM reverse primer, 200 uM dNTPs and 3 units Pfu DNA polymerase. The reactions were started with 95°C for 5 minutes, then 12 cycles of amplification. Each of the cycles contains 95°C for 1 minute, annealing for 1 minute and 72°C for 18 minutes from extension. All the mutants were confirmed by DNA sequencing.

4.2.2.2 Mutant proteins expression and purification

The mutant proteins D389A, E391A, G392A, D466A, H528A and D533A which are putative catalytic residues are expressed and purified as native NP described in

previous chapter (Fig. 4.5-4.10). They all show two forms of oligomers that are hexamer and trimer, based on the elution volume of the target protein.

I. The NP D389A mutant expression and purification

The mutant D389A was expressed and purified using similar methods as the native NP. Briefly, the Rosetta cells that contains the target mutant gene NP D389A plasmid were innocated into 10 liters of LB medium in present of 50 μg/ml Ampcilin and 34 μg/ml chloramphenicol (Duchefa Biochemic) at 37°C, and the protein was induced by 0.03mM IPTG when the OD_{600} of the cell cultures reached to 0.6. The NP protein was induced overnight at 20°C. The cells were harvested, and the cells were broken by sonication. The fusion protein was purified by amylose resin, and the MBP was removed by TEV cleavage. The D389A mutant protein is finally purified by a gel filtration column (Fig. 4.5 top), showing that the first peak is NP which is separated from MBP according to the size difference. To test the purity, each fraction was loaded on a SDS-PAGE (Fig. 4.5 bottom). 15mg of the purified trimeric NP D389A mutant protein was obtained from 10-liter cultures.

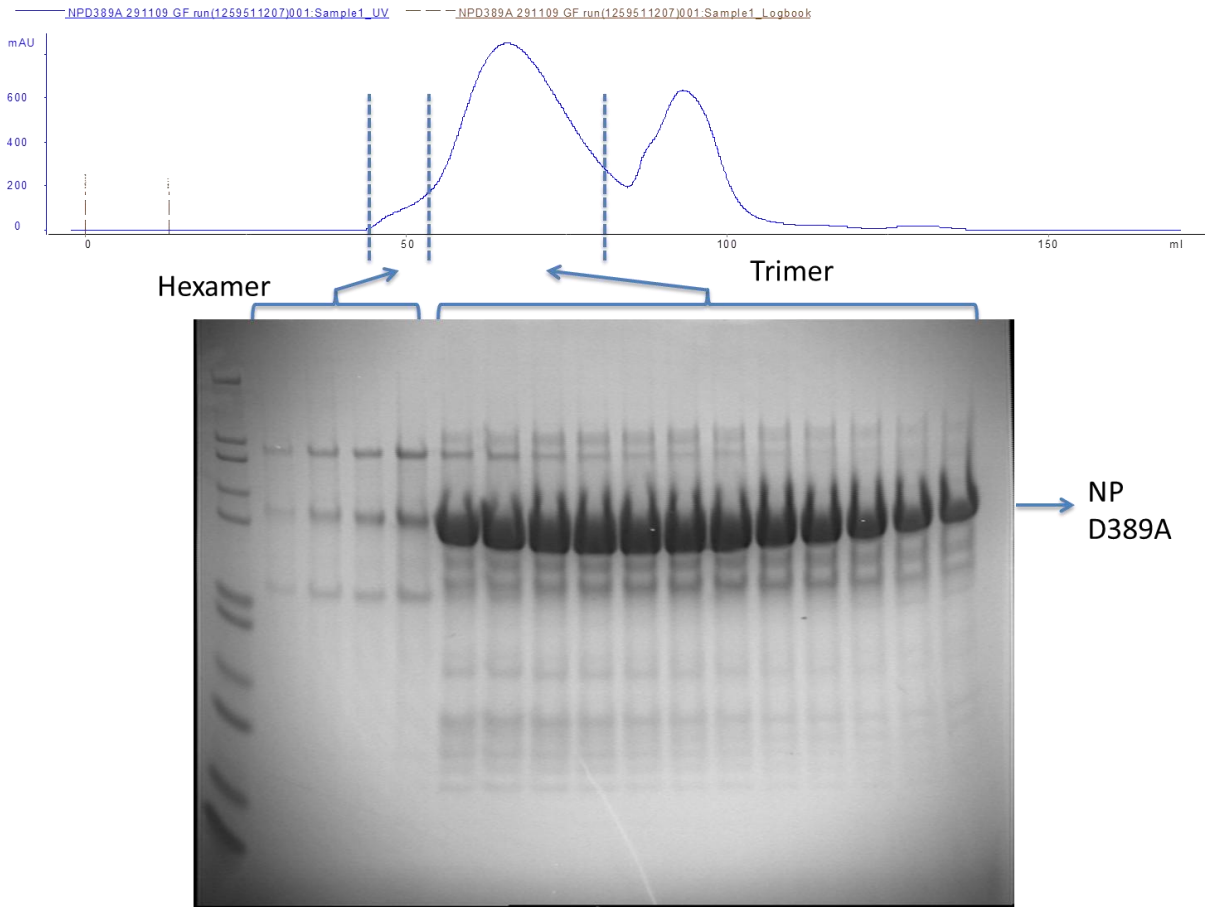


Figure 4.5 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein D389A.

II. The NP E391A mutant protein expression and purification

The mutant E391A protein was expressed and purified using similar methods as mutant D389A that is described above. The E391A mutant is finally purified by gel filtration column (Fig. 4.6 top), and the purity was test by a SDS-PAGE gel (Fig. 4.6 bottom). 7mg of the purified trimeric NP E391A mutant protein was obtained from 10 liters of culture.

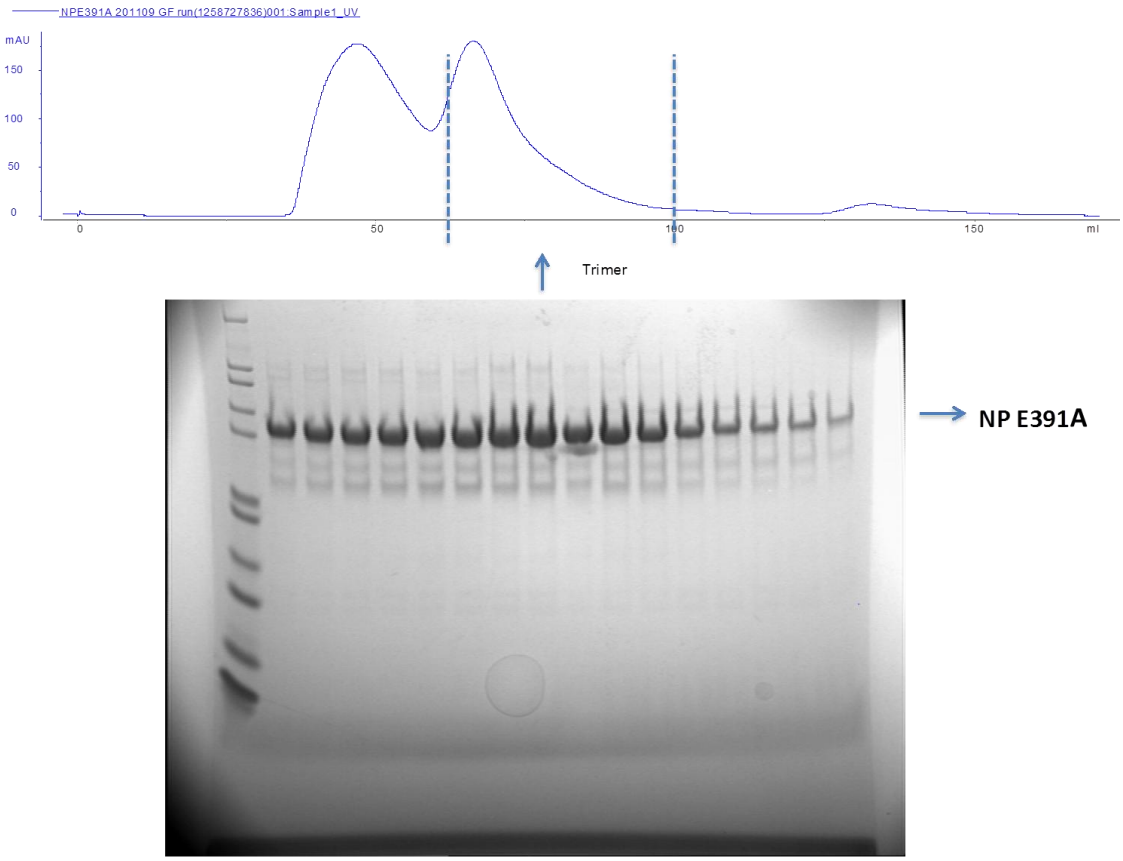


Figure 4.6 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein E391A.

III. The NP G392A mutant protein expression and purification

The mutant G392A was expressed and purified using similar methods as mutant D389A that is described above. The G392A mutant is finally purified by a gel filtration column (Fig. 4.7 top), and the purity was checked by a SDS-PAGE gel (Fig. 4.7 bottom). 6mg of the purified trimeric NP G392A mutant protein was obtained from 10 liters of culture.

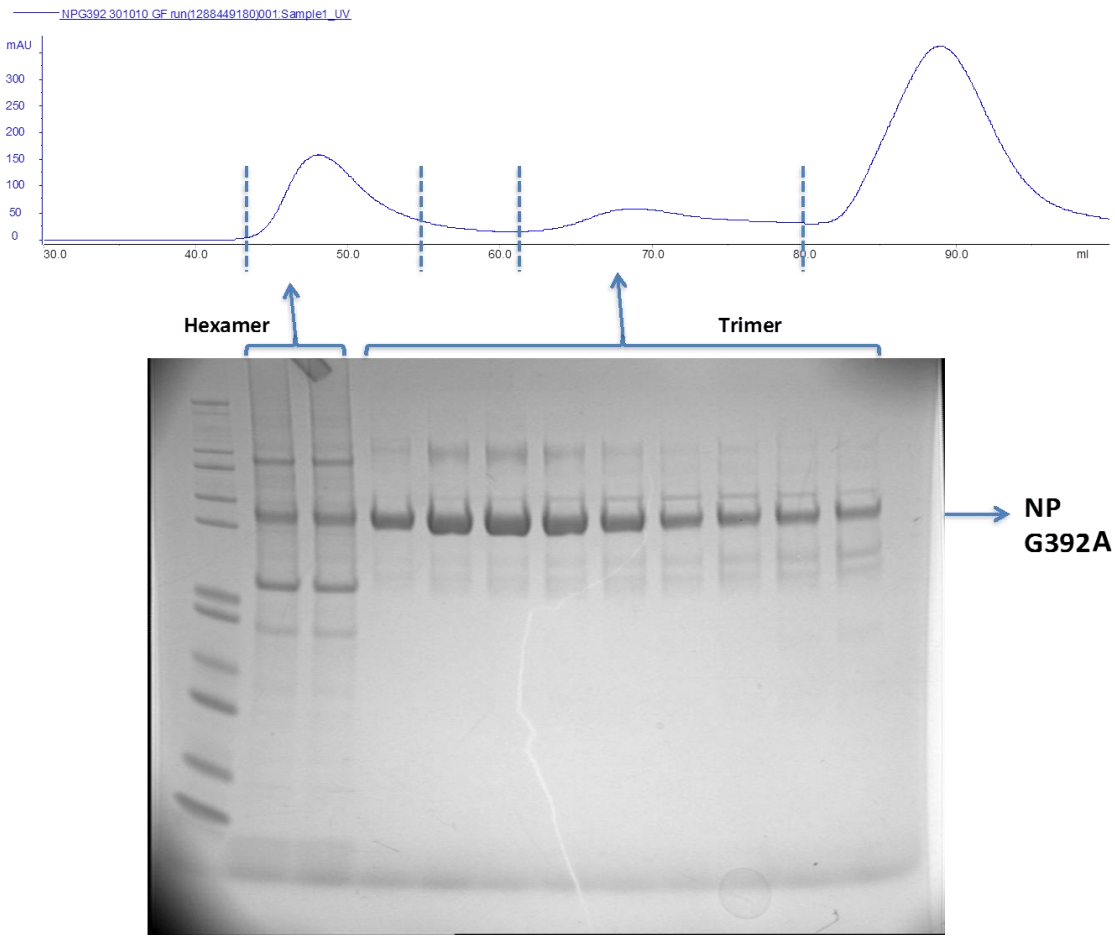


Figure 4.7 The gel filtration chromatography and the SDS-PAGE gel picture of the purified NP mutant protein G392A.

IV. The NP D466A mutant protein expression and purification

The mutant D466A was expressed and purified using similar methods as mutant D389A that is described above. The D466A mutant is finally purified by gel filtration column (Fig. 4.8 top), and the purity was determined by a SDS-PAGE gel (Fig. 4.8 bottom). 10mg of the purified trimeric NP D466A mutant protein was obtained from 10 liters of culture.

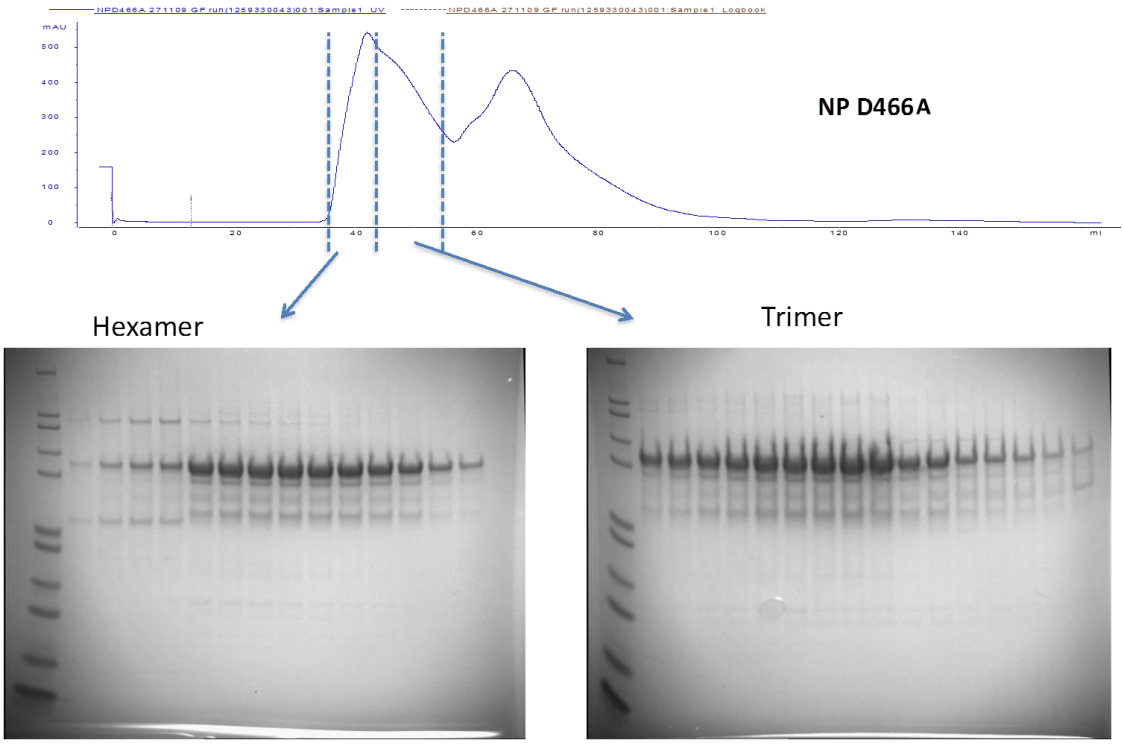


Figure 4.8 The gel filtration chromatography and the SDS-PAGE gel picture of the purified NP mutant protein D466A.

V. The NP H528A mutant protein expression and purification

The mutant H528A NP was expressed and purified using similar methods as mutant D389A that is described above. The H528A mutant NP is finally purified by a gel filtration column (Fig. 4.9 top), and the purity was checked by a SDS-PAGE gel (Fig.4.9 bottom). 12mg of the purified trimeric NP H528A mutant protein was obtained from 10 liters of culture.

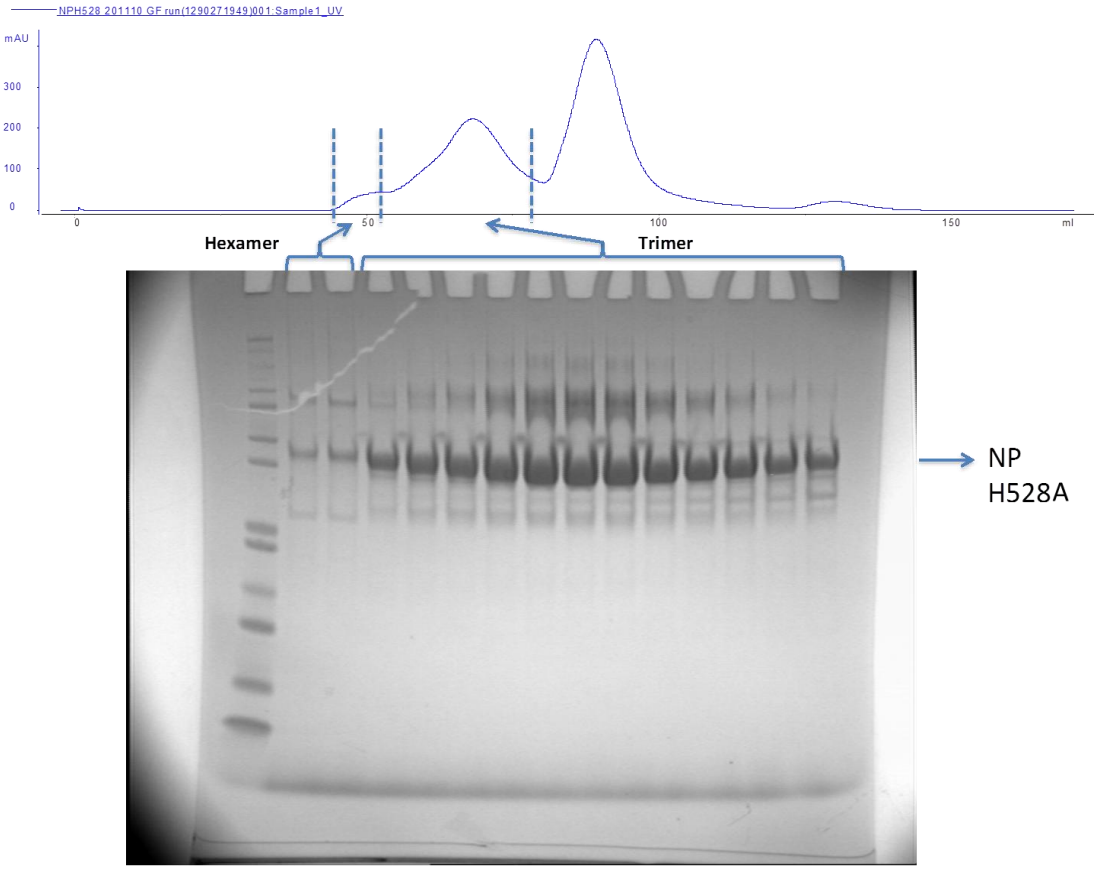


Figure 4.9 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein H528A.

VI. The NP D533A mutant NP expression and purification

The mutant D533A NP was expressed and purified using similar methods as mutant D389A that is described above. The D533A mutant is finally purified by a gel filtration column (Fig. 4.10 top), and the purity was tested by a SDS-PAGE gel (Fig. 4.10 bottom). 15mg of the purified trimeric NP D533A mutant protein was obtained from 10 liters of culture.

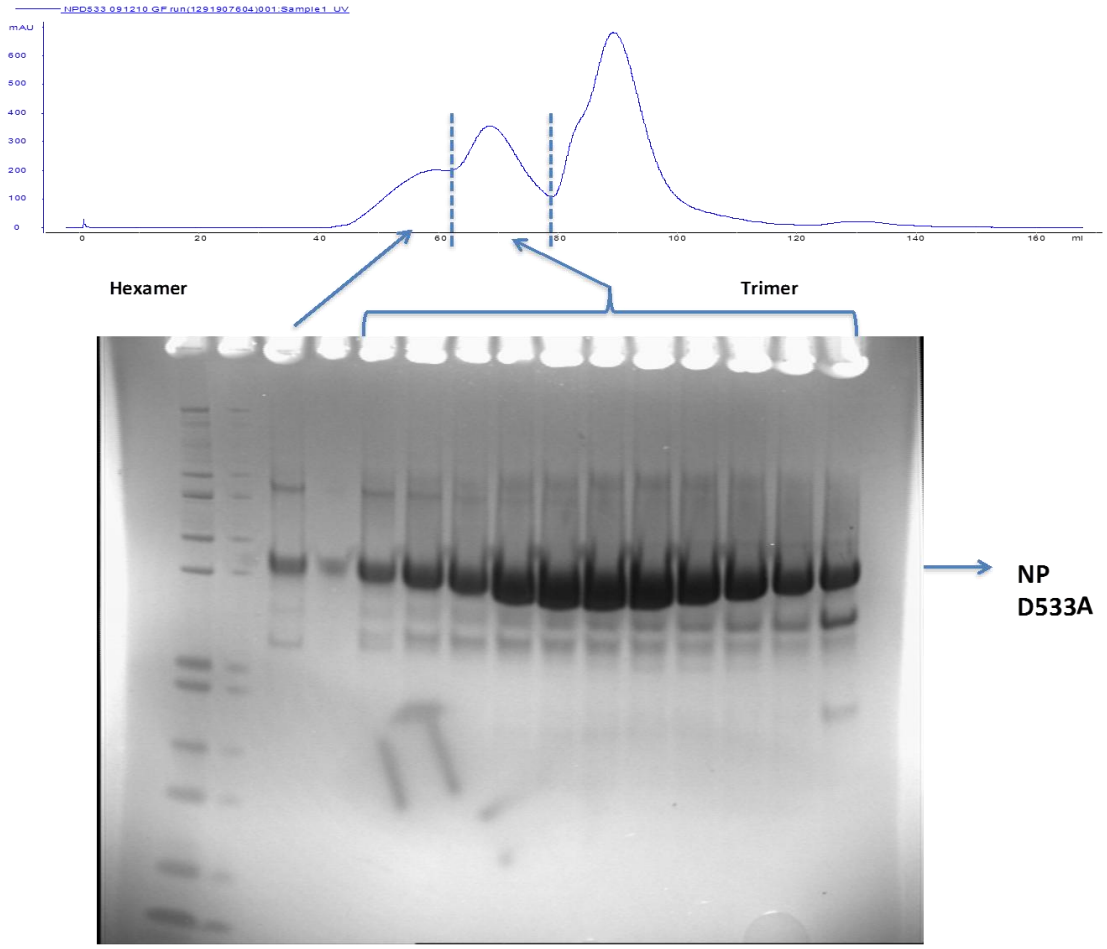


Figure 4.10 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein D533A.

4.2.3 Quanternary structures of the NP mutants D389A, E391A and D466A

All these three NP mutants D389A, E391A and D466A share the identical quanternary structure with native NP. Although these purified mutants bind more RNAs than native NP, which prohibited us to compare their quaternary structures by circular dichroism spectroscopy. The gel filtration graphs (Fig. 4.11) strongly suggest that the mutants fold properly as the native NP. This is further confirmed by our functional LASV minigenome assay showing that all the three mutants had the similar transcriptional activities as the native NP.

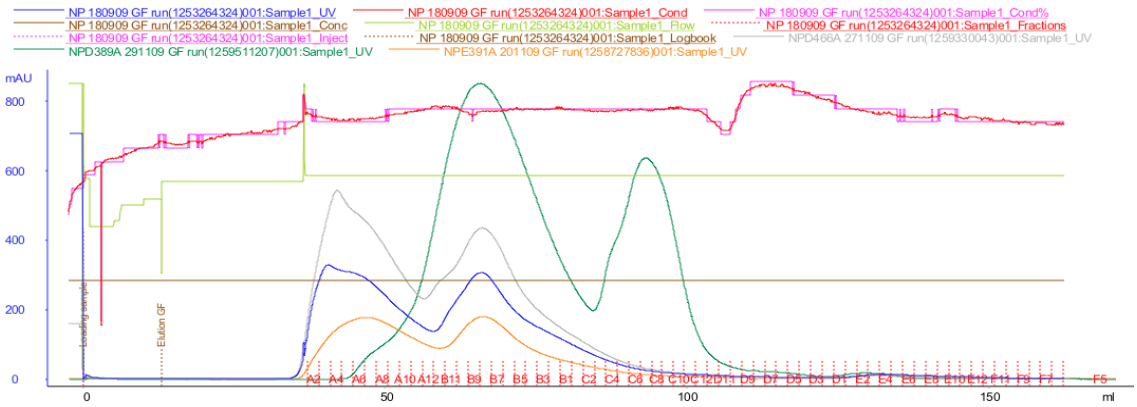


Figure 4.11 Gel filtration chromatographs of the WT and mutant LASV NP proteins. The gel filtration chromatography was performed using a column Hiload 16/60 Superdex 200 (GE healthcare) and AKTAxpress (GE healthcare). The first peak is the hexameric and the second peak is the trimeric LASV NP. The third peak is MBP. D389A (green line), E391A (orange line), D466A (grey line) and WT (blue line).

4.2.4 The in vitro activity determination of the 3'-5' exoribonuclease

As above, the structural information of the NP protein indicated that LASV NP is a 3'-5' exoribonuclease. To test this hypothsis, in vitro assays of RNA degradation were performed using both native and mutated NP proteins.

4.2.4.1 LASV NP degrades double strand DNAs

The experiment was carried out with 1kb dsDNA ladder which was incubated at 37^{o} C for an hour in the presence of 10 mM EDTA as control 1, 1kb dsDNA ladder and 10mM EDTA together with trimeric NP protein as control 2, 1kb dsDNA ladder and 10mM $MnCl_2$ together with trimeric NP protein in line 3, and 1kb dsDNA ladder with 10mM $MnCl_2$ together with hexameric NP protein in line4 (Fig. 4.12). The dsDNA ladders were digested and then separated on 1% agarose gel, and the gel was stained by ethidium bromide. The digestion result shows that with the presence of Mn^{2+} both the trimeric and hexameric NPs can degrade dsDNA.

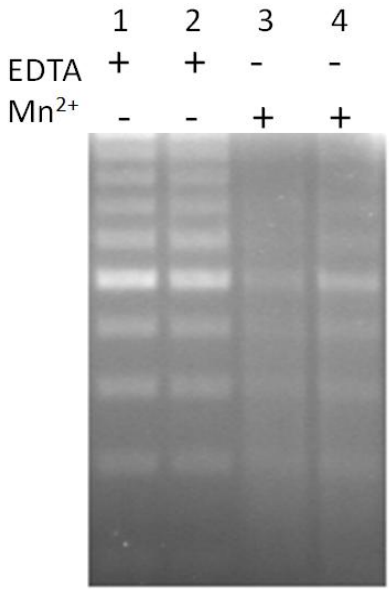


Figure 4.12 Both trimeric and hexameric LASV NP proteins degrade double strand DNA. Lane 1, control 1, DNA ladder in presence of EDTA; lane 2, control 2, DNA ladder, EDTA and trimeric NP; lane 3, DNA ladder, $MnCl_2$ and trimeric NPs; lane4, DNA ladder, $MnCl_2$ and hexameric NPs.

4.2.4.2 LASV NP degrades both ssRNA and dsRNA

In order to test the 3'-5' ribonuclease activities, different RNA substrates were prepared as below:

4.2.4.2.1 In vitro chemical and biological RNA synthesis

To synthesize single strand RNAs in vitro, the 30-nucleotide cRNA (sense) sequence 5'-CUGGGCUUACCUAUUCUCAGCUGAUGACCC-3' was derived from the LASV NP (Josiah strain) S segment (nucleotides 2186–2215 in antigenomic orientation) and chemically synthesized by Eurogenes. The 30-nucleotide vRNA (in genomic orientation) sequence 5'-GGGUCAUCAGCUGAGAAUAGGUAAGCCCAG-3' was complementary to the cRNA. The cRNA (30 nucleotides) was used as one of the three substrates for

3’–5’exoribonuclease assay. The 5’-triphosphorylated vRNA was generated by in vitro transcription of the partial dsDNA template formed by the T7 promoter sequence 5’-AATTTAATACGACTCACTATAGG-3’ and the reverse complement of the T7 promoter sequence and of the LASV (Josiah strain) S segment (nucleotides 2186–2215)

5’-CTGGGCTTACCTATTCTCAGCTGATGACCCTATAGTGAGTCGTATTAAATT-3’

using the T7 MEGAshortscript kit following the manufacturer’s instructions (Ambion). A similar strategy was used to generate the 32-nucleotide triphosphorylated cRNA with the T7 primer and LASV (Josiah strain) S segment (nucleotides 2186–2213)

5’-GGGTCATCAGCTGAGAATAGGTAAGCCCAGCCTATAGTGAGTCGTATTAA ATT-3’. A similar strategy was used to generate the 60-nucleotide vRNA corresponding to LASV (Josiah strain) S segment (nucleo- tides 2186–2213), using the partial dsDNA template formed by 5’-AATTTAATACGACTCACTATAGG-3’ and

5’-GTAAATCCCTGCAGTCGGCAGGGTTTACCGCTGGGCTTACCTATTCTCAG CTGATGACCCTATAGTGAGTCGTAT TAAATT-3’ as the template. To synthesize the capped viral mRNA transcripts corresponding to nucleotides 992–1117 of the LASV NP gene, the DNA template was PCR amplified from the NP expression plasmid with a forward primer

5’-<u>AATTTAATACGACTCACTATAGGG</u>AAAACACTGTCGTTGATCTGGAATC-3’ (underlined are T7 promoter sequences) and a reverse primer

5’-GGGTCATCAGCTGAGAATAGGTAAGCCCAGCGG-3’, and subjected to in vitro RNA synthesis using the mMESSAGE mMACHINE T7 Ultra kit (Ambion) following the manufacturer’s instruction, except that no poly(A) tail was added (see Appendix I.4).

To obtain the blunted dsRNA (5’-OH dsRNA), both chemically synthesized 30 nucleotide unphosphorylated cRNA and vRNA oligonucleotides were dissolved into 0.1 M NaCl, 1 mM EDTA and 0.1 M Tris pH 8.0 at the final concentration of 200 mM, and an equal amount of the two oligonucleotides was mixed together and annealed in a thermo-cycler as follows: 95^{o}C for 3 min, 68^{o}C for 1 min and then 4^{o}C. To generate the double 5’-triphosphorylated dsRNA, equal amounts of the triphosphorylated 5’ppp-vRNA and 5’ppp-cRNA were annealed in vitro. To make the singly 5’-triphosphorylated dsRNA, equal amounts of the in vitro synthesized 32-nucleotide 5’ppp-vRNA and the chemically synthesized 30-nucleotide unphosphorylated cRNA were annealed in vitro (see Appendix I.5).

The human 18S rRNA fragment (128 nucleotides) was generated by a T7 RNA polymerase directed in vitro RNA synthesis reaction, using the pTRI-RNA 18S control plasmid (Ambion), following the manufacturer’s instruction. A plasmid phRL-CMV that encodes the T7 promoter (T7p)-directed human β-globin gene was provided by R. Elliott and G. Blakqori. The T7 β-globin DNA fragment was purified by agarose electrophoresis after digestion of the phRL-CMV plasmid with HindIII

and SmaI. The capped human globin mRNA transcripts were generated using the T7p-globin fragment as a template and the mMESSAGE mMACHINE T7 Ultra kit from Ambion, and the poly(A) tail was added following the manufacturer's instruction.

The ssRNA markers (perfect RNA markers, 0.1–1 kb) were purchased from Novagen. The low molecular mass ssRNA marker (10–100 nucleotides) was purchased from USB. The dsRNA ladder (21–500 bp) was purchased from New England Biolabs.

4.2.4.2.2 The LASV NPs degrade single strand RNA

The first experiment was carried out in 12 ul reaction solution containing 2ul RNA ladder with the size range from 0.1kb to 1kb (perfect RNA markers, Merck). The reaction solutions were incubated with or without 7 ug NPs in 0.3 M NaCl, 10% glycerol and 20 mM Tris pH 7.5 and 10mM $MnCl_2$ at 37^{o}C for a time course of 20 minutes, 40 minutes and 80 minutes, and the three controls were incubated for 80 minutes. All these reactions were stopped by adding EDTA to a final concentration of 20mM. The samples were mixed with equal volumes of RNA loading buffer (Ambion), heated at 95 ^{o}C for 3 min, cooled on ice for 5 min, and separated in 6% polyacrylamide gel containing 8M. The gels were stained in 0.05% ethidium bromide for 25 min, visualized using the 2UV transilluminator (UVP). The gel result shows that LASV NP degrades different size of ssRNA on the range of 0.1 to 1kb (Fig. 4.13).

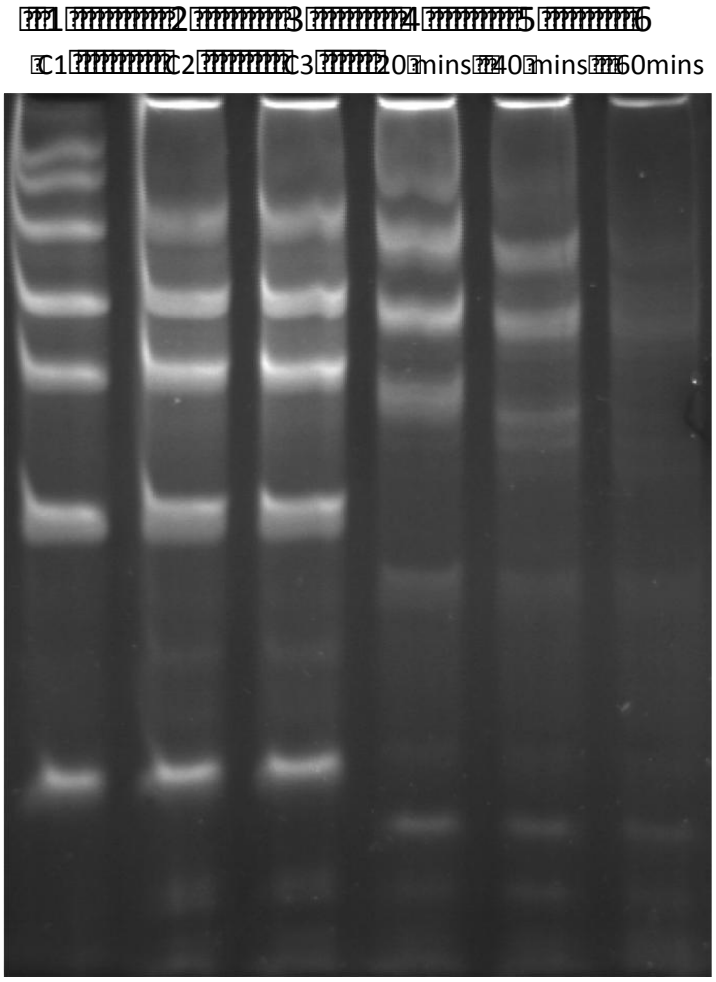


Figure 4.13 Lassa NP protein degrades different size of ssRNA. Lane1, control 1, EDTA; lane2, control 2, EDTA and NP; lane3, control 3, NP; lane 4, NP, $MnCl_2$ with 20 minutes incubation; lane5, NP, $MnCl_2$ with 40 minutes incubation; lane6, NP, $MnCl_2$ with 60 minutes incubation.

To test whether the LASV NP can discriminate viral RNA (vRNA), viral complementary RNA (cRNA) and viral message RNA (mRNA), we mixed those three substrates together and did the in vitro cleavage assays. The activities of the NP native and three catalytic residue mutants D389A, E391A and D466A were determined as shown in figure 4.14. The capped viral mRNA contains 126nt, showed as the top lane; the viral genomic sense RNA contains 60nt, showed as the middle lane; the complementary antigenomic sense RNA contains 30nt, showed as the bottom lane.

These ssRNA substrates were incubated at 37°C for one hour unless otherwise stated (line 4). The samples were then separated on 15% PAGE gel containing 8M urea. Both trimeric and hexameric WT NPs degrade these three ssRNA substrates even with 30 minutes incubation of trimeric NPs. But the trimeric mutants D389A, E391A and D466A show significant reduction of the RNase activity in vitro (Fig. 4.14). Control 1 is the mixed of the three ssRNAs without NP, whearas control 2 shows the ssRNA substrates with WT trimeric NP in the presence of 10mM EDTA.

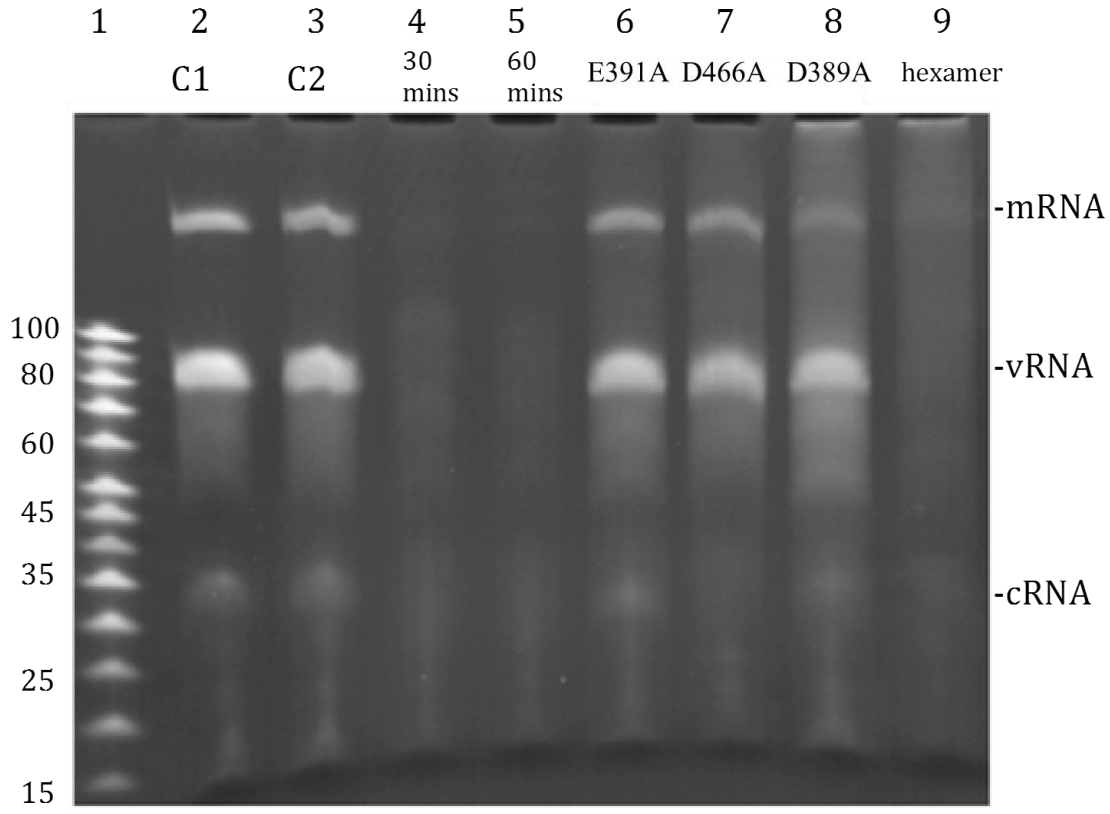


Figure 4.14 Both trimeric and hexameric WT NPs but not its catalytic mutants degrade ssRNA. Lane 1, RNA ladder; lane 2, control 1, ssRNA substrate without NP; lane 3, control 2, ssRNA, WT NP and 10mM EDTA; lane 4, ssRNA, WT NP for 30 minutes incubation; lane 5, ssRNA and WT NP for 60 minutes incubation; lanes 6-8, ssRNA, trimeric NP mutants D389A, E391A and D466A; lane 9, WT hexameric NP. The ssRNA substrates contain capped mRNA, cRNA and vRNA, which were marked as above.

LASV NPs degrade various cellular RNAs with a priority of shorter one. The comparision of the degradation efficiency between 128nt 18s rRNA and human β-globin mRNA was shown in figure 4.15. The β-globin mRNA was synthesized by an in vitro transcription reaction using mMESSAGE mMACHINE T7 ultra kit (Ambion), which incorporated the 5'cap analogue and 3' polyA sequence. The human 18S rRNA fragment was in vitro transcribed using the MEGAshortscript kit (Ambion) from the supplied control template. The controls are in present of 10 mM EDTA.

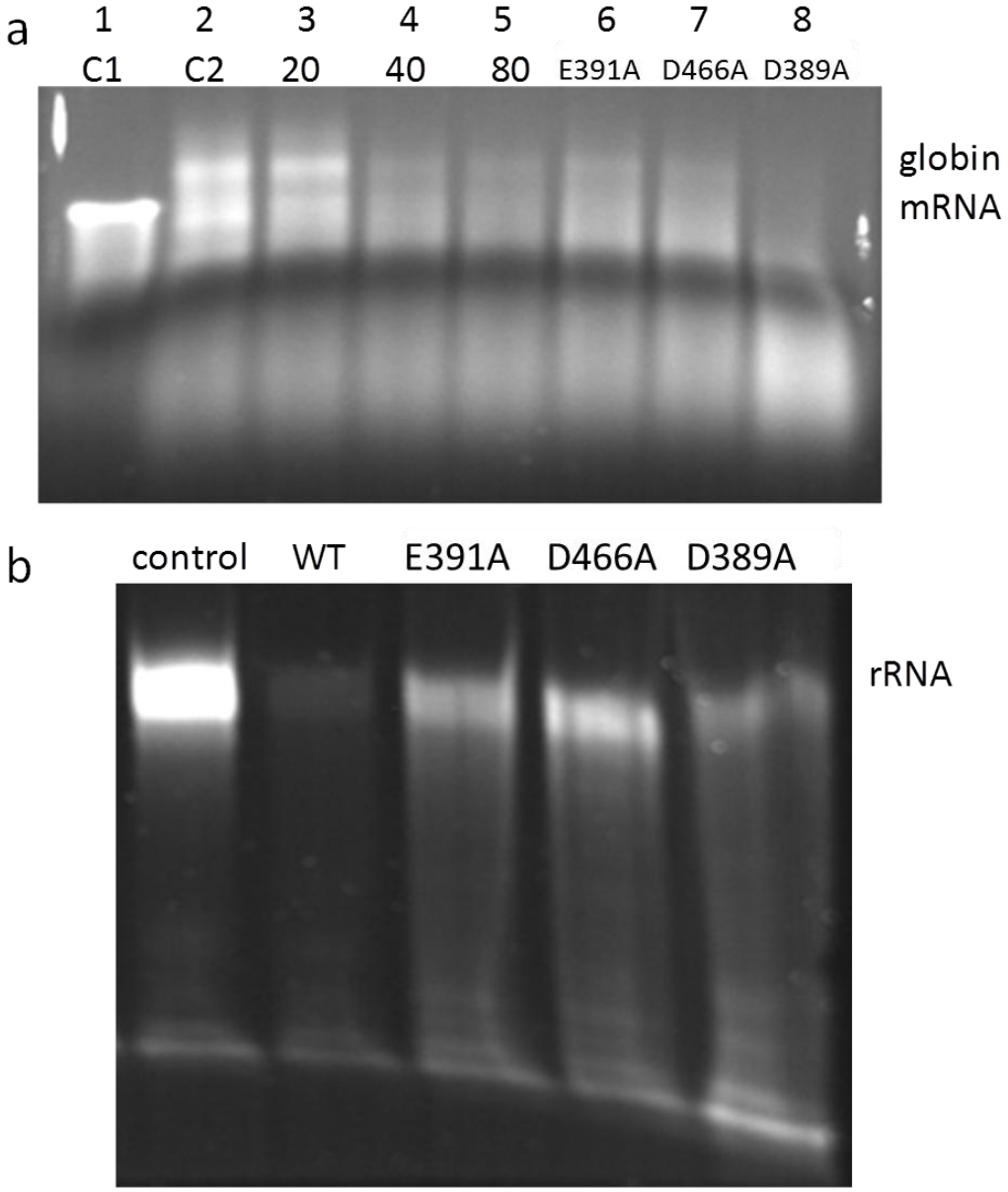


Figure 4.15 The comparision of the LASV NPs in degrading between 18s rRNA and human β-globin mRNA. a, the degradation level of human β-globin mRNA by either WT LASV NPs or its catalytic mutants; b, the degradation level of 18s rRNA by either WT LASV NPs or its three catalytic mutants.

4.2.4.2.3 LASV NP degrades double strand RNA

Further experiments were carried out to determine whether the WT LASV NP can degrade various sizes of double stranded RNA substrates. The first substrates were a dsRNA ladder which is from 21 to 500 bp (dsRNA ladder, New England Biolabs) (Fig.

4.16). 2ul of the dsRNA ladder was added to each of the reaction solutions that were incubated for 60 mins unless otherwise stated, with and without NP (WT or D389A mutant), EDTA, or $MnCl_2$, and separated by a 2% agarose gel.

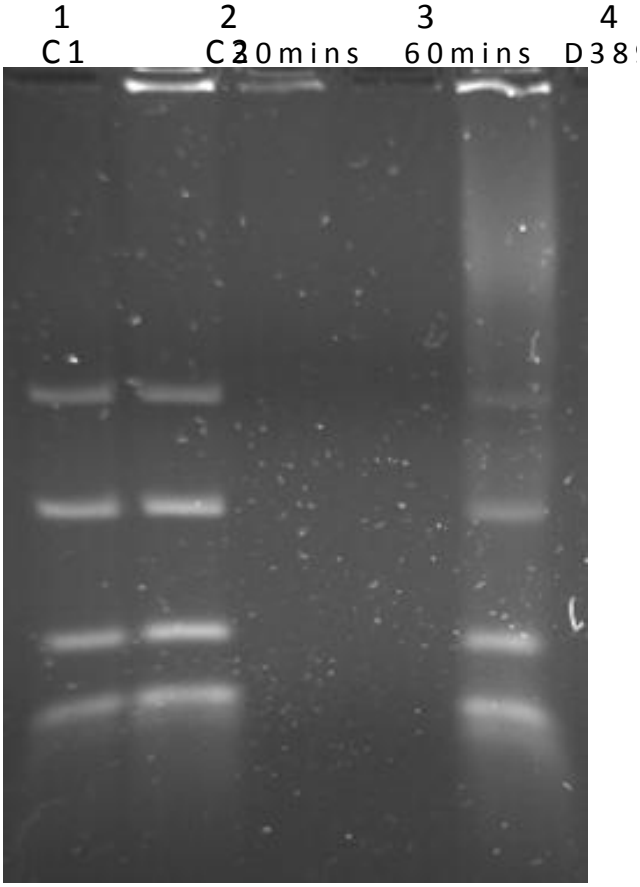


Figure 4.16 LASV NP degrades various sizes of dsRNA substrates. Lane 1, control 1, EDTA; lane 2, control 2, EDTA and WT NP; lane 3, WT NP and $MnCl_2$ for 30 mins incubation; lane 4, WT NP and $MnCl_2$ for 60 mins incubation; lane 5, NP mutants D389A and $MnCl_2$ for 60 mins.

LASV NP degrades dsRNAs regardless whether the dsRNA contains a 5' hydroxyle (5'-OH), or single triphosphaorylated 5' (5'ppp/5'-OH) or double triphosphorylated 5' (5'ppp/5'ppp) (Fig. 4.17). Although it also digests the long dsRNA mimic poly (I:C) (Fig. 4.18), the degradation efficiency is not as good as the shorter dsRNAs.

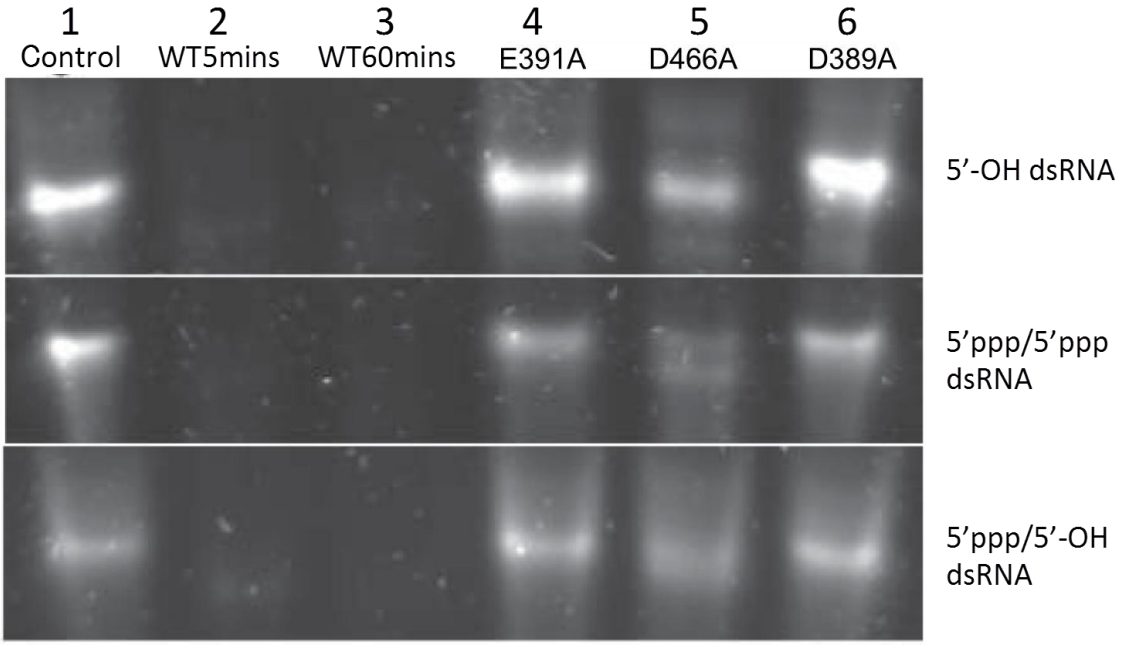


Figure 4.17 LASV NP degrades various dsRNAs. These dsRNAs contain 5’ hydroxyl dsRNA (5’-OH) (top), or single 5’ triphosphaorylated dsRNA (5’ppp/5’-OH) (bottom) or double 5’ triphosphorylated dsRNA (5’ppp/5’ppp) (middle).

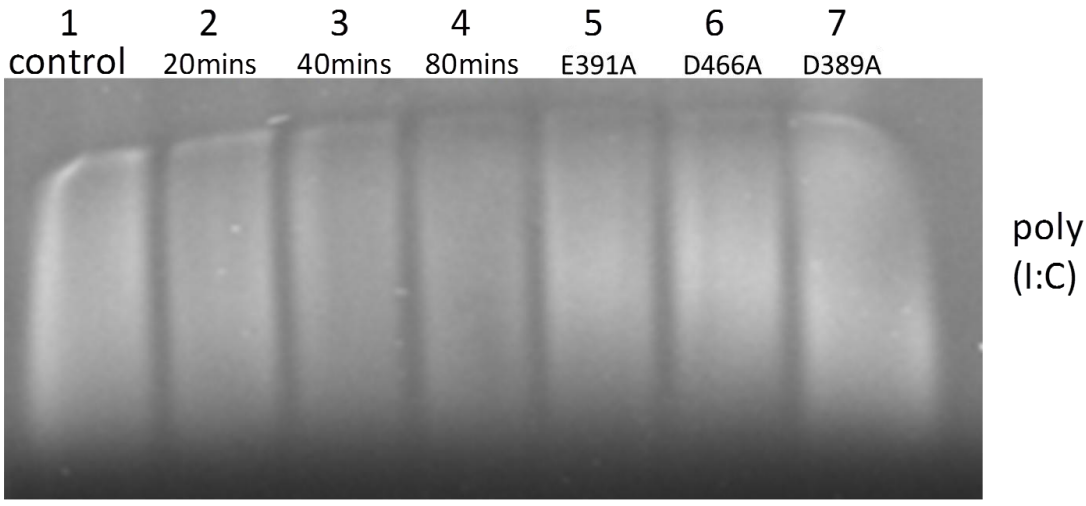


Figure 4.18. LASV NP degrades long dsRNA mimic poly (I:C). Lane 1, control without NP for 80 minutes incubation; lane 2, WT NP and poly (I:C) for 20 minutes incubation; lane 3, WT NP and poly (I:C) for 40 minutes; lane 4, WT NP and poly (I:C) for 80 minutes incubation; lane 5, poly (I:C) and mutant NP E391A for 80 minutes incubation; lane 6, poly (I:C) and mutant NP D466A for 80 minutes incubation; lane 7, poly (I:C) and mutant NP D389A for 80 minutes incubation.

4.2.4.3 The 3’-5’ exoribonuclease activity is divalent cation dependent

3’-5’ exonuclease is a metal dependent enzyme, and our structural studies have shown that the manganese is also found in the catalytic center of the crystal structure of LASV NP. Mn^{2+} is assumed to be essential for the RNase reaction. To check which metal ion is the best for the activity, we tested the RNA degradation efficiency with different divalent cations. The activity order of them is:

$Mn^{2+}>Co^{2+}>Mg^{2+}>Ca^{2+}>Zn^{2+}>Fe^{2+}>Ni^{2+}>Cu^{2+}$ (Fig. 4.19). In this experiment, three ssRNA substrates were mixed and incubated with the WT LASV NP protein at 37°C for 80 min (unless otherwise stated) in the presence of 10 mM EDTA (for control) or 10 mM of the respective divalent cation ions.

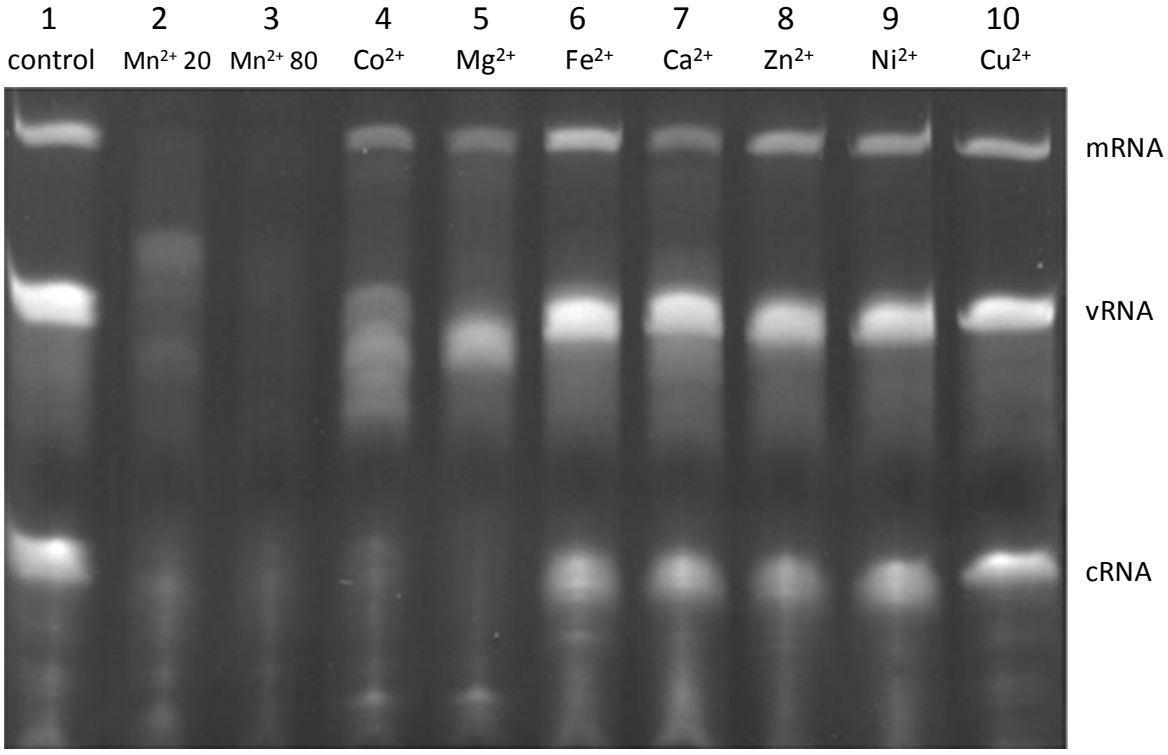


Figure 4.19 LASV NP is a divalent cation dependent 3’-5’ exoribonuclease. All the samples were incubated with WT LASV NP and three ssRNA substrates. Lane 1, control, EDTA; lane 2, Mn^{2+}, 20 minutes; lane 3, Mn^{2+}, 80 minutes; lanes 4-10, Co^{2+}, Mg^{2+}, Fe^{2+}, Ca^{2+}, Zn^{2+}, Ni^{2+}, Cu^{2+}, 80 minutes.

4.2.5 The zinc binding site of LASV NP

4.2.5.1 The zinc finger of LASV NP

Although during expression and purification, no zinc compounds were added, a zinc ion which is presented within C-terminal domain near the 3'-5' exoribonuclease active site, which was confirmed in the structure by fluorescence scanning analysis (Fig. 4.20). The residues that coordinate the zinc ion were E399A, C506A, H509A and C529A. The zinc finger fold domain is well known as the DNA-binding or RNA-binding site of the zinc finger protein. In the NP structure, the zinc binding residues are atypical CCHE (Fig. 4.20).

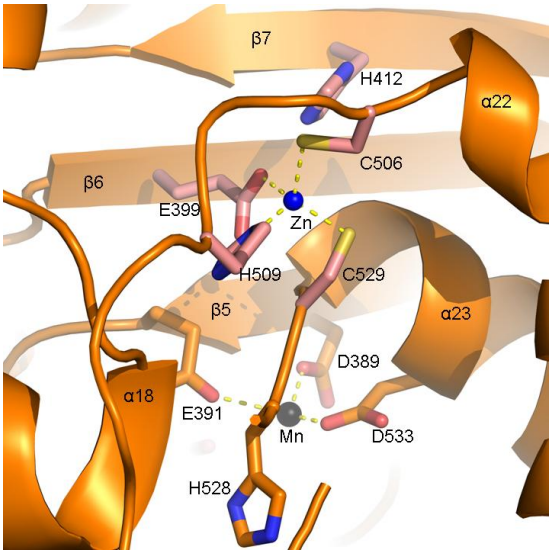


Figure 4.20 The zinc-binding site of LASV NP. The zinc ion is coordinated by residues E399A, H509A, C506A, and C529A. The zinc ion-binding site is very close to the Mn^{2+} binding site. Zn^{2+} is shown as a blue sphere and Mn^{2+} in black.

4.2.5.2 The NP mutants of zinc binding residues

The four NP mutants at zinc binding site were generated to test the RNA binding affinity (the result does not shown in this thesis) by using the similar mutagenesis as

descripted above, and the mutants were confirmed by DNA sequencing.

I. The NP E399A mutant protein expression and purification

The mutant E399A NP was expressed and purified using similar methods as the native NP. Briefly, the expression cells containing the target mutant gene NP E399A was innoculated into 10 liters of LB medium in present of 50 μg/ml Ampcilin and 34 μg/ml chloramphenicol (Duchefa Biochemic) at 37°C, and the protein was induced by 0.03mM IPTG when the OD_{600} of the cell culture reached to 0.6. The NP mutant protein was induced overnight at 20°C. The cells were harvested, and the cells were brokend by sonication, and the fusion protein was purified by amylose resin. After tev cleavage and removing MBP, the E399A mutant protein is finally purified by a gel filtration column (Fig. 4.21 top), showing that the first peak is NP which is separated from MBP according to the size difference. To test the purity, each fraction was loaded on a SDS-PAGE (Fig. 4.21 bottom). 0.3mg of the purified trimeric NP E399A mutant protein was obtained from 10 liters of culture.

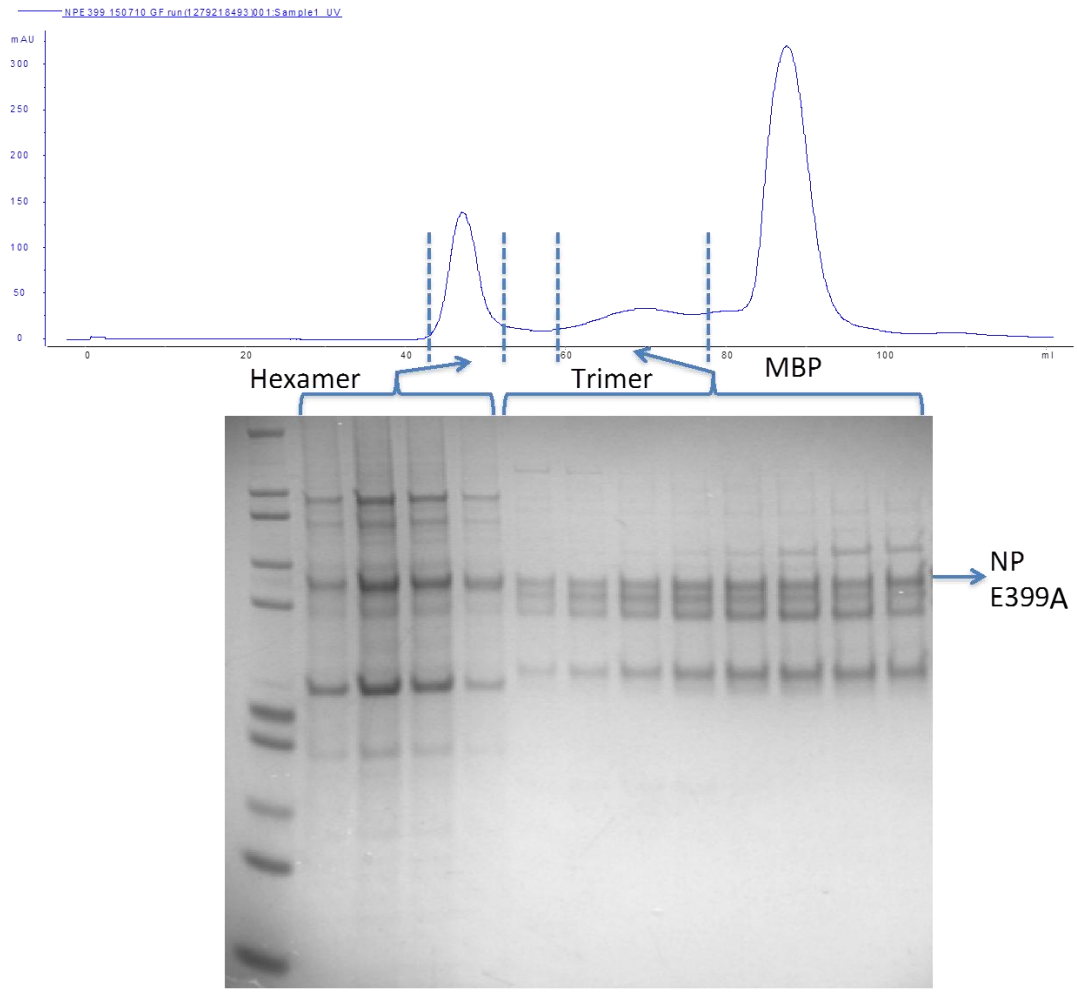


Figure 4.21 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein E399A.

II. The C506A mutant NP expression and purification

The NP mutant C506A was expressed and purified using similar methods as mutant E399A that is described above. The NP C506A mutant is finally purified by a gel filtration column (Fig. 4.22 top), and the protein purity was tested by a SDS-PAGE gel (Fig. 4.22 bottom). 0.2mg of the purified trimeric NP C506A mutant protein was obtained from 10 liters of culture.

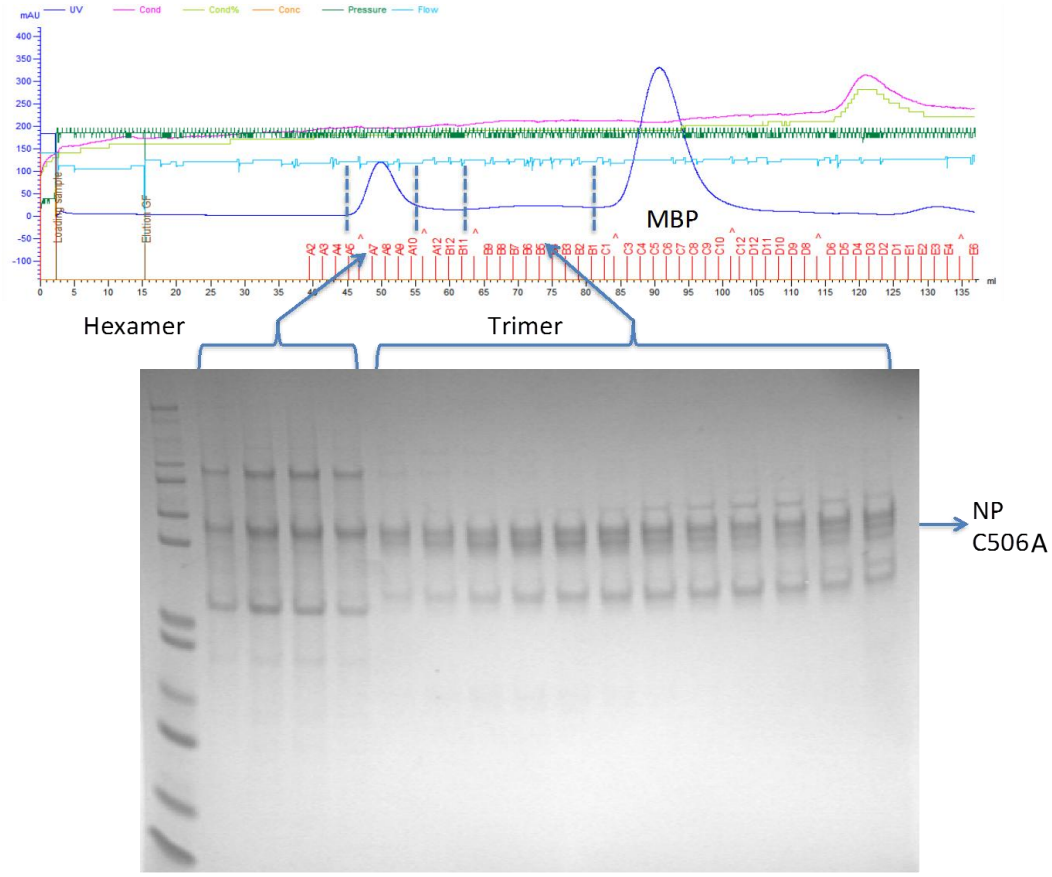


Figure 4.22 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein C506A.

III. The NP H509A mutant protein expression and purification

The NP mutant H509A was expressed and purified using the similar methods as mutant E399A described above. The NP H509A mutant is finally purified by a gel filtration column (Fig. 4.23 top), and the protein purity was checked by a SDS-PAGE gel (Fig. 4.23 bottom). 0.1mg of the purified trimeric NP H509A mutant protein was obtained from 10 liters of culture.

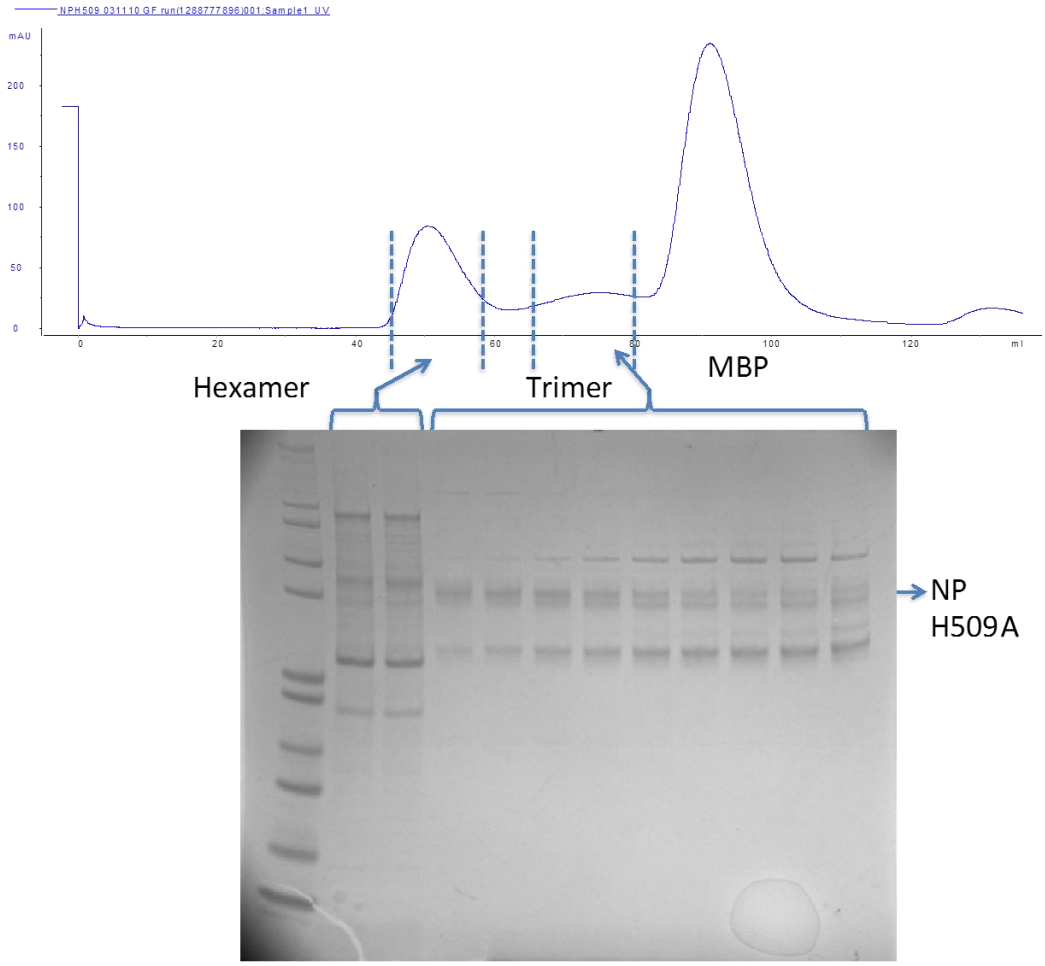


Figure 4.23 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein H509A.

IV. The NP C529A mutant protein expression and purification

The NP mutant C529A was expressed and purified using the similar methods as mutant E399A described above. The C529A mutant is finally purified by a gel filtration column (Fig. 4.24 top), and the purity was checked by a SDS-PAGE gel (Fig. 4.24 bottom). 0.1mg of the purified trimeric NP C529A mutant protein was obtained from 10 liters of culture.

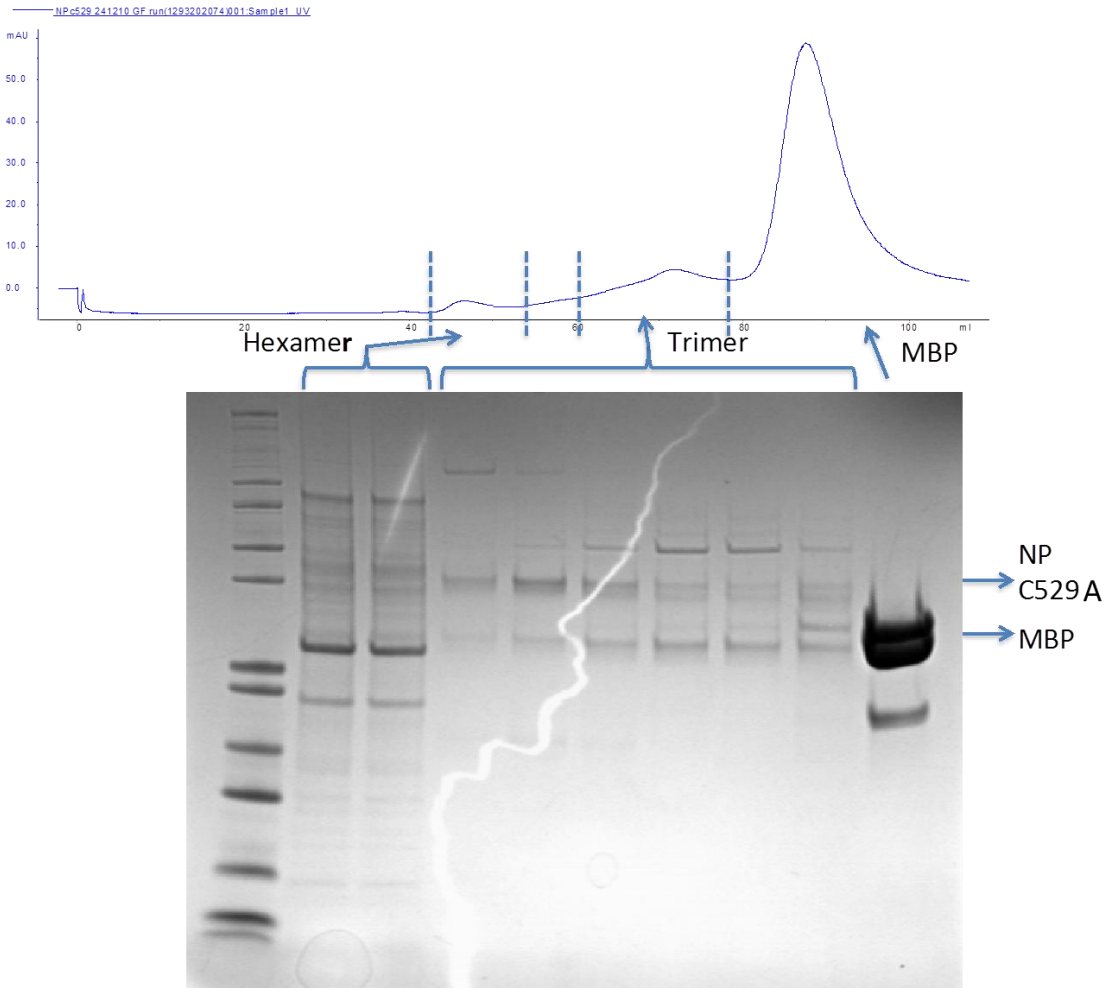


Figure 4.24 The gel filtration chromatograph and the SDS-PAGE gel picture of the purified NP mutant protein C529A.

4.2.6 The in vivo functional assays of LASV NP C-terminal domain

To probe the significance of the function of NP's exoribonuclease in suppression of type I IFN-β production, five putative catalytic residues D389A, E391A, D466A, D533A and H528A were substituted by alanine, along with the WT NP, were sent to our collaborators at Emory University for the luciferase reporter gene based assay and to examine whether NP's exoribonuclease activity is required for suppression of type I IFN (Fig. 4.25) (Qi, 2010).

Firstly, we tested whether the mutants affect the viral RNA transcription using a

minigenome assay. The assay was carried out with WT NP and its mutants transcribing the LASV RNA that encoding a Renilla luciferase reporter gene, the mammalian cell expression vector contain either native or Myc tagged NP gene. The expression level of the proteins in mammalian cell which were shown in figure 4.25a (top) indicates neither the tags nor the mutations altered the overall structure of NP. The similar increased fold of Renilla luciferase activities between WT NP and its mutants also proved that the mutations did not affect NP's function in mediating viral RNA transcription (Fig. 4.25a bottom).

To determine the role of LASV NP in immune evasion, serial experiments were carried out by co-transfecting the 293T cells with the vectors of WT NP and a group of catalytic residues that residues were substituted with alanine, the vectors contain both firefly luciferase reporter gene and the functional promoter sequence IFN-β gene (pIFN β-Luc), and the plasmid of β-gal to normalize the transfection efficiency. The aim is to reveal the inhibition of IFN-β activation by NP or its mutants. The fold induction of luciferase activity over the control shows that LASV WT NP but not its five catalytic mutants exhibit suppressive activity of either Sendai-virus induced or immunostimulatory Pichinde-virion-associated RNAs/ poly(I:C) induced IFN-β activation (Fig. 4.25b,c).

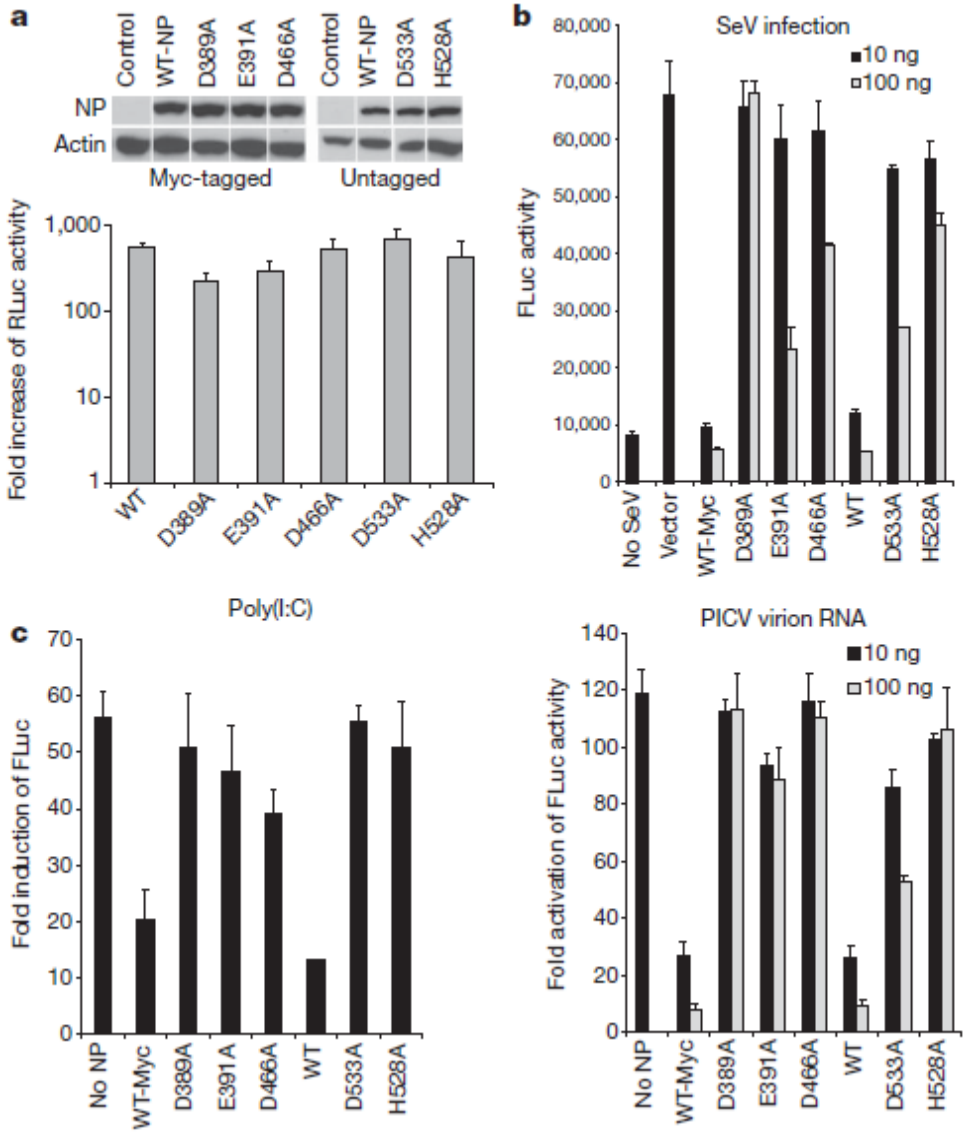


Figure 4.25 The exoribonuclease activity of NP is important for blocking the IFN induction. a, the NP catalytic mutants were expressed at the similar levels to the wild type in mammalian cells and had the similar transcriptional activities in the LASV minigenome assay; b, the NP catalytic mutants were defective in suppressing the Sendai-virus (SeV)-induced IFN induction by a LUC-based IFN-β promoter assay; c, the NP catalytic mutants were defective in suppressing the IFN production induced by the immunostimulatory RNAs poly(I:C) and Pichinde-virion-associated RNAs.

4.3 The cap binding cavity on N-terminal domain of LASV NP

4.3.1 The complex structure of the UTP/dTTP and LASV NP

The N-terminal domain has a novel fold, showing a deep cavity (Fig. 4.26). To

investigate the function of the cavity, both the dTTP and UTP compounds were incubated with WT NP at 20mM final concentration for 30 mins on ice, and the NP in complexes with dTTP and UTP were obtained. The structures were determined as described in the last chapter. The superimposition of the complex shows the triphosphate of UTP was bound in the same manner as that of dTTP at the deep cavity of LASV NP. The original F_o-F_c map contoured at 2.5 σ for both dTTP and UTP.

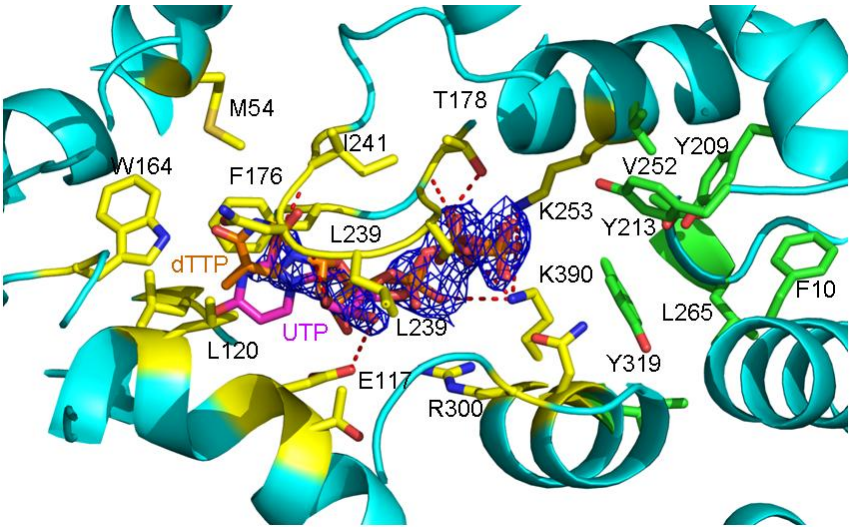


Figure 4.26 Superimposition of the UTP-NP and dTTP-NP complexes. The original F_o-F_c map of the triphosphate of UTP is shown in blue, with the carbon atoms of UTP shown in magenta and those of dTTP in orange. The deep cavity residues are colored in yellow, the entrance cavity residues are colored in green.

The triphosphate group of dTTP/UTP was anchored by salt bond formed with side chains of the conserved residues K253A, R300A, R323A and K309A. The hydrophobic pocket is composed of residues F176A, W164A, L172A, M54A, I241A, L120A and L239A. The entrance region is formed by hydrophobic residues Y319A, Y209A, Y213A, L265A and the acidic residue E266A. The acidic amino acid is known for base binding, the entrance cavity probably anchored one nucleotide from

the host mRNA. Unlike the cap snatching mechanism in influenza A virus, when PB2 hold the cap structure of the cellular mRNA, PA acts as the endonuclease to cleave the mRNA and leave 10-12 nucleotide after the cap, exposed outside the protein, the cap structure m^7G along with at least one more nucleotide are embedded within the deep cavity in NP. Although more evidence should be provided, the hydrophobic thymidine binding pocket is assumed to be the cap binding site for the cap snatching mechanism in Arenaviridae family.

4.3.2 Expression and purification of the LASV NP mutants

The NP mutants M54A, E117A, L120A, W164A, F176A, Y209A, K253A, E266A, R300A, K309A, Y319A, R323A that located around the deep cavity on N-terminal domain were generated as previously described to send for the surface plasmon resonance (SPR) experiment to identify the binding affinity between NP and various cap analogues (the experiment data does not shown in this thesis). These mutants were purified and prepared for further experiments. The gel filtration results are shown in figure 4.27- 4.38.

I. The NP M54A mutant expression and purification

The NP mutant M54A was expressed and purified using similar methods as the native NP. Briefly, the expression plasmid that contains the target mutant gene NP M54A was incubate in 10 liters of LB medium in present of 50 μg/ml Ampcilin and 34 μg/ml chloramphenicol (Duchefa Biochemic) at 37°C, and the mutant protein was induced

by 0.03mM IPTG when the OD_{600} of the cell culture reached to 0.6. The NP protein was induced overnight at 20°C. The cells were harvested, and the cells were broken by sonication. The fusion protein was purified by amylose resin. After TEV cleavage and removing MBP, the NP M54A mutant is finally purified by a gel filtration column (Fig. 4.27 top), which showed that the first peak is NP and the second peak is MBP. To test the protein purity, each fraction was loaded on a SDS-PAGE (Fig. 4.27 bottom), and the 12mg of the purified trimeric NP M54A mutant protein was obtained from 10 liters of culture.

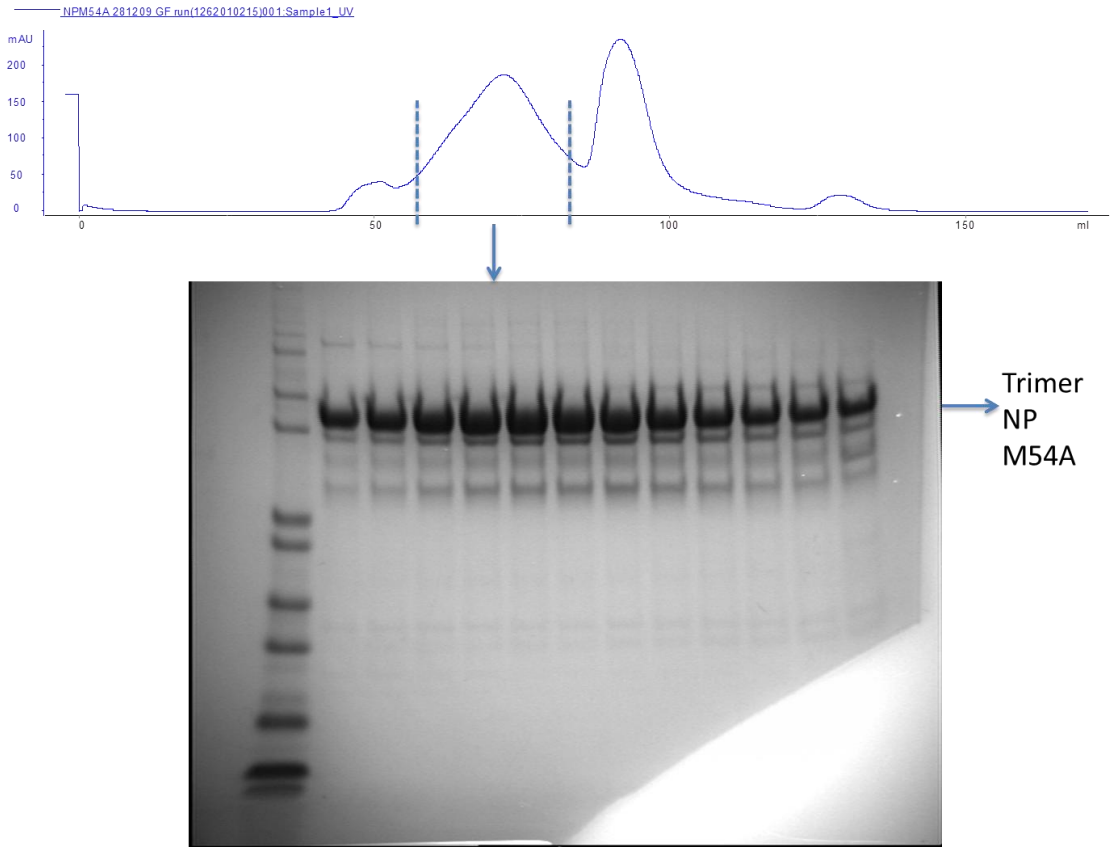


Figure 4.27 The gel filtration result and SDS-PAGE gel of the NP M54A mutant that is located around the deep cavity in N-terminal domain.

II. The NP E117A mutant expression and purification

The NP mutant E117A was expressed and purified using the similar methods as NP

mutant M54A described above. The E117A mutant is finally purified by a gel filtration column (Fig. 4.28 top), and the protein purity was determined by a SDS-PAGE gel (Fig. 4.28 bottom). 10mg of the purified trimeric NP E117A mutant protein was obtained from 10 liters of culture.

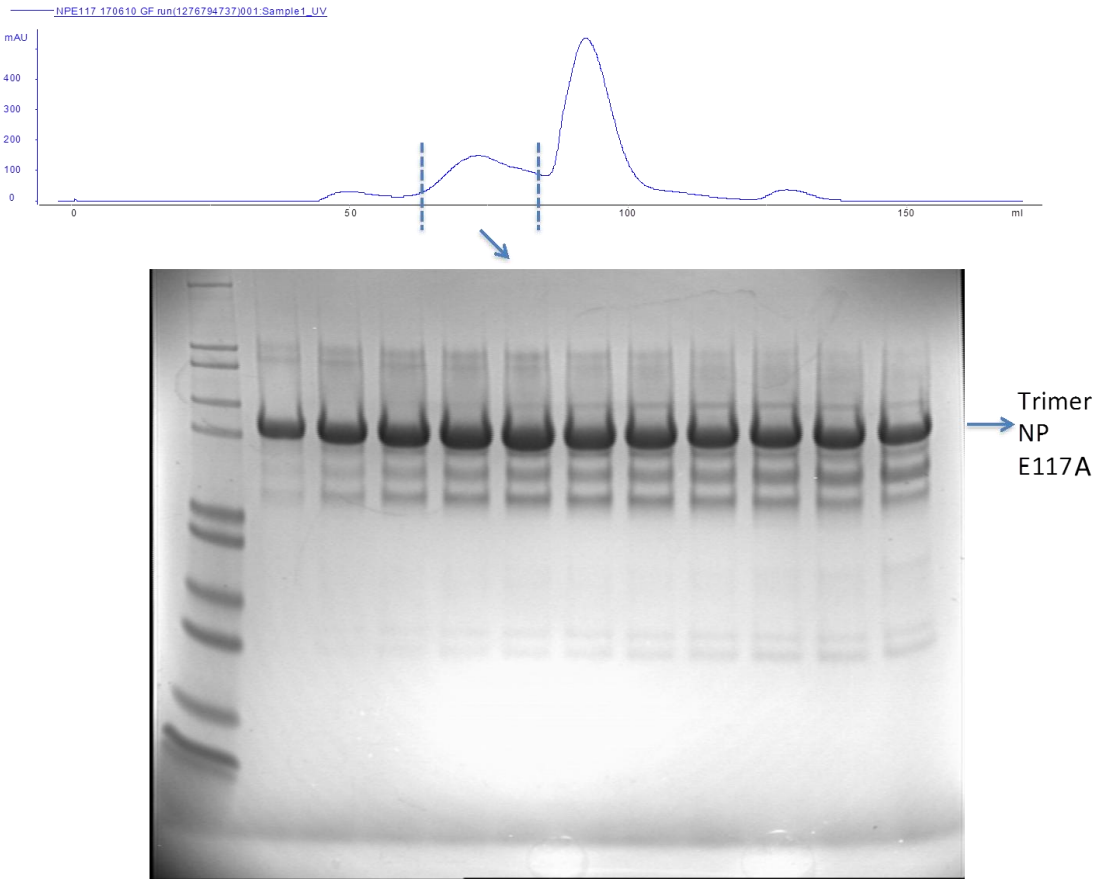


Figure 4.28 The gel filtration chromatograph and the SDS-PAGE gel of the NP E117A mutant.

III. The NP L120A mutant expression and purification

The NP mutant L120A was expressed and purified using similar methods as mutant NP M54A described above. The NP L120A mutant is finally purified by a gel filtration column (Fig. 4.29 top), and the purity was tested by a SDS-PAGE gel (Fig. 4.29 bottom). 12mg of the purified trimeric NP L120A mutant protein was obtained

from 10 liters of culture.

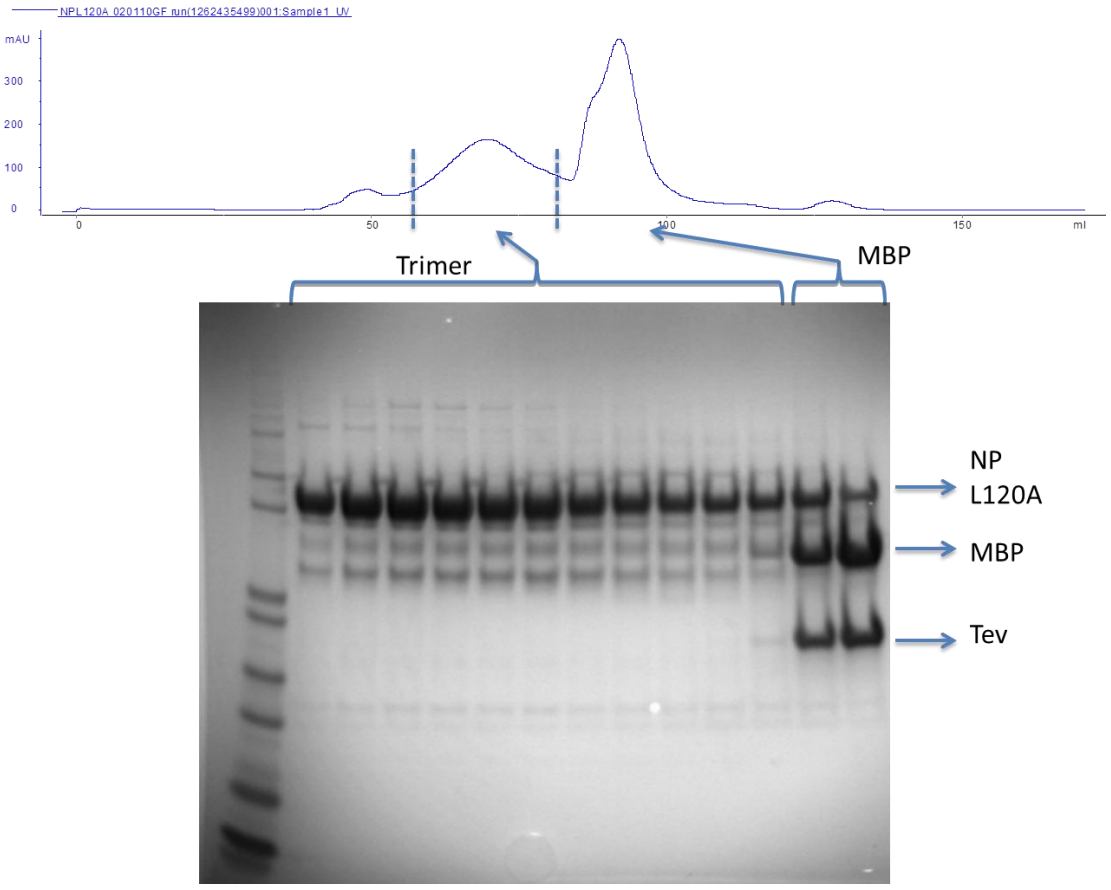


Figure 4.29 The gel filtration chromatograph and the SDS-PAGE gel of the NP L120A mutant.

IV. The NP W164A mutant expression and purification

The NP mutant W164A was expressed and purified using the similar methods as the NP mutant M54A described above. The W164A mutant is finally purified by a gel filtration column (Fig. 4.30 top), and the purity was tested by a SDS-PAGE gel (Fig. 4.30 bottom). 16mg of the purified trimeric NP W164A mutant protein was obtained from 10 liters of the cell culture.

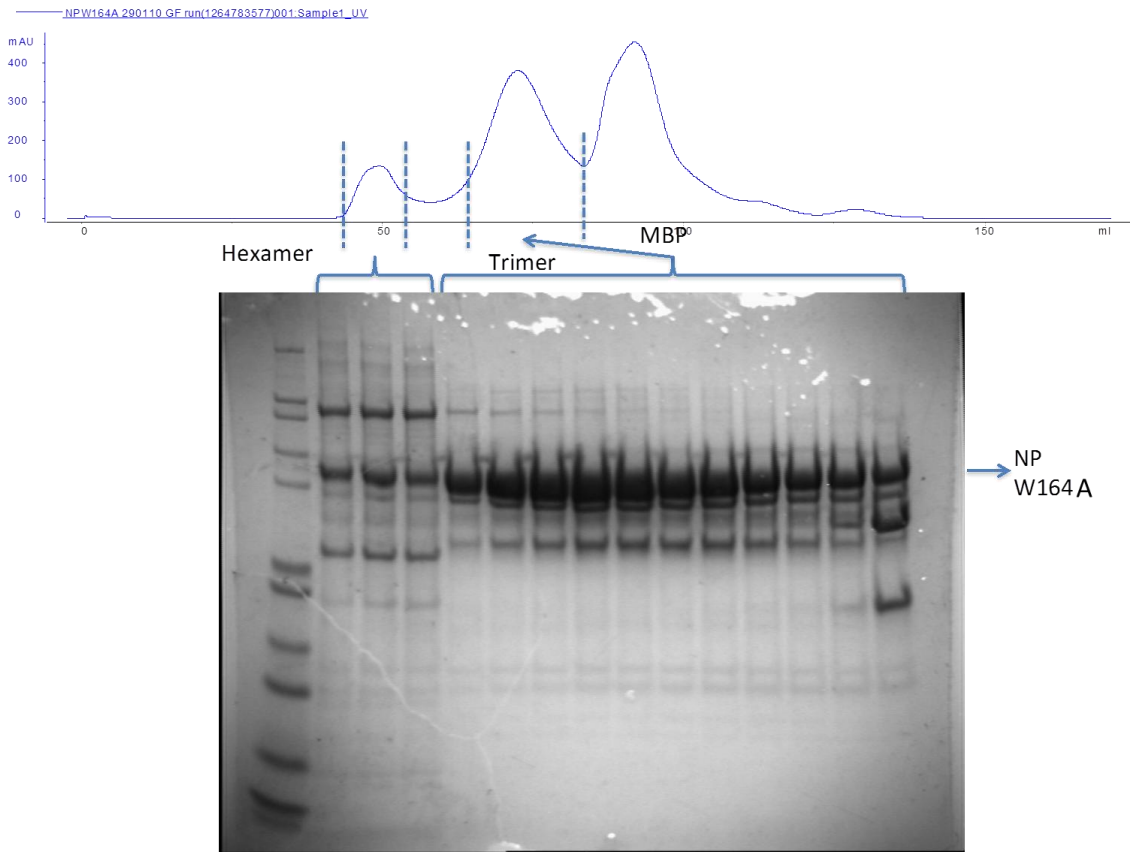


Figure 4.30 The gel filtration chromatograph and the SDS-PAGE gel of the NP W164A mutant.

V. The NP F176A mutant expression and purification

The NP mutant F176A was expressed and purified using similar methods as the NP mutant M54A described above. The NP F176A mutant is finally purified by a gel filtration column (Fig. 4.31 top), and the purity was tested by a SDS-PAGE gel (Fig. 4.31 bottom). 8mg of the purified trimeric NP F176A mutant protein was obtained from 10 liters of culture.

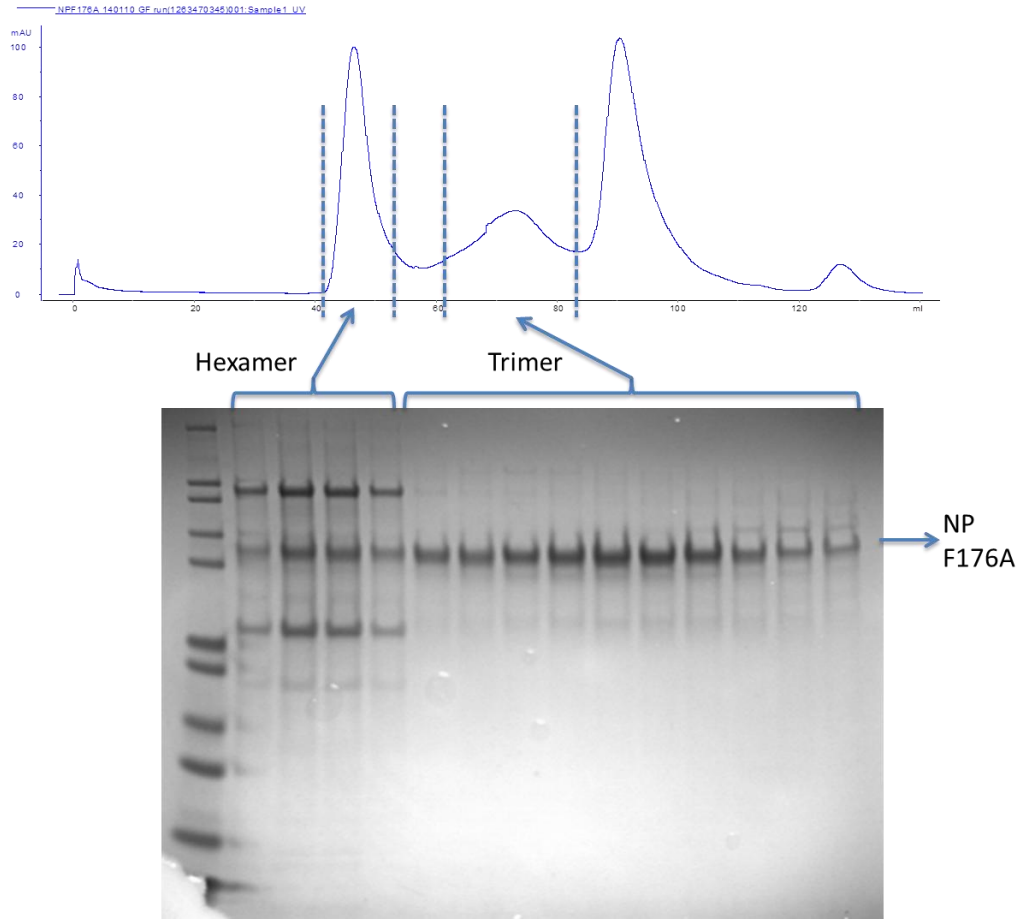


Figure 4.31 The gel filtration chromatograph and the SDS-PAGE gel of the NP F176A mutant.

VI. The NP Y209A mutant expression and purification

The NP mutant Y209A was expressed and purified using the similar methods as the NP mutant M54A described above. The Y209A mutant is finally purified by gel filtration chromatography (Fig. 4.32 top), and the protein purity was checked by a SDS-PAGE gel (Fig. 4.32 bottom). 12mg of the purified trimeric NP Y209A mutant protein was obtained from 10 liters of culture.

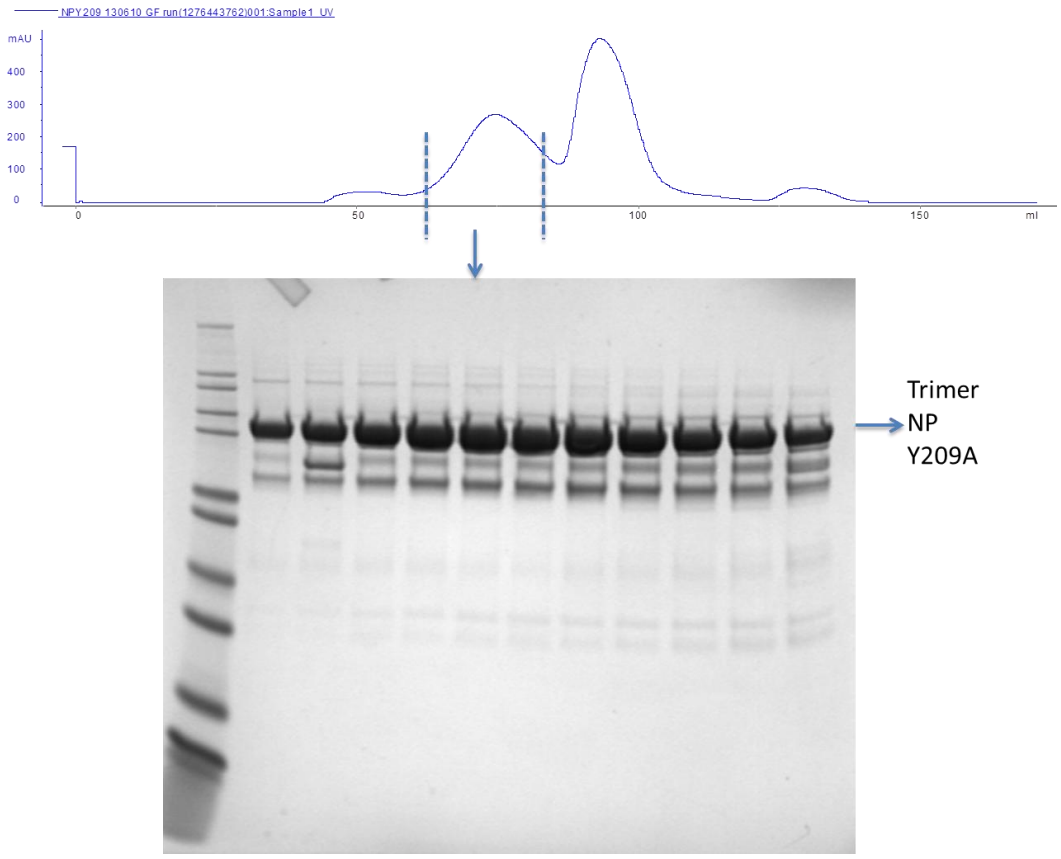


Figure 4.32 The gel filtration chromatograph and the SDS-PAGE gel of the NP Y209A mutant.

VII. The NP K253A mutant expression and purification

The NP mutant K253A was expressed and purified using the similar methods as the NP mutant M54A described above. The NP K253A mutant is finally purified by a gel filtration column (Fig. 4.33 top), and the protein purity was checked by a SDS-PAGE gel (Fig. 4.33 bottom). 17mg of the purified trimeric NP K253A mutant protein was obtained from 10 liters of culture.

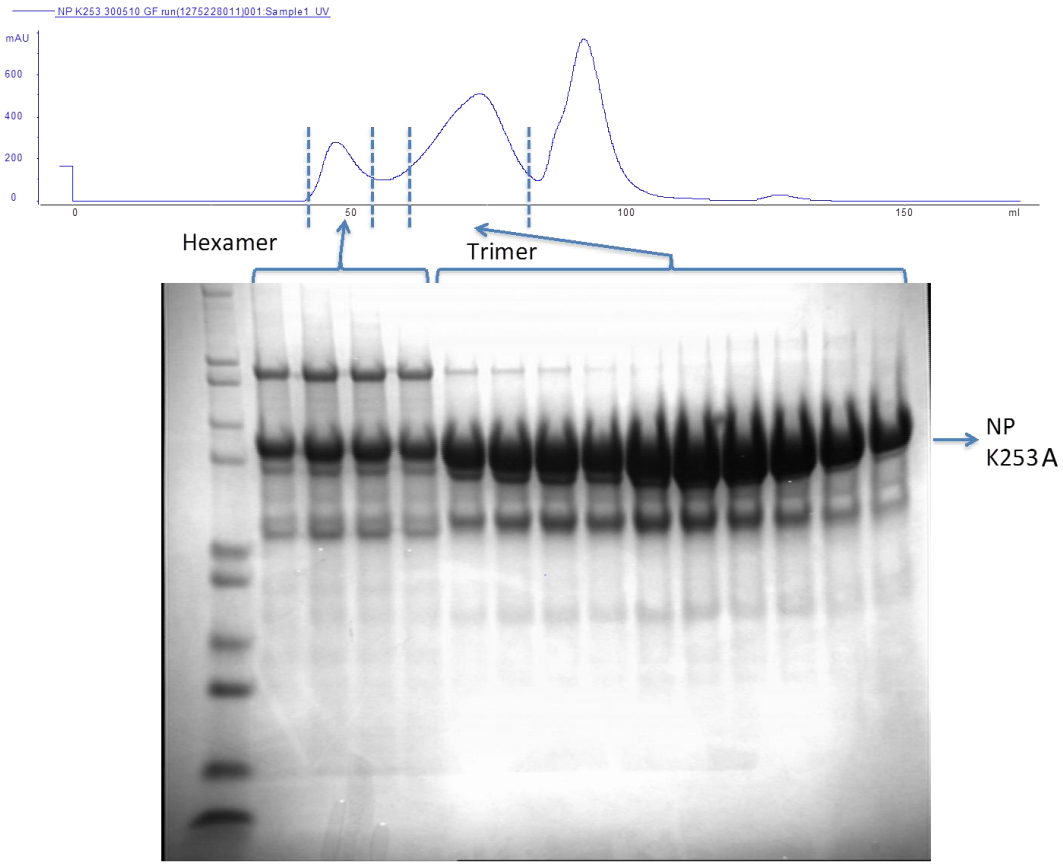


Figure 4.33 The gel filtration chromatograph and the SDS-PAGE gel of the NP K253A mutant.

VIII. The NP E266A mutant expression and purification

The NP mutant E266A was expressed and purified using the similar methods as the NP mutant M54A described above. The E266A mutant is finally purified by a gel filtration column (Fig. 4.34 top), and the protein purity was determined by a SDS-PAGE gel (Fig. 4.34 bottom). 7mg of the purified trimeric NP E266A mutant protein was obtained from 10 liters of culture.

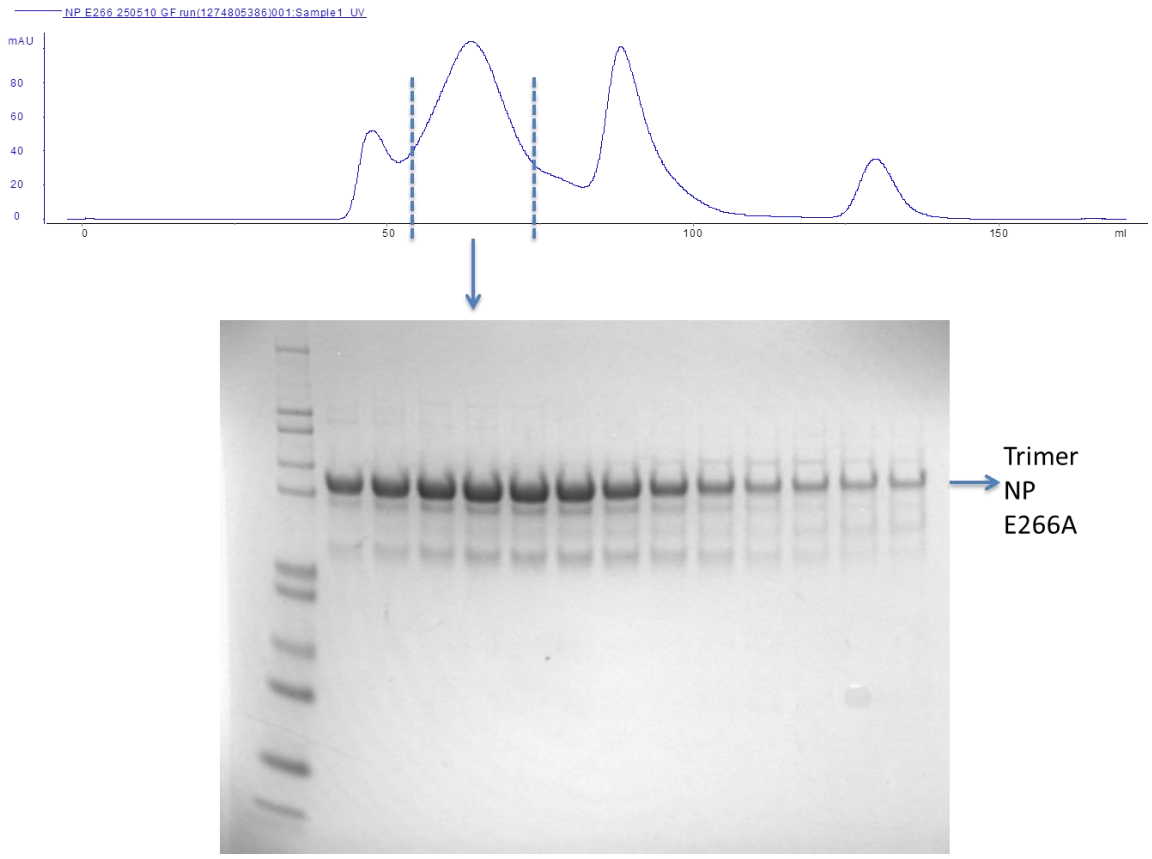


Figure 4.34 The gel filtration chromatograph and the SDS-PAGE gel of the NP E266A mutant.

IX. The NP R300A mutant expression and purification

The NP mutant R300A was expressed and purified using similar methods as the NP mutant M54A described above. The NP R300A mutant is finally purified by a gel filtration column (Fig. 4.35 top), and the protein purity was checked by a SDS-PAGE gel (Fig. 4.35 bottom). 12mg of the purified trimeric NP R300A mutant protein was obtained from 10 liters of culture.

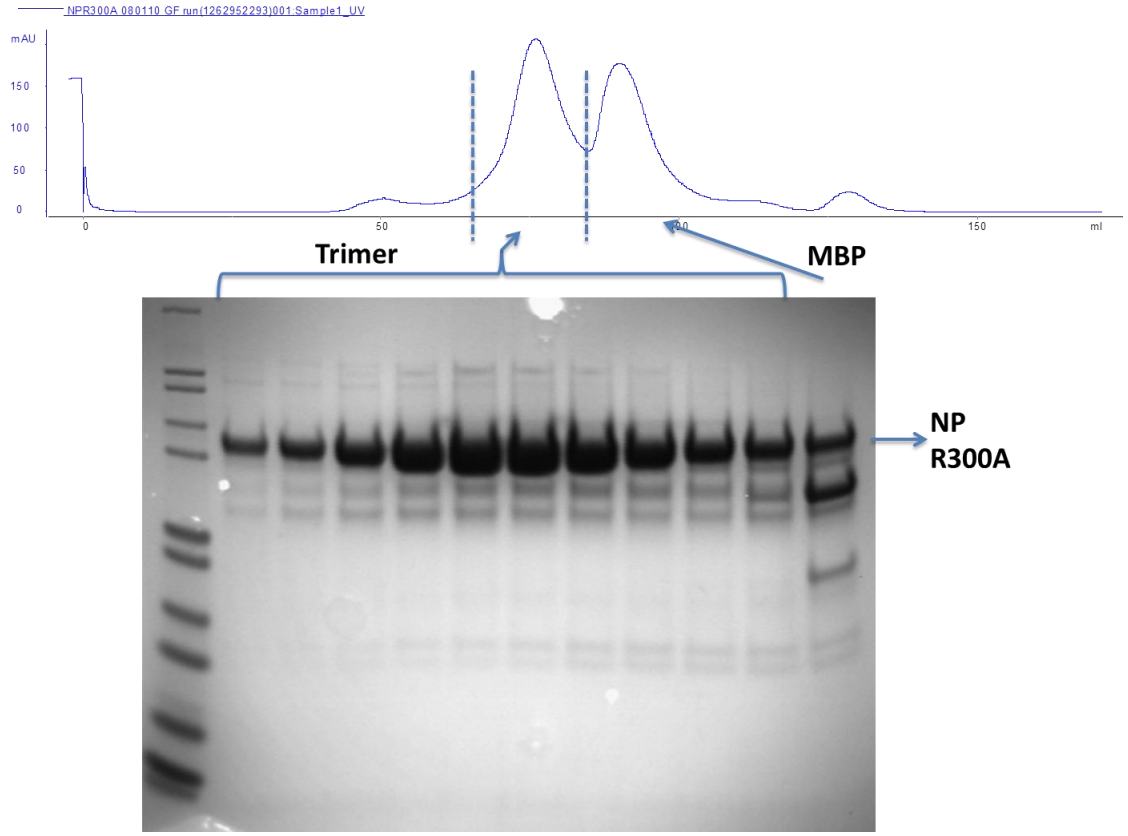


Figure 4.35 The gel filtration chromatograph and the SDS-PAGE gel of the NP R300A mutant.

X. The NP K309A mutant expression and purification

The NP mutant K309A was expressed and purified using the similar methods as the NP mutant M54A described above. The NP K309A mutant is finally purified by a gel filtration column (Fig. 4.36 top), and the protein purity was tested by a SDS-PAGE gel (Fig. 4.36 bottom). 17mg of the purified trimeric NP K309A mutant protein was obtained from 10 liters of culture.

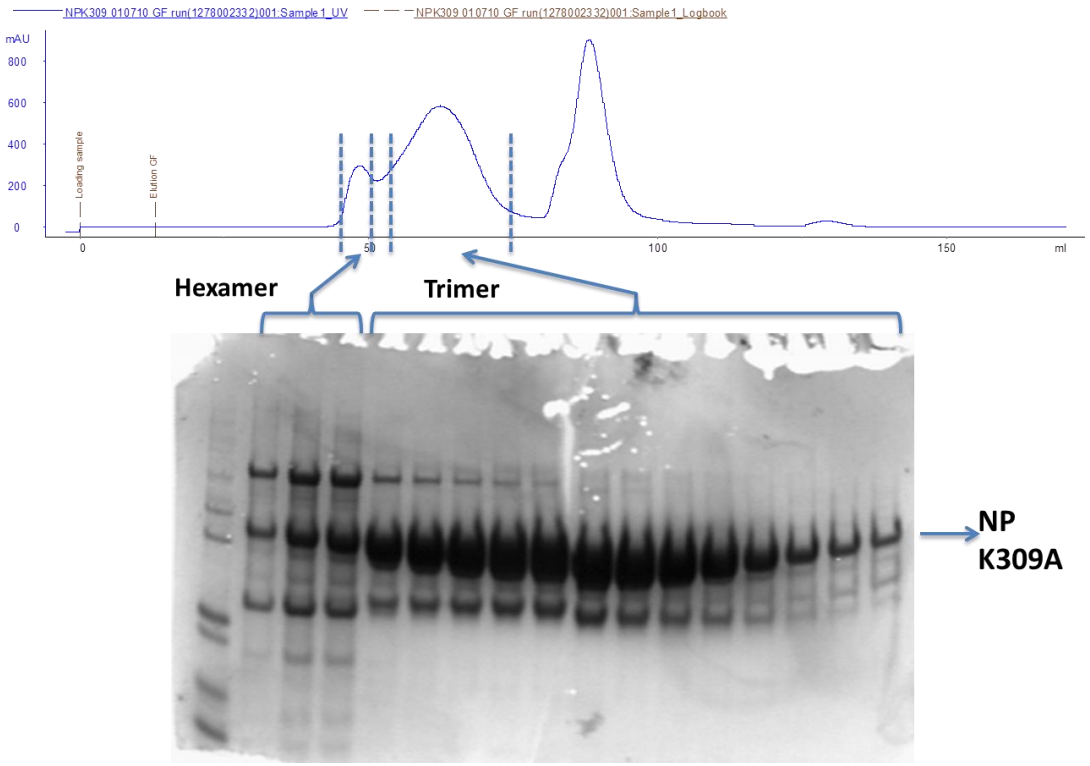


Figure 4.36 The gel filtration chromatograph and the SDS-PAGE gel of the NP K309A mutant.

XI. The NP Y319A mutant expression and purification

The NP mutant Y319A was expressed and purified using similar methods as the NP mutant M54A described above. The NP Y319A mutant is finally purified by a gel filtration column (Fig. 4.37 top), and the protein purity was tested by a SDS-PAGE gel (Fig. 4.37 bottom). 15mg of the purified trimeric NP Y319A mutant protein was obtained from 10 liters of culture.

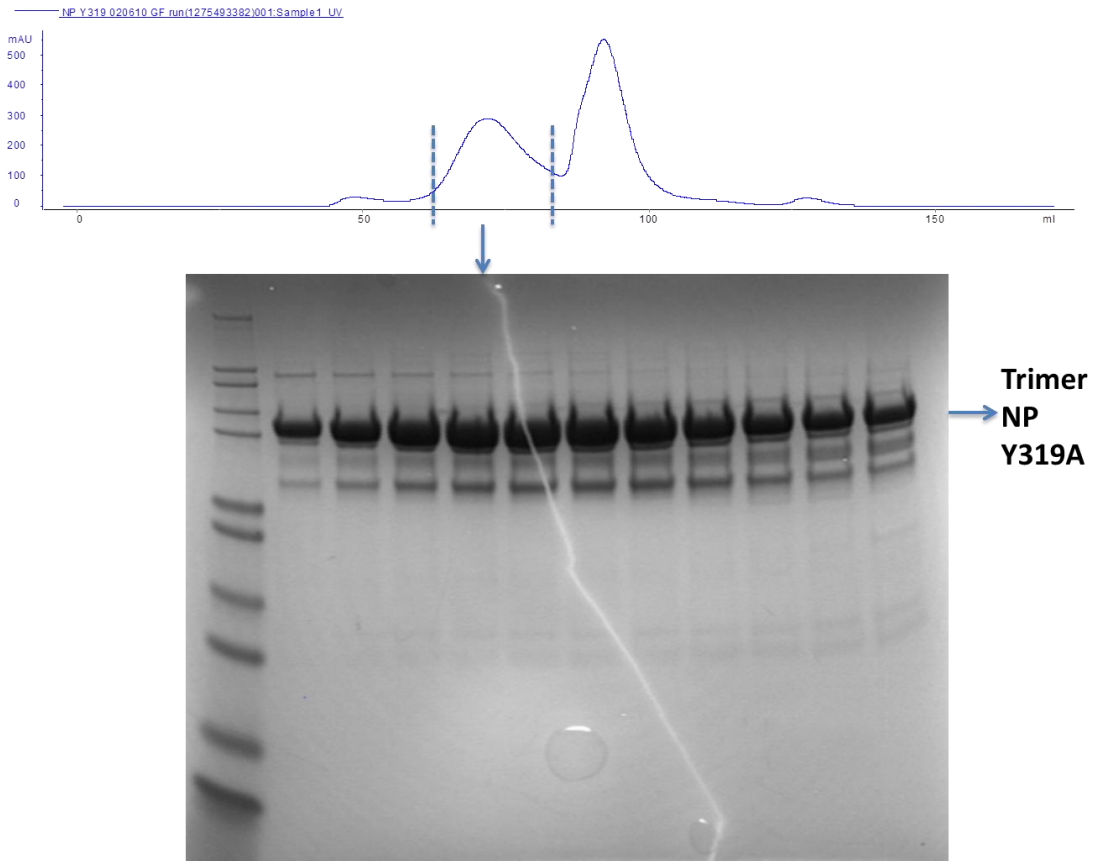


Figure 4.37 The gel filtration chromatograph and the SDS-PAGE gel of the NP Y319A mutant.

XII. The NP R323A mutant expression and purification

The NP mutant R323A was expressed and purified using the similar methods as the NP mutant M54A described above. The NP R323A mutant is finally purified by a gel filtration column (Fig. 4.38 top), and the protein purity was checked by a SDS-PAGE gel (Fig. 4.38 bottom). 17mg of the purified trimeric NP R323A mutant protein was obtained from 10 liters of culture.

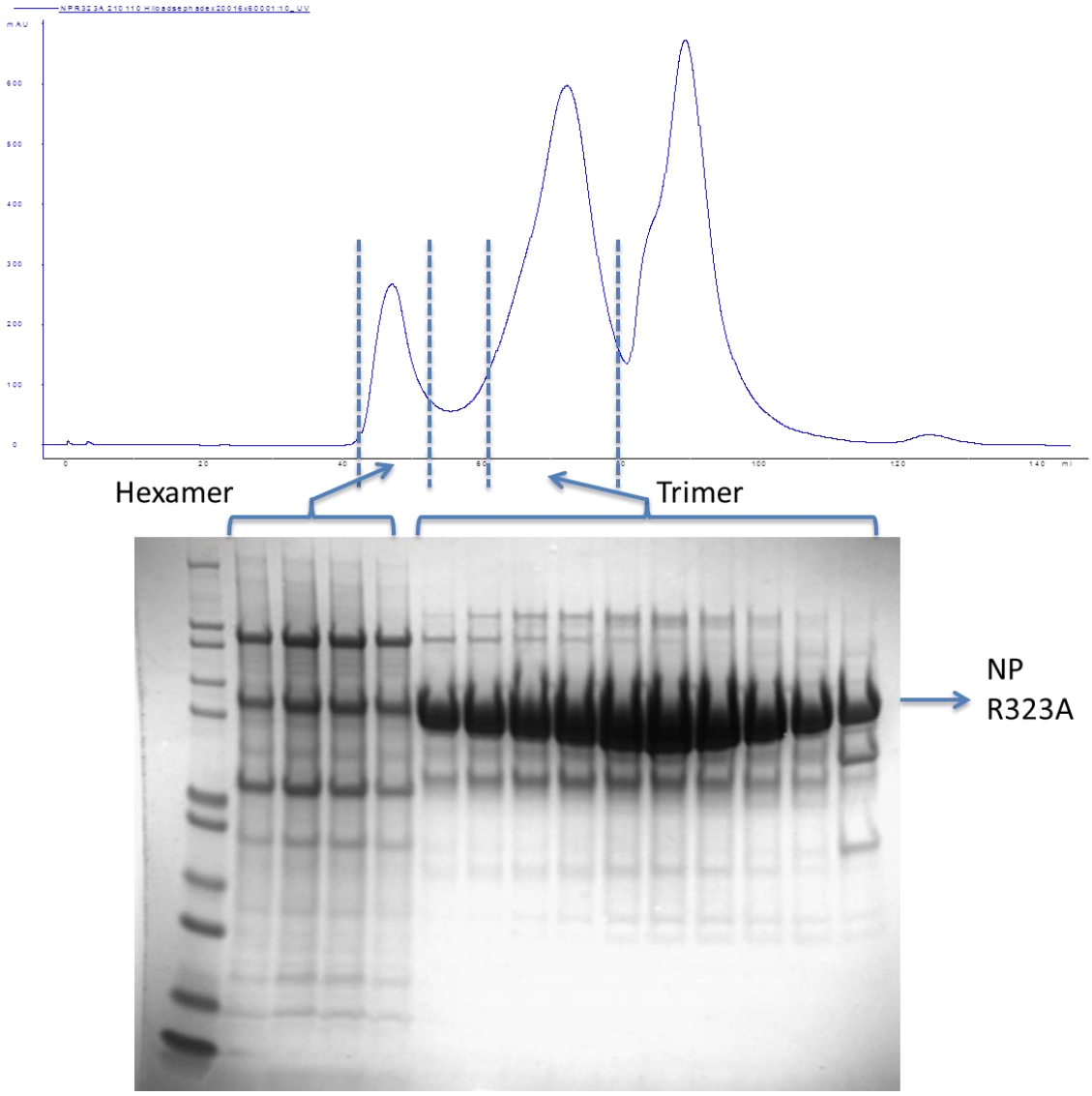


Figure 4.38 The gel filtration chomatograph and the SDS-PAGE gel of the NP R323A mutant.

4.3.3 The in vivo functional assays of LASV NP N-terminal domain

To characterize N-terminal domain's function in viral transcription, a group of mutants that resides for the cap binding in N-terminal domain were generated as previous described. The NP and the mutant genes were cloned into pCAGGS vector and were expressed at a comparable level in mammalian cells (Fig. 4.39a). The LASV minigenome assays were performed by transfecting the 293T cells with the in vitro transcribed LASV minigenome RNA containing T7-promoter directed LASV

s-segment sequence analogue and the Renilla luciferase (RLuc) reporter gene in place of NP coding sequence, together with LASV L expression plasmid, LASV WT or mutated NP expression plasmid and a β-gal expression plasmid which was used for normalizing the transfection efficiency. The increase fold of RLuc activity shows the mRNA transcription level of the MG RNA. Comparing to WT NP, the mutants K253A and E266A totally lost their function in transcribing MG RNA, whereas the mutants W164A, F176A, K309A, Y319A and R323A significantly decrease the activity. The mutants W12A, Y209A, Y213A and R300A have minor defect of this activity (Fig. 4.39b). Taken together, the mutants of the cap binding residues around the deep hydrophobic thymidine binding pocket defect viral RNA transcription activity.

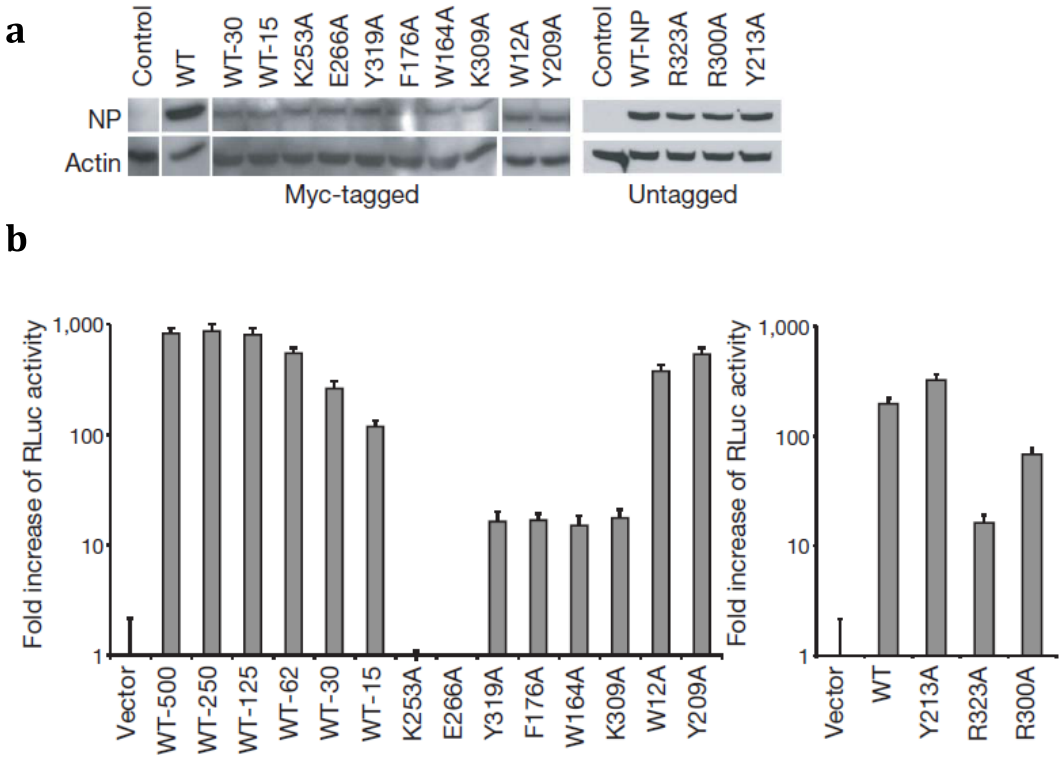


Figure 4.39 The role of the residues on the N-terminal deep cavity. a, the NP mutants were expressed at the similar levels as the wild type at 15-30 ng plasmid (WT-15, WT-30) in the transfected mammalian cells; b, luciferase based LASV minigenome transcription assay for WT NP and its mutants.

A luciferase based promoter assay for either WT NP or its N-terminal mutants to inhibit the Sendai virus induced IFN-β activation was also performed. The result shows unlike the mutants of the 3'-5' exoribonuclease catalytic residues of the C terminal domain, these mutants suppress the IFN- β activation as well as WT NP, which indicates these residues in the deep cavity of the N-terminal domain do not involve in the immune suppression (Fig. 4.40).

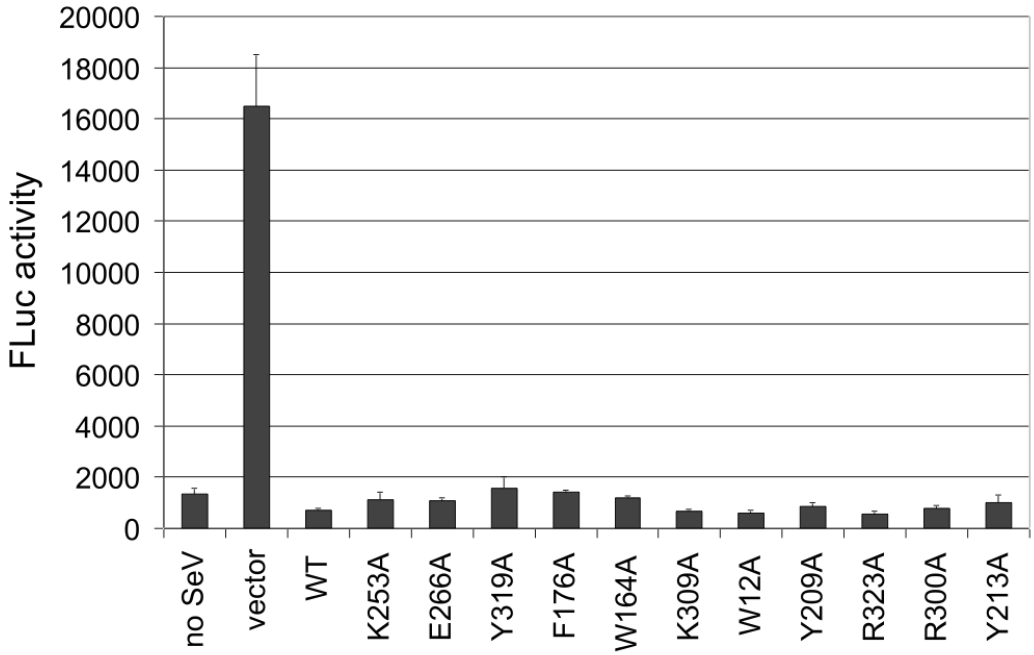


Figure 4.40 LASV NP mutants of the deep cavity on N-terminal domain are functional in suppressing Sendai virus-induced IFN-beta activation. This result suggests that these mutations do not alter the NP structure or affect its IFN suppression activity.

Chapter 5: Interaction between LASV NP and Z protein

LASV NP protein was reported to interact with Z protein in vivo, and this interaction may play an essential role in viral RNA replication and transcription (Eichler, 2004; Ortiz-Riaño, 2011). To understand the interaction mechanism, we tried to study the interaction in vitro.

5.1 Preparation of Z protein expression construct

The entire Z gene of Lassa fever virus was from a Z containing pCAGGS plasmid and subcloned into a pLOU3 vector, which is a derivative plasmid from pMAL-C2X, containing a 6His-tag at the N-terminus of MBP and a Tev proteinase cleavage site between the MBP and target protein. The process is the same as the NP gene subcloning except both the Z gene fragment and pLOU3 plasmid were treated using NcoI and XbaI. The insertion was confirmed by digesting the Z-pLOU3 plasmid with NcoI and XbaI and agarose gel electrophoresis.

5.2 LASV Z Protein expression and purification

The recombinant plasmid was transformed into *E.coli* expression strain Rosseta cells. Single colony was picked for preparation of 500 ml overnight culture. The overnight culture was inoculated into 10 liters of the LB medium containing 50 μg/ml Ampcilin and 34 μg/ml chloramphenicol and the cells grew at 37°C and 200 r.p.m.. The protein was induced by adding isopropyl-beta-d-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM when the OD_{600} of the cell culture reached at 0.6. Then the cell culture was continued to culture at 28°C overnight.

The cells were harvested by centrifugation at 8,000 *g* for 10 min at 4 °C. The cell pellet collected from 10L culture was resuspended in 100 ml of 10 mM imidazole binding buffer (10%Glycerol, 10mM imidazole, 20mM phosphate pH7.0, 0.3M NaCl) with 2 protease inhibitor tablets (Roche), 1 uM DNase (Sigma) and 1 mM Phenylmethylsulfonyl fluoride (PMSF) (Sigma). The cells were disrupted by a cell disruptor (Constant system Ltd) and centrifuged down at 20,000 rpm for 30 min to remove the cell debris. The supernatant was loaded on a Nickel column that was equilibrated using the binding buffer. The column then washed with a washing buffer (30mM imidazole, 20mM phosphate pH7.0, 0.3M NaCl, 10% glycerol). The fusion protein then was eluted with an elution buffer (10%Glycerol, 20mM phosphate pH7.0, 0.3M NaCl, 0.5M imidazole). The protein buffer was changed into the Tev cleavage buffer (1/50 PBS, 50mM Tris-HCl, 0.5mM EDTA, 1mM DTT, 0.3M NaCl, 10% Glycerol) using a desalting column (Hiprep 26/10, GE helathcare). Fractions containing HisMBP-Z fusion protein was pooled, MBP tag was cleaved off from the Z protein using 1ml His-tagged TEV protease at room temperature overnight. The cleaved sample was desalted again to remove the EDTA, and then loaded to the nickel resin again to remove the MBP. The flow through sample (Z protein) was concentrated to 5 ml and purified by a gel filtration column (Sephacryl S-200, GE healthcare) that was equilibrated by the gel filtration buffer (20mM Tris pH 7.5, 0.3M NaCl, 10% Glycerol). The protein's identity and integrity were checked by mass spectroscopy and SDS-PAGE. The SDS-PAGEs were carried out using pre-cast NuPAGE 4-12% Bis-Tris gels and Powerease 500 according to the manufacturer's

Manuel, with the protein molecular standards of Mark12 (Invitrogen). The crystallization trials have been set up.

The expression of Z protein was very successful according to the results shown in the SDS gel below (Fig. 5.1). The Z-MBP fusion protein was cleaved by TEV proteinase; the upper band is MBP protein and the below one is Z protein.

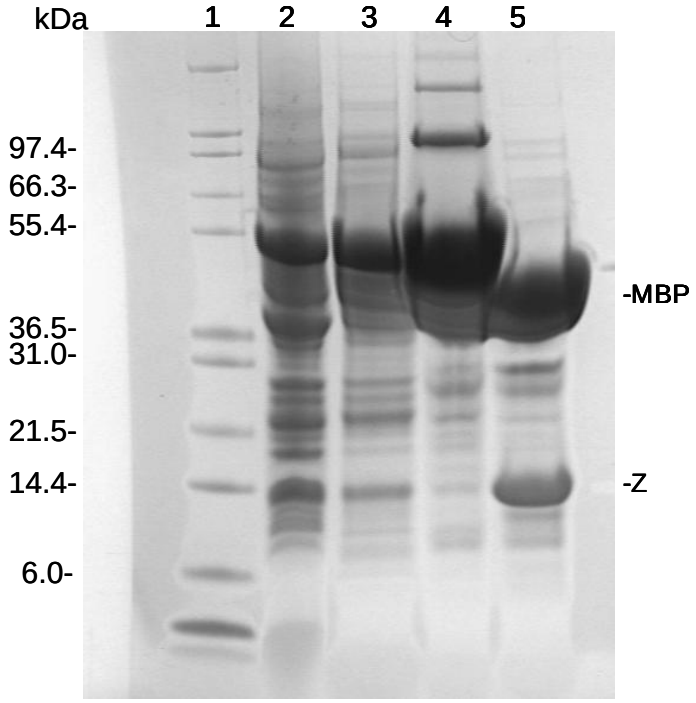


Figure 5.1 SDS-PAGE gel of Z protein. Lane 1: marker 12; lane 2: flowthrough; lane 3: 30 mM imidazole wash; lane 4: elution; lane 5: Z-MBP after TEV cleavage.

During gel filtration, Z protein was separated into two, forms which are monomer and oligomer. The oligomeric Z protein is brown color while the monomeric Z protein is colorless. The oligomeric Z protein was concentrated to 1ml with the concentration of 6.5 mg/ml, which means an approximate 6.5 mg of oligomeric Z protein was obtained from 10-litre culture. The monomeric Z protein was concentrated to 400ul with the

concentration of 4.5 mg/ml, which means around 1.8 mg of monomeric Z protein was obtained from10-litre culture (Fig. 5.2 & 5.3).

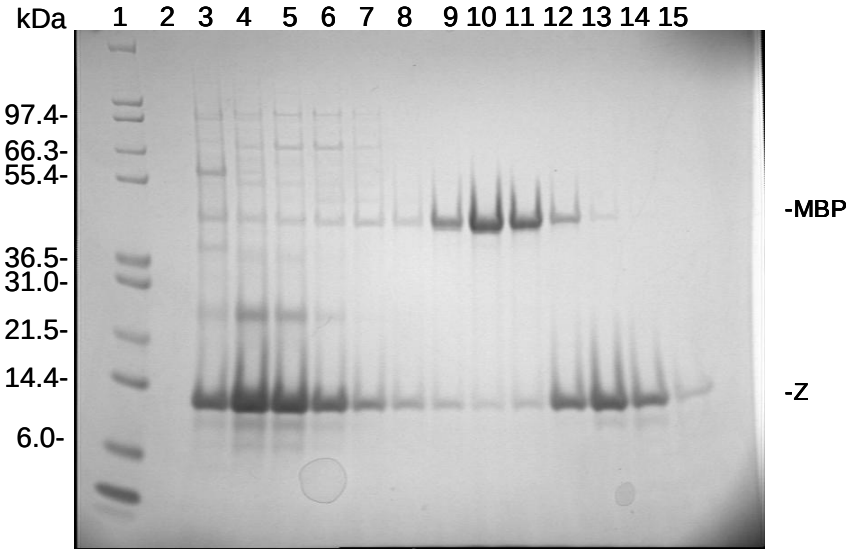


Figure 5.2 The SDS-PAGE gel of gel filtration. Line 1: marker 12, line 2-8: oligomeric Z protein- the first peak shown in figure 17, line 9-11: MBP protein-the second peak in figure 17, line 12-15: monomeric Z protein- the third peak.

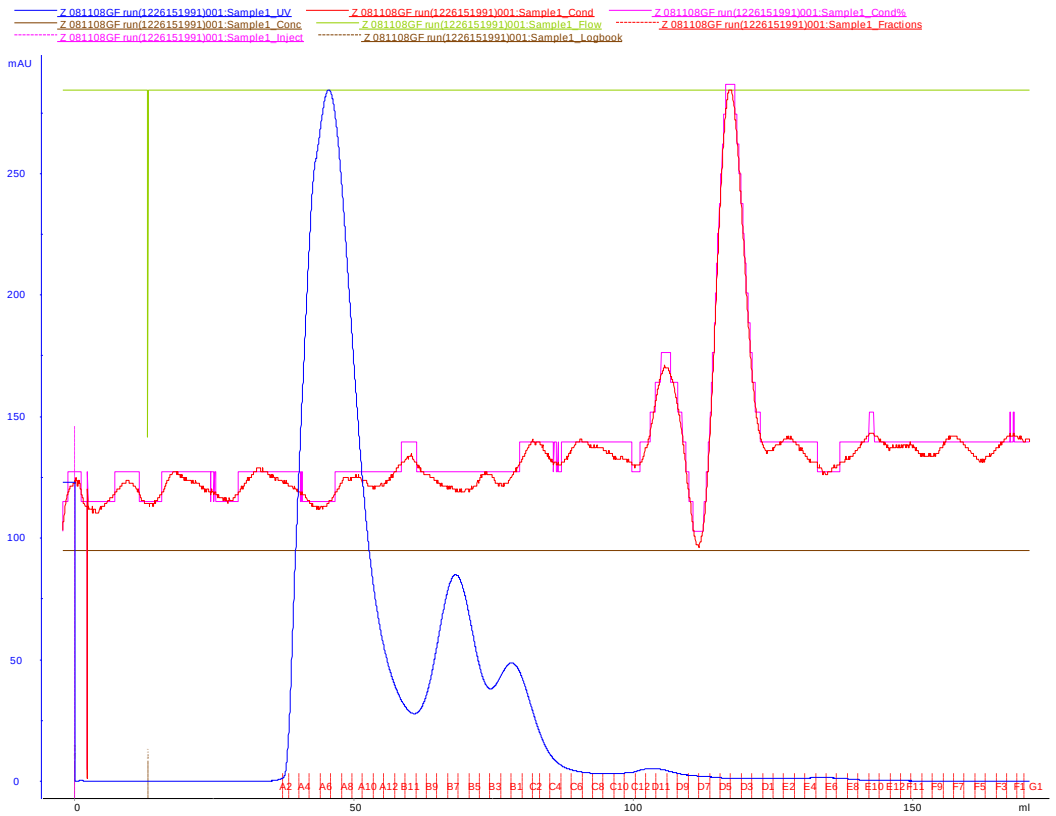


Figure 5.3 The gel filtration graph of the Z protein. The first peak is oligomeric Z protein and the third peak is monomeric Z protein. The second peak is MBP.

After gel filtration, the monomeric Z protein was quite pure, but the oligomer Z protein needs further purification. So the oligomeric Z protein was then go through the nickel column again. Some of the contamination proteins were removed, which directly go through the nickel column. The oligomeric Z protein was bound on the nickel column and was eluted with 10 mM imidazole buffer containing 20 mM tris pH7.5, 0.3 M NaCl and 10% glycerol. As shown in figure 5.4, lane 5, there is still an extra band. They may be some contaminant proteins which are bound inside the Z oligomers.

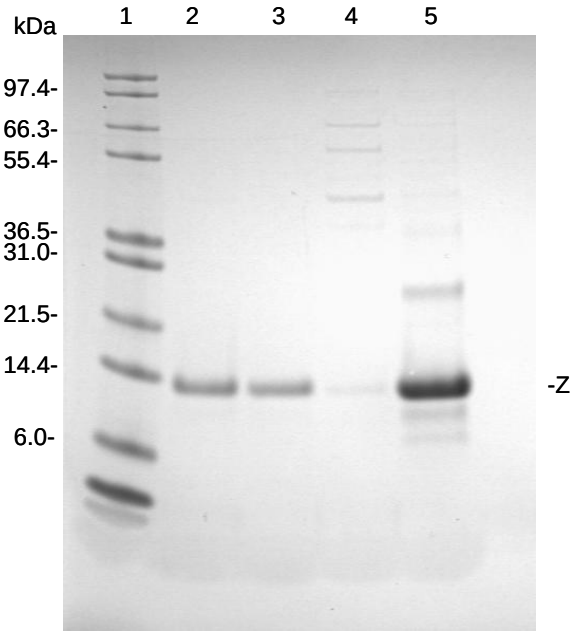


Figure 5.4 SDS-PAGE gel of purified Z protein. Lane 1: marker 12, Lane 2-3: monomeric Z protein, lane 4: flowthrough after nickel column, lane 5: elute with 10 imidazole buffer.

5.3 Characterization of LASV Z protein

To determine the molecular weight of different forms of Z protein, a gel filtration chromatography was performed using a Superdex 200 10/300 GL column (GE healthcare) and a running buffer containing 20mM Tris pH 7.5, 0.3M NaCl and 10% glycerol. The molecular weight of the monomeric and oligomeric Z was determined based on the protein elution volumes against the molecular standards. The monomeric Z appeared six peaks during the gel filtration (Fig. 5.5). indicating that monomeric Z protein can form different oligomers.

As described in chapter for LASV NP protein, we use the same equation y= -0.5764x + 19.336 to calculate the size of oligomer Z protein. According to the elution volumes and the molecular weight of each protein, monomeric Z protein was obtained

from figure 5.5. The monomeric Z protein formed different peaks in the gel filtration, indicating that the monomer can form oligomers. In the monomeric peak positions, there are three peaks, which may caused by the monomers with different sizes of RNA bound.

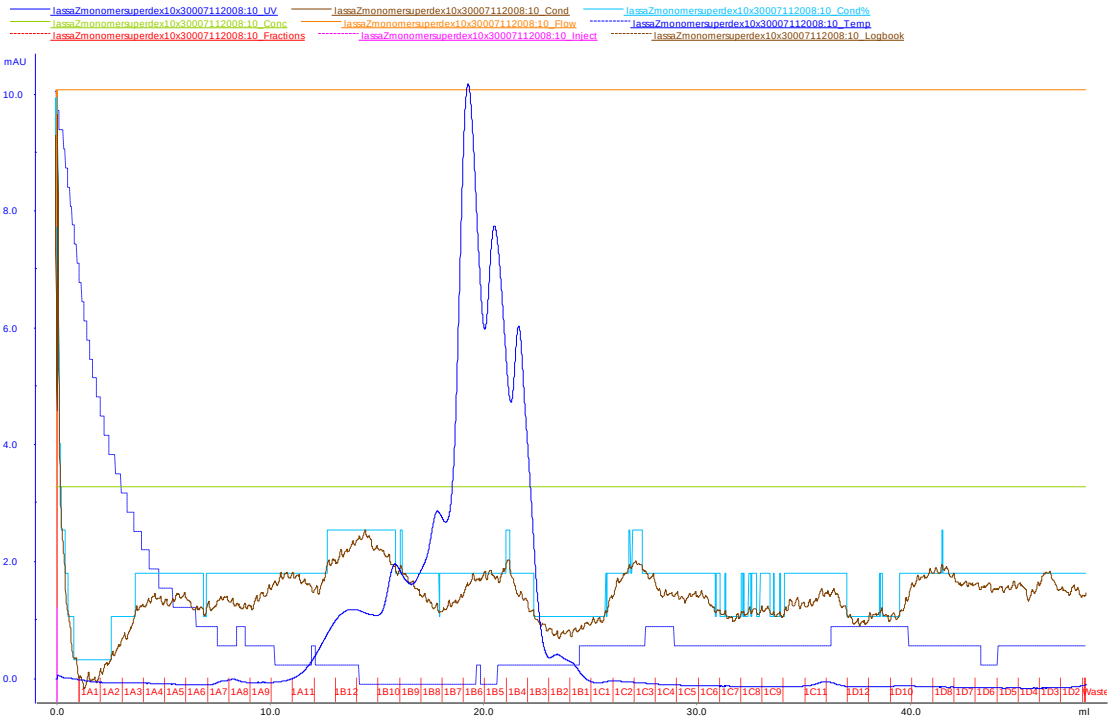


Figure 5.5 The purified monomeric Z protein from gel filtration by Superdex 200 10/300. Six peaks were formed, corresponding to different Z protein oligomers.

The elution volume of oligomeric Z is 12.3ml (Fig. 5.6), so the size of oligomeric Z protein is around 212 kDa. Because the monomeric Z protein is 11 kDa, so it was assumed about 20 monomers form the oligomeric Z protein.

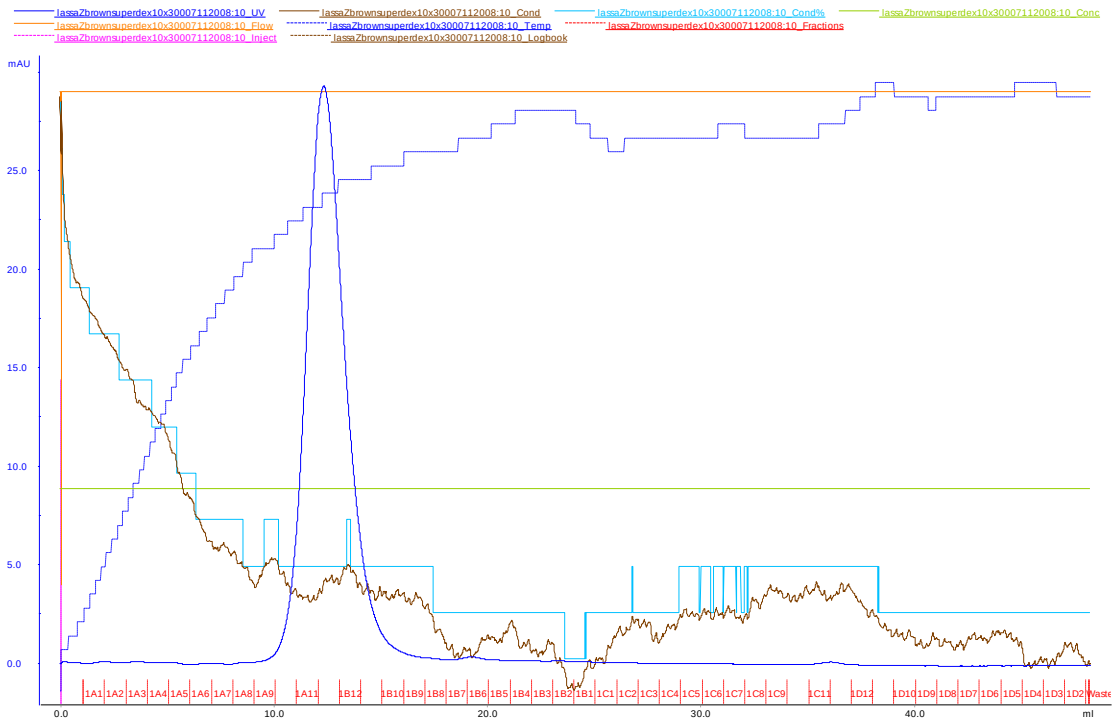


Figure 5.6 The purified oligomeric Z protein from gel filtration by Superdex 200.

5.4 Interaction between NP and Z protein

In order to test whether NP protein and Z protein have interaction in vitro, the Isothermal titration calorimetry (ITC) experiment was performed using VP-ITC microcalorimeter (MicroCal). 2ml of 3.75mg/ml NP protein and 450µl of 3.25mg/ml Z oligomer protein were loaded into VP-ITC with both proteins in same buffer containing 20 mM Tris pH7.5, 0.3 M NaCl and 10% glycerol. During the experiment, the ligand (Z protein) was titrated into the sample cell (NP protein) in precisely known aliquots (10 ul). The data were recorded and fitted using the Origin software (Fig. 5.7). In the figure, the K is binding constant and N is binding stoichiometry, while ΔS is entropy changes, and ΔH is enthalpy changes and ΔG is Gibbs energy changes.

Their relationship can be describe as:

$\Delta G=-RT\ln K=\Delta H-T\Delta S$ (R: gas constant; T: absolute temperature)

The binding constant is 3.11×10^{-5} M. Interestingly, we also determined the interaction between monomeric Z protein and NP protein using the VP-ITC and the result showed that there is no interaction between the NP and the monomeric Z protein.

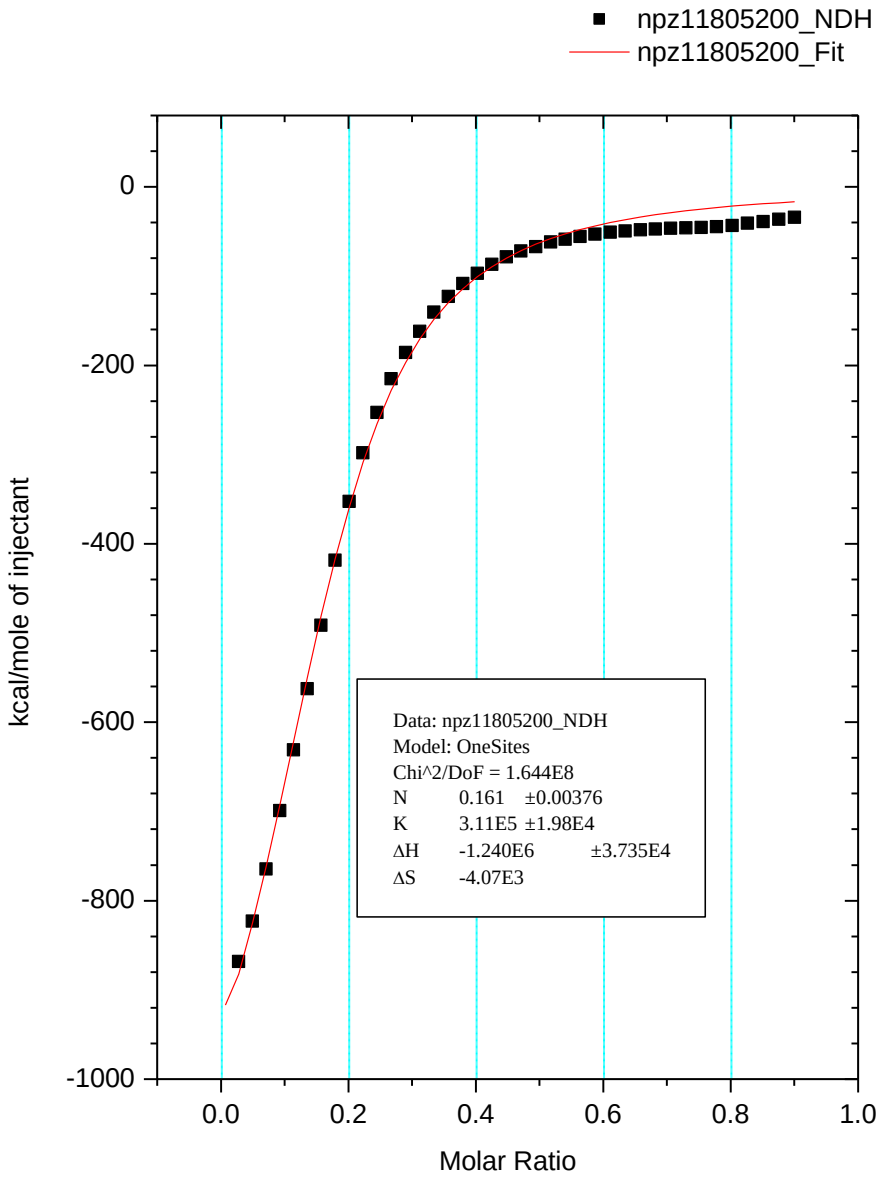


Figure 5.7 The ITC result of NP and oligomeric Z protein interaction.

Chapter 6: Discussion

6.1 The immune evasion function of LASV NP

Human cells have different receptors to detect viral infections and trigger different pathways to control the viral infections. To produce interferons is the first line for human immune system to attack viral infection. In the same time, viruses have developed different strategies to evade the host immune responses. By determination of the Lassa fever virus nucleoprotein structure and in vitro and in vivo assays, we are the first group in the world to discover that arenaviruses use 3'-5' exoribonuclease to suppress the human immune responses.

3'-5' exonucleases are a big family enzymes, which involve in many crucial processes in cells, such as the RNA maturation, DNA replication and DNA and RNA degradation. Trex1 is a human 3'-5' exonuclease, which is crucial for degrading DNAs during cell apoptosis or cell death, and genetic data and biochemical data showed that natural mutation of the Trex1 resulted in inactivation of the enzyme would lead to autoimmune disease (Yang, 2007; Stetson, 2008). As the structure of the C-terminus of the LASV NP is very similar to human Trex1, we proposed that the LASV NP can remove viral pathogen-associated pattern molecule RNAs, so that the human receptors RIG-I and MDA5 can't detect the viral infection and therefore no interferon production, It has evidences to show that the ligands for RIG-I and MDA5 are short triphophated dsRNA and long dsRNA, respectively. Our unpublished data showed that the NP could selectively recognize dsRNA and cleave the dsRNA much quicker than ssRNA. In the structure of C terminal domain of LASV NP protein, we identified that catalytic residues are D389, E391, D466, D533 and H528 which

belongs to DEDDh subfamily of exonucleases (Fig. 6.1). Exonucleases are known as two metal dependent enzymes (Steitz, 1993). However, from our structure studies, only one Mn^{2+} molecule is found in the LASV NP. A 3'-5' exonuclease, E.coli DNA polymerase was reported only one divalent cation molecule was found in the native complex, but the second metal molecule was found in the product analogy dNMP complex (Xia, 2012). We therefore proposed that the second Mn^{2+} molecule could be seen in the structure if the substrate bound on the NP.

Zinc finger fold is crucial for recognizing and binding DNA or RNA, and Zinc finger nucleases have been used as a specific tool for gene therapies (Urnoy, 2005). There is a zinc binding site near the active site of the LASV NP, which may play essential role in recognizing the dsRNA molecules. The purifications of the four mutant proteins of zinc binding site showed that their expression level in E.coli are very low (Fig. 4.21-4.24), which indicates that the residues around zinc ion may be critical for protein folding and stabilization of the structure and/or contribute to the substrate binding and specificity of the exoribonuclease activity.

The NP Tacaribe virus was reported to have no immune suppression function (Martínez-Sobrido, 2007; Lan, 2008), but our recently unpublished data shows that the NP of Tacaribe virus has the similar activity as the LASV NP in degrading dsRNA and ssRNA in vitro, and in vivo assays shows that the Tacaribe virus has the similar interferon suppression function as the LASV NP, which means that all the

arenaviruses known share the same function in immune suppression.

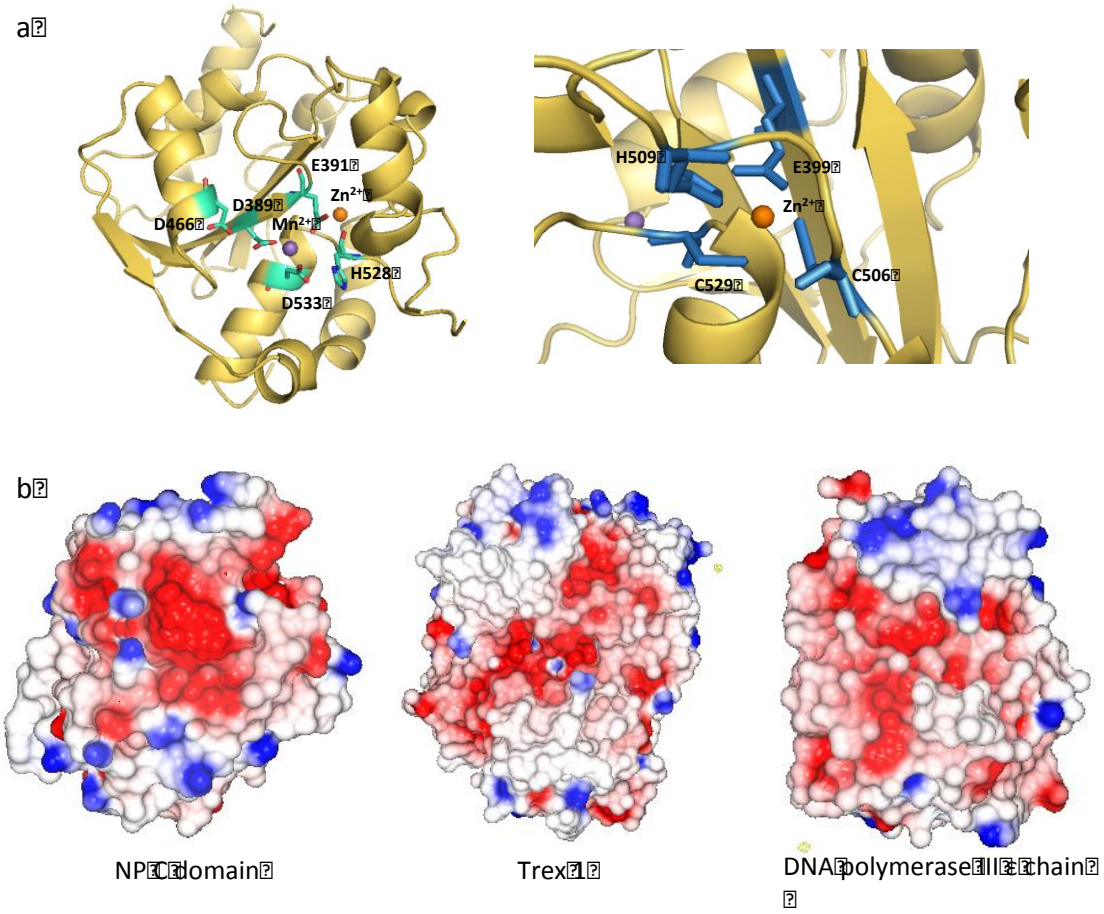


Figure 6.1 C-terminal domain of NP protein that holds the exoribonulcease activity site is strucuturally similar with many DEDDh exonucleases such as Trex1 and DNA polymerase III ε chain. a, five amino acids D389, E391, D466, H528 and D533 reside at the catalytic site, amino acids E399, C506, H509 and C529 reside around the Zn^{2+} finger fold motif; b, the electrostatics map of NP C-terminal domain, human Trex 1 (PDB code: 2IOC) and DNA polymerase III ε chain (PDB code:1J54) all present a negatively charged (colored in red) center.

6.2 The cap-binding cavity on LASV NP

The cap structure on 5' end of mRNA is essential for protein translation, it protects the mRNA from being degrading by 5' exonucleases. Unlike non-segmented group of negative strand RNA virus that the L protein has the capping activity, the segmented negative strand RNA viruses have to steal cap from the host cellular mRNA. The most

well established cap snatching model in segmented negative strand RNA virus is the heterotrimeric polymerase complex of influenza A virus which is consist of PA, PB1 and PB2, where the cap binding site is on PB2. But unlike the polymerase complex, the arenaviruses genomic RNA only encodes four proteins. The glycoprotein forms the spike on the surface to interact with host cell membrane, Z protein function as the matrix protein, that makes the cap binding site can only be located on NP or L protein. L protein is a RNA dependent RNA polymerase (RdRp) that contains three functional domains, the middle one is the RdRp, the N terminal domain is an endonuclease which is similar to PA subunit in influenza A virus. There is only one domain structure left uncharacterized and we don't know whether it is the cap-binding domain.

Although the N-terminal domain of LASV NP is structurally unlike any of the cap binding proteins, the hydrophobic thymidine-binding pocket that consist of F176, W164, L172, M54, L120, L239 and I241 shares common features with other cap binding protein such as PB2 in influenza A virus, the cap binding protein (CBC), the eukaryotic translation initiation factor (eIF4E) and the vaccinia viral protein 39 (VP39) (Fechter, 2005; Guilligay, 2008). The significant features of the cap binding protein is that there is an aromatic sandwich shape that the aromatic ring of the m^7G is anchored in the middle of two protein aromatic residues (F404/H357 in PB2, W102/W56 in eIF4E, Y20/Y43 in CBC and Y22/F180 in VP39), and there is at least one conserved negatively charged acidic amino acid (E361 in PB2) since the purine ring that contains the methylated N7 is positively charged and the m^7G counteracts with acidic

residues Glu or Asp on the protein, plus there is a general area for binding the triphosphate group (Fechter, 2005). In N-terminal domain of NP, residues K253, R300, K309 and R323 are involved in anchoring the triphosphate group (colored in orange in Fig.59) of dTTP, whereas the acidic residue E117 that is assumed to counteract with the methylated N7 on the m^7G is located at the deep end of the cavity. But unlike the typical aromatic sandwhich shape in cap binding protein, only amino acid F176 is found to horizonally align with the aromatic ring on dTTP (Fig. 6.2). This unique aromatic architecture may require residue W614 to change conformation to accommodate the m^7G.

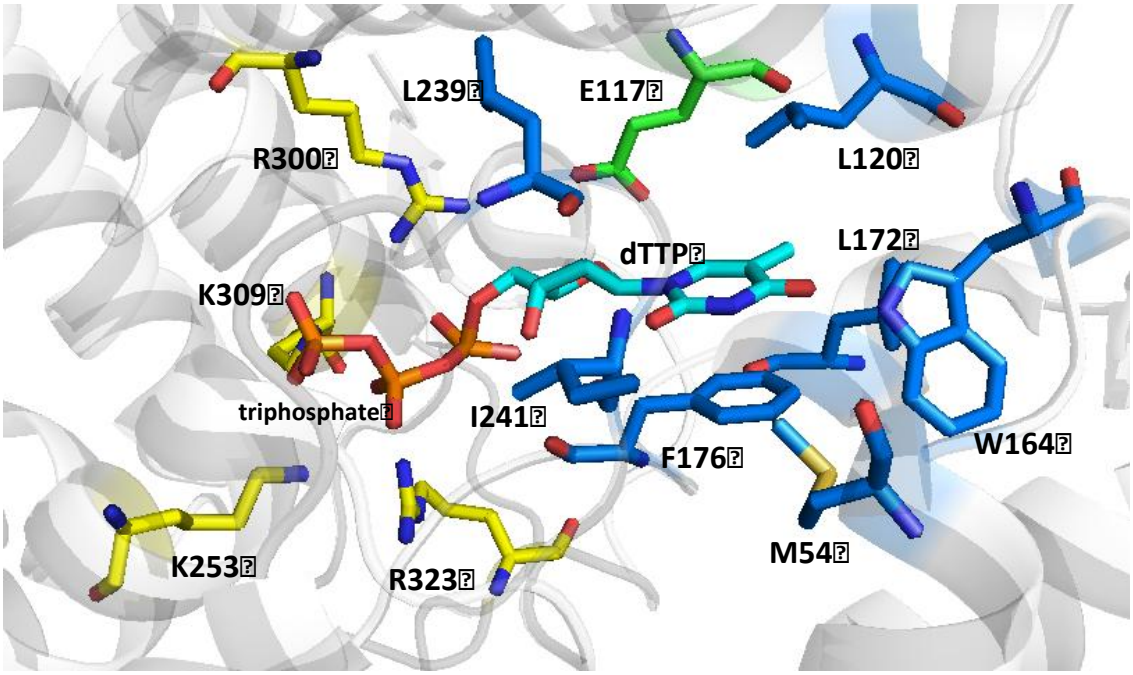


Figure 6.2 N-terminal domain of NP in complex with cap analog dTTP. Residues that anchor the triphosphate group are colored in yellow, residues that form the hydrophobic cavity are colored mainly in blue, the acidic residue E117 is colored in green.

Since influenza A virus and Lassa virus are both belong to the segmented negative strand RNA virus, comparing the cap binding protein PB2 and NP of these two virus

will be present below (Fig. 6.3). The residues that are located at the hydrophobic cavity of PB2 on influenza A virus are F323, F325, F330, F404 and F363. Residues Phe404 and H357 forms the top and the bottom of the aromatic sandwich, whereas acidic residues E361 is also present at the same cavity.

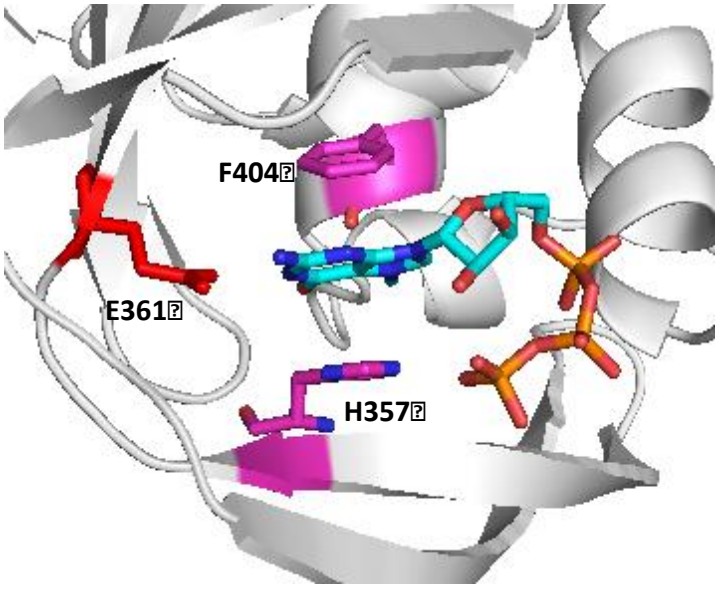


Figure 6.3 PB2 subunit of influenza A virus in complex with m^7GTP. The acidic residue E361 is colored in red, and the residues that form the sandwhich architecture are colored in magentas.

The triphosphate group can be observed from both the dTTP and UTP complex electron density map that has been described in last chapter. Along with a second hydrophobic cavity at the entrance, where a base recognizing residue E266 was found, the whole cavity within the center of N-terminal domain of NP is considered to be the entire m^7GpppN cap structure binding site which shares the common feature with other typical cap binding protein. The in vivo experiments were set to characterize the viral transcription function of NP by using the LASV minigenome replicon assay. The Renilla luciferase activity presents the LASV MG mRNA expression level with

participation of wild type NP and various mutants that substitute single amino acid on cap binding site with alanine. The various degree of defect on MG mRNA transcription shows the critical role of these residues that reside around the hydrophobic cap binding cavity in viral genomic RNA transcription. Both the structural and the functional result are consistent of the hypothesis of NP's role in cap snatching mechanism in Arenaviridae family. Of course, further studies are required to further characterize the cap binding function, such as to test the NP binding affinity to m^7GpppG and to have complex structure of the NP in complex with m^7GpppG.

6.3 Update of the recently discovered NP-RNA complex

The Saphire's research group in Scripps institute recently solved the structure of N domain of LASV NP and RNA complex (Hastie, 2011) (Fig. 6.4). This complex structure reveals a gating mechanism for RNA binding with NP. The structure of N terminal domain of NP binds to a random RNA from the expression host E.coli cell while the whole length of NP protein can't be crystallized with RNA. The difference between our model and the complex is that helix α5 and α6 will change conformation to open a gate for RNA to bind within the center cavity. The β2 which is formed by residues V24 and Y25 sits between α5 and α6 in our structural model, but in the NP-RNA complex model (Hastie, 2011) α5 and α6 are widely open, and the residues between them are absent. After RNA binding to NP, a linear form of NP binding with RNA will take place of the trimeric NP. Although more experimental evidences need to be provided, this complex structure offers a possible way for NP

encapsidating the viral genomic RNA.

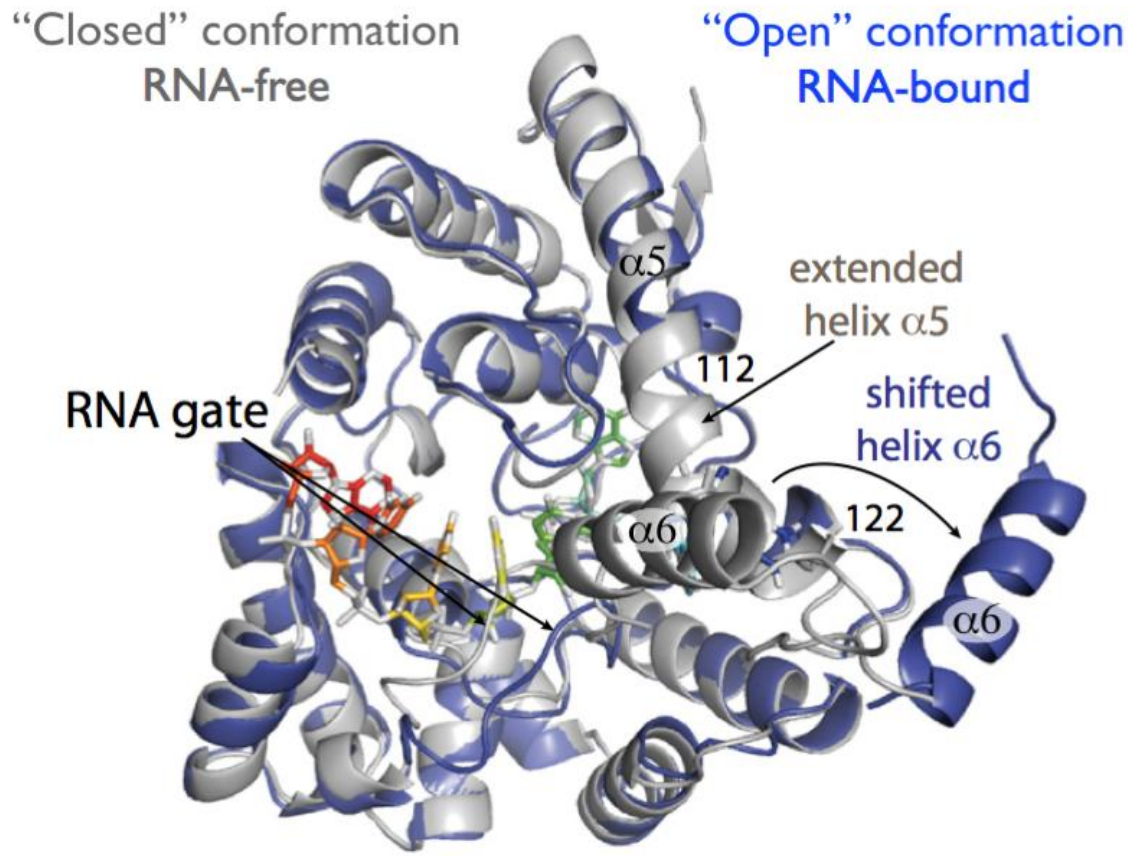


Figure 6.4 The comparision between native structure of NP N domain and NP-RNA complex. The native NP is colored in grey while the complex is colored in blue (Figure from Saphire, 2011).

6.4 Interactions between NP and other arenaviral proteins

After solving the structure of LASV NP protein, it becomes our interest to know whether and how NP interacts with other arenaviral proteins. The matrix LASV Z protein has been reported to be sufficient for mature virion budding and can modulate the activity of vRNP by interact with NP protein (Strecker, 2003; Capul, 2011). The in vitro assay ITC result showed that there is interaction between LASV NP and oligomeric Z protein (described in chapter 5). The recently published study in

Ortiz-Riaño's group also pointed out the residues on LCMV NP that are involved in type I IFN suppression overlap with the residues that play critical role in NP-Z interaction, and the mutation study showed these two functions are independently carried out (Ortiz-Riaño, 2011).

L protein of arenaviruses contains a middle RdRp domain and an endonuclease domain on its N terminal side. Although there has been studies on the interaction between NP and RdRp L protein, like the report on residue 53 and C-terminal domain of LMCV NP are critical for LMCV L protein's function during replication and transcription (Kerber, 2011), how did they interact to each other and the interaction's role in transcription remains unclear. In the future, the detailed structural information of the NP-L complex may provide more evidence in arenaviruses replication and transcription mechanism.

6.5 Conclusion and future work

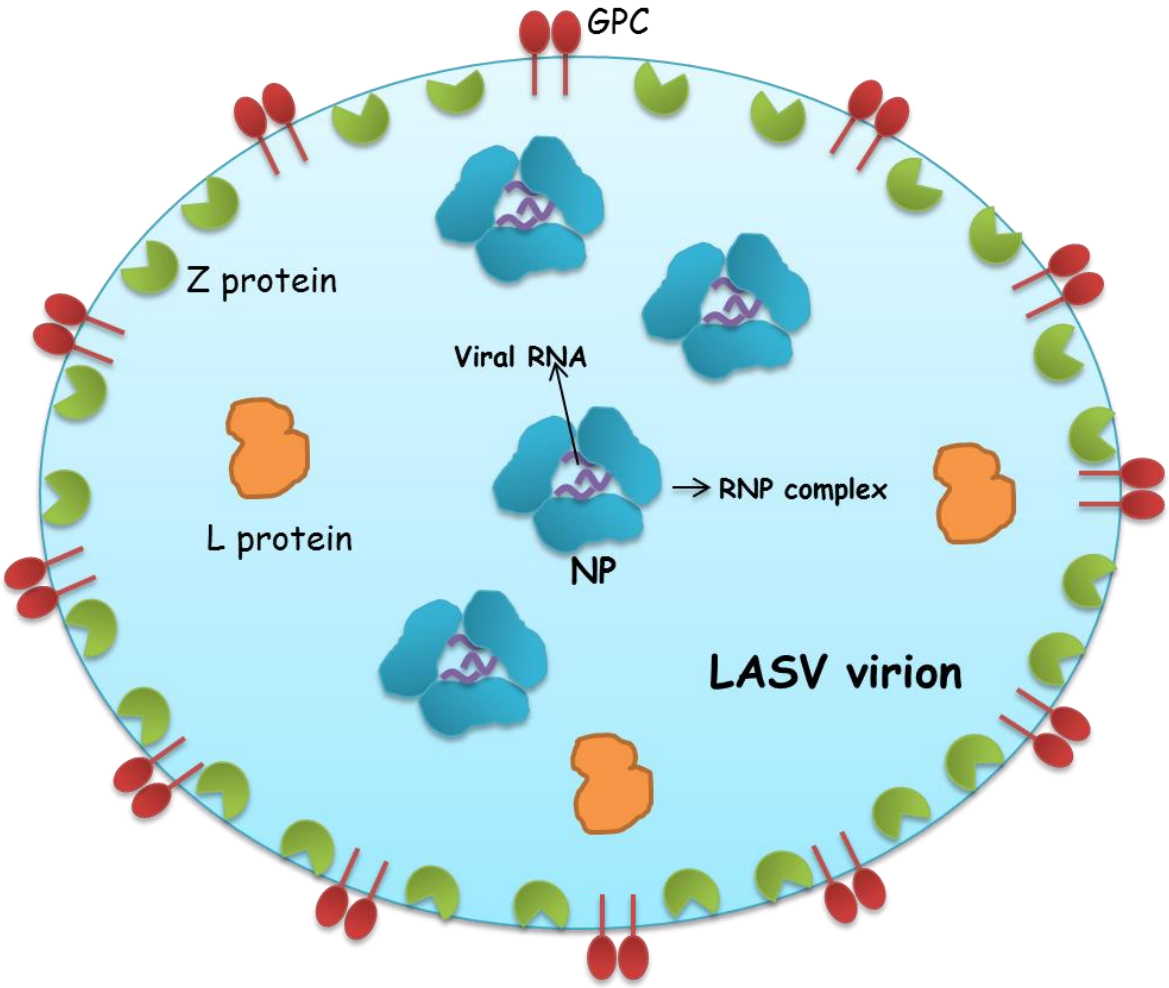
Both the structural and functional studies provide evidences that 3'-5' exoribonuclease activity in C terminal domain of Lassa nucleoprotein performs suppression of type I IFN-β induction. This is the first nucleoprotein structure in arenaviruses, the IFN suppression function shield the light in how arenaviruses escape from the host innate immune system, our research suggests that NP plays a critical role in suppressing IFN by removing PAMP RNA ligands specifically. The potential cap binding site on N terminal domain of LASV NP explains the cap snatching mechanism of arenaviruses during viral genome transcription, this cap snatching

mechanism that is shared in segmented negative strands RNA viruses provide transcriptional evidence in evolution of negative strand RNA virus. Along with the typical zinc finger ring fold on C terminal domain, these findings reveal many potential drug target based on the NP structure model that can be used not only in designing vaccine to combat with arenaviruses but also provide structural evidence for studies on other negative strand RNA viruses.

The quaternary centrosymmetric structure of LASV NP in solution was then obtained and confirmed by small-angle x-ray scattering (SAXS) and cryo-electron microscopy (cryo-EM) by Brunotte's group (Brunotte L, 2011). Although the N terminal domain of NP with host random RNA complex was solved, the way of NP encapsidating viral genomic RNA is not clear in details. The question of what specific RNA binding with NP to form the RNP complex need be studied in the further work. Taken together, these findings may provide us very good target for design drug.

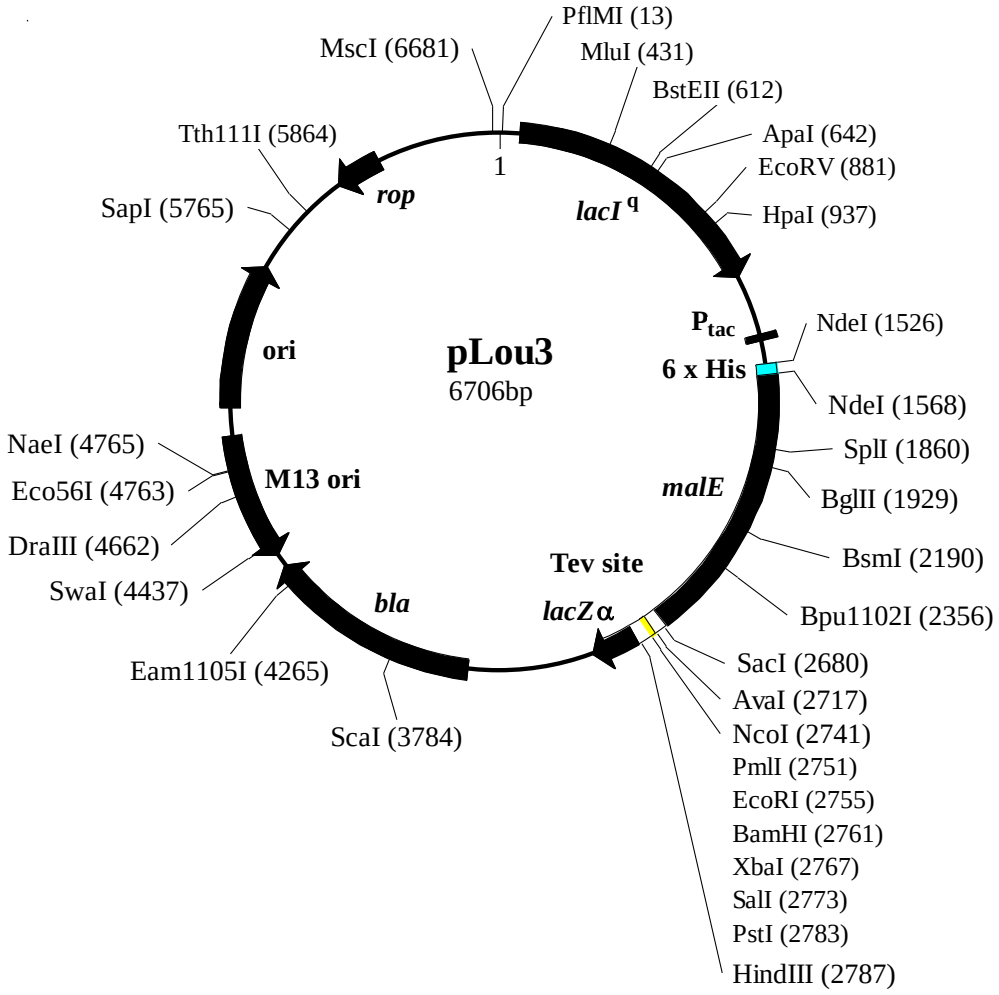
Appendix I

I.1 Diagram of LASV virion components.



I.2 Plasmid constrcution: vector pLou3

Both LASV NP and Z protein were ligated on pLou3 vector that contains the 6His-tag site at the N terminus of maltose binding protein (MBP) with the tev cleavage site between the target gene and MBP gene.



I.3 Forward and reverse primers of the mutants

Mutants	F/R primers	Primer sequences
M54	Forward	AACGGTTG GCG CGCAAGGAGAGAAGGGATGACAATGATTTGAAAC
	Reverse	TCCTTGCG CGC CAACCGTTGAACATTACTGACTTCGGAGAAGTC
E117	Forward	TCAGAACA GCA AGGCCATTAAGTGCAGGTGTCTATATGGGCAAC
	Reverse	TAATGGCCT TGC TGTTCTGATCACCTTTGACTTTAACTTCTCTAG
L120	Forward	AAAGGCCA GCA AGTGCAGGTGTCTATATGGGCAACCTAAGCTC
	Reverse	CCTGCACT TGC TGGCCTTTCTGTTCTGATCACCTTTGACTTTAAC
W164	Forward	GTGAGAGTT GCG GATGTGAA;AAATGCAGAGTTGCTCAATAATCAG
	Reverse	TTCACATC CGC AACTCTCAC;CACTCCATCTCTCCCAGCCCGAG
F176	Forward	ATAATCAG GCC GGGACCA;TGCCAAGTCTGACACTGGCATGTC
	Reverse	TGGTCCC GGC CTGATTAT;TGAGCAACTCTGCATTTTTCACATC
Y209	Forward	GATTTGGGTTTGATC GCC ACAGCAAAGTATCCCAAC
	Reverse	GATACTTTGCTGT GGC GATCAAACCCAAATCTGTC
Y213	Forward	GATCTACACAGCAAAG GCT CCCAACACTTCAGACTTAG
	Reverse	CTGAAGTGTTGGG AGC CTTTGCTGTGTAGATCAAAC
K253	Forward	GGTGCAGCTGTG GCG GCAGGAGCTTGCATGC
	Reverse	GAAGCTCCTGC CGC CACAGCTGCACCCAAGC
E266	Forward	GGCAATATGTTG GCG ACAATCAAGGTGTCAC
	Reverse	CACCTTGATTGT CGC CAACATATTGCCACCATC
R300	Forward	CTGGTGAA GCG AATCCTTATGAAAACATACTCTACAAGATTTGTTTG
	Reverse	ATAAGGATT CGC TTCACCAGGGGTGTCTGAAATGAACATTCCAAG
K309	Forward	ATACTCTAC GCG ATTTGTTTGTCAGGAGATGGATGGCCATATATTG
	Reverse	CAAACAAAT CGC GTAGAGTATGTTTTCATAAGGATTCCTTTCACCAG
Y319	Forward	GATGGATGGCCA GCT ATTGCATCAAGAACCTC
	Reverse	TCTTGATGCAAT AGC TGGCCATCCATCTCCTG
R323	Forward	ATTGCATCA GCA ACCTCAATAACAGGAAGGGCCTGGGAAAACACTG
	Reverse	TATTGAGGT TGC TGATGCAATATATGGCCATCCATCTCCTGACAAAC
D389	Forward	GACCTGGATG GCC ATTGAAGGAAGACCTG
	Reverse	CTTCCTTCAAT GGC CATCCAGGTCTTAGC
E391	Forward	GGATGGACATT GCA GGAAGACCTGAAGATC
	Reverse	CAGGTCTTCC TGC AATGTCCATCCAGGTC
G392	Forward	GACATTGAA GCA AGACCTGA AGATCCAGTGGAAATTGCCCTCTA
	Reverse	TCAGGTCTTGCTTCAATGTCCATCCAGGTCTTAGCATTTGGGTC
E399	Forward	GATCCAGTG GCA ATTGCCC TCTATCAACCAAGTTCAGGCTGCTA
	Reverse	GGGCAATTGC CACTGGATC TTCAGGTCTTC CTTCAATGTCCATC
D466	Forward	GGGGTCCGAT GCC ATAAGGAAACTCCTTG
	Reverse	GAGTTTCCTTAT GGC ATCGGACCCCTGACAG
C506	Forward	AAAGACTTA GCC CACATGCA CACAGGTGTCGTTGTTGAAAAGAAG
	Reverse	TGCATGTGGC ATAAGTCTTT ATACTGGTCC CAGACTGCAT TTTCA
H509	Forward	GCCACATG GCC ACAGGTGT CGTTGTTGAAAAGAAGAAAAGAGGC
	Reverse	*ACACCTGTGGCCATGTGGC ATAAGTCTTTATACTGGTCCCAGAC*
H528	Forward	TAACCCCT GCC TGTGCACT AATGGACTGCATCATGTTTGATGCA

	Reverse	AGTGCACAGG CAGGGGTTA TTTCCTCTTTACCGCCTCTTTTCTT
C529	Forward	CCCCTCAC GCT GCACTAAT GGACTGCATCATGTTTGATGCAGCA
	Reverse	ATTAGTGCAG CGTGAGGGG TTATTTCCTCT TTACCGCCTCTTTTC
D533	Forward	ACTAATG GCC TGCATCATGT TTGATGCAGCAGTGTCAGGA
	Reverse	ACATGATGCA GGCCATTAGT GCACAGTGAGGGGTTATTTCCTCT

I.4 In vitro ssRNA synthesis

ssRNA	Sequences/Templates/Primers	company/kits	chem/bio synthesis
30nt cRNA	sequence 5’-CUGGGC UUACCUAUUCUCAGCUGAUGACCC-3’	Eurogenes	chemical
30nt vRNA	sequence 5’-GGGUCAUCAGCUGAGAAUAGGUAAGCCCAG-3’	Eurogenes	chemical
30nt	sequence 5’-GGGUCAUCAGCUGAGAAUAGGUAAGCCCAG-3’	MEGAshortscript	biological
5'ppp vRNA		kit (Ambion)	
32nt	Template	MEGAshortscript	biological
5'ppp cRNA	5’-GGGTCATCAGCTGAGAATAGGTAAGCCCAGCCTATAGTGAGTCGTATTAAATT-3’	kit (Ambion)	
60nt	Template 5’-AATTTAATACGACTCACTATAGG-3’	MEGAshortscript	biological
5'ppp cRNA	5’-GTAAATCCCTGCAGTCGGCAGGGTTTACCGCTGGGCTTACCTATTCTCAGCTGATGA	kit (Ambion)	
	CCCTATAGTGAGTCGTAT TAAATT-3’		
126nt capped	F primer 5’-AATTTAATACGACTCACTATAGGGAAAACACTGTCGTTGATCTGGAATC-3'	mMESSAGE	biological
mRNA	R primer 5’-GGGTCATCAGCTGAGAATAGGTAAGCCCAGCGG-3’	mMACHINE T7 Ultra	
		kit (Ambion)	

I.5 In vitro dsRNA synthesis

dsRNA	Synthesized by ssRNA
5'-OH/5’-OH dsRNA	30nt cRNA (Chem); 30nt vRNA (Chem)
5'ppp/5'ppp dsRNA	5'ppp cRNA (Bio); 5'ppp vRNA (Bio)
5'ppp/5'-OH dsRNA	32nt 5'ppp vRNA (Bio); 30nt 5'-OH cRNA (Chem)

Bibliography

Abate, D.A., Watanabe, S. and Mocarski, E.S. (2004). Major human cytomegalovirus structural protein pp65 (ppUL83) prevents interferon response factor 3 activation in the interferon response. J Virol *78*, 10995-11006.

Baize, S., Marianneau, P., Loth, P., Reynard, S., Journeaux, A., Chevallier, M., Tordo, N., Deubel, V., Contamin, H. (2009). Early and strong immune responses are associated with control of viral replication and recovery in lassa virus infected cynomolgus monkeys. J Virol *83*, 5890-5903.

Beaton, A.R., Krug, R.M. (1981). Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. Nucleic Acids Res *9*, 4423-4436.

Beese, L.S., Steitz, T.A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. EMBO J *10*, 25-33.

Bentz, G.L., Liu, R., Hahn, A.M., Shackelford, J. and Pagano, J.S. (2010). Epstein-Barr virus BRLF1 inhibits transcription of IRF3 and IRF7 and suppresses induction of interferon-β. Virology *402*, 121-128.

Berthold, W., Tan, C. and Tan, Y.H. (1978). Chemical modifications of tyrosyl residue(s) and action of human-fibroblast interferon. Eur J Biochem *87*, 367-370.

Berthold, W., Tan, C. and Tan, Y.H. (1978). Purification and in vitro labelling of interferon from a human fibroblastoid cell line. J Biol Chem *253*, 5206-5212.

Borden, K.L.B., Dwyer, E.J.C., and Salvato, M.S. (1998). An arenavirus RING (zinc-binding) protein binds the oncoprotein promyelocyte leukemia protein (PML) and relocates PML nuclear bodies to the cytoplasm. J Virol *72*, 758-766.

Borden, K.L.B., Dwyer, E.J.C., Carlile, G.W., Djavani, M., and Salvato, M.S. (1998). Two RING finger proteins, the oncoprotein PML and the arenavirus Z protein, colocalize with the nuclear fraction of the ribosomal P proteins. J Virol *72*, 3819-3826.

Bouloy, M., Plotch, S.J., Krug, R.M. (1978). Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. Proc Natl Acad Sci U S A *75*, 4886-4890.

Brunotte, L., Kerber, R., Shang, W., Hauer, F., Hass, M., Gabriel, M., Lelke, M., Busch, C., Stark, H., Svergun, D.I., Betzel, C., Perbandt, M., Günther, S. (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.

Burns, J. W. & Buchmeier, M. J. (1993). Glycoproteins of the arenaviruses. The Arenaviridae, ed. Salvato, M. S (Plenum, New York), pp. 17–35

Cao, W., Henry, M. D., Borrow, P., Yamada, H., Elder, J. H., Ravkov, E. V., Nichol, S. T., Compans, R. W., Campbell, K. P. & Oldstone, M. B. (1998). Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. Science *282*, 2079 –2081.

Capul, A.A., de la Torre, J.C. and Buchmeier, M.J. (2011). Conserved residues in Lassa Fever virus Z protein modulate viral infectivity at the level of the Ribonucleoprotein. J Virol *85*, 3172-3178.

Cárdenas, W.B., Loo, Y.M., Gale, M.Jr., Hartman, A.L., Kimberlin, C.R., Martínez-Sobrido, L., Saphire, E.O., Basler, C.F. (2006). Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. J Virol *80*, 5168–5178.

Charrel, R.N., de Lamballerie, X., Emonet, S. (2008). Phylogeny of the genus Arenavirus. Curr Opin Microbiol *11*, 362-368.

Charrel, R.N., Coutard, B., Baronti, C., Canard, B., Nougairede, A., Frangeul, A., Morin, B., Jamal, S., Schmidt, C.L., Hilgenfeld, R., Klempa, B., de Lamballerie, X. (2011). Arenaviruses and hantaviruses: from epidemiology and genomics to antivirals. Antiviral Res *90*, 102-114.

Cheng, A., Wong, S.M., Yuan, Y.A. (2009). Structural basis for dsRNA recognition by NS1 protein of influenza A virus. Cell Res *19*, 187-195.

Cisneros, G. A., Perera, L., Schaaper, R.M., Pedersen, L.C., London, R.E., Pedersen, L.G., Darden, T.A. (2009). Reaction mechanism of the epsilon subunit of E. coli DNA polymerase III: insights into active site metal coordination and catalytically significant residues. J Am Chem Soc *131*, 1550–1556.

Crescenzo-Chaigne, B., van der Werf, S., Naffakh, N. (2002). Differential effect of nucleotide substitutions in the 3’ arm of the influenza A virus vRNA promoter on transcription/replication by avian and human polymerase complexes is related to the nature of PB2 amino acid 627. Virology *303*, 240–252.

Crow, Y. J. and Rehwinkel, J. (2009). Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. Hum Mol Genet *18*,

R130–R136.

Cui, S., Eisenächer, K., Kirchhofer, A., Brzózka, K., Lammens, A., Lammens, K., Fujita, T., Conzelmann, K.K., Krug, A., and Hopfner, K.P. (2008). The C-terminal regulatory domain is the RNA 50-triphosphate sensor of RIG-I. Mol Cell *29*, 169–179.

de Silva, U., Choudhury, S., Bailey, S.L., Harvey, S., Perrino, F.W., Hollis, T. (2007). The crystal structure of TREX1 explains the 39 nucleotide specificity and reveals a polyproline II helix for protein partnering. J Biol Chem *282*, 10537–10543.

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.

Duijsings, D., Kormelink, R., Goldbach, R. (2001). In vivo analysis of the TSWV cap-snatching mechanism: single base complementarity and primer length requirements. EMBO J *20*, 2545-2552.

Eichler, R., Strecker, T., Kolesnikova, L., ter Meulen, J., Weissenhorn, W., Becker, S., Klenk, H.D., Garten, W., Lenz, O. (2004). Characterization of the Lassa virus matrix protein Z: electron microscopic study of virus-like particles and interaction with the nucleoprotein (NP). Virus Res *100*, 249-255.

Emonet, S.F., de la Torre, J.C., Domingo, E., Sevilla, N. (2009). Arenavirus genetic diversity and its biological implications. Infect Genet Evol *9*, 417-429.

Fasler-Kan, E., Wunderlich, K., Hildebrand, P., Flammer, J., Meyer, P. (2005). Activated STAT 3 in choroidal neovascular membranes of patients with age-related macular degeneration. Ophthalmologica *219*, 214-221.

Fechter, P. and Brownlee, G.G. (2005). Recognition of mRNA cap structures by viral and cellular proteins. J GenVirol *86*, 1239-1249.

Fischer, S. A., Graham, M. B., Kuehnert, M. J., Kotton, C. N., Srinivasan, A., Marty, F. M., Comer, J. A., Guarner, J., Paddock, C. D., DeMeo, D. L., Shieh, W. J., Erickson, B. R., Bandy, U., DeMaria, A., Davis, Jr., J. P., Del- monico, F. L., Pavlin, B., Likos, A., Vincent, M. J., Sealy, T. K., Goldsmith, C. S., Jernigan, D. B., Rollin, P. E., Packard, M. M., Patel, M., Rowland, C., Helfand, R. F., Nichol, S. T., Fishman, J. A., Ksiazek, T. and Zaki, S. R. (2006). Transmission of lymphocytic choriomeningitis virus by organ transplantation. N Engl J Med *354*, 2235–2249.

Fisher-Hoch, S.P., McCormick, J.B. (1987). Pathophysiology and treatment of Lassa fever. Curr Top Microbiol Immunol *134*, 231-239.

Fodor, E., Crow, M., Mingay, L.J., Deng, T., Sharps, J., Fechter, P., Browniee, G.G. (2002). A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. J Virol *76*, 8989–9001.

Garcin, D., Rochat, S., Kolakofsky, D. (1993). The Tacaribe arenavirus small zinc finger protein is required for both mRNA synthesis and genome replication. J Virol *67*, 807-812.

Geisbert, T.W., Jones, S., Fritz, E.A., Shurtleff, A.C., Geisbert, J.B., Liebscher, R., Grolla, A., Ströher, U., Fernando, L., Daddario, K.M., Guttieri, M.C., Mothé, B.R., Larsen, T., Hensley, L.E., Jahrling, P.B., Feldmann, H. (2005). Development of a new vaccine for the prevention of Lassa fever. PLos Med *2*, e183.

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

Hara, K., Schmidt, F. I., Crow, M. and Brownlee, G. G. (2006). Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. J Virol *80*, 7789–7798.

Hastie, K.M., Liu, T., Li, S., King, L.B., Ngo, N., Zandonatti, M.A., Woods, V.L., Jr, de la Torre, J.C. and Saphire, E.O. (2011). Ctrystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding. Proc Natl Acad Sci USA *108*, 19365-19370.

Hodel, A.E., Gershon, P.D. and Quiocho, F.A. (1998). Structural basis for sequence-nonspecific recognition of 5'-capped mRNA by a cap-modifying enzyme. Mol Cell *1*, 443–447.

Honda, K., Yanai, H., Takaoka, A. and Taniguchi, T. (2005a). Regulation of the type I IFN induction: a current view. Int Immunol *17*, 1367-1378.

Honda, K., Yanai, H., Negishi, H., Asagiri, M., Sato, M., Mizutani, T., Shimada, N., Ohba, Y., Takaoka, A., Yoshida, N., Taniguchi, T. (2005b). IRF-7 is the master regulator of type-I interferon-dependent immune responses. Nature *434*, 772-777.

Isaacs, A., Lindenmann, J. (1957). Virus interference. I. The interferon. Proc R Soc Lond B Biol Sci *147*, 258-267.

Jiang, F., Ramanathan, A., Miller, M.T., Tang, G.Q., Gale, M. Jr., Patel, S.S.,

Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature *479*, 423–427.

Jin, H. and Elliott, R.M. (1993). Characterization of Bunyamwera virus S RNA that is transcribed and replicated by the L protein expressed from recombinant vaccinia virus. J Virol *67*, 1396–1404.

Kentsis, A., Dwyer, E.C., Perez, J.M., Sharma, M., Chen, A., Pan, Z.Q., Borden, K.L. (2001). The RING domains of the promyelocytic leukemia protein PML and the arenaviral protein Z repress translation by directly inhibiting translation initiation factor eIF4E. J Mol Biol *312*, 609-623.

Kerber, R., Rieger, T., Busch, C., Flatz, L., Pinschewer, D.D., Kümmerer, B.M. and Günther, S. (2011). Cross-species analysis of the replication complex of Old World arenaviruses reveals two nucleoprotein sites involved in L protein function. J Virol *85*, 12518-12528.

Kimberlin, C.R., Bornholdt, Z.A., Li, S., Woods, V.L. Jr., MacRae, I.J. and Saphire, E.O. (2010). Ebolavirus VP35 uses a bimodal strategy to bind dsRNA for innate immune suppression. Proc Natl Acad Sci USA *107*,314-319.

Khan, S.H., Goba, A., Chu, M., Roth, C., Healing, T., Marx, A., Fair, J., Guttieri, M.C., Ferro, P., Imes, T., Monagin, C., Garry, R.F., Bausch, D.G.; Mano River Union Lassa Fever Network (2008). New opportunities for field research on the pathogenesis and treatment of Lassa fever. Antiviral Res *78*, 103-115.

Kotenko, S.V., Gallagher, G., Baurin, V.V., Lewis-Antes, A., Shen, M., Shah, N.K., Langer, J.A., Sheikh, F., Dickensheets, H., Donnelly, R.P. (2003). IFN-lambdas mediate antiviral protection through a distinct class II cytokine receptor complex. Nat Immunol *4*, 69-77.

Kowalinski, E., Lunardi, T., McCarthy, A.A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D. and Cusack, S. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. Cell *147*, 423-435.

Labadie, K., Dos Santos Afonso E., Rameix-Welti M.A., van der Werf S., Naffakh N. (2007). Host-range determinants on the PB2 protein of influenza A viruses control the interaction between the viral polymerase and nucleoprotein in human cells. Virology *362*, 271–282.

Lan, S., McLay, L., Aronson, J., Ly, H. and Liang, Y. (2008). Genome comparison of virulent and avirulent strains of the Pichinde arenavirus. Arch Virol *153*, 1241–1250.

Lee-Kirsch, M.A., Gong, M., Chowdhury, D., Senenko, L., Engel, K., Lee, Y.A., de

Silva, U., Bailey, S.L., Witte, T., Vyse, T.J., Kere, J., Pfeiffer, C., Harvey, S., Wong, A., Koskenmies, S., Hummel, O., Rohde, K., Schmidt, R.E., Dominiczak, A.F., Gahr, M., Hollis, T., Perrino, F.W., Lieberman, J., Hübner, N. (2007). Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. Nature Genet *39*, 1065–1067.

Lehtinen, D. A., Harvey, S., Mulcahy, M. J., Hollis, T. and Perrino, F. W. (2008). The TREX1 double-stranded DNA degradation activity is defective in dominant mutations associated with autoimmune disease. J Biol Chem *283*, 31649–31656.

Lelke, M., Brunotte, L., Busch, C., and Günther, S. (2010). An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome. J Virol *84*, 1934-1944.

Lenz, O., ter Meulen, J., Klenk, H.D., Seidah, N.G. and Garten, W. (2001). The Lassa virus glycoprotein precursor GP-C is proteolytically processed by sub- tilase SKI-1/S1P. Proc Natl Acad Sci USA *98*, 12701–12705.

Leung, D.W., Prins, K.C., Borek, D.M., Farahbakhsh, M., Tufariello, J.M., Ramanan, P.,Nix, J.C., Helgeson, L.A., Otwinowski, Z., Honzatko, R.B., Basler, C.F. and Amarasinghe, G.K. (2010) Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. Nat struct Mol Biol *17*, 165-173.

Lopez, N., Jacamo, R. and Franze-Fernandez, M. T. (2001). Transcription and RNA replication of tacaribe virus genome and antigenome analogs require N and L proteins: Z protein is an inhibitor of these processes. J Virol *75*, 12241–12251.

Lu, Y., Wambach, M., Katze, M.G. and Krug, R.M. (1995). Binding of the Influenza Virus NS1 protein to double-stranded RNA inhibits the activation of the protein kinae that phosphorylates the elF-2 translation initiation factor. Virology *214*, 222-228.

Mahanty, S., Hutchinson, K., Agarwal, S., McRae, M., Rollin, P.E., Pulendran, B. (2003). Cutting edge: impairment of dendritic cells and adaptive immunity by Ebola and Lassa viruses. J Immunol *170*, 2797-2801.

Marcotrigiano, J., Gingras, A.C., Sonenberg, N. and Burley, S.K. (1997). Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. Cell *89*, 951–961.

Martínez-Sobrido, L., Zúñiga, E.I., Rosario, D., García-Sastre, A., and de la Torre, J.C. (2006). Inhibition of the type I interferon response by the nucleoprotein of the prototypic arenavirus Lymphocytic Choriomeningitis virus. J Virol *80*, 9192-9199.

Martínez-Sobrido L., Giannakas P., Cubitt B., García-Sastre A., and de la Torre JC. (2007). Differential Inhibition of Type I Interferon Induction by Arenavirus

Nucleoproteins. J Virol *81*, 12696-12703.

Martínez-Sobrido, L., Emonet, S., Giannakas, P., Cubitt, B., García-Sastre, A., and de la Torre, JC. (2009). Identification of Amino Acid Residues Critical for the Anti-Interferon Activity of the Nucleoprotein of the Prototypic Arenavirus Lymphocytic Choriomeningitis Virus. J Virol *83*, 11330-11340.

Mazza, C., Segref, A., Mattaj, I.W. and Cusack, S. (2002). Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. EMBO J *21*, 5548–5557.

McCartney, S.A., Colonna, M. (2009). Viral sensors: diversity in pathogen recognition. Immunol Rev *227*, 87-94.

Meyer, B. J. and Southern, P. J. (1993). Concurrent sequence analysis of 5' and 3' RNA termini by intramolecular circularization reveals 5' nontemplated bases and 3' terminal heterogeneity for lymphocytic choriomeningitis virus mRNAs. J Virol *67*, 2621–2627.

Morin, B., Coutard, B., Lelke, M., Ferron, F., Kerber, R., Jamal, S., Frangeul, A., Baronti, C., Charrel, R., de Lamballerie, X., Vonrhein, C., Lescar, J., Bricogne, G., Günther, S., and Canard, B. (2010). The N-terminal domain of the Arenavirus L protein is an RNA endonuclease essential in mRNA transcription. Plos Pathog *6*, e1001038.

Naffakh, N., Massin, P., Escriou, N., Crescenza-Chaigne, B., van der Werf, S. (2000). Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. J Gen Virol *81*, 1283-1291.
Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J.R., Park, S.Y. (2008). The structural basis for an essential subunit interaction in influenza virus RNA polymerase. Nature *454*, 1127-1131.

Ortiz-Riaño, E., Cheng, B.Y.H., de la Torre, J.C. and Martínez-Sobrido, L. (2011). The C-terminal region of Lymphocytic Choriomeningitis virus nucleoprotein contains distinct and segregable functional domains involved in NP-Z interaction and counteraction of the type I interferon response. J Virol *85*, 13038-13048.

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., Conlan, S., Quan, P. L., Hui, J., Marshall, J., Simons, J. F., Egholm, M., Paddock, C. D., Shieh, W. J., Goldsmith, C. S., Zaki, S. R., Catton, M. and Lipkin, W. I. (2008). A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med *358*, 991–998.

Perez, M., Craven, R. C. and de la Torre, J.C. (2003). The small RING finger protein Z drives arenavirus budding: implications for antiviral strategies. Proc Natl Acad Sci

USA *100*, 12978–12983.

Pestka, S., Krause, C.D., Walter, M.R. (2004). Interferons, interferon-like cytokines, and their receptors. Immunol Rev *202*, 8-32.

Pestka, S. (2007). The interferons: 50 years after their discovery, there is much more to learn. J Biol Chem *282*, 20047-20051.

Poch, O., Sauvaget, I., Delarue, M., Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J *8*, 3867-3874.

Polyak, S. J., Zheng, S. and Harnish, D. G. (1995). 5' termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. J Virol *69*, 3211–3215.
Rey, F.A. (2010). One protein, many functions. Nature *468*, 773-774.

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

Rey, F.A. (2010). One protein, many functions. Nature *468*, 773-775.

Rojek, J.M. and Kunz, S. (2008). Cell entry by human pathogenic arenaviruses. Cell Microbiol *10*, 828-835.

Rodrigo, W.W.S.I., de la Torre, J.C., and Martínez-Sobrido, L. (2011). Use of single-cycle infectious Lymphocytic Choriomeningitis virus to study hemorrhagic fever arenaviruses. J Virol *85*, 1684-1695.

Rubinstein, M., Rubinstein, S., Familletti, P.C., Gross, M.S., Miller, R.S., Waldman, A.A., Pestka, S. (1978). Human leukocyte interferon purified to homogeneity. Science *202*, 1289-1290.

Samji, T. (2009). Influenza A: Understanding the viral like life cycle. Yale J Biol Med *82*, 153-159.

Sanchez, V., Clark, C.L., Yen, J.Y., Dwarakanath, R. and Spector, D.H. (2002). Viable human cytomegalovirus recombinant virus with an internal deletion of the IE2 86 gene affects late stages of viral replication. J Virol *76*, 2973-2989.

Sanchez, A. B. and de la Torre, J.C. (2005). Genetic and biochemical evidence for an oligomeric structure of the functional L polymerase of the prototypic arenavirus lymphocytic choriomeningitis virus. J Virol *79*, 7262-7268.

Schumann, M., Gantke, T., Muhlberger, E. (2009). Ebola virus VP35 antagonizes PKR activity through its C- terminal interferon inhibitory domain. J Virol *83*,8993–8997.

Smallwood, S., Cevik, B. and Moyer, S.A. (2002). Intragenic complementation and oligomerization of the L subunit of the Sendai virus RNA polymerase. Virology *304*, 235-245.

Smallwood, S. and Moyer, S.A. (2004). The L polymerase protein of parainfluenza virus 3 forms an oligomer and can interact with the heterologous Sendai virus L, P and C proteins. Virology *318*, 439-450.

Steitz, T.A. and Steitz, J.A. (1993). A general two metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci USA *90*, 6498-6502.

Stetson, D. B., Ko, J. S., Heidmann, T. and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell *134*, 587–598.

Strecker, T., Eichler, R., ter Meulen, J., Weissenhorn, W., Klenk, H.D., Garten, W. and Lenz, O. (2003). Lassa virus Z protein is a matrix protein sufficient for the release of virus-like particles. J Virol *77*, 10700-10705.

Sugiyama, K., Obayashi, E., Kawaguchi, A., Suzuki, Y., Tame, J.R., Nagata, K. and Park, S.Y. (2009). Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. EMBO J *28*, 1803-1811.

Torpey, N., Maher, S.E., Bothwell, A.L., Pober, J.S. (2004). Interferon alpha but not interleukin 12 activates STAT4 signalling in human vascular endothelial cells. J Biol Chem *279*, 26789-26796.

Ulmanen, I., Broni, B.A. and Krug, R.M. (1981). Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m^7GpppNm) on RNAs and in initiating viral RNA transcription. Proc Natl Acad Sci USA *78*, 7355-7359.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646-651.

Uzé, G., Lutfalla, G., Gresser, I. (1990). Genetic transfer of a functional human interferon α receptor into mouse cells: Cloning and expression of its c-DNA. Cell *60*, 225-234.

Vandevenne, P., Sadzot-Delvaux, C., Piette, J. (2010). Innate immune response and

viral interference strategies developed by human herpesviruses. Biochem Pharmacol *80*, 1955-1972.

Vialat, P., Bouloy, M. (1992). Germiston virus transcriptase require active 40S ribosomal subunit and utilizes capped cellular RNAs. J Virol *66*, 685-693.

Vieth, S., Torda, A.E., Asper, M., Schmitz, H., Günther, S. (2003). Sequence analysis of L RNA of Lassa virus. Virology *318*, 153-168.

Volpon, L., Osborne, M.J., Capul, A.A., de la Torre, J.C., Borden, K.L. (2010). Structural characterization of the Z RING-eIF4E complex reveals a distinct mode of control for eIF4E. Proc Natl Acad Sci U S A *107*, 5441-5446.

Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H.T., Juranek, S., Sheng, G., Micura, R., Tuschl, T., Hartmann, G. and Patel, D.J. (2010). Structural and functional insights into 5′-ppp RNA pattern recognition by the innate immune receptor RIG-I. Nat Struct Mol Biol *17*, 781-787.

Xia, S., Eom, S.H., Konigsberg, W.H., Wang, J. (2012). Structural basis for differential insertion kinetics of dNMPs opposite a difluorotoluene nucleotide residue. Biochemistry *51*, 1476-1485.

Yang, Y.G., Lindahl, T., and Barnes, D.E. (2007). Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. Cell *131*, 873-886.

Yoneyama, M., Suhara, W., Fukuhara, Y., Fukuda, M., Nishida, E. and Fujita, T. (1998). Direct triggering of the type I interferon system by virus infection: activation of a transcription factor complex containing IRF-3 and CBP/p300. EMBO J *17*, 1087-1095.

Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z. & Liu, Y. (2009). Crystal structure of an avian influenza polymerase PA_N reveals an endonuclease active site. Nature *458*, 909-913.

Yumerefendi, H., Tarendeau, F., Mas, P.J., 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.

Zamoto-Niikura, A., Terasaki, K., Ikegami, T., Peters, C.J. and Makino, S. (2009). Rift valley fever virus L protein forms a biologically active oligomer. J Virol *83*, 12779-12789.

Zuo, Y., Zheng H., Wang Y., Chruszcz M., Cymborowski M., Skarina T., Savchenko A., Malhotra A., Minor W. (2007). Crystal structure of RNase T, an exoribonuclease

involved in tRNA maturation and end turnover. Structure *15*, 417–428.

Publication

I. Article

Qi, X., Lan, S., Wang, W., Schelde, L.M., Dong, H., Wallat, G.D., Ly, H., Liang, Y. & Dong, C.J. (2010). Cap binding and immune evasion revealed by Lassa nucleoprotein structure. **Nature** *468*, 779-783.

II. PDB structures

Lassa fever native nucleoprotein structure: 3MWP

NP with Mn^{2+}: 3MWT

NP in complex with dTTP: 3MX2

NP in complex with UTP: 3MX5

Paper

Cap binding and immune evasion revealed by Lassa nucleoprotein structure

Xiaoxuan Qi , Shuiyun Lan , Wenjian Wang , Lisa McLay Schelde , Haohao Dong , Gregor D. Wallat , Hinh Ly , Yuying Liang & Changjiang Dong

NATURE Vol. 468 (2010), pp. 779-783

doi:10.1038/nature09605