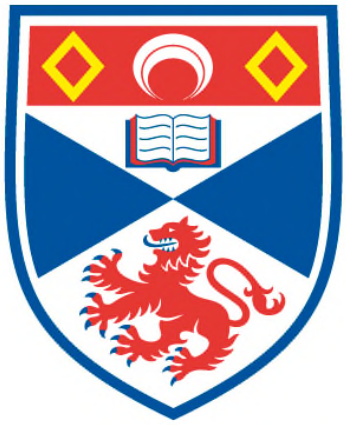
STRUCTURAL BASIS OF LASSA FEVER NUCLEOPROTEIN BINDING PATHOGEN-ASSOCIATED PATTERN MOLECULE dsRNA

Xue Jiang

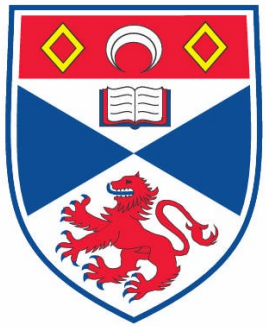
A Thesis Submitted for the Degree of MPhil
at the
University of St Andrews



2012

Structural basis of Lassa fever nucleoprotein binding pathogen-associated pattern molecule dsRNA

Xue Jiang



This thesis is submitted in partial fulfillment for the degree of MPhil

at the School of Chemistry

University of St Andrews

1th October 2012

1. Candidate's declarations:

I, Xue Jiang, hereby certify that this thesis, which is approximately 9000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2010 and as a candidate for the degree of Mphil in February 2012; the higher study for which this is a record was carried out in the University of St Andrews between 2010 and 2012.

Date …… …… …… signature of candidate …… …… …… ……

2. Supervisor's declaration:

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

Date …… …… …… signature of supervisor …… …… …… ……

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am 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. I 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. I have obtained any third-party copyright permission 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:

Embargo on both all of printed copy and electronic copy for the same fixed period of 2 years on the following grounds:

Publication would be commercially damaging to the researcher, or to the supervisor, or the University;

Publication would preclude future publication;

Date …… …… signature of candidate ……

signature of supervisor ……

A supporting statement for a request for an embargo must be included with the submission of the draft copy of the thesis. Where part of a thesis is to be embargoed, please specify the part and the reasons.

Supporting statement for embargo

My project focuses on a novel protein complex structure with dsRNA. Some competitors have published the protein's complex structure with ssRNA but not dsRNA. And the model of my protein-RNA complex could help elucidating the mechanism of its immune suppression function.

Since my result has not been sent out to any journals or publications, and the project is not finished yet, I request an embargo on my entire thesis for 2 years.

Abstract

Lassa fever virus (LASV) infects thousands of people and produces more than 5,000 deaths each year in West Africa. This severe virus is a huge threat, as it transmits between human and rodents, and no effective vaccine or drug is available currently. One key of getting control of this disease lies in the nucleoprotein (NP) of LASV, which plays an essential role in viral replication, transcription and immune suppression. The full length NP crystal structure has been solved, showing a novel structural fold and multi-functions with unusual mechanisms in immune suppression and viral RNA transcription.

The C-terminal domain of LAVS NP is a 3'-5' exonuclease, whose activity is essential for viral immune suppression. This domain alone can suppress an immune response and can degrade dsRNAs with specific preference higher than for ssRNAs. However, the detail of the mechanism is unclear. To understand the mechanism while avoiding another domain's effect (the N-terminal domain), the C-terminal domain of LASV NP was expressed and purified, and pathogen-associated pattern molecular RNAs were synthesized chemically and biologically to carry on crystallization and functional testing. The C-domain crystals in complex with a pathogen-associated pattern molecule, triphosphate 8 nucleotide dsRNA were obtained. The crystal belongs to the space group *P*3 with unit cell dimension $a=b=177.6$ Å, $c=56.49$Å, $\alpha=\beta=90°$, $\gamma=120°$. This crystal structure showed that the dsRNA binds in the 3'-5' exonuclease active site with one 3' end of the dsRNA perfectly sitting for cleavage. We are trying to figure out the detailed mechanism by mutagenesis, fluorescence-labeled RNA gel scan and band shift assays.

Acknowledgment

Thanks very much to my supervisor Dr. Changjiang Dong. It is a great honor to study with your guidance. Thank you for providing me the opportunity to this project. Your patience and profound knowledge are a great support to me.

Thanks to School of Chemistry, University of St Andrews, for the financial support and study opportunity to carry out this research.

Thanks to all the group members in CJD group and JHN group, for all the technical support and routine help.

Thanks to Dr. Timothy Smith in University of Aberdeen, for the supporting on ITC method. And thanks to Dr Janice Bramham in University of Edinburgh, for help with SPR.

Thanks to everyone who works in the BMS building for the mass spectra service and culture medium.

Index

1 INTRODUCTION 1

1.1 LASSA FEVER 1

1.1.1 Background 1

1.1.2 Transmission 1

1.1.3 Pathogenesis 2

1.1.4 Prevention and treatment 2

1.2 LASSA VIRUS 3

1.2.1 Classification 3

1.2.2 Shape 3

1.2.3 Genome 4

1.2.4 Infection mechanism 5

1.2.5 Proteins 7

1.3 NP PROTEIN 7

1.3.1 Background 7

1.3.2 Structure 8

1.3.3 Function 10

2 MATERIALS AND METHODS 13

2.1 EXPRESSION CONSTRUCT 13

2.1.1 Cloning NP C-terminal domain's nucleotides into pLOU3 (His-tag-MBP attached) 13

2.1.2 Cloning NP C-terminal domain into pHisTEV plasmid 13

2.1.3 mutagenesis of pHisTEV NP C-terminal domain. 13

2.2 PROTEIN EXPRESSION AND PURIFICATION 14

2.2.1 for the NP C-terminal domain (His-tag-MBP attached) 14

2.2.2 For the His-tagged NP C-terminal domain (native and mutations) 15

2.3 SDS-PAGE GEL 16

2.4 TRIPHOSPHATE-DSRNA CONSTRUCTION 16

2.4.1 Template DNA preparations: 16

2.4.2 Transcription reaction 16

2.4.3 RNA purification 17

2.4.4 dsRNA preparation 17

2.5 CRYSTALLIZATION 17

2.6 X-RAY CRYSTAL DATA COLLECTION 18

2.7 STRUCTURE DETERMINATION, REFINEMENT AND MODEL BUILDING 18

3 RESULT AND DISCUSSION 19

3.1 PROTEIN PURIFICATION 19
3.1.1 His-tag-MBP fusion protein 19
3.1.2 His-tagged NP C-terminal protein 22
3.2 CRYSTALLIZATION 24
3.3 CRYSTAL STRUCTURE 26
3.4 MUTAGENESIS AND MUTANT PROTEIN PURIFICATION 29
3.4.1 pHisTEV NP C-terminal domain plasmid construction 29
3.4.2 single site plasmid mutagenesis 30
3.4.3 expression and purification 31
4 FUTURE WORK 32
5 NP N-TERMINAL DOMAIN 33
5.1 MATERIAL AND METHODS 33
5.1.1 expression construct 33
5.1.2 protein expression and purification 33
5.1.3 Crystallization 34
5.2 RESULT 34
5.3 FUTURE WORK 36
6 APPENDIX I 37
7 APPENDIX II 38
8 APPENDIX III 39
9 APPENDIX IV 40
10 APPENDIX V 41
11 REFERENCES 42

1 INTRODUCTION

1.1 LASSA FEVER

1.1.1 BACKGROUND

Lassa fever is a hemorrhagic illness caused by Lassa virus. It is first described in the town of Lassa, Nigeria in 1969, when two missionary nurses died there. It spread endemically across West Africa, mainly in Nigeria, Guinea, Liberia and Sierra Leone. Lassa fever cases are also been reported in Europe, America and Asia, due to people travelling to epidemic areas. (Ogbu O*, et al.*, 2007)

Lassa virus (LAV) has high mortality and no effective vaccine. It causes 300,000 to 500,000 infections and 5000 deaths yearly, (Walter HH, *et al.,* 2003) with 15% to 20% hospitalized patients' death and 50% fatality during epidemics. The Center of Disease Control and Prevention of USA classified Lassa fever virus as a biosafety level 4 agent, for it is airborne transmission. (Illick MM*, et al.,* 2003)

1.1.2 TRANSMISSION

The wide distribution of Lassa fever is mainly caused by infected Mastomys rodents. They can instigate the virus' spreading by direct contacting human, contaminating foods with their secretions, or spreading tiny virus particles from excretions in the air which might inhaled by people. Since rodent distributes very close to human living range and lives undetectably, Lassa fever transmission becomes hard to control.

Fig 1: The Lassa fever virus host, Mastomys rodent.
(Picture from: viral hemorrhagic fever consortium website)

Lassa fever can also transmit from person to person, via blood, tissue, secretions, or excretions of an individual infected with the Lassa virus. However, casual contact, such as skin to skin without exchange of body fluids might not spread the disease. On the other hand, both men and women, young and old can be infected. And the pregnant women and infants are among the most risky group. (The center of Disease Control and Prevention, December 3, 2004)

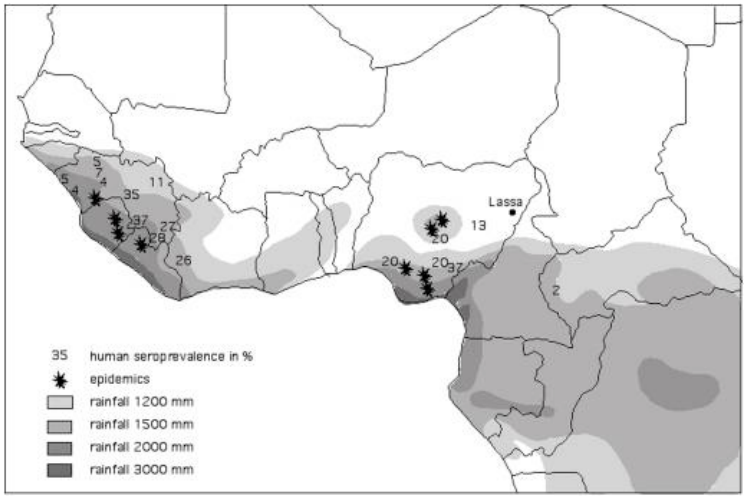


Fig 2: Risk Map of Lassa fever in West Africa (Picture from Fichet E, Rogers DJ. 2009)

1.1.3 PATHOGENESIS

The incubation period of Lassa fever ranges from 6 to 21 days, and the onset duration is 1to 4 weeks. This acute illness will harm several organs in the body, such as the liver, spleen and kidneys. The disease usually begins with fever, general weakness, and malaise, following by headache, sore throat, muscle pain, chest pain, nausea, vomiting, diarrhea, cough, and abdominal pain in a few days later. In extreme situations, patients may suffer facial swelling, fluid in the lung cavity, bleeding from mouth, nose, vagina or gastrointestinal tract, and low blood pressure. Fatal cases always end up in death within 14 days.

1.1.4 PREVENTION AND TREATMENT

Lassa fever has many kinds of unspecific symptoms, which make it hard to be diagnosed out from other diseases during the first stage. Malaria, shigellosis, typhoid fever, yellow fever and other viral hemorrhagic fevers may have similar signs as Lassa fever. Therefore, the most often detection method is using enzyme-linked immunosorbent serologic assays (ELISA), which detect IgM and IgG antibodies as well as Lassa antigen.

At an early stage, giving ribavirin, a general antiviral drug, could be effective. But there is no evidence to show ribavirin could be a prophylactic treatment. When the disease becomes severe, ribavirin is no longer efficient. General prevention is by trying all means to eliminate rodents from the human community. (World Health Organization, Fact sheet N°179)

1.2 LASSA VIRUS

1.2.1 CLASSIFICATION

Lassa virus belongs to old world arenaviruse, the Arenaviridae family. The Arenaviridae family contains many life threatening members such as the hemorrhagic fever viruses Machupo, Junin, and the Lymphocytic Choriomeningitis virus (LCMV). Arenaviruses are divided into two groups, the Old World and New World, based on their geographical distribution and genetic differences. The Old World is found in the Eastern Hemisphere in places such as Europe, Asia, and Africa. While New World is found in the Western Hemisphere, in places such as Argentina, Bolivia, Venezuela, Brazil, and the United States. (Buchmeier MJ*, et al*., 2007)

1.2.2 SHAPE

Like all other Arenaviruses, Lassa virus is enveloped and has a bisegmented single-strained RNA with a unique ambisense genomic organization, and is coated by a nucleoprotein to form a nucleocapsid. Under cryoelectron microscopy, the virion looks like a grainy particle, due to its beaded nucleocapsid. That is how arenavirus gets its name, arena in the Latin root meaning sand.

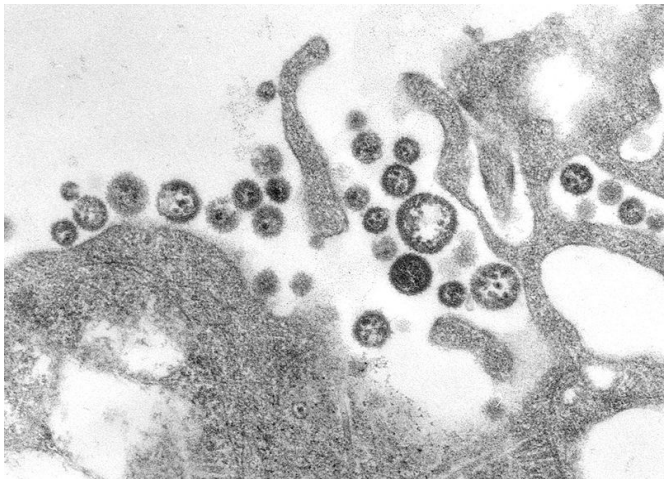


Fig 2: Transparent Electrical Microscopy micrograph of Lassa virus virions. (Picture from Ogbu O, *et al*., 2007)

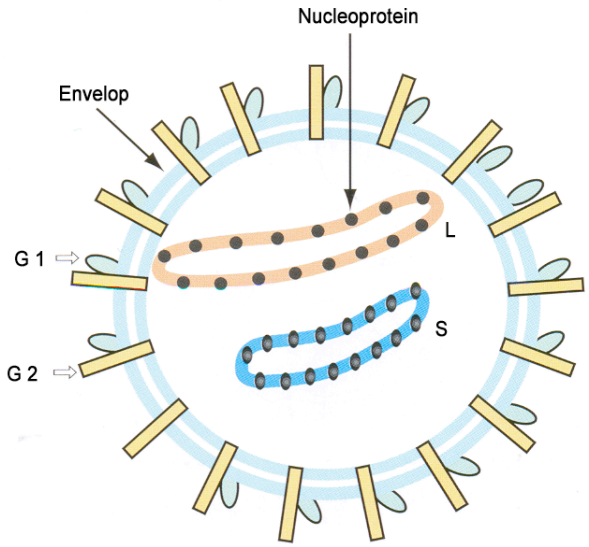


Fig 3: Lassa virus structure (Image taken from Swiss Institute of Bioinformatics, http://viralzone.expasy.org/viralzone/all_by_species/501.html)

1.2.3 GENOME

Including Lassa virus, Arenaviruses' two genomic RNA segments are large RNA (L-RNA, 7.2kb) and small RNA (S-RNA, 3.4kb), and they encode four known proteins: a nucleoprotein NP (60 kDa), a matrix protein Z (~11kDa), a RNA-dependent polymerase L (~200 kDa), and a glycoprotein GP. Their ambisense coding strategy allows the polypeptides to be synthesized in two opposite directions, with the help of a noncoding intergenic region (IGR) folding into a hairpin structure.

The small RNA encodes NP and immature GP precursor. While large RNA is responsible for L and Z. In the S segment, the NP coding region is transcribed from viral sense strand 3' to 5' into genomic complementary mRNA, while the GP coding region is transcribed from 5' to 3' into a genomic sense mRNA. The same rule applies to the L gene. The L protein's mRNA is genomic complimentary and Z protein's is genomic sense. This is the ambisense-coding stratagem that helps viral mRNAs being transcribed without overlapping. As a result of this arrangement, only after viral genomic RNA can the replication begin, opening reading frames of GPC and Z being transcribed.

For precursor GP (82kDa), during the translation in the endoplasmic reticulum, a signal peptidase cleaves it into a signal peptide and GP-C (76kDa). After translation, subtilase SKI-1/S1P cleaves this GP-C into N-terminal subunit GP-1 (40 kDa) and membrane-bound subunit GP-2 (36 kDa). (Michael DB*, et al*., 2000)

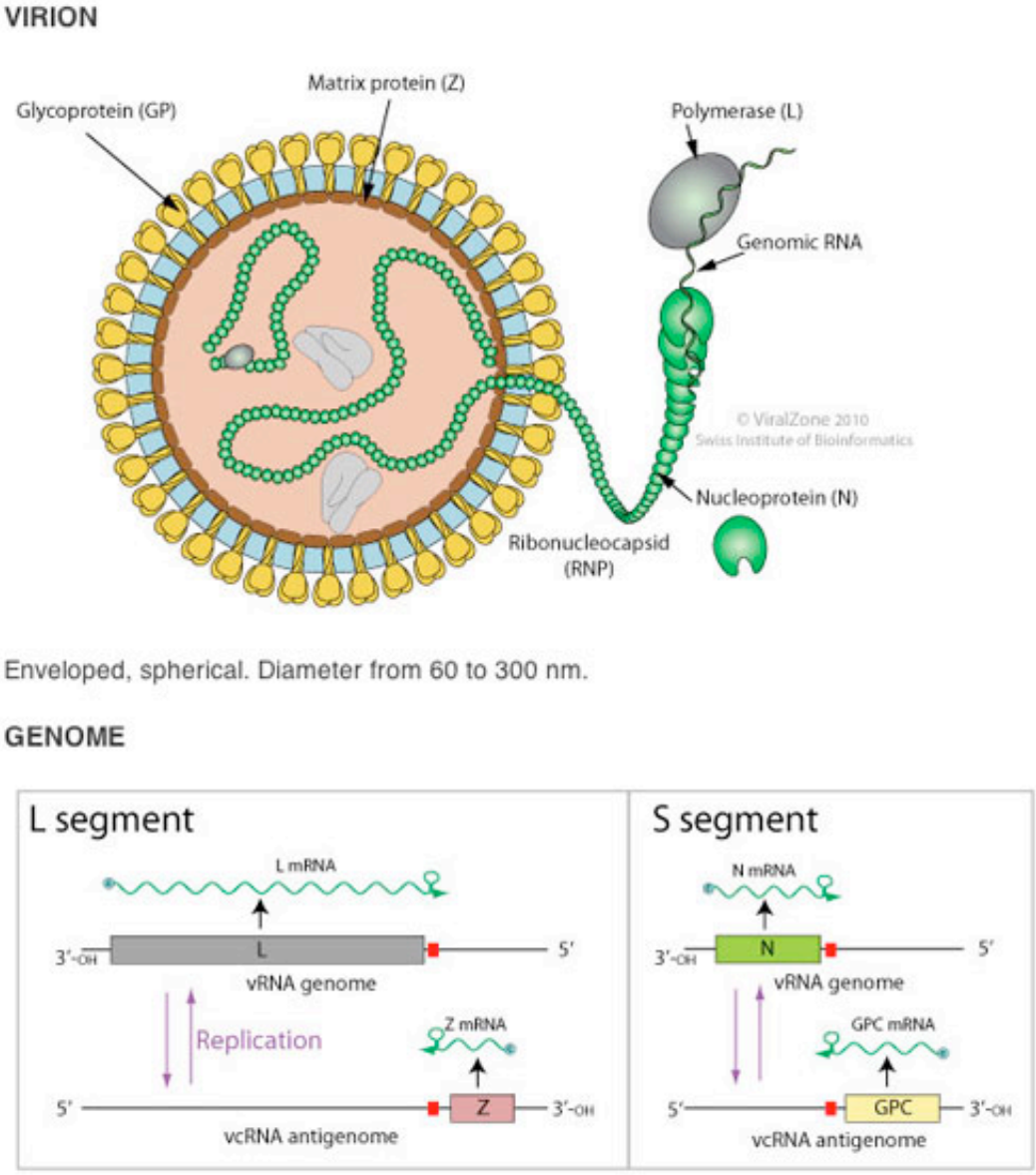


Fig 4: Lassa fever virus' virion and genome structure.
(Image taken from Swiss Institute of Bioinformatics, http://viralzone.expasy.org/all_by_species/83.html)

1.2.4 INFECTION MECHANISM

Lassa virus infects hosts mainly by contacting, but how virions break through the epithelial barrier for initializing host infection is still unclear. When it reaches the host cell, the α-dystroglycan (α-DG), a multipurpose receptor for the extracellular matrix protein, becomes the essential receptor and helps Lassa virus' endocytosis to enter the cell. The GP protein plays the role for receptor recognition. GP1 (N-terminal) locates at the glycoprotein spike to bind receptor α-DG on the cell surface. GP2 (C-terminal) binds to membrane and is responsible for transmembrane fusion. When the pH is under 5 (the ideal value), membrane fusion happens,

followed by the viral nucleocapsid entering a late endosomal partition via vesicular trafficking. This process is independent from molecular motors like clathrin, caveolin, dynamin or actin. However, some cholesterol from the membrane is required. The unusual endocytosis pathway allows Lassa virus particles being delivered to late endosomes more rapidly, thus makes infection more effective. (Schlie K*, et al.,* 2010)

Old world arenavirus LASV and LCMV share the same cell-entering mechanism. However, new world arenaviruses recognize different cell receptors and need molecular motors when entering a cell, and depend on Rab5 for delivering into early endosome, and require Rab7 as well for transmitting to late endosome (Fig 5). (Kunz S. 2009)

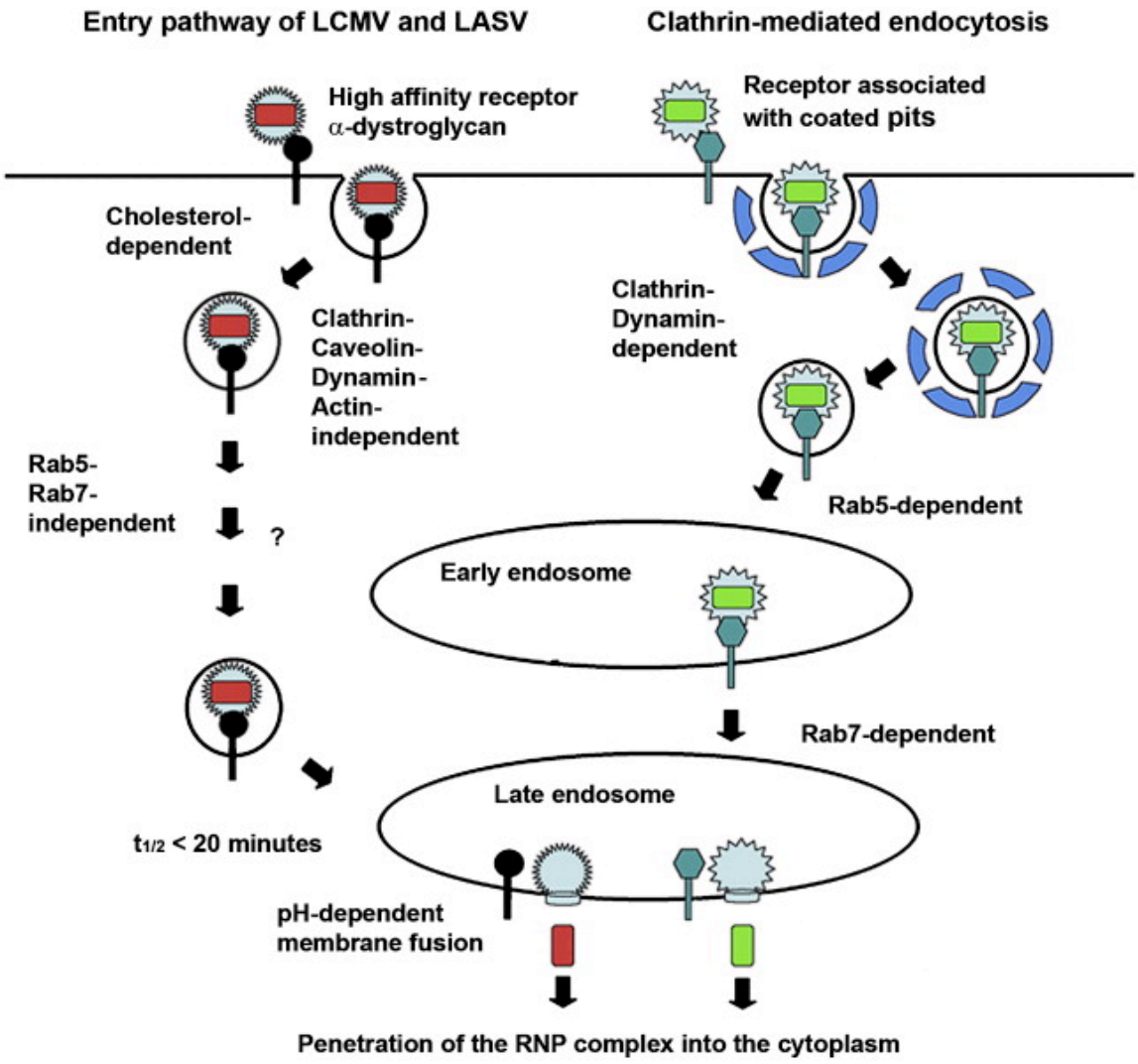


Fig 5: Different receptor binding and cell entry mechanisms between Old World and New World arenaviruses (Picture taken from Kunz S. 2009)

1.2.5 PROTEINS

Glycoprotein GP

The GP protein is encoded by the S-segment RNA. Its function is to recognize and bind to the receptor of host cell surface and fuse into the cell. The process has been introduced above.

Ring finger protein Z

The Z protein is a matrix protein, which plays essential roles in virus assembly and budding. Expressing only Z without other viral proteins can be enough to form and release the enveloped Z-containing particles, which is not notably different from Lassa virus particles in morphology and size. Z acts as the driving force during virus particle releasing. Z also has strong association with the membrane. In addition, Z is assumed to interact with NP during viral assembling. (Shtanko O, *et al.,* 2010)

RNA-dependent polymerase L

L protein is a putative RNA-dependent RNA polymerase. It controls the synthesis of mRNA terminating in the intergenic region, and noncapped genomic or antigenic RNA forming a full-length genome copy. (Lelke M*, et al.,* 2010)

The L protein consists of 2300 amino acids, which can be divided into three domains, N-terminal domain, central domain and C-terminal domain. It is believed to harbor several enzymatic functions in the N and C terminals, but this is yet unproven. Cap-snatching is supposed to be one of the L protein's significant functions. With NP, L protein can form the minimal *trans*-acting factors in genome replication and replication. Both L and NP N-terminal domain are believed to cooperate with each other in a cap-snatching mechanism. The central domain (residues 1000-1500) of L is regarded as the RNA-dependent RNA polymerase. (Lelke M*, et al*., 2010)

1.3 NP PROTEIN

1.3.1 BACKGROUND

NP is the most abundant viral protein in an LASV infected cell. NP associates with RNA to form the ribonucleoprotein (RNP) core protecting the genomic RNA and the nucleocapsid serves as a template for L

protein for RNA replication and transcription. NP also interacts with the matrix protein Z during viral assembly. (Eichler R, *et al*., 2004)

However, the most interesting function of NP is to suppress the host cell immune responses. Moreover, LAVS NP structure is the only NP structure reported among Arenaviridae family.

In human cells, there are several receptors, which play essential roles in sensing infections by detecting pathogen-associated pattern molecules (PAPM) and trigger immune response pathways. It is well known that cytoplasm receptors retinoic acid-inducible I receptor and melanoma differentiation-associated 5(MDA-5), and membrane toll-like receptors can detect PAMP molecules. PAMP, such as triphosphate dsRNAs, long chain dsRNAs or short chain dsRNAs, trigger immune response pathways to produce interferon, when virus infections occur. Virus infection usually produces dsRNA in the infected cell. Once retinoic acid-inducible I (RIG-I), melanoma differentiation-associated 5 (MDA-5) or other cellular immune receptors detect dsRNA, signaling will trigger IFNs production. LAVS NP has been demonstrated to play an essential role in immune suppression, and its 3'-5' exonuclease activity has been shown to be crucial for the function. Our hypothesis is that the LAVS NP can specifically recognize and degrade the PAMP RNA molecules generated from the virus infection. Therefore the receptors cannot detect the infection, and no immune response occurs. (Kathryn MH, *et al.,* 2011)

1.3.2 STRUCTURE

The NP coated viral genomic RNA to formed nucleocapsid, a beads-like viral particle under microscopy. The LAVS NP consists of two parts: amino (N) terminal and carboxy (C) terminal domains. A positively charged groove locates between the two domains and the viral genomic RNAs are expected to bind in the groove. All the structures of published nucleoprotein structures among all negative-stranded RNA viruses share the same organization.

In each NP protomer, a Zn^{2+} locates in the C-terminal domain, forming a zinc finger structure. As a 3'-5' exonuclease, one Mn^{2+} molecule was identified in the C-terminal domain active site. In the N-terminal domain, the electrostatic surface potential map indicates a RNA-binding cavity entrance.

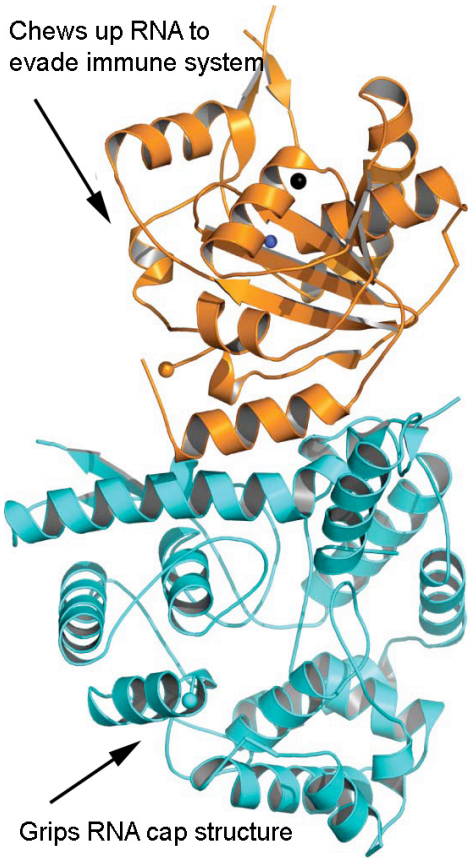


Fig 6: NP monomer structure

Yellow region represents N-domain. Blue region represents C-domain, PDB code 3MWP.

(Picture taken from Qi X, *et al.,* 2010)

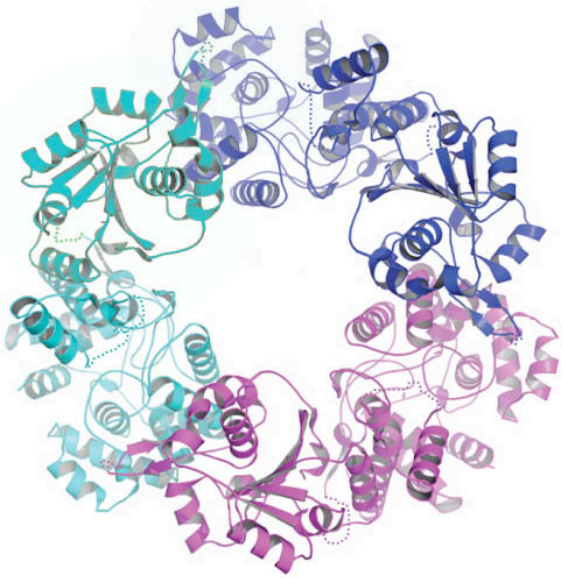


Fig 7: The ring shape form of the NP trimer

Navy, sapphire and magenta regions show three identical molecules forming a trimer, PDB code 3MWP.

(Picture taken from Qi X, *et al.,* 2010)

1.3.3 FUNCTION

The N-terminal domain of LAVS NP forms a deep cavity for cap binding. The 5' terminal m^7G cap is presented on most eukaryotic mRNAs. For most RNAs, elimination of the cap structure causes a loss of stability, especially against exonuclease degradation, and a decrease in the formation of the initiation complex of mRNAs for protein synthesis. In addition, a cap requirement has been observed for splicing eukaryotic substrate RNAs. Only mRNAs with a 'cap' structure at its 5' end can be translated into protein by the ribosome. But arenavirus cannot produce cap from their own viral mRNA, therefore Lassa virus has to snatch cellular mRNAs' caps. During the cap-snatching, L protein's N terminal domain is an endonuclease that cleaves off the 5' end of cellular mRNAs, while the NP's N terminal domain holds the 5' end cap. The holding function has not been reported in any equivalent protein among any other virus family. (Qi X*, et al*., 2010)

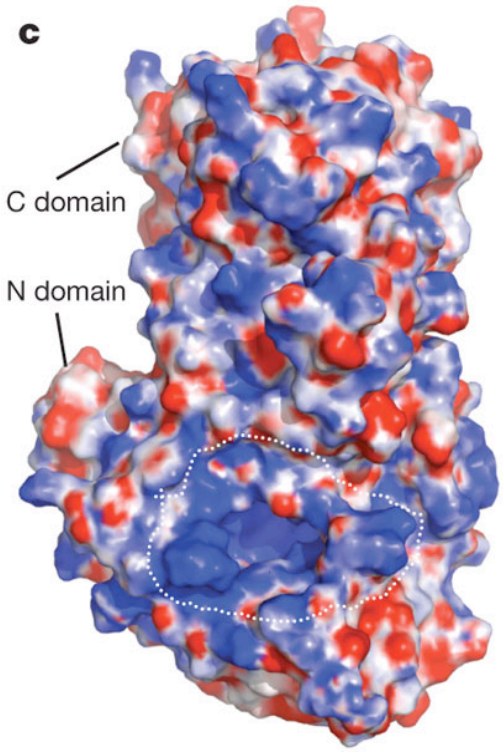


Fig 8: The electrostatic surface potential map of NP. The dashed white line represents the cap-binding cavity. Electropositive residues are blue, electronegative residues are red. (Picture taken from Qi X, *et al.,* 2010)

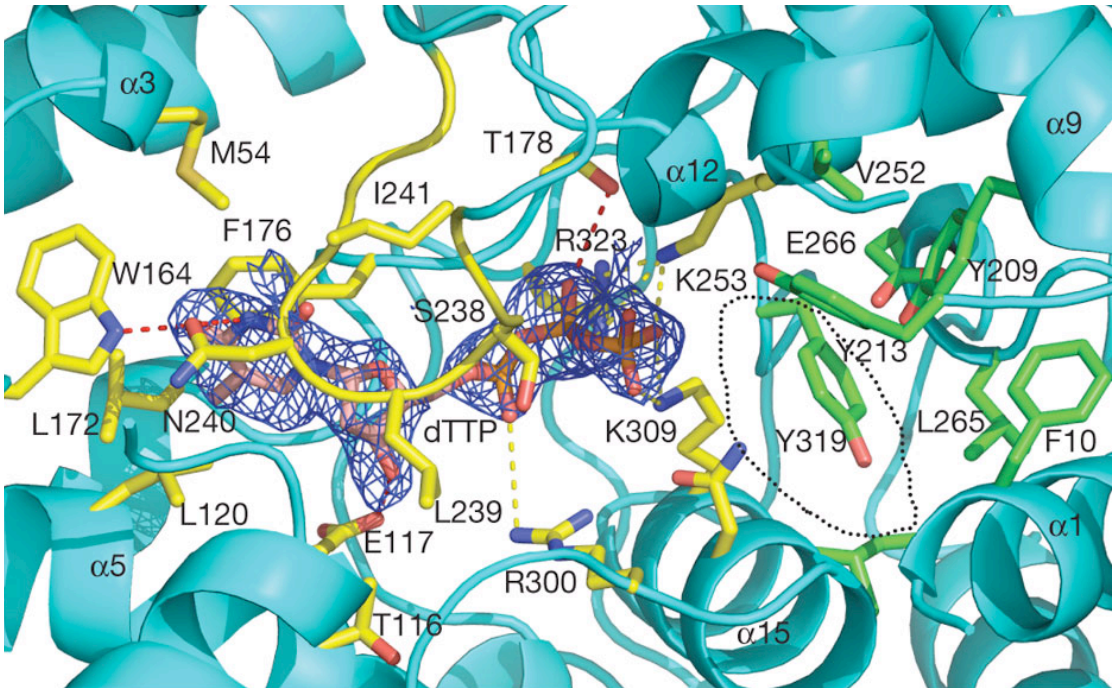


Fig 9: A cap analogue molecule dTTP located in the cap-binding site. The initial FoFc difference density map for dTTP is in blue, with pink atoms representing carbon. Yellow residues are for the deep cavity site. Green residues are the pass for another mRNA to enter, PDB code 3MX2. (Picture taken from Qi X, *et al.,* 2010)

The C terminal domain structure is similar to 3'-5' exonucleases, especially the human DNA 3'-5' exonuclease enzyme TREX1. Experiments also show the C terminal domain alone can digest RNA substrates and suppresses immune responses, which means that the C-terminal domain is the functional part for the exonucleases activity.

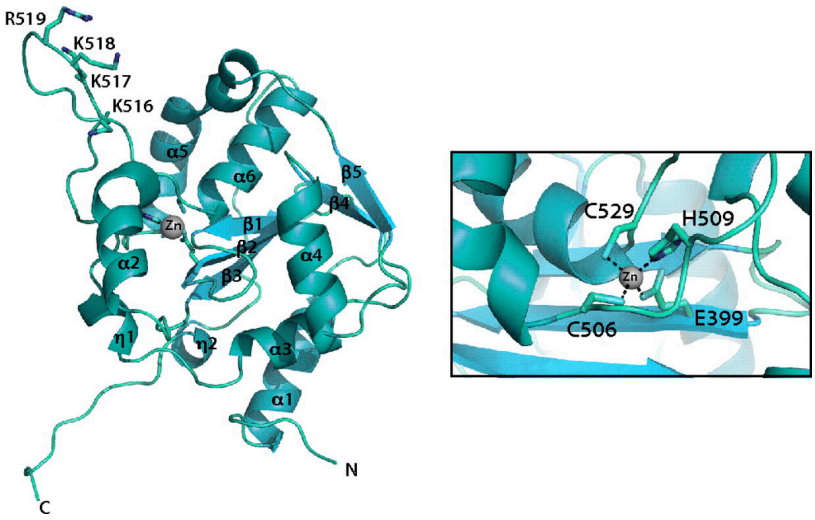


Fig 10: Structure of LAVS NP C-terminal domain. Right side shows Zn^{2+} and its coordination, PDB code 3Q7B. (Modified from Kathryn MH, *et al.,* 2011)

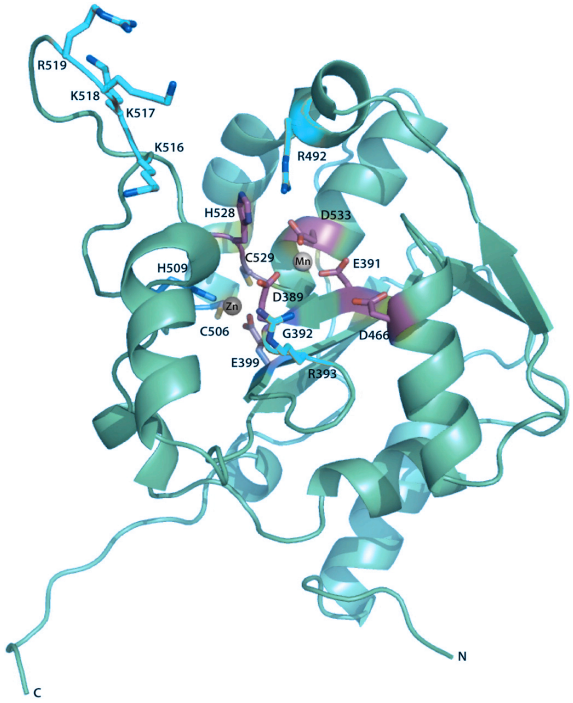


Fig 11: NP C-terminal domain with Mn^{2+}, showing the essential residues for exonuclease activity. Purple colored residues match to DEDDh exonuclease, PDB code 3Q7C (Modified from Kathryn MH, *et al*., 2011)

However, how NP binds and degrades PAMP molecules, such as dsRNA is not exactly clear, for no complex crystal has been reported. Therefore, my work focuses on expression, purification and crystallization of LAVS NP C-terminal domain in complex with PAMP molecules, figuring out the novel mechanism.

I have built an expression construct of the native NP C-domain and carried out mutagenesis on designed sites. Protein expression and purification were successful. I've set up crystallization trials and get nice crystal complex of native NP C-domain with RNA ligand. The crystals have been sent to Diamond IO3 for X-ray diffraction, and the structure of the complex is solved.

2 MATERIALS AND METHODS

2.1 EXPRESSION CONSTRUCT

2.1.1 CLONING NP C-TERMINAL DOMAIN'S NUCLEOTIDES INTO PLOU3 (HIS-TAG-MBP ATTACHED)

The nucleotides encoding the LAVS NP C-terminal domain from residue 364 to 569 was cloned into the pMAL-c2X-derived plasmid pLou3, with a 6His-tag at N terminus of MBP (maltose binding protein) and a TEV cleavage site between the MBP and the target protein. This construct was transformed into Rosetta cells (Novagen). The transformed cells have antibiotic resistance against ampicillin and chloramphenicol. Xiaoxuan Qi previously built this construct in our group.

2.1.2 CLONING THE NP C-TERMINAL DOMAIN INTO PHISTEV PLASMID

The nucleotides encoding the NP C-terminal domain were also inserted into pHisTEV plasmid. The forward and reverse primers containing the NCOI and HindIII enzyme cutting sites were synthesized in Eurogentec. The C-terminal nucleotides were amplified by PCR using a program of heating at 95°C for 2 mins, annealing at 58°C for 1 min and extending at 72°C for 2 min for 32 cycles. The pHisTEV plasmid was cut by restriction enzymes NCOI and HindIII. The NP C-terminal domain PCR product was then inserted into the pHisTEV plasmid with the help of T4 DNA ligase (10 μl reaction system in 37°C for 2 hour. See appendix Ⅳ).

This reconstructed plasmid was transformed into TAM1 cells for producing and harvesting more plasmids (transform method and plasmid extraction with miniprep method see appendix Ⅴ). Enzyme digestion was used to confirm the correct insertion (NCOI and HindIII cutting reconstructed plasmid). Afterwards, the correct plasmid was transferred into expression system Rosetta cells.

2.1.3 MUTAGENESIS OF PHISTEV NP C-TERMINAL DOMAIN.

DNA primers were designed based on QuikChange™ Site-Directed Mutagenesis and Huangting Liu's protocol, and were synthesized at Eurogenetec. (Details given in appendix I)

The mutant plasmids were generated using the primers and Promega Pfu DNA Polymerase. Five units of Dpn I were used to remove the original template plasmids by incubation at 37°C for 2 hrs.

Agarose gel electrophoresis was used to separate the C-terminal mutant plasmid DNA band from others. One micro liter of each mutant plasmid was transformed into TAM1 cells, and the transformed cells were spread on LB agar gel containing 100μg/ml ampicillin. Two colonies from each mutant were picked and each colony was inoculated into 10 ml of LB containing 100μg/ml ampicillin and cultured overnight. The mutant plasmids were extracted with Qiagen miniprep kit, and 5μl of each plasmid was sent to University of Dundee for sequencing. The correct mutant genes were transformed into Rosetta cells for expression.

2.2 PROTEIN EXPRESSION AND PURIFICATION

2.2.1 FOR THE NP C-TERMINAL DOMAIN (HIS-TAG-MBP ATTACHED)

The single colony of the Rosetta transformed cells was inoculated into 500ml LB containing 50μg/ml ampicillin and 34 μg/ml chloramphenicol and was cultured overnight in incubator at 200 rpm in 37°C. This overnight culture was then subcultured to 10 liters of LB (containing 50μl/ml ampicillin and 34 μg/ml chloramphenicol) at 200 rpm at 37°C. When OD_{600} reached 0.6, the cells were induced with 0.03mM final concentration of IPTG (isopropyl-beta-d-thiogalactopyranoside). And the protein expressions were induced at 20°C for around 20 hours.

The cells were harvested by centrifugation at 10000g for 15 mins. The Cell pellet was resuspended in loading buffer (20mM Tris-HCl, pH7.5, 10mM imidazole, 300mM NaCl, 10% glycerol), 1 tablet of EDTA-free protease inhibitor (Roch), 1μM DNase (Sigma), 1μM lysosome (Sigma) and 1mM PMSF (phenylemethylsulfonyl fluoride, Sigma). Cells had gone through cell disrupter twice for an adequate disruption. The debris is removed by centrifugation at 40000g for 30 mins. Supernatant was collected and applied to 10 ml Ni-NTA agarose (Qiagen) beads, which was pre-equilibrated with the loading buffer. The beads were washed with 40ml of loading buffer (however, even 30mM imidazole wash buffer would wash away some target protein), and were eluted with 22ml elution buffer consisting of 20mM Tris-HCl, pH 7.5, 500mM imidazole, 300mM NaCl and10% glycerol.

Then the protein buffer was changed to a buffer containing 20mM Tris-HCl, pH 7.5, 300mM NaCl and 10% glycerol, by going through a desalting column (Hiprep 26/10, GE). The His-MBP-NPC fusion protein was cleaved by TEV proteinase at room temperature for around 18 hours. Afterwards, the cleavage sample was applied to Ni-NTA agarose beads (pre-equilibrated with desalt buffer). The His-MBP was removed from the sample by Ni-NTA, while the target protein went through the beads column.

The sample was concentrated to 7.5ml for gel filtration. The gel filtration column was pre-equilibrated with GF buffer (20mM Tris-HCl, pH7.5, 300mM NaCl, 10% glycerol for high salt concentration condition, or 20mM Tris-HCl, pH7.5, 100mM NaCl, 10% glycerol for low salt concentration). The fractions containing the NP C-terminal domain were pooled. The protein was concentrated to 10mg/ml and frozen in liquid nitrogen, then stored at -80°C.

2.2.2 For the His-tagged NP C-terminal domain (native and mutations)

The His-tagged NP C-terminal domain was expressed in Rosetta *E. coli* cells. The overnight cultures grew in incubator at 37 °C and 200rpm in 500 mL LB (supplemented with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol). The overnight culture was inoculated into 10 L LB containing 50 μg/ml kanamycin, 34 μg/ml chloramphenicol and 100 μM $ZnCl_2$. When OD_{600} reached around 0.4, the protein was induced with 500 μM IPTG for around 20 hours at 25 °C and 200 rpm.

The cells were harvested by centrifugation at 10000g for 15 min. The cell pellet was resuspended in loading buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole), 1 tablet of EDTA-free protease inhibitor (Roch), 1μM DNase (Sigma), 1μM lysosome and 1mM phenylemethylsulfonyl fluoride (PMSF, Sigma). Cells had gone through cell disrupter twice for an adequate disruption. The debris is removed by centrifugation at 40000g for 30mins.

The supernatant was pooled and applied to 10 ml Ni-NTA agarose (Qiagen) beads (pre-equilibrated with loading buffer). The Ni-NTA beads were washed three times with 50 mL of wash buffer (50 mM Tris-HCl, pH 8.5, 300 mM NaCl and 30 mM imidazole), and the protein was eluted with 22ml elution buffer (50mM Tris-HCl,

pH 7.5, 500mM imidazole, 300mM NaCl). The protein was concentrated and purified by gel filtration using a Superdex 200 (GE Healthcare) column, which was equilibrated with GF buffer (20 mM Tris-HCl, pH 7.5, 300 mM NaCl and 10% glycerol). The NP C-terminal domain samples were pooled, concentrated and stored at -80 °C for binding future affinity trials.

2.3 SDS-PAGE GEL

SDS-PAGE was used for checking protein purity in specific fractions after gel filtration. The gel was pre-cast NuPAGE 4-12% Bis-Tris gel, and the power supply was Powerease 500. A15µl sample was mixed with 5µl of SDS loading buffer and incubated at 95 °C for 5 min. Protein molecular standard (Mark12) was used to indicate protein molecular weight.

2.4 TRIPHOSPHATE-DSRNA CONSTRUCTION

The triphosphate dsRNA was produced with MEGAshortscript Kit according to the manufacture's manual. The DNA templates for RNA were synthesized by Eurogentec.

2.4.1 TEMPLATE DNA PREPARATIONS:

The 50µl 100µM T7poly_sense_common and the 50 µl 100µM T7_anti_polyGC8 were mixed together, incubated at 95°C for 3min, and then cooled down to room temperature.

2.4.2 TRANSCRIPTION REACTION

The reaction takes place in a sterilized 1.5ml eppendorf tube. The 20µl reaction contains 2 µl T7 10X reaction buffer, 2µl T7 ATP solution (75 mM), 2µl T7 CTP solution (75 mM), 2µl T7 GTP solution (75 mM), 2µl T7 UTP solution (75 mM), 2µl T7 enzyme Mix, 3µl template DNA and 5µL water (nuclease-free).

The 20µl reaction system was mixed thoroughly, and incubated at 37°C overnight.

In order to degrade the template DNA, 1µl of TURBO DNase was added to the system and mixed well, with incubation at 37°C for more than 1 hour.

This reaction is terminated with 115µl nuclease-free water and 15µl ammonium acetate stop solution.

2.4.3 RNA PURIFICATION

The RNAs were extracted with an equal volume (151µl) of phenol/chloroform (water-saturated), and then with an equal volume (151 µl) of chloroform. The aqueous phase was collected into a new tube. The RNAs were precipitated by adding 2 volumes of ethanol (approximately 600µl) and mixing well and keeping at –20°C for least 15 minutes.

The RNA was pelleted by centrifugation at 4°C for 15 minutes at maximum speed (≥10,000 x g), and the supernatant was carefully removed. The RNA pellet was washed with 70% ethanol at -20 °C and dried at room temperature.

2.4.4 DSRNA PREPARATION

The synthesized RNA was dissolved in TEA buffer (10 mM Tris-HCl, pH 8, 1mM EDTA, 0.1M NaCl), around 10µl for 5 20µl-scale reactions. The RNA sample was heated at 95 °C for 3min and annealed at room temperature, forming triphosphate double strand RNA. The RNA samples were quantified by Nano drop and store at -80°C for crystallization and assays.

2.5 CRYSTALLIZATION

Crystallization conditions were screened on a SWISSCI 'MRC' 2-Well Crystallization Plates (Douglas) with the sitting drop vapor diffusion method set by the Honeybee system. Each drop contained 0.15µl protein-RNA complex and 0.15µl buffer, while 85µl buffer was loaded in each reservoir. The crystallization screens are stochastic screens, made in JHN laboratory. The NP C-terminal domain RNA complex was formed by mixing protein and dsRNA at 1:1 molar ratio (approximately 100µl 10mg/ml protein with 4.2µl 10mM dsRNA). Initial crystals were obtained in 18.5% (w/v) PEG MME 5000, 0.1M Na-citrate, pH 4.5, 2.4% (w/v) PEG MME350 at room temperature for 2 weeks. Crystallization optimization was carried out with varying the concentration of PEG MME 5000 and PEG MME350 and pH. The best crystals grew in18%-19.5% (w/v) PEG MME 5000, 0.1M Na-citrate, pH 4.2-4.5 and 2.4% (w/v) PEG MME350.

2.6 X-RAY CRYSTAL DATA COLLECTION

The crystals were frozen in liquid nitrogen using paraffin oil as a cryoprotectant. To obtain the NP C-terminal domain in complex with dsRNA and Mn^{2+}, the crystals were soaked in a solution containing additional 100 mM $MnCl_2$ and 20% glycerol for 1 min or 1.5 min or 5 min or 10 min before being frozen.

The diffraction data were collected at the Diamond Light Source UK. Crystals were screened at beamline IO3 with Pilatus 6M-F detector. Some of the good crystals were selected for data collection. Approximately 800 frames have been recorded for each crystal. The parameters used for data collection for NP C-terminal domain in complex with dsRNA are listed in Table 1.

Table 1: Data collection parameters for a crystal of NP C-terminal domain in complex with dsRNA

Wavelength (constant)	0.9762 Å
Energy	12.701 keV
Omega start	144.0 °
Oscillation range	0.1°
Exposure Time	0.1 s
Detector Distance	391.4 mm
Resolution at edge	2.0 Å
Xbeam	1222.62 mm
Ybeam	1264.49 mm

2.7 STRUCTURE DETERMINATION, REFINEMENT AND MODEL BUILDING

All the data were indexed and integrated using Mosflm (Geoff GB, *et al*., 2011), and the data were scaled using Scala (Evans P, 2006). The structure was determined by molecular replacement using Phaser (McCoy AJ. 2007). The initial search model was NP C-terminal domain (PDB code 3MWP). The models were built using Coot (Emsley P, *et al.*, 2010), and the refinements were carried out using REFMAC5 (Garib NM, *et al.,* 2007).

3 RESULT AND DISCUSSION

3.1 PROTEIN PURIFICATION

3.1.1 HIS-TAG-MBP FUSION PROTEIN

The NP C-terminal domain was successfully expressed as a His-tagged MBP fusion protein. The entire fusion protein was purified using 10 ml of nickel beads. After washing with 30 mM imidazole, we could obtain relatively pure fusion protein using 22 ml of the elution buffer as shown in Fig 12. The His-tagged MBP is cleaved off by TEV proteinase. To find the best time for cleavage, we did a time course and found that 15 hours' TEV cleavage is insufficient. We therefore always tried to cleave the MBP for 18 to 20 hours.

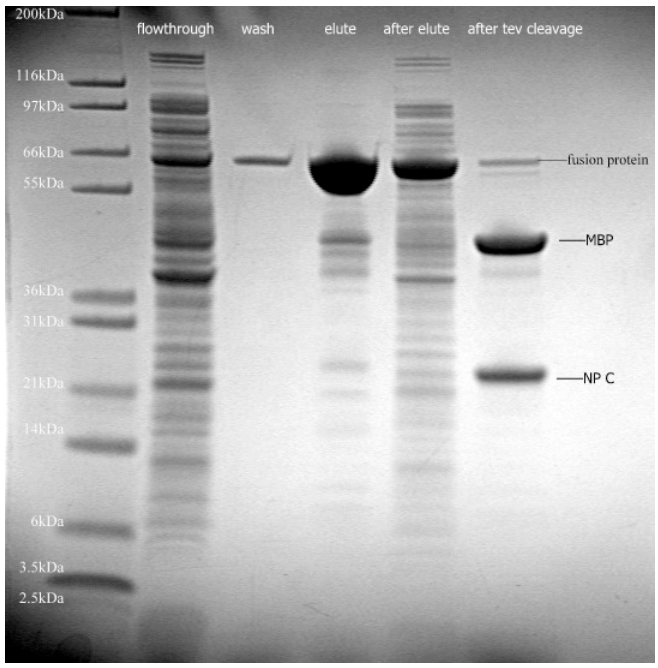


Fig 12: Purification of the MBP and N-terminal domain fusion protein and TEV proteinase cleavage.

However, gel filtration cannot separate MBP and NP C-terminal domain because their molecular weight is similar (Fig 13)

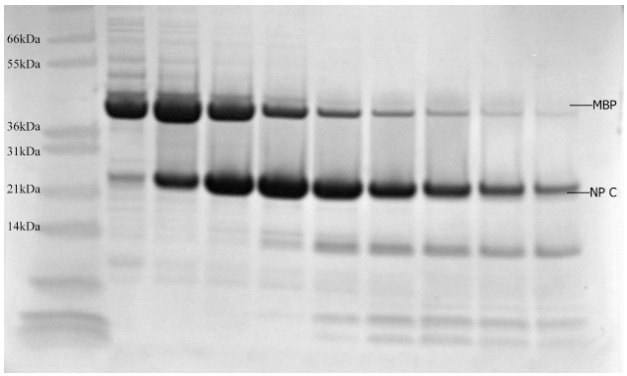


Fig 13: The C-terminal domain could not be separated from MBP by gel filtration. Lane 1 is marker, lane 2 to 9 are different fractions from the gel filtration.

In order to remove MBP from the C-terminal domain, an additional nickel column was used before gel filtration. One effective way was to connect extra Ni-NTA columns (5ml) along with the gel filtration column, and we found that two Ni-NTA columns gave us the best result (Fig 14 and Fig 15).

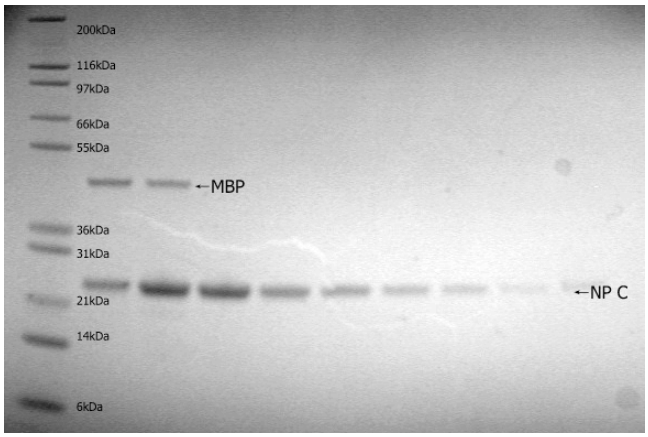


Fig. 14: NP C-terminal domain purification by a Ni-NTA column connecting to gel filtration column.

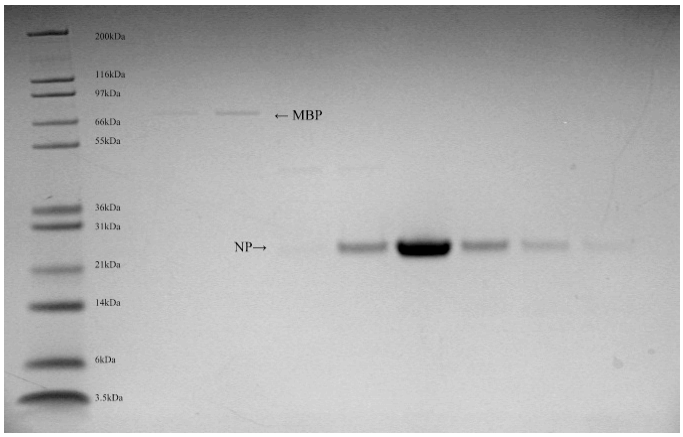


Fig 15: NP C-terminal domain purification by two Ni-NTA columns connecting to gel filtration.

We have found that it did not reduce the protein yield using the two Ni-NTA columns to remove the MBP, compared with using one Ni-NTA column.

Two different gel filtration buffers were applied in this experiment: high salt condition (20mM Tris-HCl, pH 7.5, 300mM NaCl, 10% glycerol), and low salt condition (20mM Tris-HCl, pH 7.5, 100mM NaCl, 10% glycerol)

Gel filtration result and SDS-PAGE gel picture under high salt condition:

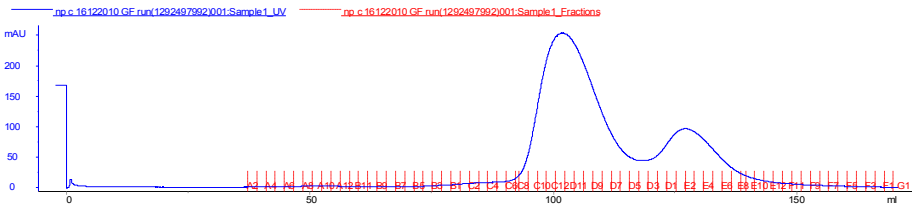


Fig.16: gel filtration graph. There are two peaks, and both of them were the NP C-terminal domain

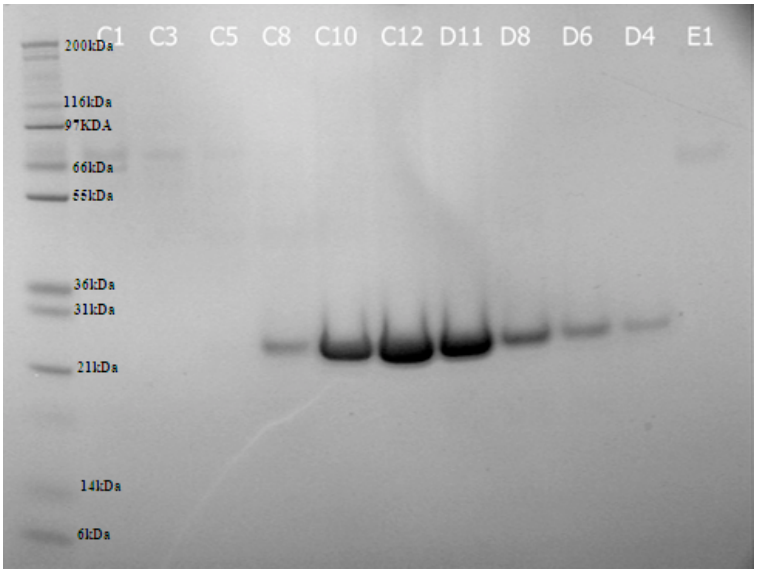


Fig. 17: NP C-terminal domain purification after gel filtration in high salt buffer.

Gel filtration result and SDS-PAGE gel picture under low salt condition:

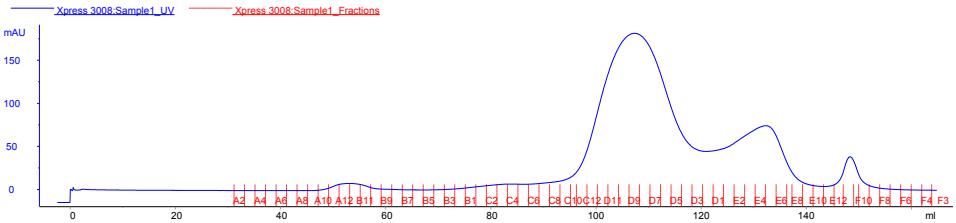


Fig.18: Gel filtration graph under low salt concentration. There are two major peaks, both of them were NP C-terminal domain. The third tiny peak shows nothing on SDS-PAGE gel, which could be just jam signal.

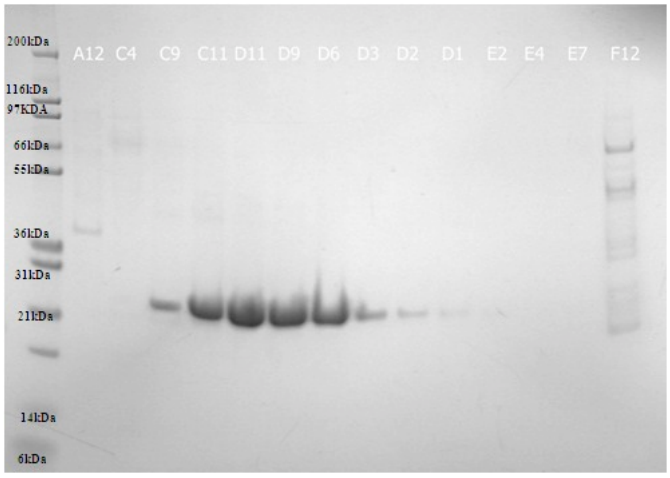


Fig 19: SDS-PAGE gel of gel filtration purification in low salt buffer

Both peaks were confirmed by SDS-PAGE and mass spectrometry as they both were the NP C-terminal domain. However, the second peak contains much less protein and we hardly got enough amount of protein from this peak after concentration. Therefore, we mainly focused on the first peak for crystallization. And the yield was around 1mg protein per liter culture in high salt buffer, or a bit less around 0.7mg protein per liter culture in low salt buffer.

3.1.2 HIS-TAGGED NP C-TERMINAL PROTEIN

The His-tagged NP C-terminal domain was expressed successfully. The protein yield is higher, around 6mg protein per liter cell culture, while the MBP-fusion one got 1mg protein per liter cell culture. But the purity of the purified protein was not as good as the His-tagged MBP fusion protein. We did not cleave the His-tag away as the protein would be used for surface plasmon resonance (SPR) assays. To improve the protein purity, we used 50 mM imidazole as washing buffer, instead of 20 mM imidazole.

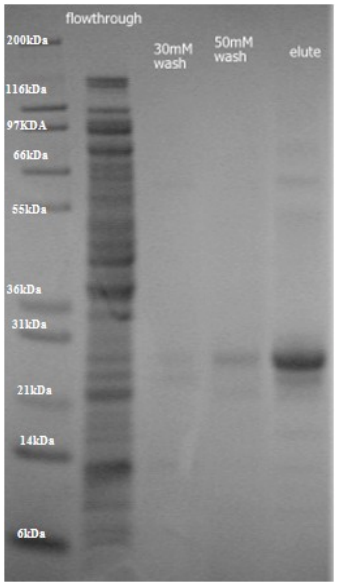


Fig.20: SDS-PAGE shows His-tagged NP C-terminal domain purification by a nickel column

The gel filtration graph showed that there is mainly one peak (Fig. 21). We obtained 5ml of the purified C-terminal domain with a concentration of 5mg/ml.

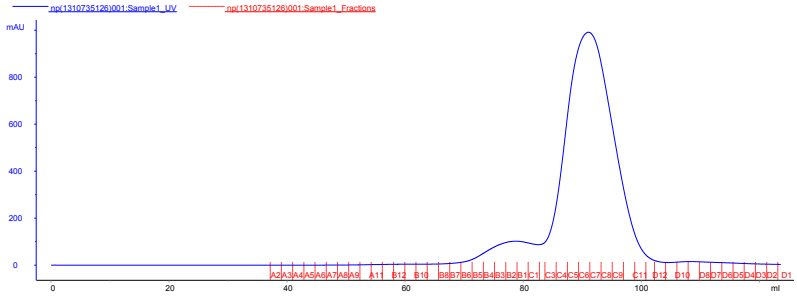


Fig. 21: gel filtration graph.

The SDS-PAGE revealed that we had obtained a reasonably pure protein (Fig 22).

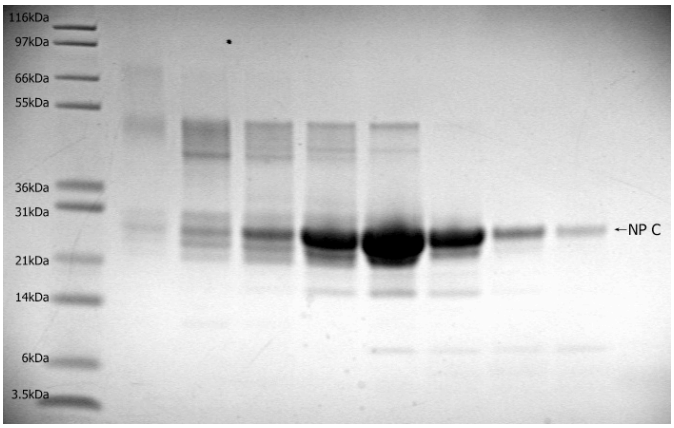


Fig 22: SDS-PAGE gel of gel filtration of His-tagged NP C-terminal domain. Lane 2 to 9 are different fractions from the gel filtration

3.2 CRYSTALLIZATION

The NP C-terminal domain, purified from a His-tagged MBP fusion protein was used for crystallization. We obtained the native crystals of NP C-terminal domain as shown in Fig 23. As our goal is to determine the NP C-terminal domain in complex with PAPM molecules, we focused on the crystals of NP C-terminal domain in complex with PAPM molecules, such as dsRNA and triphosphate dsRNAs.

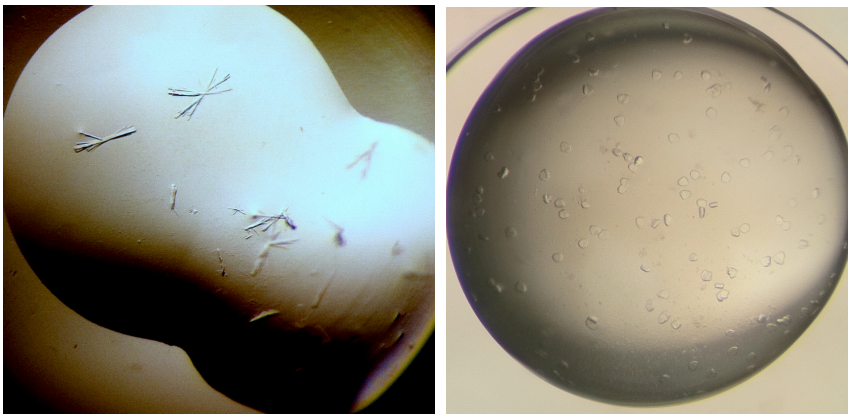


Fig 23: Crystal pictures of the native NP C-terminal protein.
Left one was screen result using honeybee system. Right one was optimization trials result.

A 6 nucleotide-length dsRNA, an 8 nucleotide-length dsRNA and a 12 nucleotide-length dsRNA were chemically synthesized in Eurogentec. Only the 8 nucleotide-length dsRNA formed crystals with NP C-terminal domain. The crystals of the NP C-terminal domain with dsRNA are shown in Fig 24. These crystals were thin. Although intensive optimization was carried out, we could not improve the crystals. We therefore tried to synthesize 8 and 12 nucleotide-length triphosphate dsRNAs with MEGA shortscript kit. We have tried to co-crystallize the triphosphate dsRNA with the C-terminal domain.

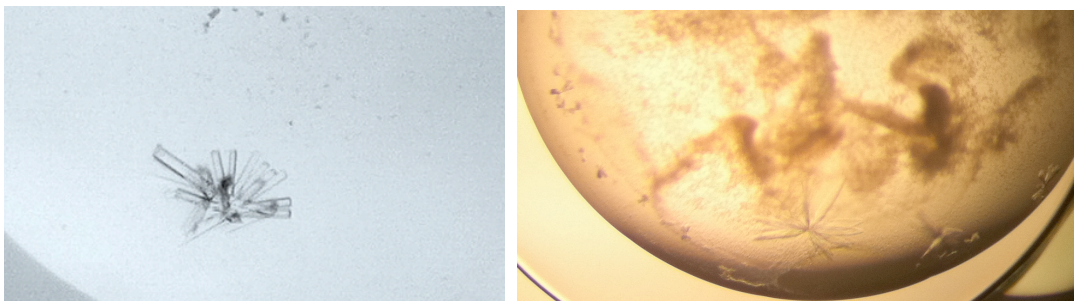


Fig 24: NP C-terminal domain protein with dsRNA 8 complex crystal
Left one was screen result using honeybee system. Right one was optimization trials result.

The crystals of the NP C-terminal domain in complex with triphosphate 8 nucleotide length dsRNA are shown in Fig 25. Those crystals can grow very fast and big in Crystal Clear P (Douglas) plates. Protein-RNA complex crystals appeared in less than 24 hours, and these crystals took about a week to grow large enough for X-ray diffraction.

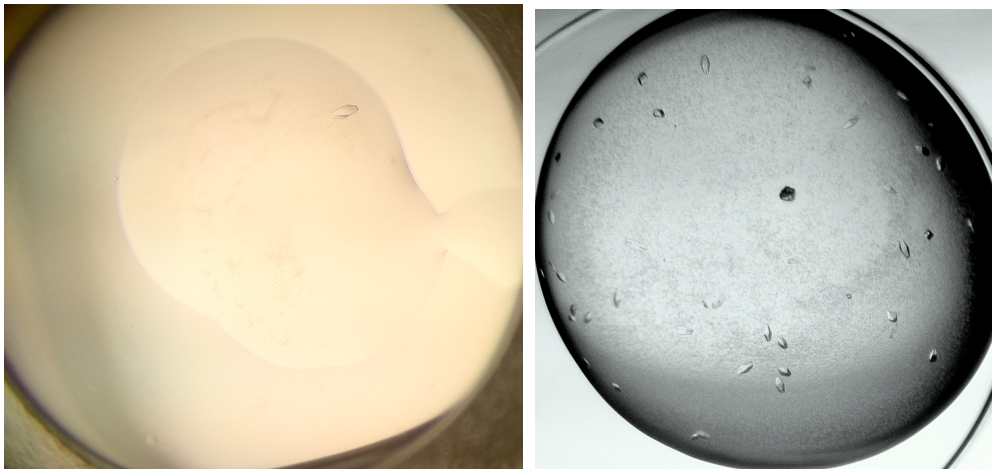


Fig 25: Crystals of NP C-terminal domain protein in complex with triphosphate dsRNA 8. The crystals that grew from the screening are shown on the left and the crystals after optimization were shown on the right.

The crystal that was used to collect data on Diamond beam line IO3 is shown in Fig 26. The rhombus shaped crystal is approximately 150µm long and 40µm wide.

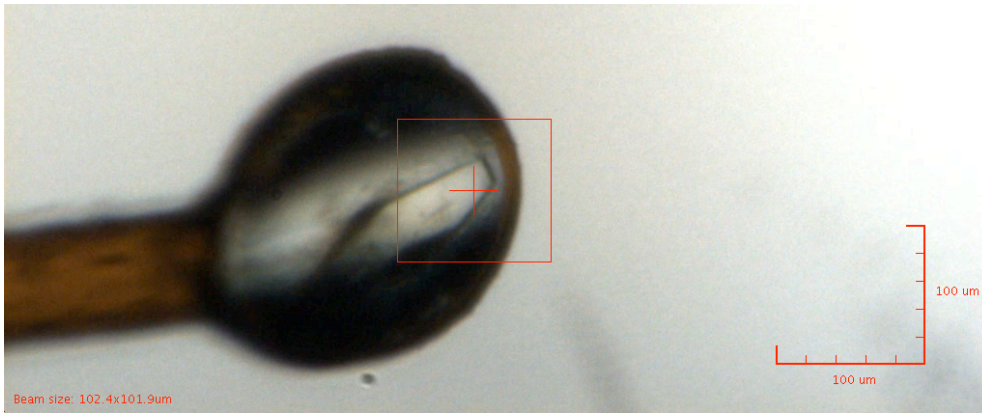


Fig 26: Crystal of NP C-terminal domain protein in complex with triphosphate dsRNA 8 in loop for X-ray diffraction

The diffraction pattern of the crystal is shown in Fig 27. The crystal diffracted to 1.73 Å and shows clearly defined diffraction spots. This diffraction pattern shows no strong ice rings and the frequency of the diffraction spots infers a large space group.

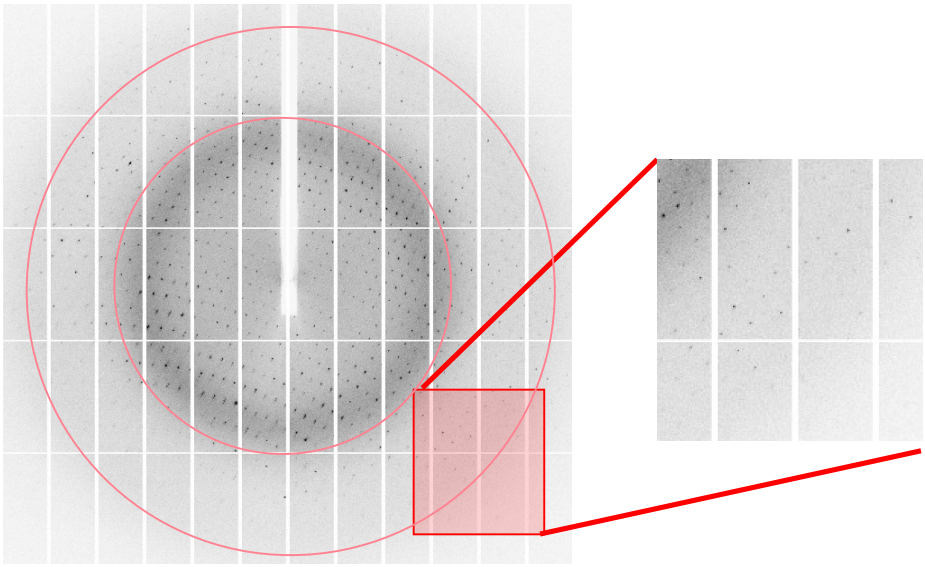


Fig 27: One of the diffraction images, collected on Diamond Light Source UK, beamline IO3. The resolution of the outer circle is 1.7 Å. Picture on the right side is the enlarged area of red square on the left side picture.

3.3 CRYSTAL STRUCTURE

The collected data were processed, and the results were listed in Table 2. After molecular replacement and a few cycles of refinement, the electron density for the dsRNA was very clear. However, we did not see electron density for the triphosphate. A dsRNA (GGGC/CCCG) was built in the electron density, which shows that the 3'-end nucleotide (cytosine) of chain GGGC is directly located in the 3'-5' exonuclease active site (Fig 28, 29).

Table 2: The crystallographic statistics of the NP C-terminal domain in complex with triphosphate dsRNA

Data Statistics	
Space Group	*P*3
Unit Cell parameters	
$a = b$	177.16 Å
c	56.49 Å
$\alpha=\beta$	90°
γ	120°
Resolution range	46.5-1.73 Å
Average redundancy	5.6 (5.5)

Unique reflections	184629 (5543)
I/σI	16.4 (4.2)
Completeness	99.7% (99.9%)
Rmerge	7.3% (39.3%)
Refinement Statistics	
Rfactor	17.80%
Rfree	19.87%
Rmsd bonds	0.03 Å / 2.45°
Number of protein atoms	9076
Mean B-factor	25.27 Å
Residues in Ramachandran core	96.83%
Residues in disallowed regions	0

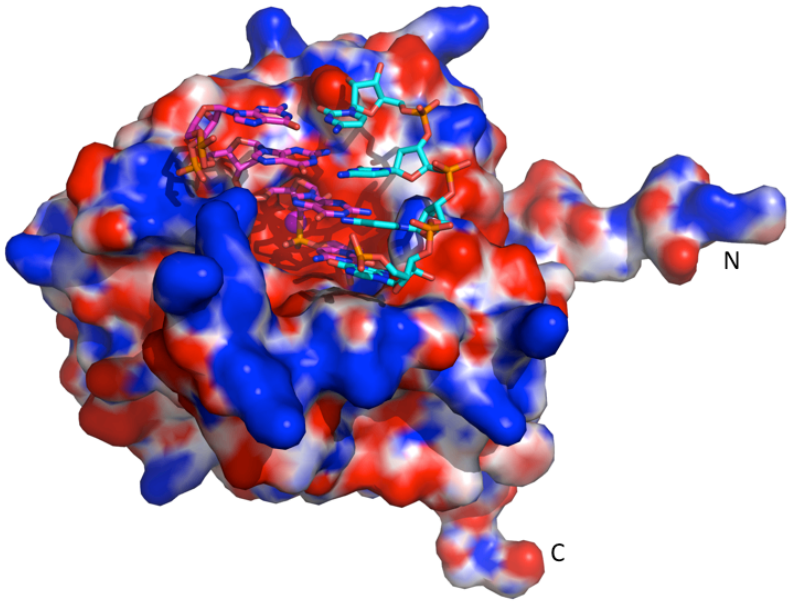


Fig 28: Electrostatic surface potential map of the NPC and dsRNA complex, and dsRNA is shown in stick. The cleaving strand of dsRNA main chain is located in a blue belt formed by positive charge residues. The blue area represents positively charged residues and the red area represents negatively charged residues.

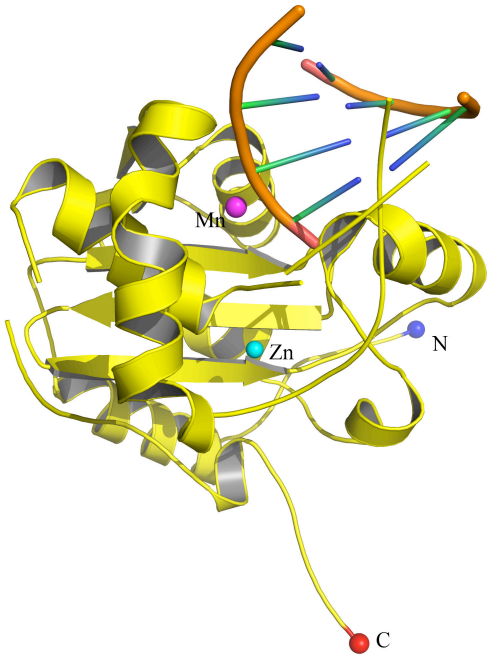


Fig 29: LAVS NPC and dsRNA complex structure in cartoon, the manganese ion shows as magentas ball and Zinc ion shows as grays ball.

From the structure we can see that one strand (5'GGGC3') of the dsRNA interacts extensively with NP residues D466, D389, Gln462, D533, R492, E391, H528 and S430, in which most of them are catalytic residues, while another strand (5'GCCC3') interacts with side chains of NP residues Y429, D426, Q425, Q422, R393 and D465. Particularly, the side chain of residue Y429 stacks against the guanidine ring (Fig 30, 31).

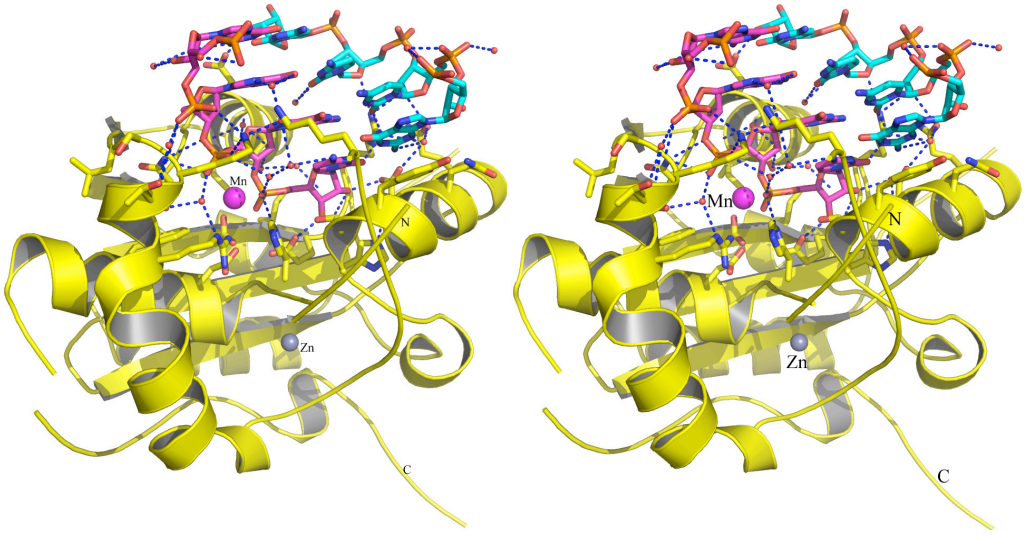


Fig 30: Close stereo view of the interaction residues in cartoon. Dotted line shows the interaction part between residues and RNA. There is 3° difference between left and right pictures.

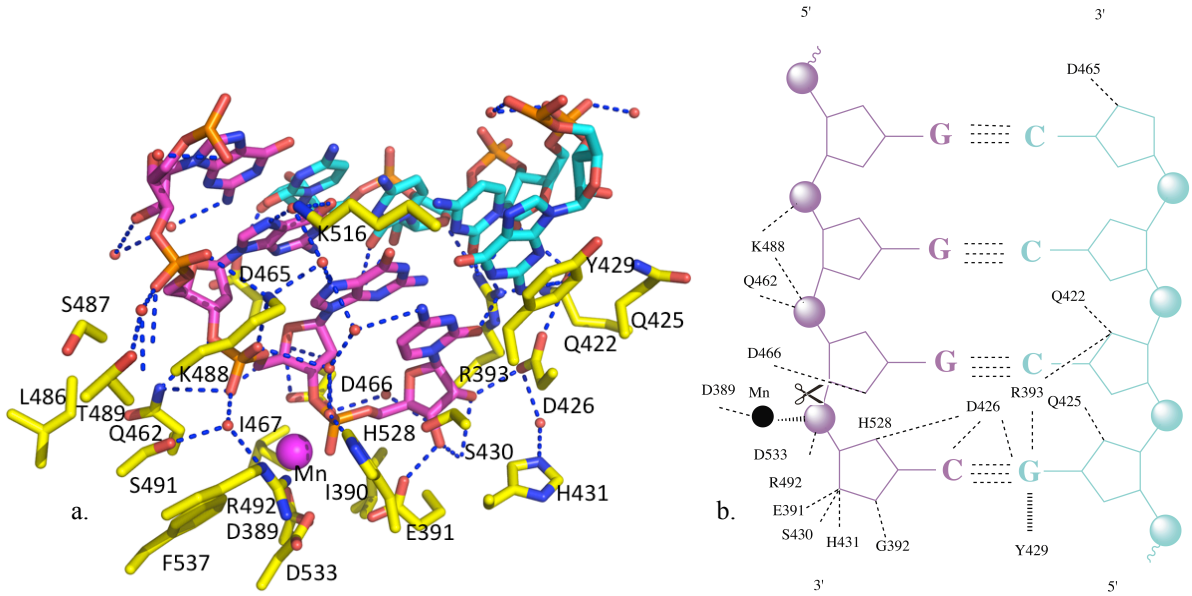


Fig 31: Interactions between LAVS NPC and triphosphate dsRNA and dsRNA cleavage activities of LAVS NP-C mutants. a, Residues are involves in the dsRNA binding and cleavage. The cleavage chain of the dsRNA in purple and the binding chain in cyan. b. Schematic representation of LAVS NPC and triphosphate dsRNA interactions. The scissile bond is indicated with a scissor.

In order to validate the functions of the binding residues in dsRNA recognition and cleavage, we have generated several alanine substitute mutant constructs (name of mutant sites and primer details were shown in appendix II), and we are purifying the mutant proteins. We will try ITC and SPR to check the mutant binding affinities with different RNA ligands, such as dsRNA, triphosphate dsRNA and ssRNA. In the same time, we will check whether the mutant proteins process the PAPM RNA differently in efficiency.

3.4 MUTAGENESIS AND MUTANT PROTEIN PURIFICATION

3.4.1 PHISTEV NP C-TERMINAL DOMAIN PLASMID CONSTRUCTION

The gene fragment encoding the NP C-terminal was inserted into pHisTEV plasmid successfully, confirmed by the restriction enzyme digestion result (Fig 32). The constructed plasmid was digested with NCO Ⅰ and HindIII. And the digested plasmid had been run through a 10% agarose gel. The gel showed that the NP C-terminal domain encoding gene band was around 600bp, corresponding to the exactly size for the NP C-terminal domain encoding gene 615bp.

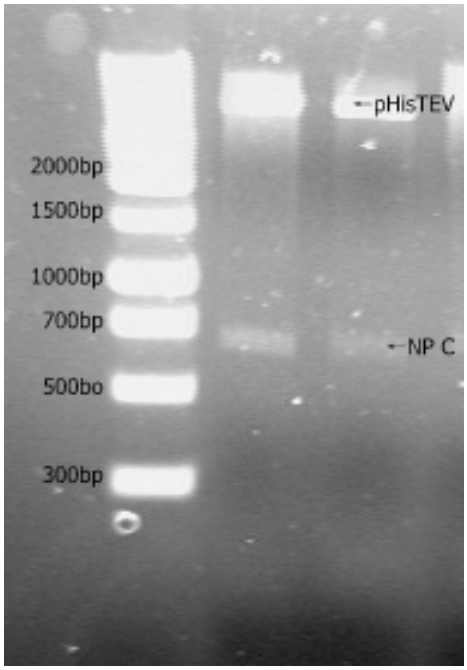


Fig 32: The electrophoresis gel of digested NPC-pHisTEV

3.4.2 SINGLE SITE PLASMID MUTAGENESIS

The PCR result of a mutation is shown in Fig 33. Those plasmids were around 6kbp, equivalent to the length of pHisTEV vector.

Fig 33: PCR result for mutations

The DNAMan software had been used to compare the mutation sequences results with NP C-terminal domain encoding gene. Only the desired constructed plasmids were selected to transfer to Rosetta cells for expression.

3.4.3 EXPRESSION AND PURIFICATION

The mutants were obtained by using the same expression and purification method as for the native pHisTEV NP C-terminal domain. The single site mutation proteins were expressed well as the native protein, around 5 to 6mg protein per liter culture. Fig 36 and 37 were S430A result features as an example of all mutants.

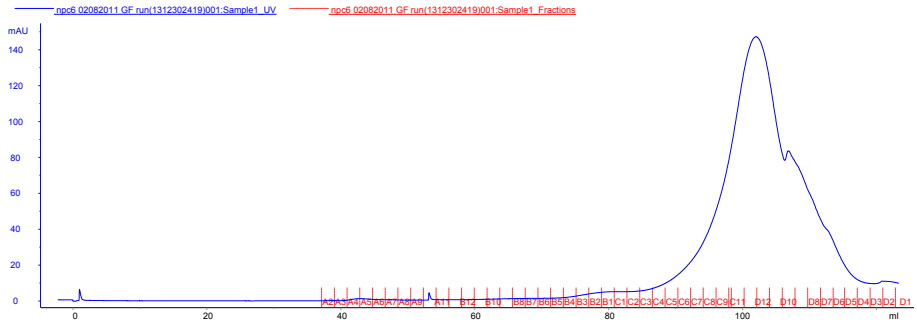


Fig 34: gel filtration graph on S430A mutant. There is one main peak of protein.

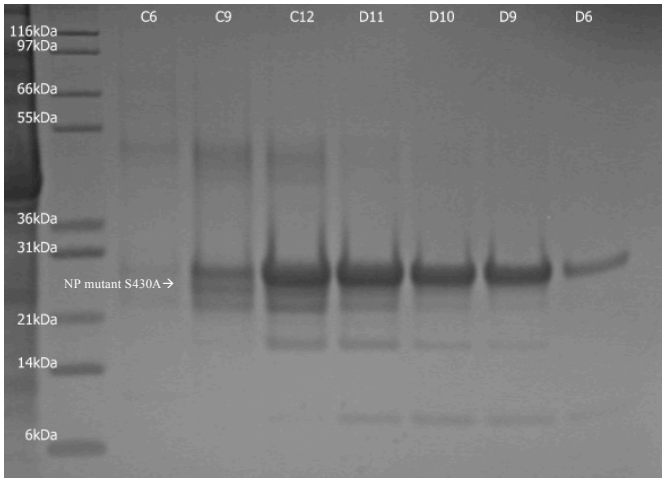


Fig 35: SDS-PAGE gel of gel filtration of S430A mutant. Lane C9 to D9 were different fractions of main peak from gel filtration.

4 FUTURE WORK

LAVS NP and other arenavirus's NPs use a totally novel way to evade the human immune responses. To find the detail of the mechanism is important for the controlling of such deadly infectious diseases caused by some arenaviruses. Our collaborators at Emory University have shown that the LAVS NP can process dsRNA much quicker than those of ssRNA (approximate 20 times faster), and our structure of dsRNA and C-terminal domain sheds an important light on the mechanism.

In order to elucidate the NP C-terminal's immune suppression function, the mutant proteins have been designed, expressed and purified. Their binding affinities with different PAPM RNAs will need to be tested in future experiments. I'll use florescence labeled RNAs to identify the cleavage efficiency of different mutants. Each mutant will interact with several designed RNAs under certain time and temperature. Then polyacrylamide gel electrophoresis will be used to separate bands and recognize the cleavage extent. On the other hand, a standard gel filtration chromatography method will be used to determine whether the NP C-terminal protein we obtained is monomer, dimer or trimer in different solvent condition.

5 THE NP N-TERMINAL DOMAIN

The N-terminal domain of LAVS plays an essential role for cap binding. To investigate the detailed mechanism, we have cloned, expressed and crystallized the N-terminal domain.

5.1 MATERIAL AND METHODS

5.1.1 EXPRESSION CONSTRUCT

The N-terminal domain consists of residues 1 to 338. The gene segment is derived from Lassa virus (Josiah strain), S genome. Similar way as dealing with the C-terminal domain, the N-terminal domain gene segment was cloned into pMAL-c2X-derived pLou3 plasmid, with a 6His-tag at the N terminus of MBP and TEV cleavage site between MBP and the target domain. This construct was transformed into Rosetta cells (Novagen). Xiaoxuan Qi previously built this construct in our group.

5.1.2 PROTEIN EXPRESSION AND PURIFICATION

The expression and purification process was very similar to those of the C-terminal domain.
The transformed cells were cultured in LB media with 50mg/ml ampicillin and 34mg/ml chloramphenicol. The cells were induced with 0.03mM IPTG when the cell OD_{600} reached 0.6. And then the cells were cultured for about 20 hours at 20°C.

The cells were harvested by centrifugation for 15min at 10000g. Cell pellet was resuspended in loading buffer (20mM Tris-HCl, pH7.5, 10mM imidazole, 300mM NaCl and 10% glycerol) with1 tablet of EDTA-free protease inhibitor (Roch), 1μM DNase (Sigma), 1μM Lysozome and 1mM PMSF (Sigma). The cells were disrupted using a cell disruptor. The cell debrides were removed by centrifugation for 30min at 40000g. The supernatant was collected and applied to 10 ml of Ni-NTA agarose (Qiagen) beads, which pre-equilibrated with loading buffer. The beads were washed with 20ml loading buffer (wash buffer with 30mM imidazole will wash away target protein), and the fusion protein was eluted with 22ml elution buffer consisting of 20mM Tris-HCl, pH 7.5, 500mM imidazole, 300mM NaCl and 10% glycerol).

The sample buffer was changed to desalt buffer (20mM Tris-HCl, pH 7.5, 300mM NaCl and 10% glycerol), using a desalting column (Hiprep 26/10, GE). The His-MBP-NPC fusion protein was cleaved by TEV proteinase at room temperature for 16 hours. The His-MBP was removed by Ni-NTA agarose beads (pre-equilibrated with desalt buffer).

The sample was concentrated to 7.5ml for the gel filtration. The gel filtration column was pre-equilibrated with GF buffer (20mM Tris-HCl, pH7.5, 300mM NaCl and 10% glycerol as high salt concentration condition, or 20mM Tris-HCl, pH7.5, 100mM NaCl and 10% glycerol for low salt concentration). The fractions containing the NP N-terminal domain were checked by SDS-PAGE gel (Invitrogen, Mark12) and pooled. The purified N-terminal domain was concentrated to 20mg/ml, frozen by liquid nitrogen and stored at -80°C.

5.1.3 CRYSTALLIZATION

Crystallization conditions were screened on MRC plates with the sitting drop vapor-diffusion method using a Honeybee robot. Drops contained 0.15µl protein solution and 0.15µl buffer with 100µl buffer in each reservoir. The crystallization screens were stochastic screens, made by JHN lab.

Initial crystals were obtained around 2~3 weeks at 20 °C in 0.5M magnesium formate dihydrate, 0.1M HEPES, pH 7.5. Optimizations were carried out, and the crystallization conditions were optimized to 0.2M~0.8M magnesium formate dihydrate, 0.1M HEPES, pH 7.1-7.9.

5.2 RESULT

The gel filtration shows that the N-terminal domain displays two peaks. Mass spectra data is needed to determine whether both peaks are the same N-terminal domain. The two peaks gain almost same amount of protein. However, only the peak two proteins formed crystals.

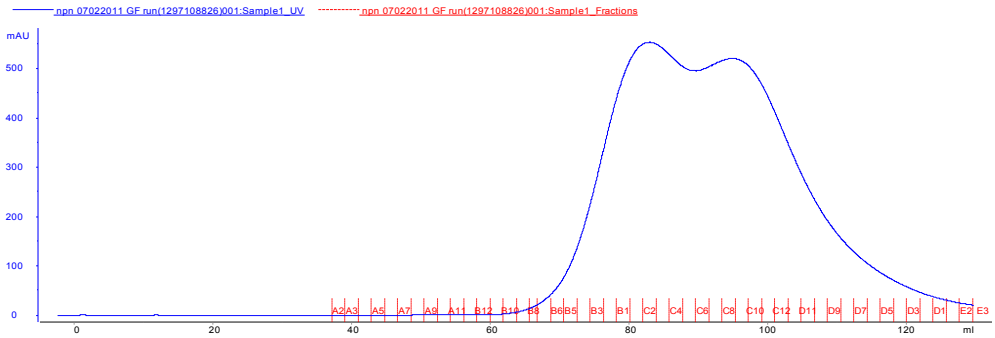


Fig 36: NP N-terminal domain gel filtration graph

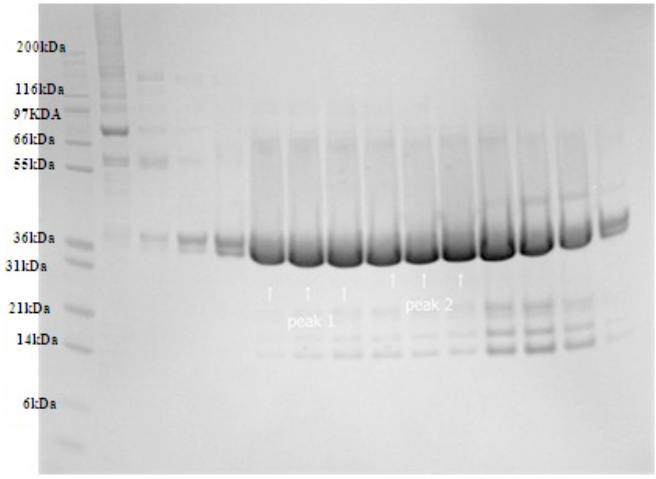


Fig 37: SDS-PAGE gel gel of gel filtration fractions of N-terminal domain. Lane 6 to 8 were the first peak, lane 9 to 11 were the second peak.

Crystals

Fig 38 shows the obtained crystals of the N-terminal domain. On the left figure is the initial crystal from the screen and the right figure shows the crystals grown under optimized conditions. We are trying to improve the crystal quality.

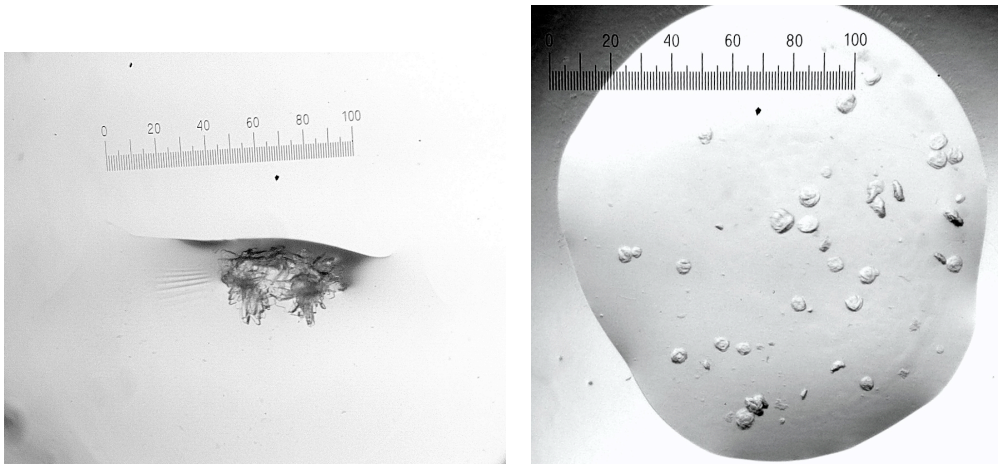


Fig 38: crystal of the NP N-terminal domain

5.3 FUTURE WORK

The N-terminal domain is very important for the cap-snatching mechanism, and this mechanism is different from that of bunyavirus or influenza virus. To find out the detailed mechanism, I will try to improve the crystal quality, test whether the crystal diffracts. If I gain good diffraction data, I will then try to evaluate the cap binding function by obtaining the m7GpppG and N-terminal domain complex structure. At the same time, I will set up the *in vitro* assays using the NP and the N-terminal domain of the L to test how the virus from the host mRNAs obtains the caps.

6 APPENDIX I

Short chain RNA sequence details

For crystallization trials

ds RNA 6：GAC-GCU

ds RNA 8: CGC-AUG-CG

ds RNA 12: GAC-GCU-AGC-GUC

ppp ds RNA GC8: GGG-CGC-CC

For protein-RNA binding affinity test

ds RNA 16 forward: UCU-CUC-UCU-CUC-UCC-C

ds RNA 16 reverse: GGG-AGA-GAG-AGA-GAG-A

ppp ds RNA GC14: GGC-GCG-CGC-GCG-CCT-ATA-GTG-AGT-CG

DNA template sequence details

(For RNA synthesis)

common: AAT-TTA-ATA-CGA-CTC-ACT-ATA-GG,

GC8: GGG-CGC-CCT-ATA-GTG-AGT-CGT-ATT-AAA-TT

GC14: GGC-GCG-CGC-GCG-CCT-ATA-GTG-AGT-CGT-ATT-AAA-TT

7 APPENDIX II

DNA primer sequence details for mutants

1_D465A_forward: CCT-GTC-AGG-GGT-CCG-CTG-ACA-TAA-GGA-AAC-TC

D465A_reverse: GAG-TTT-CCT-TAT-GTC-AGC-GGA-CCC-CTG-ACA-GG

2_K469A_forward: CCG-ATG-ACA-TAA-GGG-CAC-TCC-TTG-AAT-CAC-AAG

K469A_reverse: CTT-GTG-ATT-CAA-GGA-GTG-CCC-TTA-TGT-CAT-CGG

3_Q462A_forward: GGT-CAT-TAC-CTG-TGC-GGG-GTC-CGA-TGA-C

Q462A_reverse: GTC-ATC-GGA-CCC-CGC-ACA-GGT-AAT-GAC-C

4_R492A_forward: CAG-CAA-AAC-TGA-TTC-CGC-GAA-GTA-TGA-AAA-TGC-AG

R492A_reverse: CTG-CAT-TTT-CAT-ACT-TCG-CGG-AAT-CAG-TTT-TGC-TG

5_R468A_forward: GGG-TCC-GAT-GAC-ATA-GCG-AAA-CTC-CTT-GAA-TCA-C

R468A_reverse: GTG-ATT-CAA-GGA-GTT-TCG-CTA-TGT-CAT-CGG-ACC-C

6_S430A_forward: CAG-GAT-GCT-AAG-TAC-GCA-CAT-GGG-ATT-GAT-GTC

S430A_reverse: GAC-ATC-AAT-CCC-ATG-TGC-GTA-CTT-AGC-ATC-CTG

7_Y429A_forward: CAA-GCA-GGA-TGC-TAA-GGC-CTC-ACA-TGG-GAT-TGA-TG

Y429A_reverse: CAT-CAA-TCC-CAT-GTG-AGG-CCT-TAG-CAT-CCT-GCT-TG

8_Y429F_forward: CAA-GCA-GGA-TGC-TAA-GTT-CTC-ACA-TGG-GAT-TGA-TG

Y429F_reverse: CAT-CAA-TCC-CAT-GTG-AGA-ACT-TAG-CAT-CCT-GCT-TG

9_R492A_forward: GCA-AAA-CTG-ATT-CCG-CGA-AGT-ATG-AAA-ATG

R492A_reverse: CAT-TTT-CAT-ACT-TCG-CGG-AAT-CAG-TTT-TG

10_K516A_forward: TGT-CGT-TGT-TGA-AGC-GAA-GAA-AAG-AGG

K516A_reverse: GCC-TCT-TTT-CTT-CGC-TTC-AAC-AAC-GAC

11_K517A_forward: GTT-GTT-GAA-AAG-GCG-AAA-AGA-GGC-GG

K517A_reverse: CCG-CCT-CTT-TTC-GCC-TTT-TCA-ACA-AC

12_truncation_515_522forward:

CAT-GCA-CAC-AGG-TGT-CGT-TGT-T||GA-GGA-AAT-AAC-CCC-TCA-CTG

truncation_515_522reverse:

CAG-TGA-GGG-GTT-ATT-TCC-TC||A-ACA-ACG-ACA-CCT-GTG-TGC-ATG

8 APPENDIX III

PCR protocol based on that provided by Dr Huangting Liu.

Pfu buffer	5µl
dNTP	5µl
primer Forward	1µl
primer Reverse	1µl
template DNA	~10ng
pfu enzyme	1µl
H_2O to final volume of	50µl

	95°C	5min
12 cycles	95°C	1min
	61°C	1min (Tm -5°C)
	72°C	12min
	95°C	1min
	61°C	1min
	72°C	30min
	4°C	forever

PCR protocol based on QuikChange™ Site-Directed Mutagenesis method

Pfu buffer	5µl
dNTP	5µl
primer Forward	1µl
primer Reverse	1µl
template DNA	~20ng
pfu enzyme	1µl
H_2O to final volume of	50µl

	95°C	1min
13 cycles	95°C	30s
	55°C	1min
	68°C	12min
	4°C	forever

9 APPENDIX IV

Restriction enzyme reaction

buffer E	1µl	10 µl
BSA	1µl	
HindIII	0.3µl	
NCOI	0.3µl	
Plasmid	5µl	
H_2O	3.3µl	

Reaction in 37°C for 2 hours

Run 1% agarose gel for result checking

T4 DNA ligase protocol

vector DNA	100ng	10 µl
insert DNA	17ng	
Ligase 10Xbuffer	1µl	
T4 DNA ligase (Weiss units)	0.1~1u	
Nuclease-Free water to final volume of 10µl		

Incubate in room temperature for 3 hours, or 4°C overnight.

10 APPENDIX V

TRANSFORM PLASMID TO COMPETENT CELLS

1. At 4ºC, add 1µl of plasmid to a microtube with 50µL of TAM1/Rosetta competent cells and mix well; keep in ice for 30min

2. Heat shock at 42ºC for 90sec, and put in ice for 2min

3. Add 1ml of LB medium to transformed cells. Incubate at 200rpm, 37ºC for 1hour

4. Spin down at 13000 rpm for 2min in a microcentrifuge

5. Reject 850ml supernatant and resuspend cells pellet

6. Spread cells on LB/KANA agarose plate (TAM 1 cell) or LB/KANA/CHLO agarose plate (Rosetta). Inserted plasmid contains kanamycin resistance, and Rosetta cell itself has chloramphenicol resistance.

7. Grow cells overnight by putting the plate up side down at a 37ºC still incubator

Plasmid extraction by QIAGEN miniprep kit

1. Culture target cells overnight in 10ml LB media (supplemented with 34 µg/ml kanamycin for TAM1 cell, or 34 µg/ml kanamycin and 34 µg/ml chloramphenicol for Rosetta cell). Cell colony is picked from agarose plate.

2. Transit overnight culture to 15ml centrifuge tube. Spin down 4500rpm for 10min. Discard supernatant

3. Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube.

4. Add 250 µl Buffer P2 and gently invert the tube 4–6 times to mix.

5. Add 350 µl Buffer N3 and invert the tube immediately but gently 4–6 times.

6. Centrifuge for 10 min at 13,000 rpm in microcentrifuge.

7. Apply the supernatants from last step to the QIAprep spin column by decanting or pipetting.

8. Centrifuge for 1min. Discard the flow-through.

9. Wash the spin column by adding 0.5 ml Buffer PB and centrifuging for 1min. discard the flow-through.

10. Wash spin column by adding 0.75 ml Buffer PE and centrifuging for 30–60 s.

11. Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.

12. Place the column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 45 µl water to the center of each spin column, let stand for 1 min, and centrifuge for 1 min.

11 REFERENCE

Ogbu O, Ajuluchukwu E, Uneke CJ. **Lassa fever in West African sub-region: an overview.** Journal of vector borne diseases 2007 Mar;44(1):1-11.

Haas WH, Breuer T, Pfaff G, Schmitz H, Köhler P, Asper M, Emmerich P, Drosten C, Gölnitz U, Fleischer K, Günther S. **Imported Lassa Fever in Germany: Surveillance and Management of Contact Persons.** Clinical Infectious Diseases 2003 May 15;36(10):1254-8.

Illick MM, Branco LM, Fair JN, Illick KA, Matschiner A, Schoepp R, Garry RF, Guttieri MC. **Uncoupling GP1 and GP2 expression in the Lassa virus glycoprotein complex: implications for GP1 ectodomain shedding.** Virology journal 2008 Dec 23;5:161.

Fichet E, Rogers DJ. **Risk Maps of Lassa Fever in West Africa.** PLoS Neglected Tropical Diseases. 2009 Mar;3(3):e388.

Schlie K, Maisa A, Freiberg F, Groseth A, Strecker T, Garten W. **Viral protein determinants of Lassa virus entry and release from polarized epithelial cells.** Journal of virology 2010 Apr;84(7):3178-88.

Eichler R, Strecker T, Kolesnikova L, ter Meulen J, Weissenhorn W, Becker S, Klenk HD, Garten W, Lenz O. **Characterization of the Lassa virus matrix protein Z: electron microscopic study of virus-like particles and interaction with the nucleoprotein (NP)**. Virus research 2004 Mar 15;100(2):249-55.

Lelke M, Brunotte L, Busch C, Günther S. **An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome.** Journal of virology 2010 Feb; 84(4): 1934-44.

Buchmeier MJ, de la Torre JC, Peters CJ. **Arenaviridae: the viruses and their replication.** In Fields Virology, Vol 2. Edited by Knipe DM, Howley PM. Lippincott Williams & Wilkins, a Wolters Klumer Business; 2007:1791-1828.

Bowen MD, Rollin PE, Ksiazek TG, Hustad HL, Bausch DG, Demby AH, Bajani MD, Peters CJ, Nichol ST. **Genetic diversity among Lassa virus strains.** Journal of virology 2000 Aug;74(15):6992-7004.

Kunz S. **Receptor binding and cell entry of Old World arenaviruses reveal novel Dects of virus-host interaction.** Virology 2009 May 10;387(2):245-9.

Shtanko O, Imai M, Goto H, Lukashevich IS, Neumann G, Watanabe T, Kawaoka Y.**A role for the C terminus of Mopeia virus nucleoprotein in its incorporation into Z protein-induced virus-like particles.** Journal of virology 2010 May;84(10):5415-22.

Qi X, Lan S, Wang W, Schelde LM, Dong H, Wallat GD, Ly H, Liang Y, Dong C. **Cap binding and immune evasion revealed by Lassa nucleoprotein structure.** Nature 2010 Dec 9;468(7325):779-83.

Hastie KM, Kimberlin CR, Zandonatti MA, MacRae IJ, Saphire EO. **Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3′ to 5′ exonuclease activity essential for immune suppression.** Proceedings of the National Academy of Sciences 2011 Feb 8;108(6):2396-401.

Wang Y, Ludwig J, Schuberth C, Goldeck M, Schlee M, Li H, Juranek S, Sheng G, Micura R, Tuschl T, Hartmann G, Patel DJ. **Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I.** Nature structural and molecular biology 2010 Jul;17(7):781-7.

Cheng A, Wong SM, Yuan YA. **Structural basis for dsRNA recognition by NS1 protein of influenza A virus.** Cell research 2009 Feb;19(2):187-95.

Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. **iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.** Acta Crystallogr D Biol Crystallogr 2011 Apr;67(Pt 4):271-81.

Evans P. **Scaling and assessment of data quality**. Acta Crystallogr D Biol Crystallogr 2006 Jan;62(Pt 1):72-82.

McCoy AJ. **Solving structures of protein complexes by molecular replacement with Phaser.** Acta Crystallogr D Biol Crystallogr 2007 Jan;63(Pt 1):32-41.

Emsley P, Lohkamp B, Scott WG, Cowtan K. **Features and development of Coot**. Acta Crystallogr D Biol Crystallogr 2010 Apr;66(Pt 4):486-501.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. **REFMAC5 forthe refinement of macromolecular crystal structures.** Acta Crystallographica Section D: Biological Crystallography 2011 Apr;67(Pt 4):355-67..

Rey FA. **Virology: One protein, many functions**, Nature 2010 Dec 9;468(7325):773-5.

Leung DW, Prins KC, Borek DM, Farahbakhsh M, Tufariello JM, Ramanan P, Nix JC, Helgeson LA, Otwinowski Z, Honzatko RB, Basler CF, Amarasinghe GK. **Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35.** Nature structural & molecular biology 2010 Feb;17(2):165-72..

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, Canard B. **The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription.** PLoS pathogens 2010 Sep 16;6(9):e1001038.

McCartney SA, Colonna M. **Viral sensors: diversity in pathogen recognition.** Immunology review 2009 Jan;227(1):87-94.

Shah NR, Sunderland A, Grdzelishvili VZ. **Cell type mediated resistance of vesicular stomatitis virus and Sendai virus to ribavirin.** PloS one 2010 Jun 22;5(6):e11265.

Tayeh A, Tatard C, Kako-Ouraga S, Duplantier JM, Dobigny G. **Rodent host cell/Lassa virus interactions: evolution and expression of α-Dystroglycan, LARGE-1 and LARGE-2 genes, with special emphasis on the Mastomys genus.** Infection, genetics and evolution journal of molecular epidemiology and evolutionary genetics in infectious diseases 2010 Dec;10(8):1262-70.

Mendenhall M, Russell A, Juelich T, Messina EL, Smee DF, Freiberg AN, Holbrook MR, Furuta Y, de la Torre JC, Nunberg JH, Gowen BB. **T-705 (favipiravir) inhibition of arenavirus replication in cell culture.** Antimicrobial agents and chemotherapy 2011 Feb;55(2):782-7..

Capul AA, de la Torre JC, Buchmeier MJ. **Conserved residues in Lassa fever virus Z protein modulate viral infectivity at the level of the ribonucleoprotein.** Journal of virology 2011 Apr;85(7):3172-8.

Albariño CG, Bird BH, Chakrabarti AK, Dodd KA, Erickson BR, Nichol ST. **Efficient rescue of recombinant Lassa virus reveals the influence of S segment noncoding regions on virus replication and virulence.** Journal of virology 2011 Apr;85(8):4020-4.

Macher AM, Wolfe MS. **Historical Lassa fever reports and 30-year clinical update.** Emerging Infectious Diseases 2006 May;12(5):835-7.

Feig AL. **Studying RNA-RNA and RNA-protein interactions by isothermal titration calorimetry.** Methods in enzymology 2009;468:409-22..

Khan SH, Goba A, Chu M, Roth C, Healing T, Marx A, Fair J, Guttieri MC, Ferro P, Imes T, Monagin C, Garry RF, Bausch DG. **New opportunities for field research on the pathogenesis and treatment of Lassa fever.** Mano River Union Lassa Fever Network: Antiviral research 2008 Apr;78(1):103-15.

Martín V, Abia D, Domingo E, Grande-Pérez A. **An interfering activity against lymphocytic choriomeningitis virus replication associated with enhanced mutagenesis.** The Journal of general virology 2010 Apr;91(Pt 4):990-1003..

Pinschewer D, Perez M, de la Torre JC. **Role of the virus nucleoprotein in the regulation of lymphocytic choriomeningitis virus transcription and RNA replication.** Journal of virology 2003 Mar;77(6):3882-7.

Kuri T, Habjan M, Penski N, Weber F. **Species-independent bioassay for sensitive quantification of antiviral type I interferons.** Virology journal 2010 Feb 26;7:50..

Martínez L, Emonet S, Giannakas P, Cubitt B, García-Sastre A, de la Torre JC. **Identification of amino acid residues critical for the anti-interferon activity of the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus.** Journal of virology 2009 Nov;83(21):11330-40.

Borio L, Inglesby T, Peters CJ, Schmaljohn AL, Hughes JM, Jahrling PB, Ksiazek T, Johnson KM, Meyerhoff A, O'Toole T, Ascher MS, Bartlett J, Breman JG, Eitzen EM Jr, Hamburg M, Hauer J, Henderson DA, Johnson RT, Kwik G, Layton M, Lillibridge S, Nabel GJ, Osterholm MT, Perl TM, Russell P, Tonat K; Working Group on Civilian Biodefense. **Hemorrhagic fever viruses as biological weapons: medical and public health management.** Journal of the American Medical Association 2002 May 8;287(18):2391-405.

The Center of Disease Control and Prevention, **Lassa virus fact sheet**, 2004 Dec 3.

World Health Organization, Media centre, **Lassa virus fact sheet N°179.** Revised 2005 Apr.