Biochemical Studies of SARS-CoV-2 Non-structural Protein NSP14 and NiRAN Domain of NSP12

A thesis submitted to Indian Institute of Science Education and Research Pune in partial fulfillment of the requirements for the BS-MS Dual Degree Programme

Submitted by
Muhammed Navas P
20181077

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

Date: April 2023

Thesis Supervisor:
Prof. Saikrishnan Kayarat
Professor, Department of Biology
IISER Pune

From May 2022 to March 2023
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

CERTIFICATE

This is to certify that this dissertation entitled “Biochemical Studies of SARS-CoV-2 Non-structural Protein NSP14 and NiRAN Domain of NSP12” towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Muhammed Navas P at Indian Institute of Science Education and Research under the supervision of Prof. Saikrishnan Kayarat, Professor, Department of Biology, during the academic year 2022-2023.

Prof. Saikrishnan Kayarat

Committee:

Prof. Saikrishnan Kayarat

Dr. Gayathri Pananghat

DECLARATION

I hereby declare that the matter embodied in the report entitled "Biochemical Studies of SARS-CoV-2 Non-structural Protein NSP14 and NiRAN Domain of NSP12" are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Saikrishnan Kayarat and the same has not been submitted elsewhere for any other degree

Muhammed Navas P
20181077
BS-MS 2018 Batch
IISER PUNE

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGMENT

I would like to express my deepest gratitude to my thesis supervisor Prof. Saikrishnan Kayarat for his help, guidance, and support during the project period. I am thankful for getting the opportunity to work under his expertise. I am also grateful to Dr. Gayathri Pananghat for all the insights she provided for my project as my thesis advisor.

It was a wonderful experience to work in SK Lab with the help of both SK Lab and G3 Lab members. I could not have undertaken this journey without the support of Dr. Om Prakash Chouhan, and I am thankful and happy that I had the opportunity to work under his mentorship. I want to thank Akhilesh Pratap Singh for the training he gave me in lab techniques during my semester project before final year. I also would like to thank Ashwin Uday for helping me with RTC protein and Ishtiyaq from JBU Lab for helping me with mass spectrometry.

CONTRIBUTIONS

Contributor name	Contributor role
Saikrishnan Kayarat, Om Prakash	Conceptualization Ideas
Muhammed Navas, Saikrishnan Kayarat, Om Prakash	Methodology
—	Software
Saikrishnan Kayarat, Om Prakash	Validation
Muhammed Navas	Formal analysis
Muhammed Navas	Investigation
Saikrishnan Kayarat	Resources
Muhammed Navas	Data Curation
Muhammed Navas	Writing - original draft preparation
Muhammed Navas, Saikrishnan Kayarat	Writing - review and editing
Muhammed Navas	Visualization
Saikrishnan Kayarat	Supervision
Saikrishnan Kayarat	Project administration
Saikrishnan Kayarat	Funding acquisition

ABSTRACT

Proofreading enzymes are essential for the proper functioning of many organisms, including viruses. In coronaviruses such as SARS-CoV-2, NSP14 carries out this function with the help of the N-terminal ExoN domain using its 3’ to 5’ exoribonuclease activity. CoVs cap their genomic RNA and mRNA to evade the host immune system. The NiRAN domain of NSP12 is proposed to play a role in this capping process. Exonuclease assay results in this project show that NSP14 prefers single-stranded RNA substrates to double-stranded RNA substrates, contradictory to the recently published papers. The NiRAN domain is inactive in the absence of the rest of the NSP12 domains. Mass spectrometry confirmed that NSP12, in its intact form, is able to transfer an AMP from an ATP molecule to NSP9. Attempts to crystallize the NiRAN domain and NSP14 were not successful. The nature of different substrates that can bind to the active site of the NiRAN domain is yet to be found out. Further experiments are required to be carried out in order to unravel the influence of NSP14 on the activities of NSP12 and NSP13.

Chapter 1: Introduction

Coronaviridae is a family of viruses consisting of members with single-stranded RNA genomes, and they are the largest known RNA viruses. This large family is divided into three subfamilies *orthocoronavirinae*, *Letovirinae* and *Pitovirinae*. Murine hepatitis virus (MHV), infectious bronchitis virus (IBV), and human CoVs (HCoVs) are some of the well-known representatives of the subfamily *Orthocoronavirinae*, which is commonly referred to as coronavirus (ICTV, 2023). Coronaviruses (CoVs) have been known to humankind for many decades and are capable of infecting and causing disease in humans and other animals (Leao JC et al., 2022). A human coronavirus was first identified by Tyrrell and Bynoe in nasal washings of adult patients with common cold in the year 1965 (Tyrrell DA et al., 1965). HCoV-NL63, HCoV-OC43, and HCoV-HKU1 are some of the coronaviruses that cause mild respiratory tract illness in humans (Forni D et al., 2017). The outbreak of severe acute respiratory syndrome (SARS) in 2002 caused by a highly pathogenic coronavirus SARS-CoV increased interest in research and studies related to coronaviruses which were until then considered pathogens with little to no threat to humans (Payne S, 2017). The outbreak of middle eastern respiratory syndrome (MERS) in 2012, caused by MERS-CoV, was the second major epidemic caused by a coronavirus (Zaki A M et al., 2012). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a viral pathogen responsible for the Covid-19 pandemic, is also a coronavirus classified in the genus betacoronavirus and it shows 79% genome similarity to SARS-CoV and 50% genome similarity to MERS-CoV (Cui J et al., 2019; Lu R et al., 2020). In some patients, infection with SARS-CoV-2 causes severe symptoms of flu that can escalate to fatal conditions such as pneumonia and renal failure (Harrison A G et al., 2020).

1.1. SARS-CoV-2: a Highly Pathogenic Coronavirus

The single-stranded, positive-sense RNA genome of SARS-CoV-2, which has an approximate size of 30 kilobases, consists of 14 open reading frames (ORFs). ORF1a and ORF1b occupy the initial two-thirds of the RNA genome at the 5'-end and encode non-structural proteins (NSPs) of the virus (Khailany R A et al., 2020). The first open reading frame ORF1a is translated into polyprotein pp1a, whereas

polyprotein pp1ab is synthesized from both ORF1a and ORF1b as a result of a ribosomal frameshift event, which takes place near the stop codon of ORF1a, promoted by programmed −1 ribosomal frameshifting element (PFSE) of SARS-CoV-2 (Roman C et al., 2021). These polyproteins are cleaved by two viral proteases, NSP3 and NSP5, to form functional non-structural proteins. Polyprotein pp1a is cleaved into non-structural proteins from NSP1 to NSP11, whereas pp1ab forms nonstructural proteins from NSP1 to NSP16 (Chen Y et al., 2020; Jin Y et al., 2022). Unlike structural proteins, NSPs play functional roles such as genome replication, proofreading, and methylation, which are crucial for the replication and transmission of the virus (Yadav R et al., 2021).

CoVs are roughly spherical in shape, and the diameter of the virions ranges from 118 nm to 140 nm. They possess non-segmented genomic RNA, with sizes ranging from 25 kb to 32 kb (Payne S, 2017). The coronavirus RNA genome encodes four structural proteins, namely nucleocapsid (N), membrane (M), envelope (E), and spike (S) proteins. The nucleocapsid protein binds to the genomic RNA and packages it, while the membrane proteins self-assemble through protein-protein interaction and enclose the nucleocapsid-bound genome (Tseng YT et al., 2010; Satarker S et al., 2020). Envelope protein is a small integral membrane protein that plays roles in pathogenesis and cytotoxicity (Ye Y and Hogue BG, 2007). It also exhibits characteristics of viroporins which show activities such as virus assembly and release (Liao Y et al., 2006). Spike protein is a large glycoprotein projecting out from the virus surface, and it helps in the recognition of host cell receptors (Satarker S et al., 2020). It binds to the cell receptor and induces virus-cell membrane fusion enabling the virus to penetrate the human body and cause infection (Huang Y et al., 2020).

The entry of SARS-CoV-2 into the host body commences when the receptor-binding domain (RBD) of the spike protein binds to angiotensin-converting enzyme 2 (ACE2), a cellular receptor present in the host cells (Wu F et al., 2020; Zhou P et al., 2020). After the successful entry into the host cell, the virus releases its genome into the cytoplasm, where the ORF1a and ORF1b are translated into the respective polyproteins and are cleaved into aforementioned individual non-structural proteins (Perlman S et al., 2009). Afterward, the viral components form double-membrane

vesicles by rearranging the endoplasmic reticulum to mediate the formation of genomic and subgenomic RNAs (sgRNA) through viral genome replication. sgRNA encodes accessory proteins and structural proteins, which are translocated into ER–Golgi intermediate compartment after translation for the assembly of the SARS-CoV-2 virion. After complete assembly and incorporation of the RNA genome, virions are released from the plasma membrane (Snijder E J et al., 2006; Wu H Y and Brian D A, 2010).

1.2. The Exoribonuclease NSP14

Nucleases are enzymes that cleave nucleic acids such as DNAs and RNAs into smaller fragments by hydrolyzing phosphodiester bonds, a strong and stable biochemical bridge between nucleotides (Nishino T et al., 2002). Nucleases cleaving DNA substrates are termed DNAses, and those cleaving RNAs, RNAses (Zuo Y and Deutscher M P, 2001). Exonucleases hydrolyze the phosphodiester bonds at either end of the substrate, cleaving nucleotides one at a time, while endonucleases act internally on the substrates (Mason P A et al., 2012). Restriction endonucleases cleave DNA substrates upon recognizing a specific nucleotide sequence at the restriction site, whereas a structure-specific endonuclease (SSE) identifies the secondary structure of its nucleic acid substrate (Loenen W A et al., 2014; Dehé P M et al., 2017). Most of the exonucleases do not have any sequence specificity, and they can act either in a 3’-5’ direction or a 5’-3’ direction (Mason P A et al., 2012). Exonucleases that process in the 5’-3’ direction, such as FEN-1, help in the removal of Okazaki fragments during DNA replication (Hosfield D J et al., 1998). 3’-5’ exonucleases, such as the ε subunit of DNA polymerase III, help in proofreading of newly synthesized nucleic acid chains, acting hand in hand with polymerase enzymes (Johnson A et al., 2005).

The fourteenth non-structural protein of SARS-CoV-2, NSP14, is a bifunctional protein consisting of an N-terminal exoribonuclease domain (ExoN) and a C-terminal methyltransferase domain (MTase). NSP14 is also present in other coronaviruses, such as SARS-CoV and MERS-CoV, and they exhibit structural and functional similarity (Ogando N S et al., 2020). The SARS-CoV-2 NSP14 has a molecular mass of 60 kDa. The two major domains of NSP14 are connected by a hinge domain

which controls their flexibility (Ferron F et al., 2018). The MTase domain is an N7-guanine methyltransferase that helps in the synthesis of viral RNA cap by methylating the guanine nucleotide at the 5'-end of the cap structure (Chen Y et al., 2009). MTase is capable of methylating cap analogs in the presence of S-adenosylmethionine (SAM), a methyl donor (Jin X et al., 2013). Recent studies suggest that NSP14 could potentially suppress interferon production and interferon signaling to help the virus fight the host immune system (Yuen C K et al.,2020).

The ExoN domain is a 3'-5' exoribonuclease that cleaves RNA substrates, and it does not affect DNA substrates (Baddock H T et al., 2022). It is proposed that ExoN works together with NSP12, the RNA-dependent RNA-polymerase (RdRp) of CoVs, as a proofreader during genome replication and transcription (Subissi L et al., 2014). Studies conducted in the CoV murine hepatitis virus show that NSP14 exceptionally increases the genomic replication fidelity (Eckerle L D et al., 2007). ExoN belongs to the DEDDh superfamily of exonucleases, and the family members have four invariant acidic amino acids distributed in three separate sequence motifs (Bernad A et al.,1989; Zuo Y and Deutscher M P, 2001). The four residues D90, E92, D243, and D273 in its active site form the metal binding catalytic core and are conserved across other coronaviruses (Saramago M et al., 2021). In proteins related to NSP14, the residue H268 is proposed to deprotonate water molecules for nucleophilic attacks, functioning as a general base (Hamdan S et al., 2003). The activity of this domain is dependent on divalent cations such as Mg^{2+} or $Mn^{2+,}$ and Mg^{2+} is the most effective for catalysis. The addition of EDTA, a popular chelating agent for divalent cations, during in vitro reactions can completely stop the exonuclease activity of NSP14. Non-structural protein NSP10 is necessary for the proper functioning of the ExoN domain (Saramago M et al., 2021).

NSP10 is a single-domain zinc-finger protein that has two zinc-binding motifs, and it is found to be very crucial for the exonuclease activity of NSP14. NSP10 has a highly conserved sequence, and it is present in all coronaviruses (Joseph J S et al., 2006; Lin S et al., 2021). While acting alone on a nucleic acid substrate, the exonuclease activity of NSP14 is considerably low, and this activity is increased when NSP10 forms a complex with NSP14 (Saramago M et al., 2021). The zinc fingers of NSP10 help the protein to interact with and bind to the N-terminus of NSP14. The N-terminal

end of the ExoN domain, which stays “closed” in the absence of NSP10, opens up upon interaction with NSP10, and this conformational change is believed to increase the catalytic activity of the protein (Imprachim N et al., 2023).

1.3. NiRAN: the Nucleotidyltransferase Domain of NSP12

NSP12, also a bifunctional protein, has a C-terminal RNA-dependent RNA-polymerase (RdRp) domain and an N-terminal “Nidovirus RdRp associated nucleotidyltransferase” (NiRAN) domain. These two domains are connected through an interface domain located between them (Kirchdoerfer R N et al., 2019). The RdRp domain is indispensable for genome replication, and it is found in almost all RNA viruses (Tang X et al., 2022). NSP12 binds with one NSP7 protein and two NSP8 proteins to form the replication-transcription complex (RTC) of SARS-CoV-2, which helps in the replication and transcription of genomic RNA (Peersen O B, 2019; Kirchdoerfer R N et al., 2019). Studies suggest that NSP7 and NSP8 could be acting as a primase during replication and transcription (te Velthuis A J et al., 2012). Unlike the RdRp domain, the NiRAN domain exhibits significant divergence in sequence, and it only has four conserved motifs; preA_N, A_N, $B_{N,}$ and C_N. This 250 residue domain acts as a genetic marker for the order *Nidovirales* since it is present in all the members of the order and is not seen in any other RNA viruses (Lehmann K C et al., 2015).

The 5’-end of genomic RNA and mRNAs of CoVs carry a $^{7Me}GpppA_{2'OMe}$ cap structure which safeguards the mRNAs and promotes the translation initiation of encoded proteins. This 5’-cap increases the resemblance of viral RNAs to mRNAs synthesized by the host cells and helps the virus to evade the host immune recognition (Daffis S et al., 2010). The NiRAN domain of SARS-CoV-2 has a kinase-like fold, and it can play a potential role in the synthesis of viral RNA cap structure (Yan L et al., 2022; Dwivedy A et al., 2021). A protein kinase is an enzyme that catalyzes the modification of other proteins by the covalent addition of phosphates to them in a process termed phosphorylation (Ubersax J A et al., 2007). NiRAN domain shows structural similarity to selenoprotein-O (SelO), a selenium-containing protein with a fold similar to that of a protein kinase. But SelO

protein does not exhibit phosphorylation activity. Hence it is named a pseudokinase (Dudkiewicz M et al., 2012; Sreelatha A et al., 2018).

Recent studies show that SelO protein is capable of transferring an AMP from ATP to protein substrates (Sreelatha A et al., 2018). The NiRAN domain of NSP12 also exhibits this activity by utilizing NSP9 as the protein substrate. In the presence of divalent metal cations such as Mg^{2+} or Mn^{2+}, the NiRAN domain can transfer a nucleotide monophosphate (NMP) from a nucleotide triphosphate (NTP) to the first N-terminal residue asparagine of NSP9, in a process termed an NMPylation (Wang B et al., 2021). In a similar fashion, the NiRAN domain can also remove a pyrophosphate from the 5'-end of a 5'-triphosphate RNA substrate (5'-pppRNA) through pyrophosphorolysis and transfer the newly formed 5'-monophosphate RNA (5'-pRNA) to the N-terminal asparagine residue of NSP9 in a process termed as RNAylation. This is proposed to be the initial step of viral RNA cap synthesis. In the presence of guanosine diphosphate (GDP), the NiRAN domain catalyzes the release of 5'-pRNA from NSP9, forming the product 5'-GpppRNA, the cap core structure. In vitro experiments show that RNA is released from NSP9 only in the presence of a guanosine-based nucleotide (Park, G J et al., 2022). The MTase domain of NSP14 can methylate the 5'-GpppRNA structure to form the cap-0 structure. Subsequent methylation of the cap-0 structure by NSP16, a 2′-O-MTase, leads to the formation of the final cap-1 structure (Viswanathan T et al., 2020).

1.4. Objectives

- To check the activity of NSP14 on different substrates
- To purify and crystallize NSP14(-6)
- To check the activity of NiRAN-domain protein
- To design assays to detect pyrophosphate release during NMPylation
- To crystallize NiRAN-domain protein

Chapter 2: Materials and Methods

2.1. Bacterial Transformation

Plasmids containing the gene of interest were transformed into *E.coli* cells through the heat-shock method or electroporation.

2.1.1. Heat shock transformation

Glycerol stocks of chemically competent cells were taken out from -80°C and were thawed on ice for 10 minutes. 200 ng of plasmid was added to the thawed cell stock and was kept on ice. After 20 minutes, it was incubated at 42°C for 90 seconds and was kept back on ice. After 10 minutes, 200 µL of LB-media was added to the tube containing cells and it was incubated at 37°C for 30 minutes by shaking at 500 RPM. After incubation, the cells were plated on an LB-agar plate containing appropriate antibiotics.

2.1.2. Electroporation

NEB turbo electrocompetent cells were used for transformation using this method. 200 ng of plasmid DNA and 150 µL of chilled autoclaved milli-Q were added to the tube containing the cells. Afterward, the contents of the tube were transferred to a 1mm electroporation cuvette. After incubating the cuvette on ice for 30 to 45 mins, it was electroporated under the following conditions: 1800 V voltage, 25 µF capacitance, and 200 Ω resistance. GenePulser Xcell electroporator by Bio-Rad was used for the procedure. 200 µL LB-media was added to the cuvette within 30 seconds of electroporation, and the cuvette was incubated at 37°C for 30 minutes. afterward, the cells were plated on an LB-agar plate containing appropriate antibiotics.

2.2. Plasmid Isolation

Plasmids were transformed into NEB turbo cells and plated on LB-agar plates containing appropriate antibiotics. After incubating the plates at 37°C overnight, isolated colonies were picked from the plates and were inoculated in 10 ml LB-media containing appropriate antibiotics. Culture tubes were incubated at 37°C by shaking

at 180 rpm overnight. The next day, cell cultures were centrifuged at 4000 rpm for 20 minutes (Centrifuge 5810 R, Eppendorf), and cell pellets were collected, discarding the supernatant. Plasmids were isolated from the cell pellet using the alkaline lysis method.

2.2.1. Manual method

	Buffer name	**Composition**
1	Solution I	50 mM Tris (pH 8.0), 10 mM EDTA and 100 ug/mL RNAse
2	Solution II	0.2 N NaOH, 1% (w/v) SDS
3	Solution III	3 M CH_3COOK (pH 8.0)

Table1: Alkaline lysis solutions and composition

Firstly, the cell pellet was resuspended in 300 µL solution I in 1.5 ml tubes. Then 300 µL of freshly prepared solution II was added to it. The tubes were gently inverted 5 times to mix the sample. Solution III was added after 4 minutes, and the contents were thoroughly mixed by inverting the tubes 5 times. The sample was spun at 14000 rpm for 10 minutes, and the supernatant was collected (Centrifuge 5424 R, Eppendorf). 900 µL of chilled ethanol was added into the supernatant, and it was again spun at 14000 rpm for 20 minutes. This time, the supernatant was discarded, and 900 µL 70% ethanol was added into the tube, which was then centrifuged at 14000 rpm for 10 minutes. Then the supernatant was discarded, and the pellet was dried by incubating at 55°C for 30 minutes. Afterward, the pellet was dissolved in 50 µL milli-Q.

2.2.2 Kit-based method

Plasmid isolation kit by QIAGEN was used for this method. The cell pellet was first resuspended using 250 µL buffer P1 in a 1.5 ml tube. 250 µL buffer P2 was added into this mix, and it was mixed by inverting 5 times. After 4 minutes, buffer N3 was added, and the tube was centrifuged at 14000 rpm for 20-30 minutes. After the spin, the supernatant was loaded in plasmid isolation columns by spinning at 14000 rpm for 5 minutes. The flowthrough was discarded, and 750 µL buffer QC was added into

the column, followed by spinning at 14000 rpm for 5 minutes. After discarding the flowthrough, 50 µL buffer EB was added into the column, and the plasmid was eluted by spinning at 14000 rpm for 5 minutes.

2.3. DNA Cloning

Restriction-ligation cloning and restriction-free cloning were used to change protein tags, switch the gene from one vector to another, and introduce mutations in the gene.

2.3.1. Restriction-ligation cloning

pGEX vector backbone digested by restriction enzymes NdeI and BamHI was used for restriction-ligation cloning. The vector backbone was also treated with TSAP (Promega). Insert containing the gene was isolated from the plasmid using NdeI and BamHI restriction enzymes (NEB). The reaction mix contained 150 ng DNA insert, 50 ng vector backbone, 1uL T4 ligase (NEB), and 1x ligase buffer. The same reaction mix without DNA insert was used as a control. After incubation at 25°C for 3 hours, control and test samples were transformed into NEB Turbo cells. Transformed cells were plated on LB agar plates containing ampicillin, and the plates were incubated at 37°C overnight. The next day, isolated colonies were picked from the plates and were inoculated in 10 mL LB-media with ampicillin, followed by overnight incubation at 37°C with shaking at 180 rpm. Afterward, the cell culture was centrifuged at 5000 rpm for 30 minutes at 4°C, and the cell pellet was collected, discarding the supernatant.

Restriction-ligation was used to remove 6xHis-tag from the N-terminus of NSP14 and NiRAN-domain clones with N-terminal GST-tag.

- Clones made using this method:
    - Nsp14(-6) with N-terminal GST-tag and TEV cleavage site in pGEX vector
        - GST - TEV site - Nsp14(-6)
    - NiRAN with N-terminal GST-tag and TEV cleavage site in pGEX vector
        - GST - TEV site - NiRAN

2.3.2. Restriction-free (RF) cloning

	Component	Final Concentration
1	Pfu buffer	1x
2	dNTP mix	200 µM
3	Forward primer	0.4 µM
4	Reverse primer	0.4 µM
5	Template DNA	2 ng/µL
6	Pfu polymerase	0.5 µL in 50 µL

Table 2: Components of primary PCR reaction

	Component	Final Concentration
1	Pfu buffer	1x
2	dNTP mix	300-400 µM
3	mega-primer	20 ng/µl
4	Template DNA	2 ng/µl
5	Pfu polymerase	0.5 µL in 50 µL

Table 3: Components of secondary PCR reaction

RF cloning was used to either introduce a mutation in the gene or to insert the gene in a different vector of interest. In the first step, a mega-primer was produced through PCR reaction using a forward primer and a reverse primer (primary PCR reaction) under the following conditions: denaturation at 95°C for 30 s, annealing at 60°C for 30s and extension at 72°C for a duration depending on the size of the amplified product (1 min/kb). Mastercycler X50s by Eppendorf was used for PCR reactions.

Components for the primary PCR reaction were used as mentioned in Table 2. After the PCR reaction, the presence of an amplified product was confirmed using agarose gel electrophoresis, and the product was cleaned up using a PCR-purification kit (QIAGEN). 100 µL of PCR product was loaded in a 50 QIAquick Spin Column along with 500 µL buffer PB and the column was spun at 14000 rpm for 3 mins. After adding 750 µL buffer PE into the column, it was spun at 14000 rpm for 5 mins, and the flowthrough was discarded. Afterward, the DNA bound to the column was eluted using 50 µL buffer EB.

In the second step, the mega-primer from the primary PCR was used to amplify the complete plasmid vector (secondary PCR). Secondary PCR was carried out in the following conditions: denaturation at 95°C for 45 seconds, annealing at 51°C for 45 seconds, and extension at 72°C for a duration of 1 min/kb. Components for the secondary PCR reaction were used as mentioned in Table 3. A control was also kept during secondary PCR, in which mega-primer was not added. After the PCR, the success of the reaction was confirmed using agarose gel-electrophoresis, and the products were subjected to Dpn1 digestion at 37°C for 3 hours. Afterward, the digestion product was transformed into NEB turbo electrocompetent cells through electroporation. Plasmids were isolated from the cell culture as mentioned before.

RF cloning was used to insert the NSP9 gene into pRSFDuet-1 and to introduce D208A and D218A point mutations in the NiRAN domain of NSP12.

2.4. Protein Expression Check

Plasmids containing the gene of interest were transformed into protein expression strains of *E.coli* bacterial cells (Rosetta 2, BL21 (DE3), BL21-AI, C43, etc.) through heat shock transformation. The transformed cells were plated on an LB-agar plate containing the appropriate antibiotics, and the plate was incubated at 37°C for 12 to 14 hours. Cells were scraped from the plates the next day and inoculated into two culture tubes containing 10 mL LB-media with the appropriate antibiotics. Tubes were then incubated at 37°C by shaking at 180 rpm until the OD_{600} reached 0.6. One of the tubes was left uninduced, keeping it as a control, and the other one was induced with an appropriate amount of IPTG. Afterward, the culture tubes were

incubated at 18°C overnight at 180 rpm. The next day, the cell culture was centrifuged at 5000 rpm for 20 minutes, and the supernatant was discarded. The cell pellet collected from uninduced cell culture was resuspended in 5x TGS buffer containing 125 mM Tris, 960 mM Glycine, and 0.5% SDS (sample 1). Cell pellet collected from induced cell culture was resuspended in TGS buffer (sample 2) and lysis buffer separately. The samples were then sonicated for two minutes with 1-second “on” and 3-second “off” settings at 60% amplitude (Vibra-cell, Sonics). The cell pellet resuspended in the lysis buffer was centrifuged at 12000 rpm for 10 minutes. The supernatant was collected (sample 3), and the lysed pellet was resuspended in the TGS buffer (sample 4). 3x SDS loading dye was added to all four samples, and the samples were run on an SDS polyacrylamide gel at 200 V. The gel image was then captured using E-Gel Imager by Life Technologies.

2.5. Protein Purification

This section explains the general protocol used for protein purification. Variables specific to each protein, such as type of antibiotic, amount of IPTG used for induction, and buffer solutions used for each column, are mentioned afterward.

	Buffer name	**Composition**
1	Lysis Buffer	50 mM Tris (pH 7.5), 500 mM NaCl, 5 mM MgCl2, 10% glycerol, 15 mM Imidazole
2	Buffer A	50 mM Tris (pH 7.5), 500 mM NaCl, 15 mM Imidazole
3	Buffer B	50 mM Tris (pH 7.5), 500 mM NaCl, 500 mM Imidazole
4	Buffer GA	50 mM Tris (pH 7.5), 500 mM NaCl
5	Buffer GB	50 mM Tris (pH 7.5), 500 mM NaCl, 20 mM Glutathione
6	Dialysis D	50 mM Tris (pH 8), 50 mM NaCl, 1 mM DTT
7	Buffer B_{50}	50 mM Tris (pH 8), 50 mM NaCl, 1 mM DTT
8	Buffer B_{1000}	50 mM Tris (pH 8), 1000 mM NaCl, 1 mM DTT
9	Buffer S	25 mM Tris (pH 8), 150 mM NaCl, 0.1mM MgCl2, 1 mM DTT

Table 4: General composition of buffers used for protein purification.
Variations for each protein are mentioned below.

2.5.1. Protein expression and sonication

Plasmid containing the gene of interest was transformed into appropriate cells of *E.coli* strains, as per the result of the expression check, and the cells were plated on LB-agar plates. To grow the primary culture, cell colonies were inoculated into 100 mL LB-media. The primary culture was incubated at 37°C for 2-3 hours. Four to six conical flasks containing 1-liter LB-media each were inoculated using 10 mL of primary culture in each flask to grow the secondary culture. The flasks were incubated at 37°C until the optical density (OD_{600}) reached the optimum level. According to the expression checks, all the proteins purified were showing expression when induced at 0.6 OD. The OD was checked using a photometer (Biophotometer, Eppendorf). The secondary cultures were induced with a suitable amount of IPTG at OD 0.6 and were grown overnight, incubating at 18°C and 180 rpm. In the morning, the cell cultures were pelleted by centrifuging at 5000 rpm for 30 minutes (Avanti J-26S XP, Beckman Coulter). The cell pellet was resuspended in lysis buffer and was sonicated with 1-second “on-time” and 3-second “off-time” for 3 minutes at 4°C (Vibra-cell, Sonics). The 3-minute cycles were repeated 2-3 times until the cells were properly lysed. Afterward, the lysis solution was centrifuged at 15000 rpm for 60 minutes (Avanti J-26S XP, Beckman Coulter), and the supernatant (lysate) was collected. The lysate was passed through either a Ni-NTA column (5 mL HisTrap HP, Cytiva) or a GST column (5 mL GSTrap, Cytiva) using GE Akta Prime FPLC as the first step of protein purification.

2.5.2. Affinity chromatography: Ni-NTA column

HisTrap column was washed using 40 ml of 100% buffer B at a flow rate of 2 ml/min, followed by equilibration using 80 ml of buffer A at the same flow rate. Then the lysate collected after sonication was passed through the column at a 1.5 ml/min flow rate. After loading the lysate, 80 ml buffer A was passed to remove contaminants loosely bound to the column. The protein of interest bound to the Ni-NTA column was eluted using 10 ml each of 5%, 10%, 20%, 40%, 60%, and 100% buffer B, which were collected in 6 different fractions. 10 µL from each fraction, along with load and flowthrough, was collected in 8 different tubes and was treated with 5 µL 3x

SDS-loading dye. Samples were then loaded and run on an SDS-PAGE gel. After the electrophoresis was finished, the gel was stained using a coomassie brilliant blue solution and heated until the solution started boiling. Then the gel was kept on a rocker for 10 minutes. After properly staining the gel, the coomassie blue solution was discarded, and the gel was kept in a destain solution containing 20% ethanol and 10% glacial acetic acid. The gel with the destain solution was heated until it started boiling and kept on a rocker until it was properly destained. The gel image was captured using E-Gel Imager.

2.5.3. Affinity chromatography: GSTrap column

GSTrap column was washed using 40 ml 100% buffer GB at a flowrate of 2 ml/min, followed by equilibration using 80 ml of buffer GA at the same flow rate. Then the lysate collected after sonication was passed through the column at a 1.5 ml/min flow rate. After loading the lysate, 80 ml buffer A was passed to remove contaminants loosely bound to the column. Afterward, the GST-tagged protein bound to the column was eluted using 30 ml buffer GB, collected in three fractions of 10 ml each. The presence of a protein of interest was confirmed using SDS-gel electrophoresis.

2.5.4. Dialysis and tag cleavage

After affinity chromatography, the eluate was kept for dialysis overnight. Fractions containing the protein were pooled together and were kept for dialysis using a dialysis membrane in two liters of buffer D overnight. If the protein had any purification tag that needed to be removed, it was cleaved off during this step. Either TEV protease (1:10 protease to protein ratio) or Ulp1 protease (1:100 protease to protein ratio) was added to the protein solution depending on the cleavage site, during dialysis. The next day, the protein solution was taken out and was centrifuged at 15000 rpm for 20 minutes at 4°C (Avanti J-26S XP, Beckman Coulter) to remove any protein precipitate, and the supernatant was collected.

2.5.5. Ion-exchange chromatography: MonoQ and MonoS columns

MonoQ 10/100 GL or MonoS 10/100 GL (GE Healthcare) columns were used as anion exchange and cation exchange columns, respectively. The column was washed using 40 ml buffer B_{1000} at a 1 ml/min flow rate, followed by equilibration

using 80 ml buffer B_{50} at the same flow rate. The protein solution was then loaded in the column using a superloop at a flow rate of 1 ml/min. Afterward, the bound protein was eluted using an increasing gradient of buffer B_{1000} from 0% to 100%. Samples were collected from the eluted fractions corresponding to the peaks on the chromatogram and were run on an SDS polyacrylamide gel. Fractions containing protein of interest were pooled together, and the solution was concentrated to a volume less than 1 ml using a protein concentrator (Vivaspin Turbo, 5 kDa or 10 kDa) by centrifuging at 5000 rpm (Centrifuge 5810 R, Eppendorf).

2.5.6. Size-exclusion chromatography: Superdex 200 and Superdex 75

Superdex 75 10/300 GL or Superdex 200 10/300 GL columns were used for size-exclusion chromatography. Before loading the protein into the column, it was equilibrated using 48 ml of buffer S. Superdex 75 was operated at a flow rate of 0.8 ml/min and Superdex 200 at 0.3 ml/min. Protein solution concentrated to a volume below 1 mL was injected into the system using a syringe. The protein solution was passed through the column using the buffer S. Samples were collected from the eluted fractions corresponding to the chromatogram peaks and were analyzed using an SDS-polyacrylamide gel. Fractions containing the desired protein were pooled together, and the protein solution was concentrated using a protein concentrator (Vivaspin Turbo, 5 kDa or 10 kDa). 5 µL aliquots were made from the concentrated protein, and the aliquots were flash-frozen in liquid nitrogen and stored at -80°C.

2.5.7. Specific Conditions for Each Protein

NSP14 (WT)

Plasmid construct : [GST-TEVsite-NSP14(WT)-6xHis] in pGEX vector
Expression cells : Rosetta 2(DE3)pLysS
Antibiotics : 100 µg/mL ampicillin and 25 µg/mL chloramphenicol
Inducer : 0.5 mM IPTG
Columns used : Ni-NTA, MonoQ, and Superdex200
Protease : TEV protease

Variation from buffer composition mentioned in table 4:

- The Lysis buffer used had additional 1 mM DTT.

NSP14 (-6): Ni-NTA Column

Plasmid construct: [6xHis-GST-TEVsite-NSP14(-6)] in pGEX vector

- Conditions were the same as NSP14 (WT) protein.

NSP14 (-6): GST Column

Plasmid construct: [GST-TEVsite-NSP14(-6)] in pGEX vector

- GST column was used in the first step of purification along with buffer GA and buffer GB. Everything else remained the same as NSP14 (WT) purification.

NSP10

Expression cells : Rosetta 2(DE3)pLysS
Antibiotics : 50 µg/mL kanamycin and 25 µg/mL chloramphenicol
Inducer : 0.5 mM IPTG
Columns used : Ni-NTA, MonoQ, and Superdex75

- Buffer S was used without $MgCl_2$.

NiRAN: Ni-NTA

Plasmid construct : [6xHis-GST-TEVsite-NSP14(WT)] in pGEX vector
Expression cells : Rosetta 2(DE3)pLysS
Antibiotics : 100 µg/mL ampicillin and 25 µg/mL chloramphenicol
Inducer : 0.5 mM IPTG
Protease : TEV protease
Columns used : Ni-NTA, MonoQ, and Superdex75

- After cleaving the tag using TEV protease, the protein solution was passed through the Ni-NTA column again.

NiRAN: GST Column

Plasmid construct : [GST-TEVsite-NSP14(WT)] in pGEX vector
Expression cells : Rosetta 2(DE3)pLysS
Antibiotics : 100 µg/mL ampicillin and 25 µg/mL chloramphenicol

Inducer : 0.5 mM IPTG

Protease : TEV protease

Columns used : GST column, MonoQ, and Superdex75

NSP9

Plasmid construct : [6xHis-SUMO-NSP9] in pRSFDuet-1

Expression cells : Rosetta 2(DE3)pLysS

Antibiotics : 50 µg/mL kanamycin and 25 µg/mL chloramphenicol

Inducer : 0.5 mM IPTG

Protease : Ulp1

Columns used : Ni-NTA, MonoS, and Superdex75

- Buffer D was used with 1 mM EDTA and Buffer S without $MgCl_2$.
- pH of the buffers B_{50} and B_{1000} was 7.5

NSP12

Plasmid construct : [6xHis-SUMO-NSP12] in pRSFDuet-1

Expression cells : Rosetta 2(DE3)pLysS

Antibiotics : 50 µg/mL kanamycin and 25 µg/mL chloramphenicol

Inducer : 0.5 mM IPTG

Protease : Ulp1

Columns used : Ni-NTA, MonoQ, and Superdex75

- Lysis buffer had 0.2% CHAPS. Buffer D, Buffer $B_{50,}$ and Buffer B_{1000} were used with 1 mM EDTA. Buffer S was used without $MgCl_2$.
- After tag cleavage, the protein solution was passed again through the Ni-NTA column.

2.6. Exonuclease Assay

Activity check of NSP14 was carried out in an ExoN buffer containing 25 mM Tris (pH 7.5), 5 mM $MgCl_2$, 50 mM NaCl, 5% glycerol, and 1 mM DTT. NSP14 and NSP10, along with the nucleic acid substrate, were added into a 10 µL reaction mix containing ExoN buffer in 0.5 mL tubes. The concentration of 5’-FAM labeled substrates was 10 nM. The reaction mix was incubated at 37°C for 45 minutes. The reaction was stopped by adding 5 µL of stop buffer (95% formamide (v/v), 10 mM

EDTA, and 0.25% bromophenol blue). Afterward, the tubes were incubated at 95°C for 5 minutes. The contents of the tubes were loaded in a 20% polyacrylamide gel containing 6 M urea, and the gel was run at 200 volts. The gels were visualized and captured using the Amersham Typhoon laser scanner.

2.7. NMPylation Reaction

NMPylation reaction was carried out in a buffer containing 25 mM Tris (pH 7.5), 10 mM KCl, 5 mM $MgCl_2$, and 1 mM DTT. The composition of the 20 µL reaction mix in a 0.5 mL tube was as follows: 1x buffer, 50 nM NiRAN-domain protein, NSP12 or RTC, 150 µM NSP9, and 300 µM ATP. The reaction was started by the addition of ATP. In order to detect the pyrophosphate release using malachite assay, it was digested into phosphates using pyrophosphatase. 1 µL of pyrophosphatase (by NEB) was added to the reaction mix, and it was incubated at 37°C for 2 to 3 hours. Afterward, 50 µL of malachite-green reaction solution was added to the reaction mix, and absorbance at 630 nm was measured using a microplate reader (CLARIOstar, BMG Labtech).

To prepare the malachite green reaction solution, 0.12% malachite green dye (dissolved in 3N sulphuric acid), 7.5% ammonium molybdate, and 11% tween-20 were used and the components were mixed in a 10:2.5:0.2 ratio. A concentration gradient of NaH_2PO_4, ranging from 0 to 500 µM, was used to plot the standard curve.

2.8. Protein Crystallization

Protein solution concentrated to 10 mg/ml was used for crystallization. In the sitting-drop vapor diffusion method using 96-well plates, 60 µL reservoir solution was added into each well (reservoir chamber). Different plates were used for each screen. 100 nL protein solution along with 100 nL reservoir solution was added to the drop shelf using a Mosquito crystallization robot. Afterward, all the chambers were sealed with clear tape, and the plates were kept at 18°C for at least four days.

To reproduce crystals obtained in the sitting drop method, the reservoir solutions were manually prepared and were used for the “hanging drop method” in 24-well plates. 600 µL of manually prepared reservoir solutions were added into each

chamber. 1 µL of protein solution along with 1 µL of reservoir solution was added on a cover slip. Each well was sealed with these coverslips, and the plates were kept at 18°C for at least one week.

Chapter 3: RESULTS

3.1. Check for Positive Clones

After isolating the plasmid from NEB Turbo cells, they were digested using appropriate type II restriction enzymes and were run on 0.8% agarose gels at 90-110 volts.

3.1.1. Restriction-free cloning

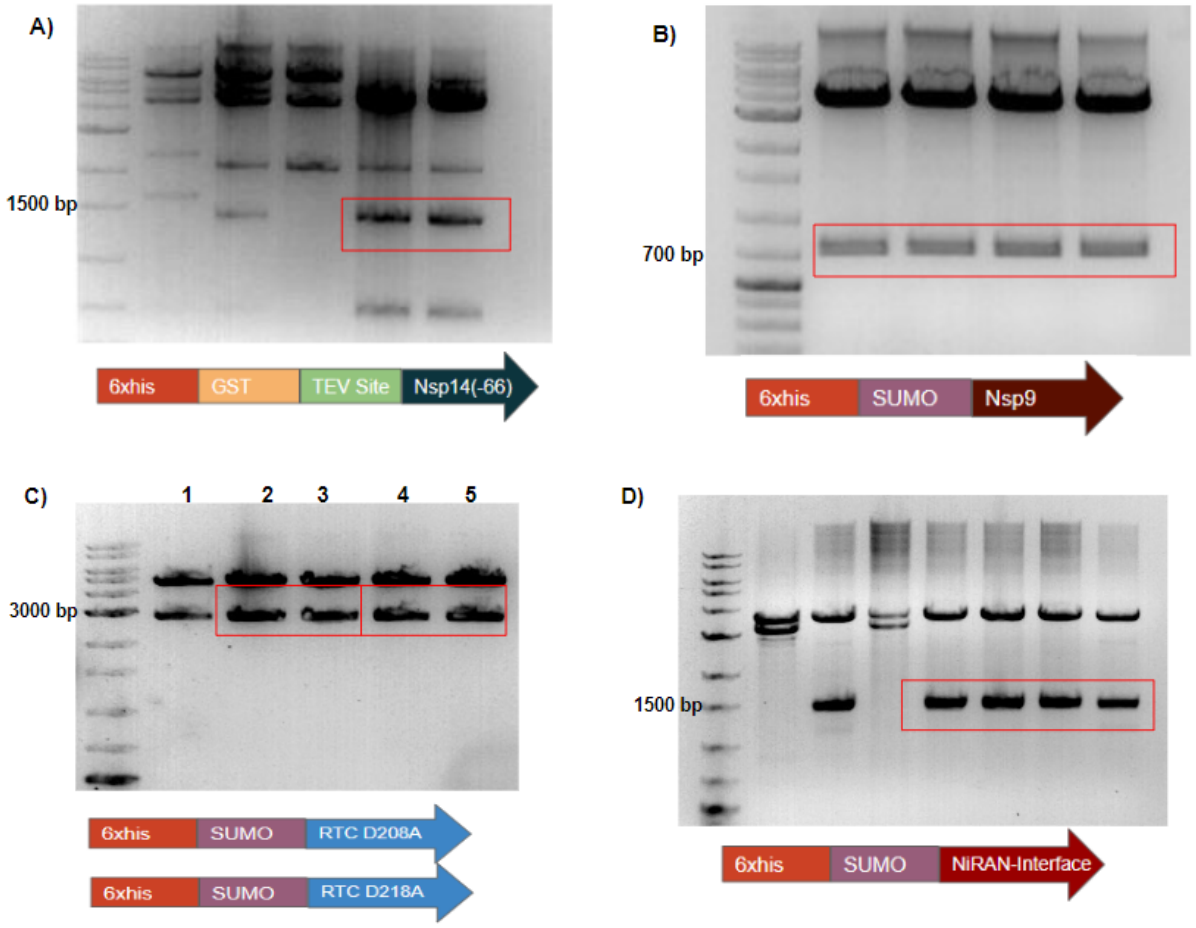

Fig. 1: Agarose gel images of the clone check (RF cloning)

Bands corresponding to positive clones are marked in red.

Fig. 1A is the clone check for NSP14(-66) construct. The first 66 amino acids of NSP14 were deleted in this clone. Gene was cloned in the pGEX vector, and plasmid was digested using NdeI and BamHI for the clone check. Fig. 1B shows the clone check for NSP9 inserted into the pRSFDuet-1 vector with N-terminal SUMO-tag.

Plasmid was digested using NcoI and XhoI. In fig. 1C, lane 1 is control, lane 2 and 3 are D208A NiRAN-domain mutants on NSP12/NSP7/NSP8 complex (RTC) and lane 4 and 5 are D218A NiRAN domain mutants. Here the gene is in pRSFDuet-1, and NdeI and XhoI were used for plasmid digestion. Fig. 1D shows positive clones for the NiRAN-Interface protein, consisting of the first 366 N-terminal residues of NSP12 in the pRSFDuet-1 vector. Plasmid was digested using NdeI and XhoI for this clone. Arrow representations in the figures show different purification tags in a 3' to 5' direction.

3.1.2. Restriction-ligation cloning

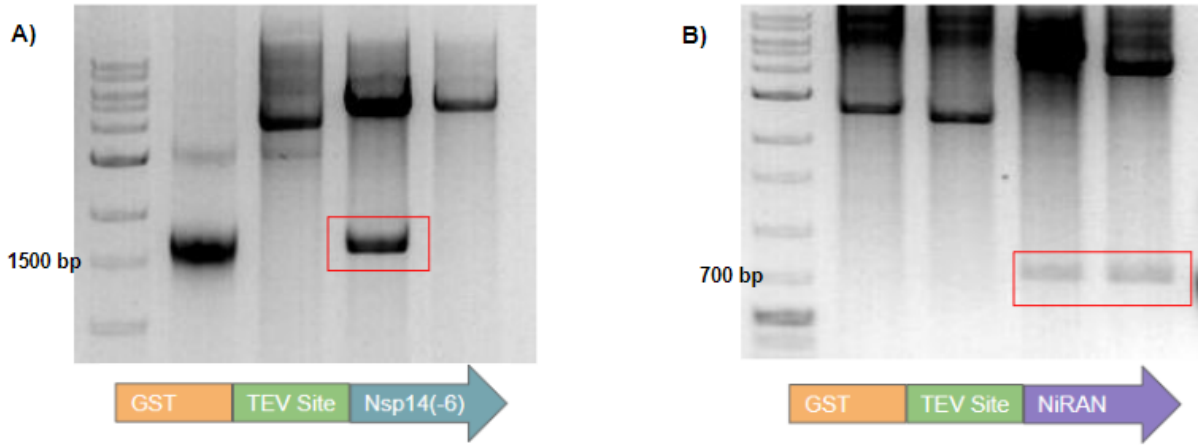

Fig. 2: Agarose gel images of the clone check (Restriction-ligation)
Bands corresponding to positive clones are marked in red.
N-terminal 6xHis-tags were removed from NSP14(-6) and NiRAN domain proteins through Restriction-ligation cloning. NSP14(-6) is an NSP14 mutant in which the first 6 amino acid residues were deleted. Both genes were cloned in the pGEX vector, and the plasmids were digested using NdeI and BamHI.

3.2. Expression Check for Different Proteins

Protein expression check was carried out using the protocol mentioned in the materials and methods. Specific details of expression checks of different proteins are mentioned in this section.

3.2.1. Optimising expression of NSP14(-66) clone

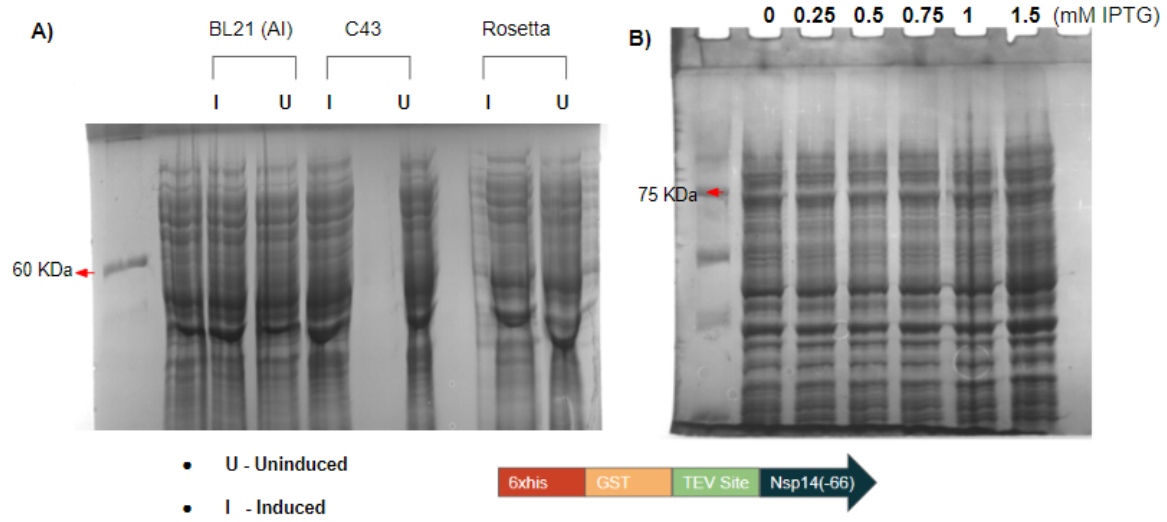

Fig. 3: Expression gel images of NSP14 (-66)

A) Expression check in different strains of *E.coli* as mentioned in the figure. BL21-AI cells were induced using 0.2% L-Arabinose. C43 and Rosetta 2(DE3) cells were induced using 0.5 mM IPTG. **B)** Expression checks in Rosetta 2(DE3) cells using different concentrations of IPTG from 0 to 1.5 mM for induction. After induction, the culture was grown at 18°C overnight. Protein is not showing expression under any of the conditions tested.

3.2.2. Expression check for NSP14(-6) clone

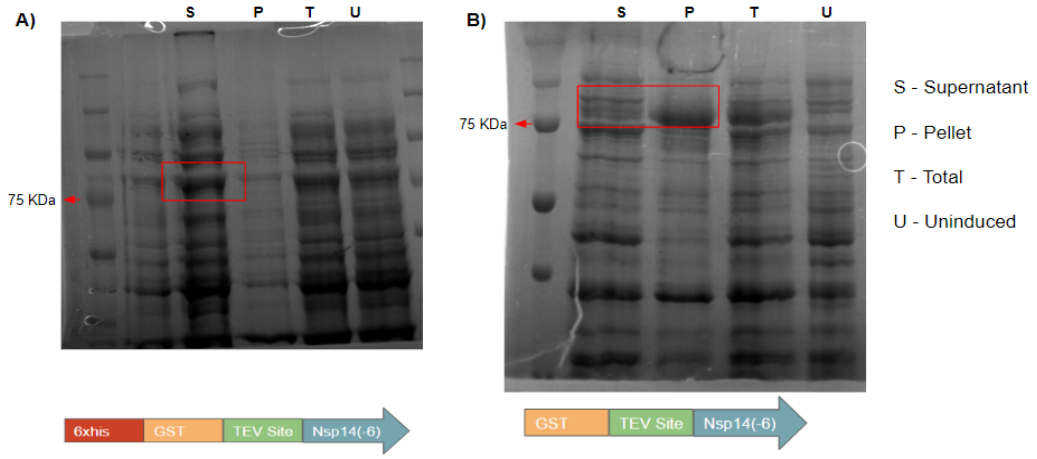

Fig. 4: Expression gel images of NSP14 (-6)

A) Expression check of NSP14(-6) clone with 6xHis tag in the N-terminal. **B)** Expression check of NSP14(-6) clone without 6xHis tag in the N-terminal.

The expression check was carried out using Rosetta 2(DE3) cells. The culture was induced using 0.5 mM IPTG at 0.6 OD and was grown at 18°C overnight after induction. The supernatant, pellet, and “total” are for induced cell cultures. “Total” consists of both supernatant and pellet. The supernatant and pellet were separated after sonication by spinning the tube at 14000 rpm for 10 minutes at 4°C. The red markings show the bands of interest.

3.2.3. Expression checks for NiRAN domain clones

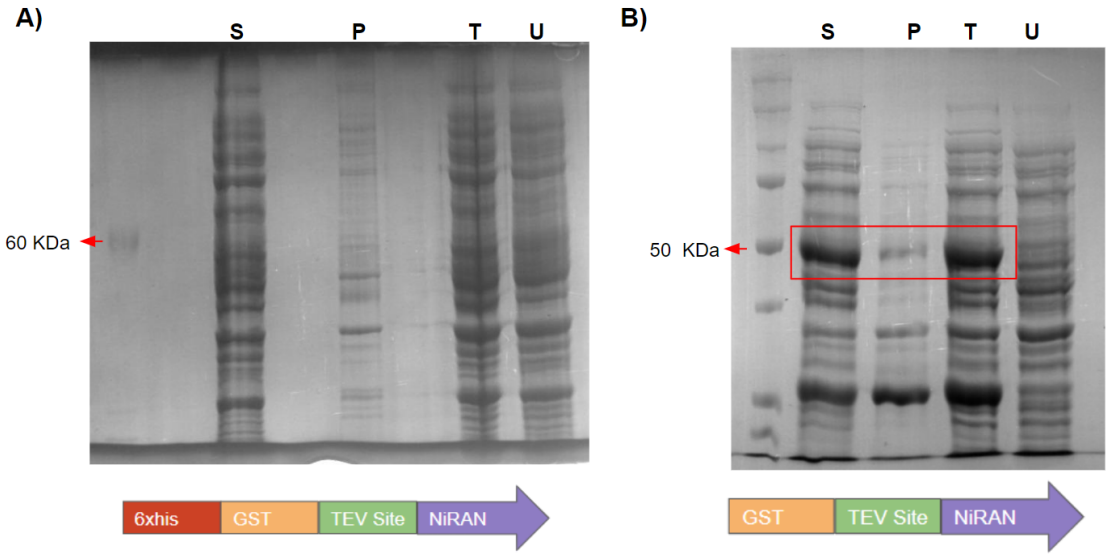

Fig. 5: Expression gel images of NiRAN

A) Expression of NiRAN domain clone with N-terminal 6xHis tag. C43 strain of *E.coli* was used for the expression check. The culture was induced at 0.6 OD using 0.5 mM IPTG. Cell culture was incubated at 18°C overnight after induction. **B)** Expression of NiRAN domain clone without N-terminal 6xHis tag. Protein was expressed in Rosetta 2(DE3) cells. The culture was induced at 0.6 OD using 0.5 mM IPTG. Cell culture was incubated at 18°C overnight after induction. Bands of interest are marked in red.

3.2.4. Expression check for NSP9

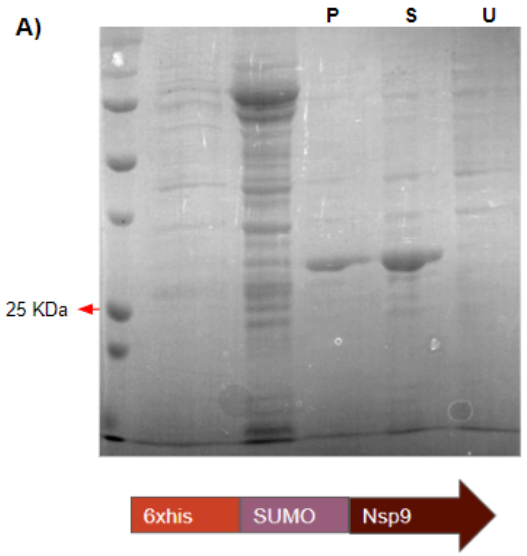

Fig. 6: Expression gel image of NSP9

NSP9 with sumo tag was used for purification since this would be helpful to maintain the native N-terminus of the protein after tag cleavage using Ulp1. Rosetta 2(DE3) cells were used for expression. The culture was induced using 0.5 mM IPTG at 0.6 OD and was incubated at 18°C overnight.

3.3. Protein Purification

Proteins were purified using the protocol explained in the materials and methods section. Purification was carried out at 4°C either in a cold cabinet or a cold room.

3.3.1. Purification of NSP14(WT)

Wildtype NSP14 cloned in pGEX vector with a GST-tag followed by a TEV cleavage site on the N-terminus and a 6xHis-tag on the C-terminus was used for this purification. (GST-TEV-NSP14-6xHis)

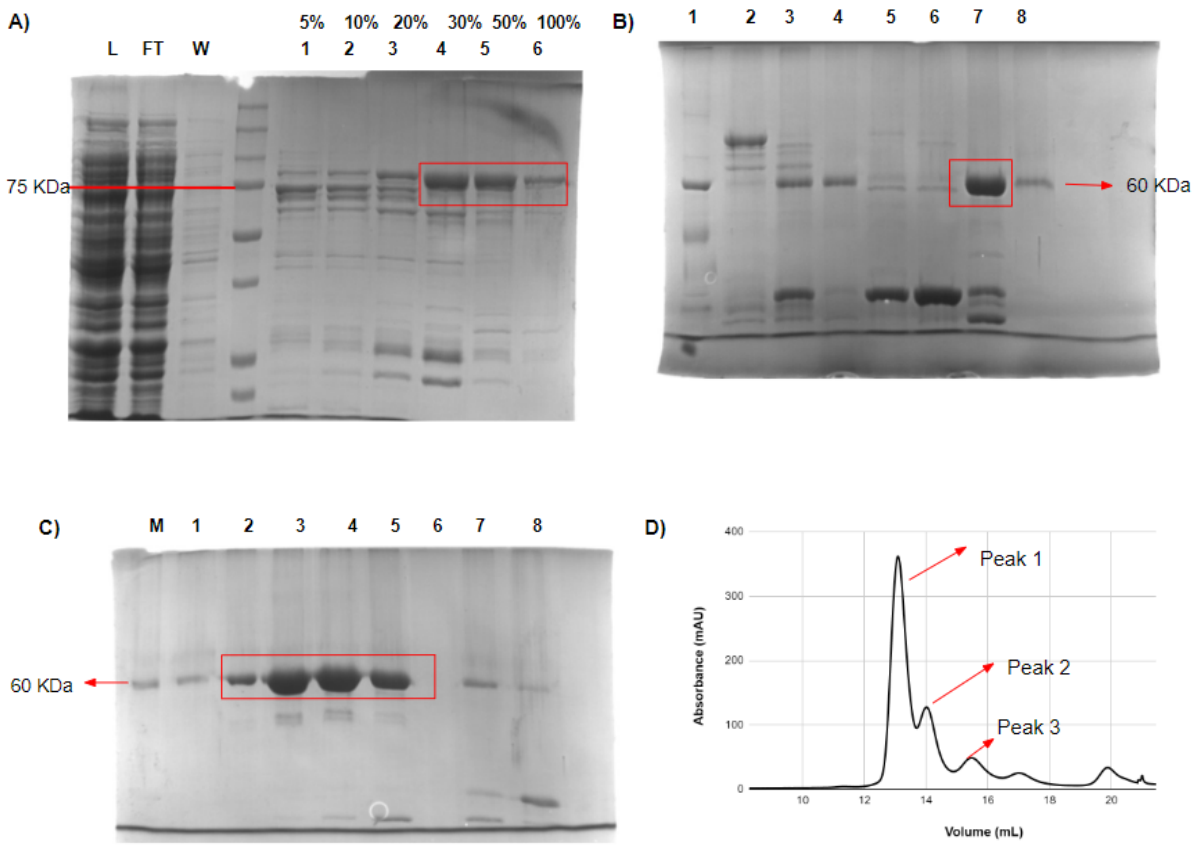

Fig. 7: Purification gel images and chromatogram of NSP14

A) SDS-PAGE gel of Ni-NTA purification. Lane "L" is the cell lysate loaded on the column, and Lane "FT" is the flowthrough. Lane "W" shows protein contents eluted during the wash step using buffer A. Lanes 1 to 6 are eluates eluted at a concentration of buffer B mentioned above each lane in percentage. **B)** SDS-PAGE gel of MonoQ column. Lane 2 shows the Ni-NTA eluate. Lane 3 shows the TEV protease cleavage product that was loaded in the MonoQ column, while lanes 5 and 6 show MonoQ eluates. Lane 7 is the flowthrough. **C)** SDS-PAGE gel of Superdex200 column. Lanes 1 to 5 correspond to the first peak of the Superdex 200 chromatogram, while lanes 7 and 8 correspond to the second and third peaks of the chromatogram. **D)** Superdex200 chromatogram showing absorbance at 280 nm. Bands of interest in each gel are marked in red.

3.3.2. Purification of NSP14(-6)

NSP14 without the first 6 residues on the N-terminus was used for this purification. Protein has a GST-tag followed by a TEV cleavage site on the N-terminal. Protein was expressed in Rosetta 2(DE3) cells.

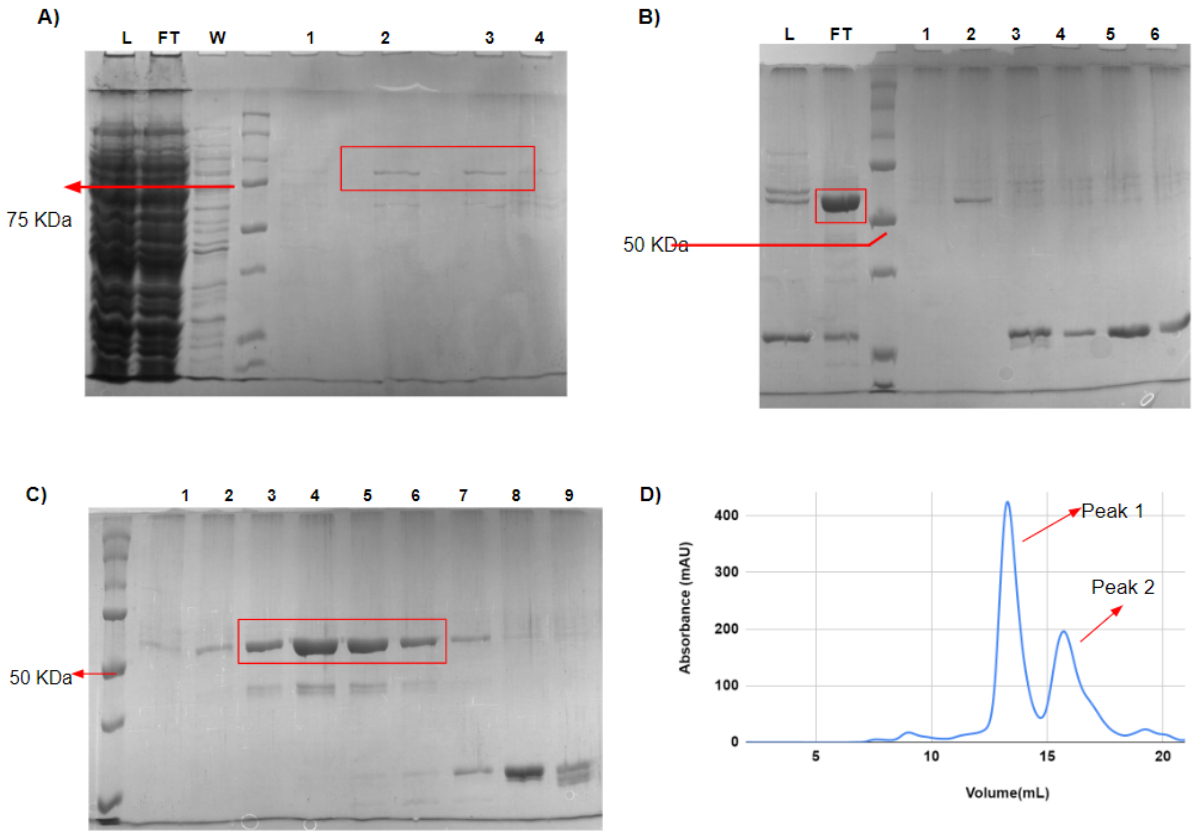

Fig. 8: Purification gel images and chromatogram of NSP14(-6)

A) The cell lysate was passed through the GSTrap column after sonication. Protein bound to the column was eluted using 100% buffer GB. Lanes 2 and 3 correspond to elution fractions containing the protein of interest. **B)** After TEV protease tag cleavage, the protein solution was passed through the MonoQ column. Protein did not bind to the column and came out with the flowthrough. Lane "FT" shows concentrated flowthrough from the MonoQ column. Lanes 1 to 6 show impurities bound to the column eluted with an increasing concentration of buffer B_{1000}. **C)** flowthrough from MonoQ was concentrated using Vivaspin Turbo 10 kDa protein concentrator to a volume below 1 mL, and the same was passed through the Superdex 200 column. Lanes 1 to 6 correspond to peak 1, and lanes 7 to 9 correspond to peak 2 of the Superdex 200 chromatogram. **D)** Superdex 200 chromatogram.

3.3.3. Purification of NSP10

NSP10 protein with N-terminal 6xHis-tag cloned in the pHIS vector was used for this purification. Protein was expressed in Rosetta 2(DE3) cells.

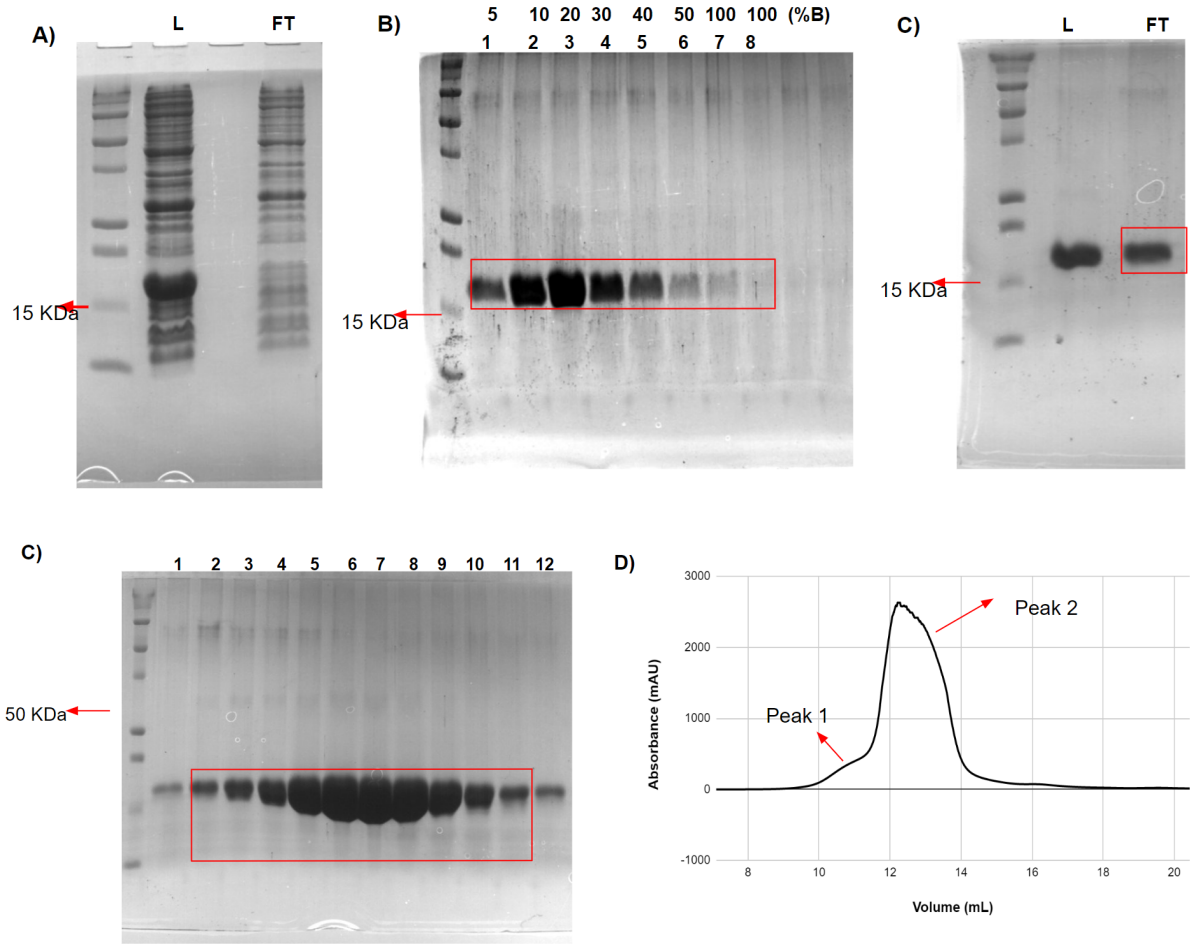

Fig. 9 Purification gel images and chromatogram of NSP10

A) Cell lysate containing the protein was first passed through the Ni-NTA column. Lane "L" shows the composition of lysate loaded in the column, and lane "FT" shows the flowthrough of the column. **B)** This gel shows the composition of different elution fractions of the Ni-NTA column. The percent of buffer B used for the elution of each fraction is mentioned above the lanes. **C)** Eluate of the Ni-NTA column containing NSP10 was loaded in the MonoQ column, but the protein did not bind to it. Lane "L" and Lane "FT" show load and flowthrough of the MonoQ column, respectively. **C)** MonoQ flowthrough was concentrated to a volume below 1 mL using 5 kDa Amicon Ultra-15 protein concentrators, and it was passed through the Superdex 75 column. Lanes 1 to 4 in the gel image in this figure correspond to peak 1 of the chromatogram, and lanes 5 to 12 correspond to peak 2. **D)** Chromatogram of Superdex 75.

3.3.4. Purification of NiRAN-domain using Ni-NTA column

NiRAN domain protein with N-terminal 6xHis-tag followed by GST-tag and TEV site cloned in pGEX vector was used for this purification. Protein was expressed in *E.coli* C43 cells.

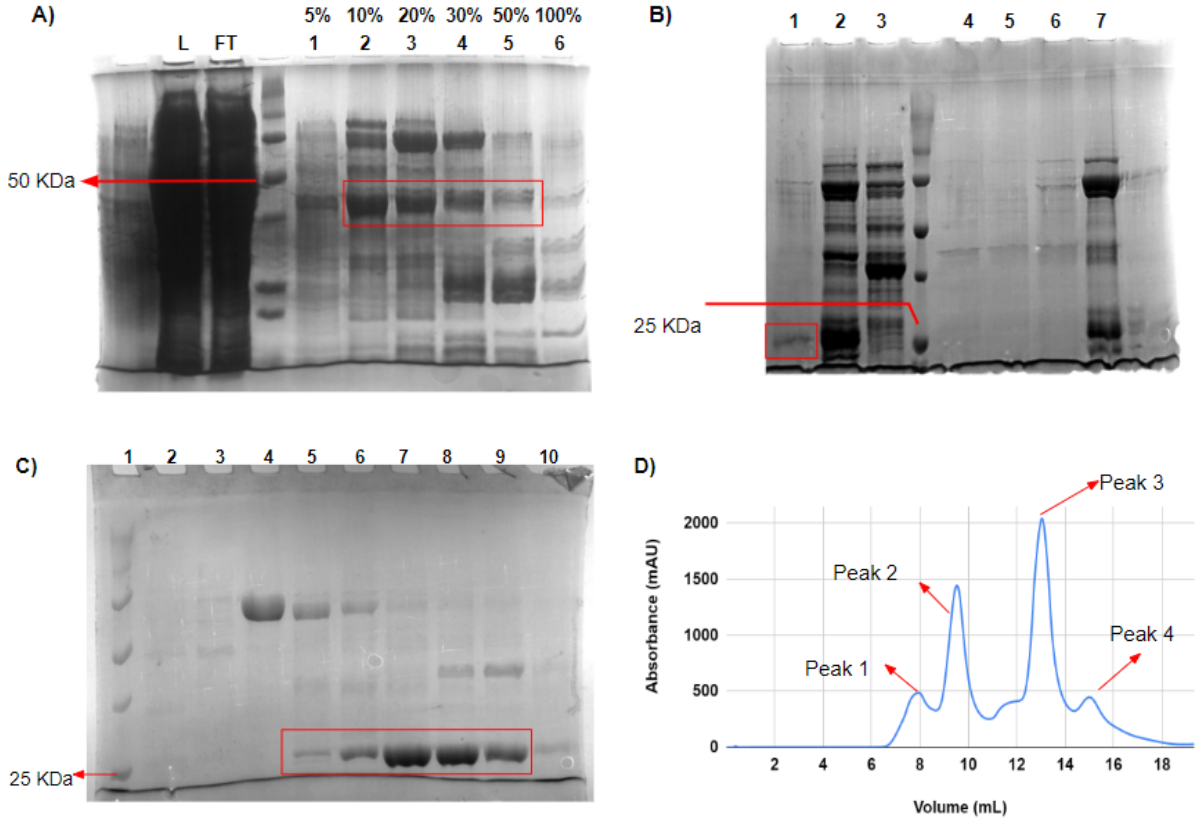

Fig. 10: Purification gel images and chromatogram of NiRAN (Ni-NTA)
A) 10% SDS-PAGE gel of Ni-NTA purification. L and FT are load and flowthrough, respectively. Lane 1 to 6 shows the fractions eluted at %B mentioned above the lanes. **B)** Gel image for second Ni-NTA column after tag cleavage by TEV protease. Lane 3 is the first Ni-NTA column eluate. Lane 2 is the protein solution after tag cleavage, which was used as the load in the second Ni-NTA step. **C)** Gel for Superdex 75 column; lanes 2 and 3 correspond to peak 1 on the chromatogram in Fig. 10D, lanes 4 and 5 correspond to peak 2, lanes 6 to 9 correspond to peak 3, and lane 10 corresponds to peak 4. Bands of interest in each gel are marked in red.

3.3.5. Purification of NiRAN-domain using GSTrap column

NiRAN domain protein with N-terminal GST-tag followed by TEV site, cloned in pGEX vector, was used for this purification. Protein was expressed in Rosetta 2(DE3) cells.

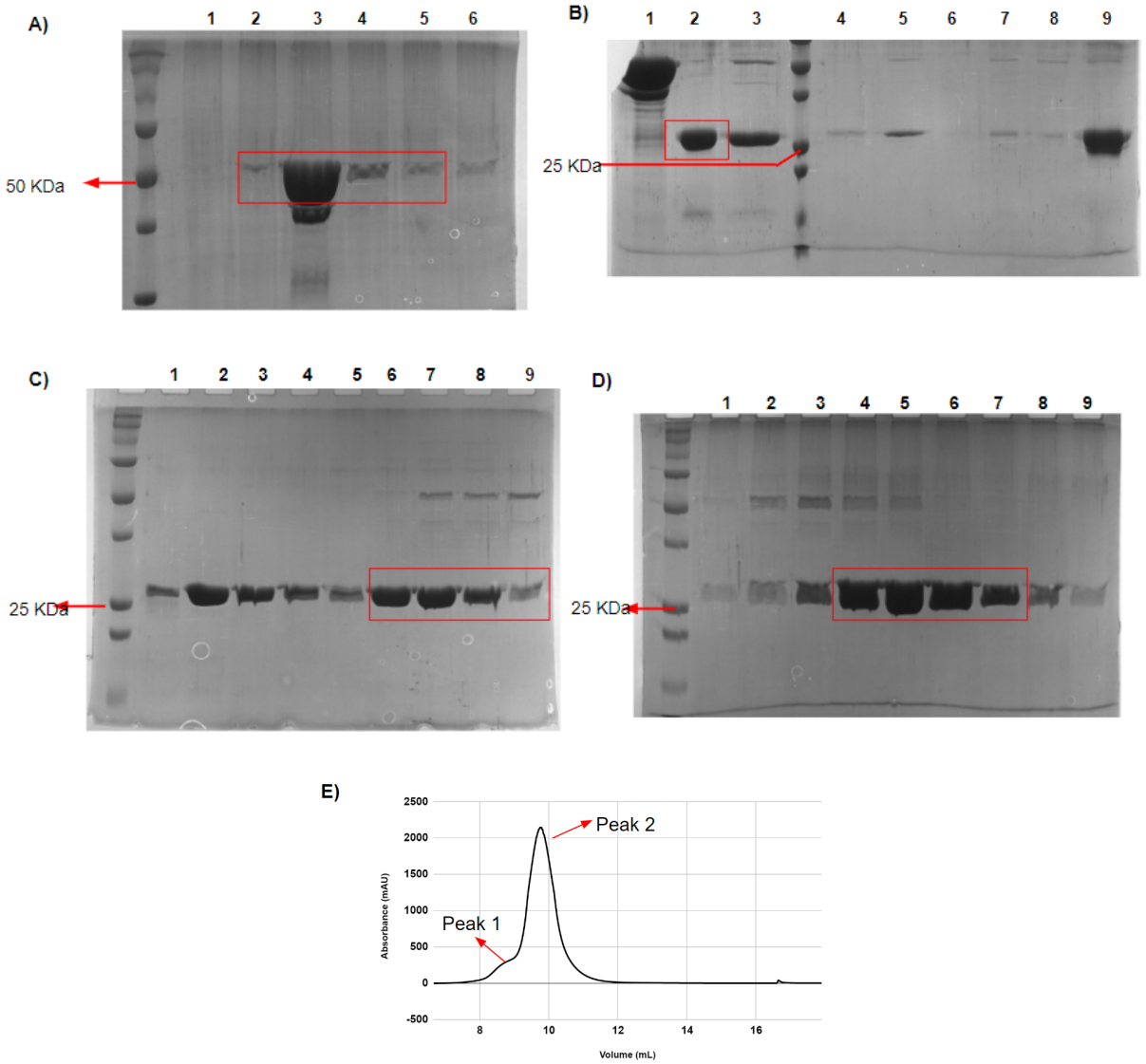

Fig. 11: Purification gel images of NiRAN (GSTrap)

A) Gel of GSTrap column. Protein bound to the column was purified with 100% buffer GB. **B)** After tag cleavage using TEV protease, the protein was passed through the GSTtrap column and Ni-NTA column connected in series. Lane 1 shows the uncleaved eluate from the first step. Lane 3 is protein after tag cleavage. Lane 2 is flowthrough. Impurities bound to the GSTrap column and Ni-NTA column were eluted separately. Lanes 4 to 6 are GSTrap eluates eluted using 100% buffer GB. Lanes 7 to 9 are Ni-NTA eluates eluted using 100% buffer B. **C)** SDS-PAGE gel for the MonoQ column. Initial 5 lanes were eluted at a conductivity of 35 mS/cm, which corresponds to the elution of GST-tag. Lanes 6 to 9 were eluted at a higher conductivity of 47 mS/cm. **D)** Gel for superdex 75 elution. Lanes 1 to 3 correspond to peak 1, and lanes 4 to 9 corresponds to peak 2 on the chromatogram **E)** Chromatogram of Superdex 75 elution. Bands of interest are marked in red.

3.3.6. Purification of NSP12

NSP12 with N-terminal 6xHis-tag followed by SUMO-tag was used for this purification. Protein was expressed in Rosetta 2 (DE3) cells.

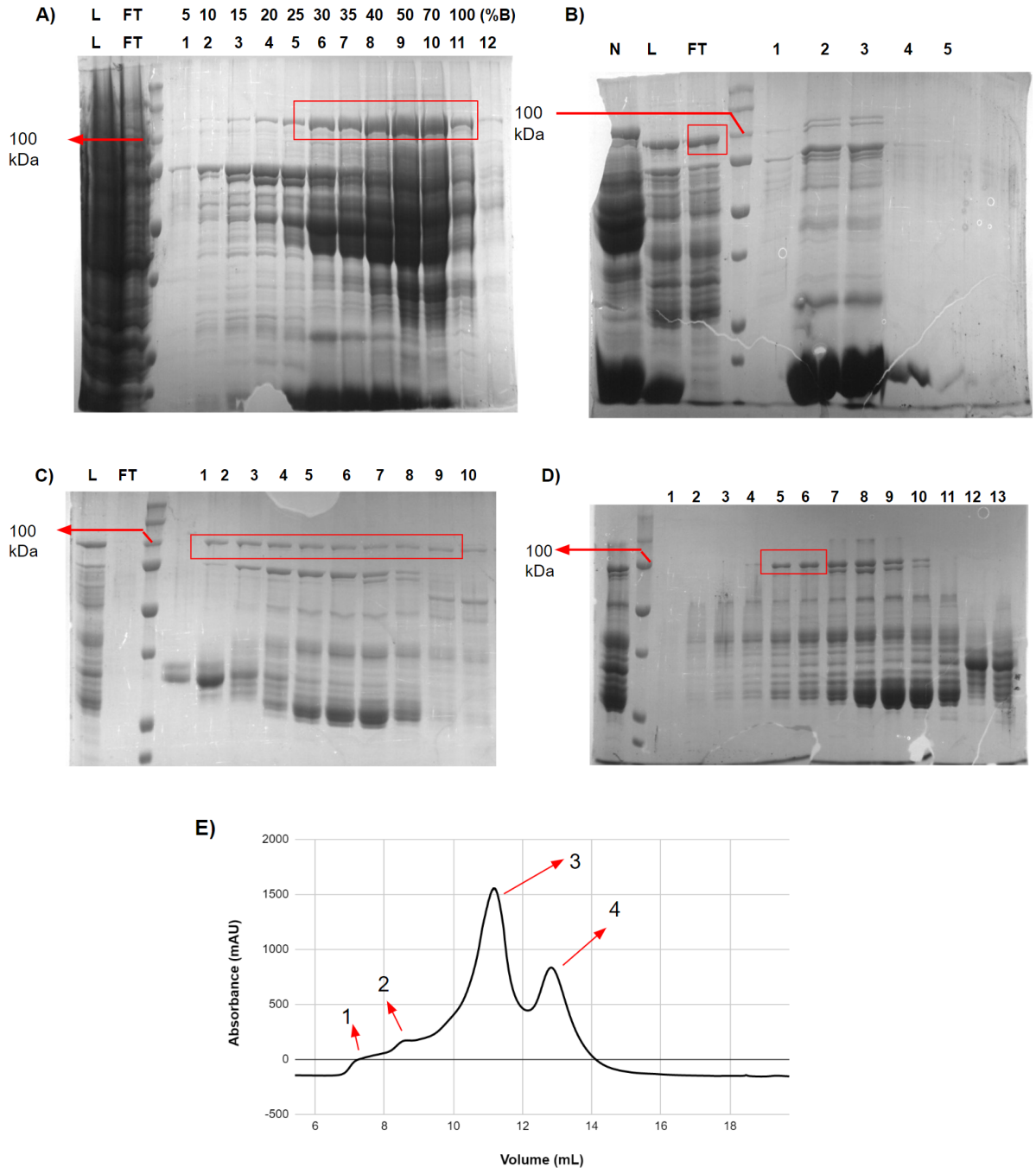

Fig. 12: Purification gel images and chromatogram of NSP12

A) Cell lysate was initially passed through the Ni-NTA column. Lanes 1 to 12 correspond to the fractions eluted at the percentage of buffer B mentioned above each lane. **B)** After cleaving the SUMO-tag using Ulp1 protease, the protein solution

was again loaded in the Ni-NTA column. Lane “N” shows the eluate of the first Ni-NTA purification before the tag cleavage. Lane “L” shows the contents of the protein solution after tag cleavage, which was loaded in the column. Lane “FT” shows the flowthrough with unbound NSP12. Lanes 1 to 5 show impurities bound to the column eluted using 100% buffer B. **C)** After the second Ni-NTA column step, the flowthrough was loaded in MonoQ. Lanes “L” and “FT” show load and flowthrough, respectively. Lanes 1 to 10 correspond to MonoQ elution fractions eluted with increasing concentrations of buffer B_{1000}. **D)** MonoQ elution fractions containing NSP12 were pooled together and concentrated into a volume of less than 1 ml, then passed through the Superdex 200 column. Lanes 1 and 2 correspond to peak 1, lanes 3 to 5 correspond to peak 2, lanes 6 to 10 correspond to peak 3, and lanes 11 to 13 correspond to peak 4 of the Superdex 200 chromatogram shown if Fig. D.

3.3.7. Purification of NSP9

NSP9 with 6xHis-tag followed by a SUMO-tag on the N-terminal, cloned in the pRSFDuet-1 vector was used for this purification. The protein was expressed in Rosetta 2(DE3) cells.

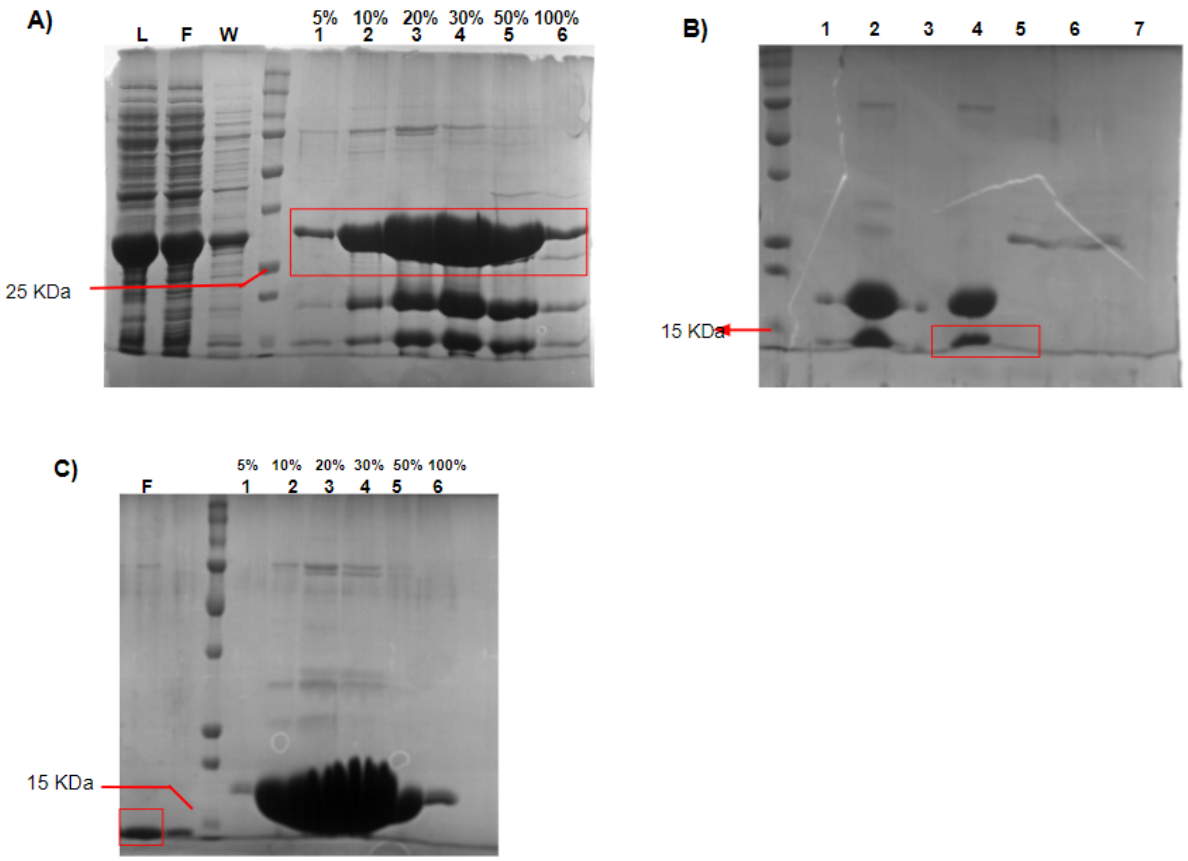

Fig. 13: Purification gel images and chromatogram of NSP9

A) First step of purification was using the Ni-NTA column. **B)** After tag cleavage using Ulp1 protease, the protein solution was passed through the MonoS column. Lane 2 and lane 4 are load and flowthrough, respectively. Lane 5 and 6 show MonoS eluate. **C)** The MonoS flowthrough containing the NSP9 was passed through the Ni-NTA column again. Lane "F" shows the flowthrough containing NSP9 protein. Lanes 1 to 6 show the fractions eluted at %B mentioned above each lane.

Bands of interest are marked in red.

3.4. Exonuclease Assay

Cleavage assay to check the activity of NSP14 was carried out in 1x ExoN buffer (25 mM Tris (pH 7.5), 5 mM $MgCl_2$, 50 mM NaCl, 5% glycerol, and 1 mM DTT) in the presence of NSP10. Amersham Typhoon laser scanner was used to capture the UREA-PAGE gel images.

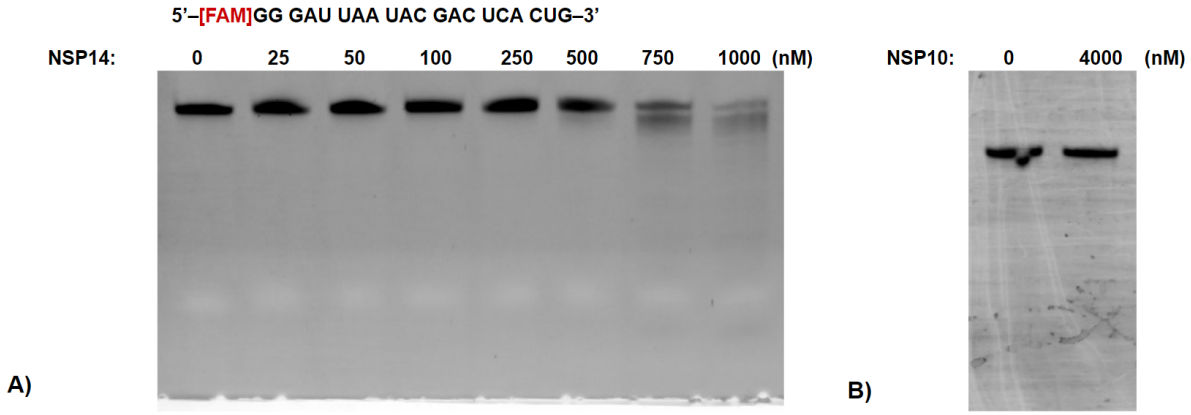

Fig. 14: Urea-PAGE images of Exonuclease assay - I

A) The gel image shows the cleavage activity of NSP14 when acted alone in the absence of NSP10. This assay was carried out using 10 nM 20-mer ssRNA with a 5'-FAM label as the substrate. The concentration of NSP14 in each reaction and the sequence of the substrate used are mentioned in the figure. **B)** Exonuclease reaction using 4000 nM NSP10, in the absence of NSP14, was carried out to check if the protein has any nuclease contamination. 10 nM 20-mer ssRNA with a 5'-FAM label was used for this.

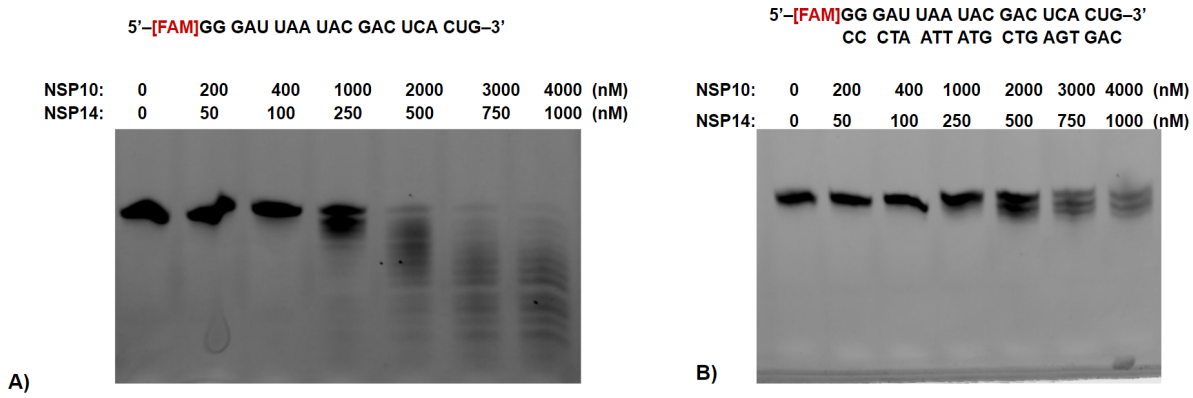

Fig. 15: Urea-PAGE images of Exonuclease assay - II

A) The gel image shows the exonuclease activity of NSP14 in the presence of NSP10. 10 nM 20-mer ssRNA with a 5'-FAM label was used as the substrate, and the concentration of NSP10 was four times that of NSP14 in each reaction. **B)** The gel image shows the exonuclease activity of NSP14 on 10 nM 20-mer dsRNA with a 5'-FAM label on one strand. Sequences of the substrates used are mentioned in the respective gel images.

3.5. NiRAN-NSP9 Binding Assay

The Superdex 75 column was used to check the binding of NSP9 to NiRAN. The NSP9 protein used in this reaction did not have the native N-terminus. 200 µL of the solution containing 25 µM NSP9 and 25 µM NiRAN was passed through the column.

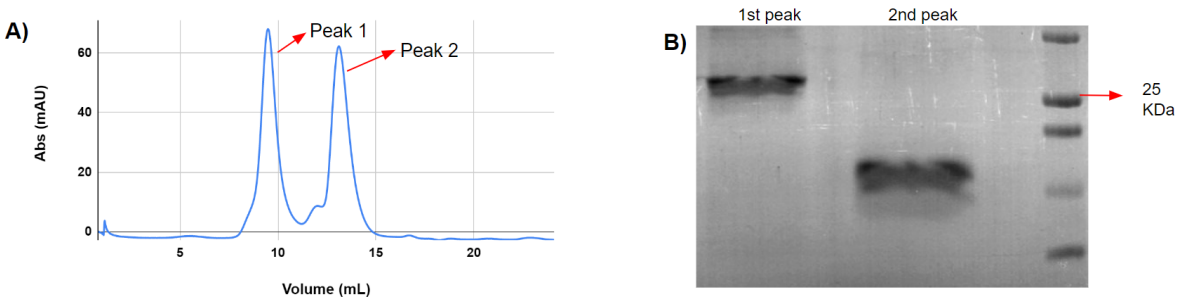

Fig. 16: Chromatogram and gel image of NiRAN-NSP9 binding assay

A) Superdex 75 chromatogram for the analytical run, showing the absorbance at 280 nm. **B)** 15% SDS-polyacrylamide gel image showing the protein contents in each

peak of the chromatogram. The first peak only has NiRAN domain protein, and the second peak only has NSP9 protein.

3.6. NMPylation Assay

This assay was carried out to check the NMPylation Activity of the NiRAN domain of NSP12. NiRAN domain protein (without the rest of NSP12), NSP12, and NSP7/NSP8/NSP12 complex (RTC) were used for this assay.

3.6.1. Malachite-green assay to detect pyrophosphate release

After carrying out the NMPylation assay as described in the materials and methods section, the reaction mix was digested using pyrophosphatase. 50 µL malachite-green solution was added to the 21 µL reaction mix, and the absorbance at 630 nm was detected using a microplate reader (CLARIOstar, BMG Labtech).

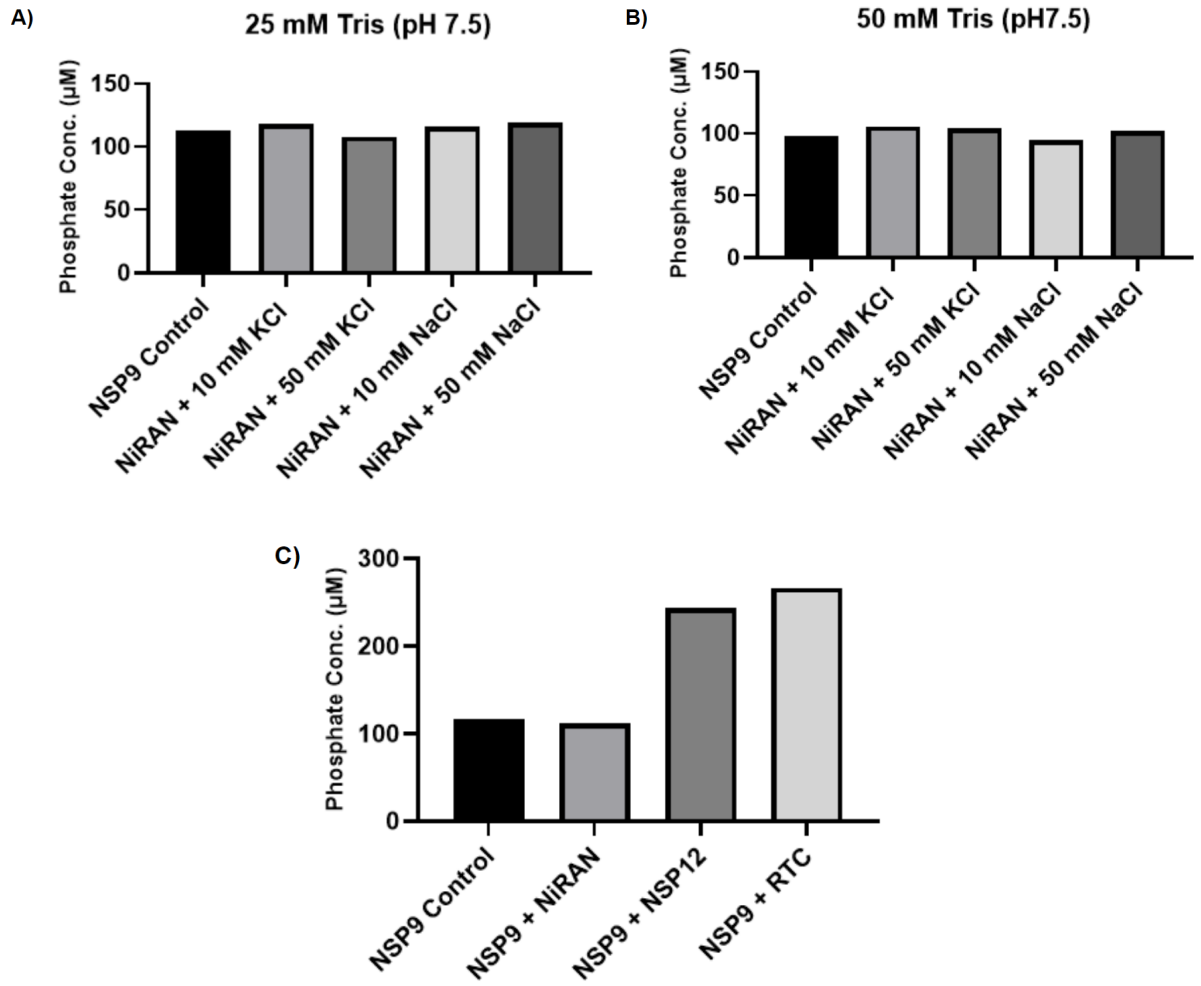

Fig. 17: Bar charts of NMPylation assay

“NSP9 Control” shows the background activity by contaminants in NSP9 protein solution. The reaction was started by the addition of 300 μM ATP to the reaction mix containing 150 μM NSP9 and 50 nM NiRAN, NSP12, or RTC. After the addition of pyrophosphatase, the reaction mix was incubated at 37°C for 2 hours.

A) This NMPylation assay was carried out in a 25 mM Tris buffer (pH 7.5). The reaction also contained 5 mM $MgCl_2$ and 1 mM DTT. The concentration of KCl or NaCl used in each reaction is mentioned below the bars. **B)** NMPylation assay carried out in 50 mM Tris (pH 7.5). Other conditions were the same as in figure A. **C)** This bar chart shows the NMPylation activity of different proteins in the presence of NSP9, as mentioned in the figure. Assay was carried out in a buffer containing 25 mM Tris (pH 7.5), 10 mM KCl, 5 mM $MgCl_2$, and 1 mM DTT.

3.6.2. Mass spectrometry to confirm NMPylation of NSP9

Mass spectrometry was used to confirm that the release of pyrophosphate observed in the malachite assay was due to the addition of AMP from ATP to NSP9.

50 nM NiRAN or RTC was added to 150 μM NSP9 in 1x NMPylation buffer (25 mM Tris pH 7.5, 10 mM KCl, 5 mM $MgCl_2$, and 1 mM DTT). The addition of 300 μM ATP started the reaction, and the sample was incubated at 37°C for 90 minutes before using it for mass spectrometry. The control reaction for NSP9 was performed in the absence of NiRAN and RTC. The procedure was carried out using ESI-TOF Mass Spectrometer by Istiyaq (JBU Lab).

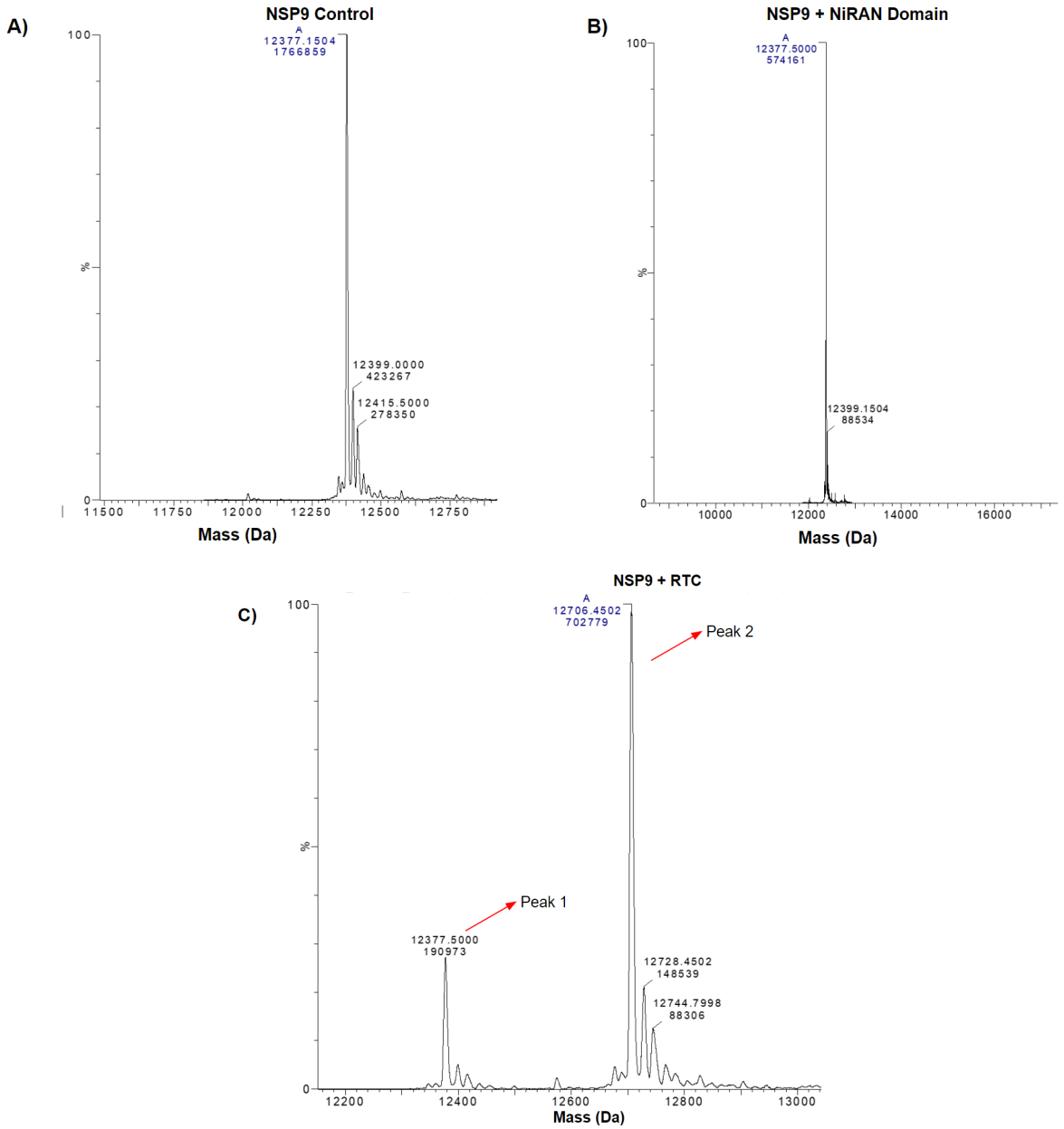

Fig. 18: Graphs of mass spectrometry.

A) Mass spectrometry for NSP9 control. The peak shows the original mass of NSP9. **B)** This figure shows the result of the NMPylation reaction in the presence of NiRAN domain protein. **C)** Result of the NMPylation reaction in the presence of RTC. Peak 1 corresponds to the mass of NSP9, and peak 2 corresponds to the mass of the AMP-NSP9 adduct.

3.7. Protein Crystallization

To crystallize NSP14(-6) and NiRAN domain, the procedure mentioned in materials and methods was followed.

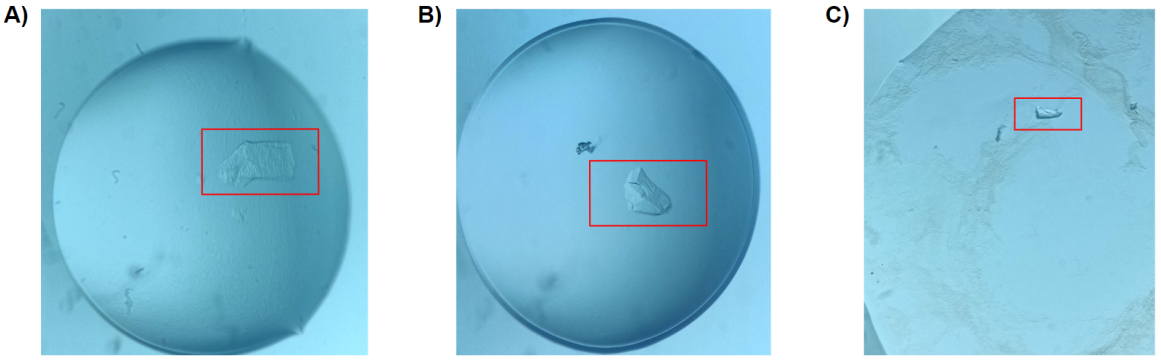

Fig. 19: Crystal images

The figure shows the crystal-like structures obtained during protein crystallization using different screens in 96-well plates.

A) NSP14(-6) protein in a condition containing 0.5 M lithium sulfate, 0.1M Tris (pH 8.5), and 25% w/v PEG 3350 **B)** NSP14(-6) protein in a condition containing 1.0 M lithium sulfate, 0.1M Imidazole malate (pH= 6.5), and 2% w/v PEG 8000. **C)** NiRAN domain protein in a condition containing 0.09M Halogen, 1M Tris bicine (pH 8.5), and 37% MDP_PEG1000_PEG3350 (Molecular Dimensions).

Chapter 4: Discussion

During the project, I mainly worked on two different non-structural proteins of SARS-CoV-2, NSP14, and NSP12 (NiRAN domain). I was able to successfully prepare different gene clones, some of which were made with the help of Dr. Om Prakash Chouhan. After confirming the sequence of each clone through DNA sequencing, I checked the expression in different conditions. Most of the proteins were successfully expressed in the Rosetta 2(DE3) strain of *E.coli* bacteria. NSP14(-66) mutant is not showing protein expression in any of the conditions tested. This mutant was cloned to study the interaction of NSP14 with NSP10.

6xHis-tags were removed from the N-terminus of the GST-tags on NSP14(-6) and NiRAN to test if it would increase the expression. Even though removing the 6xHis-tags from the aforementioned clones increased the expression of the protein, the solubility was still low for the NSP14(-6) clone. A satisfactory level of solubility was observed in the NiRAN domain protein. NSP9 was cloned into a pRSFDuet-1 vector to attach a SUMO-tag on its N-terminus since it would help to maintain the native N-terminus of the protein, a condition important for the NMPylation activity of the NiRAN domain. During protein purification, I mostly followed three chromatography steps in order: affinity chromatography, ion-exchange chromatography, and size-exclusion chromatography. I was able to isolate the protein of interest from the cell lysate to a good level of purity, except for NSP12. Even though NSP12 was showing activity, RTC purified by Ashwin Uday was also used for the assays because of this reason. Isolating the proteins using GSTrap column resulted in better purity compared to Ni-NTA column. Even though the PI of NSP9 is above 9, it did not bind to the MonoS column at pH 7.5.

When NSP14 acts alone, the nuclease activity is very low. Even though NSP10 does not cleave RNA substrates on its own, it considerably increases the activity of NSP14. Recently published papers suggest that NSP14 cleaves double-stranded RNA substrates better than single-stranded RNA substrates. But the exonuclease assays I carried out show that NSP14 cleaves ssRNA substrates far more efficiently than dsRNA substrates. NSP14(-6) mutants were made to crystallize the protein with

some of its inhibitor molecules since it was recently reported that this mutant is able to be crystallized in a condition containing 1.26 M NaH_2PO_4 and 0.14 M K_2HPO_4 (Imprachim N et al, 2023). But I was not able to crystallize the protein in this condition.

The NMPylation assays I performed show that the NiRAN domain protein isolated from the rest of the NSP12 protein is inactive on its own. NSP12 and RTC are able to add AMP from ATP to NSP9 successfully. I was able to detect a sufficient level of activity for NSP12 and RTC in the NMPylation reaction using the malachite assay. Hence I proceeded with mass spectrometry, and the mass shift of the peak in the presence of RTC suggests that AMP is being added to the NSP9 protein. The analytical run for the binding assay suggests that NiRAN is not making any complex with NSP9, but the NSP9 used for this experiment did not have the native N-terminus with the first residue of the protein as an asparagine. D208A and D218A mutants of NiRAN were cloned to study the influence of these residues on the protein activity.

I tried to crystallize NSP14(-6) and NiRAN domain proteins in 96-well plates using different screens from companies such as Molecular Dimensions. I wanted to crystallize NiRAN in order to examine if there is any structural basis for its inactivity in the NMPylation assays. Even though I was able to get a few crystals for both proteins, I was not able to reproduce the crystals while redoing the procedure using manually prepared reservoirs, even after slightly varying the conditions.

References

1. Leao, J. C., Gusmao, T. P. L., Zarzar, A. M., Leao Filho, J. C., Barkokebas Santos de Faria, A., Morais Silva, I. H., Gueiros, L. A. M., Robinson, N. A., Porter, S., & Carvalho, A. A. T. (2022). Coronaviridae-Old friends, new enemy!. *Oral diseases*, *28 Suppl 1*(Suppl 1), 858–866.
2. Payne S. (2017). Family *Coronaviridae*. *Viruses*, 149–158.
3. TYRRELL, D. A., & BYNOE, M. L. (1965). CULTIVATION OF A NOVEL TYPE OF COMMON-COLD VIRUS IN ORGAN CULTURES. *British medical journal*, *1*(5448), 1467–1470.
4. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet (London, England)*, *395*(10224), 565–574.
5. Satarker, S., & Nampoothiri, M. (2020). Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2. *Archives of medical research*, *51*(6), 482–491.
6. Tseng, Y. T., Wang, S. M., Huang, K. J., Lee, A. I., Chiang, C. C., & Wang, C. T. (2010). Self-assembly of severe acute respiratory syndrome coronavirus membrane protein. *The Journal of biological chemistry*, *285*(17), 12862–12872.
7. Ye, Y., & Hogue, B. G. (2007). Role of the coronavirus E viroporin protein transmembrane domain in virus assembly. *Journal of virology*, *81*(7), 3597–3607.
8. Liao, Y., Yuan, Q., Torres, J., Tam, J. P., & Liu, D. X. (2006). Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology*, *349*(2), 264–275.
9. Huang, Y., Yang, C., Xu, X. F., Xu, W., & Liu, S. W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta pharmacologica Sinica*, *41*(9), 1141–1149.
10. Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature reviews. Microbiology*, *17*(3), 181–192.
11. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England journal of medicine*, *367*(19), 1814–1820.
12. Forni, D., Cagliani, R., Clerici, M., & Sironi, M. (2017). Molecular Evolution of Human Coronavirus Genomes. *Trends in microbiology*, *25*(1), 35–48.
13. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269.
14. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., … Shi, Z. L. (2020). A pneumonia outbreak

associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273.
15. Perlman, S., & Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. *Nature reviews. Microbiology*, *7*(6), 439–450.
16. Snijder, E. J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J. J., van der Meulen, J., Koerten, H. K., & Mommaas, A. M. (2006). Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. *Journal of virology*, *80*(12), 5927–5940.
17. Wu, H. Y., & Brian, D. A. (2010). Subgenomic messenger RNA amplification in coronaviruses. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12257–12262.
18. Zuo, Y., & Deutscher, M. P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic acids research*, *29*(5), 1017–1026.
19. Nishino, T., & Morikawa, K. (2002). Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene*, *21*(58), 9022–9032.
20. Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G., & Murray, N. E. (2014). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic acids research*, *42*(1), 3–19.
21. Dehé, P. M., & Gaillard, P. H. L. (2017). Control of structure-specific endonucleases to maintain genome stability. *Nature reviews. Molecular cell biology*, *18*(5), 315–330.
22. Mason, P. A., & Cox, L. S. (2012). The role of DNA exonucleases in protecting genome stability and their impact on ageing. *Age (Dordrecht, Netherlands)*, *34*(6), 1317–1340.
23. Hosfield, D. J., Mol, C. D., Shen, B., & Tainer, J. A. (1998). Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell*, *95*(1), 135–146.
24. Johnson, A., & O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. *Annual review of biochemistry*, *74*, 283–315.
25. Ogando, N. S., Zevenhoven-Dobbe, J. C., van der Meer, Y., Bredenbeek, P. J., Posthuma, C. C., & Snijder, E. J. (2020). The Enzymatic Activity of the nsp14 Exoribonuclease Is Critical for Replication of MERS-CoV and SARS-CoV-2. *Journal of virology*, *94*(23), e01246-20.
26. Baddock, H. T., Brolih, S., Yosaatmadja, Y., Ratnaweera, M., Bielinski, M., Swift, L. P., Cruz-Migoni, A., Fan, H., Keown, J. R., Walker, A. P., Morris, G. M., Grimes, J. M., Fodor, E., Schofield, C. J., Gileadi, O., & McHugh, P. J. (2022). Characterization of the SARS-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *Nucleic acids research*, *50*(3), 1484–1500.
27. Subissi, L., Posthuma, C. C., Collet, A., Zevenhoven-Dobbe, J. C., Gorbalenya, A. E., Decroly, E., Snijder, E. J., Canard, B., & Imbert, I. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(37), E3900–E3909.
28. Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T., & Guo, D. (2009). Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3484–3489.

29. Ferron, F., Subissi, L., Silveira De Morais, A. T., Le, N. T. T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B., & Imbert, I. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(2), E162–E171.
30. Jin, X., Chen, Y., Sun, Y., Zeng, C., Wang, Y., Tao, J., Wu, A., Yu, X., Zhang, Z., Tian, J., & Guo, D. (2013). Characterization of the guanine-N7 methyltransferase activity of coronavirus nsp14 on nucleotide GTP. *Virus research*, *176*(1-2), 45–52.
31. Saramago, M., Bárria, C., Costa, V. G., Souza, C. S., Viegas, S. C., Domingues, S., Lousa, D., Soares, C. M., Arraiano, C. M., & Matos, R. G. (2021). New targets for drug design: importance of nsp14/nsp10 complex formation for the 3'-5' exoribonucleolytic activity on SARS-CoV-2. *The FEBS journal*, *288*(17), 5130–5147.
32. Bernad, A., Blanco, L., Lázaro, J. M., Martín, G., & Salas, M. (1989). A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell*, *59*(1), 219–228.
33. Kirchdoerfer, R. N., & Ward, A. B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature communications*, *10*(1), 2342.
34. te Velthuis, A. J., van den Worm, S. H., & Snijder, E. J. (2012). The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic acids research*, *40*(4), 1737–1747.
35. Yan, L., Huang, Y., Ge, J., Liu, Z., Lu, P., Huang, B., Gao, S., Wang, J., Tan, L., Ye, S., Yu, F., Lan, W., Xu, S., Zhou, F., Shi, L., Guddat, L. W., Gao, Y., Rao, Z., & Lou, Z. (2022). A mechanism for SARS-CoV-2 RNA capping and its inhibition by nucleotide analog inhibitors. *Cell*, *185*(23), 4347–4360.e17.
36. Daffis, S., Szretter, K. J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T. Y., Schneller, S., Zust, R., Dong, H., Thiel, V., Sen, G. C., Fensterl, V., Klimstra, W. B., Pierson, T. C., Buller, R. M., Gale, M., Jr, Shi, P. Y., & Diamond, M. S. (2010). 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*, *468*(7322), 452–456.
37. Dwivedy, A., Mariadasse, R., Ahmad, M., Chakraborty, S., Kar, D., Tiwari, S., Bhattacharyya, S., Sonar, S., Mani, S., Tailor, P., Majumdar, T., Jeyakanthan, J., & Biswal, B. K. (2021). Characterization of the NiRAN domain from RNA-dependent RNA polymerase provides insights into a potential therapeutic target against SARS-CoV-2. *PLoS computational biology*, *17*(9), e1009384.
38. Yuen, C. K., Lam, J. Y., Wong, W. M., Mak, L. F., Wang, X., Chu, H., Cai, J. P., Jin, D. Y., To, K. K., Chan, J. F., Yuen, K. Y., & Kok, K. H. (2020). SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. *Emerging microbes & infections*, *9*(1), 1418–1428.
39. Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Brooun, A., Griffith, M., Moy, K., Yadav, M. K., Velasquez, J., Buchmeier, M. J., Stevens, R. C., & Kuhn, P. (2006). Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *Journal of virology*, *80*(16), 7894–7901.
40. Lin, S., Chen, H., Chen, Z., Yang, F., Ye, F., Zheng, Y., Yang, J., Lin, X., Sun, H., Wang, L., Wen, A., Dong, H., Xiao, Q., Deng, D., Cao, Y., & Lu, G. (2021). Crystal structure of

SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic acids research*, *49*(9), 5382–5392.
41. Hamdan, S., Carr, P. D., Brown, S. E., Ollis, D. L., & Dixon, N. E. (2002). Structural basis for proofreading during replication of the Escherichia coli chromosome. *Structure (London, England : 1993)*, *10*(4), 535–546.
42. Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L., & Denison, M. R. (2007). High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *Journal of virology*, *81*(22), 12135–12144.
43. Imprachim, N., Yosaatmadja, Y., & Newman, J. A. (2023). Crystal structures and fragment screening of SARS-CoV-2 NSP14 reveal details of exoribonuclease activation and mRNA capping and provide starting points for antiviral drug development. *Nucleic acids research*, *51*(1), 475–487.
44. Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of medical virology*, *92*(4), 418–423.
45. Krichel, B., Falke, S., Hilgenfeld, R., Redecke, L., & Uetrecht, C. (2020). Processing of the SARS-CoV pp1a/ab nsp7-10 region. *The Biochemical journal*, *477*(5), 1009–1019.
46. Khailany, R. A., Safdar, M., & Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene reports*, *19*, 100682.
47. Jin, Y., Ouyang, M., Yu, T., Zhuang, J., Wang, W., Liu, X., Duan, F., Guo, D., Peng, X., & Pan, J. A. (2022). Genome-Wide Analysis of the Indispensable Role of Non-structural Proteins in the Replication of SARS-CoV-2. *Frontiers in microbiology*, *13*, 907422.
48. Yadav, R., Chaudhary, J. K., Jain, N., Chaudhary, P. K., Khanra, S., Dhamija, P., Sharma, A., Kumar, A., & Handu, S. (2021). Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells*, *10*(4), 821.
49. Peersen O. B. (2019). A Comprehensive Superposition of Viral Polymerase Structures. *Viruses*, *11*(8), 745.
50. Lehmann, K. C., Gulyaeva, A., Zevenhoven-Dobbe, J. C., Janssen, G. M., Ruben, M., Overkleeft, H. S., van Veelen, P. A., Samborskiy, D. V., Kravchenko, A. A., Leontovich, A. M., Sidorov, I. A., Snijder, E. J., Posthuma, C. C., & Gorbalenya, A. E. (2015). Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic acids research*, *43*(17), 8416–8434.
51. Ubersax, J. A., & Ferrell, J. E., Jr (2007). Mechanisms of specificity in protein phosphorylation. *Nature reviews. Molecular cell biology*, *8*(7), 530–541.
52. Dudkiewicz, M., Szczepińska, T., Grynberg, M., & Pawłowski, K. (2012). A novel protein kinase-like domain in a selenoprotein, widespread in the tree of life. *PloS one*, *7*(2), e32138.
53. Sreelatha, A., Yee, S. S., Lopez, V. A., Park, B. C., Kinch, L. N., Pilch, S., Servage, K. A., Zhang, J., Jiou, J., Karasiewicz-Urbańska, M., Łobocka, M., Grishin, N. V., Orth, K., Kucharczyk, R., Pawłowski, K., Tomchick, D. R., & Tagliabracci, V. S. (2018). Protein AMPylation by an Evolutionarily Conserved Pseudokinase. *Cell*, *175*(3), 809–821.e19.
54. Wang, B., Svetlov, D., & Artsimovitch, I. (2021). NMPylation and de-NMPylation of SARS-CoV-2 nsp9 by the NiRAN domain. *Nucleic acids research*, *49*(15), 8822–8835.
55. Roman, C., Lewicka, A., Koirala, D., Li, N. S., & Piccirilli, J. A. (2021). The SARS-CoV-2 Programmed -1 Ribosomal Frameshifting Element Crystal Structure Solved to 2.09 Å

Using Chaperone-Assisted RNA Crystallography. *ACS chemical biology*, *16*(8), 1469–1481.

56. Park, G. J., Osinski, A., Hernandez, G., Eitson, J. L., Majumdar, A., Tonelli, M., Henzler-Wildman, K., Pawłowski, K., Chen, Z., Li, Y., Schoggins, J. W., & Tagliabracci, V. S. (2022). The mechanism of RNA capping by SARS-CoV-2. *Research square*, rs.3.rs-1336910.
57. Tang, X., Shang, J., & Sun, Y. (2022). RdRp-based sensitive taxonomic classification of RNA viruses for metagenomic data. *Briefings in bioinformatics*, *23*(2), bbac011.
58. Viswanathan, T., Arya, S., Chan, S. H., Qi, S., Dai, N., Misra, A., Park, J. G., Oladunni, F., Kovalskyy, D., Hromas, R. A., Martinez-Sobrido, L., & Gupta, Y. K. (2020). Structural basis of RNA cap modification by SARS-CoV-2. *Nature communications*, *11*(1), 3718.
59. Harrison, A. G., Lin, T., & Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in immunology*, *41*(12), 1100–1115.