Biochemical Studies of SARS-CoV-2 RTC

A thesis submitted to the Indian Institute of Science Education and Research (IISER) Pune in the partial fulfilment of the requirement for the BS-MS Dual degree programme by

P S SIVAPRASAD
20191112

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

Date: March 15 2024

Under the guidance of

Supervisor: **Prof Saikrishnan Kayarat**

Professor Department of Biology, IISER Pune

From May 2023 to March 2024

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

The thesis is dedicated to

My inner self who always aspires to use scientific knowledge and skills for the betterment of humanity

Certificate

This is to certify that this dissertation entitled **Biochemical Studies of SARS-CoV-2 RTC** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune, represents work carried out by **P S SIVAPRASAD** at Indian Institute of Science Education and Research under the supervision of **Prof Saikrishnan Kayarat**, Department of Biology, during the academic year 2023-2024.

Prof Saikrishnan Kayarat

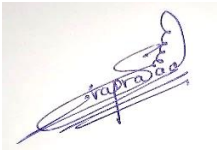
Committee:

Name of your Guide: **Prof Saikrishnan Kayarat**

Name of Your TAC: **Dr. Gayathri Pananghat**

Declaration

I hereby declare that the matter embodied in the report entitled "**Biochemical Studies of SARS-CoV-2 RTC**" are the results of the work carried out by me at the **Department of Biology**, Indian Institute of Science Education & Research (IISER) Pune, under the supervision of **Prof Saikrishnan Kayarat** and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.



Name of Student: **P S SIVAPRASAD**

Roll No: 20191112

List of Table

List of Figures

Table of Contents

Abstract

COVID-19, caused by SARS-CoV-2, has been a catastrophic pandemic. While vaccines have helped restore normalcy, reinfections persist, highlighting the urgent need for antiviral drugs. The replication transcription complex (RTC) offers a promising target due to its conservation across strains. The NiRAN (Nidoviruses RdRp associated nucleotidyltransferase) domain of Nsp 12 (RNA polymerase), which carries out the initial steps of RNA capping, has three different nucleotide binding poses. Each pose has its own significance for capping viral mRNA and NMPylation reactions (Nucleotide monophosphate addition to Nsp 9). In this investigation, using mutagenesis, I tried to delineate the molecular significance of each pose. Additionally, specific NiRAN residues influence RdRp's extension activity, highlighting their role in RTC stability and supporting my observations of feeble polymerase activity upon deletion of the NiRAN domain. These findings contribute to understanding CoV-2 replication and capping reaction. In addition, my project also provides an RTC mutant for virologists to hunt for the physiological significance of the NMPylation reaction.

.

Acknowledgements

I extend my deepest thanks to Professor Saikrishnan Kayarat for his invaluable mentorship, which has played a pivotal role in shaping my academic path. His guidance and shared wisdom have greatly enriched my learning journey. I also appreciate Dr. Gayathri Pananghat for all the help and guidance during the project.

My time at SK Lab was incredibly fulfilling, thanks to the collaborative environment of both SK Lab and G3 Lab members, which facilitated growth and learning. Under the mentorship of Ashwin Uday, I had the privilege of working on various aspects of SARS-CoV-2 replication Transcription Complex projects that challenged me and provided me with valuable insights and experiences. His guidance has been invaluable in shaping my understanding and skills in my field of study. I also wish to thank Muhammed Navas for all the technical support he provided during the project.

In closing, I wish to express my heartfelt gratitude to all those who have supported and guided me throughout my academic journey. Their contributions have been instrumental in my growth and development, and I am truly grateful for their unwavering support.

Contributions

Contributor Role	Contributor name(s)
Conceptualization of Ideas	Saikrishnan Kayarat, Ashwin Uday
Methodology	Ashwin Uday, Muhammed Navas
Software	-------
Validation	Saikrishnan Kayarat, Ashwin Uday
Formal analysis	Sivaprasad
Investigation	Sivaprasad
Resources	Saikrishnan Kayarat
Data Curation	Sivaprasad
Writing original draft preparation	Sivaprasad
Writing, review and editing	Sivaprasad, Ashwin Uday
Visualization	Sivaprasad
Supervision	Saikrishnan Kayarat
Project administration	Saikrishnan Kayarat
Funding acquisition	Saikrishnan Kayarat

Chapter 01: Introduction

Chapter 01: Introduction

SARS-CoV-2

An outbreak of pneumonia was first reported in Wuhan, China, by a novel virus in December 2019, and soon it spread all over the world. The World Health Organization (WHO) and International Committee on Taxonomy of Viruses (ICTV) designated the viruses as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and associated illness as Coronavirus Disease-2019 (COVID-19) (Gorbalenya et al.). It is the seventh coronavirus that affect humans. Severe acute respiratory syndrome (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), along with SARS-CoV-2, are severe pneumonia-causing viruses with fatality rates of 9.6%, 36%, and 2.9%, respectively. (McIntosh et al., 1967; Paules et al., 2020). Apart from these, there are four other human coronaviruses: OC43, HKU1, NL63, and 229E, which primarily induce minor symptoms of a self-limiting illness (McIntosh et al., 1967). Coronaviruses universally possess a single-stranded, positive sense polycistronic RNA, enveloped with a distinctive crown-like appearance under the microscope due to the presence of characteristic spike protein on it. SARS falls within the betacoronavirus genus of the Coronaviridae family, categorized within the Nidovirales order. Angiotensin-converting enzyme 2 (ACE2) is the primary receptor for SARS-CoV-2, with host proteases acting as cofactor for viral entry.

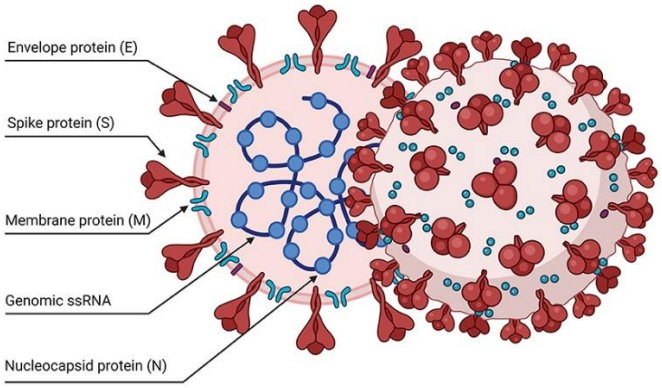


Figure 01: Schematic representation of SARS-CoV-2 (Brant et al., 2021).

Genomic Organization

SARS-CoV-2 has a single-stranded, positive sense RNA, which means the genome can be directly translated upon entry into the host cell. The typical length of the genome spans ~29.9 Kbp. The genome is unstable at higher temperatures because of higher A+U content (62 %). Like all coronaviruses, SARS-CoV-2 has a cap made up of $^{7Me}GpppA_{2'OMe}$ on the 5' end of the

genome and a ~ 30-60 nucleotide long (Median length = 47 nucleotide) poly-A tail on its 3' end of the genome (Viswanathan et al., 2020; Park et al., 2022; Tvarogová et al., 2019). The cap is essential for viruses as it facilitates the translation of viral mRNA, protects the viral genome from host exonuclease, and helps evade host immunity. Similarly, the poly-A tail also contributes to the viral genome stability and protects from host exoribonuclease action. SARS-CoV-2 has untranslated regions (UTR) on both the 5' and 3' end of the genome with 265 nucleotides long 5'UTR containing the 72 nucleotides 5' leader long transcription regulatory core sequence (TRS_L: ACGAAC) and also contains many cis-elements to modulate viral translation, subgenome mRNA synthesis and viral genome packaging (Miao et al., 2021; Wang et al.; Rangan et al., 2020). The 3' UTR is 337 nucleotide long and contains the binding site for replication transcription complex (RTC), which plays an essential role in RNA synthesis and also harbours many cis-acting elements like bulged stem-loop (BSL) and a pseudoknot (Miao et al., 2021; Cascella et al., 2022).

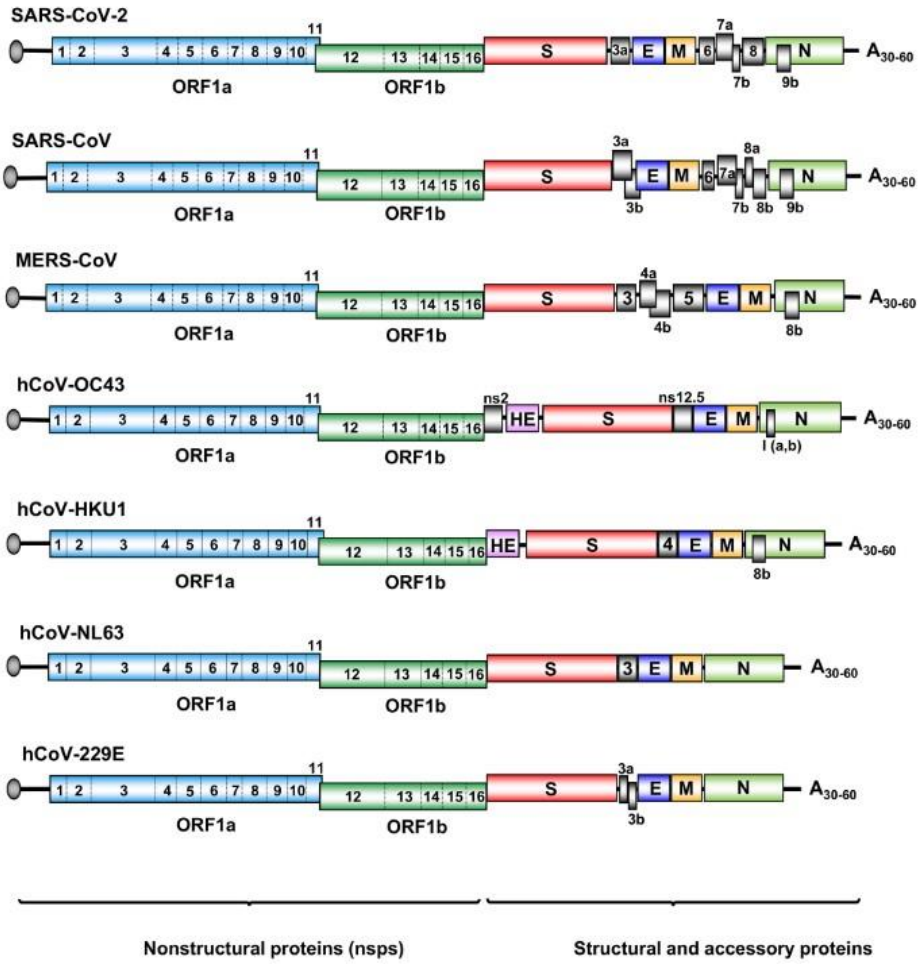


Figure 02: Schematic representation of genome organisation of various coronaviruses (Brant et al., 2021)

The genome of SARS-CoV-2 contains 14 ORFs, which encode for 16 nonstructural, four structural and nine accessory proteins (Finkel et al., 2020; Pizzato et al., 2022). At the 5' end, the Nidoviral genome harbours two sizable ORFs named ORF1a and ORF1b, constituting 2/3rd of the genome and encodes 16 Non-structural proteins (Nsps) in a polyprotein fashion. These Nsps form the major components of the replication transcription apparatus and play a role in host immune evasion. Encoded toward the 3' end of the genome lies the four structural proteins- spike (S), envelope (E), membrane (M) and nucleocapsid (N) which play essential roles in viral morphology, genome packaging and host cell attachment (Arya et al., 2021). SARS-CoV-2 also encodes nine accessory proteins termed ORF3a, 3b, 6, 7a, 7b, 8, 9a, 9b and 10 encoded in a homonymous fashion and is involved in modulating host cell metabolism and immune evasion (Brant et al., 2021).

Life Cycle

SARS-CoV-2 targets the goblet cells, ciliated cells in the respiratory tract, and endothelial cells in the lungs within the human body. The receptor binding domain (RBD) on S1 subunit of spike protein on the viral surface interacts with the human Angiotensin Converting Enzyme 2 (hACE2) receptor on host cells, leading to a conformational change in spike protein followed by the priming of spike protein by cellular protease, which cleaves the spike protein between the boundary of the extracellular domain (S1 and S2) giving rise to a metastable prefusion conformation which remains non-covalently attached (Shang et al., 2020; Shirato et al., 2018). This is followed by processing by other host proteases like TMPRSS2, which produces a second cleavage in the S2 domain within the cell surface or the endosomes, resulting in the release of the S1 domain and exposing the fusion peptide in the S2 domain and activating the fusogenic potential of spike protein leading to the fusion with plasma membrane followed by penetration of virus into the cytoplasm (Yoshimoto, 1234; Suryawanshi et al., 2021). Inside the cytoplasm, the virus unpacks its genome from the viral N-bound packed state by the action of cellular protease, which is then directly translated with the help of host ribosome machinery to produce two large polyprotein pp1a and pp1ab from ORF1a and ORF1b respectively. Here, pp1ab is expressed as a result of -1 programmed ribosomal frameshift upstream of ORF1a stop codon, thereby extending pp1a with the polyprotein from ORF1b. Sixteen non-structural proteins are released upon fifteen proteolytic cleavages by Nsp 3 (papin-like protease (PL^{pro}) or Nsp 5 (main protease (M^{pro})). In this way, the cleavage of pp1a produces 11 non-structural proteins from Nsp 1 to Nsp 11. The cleavage of pp1ab produces fifteen non-structural proteins from Nsp 1 to Nsp 10 and Nsp 12 to Nsp 16 (V'kovski et al., 2021). The released Nsp 1

mediates the shutdown of host mRNA by binding to the ribosomes, thereby suppressing cellular immune response (Schubert et al.). Most other non-structural proteins engage in viral replication transcription complex (RTC). Viral RNA synthesis occurs with the double membrane vesicles (DMV), which are formed by viral action in the host cytoplasm of infected cells. The exact mechanism of formation of DMVs is still unknown (Den Boon et al., 2010). It is speculated that host and viral protein together play a role in this process. In the replication process, the viral genome of (+) polarity acts as a template for synthesising (–) strands and vice versa. This is done by the non-structural proteins Nsp12, Nsp 8 and Nsp 7, which form the core RTC responsible for the synthesis of both (+) & (–) RNA strands and sub-genomic mRNAs (sg RNAs) (Yan et al., 2020). RNA replication and sgRNA synthesis happen inside the DMVs as they help viral RNA evade host immune sensors (Tsang et al., 2021). The RTC produces gRNA and sgRNA encoding the four structural genes and other accessory genes. The newly synthesized (+) g RNA is transported out of DMVs and can undergo several processes: directly translated to produce more non-structural proteins, they act as a template for (–) gRNA synthesis, or be packaged into the new virions. The viral assembly commences with the packing of (+) gRNA by the N protein, which results in the nucleocapsid structure that buds into the endoplasmic reticulum- Golgi intermediate compartment (ERGIC) which results in acquiring the lipid bilayer containing the viral spike, membrane and envelope proteins (Stertz et al., 2007; Klein et al., 2020).

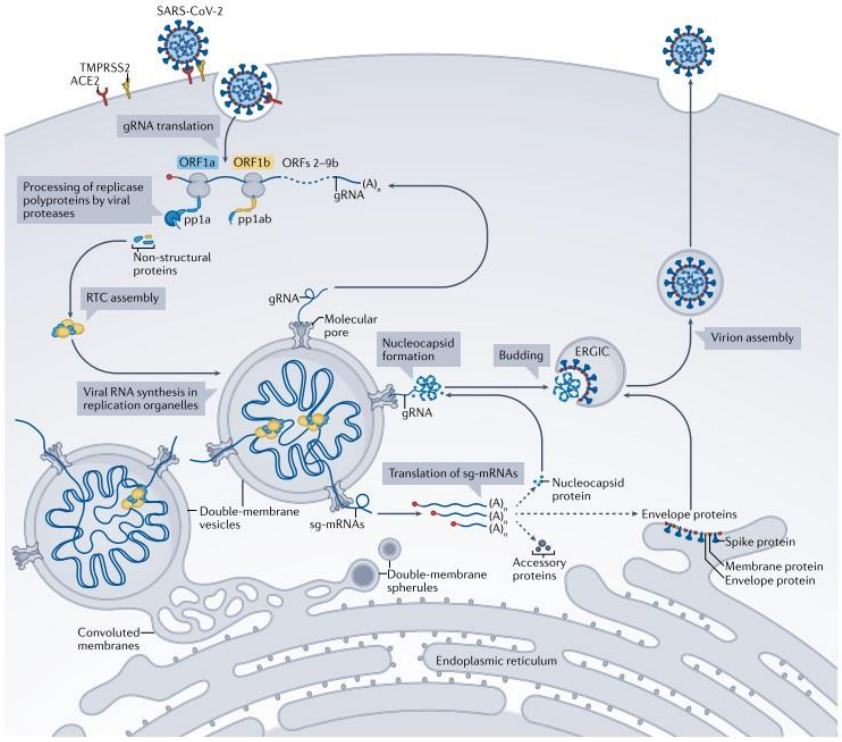


Figure 03: Schematic representation of SARS-CoV-2 life cycle (Malone et al., 2022)

Replication Transcription Complex (RTC)

Central to the replication of a virus and its life cycle is the Replication Transcription Complex (RTC). The core RTC of SARS-CoV-2 is constituted by Non-structural protein 12 (Nsp 12) bound to Nsp7/Nsp8 heterodimer and a second Nsp 8 monomer. (Yan et al., 2020). Most structures of RTC reported so far are in complex with an RNA primer-template duplex. In which the Nsp 12 contacts around 6bp of RNA duplex from 3' end of primer RNA strand with two Nsp 8 attached to the RdRp domain of Nsp 12 because of the asymmetry of Nsp 12, it needs Nsp 7 to attach to Nsp 8 in one side. Nsp 8 has a long N terminal helical structure with mostly positively charged residue that interacts non-specifically with the synthesised RNA duplex released from the RdRp active site (Gao et al., 2020; Wang et al., 2020). The addition of proteins such as Nsp 13 (helicase), Nsp 14 (exonucleases), Nsp 16 (2'-O-methyltransferase), Nsp 9 (accessory protein) and Nsp 10 (accessory protein) to the core RTC results in the assembly of extended RTC (Campagnola et al., 2022).

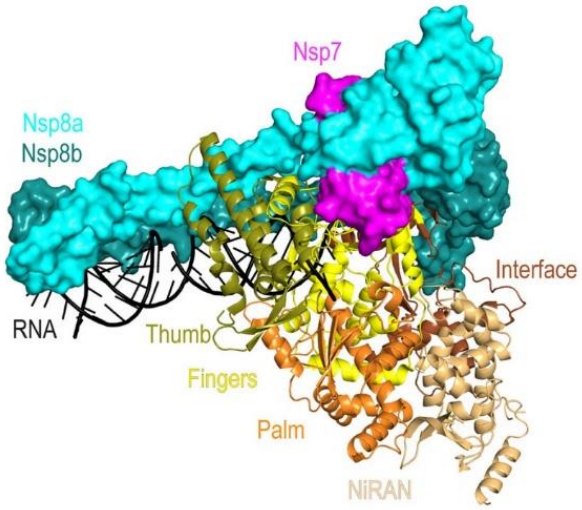


Figure 04: Schematic representation of SARS-CoV-2 RTC (Wang et al., 2021b)

The Nsp12 has three domains: a Nidoviruses RdRp associated nucleotidyltransferase (NiRAN), Interface and RNA-dependent RNA polymerase (RdRp) domain. NiRAN domain is 246 amino acids long (3-249), the Interface domain is 146 amino acids long (250-396), and the RdRp domain is 535 amino acids long (397-932) (Gao et al., 2020). The Nsp 12 enzyme contains two active sites (Wang et al., 2021b)- one is the common and well-known site involved in the polymerization of RNA molecules, and the second active site is located within the Nidoviruses RdRp-associated nucleotidyltransferase (NiRAN) domain attributed to

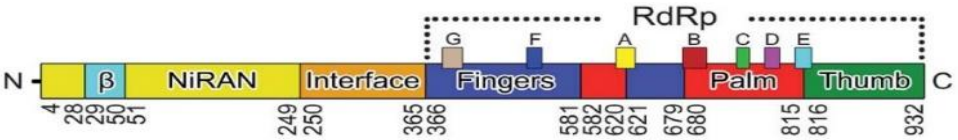


Figure 05: Schematic representation of different domains and subdomains of Nsp12 CoV (Gao et al., 2020)

capping functions. The RTC synthesises both (+) and (–) RNA strands and sub-genomic RNAs. The replication of (–) g RNA produces full-length (+) g RNA in CoV, which can either serve as a template to produce (–) strand RNA or act as a messenger molecule that can be translated to produce more polyproteins or act as genome that can be packaged into new virion particles. Transcription produces the nested set of sg RNAs, which codes for accessory and structural genes. All sg-mRNAs carry the same 3'-terminal sequence and have a common 5' leader sequence like the + g RNA strand. This common 5' leader sequence in viral transcripts allows escape from translation shut-off imposed by Nsp 1 binding to the mRNA entry channel. The structural and accessory gene at the 3' end of the (+) RNA genome acquires the leader sequence at the 5' end of the RNA strand through template switching via discontinuous synthesis during the (–) g RNA and (–) sg RNA synthesis which then subsequently act as a template for the synthesis of corresponding (+) strands. The regulation between the synthesis of full-length (–) RNA genome and (–) sg RNA is speculated to be decided by various factors like secondary structural elements in the CoV genome, genome cyclization, flanking RNA sequences, and interaction with specific protein factors dictate the balance between transcription and replication (Sola et al., 2015).

Nidoviruses RdRp Associated Nucleotidyltransferase (NiRAN)

The NiRAN domain is one of the two biomarkers of order nidovirales (Lehmann et al., 2015). The other is the Zinc-finger domain within the Nsp 13 helicase (Gorbalenya et al., 2006). Among them, NiRAN is the enzymatic biomarker essential for viral propagation. Compared to the RdRp domain, which is nearly invariant across various RNA viruses (Te Velthuis, 2014), the NiRAN domain exhibits a greater level of divergence in sequence and has four conserved motifs: preAN, AN, BN and CN (Park et al., 2022), The NiRAN domain in initial years of its discovery was hypothesized to have role in either mRNA capping or protein primed RNA synthesis or RNA ligases (Lehmann et al., 2015). Recently, in 2022, the NiRAN domain was attributed to play a pivotal role in the capping machinery of CoV. The 5' end of both genomic

RNA and sg mRNA of coronaviruses possess a $^{7Me}GpppA_{2'OMe}$ cap structure (Wang et al., 2021a; Park et al., 2022) which protects the viral nucleic acid molecules from host immune responses and also promotes translation of viral mRNA using eukaryotic machinery.

NiRAN domain is involved in two crucial reactions: nucleotidylation and de-nucleotidylation reaction, which are necessary for cap formation. In the NMPylation reaction, the nucleoside monophosphate (NMP) is transferred from nucleoside triphosphate (NTP) to Nsp 9 (protein substrate) in which the main chain amino group of N terminal asparagine of Nsp 9 attacks the α phosphorus of the NTP (present within the NiRAN pocket) to forms a phosphoramide bond with the release of pyrophosphate. Whereas in the deNMPylation reaction, the β phosphorous of GDP bound within the NiRAN active site attacks the NMPylated Nsp9 to release free Nsp 9 and GpppN.

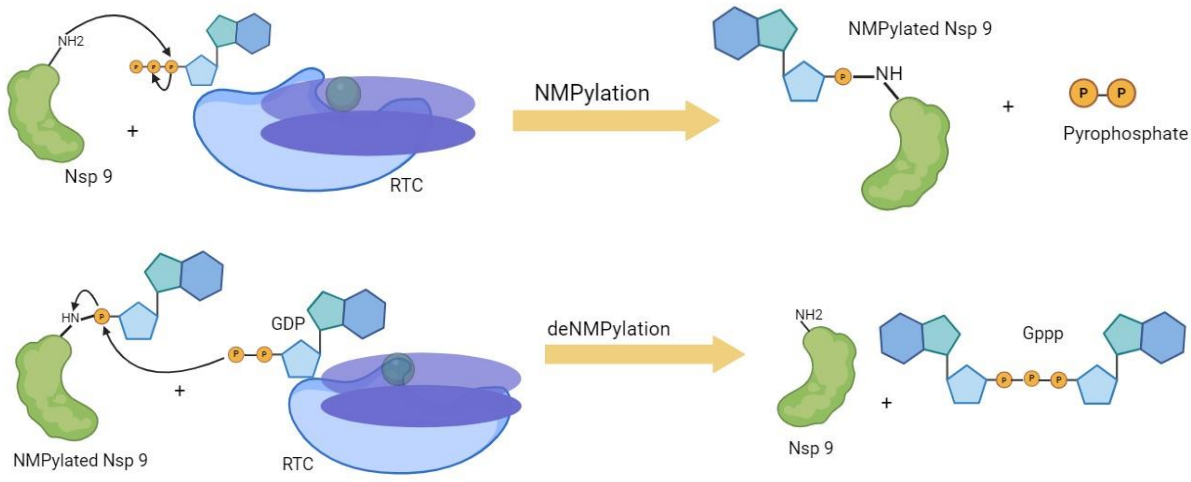


Figure 06: The schema for NMPylation and deNMPylation reactions (created using Bio Render)

When an RNA with 5' triphosphate take part in this reaction, it is termed RNAylation and deRNAylation, respectively. In the RNAylation reaction, the 5'pA-RNA is transferred from 5'pppA-RNA to Nsp9 in which the main chain amino group of N terminal amino acid (asparagine) of Nsp 9 attacks the α phosphorus of the 5'pppA-RNA to form a phosphoramide bond with the release of pyrophosphate. In the deRNAylation reaction, the β phosphorous of GDP bound within the NiRAN active site attacks the RNAylated Nsp9 to release free Nsp 9 and GpppA-RNA, forming the core cap structure. The methyltransferase domain of Nsp 14 can methylate the 5'GpppA-RNA structure to form cap 0 (7MeGpppA), and subsequent

methylation by Nsp 16 at the 2' sugar moiety leads to the formation of final cap structure $^{7Me}GpppA_{2'OMe}$.

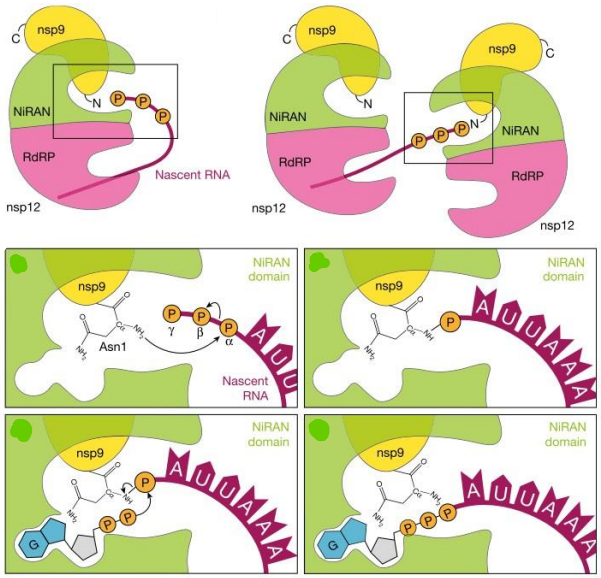


Figure 07: The model for RNA capping by NiRAN domain of SARS-CoV- 2 (Park et al., 2022)

Previously, many studies have proposed that these reactions occur within two poses of nucleotides within the NiRAN single active site. The two poses are "Base-In" and "Base-Out" (Malone et al., 2023; Park et al., 2022). Here, the Base-In pose was responsible for the de-nucleotidylation reaction, and the Base-Out pose was responsible for the nucleotidylation reaction. Recently, another team proposed an alternative model for NMPylation reaction through a newly discovered "Base Up" pose (Small et al., 2023). Through structural analyses and mutagenesis study, this project aims to delineate the molecular intricacies of the four reactions - NMPylation, deNMPylation, RNAylation and deRNAylation. The report encompasses a detailed account of experiments undertaken to explore different poses of nucleotides within the NiRAN pocket and their physiological relevance.

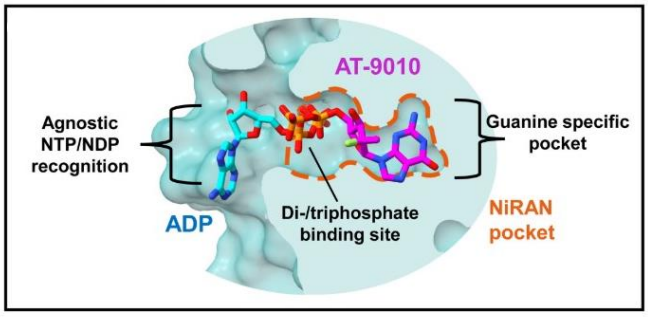


Figure 08: A schematic representation of Base-in and Base-out poses within the NiRAN pocket (Malone et al., 2023)

RNA-dependent RNA polymerase

The RNA-dependent RNA polymerase (RdRp) is a member of the superfamily of template-dependent nucleic acid polymerase and, on average, is 400 amino acids long. The conservation of sequence in RdRp is highly variable on average, with some regions showing less than 10 % conservation. In contrast, certain regions involved in nucleotide selection and catalysis are strongly conserved across various RNA viruses. Typically, RNA polymerases contain seven motifs named A, B, C, D, E, F (1-3) and G and are arranged in the order G, F1-3, A, B, C, D and E from N terminus to C terminus(Te Velthuis, 2014; Wang et al., 2020). Each of these motifs has a specific and conserved fold within the polymerase domain. The seven motifs and associated amino acid sequences are grouped into three subdomains. These sub-domains are called finger, palm and thumb by comparing the polymerase domain to the cupped right-hand (Gao et al., 2020).

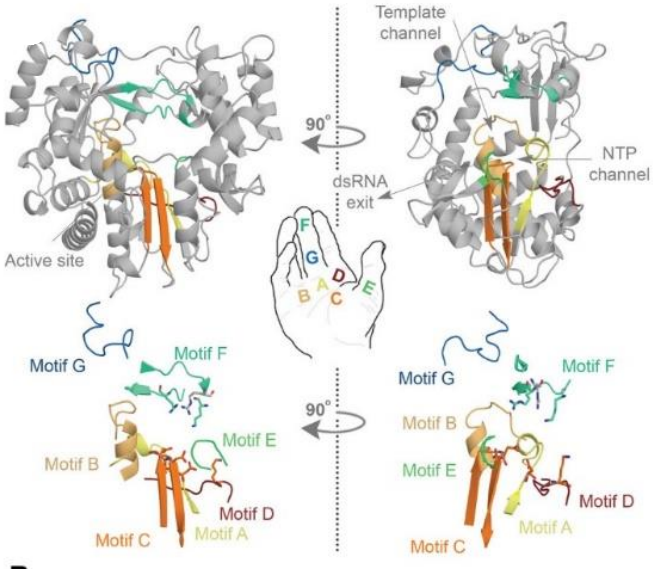


Figure 09: Shows the schematic representation of the seven motifs in the RdRp domain (Te Velthuis, 2014)

RNA viruses have smaller size variations in the genome compared to DNA viruses. The range of variation of the genome in DNA viruses is the order of magnitude 3, whereas, for RNA viruses, it is 1. This is fundamentally linked to the fact that RNA polymerases have lower fidelity in RNA synthesis. This is highlighted by the study which shows that RNA viruses can produce as much as one mutation per genome per replication (Lauber et al., 2013), which is mitigated by the production of vast amounts of progeny termed ‘quasispecies’ based on genome diversity (Holmes, 2011). This property has given rise to remarkable adaptability skills for RNA viruses, exploring new avenues of replicative mechanisms, exploring the environment

niche, etc. However, RNA viruses are still constrained by genome length as they must ensure the mutation load is below the error threshold, above which quasispecies survival is endangered (Gago et al., 2009). So, because of low fidelity RNA polymerase, its genome length is constrained. Because of constraints in length, it cannot acquire greater genome diversification to harbour a proofreading complex, which results in a low state of low-fidelity replication. This lock of fidelity, genome length and genome complexity among each other is termed the Eigen paradox. In other words, constrained genome length can be overcome by having a proofreading complex by the RNA viruses (Steinhauer et al., 1992).

Among the four classes of RNA viruses – (+) ssRNA, (-) ssRNA, dsRNA and retroviruses, only (+) ssRNA have an expanded RNA genome by comparative standards among them. And Nidoviruses belong to this class of viruses also harbour exoribonuclease machinery, which can perform the proofreading activity, but the largest CoV reported to date is planarian secretory cell nidoviruses (PSCNV), which 41.1kbp (Saberi et al., 2018) Still several orders of magnitude smaller than DNA viruses. So, nidoviruses have solved the Eigen paradox but still remain smaller genome viruses compared to viruses as a whole. This hints at the possibility that genome expansions may be decided by multiple factors, and polymerase nucleotide misincorporations may be one among them. Hence, the CoV RNA polymerase offers an exciting system for studying nucleotide misincorporation patterns.

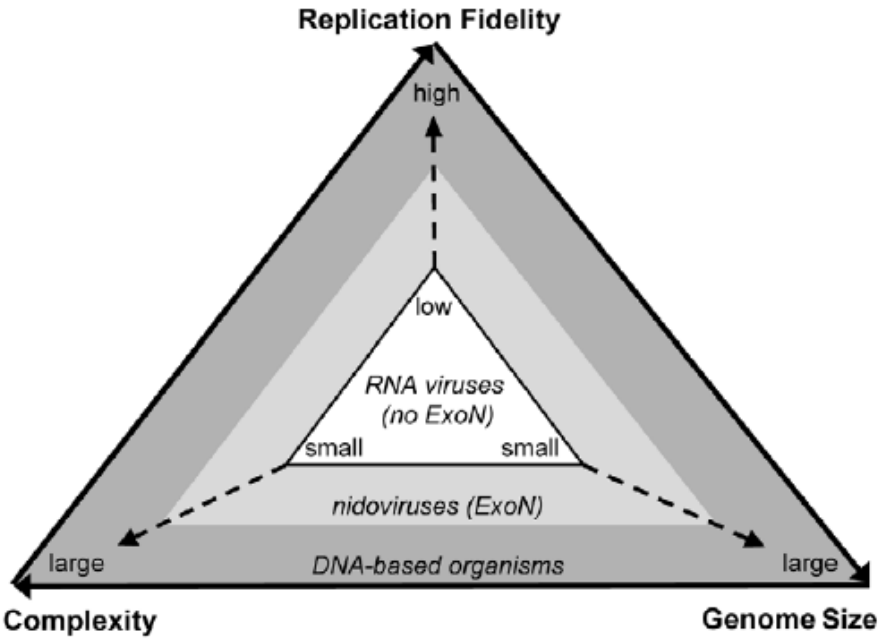


Figure 10: A schematic representation of Eigen paradox (Nga et al., 2011)

The family Coronaviridae and Roniviridae are characterised by viruses having large positive RNA genome, which brings the challenge of replicating this large genome quickly to escape the host's immune reaction. As a result, the CoV polymerases have double the rate of replication compared to the polymerase of other RNA viruses. Generally, as replication speed increases, the chance of misincorporating a wrong base increases, i.e., fidelity decreases. However, the nucleotide misincorporation level of Nsp 12 is similar to polymerases of slower polioviruses (smaller RNA viruses). This happens because Glu 547 is substituted by Ala in Nsp 12 of SARS-CoV-2; having Ala in motif F doubles the replication rate, and at the same time, the fidelity cost associated with a higher speed of replication is mitigated by having a Ser at 759 instead of Gly at 759 as in the case of polioviruses (Campagnola et al., 2022).

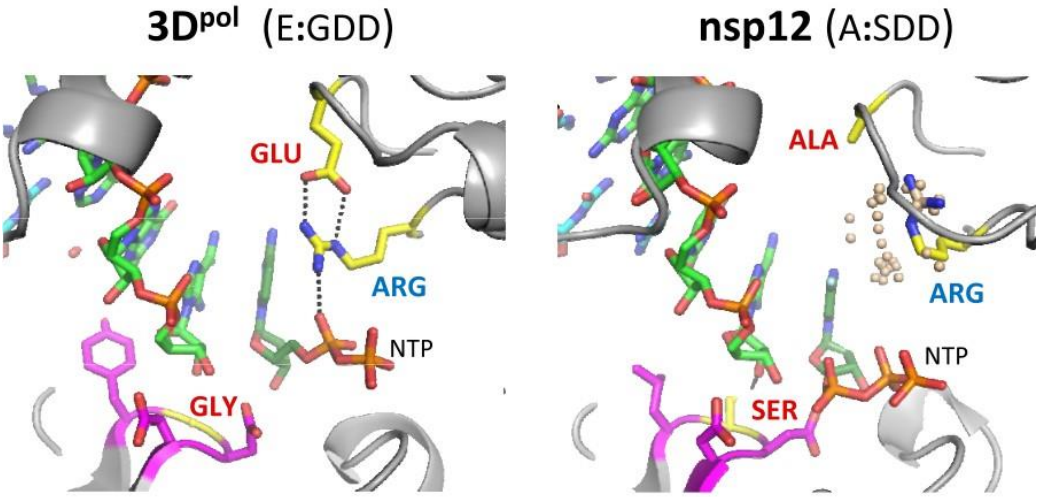


Figure 11: The figure represents the structural details of the RdRp polymerase domain of polioviruses (towards the left side) and SARS-CoV-2 (towards the right), representing the key residues that are involved in tuning the rate and fidelity of the polymerase (Campagnola et al., 2022).

Using this previous study of coronaviruses and polioviruses polymerases, we plan to study the nucleotide misincorporation by mutating the Nsp 12 of CoV RTC into Nsp12^{A547E}, Nsp 12^{S759G} and Nsp12$^{A, S\ (547,759)\ E, G}$. These mutants will change the rate, fidelity and both rate and fidelity, respectively, of CoV RTC, and studies conducted with this will shed light on how the nucleotide incorporation is affected by kinetic and fidelity variations of polymerases.

Chapter 02: Materials And Methods

Chapter 02: Materials and Methods

Cloning

Cloning of Nsp 12 Sub-domains mutants

To create the subdomain clones of Nsp 12, I used the pRSF Duet-1 (nsp7-nsp8) (nsp 12) plasmid (*Addgene Id*: 165451) as the parent vector (Madru et al., 2021). During the course of my project, I have prepared four subdomain clones, which include -14XHisInRdRp, 14XHisRdRp, 6XHisInRdRp and RTC RdRp. The constructs to be prepared require regions of Interface to RdRp (InRdRp) and RdRp alone to be cloned separately into pRSF vector retaining the 14-histidine tag and associated TEV site from the parent plasmid. To generate these constructs, we employed a Restriction-Free (RF) cloning strategy.

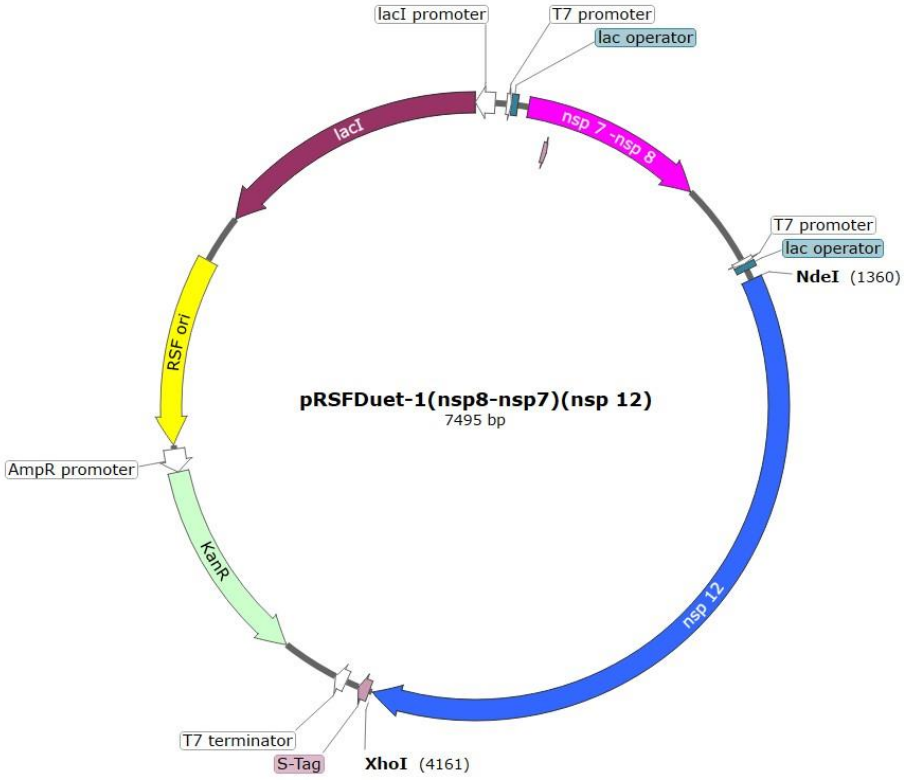


Figure 12: The diagram represents the parent plasmid

In this technique, the parent plasmid is transformed into a Dcm or Dam positive strain so that the plasmid DNA is methylated by bacterial Restriction Modification machinery. Using parent plasmid as a template, the **14XHisInRdRp** megaprimer was created through a PCR reaction via using a forward primer lacking regions complementary to nsp 7, nsp 8 & NiRAN and reverse primer containing region complementary to RdRp domain. The megaprimer for

14XHisRdRp was created in another PCR reaction using parent plasmid as the template and designing a forward primer lacking regions complementary to NiRAN, nsp 7, nsp 8 and interface domain and reverse primer containing region complementary to RdRp domain (as shown in Figure 13).

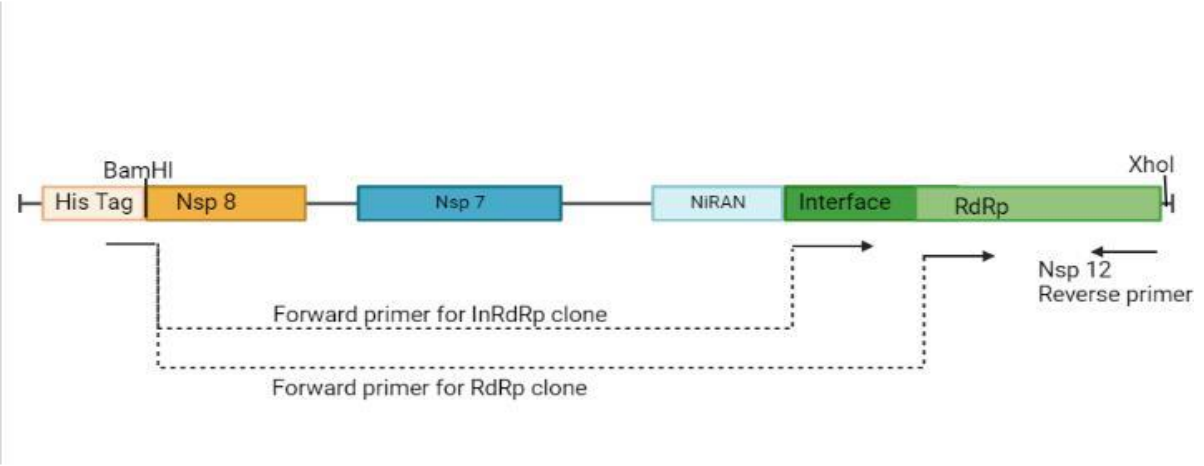


Figure 13: The diagram represents the cloning strategies for 14XHisInRdRp and 14XHisRdRp

Table 01: List of reagents and their concertation in PCR for megaprimer synthesis

Reagents	**Final Concentration**
Pfu Buffer	1X
dNTPs mix	100-250 µM
Forward primer	0.4 µM
Reverse primer	0.4 µM
Template DNA	100 ng
Pfu polymerase	0.6 µL in 50 µL

The megaprimer was purified using a column (from a Qiagen PCR purification kit) to remove proteins and buffer salt from the reaction mixture. Then, this megaprimer is used to perform Restriction Free Polymerase Chain Reaction (RF PCR) using the parent plasmid as a template. After the PCR, the samples are digested with Dpn I restriction enzyme for 6 hours to cleave

the methylated template strand and rescue the amplicon strands as they are unmethylated. Later, the digested product is transformed into electrocompetent NEB turbo cells through electroporation under 1800 V voltage, 25 μF capacitance and 200 Ω resistance. GenePulser Xcell electroporator was used for all the electroporation. After the pulsing, 150 μl of 1X LB media was added to the cells and kept for revival for 20 minutes at 37°C and later plated onto a kanamycin plate.

Whereas in **6XHisInRdRp** construct was generated using parent plasmid as a template with a forward primer containing regions complementary to InRdRp domain and RBS of pHis vector backbone and a reverse primer containing region complementary to RdRp domain and 6XHis tag (as shown in Figure 14). Then, this megaprimer is used to perform Restriction Free Polymerase Chain Reaction (RF PCR) using pHis plasmid as a template.

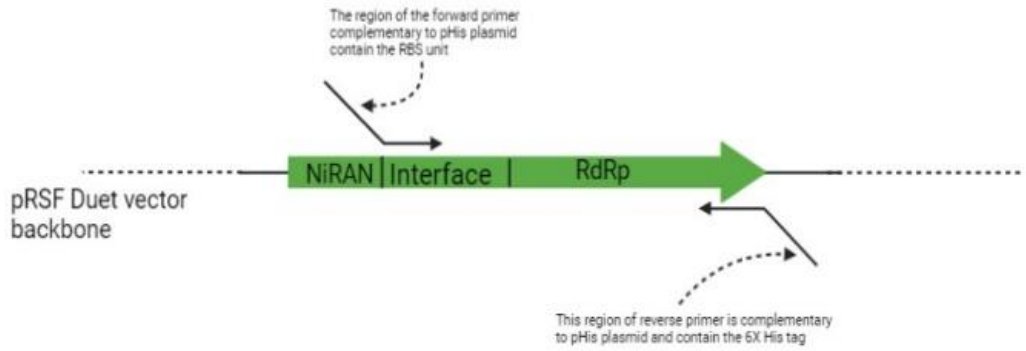


Figure 14: The diagram represents the cloning strategy of the 6XHisInRdRp monomer

In generating the construct **RTC RdRp,** the RdRp domain is co-transcribed along with Nsp 7 and Nsp 8. This construct was developed using a forward primer containing sequences complementary to the RdRp domain and an Nsp12 reverse primer. The megaprimer obtained in the first PCR was used for the second PCR with the parent plasmid as a template. But this time, the clone retains both Nsp 7 and Nsp 8 genes in it.

Cloning of NiRAN point mutants

For generating the constructs - **RTC Nsp 12^{T51A}**, **RTC Nsp 12^{N52A}**, **RTC Nsp 12$^{H,\ E(75,83)A,\ A}$** and **RTC Nsp 12^{N39A}**. I used parent plasmid as the template for all three except for **RTC Nsp 12$^{H,\ E(75,83)A,\ A}$**, in which **RTC Nsp 12^{H75A}** (created by Mohammad Navas) was used as a template. Using four different mutated primers (RTC_T51_Rev, RTC_N52A_Rev,

RTC_HE_Rev and RTC_N39A_Rev) and a common Nsp 8_primer1 all the mutants were generated (see the list of primers shown in Table 03). All protocols involving generating megaprimers, purifying the megaprimer and using it for RF PCR, later followed by digestion with Dpn I, remain the same. Except in the case of NiRAN mutants, the digested product was transformed into DH5α ultracompetent cells through heat shock at ice for 5 to 10 minutes and 37 °C for 3 minutes and immediately back to ice for 10 to 15 minutes. After the heat shock, 150 µl of 1X LB media was added to the cells, kept for revival for 60 minutes at 37°C, and later plated onto a kanamycin plate.

Table 02: List of reagents and their concertation in second PCR or RF PCR

Reagents	**Final Concentration**
Pfu Buffer	1X
dNTPs mix	100-250 µM
Megaprimer primer	50-100 ng/ µL
Template DNA	1 ng/ µL
Pfu polymerase	0.3 µL in 25 µL

Cloning of RdRp point mutants

For generating the constructs- **RTC Nsp 12^{A547E}**, **RTC Nsp 12^{S759G}** and **RTC Nsp 12$^{A, S(547,759)E, G}$** used parent plasmid as a template and using two mutated primers for generating the Nsp12 mutated megaprimers. The two mutated primers and two common nsp 12_primer3 and nsp 12 rev primers were used for cloning this construct (list of primers shown in Table 03). For generating the **RTC Nsp 12^{A547E}** megaprimer, I used A547E_F and nsp 12_rev primer; for developing the **RTC Nsp 12^{S759G}** megaprimer, I used nsp12_primer3 and S759G_rev and for generating **RTC Nsp 12$^{A, S(547,759)E, G}$** megaprimer I used A547E_F and S759G_rev. All other protocols remain similar (as discussed in the previous section) to that of NiRAN mutants.

Table 03: List of primers used for all the cloning and their sequences

Name of the Primer	Sequence of the primer in 5' to 3' direction
Nsp12_InRd_F	GTATTTTCAAGGGACGGGGGATGGATCCACCGATCTGACAAAACC
Nsp12_Rd_F	GTATTTTCAAGGGACGGGGGATGGATCCCTGAGCTTTAAAGAACTG
pHis_InRdRp_F	GTTTAACTTTAAGAAGGAGATATACATATGACCGATCTGACAAAACCG
pHis_InRdRp_R	GCTTTTAATGATGATGATGATGATGGGATCCCAGCACGGTATGCGGTG
RdRp_RTC_F	GTATAAGAAGGAGATATACAT ATGCTGAGCTTTAAAGAACTGCTGG
Nsp12_Rev	GCGGTTTCTTTACCAGACTCGAGTTACAGCACGGTATGC
RTC_T51A_Rev	GGCAACAATTGGCTTTCAGAAATTTGGC
RTC_N52A_Rev	GCGGCAACAAGCGGTTTTCAGAAATTTGGC
RTC_HE_Rev	GCAGATTGTAAATGGTTTCCGCGTGCTGATAGTTGC
RTC_N39A_Rev	CCGGCAACTTTATCGGCGTAGATATCAAATGCACG
Nsp8_primer1	CCTGCATTAGGAAATTAATAC
A547E_F	CTGAAATACGAAATTAGCGCAAAAAATC
S759G_R	GGCATCATCACCCAGAATCATCATGC
nsp12_primer3	CCGAATTGTGTTAATTGTCTG

Screening of Nsp 12 Sub-domain mutants

For screening of colonies, multiple single colonies from the test plate were inoculated on multiple 5ml 1X LB containing kanamycin antibiotics and incubated overnight. The bacteria were pelleted down by spinning at 5000 rpm for 7 minutes at 4°C. The supernatant was discarded, and the pellet was resuspended using 250 µl of solution I (details shown in Table 04), followed by adding 200 µl of solution II and mixed well by inverting the closed microcentrifuge tube 4-6 times. Later, 200 µl of solution III was mixed. Followed by spinning the sample at 15000 rpm for 15 minutes at 4°C. The supernatant was discarded, and 70% ethanol was added to the pellet, and given a short spin at 4°C. The supernatant was discarded, and the sample was dried at 55-65°C for 15 minutes and resuspended using deionized autoclaved water. The purified plasmid was kept for digestion with BamHI and Xho I overnight. And loaded onto a 1% agarose gel to check the presence of respective insert bands under UV light and the image captured by BioRad ChemDoc Imager.

Table 04: List of solutions used for plasmid purification from bacterial pellet

Buffer Name	**Composition**
Solution I	50 mM Tris (pH: 8), 10 mM EDTA and 100µg/mL RNAase A
Solution II	200 mM NaOH and 1% (w/v) SDS
Solution III	3M CH_3COOK (pH = 5.5)

Screening of NiRAN and RdRp point mutants

As the point mutations didn't alter the restriction sites within the clones, the screening of NiRAN point mutant and RdRp point mutants was done directly through sequencing the plasmid purified from the clone grown from the test plate. For the purification of plasmids from the bacterial lysate, we used a Qiagen miniprep kit. The sequencing was provided by Barcode Bioscience.

Recombinant Protein Expression check and Optimization

Protein expression check of Nsp 12 Sub-domain mutants

For checking the recombinant protein expression, the plasmid DNA of various clones obtained in the cloning step was transformed into a suitable strain through heat shock, in which 100 ng of the DNA was added to 50 µl of cells and incubated on ice for 5 to 10 minutes and given a heat shock at 37°C for 3 minutes and kept back on ice for 10 to 15 minutes. After the heat shock, 150 µl of 1X LB media was added to the transformation reaction mixture and kept for revival for 20 minutes at 37°C and later plated onto a suitable antibiotic plate. A single colony was picked for setting the primary culture containing kanamycin or a combination of kanamycin with chloramphenicol (for strains like Rosetta 2 (DE3) and Codon plus RIL). Later, the secondary culture containing respective antibiotics or its combination was inoculated with 1% of the primary culture. The secondary culture was grown till the OD 600 nm reached 0.6 to 0.8. Then, one of the secondary cultures was induced with an appropriate inducer like - 0.1 mM IPTG (Isopropyl β-D-1 thiogalacto pyranoside); after the induction, the culture was shifted to an incubator at a lower temperature of 16 °C. Post induction, the culture was kept for a long incubation period of 18 hours, and 5 ml of induced culture was taken to prepare the "Total" and "Pellet" fractions. The OD normalized volume of uninduced culture was transferred to the "Uninduced" fraction. The culture was spun to pellet the bacteria. Total and Uninduced fractions were resuspended in 5X TGS (Tris-glycine sodium dodecyl sulphate), whereas the "Pellet" fraction was resuspended in the lysis buffer (500 mM sodium chloride, 50 mM of sodium HEPES of pH 8 and 10mM Imidazole). Sonication was performed for all three fractions for 1 minute for 5 seconds on and 5 seconds off at 60 % amplitude. The pellet fraction was spun in a centrifuge at maximum speed for 15 minutes- the supernatant was transferred to another vial named "Soluble" fraction, and the obtained pellet was resuspended in 5X TGS. An aliquot of the four fractions was heated to denature the protein and loaded on a 12% SDS PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis). The gel was run at 210 V for 40 minutes and, stained by Coomassie Blue and later imaged by E-Gel Imager by Life Technologies.

Protein expression check of NiRAN and RdRp point mutants

Most parts of protocols involving transforming clone constructs to various expression strains and growing of primary and secondary cultures expect that the complex constructs (containing Nsp 7 and Nsp 8) be grown at 30 °C postinduction for 8 hours instead of 16 °C for monomer constructs. The induction was done with 0.1 mM IPTG in both cases. The remaining part of

processing the samples for denatured proteins remains the same, followed by loading it onto 15% SDS PAGE and staining it using Coomassie Blue and later imaged by E-Gel Imager by Life Technologies.

Protein Purification

Protein Purification of Nsp 12 Sub-domain mutants

The construct plasmids generated during the cloning step were transformed to their respective expression strain (which was optimized during a small culture expression check) by heat shock.

Table 05: List of buffers used for all the protein purifications

Buffer Name	**Composition of the Buffer**
Ni-NTA equilibration buffer (Sub-domain mutant)	50 mM NaHEPES (pH = 8) and 500 mM NaCl
Ni-NTA equilibration buffer (NiRAN mutant)	50 mM NaHEPES (pH = 8), 500 mM NaCl and 10 mM Imidazole
Ni-NTA elution buffer	50 Mm NaHEPES (pH = 8), 500 mM NaCl and 500 mM Imidazole
Hitrap QHP equilibration buffer (Sub-domain mutant)	50 mM NaHEPES (pH = 8) and 100 mM NaCl
Hitrap QHP equilibration buffer (NiRAN mutant)	50 mM NaHEPES (pH = 8) and 120 mM NaCl
Hitrap QHP elution buffer	50 mM NaHEPES (pH = 8) and 1000 mM NaCl
RTC SEC Buffer	20 mM Tris (pH = 8), 200 mM KCl and 1 mM $MgCl_2$
RTC Dilution Buffer	50mM Tris (pH = 8)

For protein purification of mutants obtained at the cloning step, the constructs were transformed to their respective expression strain and plated on suitable antibiotic plates. A single colony was picked for setting up the primary culture containing kanamycin or a combination of kanamycin with chloramphenicol (for strains like Rosetta 2 (DE3)) and incubated at 37 °C overnight. The next day, the primary culture was inoculated to two to three 1L of 1X LB containing appropriate antibiotics or its combination and incubated at 37°C. The secondary culture was grown till the OD at 600 nm reached 0.6 to 0.8. Then, all the secondary cultures were induced with 0.1 mM IPTG (Isopropyl β-D-1 thiogalacto pyranoside); after the induction, the culture was shifted to an incubator at a lower temperature of 16 °C. Post-induction, the culture is kept for a long incubation period of 18 hours. After the long incubation period, cultures were harvested and pelleted down using a Thermo Scientific high-speed floor centrifuge, and the pellet was flash-frozen and stored at -80 °C.

The pellet was thawed and resuspended in 60-75 mL of lysis buffer and lysed using a sonicator for 3 minutes duration with a pulse of 1 second ON and 3 seconds OFF at 60 % amplitude for three cycles. The lysate was spun at 16000 rpm for 45 minutes at 4°C using Avanti J-26XP (Beckman Coulter Life Science). The resulting supernatant fraction was loaded onto a 5 mL Ni-NTA (Sigma-Aldrich) column pre-equilibrated with Ni-NTA equilibration buffer. A 16-18 CV (column volume) wash was given to the column using Ni-NTA equilibration buffer to remove the non-specifically bound impurities to the column. The protein was eluted with increasing concentration of imidazole in a step-wise gradient – 5%B, 10%B, 20%B, 30%B, 50%B and 100%B where B is Ni-NTA elution buffer. The fraction containing the protein of interest was pooled and diluted using a dilution buffer so that the resulting NaCl concentration of the protein sample was set similarly to the NaCl concentration of the equilibration buffer of QHP. The diluted protein sample is then loaded onto a 5 mL HiTrap QHP column pre-equilibrated with equilibration buffer QHP_Sub-domain. This was followed by a 16-18 CV wash given to the column using equilibration buffer QHP_Sub-domain. The protein was eluted using increasing concentration of NaCl in a linear gradient. The fractions containing the protein of interest were pooled and concentrated using centricon (Merck Millipore) and loaded onto a size exclusion chromatography (SEC) 200 10/300 or 650 (Bio-Rad) to check the homogeneity of the protein of interest. The fractions containing pure protein were pooled and concentrated using a 30 kDa centricon. The concentrated protein was given a short spin at 15000 rpm for 7 minutes at 4°C and aliquot into 200 µL tubes and flash frozen and stored at -80 °C. The end purified protein sample was used for all assay purposes.

Protein Purification of NiRAN point mutants

For protein purification of NiRAN mutants, all protocols involving culturing of clones, pelleting down the cells, and sonicating the cells remain the same except that the Ni-NTA equilibration buffer (NiRAN mutant) contains 10 mM Imidazole whereas Ni-NTA equilibration buffer of Sub-domain mutant doesn't contain any imidazole, a change incorporated aftermath of 14XHisInRdRp purification (discussed more on Result and Discussion chapter). Another difference accounts for the equilibration buffer of HiTrap: the equilibration buffer contains 120 mM NaCl concentration for NiRAN mutant, unlike the equilibration buffer of the Sub-domain, which has only 100 mM NaCl. The increase in NaCl concentration was incorporated to remove excess Nsp 8 during the HiTrap wash step.

All other steps involving protein concentration, loading it into SEC 650 or Superdex 200, further concentrating of protein, flash freezing and storage remain the same. The result chapter encompasses details of where SEC 650 was used and where Sup 200 was used. There were no particular reasons to select one column over the other. The use of a column from the two kinds of SEC columns was determined purely by the availability of the column in the lab.

Extension Assay

Extension Assay for Nsp 12 Sub-domain mutants

The purified RdRp domain is mixed with Nsp 7 and Nsp 8 (purified by Ashwin Uday) and incubated with cyanine fluorophore labelled template primer RNA duplex along with nucleotide at 37 °C for one hour. The reaction is stopped using formamide EDTA dye and heated at 98 °C for 10 to 15 minutes to denature the protein. And loaded onto 21% Urea PAGE and ran for two hours at 230 V. The gel is visualized for the labelled primer strand at cyanine 5 wavelengths using Amersham Typhoon (Cytiva).

Table 06: The sequence of RNA primer and template used for generating the RNA duplex

Name of the RNA strand	**Sequence of the RNA strand (5' to 3')**
Primer (25 mer)	[Cy5] CAGUGAGUCGUAUUA AUCCCGUGA A
Template (45 mer)	[FAM] CUAUAUUCAUAAAGCUGCUAUUCA CGGGAUUAAUACGACUCACUG

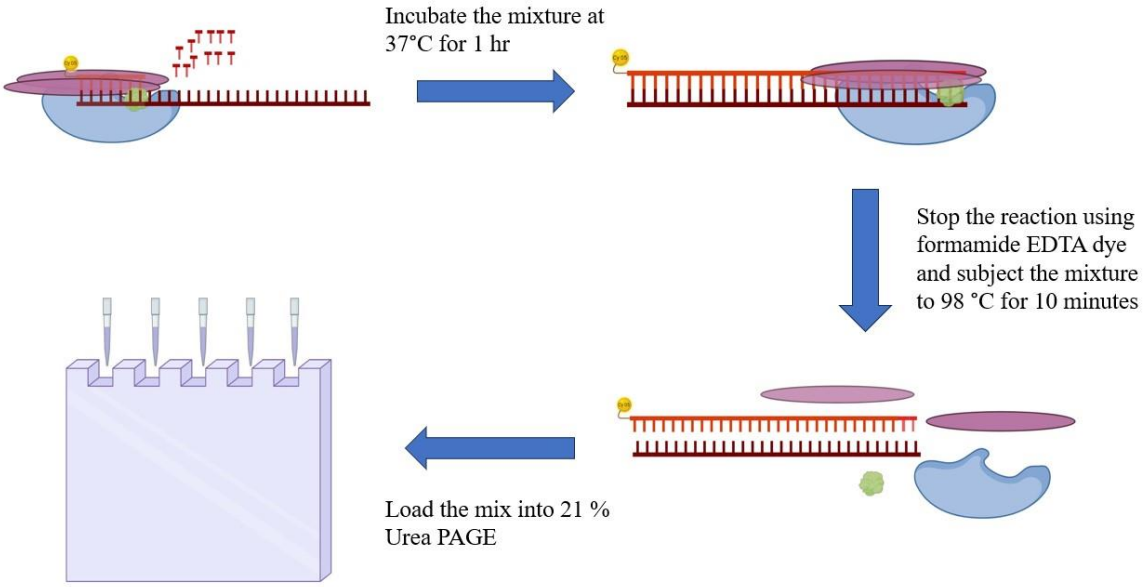


Figure 15: Represents the schematic description of the extension assay

Table 07: Concentration of various components in the extension assay

Reagents	**Concentration**
10X Extension buffer (100 mM KCl, 100 mM Tris (pH:8),20mM $MgCl_2$,10mM DTT)	1X
RNA template (L25/L45)	20 nM
14XHis RdRp or RTC	1 µM
Nsp 8	3 µM
Nsp 7	2 µM
RTC	250 nM
Nucleotides (rNTP mix)	200 µM

Extension Assay for NiRAN point mutants

The mutant RTC protein along cyanine fluorophore labelled template primer RNA duplex (the same sequence primer and template were used) was incubated with nucleotide at 37 °C for one hour. The reaction is stopped using formamide EDTA dye and heated at 98 °C for 10 to 15 minutes to denature the protein. And loaded onto 21% Urea PAGE and ran for two hours at 230 V. The gel is visualized for the labelled primer strand at cyanine 5 wavelengths using Amersham Typhoon (Cytiva).

Table 08: Concentration of various components in the extension assay of NiRAN mutants

Reagents	**Concentration**
10X Extension buffer (10 mM KCl, 10 mM Tris (pH:8),2mM $MgCl_2$,1mM DTT)	1X
RNA template (L25/L45)	20 nM
RTC Nsp 12^{T51A}, RTC Nsp 12^{N52A} RTC Nsp $12^{H, E (75,83) A, A}$ and RTC Nsp 12^{N39A}	250 nM
RTC Nsp 12 WT	250 nM
Nucleotides (rNTP mix)	200 μM

NMPylation Assay

The wild-type RTC and mutant RTC (at NiRAN) were incubated with Nsp 9 at 37 °C in the presence of one of the four canonical NTPs for 45 minutes. The reactions were stopped using SDS loading dye containing Bromophenol Blue, SDS and DTT, heated at 98 °C for 10 to 15 minutes, and loaded onto 19 % SDS PAGE. The gel was run at 200 V for 118 minutes. The mixture was stained using Coomassie Blue and destained using an acetic acid and ethanol solution. Imaged using E-Gel Imager by Life Technologies.

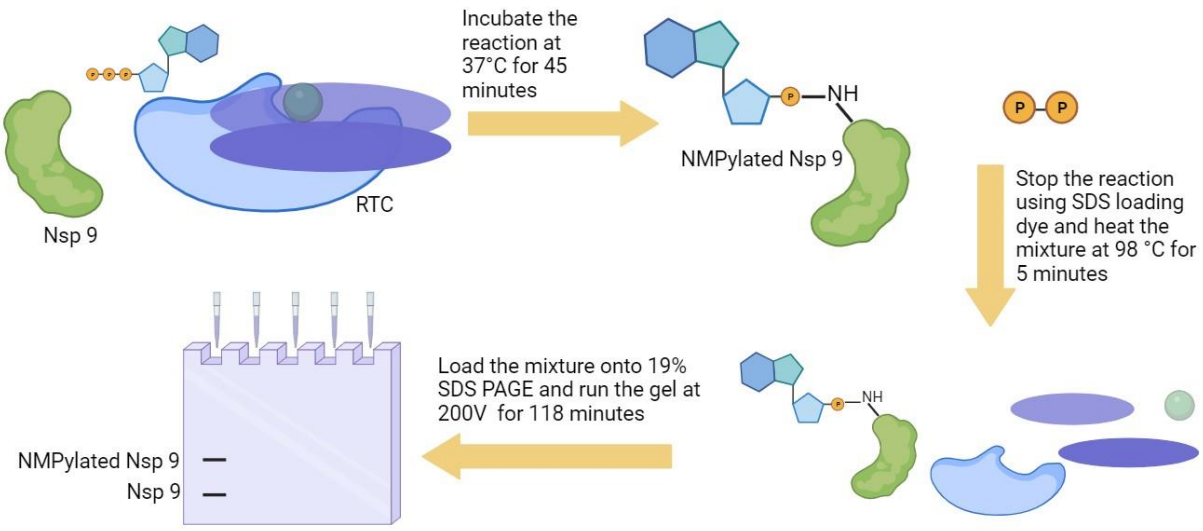


Figure 16: Represents the schematic description of the NMPylation assay

Table 09: Concentration of various components in the NMPylation assay

Reagents	**Concentration**
5X Extension buffer (50 mM KCl, 125 mM Tris (pH:8),25mM $MgCl_2$,5mM DTT)	1X
Nsp 9	7 µM
RTC Nsp 12^{T51A}, RTC Nsp 12^{N52A} RTC Nsp $12^{H, E (75,83) A, A}$ and RTC Nsp 12^{N39A}	250 nM
RTC Nsp 12 WT	250 nM
NTPs	300 µM

deNMPylation Assay

The wild-type RTC and mutant RTC (at NiRAN) were incubated with AMPylated Nsp 9 at 37 °C in the presence of GDP for 45 minutes. The reactions were stopped using SDS loading dye containing Bromophenol Blue, SDS and DTT, heated at 98 °C for 10 to 15 minutes, and loaded onto 19 % SDS PAGE. The gel was run at 200 V for 118 minutes. The mixture was stained

using Coomassie Blue and destained using an acetic acid and ethanol solution. Imaged using E-Gel Imager by Life Technologies.

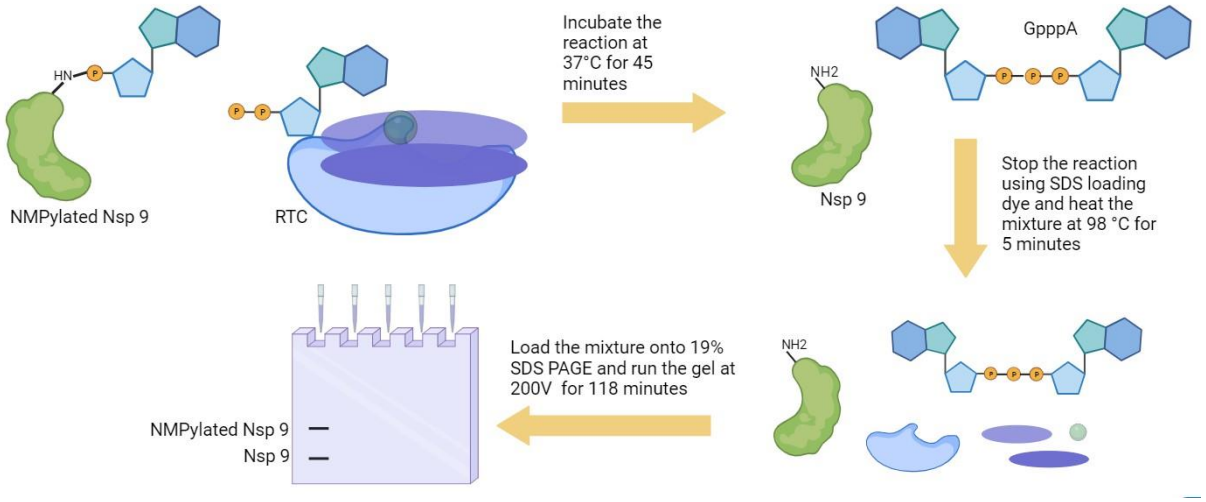


Figure 17: Represents the schematic description of the deNMPylation assay

Table 10: Concentration of various components in the deNMPylation assay

Reagents	**Concentration**
5X Extension buffer (50 mM KCl, 125 mM Tris (pH:8),25mM $MgCl_2$,5mM DTT)	1X
AMPylated Nsp 9	7 µM
RTC Nsp 12^{T51A}, RTC Nsp 12^{N52A} RTC Nsp $12^{H, E(75,83) A, A}$ and RTC Nsp 12^{N39A}	750 nM
RTC Nsp 12 WT	750 nM
GDP	1500 µM

Chapter 03 Results

Chapter 03: Results

Screening of Clones

Screening of Nsp 12 Sub-domains mutants

All clones obtained positive from restriction mapping were sent for sequencing to Barcode Biosciences and were confirmed positive.

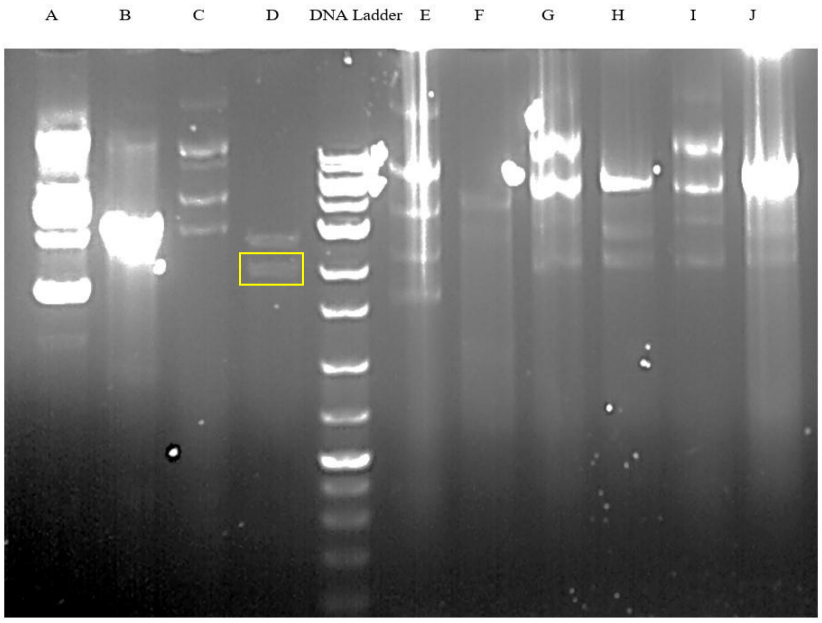


Figure 18: Screening of **6XHisInRdRp** clones labelled as 13, 14, 15, 16 & 17. Lanes A, C, E, G & I represent uncut plasmid clones. Lanes B, D, F, H & J, respectively, represent clones 13, 14, 15, 16 & 17 digested with BamHI + NdeI. Clone 14 turned out to be the positive one, which shows the required bands of 2028 bp and 2750 bp.

Five clones named 13, 14, 15, 16 and 17 were cultured, and plasmids were isolated as discussed in the section "screening of sub-domains" of the chapter "Materials and Methods". This was followed by the overnight digestion of purified plasmids by BamHI and NdeI and, loaded the digestion mixture to 1% Agarose gel and ran for 45 minutes. Clone 14, which contained the insert of the required size, i.e., 2028 bp, was transformed to NEB turbo cells by heat shock and cultured to purify plasmid construct using a Qiagen miniprep kit. The purified plasmid was sent for DNA sequencing to Barcode Bioscience. Later, clone 14 conformed to the positive for 6XHisInRdRp by sequencing.

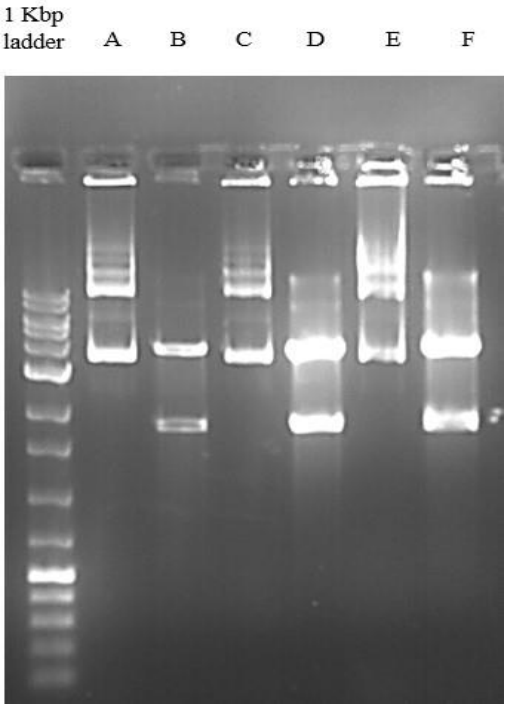


Figure 19: Screening of **14XHisRdRp** clones labelled as 1, 2 & 3. Lanes A, C & E, respectively, represent uncut plasmid clones. And Lanes B, D & F, respectively, represent clones 1, 2 & 3 digested with BamHI + XhoI. All three clones turned out to be positive, showcasing required bands of sizes- 1707 bp and 3672 bp.

Three clones named 1, 2 and 3 were cultured, and plasmids were isolated, as discussed in the screening section of the Subdomains of chapter materials and methods, followed by the digestion of purified plasmids by BamHI and XhoI overnight. Subsequently, the digestion mixture was loaded into 1% Agarose gel and ran for 45 minutes. All three clones contained the insert of the required size, i.e., 1707 bp. All three plasmid constructs were transformed to NEB turbo cells by heat shock and cultured to purify plasmid constructs using a Qiagen miniprep kit. The purified plasmids were sent for DNA sequencing to Barcode Bioscience. Later, clone 1 was confirmed positive for 14XHisRdRp. Clones 2 and 3 contained random point mutation, which altered the frame of the RdRp operon and were discarded later.

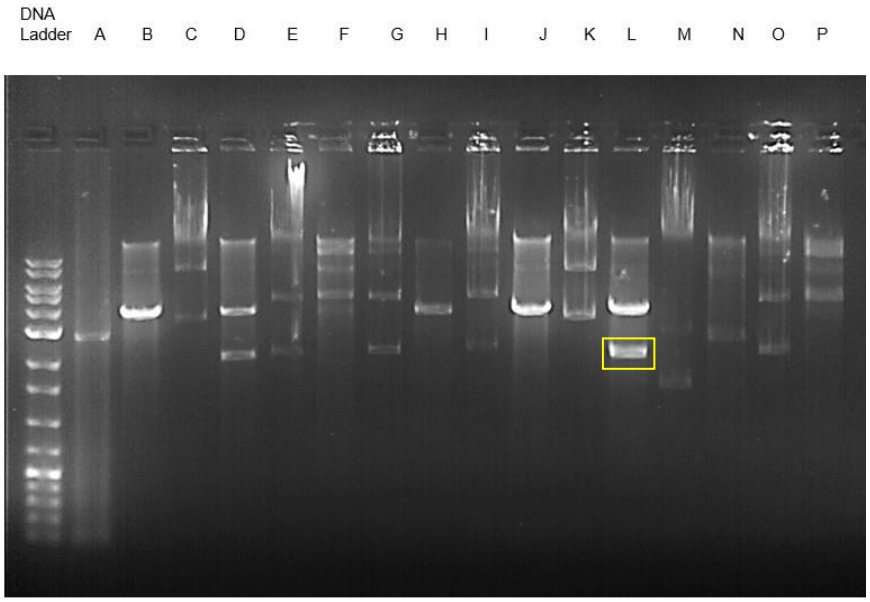


Figure 20: Screening of **14XHisInRdRp** clones labelled as 1, 2, 7, 8, 9,10 & 11. Lane A represents uncut parent plasmid; lane B represents parent plasmid + BamHI + XhoI. Lanes C, E, G, I, K, M & O, respectively, represent uncut plasmid clones. And Lanes D, F, H, J, L, N &P, respectively, represent clones 1, 2, 7, 8, 9,10 & 11 digested with BamHI + XhoI. Clone 9 has the two required bands of sizes- 2028 bp and 3672 bp.

Three clones named 1, 2, 7, 8, 9, 10 and 11 were cultured, and plasmids were isolated as discussed in the screening section of sub-domains of chapter materials and methods and followed by digestion of purified plasmids by BamHI and XhoI for overnight. Subsequently, the digestion mixture was loaded into 1% Agarose gel and ran for 45 minutes. Clone 9 contained the insert of the required size, i.e., 1707 bp. The plasmid construct of clone 9 was transformed into NEB turbo cells by heat shock and cultured to purify the plasmid construct using a Qiagen miniprep kit. The purified plasmid was sent for DNA sequencing to Barcode Bioscience. Later, clone 9 was confirmed positive for 14XHisInRdRp.

Screening of NiRAN and RdRp point mutants

As these are point mutants, this mutation didn't alter the restriction sites of the parent plasmid. We screened it directly by sending it for DNA sequencing to Barcode Bioscience. And we obtained positive clones for RTC Nsp 12^{T51A}, RTC Nsp 12^{N52A}, RTC Nsp $12^{H, E (75,83) A, A}$, RTC Nsp 12^{N39A}, RTC Nsp 12^{A547E}, RTC Nsp 12^{S759G} and RTC Nsp $12^{A, S (547,759) E, G}$.

Recombinant Protein Expression check and Optimization

Protein expression check of Nsp 12 Sub-domain mutants

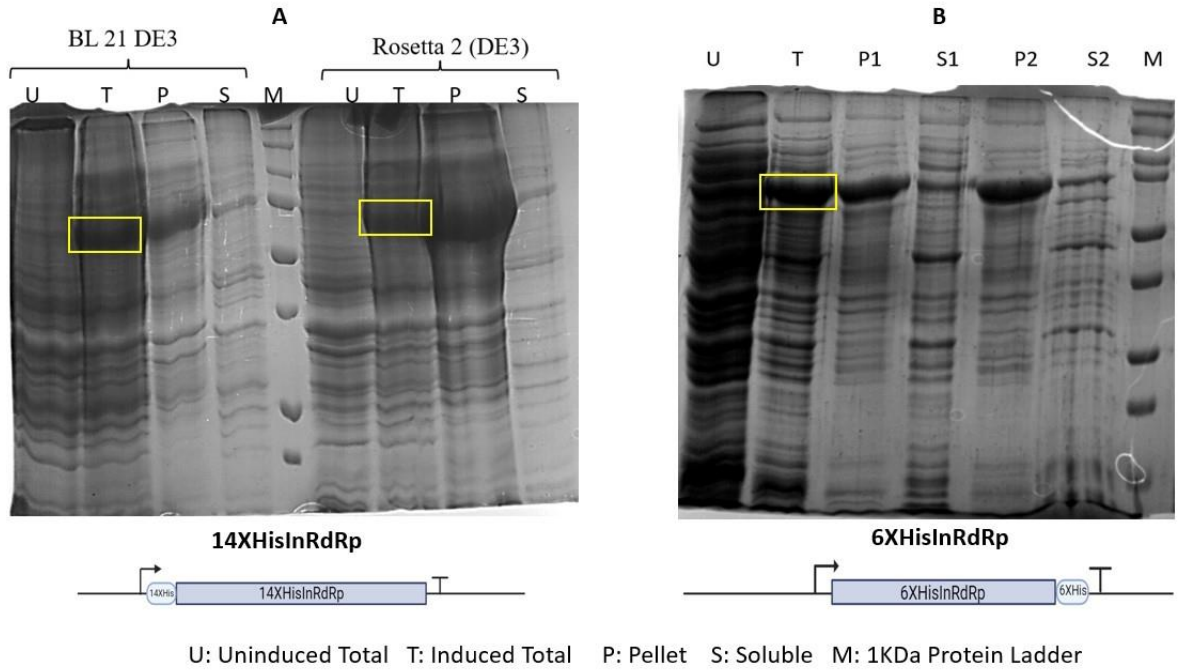


Figure 21: A) Expression check of **14XHisInRdRp** in BL21 DE3 and Rosetta 2 (DE3). The lane towards the left of the protein ladder represents data for BL21 DE3, whereas the lane towards the right of the protein ladder represents data for the Rosetta strain. The lanes U, T, P and S represent the uninduced total, induced total, and pellet and soluble fractions, respectively. The induction was done with 0.1 mM IPTG, and the induced culture was incubated at 16 °C for 16 hours post-induction. **B**) Expression check of **6XHisInRdRp** Rosetta 2 (DE3). The lane U, T, P1, P2, S1 and S2 represent the uninduced total, the induced total, and the pellet and soluble fractions, respectively. S1 and S2 represent two different lysis buffer conditions (**Lysis buffer 1 = 50 mM NaHEPES (pH = 8) + 500 mM NaCl**. And **Lysis buffer 2 = 50 mM NaHEPES (pH = 8) + 500 mM NaCl +1% CHAPSO**). The induction was done with 0.1 mM IPTG, and the induced culture was incubated at 16 °C for 16 hours post-induction.

The molecular mass for 14XHisInRdRp and 6XHisInRdRp are 81 kDa and 78 kDa, respectively. The bands are running differently from the desired size. Both buffer conditions have good protein expression and a significant amount of protein in the soluble fraction. Concluded the use of BL21 DE3 strain as the expression strain for both 14XHisInRdRp and 6XHisInRdRp constructs. To maintain uniformity, we decided to use Lysis Buffer 1 for further downstream protein purification protocols.

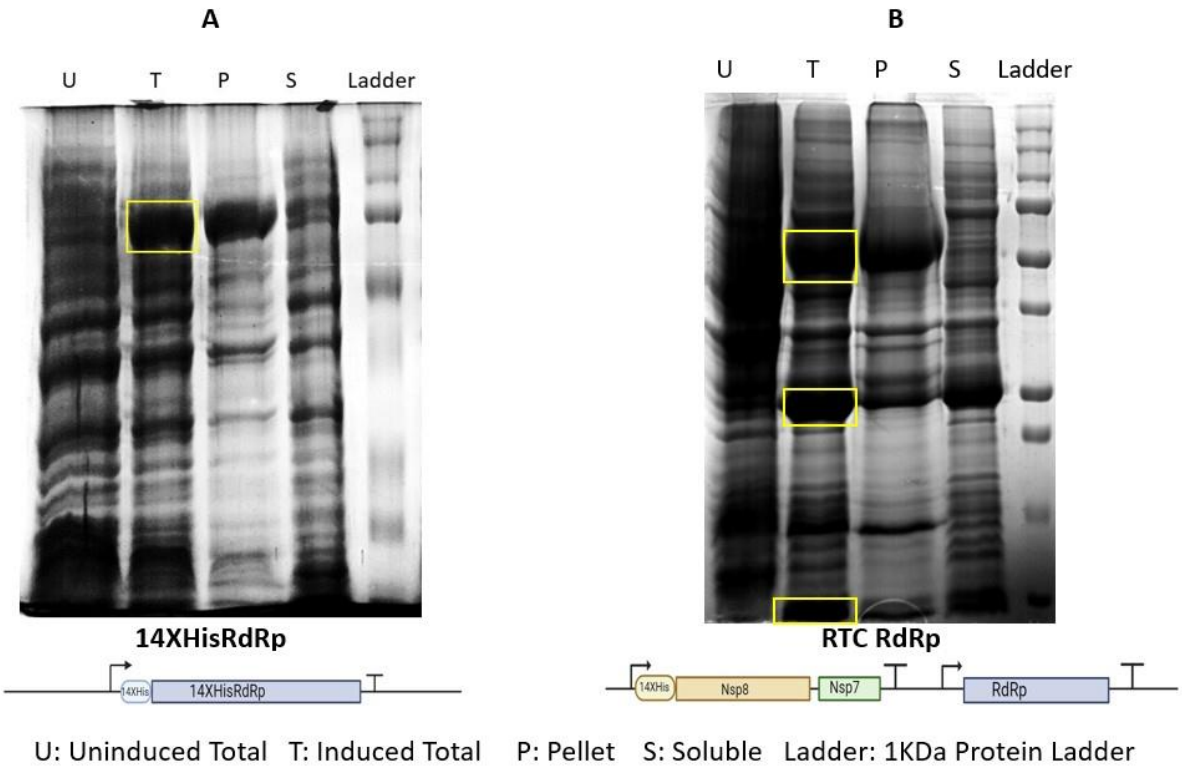


Figure 22: A) Expression check of **14XHisRdRp** in Codon plus RIL strain. The lanes U, T, P and S represent the Uninduced total, Induced total, Pellet and Soluble fractions, respectively. The induction was done with 0.1 mM IPTG and post-induction, the culture was incubated at 16 °C for 16 hours.

B) Expression check of **RTC RdRp** in Codon plus RIL strain. The lanes U, T, P and S represent the Uninduced total, Induced total, Pellet and Soluble fractions, respectively. The induction was done with 0.1 mM IPTG and post-induction, the culture was incubated at 30 °C for 8 hours. The top, middle and lower bands correspond to the RdRp domain, nsp 8 and Nsp 7.

The protocols for culturing, induction, and sample preparation were followed as described in the chapter materials and methods section for recombinant protein expression optimization. The molecular mass for 14XHisRdRp and RTC RdRp are 69 and 125 kDa, respectively. The 14XHisRdRp band is running higher compared to its actual size. RTC RdRp comprises three bands: RdRp (64 kDa), Nsp 8 (25 kDa) and Nsp 7 (10 kDa). A good level of protein expression is seen, and a significant amount of protein is present in the soluble fraction. We used Codon plus RIL strain as the expression strain for both 14XHisRdRp and RTC RdRp constructs with post-induction temperature 16°C for 14XHisRdRp and 30°C for RTC RdRp.

Protein expression check of NiRAN point mutants

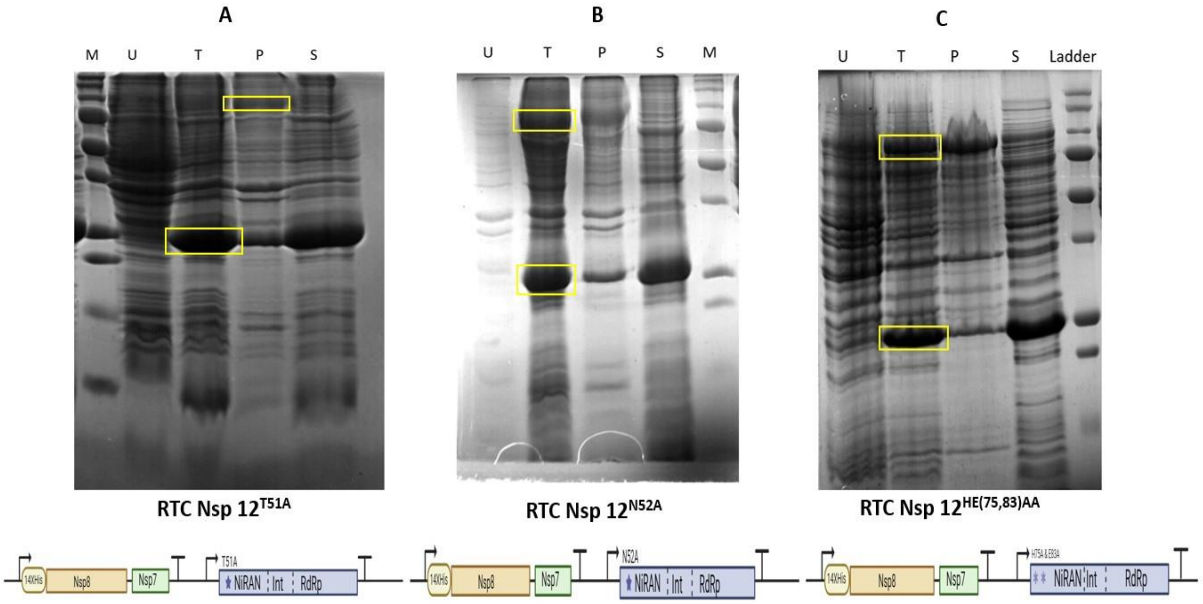


Figure 23: A) Expression check of **RTC Nsp 12^{T51A}** in Rosetta 2 (DE3) strain. The lanes M, U, T, P and S represent the 100 kDa protein ladder, including the uninduced total, induced total, pellet and soluble fractions. The induction was done with 0.1 mM IPTG and post-induction, the culture was incubated at 30 °C for 8 hours. **B)** Expression check of RTC Nsp 12^{N52A} in BL21 DE3 strain. The lanes M, U, T, P and S represent the 100 kDa protein ladder, including the uninduced total, induced total, pellet and soluble fractions. The induction was done with 0.1 mM IPTG and post-induction, the culture was incubated at 30 °C for 8 hours (Nsp 7 had run out). **C)** Expression check of RTC Nsp 12$^{H, E (75,83) A, A}$ in BL21 DE3 strain. The lanes M, U, T, P and S represent the 100 kDa protein ladder, including the uninduced total, induced total, pellet and soluble fractions. The induction was done with 0.1 mM IPTG, and post-induction, the culture was incubated at 30 °C for 8 hours.

The protocols for culturing, induction, and sample preparation were followed as described in the chapter materials and methods section for recombinant protein expression optimization. The molecular mass of RTC is 160 kDa. The RTC comprises three bands: Nsp12 (106 kDa), Nsp 8 (25 kDa) and Nsp 7 (10 kDa). A good level of protein expression is seen, and a significant amount of protein is present in the soluble fraction. Concluded, to use BL21 DE3 strain as the expression strain for both RTC Nsp 12^{N52A} and RTC Nsp 12$^{H, E(75,83)A, A}$ constructs. Rosetta 2 (DE3) strain is the expression strain for the RTC Nsp 12^{T51A} construct used to conduct the protein purification protocol. The post-induction temperature and durations remain the same for all three, i.e., 30 °C for 8 hours.

Protein Expression check of RdRp point mutants

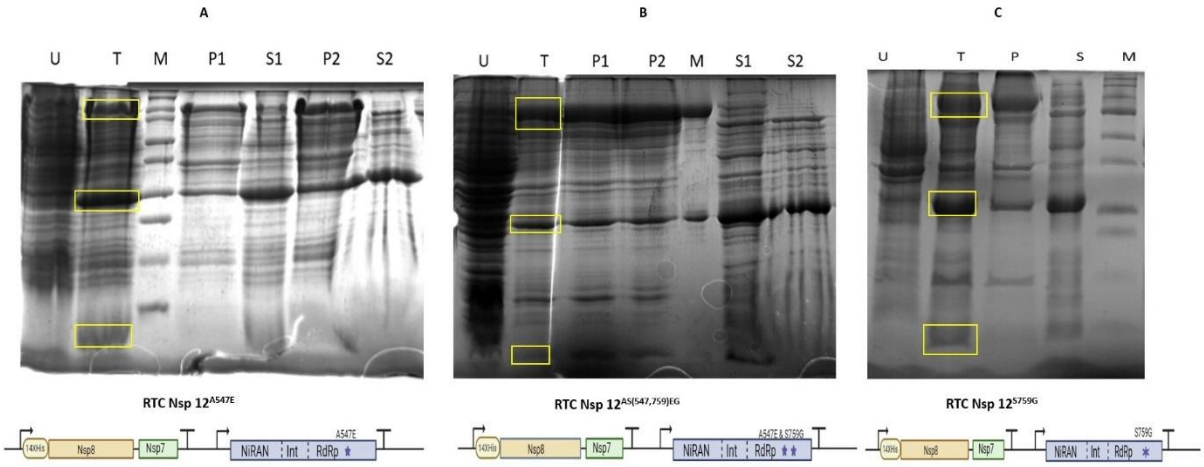


Figure 24: A) Expression check of **RTC Nsp 12^{A547E}** in Rosetta 2 (DE3) strain. The lanes U, T, P and S represent the Uninduced total, Induced total, Pellet and Soluble fractions, respectively. Here, S1 represents the lysis buffer conditions 50 mM NaHEPES (pH = 8) + 500 mM NaCl, and **S2** represents the lysis buffer conditions **50 mM NaHEPES (pH = 8) + 500 mM NaCl +1% CHAPSO**. The induction was done with 0.1 mM IPTG, and post-induction, the culture was incubated at 30 °C for 8 hours. **B)** Expression check of **RTC Nsp 12$^{A, S(547,759)E, G}$** in Rosetta 2 (DE3) strain. The lanes M, U, T, P and S represent the 100kDa protein ladder, uninduced total, induced total, pellet and soluble fractions, respectively. Here, S1 represents the lysis buffer conditions **50 mM NaHEPES (pH = 8) + 500 mM NaCl**, and **S2** represents the lysis buffer conditions **50 mM NaHEPES (pH = 8) + 500 mM NaCl +1% CHAPSO**. The induction was done with 0.1 mM IPTG, and post-induction, the culture was incubated at 30 °C for 8 hours. **C)** Expression check of **RTC Nsp 12^{S759G}** in Rosetta 2 (DE3) strain. The lanes M, U, T, P and S represent the 100 kDa protein ladder, uninduced total, induced total, pellet and soluble fractions, respectively. Here, the lysis buffer conditions used involve 50 mM NaHEPES (pH = 8) + 500 mM NaCl. The induction was done with 0.1 mM IPTG and post-induction, the culture was incubated at 30 °C for 8 hours

Rosetta 2 (DE3) strain was used as the expression strain for RTC Nsp 12^{A547E}, RTC Nsp 12^{S759G} and RTC Nsp 12$^{A, S (547,759) E, G}$ constructs to conduct the protein purification protocol. The post-induction temperature and durations remain the same for all three constructs, i.e., 30 °C for 8 hours.

Protein Purification

Protein Purification of Nsp 12 Sub-domain mutants

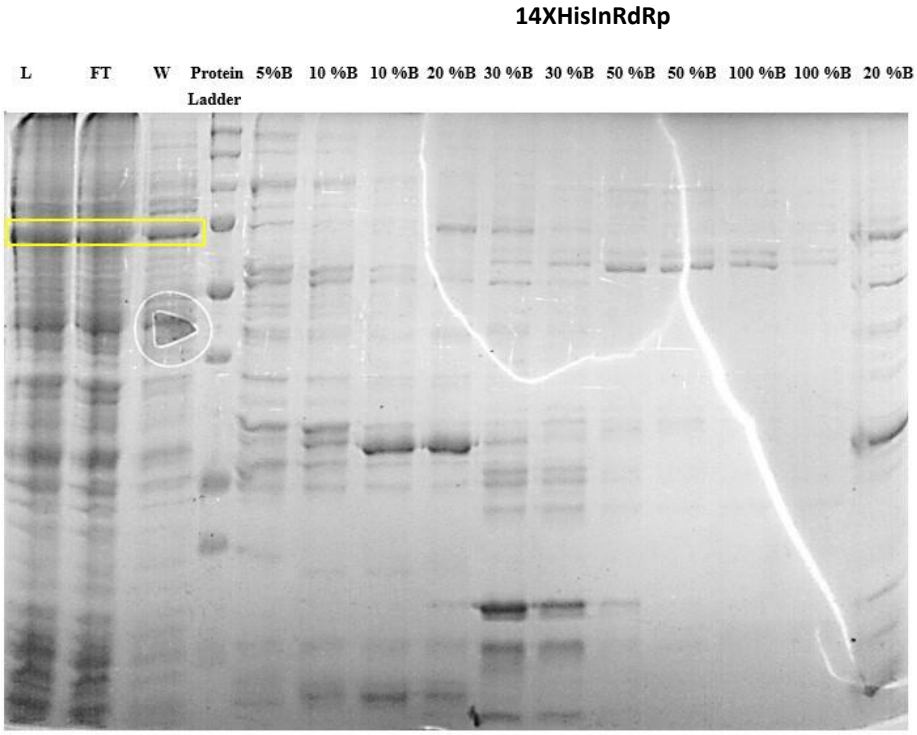


Figure 25: Ni-NTA purification gel of **14XHisInRdRp** through BL21 DE3 strain. The lane L, FT, and W represents Load, Flow through and Wash. The protein was eluted with increasing concentrations of imidazole represented by 5%B, 10%B, 20%B, 30%B, 50%B, and 100%B, where B represents the elution buffer of the NiNTA column.

The growth of culture, induction of protein of interest, pelleting down cells, sonicating the cells, spinning the sample, and loading it into the column was followed as stated in the section protein purification of the chapter Material and Methods. Here, we observe that the majority of the protein is present in the flow through and wash, which implies that the protein is not binding to the column.

To mitigate this trouble, we construct a new clone for InRdRp with a 6XHis tag in the C terminus of the gene of interest in a pHis vector backbone. The decision to use the 6XHis tag instead of the early 14XHis tag was made because of speculation that the 14XHis tag may not be folding correctly and we tried to see if a 6XHis tag at the C terminus would fix this problem. Furthermore, to remove imidazole from the equilibration buffer for all the subsequent purification of sub-domain or monomer domain mutants.

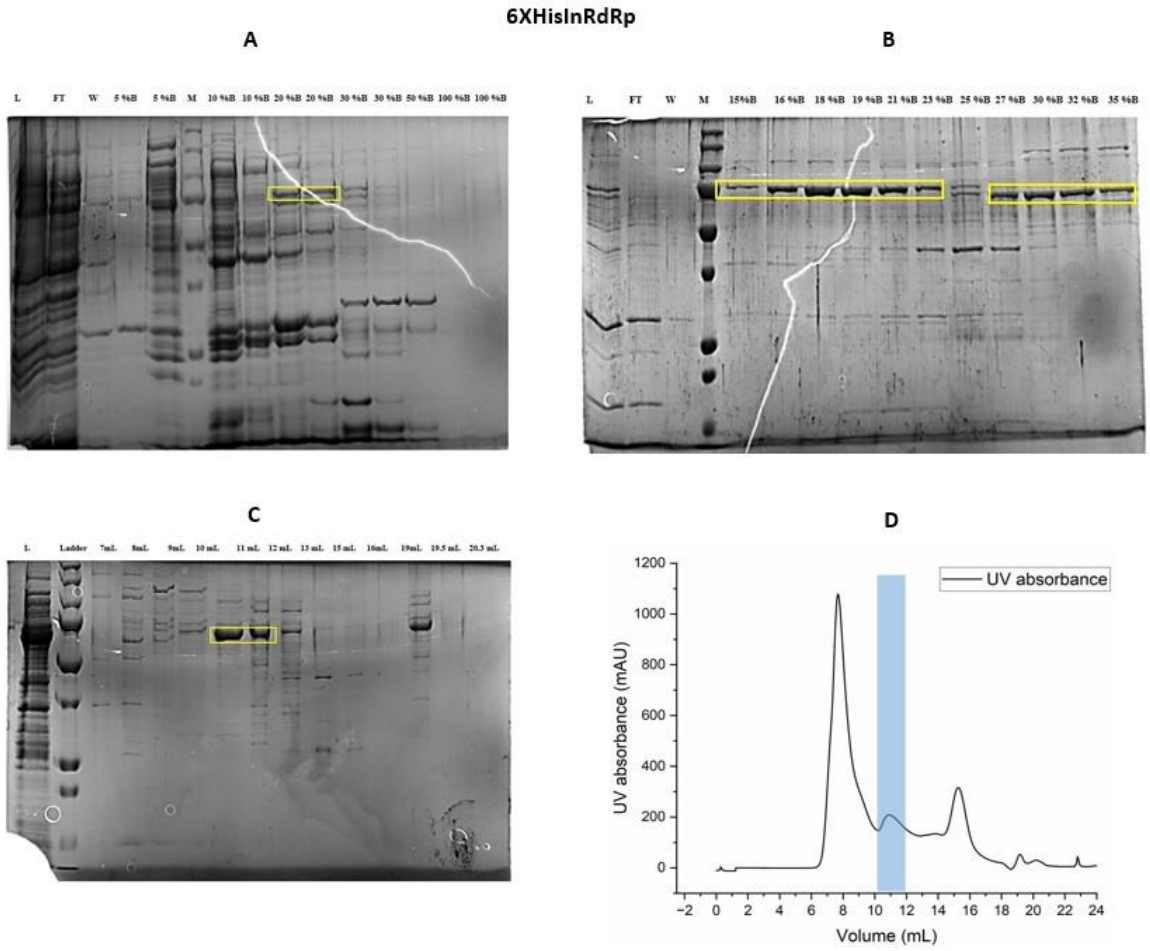


Figure 26: A) Represents the Ni NTA elution fractions of **6XHisInRdRp** from 5% B, 10% B, 20%B, 30%B, 50%B and 100 % B. Here, L, FT and W represent the load, flow through and wash. Fractions 20% and 30% were pooled, diluted to reduce the salt concentration to 100 mM NaCl and used for further downstream process. **B)** Represents the selected elution fractions of HiTrap QHP column of 6XHisInRdRp; the bounded protein was eluted using increasing concentration of NaCl. Due to confusion in band patterns, the fractions from 15%B to 23%B were concentrated and flash frozen, stored at -80 °C and used directly for assays. Selected fractions from 27%B to 35%B, pooled, concentrated, and loaded onto SEC 200. **C)** The gel represents the selected elution fractions of Size Exclusion Chromatography (SEC) 200 for 6XHisInRdRp. Fractions from 10 mL to 11 mL were pooled, concentrated using a centricon and flash frozen and stored at -80 °C. **D)** The graph plots the UV absorbance against the volume of SEC 200 for the purification of 6XHisRdRp.

Both purified samples were later used to study the extension activity, but the heavy nuclease contamination in the purified samples degraded the RNA template. We need to troubleshoot the purification and get a purer sample. I decided to try out 14XHisRdRp purification.

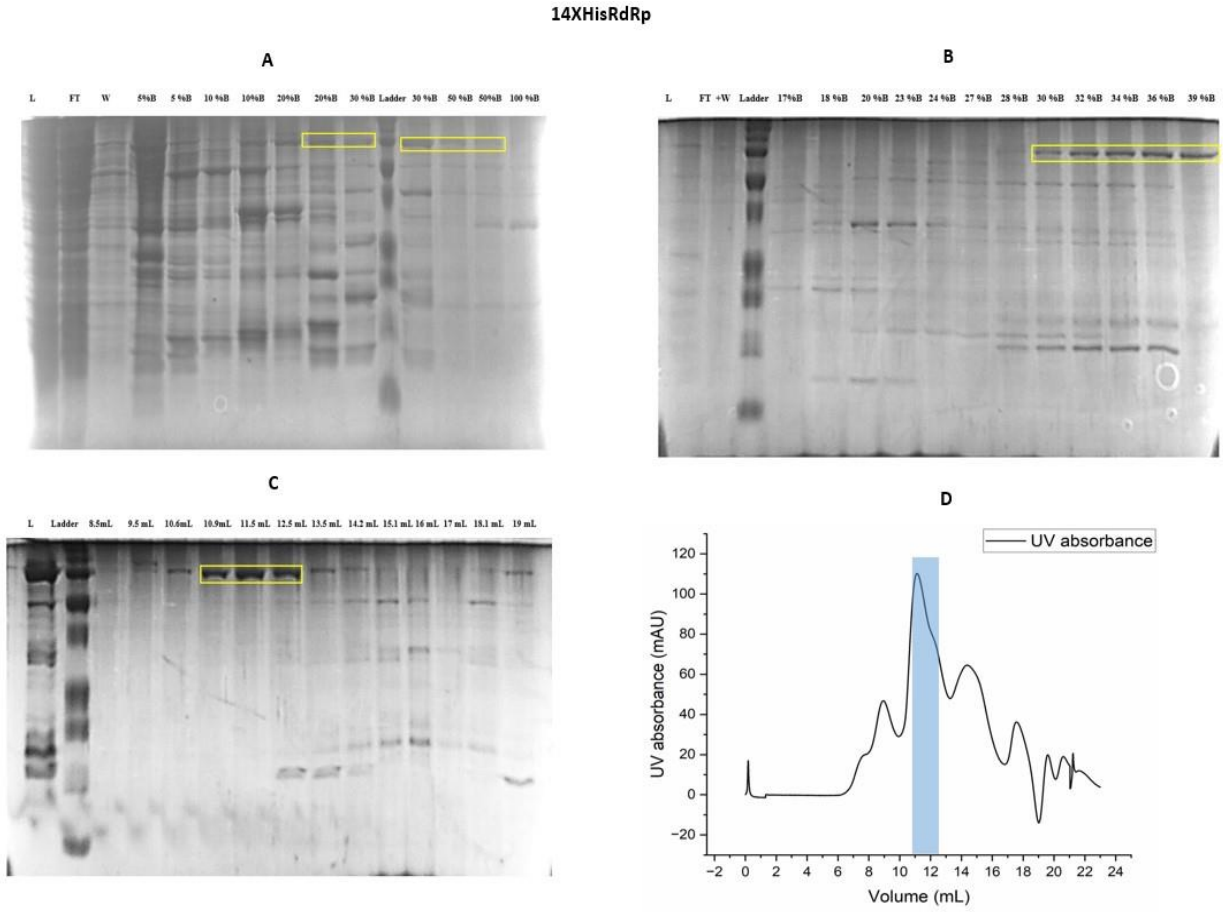


Figure 27: A) Represents the Ni NTA elution fractions of **14XHisRdRp** from 5% B, 10% B, 20%B, 30%B, 50%B and 100 % B. Here, L, FT and W represent the Load, Flow through and Wash. Fractions 20%, 30%, 30%, 50% and 50% were pooled, diluted to reduce the salt concentration to 100 mM NaCl and used for further downstream process.

B) Represents the selected elution fractions of HiTrap QHP column for 14XHisRdRp; the bounded protein was eluted using increasing concentration of NaCl. Selected fractions from 30%B to 39%B were pooled, concentrated, and loaded onto SEC 200

C) The gel represents the selected elution fractions of Size Exclusion Chromatography (SEC) 200 for 14XHisRdRp. Fractions from 10.9mL to 12.5mL were pooled, concentrated using a centricon flash frozen and stored at -80 °C.

D) The graph plots the UV absorbance against the volume of SEC 200 for the purification of 14XHisRdRp.

Here, we also observed that 14XHisRdRp was binding very well with the Ni-NTA column. The purified sample was later used to check the extension activity of the RdRp domain alone and with various accessory proteins.

Protein Purification of NiRAN point mutants

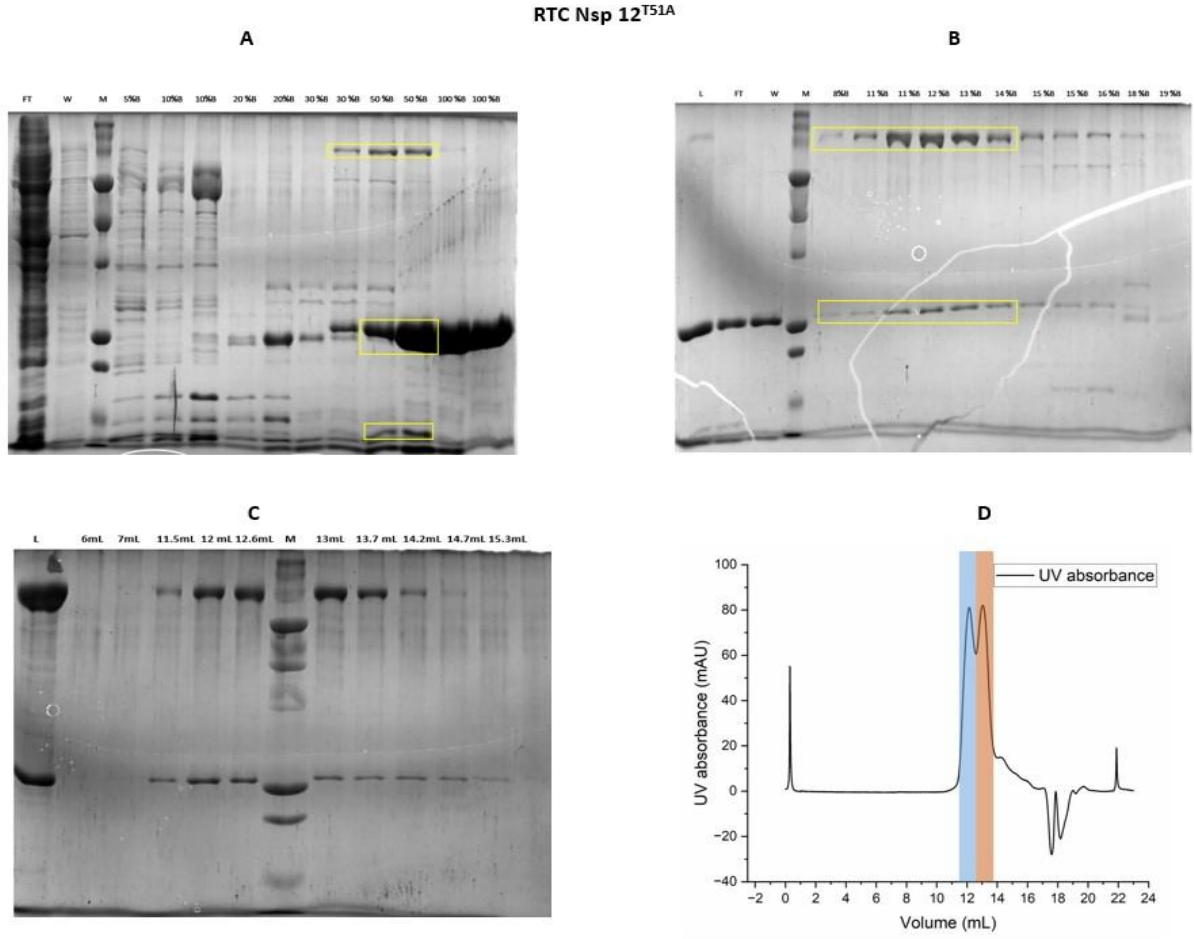


Figure 28: A) Represents the Ni NTA elution fractions of **RTC Nsp 12^{T51A}** from 5%B, 10% B, 20%B, 30%B, 50%B, 70%B and 100 % B. Here, FT and W represent the Flow through and Wash. Fractions 30% and 50% were pooled and diluted to reduce the salt concentration to 100 mM NaCl and used for further downstream process.

B) Represents the selected elution fractions of the HiTrap QHP column of RTC Nsp 12^{T51A}. The bounded protein was eluted using increasing concentrations of NaCl. Selected fractions from 8%B to 14%B pooled and concentrated the sample and loaded it onto SEC 650.

C) The gel represents the selected elution fractions of Size Exclusion Chromatography (SEC) 650 for RTC Nsp 12^{T51A}. Observed two peaks of nearly equal magnitude decide to pool fraction from 11.5 mL to 12.6 mL (peak 1 in the blue shade) and 13 mL to 14.2 mL (peak 2 in the red shade) separately and concentrate both the peaks separately using a centricon and flash frozen and stored at -80 °C.

D) The graph plots the UV absorbance against the volume of SEC 650 for the purification of RTC Nsp 12^{T51A}.

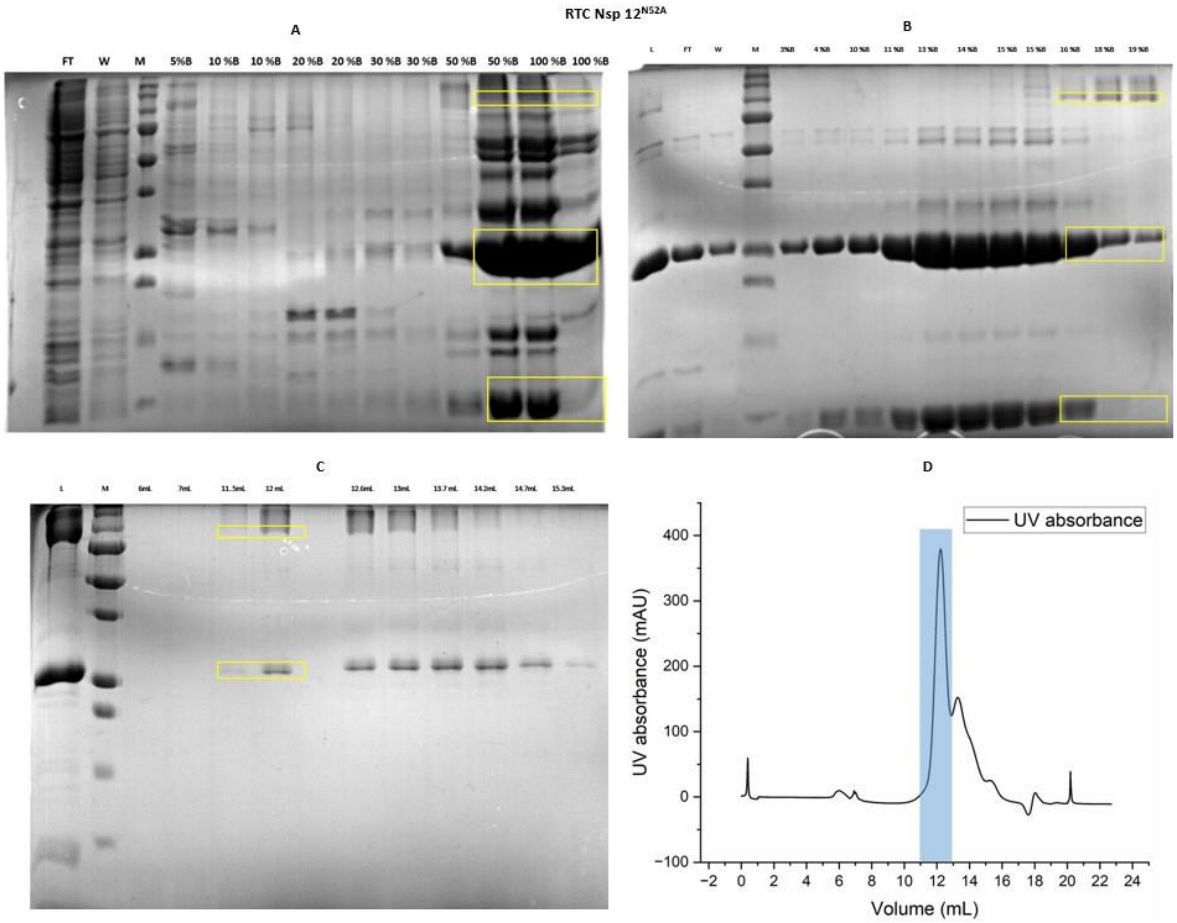


Figure 29: **A)** Represents the Ni NTA elution fractions of **RTC Nsp 12^{N52A}** from 5%B, 10% B, 20%B, 30%B, 50%B and 100 % B where B represents the elution buffer of NiNTA column. Here, FT, W represents the Flow through and Wash. Fractions 50% and 100% were pooled and diluted to reduce the salt concentration to 100 mM NaCl and used for further downstream process.

B) The gel represents the selected elution fractions of the HiTrap QHP column of RTC Nsp 12N52A. The bounded protein was eluted using increasing concentrations of NaCl. Selected fractions from 16%B to 19%B pooled, concentrated the sample, and loaded it onto SEC 650.

C) The gel represents the selected elution fractions of Size Exclusion Chromatography (SEC) 650 for RTC Nsp 12^{N52A}. Fractions from 11.5 mL to 13.7 mL were pooled, concentrated using a centricon and flash freeze and stored at -80 °C.

D) The graph plots the UV absorbance against the volume of SEC 650 for the purification of RTC Nsp 12^{N52A}.

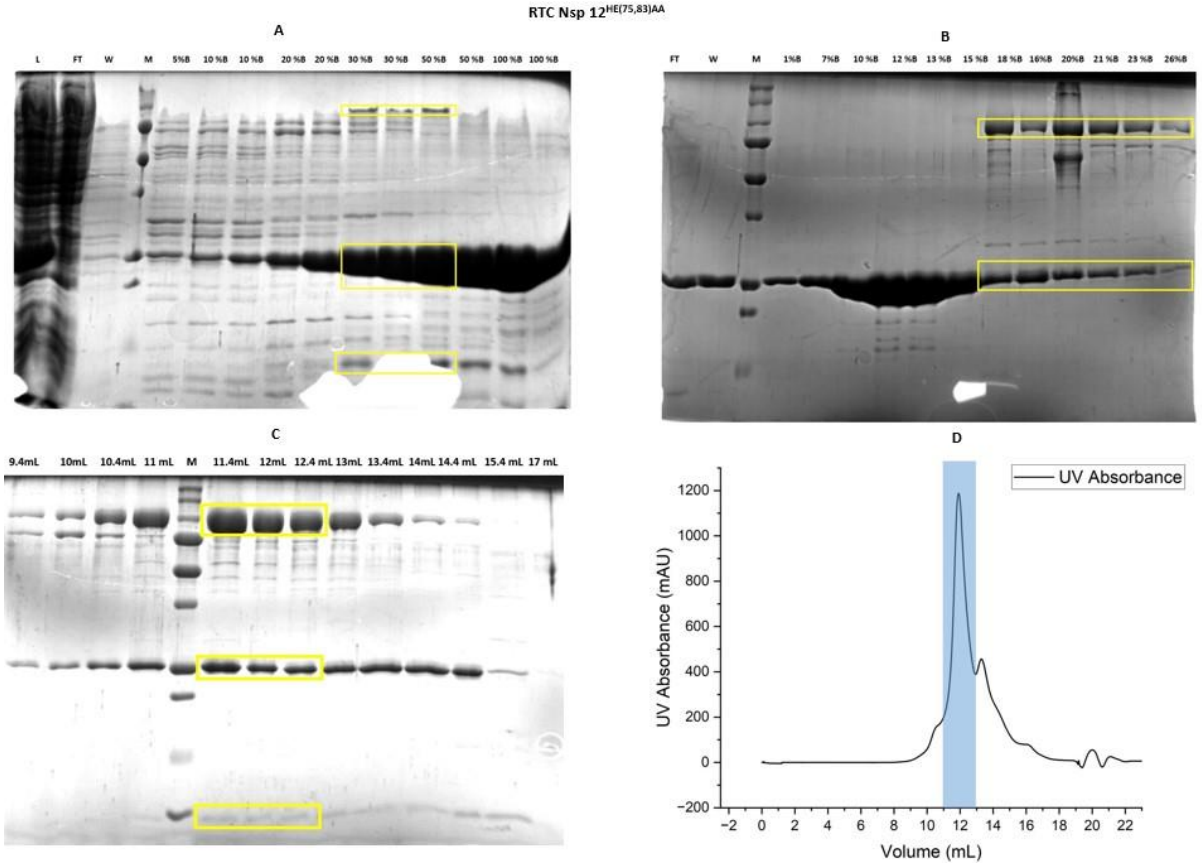


Figure 30: A) Represents the Ni NTA elution fractions of **RTC Nsp 12$^{H, E (75,83) A, A}$** from 5% B, 10% B, 20%B, 30%B, 50%B and 100 % B where B represents the elution buffer of NiNTA. Here, L, FT and W represent the elution buffer of Ni-NTA, Load, Flow through and Wash. Fractions 20 %, 30%, 30%,50 %and 50% were pooled diluted to reduce the salt concentration to 100 mM NaCl and used for further downstream process.

B) Represents the selected elution fractions of the HiTrap QHP column for RTC Nsp 12$^{H, E (75,83) A, A}$. The bounded protein was eluted using an increasing concentration of NaCl. Selected fractions from 18%B to 26%B were pooled, concentrated and loaded onto SEC 200

C) The gel represents the selected elution fractions of Size Exclusion Chromatography (SEC) 200 for RTC Nsp 12$^{H, E(75,83)A, A}$. Fractions from 11.4 mL to 12.4 mL were pooled, concentrated using a centricon and flash frozen and stored at -80 °C.

D) The graph plots the UV absorbance against the volume of SEC 200 for the purification of RTC Nsp 12$^{H, E (75,83) A, A}$.

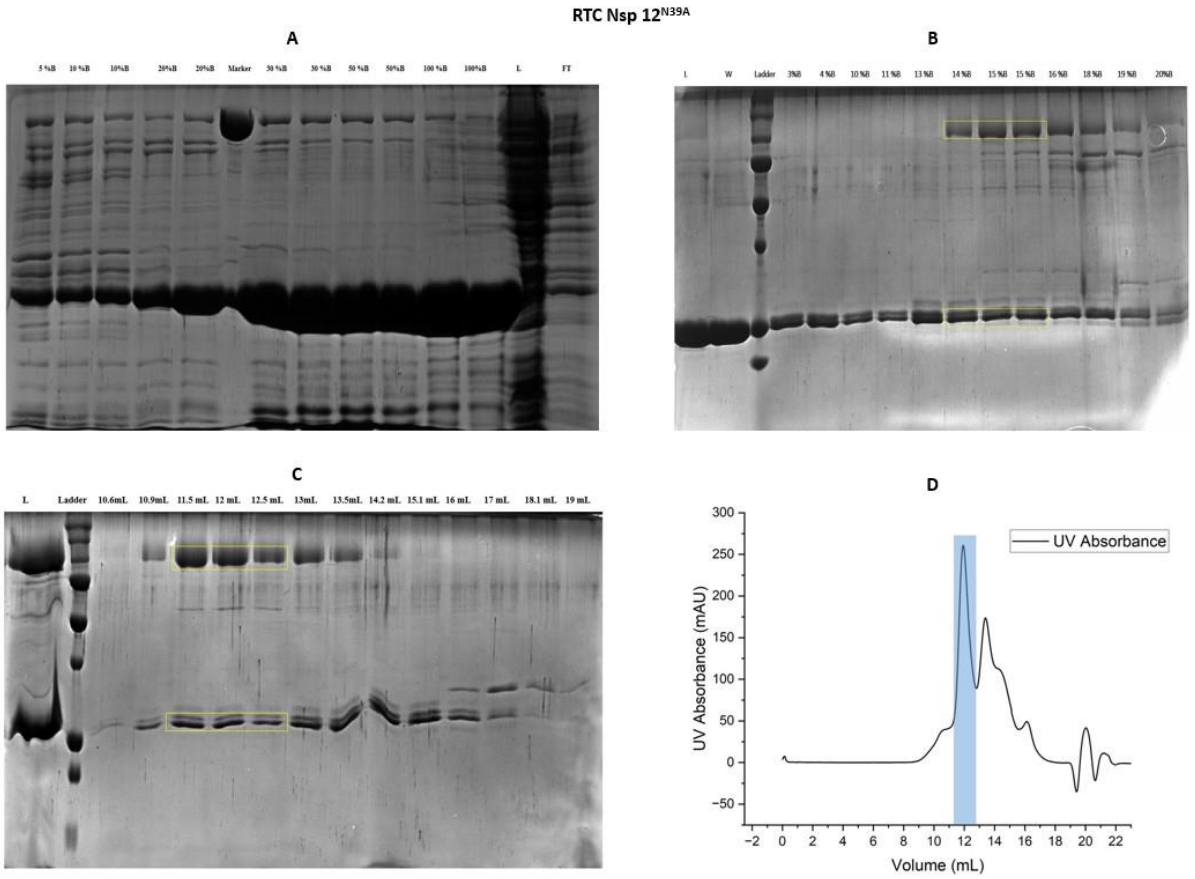


Figure 31: A) Represents the Ni NTA elution fractions of **RTC Nsp 12^{N39A}** from 5% B, 10% B, 20%B, 30%B, 50%B and 100 % B where B represents the elution buffer of NiNTA. Here, L, FT and W represent the Load, Flow through and Wash. Fractions 30%, 30%, 50 %and 50% were pooled diluted to reduce the salt concentration to 100 mM NaCl and used for further downstream process.

B) Represents the selected elution fractions of the HiTrap QHP column for RTC Nsp 12^{N39A}. The bounded protein was eluted using increasing concentrations of NaCl. Selected fractions from 14%B to 15%B were pooled, concentrated, and loaded onto SEC 200

C) The gel represents the selected elution fractions of Size Exclusion Chromatography (SEC) 200 for RTC Nsp 12^{N39A}. Fractions from 11.5mL to 12.5mL were pooled, concentrated using a centricon flash frozen and stored at -80 °C.

D) The graph plots the UV absorbance against the volume of SEC 200 for the purification of RTC Nsp 12^{N39A}.

All four NiRAN point mutants were purified and later used for NMPylation and Extension assays to study how each mutation affects the same.

Extension Assay

Extension Assay of Nsp 12 Sub-domain mutants

To check the activity 14XHisRdRp mutant. Extension assay was performed as described in section extension assay chapter materials and methods. Here, the purified mutant was incubated with NiRAN-Int (purified by Mohammad Navas) in the presence and absence of NTPs to check for a complete extension.

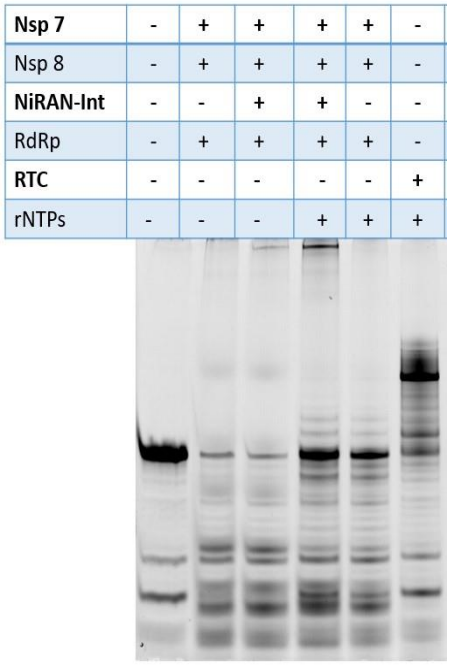


Figure 32: The gel represents the extension assay carried out by the **14XHisRdRp** domain and RTC-WT.

Here, we observe a feeble extension of RNA primer-template complex by RdRp domain + Nsp8 + Nsp 7 only in the presence of NTPs—and maximum extension by RTC-WT (positive control). Unfortunately, we don't observe even minimal extension when adding the NiRAN-Int domain. We also observe many degradation bands when the purified domain is added to the reaction mixture, most possibly caused by RNAase contamination. We plan to mitigate this by designing another construct of RdRp containing Nsp 7 and Nsp 8 in a single operon, essentially copurifying it as complex (RTC RdRp).

Next, I carried out assays to investigate how the activity of the RdRp domain is affected when the accessory proteins are varied. For this, the purified RdRp domain mutant was incubated with RNA primer-template duplex alone with Nsp 7 and Nsp 8 separately in the presence of

NTPs. The RdRp domain alone was incubated with RNA primer-template duplex in the presence of NTPs without any accessory factors.

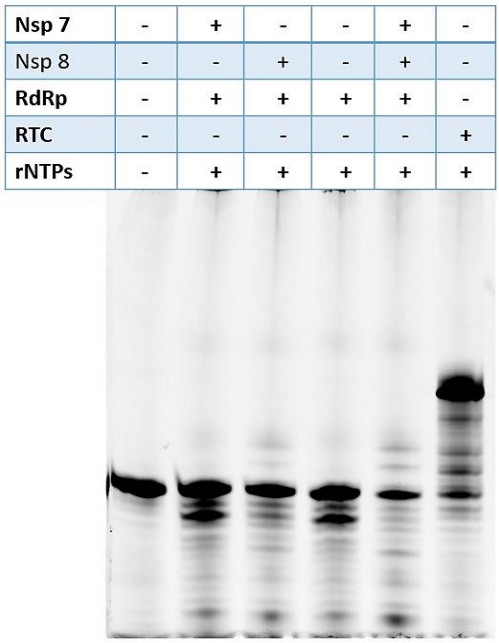


Figure 33: The gel shows the extension assay of **14XHisRdRp** carried out by varying ancillary proteins.

Observe that the RdRp domain requires at least Nsp 8 to show the basic feeble extension of the RNA primer-template complex. And can't extend the RNA primer-template duplex by itself.

Extension Assay of NiRAN mutants

To check if point mutation in NiRAN affects the RdRp activity, we decided to check the extension activity of RTC Nsp 12^{T51A} peaks 1 and 2.

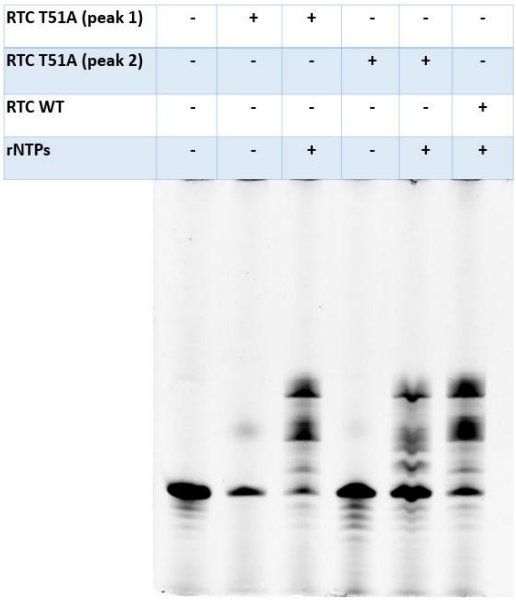


Figure 34: The gel represents the extension activity of two **RTC Nsp 12^{T51A}** peaks.

While we observed a weaker extension activity by peak 2 fraction, peak 1 fraction extended RNA duplex similar to RTC WT. Based on previous experience and studies, we came to the

conclusion that peak 2 has proteins in the wrong stoichiometry. Whereas peak 1 has it in the correct stoichiometry.

To check if point mutation and other NiRAN mutants also have any effect on the RdRp activity, we decided to check the extension activity of RTC Nsp 12^{N52A}.

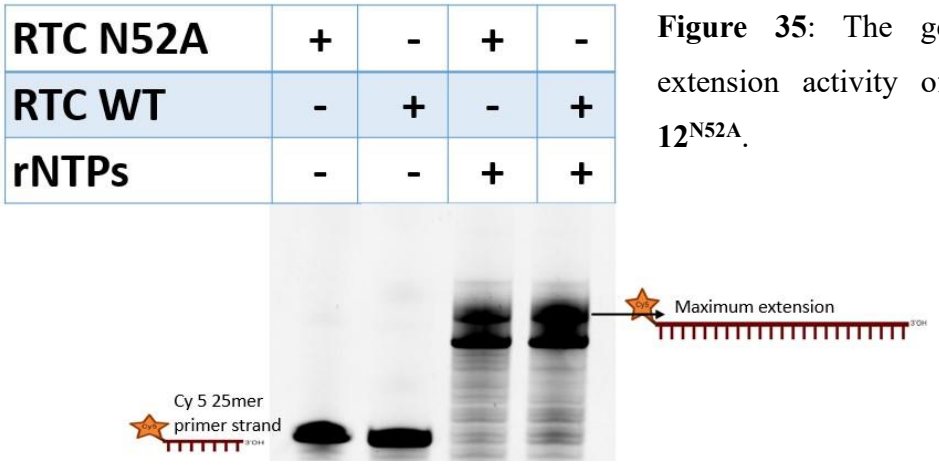


Figure 35: The gel represents the extension activity of two **RTC Nsp 12^{N52A}**.

N52A does not affect the extension activity and is comparable to RTC WT.

NMPylation And deNMPylation Assay

NMPylation and deNMPylation Assay of NiRAN mutants

T51 is a residue in the vicinity of the nucleotide in the Base-Out pose. To check if T51 interacts with the NTP in the Base-Out pose, I performed an NMPylation Assay of RTC Nsp 12^{T51A} mutant.

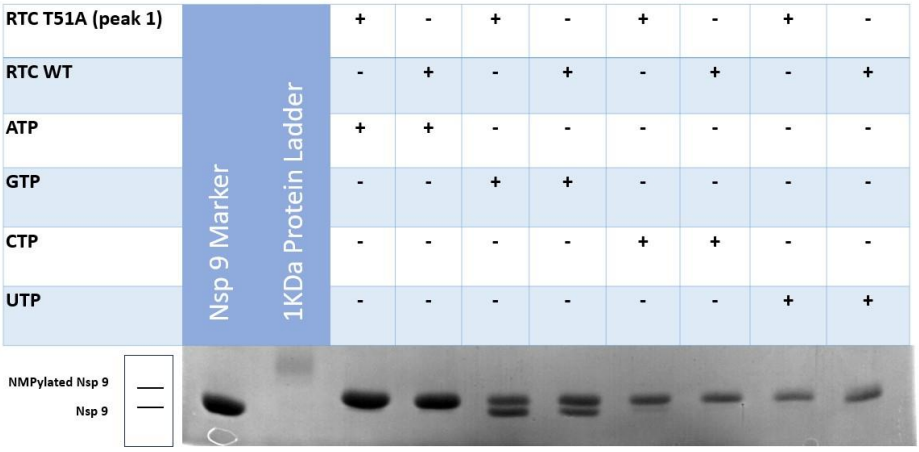


Figure 36: The gel represents the NMPylation assay for the **RTC Nsp 12^{T51A}** mutant.

Observe that the NMPylation activity of the T51A mutant is similar to RTC WT for all four NTPs. Hence, T51 does not interact with a nucleotide in the Base-Out pose.

Another residue in the vicinity of nucleotide in the Base-Out pose is N52. To check if N52 interacts with the NTP in the Base-Out pose, I performed an NMPylation Assay of RTC Nsp 12^{N52A} mutant.

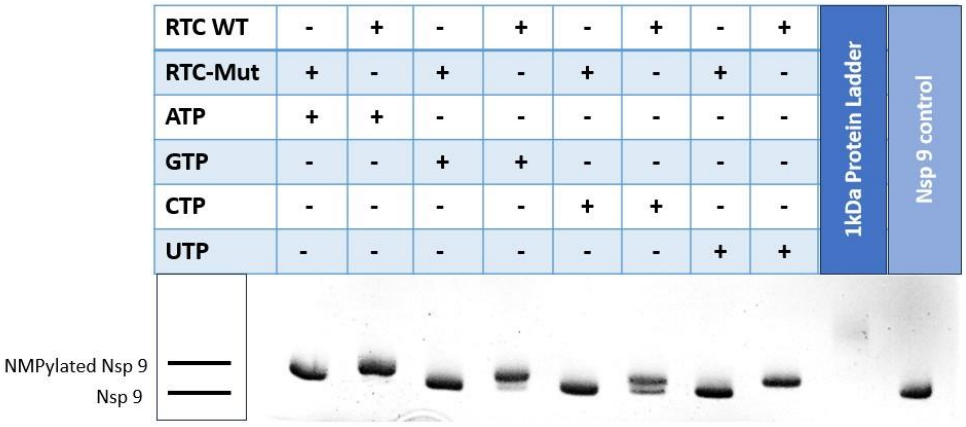


Figure 37: The gel represents the NMPylation assay for the **RTC Nsp 12^{N52A}** mutant.

We observe that the N52A mutant is dead for NMPylation activity compared to RTC WT for all four NTPs. This could mean that **Base-Out is the most significant pose for NMPylation reaction**.

A recent model published by (Small et al., 2023) attributes the Base-Up pose of nucleotide as the pose responsible for the NMPylation reaction. The shortcoming of the study was that the authors had not conducted mutagenesis experiments to validate the model. We decided to test this model by mutating H75A and E83A of the NiRAN pocket- which is responsible for the coordination bond with Mg^{2+} in the recently proposed model (Refer to diagram).

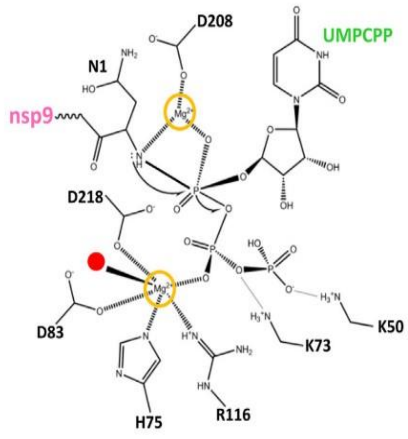


Figure 38: The recently proposed alternative model for the NMPylation reaction (Small et al., 2023).

To test this, we generated RTC Nsp $12^{H,\ E(75,83)A,\ A}$ and conducted NMPylation and deNMPylation Assay of RTC Nsp $12^{H,\ E(75,83)A,\ A}$

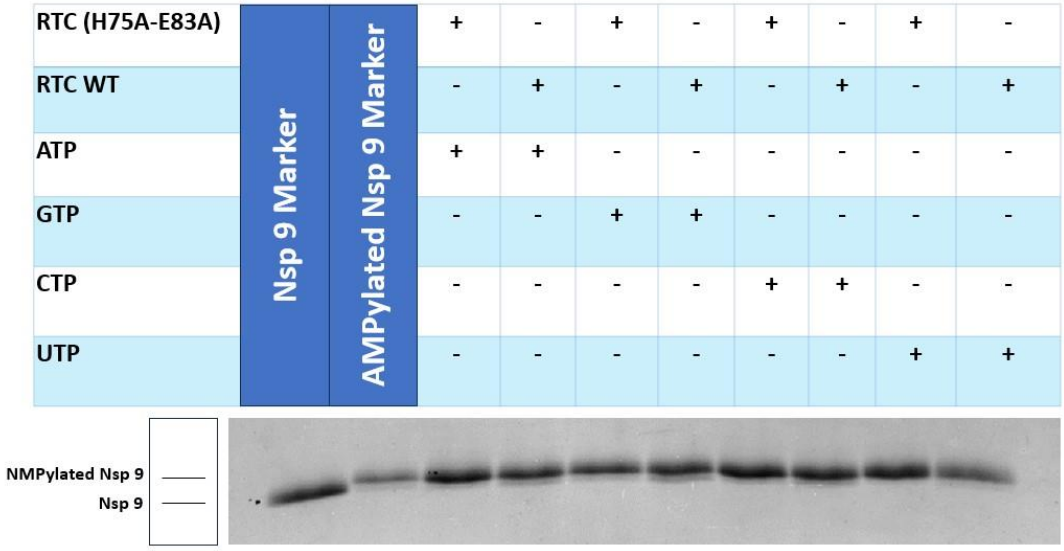


Figure 39: The gel represents the NMPylation assay for **RTC Nsp 12$^{H, E (75,83) A, A}$** mutant.

Based on the model, we expect that when H75 and E83 are mutated to alanine, the coordination with Mg^{2+} is lost, and the NMPylation reaction is affected. However, the mutant has NMPylates Nsp 9 using all four nucleotides, similar to RTC-WT. The mutant also shows deNMPylation activity similar to RTC WT. This could mean that Mg^{2+} is NOT required for the NMPylation reaction.

To further check the Base-Up model for NMPylation, We decided to mutate N39, a residue in the Up pose that interacts with the 4-oxo group of nucleotides in the Up pose.

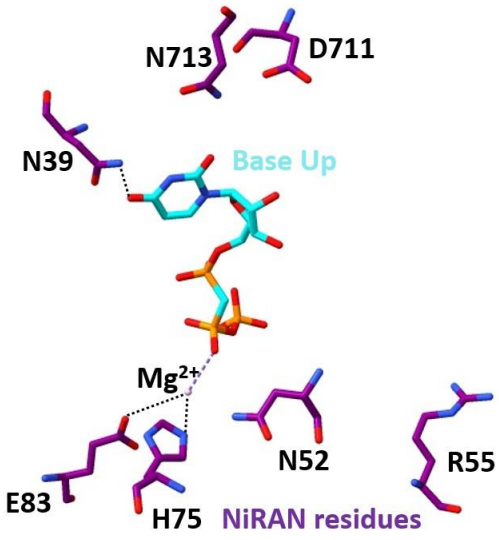


Figure 40: The recently proposed alternative model for the NMPylation reaction (Small et al., 2023)

Generated RTC Nsp 12^{N39A} and conducted NMPylation and deNMPylation Assay using the same.

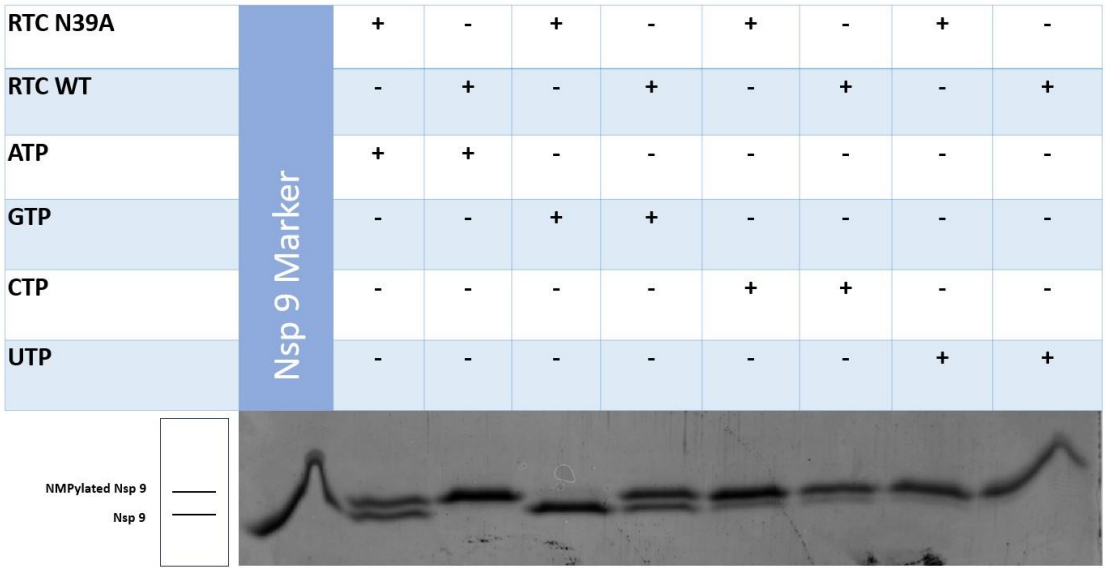


Figure 41: The gel represents the NMPylation assay for the **RTC Nsp 12^{N39A}** mutant.

Observe that the CMPylation and UMPylation activity of the N39A mutant is similar to RTC WT. Surprisingly, there was a 50 % reduction in AMPylation activity and no GMPylation activity for N39A mutant compared to RTC-WT, which is contrary to what the Bae-Up model proposes.

RTC Nsp 12$^{H, E (75,83) A, A}$ and RTC Nsp 12^{N39A} NMPylation assays tell us that the Base-Up pose of nucleotide is NOT the significant pose for NMPylation. NMPylation using NTPs proceeds through the Base-Out pose similar to the bacterial homolog of NiRAN - SelO.

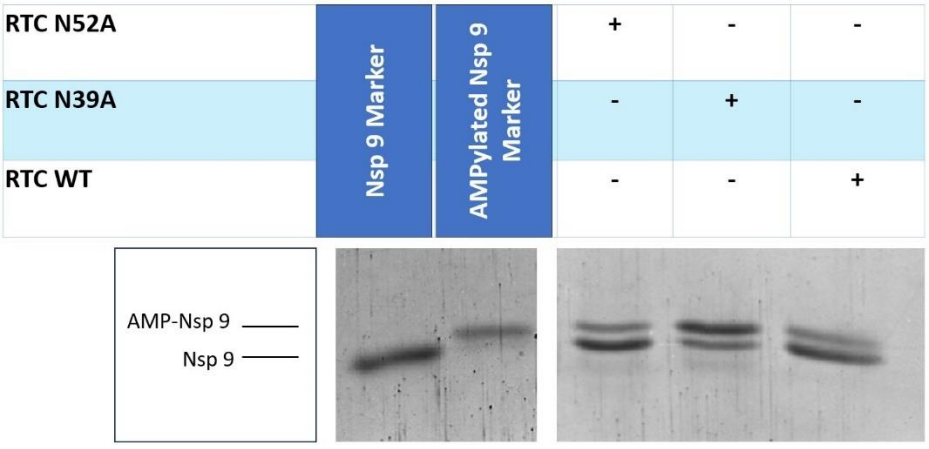


Figure 42: The gel represents the deNMPylation assay carried out for RTC Nsp 12^{N52A}, RTC Nsp 12^{N39A} mutant.

The deNMPylation assay with RTC Nsp 12^{N52A} and RTC Nsp 12^{N39A} mutants show that the deNMPylation reaction, which proceeds through the Base-In pose, is affected by N39 - a

residue in the Up pose, while NOT affected by N52 - a residue in the Out pose. This observation led to the proposal that **AMPylated Nsp 9 binds the NiRAN pocket in the Base-Up pose.**

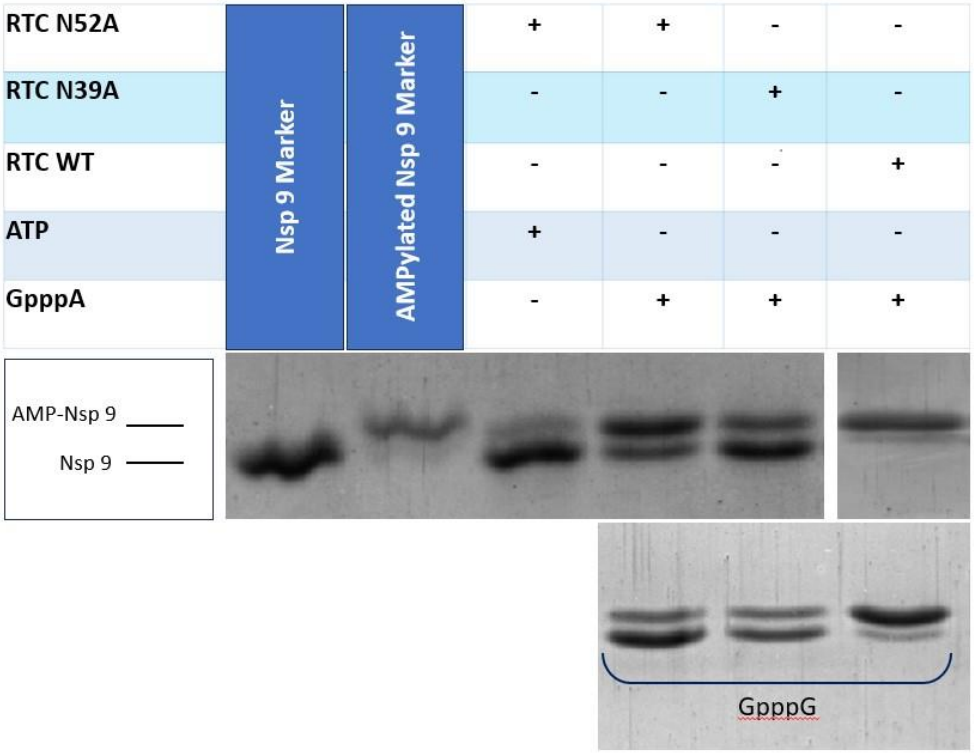


Figure 43: The gel represents the NMPylation assay using GpppA and GpppG.

Observe that the N52A, which is inactive for NMPylation of Nsp 9 using NTPs as substrate, uses both GpppA and GpppG to NMPylate Nsp 9. This, along with RTC Nsp 12^{N713A} and RTC Nsp 12^{D711A} mutants (shown by Navas in the lab), show an effect on NMPylation reaction using GpppA and GpppG while other activities are similar to RTC-WT. This observation led to the proposal that **GpppN binds in the NiRAN pocket in the Base-Up pose.**

Chapter 04 Discussion

Chapter 04: Discussion

During the course of this project, I successfully conducted the cloning and recombinant protein optimization of all the constructs that were planned. The purifications of most proteins were finished. Purified proteins were used to perform three main assays, extension, NMPylation, and deNMPylation, to delineate the biochemical intricacies of SARS-CoV-2 RTC.

The project focused on gaining more fundamental insights into NiRAN activity and its effect on the RdRp activity of SARS-CoV-2 RTC. Previous studies had suggested that the capping of nascent viral mRNA (as discussed in the introduction chapter) proceeds through the cascade of viral enzymatic reactions performed by various non-structural proteins. NiRAN domain of Nsp 12 plays a crucial role in forming core cap structure (GpppA). The NiRAN catalyses different nucleotidyltransferase reactions by different poses of nucleotides within the NiRAN pocket. The earlier studies proposed that the NMPylation reaction proceeds through the Base-Out pose while the deNMPylation reaction proceeds through the Base-In pose. A recent study by (Small et al., 2023) proposed a model for the cap formation in which the nucleotide occupies the Base-Up pose for the NMPylation reaction. The shortcoming of this study is that the authors have not conducted a mutagenesis study to validate their model. Through our structural analysis and mutagenesis study, we illustrated that the RTC Nsp $12^{H,\ E\ (75,83)\ A,\ A}$ and RTC Nsp 12^{N39A} show NMPylation activity contrary to what the Base-Up model for NMPylation proposes. RTC Nsp 12^{N52A} mutant did not show any NMPylation activity for all four nucleotides, which implies that the Base-Out pose is the most significant pose for the NMPylation reaction. Interestingly, N39 - a residue that interacts with the Up pose, affected the deNMPylation reaction; this observation led us to propose a model on how AMPylated Nsp 9 binds the NiRAN pocket- in a Base-Up fashion (as shown in the diagram below).

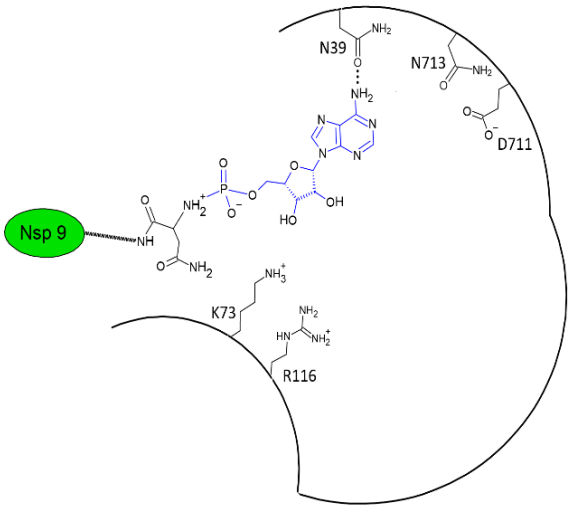


Figure 44: The proposed model for AMPylated Nsp 9 binding to the NiRAN pocket

The RTC Nsp 12^{N52A} is dead for NMPylation reaction with all four NTPs but exhibits activity using GpppN (where N = A and G) as substrate. Along with the mutagenesis study of RTC Nsp 12^{N713A} and RTC Nsp 12^{D711A} mutants conducted by Navas showed an effect on NMPylation using GpppN as substrate. This combined observation led to the proposal of a model on how GpppA binds the NiRAN pocket- in a Base-Up fashion (as shown in the diagram).

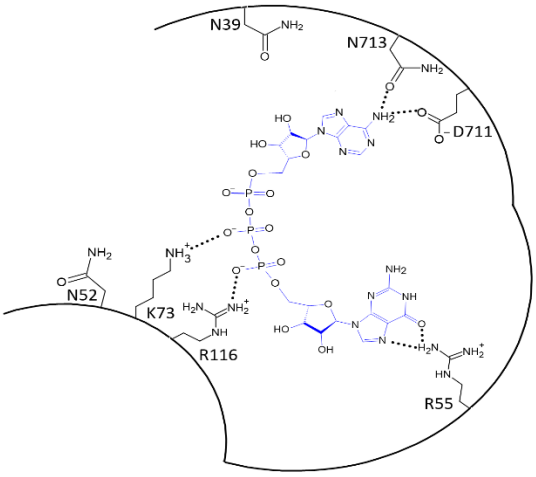


Figure 45: The proposed model for GpppA binding to the NiRAN pocket

The other question we asked was- Does the NiRAN domain affect the RdRp activity of Nsp 12? We observe that RTC Nsp 12^{N52A}, which is NMPylation inactive, does not affect the RdRp activity of Nsp 12. To investigate if the NiRAN domain of Nsp 12 affects RdRp activity, we deleted the NiRAN-Interface domain and generated the construct 14XHisRdRp. When the extension assay was performed using this construct, we saw only feeble activity, and the RdRp domain required at least Nsp 8 as an accessory factor to exhibit this minimal extension of RNA primer-template duplex. The major problem with the 14XHisRdRp was that the purified sample contained nuclease contaminations, so we are planning to either copurify the RdRp domain along with Nsp 7 and Nsp 8 or mix Nsp 7 and Nsp 8 during size exclusion chromatography. A nuclease-free RdRp domain can be used for biochemical characterization in the future.

The fidelity and rate mutants of SARS-CoV-2 RTC have to be purified hereafter to explore the difference in their nucleotide misincorporation patterns. This could possibly shed more light on the fundamental details and evolution of polymerase in large RNA viruses.

References

Arya, R., S. Kumari, B. Pandey, H. Mistry, S.C. Bihani, A. Das, V. Prashar, G.D. Gupta, L. Panicker, and M. Kumar. 2021. Structural insights into SARS-CoV-2 proteins. *J. Mol. Biol.* 433:166725. doi:10.1016/j.jmb.2020.11.024.

Den Boon, J.A., A. Diaz, and P. Ahlquist. 2010. Cytoplasmic viral replication complexes. *Cell Host Microbe*. 8:77–85. doi:10.1016/j.chom.2010.06.010.

Brant, A.C., W. Tian, V. Majerciak, W. Yang, and Z.M. Zheng. 2021. SARS-CoV-2: from its discovery to genome structure, transcription, and replication. *Cell Biosci.* 11. doi:10.1186/S13578-021-00643-Z.

Campagnola, G., V. Govindarajan, A. Pelletier, B. Canard, and O.B. Peersen. 2022. The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J. Virol.* 96. doi:10.1128/jvi.00671-22.

Cascella, M., M. Rajnik, A. Cuomo, S.C. Dulebohn, and R. Di Napoli. 2022. Features, Evaluation, and Treatment of Coronavirus (COVID-19). *StatPearls*.

Finkel, Y., O. Mizrahi, A. Nachshon, S. Weingarten-Gabbay, D. Morgenstern, Y. Yahalom-Ronen, H. Tamir, H. Achdout, D. Stein, O. Israeli, A. Beth-Din, S. Melamed, S. Weiss, T. Israely, N. Paran, M. Schwartz, and N. Stern-Ginossar. 2020. The coding capacity of SARS-CoV-2. *Nat. 2020 5897840*. 589:125–130. doi:10.1038/s41586-020-2739-1.

Gago, S., S.F. Elena, R. Flores, and R. Sanjuán. 2009. Extremely high mutation rate of a hammerhead viroid. *Science (80-.).* 323:1308. doi:10.1126/SCIENCE.1169202/SUPPL_FILE/GAGO.SOM.PDF.

Gao, Y., L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L.W. Guddat, Q. Wang, Z. Lou, and Z. Rao. 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science (80-.).* 368:779–782. doi:10.1126/science.abb7498.

Gorbalenya, A.E., S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, B.L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman, D. Penzar, S. Perlman, L.L. M Poon, D. V Samborskiy, I.A. Sidorov, I. Sola, and J. Ziebuhr. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. doi:10.1038/s41564-020-0695-z.

Gorbalenya, A.E., L. Enjuanes, J. Ziebuhr, and E.J. Snijder. 2006. Nidovirales: Evolving the largest RNA virus genome. *Virus Res.* 117:17–37. doi:10.1016/j.virusres.2006.01.017.

Holmes, E.C. 2011. What Does Virus Evolution Tell Us about Virus Origins? *J. Virol.* 85:5247–5251. doi:10.1128/JVI.02203-10.

Klein, S., M. Cortese, S.L. Winter, M. Wachsmuth-Melm, C.J. Neufeldt, B. Cerikan, M.L. Stanifer, S. Boulant, R. Bartenschlager, and P. Chlanda. 2020. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat. Commun. 2020 111*. 11:1–10. doi:10.1038/s41467-020-19619-7.

Lauber, C., J.J. Goeman, M.C. de Parquet, P. Thi Nga, E.J. Snijder, K. Morita, and A.E. Gorbalenya. 2013. The Footprint of Genome Architecture in the Largest Genome Expansion in RNA Viruses. *PLoS Pathog.* 9. doi:10.1371/journal.ppat.1003500.

Lehmann, K.C., A. Gulyaeva, J.C. Zevenhoven-Dobbe, G.M.C. Janssen, M. Ruben, H.S. Overkleeft, P.A. Van Veelen, D. V. Samborskiy, A.A. Kravchenko, A.M. Leontovich, I.A. Sidorov, E.J. Snijder, C.C. Posthuma, and A.E. Gorbalenya. 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 Res.* 43:8416–8434. doi:10.1093/nar/gkv838.

Madru, C., A.D. Tekpinar, S. Rosario, D. Czernecki, S. Brûlé, L. Sauguet, and M. Delarue. 2021. Fast and efficient purification of SARS-CoV-2 RNA dependent RNA polymerase complex expressed in Escherichia coli. *PLoS One*. 16:e0250610. doi:10.1371/JOURNAL.PONE.0250610.

Malone, B., N. Urakova, E.J. Snijder, and E.A. Campbell. 2022. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat. Rev. Mol. Cell Biol.* 23:21–39. doi:10.1038/s41580-021-00432-z.

Malone, B.F., J.K. Perry, P.D.B. Olinares, H.W. Lee, J. Chen, T.C. Appleby, J.Y. Feng, J.P. Bilello, H. Ng, J. Sotiris, M. Ebrahim, E.Y.D. Chua, J.H. Mendez, E.T. Eng, R. Landick, M. Götte, B.T. Chait, E.A. Campbell, and S.A. Darst. 2023. Structural basis for substrate selection by the SARS-CoV-2 replicase. *Nature*. 614:781–787. doi:10.1038/s41586-022-05664-3.

McIntosh, K., J.H. Dees, W.B. Becker, A.Z. Kapikian, and R.M. Chanock. 1967. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci. U. S. A.* 57:933–940. doi:10.1073/PNAS.57.4.933/ASSET/23D913F2-9185-49A3-BDB5-3AC6AE9ACC70/ASSETS/PNAS.57.4.933.FP.PNG.

Miao, Z., A. Tidu, G. Eriani, and F. Martin. 2021. Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biol.* 18:447–456. doi:10.1080/15476286.2020.1814556.

Nga, P.T., M.C. de Parquet, C. Lauber, M. Parida, T. Nabeshima, F. Yu, N.T. Thuy, S. Inoue, T. Ito, K. Okamoto, A. Ichinose, E.J. Snijder, K. Morita, and A.E. Gorbalenya. 2011. Discovery of the First Insect Nidovirus, a Missing Evolutionary Link in the Emergence of the Largest RNA Virus Genomes. *PLOS Pathog.* 7:e1002215. doi:10.1371/JOURNAL.PPAT.1002215.

Park, G.J., A. Osinski, G. Hernandez, J.L. Eitson, A. Majumdar, M. Tonelli, K. Henzler-Wildman, K. Pawłowski, Z. Chen, Y. Li, J.W. Schoggins, and V.S. Tagliabracci. 2022. The mechanism of RNA capping by SARS-CoV-2. *Nature*. 609:793–800. doi:10.1038/s41586-022-05185-z.

Paules, C.I., H.D. Marston, and A.S. Fauci. 2020. Coronavirus Infections—More Than Just the Common Cold. *JAMA*. 323:707–708. doi:10.1001/JAMA.2020.0757.

Pizzato, M., C. Baraldi, G. Boscato Sopetto, D. Finozzi, C. Gentile, M.D. Gentile, R. Marconi, D. Paladino, A. Raoss, I. Riedmiller, H. Ur Rehman, A. Santini, V. Succetti, and L. Volpini. 2022. SARS-CoV-2 and the Host Cell: A Tale of Interactions. *Front. Virol.* 1:1–29. doi:10.3389/fviro.2021.815388.

Rangan, R., I.N. Zheludev, R.J. Hagey, E.A. Pham, H.K. Wayment-Steele, J.S. Glenn, and R. Das. 2020. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: A first look. *RNA*. 26:937–959. doi:10.1261/RNA.076141.120/-/DC1.

Saberi, A., A.A. Gulyaeva, J.L. Brubacher, P.A. Newmark, and A.E. Gorbalenya. 2018. A planarian nidovirus expands the limits of RNA genome size. 14. 1–41 pp.

Schubert, K., E.D. Karousis, A. Jomaa, A. Scaiola, B. Echeverria, L.-A. Gurzeler, M. Leibundgut, V. Thiel, O. Mamp, and N. Ban. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. doi:10.1038/s41594-020-0511-8.

Shang, J., Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach, and F. Li. 2020. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* 117. doi:10.1073/pnas.2003138117.

Shirato, K., M. Kawase, and S. Matsuyama. 2018. Wild-type human coronaviruses prefer cell-surface TMPRSS2 to endosomal cathepsins for cell entry. *Virology*. 517:9–15. doi:10.1016/J.VIROL.2017.11.012.

Small, G.I., O. Fedorova, P.D.B. Olinares, B.T. Chait, S.A. Darst, E.A. Campbell, G.I. Small, O. Fedorova, P.D.B. Olinares, J. Chandanani, and A. Banerjee. 2023. Structural and functional insights into the enzymatic plasticity of the SARS-CoV-2 NiRAN domain. *Mol. Cell*. 83:3921-3930.e7. doi:10.1016/j.molcel.2023.10.001.

Sola, I., F. Almazán, S. Zúñiga, and L. Enjuanes. 2015. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu. Rev. Virol.* 2:265–288. doi:10.1146/ANNUREV-VIROLOGY-100114-055218/SUPPL_FILE/VI02_ENJUANES_FIG2.PPSX.

Steinhauer, D.A., E. Domingo, and J.J. Holland. 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene*. 122:281–288. doi:10.1016/0378-1119(92)90216-C.

Stertz, S., M. Reichelt, M. Spiegel, T. Kuri, L. Martínez-Sobrido, A. García-Sastre, F. Weber, and G. Kochs. 2007. The intracellular sites of early replication and budding of SARS-coronavirus. doi:10.1016/j.virol.2006.11.027.

Suryawanshi, R.K., R. Koganti, A. Agelidis, C.D. Patil, and D. Shukla. 2021. Dysregulation of Cell Signaling by SARS-CoV-2. *Trends Microbiol.* 29:224–237. doi:10.1016/J.TIM.2020.12.007.

Tsang, H.F., L.W.C. Chan, W.C.S. Cho, A.C.S. Yu, A.K.Y. Yim, A.K.C. Chan, L.P.W. Ng, Y.K.E. Wong, X.M. Pei, M.J.W. Li, and S.C.C. Wong. 2021. An update on COVID-19 pandemic: the epidemiology, pathogenesis, prevention and treatment strategies. *Expert Rev. Anti. Infect. Ther.* 19:877–888. doi:10.1080/14787210.2021.1863146.

Tvarogová, J., R. Madhugiri, G. Bylapudi, L.J. Ferguson, N. Karl, and J. Ziebuhr. 2019. Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3′-Terminal Adenylyltransferase Activity. *J. Virol.* 93:291–310. doi:10.1128/JVI.00291-19/ASSET/01E96363-1454-4B89-9C80-E4048C038DF3/ASSETS/GRAPHIC/JVI.00291-19-F0009.JPEG.

V'kovski, P., A. Kratzel, S. Steiner, H. Stalder, and V. Thiel. 2021. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* 19:155–170. doi:10.1038/S41579-020-00468-6.

Te Velthuis, A.J.W. 2014. Common and unique features of viral RNA-dependent polymerases. *Cell. Mol. Life Sci.* 71:4403–4420. doi:10.1007/s00018-014-1695-z.

Viswanathan, T., S. Arya, S.H. Chan, S. Qi, N. Dai, A. Misra, J.G. Park, F. Oladunni, D. Kovalskyy, R.A. Hromas, L. Martinez-Sobrido, and Y.K. Gupta. 2020. Structural basis of RNA cap modification by SARS-CoV-2. *Nat. Commun. 2020 111*. 11:1–7. doi:10.1038/s41467-020-17496-8.

Wang, B., D. Svetlov, and I. Artsimovitch. 2021a. NMPylation and de-NMPylation of

SARS-CoV-2 nsp9 by the NiRAN domain. *Nucleic Acids Res.* 49:8822–8835. doi:10.1093/nar/gkab677.

Wang, B., V. Svetlov, Y.I. Wolf, E. V. Koonin, E. Nudler, and I. Artsimovitch. 2021b. Allosteric activation of sars-cov-2 rna-dependent rna polymerase by remdesivir triphosphate and other phosphorylated nucleotides. *MBio*. 12. doi:10.1128/mBio.01423-21.

Wang, D., A. Jiang, J. Feng, K. Lan, Y. Chen, and Y.Z. Correspondence. The SARS-CoV-2 subgenome landscape and its novel regulatory features. doi:10.1016/j.molcel.2021.02.036.

Wang, Q., J. Wu, H. Wang, Y. Gao, Q. Liu, A. Mu, W. Ji, L. Yan, Y. Zhu, C. Zhu, X. Fang, X. Yang, Y. Huang, H. Gao, F. Liu, J. Ge, Q. Sun, X. Yang, W. Xu, Z. Liu, H. Yang, Z. Lou, B. Jiang, L.W. Guddat, P. Gong, and Z. Rao. 2020. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*. 182:417-428.e13. doi:10.1016/j.cell.2020.05.034.

Yan, L., Y. Zhang, J. Ge, L. Zheng, Y. Gao, T. Wang, Z. Jia, H. Wang, Y. Huang, M. Li, Q. Wang, Z. Rao, and Z. Lou. 2020. Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat. Commun.* 11:3–8. doi:10.1038/s41467-020-19770-1.

Yoshimoto, F.K. 1234. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein J.* 39:198–216. doi:10.1007/s10930-020-09901-4.