EXPRESSION OF SARS-COV-2 K51A/S54A NSP7 PROTEIN

IN *Escherichia coli* C41(DE3) and *Escherichia coli* BL21(DE3)

NURUL ALYA BINTI MAZRI

SCHOOL OF HEALTH SCIENCES

UNIVERSITI SAINS MALAYSIA

2025

EXPRESSION OF SARS-COV-2 K51A/S54A NSP7 PROTEIN IN *Escherichia coli* C41(DE3) and *Escherichia coli* BL21(DE3)

by

NURUL ALYA BINTI MAZRI

Dissertation submitted in partial fulfillment of the requirements

for the degree of

Bachelor of Health Sciences (Honours) (Biomedicine)

January 2025

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except otherwise stated and duly acknowledged. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at Universiti Sains Malaysia or other institutions. I grant Universiti Sains Malaysia the right to use the dissertation for teaching, research, and promotional purposes.

Signature,

……
(NURUL ALYA BINTI MAZRI)

Date: 27/01/2025

ACKNOWLEDGEMENT

Alhamdulillah, all praise is due to Allah, the Almighty, for granting me the strength, wisdom, and perseverance to complete this final year research project. His endless blessings have been my guiding light throughout this journey, and I would not be where I am today without His grace.

First and foremost, I would like to express my heartfelt gratitude to my supervisors, Assoc. Prof. Dr. Few Ling Ling and Assoc. Prof. Dr. See Too Wei Cun, for their unwavering support, insightful guidance, and invaluable advice during this research journey. Their expertise and encouragement have been instrumental in the successful completion of my project titled "Expression of SARS-CoV-2 K51A/S54A NSP7 protein in *Escherichia coli* C41(DE3) and *Escherichia coli* BL21(DE3)." Besides, I am deeply thankful to Mr. Mhd Houssam Tounbakjy, Ms. Shaleniprieya A/P Muniandy, and Ms. Lim Yinwun for their invaluable support, guidance, and dedication in teaching me essential research skills.

I am also deeply grateful to my family, especially my beloved mother, Mrs. Norita binti Abdul Ghani, whose prayers and unwavering support have been my greatest source of strength. Special thanks to my precious friends, Ms. Angelyn Lee Chui Yee, Ms. Nur Hasya Athirah binti Hanizam, and Ms. Hanina Sofea binti Faizal, for their encouragement, companionship, and endless support throughout this journey. From late-night pep talks to celebrating small wins together, they've made every step of this journey so much more meaningful.

Lastly, I want to take a moment to thank myself for showing up every single day, even when it was hard, for staying resilient in the face of challenges, and for believing in

my abilities when the road felt uncertain. Here's to every late night, early morning, and the endless effort I’ve poured into reaching this point. Long story short, I survived.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

%	-	Percentage
≤	-	Less than or equal to
≥	-	More than or equal to
°C	-	Degree Celsius
µL	-	Microlitre
µg/mL	-	Microgram per Millilitre
µM	-	Micromolar
mM	-	Millimolar
bp	-	Base pair
g force	-	Gravitational Force
g	-	Gram
kb	-	Kilobase
kDa	-	Kilodalton
min	-	Minute
mL	-	Millilitre
ng	-	Nanogram
pH	-	Potential of Hydrogen
Tm	-	Melting temperature
V		Volt
v/v		Volume per volume
w/v		Weight per volume

LIST OF ABBREVIATIONS

ARDS	-	Acute respiratory distress syndrome
BEVS	-	Baculovirus expression vector system
CBC	-	Complete blood count
CHO	-	Chinese Hamster Ovary
DNA	-	Deoxyribonucleic acid
E	-	Envelope protein
EUAs	-	Emergency Use Authorizations
E. coli	-	*Escherichia coli*
ExoN	-	Exonuclease
FDA		Food and Drug Administration
GRAS	-	generally recognizerd as safe
HEK 293	-	Human Embryonic Kidney
HBsAg	-	Hepatitis B surface antigen
His6 tag	-	Hexa-histidine tag
IMAC	-	Immobilized metal affinity chromatography
IPTG	-	Isopropyl B-D-1-thiogalactopyranoside
LB	-	Luria broth
LPS	-	Lipopolysaccharides
M	-	Membrane protein
MCS	-	Multiple cloning site
MERS-CoV	-	Middle East respiratory syndrome coronavirus
NiRAN	-	N-terminal nidovirus RdRp-associated nucleotidyltransferase

NMR	-	Nuclear magnetic resonance
NSPs	-	Non-structural proteins
NTPs	-	Nucleotide triphosphates
ORFs	-	Open reading frames
OGCP	-	Bovine oxoglutarate-malate transport protein
PCR	-	Polymerase chain reaction
PEI	-	Polyethyleneimine
PTMs	-	Post-translational modifications
RTC	-	Replication-transcription complex
RdRp	-	RNA-dependent RNA polymerase
RE	-	Restriction enzyme
RNA	-	Ribonucleic acid
SARS-CoV	-	Severe acute respiratory syndrome coronavirus
+ssRNA	-	Positive-sense single-stranded RNA
sgRNAs	-	Subgenomic RNAs
SEC-MALS-SAXS	-	Size-exclusion chromatography with multi-angle light scattering and small-angle X-ray scattering
SD	-	Standard deviation
SDS-PAGE	-	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TB	-	Terrific broth
TRSs	-	Transcription regulatory sequences
TRS-L	-	TRS leader
UTRs	-	untranslated regions
VLPs	-	Virus-like particles

EKSPRESI PROTEIN K51A/S54A NSP7 SARS-COV-2 DALAM *Escherichia coli* C41(DE3) DAN *Escherichia coli* BL21(DE3)

ABSTRAK

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) bergantung kepada kompleks RNA-dependent RNA polymerase (RdRp) yang terdiri daripada NSP7, NSP8, dan NSP12 untuk proses replikasi dan transkripsi. Mutasi dalam NSP7, seperti K51A dan S54A, mengganggu interaksi dengan NSP8 dan NSP12, seterusnya merosakkan aktiviti polimerase dan integriti struktur. Mutasi ini memberikan pandangan yang penting dalam pembangunan terapeutik antivirus yang menyasarkan kompleks RdRp. Kajian ini bertujuan untuk mengekspresikan protein mutan K51A/S54A NSP7 dalam *Escherichia coli* C41(DE3) dan BL21(DE3) menggunakan vektor pET-15(b). Metodologi merangkumi penyediaan konstruk plasmid pET-15(b)-mutan NSP7, transformasi plasmid ke dalam *E. coli*, dan pengoptimuman ekspresi protein. Parameter seperti kepekatan IPTG (0–1.0 mM) dan strain perumah yang berbeza telah dioptimumkan secara sistematik. Protein mutan NSP7 (K51A/S54A) berjaya diekspresikan dalam kedua-dua strain *E. coli*, dengan SDS-PAGE menunjukkan jalur yang jelas pada berat molekul yang dijangka. Ekspresi protein optimum dicapai pada kepekatan IPTG 0.5 mM. Hasil kajian menunjukkan perbezaan kecekapan ekspresi antara *E. coli* C41(DE3) dan *E. coli* BL21(DE3), di mana *E. coli* BL21(DE3) menghasilkan tahap ekspresi yang lebih tinggi. Kajian ini menyumbang kepada pemahaman tentang replikasi SARS-CoV-2 dengan menyediakan keadaan optimum untuk mengekspresikan protein K51A/S54A NSP7, sekaligus membolehkan kajian berkenaan fungsi dan struktur. Penemuan ini menyediakan asas untuk meneroka mutasi NSP7 sebagai sasaran terapeutik

dan menunjukkan keberkesanan sistem bakteria untuk menghasilkan protein virus dalam jumlah yang tinggi, berpotensi mempercepatkan penemuan ubat antivirus.

EXPRESSION OF SARS-COV-2 K51A/S54A NSP7 PROTEIN IN *Escherichia coli* C41(DE3) AND *Escherichia coli* BL21(DE3)

ABSTRACT

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) relies on the RNA-dependent RNA polymerase (RdRp) complex, comprising NSP7, NSP8, and NSP12, for replication and transcription. Mutations in NSP7, such as K51A and S54A, disrupt interactions with NSP8 and NSP12, impairing polymerase activity and structural integrity. These mutations offer insights into the development of antiviral therapeutics targeting the RdRp complex. This study aimed to express the mutant K51A/S54A NSP7 protein in *Escherichia coli* C41(DE3) and BL21(DE3) using the pET-15(b) vector. The methodologies included the preparation of the pET-15(b)-mutant NSP7 plasmid construct, transformation of the plasmid into *E. coli*, and optimization of protein expression. Parameters such as IPTG concentrations (0–1.0 mM) and different host strains were systematically optimized. The mutant NSP7 (K51A/S54A) protein was successfully expressed in both *E. coli* strains, with SDS-PAGE revealing distinct bands at the expected molecular weight. Optimal protein expression was achieved at 0.5 mM IPTG concentration. The findings highlighted differences in expression efficiency between *E. coli* C41(DE3) and *E. coli* BL21(DE3), with *E. coli* BL21(DE3) yielding higher expression levels. This research contributes to the understanding of SARS-CoV-2 replication by providing optimized conditions for expressing K51A/S54A NSP7 protein, enabling functional and structural studies. The findings offer a foundation for exploring NSP7 mutations as therapeutic targets and demonstrate the utility of bacterial systems for producing viral proteins in high yield, potentially accelerating antiviral drug discovery.

CHAPTER 1: INTRODUCTION

1.1 Background of Study

SARS-CoV-2, the virus responsible for the COVID-19 pandemic, was first identified in December 2019 in Wuhan, China. The virus spread rapidly worldwide, causing significant disruption to healthcare systems, economies, and daily life. SARS-CoV-2 is a member of the Coronaviridae family, characterized by its large RNA genome and reliance on host machinery for replication. Central to its replication process are the non-structural proteins (NSPs), which form the replication-transcription complex (RTC) necessary for copying the viral genome and synthesizing viral proteins (Hao et al., 2022). Understanding how these proteins function, particularly the NSP7-NSP8-NSP12 RNA-dependent RNA polymerase (RdRp) complex, is critical for developing strategies to inhibit viral replication.

NSP7 plays an essential role in stabilizing the interaction between NSP8 and NSP12 in the RdRp complex, a key component of the viral RTC. Together with NSP8 and NSP12, NSP7 facilitates the polymerization of RNA, allowing the virus to replicate its genome efficiently. Disrupting these interactions can impede viral replication, making NSP7 an attractive target for antiviral research. Despite its small size, NSP7 is crucial in maintaining the structural integrity and functionality of the RdRp complex, making it an important subject of study for understanding viral replication mechanisms (Wilamowski et al., 2021).

Mutations in NSP7, such as K51A and S54A, offer valuable insights into the protein's role within the RdRp complex. In the K51A mutation, lysine is replaced by alanine, which is expected to disrupt the NSP7-NSP8-NSP12 interaction, potentially weakening the overall polymerase activity. Similarly, the S54A mutation, where serine is

replaced with alanine, may affect the structural stability of the complex (Neto et al., 2024). By studying these mutations, researchers can better understand how NSP7 contributes to the overall viral replication process. Moreover, these mutations could serve as potential drug targets, as inhibitors could be designed to mimic their effects and impair the virus's ability to replicate.

To investigate the biochemical properties of mutant NSP7, it is necessary to produce large quantities of the protein for analysis. *Escherichia coli* (*E. coli*) is one of the most widely used systems for producing recombinant proteins due to its fast growth and ease of manipulation. The *E. coli* C41(DE3), a derivative of BL21(DE3), is often employed as it is more tolerant of toxic proteins and can sustain higher levels of protein expression without compromising cell viability (Baumgarten et al., 2017). The pET expression system, commonly used with *E. coli*, enables high-level protein expression under the control of the T7 promoter (Shilling et al., 2020). Given the importance of producing sufficient quantities of mutant NSP7 for structural and functional analysis, optimizing protein expression becomes crucial.

In bacterial expression systems, various parameters are commonly optimized to improve protein yield, solubility, and stability. Optimization can involve adjusting factors such as the concentration of the inducer (e.g., isopropyl β-D-1-thiogalactopyranoside (IPTG)), growth temperature, media composition, and varying host strains. Each of these factors can influence the efficiency and quality of protein expression. Lower induction temperatures, for example, can reduce the formation of inclusion bodies and enhance protein solubility, while IPTG concentration can be adjusted to balance expression levels and minimize toxicity. Besides, different *E. coli* strains can offer unique advantages in recombinant protein production, particularly for difficult-to-express proteins (Francis & Page, 2010). By comparing the expression of mutant NSP7 between different strains, it

is possible to optimize protein yield and solubility, ensuring the production of high-quality protein for downstream analyses. Therefore, this study will focus on optimizing two key parameters which are varying IPTG concentrations and selecting different *E. coli* host strains, specifically C41(DE3) and BL21(DE3). The optimization is essential not only for the successful expression of mutant NSP7 but also for advancing methods in recombinant protein production for future studies on SARS-CoV-2 and other pathogens.

Studying the expression of mutant NSP7 in *E. coli* provides valuable information for understanding its role in the RdRp complex. Successfully producing and characterizing this protein will allow for detailed structural and functional analyses, which are crucial for identifying vulnerabilities in the viral replication machinery. These insights could contribute to the development of antiviral drugs targeting NSP7, particularly in its mutant forms. Furthermore, optimizing bacterial expression systems for producing viral proteins adds to the broader field of recombinant protein expression, benefiting future studies on SARS-CoV-2 and other pathogens.

1.2 Problem statement

The COVID-19 pandemic has created a significant global economic and public health crisis, highlighting the need for effective treatments. Currently, remdesivir is the only authorized antiviral for COVID-19. It is a nucleoside analogue that targets the RdRp of the virus, causing errors in viral RNA synthesis that ultimately halt replication (Dance, 2021). However, its effectiveness remains controversial, with some clinical trials showing no impact on recovery time or mortality rates (Bertolin et al., 2021). Given the reliance of viruses on human cell machinery for replication, developing antivirals that do not harm healthy cells remains challenging. The ongoing threat of SARS-CoV-2 underscores the urgent need for new antiviral treatments.

Coronavirus replication relies on NSPs encoded by ORF1a and ORF1ab, which are initially produced as polyproteins and later cleaved into mature forms. Among them, NSP12 functions as the RdRp, but its activity is significantly enhanced by NSP7 and NSP8 cofactors. The NSP7-NSP8-NSP12 subcomplex acts as a minimal core complex for viral RNA synthesis (Peng et al., 2020). While mutations in RdRp such as P323L have been linked to altered enzyme stability and proofreading ability, the impact of mutations in NSP7 and NSP8 on the polymerase complex remains unclear. Despite evidence linking mutations in other viral proteins, like the spike protein, to disease severity, there is limited research on how changes in RdRp, NSP7, or NSP8 affect SARS-CoV-2 virulence (Reshamwala et al., 2021). Since NSP7 is crucial for the replication-transcription complex, mutations in NSP7 may impact viral replication and pathogenicity, yet data on the expression and function of mutant forms of NSP7 are lacking.

E. coli is a widely used bacterial host for recombinant protein production due to its rapid growth, cost-effectiveness, and well-established genetic manipulation tools. However, achieving high-yield expression of soluble recombinant proteins, such as mutant NSP7, in *E. coli*, is often challenging due to protein aggregation into inclusion bodies. This research faces significant hurdles in producing sufficient quantities of NSP7 for functional studies and therapeutic evaluation. Despite efforts to optimize coding sequences and employ various expression vectors, current approaches have failed to achieve the desired protein yields, limiting further biochemical characterization and exploration of the protein's therapeutic potential (Bhatwa et al., 2021).

1.3 Null Hypothesis

Optimization of IPTG concentration and host strains will not demonstrate higher protein yield and more stable expression of soluble His-tagged K51A/S54A NSP7 protein.

1.4 Alternative Hypothesis

Optimization of IPTG concentration and host strains will demonstrate higher protein yield and more stable expression of soluble His-tagged K51A/S54A NSP7 protein.

1.5 Rationale of study

The study aims to deepen the understanding of SARS-CoV-2 replication by investigating how specific mutations in NSP7 affect its function within the replication-transcription complex. The NSP7-NSP8-NSP12 subcomplex is a critical component for viral RNA synthesis, and mutations in these proteins could significantly impact the virus's replication efficiency and adaptability. By examining the effects of NSP7 mutations, this research will provide valuable insights into the molecular mechanisms that govern the replication of SARS-CoV-2 and its ability to adapt to environmental changes.

Furthermore, by characterizing the mutant form of NSP7, the aim is to identify potential vulnerabilities that can be targeted by new antiviral strategies. Understanding the relationship between NSP7 mutations and viral pathogenicity may reveal new therapeutic targets and aid in developing drugs that inhibit SARS-CoV-2 replication more effectively. The findings could contribute to broader efforts in designing antiviral agents, especially given the ongoing threat posed by SARS-CoV-2 and the need for more effective treatment options.

Lastly, this study seeks to optimize the conditions for expressing the mutant NSP7 protein in *E. coli* to obtain sufficient quantities for further research. High-yield production of soluble mutant NSP7 is essential for conducting comprehensive functional analyses and biochemical characterizations, as well as exploring the protein's therapeutic potential. Various optimization strategies, including IPTG concentration, growth temperature, media composition, and host strain selection, can impact protein expression levels and

solubility. This study focused on optimizing two key parameters, which are the IPTG concentrations and different *E. coli* host strains, specifically C41(DE3) and BL21(DE3). By evaluating these factors, the aim was to overcome challenges related to protein toxicity and inclusion body formation, ensuring the production of soluble mutant NSP7 for comprehensive functional analyses and biochemical characterizations. Addressing these optimization challenges will facilitate future studies on SARS-CoV-2 and contribute to the development of antiviral strategies.

1.6 Objectives

The general objective of the project was to express the mutant NSP7 (K51A/S54A NSP7) of SARS-CoV-2 in *E. coli* C41(DE3).

There were three specific objectives:

1. To prepare and verify the pET-15(b)-K51A/S54A-*nsp7* plasmid construct.
2. To transform pET-15(b)-K51A/S54A-*nsp7* plasmid construct into *E. coli* C41(DE3).
3. To optimize the expression of His-tag K51A/S54A NSP7 protein in *E. coli*.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction to SARS-CoV-2

Coronaviruses are a large and varied group of viruses that infect numerous animal species, including humans, and cause respiratory illnesses that range from mild colds to severe pneumonia. In recent decades, two especially dangerous strains have emerged, known as severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012. Both of these viruses caused serious respiratory diseases with high mortality rates, demonstrating how emerging coronaviruses pose a significant threat to public health (Hu et al., 2020). The COVID-19 pandemic, caused by SARS-CoV-2, continues to be a global health emergency, underscoring the urgent need for effective therapies and preventive measures.

SARS-CoV-2, the virus responsible for COVID-19, belongs to the Betacoronavirus genus and shares close genetic links to SARS-CoV and MERS-CoV. It was first identified in Wuhan, China, in late 2019 and rapidly spread worldwide due to its high transmissibility, far surpassing SARS and MERS in total cases and geographical spread (Hu et al., 2020). Genomic research indicates that SARS-CoV-2 likely originated from a bat coronavirus, showing a 96% similarity with bat betaCoV RaTG13, and may have entered the human population through intermediate animals, such as pangolins or minks. The virus itself has a round or elliptical shape and is sensitive to ultraviolet light, heat, and lipid solvents. It can survive longer at lower temperatures, which has facilitated its widespread transmission. Understanding its zoonotic origins and rapid spread is essential for ongoing research and monitoring of coronaviruses as they evolve (Cascella et al., 2023).

Within the Coronaviridae family, coronaviruses are grouped under the Orthocoronavirinae subfamily, which is further divided into four genera which are

Alphacoronavirus, Betacoronavirus, Deltacoronavirus, and Gammacoronavirus. Research indicates that bats and rodents are the main genetic sources of alpha and beta coronaviruses, while avian species are the main sources of delta and gamma coronaviruses. This extensive family of viruses infects a wide range of animals, including camels, cattle, cats, and bats, with some strains able to jump species barriers and infect humans, causing outbreaks of respiratory illnesses (Cascella et al., 2023).

SARS-CoV-2 spreads primarily through respiratory droplets expelled by infected individuals, whether symptomatic or not. Airborne transmission can also occur, especially during medical procedures that generate aerosols. Transmission through contaminated surfaces, or fomites, has been well documented, with the virus remaining viable for up to 72 hours on surfaces like stainless steel and plastic, and up to 28 days on glass under controlled conditions. However, it is less stable on porous materials. In healthcare settings, SARS-CoV-2 has been detected on various surfaces, and airborne viral particles have been found as far as 4 meters from infected patients. Although vertical transmission from mother to neonate is rare, it has been observed in a small number of cases (Cascella et al., 2023).

Upon entering the body, SARS-CoV-2 binds to host cells through the ACE2 receptor, which is abundantly expressed on pulmonary epithelial cells. The virus's spike (S) protein, crucial for cell entry, is composed of two subunits which are S1, which binds to the host cell receptor, and S2, which enables the fusion of viral and host membranes. The spike protein first binds to the ACE2 receptor and then undergoes a two-step cleavage process by host proteases. This cleavage process primes the spike protein, allowing it to fuse with the host cell membrane and initiate infection. Once inside, the virus replicates its RNA, which serves as a template to produce new viral particles (Hu et al., 2020). These particles are assembled in the endoplasmic reticulum and Golgi apparatus before being

released through exocytosis to infect nearby cells or spread through respiratory droplets (Parasher, 2020).

COVID-19 has an incubation period of about 5 to 6 days, with a range extending up to 14 days, during which the virus can still spread (Parasher, 2020). COVID-19 has shown that individuals of all ages are susceptible to infection. However, those aged 60 and older or with comorbidities like obesity, cardiovascular disease, diabetes, and chronic lung disease are at a higher risk of severe outcomes. Age is the strongest predictor of severe illness, with older adults facing a significantly higher risk of death compared to younger individuals. For instance, those aged 65 to 74 have a 60 times higher risk of death, and those over 85 have a 340 times higher risk, compared to those under 30. Additionally, people with preexisting conditions face a six times higher rate of hospitalization and a twelve times higher rate of mortality due to COVID-19, highlighting the compounding risk factors involved (Cascella et al., 2023).

Symptoms vary from mild, such as fever, body aches, and dry cough, to severe respiratory distress. Gastrointestinal symptoms like abdominal pain, vomiting, and diarrhea are also reported in some cases (Parasher, 2020). Severe cases often progress to dyspnea and acute respiratory distress syndrome (ARDS), characterized by lung inflammation and significant tissue damage. In critical cases, a "cytokine storm", an extreme immune response involving high levels of inflammatory cytokines like IL-1, IL-6, IL-8, and TNF, can cause widespread tissue damage and systemic hyperinflammation. Severe COVID-19 can lead to multi-organ complications, including damage to the heart, kidneys, and liver, along with clotting disorders and shock. Risk factors for severe illness include advanced age, obesity, male gender, and underlying conditions such as hypertension and diabetes. Genetic factors, especially those affecting interferon pathways, may also contribute to the severity of the disease. Although COVID-19 has a

range of systemic effects, its most severe impact is often on the lungs, where viral replication triggers ARDS (Lamers & Haagmans, 2022).

Effective COVID-19 diagnosis relies primarily on polymerase chain reaction (PCR) testing of nasopharyngeal swabs, which offers high specificity, though sensitivity varies with sample quality and timing. Meanwhile, antigen tests provide faster results than PCR, though with lower sensitivity, making them suitable for rapid testing in some cases. For hospitalized patients, laboratory tests such as complete blood count (CBC), metabolic panels, and D-dimer help to assess disease severity, while chest X-rays and CT scans can reveal lung involvement, though imaging alone is not diagnostic (Cascella et al., 2023).

There are two primary processes that drive COVID-19 pathogenesis which are viral replication during the initial stage of illness and a dysregulated immune or inflammatory response in later stages that results in widespread tissue damage. Currently, remdesivir is the only antiviral drug approved by the U.S. Food and Drug Administration (FDA) to treat COVID-19 by limiting viral replication early on. It also serves as an immunomodulators to manage inflammation in advanced stages. Other treatments, including ritonavir-boosted nirmatrelvir, molnupiravir, and high-titer COVID-19 convalescent plasma, have Emergency Use Authorizations (EUAs) for COVID-19 treatment. Additionally, monoclonal antibodies, specifically tixagevimab (300 mg) plus cilgavimab (300 mg), have EUAs for use as pre-exposure prophylaxis for certain individuals at higher risk (Cascella et al., 2023).

COVID-19 remains a global health challenge, with severe outcomes linked to factors such as age, underlying health conditions, and demographic disparities. Hence, understanding the mechanisms of SARS-CoV-2 infection, from viral replication to immune dysregulation, has been crucial in guiding therapeutic strategies. Currently,

antiviral treatments like remdesivir and therapies for immune modulation, such as monoclonal antibodies and immunomodulators, are used to manage the disease. Continued research and the development of targeted treatments are essential in combating the ongoing impact of COVID-19, especially as the virus continues to evolve.

2.2 SARS-CoV-2 RNA replication machinery

SARS-CoV-2, a newly identified betacoronavirus, shares 79% genome sequence identity with SARS-CoV and 50% with MERS-CoV. Its genome is a positive-sense single-stranded RNA (+ssRNA) approximately 30 kb in length, featuring a 5′ cap and a 3′ poly(A) tail, allowing it to function directly as mRNA for translating viral polyproteins (Romano et al., 2020). The genome's structure is similar to other betacoronaviruses, containing six main open reading frames (ORFs) arranged in a 5′ to 3′ orientation, such as replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M), and nucleocapsid (N). Additionally, seven ORFs encoding accessory proteins are interspersed among the structural genes (Hu et al., 2020).

SARS-CoV-2 shares over 90% amino acid similarity in structural genes with SARS-CoV, except for the spike (S) gene, which is more divergent. Meanwhile, the replicase gene occupies two-thirds of the 5′ genome and encodes the polyproteins PP1a and PP1ab, which are proteolytically cleaved into 16 non-structural proteins (NSP1-16) (Hu et al., 2020). These proteins form the replication-transcription complex (RTC), which is essential for viral genome replication and transcription. The RTC includes core enzymes such as RNA-dependent RNA polymerase (RdRp, NSP12), helicase (NSP13), and other proteins involved in RNA capping, proofreading, and modification. Additionally, structured untranslated regions (UTRs) at the 5′ and 3′ ends of the genome further regulate replication and transcription (Romano et al., 2020). Figure 2.1 illustrates the genomic arrangement of SARS-CoV-2, highlighting the ORF1a and ORF1b regions

encoding non-structural proteins (NSP1-16) and the structural proteins (S, E, M, and N), alongside accessory factors, with key functional domains and cleavage sites labeled.

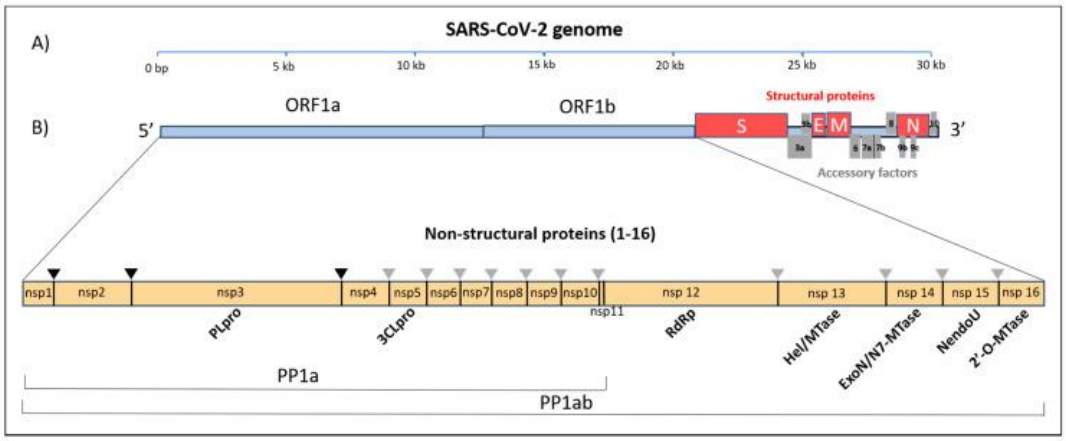

Figure 2. 1: Genomic arrangement of SARS-CoV-2 (Romano et al., 2020)

The replication process starts with the production of full-length negative-sense RNA copies from the positive-sense viral genome. These negative-sense RNAs act as templates for new positive-sense genomic RNA synthesis, which are then either translated to produce more NSPs and RTCs or are packaged into new viral particles. Besides, coronaviruses use discontinuous transcription, a special process to make smaller RNA molecules called subgenomic RNAs (sgRNAs). During the synthesis of the negative-strand RNA, the RTC can pause at specific sites called transcription regulatory sequences (TRSs) near the beginning of each gene, the ORF. After pausing, the RTC jumps to another TRS near the 5′ end of the genome, called the TRS leader (TRS-L). This process creates negative-strand sgRNAs, which are then used as templates to produce positive-strand sgRNAs. These positive-strand sgRNAs are used to translate structural and accessory proteins required for assembling new viral particles. Although structurally

polycistronic, these sgRNAs are functionally monocistronic, translating only the first ORF at the 5′ end of each sgRNA (V'kovski et al., 2020).

The RTC's activity relies on complex interactions among non-structural proteins such as NSP7, NSP8, and NSP13, which coordinate replication and transcription effectively. Certain non-structural proteins also play roles in immune evasion. For example, NSP1 suppresses host gene expression, aiding viral virulence. Additionally, the nucleocapsid (N) protein safeguards the viral genome by packaging it into a ribonucleocapsid structure, facilitating genome packaging in newly formed virions (Romano et al., 2020). Other than that, NSP13 which is an SF1B-family RNA helicase, is also crucial for coronavirus replication. Its 5′-3′ RNA unwinding activity doubles in efficiency when paired with NSP12. NSP13 forms stable interactions with the RTC, allowing researchers to study its structure within the RTC complex. Structural studies revealed that two NSP13 molecules, NSP13F and NSP13T, are bound to the RTC at fingers and thumb subdomains respectively. Both NSP13F and NSP13T interact extensively with the RdRp, but only NSP13T binds to the RNA template at its 5′ end. NSP13T, positioned downstream of the RdRp active site, moves in the opposite direction of the RdRp. This opposing movement may create a translocation conflict. As NSP13T moves forward on the template RNA, it may cause the RdRp to move backward on the product RNA, a process known as backtracking, which plays a crucial role in removing incorrectly incorporated nucleotides during coronavirus RNA synthesis (Malone et al., 2021).

Central to RNA replication and transcription is the RdRp complex, comprising NSP12, NSP8, and NSP7. While NSP12 exhibits limited catalytic activity on its own, its function is significantly enhanced by the cofactor subunits NSP7 and NSP8, which together form the minimal core complex required for efficient RNA replication.

Structurally, the RdRp of SARS-CoV-2 closely resembles that of SARS-CoV. The NSP12 subunit includes an N-terminal nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain, an interface domain, and a C-terminal RdRp domain. The NiRAN domain, connected to the RdRp domain through an N-terminal β-hairpin structure, may play a regulatory role in polymerization. The RdRp domain itself adopts a "right-hand" structure, with distinct fingers, palm, and thumb subdomains, typical of single-subunit polymerases required for RNA synthesis (Hillen et al., 2020).

The accessory subunits NSP7 and NSP8 stabilize and modulate the function of NSP12 by binding to different regions of the protein. One NSP8 subunit binds to the fingers domain, while a heterodimer of NSP7-NSP8 associates with the thumb domain. The binding site for this heterodimer lies at the interface of the “index finger” and “thumb” domains of NSP12, where an alpha-helix in the index finger loop forms a close interaction with the thumb domain, providing additional structural stability. The electrostatic surface of NSP12 is primarily negatively charged, with positively charged regions around the RNA template and NTP-binding sites. Hence, these positively charged areas attract the negatively charged RNA and NTPs, helping to align them properly for replication. In contrast, the regions where NSP7, NSP8, and the RNA exit tunnel interact with NSP12 are relatively neutral in charge. As a result, this neutrality ensures smooth and stable interactions between these components. Additionally, a second NSP8 subunit provides positively charged residues along the RNA template-binding channel. These positive charges help stabilize the negatively charged RNA as it moves through the replication machinery, ensuring the RdRp complex functions effectively (Kirchdoerfer & Ward, 2019).

The NSP7-NSP8 heterodimer is particularly important in stabilizing the RdRp complex, positioned above the NSP12 thumb subdomain. It strengthens the interface

between the thumb and finger subdomains. NSP7 plays a significant role in binding affinity with NSP12, while NSP8 makes minimal contact, though a second NSP8 copy (NSP8-2) interacts with the finger subdomain, adding further stability to the complex (Jiang et al., 2021). NSP8 not only enhances the processivity of the RdRp complex but also interacts *in vitro* with various other non-structural proteins that likely support the RTC. This suggests that NSP8 plays a crucial role in forming higher-order RTCs. These complexes coordinate multiple functions, including template unwinding involving NSP13, proofreading involving NSP10, NSP12, NSP13, and NSP14, and RNA capping involving NSP10, NSP13, NSP14, and NSP16 (Malone et al., 2021). Figure 2.2 presents two views of the cryo-EM map of the SARS-CoV-2 apo RdRp complex structure, showcasing key components, including NSP7, NSP8-1, NSP8-2, and the catalytic subunit, with labeled functional domains such as the Fingers, Thumb, Palm, NiRAN, β-hairpin, and Interface.

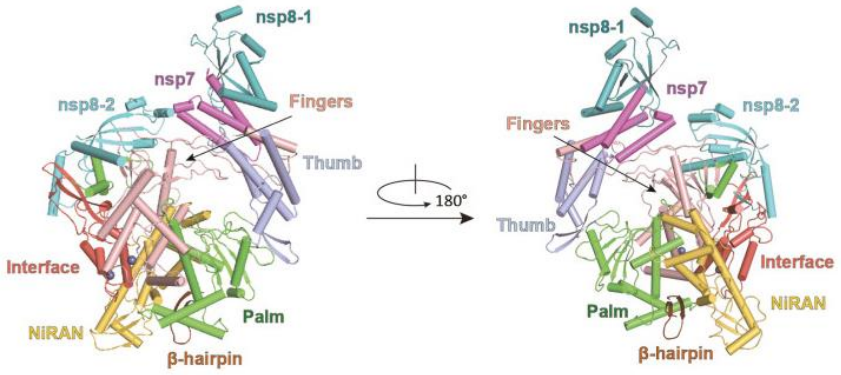

Figure 2. 2: Two views of cryo-EM map of SARS-CoV-2 apo RdRp complex structure (Jiang et al., 2021).

The RdRP which is encoded by NSP12, is essential for both genome replication and sgRNA transcription. The RdRP complex, which includes cofactors NSP7 and NSP8, shares a high degree of structural similarity with SARS-CoV and other RNA viruses,

making it a potential antiviral target. The conserved active site of RdRP allows for the possibility of repurposing antiviral drugs that are effective against other RNA viruses. One such drug, remdesivir, is incorporated into the viral RNA in place of ATP, leading to delayed-chain termination. Remdesivir is a phosphoramidate analogue that effectively inhibits the RdRp of various coronaviruses, demonstrating antiviral activity in cell cultures and animal models. It has been granted emergency use authorization for treating COVID-19 in the United States and is undergoing clinical trials worldwide (Hillen et al., 2020). Unlike classic nucleoside analogues that stop RNA synthesis immediately, remdesivir allows synthesis to continue for a few nucleotides before halting, potentially bypassing SARS-CoV-2's proofreading enzyme, NSP14 exonuclease (ExoN). In vitro studies show that remdesivir reduces SARS-CoV-2 replication and clinical studies have demonstrated limited effectiveness in severe COVID-19 cases, suggesting further research is needed (V'kovski et al., 2020). Therefore, the current study aimed to contribute to this research by investigating the expression and characterization of SARS-CoV-2 K51A/S54A NSP7 in *E. coli*, a protein critical for viral replication. By analyzing the structural and functional properties of K51A/S54A NSP7, this study could help uncover potential resistance mechanisms to remdesivir. Understanding these mechanisms may lead to strategies for improving the drug's efficacy, including the development of combination therapies or the design of next-generation antiviral drugs that can better target emerging viral variants.

In conclusion, the SARS-CoV-2 RNA replication machinery's detailed architecture and regulatory mechanisms underscore its significance in viral replication. Deepening our understanding of these processes enhances our insight into viral biology and supports the development of targeted therapeutic strategies to combat COVID-19.

2.3 Role of NSP7 in replication-transcription complex (RTC)

NSP7, a 9-kDa protein predominantly composed of alpha-helices, is a crucial component of the SARS-CoV-2 RTC, functioning within the RdRp complex alongside NSP8 and NSP12 to facilitate viral RNA synthesis. Within this complex, NSP12 serves as the catalytic core, while NSP7 and NSP8 enhance NSP12's enzymatic activity (Subong & Ozawa, 2024).

Structurally, NSP7 contains three alpha helices (α1, α2, and α3) arranged in a three-helical coiled-coil bundle, followed by a flexible C-terminal loop (residues 62–70). In certain structural arrangements, this loop is succeeded by a short, less-defined helix (residues 68–72). This structure allows NSP7 to form a stable heterodimer with NSP8, which is essential for the assembly of the RdRp complex. The N-terminal helices (α1 and α2) of NSP8 engage in hydrophobic interactions with four helices of NSP7, with NSP7's loop (residues 66–72) interposing between the NSP8 helices to enhance their interaction surface, largely through hydrophobic contacts (Wilamowski et al., 2021). Figure 2.3 illustrates the crystallized structure of the NSP7-NSP8 heterodimer which reveals that NSP7's (yellow) C-terminal helices are intercalated between NSP8's (purple) long N-terminal α1 helix and its α2 helix.

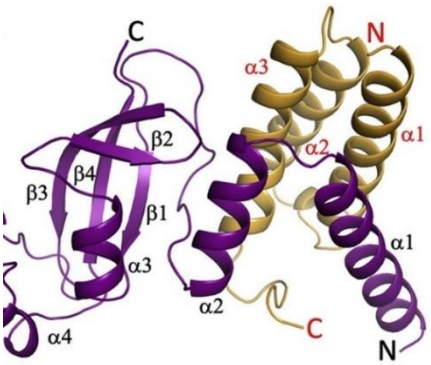

Figure 2. 3: Structure of NSP7-NSP8 complex (Wilamowski et al., 2021).

NSP7 primarily binds to the thumb domain of NSP12 (residues 812–932), while NSP8 interacts with the fingers domain (residues 250–398) of NSP12. These interactions facilitate NSP12's recruitment of NSP8, allowing stable formation of the RdRp complex and ensuring efficient viral RNA synthesis. Notably, the NSP7-NSP8 complex can adopt several conformations, including a tetrameric arrangement, reflecting its adaptability within the replication machinery (Wilamowski et al., 2021). NSP7 promotes the proper configuration of NSP7 for incorporation into the RdRp, contributing to the complex's stability and functionality.

Biophysical studies, such as size-exclusion chromatography with multi-angle light scattering and small-angle X-ray scattering (SEC-MALS-SAXS), have illuminated the specific interactions within the RdRp complex. These analyses show that NSP12 primarily stabilizes NSP8 in the fingers domain, with NSP7 or RNA likely required to anchor NSP8's position within the complex (Wilamowski et al., 2021). Additionally, NSP7's strong affinity for NSP8 and lack of affinity for RNA may facilitate RNA release from NSP8, enabling the replication machinery to progress (Neto et al., 2024).

Unique to SARS-CoV-2, the NSP7-NSP8 complex adopts a heterotetrameric configuration, consisting of a dimer of NSP7-NSP8 dimers. Hence, this leads to a stable 1:2:1 arrangement of the NSP7-NSP8-NSP12 complex, with one NSP7-NSP8 heterodimer and one NSP8 monomer associated with the NSP12 molecule. The NSP7-NSP8 heterotetramer, mediated by conserved oligomeric interfaces, is essential for both dimerization and tetramerization, distinguishing SARS-CoV-2 from the hexadecameric structure seen in SARS-CoV. This heterotetramer is crucial for RTC assembly and viral RNA synthesis (Biswal et al., 2021). Structural integrity is mutually reinforced within the NSP7-NSP8 complex, creating synergistic interactions critical for forming a functional RdRp. While the amino acid sequences of NSP7, NSP8, and NSP12 are highly conserved

across coronaviruses, specific mutations at key interaction sites can impair RdRp activity, underscoring NSP7's role in maintaining RTC stability (Neto et al., 2024).

Moreover, NSP7 contributes to the RdRp's processivity, ensuring continuous RNA strand synthesis needed to produce full-length genomes and subgenomic RNAs for viral replication and protein translation. NSP7 also aids in preserving RTC structure by interacting with other non-structural proteins, such as NSP13, NSP14, NSP15, and NSP16, which coordinate various RNA processing and synthesis functions (Biswal et al., 2021).

Mutations in NSP7, such as F49A, M52A, L56A, the triple mutation F49A/M52A/L56A, C8G, and V11A, have been shown to reduce RdRp activity. Additionally, since the NSP7–NSP8 complex is highly conserved across different coronaviruses, it offers potential as a target for antiviral drug development due to its essential functions and high conservation. Studies have demonstrated that *in vitro* mutations in viral replication complex proteins from other coronaviruses can lead to structural folding defects, impairing their function, or result in delayed viral growth. These findings suggest that NSP7 could be a valuable target for antiviral therapies, as understanding these mutations provides insights into potential therapeutic approaches for SARS-CoV-2 (Subong & Ozawa, 2024).

In this project, a SARS-CoV-2 mutant NSP7 with K51A and S54A mutations was used for protein expression. It is found that most of the residues in the viral replication complex tend to destabilize the structure when mutated to alanine. For instance, the amino acids tryptophan (W), phenylalanine (F), leucine (L), and isoleucine (I) in NSP7 show the greatest destabilizing effects when mutated to alanine, with binding energies of less than −2.0 kcal/mol on average (Subong & Ozawa, 2024). Therefore, replacing lysine and

serine with alanine may disrupt key interactions that are important for maintaining NSP7's stability and its interaction with other components of the replication complex. As mentioned, alanine substitutions often disturb critical interactions, which could lead to a decrease in the stability of NSP7 and its ability to interact properly with NSP8 and NSP12. Hence, by expressing this mutant protein, the project aimed to explore the effects of these mutations on the replication machinery and their potential as targets for therapeutic interventions in SARS-CoV-2.

2.4 Antiviral Drugs Targeting SARS-CoV-2 Replication Machinery

Nucleoside analogues represent a class of antiviral therapeutics that effectively target the RdRp of SARS-CoV-2. It mimics natural nucleosides, the building blocks of RNA, but are chemically modified to interfere with viral RNA synthesis. Among the prominent examples are remdesivir that mimics adenosine, and molnupiravir that mimics cytidine, both approved for COVID-19 treatment.

Remdesivir, an inhibitor of RdRp, received approval from the US FDA in October 2020 for treating hospitalized COVID-19 patients. Its membrane-permeable structure allows it to readily enter the cytoplasm, where it undergoes conversion into remdesivir monophosphate and subsequently into its active form, remdesivir triphosphate. It is preferentially incorporated by SARS-CoV-2 RdRp over its natural analogue, adenosine, without causing immediate pausing of RNA synthesis, unlike classical chain terminators, which abruptly stop elongation (Aiello et al., 2022). Early studies suggested that remdesivir inhibits RNA synthesis through delayed chain termination, where its nitrile group forms a steric clash with NSP12 Ser861 after incorporation at the fourth position from the 3′ end of the RNA. However, this steric inhibition is bypassed under subphysiological concentrations of nucleotide triphosphates (NTPs) *in vitro*, indicating that it may not be the primary inhibitory mechanism in living cells (Malone et al., 2021).

Like remdesivir, molnupiravir is a prodrug that converts into its active triphosphate form within cells, acting as a cytidine analogue. It inhibits SARS-CoV-2 replication through lethal mutagenesis, causing extensive genome mutations that compromise viral viability. It introduces errors into the viral RNA genome by causing mutations during SARS-CoV-2 replication (Malone et al., 2021). When incorporated into RNA, it can pair with adenine (A) or guanine (G), leading to genetic changes mutations. These mutations, such as G>A and C>U transitions, accumulate, impairing the virus's ability to replicate effectively and rendering it non-viable (Tian et al., 2022).

Molnupiravir has shown a high barrier to viral resistance in laboratory studies, meaning it remains effective even when the virus tries to mutate to escape its effects. For example, the SARS-CoV-2 NSP14 protein contains exonuclease (ExoN) activity, which corrects replication errors by removing misincorporated nucleotides. Thus, molnupiravir avoids detection by NSP14-ExoN, allowing it to evade this proofreading mechanism and introduce mutations unchecked. It is unique in being an orally administered drug, which, combined with its potent antiviral activity and resistance profile, makes it a valuable alternative to remdesivir (Malone et al., 2021).

However, the limitations of remdesivir and molnupiravir underscore the need for more effective antiviral treatments for SARS-CoV-2. Remdesivir, for instance, faces challenges in its administration, as it requires intravenous delivery, necessitating hospitalization or specialized medical settings. This limits its accessibility and practicality for widespread use. Additionally, while remdesivir reduces the duration of hospitalization, its overall efficacy in reducing mortality rates remains modest and inconsistent across studies. Its delayed action mechanism, which relies on delayed chain termination, is not always effective under physiological conditions due to the availability of sufficient NTPs that can bypass the steric inhibition caused by remdesivir.

Furthermore, adverse reactions, such as diarrhea, rash, renal impairment, and hypotension, also pose significant concerns for its use (Aiello et al., 2022).

Molnupiravir, on the other hand, raises distinct challenges. Clinical trials have shown only modest benefits in reducing hospitalization or mortality rates, particularly in patients with advanced disease. In a phase-3 study involving 1,433 patients with mild-to-moderate COVID-19 and at least one risk factor for severe illness, molnupiravir treatment within 5 days of symptom onset reduced the risk of hospitalization and death by 30% compared to a placebo. However, this reduction is lower than the results reported for remdesivir or nirmatrelvir/ritonavir in similar clinical trials. Besides, adverse effects such as diarrhea, nausea, dizziness, embryo-fetal toxicity bone and cartilage toxicity, and hypersensitivity, pose significant concerns for its use (Aiello et al., 2022). Therefore, molnupiravir is not authorized for patients under 18 years of age due to concerns about its potential impact on bone and cartilage growth. Additionally, it is not recommended for use in pregnant individuals as it may harm the fetus (Tian et al., 2022).

In a nutshell, the urgency for new antiviral treatments stems from these limitations. There is a need for broad-spectrum antivirals that target a wider range of mechanisms beyond those addressed by remdesivir and molnupiravir. Developing drugs with fewer side effects and lower mutagenic risks is critical for long-term use. Furthermore, new treatments should aim to work synergistically with existing options to improve efficacy and reduce the development of resistance. Addressing these gaps will enhance the ability to manage current and future outbreaks of SARS-CoV-2 effectively.

2.5 Overview of protein expression system

Recombinant proteins are engineered versions of native proteins, produced using recombinant DNA technology to enable the efficient and large-scale synthesis of specific

proteins. This process involves cloning a gene of interest into an expression vector, which is then introduced into an appropriate expression system. These systems facilitate the transcription and translation of the target gene, leading to the synthesis of the desired recombinant protein. Ensuring the availability of pure, high-quality recombinant proteins with consistent batch-to-batch characteristics is essential for reliable outcomes in research, diagnostics, and therapeutic applications.

The choice of expression system is a critical factor when producing recombinant proteins, as each system offers unique features suited to different applications. Factors such as the type of protein, its post-translational modifications, biological function, and required production yield influence this decision. Expression systems, including mammalian, insect, yeast, and bacterial have distinct advantages and limitations. These systems serve as essential tools in biotechnology, enabling the production of proteins for structural analysis, functional studies, and therapeutic development. Understanding the attributes and limitations of each system is vital for optimizing protein expression and ensuring its compatibility with the intended application.

Mammalian cells, such as Human Embryonic Kidney (HEK 293) and Chinese Hamster Ovary (CHO), are extensively utilized for expressing proteins requiring complex post-translational modifications (PTMs) like glycosylation, phosphorylation, and disulfide bond formation. These modifications are critical for the proper folding, stability, and biological activity of many therapeutic proteins, including antibodies and complex enzymes. Transfection methods like polyethyleneimine (PEI) or calcium phosphate are commonly used, with HEK293 achieving high transfection efficiency. The ability of mammalian systems to mimic the native cellular environment makes them indispensable for producing bioactive eukaryotic proteins. Mammalian cell platforms dominate biopharmaceutical production, with many biologics, including monoclonal antibodies,

produced in these systems due to their ability to deliver properly folded and functional proteins. However, these systems are expensive to maintain due to high media costs and extended culture durations. Additionally, achieving high yields often requires optimization of vectors and culture conditions, posing challenges in large-scale applications (Khan, 2013).

Furthermore, insect cells, including Sf9 and Sf21, represent a cost-effective alternative for expressing eukaryotic proteins, particularly those requiring PTMs and high molecular weights. The baculovirus expression vector system (BEVS) is widely employed in insect cell-based expression, facilitating the production of large quantities of proteins in a relatively short time. The baculovirus–insect cell system is widely used to produce virus-like particles (VLPs) for studying viral assembly, creating antigens for vaccines, and producing proteins for diagnostics. This system is particularly valuable for viruses like HPV and HCV, where it enables the delivery of viral structural proteins either through coinfection with multiple baculoviruses or a single virus expressing multiple proteins. A notable application is the development of HPV VLP-based vaccines, which have proven effective in clinical trials at preventing cervical infections and related abnormalities caused by HPV-16 and HPV-18 (Kost et al., 2005). These systems strike a balance between the advanced PTM capabilities of mammalian cells and the affordability of bacterial systems. However, the PTMs in insect cells can differ from those in mammalian systems, potentially impacting therapeutic protein functionality. Moreover, the development of recombinant baculovirus constructs is time-intensive and laborious, necessitating specialized expertise (Stolt-Bergner et al., 2018).

In addition, yeast cells also offer a robust platform for recombinant protein production. They combine the simplicity of bacterial systems with the ability to perform certain eukaryotic-like PTMs, making them suitable for the production of enzymes,