T.C.

INONU UNIVERSITY
INSTITUTE OF SCIENCES

CLONING AND EXPRESSION OF SARS-CoV-2 SPIKE GENE IN BACTERIA AND YEAST

MASTER THESIS

Okan ÖZŞAHİN

Department of Molecular Biology and Genetics

Thesis Advisor: Prof. Dr. Hikmet GEÇKİL

AUGUST 2023

T.C.

INONU UNIVERSITY
INSTITUTE OF SCIENCES

CLONING AND EXPRESSION OF SARS-CoV-2 SPIKE GENE IN BACTERIA AND YEAST

MASTER THESIS

Okan ÖZŞAHİN
(36193623003)

Department of Molecular Biology and Genetics

Thesis Advisor: Prof. Dr. Hikmet GEÇKİL

AUGUST 2023

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Prof. Dr. Hikmet GEÇKİL for his encouragement, valuable advice, guidance and insights. I am also grateful to my family for their continuous support. I would be remiss in not mentioning my friends Muhammed DÜNDAR (Research Assistant), Enes GÜLTEKİN (PhD student) and Rüveyda SEVER (MSc student) who have contributed to my work with their feedback and moral support. I would like also to thank İrem EKİCİ and Işın KORKMAZ, two undergraduate students from Department of Molecular Biology and Genetics at Istanbul Medeniyet University and Dicle University, respectively, for helping me with my experiments during their 2022 summer internship. I am also grateful to Scientific Research Project (BAP) Unit at İnönü University for their financial support (project code FYL-2021-2761 and 2761).

WORD OF HONOR

I declare this **" Cloning and Expression of SARS-CoV-2 Spike Gene in Bacteria and Yeast"**, which I submitted as my master thesis, was written by myself without resorting to any help that would otherwise contradict scientific morals and traditions, and that all the sources I benefited from were shown in accordance with the required methodology both in the text and in the bibliography and I proudly confirm this.

Okan Özşahin

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

DNA	**:** Deoxyribonucleic acid
RNA	**:** Ribonucleic acid
EtBr	**:** Ethidium bromide
TAE	**:** Tris-acetate-EDTA
TE	**:** Tris EDTA
kDa	**:** kiloDalton
bp	**:** Base pair
UPW	**:** Ultrapure water
DDT	**:** 1,4-Dithiothreitol
RNase	**:** Ribonuclease
NaCl	**:** Sodium chloride
KCl	**:** Potassium chloride
$MgCl_2$	**:** Magnesium chloride
TEMED	**:** Tetramethyl ethylenediamine
ddH_2O	**:** Double-distilled water
SDS-PAGE	**:** Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
Acryl/Bis	**:** Acrylamide/Bis-Acrylamide
SDS	**:** Sodium dodecyl sulphate
YNB	**:** Yeast Nitrogen Base
YPD	**:** Yeast Extract Peptone Dextrose
Leu	**:** Leucine
His	**:** Histidine
Met	**:** Methionine
Ura	**:** Uracil
LiOAc-TE	**:** Lithium acetate - Tris EDTA
PEG	**:** Polyethylene Glycol
LB	**:** Lysogeny broth
LB amp	**:** Lysogeny broth ampicillin
IPTC	**:** Isopropyl-beta-D-thioacetamide
NAOH	: Sodium hydroxide
HCI	**:** Hydrochloric acid
APS	**:** Ammonium persulphate

ÖZET

Yüksek Lisans Tezi

SARS-COV-2 Spike Geninin Bakteri ve Mayalara Klonlanması ve İfadesi
Okan ÖZŞAHİN

İnönü Üniversitesi
Fen Bilimleri Enstitüsü
Moleküler Biyoloji ve Genetik Anabilim Dalı

57+ IX sayfa

2023

Danışman: Prof. Dr. Hikmet GEÇKİL

Şiddetli Akut Solunum Sendromu Koronavirüs 2 (SARS-CoV-2), 2019 yılında Çin'in Wuhan şehrinde ortaya çıkan, koronavirüs hastalığına (COVID-19) neden olan, zarflı, pozitif anlamlı, tek sarmallı bir RNA virüsüdür. Virüs partikülleri bir RNA (mRNA), yapısal proteinler, yapısal olmayan proteinler (NSP'ler) ve yardımcı proteinler içerir. Virüs, konakçı hücrelerin yüzeyindeki reseptörler aracılığıyla hücreye girdikten sonra 4 yapısal protein; Spike, Membran, Nükleokapsid, Zarf (S, M, N, E), yaklaşık 20 yapısal olmayan protein (NSP'ler) ve bir düzine yardımcı protein (AP'ler) yapılır. Virüsün girişi, partikül oluşması ve bağışıklıktan kaçması için yapısal proteinler gerekliyken, NSP'ler ve AP'ler transkripsiyon, replikasyon ve konak hücre kontrolünde önemlidir. Virüsün en büyük protein olan Spike glikoproteini viral bağlanma ve konakçıya girişe aracılık ettiğinden aşı geliştirme, monoklonal antikor tedavileri ve antijen bazlı veya PCR tanı testleri için önemli bir hedef olmuştur. Bu çalışmada Spike geni hem bakteri hem de mayada replikasyon ve ekspresyon yapabilen bir mekik vektörüne klonlanmış ve proteinin ifade profili araştırılmıştır. Çalışmamız hem bakteri hem de mayanın Spike proteinini ifade ettiğini ve rekombinant mekik vektörünün her iki hücre tipinde de kararlı bir şekilde muhafaza edildiğini gösterdi.

Anahtar Kelimeler*:* SARS-CoV-2, COVID-19, Koronavirüsler, Spike proteini, Mekik vektörler*, Escherichia coli, Sacharomyces cerevisiae.*

ABSTRACT

Master Thesis

Cloning and Expression of SARS-CoV-2 Spike Gene in Bacteria and Fungi
Okan ÖZŞAHİN

Inonu University
Graduate School of Nature and Applied Sciences
Department of Molecular Biology and Genetics

57 + IX pages
2023

Supervisor: Prof. Dr. Hikmet GEÇKİL

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense, single-stranded RNA virus that causes coronavirus disease (COVID-19) that emerged in 2019 in Wuhan city of China. Virus particles contain an RNA (mRNA), structural proteins, non-structural proteins (NSPs), and accessory proteins. After the virus enters the cell via receptors on the surface of host cells, 4 structural proteins; Spike, Membrane, Nucleocapsid, Envelope (S, M, N, E), about 20 non-structural proteins (NSPs), and a dozen accessory proteins (APs) are made. While structural proteins are required for virus entry, assembly, and immune evasion, NSPs and APs are important in transcription, replication, and host cell control. Since the Spike glycoprotein, the largest protein of the virus, mediates viral attachment and entry into the host, it has been an important target for vaccine development, monoclonal antibody therapies, and antigen-based or PCR diagnostic tests. In this study, the Spike gene was cloned into a shuttle vector that can replicate and express in both bacteria and yeast, and the expression profile of the protein was investigated. Our study showed that both bacteria and yeast expressed the Spike protein and the recombinant shuttle vector was stably maintained in both cell types.

Keywords: SARS-CoV-2, COVID-19, Coronaviruses, Spike protein, Shuttle vectors, *Escherichia coli, Saccharomyces cerevisiae.*

1. INTRODUCTION

1.1 Pandemics and Epidemics

Pandemics are global epidemics of viral and bacterial origin, affecting many populations and covering many countries and continents around the world (Bhadoria, Gupta & Agarwal 2021). Etymologically, pandemic is a concept of Ancient Greek origin in 8th century BCE, the modern usage of the term in referral to infectious diseases dates back only to 19th and 20th centuries as the field of epidemiology evolved. It was formed by combining the words "pan" meaning “everyone” or “everywhere” and "demos" meaning people (Qiu, Rutherford, Mao & Chu 2017).

Pandemics have occurred countless times in the past, infecting, injuring, and killing millions of people. Many of agents of disease such as plague, cholera, smallpox, flu changed the life and the course of human history. While the concept of pandemics as we know today did not exist in prehistoric times, archaeological and anthropological evidence suggest the occurrence of pathogens with pandemic potential for diseases such as tuberculosis, plague (a.k.a the Black Death), and smallpox and their significant impact on prehistoric human populations. Pandemics have not only caused a great deal of human suffering, they also have caused social and political disturbance, deterioration of economies, serious demographic shifts and human migration. Thus, pandemics have been a part of human culture and life since the beginning of written history. After all, while we modern humans (*Homo sapiens*) have only been around for 200-300 thousand years bacteria and viruses have existed for 2-3 for billion years.

The term pandemic particularly gained prominence during the Spanish flu, a global influenza pandemic in 1918-1919 that caused millions of deaths. Based on specific criteria related to

the geographic spread and impact of a disease, the declaration of an outbreak as pandemic by the World Health Organization (WHO) has been influential in shaping how we understand and respond to pandemics in the modern era.

Several pandemics and epidemics have also occurred in the 21st century. H1N1 Influenza (2009-2010), often referred to as the swine flu spread globally and resulted in millions of cases and thousands of deaths. Also Ebola outbreak (2014-2016) in West Africa and Democratic Republic of Congo and Zika outbreak (2015-2016) primarily in the Americas (Brazil, Colombia, Venezuela, Mexico, Puerto Rico) happened in this century. Cholera outbreaks have been occurring in various parts of the world, including Haiti, Yemen, and East Africa. Several countries have experienced measles, a long forgotten disease, due to recent vaccine hesitancy and declining vaccinations. Dengue, a mosquito-borne viral disease like Zika, has seen periodic epidemics in various regions, including Southeast Asia and Latin America. Lassa fever outbreaks have occurred in West Africa, particularly in Nigeria, leading to significant morbidity and mortality. Although, the first cases of Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) cases were reported in the early 1980s and it has been an ongoing global health concern with millions of people affected worldwide.

Also, in the last 20 years the World experienced 3 significant outbreaks caused by newly emerging coronaviruses: Severe Acute Respiratory Syndrome (SARS) virus which emerged in China in 2002 (Xu, et al., 2004), Middle East Respiratory Syndrome (MERS) virus which emerged in the Middle East in 2012 (Alsahafi & Cheng, 2016). and COVID-19, the disease, caused by SARS-CoV-2 (Wu et al., 2020) which emerged at the end of 2019 in Wuhan, China. While the former two led to epidemics, the latter was officially declared as a pandemic by WHO and has had a profound impact by causing a global health crisis which killed millions and devastated economies.

Coronaviruses (*Coronaviridae* in Latin) are a family of RNA viruses from *Nidovirales* order. These viruses cause respiratory tract infections ranging from mild to lethal in mammals and birds. Estimates of the emergence of the first coronavirus suggest a wide time period ranging from 10,000 years to 150 million years ago (Hayman & Knox, 2021). To date, there are seven coronaviruses that infect humans. Four of them (229E, NL63, OC43, and HKU1) cause mild cases of the common cold. Indeed about 20% of seasonal common cold is caused

by these coronaviruses, the second highest after rhinoviruses. Three coronaviruses (SARS-CoV-1, MERS, and SARS-CoV-2), with epidemic and pandemic potential, are newly emerging viruses in the last quarter century (Pollett et al., 2021).

1.1.2 SARS and MERS Epidemics

The first SARS epidemic occurred between November 2002 and July 2003. It was caused by the SARS-CoV-1 virus, a closely related virus to the current SARS-CoV-2. The SARS outbreak started in the Guangdong province of China in 2002 and spread to some other countries through international travel. With a case fatality rate (CFR) 10%, SARS (SARS-CoV-1) caused 8,098 confirmed cases and 774 deaths recorded worldwide (Zizza et al., 2021).

The MERS epidemic began in 2012. The disease is caused by the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). The first case was reported in Saudi Arabia, and the virus has since been identified in several other countries in the Middle East and beyond. The epidemic has been characterized by sporadic outbreaks and has resulted in cases in various countries, with most cases reported in the Arabian Peninsula (Rabaan et al., 2020). MERS-CoV is believed to have originated in bats and then transmitted to camels before infecting humans. The transmission to humans typically occurs through close contact with infected camels or through human-to-human transmission in healthcare settings. MERS-CoV causes severe respiratory illness, with a higher CFR (%35) compared to other coronaviruses (Huang et al., 2020).

1.1.3 SARS CoV-2 pandemic

The COVID-19 pandemic, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), began in December 2019 in Wuhan, China (Wu et al., 2020). It quickly spread globally, leading the WHO to declare it as a pandemic in March 2020. COVID-19 is characterized by a wide range of symptoms, from mild respiratory illness to severe pneumonia and acute respiratory distress syndrome (ARDS). With a 0.5% CFR, it can cause various non-respiratory symptoms and has been associated with long-term effects, named “long COVID” in some individuals. The pandemic has resulted in millions of cases and deaths worldwide, significantly impacting public health, economies, and societies (Fontanet et al., 2021; Wiersinga, Rhodes, Cheng, Peacock, & Prescott, 2020). As of October 20 2023, there were about 700 million COVID-19 cases and 7 million deaths worldwide (COVID-19 Dashboard at Johns Hopkins University).

The exact origin of "SARS-CoV-2'', the virus responsible for the COVID-19 pandemic, is still under investigation. However, the prevailing theory is that the virus is of zoonotic origin, meaning it likely originated in animals and then jumped to humans (Andersen et al., 2020). Many early cases were linked to a seafood market in the city of Wuhan, China. However, further research has suggested that the market may have played a role in amplifying and spreading the virus rather than being the original source (Xiaolu et al., 2020). Bats are considered the be the natural reservoir of coronaviruses and it is suspected that the virus might have been transmitted from bats to another animal species (still unknown), which in turn transmitted the virus to humans (Zhang et al., 2020). There is a need for extensive research and collaboration, including examining the original viral genetic sequences and investigating early cases and possible animal hosts, to determine the exact origin of the virus.

1.2 Coronaviruses

Coronaviruses are single-stranded positive (+) sense RNA (mRNA) viruses belonging to the *Coronaviridae* family (Leao et al., 2022). These viruses have the largest genomes (~30 kbp) among RNA viruses. The genome has a 3'-polyA tail, two large overlapping Open Reading Frames (ORFs) at the 5' end and encodes major nonstructural proteins such as protease, helicase, nucleocapsid protein, and RNA dependent RNA (RdRp) polymerase expressed by ribosomal frame shifting.

Coronaviridae can infect a variety of animals, particularly birds and mammals. Coronaviruses are viruses that have been circulating in the human population and animals for a long time. To date, there are seven human coronaviruses belonging to either alpha coronaviruses (NL63 and 229E) or beta coronaviruses (OC43, HKU1, SARS or sometimes named as SARS-CoV-1, MERS and SARS-CoV-2). This family of viruses has their own unique morphological features. Virions are enveloped and spherical. There are envelope glycoproteins embedded in each virion and protruding from the virion. These glycoproteins surround the virion and are in the form of a fringe or "corona", thus the name coronaviruses (Burrell et al., 2017; Hu et al., 2021).

1.2.1. Classification of coronaviruses

Coronaviruses are classified based on their genetic and antigenic properties. The family *Coronaviridae* is divided into four genera (Figure 1.1): alpha (α) coronaviruses, beta (β) coronaviruses, gamma (γ) coronaviruses, and delta (δ)coronaviruses.

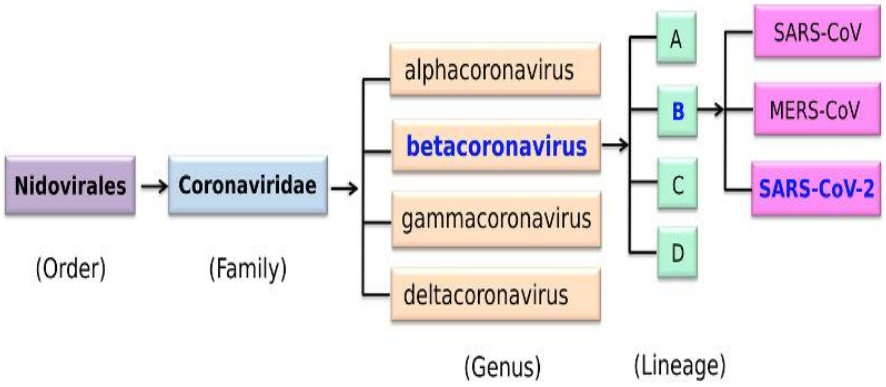

Figure 1.1: The Taxonomy of Coronaviruses (Santos-Sánchez et al., 2020).

1.2.1.1 Alphacoronaviruses (α-CoVs)

Like other coronaviruses, α-CoVs (represented by human coronaviruses NL63 and 229E) have spherical virions with a viral envelope and stick-shaped surface projections formed by trimers of the Spike (S) protein. Both α-CoVs and β-CoVs come from the bat viral gene pool (Cui et al., 2029).

1.2.1.2 Betacoronaviruses (β-CoVs)

The latest pandemic virus SARS-CoV-2 is a member of betacoronaviruses which have other four other known human coronaviruses (OC43, HKU1, SARS an MERS). The β-CoV virions are also enveloped and spherical, 120-160 nm in diameter in size with a core shell of about 65 nm. Homotrimers of S protein form large surface protrusions that create the appearance of the solar corona, thus the name coronavirus (Hu et al., 2021).

1.2.1.3 Gammacoronaviruses (γ-CoVs)

This genus primarily infects birds, and there are no known human coronaviruses in this group (Jackwood et al., 2012)

1.2.1.4 Deltacoronaviruses (δ-CoVs)

This genus includes coronaviruses that primarily infect birds, but some strains have been found in pigs and other animals (Cui et al., 2019).

1.3 SARS-CoV-2

The largest pandemic of the 21st century, still ongoing, has been caused by an emerging virus named as SARS-CoV-2. Like other coronaviruses, the virus is an enveloped virus containing a single-stranded, positive (+) sense RNA (mRNA) genome. The "International Virus Taxonomy Committee" (ICTV), which oversees virus taxonomy, has classified the virus as a strain of Severe Acute Respiratory Syndrome viruses from the family *Riboviria*, the order

Nidovirales, the suborder *Cornidovirineae*, the family *Coronoviridae*, genus, Betacoronaviruses, subgenus Sarbecoviruses (Figure 1.1) (Santos-Sánchez et al., 2020). We know more or less the morphological structures of coronavirus particles examined by cryo-electron microscopy. These viruses have a large (approximately 30,000 nucleotides), positive-strand (sense) RNA genome (Wu et al., 2020). This means that as soon as the genetic material (the genome or mRNA) of SARS-CoV-2 enters the cell, it can be directly read by the ribosomes in the cell and viral proteins can be produced. Moreover, we will see that new mRNAs are produced from this mRNA by the virus' own enzymes (as we will see later, the replicase complex consisting of several different viral proteins, including RdRp. SARS-CoV-2, like other coronaviruses, has a Spike (S) protein composed of a type I fusion machine with and envelope anchoring S2 domain and receptor binding S1 domain. The S protein forms a homotrimer, a heavily glycosylated protein complex appending from the surface of the virus or virions. Data obtained from cryo-electron microscopy structure show that this protein consists of two overlapping subunits (S1/S2). The upper rough part is the S1 subunit, the lower part is the S2 subunit embedded in the envelope. Three of complete identical Spike proteins (homotrimer) come together to form a "closed" prefusion complex. Its lower domain (i.e., S2) is involved in fusion with the cell and has a relatively conserved structure. This domain helps the virus genome (mRNA) to enter to the cell by promoting the fusion of the virus' envelope and the cell's membrane. The upper part (i.e., S1) contains the receptor binding domain (RBD), which binds when it finds the appropriate receptor (Angiotensin Converting Enzyme-2 receptor, ACE-2) in our cells, and has a highly mutation prone structure. This region is also the target of all vaccines. When the RBD of the S1 binds to ACE-2, the whole Spike protein (S) is processed with host proteases such as TMPRSS2 and furin and the closed perfusion protein complex turns into an active fusion open complex that helps the fusion of viral envelope and cell membrane for the entry of viral genome into the cell cytoplasm (Wrapp et al. 2020; Mobini et al., 2021; Rahimi et al. 2021; Velusamy et al., 2021; Nassaret al. 2021).

Compared to SARS (a.k.a. SARS-CoV-1), SARS-CoV-2 has sequence made of 6 different amino acids in the RBD, corresponding to 18 nucleotides (Wu et al., 2020; Wrapp et al. 2020). However, this mutated RBD can still bind to the ACE2 receptor effectively. The second, and perhaps the most important, difference is that SARS-CoV-2 has a polybasic “furin cleavage site” consisting of basic amino acids. This region has not been identified so far in any member of Sarbecoviruses. There are about 14 ORFs on the SARS-CoV-2 genome

which encodes about 30 proteins with structural, non-structural, and accessory functions. The single mRNA genome with 14 ORFs, encoding approximately 30 proteins is truly a remarkable characteristic of the virus, given that most mRNAs in humans are single ORFs coding for a single protein. The large part of the genome contains a single ORF (pp1a/pp1ab) between 1-7096 nucleotides located at the 5' end of the genome, which is the end of the genome that first binds to the ribosome and encodes a large protein without stop codons. Then, this large protein is converted by proteolytic cleavage into about 16 nonstructural (NSP) proteins which are also made by a process called "programmed ribosome frameshifting" (Dinman, 2006; Yan et al., 2020; Guo et al., 2021).

There are 4 structural genes genes at the 3' end of the genome that code for structural proteins; spike (S), matrix (M), envelope (E), and nucleopcapsid (N). These proteins are produced small subgenomic mRNAs with a completely different strategy called "discontinuous transcription".

Although, all these proteins (structural or NSPs) can be made from the original (+) sense mRNA genome of the virus entering the cells, the same mRNA genome can first be copied to a (-) non-sense strand and then be re-copied to (+) sense mRNA (replication) needed to make more viral proteins for more virions.

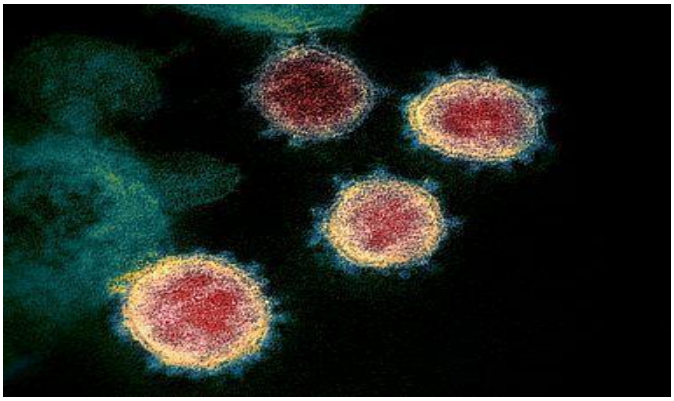

Figure 1.2: Transmission electron microscope image shows SARS-CoV-2, the virus that causes COVID-19, isolated from a patient in the U.S. Virus particles are emerging from the surface of cells cultured in the lab. The spikes on the outer edge of the virus particles give coronaviruses their name, crown-like. *NIAID-RML.*

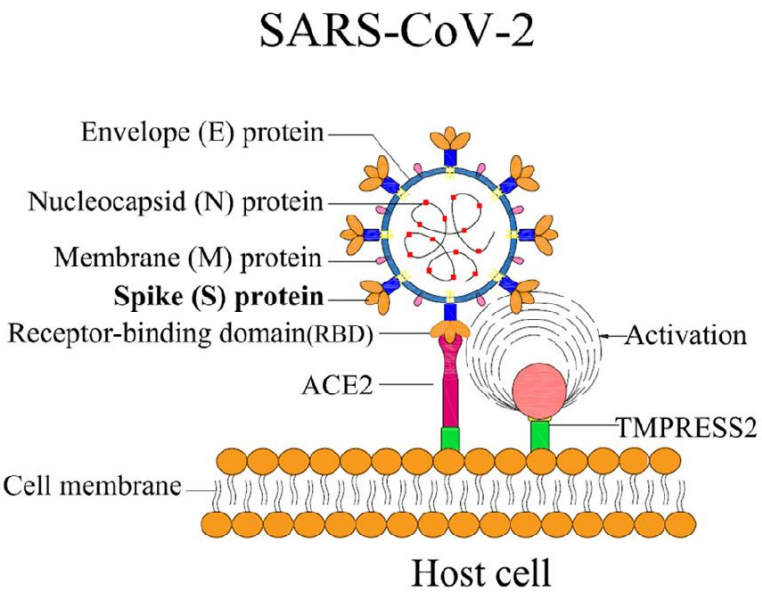

Figure 1.3: General structure of SARS-CoV-2 S-protein and its interaction with ACE-2 receptor on the host cell (Velusamy et al., 2021).

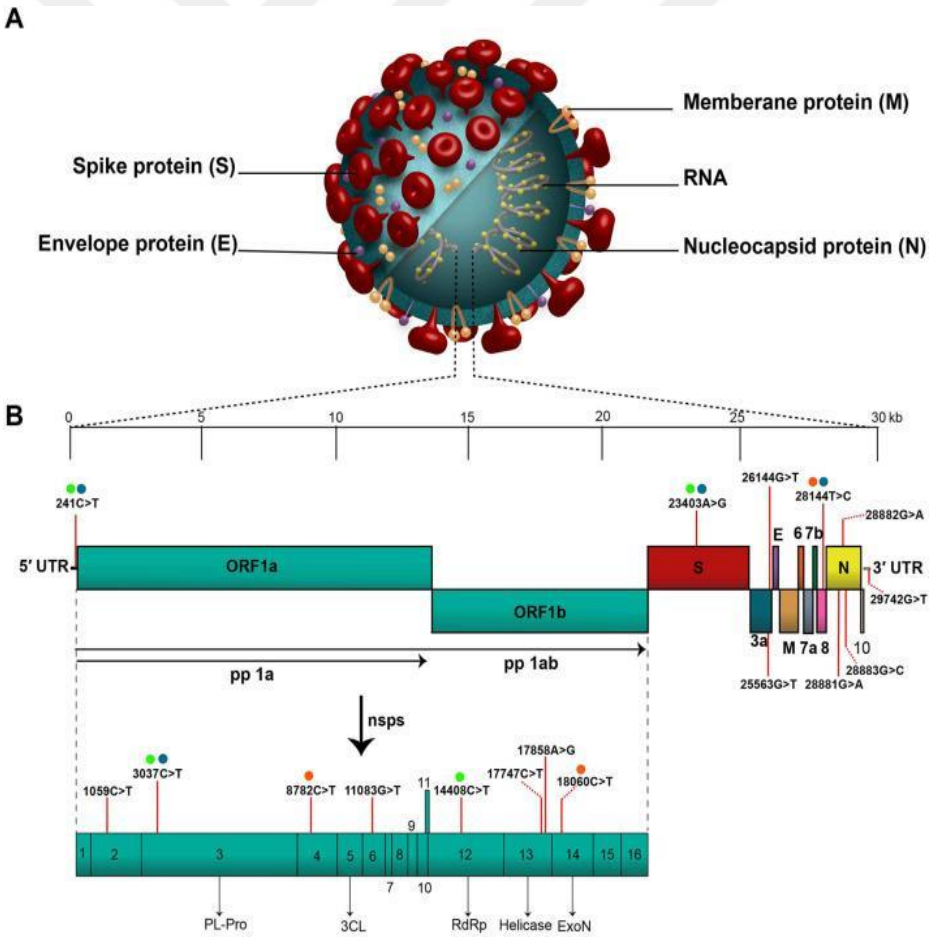

Figure 1.4: General structure of SARS-CoV-2 (A) and its genomic organization (B) (Rahimi et al. 2021).

5′

attaaaggtttataccttcccaggtaacaaaccaaccaactttcgatctcttgtagatctgttctctaaacgaactttaaaatctgtgtggctgtcactcggctgcatgcttagtgcactcacgc
agtataattaataactaattactgtcgttgacaggacacgagtaactcgtctatcttctgcaggctgcttacggtttcgtccgtgttgcagccgatcatcagcacatctaggtttcgtccgggtg
tgaccgaaaggtaagatggagagccttgtccctggtttcaacgagaaaacacacgtccaactcagtttgcctgttttacaggttcgcgacgtgctcgtacgtggctttggagactccgtggagga
ggtcttatcagaggcacgtcaacatcttaaagatggcacttgtggcttagtagaagttgaaaaaggcgttttgcctcaacttgaacagccctatgtgttcatcaaacgttcggatgctcgaactg
cacctcatggtcatgttatggttgagctggtagcagaactcgaaggcattcagtacggtcgtagtggtgagacacttggtgtccttgtccctcatgtgggcgaaataccagtggcttaccgcaag
gttcttcttcgtaagaacggtaataaaggagctggtggccatagttacggcgccgatctaaagtcatttgacttaggcgacgagcttggcactgatccttatgaagatttcaagaaaactggaaa
cactaaacatagcagtggtgtttacccgtgaactcatgcgtgagcttaacggagggggcatacactcgctatgtcgataacaacttctgtggccctgatggctaccctcttgagtgcattaaagacc
ttctagcacgtgctggtaaagcttcatgcactttgtccgaacaactggactttattgacactaagaggggtgtatactgctgccgtgaacatgagcatgaaattgcttggtacacggaacgttct
gaaaagagctatgaattgcagacaccttttgaaattaaattggcaaagaaatttgacaccttcaatggggaatgtccaaattttgtatttccctt aaattccataatcaagactattcaaccaag
ggttgaaaagaaaaagcttgatggctttatgggtagaattcgatctgtctatccagttgcgtcaccaaatgaatgcaaccaaatgtgccttt caactctcatgaagtgtgatcattgtggtgaaa
cttcatggcagacgggcgattttgttaaagccacttgcgaattttgtggcactgagaatttgactaaagaaggtgccactacttgtggttacttaccccaaaatgctgttgttaaaatttattgt
ccagcatgtcacaattcagaagtaggacctgagcatagtcttgccgaataccataatgaatctggcttgaaaaccattcttcgtaagggtggtcgcactattgcctttggaggctgtgtgttctc
ttatgttggttgccataacaagtgtgcctattgggttccacgtgctagcgctaacatagg ttgtaaccatacaggtgttgttggagaaggttccgaaggtcttaatgacaacccttcttgaaatac
tccaaaaagagaaagtcaacatcaatattgttggtgacttaaacttaatgaagagatcgccattattttggcatctttttctgcttccacaagtgcttttgtggaaactgtgaaaggttt ggat
tataaagcattcaaacaaattgttgaatcctgtggtaattttaaagttacaaaaggaaaagctaaaaaaggtgcctggaatattggtgaacagaaatcaatactgagtcctctttatgcatttgc
atcagaggctgctcgtgttgtacgatcaatttttctcccgcactcttgaaactgctcaaaattctgtgcgtgttttacagaaggccgctataacaatactagatggaatttcacagtatt cactga
gactcattgatgctatgatgttcacatctgatttggctactaacaatctagttgtaatggcctacattacaggtggtgttgttcagttgacttcgcagtggctaactaacatctttggcactgtt
tatgaaaaactcaaacccgtccttgattggcttgaagagaagtttaaggaaggtgtagagtttcttagagacggttgggaaattgttaaatttatctcaacctgtgcttgtgaaattgtcggtgg
acaaattgtcacctgtgcaaaggaaattaaggagagtgttcagacattctttaagcttgtaaataaatttttggctttgtgtgctgactctatcattattggtggagctaaacttaaagccttga
atttaggtgaaacatttgtcacgcactcaaagggattgtacagaaagtgtgttaaatccagagaagaaactggcctactcatgcctctaaaagccccaaaagaaattatcttcttagagggagaa
acacttcccacagaagtgttaacagaggaagttgtcttgaaaactggtgatttacaaccattagaacaacctactagtgaagctgttgaagctccattggttggtacaccagtttgtattaacgg
gcttatgttgctcgaaatcaaagacacagaaaagtactgtgcccttgcacctaatatgatggtaacaaacaataccttcacactcaaaggcggtgcaccaacaaaggttacttttggtgatgaca
ctgtgatagaagtgcaaggttacaagagtgtgaatatcacttttgaacttgatgaaaggattgataaagtacttaatgagaagtgctctgcctatacagttgaactcggtacagaagtaaatgag
ttcgcctgtgttgtggcagatgctgtcataaaaactttgcaaccagtatctgaattacttacaccactgggcattgatttagatgagtggagtatggctacatactacttatttgatgagtctgg
tgagtttaaattggcttcacatatgtattgttctttctaccctccagatgaggatgaagaagaaggtgattgtgaagaagaagagtttgagccatcaactcaatatgagtatggtactgaagatg
attaccaaggtaaacctttggaatttggtgccacttctgctgctcttcaacctgaagaagagcaagaagaagattggttagatgatgatagtcaacaaactgttggtcaacaagacggcagtgag
gacaatcagacaactactattcaaacaattgttgaggttcaacctcaattagagatggaacttacaccagttgttcagactattgaagtgaatagttttagtggttatttaaaacttactgacaa
tgtatacattaaaaatgcagacattggtggaagaagctaaaaaggtaaaaccaacagtggttgttaatgcagccaatgtttaccttaaacatggaggaggtgttgcaggagccttaaataaggcta
ctaacaatgccatgcaagttgaatctgatgattacatagctactaatggaccacttaaagtgggtggtagttgtgtttaagcggacacaatcttgctaaacactgtcttcatgttgtcggccca
aatgttaacaaaggtgaagacattcaacttcttaagagtgcttatgaaaattttaatcagcacgaagttctacttgcaccattattatcagctggtatttttggtgctgaccctatacattcttt
aagagtttgtgtagatactgttcgcacaaatgtctacttagctgtctttgataaaaatctctatgacaaacttgtttcaagcttttggaaatgaagagtgaaaagcaagttgaacaaaagatcg
ctgagattcctaaagaggaagttaagccatttataactgaaagtaaaccttcagttgaacagagaaaacaagatgataagaaaatcaaagcttgtgttgaagaagttacaacaactctggaagaa
actaagttcctcacagaaaacttgttacttttatattgacattaatggcaatcttcatccagattctgccactcttgttagtgacattgacatcacttttcttaaagaaagatgctccatatatagt
gggtgatgttgttcaagagggtgttttaactgctgtggttatacctactaaaaaggctggtggcactactgaaatgctagcgaaagctttgagaaaagtgccaacagacaattatataaccactt
acccgggtcagggttaaatggttacactgtagaggaggcaaagacagtgcttaaaaagtgtaaaagtgccttttacattctaccatctattatctctaatgagaagcaagaaattcttggaact
gtttcttggaatttgcgagaaatgcttgcacatgcagaagaaacacgcaaattaatgcctgtctgtgtggaaactaaagccatagtttcaactatacagcgtaaatataagggtattaaaatca
agagggtgtggttgattatggtgctagatttacttttacaccagtaaaacaactgtagcgtcacttatcaacacacttaacgatctaaatgaaactcttgttacaatgccacttggctatgtaa
cacatggcttaaatttggaagaagctgctcggtatatgagatctctcaaagtgccagctacagtttctgtttcttcacctgatgctgttacagcgtataatggttatcttacttcttcttctaaa
acacctgaagaacattttattgaaaccatctcacttgctggttcctataaagattggtcctattctggacaatctacacaactaggtatagaatttcttaagagaggtgataaaagtgtatatta
cactagtaatcctaccacattccacctagatggtgaagttatcacctttgacaatcttaagacacttctttctttgagagaagtgaggactattaaggtgtttacaacagtagacaacattaacc
tccacacgcaagttgtggacatgtcaatgacatatggacaacagtttggtccaacttatttggatggagctgatgttactaaaataaaacctcataattcacatgaaggtaaaacatttttatgtt
ttacctaatgatgacactctacgtgttgaggcttttgagtactaccacacaactgatcctagttttctgggtaggtacatgtcagcattaaatcacactaaaaagtggaaatacccacaagttaa
tggtttaacttctattaaatgggcagataacaactgttatcttgccactgcattgttaacactccaacaaatagagttgaagtttaatccacctgctctacaagatgcttattacagagcaaggg
ctggtgaagctgctaactttttgtgcacttatcttagcctactgtaataagacagtaggtgagttaggtgatgttagagaaacaatgagttacttgtttcaacatgccaatttagattcttgcaaa
agagtcttgaacgtggtgtgtaaaacttgtggacaacagcagacaaccctaaggtgtgagaagctgttatgtacatgggcacacttttcttatgaacaatttaagaaaggtgttcagataccttg
tacgtgtggtaaacaagctacaaaatatctagtacaacaggagtcaccttttgttatgatgtcagcaccacctgctcagtatgaacttaagcatggtacatttacttgtgctagtgagtacactg
gtaattaccagtgtggtcactataaacatataacttctaaagaaacttttgtattgcatagacggtgctttacttacaaagtcctcagaatacaaaggtcctattacggatgttttctacaaagaa
aacagttacacaacaaccatataaaccagttacttatataaattggatggtgttgtttgtacagaaattgaccctaagttggacaattattataagaaagacaattcttatttcacagagcaaccaat
tgatcttgtaccaaaccaaccatatccaaacgcaagcttcgataatttaagtttgtatgtgataatatcaaatttgctgatgatttaaaccagttaactggttataagaaacctgcttcaagag
agcttaaagttacatttttccctgacttaaatggtgatgtggtggctattgattataaacactacacaccctcttttaagaaaggagctaaattgttacataaacctattgtttggcatgttaac
aatgcaactaataaagccacgtataaaccaaatacctggtgtatacgttgtctttggagcacaaaaccagttgaaacatcaaattcgtttgatgtactgaagtcagaggacgcgcagggaatgga
taatcttgcctgcgaagatctaaaaccagtctctgaagaagtagtggaaaatcctaccatacagaaagacgttcttgagtgtaatgtgaaaactaccgaagttgtaggagacattatacttaaac
cagcaaataatagttaaaaattacagaagaggttggccacacagatctaatggctgcttatgtagacaattctagtcttactattaagaaacctaatgaattatctagagattaggttgaaa
acccttgctactcatggtttagctgctgttaatagtgtccctgggatactatagctaattatgctaagcctttttcttaacaaagttgttagtacaactactaacatagttacacggtgtttaaa
ccgtgtttgtactaattatatgccttatttctttacttattgctacaattgtgtactttactagaagtacaaatctagaattaaagcatctatgccgactactatagcaaagaatactgtta
agagtgtcggtaaattttgtctagaggcttcatttaattatttgaagtcacctaatttttctaaactgataaatattataatttggttttactattaagtgtttgcctaggttctttaatctac
tcaaccgctgctttaggtgttttaatgtctaatttaggcatgccttcttactgtactggttacagagaaggctatttgaactctactaatgtcactattgcaacctactgtactggttctatacc
ttgtagtgtttgtcttagtggtttagattctttagacacctatccttctttagaaactatacaaattaccatttcatctttaaatgggatttaactgcttttggcttagttgcagagtggtttt
tggcatatattcttttcactaggtttttctatgtacttggattggctgcaatcatgcaattgttttcagctattttgcagtacatttttattagtaattcttggcttatgtggttaataattaat
cttgtacaaatggccccgatttcagctatggttagaatgtacatcttctttgcatcattttattatgtatggaaaagttatgtgcatgttgtagacggttgtaattcatcaacttgtatgatgtg
ttacaaacgtaatagagcaacaagagtcgaatgtacaactattgttaatggtgttagaaggtccttttatgtctatgctaatggaggtaaaggctttgcaaactacacaattggaattgtgtta
attgtgatacattctgtgctggtagtacatttattagtgatgaagttgcgagagacttgtcactacagtttaaaagaccaataaatcctactgaccagtcttcttacatcgttgatagtgttaca
gtgaagaatggttccatccatctttacttttgataaagctggtcaaaagactttatgaaagacattctctctctcattttgttaacttagacaacctgagagctaataacactaaaggttcattgcc
tattaatgttatagtttttgatggtaaatcaaaatgtgaagaatcatctgcaaaatcagcgtctgtttactacagtcagcttatgtgtcaacctatactgttactagatcaggcattagtgtctg

5'UTR 1 265 arrow_on Filled Yellow
CDS 266 21556 arrow_on Filled Red
S 21563 21384 arrow_on Filled Yellow
ORF3a 25393 26220 arrow_on Filled Silver
E 26245 26472 arrow_on Filled Green
M 26523 27191 arrow_on Filled Black
ORF6 27202 27387 arrow_on Filled Orange
ORF7a 27394 27759 arrow_on Filled Purple
ORF8 27894 28259 arrow_on Filled Green
N 28274 29533 arrow_on Filled Red
ORF10 29558 29674 arrow_on Filled Yellow

Figure1.5: Genomic sequence of SARS-CoV-2 (Wu et al., 2020).

Given its large genome and its error-prone viral polymerase (RdRp), SARS-CoV-2 accumulates mutations that contribute to its pathogenicity. However, one surprising feature of SARS-CoV-2 is that its ExoN (NSP14, NSP10 complex) has error (i.e., mutation) correction and chain recombination activities, normally not found in other viruses (Moeller et al., 2022).

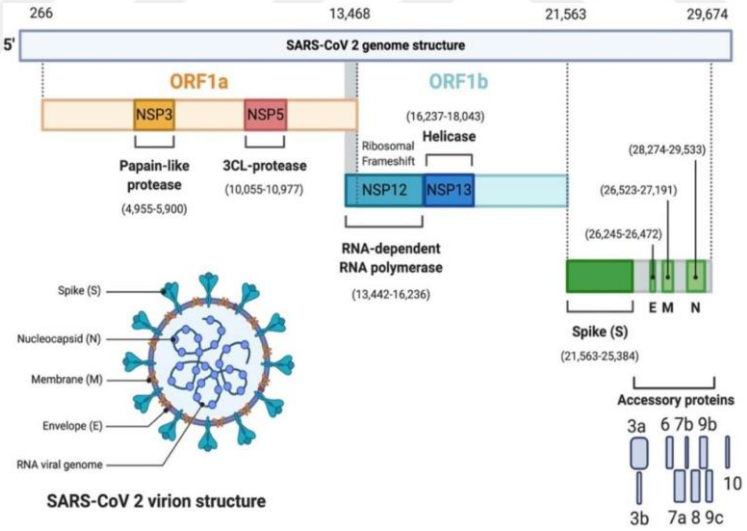

Figure1.6: Genomic construction of SARS-CoV-2. The virus genome encodes two large genes ORF1 (yellow), ORF1b (blue) coding for NSP1-NSP16. Some NSPs form a replication-transcription complex (RTC) involved in replication and transcription. NSP3 and NSP5 also encode papain-like protease (PLP) and 3CL- protease. These two proteins are involved in the breakdown of polypeptides and inhibit the host's innate immune system. NSP12 encodes for RNA-dependent RNA polymerase (RdRp). NSP15 encodes RNA helicase. Structural genes encode Spike (S), Matrix (M), Envelope (E), and Nucleocapsid (N) proteins (green) (Alanagreh, et al. 2020).

1.3.1 Proteins of SARS-CoV-2

SARS-CoV-2, like other coronaviruses structural, non-structural, and accessory proteins that are essential for its replication, assembly, interaction with host cells, infection, and immune evasion. These proteins play different roles in the life cycle of the virus and understanding

their functions is crucial for developing drugs and vaccines. SARS- CoV-2 have a 4 structural, 16 nonstructural, and half a dozen accessory proteins with unknown functions (Figure 1.6) (Alanagreh, et al. 2020; Yan et al., 2020; Guo et al., 2021).

1.3.1.1 NSPs and accessory proteins

Approximately 25 proteins of the virus are non-structural proteins (such as NSPs, e.g., 3-chymotrypsin-like protease, papain-like protease, helicase, and RdRp) and accessory proteins with unknown functions yet (Yan et al., 2020; Guo et al., 2021).

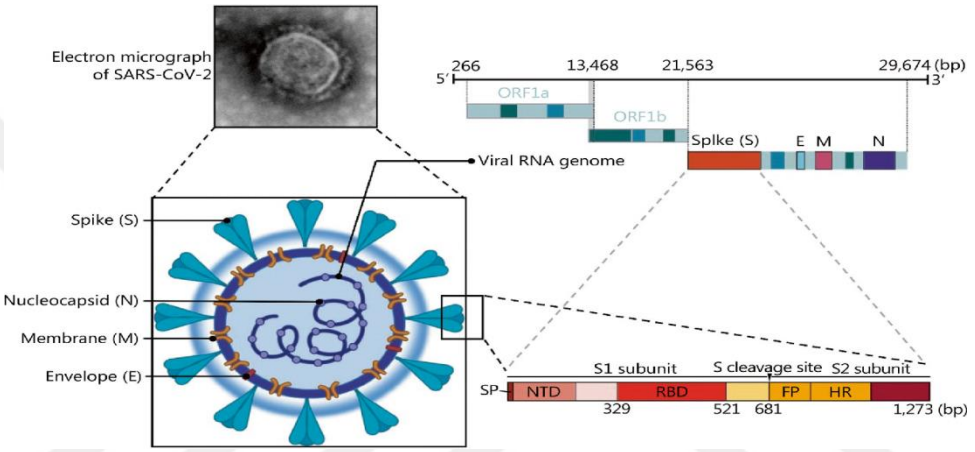

Figure 1.7: Electron micrograph and schematic view of SARS-CoV-2 (Li & Li, 2021).

The Spike (S) protein, the envelope protein (E), the membrane protein (M), and the nucleocapsid proteins are all required for the assembly, integrity and host cell invasion of viral particles (Figure 1.7) (Hu et al., 2003; Mandala et al., 2020; Li & Li, 2021). N also wraps around the mRNA genome (Cubuk et al., 2021) and may have a function in regulation of viral RNA synthesis and may interact with the M protein during virus budding (Wu et al., 2023).

1.3.1.2 Envelope protein

The envelope protein of SARS-CoV-2 is a small structural membrane protein composed of about 100 amino acids that plays a vital role in the assembly and release of new viral particles. While it is less studied compared to the spike glycoprotein, the envelope protein is essential for the life cycle of the virus (Mandala et al., 2020). The E protein interacts with other structural proteins, such as the S glycoprotein, the M protein, and, N protein to shape the viral envelope and package the viral genome (Chai et al., 2021). These interactions are

essential for stabilizing the viral envelope and incorporating the viral RNA into new viral particles and into the host cells (Kuzmin et al., 2022).

1.3.1.3 Membrane protein

The membrane (M) protein of SARS-CoV-2 is a crucial structural protein that plays a significant role in shaping the viral envelope and facilitating viral assembly and budding. The M protein is a transmembrane protein, spanning the lipid bilayer of the viral envelope. It is relatively large, consisting of about 230 amino acids. The protein is highly conserved among coronaviruses and its sequence is crucial for maintaining the overall structure and stability of the virus (Thomas & Immunity, 2020; Boson et al., 2021; Zhang et al., 2022; Mahtarin et al., 2022).

1.3.1.4 Nucleocapsid protein

The nucleocapsid (N) protein of SARS-CoV-2 is a fundamental structural protein that plays a central role in the viral life cycle. It is a highly conserved protein found in all coronaviruses, including SARS-CoV-2. The N protein is a multifunctional protein with a molecular weight of approximately 50-60 kDa. It is a phosphoprotein that binds to the viral RNA genome, forming a ribonucleoprotein complex known as the nucleocapsid (Cubuk et al., 2021). The N protein consists of two domains: the N-terminal domain (NTD) and the C-terminal domain (CTD). The NTD is involved in RNA binding, while the CTD is responsible for protein-protein interactions with other viral and host proteins (Zeng et al., 2020). The N protein interacts with other viral proteins, such as RdRp complex (i.e., replicase) to facilitate viral RNA synthesis (Dutta et al., 2020; Ye et al., 2020).

1.3.1.5 Spike glycoprotein

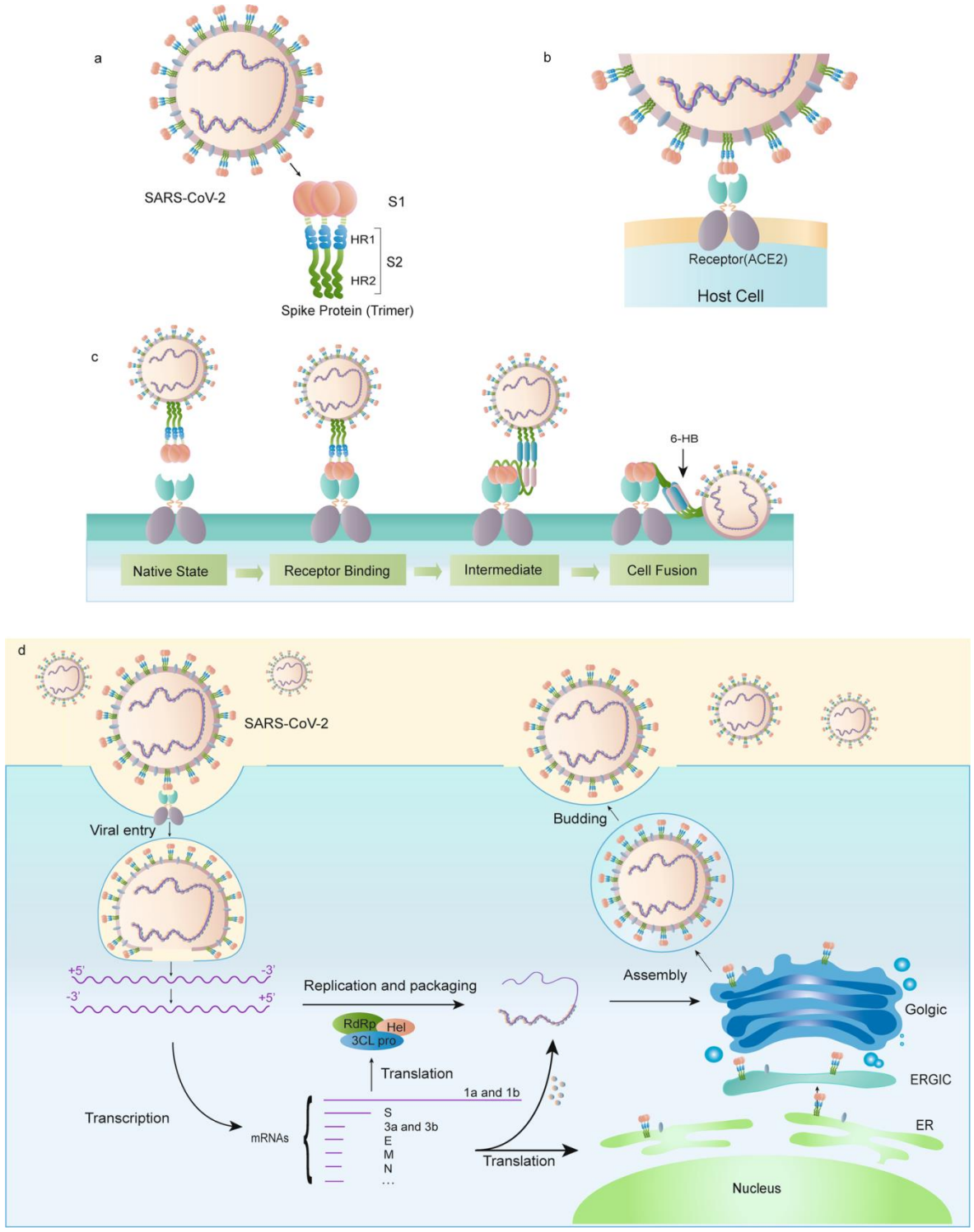

Figure 1.8: The general homotrimeric structure of the S protein (**a**), its binding to the ACE-2 receptor (**b**), the virus–cell fusion process mediated by the S (**c**), and the life cycle of SARS-CoV-2 in host cells (**d**) (Huang et al., 2020).

Spike (S) gene consists of 3822 nucleotides, encoding a 1273 amino acids protein (Figure 1.9) (Figure 1.10) with a size about 150 kDa. The S protein is a transmembrane with two domains (S1/S2). The S2 domain is embedded within the viral membrane. The S1 domain carries a RBD for the binding of the virus to a receptor (i.e., ACE-2) on the host cell (Wrapp et al. 2020). When the virus interacts with the host cell, the S protein undergoes a major structural rearrangement, helping the virus envelope to fuse with host cell membrane and initiate infectious cycle (Figure 1.8) (Huang et al., 2020; Nassaret al. 2021; Zhang et al., 2021).

```
   1 mfvflvllplvssqcvnlttrtqlppaytnsftrgvyypdkvfrssvlhstqdlflpffs
  61 nvtwfhaihvsgtngtkrfdnpvlpfndgvyfasteksniirgwifgttldsktqslliv
 121 nnatnvvikvcefqfcndpflgvyyhknnkswmesefrvyssannctfeyvsqpflmdle
 181 gkqgnfknlrefvfknidgyfkiyskhtpinlvrdlpqgfsaleplvdlpiginitrfqt
 241 llalhrsyltpgdsssgwtagaaayyvgylqprtfllkynengtitdavdcaldplsetk
 301 ctlksftvekgiyqtsnfrvqptesivrfpnitnlcpfgevfnatrfasvyawnrkrisn
 361 cvadysvlynsasfstfkcygvsptklndlcftnvyadsfvirgdevrqiapgqtgkiad
 421 ynyklpddftgcviawnsnnldskvggnynylyrlfrksnlkpferdisteiyqagstpc
 481 ngvegfncyfplqsygfqptngvgyqpyrvvvlsfellhapatvcgpkkstnlvknkcvn
 541 fnfngltgtgvltesnkkflpfqqfgrdiadttdavrdpqtleilditpcsfggvsvitp
 601 gtntsnqvavlyqdvnctevpvaihadqltptwrvystgsnvfqtragcligaehvnnsy
 661 ecdipigagicasyqtqtnsprrarsvasqsiiaytmslgaensvaysnnsiaiptnfti
 721 svtteilpvsmtktsvdctmyicgdstecsnlllqygsfctqlnraltgiaveqdkntqe
 781 vfaqvkqiyktppikdfggfnfsqilpdpskpskrsfiedllfnkvtladagfikqygdc
 841 lgdiaardlicaqkfngltvlppllтdemiaqytsallagtitsgwtfgagaalqipfam
 901 qmayrfngigvtqnvlyenqklianqfnsaigkiqdslsstasalgklqdvvnqnaqaln
 961 tlvkqlssnfgaissvlndilsrldkveaevqidrlitgrlqslqtyvtqqliraaeira
1021 sanlaatkmsecvlgqskrvdfcgkgyhlmsfpqsaphgvvflhvtyvpaqeknfttapa
1081 ichdgkahfpregvfvsngthwfvtqrnfyepqiittdntfvsgncdvvigivnntvydp
1141 lqpeldsfkeeldkyfknhtspdvdlgdisginasvvniqkeidrlnevaknlneslidl
1201 qelgkyeqyikwpwyiwlgfiagliaivmvtimlccmtsccsclkgccscgscckfdedd
1261 sepvlkgvklhyt
```

Figure1.9: Amino acid sequence of the S protein of SARS-CoV-2 (Wu et al., 2020).

```
21601 tcagtgtgttaatcttacaaccagaactcaattacccccctgcatacactaattctttcac
21661 acgtggtgtttattaccctgacaaagttttcagatcctcagttttacattcaactcagga
21721 cttgttcttacctttctttttccaatgttacttggttccatgctatacatgtctctgggac
21781 caatggtactaagaggtttgataaccctgtcctaccatttaatgatggtgtttatttttgc
21841 ttccactgagaagtctaacataataagaggctggatttttggtactactttagattcgaa
21901 gacccagtccctacttattgttaataacgctactaatgttgttattaaagtctgtgaatt
21961 tcaattttgtaatgatccatttttgggtgtttattaccacaaaaacaacaaaagttggat
22021 ggaaagtgagttcagagtttattctagtgcgaataattgcacttttgaatatgtctctca
22081 gccttttcttatggaccttgaaggaaaacagggtaatttcaaaaatcttagggaatttgt
22141 gtttaagaatattgatggttatttaaaatatattctaagcacacgcctattaatttagt
22201 gcgtgatctccctcagggtttttcggctttagaaccattggtagatttgccaataggtat
22261 taacatcactaggtttcaaactttacttgctttacatagaagttatttgactcctggtga
22321 ttcttcttcaggttggacagctggtgctgcagcttattatgtgggttatcttcaacctag
22381 gacttttctattaaaatataatgaaaatggaaccattacagatgctgtagactgtgcact
22441 tgaccctctctcagaaacaaagtgtacgttgaaatccttcactgtagaaaaaggaatcta
22501 tcaaacttctaactttagagtccaaccaacagaatctattgttagatttcctaatattac
22561 aaacttgtgccctttggtgaagtttttaacgccaccagatttgcatctgtttatgcttg
22621 gaacaggaagagaatcagcaactgtgttgctgattattctgtcctatataattccgcatc
22681 attttccacttttaagtgttatggagtgtctcctactaaattaaatgatctctgctttac
```

```
22741 taatgtctatgcagattcatttgtaattagaggtgatgaagtcagacaaatcgctccagg
22801 gcaaactggaaagattgctgattataattataaattaccagatgattttacaggctgcgt
22861 tatagcttggaattctaacaatcttgattctaaggttggtggtaattataattacctgta
22921 tagattgtttaggaagtctaatctcaaacctttgagagagatatttcaactgaaatcta
22981 tcaggccggtagcacaccttgtaatggtgttgaaggttttaattgttactttcctttaca
23041 atcatatggtttccaacccactaatggtgttggttaccaaccatacagagtagtagtact
23101 ttcttttgaacttctacatgcaccagcaactgtttgtggacctaaaaagtctactaattt
23161 ggttaaaaacaaatgtgtcaatttcaacttcaatggtttaacaggcacaggtgttcttac
23221 tgagtctaacaaaaagtttctgccttccaacaatttggcagagacattgctgacactac
23281 tgatgctgtccgtgatccacagacacttgagattcttgacattacaccatgttcttttgg
23341 tggtgtcagtgttataacaccaggaacaaatacttctaaccaggttgctgttctttatca
23401 ggatgttaactgcacagaagtccctgttgctattcatgcagatcaacttactcctacttg
23461 gcgtgtttattctacaggttctaatgtttttcaaacacgtgcaggctgtttaataggggc
23521 tgaacatgtcaacaactcatatgagtgtgacatacccattggtgcaggtatatgcgctag
23581 ttatcagactcagactaattctcctcggcgggcacgtagtgtagctagtcaatccatcat
23641 tgcctacactatgtcacttggtgcagaaaattcagttgcttactctaataactctattgc
23701 catacccacaaattttactattagtgttaccacagaaattctaccagtgtctatgaccaa
23761 gacatcagtagattgtacaatgtacatttgtggtgattcaactgaatgcagcaatctttt
23821 gttgcaatatggcagtttttgtacacaattaaaccgtgctttaactggaatagctgttga
23881 acaagacaaaaacacccaagaagtttttgcacaagtcaaacaaatttacaaaacaccacc
23941 aattaaagattttggtggttttaatttttcacaaatattaccagatccatcaaaaccaag
24001 caagaggtcatttattgaagatctacttttcaacaaagtgacacttgcagatgctggctt
24061 catcaaacaatatggtgattgccttggtgatattgctgctagagacctcatttgtgcaca
24121 aaagtttaacggccttactgttttgccacctttgctcacagatgaaatgattgctcaata
24181 cacttctgcactgttagcgggtacaatcacttctggttggacctttggtgcaggtgctgc
24241 attacaaataccatttgctatgcaaatggcttataggtttaatggtattggagttacaca
24301 gaatgttctctatgagaaccaaaaattgattgccaaccaatttaatagtgctattggcaa
24361 aattcaagactcactttcttccacagcaagtgcacttggaaaacttcaagatgtggtcaa
24421 ccaaaatgcacaagctttaaacacgcttgttaaacaacttagctccaattttggtgcaat
24481 ttcaagtgttttaaatgatatcctttcacgtcttgacaaagttgaggctgaagtgcaaat
24541 tgataggttgatcacaggcagacttcaaagtttgcagacatatgtgactcaacaattaat
24601 tagagctgcagaaatcagagcttctgctaatcttgctgctactaaaatgtcagagtgtgt
24661 acttggacaatcaaaaagagttgatttttgtggaaagggctatcatcttatgtccttccc
24721 tcagtcagcacctcatggtgtagtcttcttgcatgtgacttatgtccctgcacaagaaaa
24781 gaacttcacaactgctcctgccatttgtcatgatggaaaagcacactttcctcgtgaagg
24841 tgtctttgtttcaaatggcacacactggtttgtaacacaaaggaatttttatgaaccaca
24901 aatcattactacagacaacacatttgtgtctggtaactgtgatgttgtaataggaattgt
24961 caacaacacagtttatgatcctttgcaacctgaattagactcattcaaggaggagttaga
25021 taaatattttaagaatcatacatcaccagatgttgatttaggtgacatctctggcattaa
25081 tgcttcagttgtaaacattcaaaaagaaattgaccgcctcaatgaggttgccaagaattt
25141 aaatgaatctctcatcgatctccaagaacttggaaagtatgagcagtatataaaatggcc
25201 atggtacatttggctaggttttatagctggcttgattgccatagtaatggtgacaattat
25261 gctttgctgtatgaccagttgctgtagttgtctcaagggctgttgttcttgtggatcctg
25321 ctgcaaatttgatgaagacgactctgagccagtgctcaaaggagtcaaattacattacac
25381 ataaacgaacttatggatttgtttatgagaatcttcacaattggaactgtaactttgaag
```

Figure 1.10: Nucleotide sequence of the S protein of SARS-CoV-2 (Wu et al., 2020).

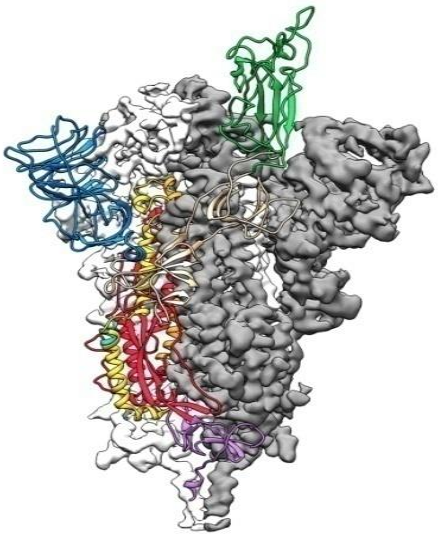

Figure 1.11: Three-dimensional structure of the SARS-CoV-2 S protein at atomic-level. The receptor binding domain (RBD) is colored green (Wrapp et al. 2020).

The SARS-CoV-2 spike protein has a signal peptide (1-13) at its N-terminus. The S1 subunit (14-685) and S2 subunit (686-1273) are responsible for receptor binding and cell membrane fusion, respectively. The N-terminal domain (14-305) and receptor binding domain (RBD, 319-541) are located within the S1 subunit. The S2 subunit contains the fusion peptide (FP) (788-806), heptapeptide repeat sequence 1 (HR1) (912-984), HR2 (1163-1213), TM domain (1213-1237), and cytoplasmic domain (1237–1273) (Figure 1.12) (Nassaret al. 2021).

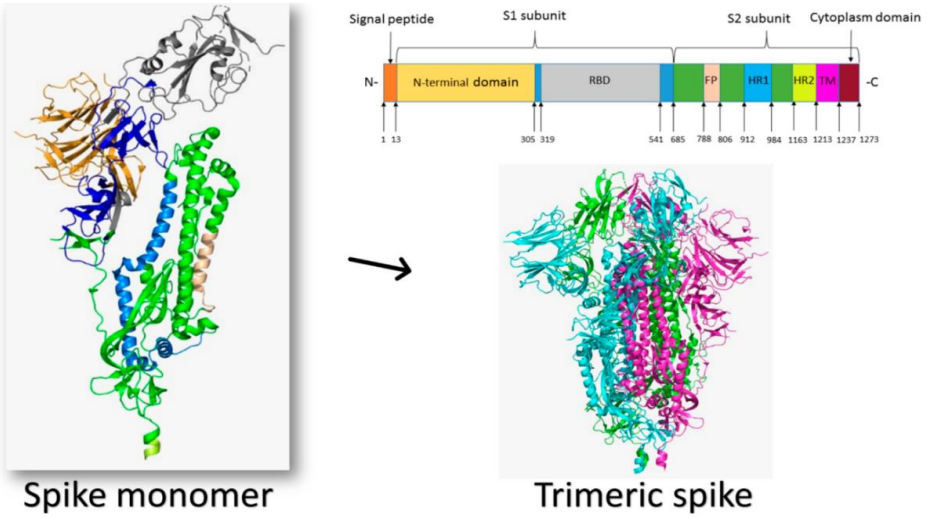

Figure 1.12: The structure of the SARS-CoV-2 S protein. The monomeric spike is illustrated as a bar by coloring different areas from the N-terminus to the C-terminus. The S protein is homotrimeric and forms a crown structure on the virus (Nassaret al. 2021).

The aim of this study was to clone the SARS-CoV-2 S gene in bacteria and fungi using shuttle vectors to determine and analyze its expression profile.

2. MATERIALS AND METHODS

2.1 Reagents

The chemicals used in cell growth, agarose gel electrophoresis, SDS-PAGE, and other tests were all reagent quality. Acrylamide, bisacrylamide, Ammonium persulfate (APS), methanol and Brilliant Blue R-250 from Merck Co.; TEMED, SDS, Isopropyl alcohol, Bromophenol blue and glycerol from Sigma Co.; Tris was from Ambresco Co.6X loading dye (Fermentas), DNA Marker BLUE 1kp, protein markers were from Abcam (ab 116029), pcDNA3.1-SARS2-Spike (Addgene Cat No: 145032) and pRG201 (Addgene Cat No: 64394) plasmids were purchased from Addgene. Ultrapure water (UPW) was used in the preparation of all solutions throughout the study.

2.2. Bacteria and yeast strains

The strains of BL21(DE3) and *E. coli* DH5α type were obtained from Thermo Scientific. The yeast strains *S. cerevisiae* BY4741 type were obtained from American Type Culture Collection.

2.2.1 Identification of bacterial species

2.2.1.1 Plasmid isolation

Bacteria, bearing the plasmids (pcDNA3.1-SARS2-Spike and pRG201), were supplied from Addgene in deep agar form. The bacteria were inoculated on LB amp agar plates and incubated at 37°C overnight. A single colony was taken from these plates and transferred into liquid LB amp medium and incubated at 37°C overnight (about 10 h). Plasmid isolation was performed with a commercial kit (“ROCHE Genopure Plasmid Midi Kit, Catalog number 03 143 414 001 “) according to the supplier's instructions with minor modifications.

“Genopure Plasmid Midi Kit”

1 Black Suspension Buffer

2 Black RNase A

3 White Neutralization Buffer

5 Red Equilibration Buffer

6 Blue Wash Buffer

7 Yellow Elution Buffer

8 NucleoBond AX 100 Columns.

Plasmid isolation protocol

1. 30 ml of bacterial culture was taken and transferred to the falcon tube. It was centrifuged at 9,000 rpm for 10 min at +4 °C. The supernatant was discarded. 4 ml of suspension buffer was added to the pellet and gently mixed.
2. 4 ml of lysis buffer was added, gently mixed, and incubated for 5 min at room temperature.
3. 4 ml of cold neutralization buffer was added to the suspension, mixed thoroughly until a homogeneous suspension was formed and incubated on ice for 5 min.
4. The suspension was centrifuged at 9,000 rpm for 45 min at +4 °C.
5. The spin column was fixed into a Falcon tube and 2.5 ml of equilibration buffer was added, allowed to flow, and the flow was discarded.
6. The supernatant from step 4 was loaded onto the fixed column and the flow was again discarded.
7. The column was washed with 5 ml of washing buffer twice.
8. 5 ml of warm (about 50-60 °C) elution buffer was added to the column and the flow was collected into a different tube.
9. 3.6 ml of 90% cold ethanol was added to the mixture containing the plasmid and centrifuged at 9000 rpm for 60 min at +4°C. The supernatant was discarded.
10. 3 ml of 70% cold ethanol was added onto the pellet and centrifuged at 9000 rpm +4°C. The supernatant was discarded and the tubes were allowed to dry at room temperature for 15 min to evaporate ethanol.
11. To the plasmid DNA pellet 100 µl TE buffer was added gently dissolved and spectrophotometeric nanodrop measurements of the plasmid solutions were made for the purity and quantity of DNA.

2.3 Construction of recombinant plasmids

Plasmids pcDNA3.1-SARS 2-Spike and plasmid pRG 201 were used for the construction of new recombinant shuttle vector. Both plasmids double digested with XbaI and SacI restriction enzymes from their multi cloning sites (MCS) and the fragment were run on in agarose gel electrophoresis (minipreps) and they were isolated from the agarose gel using a commercial kit ("Thermo Scientific GeneJET Gel Extraction Kit 00532467") for the purification of the Spike gene from pcDNA3.1-SARS 2-Spike plasmid and the larger fragment of pRG201 resulting from double digests. These fragments were ligated to construct the recombinant shuttle vector named pRG-Spike in this study.

2.3.1 Restriction digestion of plasmids

Before the isolated plasmids were exposed to the ligation reaction, they were cut from the determined regions with appropriate restriction enzymes. Below are the components of the restriction reaction and the reaction conditions.

pcDNA3.1-SARS 2-Spike (pcDNA3.1-SARS2-Spike (Addgene Cat No: 145032))

- 2 µl plasmid solution
- 1.5 µl XbaI enzyme
- 1 µl SacI enzyme
- 1 µl Tango buffer 10X
- 4.5 µl UPW

Total volume to 10 µl 4 h 37 °C after 65°C 10 min.

pRG 201(pRG201 (Addgene Cat No: 64394))

- 2 µl plasmid solution
- 1.5 µl XbaI enzyme
- 1 µl SaCI enzyme
- 1 µl Tango buffer 10X
- 4.5 µl UPW

Total volume to 10 µl 4 h 37 °C after 65°C 10 min.

2.3.2 Agarose gel electrophoresis

Plasmids digested with particular restriction enzymes were run in 1% agarose gels at 90 Volts for 60 min.

1% agarose gel

- 0.25 g agarose
- 25 ml 1X TAE
- 1 µl EtBr

The solution 1% agarose was heated in microwave oven until a clear uniform solution was formed and then the solution was allowed to cool. Then, EtBr was added and the solution was poured into gel tank. After the gel was solidified, markers and samples were added into the wells and the gel electrophoresis were run.

2.3.3 Plasmid extraction

Plasmids digested with restriction enzymes were isolated from agarose gel using a commercial kit (“Thermo Scientific GeneJET Gel Extraction Kit 00532467”)

“Thermo Scientific GeneJET Gel Extraction Kit”

- Binding Buffer
- Wash Buffer
- Elution Buffer (10 mM Tris-HCl, pH 8.5)
- GeneJET Purification Columns (preassembled with collection tubes)
- Ethanol 100%
- Microcentrifuge
- 1.5 ml or 2 ml microcentrifuge tubes
- Heating block (ESCO)

1) A gel piece containing the DNA fragments of interest was taken into a microcentrifuge tube and weighed for purification of DNA.

2) A binding buffer was added at a ratio of 1:1 (w/v) and was mixed gently until the gel dissolved.

3) The solution was transferred to a centrifuge column and was centrifuged at 16,000 x*g* for 1 min and the fluid beneath the tube was discarded.

4) 700 µl Wash buffer was added to the column, centrifuged at 16,000 x*g* for 1 minute and the fluid beneath the tube was discarded.

5) The column was centrifuged empty at 16,000 x*g* for 1 min.

6) 50 µl of elution buffer was added to the middle of the column and centrifuged at 16,000 x*g* for 1 min. The solution was taken into a new microcentrifuge tube and stored at -20 °C for use.

2.3.4 Ligation

The desired plasmid was obtained by ligation reaction of plasmid fragments isolated from agarose gel into a shuttle vector, resulting a recombinant plasmid.

Ligation

- 3 µl digested pRG plasmid
- 6 µl Spike plasmid fragment
- 2 µl T4 Ligase Buffer 10X
- 2 µl T4 Ligase
- 0.25 µl DDT
- 7.75 µl UPW
-

The ligation solution was incubated overnight at room temperature, then at 65°C for 10 min.

2.4 Bacteria Growth Medium

2.4.1. Bacterial stocks

The main strains were taken from the -80°C stock and grown in LB medium. The culture was incubated at 37°C for 12 h. For stock solutions, glycerol was added at 1:1 ratio to cultures and stored at -20°C.

2.4.2 LB media

25 g LB broth Microbiology (Merck 1.10285.0500)

The pH of the medium was adjusted to 7.0 at room temperature. The content was dispensed into 50 ml flasks and autoclaved. The medium in the flasks was clear yellowish in appearance.

2.4.3. LB agar

25 g LB Microbiology Agar (Merck 1.10283.0500)

The medium was autoclaved at 121 °C for 15 min and poured into petri dishes and was kept at room temperature to solidify.

2.5 Preparation of component cells

Transformation competent cells were prepared using a standard protocol (Green & Sambrook, 2018).

2.5.1 CCMB80 transformation buffer (500 ml)

- 1 M Potassium acetate pH 7.0 5ml
- $CaCl_2.2H_2O$: 5.9 g
- $MnCl_2.4H_2O$ 2.0 g
- $MgCl_2.6H_2O$ 1.0 g
- Glycerol %100 50 ml

The buffer was autoclaved at 121 °C for 15 min and stored at +4 °C to ensure homogeneity and preserve the competent cells.

2.5.2 SOB medium preparation (1 Liter)

- 20 g Tryptone
- 5 g Yeast Extract Powder
- 0.05 g NaCl

The solution was mixed until there is a complete dissolve. 10 mL of a 250 mM solution of KCl was added (this solution is made by dissolving 1.86 g of KCl in 100 mL of deionized H2O.) pH of the medium was adjusted to 7.0 with 5 N NaOH and the final volume was adjusted to 1 L with deionized H_2O. The final solution was sterilized by autoclaving for 20 min. Before use, 5 mL of a sterile solution of 2 M $MgCl_2$ was added (this solution is made by dissolving 19 g of $MgCl_2$ in 90 mL of deionized H_2O. The volume of the solution was adjusted to 100 mL with deionized H_2O and sterilized by autoclaving for 20 min.

2.5.3 SOB agar preparation

To SOB medium above agar was added at 2%.

2.5.4 Preparation of competent cells of DH5α and BL21

1) A single colony was taken from cells grown on LB agar and inoculated on SOB agar and incubated for 24 h at 23°C.
2) A single colony was taken from SOB agar plates and inoculated into SOB medium. It was incubated at 23°C for 16 h (125 ml DH5 alpha, 125 ml BL21).
3) 1 ml was taken and inoculated into two 250 ml SOB media and incubated at 20°C for 16 h.
4) 50 ml cultures were transferred into 5 falcon tubes and were centrifuged at 3,000 rpm at +4 °C for 10 min.
5) The supernatant was discarded and 4 ml of cold CCMB80 buffer was added to the pellets. Pipetting was done until the pellets were thoroughly dissolved. Then, 12 ml of CCMB80 buffer was added to the suspension and mixed until well-suspended.
6) The content in Falcon tubes were combined in 2 tubes for each bacterial strain (i.e., 2 for DH5 alpha, 2 for BL21).
7) The cultures were incubated in ice for 20-30 min.
8) 5 ml of cold CCMB80 buffer was added to each tube and centrifuged at 3,000 rpm for 10 min at +4 °C and the supernatant was discarded.
9) The pellets were aliquoted in 100 µl portions, held on ice for 20-30 min and stored at -80 °C.

2.5.5 SOC media (1 Liter)

The SOC medium was the same as SOB, except that it included 1 M 20 ml glucose solution.

2.5.6 LB amp medium (1 liter)

25 g of LB broth (Merck 1.10285.0500) was dissolved dH_2O and the final volume was made to 1 L with water. The pH of the medium was adjusted to 7.0 at room temperature. After the media was cooled, ampicillin at 100 µg/ml concentration (i.e., 1 ml ampicillin from a 100 mg/ml stock ampicillin solution was added to 1 liter broth). The medium was dispensed into 125 ml sterile flasks at 20 ml volumes.

2.5.7 LB amp agar

LB amp plates contained 2% agar.

2.5.8 Transformation of competent bacterial cells

Transformation of the recombinant plasmid (pRG-Spike) into the competent DH5α *E.coli*.

1. 100 µl of DH5α cells taken from -80 °C were placed on ice and allowed to dissolve (about 5-10 min).
2. The sell suspension were mixed with 5 µl of recombinant plasmid, the cells were placed in ice, and incubated for 30 min.
3. Cells were heat-shocked in a water bath at 42 °C for 45 sec.
5. They were transferred on ice and incubated for 5 min.
6. 950 µl of SOC medium was added to the cells and the mixture was incubated at 37 °C for 1 h.
7. The cell mixture was centrifuged at 2,000 rpm for 5 min and 700 µl of supernatant was discarded.
8. The remaining supernatant + pellet was spread at 50µl, 100µl and 150µl volumes over the LB amp plates and the plates were incubated overnight at 37°C.
9. The observable colonies (particularly those with satellite colonies) were selected for recombinant plasmid isolation and minipreps.

2.6 Isolation Prg-Spike recombinant plasmid

Plasmid isolation was performed with a commercial kit (“QIAprep Spin Miniprep Kit (50) Cat. No.27104”) using the supplier's instructions.

“QIAprep Spin Miniprep Kit”

“QIAprep Spin Columns”

Buffer P1

Buffer P2

Buffer N3

Buffer PB

Buffer PE (concentrate)

Buffer EB

Collection Tubes (2 ml)

2.6.1 “QIAprep” plasmid isolation protocol”

1. 5 ml of cells grown overnight were taken and centrifuged at 4000 rpm for 10 min.
2. Supernatant was discarded, 250 µl of P1 buffer was added to the pellet and waited for 5 min. Then 250 µl of P2 buffer was added, the tube was inverted 6-8 times.
3.After adding the buffers, the pellet and buffer mixture was transferred to the microcentrifuge tube.
4. On top of the mixture, 350 µl of N3 buffer was added and inverted.
5. The mixture was centrifuged at 13,000 rpm for 10 min. The supernatant was then transferred to the spin column.
6. The column was centrifuged at 13,000 rpm for 60 sec and the flow was discarded.
7. 0.5 ml of PB buffer was added to the column and centrifuged at 13000 rpm for 60 sec. Then the fluid part was discarded.
8. 0.75 ml of PE buffer was added to the column, centrifuged at 13,000 rpm for 60 sec, and the liquid was discarded.
9. The column was centrifuged empty for 60 sec at 13,000 rpm.
10. The column was placed into a 1.5 ml microcentrifuge tube and 50 µl of EB buffer was added and centrifuged at 13,000 rpm for 1 min. The eluted part was taken and maintained at -20 °C.

2.6.2 Restriction digestion of the Prg-Spike plasmid

After plasmid isolation, a nanodrop spectrophotometer measurements of the plasmids were made and the purity and amount of DNA were determined. The plasmids were digested with restriction enzymes and the presence of recombinant plasmids was confirmed by minipreps.

Restriction Digestion

- 2 µl new recombinant plasmids (Prg-Spike)
- 1.5 µl XbaI enzyme
- 1 µl SaCI enzyme
- 1 µl Tango buffer 10X
- 4.5 µ UPW

2.6.3 Agarose gel electrophoresis of the recombinant Prg-Spike plasmid

Restriction digestion was carried for 4 h at 37 °C and then enzymes were deactivated by holding the solution at 65°C 10 min. Plasmids cut with the restriction enzymes were run at 90 Volts for 60 min in 1% agarose gel.

2.6.4 Transformation of the newly constructed plasmid (i.e., recombinant pRG-Spike) into the competent BL21 cells.

1. 100 µl of Bl21 cells taken from -80 °C were placed on ice and allowed to dissolve (about 5-10 min).
2. To the cell mixture 5 µl recombinant plasmid was added, placed in ice and incubated for 30 min.
3. The cell mixture was heat shocked in a water bath at 42 °C for 45 sec.
5. The mixture was transferred onto ice and incubated for 5 min.
6. 950 µl of SOC medium was added to the cells, the mixture was incubated at 37 °C for 1 h.
7. The cells were centrifuged at 2000 rpm for 5 min and 700 µl of supernatant was discarded.
8. The remaining supernatant + pellet was plated at 50 µl, 100 µl and 150 µl volumes into LB amp plates and incubated overnight at 37°C.
9. The observable colonies (particularly those with satellite colonies) were selected for recombinant plasmid isolation and minipreps.

2.7 Isolation of spike protein from transformed BL21 cell

1) Single colonies were picked from transformed BL21 plates, inoculated into 50 ml of LB amp medium, and incubated at 37 °C overnight.
2) 1.5 ml of cell cultures of non-transformed and transformed BL21 were inoculated into two 20 ml LB amp containing flasks and the media were incubated at 37°C for 10 h.
3) When the OD600 value was 1.4, 2 µl 1 M IPTG was added.
4) The media were incubated at 16°C for 12 h.
5) The medium was divided into sterile centrifuge tubes in equal volume and centrifuged for

20 min at 16000 g at +4°C.

6) After centrifugation, the supernatants were discarded and the pellets were thoroughly dissolved by adding lysis buffer on ice and cells were sonicated for 2 min at 50% capacity, in 3 repetitions.

7) The pellet suspension was taken into 2 ml microcentrifuge tubes and was centrifuged at 18,000 rpm for 1 h at +4°C and stored at -80°C.

2.7.1 Analysis of spike protein using SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is an analytical method that separates protein mixtures according to their molecular mass in the presence of an electric field. Sodium dodecyl sulfate (SDS) molecules help the isolation and identification of proteins. The SDS-PAGE procedure consists of electrophoresis through a porous polyacrylamide gel matrix, separating the molecular masses of proteins in the presence of electric field after initial denaturation with an anionic detergent (SDS) that imparts a negative charge to the proteins.

40 µl of the samples from protein isolation were taken and 2x Laemmli Buffer (Biorad) was added at 1:1 ratio and boiled at 95 °C for 5 min. Then, 30 µl of dye (coomassie blue) was added. To each well on the SDS-PAGE gel about 10 µl of the sample mixture was loaded (Table 2.1).

Table 2.1: SDS-PAGE Gel Components

%10 Runnig Gel	**(µl)**
ddH_2O	4000
30% Acryl/Bis	3340
Tris-HCl (1.5 M, pH 8.8)	2500
SDS (20%)	50
Ammonium persulfate (10%) (APS)	50
TEMED	5
%5 Stacking Gel	**(µl)**
ddH_2O	4176
30% Acryl/bis 37.5:1	1012
Tris-HCl (1.0 M, pH 6.8)	750
SDS (20%)	30
Ammonium persulfate (10%)	30
TEMED	3

2X SDS gel-loading buffer

- 100 mM Tris-Cl (pH 6.8)
- 4% (w/v) SDS
- 0.2% (w/v) bromophenol blue
- 20% (v/v) glycerol

%20 SDS 100 ml solution

20 g sodium dodecyl sulfate was dissolved in 80 mL of dH_2O and the final volume made to 100 ml with dH_2O.

10% Ammonium persulfate

0.1 g of ammonium persulfate was dissolved in 1 ml of ddH_2O.

Tris-HCl

1 M Tris base solution: 12.11 g of Tris base was dissolved in 80 mL ddH_2O and total volume was completed to 100 ml. The pH of solution adjusted to 6.6 with 1M HCI.

1.5 M Tris base solution: 18.7 g of Tris base was dissolved in 80 mL ddH_2O and total volume was completed to 100 ml. The pH of solution was adjusted to 8.8 with 1M HCI.

Running Buffer (10X) 1 Liter pH: 8.3

- Tris Base (MWt 121.14g/mol) 30.3 g
- Glysine (MWt 75.07 g/mol) 144.4 g
- SDS (MWt 288.38) 10 g

The final volume was completed to 1 liter with ddH_2O and the pH was adjusted to 8.3.

2.8 Transformation of Yeast (BY4741) Cells

2.8.1 YPD media (200 ml)

- %2 Peptone
- %2 d- Glucose
- %1 Yeast Extract

The medium was dispensed into 50 ml flasks and autoclaved.

2.8.2 YPD agar (200 ml)

- %2 Peptone
- %2 d-Glucose
- %1 Yeast Extract
- %2 Agar

The medium was autoclaved at 121 °C for 15 min and poured into petri dishes which were kept at room temperature to solidify before paraffin sealing and storing them upside down in a refrigerator at 4 °C.

2.8.3 YNB media (250 ml)

- 1.67 g YNB
- 5g Dextrose
- 2.5 mg Ura
- 25mg Leucine
- 12.5 mg His

The medium was dispensed into 50 ml flasks and autoclaved.

2.8.4 YNB agar (250 ml)

- 1.67 g YNB
- 5g Dextrose
- 2.5 mg Uracil
- 25mg Leucine
- 12.5 mg Histidine
- 2 g Agar

The medium was autoclaved at 121 °C for 15 min and poured into petri dishes which were kept at room temperature to solidify before paraffin sealing and storing them upside down in a refrigerator at 4 °C.

2.8.5 Preparation of LiOAc-TE buffer and PEG mixture

- 1X TE-LiOAc Solution
- 10mM Tris-HCl, pH 8.0, with 1mM EDTA and 0.1 M Lithium Acetate (BLD Pharm Cat No: BD126427)
- PEG Mix
- 40% PEG (4000 or 3350) in 1X TE-LiOAc solution

The solution was autoclaved and stored at +4 °C.

2.8.6 Yeast transformation

1) Cells from the BY4741 cell line were inoculated on YPD agar and incubated for 12 h at 30 °C.

2) After the cells have grown, a single colony is taken with the help of a loop and inoculated into YPD media and incubated at 30 °C for 12 h.

3) 2.5 ml of cells grown in YPD media inoculated into YPD media and when the OD600 of cultures was 0.5 they were used for transformation.

4) Cultures were centrifuged for 5 min at 3000 rpm.

5) Pellets were washed with 1ml 1X LiOAc-TE solution.

6) Cultures were centrifuged again for 5 min at 3000 rpm and the pellet was resuspend in 100 µl 1X LiOAc-TE solution left in shaker at 30° C for 1 h.

7) Transformation mix was 100 µl cells + 2 µl recombinant plasmid + 5µl carrier (bait) DNA (ssDNA) and the mixture was incubated at 30°C for 30 min.

8) 700 µl PEG added and mixed well and incubated at 30°C for 1 h.

9) The suspension was heat shocked in a 42°C water bath for 14 min.

10) The suspension spined down at 4000 rpm for 2 min to remove PEG.

11) The pellet was resuspended in 10 µl UPW, let to mix in a shaker at 30°C for 20 min

12) The mixture diluted and plated.

The plated petri dishes were incubated overnight at 30°C. Colonies were observed after incubation. A single colony was taken with the help of a loop and planted in YNB medium incubated overnight at 30°C and protein isolation was run.

2.8.7 Isolation of spike protein from yeast cells

2.8.6.1 Preparation "Laemli Buffer" (10 ml 1X) (Uk, 1970)

- %100 glycerol 500 µl
- %20 SDS 1000 µl
- %0.1 Bromfenolblue 25 µl
- β-merkaptoetanol 400 µl
- Tris-HCl pH: 6.8 600 µl
- dH_2O 7475 µl

2.8.6.2 Protein isolation from yeast cells (Kushnirov, 2000)

1) 50 ml of yeast culture was taken and transferred to the falcon tube.

2) Centrifugation was done at 4000 rpm for 5 min at +4 °C. The supernatant was discarded and 100 µl of UPW was added to the pellet, well suspended.

3) On the suspended pellet, 0.2 M 100 µl NaOH solution was added, mixed, and pellet formation was observed.

4) Then 100 µl of 1X Laemmli Buffer was added to the mixture. The cap of the tube was tightly closed.

5) The tubes were placed in the boiling water and allowed to boil for 5 min.

6) The mixture was centrifuged at 5000 rpm for 5 min and the supernatant was used for SDS-PAGE (Table 2.2).

Table 2.2: SDS-PAGE Gel

%10 Running Gel	(µl)
ddH_2O	4000
30% Acryl/Bis	3340
Tris-HCl (1.5 M, pH 8.8)	2500
SDS (20%)	50
Ammonium persulfate (10%) (APS)	50
TEMED	5
%5 Stacking Gel	(µl)
ddH_2O	4176
30% Acryl/bis 37.5:1	1012
0.1 M Tris-HCl pH: 6.8	750
SDS (20%)	30
Ammonium persulfate (10%)	30
TEMED	3

2X SDS PAGE gel-loading buffer

- 100 mM Tris-Cl (pH 6.8)
- 4% (w/v) SDS
- 0.2% (w/v) bromophenol blue
- 20% (v/v) glycerol

%20 SDS 100 ml solution

20 g sodium dodecyl sulfate was dissolved in 80 mL of dH_2O, then the solution was adjusted to final 100 ml volume with dH_2O.

10% Ammonium persulfate (1 ml)

0.1 g of ammonium persulfate was dissolved in 1 ml of ddH_2O.

Tris-HCl and Tris-base solutions

1M Tris base solution: 12.11 g of Tris base was dissolved in 80 mL ddH_2O and total volume was completed to 100 ml. The pH of solution adjusted to 6.8 with 1M HCI.

1.5 M Tris base solution: 18.7 g of Tris base was dissolved in 80 mL ddH_2O and total volume was completed to 100 ml. The pH of solution was adjusted to 8.8 with 1M HCI.

Running Buffer (10X) (1 Litter, pH 8.3)

- Tris Base (m/w 121.14 g/mol) 30.3 g
- Glysine (m/w 75.07 g/mol) 144.4 g
- SDS (m/w 288.38) 10 g

The content was dissolved in 800 ml ddH_2O, pH was adjusted to 8.3 and the final volume was made to1 L with ddH_2O.

3. RESULTS

3.1 Bacterial and Yeast Cultures

E. coli DH5α, *E. coli* BL21(DE3) and *S. cerevisiae* BY4741 cell lines were taken from -20°C and inoculated into petri dishes. *E. coli* DH5α and *E. coli* BL21(DE3) was incubated at 37°C and *S. cerevisiae* BY4741 was incubated at 30°C until the colonies were observable.

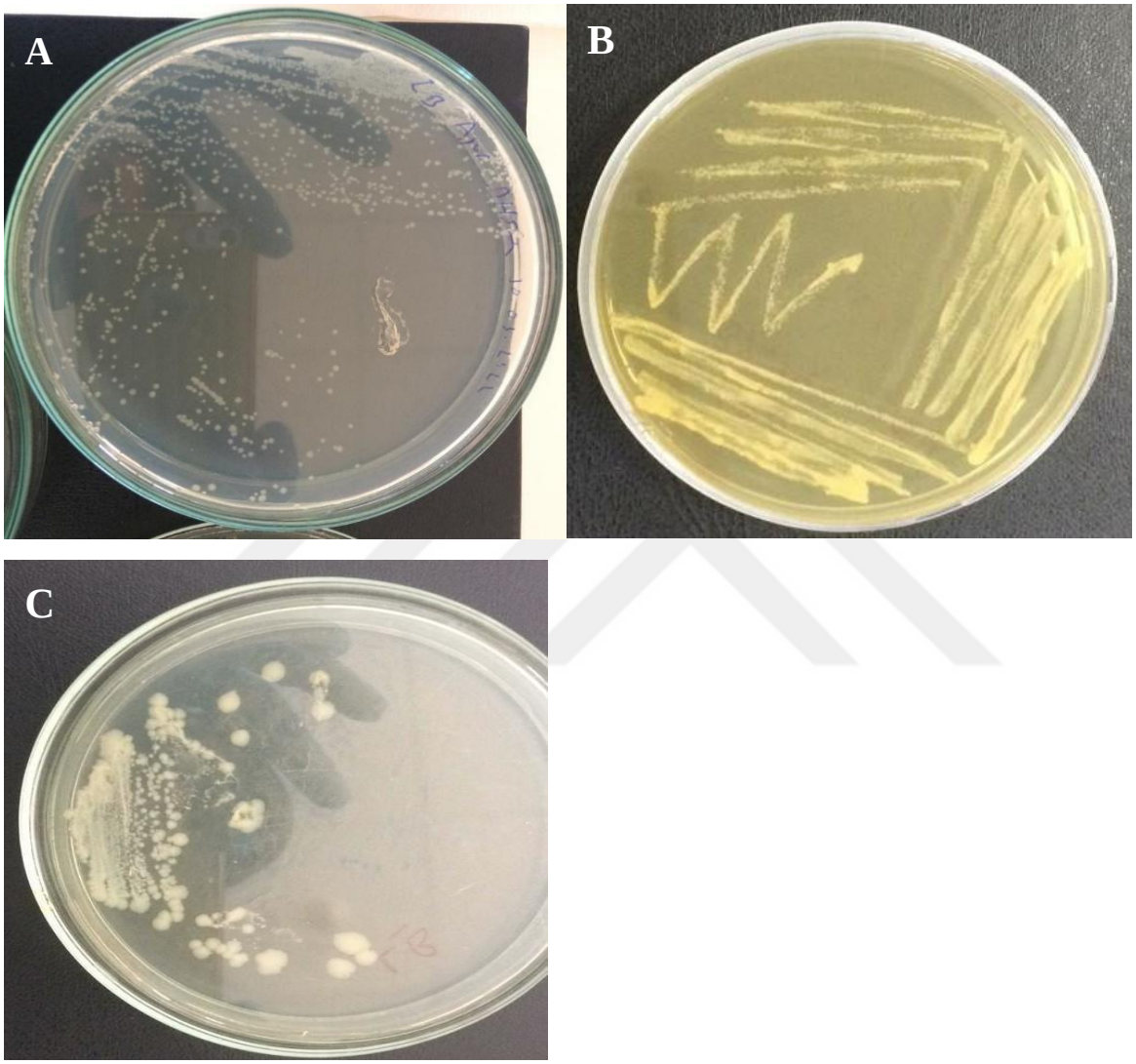

Figure 3.1: Single-cell colony formation of DH5α (A), BY4741 (B) and BL21 (C).

3.2 Plasmid Isolation

The plasmids used were pcDNA3.1-SARS2-Spike (Addgene Cat No: 145032) (Shang et al., 2020) and pRG201 (Addgene Cat No: 64394) (Gnügge, Liphardt, & Rudolf, 2016). Plasmids were isolated using ROCHE Genopure Plasmid Midi Kit according to manufacturer's manual. The purity of plasmids (DNA) was quantified by using a Denovix DS-11 FX+ Spectrophotometer.

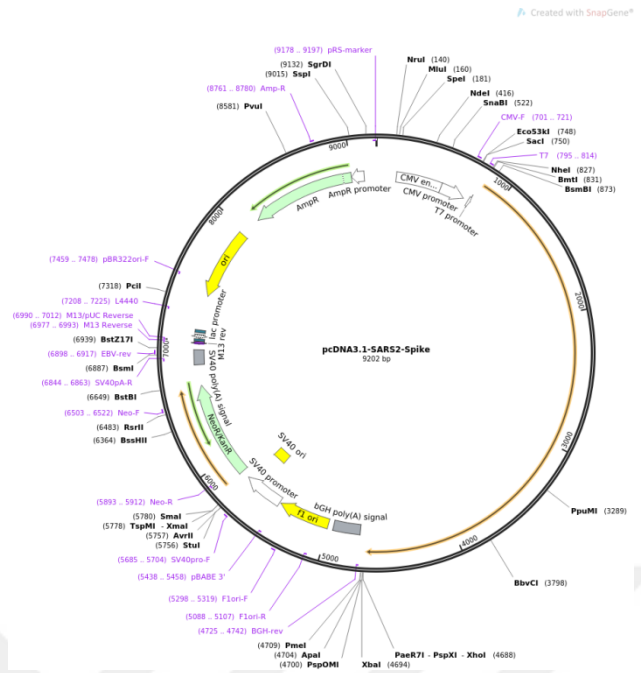

Figure 3.2: "pcDNA3.1-SARS2-Spike ("SARS-CoV-2 spike protein with C9 tag at C-terminal expressed in mammalian cells") (Shang et al., 2020).

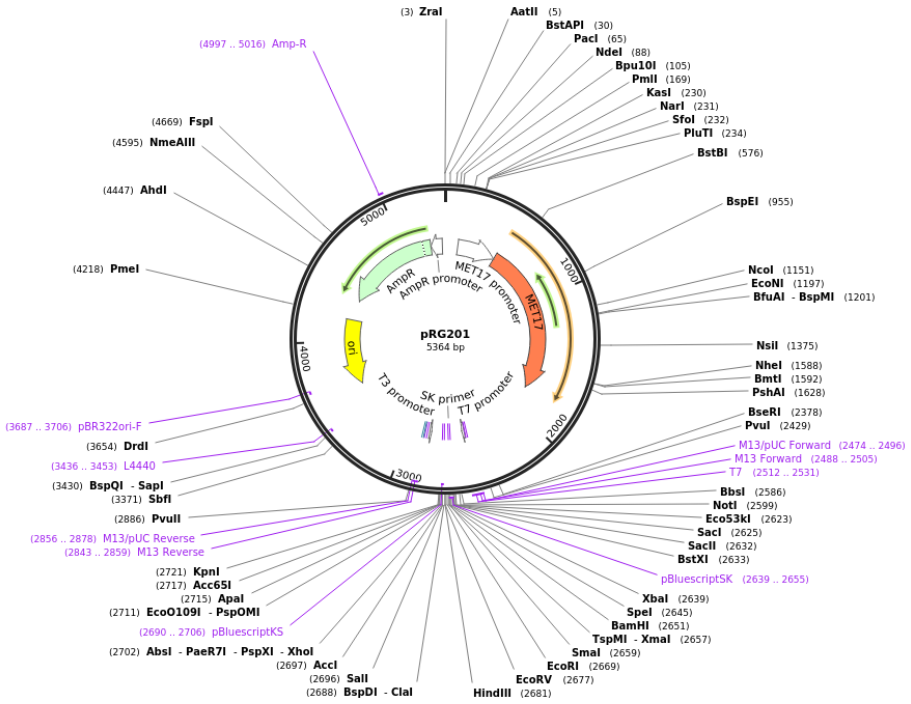

Figure 3.3:"pRG-201 (Empty Backbone) Integrative E. coli/S. cerevisiae shuttle vector" (Gnügge, Liphardt, & Rudolf, 2016).

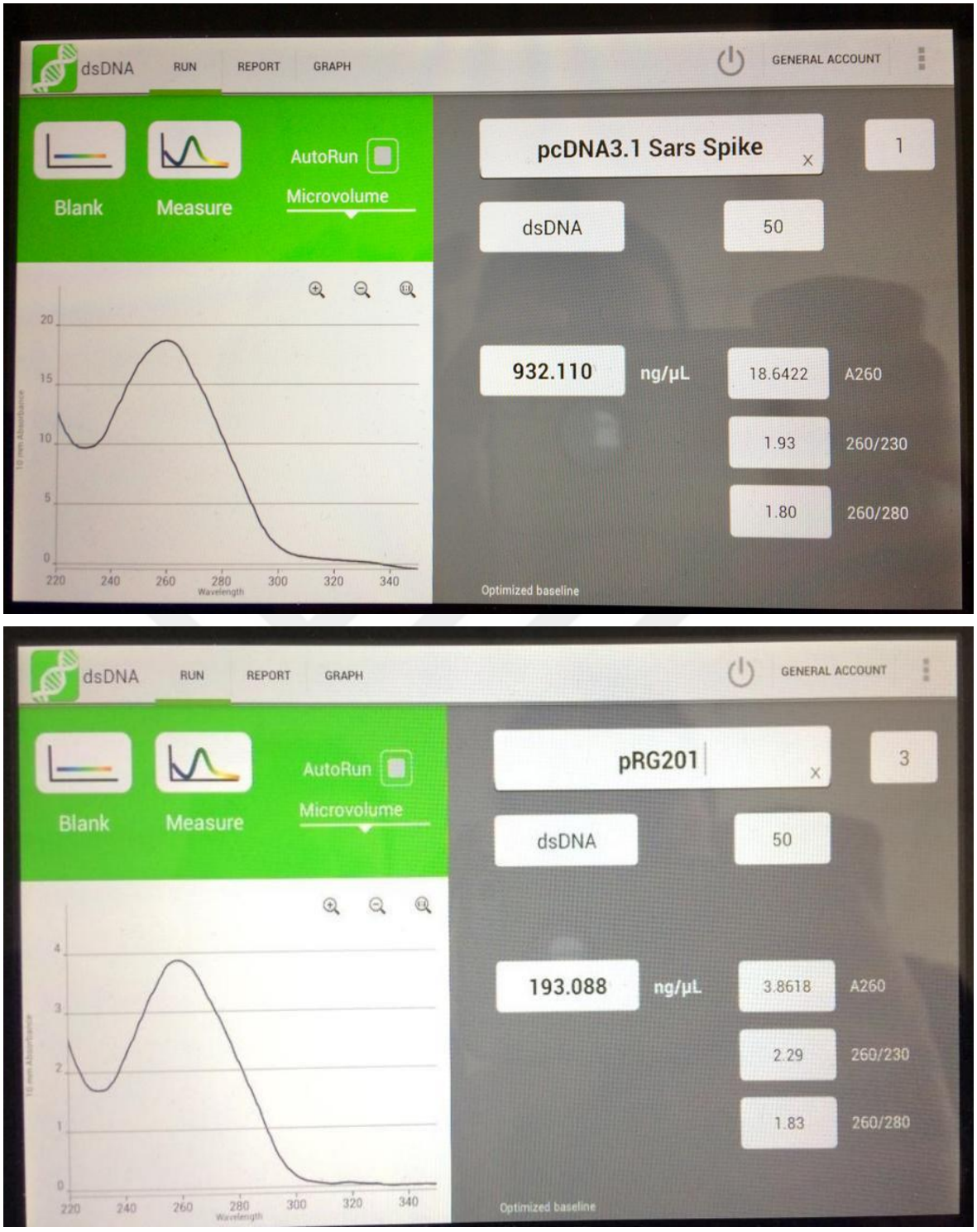

Figure 3.4: Purity and quantity of plasmids pcDNA3.1-SARS 2-Spike (above panel) and pRG201.

3.3 Restriction, Ligation and Gel Electrophoresis.

Plasmids pcDNA3.1-SARS 2-Spike (Figure 3.2) and pRG201 (Figure 3.3) were restriction double digested with the same enzymes (XbaI and SaCI) at the multi cloning site (MCS). The larger fragment of pRG201 (5348 bp) which is the result of the removal of 16 bp

sequence during the digestion of whole pRG201 plasmid and the Spike gene (3944 bp) on the pcDNA3.1-SARS 2-Spike plasmid were purified from gel by using a "Thermo Scientific GeneJET Gel Extraction Kit''. The fragments were combined to construct a recombinant shuttle vector with a size 9295 bp and expression capacity in both bacteria and yeast. The result showed that (Figure 3.3) the Spike gene was successfully integrated into the region between the XbaI and SaCI restriction sites in MCS, resulting a recombinant plasmid named as pRG-Spike in this study (Figure 3.9).

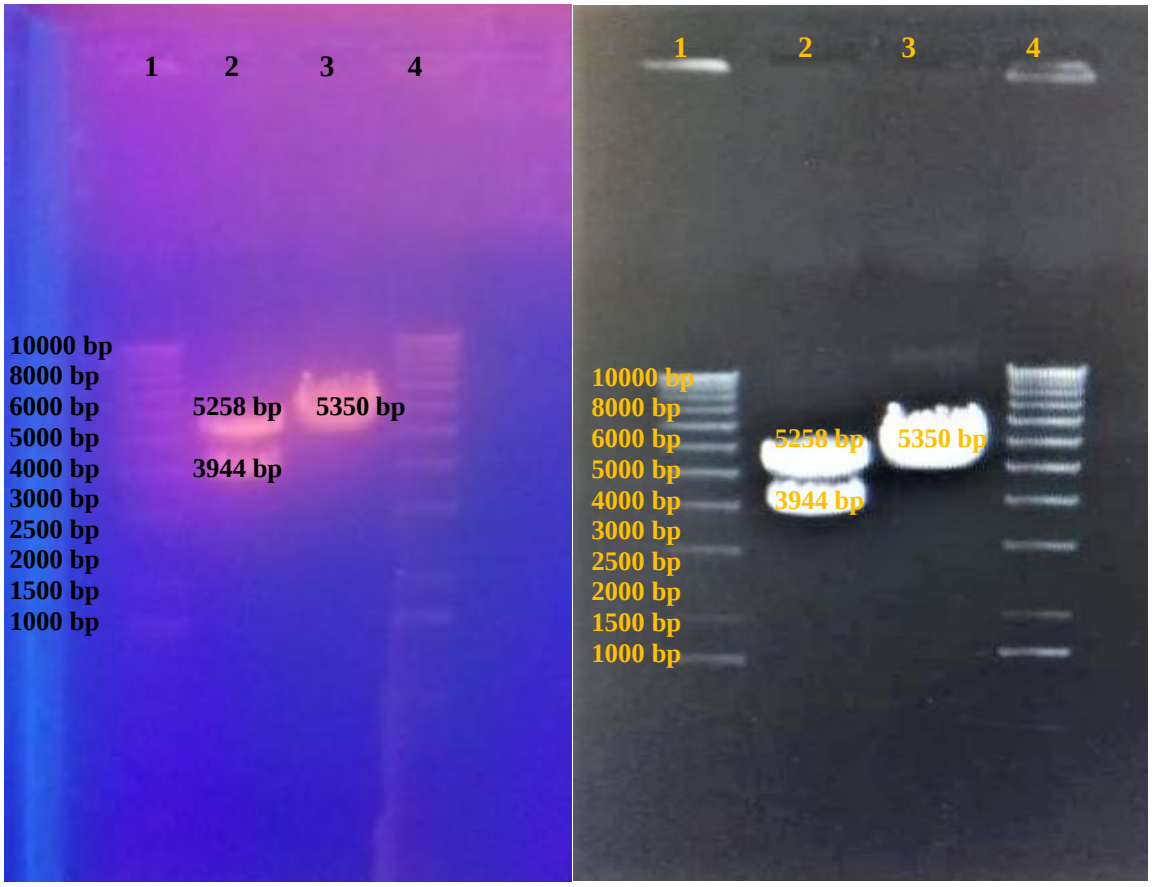

Figure 3.5: Double (SacI and XbaI) digests of pcDNA3.1-SARS 2-Spike and PR201 DNA fragments. Line 1 are DNA markers, line 2 is pcDNA3.1-SARS 2-Spike plasmid (9202 bp), line 3 is pRG201 plasmid (5364 bp) and line 4 is also DNA markers.

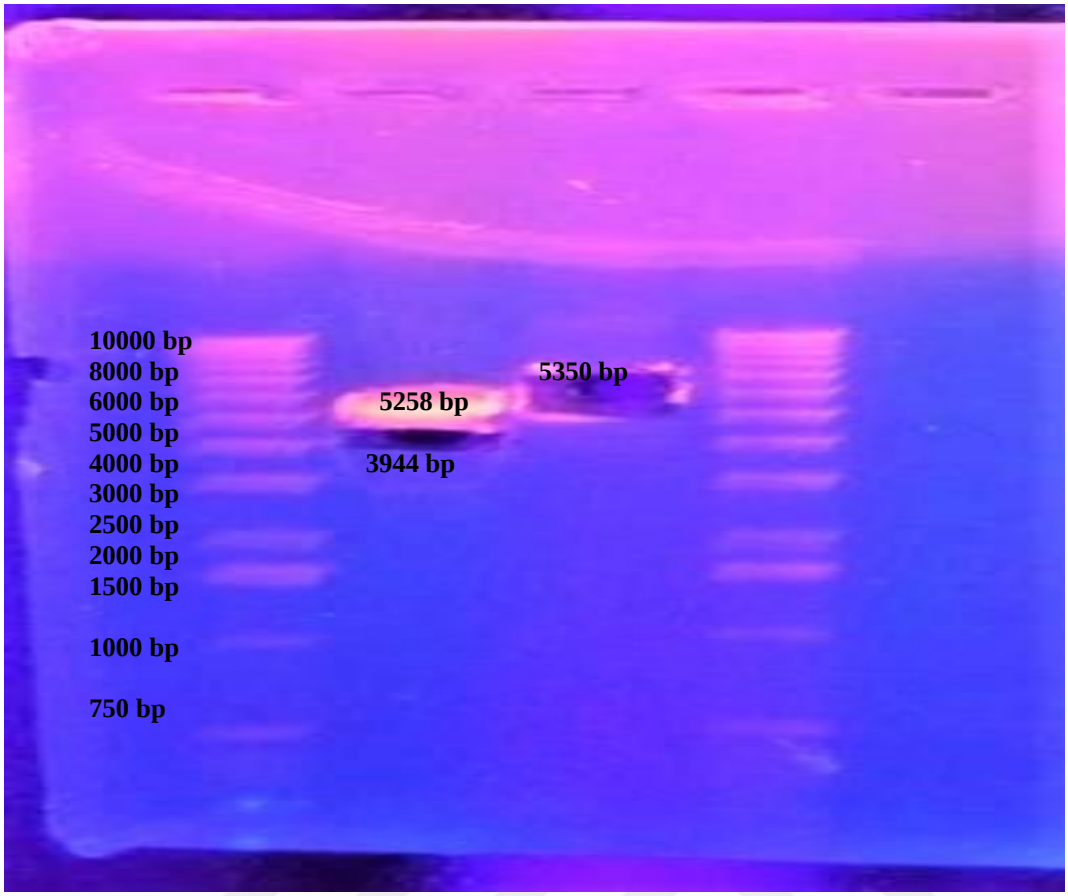

Figure 3.6: Plasmid fragments cut from the gel with a scalpel.

The purity of DNA (plasmids) isolated from the gel were measured with a Denovix DS-11 FX+ Spectrophotometer.

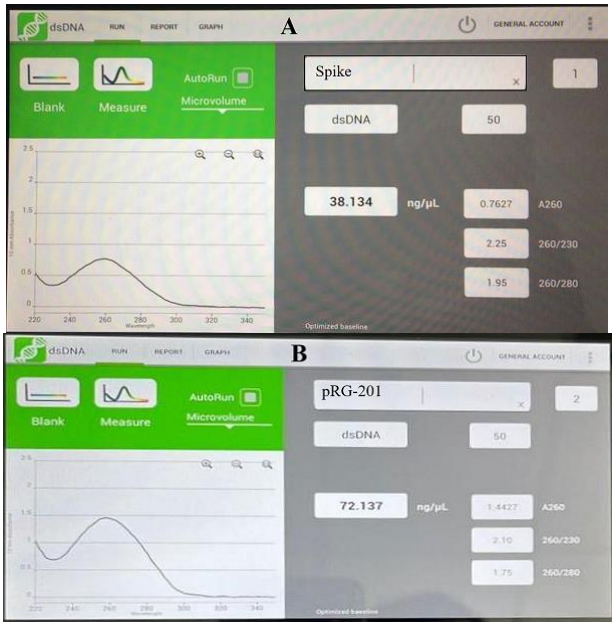

Figure 3.7: The purity and the quantity of isolated Spike gene (A) and pRG201 (B).

To confirm the newly constructed recombinant shuttle vector (Prg-Spike) obtained after the ligation reaction of the plasmid fragments, first ligated plasmids were measured with a Denovix DS-11 FX+ Spectrophotometer. The plasmids (donor and acceptor) then digested with XbaI and SacI restriction enzymes and the restriction fragments were purified from the gel.

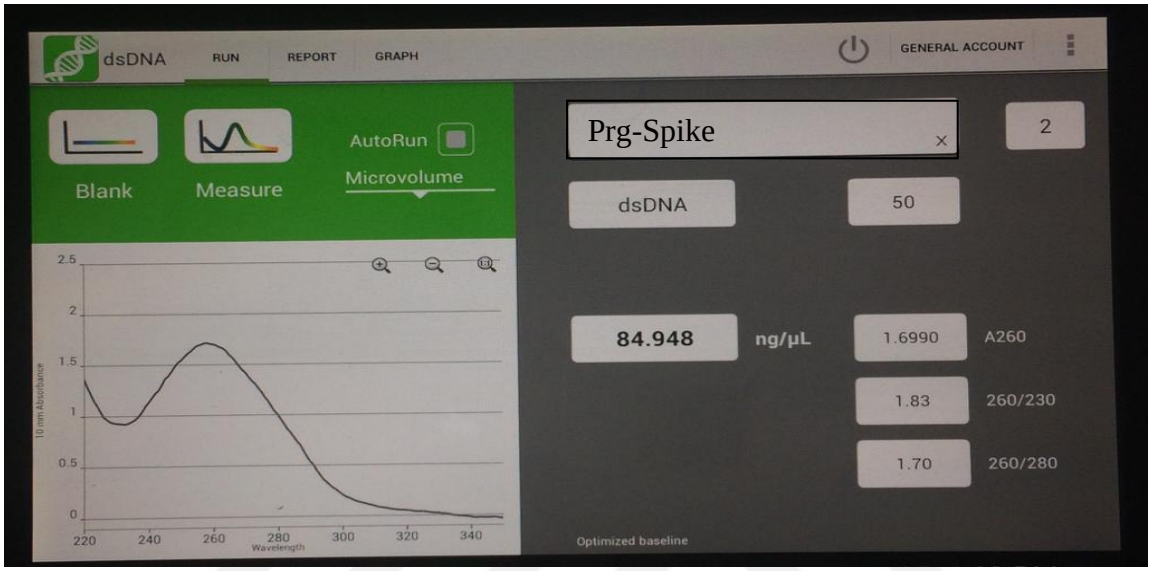

Figure 3.8: The purity and the quantity of constructed "Prg-Spike" plasmid.

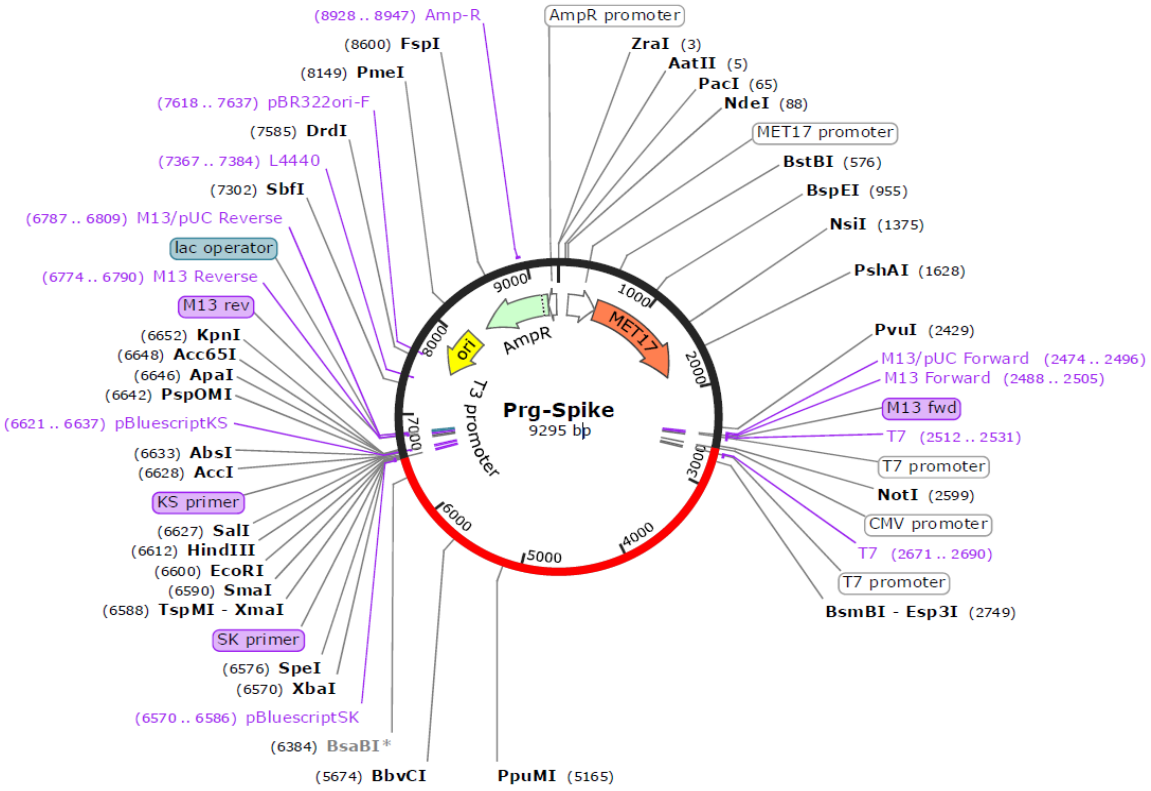

Figure 3.9: The recombinant plasmid “pRG-Spike” constructed in this study. The red line indicates the location where the Spike gene is cloned.

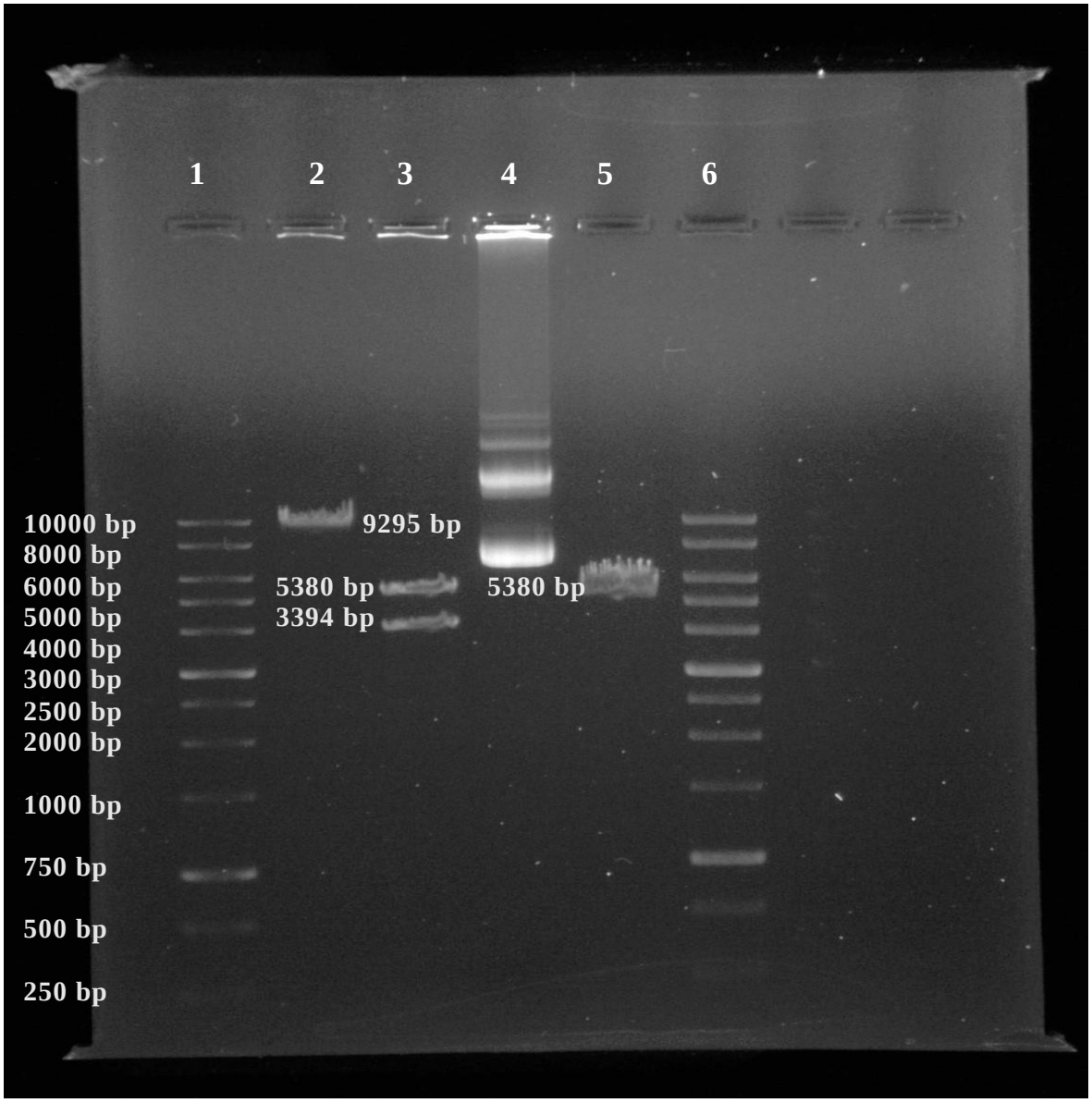

Figure 3.10: The restriction digest of the recombinant "pRG-Spike" plasmid (line 3). Line 1 are DNA markers, Line 2 is pRG-Spike with XbaI digest, line 4 is uncut pRG-Spike, line 5 is pRG201 (acceptor plasmid) double digested with SacI and XbaI, and line 6 are also DNA markers.

3.4 Growing Bacterial Cultures

Bacterial stocks at -20°C previously prepared were grown on LB agar by incubation at 37°C for 12 h. Then, a single colony was taken into LB medium and incubated at 37°C for 12 h.

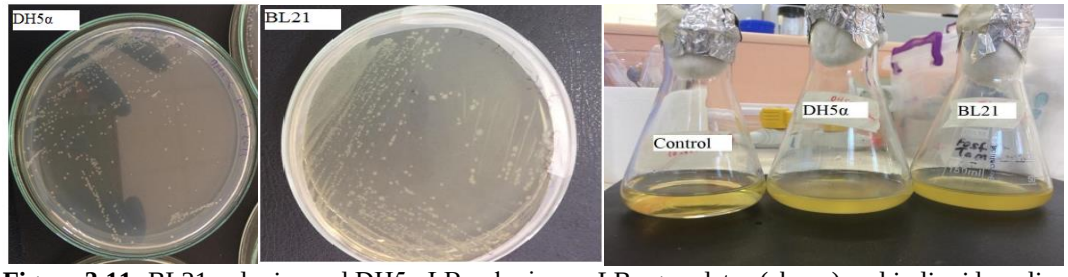

Figure 3.11: BL21 colonies and DH5α LB colonies on LB agar plates (above) and in liquid media.

3.4.1 Preparation of Cells for Transformation

Cells were prepared to make them transformation competent according to the ‘‘Hanahan Competent Cell Protocol Prepared’’. They were stored at –80°C for future use. The constructed new recombinant plasmid (Prg-Spike) was transformed into these competent cells.

3.4.2 Transformation of recombinant Prg-Spike plasmid into competent *E.coli* DH5α cells

The newly constructed 5 µl of recombinant Prg-Spike plasmid was transfard into an Eppendorf tube with 100 µl *E. coli* DH5α competent cells. Then, as given in Materials and Methods the transformants were selected based on their growth and satellite forming characteristics. Selected colonies were streaked on LB amp agar plates (Figure 3.12) and single colonies were selected to incubate in LB apm medium overnight 37°C. Plasmids were isolated from the fresh cultures grown at 37°C for 12h.

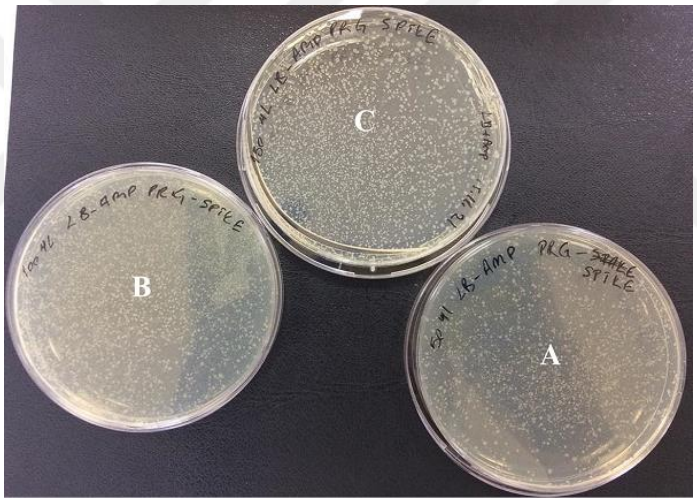

Figure 3.12: Prg-Spike plasmid bearing DH5α cell colonies on LB amp agar. Colonies formed from 50 µl (A), 100 µl (B), and 150 µl (C) cell culture.

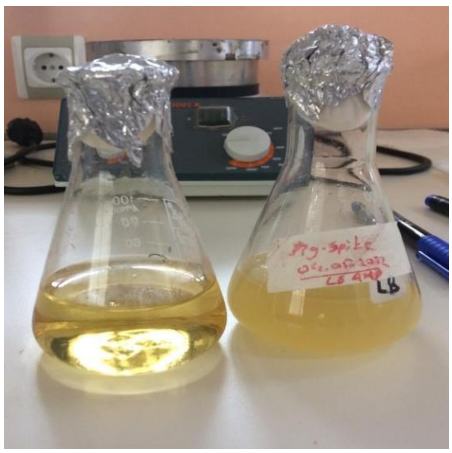

Figure 3.13: DH5α cells bearing Prg-Spike plasmid in LB amp medium. (left: control).

3.4.3 Isolation of Prg-Spike plasmid from transformed DH5α cells

Plasmid isolation was performed from transformed DH5α cells. Isolations were made with the "QIAprep Spin Miniprep Kit (50)". The plasmids obtained after isolation were first measured with a Denovix DS-11 FX+ Spectrophotometer, and their purity and amount were determined. Then, they were cut with XbaI and SacI restriction enzymes, and their size was differentiated by agarose gel electrophoresis with the help of electric field. In this way, the accuracy of the plasmid was confirmed.

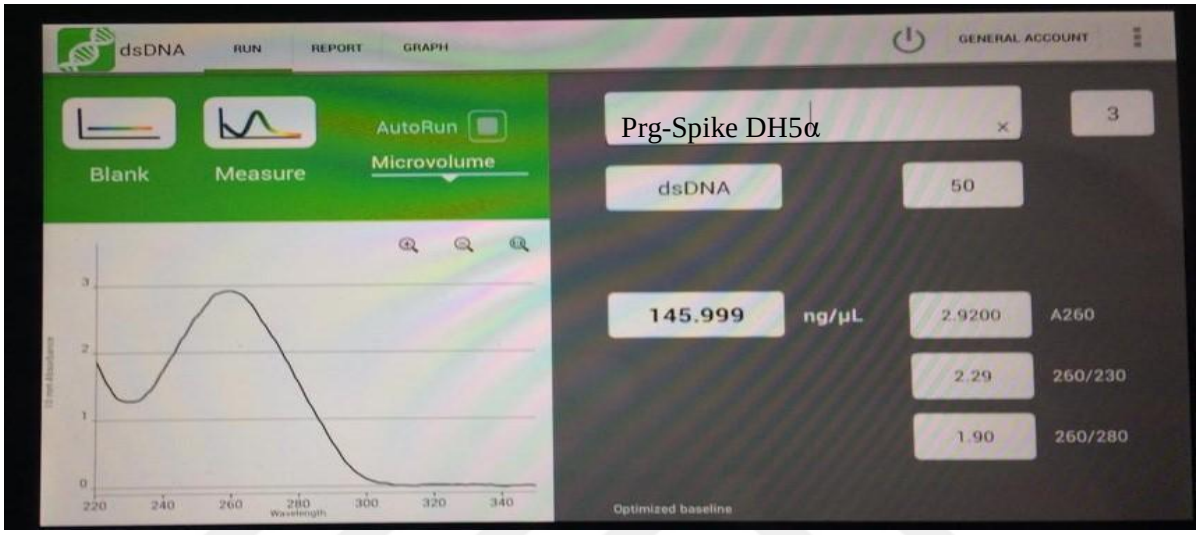

Figure 3.14: Isolation of Prg-Spike plasmid from DH5α transformed cell. The purity of the plasmids was confirmed by absorbance ratios of 260/230 and 260/280.

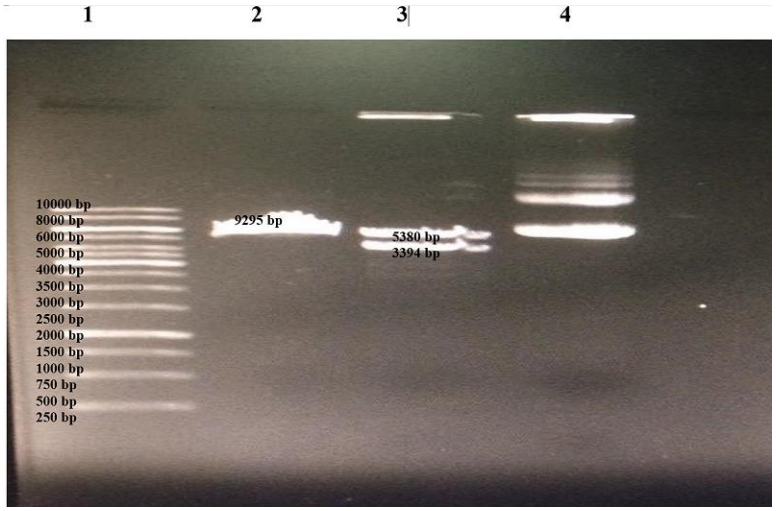

Figure 3.15: Line 1 are DNA markers, Line 2 is pRG-Spike with XbaI digest, (9295 bp). The restriction digest of the recombinant "pRG-Spike" plasmid (line 3). Line 4 pRG-Spike non-digest plasmid (9295 bp) in supercoil structure.

3.4.4 Transformation of Prg-Spike plasmid into competent *Escherichia coli*BL21 cells

The high copy Prg-Spike plasmid from DH5α cells was transformed into competent BL21 cells. BL21 cell line is a special cell line with T7 RNA polymerase activity. In this way, Spike protein was produced with the Spike gene in our plasmid. The transformation was

done on the principle of heat shock. Transformed cells were first inoculated on LB amp agar plates. Selection was made by antibiotic resistance. The seeded cell plates were incubated at 37°C for 12 h. A large number of colonies formed after incubation. A single colony was taken from the created colonies by half of the loop, inoculated into LB amp medium and incubated for 12 h at 37°C. Cells were observed to grow after incubation.

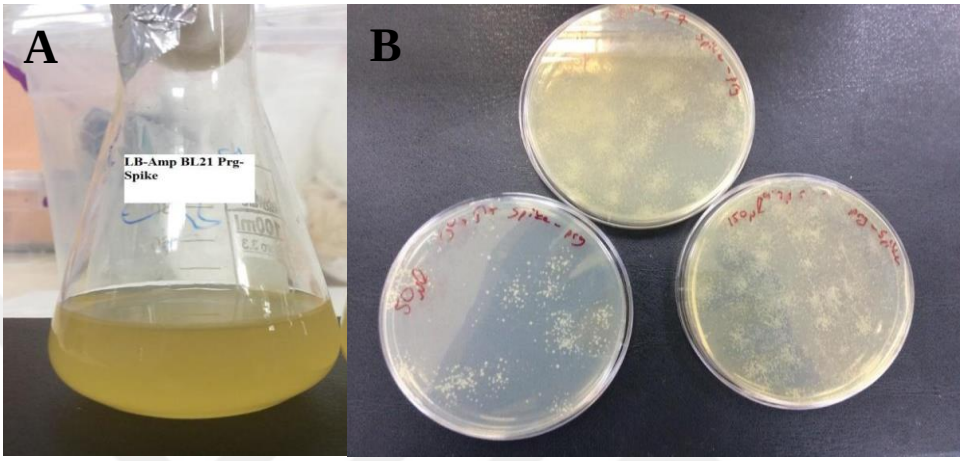

Figure 3.16: (A) BL21 growing on LB amp medium, (B) BL21 growing on LB amp agar.

3.4.5 Isolation of spike protein from transformed BL21 cells and its analysis by SDS-PAGE

Spike protein was obtained from transformed BL21 cells. First, T7 RNA polymerase was activated by stimulating the transformed cells with IPTG. Afterwards, the cells activated the cloned Spike gene present in the shuttle vector and protein production was ensured. Cells were first treated with Lysis buffer to isolate the produced Spike protein. Afterwards, they were exposed to ultrasonic waves with the help of a sonication device, and the cell membrane was broken. Finally, other cellular structures were precipitated, except for the Spike protein, by centrifugation, and protein was isolated from the supernatant.

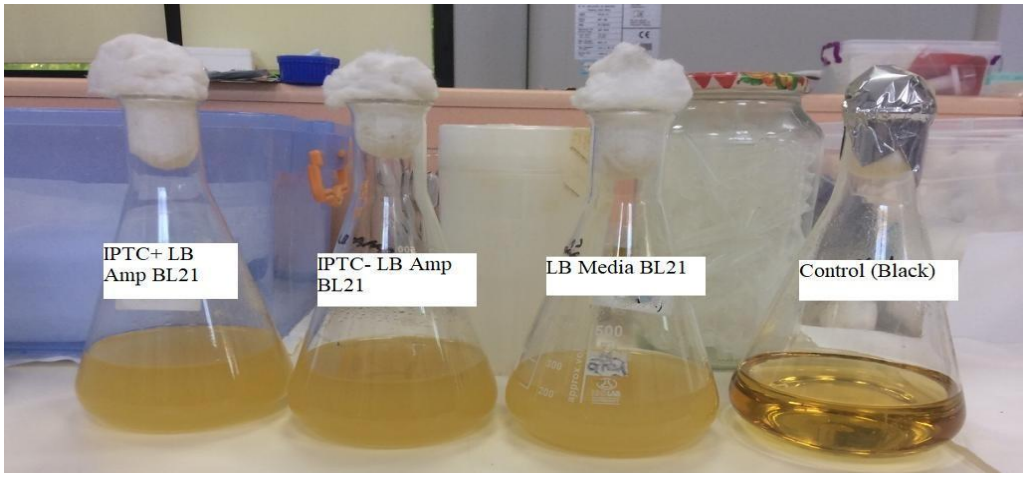

Figure 3.17: Cell cultures.

Figure 3.18: The sonication apparatus.

The total proteins obtained were boiled in 2X Leammli Sample Buffer (Bio-Rad) for 5 min at 90°C. Then SDS-PAGE installation was done. Loadings were done in 3 repetitions (Figure 3.15). Loading was done as BL21 normal strain, BL21 pRG-Spike IPTC- and BL21 pRG-Spike IPTC+, respectively. The gel was run for 1 h at 90 Volts. As a result, Spike protein was observed in the gel in IPTC+ able BL21.

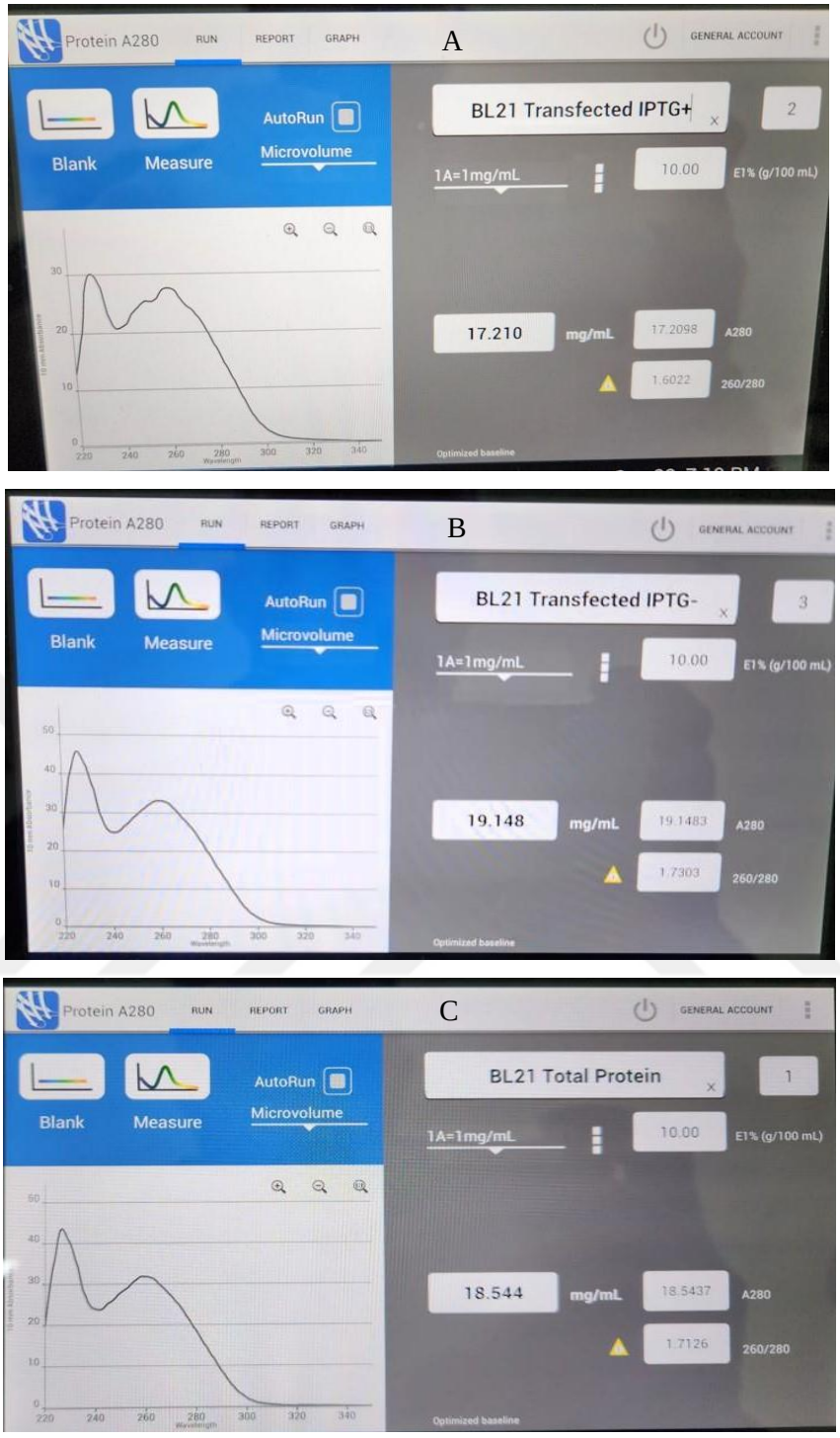

Figure 3.19: Total protein in IPTG induced (A) and non-induced (B) BL21 cells. (C) BL21 total protein amount.

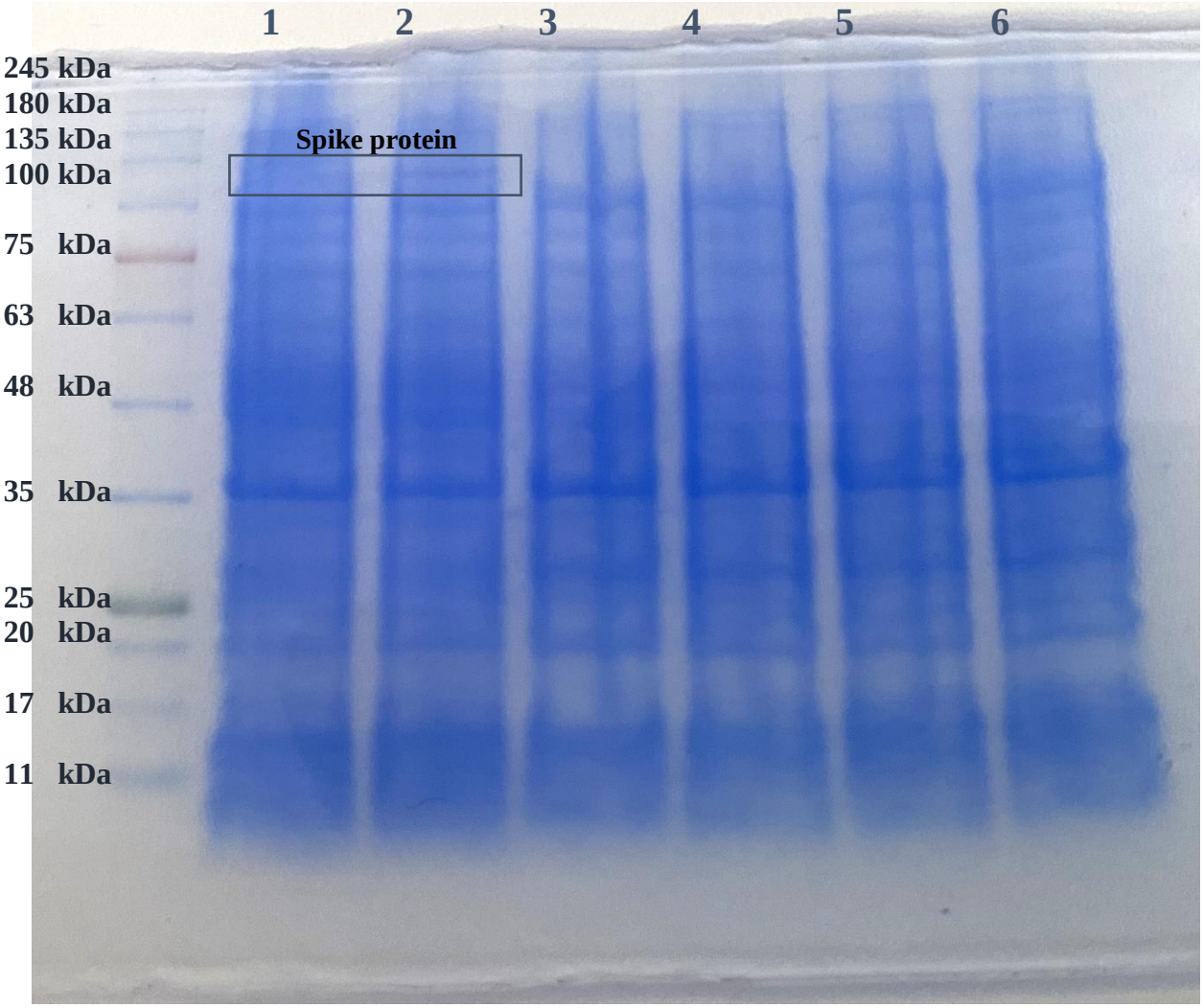

Figure 3.20: Gel image as a result of SDS-PAGE. Lines 1 and 2: total proteins of cells transformed into BL21 cells with Prg-Spike plasmid and stimulated with IPTG; lines 3 and 4 lines: total proteins of cells transformed into BL21 cells with Prg-Spike plasmid and not stimulated with IPTG; lines 5 and 6: total proteins of wild-type BL21 cells.

3.5 Transformation pRG-Spike Plasmid into Plasmid of Yeast BY4741

The pRG-Spike plasmid obtained as a result of cloning has the ability to express both bacteria and yeast. The cell line used for yeast transformation is *Saccharomyces cerevisiae* BY4741. This cell line cannot grow when it is not in Met Ura Leu and His medium. The YNB media we prepared contain Leu, His and Ura. Met is our selective maker. Since there is a Met17 promoter in the plasmid, cells that receive the plasmid will grow, but those that do not, could not reproduce because they could not produce Met. In this way, colony selection was made.

The transformation process was done by the Lithium Acetate method. Transformed cells were first seeded on YNB Ura^+, Leu^+ and His^+ agar. It was incubated at 30°C for 12 h. colonies formed. Then, YNB Ura^+, Leu^+ and His^+ medium were seeded and incubated at 30°C for 2 h. Cell proliferation was observed.

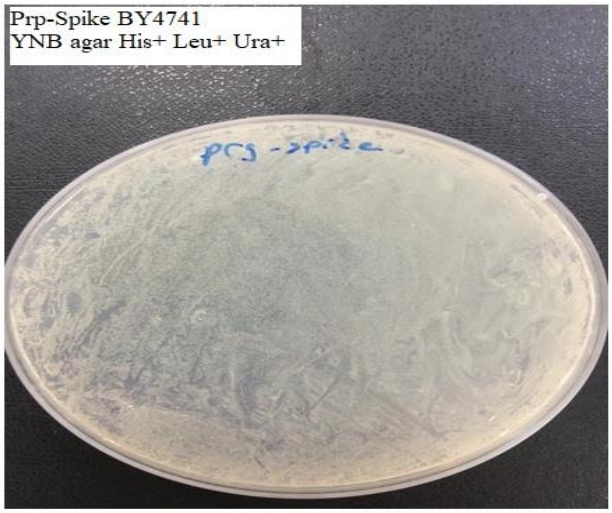

Figure 3.21: BY4741 cells transformed into pRG-Spike plasmid YNB Ura^+, Leu^+ and His^+ agar.

Figure 3.22: BY4741 cells transformed into pRG-Spike plasmid YNB Ura^+, Leu^+ and His^+ medium.

3.5.1 Isolation of spike protein from transformed yeast BY4741 and its analysis by SDS-PAGE

Spike protein was isolated from the obtained transformed yeast cell culture based on a protein extraction from yeast (Kushnirov, 2000) method, adapted for this study. Transformed yeast cells obtained first were taken from solid YNB His^+ Ura^+ Leu^+ medium was inoculated into YNB His^+ Ura^+ Leu^+ 20 ml medium in a 50 ml flask. The culture was incubated at 30°C for 12 h (Figure 3.20). 50 ml of yeast culture was transferred to the Falcon tube and centrifuged at 4000 rpm for 5 min and boiled in "1X Leammli buffer" for 5 min (Uk, 1970),

the total cell protein was electrophoresed using SDS-PAGE gel (Figure 3.23).

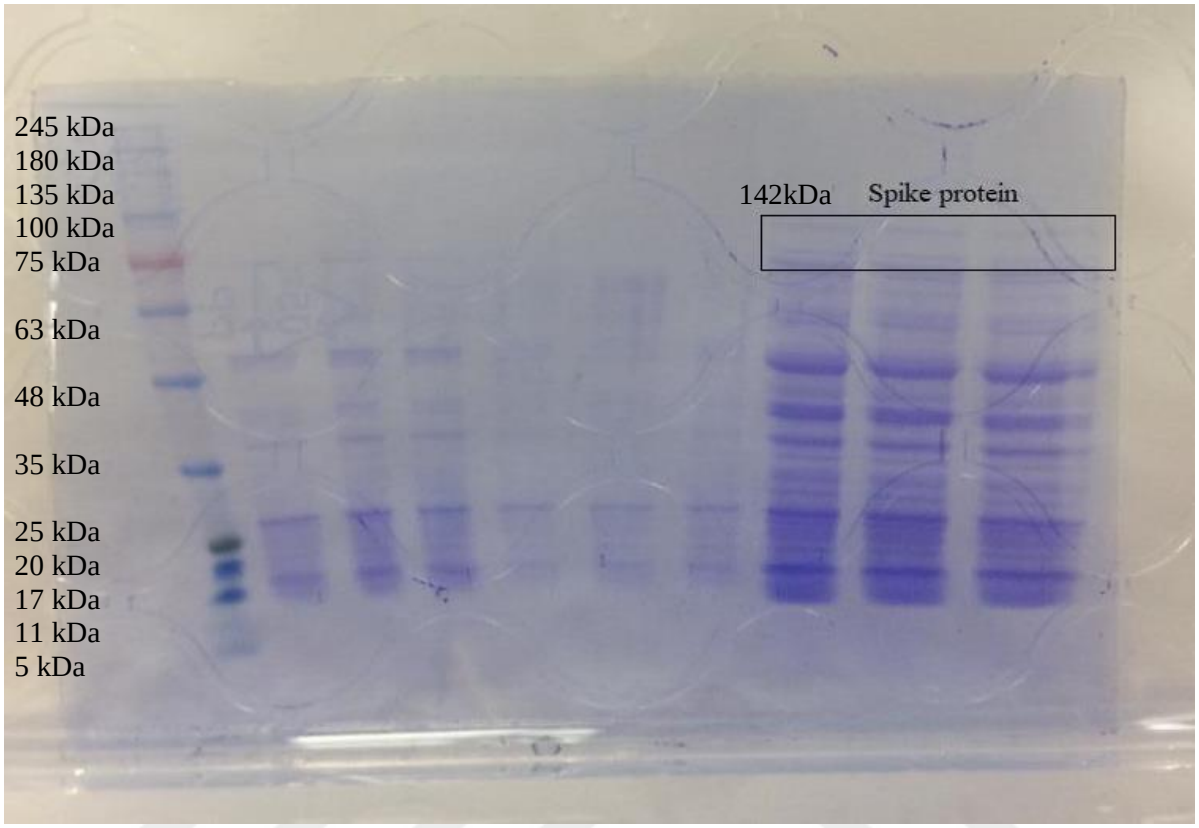

Figure 3.23: SDS-PAGE Gel image of Spike protein from transformed yeast cells.

4. DISCUSSION

The main goal of this work was to recombinantly produce and characterize the SARS-CoV-2 Spike protein in bacteria and fungi. A recombinant shuttle vector was constructed which can be expressed in bot bacteria (*E. coli* BL21 and *Saccharomyces cerevisiae* BY4741) to produce the Spike (S) protein of SARS-CoV-2, which is the main target of all vaccines to date. Producing the S protein at scales in microorganisms such as bacteria and yeast may have various advantages and applications. Given that bacteria and fungi can easily be manipulated and grown under low cost conditions, producing the S protein in these organisms have its advantages. Such production at scales may help to understand the basic structure, function of this protein and thus design better antiviral drugs, vaccines, and monoclonal antibodies. Also, the protein can be used in a rapid, cheap, and easy way in antigen-based test kits and in diagnostic tests such as lateral flow and ELISA, compared to time consuming, expertise needed and expensive PCR based test.

There have been similar studies using recombinant DNA technology and producing the Spike gene in bacteria for formulization of vaccines during the early stages of the COVID-19 pandemic in 2021 (Keech et al., 2020; Polack et al., 2020).

Thus, cloning the SARS-CoV-2 Spike gene in a shuttle vector for expression of the Spike protein in both bacteria and yeast at scalable level could serve for various applications, such as for vaccine development, diagnostic tests, studying the structure and function of the protein, comperative studies on post-translational modifications as yeast, shares some eukaryotic features with human cells. Also the protein can be used to facilitate drug screening and testing of potential antiviral compounds. Not only purified protein, but also both bacterial and yeast systems expressing this protein can be used to screen for inhibitors to understand the effects of various drugs or chemicals on the Spike protein expression. The recombinant Spike protein produced can be used as an antigen for serological tests, for antibody development, and immune response studies. Research on the Spike protein in different hosts can inform the design variant-specific vaccines to address emerging strains of SARS-CoV-2. In summary, cloning the SARS-CoV-2 Spike gene into a shuttle vector for expression in both bacterial and yeast systems opens up various avenues of research and development for understanding the viral pathogenesis, diagnosis, and medical interventions.

REFERENCES

Alanagreh, L. A., Alzoughool, F., &Atoum, M. (2020). The human coronavirus disease COVID-19: its origin, characteristics, and insights into potential drugs and its mechanisms. *Pathogens*, *9*(5), 331.

Alsahafi, A. J., & Cheng, A. C. (2016). The epidemiology of Middle East respiratory syndrome coronavirus in the Kingdom of Saudi Arabia, 2012–2015. *International Journal of Infectious Diseases*, *45*, 1-4.

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, *26*(4), 450-452.

Arya, R., Kumari, S., Pandey, B., Mistry, H., Bihani, S. C., Das, A., Kumar, M. (2021). Structural insights into SARS-CoV-2 proteins. *Journal of molecular biology*, *433*(2), 166725.

Bhadoria, P., Gupta, G., & Agarwal, A. (2021). Viral pandemics in the past two decades: an overview. *Journal of family medicine and primary care*, *10*(8), 2745.

Boson, B., Legros, V., Zhou, B., Siret, E., Mathieu, C., Cosset, F.L., Denolly, S. (2021). The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein, allowing assembly of virus-like particles. *Journal of Biological Chemistry*, *296*.

Burrell, C. J., Howard, C. R., & Murphy, F. A. virology, (2017). Coronaviruses. *Fenner and white's medical virology*, 437.

Chai, J., Cai, Y., Pang, C., Wang, L., McSweeney, S., Shanklin, J., & Liu, Q. (2021). Structural basis for SARS-CoV-2 envelope protein recognition of human cell junction protein PALS1. *Nature Communications*, *12*(1), 3433.

COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (**JHU**). Johns Hopkins University. 2023.

Cubuk, J., Alston, J. J., Incicco, J. J., Singh, S., Stuchell-Brereton, M. D., Ward, M. D., & Holehouse, A. S. (2021). The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature communications*, *12*(1), 1936.

Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature reviews microbiology*, *17*(3), 181-192.

Dinman, J. D. (2006). Programmed ribosomal frameshifting goes beyond viruses: organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift. *Microbe (Washington, DC)*, *1*(11), 521.

Dutta, N. K., Mazumdar, K., & Gordy, J. (2020). The nucleocapsid protein of SARS–CoV-2: a target for vaccine development. *Journal of virology*, *94*(13)

Fontanet, A., Autran, B., Lina, B., Kieny, M. P., Karim, S. S. A., & Sridhar, D. (2021). SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet*, *397*(10278), 952-954.

Gnügge, R., Liphardt, T., & Rudolf, F. (2016). A shuttle vector series for precise genetic engineering of Saccharomyces cerevisiae. *Yeast*, *33*(3), 83-98.

Green, M. R., & Sambrook, J. (2018). The Hanahan method for preparation and transformation of competent *Escherichia coli*: high-efficiency transformation. *Cold Spring Harbor Protocols*.

Guo, G., Gao, M., Gao, X., Zhu, B., Huang, J., Luo, K., & Lou. (2021). SARS-CoV-2 non-structural protein 13 (nsp13) hijacks host deubiquitinase USP13 and counteracts host antiviral immune response. *Signal transduction and targeted therapy*, *6*(1), 119.

Hayman, D. T., & Knox, M. A. (2021). Estimating the age of the subfamily Orthocoronavirinae using host divergence times as calibration ages at two internal nodes. *Virology*, *563*, 20-27.

Hu, B.,Guo, H., Zhou, P., & Shi, Z.L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, *19*(3), 141-154.

Hu, Y., Wen, J., Tang, L., Zhang, H., Zhang, X., Li, Y., & Yang (2003). The M protein of SARS-CoV: basic structural and immunological properties. *Genomics, proteomics & bioinformatics*, *1*(2), 118-130.

Huang, Y., Yang, C., Xu, X.-f., Xu, W., & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 1141-1149.

Jackwood, M. W., Hall, D., & Handel, A. (2012). Molecular evolution and emergence of avian gammacoronaviruses. *Infection, Genetics and Evolution*, *12*(6), 1305-1311.

Keech, C., Albert, G., Cho, I., Robertson, A., Reed, P., Neal, S., & Glenn, G. M. (2020). Phase 1–2 trial of a SARS-CoV-2 recombinant spike protein nanoparticle vaccine. *New England Journal of Medicine*, *383*(24), 2320-2332.

Kushnirov, V. V. (2000). Rapid and reliable protein extraction from yeast. *Yeast*, *16*(9), 857-860.

Kuzmin, A., Orekhov, P., Astashkin, R., Gordeliy, V., & Gushchin, I. (2022). Structure and dynamics of the SARS-CoV-2 envelope protein monomer. *Proteins: Structure, Function, and Bioinformatics*, *90*(5).

Leammli, UK. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*, 680-685.

Leao, J. C., Gusmao, T. P. d. L., Zarzar, A. M., Leao Filho, J. C., Barkokebas Santos de Faria, A., Morais Silva, I. H., & Carvalho, A. D. A. T. (2022). *Coronaviridae:* Old friends, new enemy. *Oral Diseases*, *28*, 858-866.

Li, D.D., & Li, Q.H. (2021). SARS-CoV-2: vaccines in the pandemic era. *Military Medical Research*, *8*(1), 1.

Mahtarin, R., Islam, S., Islam, M. J., Ullah, M. O., Ali, M. A., & Halim, M. A. (2022). Structure and dynamics of membrane protein in SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, *40*(10), 4725-4738.

Mandala,V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., & Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nature structural & molecular biology*, *27*(12), 1202-1208.

Moeller, N. H., Shi, K., Demir, Ö., Belica, C., Banerjee, S., Yin, L., ... & Aihara, H. (2022). Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease ExoN. *Proceedings of the National Academy of Sciences*, *119*(9), e2106379119.

Nassar, A., Ibrahim, I. M., Amin, F. G., Magdy, M., Elgharib, A. M., Azzam, E. B., & Elfiky, A. A. (2021). A review of human coronaviruses' receptors: the host-cell targets for the crown bearing viruses. *Molecules*, *26*(21), 6455.

Mobini, S., Chizari, M., Mafakher, L., Rismani, E., & Rismani, E. (2021). Structure-based study of immune receptors as eligible binding targets of coronavirus SARS-CoV-2 spike protein. *Journal of Molecular Graphics and Modelling*, *108*, 107997.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., & Qian, Z. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature communications*, *11*(1), 1620.

Qiu, W., Rutherford, S., Mao, A., Chu, C, (2017). The pandemic and its impacts. *Health, culture and society*,*9*, 1-11.

Polarck, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., & Gruber, W. C. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England journal of medicine*, *383*(27), 2603-2615

Pollett, S., Conte, M. A., Sanborn, M., Jarman, R. G., Lidl, G. M., Modjarrad, K., & Maljkovic Berry, I. (2021). A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. *Scientific Reports*, *11*(1), 17365.

Rabaan, A. A., Al-Ahmed, S. H., Haque, S., Sah, R., Tiwari, R., Malik, Y. S., …&Rodriguez-Morales,A.J. (2020). SARS-CoV-2, SARS-CoV, and MERS-COV: a comparative overview. *Infez Med*, *28*(2), 174-184.

Rahimi, A., Mirzazadeh, A., &Tavakolpour, S. J. G. (2021). Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics*, *113*(1), 1221-1232.

Rahman, M. S., Hoque, M. N., Islam, M. R., Islam, I., Mishu, I. D., Rahaman, M. M., &**Hossain, M. A.** (2021). Mutational insights into the envelope protein of SARS-CoV-2. *Gene reports*, *22*, 100997.

Santos-Sánchez, N. F., & Salas-Coronado, R. (2020). Origin, structural characteristics, prevention measures, diagnosis and potential drugs to prevent and COVID-19. *Medwave*, *20*(8).

Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. *Virology journal 16*(1), 1-22.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221-224.

Thao, T.T.N., Labroussaa, F., Ebert, N., V'kovski, P., Stalder, H., Portmann, J.,& Thiel, V. (2020). Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*, *582*(7813), 561-565.

Thomas, S. (2020). The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter SemiSWEET. *Pathogens and Immunity*, *5*(1), 342.

Velusamy, P., Kiruba, K., Su, C.-H., Arun, V., Anbu, P., Gopinath, S. C., & Vaseeharan, B. (2021). SARS-CoV-2 spike protein: Site-specific breakpoints for the development of COVID-19 vaccines. *Journal of King Saud University-Science 33*(8), 101648.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260-1263.

Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J., & Prescott, H. C. (2020). Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA*, *324*(8), 782-793.

Wu, W., Cheng, Y., Zhou, H., Sun, C., & Zhang, S. (2023). The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. *Virology Journal*, *20*(1), 1-16.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature 579*(7798), 265-269.

Xiaolu, T., Changcheng, W., Xiang, L., Yuhe, S., Xinmin, Y., Xinkai, W., & Jian, L.(2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review7*(6), 1012-1023.

Xu, R. H., He, J. F., Evans, M. R., Peng, G. W., Field, H. E., Yu, D. W., ... & Schnur, A. (2004). Epidemiologic clues to SARS origin in China. *Emerging infectious diseases*, *10*(6), 1030.

Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., & Lou, Z. (2020). Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nature communications*, *11*(1), 5874.

Ye, Q., West, A. M., Silletti, S., & Corbett, K. D. (2020). Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Science*, *29*(9), 1890-1901.

Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., & Jin, T.(2020). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochemical and biophysical research communications*, *527*(3), 618-623.

Zhang, J., Xiao, T., Cai, Y., & Chen, B. (2021). Structure of SARS-CoV-2 spike protein. *Current opinion in virology*, *50*, 173-182.

Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Current Biology*, *30*(7), 1346-1351. e1342.

Zhang, Z., Nomura, N., Muramoto, Y., Ekimoto, T., Uemura, T., Liu, K., & Shimizu, T. (2022). Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nature communications*, *13*(1), 4399.

Zizza, A., Recchia, V., Aloisi, A., & Guido, M. (2021). Clinical features of COVID-19 and SARS epidemics. A literature review. *Journal of Preventive Medicine and Hygiene*, *62*(1), E13.

CURRICULUM VITAE

Name- Surname: Okan Özşahin

EDUCATION STATUS:

Primary school: Atatürk Primary School/ Malatya

High school: Hacı Ahmet Akıncı Anadolu High school/ Malatya

University Bachelor: Cumhuriyet University / Sivas

Master's Degree: İnonü University / Malatya

Native Language: Turkish

Other Language: Intermediate English

Computer skills: Word, Exel, Paint, Power Point and Snap gene Software.

PROJECTS: Research scholarship through a TÜBİTAK (Turkish Scientific and Technical Research Institute) 1001 Project. Project no 120Z962: Developments of SARS-CoV-2 (COVID-19) Specific Rapid Diagnostic Strip Test Platforms (2023).