

THE SEARCH FOR POTENTIAL ANTIVIRALS OF SARS-  
COV-2 SPIKE PROTEIN: A VIRTUAL SCREENING STUDY

BASMA MOHAMED REHAN

FACULTY OF SCIENCE  
UNIVERSITI OF MALAYA  
KUALA LUMPUR

2023

**THE SEARCH FOR POTENTIAL ANTIVIRALS OF  
SARS-COV-2 SPIKE PROTEIN: A VIRTUAL  
SCREENING STUDY**

**BASMA MOHAMED T REHAN**

**DISSERTATION SUBMITTED IN PARTIAL  
FULFILMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF MASTER OF BIOTECHNOLOGY**

**I INSTITUTE OF BIOLOGICAL SCIENCES  
FACULTY OF SCIENCE  
UNIVERSITI OF MALAYA  
KUALA LUMPUR**

**2023**

**UNIVERSITY OF MALAYA**  
**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: **BASMA MOHAMED REHAN**

Matric No: **17221626**

Name of Degree: **MASTER OF SCIENCE BIOTECHNOLOGY**

Title of Dissertation (“this Work”): **THE SEARCH FOR POTENTIAL ANTIVIRALS OF SARS-CoV-2 SPIKE PROTEIN: A VIRTUAL SCREENING STUDY**

Field of Study: **BIOINFORMATICS**

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature

Date: 6/04/2023

Subscribed and solemnly declared before,

Witness’s Signature

Date: 6/4/2023

Name:

Designation:

# THE SEARCH FOR POTENTIAL ANTIVIRALS OF SARS-COV-2 SPIKE PROTEIN: A VIRTUAL SCREENING STUDY

## ABSTRACT

The COVID-19 outbreak caused by a new coronavirus emerged in December 2019 and has since spread across the globe. The spike protein of SARS-CoV-2 is the main target in research therapeutic to control viral infection and pathogenesis. Inhibition of spike protein can cause structural and functional changes, subsequently blocking the entry of the coronavirus into host cells. The in-silico virtual screening is an alternative to traditional screening in drug discoveries with a significant reduction to cost and time. This study aimed to screen NuBBE database to identify a potential inhibitor against spike protein. A library of natural-based compounds was docked on spike protein obtained from the protein data bank. Ten compounds were selected based on their binding affinity for further analysis. The ligand bound in close proximity to the positive control binding site was chosen as a promising compound and subjected for molecular dynamics simulation and ADMET scoring. the result indicated that (spike-Lig126) has a stable conformation with hydrogen bonds formation and ADMET properties. The inhibitory properties of Lig126 require further investigation in the laboratory to prove it and it could finally be upgraded to an anti-SARS drug.

**Keywords:** SARS-CoV-2, spike protein, structure-based virtual screening, antiviral, COVID-19, in-silico drug discovery.

## PENCARIAN ANTIVIRAL BERPOTENSI PROTEIN SPIKE SARS-COV-2

### MELALUI SARINGAN MAYA

#### ABSTRAK

Wabak COVID-19 yang disebabkan oleh coronavirus baharu muncul pada Disember 2019 dan sejak itu telah merebak ke seluruh dunia. Protein spike SARS-CoV-2 adalah sasaran utama dalam terapeutik penyelidikan untuk mengawal jangkitan virus dan patogenesis. Perencatan protein spike boleh menyebabkan perubahan struktur dan fungsi, seterusnya menyekat kemasukan coronavirus ke dalam sel perumah. Pemeriksaan maya dalam silico adalah alternatif kepada pemeriksaan tradisional dalam penemuan dadah dengan pengurangan ketara kepada kos dan masa. Kajian ini bertujuan untuk menyaring pangkalan data NuBBE untuk mengenal pasti perencat yang berpotensi terhadap protein spike. Sebuah perpustakaan berasaskan sebatian semula jadi didok pada protein spike yang diperoleh daripada bank data protein. Sepuluh sebatian telah dipilih berdasarkan keupayaan pengikatannya untuk analisis selanjutnya. Ligan yang terikat berdekatan dengan tapak pengikat kawalan positif telah dipilih sebagai sebatian yang menjanjikan dan tertakluk kepada simulasi dinamik molekul dan skor ADMET. Lig126 didapati membentuk ikatan H yang stabil dengan protein spike dan mempamerkan sifat keserupaan dadah mengikut analisis ADMET. Sifat perencatan Lig126 memerlukan kajian lanjutan di makmal untuk membuktikan, dan memberikan petunjuk dan data penting untuk ligan yang diperhatikan.

**Kata Kunci:** SARS-CoV-2, protein spike, saringan maya berasaskan struktur, antivirus, COVID-19.

## ACKNOWLEDGEMENTS

Firstly, I want to thank GOD for his blessing to complete this project.

Secondly, I would like to give my thanks and deepest appreciation to Professor. Dr. Saharuddin Bin Mohamad, for providing invaluable encouragement and guidance. He was patient and was always there when I needed his support and advice.

A special thanks to the coordinator of my master program Dr. Nikman Adli Bin Nor Hashim for his supporting and assisting.

Also, Sincere thankful to my parents, family and friends for all the support. Also, thanks for Libyan government for giving me the opportunity to learn at the Universiti Malaya.

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>iii</b>
<b>ABSTRAK</b> .....	<b>ivv</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>v</b>
<b>TABLE OF CONTENTS</b> .....	<b>vi</b>
<b>LIST OF FIGURES</b> .....	<b>vix</b>
<b>LIST OF TABLES</b> .....	<b>vix</b>
<b>LIST OF SYMBOLS AND ABBREVIATIONS</b> .....	<b>xix</b>
<b>LIST OF APPENDICES</b> .....	<b>xxiv</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>1</b>
1.1 Overview .....	1
1.2 Objectives.....	2
<b>CHAPTER 2: LITERATURE REVIEW</b> .....	<b>3</b>
2.1 The SARS-CoV-2 Virus .....	3
2.1.1 Background of COVID-19.....	4
2.1.2 The Origin and Evolution of SARS-CoV-2 .....	5
2.1.3 SARS-CoV-2 Structural Proteins.....	6
2.1.4 The potential Targets For Antiviral Therapy Against SARS-CoV-2 .....	7
2.2 Spik Protein.....	7
2.2.1 The Structural Analysis Of SARS-CoV-2 Spike Protein.....	10
2.2.2 The structure of S1 .....	10
2.2.3 The structure of S2.....	11
2.2.4 Viral Fusion.....	13
2.2.4 Spike Protein Functions.....	13

2.3 COVID-19 Infections.....	14
2.4 COVID-19 Treatments.....	15
2.5 Natural Compounds .....	16
2.6 The Database Of Natural Products.....	17
2.7 Structure-Based Virtual Screening.....	18
2.8 Molecular Docking and Molecular Dynamics Simulations .....	19
<b>CHAPTER 3: METHODOLOGY.....</b>	<b>21</b>
3.1 Workstation Equipment .....	21
3.2 Protein Preparation.....	23
3.3 Ligand Library Preparation .....	23
3.4 Virtual Screening .....	24
3.4.1 Positive Control Preparation .....	24
3.5 Molecular Docking.....	25
3.6 Molecular Dynamics Simulations.....	26
3.6.1 Protein MD Simulation .....	26
3.6.1Complex MD Simulation .....	26
3.7 Prediction of ADMET for Positive Control and Selected Compound.....	28
<b>CHAPTER 4: RESULTS.....</b>	<b>29</b>
4.1 Protein Preparation.....	29
4.2 Ligand Library Preparation .....	32
4.3 Virtual Screening .....	32
4.3.1 Molecular Docking.....	34
4.3.2 protein-ligand complexes analysis .....	34
4.4 Molecular Dynamics Simulations .....	42
4.5 Prediction of ADMET for Positive Control and Selected Compound.....	43



<b>CHAPTER 5: DISCUSSION</b> .....	<b>48</b>
5.1 Protein Preparation.....	48
5.2 Ligand Library Preparation.....	49
5.3 Virtual Screening .....	50
5.4 Molecular Dynamics Simulations.....	51
5.5 ADMET .....	52
<b>CHAPTER 6: CONCLUSION</b> .....	<b>54</b>
<b>REFERENCES</b> .....	<b>55</b>

Universiti Malaya

## LIST OF FIGURES

Figure 2.1	:Structure of SARS-CoV-2 virus.....	4
Figure 2.2	:SARS-CoV-2 spike schematic .....	9
Figure 2.3	:The S protein monomer in the open conformation.....	9
Figure 2.4	:Schematic illustrate the attachment and fusion of S protein. ....	12
Figure 2.5	:An illustration of virtual screening .....	20
Figure 3.1	:Workflow of structure-based virtual screening .....	22
Figure 4.1	:Structure of SARS-CoV-2 spike protein obtained from the PDB.....	31
Figure 4.2	:Structure of chain A of SARS-CoV-2 spike protein .....	31
Figure 4.3	:Number of selected natural compound from NuBBE database.....	33
Figure 4.4	:positive control .....	33
Figure 4.5	:Protein-ligand complexes of selected ligands and positive control.....	39
Figure 4.6	:(A) protein-lig126 interaction, (B) protein-positive interaction.....	40
Figure 4.7	:Interactions of protein-lig126 complex.....	41
Figure 4.8	:Interactions of protein-positive complex.....	41
Figure 4.9	:potential energy, Temperature and pressure .....	44
Figure 4.10	:Trajectory plots. (A) RMSD, (B) RMSF (C) H-bond for complex.....	45

## LIST OF TABLES

Table 4.1	:Summary of non-mutated S protein PDBs available in protein data bank.	30
Table 4.2	:Used criteria to filter NuBBE database under rule of five.....	33
Table 4.3	:The top ten binding affinity compounds.....	35
Table 4.4	:Protein-ligand interaction of the positive control and selected ligands .....	35
Table 4.5	:Information on selected ligands. ....	36
Table 4.6	:ADEMT information for Lig-120 and positive. ....	47

Universiti Malaya

## LIST OF SYMBOLS AND ABBREVIATIONS

$\alpha$	:	Alpha
Å	:	Angstrom
%	:	Percent
3D	:	Three dimensional
K	:	Kelvin
ADMET	:	Absorption, Distribution, Metabolism, And Excretion
ACE2	:	Angiotensin-Converting Enzyme 2
Ala	:	Alanine
Arg	:	Arginine
Asn	:	Asparagine
Asp	:	Aspartic acid
BBB	:	Blood-brain barrier
CADD	:	Computer aided drug design
COVID-19	:	Coronavirus disease 2019
D	:	Dalton
E	:	Envelope
FDA	:	U.S. Food and Drug Administration
FP	:	Fusion Peptide
GI	:	Gastrointestinal
Gly	:	Glycine
Gln	:	Glutamine
Glu	:	Glutamic acid
GROMACS	:	GRoningen Machine for Chemical Simulations
GPU	:	Graphics processing unit

H	:	Hydrogen
H-bond(s)	:	Hydrogen bond(s)
His	:	Histidine
HTS	:	High-throughput screening
HR1/HR2	:	Heptad Repeats
HuNAbs	:	Human Neutralizing Antibodies
kcal/mol	:	kilocalorie per mole
kJ/mol	:	kiloJoule per mole
Met	:	Methionine
Leu	:	Leucine
LGA	:	Lamarckian genetic algorithm
logP	:	lipophilicity
Lys	:	Lysine
NMR	:	Nuclear magnetic resonance
Na <sup>+</sup>	:	Sodium ion
nm	:	Nanometer
NP	:	Natural product
NPT	:	Number of particles, pressure and temperature
ns	:	nanosecond
NTD	:	N-terminal domain
NuBBE	:	Nuclei of Bioassays, Ecophysiology, Biosynthesis of Natural Products
NVT	:	Number of particles, volume and temperature
NTD	:	N-terminal domain
PDB	:	Protein Data Bank
Phe	:	Phenylalanine
Pro	:	Proline

Ps	:	Picosecond
PME	:	Particle Mesh Ewald
R-factor	:	residual factor
RMSD	:	Root-mean-square deviation
RMSF	:	Root-mean-square fluctuation
RNA	:	Ribonucleic acid
SARS-CoV2	:	The Severe Acute Respiratory Syndrome Coronavirus-2
SBDD	:	Structure-based drug discovery
SBVS	:	Structure Based Virtual Screening
Ser	:	Serine
SMILES	:	simplified molecular input line entry system
ssRNA	:	Single-stranded RNA
Thr	:	Threonine
Tyr	:	Tyrosine
TMPRSS2	:	Transmembrane serine protease 2
Val	:	Valine
VS	:	Virtual Screening
WCS	:	Whole-cell screening
WHO	:	World Health Organization

## LIST OF APPENDICES

Appendix A: Spike protein sequence obtained from Protein Data bank PDB...6

Appendix B: Summary of 514 natural compound found in NuBBE database...7

Universiti Malaya

## CHAPTER 1: INTRODUCTION

### 1.1 Overview

The epidemic of novel coronavirus disease COVID-19 was caused by a new coronavirus which found in 2019, December and now has spread worldwide and turned into a global pandemic (Chan et al., 2020). The severe acute respiratory coronavirus 2 and member of  $\beta$  coronavirus family, was immediately identified as the cause behind COVID-19 (Lai et al., 2020). Because of the high mortality rate and the World Health Organization's designation of SARS-CoV-2 as a pandemic, coronavirus disease 2019 have a high hospitalization rate, according to epidemiological data from Yale and BioRender research team. The coronavirus is spherical in shape, having spike proteins around them. These proteins are responsible for virus replication in human host cells. After attaching with human cells, spike proteins undergo structural changes, resulting in a fusion of viral particle membrane with the human host cell membrane. Thus, the viral RNA enters the host cell and produces more viruses after copying its genome. These steps of SARS-CoV-2 entry mechanism make contribution to its fast spreading, severity in symptoms and greater death rates (Shang et al., 2020). This indicates, the spike protein is important in the viral infection and therefore is suitable to be used as target for the drug development (Wrapp et al., 2020). Although the virus first surfaced in December 2019, the consequences of the pandemic are still being felt. However, it now attacks at a higher rate than other respiratory viral diseases due to the identification of new SARS CoV-2 strains that are more contagious than the original strain. (John E and Bennett MD., 2020). To control this breakout, numerous researchers from around the world are working to create or repurpose antiviral medications using computational and experimental techniques.



In 2021, a research study on antiviral activities of guanidine alkaloids against SARS-CoV2 was reported which utilizes in-silico structure-based virtual screening (El-Demerdash et al., 2021). Protease, nucleocapsid phosphoprotein, non-structural glycoprotein, and spike protein are analyzed proteins. The binding affinities against all compounds depicted promising results. The MD simulations were also used to verify the most potent compound's stability against target proteins. The simulation results open access to a further analysis against the prevention of SARS-CoV-2 (El-Demerdash et al., 2021). In this project, a virtual screening analysis was conducted on the SARS-CoV-2 spike protein against the natural based compounds from NuBBE database to identify a Natural compound with a good binding affinity compared to reported positive control. the selected compound complex was then simulated to study its conformation stability. This in-silico study aimed to search for SARS-CoV-2 spike proteins potential natural-based antivirals through virtual screening additionally, this helps to rapidly identify the suitable compound and helps to save over all drug discovery time and cost.

## **1.2 Objectives**

1. To identify possible antiviral compounds from NuBBE database with the ability to bind on spike protein of SARS-CoV-2 by applying in-silico structure-based virtual screening approach.
2. To analyse the protein-ligand interaction generated between the potential inhibitor and the spike protein.
3. To evaluate the stability of complex (protein-ligand) and its respective binding energy.

## CHAPTER 2: LITERATURE REVIEW

### 2.1 The SARS-CoV-2 Virus

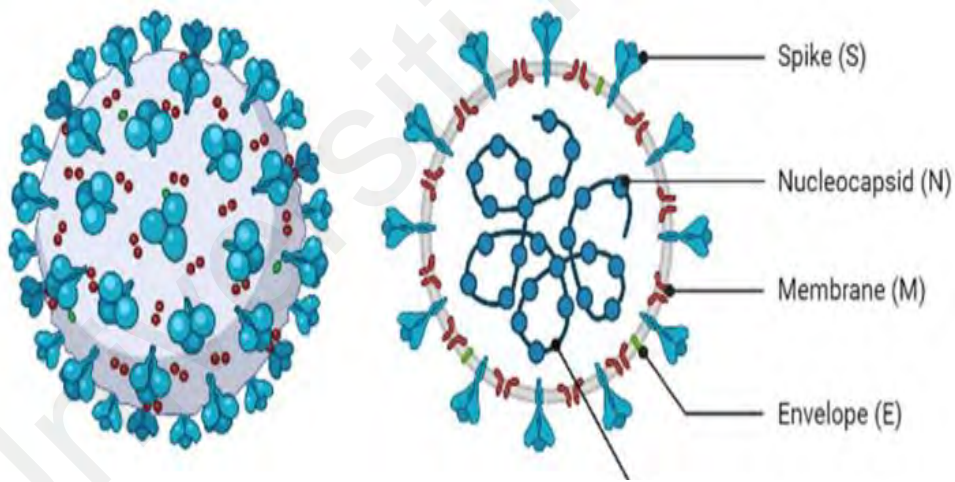
The first novel infectious disease discovered was severe acute respiratory syndrome (SARS). In November 2002, this acute and often severe respiratory ailment first appeared in province of China. In April 2003, a worldwide effort coordinated by WHO resulted in the discovery of a new coronavirus. Sars coronavirus 2 responsible for the outbreak. coronavirus illness causes in 2019 (COVID19). The pandemic's effects are still being seen even though the virus first appeared in 2019, December. However, the discovery of new coronavirus strains that are more transmissible than the initial strain has jeopardized this comeback (Harrison et al., 2020).

The SARS-CoV-2 is an encapsulated single-stranded RNA. structural and non-structural proteins encoded by its genome RNA, replicas of the polyprotein and structural proteins. A zoonotic virus (SARS-CoV) dwells in bats, which act as its natural reservoir. However, before infecting people, HPV may also infect intermediate hosts like small animals (like palm civets). A range of human cell types are susceptible to SARS-CoV infection and reproduction, which can lead to severe pathogenic changes (Naqvi et al., 2020). Target cells are accessed by SARS-CoV-2 through endosomes. The cell receptor angiotensin-converting enzyme 2 interacts to S protein (ACE2). The ACE2 is then transported to the endosomes, where endosome acid proteases break the S protein to trigger its fusion activity. Viral proteinases split the pp1a and 1ab viral replication polyproteins into smaller products once the viral genome has been released and translated. Genomic RNA templates are created using the comprehensive negative-strand template. Genomic RNA is converted into viral nucleocapsids in the cytoplasm, and N protein is subsequently budded into the ERGIC (endoplasmic reticulum Golgi intermediate compartment) (Romano et al., 2020).

### 2.1.1 Background of COVID-19

In December 2019, China was the first country deduced the novel coronavirus that caused several respiratory diseases, including pneumonia. The World Health Organization (WHO) suggested a language to characterize the virus when it was firstly formerly as new COVID-19. SARS coronavirus 2 have been identified as the illness's causative agent (severe acute respiratory syndrome coronavirus 2). The High Consequence Infectious Disease called COVID-19 is airborne and contagious. (Ciotti et al., 2020).

People across the world are propagating SARS-CoV-2, according to the WHO's dashboard for daily status reports. Vaccines are currently widely accessible., the viral disease cannot be treated with antibiotics. (U. K. H. S., 2022).



**Figure 2.1: Structure of SARS-CoV-2 virus.**

### 2.1.2 The Origin and Evolution of SARS-CoV-2

Through bioinformatics research, it was shown that SARS-CoV-2 has characteristics with the coronavirus family. It belongs to the beta coronavirus 2B lineage (Lai et al., 2020). Quickly in China, Scientists were sequencing the genome of virus for a number of COVID-19 patient . The findings indicated that 79.5% similarity was found with SARS-CoV which is the SARS-CoV2 a newfangled strain. This novel beta is considered to be the source of this disease(Zhou et al., 2020). The sequence of virus genome was aligned with with previously published beta coronavirus. This virus and the BatCov RaTG13 have the closest relationships, with a 96 percent identity. These studies suggest that SARS-CoV-2 may have evolved spontaneously from the bat coronavirus RaTG13 and may have come from bats (Wang et al., 2020; Zhang et al., 2020). The SARS-CoV-2 virus belongs to genus “Betacoronavirus”. SARS2-COV-2 shows similarities with the SARS-COV with respect to genome (structure), pathogenesis and tissue tropism. SARS-CoV-2 is more transmissible virus and yet variations regarding immunity responses are not cleared (Harrison et al., 2020).

Coronaviruses belong to the RNA enveloped viruses and the main cause of chronic and acute respiratory diseases in animals and humans. The total six species of coronavirus, four of them show common symptoms as “Cold” and two of them cause human diseases named as MERS-CoV (Middle East respiratory syndrome coronavirus) and SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus) (Peiris et al., 2003).

In 2002 to 2003, the globally epidemic caused by SARS-CoV in 26 countries, it affected greater than 8000 patients and near about 800 deaths. The similarities between SARS-CoV and SARS-CoV2 have been proved in 2013 studies (Graham et al., 2013). Though, recurrent epidemics of MERS, SARS and COVID-19 have depicted that the CoVs can transmit from other species to humans (Kim et al., 2020).SARS-CoV-2 shows closed relationship against two species as SARS-CoV and BAT-derived SARS

coronaviruses regarding phylogeny distancing (Lu et al., 2020). The similarity between SARS-CoV-2 and SARS-CoV shows 76% similarity index and 96% similarity in their main proteases while the other similarities between these two viruses are: symptoms development and way of infection (Entry mechanism) (Verma et al., 2020). The receptor is also same for both coronaviruses: (SARS-CoV and SARS-CoV2) as ACE2 (Angiotensin-Converting Enzyme 2), which shows similarity in tropism and mechanism of entry (Lan et al., 2020).

### **2.1.3 SARS-CoV-2 Structural Proteins**

The SARS-CoV-2 virus has a single-stranded envelope RNA. Metagenomic next-generation sequencing was applied for identifying full 29,881-bp genome (GenBank accession number MN908947), encoding 9860 amino acids (Lu R et al., 2020). Gene segments express structural and nonstructural proteins. The S, E, M, and N genes encode structural proteins, whereas the ORF area encodes nonstructural proteins such as 3-chymotrypsin-like protease, papain-like protease, and RNA-dependent RNA polymerase (Chan et al., 2020). Structurally, this virus consists of 4 structural proteins, which are spike glycoprotein (S), nucleocapsid protein (N), envelop protein (E), transmembrane glycoprotein (M). In contrast, the N protein holds the genome (Arya et al., 2021). The SARS-CoV-2 surface is covered with many glycosylated S proteins that bind to the angiotensin-converting enzyme 2 (ACE2) receptor to facilitate viral cell entry (Letko et al., 2020). When the S protein binds to the receptor, TM protease serine 2 (TMPRSS2), a type 2 TM serine protease present on the host cell membrane, the virus infects a cell. These proteins are potential targets for medical treatment since they are crucial to the viral life cycle (Morse et al., 2020). Spike protein is one of structural proteins that crucial for virus assembly and structure in addition to the pathogenicity of viruses (Gralinski et al., 2018). The S glycoprotein has a strong impact on the pathogenic phenotype and viral tropism. It has been established that the S protein is the main mediator of SARS-CoV-2

binding to ACE2 receptor in human cells and causing membrane fusion, which plays a crucial role in viral entrance into cells (Walls et al., 2020). The M protein is crucial for stabilizing the viral structure, envelope development, viral budding, and viral release in SARS-CoV. In order to aid in the viral replication, the nucleocapsid (N) protein attaches to the viral RNA. It has been shown that the E protein has a virulence domain that causes the SARS-CoV infection to cause immunopathology (Jimenez Guardeno et al., 2014).

#### **2.1.4 The Potential Targets for Antiviral Therapy Against SARS-CoV-2**

A complete analysis of the current developments of developing interventional methods and possible targets based on the categories of "target virus" and "target host" was also undertaken. The specifics of their crystal structures may play significant roles in the SARS-CoV-2 replication cycle, even more significant than the typical types among coronaviruses, but might determine the efficacy of the antiviral approaches. SARS-CoV-2 structural proteins, particularly S protein, remain the most promising direct antiviral targets (Cerutti et al., 2021). Certain viral Nsps are required for virus replication as well as virus–host interactions, and they might be used as indirect targets for antiviral treatments. As a result, deciphering the precise unique structures of viral proteins as well as the diseased biological pathogenesis might lead to the discovery of new COVID-19 treatment targets.

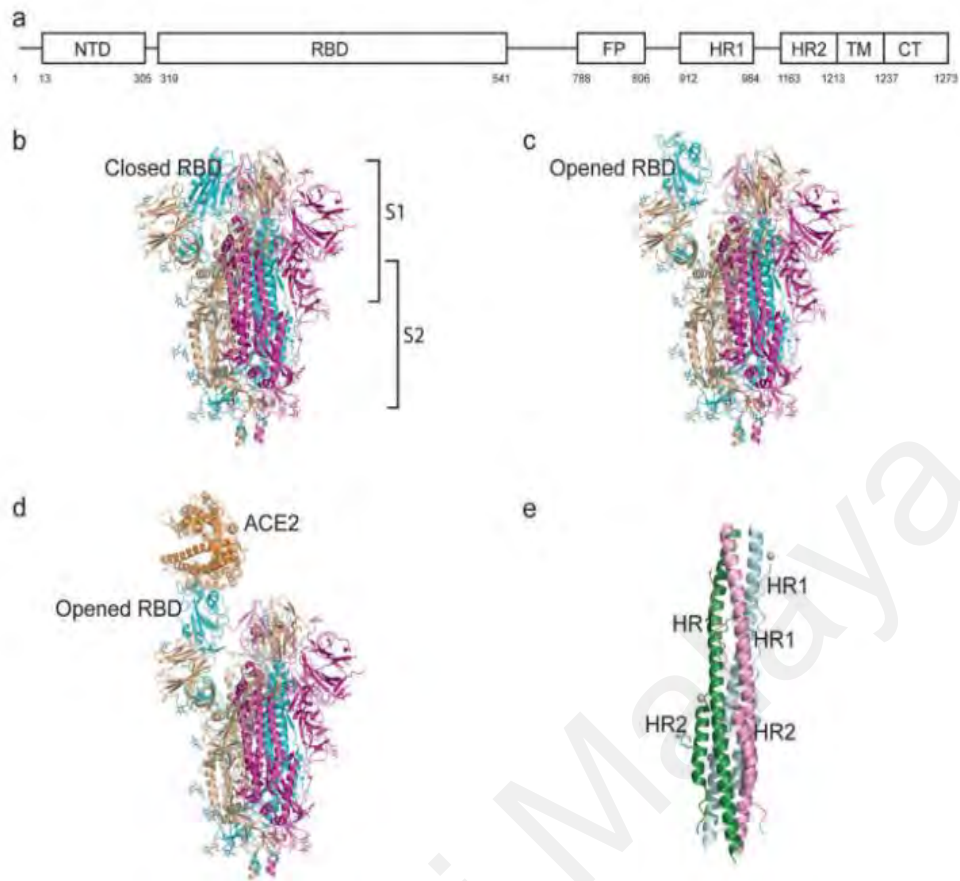
## **2.2 Spike Protein**

The Coronaviridae family exhibits varying degrees of conservation of the S protein. Protein's three polypeptide chains join together after synthesis to create a homotrimeric assembly. Additionally, every monomeric S protein bends in a three-dimensional formation with the head, stalk, and cytoplasmic tail constituting its three different topological domains (Casalino et al., 2020). During contact with membrane bound ACE2 receptor, the RBD of S1 subunit undergoes a down to up conformational change, enhancing cell recognition and binding. The general mechanical stability of homotrimeric

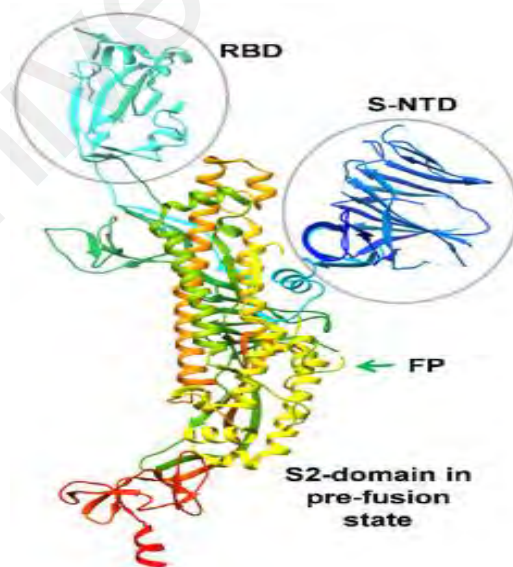
spikes is also greatly improved by RBD (Moreira et al., 2020). The S2 subunit's fusion peptide (FP), central helix (CH), connecting domain (CD), and heptad repeat (HR1/2) domains facilitate the integration of the viral membrane with the host cells membrane. In reality, entry into host cells is caused by the receptor binding motif of the S1 subunit's RBD interacting with the ACE2 receptor. The S protein has substantial N-linked glycosylation, which contributes significantly to the protein's contact area and exhibits dynamic changes that rely on the protein's conformation. N-glycans play a function in regulating RBD conformation in addition to shielding, particularly at the N165 and N234 sites, which may be used as a therapeutic target (Casalino et al., 2020). Moreover, the Cryo-electron microscopy has determined the structure of the SARS-CoV-2 spike protein (Walls et al., 2020).

The spike protein plays a critical function in the interaction with the host receptor ACE2. The spike protein participates in the infection's first entrance process. Trimeric spike protein binds to ACE2 receptors on the viral surface protein, fusing the membranes and allowing the virus to enter the cell.

While the S2 domain of the spike protein is crucial for membrane fusion, the S1 domain of the spike protein plays a critical role in receptor binding. The RBD domain of spike protein has a motif against receptor binding, which directly binds with ACE2 receptor of the host (Huang et al., 2020). The spike protein is mainly used as a drug target and for vaccine development (Chellapandi, 2020; Huang et al., 2020). All vaccines are designed according to the spike protein. The FDA approved vaccines as Moderna and Pfizer/BioNTech have been designed according to spike protein (Xia, 2021). The spike protein was predicted as the immunodominant viral agent. After evaluation of COVID-19 patients, it is revealed that binding and neutralizing the antibodies mainly target the RBD of S1 (Creech et al., 2021).



**Figure 2.2: SARS-CoV-2 spike schematic. This figure is adapted from Yuan Huang (2020).**



**Figure 2.3: The S protein monomer .This figure is adapted from Rimanshee Arya et al., (2020).**



### **2.2.1 The Structural Analysis of spike.**

The S protein is anchored in the viral membrane and has a molecular weight of 180–200 kDa. It normally has a metastable, prefusion shape; also, the S protein undergoes a dramatic structural change when it contacts with the host cell, allowing the virus to meld with the cell surface. These spikes are disguised with polysaccharide molecules to prevent immune system identification upon entry (Watanabe Y et al., 2020). With 1273 amino acids and the N-terminal signal peptide, SARS-CoV-2 S is a large virus (amino acids 1–13). Visually, S protein trimers surround each virus particle with a characteristic halo like a crown. The bulbous head and stalk area made up of the subunits S1 and S2, according to the coronavirus monomer's structure (Tang T et al., 2020). The structural properties of the SARS-CoV-2 trimeric S protein has been determined by cryo-electron microscopy, showing the many conformations and functions of the S RBD domain in both open and closed states (2020). The CoV S protein occurs as an unlikely to face in its original state. Cellular proteases divide the S protein into the S1 and S2 subunits, it functions similarly to certain other coronaviruses (Du L et al., 2007 & Hoffmann M et al., 2020).

### **2.2.2 The structure and function of S1 domain**

The binding of virus particles to host cells receptors marks the beginning of viral infection. Therefore, the receptor recognition is a key factor in determining viral entry and a potential for therapeutic development. The binding interface's atomic details show that the SARS-CoV-2-CTD has undergone significant changes. Crucial roles are played by mutations of critical residues in strengthening the interaction with ACE2. The RBD section of SARS-CoV-2 and SARS-CoV have 73–76% identical in sequence, making them important targets for neutralising antibodies (nAbs). Nine ACE2 contacting residues in the CoV RBD are completely intact, while four are just partially so. The majority of residues necessary for ACE2 binding in the SARS-CoV S protein are preserved in the

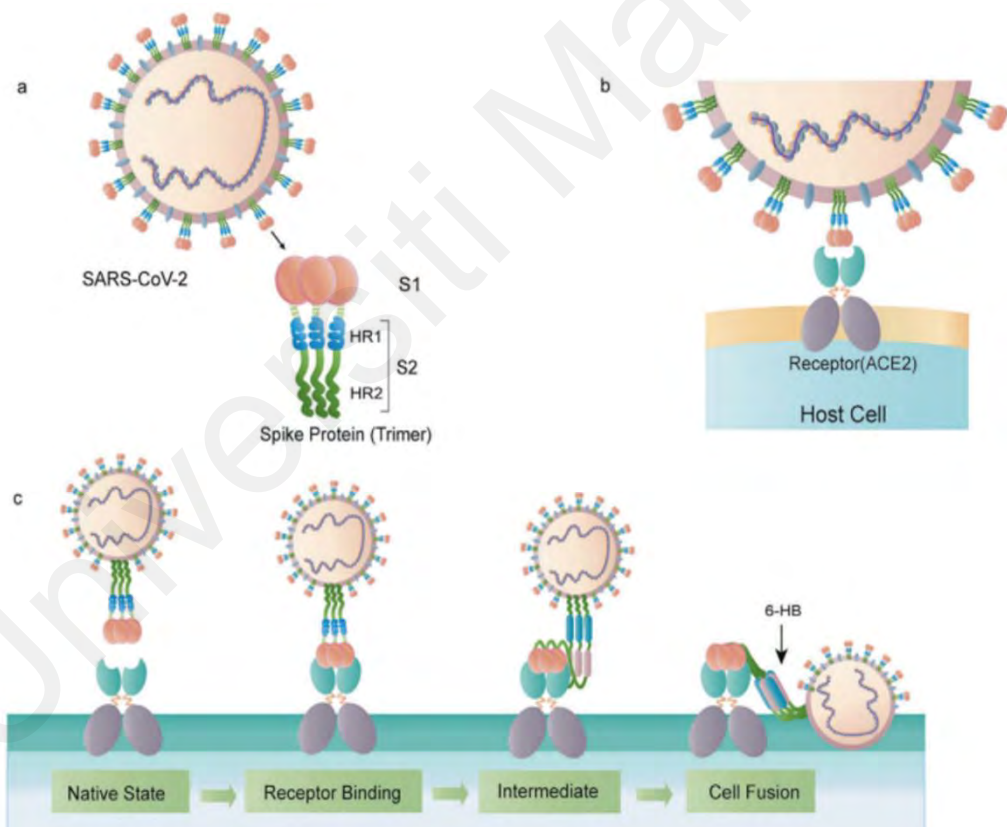
SARS-CoV-2 S protein, as determined by an examination of their RBMs. Nonetheless, other studies shown that there are antigenic differences between SARS-CoV and SARS-CoV-2, as shown by the inability of murine monoclonal antibodies (mAbs) as well as the polyclonal antibodies against SARS-RBD to bind with the S protein of SARS-CoV-2 (Wang Q et al., 2020). Because broad-spectrum anti-CoV drugs are notoriously unpredictable, S1 RBD may not have been the optimal target for treatment. SARS-CoV RBD-specific antibodies were unable to avoid disease produced by the S protein of SL-CoV-SHC014, suggesting that the S1 RBD may not have been the greatest target for medicine (Xia et al., 2020).

### **2.2.3 The structure and function of S2 domain**

The S2 subunit, which fuses HR1, HR2, FP, TM, and cytoplasmic domains, are in charge of viral fusion and entry (CT). FP is a brief stretch of 15–20 conserved amino acids of the virus family, made primarily of hydrophobic residues like glycine (G) or alanine (A), which anchor to the target membrane when the S protein assumes the hairpin conformation. The repeating heptapeptide HPPHCPC, in which H is a hydrophobic and P is the polar, while C is still another charged residue. HR2 is situated at the N-terminus of the TM domain (Robson, 2020). A hydrophobic FP's C-terminus is where HR1 is found. The downstream TM region of the S protein interacts to the viral membrane through the S2 subunit's CT tail (Tang et al., 2020). When RBD binds to ACE2, S2 changes the conformation of ACE2 by introducing FP into the target cell membrane, exposing the HR1 domain's hairpin coiled-coil, and establishing contact between the HR2 domain and HR1 trimer to produce 6-HB (Xia S et al., 2019).

The HR2 domain generates both a complicated helix and a flexible loop made up of a flexible loop to interact with the HR1 domain. The HR1 and HR2 domains make a number of solid connections in the fusion core region of CoVs' postfusion hairpin structure (HR1core and HR2core sections, respectively). According to Chen (2020), because

ACE2 has a conserved primary structure across species, including amphibians, birds, and mammals, the interaction between the S protein and ACE2 can be used to identify intermediate hosts of SARS-CoV-2. The ACE2 and SARS-CoV-2 S binding affinities in mammals, birds, snakes, and turtles were studied by Luan et al., 2020. They discovered that while ACE2 from snakes and turtles was unable to interact with SARS-CoV-2 S RBD, ACE2 from Bovidae and Cricetidae did. Viral attachment to host cells is facilitated by the interaction between the S protein and ACE2 through S1 subunit's RBD region. As shown by the fact that SARS-CoV S binds to ACE2 with a KD of 325.8 nM. Researcher ensures that SARS-CoV-2 differs from SARS-CoV in S by 24%, but only in RBD by 23% (Wan Yet al., 2020).



**Figure 2.4: Schematic illustration the attachment and fusion of SARS-CoV-2 S protein from Yuan (2020).**

#### **2.2.4 Viral Fusion**

Fusion of the viral membrane with the host cell's membrane results in the viral genome's production into the host cell. The SARS-CoV-2 S1 and S2 subunits are separated before to fusion. Human proteases split the S protein into two subunits, S1 and S2 to viral fusion occurs (Tortorici & Veelsler, 2019). SARS-CoV-2 S possesses several furin cleavage sites, which increases the likelihood that it will be cleaved by furin-like proteases and hence boosts its infectiousness (Naqvi et al., 2020). As seen in the 1997 avian influenza outbreak in Hong Kong, the furin-like cleavage domain is also present in highly pathogenic influenza viruses. It is associated to their virulence (Carbo et al., 2020).

6-HB creation is necessary for viral fusion. The FP domain at the N-terminus of SARS-CoV-2 and the two HR domains on S2 are indispensable for viral fusion (Kawase et al., 2019). The SARS-CoV-2 FP is made visible once the S protein is broken down, which leads to viral fusion. Under the influence of particular ligands, the fusion protein undergoes a conformational shift. Eventually, it enters the host cell's membrane (Harrison SC.2001, Coutard et al., 2020).

#### **2.2.5 Spike Protein Functions**

One of the main components of infection is the S protein which found on the virus's surface. Virus entrance is facilitated by a trimetric class I TM glycoprotein and it present in various HCoVs, HIV, influenza virus, paramyxovirus, and Ebola virus glycoprotein (Moreira et al., 2020). At the surface of the host cell, spike protein attaches to the humans angiotensin-converting enzyme 2 (ACE2). The transition between the S1 and S2 subunits minimizes the distance between the viral membrane and host cell membrane, allowing the virus to inject its genome into the host cell by fusion of the host cell, which is subsequently translated by host cell ribosomes when the S1 subunit connects to the receptor. During viral infection, the serine protease TMPRSS2 activates the S protein by

cleaving it into S1 and S2 subunits, activating the membrane fusion domain (Shang et al., 2020).

S protein mediates receptor identification, cell attachment, and fusion (Walls et al., 2020). The mechanism at the heart of the S protein's connection to the receptor (Walls AC et al., 2020) is the S protein's trimmer on the surface of the viral envelope. While the S2 domain directly associated with virus fusion, the S1 domain primarily contains the RBD. Therefore, it is primarily in charge of the virus's binding to the receptor (Cui et al., 2019).

### **2.3 COVID19 Infection**

The SARS-CoV-2 virus spread widely worldwide. COVID-19 causes symptoms such as fever and cough, throat discomfort with muscular ache, chest pain, confusion, dyspnea, ageusia, anosmia, and headache. These severe symptoms lead to life-fatal respiratory deficiency and also disturb the heart, liver, kidney, and nervous system. In contrast, the diagnosis of SARS-CoV-2 infection is muddled with other viral infections such as upper tract infection and influenza (Kevadiya et al., 2021). 2020 research study illustrates a scientific approach to the dynamics of identified Covid19 diseases under conditions of epidemic. The study demonstrated empirically determined everyday infections. A preliminary analysis based on the fundamental theories of how infections spread led them to two different accelerating regimes where as the COVID-19 exhibits an infectious disease. To reduce the infection rate, measures like household lockdown is essential. As a result, one week following lockdown the typical COVID-19 incubation period is discovered to represent the pivot point between slow and fast infection rates. Infections reach their peak after this transition point and the slow growth rate, after which the infection rate begins to decline. Because many families are getting infected by people who are already unwell from the prior periods, a rise may be observed after the lockdown

pivotal stage. Nevertheless, due to lockdown, diseases that have been identified continue to decline quickly, making this peak a singularity (Clara-Rahola, 2020).

## **2.4 COVID19 Treatments**

In a recent study, therapeutic and prophylaxis technique has been observed which prophylactic and therapeutic direction of EIDD2801 (antiviral agent: oral) is in phase 3 clinical trials. It halted the SARS-CoV-2 of replication cycle in vivo. Thus, it can be considered a potent treatment against COVID-19 while there needs more investigation (Wahl et al., 2021). Another research conducted by Zhou, described the HuNAb therapy destroys of the SARS-CoV-2 replication and damage in lungs, while the strong infection in nasal turbinates is cured in a maximum of three days, This therapy also needs more investigation (Zhou et al., 2021). Plasma therapy in recent studies is also described as a good treatment, but in the case of CoVs variants, there is no evidence of decreased vulnerability to neutralizing the antibodies (Kemp et al., 2021).

Some molecular docking techniques were utilized and identified drugs such as saquinavir, remdesivir, ritonavir, cefuroxime axetil, delavirdine, prevacid (Al-Khafaji et al., 2021). While the clinical and experimental procedures have not done yet. A computational study was preformed to indicate that Fisetin, Quercetin, and Kamferol compounds have the ability to disrupt the binding with human receptor (Pandey et al., 2020).

Vaccines made from recombinant or inactivated proteins methods are dependable and secure, but their immunogenicity is often limited unless they are combined with an adequate adjuvant (Weinberger, 2018). Recombinant spike protein vaccines provided better protection against the coronaviruses that cause Middle East respiratory symptoms (MERS) according to Adney et al. (2019).

Human vaccinations to hepatitis B, seasonal and pandemic influenza, and other illnesses have shown the safety and effectiveness of Advax adjuvant (Gordon et al., 2016; Sajkov et al., 2014).

## **2.5 Natural Compounds**

Natural products act as chemical scaffolding for derivatization, resulting in new molecules with enhanced pharmacological properties, which are small compounds synthesized by living organisms. In the research study of Zigolo et al.(2021) the evaluation of natural compounds has been done. The natural compounds are primarily obtained from medicinal plants(Zigolo et al., 2021). These compounds were used as inhibition of COVID-19 by using molecular dynamics approach (Antonela et al., 2020). By assessing the literature review from 2010-2020, the natural products use has been identified against coronaviruses was 3CLpro (Cysteine protease). By investigating more articles, it is observed that some present drugs as natural products being used and potent lead substance against the SARS-CoV-2 treatment. The natural compounds which have been identified are prominently Savinin, Lycorine, and 8-hydroxy. Several other techniques have been studied and used by researchers to identify new potent lead natural products. The technique was observed as an ethno pharmacological method and molecular docking etc. Though, the selection of compounds for further analysis/clinical trials should be prudently considered. The need for in vitro as well as in-silico experiments still remains for these natural products (Verma et al., 2020). In-silico research like those by Arifuzzaman et al. (2022) that also looked for a potential SARS-CoV-2 inhibitor employed NuBBE. In order to find a lead against SARS-CoV-2, a natural chemical was explored for utilising NUBBE in a structural-based virtual screening study (Singam et al., n.d.).(Arifuzzaman et al., 2022)

## 2.6 Databases of Natural Products

The Natural Products & Biological Sources (NPBS) database is a chemical data source that includes relational information about natural products and biological sources that has been carefully chosen from natural substances.

The biodiversity of NPs in specific geographic areas, which are typically delineated by country political borders, is the subject of numerous national efforts. Although these databases primarily focus on plants, they can also include NP produced by insects, microorganisms, and animal toxins. In this section, the databases are listed from West to East according to their geographic location. In the final section, collections of NPs from organisms in marine and ocean environments are described. Over 400 distinct NPs from plants, fungi, and propolis from Mexican flora and fauna are available for full download in the 2019 database BIOFAQUIM, along with information about the species from which the compounds were extracted and their geographic location (Pilon et al., 2017). The first NP library from Brazilian biodiversity is called NUBBEDB, which stands for Nuclei of Bioassays, Ecophysiology and Biosynthesis of Natural Products (Pilon et al., 2017). Over 2000 NPs are currently present, along with highly curated, high-quality metadata and simple download options for the entire or selected data. Although all datasets also contain smaller amounts of hydrophilic molecules, NPs typically tend to be lipophilic, which allows for greater membrane penetration (Maria and Christoph., 2020). Nuclei of Bioassays, Ecophysiology and Biosynthesis of natural products (NuBBE) database (Saldívar-González et al., 2019).

The NuBBE database was created with the intention of providing chemical descriptions and molecular descriptors of the natural products examined in NuBBE labs for research in medicinal chemistry and molecular modeling. In this research, NuBBE database has been chosen as it is a valuable resource for studies on dereplication and the generation of novel drugs. As a result, attempts had been made to increase its content and



incorporate a wider variety of natural sources in order to establish it as a thorough compilation of the biogeochemical data on biodiversity (Pilon et al., 2017). This open-source database, which can be accessed at <http://nubbe.iq.unesp.br/portal/nubbe-search.html>, is without a doubt going to be an invaluable resource for researchers all over the world.

## **2.7 Structure-Based Virtual Screening**

An early-stage drug discovery campaign uses the computational method known as "structure-based virtual screening" (SBVS) to explore a library of chemical compounds for new bioactive chemicals that can be utilized to block a specific therapeutic target. It uses the three-dimensional (3D) structure of the biological target to dock a variety of bioactive molecules into binding site and choose a selected group based on their binding affinity additional analysis.

Virtual screening is also known as High Throughput Screening (HTS) which is robust, cost-effective, and useful for drug design. The computer-aided drug design (CADD) is widely being used in the drug discovery process (Sabe et al., 2021). Virtual screening (VS) has become one of the promising techniques for drug design to identify a new lead candidate by the screening big compound collections targeting a biological receptor. There are two types of virtual screening methods in the computer aided drug discovery approach, which are structure-based virtual screening (SBVD) and ligand-based virtual screening (LBVS). Molecular docking and molecular dynamics simulation methods are being used to perform Structure based virtual screening approach (Sabe et al., 2021; Maia et al., 2020). Both methodologies were general forms of computer-aided drug design (CADD) approaches that can integrate wet-lab techniques to elucidate the causes of drug resistance in discovering de novo and novel antibiotics for known and new targets. Structure based virtual screening (SBVD) is a method which is based on the docking of ligands to the active site of the target protein, which includes the evaluation of databases

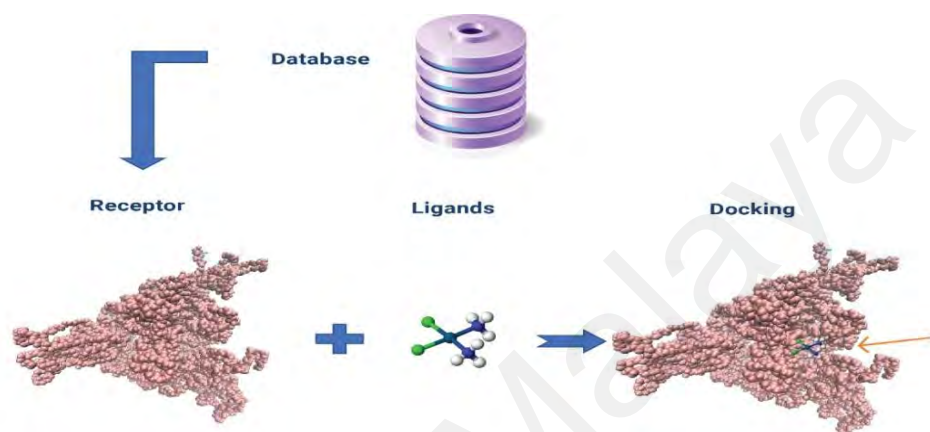
containing simulations of ligands and the target protein (Yu & Mackerell, 2017). Molecular docking and molecular dynamics (MD) simulation are two prominent methodologies used in the structure-based design and development that are also utilized in the virtual screening methodology (Lionta et al., 2014).

## **2.8 Molecular Docking and Molecular Dynamics Simulations**

Molecular docking method is crucial for designing drugs and studying structural molecular biology. Given their three-dimensional structures, the aim of this method is to forecast the most probable 'binding scenarios' between a protein and a ligand. In the execution of structure-based virtual screening, molecular docking is the necessary technique to determine the orientation and conformation of a ligand inside the receptor by analyzing their binding affinities (Pinzi & Rastelli, 2019). In a drug development program, docking is also utilized for virtual screening of novel ligands based on biological structures in order to identify hits, generate leads, or improve the potency or other characteristics of leads. Therefore, docking is an SBDD strategy that is crucial in the rational design of novel medicinal compounds (Rudrapal & Chetia, 2020).

Whereas molecular dynamics (MD), looks at the actual motions of atoms and molecules. The atoms and molecules are allowed to interact for a defined period of time, revealing information on the dynamic evolution of the program (Rapaport et al., 2004). Due to flexible entities where the recognition process of molecules in docking techniques may be derived from its structural rearrangements, the use of molecular dynamics simulation to molecular docking approaches is reliable for studying the control of ligands in binding to their target protein (Hospital et al., 2015). Thus, in general, molecular docking entails predicting and evaluating protein-ligand interactions, whereas Molecular dynamics simulation aims to the forces of molecule during a specified time period in accordance with Newton's third law of motion (Zhao & Caflisch, 2015). Molecular dynamics is a valuable method that is used to analyze ligand-receptor interactions at an

atomic level. In performing structure-based drug designing, the atomic level analysis is done by building conformational joints with respect to their dynamic properties and flexibilities. The MD simulation is reliable for simulating the molecular forces of proteins (Zhao & Caflisch, 2015).



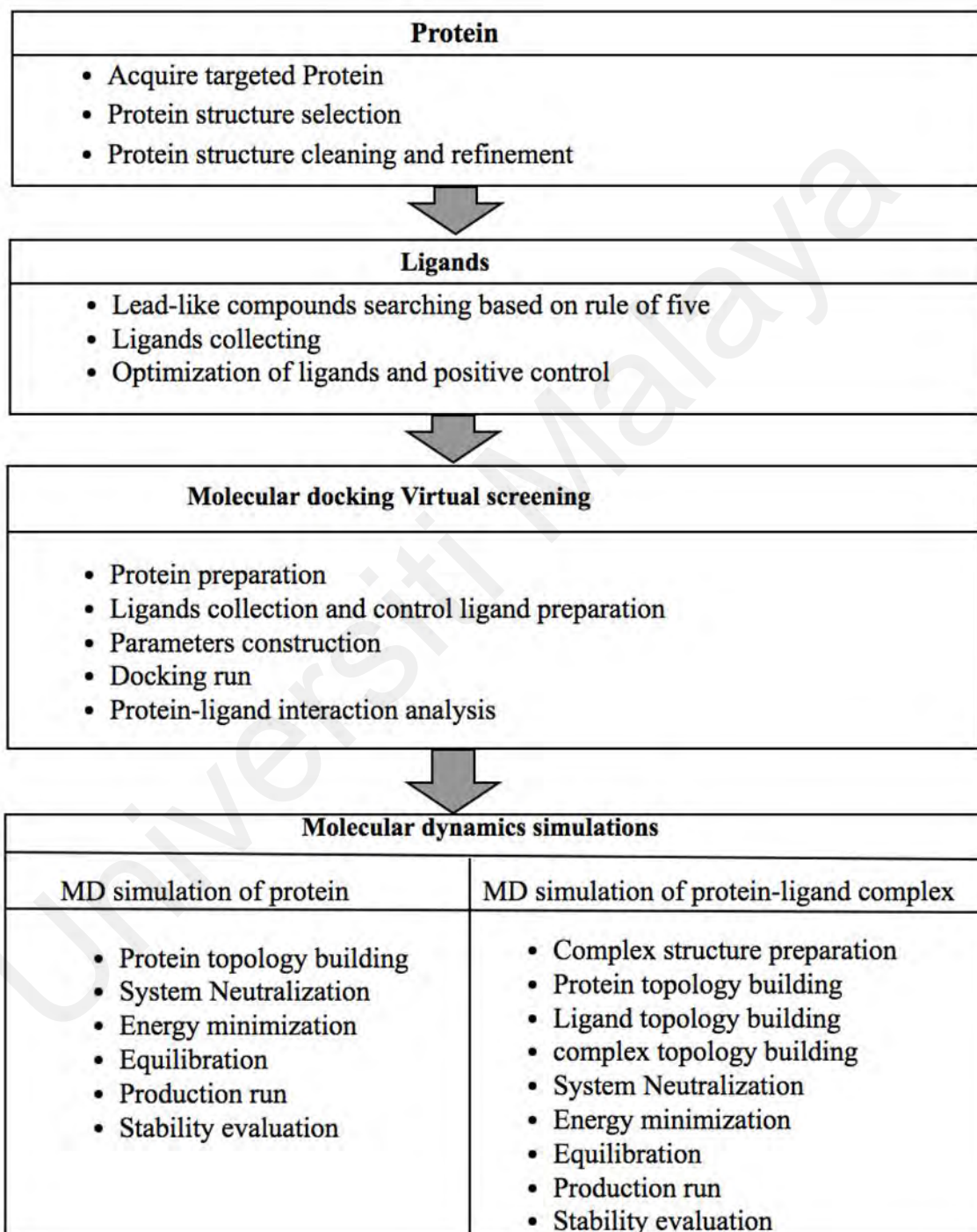
**Figure 2.5: An illustration of virtual screening steps.**

## CHAPTER 3: METHODOLOGY

This project is performed to identify a possible inhibitor for spike protein of SARS-CoV-2 virus from NuBBE database. Structure-based virtual screening has many steps which are molecular docking and molecular dynamics simulations. The workflow of this study is illustrated in Figure 3.1.

### 3.1 Workstation Equipment

To carry out this study, an operating system of Ubuntu Linux version 20.04 LTS (64-bit) was used, along with 32 gigabytes of Random Access Memory (RAM). This PC contains GeForce GTX780 Ti as its graphics processing unit (GPU) and had 12 Intel® Core™ i7-4930K 3.40 gigahertz (GHz) microprocessors.



**Figure 3.1** Work flow of structural-based virtual screening.

### **3.2 Protein Preparation**

The non-mutated three-dimensional (3D) structure of the SARS-CoV-2 spike protein was obtained from the Protein Data Bank (PDB) (Burley et al., 2019). Several structures were analysed and compared based on variants, their resolution, mutations, the number of residues, and the number of chains. After downloading the structure, the three protein chains were analysed, and one was selected using PyMOL tools. It is a tool for visualizing, editing, and analysing 3D molecular structures, among other things (Yuan et al., 2017). The protein structure was then optimized with UCSF Chimera (Pettersen et al., 2004), and AutoDock4 was used to prepare it for molecular docking. The missing atoms were replaced, and the water molecule was removed while hydrogen atoms were added. Hydrogens that are not polar were combined in the protein structure. In addition to the AutoDock Tools graphical user interface, the Kollman charges were added to the structure through the use of a Python script in the AutoDock Tools module (Goodsell et al., 2021).

### **3.3 Ligand Library Preparation**

The natural compound library was fetched from (NuBBE) database (Saldívar-González et al., 2019). NuBBE database is a natural database that provides various natural sources isolated from plants in 3D structures. When choosing ligands, Lipinski's rule of five was applied (Lipinski, 2004). Which is physicochemical properties of drug-like compounds. The parameters were as follows: the molecular mass was  $\leq 350$  Dalton, the high lipophilicity (LogP) was  $\leq 3.5$ , less than 4 hydrogen bond donors, less than 6 hydrogen bond acceptors. Ligands were then optimized by adding Gasteiger charges and merging non-polar hydrogens, and it detected a rotatable bond. Subsequently, Mol2 files were converted to PDBQT format via AutoDock Tools python scripts.

### **3.4 Virtual Screening**

In the early stages of drug discovery, the structure-based virtual screening technique is used to design drugs computationally. In this technique, natural compound from NuBBE database were screened against spike protein. The 3D structure of spike protein was obtained by binding a natural compound library into the protein. Virtual screening was performed using UNIX shell commands and Python scripts from the AutoDock suite of programs. First, the working directory was created and set as the root directory. Next, the environment was set up to allow source script access to python, adt, autodock4, and autogrid4.

Prior to beginning virtual screening, all necessary files were prepared, including ligand4 (which contains the entire ligand library's MOL2 files and PDBQT files), the positive control PDB file, the protein control docking parameter file (dpf), the list of ligand file that had been prepared for spike proteins, and the Grid Parameter File, in which the grid box was applied as a blind docking which cover the entire structure, because to identify the possible binding affinity for library collection. The grid coordinates were set to  $x= 197.044$ ,  $y= 222.794$ , and  $z= 210.385$ , and the dimension of x, y, and z points was 82,100,126. The spacing between grid boxes was set to 1, and then atomic affinity maps for a ligand library were calculated using AutoGrid's graphical user interface by reading the AutoGrid Parameter File (GPF).

#### **3.4.1 Positive Control Preparation**

The Quercetin compound was selected to be used as a control ligand based on its viral activity, and the binding affinity had been shown with spike protein in the study of Pandey et al. (2020). In addition, the previous study reported that Quercetin has interaction with ILE 870, ASP 867, ALA 1056, PRO 1057, GLY 1059, HIS 1058, SER 730, MET 730, MET 731, LYS 733, VAL 860, LEU 861, PRO 863 and residues of the S2 domain of

spike protein. In addition, the clinical trial study reported that Quercetin could be used against SARS-CoV-2 spike protein as a nasal or throat spray (Williamson & Kerimi., 2020). The PDB file of Quercetin was obtained from the PubChem database (Kim et al., 2016). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8177144/>. Once the ligand was obtained, further optimization was performed using the PyMOL tool and converted to PDB format. The docking parameter file (dpf) was followed, and the positive control ligand-protein was converted from PDB to PDBQT format using a Python script in the AutoDock Tools module.

### **3.5 Molecular Docking**

The docking was performed using Python scripts using the computational suite AutoDock4 (Morris et al., 2009). AutoDock is a collection of automated docking algorithms designed to predict the binding of small molecules as potential drug candidates to the receptor (target protein). Therefore, it employs the Lamarckian genetic algorithm (LGA) method. This instrument will be used to determine the binding location and conformation of compounds and their potential binding affinity. First, unique docking parameter files (DPF) were generated for each ligand, with the same genetic algorithm parameters applied to all DPF files. The number of genetic algorithms used in the search was 100, the number of evaluations was 2.5 million, the population was 150, and the *rmstol* was 2.0. The docking was completed once the DPF files were generated, and the result files were saved for further analysis. Then the saved files were summarized and sorted using AutoDock Tools scripts according to their binding energy value. The interactions of selected ligands will also be analyzed with the visualization tool UCSF chimera. The complexes with the most negative binding energy value will be chosen for molecular dynamics simulations.



### **3.6 Molecular Dynamics (MD) Simulations**

#### **3.6.1 Protein MD Simulations**

The protein simulation was performed to evaluate the flexibility and stability of protein PDB structure. Possibly, the residues or atoms may be missed during minimization or dockings. primarily, the PDB of the spike was prepared, followed by setting up the system, then the topology of the protein was generated by using CHARMM36 force field and TIP3P water model then, the periodic boundary condition was set in cubic with 1.0 Å distance, succeeded by neutralization and energy minimization was conducted with 250,000 number of steps. The equilibration was done based on NVT and NPT, which were set at 300 K temperature and 1-bar pressure. After that protein MD simulation run was performed by following setting 50 ns of duration and 25,000,000 steps at 300 K controlled temperature and 1 bar pressure. In order to evaluate the stability of protein PDB structure, a graph of the temperature and pressure was generated. After the completion of MD simulations, the behavior of protein was analysed. Therefore, numerous dimensions will be analysed as root mean square deviation RMSD value, root means square fluctuation RMSF.

#### **3.6.2 Complex MD Simulations**

The molecular dynamics simulation was carried out in order to evaluate the stability of selected complexes (S\_NuBBE126) and their binding energy. The GROMACS software contains a pre-processor, molecular dynamics, and energy minimization programs which can be used with random processors (Abraham et al., 2015). GROMACS stands for Groningen Machine for Chemical Simulation, developed in the early 1990s at the University of Groningen. GROMACS is a flexible, fast, and free-of-cost for MD simulations while it caters for many force fields such as OPLS, GROMOS, ENCAD, and AMBER (Berendsen et al., 1995; Kutcher et al., 2019). Preparation of PDB files: The residues or atoms may be missed during minimization or docking. Therefore, the

receptor-ligand complex pdb is checked for missing residues and separated the protein structure and ligand into two files, and then the files are saved for topology building. In the system topology building, the topology files for the designated bound and unbound spike proteins were generated by using GROMACS `pdb2gmx` commands using the (CHARMM36) force field. In comparison, the ligand parameters were obtained from SwissParam software (<http://www.swissparam.ch>). Then the complex was rebuilt, and the topology file contained force field parameters and information related to the arrangement of protein-ligand complexes. After that, the periodic box was generated, which is the cubic periodic box along with 1.0 distance dimensions set for the system. The TIP3P water model was selected for salvation.

In simulations, system neutralization, or the neutralization of a complete system, plays an essential part. The total charge on the system was calculated to neutralize the system following the addition of water molecules. A total of 6  $\text{Na}^+$  and 0  $\text{Cl}^-$  ions were added to make the system's overall charge equal to zero. After neutralization, energy minimization of the system was performed to eliminate clashes and improper geometry. The system's energy was minimized by 50000 steps of energy minimization using the steepest descent algorithm (Chambliss & Franklin, 2020). The controlled temperature and pressure were set to 10000 ps at a constant temperature of 300K by the V-rescale thermostat (Bussi et al., 2007; Chauhan & Poddar, 2019) and 1 bar by Parrinello-Rahman (Parrinello & Rahman, 1981). The LINCS algorithm is used to constrain the bonds involving H. (Panwar & Kumar, 2021). Using the volume coupling NVT (N: constant number, V: volume, T: temperature) equilibration stage, the system's temperature was fixed. During the pressure coupling NPT (N: constant number, P: pressure, T: temperature) equilibration stage, the system pressure was stabilized. After completing the equilibration stage, the MD simulations were run for 50 ns with a scale of 25,000,000 steps at a temperature of 300 K and a pressure of 1 bar under constant pressure and temperature

conditions. The frames of the trajectory were then extracted. The simulation run constructed graphs of temperature, pressure, potential energy. The UCSF Chimera visualization tool was used to analyze and display the MD simulation trajectory data. The root mean square deviation (RMSD) value is used to monitor the stability of the chosen carbon alfa C, and the root mean square fluctuation (RMSF) value is used to identify amino acid fluctuations and assess protein flexibility. Hydrogen bond plot was generated to analyze the hydrogen bonds that generated during the simulation. The UCSF Chimera visualization tool was used to analyze and display the MD simulation trajectory. The RMSD plot will determine the system's stability, and the RMSF plot will show the amino acid fluctuations.

### **3.7 ADMET Profiles**

Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) was Applied to study the selected compound's and positive control's behavior inside the human body. In-silico ADMET profile was generated for the selected ligand and positive control by web servers SwissADME (Daina et al., 2017) and pkCSM (Pires et al., 2015). The input was compound smiles.

## CHAPTER 4: RESULTS

### 4.1 Protein Preparation

In this research the target protein was set to unmutated spike protein of SARS-CoV-2. Table 4.1 summarizes the availability of 3D structure of unmutated SARS-CoV-2 spike protein. In the PDB database, the complete 3D structure of SARS-CoV-2 spike protein, was only available by cryo-electron microscopy experiment with all the structure presence in 3 chains (chains A, B, and C) (Walls et al., 2020) with the number of residues for each chain with 1270 residues. The native state of spike protein in the SARS-CoV-2 virus exist as perfusion state with either open or closed conformation(Walls et al., 2020). Since S protein observed in highly pathogenic human coronaviruses because it exist in partly opened conformation (Gur et al., 2020), open conformation is highly suggested for drug discovery studies than the closed conformation. Due to the facts mentioned above the 3D structure of open conformation spike protein with the PDB code of 6VYB (3.20 Å) was selected for the study (Figure 4.1).

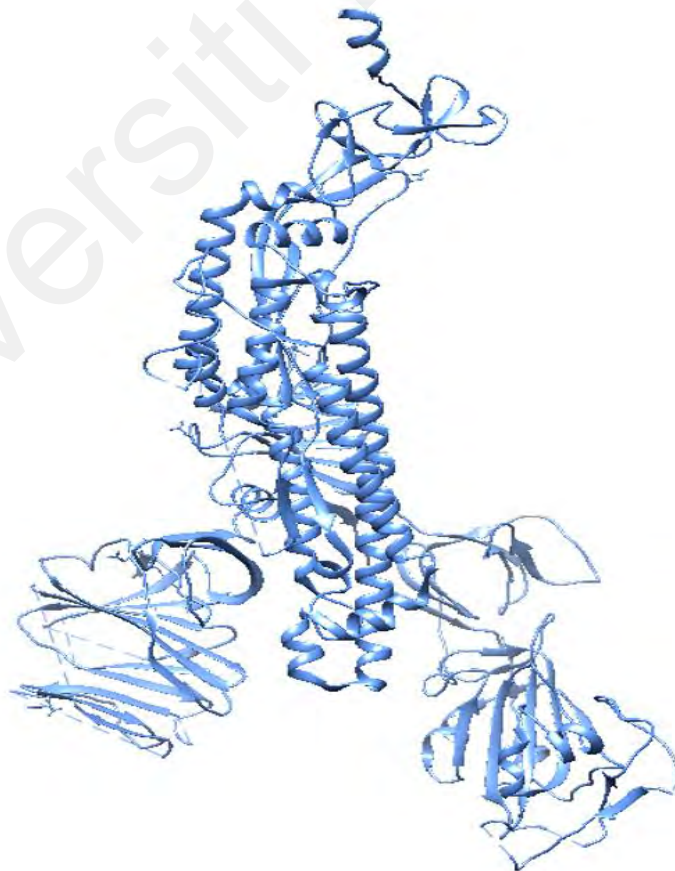
Once the structure was downloaded, the visualization of structure was done by PyMOL. All clashes are removed by PyMOL. In this study, chains B and C were deleted then chain A was extracted to be studied using PyMOL. The final protein structure is minimized via UCSF Chimera (Figure 4.2). The protein structure undergoes docking preparation by adding hydrogen atoms and merging non-polar hydrogen while the Kollman charges were added up to 3.52. The output file was then saved in PDBQT format.

**Table 4.1: Summary of non-mutated S protein PDBs available in protein data bank.**

<b>PDB</b>	<b>Methods</b>	<b>Conformation</b>	<b>Resolution</b>	<b>Length</b>	<b>Chains</b>
7EDJ	Electron microscopy	Open	3.30 Å	1286	A, B, C
7EDI	Electron microscopy	Open	3.20 Å	1286	A, B, C
7EDH	Electron microscopy	Open	3.30 Å	1286	A, B, C
7EDG	Electron microscopy	Open	3.60 Å	1286	A, B, C
7EDF	Electron microscopy	Open	3.20Å	1286	A, B, C
6ZP0	Electron microscopy	Open	3.20 Å	1286	A, B, C
6XR8	Electron microscopy	Closed	3.00Å	1247	A, B, C
6ZWV	Electron microscopy	Open	3.50Å	1310	A, B, C
7BNM	Electron microscopy	Closed	3.50Å	1273	A, B, C
7BNN	Electron microscopy	Open	3.50Å	1177	A, B, C
6ZGH	Electron microscopy	Open	6.80Å	1287	A, B, C
7MTD	Electron microscopy	Open	3.50Å	1287	A, B, C
7MTE	Electron microscopy	Closed	3.20Å	1281	A, B, C
7N1T	Electron microscopy	Closed	3.11Å	1305	A, B, C
7N1U	Electron microscopy	Closed	3.14Å	1305	A, B, C
7N1V	Electron microscopy	Closed	3.21Å	1305	A, B, C
7N1W	Electron microscopy	Closed	3.33Å	1305	A, B, C
7N1X	Electron microscopy	Closed	4.00Å	1305	A, B, C
7N1Y	Electron microscopy	Open	4.30Å	1305	A, B, C
6VYB	Electron microscopy	Open	3.20Å	1281	A, B, C
6VXX	Electron microscopy	Closed	2.80Å	1281	A, B, C



**Figure 4.1: Structure of SARS-CoV-2 spike protein obtained from the PDB database. PDB ID:6VYB Chain A colored blue, chain B colored orange and chain C colored green.**



**Figure 4.2: Structure of chain A of SARS-CoV-2 spike protein after deleted all clashes, chain B& C.**

## **4.2 Ligand Library Preparation**

NuBBEDB is a database of natural products that includes compounds and secondary metabolites from bacteria, fungi, insects, marine organisms, and plants. A modified version of Lipinsk's rule of five (Table 4.2) was applied to the database extracting 514 compounds out of 2200 in the NuBBE database as shown in Figure 4.3. The 3D structure of the selected 514 compounds in MOL2 files were downloaded and converted to PDBQT files for further processing. The summary of ligand library is shown in the appendix. Quercetin PubChem CID: 5280343 was reported to show a stable complex with spike protein of SARS-CoV-2 (Pandey et al., 2020). Thus, Quercetin was selected as positive control for the docking analysis (Figure 4.4).

## **4.3 Virtual Screening**

The ligand library of 514 natural compounds and positive control were docked against chain A of spike protein for 100 runs each ligand. The top ten ligands with the lowest binding energy value were analysed and compared with positive control (Table 4.3).

**SOURCE**

- ALL
- SEMISYNTHESIS
- BIOTRANSFORMATION PRODUCT
- ISOLATED FROM A PLANT
- ISOLATED FROM A MICROORGANISM
- ISOLATED FROM A MARINE ORGANISM
- ISOLATED FROM ANIMALIA

MOLAR MASS: 350

MONOISOTOPIC MASS: [ ]

CLOGP: 3.5

TPSA: [ ]

LIPINSKI VIOLATIONS: [ ]

H-BOND ACCEPTORS: 6

H-BOND DONORS: 4

Search compound(s)

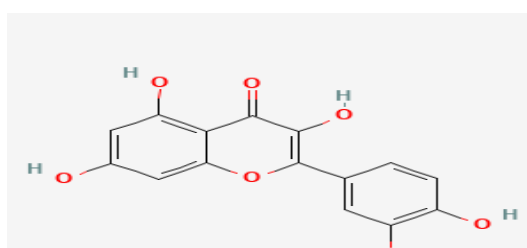
**514 compounds**

Download all molecular structure files of the results: [.mol2](#)

**Figure 4.3: Number of selected natural compound from NuBBE database.**

**Table 4.2: Used criteria to filter NuBBE database under rule of five.**

Criteria	Lipinski's rule of five	selection
Molecular mass	Less than 500 Da	0 –300 Da
Lipophilicity (logP)	Not exceed 5	0 – 3.5
H bond acceptors	No more than 10	1 – 6
Hydrogen bond donors	No more than 5	1 – 4
source	-	Plant



**Figure 4.4: Selected positive control (Quercetin).**



### 4.3.1 Molecular Docking

The bind docking preformed 100 run for each ligand from ligand library. The result files were dlg files, which were analysed and sorted into summary list files based on their binding energy, from the lowest to the highest value of binding energy. Out of 515 natural compounds, the top 10 ligands that show the lowest binding energy value were selected for further analysis. All the selected ligands show better binding energy compared to positive control (Table 4.3). The top ten of the docking analysis results show a better binding affinity with lower binding energy compared to the positive control (Table 4.3). Table 4.4 summarized the docking result for the top 10 best docked complex and Table 4.5 summarize the information of the top 10 ligands. Figure 4.5 shows the structure of the protein-ligand complex from the docking analysis.

### 4.3.2 Protein- Ligand Interactions Analysis

Table 4.3 show the top 10 of lowest binding energy of ligand-protein complex form the docking analysis together with quercetin as a positive control for comparison. Most of selected ligands formed a H bond with different amino acid of protein, while only 3 of them formed 2 H bond (Figure 4.5). Positive control complex result shows that four H bond formed with ser975, Met740, Leu799 and Arg1000 residues (Table 4.5). Out of the 10 selected ligands-protein complexes, protein complexes with lig126 and lig564 show similar binding site location with the positive control (quercetin-protein complex) (Figure 4.5 F and I). In addition, both formed two hydrogen bonds with the same residues (LYS733) and (Gly1059) with the distance of 1.795 Å and 2.242 Å for lig126 and 1.912 Å, 2.400 Å for Lig 564 respectively (Table 4.5). since ligand126 had lower binding energy than lig564 (Table 4.5) the lig126-complex was selected for molecular dynamics analysis. Figure 4.6 and 4.7 showed the interaction of complexes for lig126 and positive control.

**Table 4.3 The top ten binding affinity, sorted from the most negative binding energy value.**

Positive (Quercetin)	PubChem CID	
	5280343	-6.37
1	Lig 585	-7.8700
2	Lig 1576	-7.8400
3	Lig 1769	-7.4300
4	Lig 1507	-7.2600
5	Lig 2057	-7.1400
6	Lig 126	-7.0300
7	Lig 1768	-7.0200
8	Lig 991	-7.0200
9	Lig 564	-6.9900
10	Lig 1370	-6.9600

**Table 4.4: Protein-ligand interaction of the positive control and selected ligands**

Ligand ID	Total number in cluster (out of 100)	Binding energy (kcal/mol)	Number of H bond	The residues formed H bond with	Distance (Å)	S protein domain
Positive	78	-6.37	4	ser975 Met740 Leu977 Arg1000	2.222 1.652 2.088 2.589	S2
585	85	-7.8700	1	Phe140	1.608	S1
1576	46	-7.8400	1	Lys356	1.974	S1
1769	54	-7.4300	1	His 1048	1.936	S2
1507	48	-7.2600	1	Tur274	1.864	S1
2057	92	-7.1400	0	-	-	S2
<b>126</b>	<b>66</b>	<b>-7.0300</b>	<b>2</b>	<b>Lys733</b> <b>Gly1059</b>	<b>1.795</b> <b>2.242</b>	<b>S2</b>
1768	62	-7.0200	2	Lys1038 Ser1037	1.999 1.883	S2
991	56	-7.0200	1	Ser555	1.872	S1
564	57	-6.9900	1	Lys 733 Gly1059	1.912A 2.400	S2
1370	68	-6.9600	1	Val42	2.248	S1

Table 4.5: Information on selected ligands.

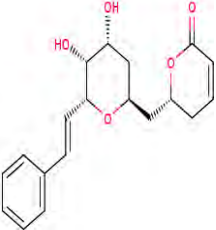
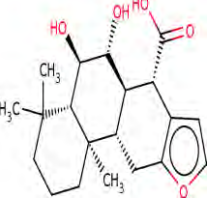
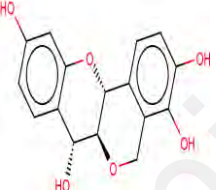
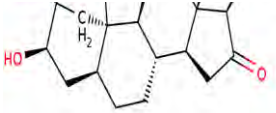
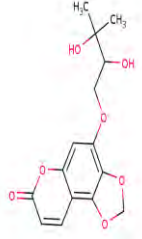
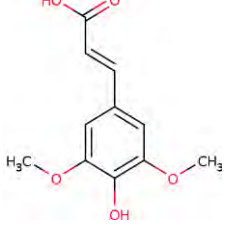
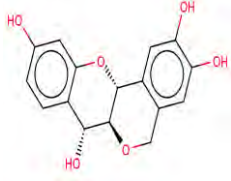
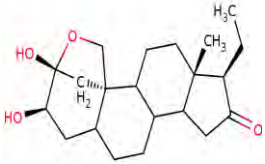
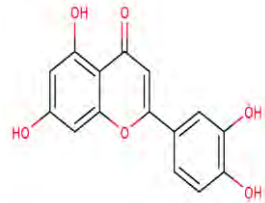
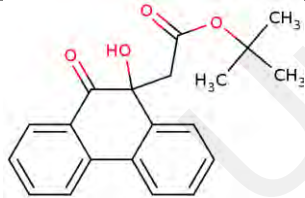
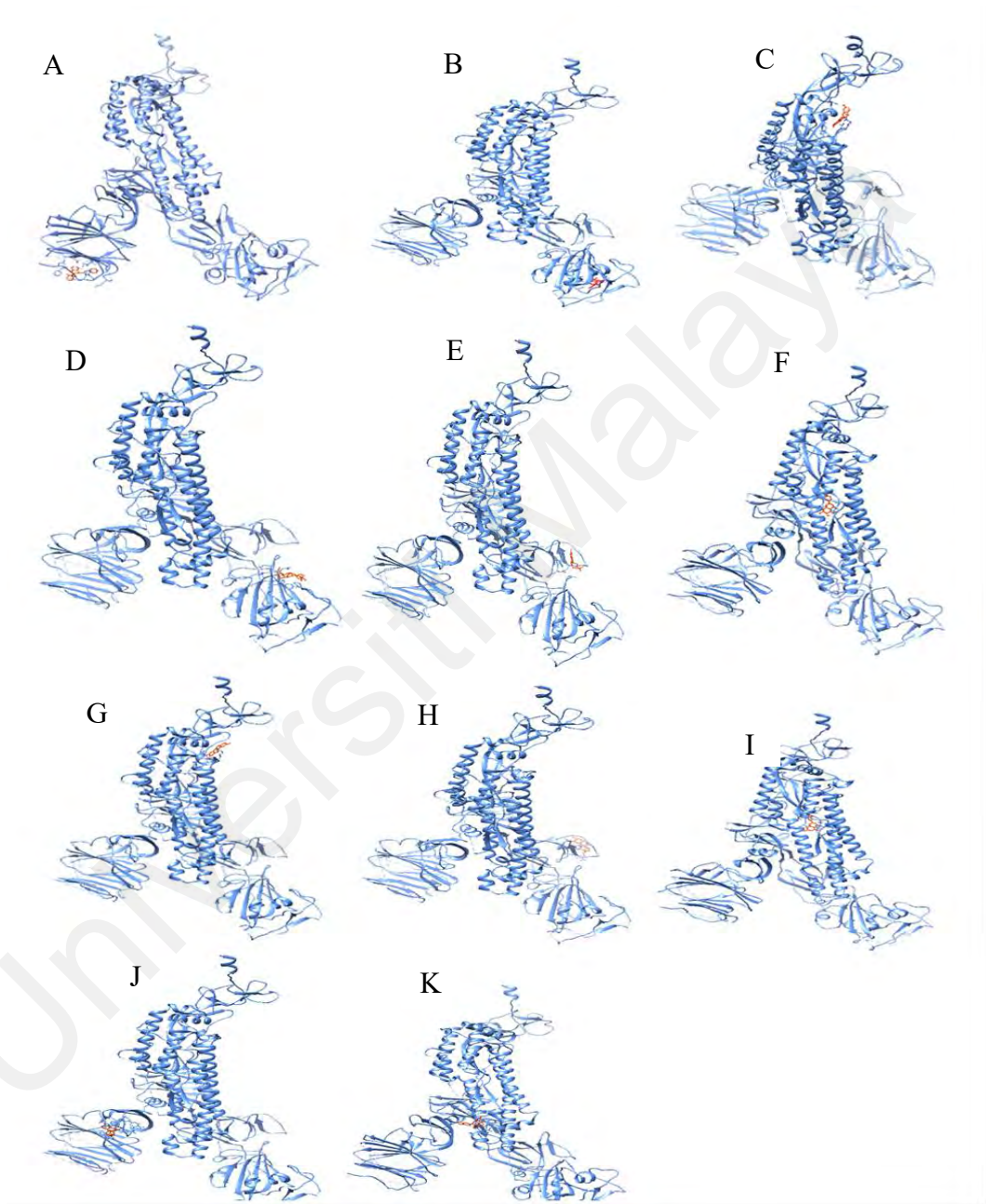
Ligand ID	Chemical Structure	Common Name	plant Species	Biological ties	Chemical Class
585		ryptopyranmoscatone A3; (6R)-5,6-dihydro-6-[(2'S,4'R,5'R,6'R,7'E)-2',6'-epoxy-4',5'-dihydroxy-8'-phenyl-7'-octenyl]-2-pyrone	<i>Citrus sp. auraceae</i> <i>Cryptocarya moschata</i> ; <i>Sao Paulo-SP</i> ;	None associated	Aromatic derivatives Styryl Pyrone
1576		6 $\alpha$ ,7 $\beta$ -Dihydroxyvouacapan-17 $\beta$ -oic acid	<i>Fabaceae</i> <i>Pterodon polygalaeflorus</i> ; <i>N/A-MG</i> ;	Cytotoxic	Terpenes Diterpenoid
1769		Mopanol	<i>Peltogyne catingae</i> ; <i>N/A-ES</i> ; <i>Peltogyne confertiflora</i> ; <i>N/A-AM</i> ;		Flavonoids Anthocyanidin

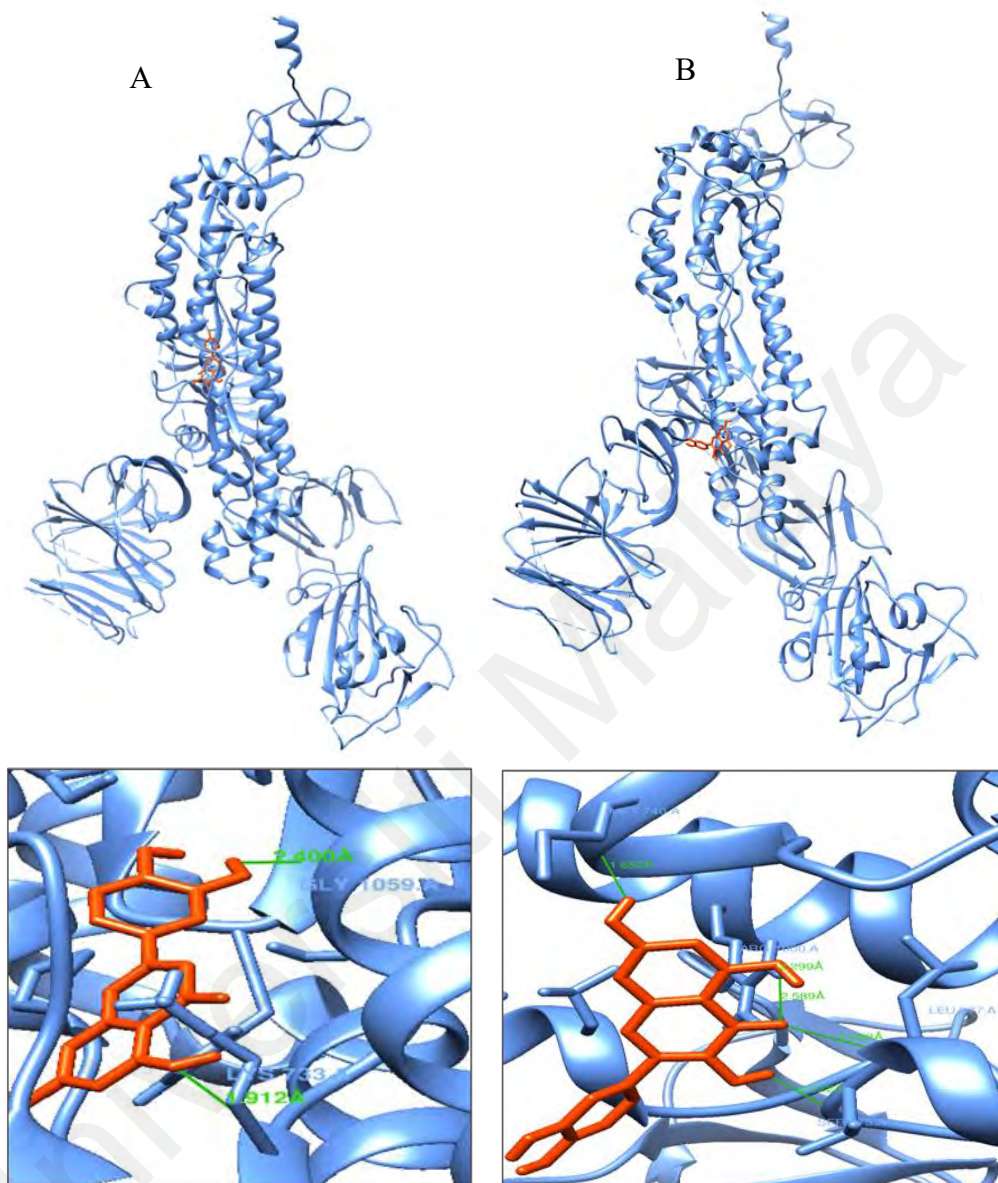
Table 4.5 (continued)

		<p><i>Clusia</i> n-16-one<math>2\beta,19</math>- hemiketal</p>	<p><i>clausenii</i>; Rio Claro-SP;</p>	
<p>2057</p>		<p>7-(2,3- Dihydroxy-3- methylbutylloxy) -5,6- methylenedioxy- coumarin</p>	<p><i>Pterocaulon</i> <i>lanatum</i>; <i>Campinas</i>- <i>SP</i>; <i>Pterocaulon</i> <i>balansae</i>; <i>Campinas</i>-<i>SP</i>;</p>	<p>Aromatic derivatives Coumarin</p>
<p>126</p>		<p><b>Sorbifolin</b></p>	<p><i>Fabaceae</i> <i>Pterogyne nitens</i> ; Sao Paulo - SP</p> <p><b>Inhibition of Myeloperoxidase</b></p> <p><b>Inhibition of Glycosidase</b></p>	<p><b>Flavonoids Flavone</b></p>

1768		peltogynol	<i>Peltogyne catingae</i> ; <i>N/A-ES</i> ; <i>Peltogyne confertiflora</i> ; <i>N/A-AM</i> ;	Flavonoids Anthocyanidin
991		2 $\alpha$ ,3 $\beta$ - Dihydroxypregnan- 16-one 2 $\beta$ ,19- hemiketal	<i>Meliaceae</i> <i>Trichilia clausenii</i> ; <i>Rio Claro-SP</i> ;	Terpenes Steroids
564		Luteolin; 2-(3, 4-dihydroxyphenyl)-5, 7-dihydroxy-4H-chromen-4-one	<i>Verbenaceae</i> <i>Vitex polygama</i> ; <i>Pocos De Caldas-MG</i>	Inhibition of Cathepsin V Antitrypanosomal
1370		9,10-Dihydro-9-hydroxy-9-(tert-butoxycarbonylmethyl)-10-oxophenanthrene	<i>Toona ciliata</i> ; <i>Vicosa-MG</i> ;	Aromatic derivatives



**Figure 4.5: Protein-ligand complexes of selected ligands and positive control. (A-J) selected complexes sorted from the lowest binding energy value. (K) spike-positive control complex.**



**Figure 4.6: (A) Protein-ligand interaction of spike-126 complex, the zoomed-in image illustrated two hydrogen bonds interaction (green lines). (B) Protein-ligand interaction of spike-positive complex the zoomed-in image illustrated four hydrogen bonds in**

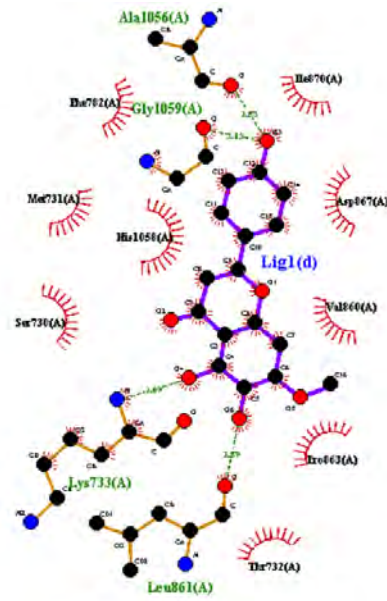


Figure 4.7: Interactions of spike\_Lig126 complex.

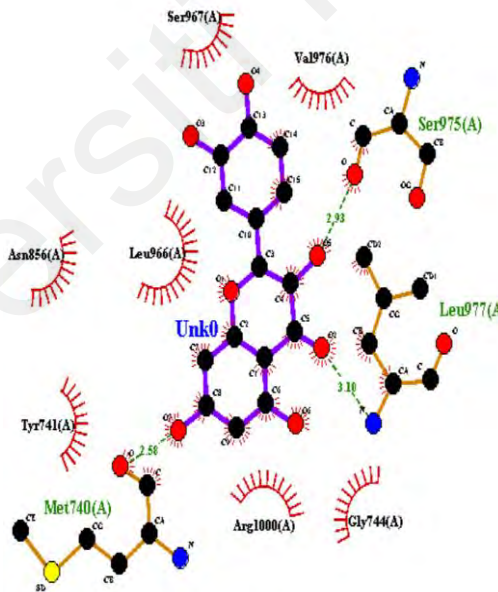


Figure 4.8: Spike-positive complex interactions



#### 4.4 Molecular Dynamics (MD) Simulation

Lig126-spike complex was selected for MD simulation to verify its stability. The MD simulations of unbounded protein and Lig126-protein complex were run for 50ns time scale. MD simulation performance of the unbounded protein and Lig126-protein complexes were assessed, and the corresponding graphs were generated by Grace and excel as presented in Figures 4.9 and 4.10, respectively.

Energy and trajectory results of both MD simulations presented a similar result which the protein system minimization reached  $-9.11217 \times 10^6$  kJ/mol while that of  $-9.6966 \times 10^6$  kJ/mol in the complex, Figures (4.9A1&2). The temperature and pressure reached 299.999 and 0.9229 bar consecutively. This indicates that throughout the simulation period, the system is stable and able to achieve the set temperature and pressure. Figures (4.9 B&C).

The RMSD plot of Lig126 ligand, spike protein, and Lig126-spike complex are shown in figure (4.10.A). The carbon-alpha ( $C\alpha$ ) (orange) showed three different phases were a gradual increment from 0-20ns, a slight decrement from  $\sim 20$  ns to  $\sim 30$  ns, a stable trend from 30 to 50 ns. The Lig126-spike complex (blue) illustrates slight fluctuations in the first 10 ns but then maintained a stable trend throughout the remainder of the simulation. While the ligand126 (green) showed a steady trend throughout the simulation period.

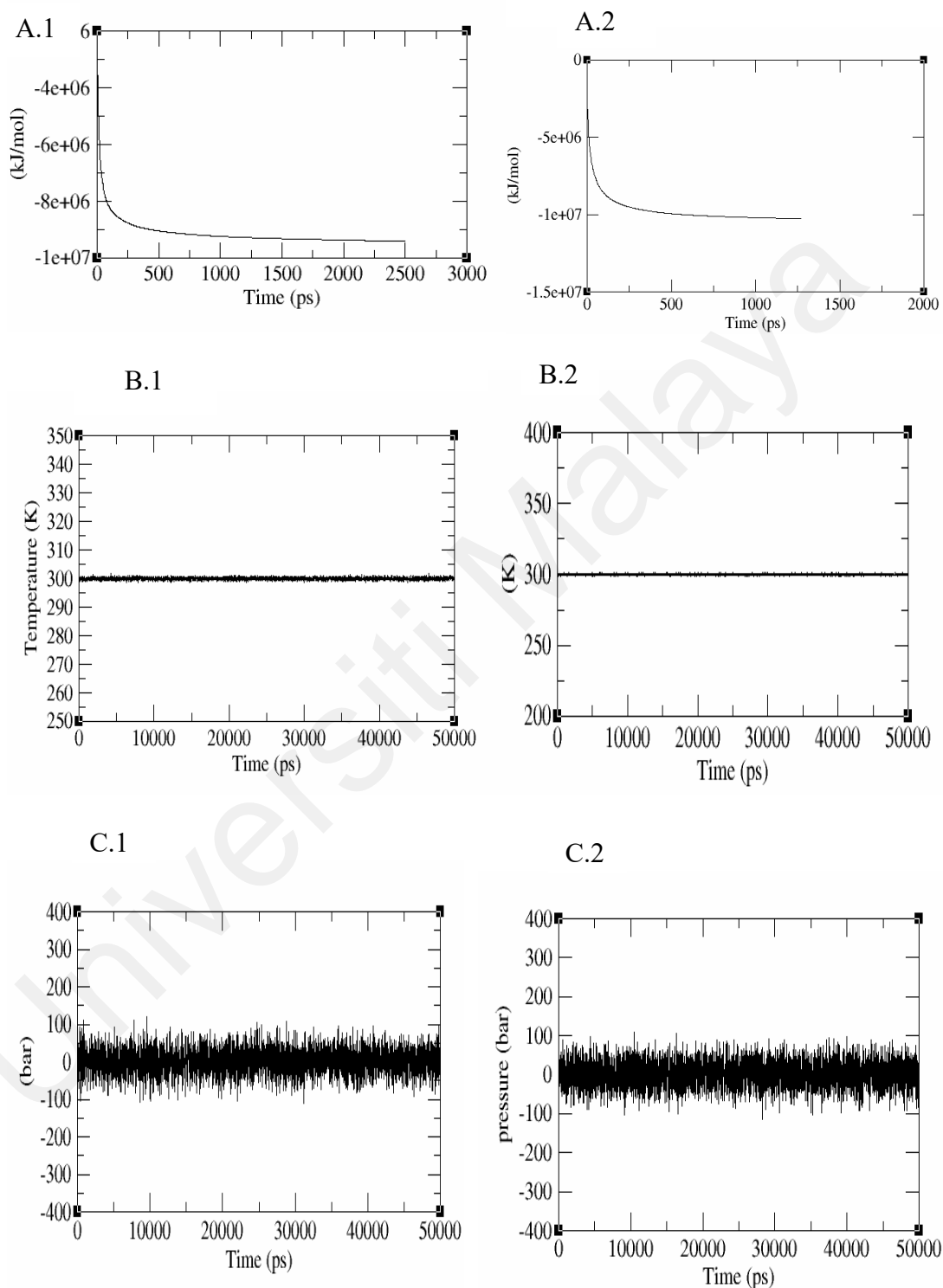
RMSF plot refer to the protein was very flexible especially at around 750 and 1000 residues. RMSF plot of undocked protein showed two significant fluctuation regions which are between 730 -790 residues and 950-1100 residues. Undocked protein shows higher fluctuation trend than the docked protein, indicating that the protein become more stable after binding to Lig126, and huge conformational and structural changes of the docked structure (Figer4.10 B). The protein-ligand complex H bond plot showed five

hydrogen bonds formed in the initiation time, and 2 of them were maintained continuously during the 50 ns simulation (Figer4.10 C).

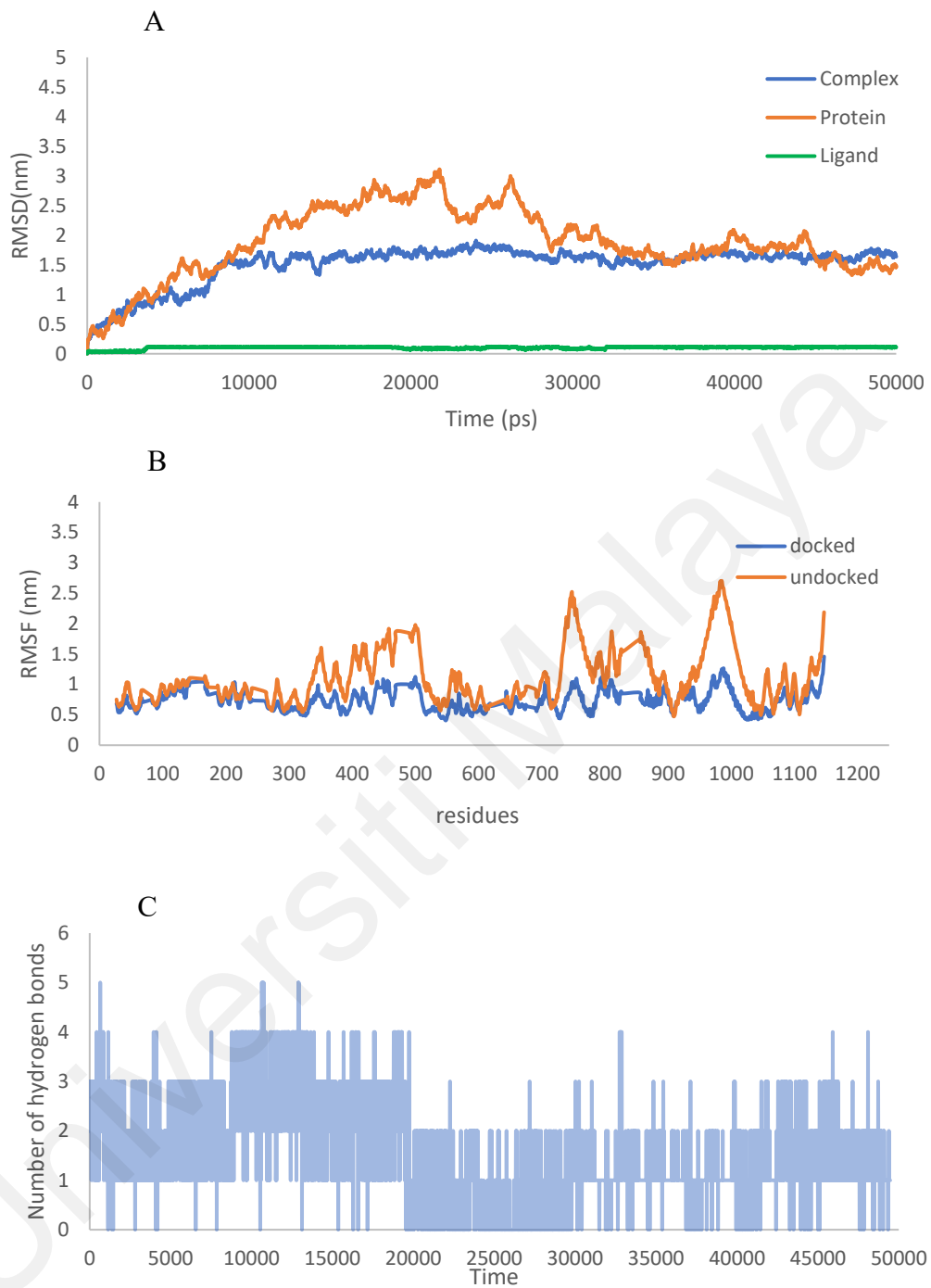
#### **4.5 Prediction of ADME by computational analysis**

The chemical absorption, distribution, metabolism, excretion and toxicity (ADMET) of the promising candidate as well as the positive control were generated by swissADEM online software. The pharmacokinetics result showed similarity results which both had a high gastrointestinal GI absorption average and had no permeant of blood brain barrier (BBB). Both had inhibitor activity for CYP1A2, CYP2D6 and CYP3A4. The skin permeability (Log Kp) was under acceptable range which are -6.01cm/s and -7.05 cm/s for Lig126 and positive respectively. Additionally, AMES toxicity test for both showed non carcinogenic effect. Both of them are drug likeness (Table 4.6).

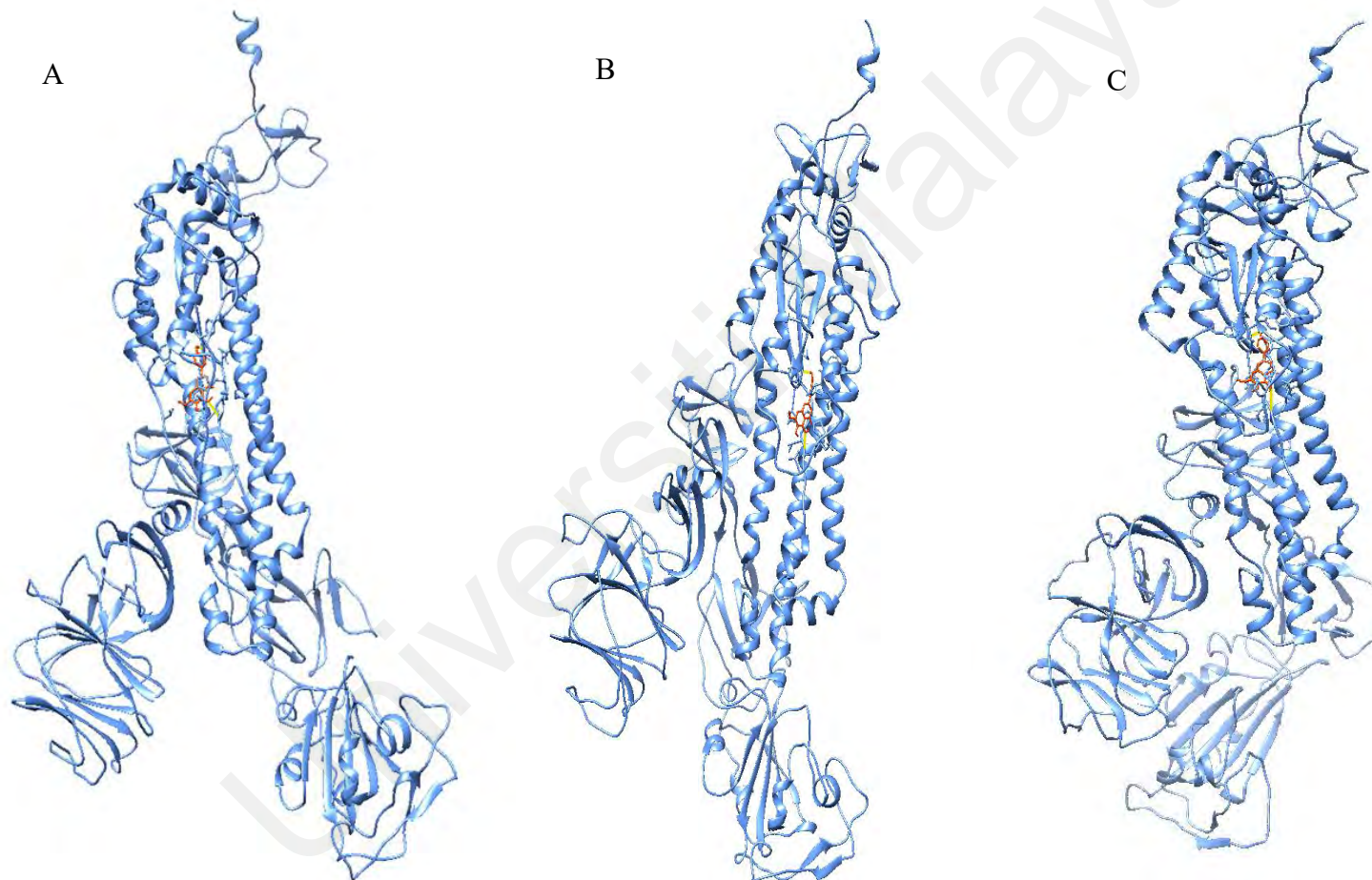
Universiti Malaysia



**Figure 4.9: Graphs for simulation result of spike protein and protein-Lig126 complex. The potential energy (A.1) for protein & (A.2) for complex. Temperature (B.1) for protein & (B.2) for complex, pressure (C.1) for protein & (C.2) for complex.**



**Figure 4.10: Trajectory graphs generated to study fluctuation and stability during 50n MD simulation. (A) RMSD for protein, ligand and complex. (B) RMSF plot for docked and undocked protein. (C) protein- ligand H-bond complex.**



**Figure 4.11: Lig126-spike complex at different time during 50 (ns) MD simulation. (A) at 1 (ns), (B) at 20 (ns) and (C) at 50 (ns). Spike protein (blue) ligand126 (red) hydrogen bonds (yellow).**

**Table 4.6 ADEMT for Lig120 and Positive.**

Parameters	Lig-126	Positive
GI Absorption	High	High
BBB Permeant	No	No
CYP2c19	No	No
CYP2D6	Yes	Yes
Log Kp	-6.01cm/s	-7.05 cm/s
AMES Toxicity	Non-toxic	Non-toxic

Universiti Malaysia

## CHAPTER 5: DISCUSSION

### 5.1 Protein Preparation.

The principal targets in the search for SARS-CoV-2 virus inhibitors are spike protein, spike protein-ACE-2 receptor complex, ACE-2 receptor, RdRp, 3CLpro, PLpro, and N-protein RNA-binding domain(Zhang & Tang, 2021). In this study, SARS-CoV-2 spike protein was selected because of its importance in viral entry to host cell(Huang et al., 2020). The spike protein of SARS-CoV-2 consist of S1 subunit that recognizes and binds to the host receptor and S2 subunit that mediates viral cell membrane fusion(Du et al., 2009; Tian et al., 2020). Several SARS-CoV-2 spike protein structures already in the protein data bank from different variants were discovered during the protein structure acquisitions. While some were closed or open confirmations before fusion, others were post-fusion. According to a prior study, S glycoprotein trimmers appear to exist in partially opened forms in highly pathogenic human coronaviruses (Shang et al., 2020) and the most dangerous coronaviruses will exhibit S glycoprotein trimmers spontaneously sampling closed and open conformations, as (Gur et al., 2020). This data suggests that the semi-open state transition is a crucial stage in the host cell recognition of SARS-CoV-2 and that this intermediate may be a promising drug binding target. Therefore, the spike protein's open confirmation structure was the intended target.

Protein Data Bank is a primary database of protein 3D structures containing nucleic acids and proteins data from various organisms that were Determined using several experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR), spectroscopy, and electron microscopy(Burley et al., 2019; Liu et al., 2018). X-ray structure of target protein is commonly used in-silico studies. Since X-ray crystallography structure of the completed structure of spike protein was unavailable in the protein data bank due to its length. In this study three-dimensional structure of an

open structure spike protein with PDB ID:6VYB, and 3.20Å resolution from electron microscopic analysis of open was selected(Walls et al., 2020).

## 5.2 Ligand Library Preparation

Nuclei of Bioassays, Ecophysiology and Biosynthesis of Natural Products database (NuBBE) are a natural sources library of small molecules (natural compounds) in a variety of chemical classes and secondary metabolites of plants, fungi, insects, marine organisms, and bacteria in three-dimensional structures. This database has been used for drug development studies (do Carmo Santos et al., 2018). In field of searching for SARS-COV-2 inhibitors, this database has been screened against Main Protease (Arifuzzaman et al., 2022; Singam et al., n.d.). NuBBE was used to obtain the natural ligand library.

Lipinski's, which is physicochemical properties of drug-like compounds. They were mainly used to assess the likelihood of a molecule with a specific pharmacological or biological action being employed as a potential drug. As Lipinski claims, any compound chosen as a possible medication should follow Lipinski's rule, which refers to the physicochemical characteristics of drug-like compounds used to determine the possibility that a molecule with a certain biological action would be identified as possible drug (Lipinski et al., 2001). Also, Lipinski's stipulates each chosen ligand as a promising inhibitor for drug discovery should have: a molecular weight less than 500 Dalton, lipophilicity less than 5, a hydrogen bond not more than 5 for donors, and 10 for acceptors (Lipinski et al., 2001). Small molecules that follow Lipinski's boosted its potential to be an antiviral with an acceptable ADMET profile. Therefore, applying these criteria to natural ligands boosted its potential to be a successful antiviral agent.



## 5.3 Virtual Screening

### 5.3.1 Molecular Docking

Virtual screening is a computational approach to screening ligand libraries (natural compounds) against target protein (spike protein). Molecular docking is often used to investigate binding interactions of a potential medication with target protein (Santos E R., 2019). Blind docking was conducted to detect possible binding sites for the ligand with the S protein, and to study the binding affinity. Using energy estimates, these binding sites were ranked according to their lowest binding energies. The binding energy demonstrates the capacity of a ligand to bind to a protein. The most negative value for binding energy refers to the energy required for ligand binding being lower. And the ligand's affinity for the protein is higher. The top ten were selected regarding their binding energy value (Arifuzzaman et al., 2022; Gahlawat et al., 2020; Maia et al., 2020). Table 4.3 shows that the binding energy value of the selected ligands was almost the same, with a slight difference ranging from -7.8 to -6.9 kcal/mol. However, all of them were lower than the positive control compound (-6.37 kcal/mol), with a slight difference between them, about -7 kcal/mol. Indicates that they almost have a good affinity. They have the ability to inhibit spike protein.

The protein-ligand interaction analysis for selected ligands showed that most ligands formed one H bonds with different sites of the protein with a slight difference in binding energy value. These hydrogen bonds are necessary to maintain the structural stability of the complex. In comparison, Lig126 and lig564 showed tow hydrogen bonds interactions with the same residues LYS733 and G1056. This was close to where the positive control bound ser975 &, Met740, Leu79, Arg1000 and it leads us to a possible hypothesis that this site is the protein's active site. Although Lig126 and Lig564 showed similarities in binding site and h bonds formation, Lig126 showed a lower binding energy value (Pandey et al., 2020). Thus, Lig126 was selected as promising compound four further analysis.

Figure 4.7 showed that Lig126 interacted with the S2 domain of the spike protein by forming hydrogen bonds with the residues ALA1056, GLY1059, and LYS733 and by interacting hydrophobically with the residues Thr732, pro863, val1860, Asp876, leu870, his 1058, phe782, ser730, and met731. Figure 4.8 showed the positive control interacted with S2 domain of spike protein with ser975, Met740, and Leu977 and displayed hydrophobic interaction with Gly744, Val976, Ser967, Leu966, Asn856, Tyr741, Arg1000.

#### **5.4 Molecular Dynamics Simulations**

Molecular dynamics simulations are a valuable method used to analyze ligand-receptor interactions at the atomic level in-silico drug design(Singh et al., 2021). Based on the protein ligand interaction analysis and binding energy value spike-Lig126 complex was simulated for 50ns duration to study the complex conformation stability and hydrogen bonds formation. The systems' energy minimization was effective, as indicated by the consistent convergence of the potential energy trend during the simulation. Additionally, the systems' temperatures stayed constant such that they did not exceed or fall below the 300 established controlled temperature 300 K, the pressure were measured 1 bar , as stated in the simulation (figure4.9).

At the start of the simulation, the RMSD graph exhibited a high value, which stabilized by the end. After 5ns of simulation time, the RMSD for the ligand demonstrated stability lower than 0.1nm. As opposed to protein, which demonstrated stability after 30 ns of simulation. Comparatively, the complex exhibited far less translational motion than the protein, which exhibited stability after ten ns at 1.7 nm throughout the duration of the simulation. The simulation's beginning wasn't in the steadiest state. That finding of the complex's stable pose was aided by the simulation. With possible interactions, the ligand that was present in the protein stabilized . Further, the protein-ligand complex structures

at 0 ns, 20 ns, and 50 ns were generated to determine if the protein was still compact (Pandey et al., 2020).

The root means square fluctuation (RMSF) graph for undocked and docked proteins revealed substantial fluctuation in three regions of the C terminal of the S1 domain, 750-800 and 950-1050 residues. A considerable variance indicated a significant degree of flexibility in RMSF. The graph demonstrated that the docked protein fluctuated less than the undocked protein, particularly in the region where the ligand was bound. Indicated that despite the protein's extreme flexibility, the addition of the ligand boosted its stability (Pandey et al., 2020).

In order to assess the molecular interaction, H bond plots were created shown in Figure (4.10). At the starting time three H-bonds exist between the protein and ligand, while two H-bonds observed continually throughout the simulation. Additionally, these two hydrogen bonds. these two hydrogen bonds were also observed in the three structures viewed at various moments during the MD simulation. appeared in the three structures that were viewed using MD simulations at different times (figure 4.11). This demonstrates that protein formed a robust connection with ligand126. Hydrogen bonds formation between Lig126 and protein were demonstrated at two parts of current study. in protein ligand interaction analysis after molecular docking and during the complex simulations.

## **5.5 ADMET Profile**

The outcome revealed that the profiles of compound126 and positive were nearly identical. However, they would be unable to pass through the blood-brain barrier. Less toxicity to the brain results from this barrier preventing therapeutic drugs in the blood from entering the brain. IG would be an oral medication or nasal spray because it has a high potential GI absorption rate (Safitri et al., 2020). The toxicology measures the harm a substance causes to the human body. According to the AMES test, Lig126 is not mutagenic and is capable of passing both in vitro and in vivo toxicity tests. A study was

reported that Lig126 (Sorbifolin) have the ability to be anti-diabetic by inhibiting alpha glucosidase (Safitri et al., 2020).

Universiti Malaya

## CHAPTER 6: CONCLUSION

In conclusion, the plant-based compound Lig126, showed antiviral activity to inhibit spike protein. that can be used as first step for discovering a new drug, against SARS-CoV-2 virus. Using a structure-based virtual screening technique, a molecule with the potential to inhibit the SARS-CoV-2 spike protein was identified. This in-silico computational approach offers inhibitors identification and save the overall drug discovery cost and time. By virtual screening of NuBBE database, ten possible compounds with the ability to bind on spike protein were identified. Based on their binding affinity compared to positive control. The result showed Lig585, Lig1576, lig1769, Lig1507, Lig126, Lig1768, Lig991, Lig564, Lig1370) have a good binding affinity compared to positive control. By analyzing complexes interaction of selected compounds and positive control, a promising compound (Lig126) was selected for further analysis based on its binding affinity and preferred residues. The result showed Lig126, Lig564 exhibited identical binding preferences for the same spike protein residues. These residues close to the positive control binding site. Lig126 showed higher binding affinity than Lig654. MD simulation revealed the dynamics of the protein-Lig126 complex over 50 ns compared to protein simulation. By applying MD simulations for promising ligand complex with spike protein, the binding interaction analysis of complex were done second time. And by comparing the docked and undocked spike protein simulation trajectories the evaluate of complex stability was done . The result showed that the protein very flexible and become more stable with Lig126 complex. The Lig126 remained bounded to Lys733, Gly1059 residues of spike protein with hydrogen bonds throughout the 50ns simulation. ADMET predicted that promising compound consisted of drug-likeness property. To verify current study, more research should be conducted using in vitro and in vivo experiments to determine the inhibitory impact of the compound126.

## REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Adney, D. R., Wang, L., Van Doremalen, N., Shi, W., Zhang, Y., Kong, W. P., Miller, M. R., Bushmaker, T., Scott, D., de Wit, E., Modjarrad, K., Petrovsky, N., Graham, B. S., Bowen, R. A., & Munster, V. J. (2019). Efficacy of an adjuvanted middle east respiratory syndrome coronavirus spike protein vaccine in dromedary camels and alpacas. *Viruses*, 11(3). <https://doi.org/10.3390/v11030212>
- Agency, U. K. H. S. (2022, May 17). Covid-19: Background information. GOV.UK. Retrieved June 11, 2022, from <https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information>
- Al-Khafaji, K., AL-Duhaidahawi, D., & Taskin Tok, T. (2021). Using integrated computational approaches to identify safe and rapid treatment for SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 39(9), 3387–3395. <https://doi.org/10.1080/07391102.2020.1764392>
- Antonela, M., Rivero, M., & Ramiro, H. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information . January.
- Aoe, T. (2020). Pathological aspects of COVID-19 as a conformational disease and the use of pharmacological chaperones as a potential therapeutic strategy. *Frontiers in pharmacology*, 11, 1095.
- Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., Peisach, E., ... Ioannidis, Y. E. (2019). Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1), D520–D528. <https://doi.org/10.1093/nar/gky949>
- Carbo, E. C., Sidorov, I. A., Zevenhoven-Dobbe, J. C., Snijder, E. J., Claas, E. C., Laros, J. F. J., Kroes, A. C. M., & de Vries, J. J. C. (2020). Coronavirus discovery by metagenomic sequencing: a tool for pandemic preparedness. *Journal of Clinical Virology*, 131(August), 104594. <https://doi.org/10.1016/j.jcv.2020.104594>
- Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., ... & Amaro, R. E. (2020). Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS central science*, 6(10), 1722-1734.
- Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., ... & Shapiro, L. (2021). Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell host & microbe*, 29(5), 819-833.

- Chan, J. F. W., Kok, K. H., Zhu, Z., Chu, H., To, K. K. W., Yuan, S., & Yuen, K. Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes and Infections*, 9(1), 221–236. <https://doi.org/10.1080/22221751.2020.1719902>
- Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., & Huang, T. H. (2014). The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral research*, 103, 39–50.
- Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W. C., Wang, C. Bin, & Bernardini, S. (2020). The COVID-19 pandemic. *Critical Reviews in Clinical Laboratory Sciences*, 0(0), 365–388. <https://doi.org/10.1080/10408363.2020.1783198>
- Clara-Rahola, J. (2020). An empirical model for the spread and reduction of the CoVid19 pandemic. *Estudios de Economia Aplicada*, 38(2). <https://doi.org/10.25115/EEA.V38I2.3323>
- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., & Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research*, 176(February), 104742. <https://doi.org/10.1016/j.antiviral.2020.104742>
- Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), 181–192. <https://doi.org/10.1038/s41579-018-0118-9>
- Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nature Reviews Microbiology*, 7(3), 226–236.
- El-Demerdash, A., Metwaly, A. M., Hassan, A., El-Aziz, T. M. A., Elkaeed, E. B., Eissa, I. H., Arafa, R. K., & Stockand, J. D. (2021). Comprehensive virtual screening of the antiviral potentialities of marine polycyclic guanidine alkaloids against sars-cov-2 (Covid-19). *Biomolecules*, 11(3), 1–25. <https://doi.org/10.3390/biom11030460>
- Gordon, D. L., Sajkov, D., Honda-Okubo, Y., Wilks, S. H., Aban, M., Barr, I. G., & Petrovsky, N. (2016). Human Phase 1 trial of low-dose inactivated seasonal influenza vaccine formulated with Advax™ delta inulin adjuvant. *Vaccine*, 34(33), 3780–3786.
- Graham, R. L., Donaldson, E. F., & Baric, R. S. (2013). A decade after SARS: strategies for controlling emerging coronaviruses. *Nature reviews. Microbiology*, 11(12), 836–848. <https://doi.org/10.1038/nrmicro3143>
- Gralinski, L. E., Sheahan, T. P., Morrison, T. E., Menachery, V. D., Jensen, K., Leist, S. R., ... & Baric, R. S. (2018). Complement activation contributes to severe acute respiratory syndrome coronavirus pathogenesis. *MBio*, 9(5), e01753-18
- Harrison, A. G., Lin, T., & Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in Immunology*, 41(12), 1100–1115. <https://doi.org/10.1016/j.it.2020.10.004>

- Huang, Y., Yang, C., Xu, X. F., Xu, W., & Liu, S. W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*, 41(9), 1141-1149.
- Jimenez-Guardeño, J. M., Nieto-Torres, J. L., DeDiego, M. L., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodríguez, C., & Enjuanes, L. (2014). The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS pathogens*, 10(8), e1004320
- Kawase, M., Kataoka, M., Shirato, K., & Matsuyama, S. (2019). Biochemical Analysis of Coronavirus Spike Glycoprotein Conformational Intermediates during Membrane Fusion. In *Journal of Virology* (Vol. 93, Issue 19). <https://doi.org/10.1128/jvi.00785-19>
- Kemp, S. A., Collier, D. A., Datir, R. P., Ferreira, I. A. T. M., Gayed, S., Jahun, A., Hosmillo, M., Rees-Spear, C., Mlcochova, P., Lumb, I. U., Roberts, D. J., Chandra, A., Temperton, N., Baker, S., Dougan, G., Hess, C., Kingston, N., Lehner, P. J., Lyons, P. A., ... Gupta, R. K. (2021). SARS-CoV-2 evolution during treatment of chronic infection. *Nature*, 592(7853), 277–282. <https://doi.org/10.1038/s41586-021-03291-y>
- Kevadiya, B. D., Machhi, J., Herskovitz, J., Oleynikov, M. D., Blomberg, W. R., Bajwa, N., Soni, D., Das, S., Hasan, M., Patel, M., Senan, A. M., Gorantla, S., McMillan, J. E., Edagwa, B., Eisenberg, R., Gurusurthy, C. B., Reid, S. P. M., Punyadeera, C., Chang, L., & Gendelman, H. E. (2021). Diagnostics for SARS-CoV-2 infections. *Nature Materials*, 20(5), 593–605. <https://doi.org/10.1038/s41563-020-00906-z>
- Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell*, 181(4), 914-921.
- Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, 55(3), 105924. <https://doi.org/10.1016/j.ijantimicag.2020.105924>
- Li, C., He, Q., Qian, H., & Liu, J. (2021). Overview of the pathogenesis of COVID 19. *Experimental and Therapeutic Medicine*, 22(3), 1-10.
- Limburg, H., Harbig, A., Bestle, D., Stein, D. A., Moulton, H. M., Jaeger, J., Janga, H., Hards, K., Koepke, J., Schulte, L., Koczulla, A. R., Schmeck, B., Klenk, H.-D., & Böttcher-Friebertshäuser, E. (2019). TMPRSS2 Is the Major Activating Protease of Influenza A Virus in Primary Human Airway Cells and Influenza B Virus in Human Type II Pneumocytes. *Journal of Virology*, 93(21). <https://doi.org/10.1128/jvi.00649-19>
- Lionta, E., Spyrou, G., Vassilatis, D., & Cournia, Z. (2014). Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry*, 14(16), 1923–1938. <https://doi.org/10.2174/1568026614666140929124445>
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., ... & Wang, X. (2020). Structure of the



SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *nature*, 581(7807), 215-220.

Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H., & Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. *Expert Opinion on Drug Discovery*, 13(1), 23–37. <https://doi.org/10.1080/17460441.2018.1403419>

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395:565–74.

Millet, J. K., & Whittaker, G. R. (2018). Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. *Virology*, 517(December 2017), 3–8. <https://doi.org/10.1016/j.virol.2017.12.015>

Mohammadi, M., Rupa, F. H., Khan, M. F., Rashid, R. B., & Rashid, M. A. (2022). Identification of Natural Compounds with Anti-SARS-CoV-2 Activity using Machine Learning, Molecular Docking and Molecular Dynamics Simulation Studies. *Dhaka University Journal of Pharmaceutical Sciences*, 21(1), 1-13.

Moreira, R. A., Guzman, H. V., Boopathi, S., Baker, J. L., & Poma, A. B. (2020). Characterization of structural and energetic differences between conformations of the SARS-CoV-2 spike protein. *Materials*, 13(23), 1–14. <https://doi.org/10.3390/ma13235362>

Morse JS, Lalonde T, Xu S, Liu WR. Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *Chembiochem*. 2020;21:730–8.

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., Atif, S. M., Hariprasad, G., Hasan, G. M., & Hassan, M. I. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1866(10), 165878. <https://doi.org/10.1016/j.bbadis.2020.165878>

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., Xiang, Z., Mu, Z., Chen, X., Chen, J., Hu, K., Jin, Q., Wang, J., & 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). <https://doi.org/10.1038/s41467-020-15562-9>

Pandey, P., Rane, J. S., Chatterjee, A., Kumar, A., Khan, R., Prakash, A., & Ray, S. (2021). Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. *Journal of Biomolecular Structure and Dynamics*, 39(16), 6306-6316.

Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K., Yuen, K. Y., & SARS study group (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet (London, England)*, 361(9366), 1319–1325. [https://doi.org/10.1016/s0140-6736\(03\)13077-2](https://doi.org/10.1016/s0140-6736(03)13077-2)

- Pilon, A. C., Valli, M., Dametto, A. C., Pinto, M. E. F., Freire, R. T., Castro-Gamboa, I., Andricopulo, A. D., & Bolzani, V. S. (2017). NuBBEDB: An updated database to uncover chemical and biological information from Brazilian biodiversity. *Scientific Reports*, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-07451-x>
- Pinzi, L., & Rastelli, G. (2019). Molecular docking: Shifting paradigms in drug discovery. *International Journal of Molecular Sciences*, 20(18). <https://doi.org/10.3390/ijms20184331>
- Rapaport, D. C., & Rapaport, D. C. R. (2004). *The art of molecular dynamics simulation*. Cambridge university press
- Robson, B. (2020). COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance. *Computers in Biology and Medicine*, 121, 103749. <https://doi.org/10.1016/j.combiomed.2020.103749>
- Rudrapal, M., & Chetia, D. (2020). Virtual Screening, Molecular Docking and QSAR Studies in Drug Discovery and Development Programme. *Journal of Drug Delivery and Therapeutics*, 10(4), 225–233. <https://doi.org/10.22270/jddt.v10i4.4218>
- Safitri, A., Sari, D. R. T., Roosdiana, A., & Fatchiyah, F. Molecular Docking Approach of Potential Alpha Glucosidase Inhibitors from Extracts Compounds of *Ruellia tuberosa* L. *Journal of Smart Bioprospecting and Technology P-ISSN*, 2686, 0805.
- Saldívar-González, F. I., Valli, M., Andricopulo, A. D., Da Silva Bolzani, V., & Medina-Franco, J. L. (2019). Chemical Space and Diversity of the NuBBE Database: A Chemoinformatic Characterization. *Journal of Chemical Information and Modeling*, 59(1), 74–85. <https://doi.org/10.1021/acs.jcim.8b00619>
- Santos, L. H., Ferreira, R. S., & Caffarena, E. R. (2019). Integrating molecular docking and molecular dynamics simulations. *Docking screens for drug discovery*, 13-34.
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*, 117(21), 1–8. <https://doi.org/10.1073/pnas.2003138117>
- Singam, E. R. A., Merrill, M. A. La, Durkin, K. A., & Smith, M. T. (n.d.). 2 , 3 \*
- Singh, S., Baker, Q. B., & Singh, D. B. (2022). Molecular docking and molecular dynamics simulation. In *Bioinformatics* (pp. 291-304). Academic Press.
- Tang, X., Wu, ChangchengLi, X., Song, Y., Yao, X., Wu, X., Dung, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., & Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, 1–24. <https://academic.oup.com/nsr/advance-article-abstract/doi/10.1093/nsr/nwaa036/5775463>
- Tian, X., Li, C., Huang, A., Xia, S., Lu, S., Shi, Z., ... & Ying, T. (2020). Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human

monoclonal antibody. *Emerging microbes & infections*, 9(1), 382-385.

Tortorici, M. A., & Veessler, D. (2019). Structural insights into coronavirus entry. In *Advances in Virus Research* (1st ed., Vol. 105). Elsevier Inc. <https://doi.org/10.1016/bs.aivir.2019.08.002>

Verma, S., Twilley, D., Esmear, T., Oosthuizen, C. B., Reid, A. M., Nel, M., & Lall, N. (2020). Anti-SARS-CoV natural products with the potential to inhibit SARS-CoV-2 (COVID-19). *Frontiers in Pharmacology*, 11, 561334.

Wahl, A., Gralinski, L. E., Johnson, C. E., Yao, W., Kovarova, M., Dinnon, K. H., Liu, H., Madden, V. J., Krzystek, H. M., De, C., White, K. K., Gully, K., Schäfer, A., Zaman, T., Leist, S. R., Grant, P. O., Bluemling, G. R., Kolykhalov, A. A., Natchus, M. G., ... Garcia, J. V. (2021). SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature*, 591(7850), 451–457. <https://doi.org/10.1038/s41586-021-03312-w>

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veessler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2), 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>

Wang, Y., Grunewald, M., & Perlman, S. (2020). Coronaviruses: an updated overview of their replication and pathogenesis. *Coronaviruses*, 1-29.

Weinberger, B. (2018). Adjuvant strategies to improve vaccination of the elderly population. *Current Opinion in Pharmacology*, 41(Table 1), 34–41. <https://doi.org/10.1016/j.coph.2018.03.014>

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260–1263. <https://doi.org/10.1126/science.aax0902>

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., Ying, T., Liu, S., Shi, Z., Jiang, S., & Lu, L. (2020). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cellular and Molecular Immunology*, 17(7), 765–767. <https://doi.org/10.1038/s41423-020-0374-2>

Yu, W., & Mackerell, A. D. (2017). Computer-aided drug design methods. *Methods in Molecular Biology*, 1520, 85–106. [https://doi.org/10.1007/978-1-4939-6634-9\\_5](https://doi.org/10.1007/978-1-4939-6634-9_5)

Zhao, H., & Caflisch, A. (2015). Molecular dynamics in drug design. *European Journal of Medicinal Chemistry*, 91, 4–14. <https://doi.org/10.1016/j.ejmech.2014.08.004>

Zhang, Y., & Tang, L. V. (2021). Overview of Targets and Potential Drugs of SARS-CoV-2 According to the Viral Replication. *Journal of proteome research*, 20(1), 49–59. <https://doi.org/10.1021/acs.jproteome.0c00526>

Zigolo, M. A., Goytia, M. R., Poma, H. R., Rajal, V. B., & Irazusta, V. P. (2021). Virtual screening of plant-derived compounds against SARS-CoV-2 viral proteins using computational tools. *Science of the Total Environment*, 781, 146400. <https://doi.org/10.1016/j.scitotenv.2021.146400>