THERAPEUTIC OPTIONS OF COVID-19 MANAGEMENT: COMPUTATIONAL APPROACH TARGETING MAIN PROTEASE (M^{PRO})

M.Sc. Thesis

2022

Submitted to

Central Department of Biotechnology

Tribhuvan University

Kirtipur, Kathmandu, Nepal

For partial fulfillment of M.Sc degree in Biotechnology

By

Samiran Subedi

Registration No: 5-2-37-1013-2013

Supervisors

Senior Scientist Dr. Pramod Aryal

Prof. Dr. Rajani Malla

ACKNOWLEDGEMENT

I would like to express my special thanks and gratitude to my supervisor and my mentor Dr. Pramod Aryal for his invaluable guidance and perpetual support. He continuous dedication showed me the right way whenever I ran into any problems during my entire thesis period. He taught me the true meaning and importance of SCIENCE.

I would also like to thank Prof Dr. Rajani Malla for her love and care which has always motivated me towards success. This research became successful as Prof. Dr. Malla, being supervisor showed generosity in supporting the research and gave me advise whenever required.

I am also thankful to Prof. Dr. Rameshwor Adhikary and RECAST for their contributions in managing the accommodation during the tenure of research. Also, I am grateful to Nepal Army for managing the logistics.

I am also thankful to Prof Dr. Krishna Das Manandhar, HOD of Central Department of Biotechnology, Tribhuvan University for letting me complete my thesis works at TU.

Sincere thanks to my colleagues Ms. Bisheshta Nepal, Ms. Kabita Kandel, Ms. Suja Maharjan, Ms. Guheshwori Chataut, Mr. Devraj Mainali and Mr. Siddartha Gautam for their noble contributions and assistance in my thesis works.

I would also like to acknowledge my seniors Mrs. Manju Pun, Ms. Sita Ghimire, Ms. Sabina Thapa Magar, Mrs. Pooja Pathak and Ms. Rita Kumari Oli for assisting me during a time of need.

Last but not the least, I must express my profound gratitude to my parents and my wife Aastha Acharya for providing me continued help and support throughout all these years and through the process of research works and writing this thesis.

Samiran Subedi

Registration No: 5-2-37-1013-2013

LIST OF ABBREVIATIONS

AA	Amino Acid
ACE-2	Angiotensinogen converting enzyme-2
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
ARDS	Acute Rrespiratory Distress Syndrome
BE	Binding Energy
CDC	Centers for Disease Control and Prevention
cFID	Consensus Induced-Fit Docking
CLEVER	Chemical Library Editing, Visualizing and Enumerating Resource
COBRA	Constraint-based Reconstruction and Analysis
CoV	Corona virus
COVID	Corona virus disease
COX-2	Cyclooxygenase-2
CSDD	Center for the Study of Drug Development
CT	Computed Tomography
DDI	Drug-drug Interaction
DNA	Deoxyribonucleic Acid
DPP4	Dipeptidyl Peptidase 4
ECMO	Extracorporeal Membrane Oxygenation
ED	Ensemble Docking
FDA	Food and Drug Administration
Gbk	Gene Bank
HTS	High Throughput Sequencing
LBVS	Ligand-Based Virtual Screening
kD	Kilodaltons
MD	Molecular Dynamics
MERS	Middle East Respiratory Symptom
MM	Molecular Mechanics

NCBI	National Center for Biotechnology Information
nm	Nanometer
NMR	Nuclear magnetic resonance
nsp	Non-structural proteins
ORF	Open reading frame
PAINS	Pan Assay Interference Compounds
PDB	Protein Data Bank
PSA	Polar Surface Area
PTFE	Polytetrafluoroethylene
PVC	Polyvinyl chloride
QM	Quantum mechanics
RBD	Receptor Binding Domain
RNA	Ribonucleic acid
RT-PCR	Real time Polymerase chain reaction
SAM	S-Adenosyl Methionine
SARS	Severe acute respiratory syndrome
SBVS	Structure Based Virtual Screening
~~Tox~~	~~Toxicity~~
TPP	Thiamine pyrophosphate
tPSA	Topological Polar Surface Area
UTR	Untranslated region
VS	Virtual Screening
WHO	World Health Organization

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

ABSTRACT

Main protease (M^{pro}) enzyme of SARS-CoV-2 is involved in the digestion of viral polyproteins and considered as an attractive drug target for antiviral drug design. This study aims to carry out the molecular docking, density functional theory (DFT) studies, prediction of ADMET properties of selected potential antiviral molecules, study the efficacy of selected hits on predicted mutated Main protease (M^{pro}) enzyme, and the drug-drug interaction of potential hits. The study provides an insight into biomolecular interactions to understand the inhibitory mechanism and the spatial orientation of the ligands and further, identification of key amino acid residues within the substrate-binding pocket that can be applied for structure-based drug design. Molecular docking of FDA approved drugs (1167) screened through stringent parameters and our formula for preference index identified tadalafil as competitive inhibitor of proteolytic function of main protease. In the docking studies, tadalafil exhibited superiority in binding with the crystal structure of M^{pro} over the other selected molecules in this study, suggested by the crucial roles of CYS145, HIS163, and GLU166 in the interaction within the active site of COVID-19 M^{pro}. Further, the binding affinity in the predicted mutated structures of main protease was also found to be higher than wild type suggesting its broad use. The thermodynamic properties and molecular orbital properties investigated for this compound followed the required drug like property of a probable drug. The original library of FDA compounds was docked with human cytochrome P450 3A4 and the inhibitory activity of drospirenone suggested its impeding caliber in clearance of tadalafil. Overall, the present *in silico* study indicated the direction to combat COVID-19 using FDA-approved drugs as therapeutic alternative for viral inhibition or symptom management, which do not need much toxicity studies and could also serve as starting points for lead optimization in drug discovery.

Keywords: ADMET, COVID-19, drug discovery, M^{pro}, molecular docking, therapeutic alternative

1. INTRODUCTION

1.1 Background

The occurrence of new strains virus after mutation is ever-going process in the nature. The occasional outbreaks of pandemic following the development of different strains of viruses is also seen time and again. Never before the outbreak of corona virus was declared a pandemic. But World Health Organization (WHO) declared the outbreak of corona virus disease – 2019 (COVID-19) as a pandemic. The first case of COVID-19 was detected on 31st December 2019 in Wuhan city of China and since then, it has spread to 213 countries and few territories of the world in just over six months mainly due to human-to-human transmission (Khan *et. al.*, 2020).

The modes of transmission of the COVID-19 including airborne, droplets, respiratory secretions and direct contact with the infected patient made it responsible for the rapid spread of this disease all over the world (Guan *et. al.,* 2020). COVID-19 shows from mild symptoms of common cold such as dry cough, running nose, high fever, and body ache to pneumonia in severe cases like the other corona viruses that have previously caused Middle East Respiratory Syndrome (MERS) in 2012 and severe acute respiratory syndrome (SARS-CoV) in 2002–2003. The patients with underlying co-morbidities such as hypertension, diabetes and cardiovascular diseases suffer major complications and mortality rate is high among these group of patients along with elderly people (Wu *et. al.,* 2020). As of 25th September 2020 the global death numbers have reached 4,754,042 (https://www.worldometers.info/coronavirus/). Hence, the requirement for development of effective vaccines or potential treatment is must to combat this global health concern.

In this time of need, repurposing the existing drug library used in the humans can be an avenue forward when the production of vaccines with higher efficiency may take some time for its development (Wouters *et. al*., 2020). Drug repurposing also called as, drug repositioning, has high probability of success rates because it explores the existing therapies for new therapeutic purposes and therefore, information related to

drug's synthetic methodology, safety and toxicology is available beforehand. Thus, drug repurposing potentially reduces the cost and shortens the time for drug discovery in comparison to the conventional *de novo* drug discovery. In addition, use of computational biology approach (March-Vila *et. al.*, 2017) could help in identifying the crucial lead target proteins against which library of approved drugs and other small drug like molecules can be virtually screened to narrow down starting molecules which could be then incrementally modified for development of probable drug candidates either for symptom management or inhibition of the virus.

1.2 Current studies

Vaccines usually take more than 15 years for their production but SARS-CoV-2 vaccines are being tested in a short span of time (Lurie *et. al.,* 2020). The concern for efficacy and safety of the vaccine has to be shown through large, long-term clinical trials for regulatory approval (Buckland, 2005). However, in this time of need the unprecedentedly accelerated development of COVID-19 vaccines has raised concerns whether to approve them with incomplete data, regarding the long-term safety of various populations (Ella and Mohan, 2020).

The vaccine does not always confer full protection against the virus. So alternative treatment regimens are needed to lower the complications of the disease. Many treatments against SARS-CoV-2 are supportive and symptomatic like ventilation, corticosteroids like dexamethasone and immunomodulatory drugs (Rochwerg *et. al.,* 2020). Several antiviral drugs need to be repurposed for this global health concern. WHO has supported the use of certain antiviral drugs such as remdesivir, ritonavir, lopinavir, chloroquine for their repurposing capabilities in COVID-19 (WHO solidarity trial consortium, 2020).

In this research, the works are solely based on application of various computational techniques concerning CADD, to ease the drug repurposing process for development of potential drug candidate capable of inhibition or symptom management and possibly prevent the developed lead molecules from being unsuccessful in the later stages of pre-clinical trials.

1.3 Hypothesis

1.3.1 Null hypothesis:

Potential lead molecules will not be repurposed against the SARS-CoV-2 virus.

1.3.2 Alternative hypothesis:

Potential lead molecules will be repurposed against the SARS-CoV-2 virus.

1.4 Objectives

1.4.1 General Objective

- To repurpose novel lead molecules against the SARS-CoV-2 causing the global health concern.

1.4.2 Specific Objectives

- To identify probable drug targets
- To create a ligand database for virtual screening
- To select 3D protein structures for virtual screening
- To perform molecular docking against the drug target
- To look for lead molecules that could potentially be repurposed as drug candidate for SARS-CoV-2
- To study the potential of developed drugs in mutated strains of virus
- To study the thermodynamic and molecular orbital properties of lead compounds
- To study the drug-drug interaction of lead compounds

1.5 Rationale

The global health crisis after this pandemic has resulted in loss of numerous lives and impeded the global economic growth. The present study is focused on developing a robust method for target identification, lead identification through virtual screening and fostering the probability of effective lead development for proper treatment or symptom management of the COVID-19 disease.

1.6 Scope of study

The present study focuses on repurposing the potential lead molecules among the FDA approved library against main protease in SARS-CoV-2. Also, the efficacy of lead compound in predicted mutations of the same protein target and the drug-drug interaction studies of potential lead compound have been sought.

2. LITERATURE REVIEW

2.1 Review of literature related to corona viruses

2.1.1 History of Corona viruses

Tyrell and Bynoe in 1965 isolated a human coronavirus was for the first time from the nasal secretions of a male child with a common cold (Singhal, 2020). The morphological similarity of the isolated virus with solar corona (crown like) as observed by the electron microscope gave those viruses the name 'corona virus'. Such appearance is mainly due to the presence of spike- glycoprotein which radiates from the viral surface (Esakandari *et. al.,* 2020).

The crossover of animal betacorona viruses to humans has resulted in severe diseases and health havoc in the past two decades on two separate events. A new coronavirus of the β genera and with origin in bats somehow transferred over to humans via the intermediary host of palm civet cats in the Guangdong province of China in 2002-2003 marking the first event of such infection by beta coronaviruses (Hui *et. al.,* 2014). This particular virus, designated as severe acute respiratory syndrome coronavirus affected 8422 people mostly in China and Hong Kong and caused 916 deaths (mortality rate 11%) before being contained (Chan-Yeung and Xu, 2003). The number of cases of SARS increased substantially in the year 2003 in China and later spread globally (Zhong *et. al.,* 2003). The international spread of SARS-CoV in 2003 was attributed to its strong transmission ability under specific circumstances and the insufficient preparedness and implementation of infection control practices. The second event of the beta corona virus infection, was marked by the emergence of the Middle East respiratory syndrome coronavirus (MERS-CoV), also of bat origin, in Saudi Arabia almost a decade later in 2012, with dromedary camels as the intermediate host and affected 2494 people and caused 858 deaths (fatality rate 34%) (**https://www.who.int/emergencies/mers-cov/en/**). The recent betacorona infection, known as COVID-19, caused by SARS-CoV 2 marks the third event within two decades where human health is seriously compromised (Wang *et. al.,* 2020).

2.1.2 Corona viruses: Structure, genome, and classification in general

Corona viruses are large roughly spherical with unique surface projections (Goldsmith *et. al.,* 2004) with variable sizes, generally averaging a diameter of 120 nm, with extreme sizes known from 50 nm to 200 nm and an average total molecular weight of 40,000 kilo daltons (Masters, 2006). The corona viruses are protected by lipid bilayer envelop, membrane proteins and nucleocapsid when outside the host cell (Neuman *et. al.,* 2011).

The membrane (M), envelope (E) and spike (S) **structural proteins** are anchored in the lipid bilayer which comprises the viral envelop (Lai and Cavannagh, 1997) . The E and M protein are the structural proteins that combined with the lipid bilayer shape the viral envelope and maintain its size whereas S proteins are needed for interaction with the host cells (Fehr and Pearlman, 2015).

Corona viruses are enveloped viruses with a positive-sense single-stranded RNA genome and a nucleocapsid of helical symmetry with genome sizes ranging from approximately 26 to 32 kilobases and are one of the largest among RNA viruses (Woo *et. al.,* 2010). The genome organization for a coronavirus is 5'-leader-UTR-replicase (ORF1ab)-spike (S)-envelope (E)-membrane (M)-nucleocapsid (N)-3'UTR-poly (A) tail. The open reading frames 1a and 1b, which occupy the first two-thirds of the genome, encode the replicase polyprotein (pp1ab) which self cleaves to form 16 non-structural proteins (nsp1–nsp16) (Fehr and Pearlman, 2015). The latter reading frames encode for four major structural proteins namely spike, envelope, membrane and nucleocapsid. In between these reading frames, the reading frames coding for the accessory proteins are dispersed and the role and number of these coded accessory proteins varies accordingly with the specific corona virus (Snijder *et. al.,* 2003).

Coronaviruses constitute the subfamily *Orthocoronavirinae*, in the family **Coronaviridae**, order **Nidovirales**, and realm **Riboviria** (Fan *et. al.,* 2019). Coronaviruses are grouped into four sub- families namely alpha, beta, gamma and delta. The alpha and beta coronaviruses originate from mammals, while gamma and

delta coronaviruses have been identified in pigs and birds (Li *et. al.,* 2020). Beta coronaviruses, otherwise referred to as bat corona viruses, cause severe infections whereas alpha corona viruses cause asymptomatic or mildly symptomatic diseases (Velavan and Mayer, 2020). SARS CoV 2, is closely related to the B lineage of the beta-coronaviruses, which are known to cause severe disease and fatalities (Petrosillo *et. al.,* 2020).

2.2 Review of literature related to SARS-CoV 2

2.2.1 Origin and Spread of SARS-CoV 2

2.2.1.1 A global scenario

Wuhan, capital city of Hubei province and a major transportation hub of China started presenting to local hospitals with severe pneumonia among adults of unknown cause in December 2019 (Singhal et al, 2020). Many of the initial cases had a common exposure to the Huanan wholesale seafood market where trading of the live animals was also done (Wang et al, 2020). The activation of the surveillance system (put into place after the SARS outbreak) took place and respiratory samples of patients were sent to reference laboratories for etiological investigations of the then suspected pneumonia cases (Wang *et. al.,* 2020). This outbreak was notified to the World Health Organization on december 31st 2019 by China and on 1st January the closing of sea food market in Huanan took place (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/**.)**The virus was identified as a coronavirus with > 70% similarity to the SARS-CoV and >95% homology with the bat coronavirus(Huang *et. al.,* 2020).

The origin of the virus was traced back to Huanan sea food market when environmental samples from the Huanan sea food market also tested positive for the virus, implying that virus originated from there(https://www.xinhuanet.com/2020-01/27/c_1125504355.htm). The fact that human-to-human transmission was occurring was suggested when the number of cases infected with the virus started increasing exponentially, some of which did not have exposure to the Huanan sea food

market (Huang *et. al.,* 2020). The first fatal case was reported on 11th Jan 2020and the massive migration of Chinese during the Chinese New Year fueled the epidemic to spread on alarming rates. Cases in other countries (Thailand, Japan and South Korea in quick succession) and in other provinces of China, were rising and these cases are believed to have spread from the people who were returning from Wuhan. Transmission to healthcare workers caring for patients was described on 20th Jan, 2020. By 23rd January, the 11 million population of Wuhan was placed under lock down with restrictions of entry and exit from the region and soon the extension of lockdown was also implemented to other cities of Hubei province.

Local human-to-human transmission was occurring in the countries other than China and this assumption was supported by the fact that Cases of COVID-19 in these countries started being reported in those with no history of travel to China (Camilla *et. al.,* 2020). Airports in different countries including Nepal put in screening mechanisms to detect symptomatic people returning from China and placed them in isolation and testing them for COVID-19. Soon it was established that the infection could be transmit rapidly among peoples from contact with asymptomatic people and also before onset of symptoms, creating a sense of havoc amongst the public. Thus, in countries including Nepal, whose citizens were evacuated from Wuhan through special flights or had travelers returning from China, placed all people symptomatic or otherwise in isolation for 14 days and tested them for the possible detection of the infection of the virus. The spread of the virus has now occurred among 220 countries around the world. As of December 19, 2020 the total number of cases detected worldwide have reached a number of 76,081,328 with a reported death toll of 1682465 and 53,340,995 people recovering from the virus infection(https://www.worldometers.info/coronavirus**/**).

2.2.1.2 SARS-CoV-2: Origin, spread and current statistics in Nepal

On Jan 13, 2020, a 32-year-old man, a Nepalese student at Wuhan University of Technology, Wuhan, China, with no history of comorbidities, returned to Nepal (Bastola et al, 2020). The patient visited the outpatient department of Sukraraj Tropical and Infectious Disease Hospital, Kathmandu, with a cough. He had become ill

on Jan 3, 6 days before he flew to Nepal with his claims that he had no exposure to the so-called wet market in Wuhan. Upon real-time RT-PCR assays at the WHO laboratory in Hong Kong the throat swabs obtained from the patient tested positive for COVID-19 (Bastola *et. al.,* 2020).

On 16 February, 175 Nepalese nationals were flown into Nepal from China after exit screening and the repatriated citizen's transportation was managed under quarantine from Tribhuwan International airport to quarantine site at Kharipati, Bhaktapur district, east of Kathmandu where these individuals were placed under 24-h surveillance(https://heoc.mohp.gov.np/update-on-novel-corona-virus-2019_ncov/). Samples of all 175 persons, sampled on 16 February, were reported negative on 19 February. A dedicated ambulance service was available to one of three hospitals—Sukraraj Tropical and Infectious Disease Hospital, Patan hospital or Armed Police Force hospital in case of a positive test or in case of onset of the symptoms amongst these individuals(https://heoc.mohp.gov.np/update-on-novel-corona-virus-2019_ncov/).

On May 16, 2020 Nepal reported the first death from COVID-19 on a female who had recently given birth on May 8 (https://www.aa.com.tr/en/asia-pacific/nepal-registers-its-first-death-from-covid-19/1843635). Despite the lockdown measures applied on different occasions, the spread of corona virus in the different parts of countries is still rising and continues to mount pressure on the government authorities for proper control and surveillance of COVID-19.

Nepal stands on the 40th position on the worldwide scenario with total number of infections suggested by the statistics of 252474 reported cases with 1765 deaths and a population of 241392 detected cases have already recovered as of December 19, 2020 **(**https://www.worldometers.info/coronavirus/). Considering the possibility of travel of infected cases within the internal borders of the country, high vigilance coupled with a strong response plan is required to address the current risk of COVID-19. In this regard, a directive from the Government of Nepal is still required for educating the public to respond to the outbreak through necessary precautions and inform travelers about possible risk despite the current efforts of the government which does not support the confinement of the spread of disease in today's context

(Shrestha *et. al.,* 2020)). There is necessity to identify and contain suspected cases from the site of origin. COVID-19 and past outbreak scenarios should be a learning experience for Nepal not only on emergency management but also towards developing a strong surveillance system and taking preventive actions for similar events as we cannot deny the fact that these types of serious complications won't emerge in the future.

2.3 SARS-CoV-2

Coronaviruses are enveloped, positive sense, non-segmented RNA viruses. They belong to the Coronaviridae family of Nidovirales order (Huang *et. al.,* 2020). They have been classified into four genera such as α- coronaviruses, β-coronaviruses, γ-coronaviruses, and δ-coronaviruses (Wu, Peng et al. 2020). Different variants of SARS-CoV such as bovine coronavirus (BCoV), bat coronavirus (HKU4), and human coronavirus OC43, SARS-CoV-2 belong to β-coronaviruses. They are transmitted through zoonotic transmission and spreads when people are in close contact.

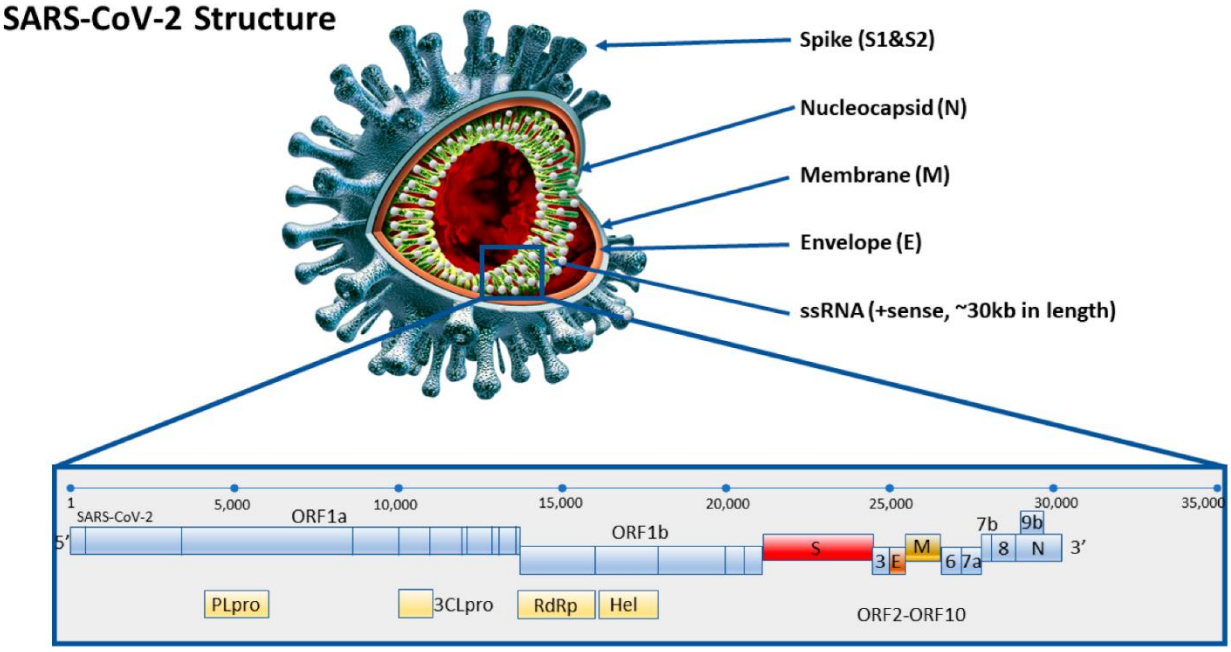


Fig 1 : SARS-CoV-2 structure (Kubina and Dziedzic 2020).

Coronaviruses are enveloped, positive sense, non-segmented RNA viruses. They belong to the Coronaviridae family of Nidovirales order (Huang, Wang et al. 2020).

They have been classified into four genera such as α- coronaviruses, β-coronaviruses, γ-coronaviruses, and δ-coronaviruses (Wu, Peng et al. 2020). Different variants of SARS-CoV such as bovine coronavirus (BCoV), bat coronavirus (HKU4), and human coronavirus OC43, SARS-CoV-2 belong to β-coronaviruses. They are transmitted through zoonotic transmission and spreads when people are in close contact. The infection of the SARS-CoV-2 grows in an alarming exponential rate (Hellewell, Abbott et al. 2020). SARS-CoV-2 shares more than 90% enzymes and proteins' sequence identity to SARS-CoV and MERS-CoV. It has more than 82% identical gene sequence with the mentioned strains. SARS-CoV-2 contains four structural proteins which are termed as spike protein (S), envelope protein (E), membrane protein(M), and nucleocapsid protein (N) (Naqvi, Fatima et al. 2020).

2.3.1 Genetic Sequence of SARS-CoV-2

Coronaviruses (SARS-Cov-2) has total genome of 26 to 32 kilobases long with assumed to have 14 open reading frames that encodes total of 27 proteins with about 7096 resides of long polyprotein (Wu, Peng et al. 2020). Among the structural proteins, spike proteins play significant role in transmission, binding and fusion in the host membrane. Compares to other RNA viruses, it has the highest number of GC that ranges from 32% to 43%. The virus genome has 5' cap and 3' poly-A tail which enables it to act as an mRNA for translation of the proteins involved in the replication (Rahimi, Mirzazadeh et al. 2020). The open reading frame consists of replicase and protease enzymes where the protease is present in 1a to 1b. The major proteins S, E, M and N follow the 5'-3' order in the ORF region (Naqvi, Fatima et al. 2020). The genes are organized in such a fashion that the 5' leader sequence includes *ORF1/ab* followed by *S-ORF3a-E-M-ORF6a-ORF7a-ORF7b-ORF8-N-ORF10-3*'. It is found that SARS-CoV-2 as compared to other β-CoVs lacks hemagglutinin-esterase gene. The largest ORF is the *ORF1a/b* region which is home to 16 nonstructural proteins and the remaining approximately one third genome lies near the 3' terminal with accessory proteins (Mirzazadeh *et. al.,* 2020). ORFs *1a* and *1b* are significantly important than other ORFs because it codes for polyproteins pp1a and pp1b which are processed by proteases to produce 16 conserved proteins (Naqvi, Fatima *et al*., 2020). The 3' terminal has the

four structural proteins and the accessory proteins known as 3a, 3b,p6,7a,7b,8b,9b, and ORF14 (Wu *et. al.,* 2020).

Several studies have reported the mutation in SARS-CoV-2 gene. The most known genes that have mutations include *ORF1ab, ORF3a, ORF6, ORF7, ORF8, ORF10, S, M, E, and N.* Among these genes, most of the genes in *ORF1ab,* and S are found to have notoriously higher rate of mutations. These genes include *nsp1, nsp2, nsp3, nsp12,* and *nsp15.* The prevalence of mutations in these genes have shown strong bias in Uracil in the genome (Alouane *et. al.,* 2020, and Chakraborty *et. al.,* 2020).

2.3.2 Proteins of SARS-CoV-2

The 29.3 kilobases of 2019 Coronavirus gene consits of 1273 amino acids with spike glycoprotein (S), envelope protein (E), membrane protein (M) , nucelocapsid protein (N), receptor binding motif (RTF), M pro, NSP7, NSP13. Some of the membrane proteins that are present in SARS-CoV-2 includes RNA polymerase, 3-chemotrypsin-like protease, papain-like protease, helicase , and accessory proteins.

2.3.2.1 Structural Proteins of SARS-CoV-2

2.3.2.1.1 Spike Protein

The spike protein of SARA-CoV-2 contains about 1273 amino acids and it has 141778 kDa molecular weight (*Chan et. al.,* 2020). It has 50-100 timers of spikes where each spike protein consists of an ectodomain element, transmembrane element, and short C fragment. There are two ectodomains : (a) S1 – facilitates the receptor binding and (b) S2- stem that forms a timer. These protruding projections from the surface are responsible for binding and fusing with the membrane (Anderson *et. al.,* 1983). S protein plays vital role in viral entry into the host cell as it has high affinity to the angiotensin-converting enzyme 2 (ACE2) receptors which are expressed in the pneumocytes in human. The presence of glutamine 493, asparagine 501, leucine 455, phenylalanine 486 and serine 494 amino acids in spike protein helps in binding with

ACE2 receptors. Along with these amino acids, receptor binding domain (RBD) present in S1 subunit recognizes and binds to human ACE2.

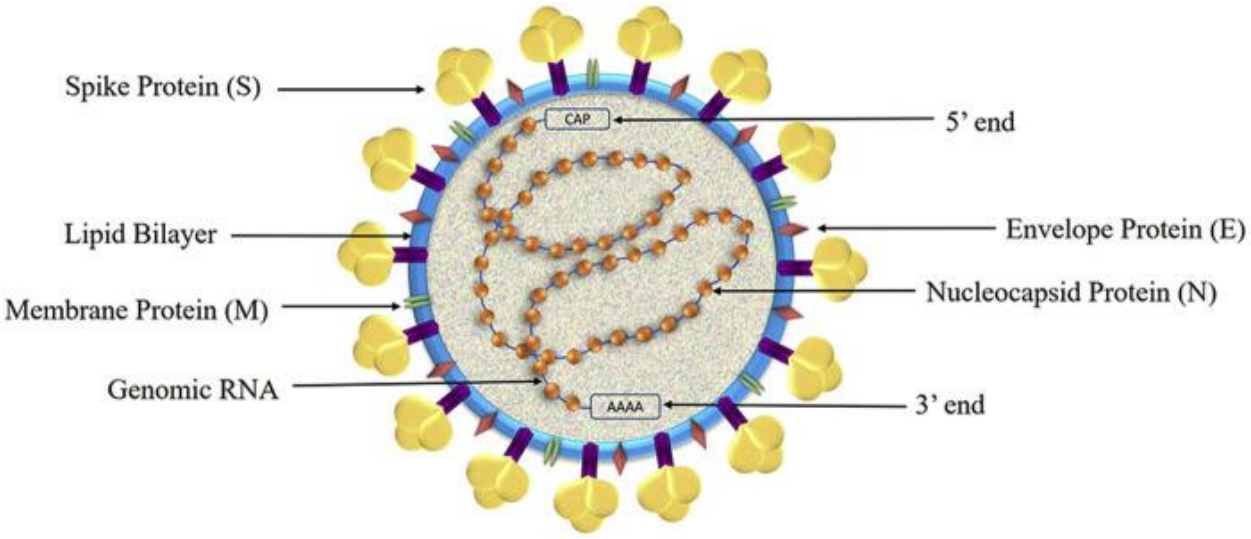


Fig 2: Representation of SARS-CoV-2 (Satarker and Nampoothiri 2020)

2.3.2.1.2 Nucleocapsid Protein

The N protein in SARS-CoV-2 known as the nucleocapsid protein is responsible for the interaction with cellular processes and fusion of virus (Huang *et. al.,* 2004). N protein consists of a N terminal domain (NTD) and C terminal domain (CTD) and a serine rich linker (SR) present between them . Repication transcription complexes (RTC) belonging to N protein plays key role in viral genome synthesis (Gerber *et. al.,* 2019). It is found to inhibit the type I Interferon (IFN) which decreases the immune responses in the host cell. The increase amount of N protein increases the viral RNA synthesis once it comes in contact with Heterogenous nuclear ribonucleoprotein (hnRNPA1) (Luo, Chen et al. 2005). The presence of Arginine at 94 position and Tyrosine at 122 are reported to be essential to bind to SARS-CoV RNA (McBride *et. al.,* 2014). Host samples express the N protein during early stages of infection. N protein is involved in various activities that are essential for the proliferation and functioning of virus.

2.3.2.1.3 Envelope Protein

Envelope Protein (E) has 76-109 amino acids with 8-12 kDa in size (Fung and Liu, 2018). It is integral membrane protein that is composed of NTD, hydrophobic domain and C terminal chain. E protein are responsible for the production of viroporins that are vital for viral assembly, release of virus, mediate pathogenicity and cause cytotoxicity (Ye and Hogue 2007). The tail of E protein is present in cytoplasm of the host cell and targets *cis*- Golgi complex region. The last four amino acids has a postsynaptic density protein /Disc Large / Zonula occulans-1 (PDZ) binding motif whose presence aids in the disruption of the lungs epithelium.

2.3.2.1.4 Membrane Protein

The membrane protien (M) encompasses the highest percentage among all proteins in the virus (Alsaadi and Jones, 2019). M protein is found to exist in two forms - long and compact form. Spike protein is present above these two forms. M protein plays key role in translation and produces the virions after interacting with E protein. M protein inhibits the Nuclear Factor Kappa B activation by interacting with I Kappa B Kinase that enhances the viral pathogenicity. Among the terminals of M protein, C termianl blocks the interaction of 3-phosphoinositide-dependent protein kinase 1 (PDK1) with Protein Kinase B (PKB) that ultimately leads to the apoptosis (Tsoi, Li et al. 2014). Along with this, M protein is involved in the activation of β interferons (IFN-β) (Wang and Liu, 2016). M protein is most responsible for increasing the inflammatory responses in the host and the formation of ribonucleoprotiens.

2.3.2.2 Non-Structural Proteins of SARS-CoV-2

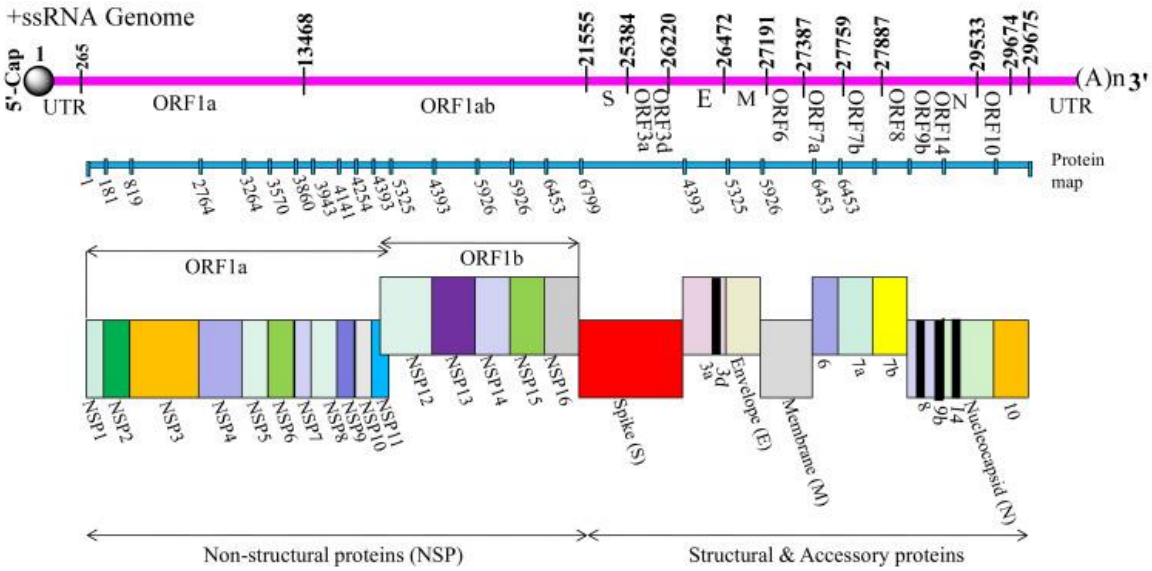


Fig 3 : Different Proteins of SARS-CoV-2 (Yadav, Chaudhary et al. 2021)

There are about 7096 residues long polyprotein that contains many structural and non-structural proteins. Non structural proteins (NSPs) are involved in viral pathogenicity where they participate in transcription regulation, helicase activity, immunomodulation, activation of certain genes, and oppose the antiviral response (Naqvi, Fatima et al. 2020). From the reports publishes by Naqvi , Fatima et al 2020, the functions of some of the NSPs are :

a. Nsp1 : Mediates the RNA replication and processing.

b. Nsp2: Essential for modulation of viral replication.

c. Nsp3: Seperates the translated polyprotein into the distinct proteins.

d. Nsp4: Helps to attach the viral transcription- replication complex with the modified ER membrane

e. Nsp5: Processes viral polyprotein during replication

f. Nsp6: Aids in the initial induction of autophagosomes from host ER.

g. Nsp9: Acts as an ssRNA-binding protein in viral replication

h. Nsp10: Helps in mRNA cap methylation.

i. Nsp12: Involved in replication and transcription of viral RNa genome

j. Nsp 16 : Mediates mRNA cap 2'-O-ribose methylation to the 5'-cap structure of viral mRNAs.

Among the mentioned NSPs, some of them cause the cleavage of human proteins which increases the pathogenicity of coronavius. Such NSPs are NSP2 and NSP5.

- Nsp2 : Non structural protein 2 is the second polyprotein which is found to bind with two host proteins known as prohibitin 1 and prohibitin 2 (PHB1 and PHB2). NSP2 is proposed to interrupt intracellular host siganlling once the cell is infected by SARS-CoV (Cornillez-Ty *et. al.,* 2009). PHB1 and PHB2 proteins in NSP2 is found to play key role in cell cycle progression, cell migration, cellular differentiation, apoptosis, and mitochondrial biogenesis. It is proposes that nsp2 interact with nsp3 to mediate host-specific functions. In case of the mutation rate, it is observed that nsp2 is highly conserved (https://www.proteogenix.science/product/nsp2/).

- Nsp5: Non structural protein Nsp5 is a cysteine protease which participates in viral polyprotein. It is vital for viral replication. It cleaves at different sites to produce mature and intermediate NSPs. This protein belong to the chymotrypsin family which cleaves peptides at P2-P1-P1' residues leucine-glutamine-alanine/serine where the cleavage occurs between P1 and p1' residues (Scott *et. al.,* 2021). Theraupetic appraoches are extensivley researched to target the Ndp5 as they are involved in mRNA processing pathways, sytokine response, cytoskeleton organization and apoptosis. Nsp5/Main Protease (Mpro) is virally encoded protease which cleaves the polyprotein to release Nsps.

2.4 Comparison of SARS-CoV 2 with other corona viruses

2.4.1 Phylogeny

The outbreaks of the SARS-CoV (2002–2003) in China and the Middle East Respiratory Syndrome Coronavirus (2012) in Saudi Arabia can be compared with the recent outbreak of SARS-CoV 2 on the basis of their zoonotic transmission and some similarities in clinical features (Hui et al, 2014). However, phylogenetic analysis of the receptor-binding domain (RBD) of betacoronavirus lineages indicates that SARS-CoV 2 closely belongs to two bat-derived SARS-like coronaviruses (bat-SL-CoVZC45 and bat-SL-CoVZXC21) with 88–89% similarity compared to the 50% and 79% similarity in relation to the SARS-CoV and MERS-CoV respectively (Lai et. al., 2020). The intermediate host through which the SARS-CoV 2 transmitted to human is still debatable in comparison to the palm civets and camels being the intermediate host in the SARS and MERS respectively. Literatures also suggest a different viral evolution of SARS-CoV 2 from that of SARS and MERS involving bat as wild reservoirs (Wu *et. al.,* 2020; Benvunoto *et. al.,* 2020 and Chan *et. al.,* 2020).

There are about 380 amino acid substitutions between SARS-CoV-2 and SARS-like coronaviruses, mostly concentrated in the non-structural protein genes, while 27 mutations have been found in genes encoding the viral spike protein S responsible for receptor binding and cell entry as revealed by the genomic comparisons between SARS-CoV-2 and SARS (Wu *et. al.,* 2020). These mutations might contribute to the apparent lower pathogenicity of SARS-CoV-2 compared with SARS-CoV, but further studies into this subject is essential (Benvunoto *et. al.,* 2020). The similarity among the external sub-domain of receptor binding domain (RBD) amongst the SARS-CoV-2 and SARS indicates the use of ACE2 binding for the entry of viral particles into the host cell (Wan *et. al.,* 2020).

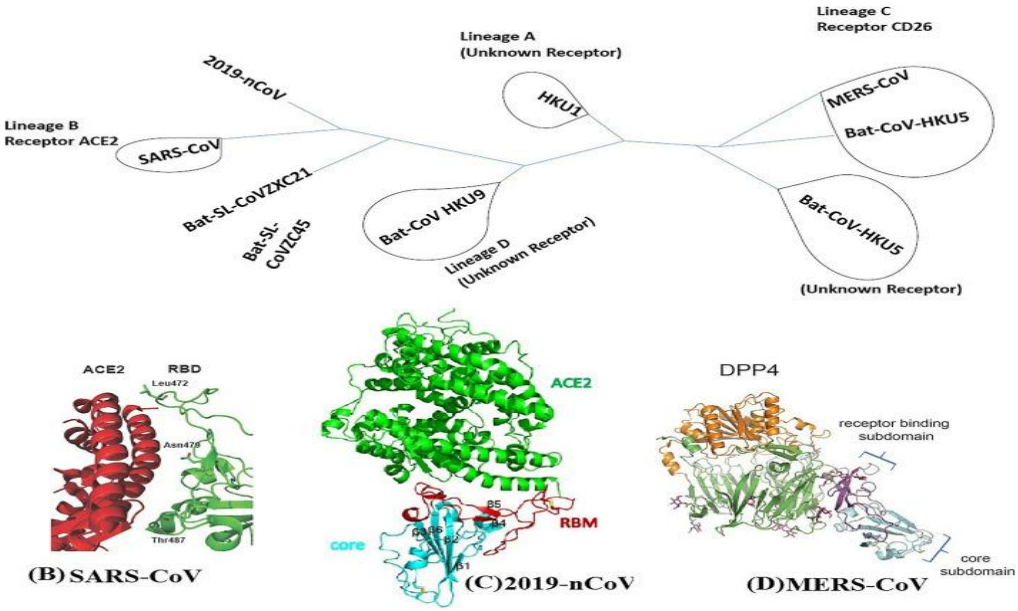


Figure 4: a) Phylogenetic illustration of RBD in various betacorona virus b) c) and d) The structure of RBD in SARS, SARS-CoV-2 and MERS respectively.

2.4.2 Pathogenicity

The same human cell receptor Angiotensin Converting enzyme 2 (ACE-2) is employed by SARS-CoV-2 and SARS for their entry into to the host cell whereas MERS-CoV uses dipeptidyl peptidase 4 (DPP4) to enter host cells (Wan *et. al.,* 2020). The works of Wan et al suggests SARS-CoV-2 binds ACE2 more efficiently than the 2003 strain of SARS-CoV, although less efficiently than the 2002 strain. Further, they have predicted that a single nucleotide mutation on the RBD of SARS-CoV-2, if it occurs, could further increase its pathogenicity.

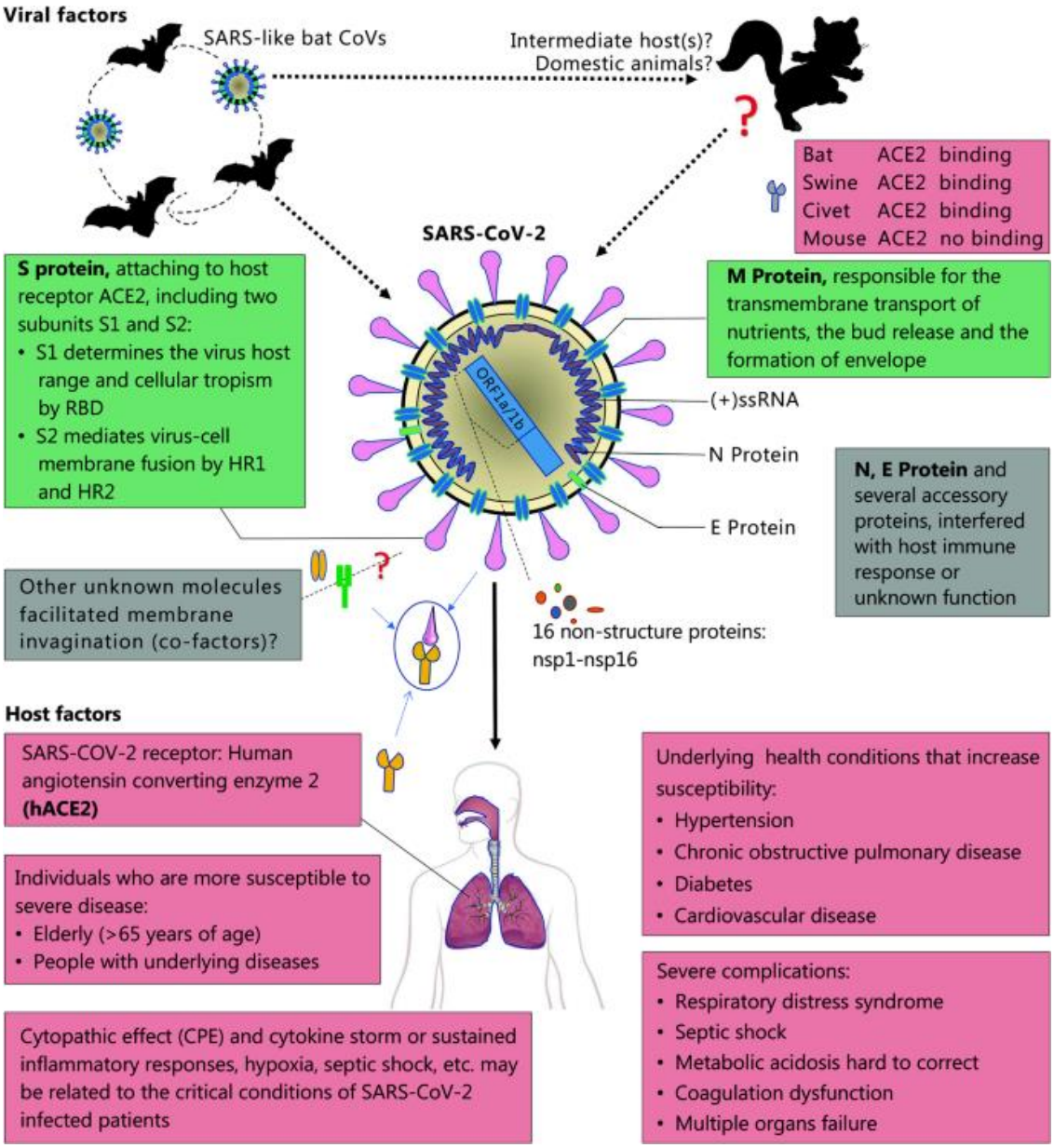


Fig 5: Pathogenicity of SARS-CoV-2 (Guo et .al., 2020)

2.4.3 Transmissibility

The reproductive number (R_0) reflecting the transmissibility of virus was found in the range of 2 and 2.5 for SARS-CoV-2 which is higher than that for SARS (1.7–1.9) and MERS (<1), as estimated by the World Health Organization, which in turn favors the higher pandemic potential possessed by SARS-CoV-2(Chen *et. al.,* 2020; Lui *et. al.,* 2020 and Wu *et. al.,* 2020). There might be certain fluctuations in the values of R_0 since the estimation of R_0 depends on the estimation method used, and the current estimate can be biased by insufficient data and the short onset times of the diseases,

as stated by Liu and colleagues. The fatality rate for novel corona virus is estimated to be 2.3% (Chen, 2020) which is a lower value in comparison to SARS (9.5%) and significantly lower when compared to MERS (34.4%) (Munster et al, 2020).

2.4.4 Clinical features

The onset of dyspnea within 5 days, ARDS within 8 days in 30% of cases, and the need for invasive mechanical ventilation and extracorporeal membrane oxygenation (ECMO) in 17% and 4% of cases, respectively are the major indications relating to the unfavorable clinical course in case of patients diagnosed with COVID-19 (Wang *et. al.,* 2020). The clinical course of MERS seems to be characterized by a more frequent development of ARDS and the need for invasive life support, especially amongst smokers and elderly patients whereas the earlier mentioned findings of clinical manifestations for COVID-19 are in conjuction with SARS percentages (Azhar *et. al.,* 2019). A direct renal cytopathic effect induced by the virus involving DDP4 receptors facilitating the entry of virus, which are largely represented in tubules and glomeruli accounts, for the acute kidney injury in MERS (Cha *et. al.,* 2015) which seldom occurs in SARS and SARS-CoV-2.

2.4.4.1 Clinical manifestations of SARS-CoV 2

2.4.4.1.2 Signs and symptopms

Regarding the four comprehensive studies on the epidemiological and clinical characteristics of SARS-CoV-2 infected patients the most common signs and symptoms of patients are fever and cough (Chen *et. al.,* 2020, Guan *et. al.,* 2020 Huang *et. al.,* 2020 and Wang *et. al.,* 2020). Fatigue was another symptom amongst the patients as complained by 96% of patients (n=138) in one study by Huang and colleagues but was less outstanding (18%, n=44) in another report by Wang and colleagues.

A combinational analysis of the common recorded signs or symptoms of the reported cases found that cough was observed in around 68% of the SARS-CoV-2 infected patients; the number of patients with fever is relatively more (90%) compared to

cough (Zheng, 2020). In addition, shortness of breath or dyspnea, muscle ache, headache, chest pain, diarrhea, haemoptysis, sputum production, rhinorrhoea, nausea and vomiting, sore throat, confusion, and anorexia were also observed in a proportion of the patients (Guan *et. al.,* 2020, Huang *et. al.,* 2020 and Wang *et. al.,* 2020).

Pneumonia is a typical characteristic of the SARS-CoV-2 infected patient, now termed as Coronavirus Disease 2019 (COVID-19), demonstrated on diagnosis methods like computer tomographic (CT) scan or chest X –ray (Zhong *et. al.,* 2020 and Chan *et. al.,* 2020). The quick development of acute respiratory failure and other serious complications was accompanied by the acute respiratory infection symptoms displayed by the patient in the early stages of the disease (https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected). All of the first three patients reported by the China Novel Coronavirus Investigating and Research Team developed severe pneumonia and two among these three patients showed a common feature of fever and cough (Zhu *et. al.,* 2020). 75% within 99 patients demonstrated bilateral pneumonia and the remaining 25% unilateral pneumonia in the chest X-ray and CT imaging as shown in a study (Chen *et. al.,* 2020). Overall, 14% of the patients showed multiple mottling and ground-glass opacity as presented on the same study. One of the early cases of coronavirus infection in the United States detected pneumonia on the tenth day of infection with basilar streaky opacities in both lungs as shown by chest radiography (Holshue *et. al.,* 2020).

2.4.4.1.3 Modes of transmission

The wide spread of SARS-CoV-2 infection within a short period time was boosted by the fact that SARS-CoV-2 Spike (S) protein had 10- to 20-fold higher affinity to human angiotensin-converting enzyme 2 (ACE2) receptor than that of SARS-CoV based on the Cryo-EM structure analysis of S proteins, implying the highly contagious and more infectious nature of SARS-CoV-2 than initially suspected (Wrapp *et. al.,* 2020). The entry of SARS-CoV-2 into host cells is facilitated upon the recognition and binding of S

protein to ACE2 receptor of the host cells, similar to SARS (Zhou et al, 2020). The high affinity of S protein to ACE2 receptor as shown in studies likely contributes to the quick spreading of virus (Wrapp *et. al.,* 2020). The human organs with high ACE2 expression level, such as lung alveolar epithelial cells and enterocytes of the small intestine, are potentially the target of SARS-CoV-2 as indicated by the fact that ACE2 receptor are the human receptors for the virus (Zou *et. al.,* 2020).

Current knowledge for SARS-CoV-2 transmission is largely based on the knowledge for the transmission regarding the similar coronaviruses, particularly SARS-CoV and MERS-CoV, where droplets, contact and fomites are responsible for human-to-human transmission (Zheng *et. al.,* 2020). Although the well-established route of viral infection so far, is believed to be through respiratory droplets and contact with COVID-19 affected individuals so far (Signorelli *et. al.,* 2020). The possible modes of transmission are discussed below under the following sub-headings.

2.4.4.1.3.1 Saliva

The abundance of Angiotensin converting enzyme 2 (ACE2) in the epithelial cells of the oral mucosa with the fact that high expression of this receptor cells is seen on the tongue (95.86%) than in the buccal and gingival tissues outlines the role of saliva in the possible spread of the SARS-CoV-2 infection (Xu et al, 2020). A viral culture test to assess the presence of viral RNA in the saliva for possible transmission is suggested (To et al, 2020). The role of saliva can be of both friend (diagnostic tool for viral detection) and of foe (a more common mode of transmission via aerosols and salivary droplets) (Han et al, 2020).

2.4.4.1.3.2 Surface contact

Infected surfaces may still help to transmit the virus despite the control of person-to-person transmission of SARS-CoV-2. The indirect contact with the surfaces or objects contaminated by the infected person may cause the spread of this zoonotic virus, as this virus displays the ability of staying for longer durations on various surfaces making it contaminated for hours and days. Also, the large viral droplets expired by the infected person may also get deposited on the surface in their close proximity. The

human coronavirus 229E stays infectious on common touch surface materials like polytetrafluoroethylene (Teflon; PTFE), polyvinylchloride (PVC), ceramic tiles, glass, silicone rubber, and stainless steel and are inefficient on a variety of copper alloys (Warnes et al, 2020). The viability of SARS-CoV-2 on plastic and stainless steel for a maximum of 72 hours, and eventually a decline in the virus stability and titer is displayed in a study (van Dormaelan et al, 2020). Another study shows that on inanimate surfaces, another coronavirus strain (HCoV-229E) is capable to survive for up to nine days (Carraturo *et. al.,* 2020).

2.4.4.1.3.3 Aerosols and Airborne

The possible route of spread of virus maybe aerosols and airborne transmission helped by coughing, sneezing, or even talking (Mukhra *et. al.,* 2020). The droplets and aerosols produced by infected individuals on the regular basis may play an important role in the transmission of the virus (Klompas *et. al.,* 2020) and the smaller size of the droplet nuclei particles of <5µm diameter are considered to be the carriers as these can persist in the air for longer durations and can be transported to distances greater than one meter (Moraswka *et. al.,* 2020). The infectious virus can remain in the aerosols for approximately 3 hours with a gradual reduction from 103.5 to 102.7 TCID50 per litre of air under controlled conditions (Moraswka *et. al.,* 2020). A viral transmission rate of approximately 41.3% through respired air by infected patients in a nosocomial environment suggested an urgent need of consideration of the airborne passage of transmission (Wang *et. al.,* 2020). The possibility of air borne route of transmission of viral particles except the indoor confined environment was also suggested by WHO on a update released on 29th June 2020(https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions). Moreover, the possible transmission of SARS-CoV-2 could be through airborne dust as various studies suggest that microorganisms in the airborne dust particles may be associated to infectious diseases (Yu *et. al.,* 2004).

2.4.4.1.3.4 Faeco-Oral transmission

The gastrointestinal manifestations in COVID-19 patients outlines the probability of shedding viral RNA or infectious virus in the stool and fecal matters of the infected individual (Mukhra *et. al.,* 2020). The fecal-oral route as a probable route of transmission is fueled by the fact where presence of viral RNA in feces, on toilet seat and washbasin sink samples is reported (Young *et. al.,* 2020). The capability of the virus to survive in fecal samples for up to 4 days outlines the possibility of the fecal-oral route of transmission of SARS-CoV-2 (Amirian, 2020).

In a study conducted by Zhu et. al., rectal swab tested positive despite the negative nasopharyngeal swab tests among confirmed COVID-19 patients. 8 out of 10 pediatric patients continued to discharge the infectious viral RNA in feces, even after becoming asymptomatic and recovering from respiratory illness, displaying probable fecal-oral route of transmission in the same study. Likewise, other studies have reported presence of viral RNA in stools even after negative test results of the respiratory samples (Xu *et. al.,* 2020 and Chen *et. al.,* 2020). However, the fecal-oral route of transmission is still debatable due to the lack of strong evidence for possibility of viral replication in fecal swabs.

2.4.4.1.6 Other possible modes of transmission

The sexual route of transmission for SARS-CoV-2 virus (Patri *et. al.,* 2020) surfaced following the reports of fecal-oral transmission (Peng *et. al.,* 2020). Both direct and indirect exposure of oral-anal contact within few days even after recovery, with the shedding the viral RNA still seen on some cases in feces, such sexual behaviors could result in alternative ways of transmission of the virus (Patri *et. al.,* 2020). The vertical transmission of SARS-CoV-2 virus from mother to fetus or neonate is still debatable. The shedding of the RNA particles on the milk of lactating mothers could hint on the maternal to neonate transmission but this needs to be validated through further studies (Grob *et. al.,* 2020).

Conjunctival secretions might be another possible mode of viral transmission as conjunctival mucosa and respiratory tract are connected by the nasolacrimal duct and

share the same ACE-II receptor cell of SARS-CoV-2 (Sun *et al*., 2020). Lu *et al.* have raised the possibility of transmission through ocular surface. More research is suggested to develop a detailed understanding of the transmission mechanism through tear secretions and ocular surfaces.

Another major concern regarding the transmission of COVID-19 is via asymptomatic carriers who are responsible for transmitting the virus during the incubation period (Ye and Hogue, 2020). There have been reports of asymptomatic carriers of SARS-CoV-2, a case in Wuhan, where a 10 years old boy behaved as an asymptomatic carrier of virus, showing lung infiltrates on CT scans (Chan *et. al.,* 2020). These findings underline the role of asymptomatic carriers as one of the probable sources of SARS-CoV-2 infection posing a great threat to increase the infection and further calling out for an immediate evaluation of transmission dynamics of the epidemic.

2.5 Evolution of SARS- CoV-2

The identification of the origin, native host(s) and evolution pathway of the virus that causes an outbreak of a pandemic is very critical. The molecular mechanism of its cross-species spread can be understood and proper control measures to prevent it from further spreading can be implemented by having a good knowledge of the evolution pathways of that pandemic causing virus (Xu *et. al.,* 2020). The seven conserved replicase domains in ORF1ab has 94.6% sequence identity in amino acids between SARS-CoV-2 and SARS-CoV reflects the close similarity between these two viruses despite the fact that some of the six major ORFs of SARS-CoV-2 genes share less than 80% identity in nucleotide acids to SARS-CoV (Zhou *et. al.,* 2020).

Palm civets were considered to transmit SARS-CoV (Guan *et. al.,* 2003) and this was further supported by the study where three of the four patients had the record of contact with palm civets during the two small-scale of SARS outbreaks occurred in late 2003 and early 2004(Wang *et. al.,* 2005 and Song *et. al.,* 2005). However, a deep investigation based on the genome sequence of isolated viruses showed that SARS-CoV-like virus in civet had not been circulating for long (Shi and Hu, 2008) and the isolation of coronaviruses with high similarity to the human SARS-CoV or civet SARS-

CoV-like virus from horseshoe bats, established the bats as the potential natural reservoir of SARS-CoV whereas masked palm civets are the intermediate host (Li e*t. al.,* 2005 and Hu *et. al.,* 2017).

An intermediate host between bats and human might exist despite the evidences pointing the evolution of SARS-CoV-2 from bat virus (Lu *et al,* 2020). Lu *et. al.* raised four major questions for such speculation. First, most bat species in Wuhan were hibernating in late December during the spread of virus. Second, no evidence suggested the selling of finding of bats in Huanan Seafood market. Third, the sequence identity between SARS-CoV-2 and bat-SL-CoVZC45 or bat-SL-CoVZXC21, the closest relatives in their analyses, is lower than 90% and finally there is an intermediate host for other human-infecting coronaviruses that origin from bat. For instance, the intermediate hosts for SARS-CoV and MERS-CoV are masked palm civet and dromedary camels respectively (Alagaili *et. al.,* 2014 and Guan *et. al.,* 2003).

Pangolin was later believed to be a potential intermediate host for SARS-CoV-2. Anti-smuggling operations from Guangdong and Guangxi Customs of China respectively, from which samples of Malytan pangolins were retrieved, upon analysis, yielded novel coronaviruses representing two sub-lineages related to SARS-CoV-2(Lam *et. al.,* 2020). The similarity of these identified coronaviruses from pangolins to SARS-CoV-2 is approximately 85.5% to 92.4% in genomes, lower than that to the bat coronavirus RaTG13 (96.2%) (Zhou *et. al.,* 2020). However, the receptor-binding domain of S protein from one sub-lineage of the pangolin coronaviruses shows 97.4% similarity in amino acid sequences to that of SARS-CoV-2, even higher than that to RaTG13 (89.2%) (Lam *et. al.,* 2020). Interestingly, the pangolin coronavirus and SARS-CoV-2 share identical amino acids at the five critical residues of RBD of S protein, while RaTG13 only possesses one (Lam *et. al.,* 2020). The discovery of coronavirus from pangolin close to SARS-CoV-2 suggests that pangolin is a potential intermediate host and the roles of bat and pangolin as respective natural reservoir and intermediate host respectively still needs further assessment.

2.6 Review of literature related to Virtual Screening

2.6.1 Virtual Screening

High-Throughput Screening (HTS), where the experimental screening of large libraries of chemical scaffolds against selected panels of molecular targets, is the gold standard for discovering biologically active hits against any kind of molecular targets (Sliwoski *et. al.,* 2013). However, their use for drug discovery is limited as the high costs are required to operate and maintain these screening platforms. Moreover, with the advent of recent technological advances and the rapid increase of structural, chemical, and biological data available on an ever-growing number of therapeutic targets, the utilization of in silico approaches now facilitates the virtual screening of millions of compounds in less time cutting the initial costs of hit identification and improving chances of finding the desired drug candidates (Song *et. al.,* 2009, Macalino *et. al.,* 2015).

Virtual screening is the process where computational techniques are utilized to select suitable chemical structures from a library of structures that are potentially best fit to interact with protein receptors or enzymes to aid in the process of drug discovery. Virtual screening has now been an integral part of drug discovery as it saves time and cuts expenses compared to high throughput screening and the accuracy it provides with recent technological advances (Pinzi and Rasteli, 2019). It can be used for selection of in-house database compounds for screening and many other purposes.

2.6.1.1 Virtual Screening Types

The usefulness of virtual screening in drug discovery in recent times has been elevated especially for purposes like lead identification, lead optimization and scaffold hopping. Virtual screening is fast and inexpensive compared to high throughput screening for discovery of new drugs (John *et. al.,* 2010). Basically, there are two categories of computational techniques for virtual screening namely ligand based virtual screening and structure based virtual screening. In the current study, our focus in on the structure based virtual screening where we have capitalized the structural revelations

of corona virus main protease for drug discovery aspects. Ligand based virtual screening also known as ligand similarity methods are based on the fact that ligands with similar physio-chemical structures to active ligands may have enhanced drug like properties against target compared to random ligands, whereas structure based drug design also known as ligand docking methods leverages the use of available protein three- dimensional (3D) structures and preferred ligands to study the binding poses and affinities through docking studies (Hamza *et. al.,* 2012).

2.6.1.1.1 Ligand-Based Virtual Screening (LBVS)

Ligand-based virtual screening methods particularly capitalizes on the information extracted from known active ligands for both lead identification and optimization rather than the three-dimensional structure of a target protein and binding modes and affinities with selected ligands. Ligand-based methods are the ones preferred when no 3D structure of the target protein is available (Hamza *et. al.,* 2012). In drug discovery, when the protein structure for the target of interest are hard to construct, one often knows that established set of ligands are active against the target of interest. In such cases, ligand-based virtual approaches may be the only choice, i.e. finding new ligands by evaluating the similarity between candidate ligands and the known active compounds already showing potent action against the anomalies or diseased states (Tresadern *et. al*., 2009). However, there must be at least one known active compound against the biological target for this method to work. The known active compounds can be gathered in order to select a query for virtual screening or alignment in ligand-based design employing these methods.

The ligand based virtual screening methods must address the factors like an efficient similarities measure and a reliable scoring function (Hamza *et. al.,* 2010). Despite the various limitations of this computational technique, this method is particularly used against G-protein coupled receptors targets (Rai et al, 2010) or the protein structures resolved in the apo-form (Jegerschold *et. al.,* 2008).

2.6.1.1.2 Structure-Based Virtual Screening (SBVS)

Structure-based virtual screening (SBVS), often referred to as target-based virtual screening (TBVS), is concerned with the prediction of the best fit between ligands against a molecular target forming a complex. Following this, the ligands are ranked accordingly on the basis of their affinity to the target, and the most promising compounds generally lie, at the top of the list. The 3D structure of the target protein is a pre-requisite for the SBVS so that the interactions between the target and each chemical compound can be predicted *in-silico* (Liu *et. al.,* 2018). In this strategy, the compounds are selected from different databases available and sorted according to their interaction with the binding site.

Molecular docking is one of the reliable among the various available techniques of SBVS relating to its low computational cost and better results offered (Meng *et. al.,* 2011). Despite the development of this technique in 1980s (Kuntz *et. al.,* 1982), the technique was widely used in 1990s owing to the advances in computational power and accessibility of the structural data of the target proteins. A reliable scoring function, which can best explain the affinity of ligands and molecular targets, is the critical component of docking procedure (Leelanada *et. al.,* 2016). SBVS offers many advantages such as reduction of time and funds, testing a synthetic compound even before its synthesis in laboratory and so on (Maia *et. al.,* 2020). The difficulty of parameterization of the complexity of ligand-receptor binding interactions and the inability of accurate prediction of correct binding poses and classification of compounds marks the drawbacks of SBVS (Lionta et al, 2014). Nonetheless, many studies using SBVS has been employed in the recent years (Carpenter *et. al.,* 2018; Dutkiewicz and Mickstacka, 2018 and Nunes *et. al.,* 2019) drug discovery research.

2.6.1.2 Steps of Structure Based Virtual Screening

Obtaining the 3D target structural information of interest is the first step of SBVS. Various sources such as experimental data (X-ray, NMR or neutron scattering spectroscopy), homology modeling, or from Molecular Dynamics (MD) simulations may supply with the required structure of the target protein in three-dimensional

form. Numerous fundamental issues that should be given priority when considering a biological target for SBVS; for instance, the druggability of the receptor, the choice of binding site, the selection of the most relevant protein structure amongst the available structures (), incorporating receptor flexibility, suitable assignment of protonation states, and consideration of water molecules in a binding site, to name a few (Reddy *et.al.,* 2007). Another important factor is the accurate prediction of ligand binding sites on biological targets. The wise use of the set of compounds in a library to be screened in the VS has to be planned according to the target protein in question, and the preprocessing steps in those libraries must meet the criteria to assign the proper stereochemistry, tautomeric, and protonation states.

After the successful preparation of ligand library and receptor preparation, the individual compound in the library is subjected to the docking in the target binding site using a certain docking program. Docking provides the information regarding the ligand-protein complex structure. The free energy of binding between the ligand and protein in consideration can be now calculated using scoring function. Docking and scoring generates the ranked probable compounds followed by post processing that can be allowed for further experimental assays (Koppen, 2009).

2.6.1.2.1 Protein preparation

The reasonable starting structures for both the protein and the ligand determine the outcome of SBVS procedure. There is presence of several heavy atoms, water molecules, cofactors, activators, ligands, and metal ions as well as several protein subunits in a PDB file. Moreover, the structure may lack proper information regarding the bond orders, topologies, or formal atomic charges. There may be inappropriate arrangement of terminal amide groups and asparagine residues because the X-ray structures cannot unambiguously distinguish between O and NH_2 groups. Similarly, in most cases Ionization and tautomeric states lack proper assignment. There may be possibility of missing residue side chains or larger loops due to low resolution of a particular protein area, with a possibility of steric clashes (Lionta *et. al.,* 2014).

To address the above-mentioned structural problems, several protein preparation schemes have been proposed. Available software like PROPKA (Li *et. al.,* 2005), H++ (Anandakrishnan *et. al.,* 2012), SPORES (ten Brink and Exner, 2010) can be used to calculate the protonation states of protein in the respective structures. Similarly, for assigning hydrogen atoms and optimizing protein hydrogen bonds according to an optimal hydrogen bond network, PDB2PQR (Dolinsky *et. al.,* 2007) software is used. The next steps address the assignment of partial charges, capping of residues, treating metals, filling in missing loops and missing side chains, and minimizing the protein structure to relieve steric clashes (Young *et. al.,* 2007). In addition, a decision needs to be made regarding whether water molecules will be left in or removed from the binding site. Several methods have been developed to tackle this challenging problem (Michel *et. al.,* 2009). With the role of protein preparation in the docking performance (Sastry *et. al.,* 2013), it is wiser to employ best preparation schemes for SBVS rather than using minimally prepared PDB structures.

2.6.1.2.2 Binding site Identification

The accurate prediction of binding site is essential prerequisite for successful SBVS scheme. There may be different approaches for the prediction of binding sites. The static approaches involve computational solvent mapping with chemical probes to map the binding hotspots on the 3D protein target (from X-ray, MD). Some common examples include FTMap (Ngan *et. al.,* 2012), Fpocket (Le Guilloux *et. al.,* 2009) and so on. Another approach involves the evolution of probe and protein dynamically with time which may include more than one probe. The generated interaction free energies can then be used to find out the appropriate binding sites. MDMix (Seco *et. al.,* 2009) is a program following this approach. The third method for identification of binding sites is using water molecules as probes to predict putative binding sites in proteins as used in JAWS (Michel *et. al.,* 2009).

2.6.1.2.3 Compound Library Preparation

Small drug-like molecules, often freely available or available through purchase or synthesis, constitutes the compound database and the preparation of this library is another important part of SBVS process. The desirable properties such as stability and solubility in aqueous media, existence of appropriate functional groups to interact with biological protein targets through different chemical interactions and absence of toxicity and undesirable moieties is a prerequisite for the small drug-like molecules. 'Drug-likeness' is often associated with many rules, with the most popular being the "Lipinski Rule of Five" (Lipinski *et. al.,* 2001), where molecular weight lower than 500, lipophilicity (logP) lower than 5, less than five hydrogen bond donors, and less than 10 hydrogen bond acceptors for drug-like molecules. Many natural product drugs as well as 50% of marketed drugs do not follow the "Rule of Five" (Zhang *et. al.,* 2007). Subsequently, the efforts to enhance the predictions of druglikeness have spawned many extensions and considerations to the Lipinski Rule of Five.

A drug-like molecule may lie within a range of −0.4 to +5.6 for logP, 40 to 130 for molar refractivity, 180 to 500 for molecular weight, number of atoms from 20 to 70 (includes H-bond donors (e.g., OH's and NH's) and H-bond acceptors (e.g.; N's and O's), polar surface area less than 140 $Å^2$, and/or fewer than ten rotatable bonds (Veber *et. al.,* 2002). A new approach based on the values of calculated partition coefficient (ClogP) and topological polar surface area (TPSA) called "Pfizer's Rule of 3/75", strongly supports the fact that compounds with ClogP lower than 3 and TPSA higher than 75 are approximately 2.5 times more likely to be safe in in vivo assays (Hughes *et. al.,* 2008). The aqueous solubility measured as logS should be greater than -5.7, the apparent Caco-2 cell permeability should be faster than 22 nm/s, and the number of primary metabolites should be less than 7; constitute "Jorgensen Rule-of-Three", as another parameter for assessing drug- like properties of small molecules (Kerns and Di, 2008).

Selection of compounds with lead-like properties, exclusion of known toxicophores or metabolically liable moieties and avoiding the random docking (Blind docking) of huge compound collections such as PubChem (30M compounds) and ChemSpider (26M

compounds) which is both time consuming and computationally demanding, is must in compound selection step. There are several online tools for efficient filtration of compounds from a large set of library against criteria such as adsorption, distribution, metabolism, excretion, and toxicological (ADMET) properties. Chembioserver (Athanasiadis *et. al.,* 2012), a publicly available online application, facilitates compound preparation prior to (or after) VS computations by utilizing its many functions, such as basic search, filtration (steric clashes and toxicity), advanced filtering based on custom chosen physicochemical properties clustering (according to structure and compound physicochemical properties providing representative compounds for each cluster), customized pipeline and visualization of compound' properties through property graphs and hence, increase the efficiency and the quality of compounds that proceed to in vitro assays.

The realistic bond lengths and angles, in realistic 3D representations should be a priority for SBVS. Furthermore, the absence of accompanying fragments such as counter-ions, metals and solvents is must in probable drug like molecule, along with the assignment of correct bond order and filled valences, partial charges (if present), an appropriate protonation state at physiological pH or at the pH of the interest and proper designation of tautomeric states (Kalliokosky et. al., 2009). Docking programs usually include ligand preparation software such as Autodock Tools (Morris et. al., 2009).

2.6.1.2.4 Library Design

A custom-made library instead of freely available online libraries (Lavecchia and Di Giovanni, 2013) may be fruitful for effective SBVS procedure. CLEVER (Song *et. al.,* 2008), online tool for supporting the manipulation of chemical library, enumeration of combinatorial chemical library by user-specified chemical components, chemical format conversion, as well as chemical compound analysis and filtration with respect to drug-likeness, lead-likeness, and fragment-likeness based on the physicochemical properties computed from the derived molecules, may have usefulness in required library design. Another tool for library design is e-LEA3D, performing design based on

a user-generated scaffold and reactants coming, for instance, from a chemical vendor (Douguet, 2010).

2.6.1.2.5 Docking and scoring functions

Docking programs currently in use are AutoDock (Morrris *et. al.,* 2009), Dock (Ewing *et. al.,* 2001), FlexX (Rarey *et. al.,* 1996), Glide (Friesner *et. al.,* 2004) and so on (Cheng *et. al.,* 2012). The predictive study in relation to the interaction of protein and ligand is accompanied by the scoring in SBVS to prepare a rank of compounds after molecular docking. The identification of ligand conformational space as a method of conformational search employed by docking programs includes systematic methods, random or stochastic torsional searches about rotatable bonds and molecular dynamics simulation methods and energy minimization for finding energy profiles of a molecule (Guido *et. al.,* 2008).

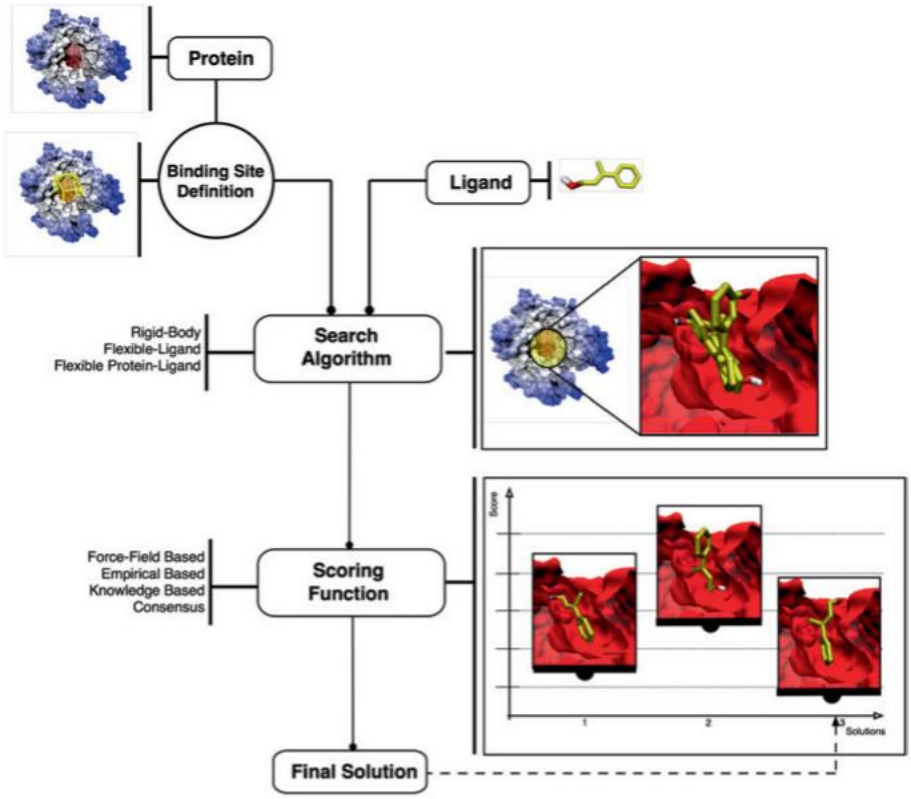


Fig 6: Schematic diagram of molecular docking protocols (Brás et al. 2014)

Scoring functions are utilized with a view to estimate the free energy of association of a ligand to a specific biological target based on a generated docked pose after docking different ligands of a database that ultimately brings about the rank of compounds (Cheng *et. al.,* 2012). The sum of the strength of different chemical interactions like intermolecular van der Waals, electrostatic interactions, and hydrogen bonding between all atoms of the two binding partners in the ligand-protein complex along with the contributions from solvation and entropy parameters can be used to calculate the binding energy in force-field based functions. Furthermore, ChemScore and Fresno are a few examples of functions contained in empirical scoring functions which are based on enumerating the number of various types of interactions between the two binding partners, i.e., hydrophobic contacts, number of hydrogen bonds and number of rotatable bonds immobilized in complex formation (Lionta *et. al.,* 2014). HYDE (Schieder *et. al.,* 2012) is another function addressing the hydrogen bonds, hydrophobic effects and desolvation parameters. The development of accurate scoring function determines the efficiency of SBVS scheme. The incorporation of different parameters like protein flexibility as done by Spitzer *et. al,* may be a way to improve the docking and scoring functions.

2.6.1.2.6 Post Processing

Post-processing of probable small drug-like molecules after performing VS/docking exercise before selection of the ones that will give more satisfactory results in the experimental test phase, as a rate-limiting step in SBVS, is often a challenge to a computational chemist. The issues like unrealistic poses, steric clashes, twisted amides, imperfect hydrogen-bonding network, and poses based on shape complementarity are an outcome of using simplified scoring functions and inappropriate sampling of the conformational space for the ligand (Waszkowycz, 2008). As a result, unreasonably high score due to the poses leads to reduce the efficiency of SBVS. Hence, correct visual inspection of all docking poses is normally required by the medicinal chemist so as to select the appropriate compound set for further down streaming processes. Appreciable efforts have been developed to raise the efficiency and the quality of compound selection (Athanasiadis *et. al.,* 2012).

A directional approach, known as, MIECs (molecular interaction energy components) accurately predicts the binding affinities of ligand-protein complexes and differentiates the binding and non-binding ligands. (Ding *et. al.*, 2013). Malmstrom et al. illustrated that free energy of binding (FEB) rescoring may assist proper lead identification resulting from a SBVS exercise as it employs a set of predicted binders and non-binders employing the concept of free energy of binding (FEB) rescoring and other empirical docking scores. Alternatively, Vdw filtering from online tools like ChemBioServer (Athanasiadis *et. al.,* 2012) may help discard the compounds with steric clashes and poses far from energy minimum followed by heirarchial clustering to identify compounds with similar physicochemical properties (Lionta *et. al.*, 2014).

2.6.1.2.7 Advancements in SBVS

The limited information on protein dynamics, provided by a single conformation of the protein, which is again dependent upon the crystallization conditions, interferes with efficiency of SBVS as the selection of ideal crystal structure of the receptor target will be in scrutiny. Furthermore, conformational changes arising in both the protein and the ligand, after the ligand binding also affects the efficiency of SBVS. Hence, although some crystal structures which are preferable at starting point for SBDD, in some cases they provide misleading information after the induced fit effects of proteins upon binding of a ligand. Hence the requirement to include receptor flexibility in docking programs is necessary, as it provides a more realistic picture of the modeled system (Korb *et. al.,* 2012). Despite focusing on one protein conformation and taking into account the concept of protein flexibility, one may use an ensemble or group of representative structures to dock candidate ligands. Ensemble Docking is based on docking a single ligand library against multiple rigid receptor conformations, on contrary to the standard single rigid receptor docking methods and helps pinpoint the realistic structural variation of the protein-target (Craig *et. al.,* 2010).

Despite the inclusion of accounting for any type of protein motion, the higher demand for computational power due to generation of higher number of protein conformations in the ensemble and inaccurate outputs from not so reliable scoring

functions leading to false prediction of ligand-protein interaction may be downside of this method (Osguthorpe *et. al.,* 2012).

Another alternative docking protocol is Consensus induced fit docking (cFID), which entails the protein binding site to adapt to different ligands during SBVS as practiced by Kadid *et al.,* on COX-2. It can also be used for screening the improvements in terms of correct identification of accurately docked poses as this method focuses on information about predicted binding modes rather than predicted binding energies (Houston and Walkinshaw, 2013). The success rate for pose prediction in consensus induced fit docking is higher as compared to other individual docking program alone as stated by Houston *et. al*. 2013.

2.7 Review of literature related to Mutation

There is observation of frequent and significant mutations in viruses which can be largely accounted to three distinct phenomena namely large populations, their short generation times, and high mutation rates. The positive selection of every mutation aids in the evasion of the immune system of the host, and is passed to the next generation for wide distribution. The absence of proofreading function of viral RNA-polymerase also results in high mutation rates**.** (Shao *et al.,* 2017). Successive cell divisions help in the accumulation of replication errors. The mutation rate is greater than zero in nature indicating, certain few base pairs are always more likely to get mutated than the initial genome in each cell. (Strippoli *et al.,* 2005). A change in single base pair of DNA is called point mutation. The evolution of viruses and ultimate production of new viral subtypes depends upon the mutations in the antigenic sites because of individual point mutations (Carrat and Flahault, 2007).

Such point mutations can cause mutations in DNA even in the absence of the external factors, ie, the spontaneous mutations. Quantum tunneling, is a phenomenon that can account to the cause of these point mutations. In accordance with the wave –particle dual nature of particular particle, in the sub atomic level, it is believed by the quantum biologists that the electrons behave like wave like particles such that electrons tunnel

through an obstacle rather than climbing over it (https://the-gist.org/2017/02/schrodingers-mutations/).

More specifically, proton tunneling, a quantum phenomenon, has been identified responsible for the spontaneous mutation inside the genome by various simulation methods and quantum approaches. The proton disappeared from one position and reappeared nearby spontaneously. The hydrogen bonds between two strands of DNA helix, under certain conditions behave like waves and spread to many locations at once by proton tunneling phenomenon. This fact largely accounts to the spontaneous mutation due to occasional finding of these atoms on wrong strand of DNA (https://www.surrey.ac.uk/news/new-study-reveals-quantum-physics-can-cause-mutations-our-dna).

Spontaneous mutations are an outcome of quantum events. The basic knowledge of various quantum jump events can predict the subsequent errors during replication. But considering Heisenberg uncertainty principle, which is similar to the genetic information uncertainty principle, the prediction of errors with a view to predict the mutations, more likely to occur in future spontaneously cannot be done with utmost certainty (Mcfadden & Al-Khalili, 1999; Strippoli *et. al.,* 2005).

Substitution rates of transitions are higher in comparison to transversions than expected by chance. The hypotheses supporting the above-mentioned statement are mutational hypothesis and selective hypothesis. The mutational hypothesis states that the transition mutation rates of polymerases are higher in compared to transversion rates. Another hypothesis, the selective hypothesis tells that the natural selection disfavors transversion, favoring transitions (Stoltzfus & Norris, 2016).

2.8 Review of literature related to Density Function Theory

2.8.1 Density functional theory

The density functional theory seems to be the least popular computational method applied to complement virtual screening. This may indicate that this technique is

highly specific to limited occasions whereby certain intricate inter-molecular interactions and reaction mechanisms need to be established (Gross and Dreizler, 2013). The main challenge with this method is that the size of enzyme systems currently excludes calculations of the whole system in reasonable timeframes (Gross *et.al.,* 2013). Hence, smaller model systems are studied or hybrid methods (QM/MM (Senn *et. al.,* 2009) or ONIOM (Verven *et. al.,* 2006)) are preferred. The results from such calculations are typically more closely comparable to the experimental values and the limited frequency of their application does not diminish its importance in CADD. The emphasis of this technique leans on accuracy at the expense of computational cost and efficiency, and this may explain the lack of appreciation when compared to other CADD techniques. This could change in the future as QM/MM (Senn *et. al.,* 2009) and ONIOM (Verven *et. al.,* 2006) methods become more popular and as improvements in hardware and software happen.

2.9 Review of literature related to Molecular dynamics (MD) simulations

2.9.1 Molecular dynamics simulations

Molecular dynamics simulations are some of the most integral calculations that follow virtual screening simulations. They should be therefore considered as an advanced technique, complementary to docking. They can also be applied prior to docking for conformational sampling and clustering on a protein molecule to cater for the conformational dynamics relating to ligand binding (Menchon *et. al.,* 2018). More prominently, MD simulations are employed to filter and validate the results obtained from protein-ligand complex docking runs by establishing the stability of the complex, interatomic interactions, the rate of system fluctuation, and binding free energies (Menchon *et. al.,* 2018). However, molecular dynamics simulations are not performed in all virtual screening studies, thus their application depends on the study design and aspects under investigation. The prominently used softwares are GROMACS (http://www.gromacs.org**/**), Amber (http://ambermd.org/) and so on for performing molecular dynamics simulations.

2.10 Review of literature related to human Cytochrome P450 enzymes

Cytochrome P450 enzymes function to catalyze a wide range of reactions, many of which are critically important for drug response (Guengerich, 2021). Human cytochrome P450 enzymes catalyze a great number of metabolic reactions that have important effects on the biological activities (physiologic, therapeutic, and/or toxic) of xenobiotics such as drugs, natural products, general chemicals (e.g., environmental chemicals such as pesticides, pro-carcinogens), and physiological compounds. Many isoforms of cytochrome P450 exist but most reactions are undertaken by CYP2C9, CYP2C19, CYP2D6 and CYP3A4.

2.10.1 Cytochrome P450 3A sub family

CYP3A is one of the most important human enzymes as approximately 60% of oxidized drugs are bio-transformed, at least in part, by its involvement. The human locus for the CYP3A gene cluster is located on chromosome 7q21-q22. It comprises two pseudogenes (CYP3AP1 and CYP3AP2) and four genes (CYP3A4, CYP3A5, CYP3A7, and CYP3A43). Each of the four functional genes at the CYP3A locus comprises 13 exons. The CYP3A4 gene is located upstream from CYP3A43, which is furthest from the centromere and is the only gene in the cluster that is in the opposite orientation (Finta and Zaphiropoulos, 2000). A pseudogene (CYP3AP2) separates CYP3A4 from CYP3A7, and another pseudogene (CYP3AP1) separates CYP3A7 from CYP3A5.The vast number of SNPs that affect each gene, especially CYP3A4 and CYP3A5, are found in all parts of the gene structure, including 5′ upstream regions, introns, exons, and the 3′ untranslated regions (UTRs) and not in localized positions (Lee *et. al.,* 2004).

CYP3A4, responsible for the bulk of metabolic activities ascribed to the CYP3A subfamily, is the most studied isoform in the CYP3A subfamily and is predominantly found in the small intestine and adult liver.

2.10.2 Relevance of CYP3A4 to drug discovery

2.10.2.1 Adverse drug-drug interaction

Administration of cocktail of drugs is common way to tackle disease these days. The proper evaluation of not only the metabolic profile of a drug candidate but also its potential interactions with clinically used other drugs is regarded as the critical part of drug discovery and development. One example of adverse drug–drug interaction (DDI) results in the impaired activation, metabolism, or elimination of another drug, resulting in increased systemic exposure of the affected drug and a higher risk of toxicity followed by low elimination rate. Terfenadine and CYP3A4 inhibitors showed drug-drug interaction, which led to the withdrawal of the former from the market (Honig *et. al.,* 1993). Terfenadine is a prodrug with antihistamine activity that can cause cardiotoxicity where the parent drug is biotransformed by CYP3A4 to a nontoxic and pharmacologically active metabolite in the liver. However, patients who were taking CYP3A4 inhibitors (e.g. ketoconazole) concurrent with terfenadine reportedly experienced cardiac arrhythmias.

The enhanced clearance of therapeutic agents, which decreases the bioavailability and efficacy of the drug(s) and necessitates dose adjustment while monitoring toxicity levels, is another adverse effect of DDIs. The co-administration of temsirolimus with dexamethasone in patients having advanced renal cell carcinoma results in acute reduction of plasma concentration of temsirolimus (Staeheler *et. al.,* 2010). This is a suitable illustration of such a DDI. The decrease in the systemic concentration of temsirolimus can be accounted to the strong CYP3A4 inducing capability of dexamethosone.

3. MATERIALS AND METHODOLOGY

3.1 Search for viral protein targets and Development of leads

3.1.1 Selection of viral protein and obtaining their genomic sequences

Among the important viral proteins involved in the completion of viral life cycle of SARS-CoV-2 virus, main protease against whom new leads are sought was taken as the target protein for further experimentation. The whole genome sequences of this protein of SARS-CoV-2, which are published in NCBI were taken for carrying out alignment and structural analysis with closely related SARS and MERS viruses.

3.1.2 Alignment studies for SARS-CoV-2 Main protease with other corona viruses

Different alignment studies based on sequence comparison were performed for the sequence of SARS-CoV-2 main proteases deposited in the NCBI database to explore the close association and relation to other corona viruses like SARS and MERS.

3.1.2.1 Alignment search using BLAST tool for SARS-CoV-2 main protease genome

The sequence for SARS-CoV-2 Mpro was searched for in NCBI database and BLAST tool in the same database was employed for alignment studies with the intention to find the matches in sequences to SARS and MERS.

3.1.2.2 Multiple Sequence Alignment search CLUSTAL OMEGA for SARS-CoV-2 main protease genome

The protein sequences of SARS-CoV-2 main protease, SARS main protease and MERS main protease were downloaded in FASTA sequences, and the Clustal Omega tool was used for performing the multiple sequence analysis and construction of phylogenetic tree amongst the three strains in consideration.

3.1.2.3 Structural Alignment analysis for SARS-CoV-2 main protease genome

The protein databank (pdb) structures were downloaded for the SARS-CoV-2 main protease, SARS main protease and MERS main protease and these structures were structurally aligned using Chimera software version 1.14. employing sequence and align function inbuilt in the software.

3.2 Selection of dockable cystal structures of viral protein target and its validation

The dockable crystal structure was selected for SARS CoV 2 M^{pro} and the Z-score analysis and Ramachandran plot analysis was done in proSA web server and SAVES V 5.0 web server respectively.

3.3 Molecular docking simulation

3.3.1 Obtaining the dockable crystal structures of the target protein

Among the myriad of viral proteins essential to SARS-CoV 2, main protease was selected as target protein. The X-ray diffraction structure of SARS-CoV-2 main protease with bound Michael acceptor inhibitor (N3) (PDB id: 6LU7) (Jin *et. al., 2020*) was obtained from Protein Data Bank (rcsb.docking.org) in PDB format.

3.3.2 Preparation of ligand database

In the present work, structure based virtual screening was performed. Structure based Virtual Screening was done with a primary aim to avoid partiality among the class of molecules and a broad chemical database with structural diversity may include diverse class of molecules that may be sufficient to develop potential leads against the chosen viral target. In the context of global health crisis, opting to a library of compounds which is already suitable for use in human with limited side effect profiles would save both the time and efforts to address this urgency. In this study, a ligand database

containing xxx molecules was prepared that included FDA approved compounds from ZINC database (Irwin and Shoichet, 2005).

3.3.3 Protein and ligand preparation

The dockable protein structures and ligands need certain preparation before docking simulations. Prior to molecular docking, the proteins and ligands need to be prepared for efficient and more accurate docking results using different tools. Protein preparation is done by adding hydrogen atoms, merging non-polar bonds, adding Gasteiger charges and finally converted to pdbqt format that is required for performing docking studies in mgltools (http://mgltools.scripps.edu/). Similarly, ligand preparation was done in Openbabel GUI (O'Boyle *et. al.*, 2011) available in PyRx 0.9.8 setup by energy minimization using forcefield and converted to pdbqt file format, a useable file format for docking afterwards.

3.3.4 Setting reference values for docking

The native ligand (N3) was removed from the binding pocket of the protein databank structure(6LU7) in PyMol and docked back in their binding sites, a process called re-docking. The highest binding energy thus calculated from the docking studies was taken as a reference value for selecting probable leads from the chemical database, docked against the same active binding sites of the same protein databank structure of main protease.

3.3.5 Structure based Virtual Screening

AutoDock Vina was used to perform virtual screening in a platform, PyRx (Dallakyan and Olson, 2015) against the target protein (main protease)with the selected ligand database. The conformation with the lowest docked energy was chosen after the docking interactions and visualized and analyzed using Discovery Studio.

3.3.6 Preparation of ligand database from top hits obtained from docking studies for final docking

Following the docking operations of the chosen ligand dataset to the target protein in consideration, the top hits were identified above the reference value i.e., the value of reference inhibitor Michael acceptor inhibitor N3 in this case.

3.3.6.1 Obtaining the 3D SDF of individual hits and merging them to prepare final ligand dataset

The obtained lead compounds after docking studies were searched in the PubChem database for 3D sdf structures and these compounds were downloaded and renamed and merged using the Osiris datawarrior tool to prepare a final dataset of ligands.

3.3.7 ADME/TOX filter

The process of drug discovery and development are very expensive processes and take a lengthy span of time. Many promising lead compounds initially regarded better, fail to reach the clinical stage of drug development mainly in relation to unacceptable pharmacokinetic properties as well as toxicity hazards. The toxic profiles and drug likeness of potential lead compounds obtained from docking studies were predicted using OSIRIS program (Nisha *et. al.*, 2016). OSIRIS calculates various drug relevant properties like molecular weight, cLogP, cLogS, Druglikeness, total polar surface , toxicities like mutagenicity, tumorigenicity, reproductive effects and irritant effects and rotatable bond counts in the lead molecules on the basis of functional groups and chemical bonds present in their structures (Sander *et. al.*, 2015).

3.3.8 Final Structure based Virtual screening

AutoDock Vina was used to perform virtual screening in a platform, PyRx (Dallakyan and Olson, 2015) against the target protein (main protease)with the selected ligand database which contains 3D structures from PubChem database. The hits obtained

yielding the higher binding energy than the reference inhibitors were further considered and analyzed for different properties.

3.4 Analysis of docking results

3.4.1 Binding interaction of ligands with protein target

The docking of a lead compound to the target protein involves certain chemical interactions which impart to the stability of the lead molecule in that association. The amino acids residues involved in making possible interactions with the protein structure were looked for using the pymol software version 4.6.0.

3.4.2 Bond length and bond types responsible for drug ligand interaction

The interaction of chemical lead with its target protein always depends upon the different kinds of feasible chemical interactions. A ligand is bound to its target protein with different bond length and different types of interactions which were analyzed by the using Discovery Studio Visualizer.

3.4.3 Hydrophobic bond interactions

Different types of bonding interaction play a crucial role in establishing the stable association of a ligand to the binding pocket of the enzyme. Among them, one of the essential bond types is hydrophobic interaction which can be studied by using Ligplot software version.

3.5 Mutagenesis studies

3.5.1 Preparation of probable mutated viral proteins

The essential amino acid residues in viral genome responsible for the binding interactions were sought for and these amino acid residues were mutated to feasible

residues. The codons coding those amino acids were noted and possible feasible mutations were carried out using Pymol software version 4.6.0.

3.5.2 Docking simulations and analysis involving the mutated protein

3.5.2.1 Structure based virtual screening

The different mutated protein structures were subjected to docking operation in the PyRx 0.9.8 setup against the top hit compounds.

3.5.2.2 Change in binding interaction of ligands with mutated protein target

The docking of a lead compound to the mutated target protein involves certain chemical interactions which impart to the stability of the lead molecule in that association. The amino acids residues involved in making possible interactions with the mutated protein and change in binding energy in respect to its native form was noted.

3.5.2.3 Bond length and bond types responsible for drug ligand interaction amongst the mutated proteins

The interaction of chemical lead with its target protein always depends upon the different kinds of feasible chemical interactions. A ligand may be bound to its mutated protein with different bond length and different types of interactions which were analyzed by the using Discovery Studio Visualizer.

3.5.2.4 Hydrophobic bond interactions amongst the mutated strains

Different types of bonding interaction play a crucial role in establishing the stable association of a ligand to the binding pocket of the enzyme. Among them, one of the essential bond types is hydrophobic interaction which can be studied by using Ligplot software version for the mutated strains.

3.6 DFT calculations

The selected structures of probable drugs were transferred to the Gaussian 03W (Vittinhoff et al, 2007) software for optimization with the B3LYP/6–31 G method based on density-functional theory (DFT), to extract a set of a quantum chemistry descriptors, namely the energies of the highest occupied molecular orbit HOMO and the lowest vacant LUMO, the total energy of the molecule (Et) and the dipole moment.

3.7 Drug-drug interaction studies

3.7.1 Molecular docking simulation

3.7.1.1 Obtaining the dockable crystal structures of the target human protein involved in drug clearance

Among the different classes of cytochromes involved in drug clearance of top hit compound, the X-ray diffraction structure of Cytochrome P450 3A4 with bound inhibitor (PDB id: 3UA1)(Sevriouka *et. al., 2012*) and cytochrome P450 2D6 with bound inhibitor (PDB id: 3QM4) (Wang *et .al., 2012*) was obtained from Protein Data Bank (rcsb.docking.org) in PDB format.

3.7.1.2 Preparation of ligand database

There are different ways to predict which type of compound will interact with a given protein structure. It is essential to look up for the possible human proteins with which a developing drug is capable of interaction and possibly harmful ones. The smiles code of top hit compound was used to find out possible interacting human proteins using the Swiss prediction online server. The same server also gives the same class of molecules capable of interaction with that same human protein as our top hit compound. A total of 385 compounds were selected. Apart from that, the initial library of FDA compounds was also taken for further experimentations.

3.7.1.3 Obtaining the 3D SDF of compounds and merging them to prepare final ligand dataset

The obtained lead compounds from Swiss prediction server were searched in the PubChem database for 3D sdf structures and these compounds were downloaded and renamed and merged using the Osiris datawarrior tool to prepare a final dataset of ligands.

3.7.1.4 ADME/TOX filter

OSIRIS datawarrior was used to filter the compounds by calculating various drug relevant properties like molecular weight, cLogP, cLogS, Druglikeness, total polar surface , toxicities like mutagenicity, tumorigenicity, reproductive effects and irritant effects and rotatable bond counts in the lead molecules on the basis of functional groups and chemical bonds present in their structures (Sander *et. al.*, 2015).A library of xx ligands was chosen for further docking studies.

3.7.2 Protein and ligand preparation

The dockable protein structures and ligands need certain preparation before docking simulations. Protein preparation is done by adding hydrogen atoms, merging non-polar bonds, adding Gasteiger charges to protein structures and finally converted to pdbqt format that is required for performing docking studies in mgltools (http://mgltools.scripps.edu/). Similarly, ligand preparation was done in Openbabel GUI (O'Boyle *et. al.*, 2011) available in PyRx 0.9.8 setup by energy minimization using energy forcefield and converted to pdbqt file format, a useable file format for docking afterwards.

3.7.3 Setting reference values for docking

The native ligand (N3) was removed from the binding pocket of the protein databank structure(6LU7) in PyMol and docked back in their binding sites, a process called re-docking. The highest binding energy thus calculated from the docking studies was taken as a reference value for selecting probable leads from the chemical database,

docked against the same active binding sites of the same protein databank structure of main protease.

3.7.4 Structure based Virtual Screening

AutoDock Vina was used to perform virtual screening in a platform, PyRx (Dallakyan and Olson, 2015) against the target protein (main protease)with the selected ligand database. The conformation with the lowest docked energy was chosen after the docking interactions and these structures were capable of interfering with the drug clearance mechanism of top hits.

4. RESULTS AND DISCUSSION

4.1 Search for viral protein targets

The global health pandemic and urgency situation caused by the rapid outbreak of SARS-CoV 2 virus throughout every nook and corner of the world has led to the need for rapid development of potential leads against this infectious disease. Amongst the myriads of viral protein encoded in the genome of the SARS-CoV 2 like envelop protein (E protein), membrane protein (M protein), Spike protein (S protein), RNA dependent RNA polymerase (RdRp) and several non-structural proteins(nsp) which are responsible for completing the viral life cycle. Recent research has been focused on developing leads for the SARS-CoV 2 targeting different proteins and their functions. The transmembrane S glycoprotein of SARS-CoV 2 directly interacts with the human angiotensin converting enzyme 2 (ACE2) and this interaction is found to be crucial in the entry of viral pathogens to the human host. Despite the indispensability of S protein in viral entrance to host cell and the fact that targeting S-protein could be of huge potential (Feng *et. al.,* 2020), a complete understanding of corona virus pathogenesis could only be achieved after obtaining detailed insights into the receptor recognition and membrane fusion (Gui Song *et. al.,* 2017) which is still not clearly described.

The nucleocapsid protein is one of the major structural proteins with its role focused on production of ribonucleoprotein complex upon the binding the viral RNA genome. It has an essential role in the regulation of viral RNA synthesis and has functional importance in fundamental aspects of the CoV life cycle, such as encapsidation and replication of virus genomes (Schelle *et. al.,* 2005). Moreover, its ability to regulate the cellular processes during viral pathogenesis, including actin reorganization, host cell cycle progression and apoptosis (Du *et. al.,* 2008; Surjit *et. al.,* 2006) emphasizes its importance to the viral pathogenesis. Besides these roles, there is less variability in the viral N-gene sequence, and thus, it is a genetically stable protein, an essential prerequisite of an effective drug target candidate (Chang *et. al.,* 2014). Although,

there are many facets to be considered for the drug development against this protein and is arduous task considering all the factors.

The M and E proteins are involved actively in virus morphogenesis and assembly and these proteins are responsible for making a protective cover outside the viral genome (Saxena *et. al.,* 2020). Despite the promising results of in vitro studies including si-RNA targeted mechanisms to inhibit M and E proteins, the optimal delivery methods to human are of concern lately (Saxena *et. al.,* 2020).

One of the best characterized drug targets among coronaviruses is the main protease (Mpro, also called 3CLpro) (Anand *et. al.,* 2003) and along with the papain-like protease(s), this enzyme acts to process the polyproteins that are translated from the viral RNA (Hilgenfeld *et. al.,* 2014). The Mpro operates at no less than 11 cleavage sites on the large polyprotein 1ab (replicase 1ab, ~790 kDa); the recognition sequence at most sites is LeuGln↓(Ser,Ala,Gly) (↓ marks the cleavage site). Inhibiting the activity of this enzyme would block viral replication and no human proteases with a similar cleavage specificity are known, inhibitors are unlikely to be toxic (Zhang and Wilkinson, 2020). Regarding all these facts and consideration, the main protease was further chosen as a drug target.

4.1.1 Selection of viral protein and obtaining their genomic sequences

The main protease, also known as non-structural protein 5 (nsp 5) or chymotrypsin-like protease (3CL-pro) was chosen as a protein target and its genome sequence was obtained from the NCBI database along with its protein sequences as well. First, the protein sequence was retrieved, and its genome was then obtained from that protein sequence using online server. These protein sequences provide the basic idea about the preferred codons, and it was seen that Mpro consisted of 306 amino acid sequences on it.

4.1.2 Alignment studies for SARS-CoV-2 Main protease

Different alignment studies based on sequence comparison using different algorithms and parameters were performed for the sequence of SARS-CoV-2 main proteases deposited in the NCBI database to explore the close association and relation to other corona viruses like SARS Mpro and MERS Mpro proteins.

4.1.2.1 Alignment search using BLAST tool for SARS-CoV-2 main protease genome

The primary sequences of SARS-CoV-2 Mpro was compared with that of SARS-CoV Mpro and MERS-CoV Mpro and it was found that SARSCoV-2 Mpro shares identity percentages of 96.08 and 50.65 with that of SARS-CoV Mpro and MERS-CoV Mpro respectively.

Table1. p-BLAST results for SARS-COV Mpro and MERS-COV with reference to SARS-COV-2 Mpro(6LU7).

Subject sequence	**Max score**	**Total score**	**Query coverage**	**E-value**	**Percentage identity**
SARS-COV Mpro	623	623	100	0.0	96.08
MERS-COV Mpro	322	322	100	3e-112	50.65

From the above data, with a query coverage of 100%, it is found that the sequences of the SARS CoV 2 resembled more with SARS-CoV Mpro in comparison to the MERS CoV. Mainly mutations that have arised from the SARS-CoV may be responsible for the difference in sequences among these three strains of the viruses.

4.1.2.2 Multiple Sequence Alignment Studies

The multiple sequence alignment studies for SARS, MERS and SARS-CoV 2 viruses was performed using the Clustal Omega tool. The FASTA sequences for the strains SARS (PDB ID:1WOF), MERS (PDB ID:5WKJ), SARS CoV 2(PDB ID:6LU7) were used for the study. From the sequence alignment studies, the results show the conservation of residues in the three domain-I (residues 1–99), II (residues 100–182) and III (residues 199–306) of the SARS-CoV-2 Mpro enzyme which includes catalytic dyad residues His41- Cys145 located in a cleft between domains I and II along with canonical oxyanion hole residues Gly143, Ser144 and Cys145 (Fig. 1). Domains I and II have a typical chymotrypsin-like two-β-barrel fold. Domain III possesses a globular structure consisting of five α-helices which is connected to Domain II by a loop of residues 183–198.

```
CLUSTAL O(1.2.4) multiple sequence alignment

5WKJ_1|Chain        MHHHHHHSGLVKMSHPSGDVEACMVQVTCGSMTLNGLWLDNTVWCPRHVMCPADQLSDPN 60
6LU7_1|Chain        -------SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPN 53
1WOF_1|Chains       --GPLGSSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPN 58
                          **: **:.***.**.********: ********:.*:*****:* :::: :**

5WKJ_1|Chain        YDALLISMTNHSFSVQKHIGAPANLRVVGHAMQGTLLKLTVDVANPSTPAYTFTTVKPGA 120
6LU7_1|Chain        YEDLLIRKSNHNFLVQA---GNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQ 110
1WOF_1|Chains       YEDLLIRKSNHSFLVQA---GNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQ 115
                    *: ***  :**.* **    . .:***:**:**. :*:*.**.:**.** *.*. ::**

5WKJ_1|Chain        AFSVLACYNGRPTGTFTVVMRPNYTIKGSFLCGSCGSVGYTKEGSVINFCYMHQMELANG 180
6LU7_1|Chain        TFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTG 170
1WOF_1|Chains       TFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTG 175
                    :********* *:*.:  .****.******* *******:. : . :.*****:*** .*

5WKJ_1|Chain        THTGSAFDGTMYGAFMDKQVHQVQLTDKYCSVNVVAWLYAAILNGCAWFVKPNRTSVVSF 240
6LU7_1|Chain        VHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDF 230
1WOF_1|Chains       VHAGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFLNRFTTTLNDF 235
                    .*:*: ::*.:** *:*:*. *.  **.  ::**:******::**  **::   *:: .*

5WKJ_1|Chain        NEWALANQFTEFVGTQ--SVDMLAVKTGVAIEQLLYAI-QQLYTGFQGKQILGSTMLEDE 297
6LU7_1|Chain        NLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDE 290
1WOF_1|Chains       NLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDE 295
                    *  *:  ::  :.  :   :. *:.:**:*: ::  :: : * .*::*: ****::****

5WKJ_1|Chain        FTPEDVNMQIMGVVMQ     313
6LU7_1|Chain        FTPFDVVRQCSGVTFQ     306
1WOF_1|Chains       FTPFDVVRQCSGVTFQ     311
                    *** **  *  **.:*
```

Figure7: Sequence alignments for MERS-COV Mpro, SARS-COV-2 Mpro and SARS-CoV Mpro

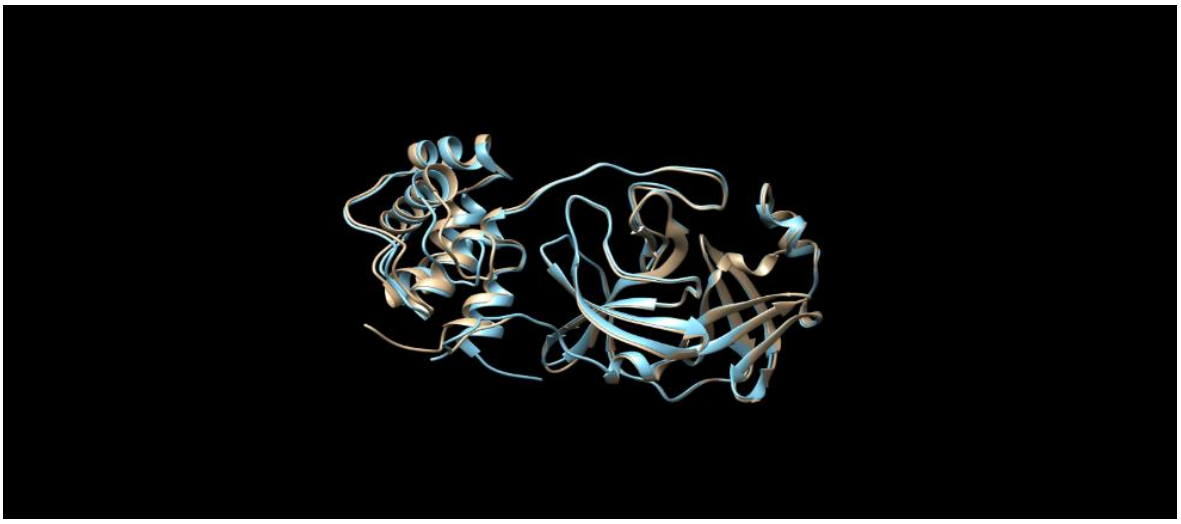
The results from the multiple sequence alignments were then intrepeted to study the phylogenetic relationship between these three closely related strains of the virus. It appears as if the association of SARS-CoV 2 virus is more closely related to SARS virus

when compared to the MERS corona viruses. A phylogenetic tree constructed to study such relationship also showed the above mentioned result as illustrated by figure 2.

Figure 8: Phylogenetic tree generation for SARS-COV, MERS and SARS-COV 2

4.1.2.3 Structural Alignment analysis

The protein databank (pdb) structures were downloaded for the strains SARS (PDB ID:1WOF), MERS (PDB ID:5WKJ), SARS CoV 2(PDB ID:6LU7) and these structures were structurally aligned using Chimera software version 1.14. employing sequence and align function inbuilt in the software. The pairwise structural superimposition of SARS-CoV-2 Mpro and SARS-CoV Mpro yields a Q-score of 0.837 while a Q-score of 0.758 was obtained when SARS-CoV-2 Mpro was superimposed on MERS-CoV Mpro. The pairwise structural imposition of SARS-CoV-2 Mpro and SARS-CoV Mpro yields an RMSD value of 0.517 Å (Fig. 3) while an RMSD value of 0.817 Å was obtained when SARS-CoV-2 Mpro was superimposed on MERS-CoV Mpro (Fig. 4). This indicates that SARS-CoV-2 Mpro is structurally closer to SARS-CoV Mpro as compared to MERS-CoV Mpro because lower the RMSD values higher the similarity in structural superposition.



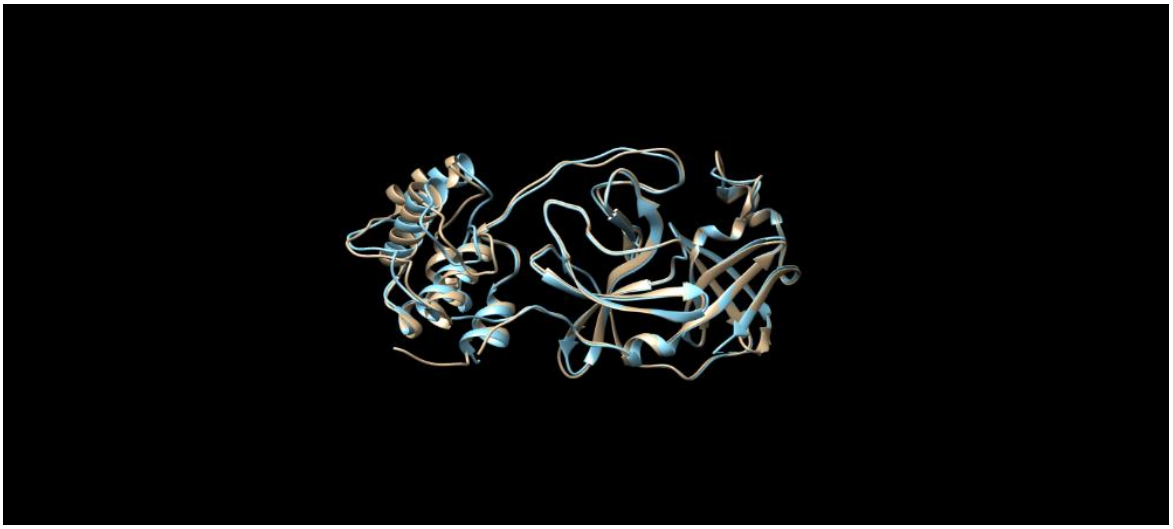


Figure 10: Structural superposition between 6LU7(tan)and 5WJK(cyan) using UCSF chimera.

Similarly, the sequence similarity after superposition of SARS-CoV-2 Mpro and SARS-CoV Mpro was 93.79% (figure 5) and that of SARS-CoV-2 Mpro and MERS-COV was found to be 46.69% (figure 6) as shown by MutaAlign tool in Chimera.

Figure 11: Sequence alignment following structural superposition of 6LU7 and 1WOF using MutaAlign viewer in chimera.

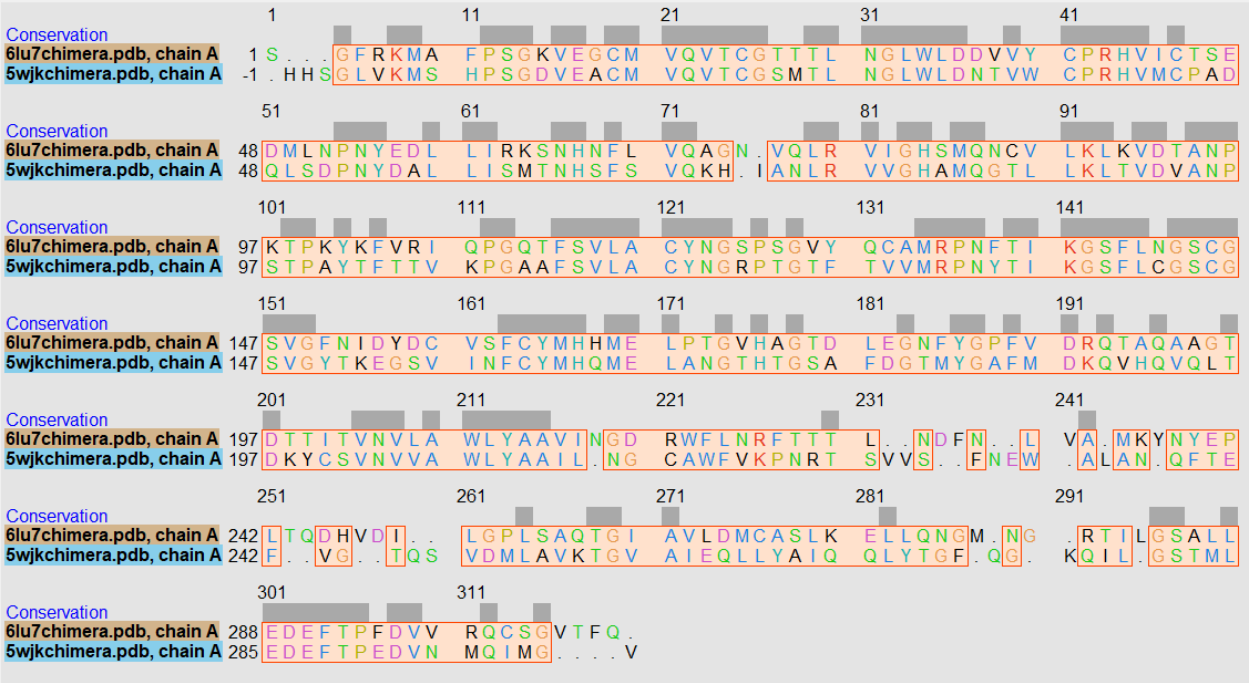


Figure 12: Sequence alignment following structural superposition between 6LU7 AND 5WKJ given by MutaAlign Viewer in UCSF Chimera.

Moreover, the above-mentioned color-coded diagrams represent the conserved residues among the SARS and SARS CoV 2 (fig 5) and that of SARS CoV 2 with MERS (fig 6). The sequence similarity percentage is slightly lower than that of the BLAST results when done by the MutaAlign Viewer software inbuilt in chimera v1.14. These might be due to the difference in respective algorithms employed by the softwares for sequence similarity searches and number of possible other reasons which could not be explained. Furthermore, this comparative study at the sequence and structural level gives the basic idea about the similarities among the viruses and these findings can be further exploited for the development of broad-spectrum inhibitors that are capable of inhibition of viruses in host cells.

4.2 Protein Tertiary Structure Analysis

4.3.1 Z-score and energy plot analysis

The assessment of the accuracy and reliability of experimental and theoretical models of protein structures is a necessary task that needs to be addressed before carrying out any kind of computational experimentations involving the drug discovery (Berman *et.*

al., 2006). The crystal structure selected for molecular docking procedures in our study was SARS CoV 2 M^{pro} (PDB id: 6LU7). The structure deposited on the database was further analysed for its z-score online server proSA (https://prosa.services.came.sbg.ac.at/prosa.php) and the Ramachandran plot analysis for the deposited structure was done using another server SAVES v 5.0(https://prosa.services.came.sbg.ac.at/prosa.php).

The *z*-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. Z-score, considered as a standard parameter of quality assessment (Wiederstein and Sippl, 2007), outside a range characteristic for native proteins indicate erroneous structures This plot can be used to check whether the *z*-score of the protein in question is within the range of scores typically found for proteins of similar size belonging to one of these groups (Sippl *et. al.,* 1993).

For our crystal structure in consideration, the web server displayed a z-score of -7.34. The second plot (Figure b), where structures from different sources (X-ray, NMR) are distinguished by different color codes, used to check whether the Z-score of the query structure is within the range of scores typically found for the similar sized native proteins, showed that z-scores for our protein structure in consideration was within the range of scores typically found for comparable proteins of same size. The first plot (Figure a) shows the local model quality by plotting energies as a function of amino acid sequence position where erroneous or inaccurate parts of the input structure are displayed by positive values. For our protein structure, some positive values for some residues were seen signifying some erroneous parts in crystal structure used.

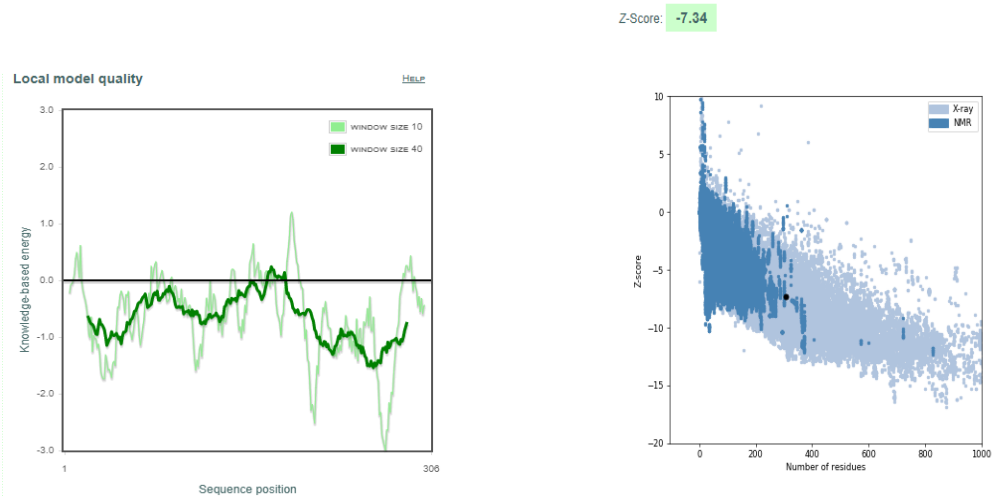


Figure 13: The Z score plots and energy plot for predicting 3D structure of a protein based on amino acids sequences.

a) *Energy plot for amino acid residues for the 3D-structure of SARS-CoV M^{pro} predicted by proSA*
b) *Z-Score plot of model predicted by proSA with structures available in database.*

4.3.2 Ramachandran plot analysis

The main basic requirement for a protein model is correct stereochemistry. Anomalies, such as phi/psi angle combinations that are placed in disallowed regions, steric collisions, and unfavorable bond lengths and angles are analyzed by programs such as PROCHECK (Laskowski, 1993) and WHATCHECK (Hooft *et. al.,* 1996). These programs also analyze these stereochemical features of the residues in the model and give an evaluation of the overall quality of a model or structure (Bhattacharya *et. al.,* 2007). Highlighting unrealistic conformations within the model by analysis of bond geometry by looking at Ramachandran plots is important. Certain conformations of phi and psi angles are forbidden in protein structures because they result in steric hindrance, or clashes between atoms. A good model will generally have 90% of its residues in the allowable regions of a Ramachandran plot (Laskowski, 1993).

For our protein crystal structure in consideration more than 90% of the residues were in allowable regions indicating that there were only few phi/psi angle combinations in the disallowed region as depicted in the figure below.

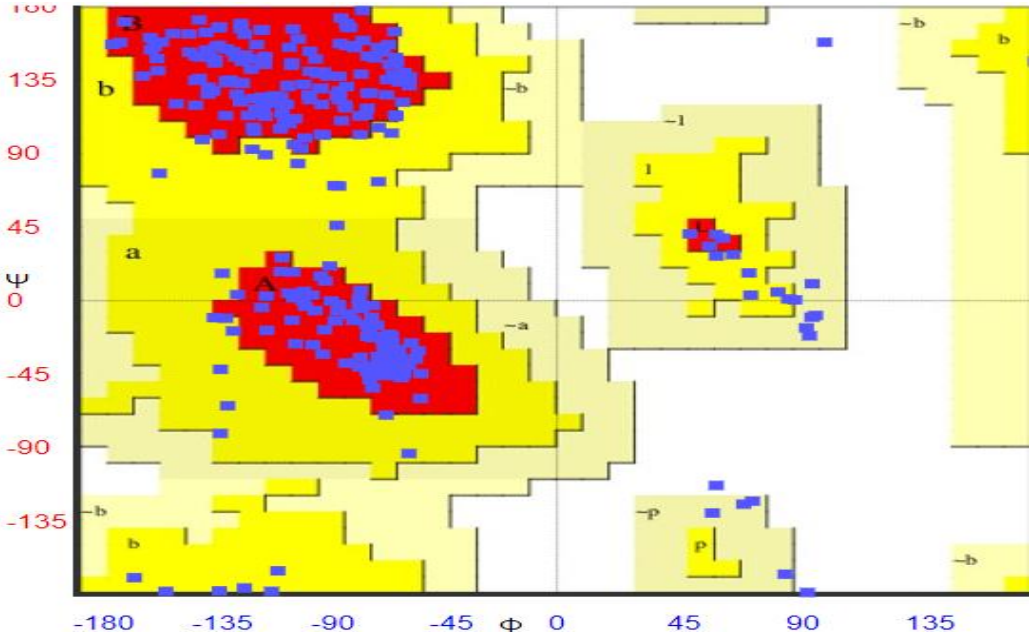


Figure 14: Ramachandran plot analysis using SAVES V 5.0 for amino acid residues of PDB id- 6LU7

4.3 Molecular docking simulation

4.3.1 Target protein preparation

Main protease being selected as target protein amongst the myriad of viral proteins essential to SARS-CoV 2. The X-ray diffraction structure of SARS-CoV-2 main protease with bound Michael acceptor inhibitor (N3) (PDB id: 6LU7) (Jin *et. al., 2020.*) was obtained from Protein Data Bank (rcsb.docking.org) in PDB format. Before docking simulations in the PyRx platform under Autodock Vina tool, the conversion of the PDB format into Pdbqt format is must. The water molecules present in the crystal structure deposited in RCSB site were deleted and the hydrogen atoms must be added using the same tool. The water molecules being loosely bound and not being involved in binding in many cases and deleting these molecules further makes the computation

easier and helps clear the possible binding pocket of the water molecules that would distort the pose search. Having said this, if any water molecule is found to be involved in possible binding interaction and those interactions to protein are of importance those water molecules should be preserved (Wong *et. al.,* 2010). Whether displacing a given structural water will improve binding depends on the balance between the enthalpy of contacts of the water molecule and the entropy change due to transferring the water to bulk solvent (Dunitz *et. al.,* 1994). In our study, no such water molecule was found with essential binding interactions that needed preserving, so all of the water molecules were deleted during the protein preparation.

Hydrogen bonds play a major role in the stabilization of protein-ligand complexes and the capability of a functional group to form them depends on the position of its hydrogen atoms (Lippert *et. al*., 2009). It is therefore important to identify hydrogen bonds and their properties that can be achieved by an accurate knowledge of the positions of hydrogen atoms in proteins. Several degrees of freedom are introduced by the high mobility of hydrogen atoms: Tautomeric states, where a hydrogen atom is capable of changing its binding partner, protonation states, where the number of hydrogen atoms at a functional group may change and torsional changes where the position of the hydrogen atom is rotated around the last heavy-atom bond in a residue (Lippert *et. al.,* 2010). Also, side-chain flips in glutamine and asparagine and histidine residues, which are common crystallographic ambiguities must be identified before structure-based calculations can be conducted. Further, the hydrogen atoms must be added before any pharmaceutical research as the protein structures may contain ambiguities that result from the experimental method, especially if it is X-ray crystallography, which has been used to determine most of the publicly available structures in the Protein Data Bank (PDB) (Berman et. al, 2000).

The merging of non-polar bonds if any present were also conducted. The docking conformation and the energy score of the resulting complex depends on the calculation of accurate partial charges on the protein and ligand both, possibly leading to more accurate estimation of complex geometry and binding energy (Tsai *et. al.,* 2008). There are several charge calculation methods which lead to significant differences in the partial charges assigned to the different atoms (Bikadi *et.al.,* 2009).

The Gasteiger charges are determined on the basis of electronegativity equilibration and is the measure of partial charges present in an protein tertiary structure (https://www.researchgate.net/post/What_is_the_differences_between_Kollman_and_Gasteiger_charges). Gaisteiger charges computed for our particular protein structure which was found to be -3.0036. The conversion of the dockable pdqt format from the respective protein tertiary structure comprises the basis for any docking simulation studies.

4.3.2 Ligand database preparation

The progressive optimization of the pharmacological properties of ligands (potential leads against biological target) and potency of these compounds are the focal points of early-stage drug discovery which can be achieved by the identification of lead compounds showing pharmacological activity against a biological target (Lionta *el. al.,* 2014). The selection of ligand database is always a crucial and determining factor for any kind of appropriate drug discovery against a selected target and there are different classes of compounds from natural to synthetic compounds, with different chemical and structural properties capable of being used as drug against any disease or viral infection (Glaab *et. al*., 2016). Depending on the goal and type of the study (e.g. drug development, toxin identification, pesticide development), the ligand library may contain already known drug substances for repositioning, synthetic substances similar to lead or drug compounds for subsequent structural optimization or other natural or xenobiotic compounds (Schames *et. al.,* 2004).

In recent years, drug repurposing methodology has come out as a successful alternative to fasten the drug development process against rapidly spreading emerging infections such as the one of present pandemic caused by SARS-CoV-2 (Kumar *et. al,* 2020). The approach of drug repurposing has successfully led to the discoveries of potential drug candidates against several diseases such as Ebola disease, hepatitis C virus, and zika virus infection (He *et. al.,* 2015 and Barrows *et. al.,* 2016). Recently several repurposing studies on SARS-CoV-2 have been performed using clinically approved drugs mainly the FDA approved ones (Fan et al, 2020), among

which a very new study comes out on a clinical trial of Lopinavir–Ritonavir drug for COVID-19(Cao *et. al,* 2020)., which was indicated on top of other drug repurposing study conducted by Kumar *et. al.*

The focus of our research was to find out the appropriate drug molecule against the pandemic causing SARS-CoV 2, Main protease, in particular as soon as possible. Opting to FDA approved ligand database, not only gave us the privilege to repurpose the ligand library (Sun *et.al.,* 2016) but also the resulting lead compounds could be used directly in human as these were the pool of drugs being used to treat different human diseases and anomalies. Based on these grounds, a ligand library of FDA approved 1167 compounds were chosen without any biasness from the zinc 15 database for development of putative leads against the SARS-CoV 2.

4.3.3 Identification of Active Binding Site

The ligand binding sites were identified from the crystal structure of SARS-CoV 2 Main Protease (M^{pro}) of (PDB id: 6LU7) available in RCSB protein data bank. The reference inhibitor N3 bound in the crystal structure exhibited certain binding interactions with the respective protein. The same set of amino acid residues involved in establishing a stable conformation were elucidiated using another software PyMoL. Several other works by Gimeno et al, Tang et al, Lin et al and Zhang et al were also consulted for choosing the active amino acid residues involved in protein ligand interaction. The enzyme has a catalytic dyad comprised of His41 and Cys145.Also for the predicted 3D-structures of M^{pro} , 3DLigandSite, a web server that superimposes the ligands bound to the structures similar to the query and thus predicts the binding site (Wass *et. al.*, 2010), was used to predict the ligand-binding sites but the results from this site which involved the name of residues was somewhat inconsistent with those from the literatures and those generated using PyMol.

With regard to the recent literatures and the protocol of finding amino acid residues by PyMol, a set of amino acid residues were chosen as ligand binding site for molecular docking simulations.

4.3.4 Pre-Virtual Screening of the FDA ligand database

Docking was performed in Pyrx 0.9.8 platform (https://pyrx.sourceforge.io/) which is an an easy-to-use user interface, using AutoDock Vina. The target protein Main protease also known as 3CL-^{pro} was docked to screen ligands with number of modes set at 16 and exhaustiveness set at a value 32. Following the docking simulations, the Michael acceptor inhibitor N3 exhibited a binding affinity of -8Kcal/mol. It is reported that the higher the negative value of binding affinity, the stronger is the binding of the ligand in the target (Dallakyan and Olson, 2015), thus the compounds showing binding energy greater in negative value than the reference inhibitor N3 were taken in consideration for further analysis. Out of 1167 ligands, only 52 ligands exhibited equal or higher binding energy (BE) than the native ligand N3 towards M^{pro} in SARS-CoV 2 (PDB id:6LU7). This indicated that these ligands could potentially interfere with the functioning of main protease in cleaving the polyprotein sequence to functional proteins thereby inhibiting the replication and transcription of the virus.

4.3.5 Final Ligand Library Preparation

The top hit 52 compounds from the initial library of ligands were further taken for further molecular docking simulations. The 3D structure of the top hits downloaded from pubchem database (https://pubchem.ncbi.nlm.nih.gov/) were further essential for other downstream processing like DFT analysis and molecular docking simulation studies. These compounds after being individually downloaded and appended using Osiris datawarrior comprised of pool of potential leads.

4.3.6 In-silico ADME/Tox tests for possible hits

All the ligands with higher negative binding energy than the reference inhibitor could now be used for further molecular docking simulations. But ADME/tox properties relating to absorbtion, distribution, metabolism and excretion and toxicity, generally carried out at the end of drug discover pipeline, play an important role as they account for the failure of 60% of drug molecules during the drug development

process(https://www.sciencedirect.com/topics/pharmacology-toxicologyandpharmaceuticalscience/adme#:~:text=6.7%20ADME%2DTox%20Prediction&text=ADMET%20properties%20play%20an%20important,during%20the%20drug%20development%20process.&text=Such%20drugs%20may%20show%20poor,might%20be%20toxic%20in%20nature). Thus, performing ADMET in earlier step could be a smart move. The primary reason for leads obtained from any research not being able to pass the clinical trials is their inability to reach the target protein and perform its predicted function, and also the toxicity imparted by those compounds(Hughes *et al.*, 2011). Thus, ADMET and physiochemical properties evaluation in the early stages of drug discovery seem to be a wiser choice. The toxic profiles and druglikeness were studied for the pre-screened ligands using OSIRIS.

The selection criteria include the following values for molecular descriptors:

- Mol weight – 200 to 500 Daltons
- cLogP – -3 to 6
- cLogS – greater than -4 i.e. -4 to -2
- Hydrogen bond donors – 0 to 5
- Hydrogen bond acceptors – 0 to 10 (Lipinski, 2004)
- Topological Polar Surface Area (tPSA) – 0 to 120
- Rotatable bonds – 10 or less (Veber *et. al.*, 2002)
- Druglikeness – Positive value

The molecular weight of the drug is another important factor to be considered for drug discovery. Typically, the compounds with low molecular weight are capable of modulation different biochemical process to diagnose, prevent or treat diseases and the drug like compounds have a molecular weight of less than 550 daltons with some exceptions (Ngo *et. al.,* 2018).

The logP value of a compound, which is the logarithm of its partition coefficient between n-octanol and waterlog($c_{octanol}/c_{water}$), is well established measure of the compound's hydrophilicity. Low hydrophilicities and therefore high logP values cause poor absorption or permeation. It has been shown for compounds to have a

reasonable probability of being well absorbed, their clogP similarity must not be greater than 5.0. Lipophilicity is an important physicochemical parameter that contributes to absorption, distribution, metabolism, excretion, and toxicity of a drug and very high levels of lipophilicity can also lead to toxicity and metabolic clearance (Arnot et.al, 2012). This molecular property has been shown to be important in solubility, oral availability, transport, penetration of the blood-brain-barrier and receptor binding as well (Hansch et al, 1962 & 1987).

The aqueous solubility of a compound significantly affects its absorption and distribution characteristics. Any drug to be absorbed must be present in the form of solution at the site of absorption (Savjani et al, 2012). Typically, a low solubility goes along with a bad absorption and therefore the general aim is to avoid poorly soluble compounds. Moreover, solubility is one of the important parameters to achieve desired concentration of drug in systemic circulation for achieving required pharmacological response (Sharma et al, 2009) and those drugs which are poorly water soluble, often require high doses to reach therapeutic plasma concentrations after oral administration (Vemula et al, 2010).

Hydrogen bonds play an important role on the specificity of ligand binding (Wade et al, 1989). To cross the membrane, the hydrogen bond to water molecules must be stripped off and too many hydrogen bond donor / acceptor groups make this process energetically very expensive. The number of rotable bond counts measures the molecular flexibility and is in turn associated with the oral bioavailabilty of drugs (Veber et al, 2002) and most of the drug like compounds have 1-10 rotatable bonds (Khanna et al, 2010). Topological polar surface area (TPSA) of a molecule can be best described as the surface sum over all polar atoms or molecules, primarily oxygen and nitrogen, also including their attached hydrogen atoms. It relates the drug's ability to permeate cells and molecules with TPSA greater than 140 angstroms are considered poor in permeating the cells and hence poorer drugs (Pajouhesh et al, 2005).

Various drug relevant molecular descriptors like molecular weight, cLogP, cLogS, Druglikeness, and toxicities like mutagenicity, tumorigenicity, reproductive effects and

irritant effects can be calculated using OSIRIS for the lead molecules on the basis of functional groups present in their structures (Sander *et. al.*, 2015). ADMET filters using OSIRIS narrowed these 52 pre-screened FDA compounds down to 6 ligands that were now eligible for further docking simulations. (Appendix 8.2).

4.3.6 Virtual Screening of the final FDA ligand database

The target protein Main protease also known as $3CL^{pro}$ or M^{pro} was docked in Pyrx 0.9.8 platform (https://pyrx.sourceforge.io/) which is an easy-to-use user interface, using AutoDock Vina, to screen ligands which were filtered by OSIRIS datawarrior for their toxicity and druggability parameters, with number of modes and exhaustiveness both being set at a value 32. Out of the six compounds obtained from the pre-screening and filtering of drugability parameters, only four compounds exhibited a higher negative value of binding energy than the reference Michael acceptor inhibitor, N3(-8.0 Kcal/mol). The following table represents the list of four compounds with binding affinity greater in negative value as compared to the reference inhibitor:

Table 2: Binding affinity of prioritized hits against main-protease.

Ligands	**Zinc id**	**Pubchem cid**	**Binding affinity against M^{pro}(kJ/mol)**
Tadalafil	ZINC3993855	110635	-9.3
S-paliperidone	ZINC18324766	9823781	-8.7
Dolutegravir	ZINC58581064	135400189	-8.6
Vardenafil	ZINC18324776	54726191	-8.3
N3(reference inhibitor)	-	-	-8.0

4.4 Analysis of docking results

4.4.1 Protein ligand interaction

The final four leads obtained from the six ligands pre-screened from a library of FDA approved compounds followed by ADMET filter were further selected for studying their interaction with the protein in consideration. Following the docking results, the ligand and macromolecule were viewed in PyMol for determining the important amino acid residues involved in binding the ligand in the active binding cavity of the crystal structure. For this the protein ligand interaction within 5 angstroms of the active binding site of the protein structure was considered. The following table includes the interaction (within 5 angstroms) of amino acid residues with the top four hits.

Table 3: Ligand-protein interaction of prioritized compounds

Ligands	**Amino acid residues (within 5A°)**
Tadalafil	His41, Met49, Tyr54, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, His172, Asp187, Arg188, Gln189
S-paliperidone	Thr25, Thr26, Leu27, His41, Met49, Asn142, Gly143, Ser144, Cys145, His164, Met165, Glu166, Pro168, Arg188, Gln189, Thr190, Ala191, Gln192
Dolutegravir	Thr25, Thr26, His41, Met49, Leu141, Asn142, Gly143, Ser144, Cys145, His144, His163, His164, Met165, Glu166, Leu167, Pro168, Val186, Arg188, Thr190, Gln192

Vardenafil	Thr25, Thr26, Leu27, His41, Cys44, Thr45, Ser46, Tyr54, Phe140, Leu141 Asn142, Gly143, Ser144, Cys145, His144, His163, His164, Met165, Glu166, His172, Arg188, Thr190, Gln192

From the above table two residues mainly Histidine at position 41 and Cysteine at position 145 are involved in binding the ligand to the protein. This interaction can be justified by the fact that these two residues are indeed the catalytic dyad involved in showing possible interactions with different classes of ligands (Gurung et. al, 2020). Likewise, other residues like histidine, methionine, glutamate, and arginine at position 164, 165, 166 and 188 respectively are also offering significant number of interactions with all four ligands. These amino acid residues from the same protein are also found to be interacting with other class of ligands in the works of Gurung et. al and Shamsi et. al. The involvement of these different amino acid residues and their possible binding interactions of different sorts must have been responsible for providing these above-mentioned compounds with better binding affinity than that of the reference Michael acceptor inhibitor, N3.

4.4.2 Bond length and interaction types responsible for drug ligand interaction

The protein ligand interaction can be further illustrated by studying the nature of interaction and bond length responsible for stabilizing the protein ligand interaction. Different types of interactions exist between the amino acid residues of the protein and ligands. The following tables depicts the bond length and interaction type amongst the amino acid residues of protein and lead compounds obtained after molecular docking simulations.

Table 4: Nature of chemical interaction between dolutegravir and main protease

Amino acid residues	**Bond length(A°)**	**Interaction type**	**Chemical group of ligand involved**
His41	4.71	Pi-cation	Substituted benzene
Gly143	1.88	Conventional hydrogen bond	Oxo- group
Cys145	4.42	Pi-sulfur	Substituted benzene
Glu166	2.83	Conventional hydrogen bond	Oxo- group
Arg188	3.49	Halogen	fluorine
Thr190	3.14	Halogen	fluorine
Gln192	2.73	Conventional hydrogen bond	fluorine

Table 5: Nature of chemical interaction between tadalafil and main protease

Amino acid residues	**Bond length(A°)**	**Interaction type**	**Chemical group of ligand involved**
His41	4.17	Pi-cation	Substituted cyclopentane
His41	5.31	Pi-alkyl	Substituted benzene
Met49	4.47	Pi-alkyl	benzene
Met49	4.91	Pi-alkyl	Substituted cyclopentane
Phe140	3.60	Carbon hydrogen bond	Substituted benzene
Leu141	3.22	Carbon hydrogen bond	Substituted benzene
Gly143	2.21	Hydrogen bond	Substituted benzene

Ser144	3.71	Carbon hydrogen bond	Substituted benzene
Glu166	2.64	Hydrogen bond	Substituted benzene

Table 6: Nature of chemical interaction between S-paliperidone and main protease

Amino acid residues	**Bond length(A°)**	**Interaction type**	**Chemical group of ligand involved**
Thr26	3.04	halogen	fluorine
His41	4.02	Pi-cation	Substituted benzene
His41	4.97	Pi-pi t shaped	Substituted benzene
Gly143	1.97	Conventional hydrogen bond	Substituted furan
Ser144	3.08	Conventional hydrogen bond	Substituted furan
Cys145	4.14	Carbon hydrogen bond	Substituted benzene
Cys145	3.60	Carbon hydrogen bond	Substituted furan
Glu166	3.37	Carbon hydrogen bond	carbon
Gln189	3.75	Carbon hydrogen bond	cyclohexane
Thr190	2.11	Conventional hydrogen bond	Substituted benzene

Table 7: Nature of chemical interaction between vardenafil and main protease

Amino acid residues	**Bond length(A°)**	**Interaction type**	**Chemical group of ligand involved**
His41	2.48	Conventional hydrogen bond	oxygen
His41	5.34	Pi-sulfur	Substituted benzene
Met49	3.95	alkyl	methyl
Met49	5.46	alkyl	methyl
Gly 143	2.72	Conventional hydrogen bond	oxygen
Cys145	3.69	Pi donor hydrogen bond	Substituted benzene
His163	5.36	alkyl	methyl
His164	2.96	Conventional hydrogen bond	nitrogen
Met165	5.05	Pi-alkyl	Substituted benzene
Met165	4.13	Pi-alkyl	cyclopentane

The molecular interaction in case of compound dolutegravir is facilitated through conventional hydrogen bonding with three residues Gly143, Glu166 and Gln192, halogen bonding with two residues Arg188 and Thr190, pi-cation interaction involving His41 and pi-sulfur interaction involving Cys145. Similarly, Tadalafil formed two hydrogen bonds with residues Gly143 and Glu166, carbon hydrogen interaction with Phe140, Leu141 and Ser144, pi-alkyl interaction with His41 and Met49 and pi-cation interaction with His41. Another compound S-paliperidone exhibited molecular interaction by conventional hydrogen bonding with residues Ser144, Cys145 and Thr190, carbon hydrogen bonding with Cys145, Glu166 and Gln189, halogen bonding with Thr26, pi-cation interaction with His41 and pi-pi T-shaped interaction with His41. Likewise, the molecular interaction displayed by another compound vardenafil was

based on the hydrogen bonding with His41, Gly143 and His164, alkyl bonding with Met49, pi-alkyl interaction with Met165, pi-sulfur interaction with Sulfur atom of His41 and pi-donor hydrogen bonding interaction with Cys145.

4.4.3 Hydrophobic bond interactions

The tendency of hydrocarbons (or of lipophilic hydrocarbon-like groups in solutes) to form intermolecular aggregates in an aqueous medium, and analogous intramolecular interactions constitute the hydrophobic interactions (http://goldbook.iupac.org/terms/view/H02907). Hydrophobic interactions describe the relation between water and non-polar molecules having a long chain of carbon that usually do not interact with water molecules(https://chem.libretexts.org/Bookshelves). Hydrophobic contacts are important for the folding of proteins keeping them stable, biologically active, and reduce the undesirable interactions with water. The hydrophobic interaction can increase the binding affinity between the drug target interfaces (Davis *et. al.,* 1999). An increase in the number of hydrophobic atoms in the active core of drug -target interface further increases the biological activity of the drug lead.

The hydrophobic interactions exhibited in the protein- ligand interactions were studied using Ligplot software. The compound dolutegravir showed hydrophobic interaction via eleven residues Thr25, Thr26, His41, Met49, Asn 142, Gly143, His164, Met165, Glu166, Gln189 and Thr190. Similarly, the amino acid residues including Leu27, His41, Asn142, Cys145, His164, Met165, Glu166, Arg188, Gln189and Ala191 were involved in hydrophobic interactions with the compound S-paliperidone. Likewise, the compound tadalafil displayed hydrophobic interactions via the amino acid residues namely His41, Met49, Phe140, leu141, Asn142, Ser144, Cys145, His163, His164, Met165, Asp187, Arg188 and Gln189. Similarly, the compound vardenafil showed hydrophobic interaction with residues like Thr25, Leu27, Ser46, Met49, Tyr 54, Phe140, Leu141, Asn142, Gly143, Cys145, Met 165, His164, Glu166, Asp187, Arg188 and Gln189. The higher binding affinity of the compound tadalafil compared to others may also have been attributed to a certain extent by these hydrophobic interactions. The presence of hydrophobic interactions involving hydrophobic amino

acids including Phe140 and Leu141 (absent in other three compounds) may have been responsible to strengthen the drug protein interaction to some extent mainly in tadalafil.

4.5 Mutagenesis studies

Mutations often lead to changes in the structure of an encoded protein or to a decrease or complete loss in its expression and the change in the DNA sequence affects all copies of the encoded protein, mutations can be particularly damaging to a cell or organism (https://www.ncbi.nlm.nih.gov/books/NBK21578/). In contrast, any alterations in the sequences of DNA in cases of virus may confer greater infectivity as implicated in the case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein that acquired a D614G mutation early in the pandemic that further enhanced the infectivity and is now the globally dominant form (Weissman *et. al.,* 2020). This is just a mere example of how possible mutations in viral genome can strengthen their infectivity and pathogenesis. With several permutations and combinations possible for mutation in the viral genome affecting the encoded protein, it is a wider choice to integrate mutagenesis studies during the drug development studies.

4.5.1 Preparation of probable mutated viral proteins

The generally high mutation rates of RNA viruses provide them with the ability to escape host immune responses, improve their virulence, and even change tissue tropism (Duffy et al, 2018; Rasschaert *et. al.,* 1990). Though the extinction events are observed with RNA viruses like Influenza A H1N1 strains (Carter *et. al.,* 2012) and SARS-CoV strains (He *et. al.,* 2004), one can never predict the outcome of mutations in RNA viruses. Despite the genomic datas suggesting a low rate of mutations in SARS-CoV 2(Rice *et. al.,* 2020), SARS-CoV 2 M^{pro} on the other hand, has endured a relatively higher rate of non-synonymous mutations at residue position 15 (G15D/S) in domain I, residue position 157 (V157I/L) in domain II, and at position 184 (P184L/S) within the interdomain linker region (Amamuddy *et. al.,* 2020).

Previously, studies in SARS-CoV have foud that the S1 subsite (residues PHE140, HIS163, MET165, GLU166, and HIS172), an oxyanion hole (main-chain amides of residues GLY143, SER144, and CYS145), and an oxyanion loop (residues SER139-LEU141) are important players in stabilizing the binding site in the active form of M^{pro} in SARS-CoV (Yang *et. al.*, 2003; Hu *et. al.,* 2009 and Cheng *et. al.,* 2010). From the alignment studies conducted in our study, all these above-mentioned residues were found to be conserved in SARS-CoV 2 M^{pro} and it was also found that residues like GLY143, SER 144 and GLU166 were actively involved in different types of interactions with top four leads from our study.

Further, the residues namely GLY143, SER 144 and GLU166 from the crystal protein structures were selected for mutagenesis studies and their preferred codons in the SARS-CoV 2 viral sequence were retrieved which were GGC, AGC and GAG for GLY143, SER 144 and GLU166 respectively. The conversion of nucleotides from Guanine to Adenine and that from Cytosine to Thymine were more energetically favourable. With this consideration, GLY143 (GGC) was mutated to Serine (AGC, AGT), Aspartic acid (GAC, GAT) and Aspargine(AAC) whereas SER144(AGC) was mutated to Aspargine (AAC, AAT) and GLU166 (GAG) to Lysine (AAG, AAA). Following the mutagenesis protocol in PyMol, the new crystal structure of mutated proteins was selected on the basis of higher percentage of mutation probability and lower steric hindrance criteria.

4.5.2 Docking simulations and analysis involving the mutated protein

4.5.2.1 Structure based virtual screening: Binding affinity of ligands with mutated protein target

Table 8: Binding affinity of top hits on mutated main proteases

Ligand involved	Mutation	Binding energy (Wild type)(Kcal/mol)	Binding energy (Mutated protein)(kcal/mol)
Tadalafil	E166K	-9.3	-8.5

Tadalafil	G143D	-9.3	-8.8
Tadalafil	G143S	-9.3	-9.7
Tadalafil	G143N	-9.3	-9.6
S-paliperidone	G143D	-8.7	-8.8
S-paliperidone	G143S	-8.7	-8.3
S-paliperidone	G143N	-8.7	-8.8
S-paliperidone	S144N	-8.7	-8.6
Dolutegravir	G143D	-8.6	-8.5
Dolutegravir	G143S	-8.6	-8.1
Dolutegravir	G143N	-8.6	-8.4
Vardenafil	G143D	-8.3	-8.9
Vardenafil	G143S	-8.3	-8.3
Vardenafil	G143N	-8.3	-8.9

Tadalafil exhibited a binding energy of -9.3kcal/mol with the wild type of SARS-CoV 2 M^{pro}. The mutation at single residues E166K, G143D, G143S and G143N in the protein structure yielded a binding energy of -8.5, -8.8, -9.7 and -9.6Kcal/mol respectively for tadalafil. There is decrease of binding affinity compared to the wild type in case of E166K and G143D mutations whereas slight increase in binding affinity in case of G143S and G143N mutations. The mutation of glutamate to lysine and aspartic acid may have contributed to somewhat weaker interactions to the ligand thereby decreasing the binding affinity in these mutations but the mutations of glutamate to serine and aspargine may have offered stronger interactions with the ligands thereby increasing the binding affinity of the respective compound. Nevertheless, in all these four possible mutations the binding affinity of the compound tadalafil is still higher in negative value than the reference Michael acceptor inhibitor N3, thus if the viral protein mutates according to our predicted analysis, this compound is still capable of inhibiting the SARS-CoV 2 M^{pro}.

S-paliperidone exhibited a binding energy of -8.7 kcal/mol with the wild type of SARS-CoV 2 M^{pro}. The mutation at single residues S144N, G143D, G143S and G143N in the protein structure yielded a binding energy of -8.6, -8.8, -8.3 and -8.8 Kcal/mol

respectively for S-paliperidone. S144N and G143S mutations led to the decrease in binding energy of the interaction with this ligand but mutations like G143D and G143N accompanied with stronger binding interactions. Despite the marked decrease in binding affinities for first two mutations, this drug could still inhibit the SARS-CoV 2 M pro. Similarly, Dolutegravir with a binding energy of -8.6 kcal/mol with the wild type of SARS-CoV 2 M^{pro} following mutations at single residues G143D, G143S and G143N in the protein structure yielded a binding energy of -8.5, -8.1, and -8.4 Kcal/mol respectively. This compound too showed the prospects in inhibiting the SARS-CoV 2 M^{pro} even after the predicted mutations in only two mutations namely G143D and G143S. In case of G143N mutation, this compound showed lower binding strength than the reference inhibitor thereby rendering its inactivity to inhibit the viral protein if the predicted mutation is to occur. The mutations at single residues G143D, G143S and G143N in the protein structure yielded a binding energy of -8.9, -8.3, and -8.9 Kcal/mol respectively for vardenafil which had a binding strength of -8.3kcal/mol with the wild type. Vardenafil could also inhibit the predicted mutated structures of the SARS-CoV 2 M^{pro}. The increase in binding strength of G143D and G143N mutated viral protein structure to the ligand in consideration with respect to the wild type suggested that these possible mutations might be responsible in imparting stronger interaction with the ligand thereby increasing the inhibitory effect of ligand in these mutated protein structures.

4.5.2.2 Ligand protein interaction involving the mutated residues

The mutation in important amino acid residues of the protein structure responsible for interacting with different ligand results in the changes in types of interaction thereby altering the binding strength of the mutated protein structure to the same ligand. The bond length and bond type responsible for molecular interactions were compared amongst the wild and mutated protein structure as shown in the table below.

Table 9: Ligand protein interaction of top ligands in wild and mutated protease strains

Ligand	Wild type residue	Bond length	Bond type	Mutated residue	Bond length	Bond type
Tadalafil	GLU166	2.64	Hydrogen bond	LYS166	2.25	Hydrogen bond
Tadalafil	GLY143	2.21	Hydrogen bond	SER143	2.97	Hydrogen bond
Tadalafil	GLY143	2.21	Hydrogen bond	SER143	3.09	Pi donor hydrogen bond
Tadalafil	GLY143	2.21	Hydrogen bond	ASP143	2.28	Hydrogen bond
Tadalafil	GLY143	2.21	Hydrogen bond	ASN143	3.10	Hydrogen bond
S-paliperidone	GLY143	1.97	Hydrogen bond	SER143	2.08	Hydrogen bond
S-paliperidone	GLY143	1.97	Hydrogen bond	ASP143	2.09	Hydrogen bond
S-paliperidone	GLY143	1.97	Hydrogen bond	ASN143	2.06	Hydrogen bond
S-paliperidone	SER144	3.08	Hydrogen bond	ASN143	-	-
Vardenafil	GLY143	2.72	Hydrogen bond	SER143	-	-
Vardenafil	GLY143	2.72	Hydrogen bond	ASP143	2.67	Hydrogen bond
Vardenafil	GLY143	2.72	Hydrogen bond	ASN143	2.63	Hydrogen bond
dolutegravir	GLY143	1.88	Hydrogen bond	SER143	-	-

dolutegravir	GLY143	1.88	Hydrogen bond	ASP143	2.02	Hydrogen bond
dolutegravir	GLY143	1.88	Hydrogen bond	ASN143	-	-

Single mutations on important amino acid residues allowed us with the leverage of comparing their individual role in molecular interaction both in wild and mutated viral proteins. The interactions involving the mutated residues were mainly focused and their implications on the binding affinity of the molecular interaction was further illustrated. For example, the increase in binding strength of Tadalafil in case of G143S mutation compared to the wild protein type may be supported by the fact that Ser143 residue in the mutated protein introduces another additional interaction pi-donor hydrogen bond with bond length of 3.09 angstroms with the ligand. Similarly, in case of G143N mutation, the hydrogen bond interaction of GLY143 (wild type protein) with tadalafil had a shorter bond length than the hydrogen bond interaction of ASN143 residue (mutated residue). Despite this the higher binding affinity of ligand and protein was seen which may be due to the other possible interactions imparted by other important amino acid residues in the mutated protein which could not be explained properly.

The decrease in binding affinity of S144N mutation regarding S-paliperidone can be attributed to the fact that following the indicated mutation, ASN144 was no longer capable of forming any type of interaction with the ligand whereas SER144 in the wild type contributed to the overall binding potential by offering a hydrogen bond interaction. Likewise, G143S mutations led to the increase in the bond length from 1.97 angstroms to 2.08 angstroms, the nature of hydrogen bond interaction remaining same in the wild and mutated protein, supporting the decrease in binding affinity in the mutated residue compared to the wild type regarding S-paliperidone. Although increase in bond length following the G143D and G143N mutations in regard to S-paliperidone was seen, other possible interactions of high caliber may have contributed to a stronger binding strength which could not be explained properly.

The decrease in binding affinity of G143N and G143S mutations with dolutegravir can shed light into the fact that following the indicated mutations, ASN143 and SER143 no longer formed any type of interaction with the ligand in contrast to the hydrogen bond interaction offered by GLY143 in the wild type. The decrease in bond lengths of the possible hydrogen bond interactions following the G143D and G143N mutations involving vardenafil comply with their corresponding increase in binding affinity with the ligands because shorter the bond length, higher the strength of binding between protein and ligand (Petrucci *et. al.*, 2007).

4.5.2.3 Hydrophobic bond interactions amongst the mutated strains and ligands

After the molecular docking simulations of the mutated proteins with the ligands, the possible hydrophobic interactions imparted by these mutated amino acids were compared with the wild type. The amino acid residues GLY143 and GLU166 of wild type both showed hydrogen bonding while interacting with tadalafil. The residues ASN143 and SER143 both showed hydrophobic bond interactions following the G143N and G143S mutations respectively with tadalafil but on the other hand residues ASP143 and LYS166 showed hydrogen bonding following G143D and E144K mutations respectively. The amino acid residue Gly143 of wild type showed hydrophobic interaction with dolutegravir and the hydrophobic interaction was only retained in the G143N mutation of the three mutations (G143S, G143N AND G143D) where ASN 143 exhibited hydrophobic interactions with the same ligand. None of the three mutations G143S, G143N AND G143D could retain the hydrophobic interactions as imparted by GLY143 in the wild type while interacting with Vardenafil. All the mutated residues ASP143, SER143 and ASN143 showed hydrogen bonding with Vardenafil and S-paliperidone following G143D, G143S and G143N mutations respectively with both ligands. Hydrophobic interaction was established between the residue ASN144 and S-paliperidone following the S144N mutation.

Glycine having hydrophobic side chains when mutated to Serine which is uncharged but with polar hydroxyl groups (Lodish *et. al.,* 2000) participated in the hydrogen bonding as seen in G143S mutations where mutated serine residue showed hydrogen

bonding with ligands like S-paliperidone and vardenafil. Aspargine which is uncharged but having polar amide groups with extensive hydrogen binding capacities (Faltejskova *et. al.,* 2020) formed hydrogen bond with S-paliperidone following G143N mutations but the formation of hydrophobic bonds by ASN 143 after G143N mutation with tadalafil, vardenafil and dolutegravir could not be explained. Similarly, the hydrogen bonding nature of ASP143 following G143D mutations with all four ligands could not be clarified properly.

4.6 Narrowing down the top hits

4.6.1 Cross reactivity with human proteins: MAT1A as a reference

Since these top hit compounds obtained from screening from library of FDA approved compound, studying their cross reactivity with important human proteins can be a tactful strategy assisting the performance of the drug when used in human. MAT (methionine adenosyltransferase, also known as *S*-adenosylmethionine synthetase) is important human enzyme that synthesizes SAM (*S*-adenosylmethionine), the principal methyl group donor and precursor for polyamine and glutathione synthesis (Lu *et. al.,* 2008; Mato *et. al.,* 1997). MAT catalyzes the transfer of the adenosyl group from ATP to the sulfur atom of Met (L-methionine), in an unusual two-step reaction cleaving at both ends of the ATP triphosphate chain (Markham *et. al.,* 1987). The MAT1A gene product is expressed in adult liver, where most transmethylation reactions take place highlighting the importance of MAT1A in maintaining the normal human metabolism (Shafqat *et. al.,* 2013). All the four top hit compounds were checked for their inhibition of human MAT1A and unfortunately all four compounds showed higher binding affinity than native ligand SAM following the molecular docking interactions. Tadalafil and Vardenafil showed a binding affinity of -8.5kcal/mol whereas S-paliperidone and dolutegravir showed a binding affinity of -9.4 kcal/mol as compared to the lower binding affinity of native ligand SAM (-7.9 Kcal/mol).

However, these drugs being FDA approved and being used in human for different diseased conditions could be used as the SAM requirements in human can be

replenished from external administration but with proper care. Further, SAM as hepatoprotective agent is indispensable to people with poor liver health (Agrimi et al, 2004) and this fact can't be ignored during the use of these compounds as probable drugs against SARS-CoV M^{pro}.

4.6.2 Selection of the best drug candidate

Amongst, the four top hits obtained from molecular docking simulations, these compounds were further analyzed for their molecular interaction, efficacy in the predicted mutated viral proteins and cross-reactivity with human proteins. Preference index was given to all the drugs based on the molecular descriptors like hydrogen bond donors, hydrogen bond acceptors and number of rotatable bonds in the potential drug molecule. An empirical formula was set and a total score of five was assigned and a hypothesis of higher the score, higher the probability of being active drug candidate was set.

Table 10: Selection of prioritized compounds on basis of binding affinity and preference index

Ligand	**Binding affinity with SARS CoV 2 M^{pro}**	**Binding affinity with MAT1A**	**Preference index**
Tadalafil	-9.3	-8.5	1.8
S-paliperidone	-8.7	-9.4	2.4
dolutegravir	-8.6	-9.4	2.6
vardernafil	-8.3	-8.5	2.4

S-paliperidone exhibited the binding affinity of -8.7 Kcal/mol with the SARS CoV 2 M^{pro} compared to -8.0 kcal/mol of the reference inhibitor N3. Further, it showed a binding affinity of -9.4 kcal/ mol with MAT1A compared to binding strength of -7.9kcal/mol of native ligand SAM. Though, the preference index of 2.4 suggested in favour of its potency, acute toxicity issues regarding use in human (https://pubchem.ncbi.nlm.nih.gov/compound/Paliperidone#section=GHS-

Classification) overshadowed the use of this drug. Likewise, Dolutegravir with a binding energy of -8.6 kcal/mol against the SARS-CoV 2 M^{pro} (-8.0 kcal/mol of reference inhibitor) and a binding energy of -9. 4 kcal/mol (-7.9 kcal/mol of native ligand SAM) against MAT1A was also discarded following its poor performance in inhibiting the predicted mutated viral proteins in our studies despite its higher score in the preference index. Amongst the two remaining drug candidates, both tadalafil and vardenafil showed a binding affinity of -8.5kcal/mol (-7.9 kcal/mol of native ligand SAM) against MAT1A which is comparatively lower compared to that displayed by S-paliperidone and Dolutegravir. The preference score for Vardenafil is higher but there is significant difference in the inhibition of viral protein as displayed by molecular docking results where tadalafil exhibits higher binding energy with SARS-CoV M^{pro}. Further, from mutation analysis we did in our study, Tadalafil holds higher grounds in terms of inhibiting the mutated viral proteins. Regarding all these facts, tadalafil was selected as a principal lead and further studies were focused on Tadalafil.

4.6.3 Prospects of using Tadalafil in SARS-CoV 2 symptom management

Tadalafil, FDA approved drug, a selective inhibitor of phosphodiesterse type 5 (PDE5) with proven clinical efficacy and excellent safety against the treatment of pulmonary arterial hypertension (PAH) and erectile dysfunction (ED), has emerging prospects of being explored for the enhancement of cGMP mediated range of vascular and non-vascular conditions (Tzoumas *et. al.,* 2020). PDE5 inhibitors can inhibit neointimal formation and platelet aggregation via NO/cGMP/PKG pathway (Yang *et. al.*, 2019) which can be useful to manage the venous thromboembolism whose mitigation is most to reduce complications in SARS-CoV-2 according to World Health Organization (Isidori *et.al.*, 2020). A hypothesis can be posted such that the administration of Sildenafil as an adjunct to other medications can be helpful in SARS-CoV-2, as it counteracts the AngII-mediated downregulation of AT-1 receptor, acts on monocyte switching, thus reducing pro-inflammatory cytokines, interstitial infiltration and the vessel damage responsible for alveolar hemorrhage-necrosis and inhibits the transition of endothelial and smooth muscle cells to mesenchymal cells in the

pulmonary artery, preventing clotting and thrombotic complications (Isidori *et.a.l*, 2020).

Patients with previous cardiovascular conditions mainly pulmonary arterial hypertension seem to be at higher risk for developing severe forms of COVID-19 (Madjid *et.al.*, 2020). In a study conducted by Nuche *et. al*, a total of 7 patients out of 10 diagnosed with COVID-19 with previous cardiovascular condition (PAH) required hospitalization with mean length of stay of 10 days but none in intensive care unit. Five patients developed pneumonia with ARDS features seen in two patients and five patients needed oxygen therapy (Nuche *et.al.*, 2020). The low ACE2 levels in PAH patients could play as a protective factor at an initial infective stage blocking SARS-CoV 2 entrance and the use of phosphodiesterase-5 inhibitors (tadalafil and sildenafil) along with calcium-channel blockers based on their vasodilating properties can be proposed for managing the hypoxic vasoconstriction and ventilation/perfusion balance in COVID-19 patients (Nuche *et.al*, 2020).

Further, a phase 2 trial has been started in Chile using Sidenafil, another class of PDE5 inhibitor in COVID-19 treatment administered 25 mg every 8 hours orally for upto seven days in patients who show hypoperfusion of healthy lung areas in a substraction computed tomography angiography within 24 hours of admission to the hospital(https://clinicaltrials.gov/ct2/show/NCT04489446?term=sildenafil&cond=SARS-CoV+Infection&draw=2&rank=1). Another phase 3 trial (NCT04304313) has been started with primary outcome measure to study rate of disease remission in China with administration of Sildenafil 0.1g/day for 14 days in 10 COVID-19 patients(https://clinicaltrials.gov/ct2/show/NCT04304313?term=sildenafil&cond=Covid19&draw=2&rank=2). No human clinical trials were done using tadalafil against SARS CoV 2 infections but from our studies and relating to recent literatures, it is found that PDE5 inhibitors has promising prospects in terms of symptom management against the SARS CoV 2 infection. Thus, the repurposing of PDE5 inhibitors (tadalafil) for symptom management of SARS-CoV 2 infection can be cheap, easily available, and non-experimental treatment adjunct along with other medications to stop the COVID-19 from progressing to severe final stages, in which current approaches of treatment regrettably not always helpful.

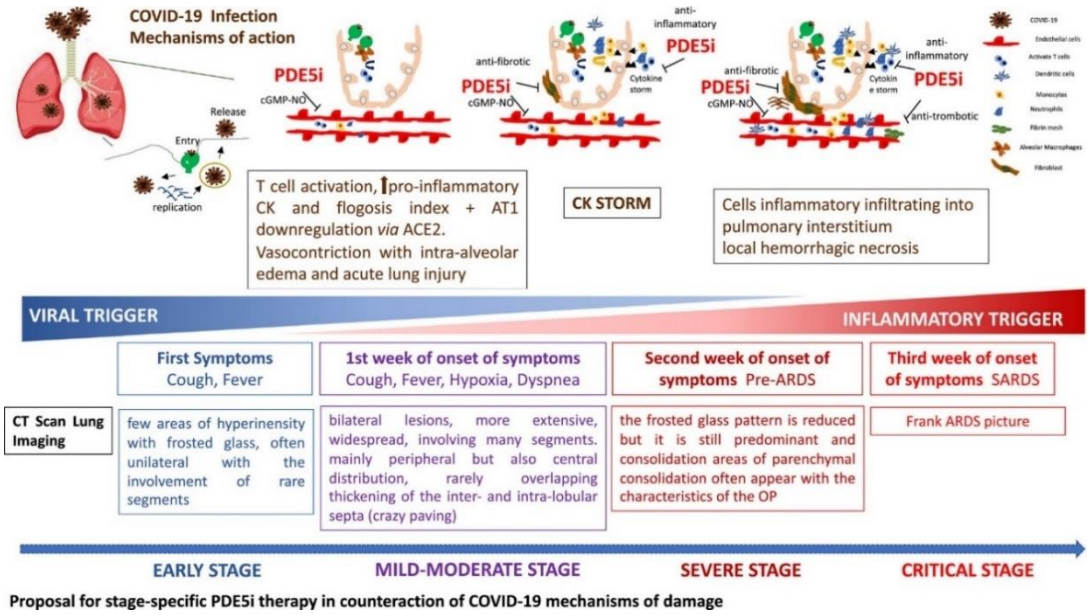


Figure 14: Proposal for PDE5i therapy in relation to COVID-19 stage specific progression. (doi: 10.1111/andr.12837)

4.7 DFT Analysis

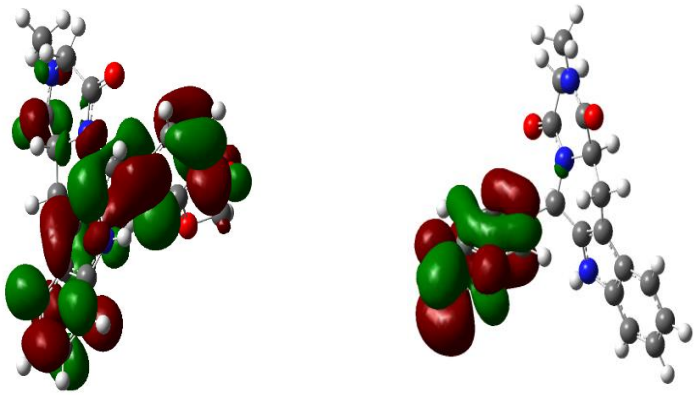
4.7.1 Dipole moment and total energy of the molecule

The dipole moment being widely used parameter, which has been shown to explain observable chemical and physical properties of molecules in many different contexts has application in drug discovery attracts high interest (Wojciechowski *et. al.,* 2014 and Ioakimidis *et. al.,* 2008). As compounds with large dipole moments are generally more soluble in water and less likely to be absorbed through lipophilic membranes, the dipole moment has been useful in the assessment of cell permeability and oral bioavailability of drugs (Ioakimdis *et. al.*, 2008 and Matuszek *et. al.,* 2016). The dipole moment for the selected compound tadalafil showed a value of 3.1156 Debye and this is in conjunction with the fact that drug like molecules have a dipole moment less than 10 Debye (Periera *et. al.,* 2018). Since dipole moment directly relates with cell permeability and oral biovailabilty of drugs, proper analysis of dipole moment can also act as proper descriptor for the estimation of drug loading capacities as well (Wu *et. al.,* 2016). Similarly, the molecule showed a total energy of -1314.03278294 a.u (in Hartree). The higher total energy of the molecule suggests the less stable and more

reactive nature of the molecule. Further, free energy is a pivotal criterion to represent the interactions of binding partners where both the sign and magnitude are important to express the likelihood of bimolecular events occurring and greater negative values indicate improved thermodynamic properties (Moniruzzaman et al, 2018). In this study, it is found that the values are negative meaning the binding will occur spontaneously without any extra energy expenditure.

4.7.2 Molecular orbital properties

Highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) were calculated using DFT analysis to assess their chemical reactivity. HOMO energy proposes the region of the small molecules, which can donate electron during the complex formation, while LUMO energy signifies the capacity of the molecule to accept the electrons from the partner protein (Banavath *et. al.,* 2014). The HOMO and LUMO calculated for tadalafil was -0.18311 and -0.0899 eV respectively and the HOMO-LUMO gap energy calculated was 0.09321 electron volts. Large HOMO-LUMO gap related to high kinetic stability and low chemical reactivity and small HOMO-LUMO gap is important for low chemical stability, because addition of electrons to a high-lying LUMO and/ or removal of electrons from a low-lying HOMO is energetically favorable in any potential reaction (Aihara *et. al.,* 1999). In case of our ligand, the low HOMO-LUMO gap signifies the higher chemical reactivity. Moreover, the difference in HOMO and LUMO energy, known as HOMO-LUMO gap energy, also signifies the electronic excitation energy that is necessary to compute the molecular reactivity and stability of the compounds (Zheng *et. al.,* 2013).



(a)HOMO *(b)LUMO*

Fig 15: Molecular Orbital Properties (a) HOMO and (b) LUMO

4.7 Drug-drug interaction studies

4.7.1 Molecular docking simulation

4.7.1.1 Selection of target human cytochromes involved in drug clearance

Among the different classes of cytochromes involved in drug clearance of tadalafil, PDE5 inhibitors including tadalafil are metabolized in human liver by the cytochrome P450 enzyme family with contributions from the CYP2C9, CYP2C19 and C2D6 pathways (Mehrotra et al, 2007). Thus, the X-ray diffraction structure of Cytochrome P450 3A4 with bound inhibitor bromoergocrptine (PDB id: 3UA1 with resolution 2.15 angstroms) (Sevriouka *et. al.,* 2012*)* and cytochrome P450 2D6 with bound inhibitor prinomastat (PDB id: 3QM4 with resolution 2.85 angstroms) (Wang *et. al.,* 2012)were further retrieved from Protein Data Bank (rcsb.docking.org) in PDB format amongst the various available crystal structures for studying the drug – drug interaction of tadalafil with other drugs. The reason for selecting these two cytochrome proteins was due to their importance in the metabolism of our principal target drug tadalafil. These crystal structures were selected due to their better resolution than other crystal structures deposited on the database.

4.7.1.2 Preparation of ligand database

The preparation of ligand database constitutes the basis for the success of any docking studies. For the molecular docking studies, the former ligand library of FDA compounds which was used to dock against the SARS-CoV M^{pro} was again utilized to dock against the crystal structure of the cytochromes CYP3A4 and CYP2D6. The same ligand library of drugs was considered because the library consisted of commonly approved drugs used in human and any drug inhibiting or interfering with the activity of cytochromes 3A4 and 2D6 from that drug pool could interfere with the clearance of tadalafil which might create serious complications in the patients. Although concomitant use of PDE5Is including tadalafil with nitrates, sGC stimulators and α-adrenoreceptor antagonists is contraindicated due to the risk of a potent hypotensive interaction (Webb *et. al.,* 1999), PDE5Is do not have synergistic effects on BP with other antihypertensive agents, such as ACEIs, ARBs, beta-blockers, CCBs, or thiazide diuretics (Kloner *et. al.,* 2003). There are various literatures which suggests the contraindications of different classes of drugs while using concomitantly with tadalafil, our sole purpose was to see if any drug from the selected ligand database could interfere with the clearance mechanism of tadalafil. Another ligand database was prepared from the group of chemicals which were known actives of PDE5 capable of interacting with PDE5 with which tadalafil also interacts, displayed when the smiles code of tadalafil was searched on online server (http://swisstargetprediction.ch/). A total of 288 molecules were prioritized, their 3D structure was downloaded, and their physiochemical parameters are drug ability parameters were set as previously discussed in section 4.3.6 of this study. After this, a final library of 64 compounds were further selected for docking studies with the cytochromes CYP3A4 and CYP2D6 X-ray crystal structures deposited on the database.

4.7.3 Virtual Screening of the ligand databases

4.7.3.1 Virtual Screening results against Cytochrome P450 3A4

Virtual screening was performed in Pyrx 0.9.8 platform (https://pyrx.sourceforge.io/) which is an an easy-to-use user interface, using AutoDock Vina. The target protein

Cytochrome P450 3A4 was docked to screen ligands using two different ligand libraries. Following the docking simulations with the ligand library which included FDA approved compounds against the Cytochrome P450 3A4, the reference inhibitor bromo-ergocrptine exhibited a binding affinity of -13.3 Kcal/mol. From a library of 1167 compounds, only one compound namely drospirenone exhibited a binding energy of -13.4kcal/mol. It is reported that the higher the negative value of binding affinity, the stronger is the binding of the ligand in the target (Dallakyan and Olson, 2015), thus the compounds showing binding energy greater in negative value than the reference inhibitor bromoergocryptine, may be capable of inhibiting the cytochrome in consideration which in turn might impede the drug metabolism process of tadalafil (Tzoumas *et. al.,* 2019).

Another ligand-based drug design study against the same cytochrome employing a different set of 64 compounds was done. All the compounds which were considered for molecular docking simulations yielded a lower binding strength than that of the reference inhibitor bromoergocryptine (-13.3kcal/mol) highlighting the absence of compounds in this dataset which could potentially interfere with the functioning of the cytochrome P450 3A4 families of enzyme.

Tadalafil is changed to desmethylene tadalafil (catechol metabolite) by CYP3A4 (Takahiro *et.al.*, 2015), the catechol metabolite undergoes extensive methylation to form methyl catechol metabolite which further undergoes glucoronidation reaction to form methyl catechol glucuronide conjugate (a major circulating metabolite 13000 times less potent that tadalafil for PDE5) (Forgue *et.al.*, 2006) which is then excreted through faeces (61%) and urine (36%). Thus, the significance of cytochrome p450 3A4 enzymes is immense and from our docking studies a compound called drosiperinone, a synthetic progestin which is an ingredient of oral contraceptive (https://pubchem.ncbi.nlm.nih.gov/compound/68873), is found to impede the drug efficacy of tadalafil. Further, studies have shown that drospirenone may increase the excretion rate of tadalafil which could result in a lower serum level and potentially a reduction in efficacy(https://go.drugbank.com/drugs/DB00820). Thus, from our findings it is suggested that the co-administration of these two drugs are not advisable if tadalafil is to be used as a drug for symptom management in SARS-CoV infection.

4.7.3.2 Virtual Screening results against Cytochrome P450 2D6

Virtual Screening against the cytochrome p450 2D6 was also done in the PyRx set up 0.9.8 employing Autodock tools using two ligand databases, one including the FDA compounds (ligand library 1) and other ligand library made up of compounds which could interact with Human PDE5 enzyme isoforms (ligand library 2). Following the docking interactions, the reference inhibitor primonastat exhibited a binding energy of -11kcal/mol. A total of 55 and 14 compounds from ligand library 1 and 2 respectively exhibited a greater binding affinity than the primonastat suggesting a requirement for thorough assessment of these compounds for their possible intervention in the efficacy of the tadalafil. However, the principal cytochrome involved in clearance of tadalafil being cytochrome P450 3A4 isoforms and the role of cytochrome P450 2D6 not so clear (Ring *et. al.,* 2004; Philips *et. al.,* 2004), the further studies regarding the interaction and their implications on performance of tadalafil was not addressed in this study.

5. SUMMARY

The *in-silico* studies to combat COVID-19 is today's necessity, using FDA-approved drugs as promising agents for viral inhibition or symptom management, which do not need much toxicity studies and could also serve as starting points for lead optimization in drug discovery. The world is in immediate need of novel treatment regimens to cope against the emerging threats of COVID-19, probable cause for numerous deaths around the globe. Computational techniques including molecular docking were employed to screen out the ligands from database collection of FDA compounds that could potentially act as competitive inhibitor against the target protein with proteolytic function under study. With this view, the present study was conducted taking 1157 compounds from FDA approved library and molecular docking was done against the target protein, SARS-CoV-2, main protease. 52 compounds yielded higher binding affinity than the reference compound N3 which upon pharmacokinetic properties and redocking were narrowed down to 4 compounds namely s-paliperidone, dolutegravir, tadalafil and vardenafil. Tadalafil was selected as the top hit based on its binding affinity, low hepatotoxicity, increased binding affinity on mutated SARS-CoV-2 main protease and our own preference index prepared on basis of important molecular descriptors. The thermodynamic properties and molecular orbital properties of the developed hit supported the drug like properties of the compound. Further, the drug-drug interaction of developed hit was studied with a view to find the probable FDA drugs which may impede the metabolism of our developed hit. With the use of structure based virtual screening, density function theory analysis, mutation prediction and efficacy of developed hits and drug-drug interaction studies, this study provides us the framework to develop the potential therapeutics for main protease of COVID-19 encompassing the different genres of computational chemistry.

6. CONCLUSION

Although FDA has issued an emergency use authorization on vaccines against COVID-19, drugs for anti-SARS-CoV-2 are still extremely important. The role of SARS-CoV-2 main protease in the polyprotein processing of a nascent viral polypeptide depicts it as a probable protein target. With the advent of computational power, opting to computer aided drug designing for pre-screening of compounds would mitigate the chances of those compounds failing in experimental setup. Based on the hypothesis that some FDA approved compound may be capable of covalent inhibition of SARS-CoV-2 M^{pro}, the molecular docking was performed. The FDA approved compounds were narrowed based on their ADME/T and virtual binding energy that exhibited higher binding efficiencies than reference inhibitor N3. Tadalafil was selected as the probable drug candidate despite its considerable hepatotoxicity. Moreover, the drug's use in pulmonary arterial hypertension suggests the ongoing potential use of drug in humans. In addition, the hypothesis set to study the thermodynamic properties and molecular orbital properties of developed hit, assessing the drug like properties of the hit, along with the drug-drug interaction study rendered the strong support for its prospects to start the lead optimization in drug discovery process. The workflow used in the study with certain modifications can be also helpful in computational part of drug discovery process in other disease models as well.

7. RECOMMENDATIONS

Thus, it is recommended that the interactions between the top hit (tadalafil) and main protease to be verified by carrying out the molecular dynamics (MD) simulations. This molecule can be further pursued as probable drug candidate for drug development. Enzyme inhibition kinetic studies along with animal testing and toxicity testing can be the way forward to strengthen the candidacy of this compound as potential suitor having therapeutic usage against SARS-CoV-2 M^{pro}.

8. BIBLIOGRAPHY

Alagaili, A. N., Briese, T., Mishra, N., Kapoor, V., Sameroff, S. C., Burbelo, P. D., de Wit, E., Munster, V. J., Hensley, L. E., Zalmout, I. S., Kapoor, A., Epstein, J. H., Karesh, W. B., Daszak, P., Mohammed, O. B., & Lipkin, W. I. (2014). Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *mBio, 5(2),* e00884-14. https://doi.org/10.1128/mBio.00884-14

Alsaadi ,E.. and Jones, I.,(2019) Membrane binding proteins of coronaviruses. *Futur Med [Internet].*

Apr 1 [cited 2020 Jun 10]; 14 (4), 275–86.

Amirian E. S. (2020). Potential fecal transmission of SARS-CoV-2: Current evidence and implications for public health. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases, 95, 363–370. https://doi.org/10.1016/j.ijid.2020.04.057

Anandakrishnan, R., Aguilar, B., & Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic acids research, 40(Web Server issue), W537–W541. https://doi.org/10.1093/nar/gks375

Athanasiadis, E., Cournia, Z., & Spyrou, G. (2012). ChemBioServer: a web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery. *Bioinformatics (Oxford, England), 28(22*), 3002–3003. https://doi.org/10.1093/bioinformatics/bts551

Azhar, E. I., Hui, D., Memish, Z. A., Drosten, C., & Zumla, A. (2019). The Middle East Respiratory Syndrome (MERS*). Infectious disease clinics of North America, 33(4),* 891–905. https://doi.org/10.1016/j.idc.2019.08.001

Bastola, A., Sah, R., Rodriguez-Morales, A. J., Lal, B. K., Jha, R., Ojha, H. C., Shrestha, B., Chu, D., Poon, L., Costello, A., Morita, K., & Pandey, B. D. (2020). The first 2019 novel coronavirus case in Nepal. The Lancet. *Infectious diseases, 20(3)*, 279–280. https://doi.org/10.1016/S1473-3099(20)30067-0

Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S., Angeletti, S., & Ciccozzi, M. (2020). The 2019-new coronavirus epidemic: *Evidence for virus evolution. Journal of medical virology, 92(4),* 455–459. https://doi.org/10.1002/jmv.25688

Buckland B. C. (2005). The process development challenge for a new vaccine. *Nature medicine*, *11*(4 Suppl), S16–S19. https://doi.org/10.1038/nm1218

Carpenter, K. A., Cohen, D. S., Jarrell, J. T., & Huang, X. (2018). Deep learning and virtual drug screening. *Future medicinal chemistry, 10(21),* 2557–2567. https://doi.org/10.4155/fmc-2018-0314

Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. *Vaccine, 25(39–40).* 6852–6862.

Carraturo, F., Del Giudice, C., Morelli, M., Cerullo, V., Libralato, G., Galdiero, E., & Guida, M. (2020). Persistence of SARS-CoV-2 in the environment and COVID-19 transmission risk from environmental matrices and surfaces. Environmental pollution *(Barking, Essex : 1987), 265(Pt B),* 115010. https://doi.org/10.1016/j.envpol.2020.115010

Cha, R. H., Joh, J. S., Jeong, I., Lee, J. Y., Shin, H. S., Kim, G., Kim, Y., & Critical Care Team of National Medical Center (2015). Renal Complications and Their Prognosis in Korean Patients with Middle East Respiratory Syndrome-Coronavirus from the Central MERS-CoV Designated Hospital. *Journal of Korean medical science, 30(12),* 1807–1814. https://doi.org/10.3346/jkms.2015.30.12.1807

Chan, J. F., Yuan, S., Kok, K. H., To, K. K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C., Poon, R. W., Tsoi, H. W., Lo, S. K., Chan, K. H., Poon, V. K., Chan, W. M., Ip, J. D., Cai, J. P., Cheng, V. C., Chen, H., Hui, C. K.,Yuen, K. Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet (London, England), 395(10223),* 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

Chan, J. F., Yuan, S., Kok, K. H., To, K. K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C., Poon, R. W., Tsoi, H. W., Lo, S. K., Chan, K. H., Poon, V. K., Chan, W. M., Ip, J. D., Cai, J. P., Cheng, V. C., Chen, H., Hui, C. K.,Yuen, K. Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-

to-person transmission: a study of a family cluster. *Lancet (London, England), 395(10223),* 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

Chan, J. F.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C., Poon, R. W., Tsoi, H. W., Lo, S. K., Chan, K. H. (2020). "Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan." *Emerging microbes & infections 9(1):* 221-236.

Chan-Yeung, M., & Xu, R. H. (2003). SARS: epidemiology. *Respirology (Carlton, Vic.), 8 Suppl(Suppl 1),* S9–S14. https://doi.org/10.1046/j.1440-1843.2003.00518.x

Chen J. (2020). Pathogenicity and transmissibility of 2019-nCoV-A quick overview and comparison with other emerging viruses. *Microbes and infection, 22(2),* 69–71. https://doi.org/10.1016/j.micinf.2020.01.004

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet (London, England), 395(10223),* 507–513. https://doi.org/10.1016/S0140-6736(20)30211-7

Chen, Y., Chen, L., Deng, Q., Zhang, G., Wu, K., Ni, L., Yang, Y., Liu, B., Wang, W., Wei, C., Yang, J., Ye, G., & Cheng, Z. (2020). The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *Journal of medical virology, 92(7),* 833–840. https://doi.org/10.1002/jmv.25825

Cornillez-Ty, C. T., et al. (2009). "Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling." Journal of virology 83(19): 10314-10318.

Craig, I. R., Essex, J. W., & Spiegel, K. (2010). Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. *Journal of chemical information and modeling, 50(4),* 511–524. https://doi.org/10.1021/ci900407c

Ding, B., Wang, J., Li, N., & Wang, W. (2013). Characterization of small molecule binding. I. Accurate identification of strong inhibitors in virtual

screening. *Journal of chemical information and modeling, 53(1),* 114–122. https://doi.org/10.1021/ci300508m

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., & Baker, N. A. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research, 35(Web Server issue),* W522–W525. https://doi.org/10.1093/nar/gkm276

Donigan, K. A., McLenigan, M. P., Yang, W., Goodman, M. F., & Woodgate, R. (2014). The Steric Gate of DNA Polymerase ι Regulates Ribonucleotide Incorporation and Deoxyribonucleotide Fidelity. *The Journal of Biological Chemistry, 289(13),* 9136. https://doi.org/10.1074/JBC.M113.545442

Douguet D. (2010). e-LEA3D: a computational-aided drug design web server. *Nucleic acids research*, *38*(Web Server issue), W615–W621. https://doi.org/10.1093/nar/gkq322

Dutkiewicz, Z., & Mikstacka, R. (2018). Structure-Based Drug Design for Cytochrome P450 Family 1 Inhibitors. *Bioinorganic chemistry and applications*, *2018*, 3924608. https://doi.org/10.1155/2018/3924608

Ella, K. M., & Mohan, V. K. (2020). Coronavirus Vaccine: Light at the End of the Tunnel. *Indian pediatrics*, *57*(5), 407–410. https://doi.org/10.1007/s13312-020-1812-z

Esakandari, H., Nabi-Afjadi, M., Fakkari-Afjadi, J., Farahmandian, N., Miresmaeili, S. M., & Bahreini, E. (2020). A comprehensive review of COVID-19 characteristics. *Biological procedures online, 22,* 19. https://doi.org/10.1186/s12575-020-00128-2

Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design, 15(5),* 411–428. https://doi.org/10.1023/a:1011115820450

Fan, Y., Zhao, K., Shi, Z. L., & Zhou, P. (2019). Bat Coronaviruses in China. *Viruses, 11(3),* 210. https://doi.org/10.3390/v11030210

Fehr, A. R., & Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Methods in molecular biology (Clifton, N.J.)*, *1282*, 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1

Finta, C., & Zaphiropoulos, P. G. (2000). The human cytochrome P450 3A locus. Gene evolution by capture of downstream exons. *Gene, 260(1-2),* 13–23. https://doi.org/10.1016/s0378-1119(00)00470-4

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry, 47(7),* 1739–1749. https://doi.org/10.1021/jm0306430

Fung, T. S., and Liu. D. (2018). "Post-translational modifications of coronavirus proteins: roles and function." *Future virology 13(6),* 405-430.

Goldsmith, C. S., Tatti, K. M., Ksiazek, T. G., Rollin, P. E., Comer, J. A., Lee, W. W., Rota, P. A., Bankamp, B., Bellini, W. J., & Zaki, S. R. (2004). Ultrastructural characterization of SARS coronavirus. *Emerging infectious diseases, 10(2),* 320–326. https://doi.org/10.3201/eid1002.030913

Groß, R., Conzelmann, C., Müller, J. A., Stenger, S., Steinhart, K., Kirchhoff, F., & Münch, J. (2020). Detection of SARS-CoV-2 in human breastmilk. *Lancet (London, England), 395(10239),* 1757–1758. https://doi.org/10.1016/S0140-6736(20)31181-8

Gross,E.K., Dreizler,R.M.(2013), Density Functional Theory, 337, *Springer Science & Business Media, 28(6), 4-8*

Guan, W,J., N,i Z,Y., Hu, Y, Liang W,H., Ou, C,Q., & He, J,X,.(2020) . Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S., & Poon, L. L. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science *(New York, N.Y.), 302(5643), 276–278.* https://doi.org/10.1126/science.1087139

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S., & Poon, L. L. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science *(New York, N.Y.), 302(5643),* 276–278. https://doi.org/10.1126/science.1087139

Guengerich F. P. (2021). Inhibition of Cytochrome P450 Enzymes by Drugs-Molecular Basis and Practical Applications. Biomolecules & therapeutics, *10.4062/biomolther.2021.102. Advance online publication.* https://doi.org/10.4062/biomolther.2021.102

Guido, R. V., Oliva, G., & Andricopulo, A. D. (2008). Virtual screening and its integration with modern drug design technologies. *Current medicinal chemistry, 15(1),* 37–46. https://doi.org/10.2174/092986708783330683

Hamza, A., Wei, N. N., & Zhan, C. G. (2012). Ligand-based virtual screening approach using a new scoring function. *Journal of chemical information and modeling, 52(4),* 963–974. https://doi.org/10.1021/ci200617d

Health Emergency Operation Center, Health Emergency and Disaster Management Unit (HEDMU), Ministry of Health and Population, Government of Nepal. Heoc.mohp.gov.np (2020). Resource materials on novel coronavirus (2019-nCoV) – health emergency operation Center. [online] Available at:https://heoc.mohp.gov.np/update-on-novel-corona-virus-2019_ncov/

Hellewell, J., et al. (2020). "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts." *The Lancet Global Health* 8(4): e488-e496.

Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S. I., Kim, L., Tong, S., Lu, X., Lindstrom, S., Pallansch, M. A., … Washington State 2019-nCoV Case Investigation Team (2020). First Case of 2019 Novel Coronavirus in the United States. *The New England journal of medicine, 382(10),* 929–936. **h**ttps://doi.org/10.1056/NEJMoa2001191

Honig, P. K., Wortham, D. C., Zamani, K., Conner, D. P., Mullin, J. C., & Cantilena, L. R. (1993). Terfenadine-ketoconazole interaction. Pharmacokinetic and electrocardiographic consequences. *JAMA, 269(12),* 1513–1518.

Houston, D. R., & Walkinshaw, M. D. (2013). Consensus docking: improving the reliability of docking in a virtual screening context. *Journal of chemical information and modeling, 53(2),* 384–390. https://doi.org/10.1021/ci300399w

Hu, B., Zeng, L. P., Yang, X. L., Ge, X. Y., Zhang, W., Li, B., Xie, J. Z., Shen, X. R., Zhang, Y. Z., Wang, N., Luo, D. S., Zheng, X. S., Wang, M. N., Daszak, P., Wang, L. F., Cui, J., & Shi, Z. L. (2017). Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS pathogens, 13(11), e1006698. https://doi.org/10.1371/journal.ppat.1006698

Huang, C., et al. (2020). "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." *The lancet 395(10223):* 497-506.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet (London, England), 395(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Huang, Q., et al. (2004). "Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein." *Biochemistry 43(20):* 6059-6063.

Hughes, J. D., Blagg, J., Price, D. A., Bailey, S., Decrescenzo, G. A., Devraj, R. V., Ellsworth, E., Fobian, Y. M., Gibbs, M. E., Gilles, R. W., Greene, N., Huang, E., Krieger-Burke, T., Loesel, J., Wager, T., Whiteley, L., & Zhang, Y. (2008). Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorganic & medicinal chemistry letters, 18(17),* 4872–4875**.** https://doi.org/10.1016/j.bmcl.2008.07.071

Hui, D. S., Memish, Z. A., & Zumla, A. (2014). Severe acute respiratory syndrome vs. the Middle East respiratory syndrome. *Current opinion in pulmonary medicine, 20(3),* 233–241. https://doi.org/10.1097/MCP.0000000000000046

Jegerschöld, C., Pawelzik, S. C., Purhonen, P., Bhakat, P., Gheorghe, K. R., Gyobu, N., Mitsuoka, K., Morgenstern, R., Jakobsson, P. J., & Hebert, H. (2008). Structural basis for induced formation of the inflammatory mediator prostaglandin E2. *Proceedings of the National Academy of Sciences of the United States of America, 105(32),* 11110–11115. https://doi.org/10.1073/pnas.0802894105

John, S., Thangapandian, S., Sakkiah, S., & Lee, K. W. (2011). Discovery of potential pancreatic cholesterol esterase inhibitors using pharmacophore modelling, virtual screening, and optimization studies. *Journal of enzyme inhibition and medicinal chemistry, 26(4),* 535–545. https://doi.org/10.3109/14756366.2010.535795

Kalid, O., Toledo Warshaviak, D., Shechter, S., Sherman, W., & Shacham, S. (2012). Consensus Induced Fit Docking (cIFD): methodology, validation, and application to the discovery of novel Crm1 inhibitors. *Journal of computer-aided molecular design, 26(11),* 1217–1228. https://doi.org/10.1007/s10822-012-9611-9

Kalliokoski, T., Salo, H. S., Lahtela-Kakkonen, M., & Poso, A. (2009). The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. *Journal of chemical information and modeling, 49(12),* 2742–2748. https://doi.org/10.1021/ci900364w

Kerns, E.H., Di L. (2009) Concepts, Structure Design and Methods from ADME to Toxicity Optimization. 1. *Academic Press*; Drug-like Properties

Khan, S. A., & Al-Balushi, K. (2021). Combating COVID-19: The role of drug repurposing and medicinal plants. *Journal of infection and public health*, *14*(4), 495–503. https://doi.org/10.1016/j.jiph.2020.10.012

Klompas, M., Baker, M. A., & Rhee, C. (2020). Airborne Transmission of SARS-CoV-2: Theoretical Considerations and Available Evidence. *JAMA,* 10.1001/jama.2020.12458. Advance online publication. https://doi.org/10.1001/jama.2020.12458

Korb, O., Olsson, T. S., Bowden, S. J., Hall, R. J., Verdonk, M. L., Liebeschuetz, J. W., & Cole, J. C. (2012). Potential and limitations of ensemble docking. *Journal of chemical information and modeling, 52(5),* 1262–1274. https://doi.org/10.1021/ci2005934

Köppen H. (2009). Virtual screening - what does it give us?. Current opinion in drug discovery & development, 12(3), 397–407.

Kubina, R., and A. Dziedzic (2020). "Molecular and serological tests for COVID-19. A comparative review of SARS-CoV-2 coronavirus laboratory and point-of-care diagnostics." *Diagnostics 10(6):* 434.

Kumar, R., Kumar, A., Långström, B., & Darreh-Shori, T. (2017). Discovery of novel choline acetyltransferase inhibitors using structure-based virtual screening. *Scientific reports, 7(1),* 16287. https://doi.org/10.1038/s41598-017-16033-w

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of molecular biology*, *161*(2), 269–288. https://doi.org/10.1016/0022-2836(82)90153-x

Laamarti, M., et al. (2020). "Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations." *PloS one 15(11):* e0240345.

Laha, S., et al. (2020). "Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission." *Infection, Genetics and Evolution 85*: 104445.

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

Lai, M. M., & Cavanagh, D. (1997). The molecular biology of coronaviruses. *Advances in virus research, 48,* 1–100. https://doi.org/10.1016/S0065-3527(08)60286-9

Lam T.Y., Shum M.H., Zhu H.C., Tong Y.G., Ni X.B., Liao Y.S. (2020) Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. *bioRxiv.* doi: https://doi.org/10.1101/2020.02.13.945485

Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Current medicinal chemistry, 20(23),* 2839–2860. https://doi.org/10.2174/09298673113209990001

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics, 10,* 168. https://doi.org/10.1186/1471-2105-10-168

Lee, S. J., Usmani, K. A., Chanas, B., Ghanayem, B., Xi, T., Hodgson, E., Mohrenweiser, H. W., & Goldstein, J. A. (2003). Genetic findings and functional studies of human CYP3A5 single nucleotide polymorphisms in different ethnic

groups. *Pharmacogenetics, 13(8),* 461–472. https://doi.org/10.1097/00008571-200308000-00004

Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. *Beilstein journal of organic chemistry, 12,* 2694–2718. https://doi.org/10.3762/bjoc.12.267

Li, H., Liu, S. M., Yu, X. H., Tang, S. L., & Tang, C. K. (2020). Coronavirus disease 2019 (COVID-19): current status and future perspectives. *International journal of antimicrobial agents, 55(5),* 105951. https://doi.org/10.1016/j.ijantimicag.2020.105951

Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins, 61(4),* 704–721. https://doi.org/10.1002/prot.20660

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S., & Wang, L. F. (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science (New York, N.Y.), 310(5748)*, 676–679. https://doi.org/10.1126/science.1118391

Lionta, E., Spyrou, G., Vassilatis, D. K., & 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

Lionta, E., Spyrou, G., Vassilatis, D. K., & 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

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews, 46(1-3), 3–26. https://doi.org/10.1016/s0169-409x(00)00129-0

Liu T., Hu J., Kang M., Lin L., Zhong H & Xiao J.(2020). Transmission dynamics of 2019 novel coronavirus (2019-nCoV) *bioRxiv* doi: 10.1101/2020.01.25.919787. 01.25.919787.

Liu, S., Alnammi, M., Ericksen, S. S., Voter, A. F., Ananiev, G. E., Keck, J. L., Hoffmann, F. M., Wildman, S. A., & Gitter, A. (2019). Practical Model Selection for Prospective *Virtual Screening. Journal of chemical information and modeling, 59(1),* 282–293. https://doi.org/10.1021/acs.jcim.8b00363

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J.and Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet (London, England), 395(10224),* 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

Luo, H., et al. (2005). "The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1." *FEBS letters 579(12):* 2623-2628.

Lurie, N., Saville, M., Hatchett, R., & Halton, J. (2020). Developing Covid-19 Vaccines at Pandemic Speed. *The New England journal of medicine, 382(21),* 1969–1973. https://doi.org/10.1056/NEJMp2005630

Lyons, D. M., & Lauring, A. S. (2017). Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. *Molecular Biology and Evolution, 34(12),* 3205. https://doi.org/10.1093/MOLBEV/MSX251

Macalino, S. J., Gosu, V., Hong, S., & Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research, 38(9),* 1686–1701. https://doi.org/10.1007/s12272-015-0640-5

Maia, E., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Frontiers in chemistry,* 8, 343. https://doi.org/10.3389/fchem.2020.00343

Malmstrom, R. D., & Watowich, S. J. (2011). Using free energy of binding calculations to improve the accuracy of virtual screening predictions. *Journal of chemical information and modeling,* 51(7), 1648–1655. https://doi.org/10.1021/ci200126v

Masters P. S. (2006). The molecular biology of coronaviruses. *Advances in virus research*, 66, 193–292. https://doi.org/10.1016/S0065-3527(06)66005-3

McBride, R., et al. (2014). "The coronavirus nucleocapsid is a multifunctional protein." *Viruses* 6(8): 2991-3018.

Mcfadden, J., & Al-Khalili, J. (1999). A quantum mechanical model of adaptive mutation. *BioSystems, 50,* 203–211.

Menchon, G., Maveyraud, L., & Czaplicki, G. (2018). Molecular Dynamics as a Tool for Virtual Ligand Screening. *Methods in molecular biology (Clifton, N.J.), 1762,* 145–178. https://doi.org/10.1007/978-1-4939-7756-7_9

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design, 7(2),* 146–157. https://doi.org/10.2174/157340911795677602

Michel, J., Tirado-Rives, J., & Jorgensen, W. L. (2009). Prediction of the water content in protein binding sites*. The journal of physical chemistry. B, 113(40),* 13337–13346. https://doi.org/10.1021/jp9047456

Middle East Respiratory Syndrome Coronavirus. Available at: https://www.who.int/emergencies/mers-cov/en/. Accessed 16 Feb 2020.

Morawska, L., & Milton, D. K. (2020). It Is Time to Address Airborne Transmission of Coronavirus Disease 2019 (COVID-19). Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 71(9), 2311–2313. https://doi.org/10.1093/cid/ciaa939

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry, 30(16),* 2785–2791. https://doi.org/10.1002/jcc.21256

Mukhra, R., Krishan, K., & Kanchan, T. (2020). Possible modes of transmission of Novel coronavirus SARS-CoV-2: a review. *Acta bio-medica : Atenei Parmensis, 91(3),* e2020036. https://doi.org/10.23750/abm.v91i3.10039

Munster, V. J., Koopmans, M., van Doremalen, N., van Riel, D., & de Wit, E. (2020). A Novel Coronavirus Emerging in China - Key Questions for Impact Assessment. *The New England journal of medicine, 382(8),* 692–694. https://doi.org/10.1056/NEJMp2000929

Naqvi, A. A. T., et al. (2020). "Insights into SARS-CoV-2 genome, structure, evolution,

pathogenesis and therapies: Structural genomics approach." Biochimica et Biophysica Acta
(BBA)-Molecular Basis of Disease 1866(10): 165878.

Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., Droese, B., Klaus, J. P., Makino, S., Sawicki, S. G., Siddell, S. G., Stamou, D. G., Wilson, I. A., Kuhn, P., & Buchmeier, M. J. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *Journal of structural biology, 174(1),* 11–22. https://doi.org/10.1016/j.jsb.2010.11.021

Ngan, C. H., Bohnuud, T., Mottarella, S. E., Beglov, D., Villar, E. A., Hall, D. R., Kozakov, D., & Vajda, S. (2012). FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic acids research*, *40*(Web Server issue), W271–W275. https://doi.org/10.1093/nar/gks441

Nunes, R. R., Fonseca, A., Pinto, A., Maia, E., Silva, A., Varotti, F. P., & Taranto, A. G. (2019). Brazilian malaria molecular targets (BraMMT): selected receptors for virtual high-throughput screening experiments. *Memorias do Instituto Oswaldo Cruz*, *114*, e180465. https://doi.org/10.1590/0074-02760180465

Osguthorpe, D. J., Sherman, W., & Hagler, A. T. (2012). Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chemical biology & drug design, 80(2),* 182–193. https://doi.org/10.1111/j.1747-0285.2012.01396.x

Patrì, A., Gallo, L., Guarino, M., & Fabbrocini, G. (2020). Sexual transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): A new possible route of infection?. *Journal of the American Academy of Dermatology, 82(6),* e227. https://doi.org/10.1016/j.jaad.2020.03.098

Peng, X., Xu, X., Li, Y., Cheng, L., Zhou, X., & Ren, B. (2020). Transmission routes of 2019-nCoV and controls in dental practice. *International journal of oral science, 12(1),* 9. https://doi.org/10.1038/s41368-020-0075-9

Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: are they closely related?. Clinical microbiology and infection : *the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 26(6),* 729–734. https://doi.org/10.1016/j.cmi.2020.03.026

Pinzi, L., & Rastelli, G. (2019). Molecular Docking: Shifting Paradigms in Drug Discovery. *International journal of molecular sciences, 20(18),* 4331. https://doi.org/10.3390/ijms20184331

Rahimi, A., et al. (2020). "Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection." *Genomics.*

Rai, B. K., Tawa, G. J., Katz, A. H., & Humblet, C. (2010). Modeling G protein-coupled receptors for structure-based drug discovery using low-frequency normal modes for refinement of homology models: application to H3 antagonists. *Proteins, 78(2),* 457–473. https://doi.org/10.1002/prot.22571

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, *261*(3), 470–489. https://doi.org/10.1006/jmbi.1996.0477

Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N., & Sastry, G. N. (2007). Virtual screening in drug discovery -- a computational perspective. *Current protein & peptide science*, *8*(4), 329–351. https://doi.org/10.2174/138920307781369427

Rochwerg, B., Agarwal, A., Siemieniuk, R. A., Agoritsas, T., Lamontagne, F., Askie, L., Lytvyn, L., Leo, Y. S., Macdonald, H., Zeng, L., Amin, W., Burhan, E., Bausch, F. J., Calfee, C. S., Cecconi, M., Chanda, D., Du, B., Geduld, H., Gee, P., Harley, N., … Vandvik, P. O. (2020). A living WHO guideline on drugs for covid-19. *BMJ (Clinical research ed.)*, *370*, m3379. https://doi.org/10.1136/bmj.m3379

Schneider, N., Hindle, S., Lange, G., Klein, R., Albrecht, J., Briem, H., Beyer, K., Claußen, H., Gastreich, M., Lemmen, C., & Rarey, M. (2012). Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. *Journal of computer-aided molecular design, 26(6),* 701–723. https://doi.org/10.1007/s10822-011-9531-0

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W., Jr (2013). Computational methods in drug discovery. *Pharmacological reviews, 66(1),* 334–395. https://doi.org/10.1124/pr.112.007336

Song, C. M., Lim, S. J., & Tong, J. C. (2009). Recent advances in computer-aided drug design. *Briefings in bioinformatics, 10(5),* 579–591. https://doi.org/10.1093/bib/bbp023

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design, 27(3),* 221–234. https://doi.org/10.1007/s10822-013-9644-8

Satarker, S. and M. Nampoothiri (2020). "Structural proteins in severe acute respiratory syndrome coronavirus-2." Archives of medical research 51(6): 482-491.

Scott, B. M., et al. (2021). "Predicted Coronavirus Nsp5 Protease Cleavage Sites in the Human Proteome: A Resource for SARS-CoV-2 Research." *bioRxiv.*

Seco, J., Luque, F. J., & Barril, X. (2009). Binding site detection and druggability index from first principles. *Journal of medicinal chemistry*, *52*(8), 2363–2371. https://doi.org/10.1021/jm801385d

Siddell, S., et al. (1983). "Coronaviridae." Intervirology 20(4): 181-189.

Senn, H. M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewandte Chemie (International ed. in English), 48(7),* 1198–1229. https://doi.org/10.1002/anie.200802019

Shao, W., Li, X., Goraya, M. U., Wang, S., & Chen, J.-L. (2017). Evolution of Influenza A Virus by Mutation and Re-Assortment. *International Journal of Molecular Sciences, 18(8),* 1650. https://doi.org/10.3390/ijms18081650

Shi, Z., & Hu, Z. (2008). A review of studies on animal reservoirs of the SARS coronavirus. *Virus research, 133(1),* 74–87. https://doi.org/10.1016/j.virusres.2007.03.012

Signorelli, C., Odone, A., Riccò, M., Bellini, L., Croci, R., Oradini-Alacreu, A., Fiacchini, D., & Burioni, R. (2020). Major sports events and the transmission of SARS-CoV-2: analysis of seven case-studies in Europe. Acta bio-medica *: Atenei Parmensis, 91(2),* 242–244. https://doi.org/10.23750/abm.v91i2.9699

Singhal T. (2020). A Review of Coronavirus Disease-2019 (COVID-19). *Indian journal of pediatrics, 87(4),* 281–286. https://doi.org/10.1007/s12098-020-03263-6

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J., & Gorbalenya, A. E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of molecular biology, 331(5),* 991–1004. https://doi.org/10.1016/s0022-2836(03)00865-9

Song, C. M., Bernardo, P. H., Chai, C. L., & Tong, J. C. (2009). CLEVER: pipeline for designing in silico chemical libraries. *Journal of molecular graphics & modelling, 27(5),* 578–583. https://doi.org/10.1016/j.jmgm.2008.09.009

Song, H. D., Tu, C. C., Zhang, G. W., Wang, S. Y., Zheng, K., Lei, L. C., Chen, Q. X., Gao, Y. W., Zhou, H. Q., Xiang, H., Zheng, H. J., Chern, S. W., Cheng, F., Pan, C. M., Xuan, H., Chen, S. J., Luo, H. M., Zhou, D. H., Liu, Y. F., He, J. F.,and Zhao, G. P. (2005). Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proceedings of the National Academy of Sciences of the United States of America, 102(7),* 2430–2435. https://doi.org/10.1073/pnas.0409608102

Staehler, M., Haseke, N., Khoder, W., & Stief, C. G. (2010). Profile of temsirolimus in the treatment of advanced renal cell carcinoma. *OncoTargets and therapy, 3,* 191–196. https://doi.org/10.2147/ott.s7657

Stoltzfus, A., & Norris, R. W. (2016). On the Causes of Evolutionary Transition:Transversion Bias. *Molecular Biology and Evolution, 33(3),* 595. https://doi.org/10.1093/MOLBEV/MSV274

Schrödinger's Mutations – theGIST. (n.d.). Retrieved September 11, 2021, from https://the-gist.org/2017/02/schrodingers-mutations/

Strippoli, P., Canaider, S., Noferini, F., D'Addabbo, P., Vitale, L., Facchin, F., Lenzi, L., Casadei, R., Carinci, P., Zannotti, M., & Frabetti, F. (2005). Uncertainty principle of genetic information in a living cell. *Theoretical Biology & Medical Modelling, 2,* 40. https://doi.org/10.1186/1742-4682-2-40

ten Brink, T., & Exner, T. E. (2010). pK(a) based protonation states and microspecies for protein-ligand docking. *Journal of computer-aided molecular design, 24(11),* 935–942. https://doi.org/10.1007/s10822-010-9385-x

The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. China CDC Wkly. 2020;*8*:113–

122**.** http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51

To K.W., Tsang O.Y., Chik-Yan. C. (2020) Consistent detection of 2019 novel coronavirus in saliva. *Clin Infect Dis.*s(5)4–6. https://doi.org/10.1093/cid/ciaa149

Tresadern, G., Bemporad, D., & Howe, T. (2009). A comparison of ligand based virtual screening methods and application to corticotropin releasing factor 1 receptor. *Journal of molecular graphics & modelling, 27(8),* 860–870. https://doi.org/10.1016/j.jmgm.2009.01.003

Tsoi, H., et al. (2014). "The SARS-coronavirus membrane protein induces apoptosis via interfering

with PDK1–PKB/Akt signalling." *Biochemical Journal 464(3),* 439-447.

van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., de Wit, E., & Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *The New England journal of medicine, 382(16),* 1564–1567. https://doi.org/10.1056/NEJMc2004973

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, *45*(12), 2615–2623. https://doi.org/10.1021/jm020017n

Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical medicine & international health: TM & IH*, *25*(3), 278–280. https://doi.org/10.1111/tmi.13383

V'kovski, P., et al. (2019). "Determination of host proteins composing the microenvironment of

coronavirus replicase complexes by proximity-labeling." *Elife* **8**: e42037.

Vreven, T., Byun, K. S., Komáromi, I., Dapprich, S., Montgomery, J. A., Morokuma, K., & Frisch, M. J. (2006). Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *Journal of chemical theory and computation, 2(3).* 815–826. https://doi.org/10.1021/ct050289g

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA, 323(11),* 1061–1069. https://doi.org/10.1001/jama.2020.1585

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of virology, 94(7),* e00127-20. https://doi.org/10.1128/JVI.00127-20

Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *Lancet (London, England), 395(10223),* 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9

Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Liu, Y. F., Guo, R. T., Liu, X. N., Zhan, L. H., Zhou, D. H., Xu, J. (2005). SARS-CoV infection in a restaurant from palm civet. *Emerging infectious diseases, 11(12),* 1860–1865. https://doi.org/10.3201/eid1112.041293

Wang, Y. and L. Liu (2016). "The membrane protein of severe acute respiratory syndrome coronavirus functions as a novel cytosolic pathogen-associated molecular pattern to promote beta interferon induction via a Toll-like-receptor-related TRAF3-independent mechanism." *MBio 7(1):* e01872-01815.

Warnes, S. L., Little, Z. R., & Keevil, C. W. (2015). Human Coronavirus 229E Remains Infectious on Common Touch Surface *Materials. mBio, 6(6),* e01697-15. https://doi.org/10.1128/mBio.01697-15

Waszkowycz B. (2008). Towards improving compound selection in structure-based virtual screening. *Drug discovery today, 13(5-6),* 219–226. https://doi.org/10.1016/j.drudis.2007.12.002

WHO Solidarity Trial Consortium, Pan, H., Peto, R., Henao-Restrepo, A. M., Preziosi, M. P., Sathiyamoorthy, V., Abdool Karim, Q., Alejandria, M. M., Hernández García, C., Kieny, M. P., Malekzadeh, R., Murthy, S., Reddy, K. S., Roses Periago, M., Abi Hanna, P., Ader, F., Al-Bader, A. M., Alhasawi, A., Allum, E., Alotaibi, A., Swaminathan, S. (2021). Repurposed Antiviral Drugs for Covid-19 - Interim

WHO Solidarity Trial Results. *The New England journal of medicine, 384(6),* 497–511. https://doi.org/10.1056/NEJMoa2023184

WHO. Clinical management of severe acute respiratory infection when Novel coronavirus (nCoV) infection is suspected: interim guidance. https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected (Accessed on 28 Feb 2020) 2020Wu, A., et al. (2020). "Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China." *Cell host & microbe 27(3):* 325-328.

Woo, P. C., Huang, Y., Lau, S. K., & Yuen, K. Y. (2010). Coronavirus genomics and bioinformatics analysis. *Viruses, 2(8), 1804–1820.* https://doi.org/10.3390/v2081803

World Health Organization. Situation reports. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

Wouters, O. J., Shadlen, K. C., Salcher-Konrad, M., Pollard, A. J., Larson, H. J., Teerawattananon, Y., & Jit, M. (2021). Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. *Lancet (London, England)*. *397*(10278), 1023–1034. https://doi.org/10.1016/S0140-6736(21)00306-8

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 (New York, N.Y.), 367(6483),* 1260–1263. https://doi.org/10.1126/science.abb2507

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet (London, England), 395(10225),* 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9

Wu, C., Chen, X., Cai, Y., Xia, J., Zhou, X., Xu, S., Huang, H., Zhang, L., Zhou, X., Du, C., Zhang, Y., Song, J., Wang, S., Chao, Y., Yang, Z., Xu, J., Zhou, X., Chen, D., Xiong, W., Xu, L., … Song, Y. (2020). Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019

Pneumonia in Wuhan, China. *JAMA internal medicine*, *180*(7), 934–943.https://doi.org/10.1001/jamainternmed.2020.0994

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., & Jiang, T. (2020). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell host & microbe, 27(3),* 325–328**.** https://doi.org/10.1016/j.chom.2020.02.001

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., Li, T., & Chen, Q. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *International journal of oral science, 12(1),* 8. https://doi.org/10.1038/s41368-020-0074-x

Yadav, R., et al. (2021). "Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19." Cells 10(4): 821.

Ye, Y. and B. G. Hogue (2007). "Role of the coronavirus E viroporin protein transmembrane domain in virus assembly." *Journal of virology 81(7):* 3597-3607.

Young, T., Abel, R., Kim, B., Berne, B. J., & Friesner, R. A. (2007). Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. Proceedings of the *National Academy of Sciences of the United States of America, 104(3),* 808–813. https://doi.org/10.1073/pnas.0610202104

Zhang, M. Q., & Wilkinson, B. (2007). Drug discovery beyond the 'rule-of-five'*. Current opinion in biotechnology,* 18(6), 478–488. https://doi.org/10.1016/j.copbio.2007.10.005

Zheng J. (2020). SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. *International journal of biological sciences, 16(10),* 1678–1685. https://doi.org/10.7150/ijbs.45053

Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y., Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris, & Guan, Y. (2003). Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet (London, England), 362(9393),* 1353–1358. https://doi.org/10.1016/s0140-6736(03)14630-2

Zou, X., Chen, K., Zou, J., Han, P., Hao, J., & Han, Z. (2020). Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Frontiers of medicine, 14(2),* 185–192. https://doi.org/10.1007/s11684-020-0754-0

9.APPENDICES

9.1 Data sources for metabolic model reconstruction and refinement

9.1.1 DNA sequence and genome annotation databases

EMBL	http://www.ebi.ac.uk/embl/	General nucleotide sequence database
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/	General nucleotide sequence database
Integr8	http://www.ebi.ac.uk/integr8/	Integrated information on complete genomes
CMR	http://cmr.jcvi.org/	Integrated information on complete prokaryotic genomes
IMG	http://img.jgi.doe.gov/	Integrated system for analysis and annotation of microbial genomes
SEED	http://seed-viewer.theseed.org/	Integrated system for analysis and annotation of genomes using functional subsystems

9.1.2 Protein and enzyme databases

ENZYME	http://www.expasy.ch/enzyme/	Enzyme nomenclature database providing extensive information on all enzymes with an associated EC number
UniProt	http://www.ebi.ac.uk/uniprot/	Universal Protein Resource gathering protein sequences and annotations from SwissProt

		(manually reviewed), trEMBL (computer annotated), and PIR
TransportDB	http://www.membranetransport.org/	Predictions of membrane transport proteins for fully sequenced genomes
PSORTdb	http://db.psort.org/	Repository of experimentally determined and predicted protein localizations
Prolinks	http://prolinks.mbi.ucla.edu/	Database of predicted functional links between proteins
STRING	http://string.embl.de/	Database of known and predicted protein–protein interactions

9.1.3 Metabolic databases

CheBI	http://www.ebi.ac.uk/chebi/	Database on small molecules of biological interest
Pubchem	http://pubchem.ncbi.nlm.nih.gov/	Database on small molecules
LipidMaps	http://www.lipidmaps.org/	Database on lipid metabolites
Reactome	http://www.reactome.org/	Curated database of biological pathways
KEGG	http://www.genome.jp/kegg/	Suite of databases comprising information on compounds, reactions, pathways, genes/proteins
BioCyc	http://www.biocyc.org/	Collection of organism-specific pathway/genome databases, including a curated multiorganism pathway database: MetaCyc

UniPathway	http://www.grenoble.prabi.fr/obiwarehouse/unipathway/	Curated resource of metabolic pathways linked to UniProt enzyme database
UM-BBD	http://umbbd.msi.umn.edu/	Database on microbial biocatalytic reactions and biodegradation pathways

9.1.4 Experimental data repositories

IntAct	http://www.ebi.ac.uk/intact/	Repository of reported protein interactions
DIP	http://dip.doe-mbi.ucla.edu/	Database of experimentally determined interactions between proteins
Array Express	http://www.ebi.ac.uk/aerep/	Public repository of microarray data
GEO	http://www.ncbi.nlm.nih.gov/geo/	Public repository of microarray data
ASAP	http://asap.ahabs.wisc.edu/	Repository of results of functional genomics experiments for selected bacterial species
E. coli multi-omics DB	http://ecoli.iab.keio.ac.jp/	Comprehensive dataset of transcriptomic, proteomic, metabolomic, and fluxomic experiments for *E. coli* K12
Systomonas	http://www.systomonas.de/	Repository of 'omics' datasets and molecular networks for Pseudomonads species
PubMed	http://www.pubmed.org/	Database on biomedical literature

9.1.5 Metabolic Model Repositories

BiGG	http://bigg.ucsd.edu/	Repository of reconstructed genome-scale metabolic models
BioModels	http://www.ebi.ac.uk/biomodels/	Database of mathematical models of biological systems

9.2 Results from OSIRIS property explorer for those passing ADME/Tox filters

Pubchem Ids	Total MW	cLogP	cLogS	H-Acceptor	H-Donor	Polar Surface Area	Druglikeness
444031	426.585	3.9043	-3.99	4	1	55.66	3.7859
54726191	419.383	0.128	-3.911	8	2	99.18	5.5879
82153	434.502	1.3033	-2.579	6	2	93.06	0.79255
9823781	426.490	2.6847	-3.712	7	1	82.17	7.1426
110635	389.410	2.396	-3.922	7	1	74.87	6.7598
135400189	488.611	2.2165	-3.61	10	1	117.51	6.4659

9.3 List of target proteins with sources

Proteins	Pdb Ids	Ligand	Organisms	References
SARS-CoV-2 M^{pro}	6LU7	N3	SARS-CoV-2	Jin *et.al*
Human MAT1A	6SW5	SAM	human	Rehse *et.al*
Human cytochrome P450 3A4	3UA1	bromoergocryptine	Human	Sevriouka *et.al*
Human cytochrome P450 2D6	3QM4	Prinomastat	Human	Wang *et.al*