Insilico Drug Development Against RNA Dependent RNA Polymerase of SARS-CoV-2

M.Sc. Thesis

2023

Submitted to

Central Department of Biotechnology

Tribhuvan University

Kirtipur, Kathmandu, Nepal

For partial fulfillment of M.Sc degree in Biotechnology

By

Siddartha Gautam

Registration No: 5-3-28-300-2017

Supervisors

Senior Scientist Dr. Pramod Aryal

Asst. Prof. Alina Shri Sapkota

ACKNOWLEDGEMENT

I would like to extend my heartfelt appreciation and gratitude to Dr. Pramod Aryal, my supervisor and mentor, for his invaluable guidance and unwavering support throughout my thesis period. His continuous dedication and expertise have been instrumental in steering me in the right direction whenever I encountered challenges. Under his tutelage, I gained a deeper understanding of the true meaning and significance of science.

I would like to express my gratitude to Prof. Dr. Rajani Malla for her care and motivation which have been instrumental in keeping me focused on the path towards success. Her guidance and advice have been invaluable in helping me overcome challenges, and her generosity in supporting the research has been greatly appreciated.

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. Samiran Subedi 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 Richa Giri for providing me continued help and support throughout all these years and through the process of research works and writing this thesis.

Siddartha Gautam

Reg no: 5-3-28-300-2017

LIST OF ABBREVIATIONS:

AA	Amino Acid
ACE	Angiotensin-converting enzyme
Ac	Acidic
ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
ADRP	ADP-ribose-1'-phosphate
ARDS	Acute Respiratory Distress Syndrome
CADD	Computer Aided Drug Design
CoV	Corona virus
COVID	Corona virus disease
DMV	Double membrane vesicle
EKG	Electrocardiogram
ER	Endoplasmic reticulum
ERGIC	Endoplasmic reticulum Golgi intermediate compartment
ExoN	Exoribonuclease
FDA	Food and Drug Administration
GTP	Guanosine triphosphate
HCoV	Human Corona virus
kb	Kilobase
kDa	Kilodaltons
kJ	Kilojoules

LBVS	Ligand-Based Virtual Screening
MERS	Middle East respiratory syndrome
MD	Molecular Dynamics
MDA	Melanoma differentiation associated protein
Mpro	Main protease
MRI	Magnetic resonance imaging
MTase	Methyltransferase
NAB	Nucleic acid binding
NCBI	National Center for Biotechnology Information
NiRAN	Nidovirus RdRp associated nucleotidyl transferase domain
NMR	Nuclear magnetic resonance
Nsp	Non-structural protein
ORF	Open reading frame
PI	Preference Index
PLPro	Papain-like protease
PDB	Protein Database
QSAR	Quantitative structure-activity relationships
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RTC	Replication/Transcription Complexes
RTP	Remdesivir triphosphate
SAM	S-Adenosyl Methionine
SARS	Severe acute respiratory syndrome

SBVS	Structure Based Virtual Screening
sg	Sub-genomic
SUD	SARS-Unique domain
TMPRSS	Transmembrane serine protease
Tox	Toxicity
TPSA	Topological Polar Surface Area
TRS	Transcriptional Regulatory Sequences
Ubl	Ubiquitin-like
USD	United Sates dollar
UTR	Untranslated region
VHTS	Virtual High Throughput Screening
WHO	World Health Organization

LIST OF TABLES:

S.No	Table number	Page number	Name of the table
1	1		Functions of nsps in coronaviruses
2	2		Various stages of clinical manifestation of disease in human.
3	3		Parameters of pharmacokinetic properties for screening of compounds.
4	4		Binding energy of potential hits against SARS-CoV-2 RdRp
5	5		Binding energy of potential hits against target compounds.
6	6		Amino acid residues within 5 Å of compound 17
7	7		Amino acid residues within 5 Å of compound 1016
8	8		Amino acid residues within 5 Å of compound LAS34154490
9	9		Types of interaction of compound 17 with active site amino residues of SARS-CoV-2 RdRp and bond distance data
10	10		Types of interaction of compound 1016 with active site amino residues of SARS-CoV-2 RdRp and bond distance data
11	11		Types of interaction of LAS34154490 with active site amino residues of SARS-CoV-2 RdRp and bond distance data
12	12		Binding energy against P323L RdRp
13	13		Binding energy of lead compounds with hMAT1A

LIST OF FIGURES:

S.No	Table number	Page number	Name of the table
1	1		Genomic organizations of representative α, β and γ CoVs
2	2		Virion structure of SARS-CoV-2 comprising structural proteins and viral RNA
3	3		Genomes of SARS-CoV, MERS-CoV and SARS-CoV-2.
4	4		SARS-CoV-2 life cycle in host cell.
5	5		LBVS. The chart shows various methods utilized in LBVS.
6	6		SBVS. General schematic of structure based virtual screening.
7	7		6M71|SARS-Cov-2 RNA-dependent RNA polymerase in complex with cofactors
8	8		The Z score plots and energy plot for predicting 3D structure of a protein based on amino acids sequences a) Z-Score plot of 6M71 predicted by proSA with structures available in database b) energy plot of 6M71 predicted by ProSA
9	9		Ramachandran plot of SARS-CoV-2 RdRp electron microscope structure.
10	10		Amino acid residue with 5 Å of GTP bonded to 7DIY; blue colored molecule is native ligand, and the sphere corresponds to magnesium ion.
11	11		Receptor ligand complex of comp 17 shown as sticks with SARS-CoV-2 RdRp depicted by surface and amino acids residue of RdRp shown as lines within 5 Å of compound 17.
12	12		Receptor ligand complex of comp 1016 shown as sticks with SARS-CoV-2 RdRp depicted by surface and amino acids residue of RdRp shown as lines within 5 Å of compound 1016.

13	13		Receptor ligand complex of comp LAS34154490 shown as sticks with SARS-CoV-2 RdRp depicted by surface and amino acids residue of RdRp shown as lines within 5 Å of compound LAS34154490.
14	14		Receptor ligand interaction and hydrophobic interaction of compound 17 with SARS-CoV-2 RdRp imaged from Discovery Studio Visualizer.
15	15		Receptor ligand interaction and hydrophobic interaction of compound 1016 with SARS-CoV-2 RdRp imaged from Discovery Studio Visualizer.
16	16		Receptor ligand interaction and hydrophobic interaction of LAS34154490 with SARS-CoV-2 RdRp imaged from Discovery Studio Visualizer.

Table of Contents

ACKNOWLEDGEMENT ..... **ii**
LIST OF ABBREVIATIONS ..... **iii**
LIST OF TABLES ..... **vi**
LIST OF FIGURES: ..... **vii**
ABSTRACT ..... **xii**
INTRODUCTION: **1**
Background 1
Current Studies 3
Hypothesis 4
Null Hypothesis 4
Alternative Hypothesis 4
Objective 4
General Objective 4s
Specific Objective 4
Rationale 5
Scope of study 5
LITERATURE REVIEW **6**
Coronaviruses 6
Introduction 6
Genomic Organization 7
SARS-CoV-2 8
Virion structure 9
Genome Organization of SARS-CoV-2 11
SARS-CoV-2 infection cycle 12
Attachment and Entry 13
Expression of Replicase protein 13
RdRp complex 15
Replication and Transcription 16
Assembly and Release 17
Understanding the global pandemic 17
Transmission 18
Clinical manifestations of SARS CoV-2 19
RdRp Mutation 21

Introduction to Computer-aided drug design (CADD) 22
Virtual screening 22
Ligand based virtual screening 23
Structure based virtual screening (SBVS 25
Attaining Protein structure and their preparation 26
Identification of Binding site 26
Compound database preparation 27
Docking and Scoring 28
Improving Pose/Compound Selection After Docking 29
METHODS AND METHODOLOGY 30
Target identification and Retreival of 3D protein structure 30
Protein structure Validation 30
Target protein Preparation and Acquisition of Binding sites 30
Compound database preparation 31
Scope of library 31
Obtaining molecular structures 31
Calculating molecular properties 31
Generating energy-minimized conformations 32
Preparing input files 32
Molecular docking Studies 32
Docking with 6M71 32
Redocking of Potential hits 33
Screening with ExoN, RTP and PI 33
Analysis of Protein-ligand interaction 33
Receptor-ligand complex visualization and Prediction of amino acids responsible for receptor-ligand interaction 33
Analysis of interactions responsible for receptor/ligand binding 34
Docking with mutated RdRp 34
Interaction with human proteins 34
RESULTS AND DISCUSSION 35
Selection of Protein Target 35
Target Protein structure validation 37
Z-score and Energy plot 37
Ramachandran plot 39
Target protein preparation and Acquisition of Binding sites 40
Compound database preparation 41

Molecular docking studies 44
Docking with 6M71 44
Redocking of potential hits 44
Screening with ExoN, RTP and PI 45
Analysis of protein-ligand interaction 48
Receptor-ligand complex visualization and Prediction of amino acids responsible for receptor-ligand interaction 48
Analysis of interactions responsible for receptor/ligand binding 52
Interaction with mutated RdRp (P323L) 59
Interaction with human protein hMAT1A 60
SUMMARY 61
CONCLUSION 62
RECOMMENDATIONS 62
REFERENCES 63

ABSTRACT:

The outbreak of SARS-CoV-2 was one of the major global health concern and quest for potential therapeutic options is still important due to ever evolving nature of the virus. Due to the sensitive nature of handling live virus, the in silico tools and resources, collectively termed as CADD (Computer Aided Drug Design) would be an alternative approach for high-throughput screening of probable drug molecules based on the molecules activity, selectivity, and pharmacokinetic properties. The RNA dependent RNA polymerase (RdRp) of SARS-CoV-2 is a crucial enzyme involved in viral replication, making it an attractive target for drug development. In this study, we conducted molecular docking studies of library of FDA-approved compounds generated by screening the molecules using strict druggability, ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) criteria, non-interference with human hepatic s-adenosylmethionine (SAM) biosynthesis and resistance against viral proofreading activity for determining potential inhibitors of RdRp. Our results identified two molecules, compound 17 and LAS34154490 as potential inhibitors of SARS-CoV-2 RdRp, which belong to class preplated microcyclics and nucleoside mimetics, respectively. Our results suggest that these drugs have strong binding affinity and stable interactions with the active site of RdRp. In conclusion, our findings highlight the potential of drug repurposing as a strategy for identifying inhibitors of RdRp for rapid safety test as these are FDA approved and may provide valuable option for the treatment of COVID-19.

Keywords: CADD, RdRp, ADME/Toxicity, molecular docking, preplated microcyclics, kinase inhibitor, nucleoside mimetics, COVID-19

INTRODUCTION:

Background:

According to WHO, more than 278 million cases and approximately 5.4 million deaths have been reported globally till 26 December 2021, due to COVID 19 infection. (https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---28-december-2021). Current condition reflects the massive spread of disease and various measure such as quarantine and lockdown are being implemented in order to mitigate the spread within population (Kucharski *et. al.,* 2020). However, lockdown and quarantine have further posed severe problems in other sectors such as transport and economy (Sharif *et. al.,* 2020, Guan *et. al.,* 2020). Thus, it is necessary to search for countermeasures which could help us to tackle the problem posed by the pandemic.

Development of effective vaccine is a complex, lengthy and costly process. From concept to getting license, vaccine development is a multistep process (Leroux-Roels *et. al.,* 2011). Additionally, cost estimated for a vaccine development is high with reports suggesting 200- 500 million USD and can take approximately 15 years or more (Plotkin *et. al.,* 2017) Currently available vaccines are not for long term and recent data suggests that they provide protection against major infection and death up to 6 months (Nordstrom *et. al.,* 2022). Furthermore, drug development is another sector which can help alleviate the threat posed by covid 19 infection. However, cost of developing new drug is estimated from USD 92 million cash (USD 161 million capitalized) to USD 883.6 million cash (USD 1.8 billion capitalized) (Morgan *et. al.,* 2011). In order to minimize the constraints of time and cost, an easy and effective technology of drug discovery is required.

Combination of computational methods for drug discovery and development with *in vitro* experimental methods will contribute hugely to overcome these constraints (Sliwoski *et. al.,* 2014). With meteoric progress in field of computer technologies, drug screening and design has massively benefitted by significantly reducing cost and time of drug development (Lin *et. al.,* 2020). Basically, along with previous genomic and

proteomic experimental data and bioinformatic tools, key genes and their translated protein can be identified which can be targeted for drug screening and design (Yamanishi *et. al.,* 2008) (Bakheet *et al.,* 2009). In order to reinforce the experiments, protein structure predictions algorithms can aid to construct near precise protein structures (Moult *et al.,* 2018). Molecular simulation along with multiscale models allows examination of biological processes of target proteins and provide efficient data on structure and dynamics governing those processes in varying degree of resolution (Ayton *et. al.,* 2007). This data along with x-ray crystallography studies can be useful to predict binding site of drug and the probable drug action mechanism (Handing *et. al.,* 2018). After this, screening of various ligands can be done through chemical libraries (Shoichet, B.K., 2004) and thus screened putative drug candidates can be docked onto target protein using bioinformatic tools such as Autodocktools, Autodock Vina (Forli *et. al.,* 2016). Aside from virtual screening of putative drug candidates, repurposing the existing drugs is of great interest (Ashburn & Thor, 2004). Many drug-like molecules and antivirals are being investigated to target SARS CoV2 RNA dependent RNA polymerase (RdRp) with the help of computer aided drug discovery (CADD) approach (Baby *et. al.,* 2020).

The foremost process in CADD approach is to identify the biomolecular target which do not share homology to any human proteins or other biomolecule as those drugs may target against the human protein as well. Biological processes such as SARS CoV2 protease activity (Ulrich and Nitsche, 2020), viral structural assembly and immune regulation in host (Peng *et. al.,* 2020) and replication process (Wu *et. al.,* 2020) are being considered as a template target to develop new leads. Viral replication is essential biological function for survival and amplification of virus in host and thus can be utilized as a target to discover potential therapeutic agents using computational approach (Wu *et. al.,* 2020).

In order to express and replicate unusually large genome (two to three-fold larger) compared to other viruses, Coronaviruses utilizes numerous RNA-synthesizing and RNA-processing enzymes. RNA-dependent RNA polymerase (RdRp) which functions as a main enzyme for viral replication and transcription, synthesizes all viral RNA essential for survival of virus in host and is proven to be a main target for several

antiviral drugs (Malone *et. al.,* 2022). Thus, by using existing databases of drug molecules and further screening them to obtain potential leads, we can target the RdRp protein with these potential leads to determine therapeutic agents posing threat to survival of SARS CoV-2 in human host.

Currently, Various molecules such as remdesivir, favipiravir, sofosbuvir (Elfiky, 2021), Plant-derived natural polyphenols theaflavin, hesperidin (Singh *et. al.,* 2021) have been identified virtually as lead antiviral drugs against SARS CoV-2 RdRp inhibition. Besides this, researches for identification of other potential drug candidates are ongoing globally. These drug candidates were determined via insilico approaches which includes array of various processes. As initial screening, the drug candidates must fulfill the lipinski's Rule of Five. Upon meeting the requirements of screening parameters, these molecules were additionally subjected to molecular docking process where binding energy of these molecules to the active site of target was calculated. The binding energy of these molecules were compared to native ligands of protein and molecules with relatively higher binding energy were further considered for additional insilico approaches. Furthermore, various intrinsic properties such as H-bond donor and acceptor, total polar surface area, rotatable bond count of molecule were taken as consideration to empirically calculate residence time of the molecule in active site of the drug.

Current Studies:

Drug repurposing, the use of existing drugs to treat new indications, has gained significant attention in the search for effective treatments for SARS-CoV-2. One potential target is the RNA dependent RNA polymerase (RdRp) of the virus, which is crucial for replication. Several studies have explored the repurposing of drugs such as remdesivir, favipiravir, and ribavirin for use against SARS-CoV-2 by targeting RdRp. For instance, a study reported the potential of remdesivir as an RdRp inhibitor (Shannon *et. al.,* 2020), while another study investigated the potential of favipiravir, ribavirin and sofosbuvir in targeting RdRp (Elfiky, 2021). Additionally, EIDD-2801, a ribonucleoside analog prodrug, has also been explored as a RdRp inhibitor. EIDD-2801

showed broad-spectrum antiviral activity against several coronaviruses, including SARS-CoV-2, through targeting RdRp (Wahl *et. al.,* 2020).

In addition to these drugs, researchers have also explored the potential of natural products as RdRp inhibitors for SARS-CoV-2. For example, molecular docking of compounds from natural products, such as ginger, garlic, and curcumin, with RdRp showed that several of these compounds exhibited strong binding affinity with RdRp, suggesting their potential as RdRp inhibitors (Singh *et. al.,* 2020).

Overall, while drug repurposing targeting RdRp of SARS-CoV-2 shows promise, further studies are needed to determine the efficacy and safety of these drugs in clinical trials. This research solely utilizes various computational techniques related to computer-aided drug design (CADD) to facilitate the drug repurposing process. The aim is to develop potential drug candidates that can either inhibit or manage symptoms, and to prevent lead molecules from failing in later stages of drug discovery.

Hypothesis:

Null Hypothesis:

Potential drug candidate against RdRp of SARS CoV-2 will not be developed

Alternative Hypothesis:

Potential drug candidate against RdRp of SARS CoV-2 will be developed

Objective:

General Objective:

To discover new lead molecules against SARS CoV-2 for potential new drug discovery

Specific Objective:

- To identify drug target necessary for survival of virus in host
- To generate a ligand library using rigorous screening parameters of druggability
- To perform Molecular docking and analyzing ligand protein interactions

- To perform Molecular Dynamic Simulation

Rationale:

Despite of availability and visibly astounding efficacy of vaccines against SARS-CoV-2, it is evident that COVID-19 continues to spread and impact communities world-wide. Various reasons such as hesitancy to get vaccinated, shortage of vaccines in developing and under-developed countries, emergence of mutant variants and diminished effectivity of COVID vaccine in immune-suppressed individuals owes to the continuing spread of the pathogen. Thus, it is predicted that the virus will become endemic (Hall et al., 2021) and will lead to reduced vaccine efficacy in future due to evolution of SARS-COV-2 along with human host. Such emerging concerns necessities the search of drugs against SARS-COV-2 to combat any sudden or unpredicted situations which might arrive in future by only relying on vaccine development. In order to do that, this research focuses on how we could generate potential drug candidates against the pathogen via computational approaches in economical and robust manner.

Scope of study:

The main focus of this study is to identify potential lead molecules that can target RdRp of SARS-CoV-2. The study involves conducting comprehensive screening processes using insilico tools and resources and filtering screened hits using different analytical approaches.

LITERATURE REVIEW:

Coronaviruses:

Introduction:

Coronaviruses (CoVs) are enveloped, positive-sense and single stranded RNA viruses and are highly diverse (Zumla *et. al.,* 2016). They belong to order Nidovirales which are further divided into four families; *Coronaviridae*, *Mesoniviridae*, *Arteriviridae* and *Roniviridae*. *Coronavirinae* which is one of subfamilies in *coronaviridae* family are further classified under four genera including alfa(α), beta(β), gamma(γ), delta(δ) coronaviruses. These viruses were classified according to the phylogenetic clustering studies (C. S. G. of the international,2020).

Coronaviruses have unusually large RNA genomes, containing upto 33.5 kilobase (kb). Different nidoviruses use similar strategies to organize, express and replicate their genomes. Some of the common features in viruses belonging to *Nidovirales* order encompasses: (a)Highly conserved genome which consists of large replicase coding region followed by other structural and accessory coding sequences, (b) ribosomal frameshifting leads to expression of numerous non-structural genes, (c)large replicase-transcriptase polyprotein with distinct or uncommon functions, (d)3'nested sub-genomics mRNAs for expressing downstream genes (Fehr and Perlman, 2015).

Coronaviruses are responsible for several diseases impacting respiratory, hepatic, enteric and neurological systems to varying degree of severity among humans and animals (Chan *et. al.,* 2013). Usually, Human CoV infections have caused less respiratory infections with HCoV-OC43, HCoV-229E, HCoV-NL63, and HCoV-HKU1 responsible for mild respiratory illness (Channappanavar *et. al.,* 2014). However, emergence of two novel CoVs, severe acute respiratory syndrome CoV (SARS-CoV) and Middle East respiratory syndrome CoV (MERS-CoV) in past decades has caused serious human diseases affecting respiratory system (Cheng *et. al.,* 2007) (Chan *et. al.,* 2015). More than 8000 peoples were affected globally resulting in 800 death cases during the SARS-CoV epidemic and MERS-CoV recorded 857 cases of infection with 334 deaths (Gretebeck and Subbarao, 2015). Among the members of family of CoVs that

infects human, SARS-CoV-2 has been registered as seventh member till now. The main symptoms of COVID-19 included fever, fatigue, and cough, which are similar to that of SARS-CoV and MERS-CoV infected cases (He *et. al.,* 2020).

Genomic Organization:

The Genome size of coronaviruses is of ~30 kb and is characteristically non-segmented, positive sense RNA molecule, establishing them as largest known RNA viruses (Enjuanes *et. al.,* 2000). The genome possesses 5' cap structure with leader sequences and untranslated region (UTR) (Morris and Geballe, 2000) and 3' untranslated region upstream from the 3' -terminal poly(A) tail which allows whole genome to act as an mRNA for translation of the replicase polyproteins (Brian and Baric, 2005). The large portion of genome, about 20 kb, encodes for non-structural proteins (nsps) which are involved in various functioning such as proteolytic cleavage of gene 1 polyprotein products, replication of genome, and transcription. ORFs 1a and ORF 1ab is translated from gene 1, with 1ab translated due to pseudoknotted structures and slippery sequence at ORF1a/1b junction which is responsible for -1 ribosomal frameshifting mechanism (Brown and Brierley, 1995). Downstream to replicase polyproteins coding region is structural and accessory proteins coding genes with each of them carrying transcriptional regulatory sequences (TRSs) which are necessary for expression of these genes (Fehr and Perlman, 2015). Accessory genes are scattered among structural genes that codes for Spike(S), Envelope(E), Membrane(M) and Nucleocapsid(N) protein. The accessory proteins, which are nonessential for replication in cell culture have been shown to have important roles in viral pathogenesis (Zhao *et. al.,* 2012).

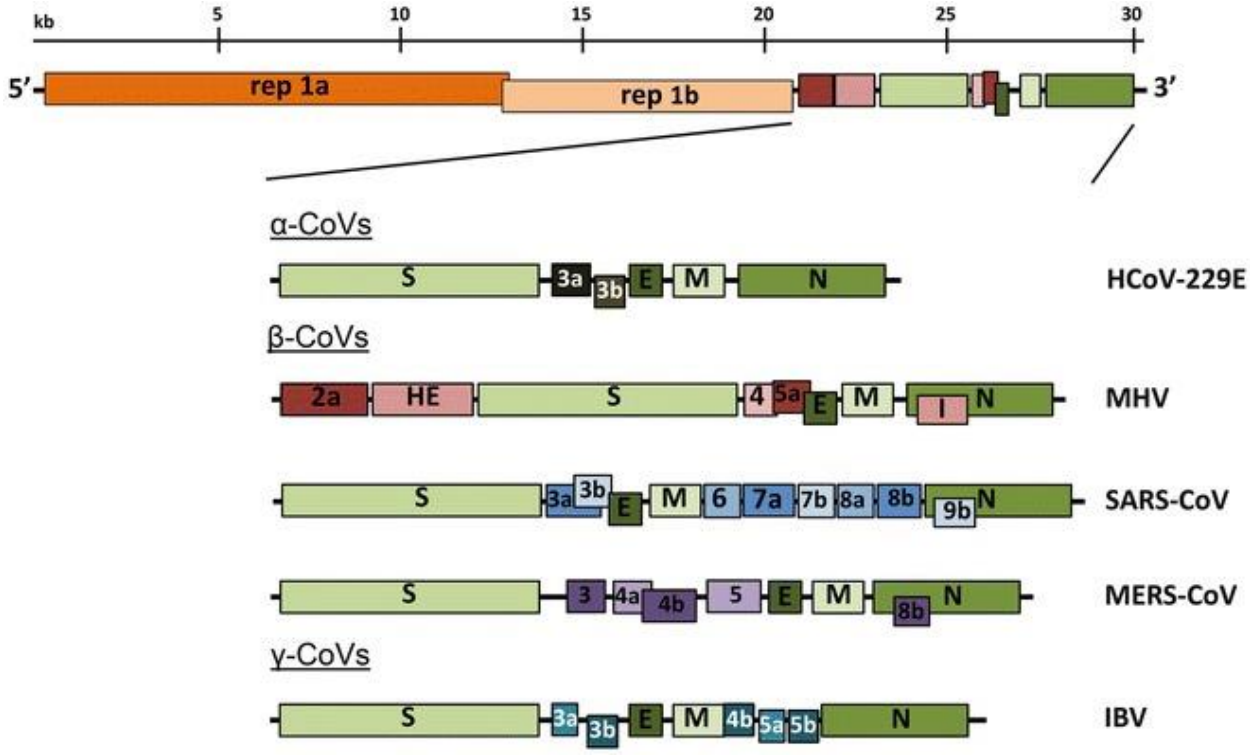


Fig1: Genomic organizations of representative α, β and γ CoVs(Fehr and Perlman, 2015).

SARS-CoV-2:

Severe acute respiratory syndrome coronavirus2 (SARS-CoV-2) is a novel coronavirus which is responsible for Coronavirus disease 19 (COVID-19) (Zhou *et. al.,* 2020) (Zhu *et. al.,* 2020). Similar to other coronaviruses (order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*), SARS-CoV-2 is an enveloped virus with non-segmented, positive-sense, single-stranded RNA genome with approximate size of 30kb and is a member of genus *betacoronavirus*, jointly with SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) with whom it shares 80% and 50% homology respectively (Kim *et. al.,* 2020) (Zhou *et. al.,* 2020). Initially, Coronaviruses (CoVs) were perceived as cause of enzootic infections in birds and mammals. Nonetheless, repeated occurrence of SARS, MERS and recent COVID-19 outbreaks have clearly indicated the ability of these viruses to cross species barrier and mediate human infections as well (Menachery *et. al.,* 2017). Although the intermediate animal host of SARS-CoV-2 is yet unknown, the potential reservoir is believed to be bats, which carry the virus with no signs of disease (Li *et. al.,* 2020).

Virion structure:

Under the investigation of electron microscopy, SARS-CoV-2 observed in infected cell post 3 days of infection, revealed crown like structure with virion size spanning from 70-90 nm diameter and were observed mainly in vesicles (Park *et. al.,* 2020). The helical nucleocapsid (N) consisting viral RNA is confined within host membrane-derived lipid bilayer which has fixed spike(S), membrane(M) and envelope(E) surface viral protein (Finlay *et. al.,* 2004).

The S protein is approximately 150 kda in size and is heavily N-link glycosylated which acts as signal sequence to gain access to Endoplasmic reticulum (ER). The virus surface is characterized by unique spike structure made by homotrimers of S protein (Beniac *et. al.,* 2006). This trimeric S protein is a Class I fusion protein that facilitates attachment to host receptor (Collins *et. al.,* 1982). Mostly in coronaviruses S protein is cleaved into two functional polypeptides, S1 and S2 via host cell furin-like protease (Abraham *et. al.,* 1990). S1 constitutes of large receptor-binding domain, while S2 functions as skeletal support in the form of stalk of the spike molecule (De Groot *et. al.,* 1987).

M protein are small (approx. 25-30 kda) structural proteins in the virion and are present in abundance comprising three transmembrane domains (Armstrong *et. al.,* 1984). The M protein consists of N-terminal ectodomain which is glycosylated and comparatively larger C-terminal endodomain which gives shape to virion structure (Nal *et. al.,* 2005). M protein is dimeric in virion and aids in sustaining the membrane curvature and binding to the nucleocapsid (Neuman *et. al.,* 2011).

E protein is highly divergent and present in small amount. The size of E protein ranges from 8.4-12 kda (Godet *et. al.,* 1992). Although membrane topology is not resolved fully, various researches predict E protein as transmembrane protein with N-terminal ectodomain and C-terminal endodomain. The E protein has an ion channel activity which helps in viral pathogenesis of SARS-CoV and probably SARS-CoV-2. Beside this, E protein also has an important role in viral assembly and release (Nieto-Torres *et. al.,* 2014) (Mukherjee *et. al.,* 2020).

The N protein is present in nucleocapsid and binds to RNA genome. It comprises of N-terminal and a C-terminal domain, each of which binds to RNA using different mechanisms (Chang *et. al.,* 2006) (Hurst *et. al.,* 2009). N protein is phosphorylated heavily and upon phosphorylation, induces structural changes that enhances affinity of N protein for Viral RNA (Stohlman and Lai, 1979). The binding of N protein to viral genome gives beads on a string structure. Two distinct RNA substrates have been recognized for N protein; TRSs and the genomic packaging signal. Genomic packaging signal binds to C-terminal RNA binding domain (stohlman *et. al.,* 1988) (Kuo and Masters, 2013). The N-protein binds to M protein and nsp3 which is a crucial constituent of replicase complex. Ultimately, the interactions with nsp3 which help tether the viral genome to replicase-transcriptase complex (RTC) and M protein facilitates in packaging of encapsidated genome into the viral particles (Hurst *et. al.,* 2013) (Sturman *et. al.,* 1980).

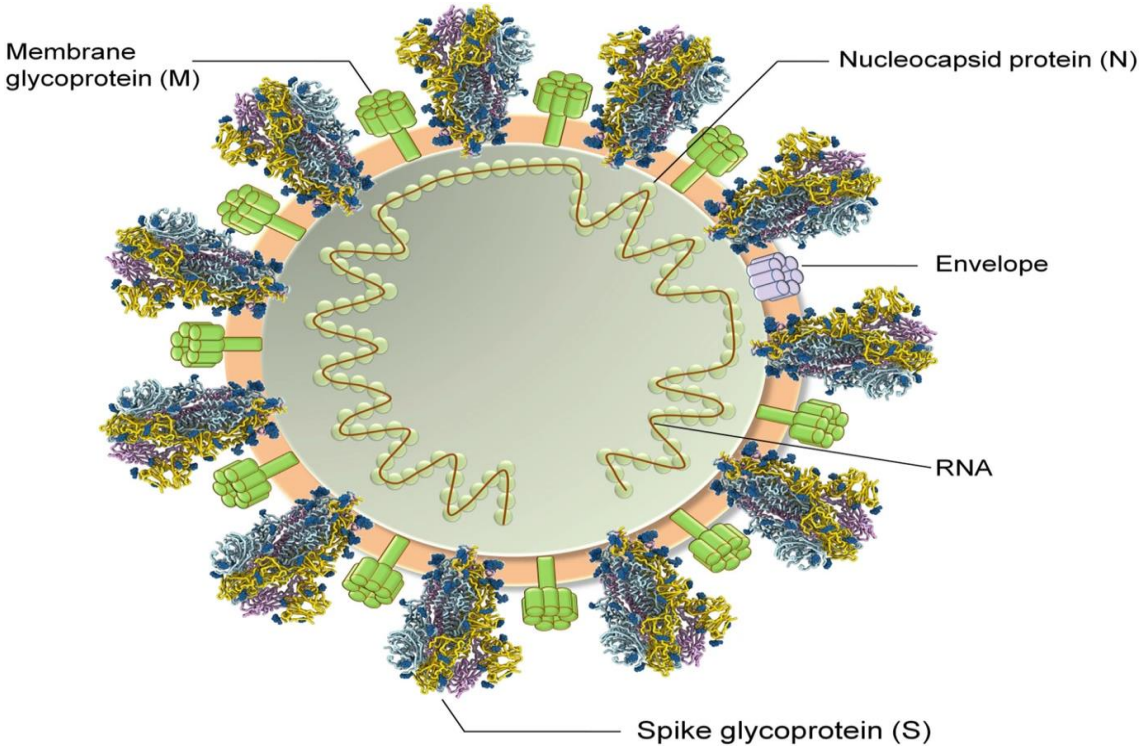


Fig 2: Virion structure of SARS-CoV-2 comprising structural proteins and viral RNA (Kumar et. al., 2020)

Genome Organization of SARS-CoV-2:

The SARS-CoV-2 genome is similar to that of typical CoVs and contains at least ten open reading frames (ORFs). The genome size of coronavirus ranges from 26 to 32 kb consisting 6–11 open reading frames (ORFs) encoding polyproteins made up of 9680 AA (Amino Acid) (Guo *et. al.,* 2020). The first ORF encodes for 16 nonstructural proteins and occupies 67% of the genome and the remaining ORFs codes for structural and accessory proteins. Unlike some other beta-coronaviruses, SARS-CoV-2 genome is devoid of hemagglutinin-esterase gene. Nonetheless, untranslated regions (UTRs) flanks were observed at both 5′ and 3′ end of 265 and 358 nucleotides respectively. There was no any notable contrast in ORFs and NSPs coding region between SARS-CoV-2 and SARS-CoV when sequence variation data was revealed. Some of the proteins coded by nsp genes encompasses two viral cysteine proteases including papain-like protease (nsp3), chymotrypsin-like, 3C-like, or main protease (nsp5), RNA-dependent RNA polymerase (nsp12), helicase (nsp13), and others which play integral role in SARS-CoV-2 transcription and replication (Chan *et. al.,* 2020). Besides that, additional region coding four structural proteins and other accessory protein completes SARS-CoV-2 genome (Li *et. al.,* 2020). By using combination of Sanger, Illumina, and Oxford nanopore sequencing, whole genomic sequences of SARS-CoV-2 were mapped (Lu *et. al.,* 2020).

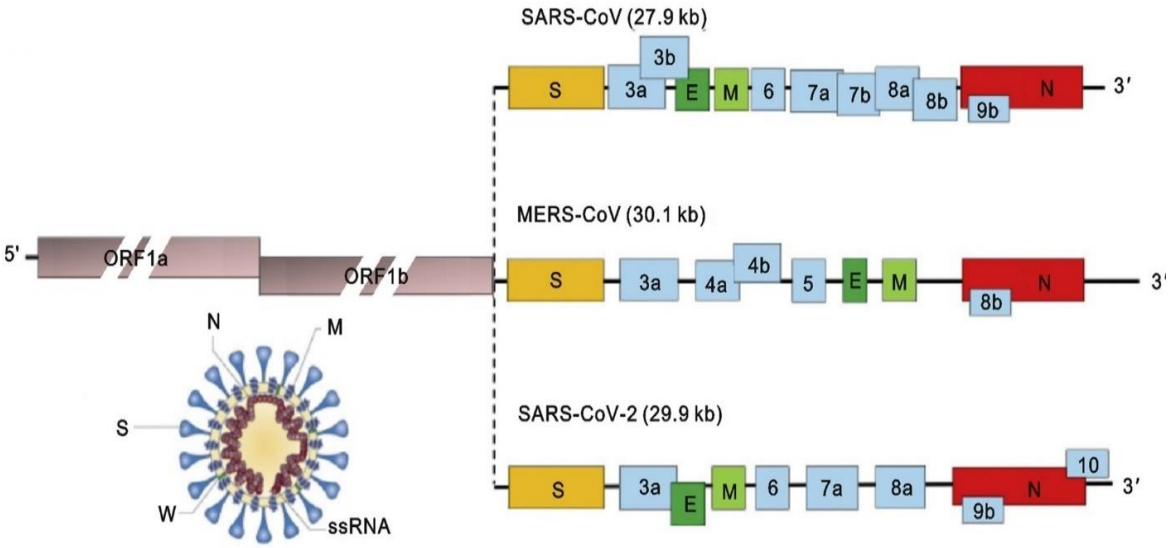


Fig3: Genomes of SARS-CoV, MERS-CoV and SARS-CoV-2 (Li et. al., 2020)

SARS-CoV-2 infection cycle:

SARS-CoV-2 utilize multiple host components, which are determinants of viral tropism, to infect the host cells. By entering the cell and translating their viral genome, they initiate a cytoplasmic replication cycle that coordinates numerous strategies for ultimate release of viral progeny while extensively relying on host machineries and metabolism (Malone *et. al.,* 2022).

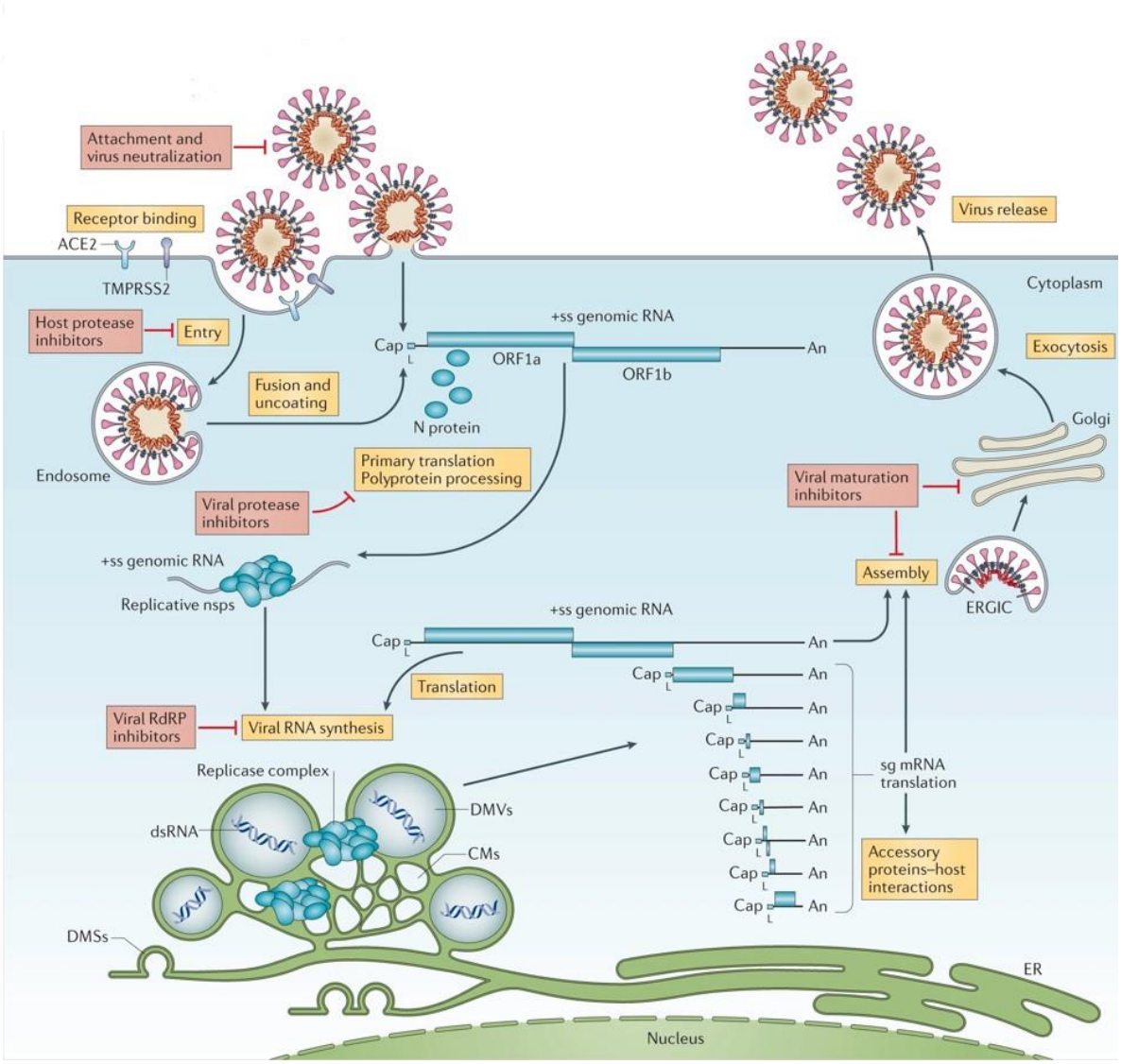


Fig 4: SARS-CoV-2 life cycle in host cell (V'kovski et. al., 2021)

Attachment and Entry:

S protein is central for attachment and entry of virus (Du *et. al.,* 2009) as it enables binding to host receptor and membrane fusion and is important in dictating host tropism and transmission capacity (Lu *et. al.,* 2015) (Li, 2015). The fusogenic spike protein with two domains S1 and S2 functionally aid in receptor binding and fusion with cell membrane (He *et. al.,* 2004). Like SARS-CoV, SARS-CoV-2 uses the angiotensin-converting enzyme 2 (ACE2) receptor for cell entry and Transmembrane serine protease 2 (TMPRSS2) for S protein priming (Hoffmann *et. al.,* 2020). Additionally, biophysical and structural studies disclosed that S protein of SARS-CoV-2 exhibits 10–20 fold higher affinity for ACE2 when compared to SARS-CoV (Wrapp *et. al.,* 2020). This higher binding affinity corelates with the evidence of higher spread of SARS-CoV-2 in human populations. The interaction between S protein and ACE2 leads to conformational change in S protein due to which E protein coalesces with host cell membrane which is followed by entry of viral RNA via endosomal pathway (Coutard *et. al.,* 2020) (Matsuyama and Taguchi, 2009). SARS-CoV-2 does not enter cell using other CoV receptors such as aminopeptidase N and dipeptidyl peptidase 4 (Zhou *et. al.,* 2020).

Expression of Replicase protein:

Following entry of viral genome into host cell cytoplasm, next step is translation of replicase gene from genomic RNA. Two large replicase ORFs; ORF1a and ORF1b is translated using host ribosomes (Wu *et. al.,* 2020). The translated products are replicase polyproteins pp1a and pp1ab which are 4405 and 7096 AA residues, respectively. Translation of pp1ab occurs due to pseudoknotted structures and slippery sequence at ORF1a/1b junction which is responsible for -1 programmed ribosomal frameshifting mechanism. The predicted efficiency of frameshifting is 45-70% which leads to overexpression of ORF1a-encoded proteins compared to ORF1b-encoded proteins (Finkel *et. al.,* 2021). After translation, pp1a and pp1ab undergoes 15 proteolytic cleavages executed by viral coded chymotrypsin-like or main protease (Mpro) and papain-like protease (PL-pro) to yield 16 non-structural proteins which are listed in Table 1(Ziebuhr *et. al.,* 2000) (Snijder *et. al.,* 2003).

Table 1: Functions of nsp genes in coronaviruses

Protein	Function
nsp1	Assists degradation of cellular mRNA and obstructs translation in host cell, thus blocking innate immune response
nsp2	Function yet unknown, binds to prohibitin proteins
nsp3	Large, transmembrane protein with multiple domains activities: Communicates with N protein via Ubl1 and Ac domains, Enhances cytokine expression through ADRP activity, PLPro/Deubiquitinase domain, cleaves viral polyprotein & hinders with host innate immune response Domains with unknown functions includeUbl2, NAB, G2M, SUD, Y domains.
nsp4	Potential transmembrane scaffold protein, integral for correct structure of DMVs
nsp5	Known as Mpro, which cleaves viral polyprotein
nsp6	Probable transmembrane scaffold protein
nsp7	Along with nsp8 makes hexadecameric complex with nsp8, might function as processivity clamp for RNA polymerase
nsp8	Along with nsp7 forms hexadecameric complex, might function as processivity clamp for RNA polymerase; may act as primase
nsp9	Binds RNA
nsp10	Forms heterodimer with both nsp16 and nsp14 and acts as Cofactor which prompts ExoN and 2-O-MT activity
nsp12	RdRp
nsp13	RNA helicase, 5’ triphosphatase

nsp14	N7 MTase) and 3’-5’ exoribonuclease, ExoN; N7 MTase adds 5’ cap to viral RNAs, ExoN activity for viral genome proofreading
nsp15	Viral endoribonuclease, NendoU
nsp16	2’-O-MT; guards viral RNA from MDA5 recognition

Note: Ubl (ubiquitin-like), Ac (acidic), ADRP (ADP-ribose-1’-phosphate), PLPro (papain-like protease), NAB (nucleic acid binding), SUD (SARS-unique domain), DMVs (double-membrane vesicles), Mpro (main protease), RdRp (RNA-dependent RNA polymerase), MTase (methyltransferase), ExoN (Viral exoribonuclease), NendoU (Viral endoribonuclease), 2’-O-MT (2’-O-Methyltransferase), MDA5 (Melanoma differentiation associated protein 5) (Malik, 2020)

Translation of host mRNA is blocked by nsp1 (Thoms *et. al.,* 2020) (Schubert *et. al.,* 2020). The other non-structural proteins form protein complexes termed as replication-transcription complexes (RTCs) whose role is crucial in synthesis of viral RNA. Primarily, proteins coded by nsp12, nsp13, nsp14 and nsp16 directs replication and transcription of viral RNA. Along with the assistance of nsp7 and nsp8 proteins, Nsp12, which is protein containing RdRp domain, forms holoenzyme RdRp (holo-RdRp) and catalyzes the RNA synthesis in host cell (Malone *et. al.,* 2022). Other subunits in RTCs play their part in regulating host innate immune response and modifying cell organelles into distinct double-membrane vesicles called ‘replication organelles’ which induces viral RNA synthesis (Cortese *et. al.,* 2020) (Snijder *et. al.,* 2020).

RdRp complex:

Multiple NSP proteins are utilized for viral replication and transcription which are assembled as breakdown product of ORF1a and ORF1ab polyproteins (Ziebuhr, 2005). Among these NSPs, RNA dependent RNA polymerase (RdRp) plays critical role in replication of viral genome by catalyzing viral RNA synthesis. For enzyme function of this protein, two additional cofactors coded by nsp7 and nsp8 are required and collectively is called RdRp complex (Subissi *et. al.,* 2014).

Cryo-EM structure of SARS-CoV-1 RdRp which was published shortly before CoVid-19 pandemic helped in understanding the structural aspects of polymerase of coronaviruses (Kirchdoerfer and Ward, 2019). Currently, various researches have successfully provided X-ray and cryo-EM structure of RdRp complex. The first structural insights of RdRp complex of novel SARS-CoV-2 were resolved in April via cryo-EM method (Gao *et. al.,* 2020). Additionally, further studies were performed which showed similar data (Peng *et. al.,* 2020) (Yin *et. al.,* 2020). These studies showed that RdRp of SARS-CoV-2 exhibited high similarity with RdRp of SARS-CoV-1 and for enzyme activity forms heterotetrameric complex of nsp12, nsp7 and nsp8. Furthermore, the structure of SARS-CoV-2 RdRp disclosed N-terminal portion of NiRAN domain which was formerly unclear to scientific community (Hillen, 2021).

Using bacterial expression system, full length nsp12 of SARS-CoV-2 virus was cloned and incubated along with nsp7 and nsp8 to form complex which was then purified for further structural studies (Gao *et. al.,* 2020). Structurally nsp12 consists of two main domains: polymerase domain from amino acid residue S367 to F920 and a nidovirus-specific N-terminal extension domain from amino acid residue D60 to R249 which maintains a nidovirus RdRp-associated nucleotidyltransferase (NiRAN) architecture (Lehman *et. al.,* 2015). These two domains are interconnected via an interface domain. Polymerase domain comprises three subdomains: a fingers subdomain,a palm subdomain, and a thumb subdomain. The active site in RdRp domain is formed by the conserved polymerase motifs A to G in the palm domain and configured like other RNA polymerases (Gao *et. al.,* 2020). The nsp7-nsp8 pair structure is conserved and analogous to SARS-CoV nsp7-nsp8 pair ((Kirchdoerfer and Ward, 2019).

Replication and Transcription:

Formation of RTCs is essential for viral RNA synthesis and subsequently responsible for replication of virus and transcription of sub-genomic mRNAs(sg-mRNAs). RTC uses genomic RNA as a template to produce full length genome compliment and a set of minus-strand sgRNAs which are acquired from genomic region downstream of the replicase gene. These minus strand sgRNAs serve as template for sub-genomic mRNAs synthesis whereas the genome complement directs production of new genomic RNA.

The sg-mRNAs are responsible for producing structural proteins which are essential for virion assembly and release (Malone *et. al.,* 2022). Beside this, sg-mRNAs are used to express accessory proteins, which plays crucial roles in modulating host innate immune responses (Wang *et. al.,* 2021) (Kim *et. al.,* 2020).

Assembly and Release:

Structural proteins S, E and M are translated after replication and transcription and incorporated into endoplasmic reticulum where it moves along the secretory pathway into the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) (Krijnse-Locker *et. al.,* 1994) (Tooze *et. al.,* 1984). The newly synthesized genomic RNA encapsidated by N proteins buds into the lumen of ERGIC (Cohen *et. al.,* 2011; Perrier *et. al.,* 2019) which contains viral structural proteins, forming mature virion particles (de Haan *et al.,* 1998). When in the ER/ERGIC, virion particles move to the golgi apparatus and trans-Golgi network to undergo glycosylation and other post translational modifications (Fung and Liu, 2018). After this, Virions leave the infected cell via lysosomal trafficking which is a different route compared to other enveloped viruses (Ghosh *et. al.,* 2020).

Understanding the global pandemic:

The rapid spread of coronavirus has caused global pandemic and impacted the globe negatively. In less than half a year the virus has spread nearly worldwide and several countries are facing the second outbreaks with emergence of new variants (Dong *et. al.,* 2020; Wu *et. al.,* 2020; Xu and Li, 2020). The infectious agent responsible for the pandemic was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on 11th of February 2020 by WHO) (Gorbalenya *et. al.,* 2020; Zhu *et. al.,* 2020) which caused coronavirus 2019 (COVID-19). It is a member of same family of virus as SARS-CoV and MERs CoV but unlike the disease caused by these viruses is characterized with high spread, mortality and morbidities (Dong *et. al.,* 2020a; Zhu *et. al.,* 2020). The disease was first reported in wuhan, china in december of 2019 and has spread all over the world via travel and community-based contacts (WHO 2020; Wu *et. al.,* 2020; Zhu *et. al.,* 2020).

The virus was swiftly recognized as a novel coronavirus showing genetic similarities to SARS-CoV-1 (Lu *et. al.,* 2020; Wan *et al.* 2020). Due to its rapid spread and high fatality, it was termed "the first pandemic of the 21st century" by WHO (Dong *et. al.,* 2020; WHO 2020; Wu *et. al.,* 2020). Owing to the above reasons, it was paramount to devise swift, sensitive and definite diagnostic and serological testing methods for coping with and prevention of COVID-19 spread. Despite the good response in developing new testing methods, global conditions exhibit necessity of additional therapeutic interventions to control disease spread and speeding patients' recovery (WHO 2020; Tao *et. al.,* 2020). The infection has taken its toll on more than 12.7 million people worldwide and more than 560,000 fatalities has been reported with those numbers ever increasing (Dong *et. al.,* 2020) and beside that it has affected our daily lifestyle, habits, work and family as well (Wu *et. al.,* 2020; Zhu *et. al.,* 2020).

Transmission:

The transmission of infection from a person to another is still under research. However, the prime mode of transmission of virus occurs via fluid droplets secreted by respiratory system of the infected individuals which is relayed onto healthy individual while sneezing, coughing or talking without covering the mouth and the nose. The virus might also prevail in the air from expelled droplets of infected person and spread to healthy individual who comes in contact with them in confined area (Gandhi *et. al.,* 2020; Meselson, 2020; Morawska and Cao, 2020). Beside this there are other modes of transmissions such as:

- Direct contact with infected individuals (Li *et. al.,* 2020).
- Via contaminated surfaces or objects such as plastic, stainless steel and others with reports suggesting virus is stable on plastic and stainless steel upto 72 hrs, more than 4 hrs on copper and upto 24 hrs on cartons but whether the virus is viable for infection is still unclear (Rubens *et. al.,* 2020; van Doremalen *et. al.,* 2020)
- Touching nose, eyes and mount with contaminated hands
- Touching or smelling excreta of infected person

- And fecal microbiota isolated from infected person according to FDA guidelines (Machhi *et. al.,* 2020)

Clinical manifestations of SARS CoV-2:

Table 2**:** Various stages of clinical manifestation of disease in human. (Machhi *et. al.,* 2020)

Organ system of concern	**Early manifestation of disease**	**moderate manifestation of disease**	**severe manifestation of disease**	**Diagnostic signs**
Lung/Respiratory Pulmonary	Cough, Sore throat, Rhinorrhea, Sneezing, Dry Cough	Pneumonia, Dyspnea, Moderate hypoxemia	Severe hypoxemia, Acute respiratory distress syndrome (ARDS), Respiratory failure and death (if untreated)	Decreased % pO2, Chest X-rays show ground glass opacities
Brain/Neurological	Hyposmia-Anosmia, Hypogeusia Ageusia, Visual Disturbance, Fatigue, Somnolence	Headaches, Nausea and vomiting, Dizziness, Myalgia, Ataxia, Encephalopathy	Cerebrovascular disease (large vessel strokes), Seizures, Meningoencephalitis, Neuropathy, Guillain Barre Syndrome, Neurogenic ARDS, Coma	Elevated creatine kinase with myalgia, Brain MRI show hyperintensities in regions with infarction or encephalitis, SARS-CoV-2 detection in cerebrospinal fluid or brain tissues in some patients

Gastrointestinal	Nausea, Vomiting, Diarrhea, Heartburn	Loss of appetite, Abdominal pain and bloating	Gastrointestinal bleeding, GI viral dissemination	Elevated liver enzymes and bilirubins, SARS-CoV-2 detection in stool samples
Heart/Cardiac	Chest pain, Arrhythmia, sinus tachycardia	Cardiac inflammation, immunocytic infiltration	Cardiomayopathy, Acute heart failure	Elevated cardiac enzymes, Abnormal EKG (Prolonged QTc intervals, elevated ST), Cardiac-specific troponin and brain natriuretic peptide
Kidney/Renal	Proteinuria, Hematuria	Acute renal injury	Renal failure	Tubular necrosis and SARSCoV-2 detection in kidney
Blood vessels/ Vascular	Blood coagulation	Arterial or venous thromboembolism, Cytokine storms	Pulmonary embolism, Large vessel occlusions, Disseminated intravascular coagulation	Elevated D-dimer, interleukin-6, other cytokines, ferritin, and lactate dehydrogenase. Prolonged PT/PTT
Mental/Psychiatric	Depressed mood, Anxiety,	Depression,	Exacerbation of neurological or psychiatric	Elevated plasma calcium and phosphorus

	Insomnia, Anger, Fear	Post-traumatic stress disorder	disorders (e.g., Alzheimer's or Addiction)	(indicative of stress)

COVID-19 infection gradually negatively impacts immune system of host on its later stage, and immune system responds by significantly increasing cytokine release also known as "cytokine storm" syndrome (Moore and June, 2020). this is marked by increased neutrophil-to-lymphocyte ratio (NLR) which is caused due to SARS-CoV-2 infecting monocytes, macrophages, dendritic cells, and lymphocytes (Grifoni *et. al.,* 2020; Park, 2020; Yuki *et. al.,* 2020). Cytokines are released as a self-defense mechanism to combat against infection and helps in recruiting immune cells to fight off against viral infection (Moore and June, 2020). However, in SARS-CoV-2 infections, cytokine egress is very high which subsequently results in elevated leukocyte recruitment to multiple body organs and specifically in lung cells which leads to acute respiratory distress syndrome (ARDS) (Zhang *et. al.,* 2020).

RdRp Mutation:

Several common mutations have been found in the RdRp coding region polyprotein gene among SARS-CoV-2 isolates worldwide. One significant mutation is the 14408C>T transition, which has been detected in more than 7,000 isolates from various continents. An initial examination of 137 SARS-CoV-2 genomes from North America and Europe showed that this particular mutation was associated with higher number of mutations. This substitution of proline to leucine (P323L) resulting from the 14408C>T mutation has been proposed to make the RdRp protein structure more rigid which may have impacted on the way RdRp interacts with other elements of the transcription/replication mechanism or the RNA template, resulting in an altered rate of mutation (Pachetti *et. al.,* 2020).

Introduction to Computer-aided drug design (CADD):

Due to limitation of human resources, money and time, drug discovery is considered an exhaustive process (Song *et. al.,* 2009). From discovery of new drug candidate to clinical trials, an estimated time of 14 years (Myers and Baker, 2001) and cost from 800 million US dollars (DiMasia *et. al.,* 2003) to 2.558 billion US dollars (DiMasia *et. al.,* 2016). A typical drug discovery process entails multiple steps of identification of putative drug candidate, validation of drug target, hit to lead discovery, lead optimization and further preclinical and clinical trials (Vohora and Singh, 2018). Regardless of high investments and time expended, only 13% of drug are successful in clinical trials (Zhong *et. al.,* 2018) and 40-60% failure at later stage owes to deficiencies in pharmacokinetic properties; absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) (Hou and Xu, 2004). Contemplating these circumstances, it was paramount to search for new tools to facilitate drug discovery process (Klee, 2006).

Eventually, a new tool was developed using computational methods to ease discovery of new molecular entities and reduce later phase failures in drug discovery via early assessment of activity, selectivity and pharmacokinetic properties (Song *et. al.,* 2009). This insilico tools and resources employed was collectively termed as Computer-aided drug design (CADD) and its use were highlighted in the early 1970s to modify biological activity of insulin (Blundell *et. al.,* 1972) and to drive synthesis of human haemoglobin (Beddell *et. al.,* 1976). Main components of CADD include virtual screening (VS), virtual high throughput screening (vHTS), homology modelling, quantitative structure-activity relationship (QSAR), three-dimensional (3D) pharmacophore mapping and molecular docking (Hassan Baig *et. al.,* 2016; Surabhi and Singh, 2018; Veselovsky and Ivanov, 2003). Out of these components, virtual screening plays important role as an alternative for experimental high-throughput screening for hit identification and optimization (Shoichet, 2004; Sun, 2008).

Virtual screening:

Virtual screening has poised as an attractive alternative approach to execute screening of multitude of compounds to generate hits otherwise considered taxing via

experimental methods in pharmaceutical industry (Shoichet, 2004). Various computational tools are at our disposal to carry out virtual screening which is broadly classified into ligand-based virtual screening (LBVS) or structure-based virtual screening (SBVS). In LBVS method, information given by known set of ligands which binds to desired target are used to recognize new set of compounds in ligand databases with comparable properties (Reddy *et. al.,* 2007). Different techniques such as similarity and substructure searching (Mestres and knegtel, 2000), pharmacophore matching (Mason *et. al.,* 2001) or 3D shape matching (Srinivasan *et. al.,* 2002) are utilized to identify new drug like candidates. Whereas SBVS method utilizes the known 3D structure of protein target obtained via Nuclear magnetic resonance (NMR) and X-Ray crystallography studies, which acts as a major source for further molecular docking processes to identify novel drug candidates and furthermore these lead compounds are optimized for further studies (Walters *et. al.,* 1998).

Ligand based virtual screening:

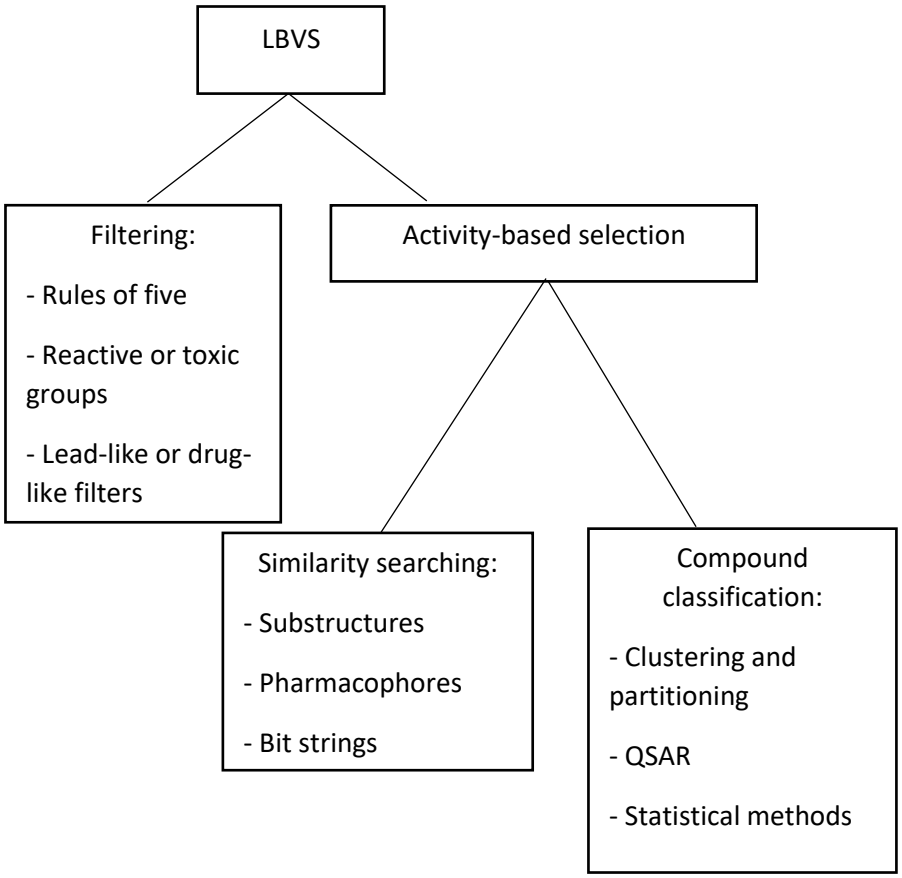


Fig 5: LBVS. The chart shows various methods utilized in LBVS (Stahura and Bajorath, 2005)

In contrast to Structure based approaches where 3D structure of target is known, LBVS method utilizes information corresponding to structure and molecular descriptors (physiochemical properties) of known ligands which are active against the target of interest and via similarity based computational tools helps anticipate new compounds with similar activity (Macalino *et. al.,* 2015). The newly identified drug candidates are ranked according to the similarity score obtained, which are based on the algorithm used (Reddy *et. al.,* 2007). The basic assumption of LBVS is similar compounds have similar effects and if one or more compounds of known activity against a target receptor already exist, it is possible to search for similar but more effective new drug candidates from the available compound databases (Lengauer *et. al.,* 2004). According to different researches, LBVS combined with other methods has shown promising result in designing drug targeting G-protein coupled receptors targets (Rai et al, 2010), Human ATP binding cassette (ABC) transporter (Sager *et. al.,* 2012) and Mitochondrial enzyme NADH:quinone oxidoreductase (PfNDH2) (Sharma *et. al.,* 2012).

Structure based virtual screening (SBVS):

The 3D structure of target protein or receptor acquired from X-ray crystallography, NMR or neutron scattering spectroscopy, homology modeling, or molecular dynamic (MD) simulation acts as a starting point for SBVS (Lionta *et. al.,* 2014).

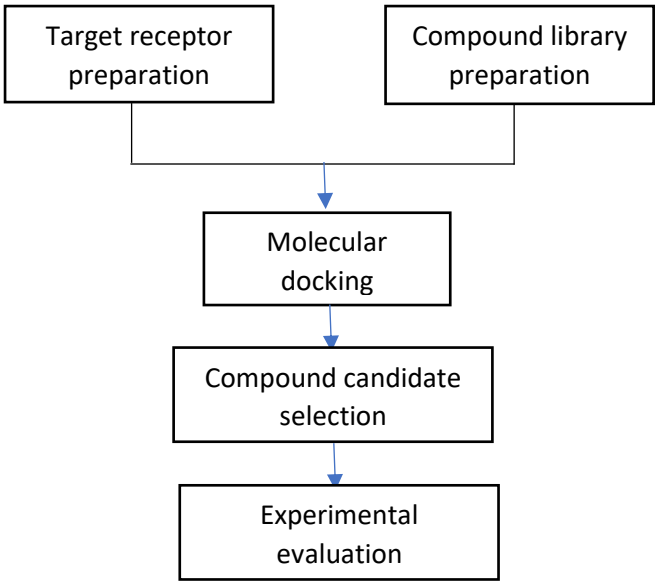


Fig 6: SBVS. General schematic of structure based virtual screening (Li and shah, 2017)

Upon availability of receptor 3d structure, ligand binding site is searched or predicted through native ligand interaction with receptor and then ligands present in different database are docked into the binding site to evaluate the affinity of these molecules with the desired target receptor (Villoutreix *et. al.,* 2009). Ligand Binding site is a cavity or protuberance in target protein which acts as molecular hotspots for probable electrostatic, hydrophobic and other molecular interactions. Receptor/ligand interactions such as steric and electrostatic are basis of identifying the best lead molecules which are further validated using biochemical assays (Anderson, A. C., 2003). Principally SBVS has three major components; 3D structure of receptor, chemical compound library and molecular docking (Li and Shah, 2017). These components are further utilized to establish an efficient screening procedure which are as follows.

Attaining Protein structure and their preparation:

Advancement in various tools and techniques such as X-ray and NMR has largely assisted in generating 3D structure of target protein, which are then stored and made available in protein data bank (PDB) format (Wang *et. al.,* 2018). When resolved 3D structure of protein target is unavailable, various computational methods such as ab initio modeling (Lee *et. al.,* 2017), threading (Lemer *et. al.,* 1995), and homology modeling (Vyas *et. al.,* 2012) has proved to predict structure of proteins from their corresponding sequences. After acquiring the 3D structure, protein is further prepared which involves analysis of quality of structure, addition of hydrogen atoms and prediction of accurate protonation state, addition or removal of water molecules, metal ions, cofactors, counterions, removal of subunits not required for ligand binding and re-orientations of hydroxyl groups (Pitt *et. al.,* 2013). Various computational tools are utilized for preparation of protein such as Sybyl and Maestro (Madhavi Sastry *et. al.,* 2013) (Pitt *et. al.,* 2013). The significance of protein preparation is directly correlated to enhancement in docking performance (Madhavi Sastry *et. al.,* 2013).

Identification of Binding site:

Binding site or pocket is a cavity where ligand adheres to the target thus producing desired effect. It is imperative to recognize the binding site of target protein to conduct further downstream processes of SBVS (Batool *et. al.,* 2019). The information of binding site can be derived from the site-directed mutagenesis study or X-ray crystallographic structures of proteins co-crystallized with substrates or inhibitors (Pan *et. al.,* 2017). Upon unavailability of experimental data, various programs are utilized to predict probable binding sites which are usually geometry-based and/or energy based (R Laurie and Jackson, 2006). Additionally, Arrays of software and webservers such as CASTp (Binkowski *et. al.,* 2003), MetaPocket (Huang, B., 2009), Q-SiteFinder (Laurie and Jackson, 2005), NSiteMatch (Sun and Chen, 2017), DoGSite Scorer (Volkamer *et. al.,* 2012), DEPTH (Tan *et. al.,* 2011), and MSPocket (Zhu and Pisabarro, 2011) are available which helps to predict the putative binding sites of the target proteins.

Compound database preparation:

Adequate database of compounds is necessary for SBVS, which includes small drug like molecules which are freely available or can be purchased (Hergenrother, P. J., 2006). Various free chemical databases are present such as ChEMBL (>1.6 million distinct compounds) (Gaulton *et. al.,* 2012), ZINC (230 million available compounds) (Irwin and Shoichet, 2005), PubChem (111 million pure and characterized chemical compounds) (Kim *et. al.,* 2016), ChemSpider (25 million unique chemical compounds) (Pence and Williams, 2010), and DrugBank (14528 drug molecules) (Wishart *et. al.,* 2008). Compounds from these databases can be screened for desirable properties such as stability and solubility in aqueous media, presence of appropriate functional groups to interact with biological targets and devoid of toxic and undesirable moieties to identify potential drug candidate (Chandrasekaran *et. al.,* 2018). Screening is done to ensure 'druglikeness' of molecule considering various rules and parameters such as lipinski's rule of five, veber's rule, ADMET parameters and other risk parameters such as acute rat toxicity, carcinogenicity, serum glutamic oxaloacetic transaminase elevation, hepatotoxicity, and inhibition of 3A4 oxidation of midazolam (Veber *et. al.,* 2002) (Wang *et. al.,* 2018).

'Druglikeness' is used in drug design to predict how much a molecule has "drug-like" properties with respect to factors like bioavailability. Approach to measure drug-likeness is property-based filters/rules, which define acceptable boundaries of certain molecular physicochemical properties for drugs and/or drug candidates. Lipinski's rule of 5 describes four physiochemical properties to measure the drug-likeness of a molecule within the permissible limits such as molecular weight lower than 500, lipophilicity (logP) lower than 5, less than five hydrogen bond donors, and less than 10 hydrogen bond acceptors (Lipinski *et. al.,* 2001). Veber's rule extends it further including rotable bond count and TPSA where rotable bond count should be less than 10 and topological polar surface area should be less than 140$Å^2$ (Veber *et. al.,* 2002).

Compound with lead like properties, devoid of toxic and metabolically liable moieties should be filtered from large number of compounds available in different molecular databases as it is time consuming and computationally demanding to perform 'blind

docking' which additionally results in unnecessary data and creates complexity in choosing compounds (Blagg, J., 2006). Recently, various computational tools are present which facilitate to efficiently filter a library of compounds against desired criteria of pharmacological and ADMET properties. DataWarrior, an open-source drug discovery tool serves as a powerful application for this purpose (Lopez-Lopez *et. al.,* 2019).

Docking and Scoring:

Molecular docking is a computational technique that simulates ligand and receptor interaction at molecular level and allows ranking of ligands by quantifying their binding affinity towards the receptor using various scoring functions. It predicts various possible binding modes of a ligand in a receptor binding site and evaluates affinity according to its conformation and complementarity with the properties present in the binding pocket (Hung and Chen, 2014). Numerous docking programs are available to date, including AutoDock (Morris *et. al.,* 2009), Dock (Ewing *et. al.,* 2001), FlexX (Bursulaya *et. al.,* 2003), Glide (Friesner *et. al.,* 2004), Gold (Jones *et. al.,* 1997). There are two types of molecular docking: flexible-ligand search docking and flexible-protein docking. The flexible-ligand search docking method typically uses three algorithms - systematic, stochastic, and simulation methods - to account for ligand flexibility (Sousa *et. al.,* 2006). On the other hand, flexible-protein docking primarily utilizes Monte Carlo (MC) and molecular dynamic (MD) methods (Oshiro *et. al.,* 1995; Hart and Read, 1992) to address protein flexibility.

A scoring function plays a critical role in enabling a docking program to explore the ligand-binding site. Once a promising binding conformation is identified, the scoring function evaluates the binding affinity. As such, the impact of scoring functions on docking is substantial. To train a scoring function, a training dataset of similar compounds with known experimental binding affinity is used. Scoring functions fall into four main categories: force field, empirical, knowledge-based, and machine learning (ML) (Moitessier *et. al.,* 2008; Huang *et. al.,* 2010). Force field scoring functions are calculated by estimating intermolecular interactions such as electrostatic and Van der Waals forces between the binding partners whereas

Empirical scoring functions rely on the number of atoms in the ligand and target protein and are used for affinity and pose prediction. This method includes hydrophobic and hydrophilic forces, hydrogen bonding, and entropy (Guedes *et. al.,* 2018). A statistical approach called multiple linear regression is used to fit scoring-function coefficients. Knowledge-based scoring functions are based on statistical potentials of intermolecular interactions. This method assumes that frequently occurring functional groups or types of atoms are energetically favorable and contribute to binding affinity (Muegge, 2006). In contrast to classical scoring functions, ML methods are not limited to a predefined functional form among structure features and binding affinity values (Li *et. al.,* 2018). ML methods are dynamic techniques for constructing and optimizing models to predict binding pose and affinity. Recently, the development of novel scoring functions using ML has gained popularity (Hecht and Fogel, 2006).

Improving Pose/Compound Selection After Docking:

In SBVS, a rate-limiting step is often the requirement for a computational chemist with expertise to manually analyze the compounds resulting from a virtual screening or docking exercise before selecting those that are suitable for further experimental testing. This is necessary because the use of simplified scoring functions and inadequate sampling of the ligand's conformational space can result in unrealistic docking poses, such as intra-ligand steric clashes, twisted amides, E/Z esters, imperfect hydrogen bonding networks, and poses based on shape complementarity, which may lead to an undesirably high score and need to be eliminated. As a result, medicinal chemists frequently need to visually inspect thousands of docking poses to identify the best compounds for testing. Significant efforts have been made to enhance the efficiency and quality of the compound selection process (Athanasiadis *et. al.,* 2012).

METHODS AND METHODOLOGY:

Target identification and Retreival of 3D protein structure:

Among various nsp of SARS-CoV-2, RNA dependent RNA polymerase (RdRp) plays critical role in replication of viral genome by catalyzing viral RNA synthesis. Blocking the action of RdRp will restrict the proliferation of virus within host cell, thus making it a suitable target for drug discovery process. The electron microscopy structure resolved in 2.90 Å of SARS-CoV-2 RdRp (PDB ID:6M71) was retrieved from RCSB protein data bank(https://www.rcsb.org/structure/6M71). Similarly, X-Ray diffraction structure of hMAT1A (6SW5) and SARS-CoV-2 Nsp14 ExoN domain (PDB ID:7DIY) were also retrieved from RCSB protein data bank.

Protein structure Validation:

Protein structure validation was done by z-score and energy plot analysis using ProSA web server tool and Ramchandran plot analysis using PROCHECK web interface.

Target protein Preparation and Acquisition of Binding sites:

Protein structures retrieved from RCSB protein database was prepared using PyMol and AutoDock tools. Protein was first visualized and the cofactors and water molecules were removed from the protein usin PyMol. The additional target protein preparation includes removal of water, addition of hydrogen, merging non-polar hydrogen and computing Gasteiger charge using AutoDock tools. Furthermore, pdbqt format of target was prepared using AutoDock tools. Binding site for SARS-CoV-2 RdRp and SARS-CoV-2 Nsp14 ExoN was acquiring through literature review, whereas for hMAT1A was acquired using 3D ligand site (http://www.sbg.bio.ic.ac.uk/3dligandsite/).

Compound database preparation:

The preparation of a compound library for CADD involves several steps:

Scope of library:

The choice of compounds to include in a CADD library depends on the specific goals of the study. It could be natural products, small molecules, or compounds from a particular therapeutic area.

Obtaining molecular structures:

There are several databases including ZINC, ASINEX and UORSY available for obtaining the 2D or 3D structures of compounds

Calculating molecular properties:

Several molecular properties are used in CADD, including molecular weight, LogP, solubility, topological polar surface area, and hydrogen bond donors/acceptors, druglikeliness, rotable bond counts and toxicity profile. This was done using OSIRIS data warrior software. Stringent values were set to identify potential hits. Parameters used for screening is as follows:

Table 3: Parameters of pharmacokinetic properties for screening of compounds.

Molecular properties	**Value**
Total molecular weight	200 to 500 dalton
Clogp	-3 to 6
clogs	-4 to -2
Hydrogen bond acceptors	0 to 10
Hydrogen bond donors	0 to 5
Topological polar surface areas	<120
Druglikeliness	Positive value

Mutagenic	No
Tumorigenic	No
Reproductive effective	No
Irritant	No
Rotable bond count	0 to 10

Additionally, 3d structure of 2d ligands were also made using OSIRIS data warrior.

Generating energy-minimized conformations:

Energy minimization helps to ensure that the input structures are accurate and consistent. Software tools including Open Babel and PyRx 0.9.8 were used for energy minimization.

Preparing input files:

Once the molecular properties have been calculated and energy was minimized, inputs files for further docking was generated using Open Babel tool present in PyRx 0.9.8.

Molecular docking Studies:

Docking with 6M71:

Molecular docking was performed using using AutoDock Vina Wizard in PyRx 0.9.8 platform against SARS-CoV2-2 RdRp with prepared library of compounds. Molecular docking of ligand and protein was done along with native ligand GTP using PyRx platform to facilitate comparative study of binding energy with respect to that of native ligand. The docking was performed using the parameters; numbers of exhaustiveness of 16 and number of modes of 32.

Redocking of Potential hits:

Molecular hits generated from first molecular docking were individually downloaded and docked again against target protein using parameters; numbers of exhaustiveness of 32 and number of modes of 32.

Screening with ExoN, RTP and PI:

Hits generated from redocking were docked against ExoN domain of Nsp14 of SARS-CoV-2 in which GTP was again used as native ligand. Furthermore, Binding energy against RdRp were compared with RTP to further screen the potential hits. In addition to that, performance indexes of each compound were calculated based on number of hydrogen-bond acceptors/donors and rotatable bond counts.

{(H-acceptor + H-donor + Rotatable bonds count) * 5}/ 25

Screening of molecular hits were done based upon the binding energy with ExoN domain of Nsp14 of SARS-CoV-2, comparison with binding energy of RTP against RdRp and Performace index value.

Analysis of Protein-ligand interaction:

Receptor-ligand complex visualization and Prediction of amino acids responsible for receptor-ligand interaction:

The lead compounds screened from docking and PI analysis were further studied using PyMol for protein-ligand complex visualization and prediction of amino acids responsible for protein-ligand interaction. For Protein-ligand complex visualization output file of ligand and protein generated by the PyRx after the successful docking performance were used. For, prediction of amino acid responsible for receptor-ligand interaction, PyMol was used to identify and visualize amino acid present within 5 Å of the ligand.

Analysis of interactions responsible for receptor/ligand binding:

In order to understand the mechanism responsible for higher affinities receptor/ligand binding, different interactions leading to bond formation and bond distance were analyzed using Discovery studio visualizer. hydrogen bond, electrostatic and hydrophobic interactions were analysed in 3d representation. For studying hydrophobic interaction PDB file were generated and protonation was done. For analyzing bond types and distance pdbqt files generated as output from molecular docking were used.

Docking with mutated RdRp:

Mutated RdRp structure from substitution of proline to leucine (P323L) resulting from the 14408C>T mutation was prepared via PyMol which was used as source for performing docking studies with lead compounds.

Interaction with human proteins:

For checking the cross reactivity of lead molecules with human proteins, lead molecules were docked against hMAT1A protein. SAM was taken as native ligan for this study.

RESULTS AND DISCUSSION:

Selection of Protein Target:

Identifying a lead target protein is a crucial step in drug discovery and development. The process involves various factors that need to be considered to ensure the target protein is suitable for drug development. The target protein should be well-validated with strong scientific evidence that it plays a critical role in the disease or condition being targeted and should be specific to the disease or condition being targeted and not have significant effects on other essential biological processes (Gilson *et. al.,* 2016). Additionally, the target protein should have a suitable structure and biochemical properties that enable the development of drugs that can effectively bind to and modulate its activity and should not have significant side effects or toxicity concerns that would limit its suitability as a drug target (Anighoro *et. al.,* 2014).

The RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 has been identified as a key target for the development of antiviral drugs against COVID-19 The RdRp of SARS-CoV-2 is a promising target for antiviral drug development due to its critical role in viral replication, unique structural features, and potential for broad-spectrum antiviral activity. Here are some of the bases for targeting SARS-CoV-2 RdRp for computer-aided drug discovery:

- Essential role in viral replication: The RdRp of SARS-CoV-2 is a critical enzyme required for viral replication, making it an attractive target for the development of antiviral drugs (Kabinger *et. al.,* 2021).
- High specificity: The RdRp of SARS-CoV-2 is highly specific to the virus and is not present in human cells, making it an ideal target for the development of specific antiviral drugs with minimal side effects (Vincenti *et. al.,* 2021).
- Similarity to other viral RdRps: The RdRp of SARS-CoV-2 shares significant similarities with RdRps of other RNA viruses, providing a basis for the development of broad-spectrum antiviral drugs (Ahn *et. al.,* 2020).

- Potential for oral administration: RdRp inhibitors have the potential for oral administration, which is an advantage over other antiviral drugs that require intravenous administration (Hashemian *et. al.,* 2022).

Hence, 3d structure of SARS-CoV-2 RdRp (PDB id: 6M71) obtained via electron microscopy resolved at 2.9 angstrom was retrieved from RCSB (Gao *et. al.,* 2020) for further insilico analysis. Furthermore, in order to check the cross reactivity of potential leads with human proteins, human hepatic hMAT1A was considered. An association has been observed between the reduced biosynthesis of S-adenosylmethionine (SAM) in the liver and an increase in the production of pro-inflammatory cytokines and mediators (Avila et. al, 2005; Grimble and Grimble, 1998). Thus, the protein targets hMAT1A (PDB ID: 6SW5; Panmanee et al., 2020) was chosen for molecular docking to identify non-interference of potential leads targeting RdRp of SARS-CoV-2. Protein 3d structure were obtained from RCSB database.

Proofreading activity of nsp14 of SARS-CoV-2 involves its exonuclease activity and its ability to prevent mutagenesis during viral replication. The nsp14 protein, also known as nonstructural protein 14, contains both exoribonuclease and endoribonuclease activities, and is responsible for proofreading viral RNA during replication, correcting errors that occur during synthesis. This activity is crucial for maintaining the stability and fidelity of the viral genome (Ogando *et. al.,* 2020). Thus, nsp14-exoribonuclease domain (PDB id: 7DIY; Lin *et. al.,* 2021) was taken for molecular docking to filter potential hits depending on their binding affinity to the protein.

Fig 7: 6M71|SARS-Cov-2 RNA-dependent RNA polymerase in complex with cofactors

Target Protein structure validation:

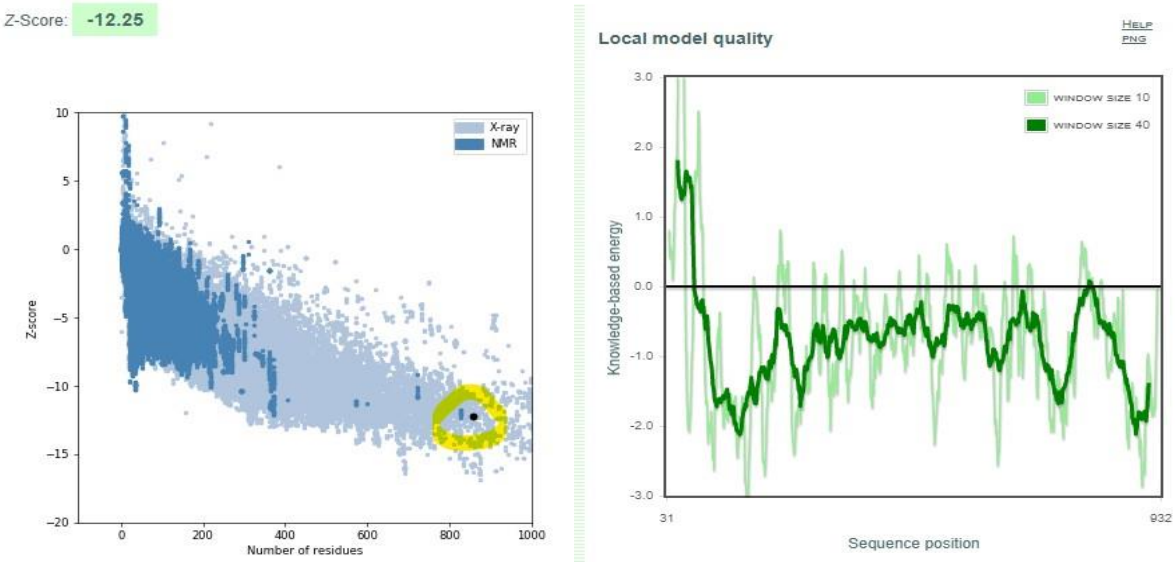
Z-score and Energy plot:

Understanding biological processes at a molecular level relies heavily on having a structural model of a protein. With recent advancements in experimental technology, large-scale structure determination pipelines have emerged, allowing for the rapid characterization of protein structures. Consequently, there is now an enormous amount of experimental structural information available. Additionally, computational methods for predicting unknown structures have added a plethora of structural models to the mix. In fact, the latest issue of the Nucleic Acids Research (NAR) web server lists approximately 50 tools in the "3D Structure Prediction" category (Fox *et. al.,* 2006). Routinely evaluating the accuracy and reliability of both experimental and theoretical models of protein structures is a crucial task that must be carried out to sustain the integrity, consistency, and reliability of public structure databases (Berman *et. al.,* 2006).

ProSA is a web server tool that has extensive use in detecting potential errors in 3D models of experimentally determined protein structures (Sippl, 1993). ProSA displays the quality score for a particular input structure in a plot that presents the scores of

all protein chains currently present in the Protein Data Bank (PDB) that have been experimentally determined (Berman *et. al.,* 2000). The website shows two traits of the input structure, namely its residue energies plot and its z-score (Wiederstein and Sippl, 2007). The z-score serves as an indicator of the model's overall quality and quantifies the difference between the total energy of the structure and the energy distribution obtained from random conformations. The z-score plot can help determine if the z-score of the protein being analyzed falls within the typical range of scores observed for proteins of similar size in one of the groups. The plot of energies as a function of amino acid sequence position provides information on the local quality of the model. Typically, areas of the model with positive values indicate problematic or erroneous regions (Sippl, 1995).

Electron microscopy structure of SARS-CoV-2 RdRp (PDB id: 6M71) was used as our target macromolecule for molecular docking studies. The structure was analysed using ProSA web server and subsequently z score and energy plot was determined.



(a) (b)

Fig 8: (a) *Z-Score plot of 6M71 predicted by proSA with structures available in database (b) Energy plot of 6M71 predicted by ProSA*

The z-score value of 6M71 was found to be -12.25 which lies within the range of z-score of native proteins. Additionally, energy plot showed amino acid of required

region occupying area with negative values. Thus, no significant error was recognized and thus the protein structure was validated.

Ramachandran plot:

The Ramachandran plot depicts the statistical distribution of backbone dihedral angle combinations ϕ and ψ. Essentially, the allowed regions of the plot indicate the feasible values of Phi/Psi angles for an amino acid, X, in an ala-X-ala tripeptide (Ramachandran *et. al.,* 1963). Analyzing the distribution of Phi/Psi values in a protein structure can be utilized as a means of validating its structure (Ramakrishnan *et. al.,* 2007). PROCHECK is web interface that is used to analyze unfavorable bond lengths and angles, as well as steric collisions and typically, a Ramachandran plot will indicate that a model is of good quality if around 90% of its residues fall within the allowable regions (laskowski *et. al.,* 1993).

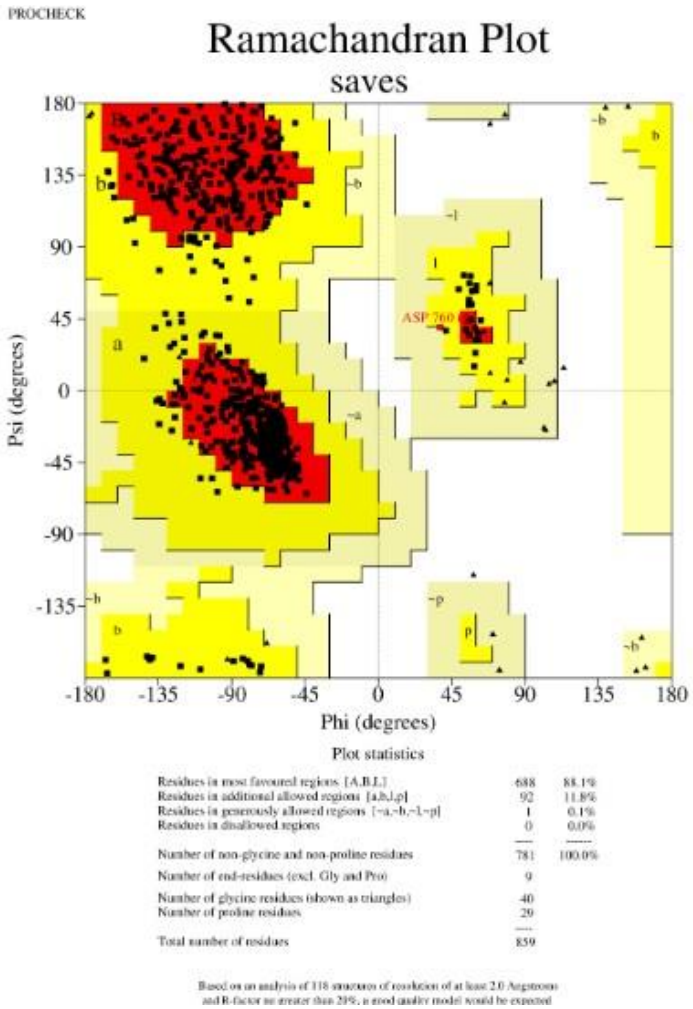


Fig 9: Ramachandran plot of SARS-CoV-2 RdRp electron microscope structure.

The plot showed that 88.1% of amino acid were present in most favoured region and 11.8 % residues were in additional allowed region. Thus, the protein structure model is acceptable to use for further molecular docking processes.

Target protein preparation and Acquisition of Binding sites:

The target protein structure in PDB formats were visualized via PyMol, which is a free molecular graphics tool. Additionally, PyMol was also further used in target protein processing and studying protein-ligand interaction (Yuan *et. al.,* 2017). Protein structure of SARS-CoV-2 RdRp (PDB id: 6M71) were fetched in PyMol and subsequently, the cofactors and water molecules were removed from the protein. Similarly, for hMAT1A (PDB ID: 6SW5) native ligands, ions, cofactors and water molecules were removed for further computational analysis. However, for nsp14-exoribonuclease domain (PDB id: 7DIY), a slightly different method was opted such that except for Mg2+ ion, all other cofactors, water molecules, ligands and other ions were removed. Retaining mg2+ ion for downstream molecular docking process is due to reliance of exonuclease activity of Nsp14 on metal ion, preferably Mg2+ (Tahir, 2021). After processing these protein structures, the molecules were exported in PDB formats for additional preparatory process. The additional target protein preparation includes removal of water, addition of hydrogen, merging non-polar hydrogen and computing Gasteiger charge using AutoDock tools which ensured the compatibility of protein for molecular docking studies (Morris *et. al.,* 2009). The target proteins were first saved in pdb file formats and then pdbqt format of target was prepared for performing molecular docking (Rizvi *et. al.,* 2013) using PyRx platform (Dallakyan and Olson, 2015).

After preparation of protein binding sites of protein were identified. For RdRp (PDB id:6M71), the binding sites were determined to be 545LYS, 555ARG, 623ASP, 682SER, THR687, ASN691, SER759, ASP760, ASP761 where 759-761 residues (SER, ASP, ASP) were conserved in most RdRp of coronaviruses (Gao *et. al.,* 2020). Similarly, binding site of nsp14-exoribonuclease (PDB id: 7DIY) was identified to be 90ASP, 92GLU, 191GLU, 243ASP, 268HIS, 273ASP and mg2+ ion (Tahir, 2021). However, for hMAT1A

(PDB ID: 6SW5), binding site were predicted and analyzed using 3D Ligand server and PyMol software. The amino acid residues within 5 angstroms of native ligands were identified using PyMol software, which helped in validation of computationally derived binding sites. The active sites of hMAT1A consists of 55ALA, 70GLU, 112GLN, 113GLN, 114SER, 133GLY, 134ASP, 289LYS, 291ASP residues.

A gridbox was made using PyRx 0.9.8 software which set the boundary of docking. The respective X, Y and Z coordinates and dimensions of gridbox was found to be Center; 115.067, 117.368 and 132.020 and dimensions; 13.887 Å, 16.098 Å and 25.225 Å for 6M71, centre; - 24.414, -3.368, -6.366 and dimensions; 23.800 Å, 18.674 Å, 10.200 Å for 7DIY, Center; 31.096, -0.571 and 24.983 and dimensions; 27.625 Å, 57.151 Å and 32.622 Å for 6SW5.

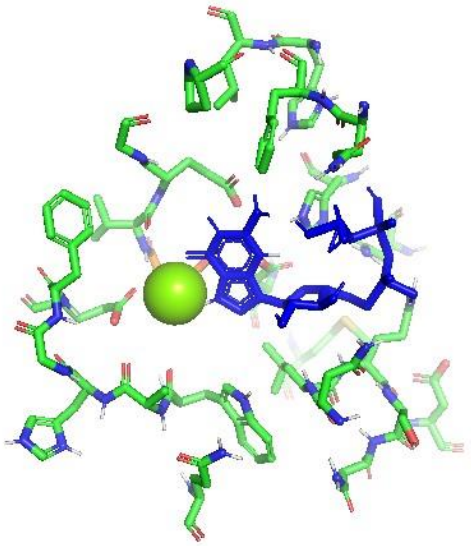


Fig 10: Amino acid residue with 5 Å of GTP bonded to 7DIY; blue colored molecule is native ligand, and the sphere corresponds to magnesium ion

Compound database preparation:

In order to prepare compound library against SARS-CoV-2 RdRp, various classes of compounds including natural products, small molecules, synthetic compounds, antivirals, nucleoside analogues were considered. ZINC15 (www.zinc15docking.org) database was used to prepare library of natural products and 21,332 compounds

under the categories such as FDA, Biogenic-FDA, in Man, in World, in Trial and *in vivo* were retrieved as such. UORSY database (https://uorsy.com/) was used to make a library of compounds including steroids, covalent modifiers, PPI modulators, Antivirals and general kinase inhibitors. In addition to that, ASINEX (https://www.asinex.com/) database was utilized to construct compound library including steroids, glycomimetics, building blocks, kinase inhibitors, antivirals, preplated microcyclics, nucleoside memetics, RNA targeting small molecules. These compounds were obtained in 2D or 3D structures from the databases in sdf file formats.

These compounds in sdf formats were screened based on 'Lipinski rules of 5' (Lipinsky *et. al.,* 2001) and veber's rule, which extends Lipinski rule to incorporate topological polar surface area and rotable bond count (Veber *et. al.,* 2002). 361 molecules were obtained upon screening 21,332 compounds from ZINC databases. In addition to that, 12,632 and 8,383 compounds were filtered from large drug library obtained from ASINEX and UORSY database respectively. Thus, a total of 21,376 compounds fulfilling the molecular properties parameters were obtained from three different databases which were then saved in dwar file format. Additionally, compound with 2d structure were converted to 3d structure. The screening and conversion to 3d structure was done using the OSIRIS data warrior (Sander et. al, 2015).

This screening of large database was possible due to use of more stringent parameters with slight modification in Lipinski and veber's rules, such as topological polar surface area of less than 120 and molecular weight from 200 to 500 dalton, toxicity and drug likeliness. These screening process filtered the ligand based on pharmacokinetic properties of the molecules otherwise known as ADME/ToX. Pharmacokinetics study helps to assess the journey of drug from its point of entry to site of action. Pharmacokinetic properties are defined as absorption, distribution, metabolism, excretion and toxicity (ADMET) (Schneider, 2013).

The Limitation of development of new drugs is largely due unfavorable pharmacokinetic properties. As such, *Insilico* tool helps to reduce these problems by predicting the ADMET properties of drugs. The significance of molecular weight lies in drug absorption which occurs via diffusion using paracellular and transcellular pathway. In paracellular pathway, limited molecules with molecular weight < 250 Da

are transported via tight junction and in transcellular pathway which is a major route of drug absorption, molecule with molecular wt > 300 and logP > 0 are absorbed through lipid membrane (El-Kattan and Varma, 2012).

ClogP or cumulative n-octanol/water partition coefficient of a chemical is correlated with the lipophilicity of organic substances, which plays a key role in the modulation of many key ADME processes. It has significant influence on various pharmacokinetic properties such as absorption, permeability and route of drug clearance (Schneider, 2013).

ClogS corresponds to 10-based logarithm of the solubility of a molecule measured in mol/L and correlates to water solubility of a drug. It has important role in drug uptake and elimination with poor soluble drugs eliminated before entering circulation and thus exhibiting no pharmacological activity (Delaney, 2005; Williams *et. al.,* 2013).

Hydrogen bonding donors/acceptors acts as a driving force which plays critical role in partitioning of biologically active compounds. This molecular property reflects the interaction between H-bond acceptor target and H-bond donor ligand or vice-versa. Additionally, In order to cross biological membrane, H-bonds between drug and aqueous environment must be broken. And so, it is unfavorable for a compound to make many H-bonds as it inversely affects degree of permeability and absorption (Chandrasekaran *et. al.,* 2018).

Topological polar surface area (TPSA) and Rotatable bond count plays important role in membrane permeation. Reduced molecular flexibility is measured by the number of rotatable bonds, and low polar surface area are found to be important predictors of good oral bioavailability (Veber *et. al.,* 2002).

After screening, energy-minimized structure of the filtered compounds using Universal Force Field was generated and they were further converted to pdbqt file format for molecular docking using PyRx (Dallakyan and Olson, 2015; EI Aissouq *et. al.,* 2021).

Molecular docking studies:

Docking with 6M71:

The probable inhibitory action of screened ligands was studied by performing molecular docking against SARS-CoV-2 RdRp. Due to large number of molecular compounds, natural products from zinc15, ASINEX and UORSY database were docked in turns. Molecular docking of 6M71 was done along with native ligand GTP using PyRx platform to facilitate comparative study of binding energy with respect to that of native ligand. Docking of natural product library along with native ligands resulted in no compounds with binding affinity higher than that of GTP. For UORSY ligand library, molecular docking revealed 2 compounds with binding affinity higher than that of native ligands. Furthermore, for ASINEX ligand library 56 compounds with binding affinity higher than that of GTP was obtained. Thus, from 21,376 screened ligands, only 58 compounds were revealed as potential hits against SARS-CoV-2 RdRp. However, upon visualizing the ligands structure in PyMol, structures of many ligands were disfigured or consists only 2D coordinates. The disfigurement of structure, especially in preplated microcyclics, might be due to large ring structure.

Redocking of potential hits:

To address the problems posed in preliminary docking studies, the potential hits from preliminary docking were identified and docked again against RDRP with parameter of 32 in both exhaustiveness and number of modes. While doing redocking all the hits from early docking studies were individually downloaded and their 3D structure were checked using visualization tool. Subsequently, upon completion of docking, structures of each ligand were visualized through PyMol to check significant changes in their structure. After redocking studies and visualization in protein-ligand interaction tool, only 7 compounds with their intact 3D structure showed binding affinity higher than that of native ligand. These 7 potential hits were used for further screening processes.

Table 4: Binding energy of potential hits against SARS-CoV-2 RdRp

ligand	Ligand source	ZINC ID	Binding energy against RdRp (kJ/mol)
Compound 17	Asinex preplated microcyclic	ZINC643558455	-8.4
Compound 1016	Asinex preplated microcyclic	ZINC643554068	-8.3
Compound 353	Asinex preplated microcyclic	ZINC643557756	-8.1
Compound 675	Asinex preplated microcyclic	ZINC643559409	-8.1
Compound 754	Asinex preplated microcyclic	ZINC643560167	-8.1
Compound 1018	Asinex preplated microcyclic	ZINC643587580	-8.1
LAS34154490	Asinex nucleoside mimetics	ZINC257274608	-8,1
Guanosine triphosphate (GTP)	PubChem	ZINC53684323	-8.0

Screening with ExoN, RTP and PI:

For further screening of compounds, 7 potential hits were again docked against ExoN domain of Nsp14 of SARS-CoV-2 in which GTP was again used as native ligand. Furthermore, Binding energy against RdRp were compared with RTP to further screen the potential hits. In addition to that, performance indexes of each compound were calculated based on number of hydrogen-bond acceptors/donors and rotatable bond

counts. Performance index gives measure of permeability through membrane, rate of absorption, molecular flexibility and residence time of molecule with in the protein target.

Table 5: Binding energy of potential hits against target compounds.

Ligands	**Binding energy (kJ/mol)**		
	RdRp	**ExoN**	**PI**
Compound 17	-8.4	-2.3	2.6
Compound 1016	-8.3	-6.2	2.6
Compound 353	-8.1	-5.0	2.0
Compound 675	-8.1	-0.9	2.8
Compound 754	-8.1	-4.8	2.6
Compound 1018	-8.1	-2.6	2.0
LAS34154490	-8.1	-5.6	3
GTP	-8.0	-5.5	
RTP	-8.3		

The above result shows the binding affinities and performance indexes of 7 potential hits. These results were used as basis to further screen the compounds to obtain most viable leads. Upon examining the binding energies against RdRp and Nsp14 ExoN compound 17 and compound 1016 showed the highest binding energy of -8.4 kJ/mol and -6.2 kJ/mol respectively. Compound 17 also shows higher binding energy against RdRp compared to -8.3 kJ/mol of Remdesivir triphosphate (RTP), which has been widely used in treatment of Covid 19 disease (Beigel *et. al.,* 2020). Besides that, compound 1016 could also be a promising drug candidate as it has high binding energy of -8.3 kJ/mol against RdRp and -6.2 kJ/mol against Nsp14-ExoN and its binding affinity is similar to that of RTP against SARS-CoV2 RdRp. Both compounds also have high

performance index suggesting that these molecules can be used as potential leads targeting RdRp of SARS-CoV-2.

Beside these two compounds, LAS34154490 can also be seen as potential lead molecule to treat Covid 19 disease. LAS34154490 was identified as nucleoside mimetic obtained from ASINEX library. Although, the binding energy of LAS34154490 lower than that of compound 17 and compound 1016 but it still has higher binding affinity of -8.1 kJ/mol compared to -8.0 kJ/mol of GTP and thus can compete with GTP for binding to SARS-CoV-2 RdRp. in addition to that, it also shows higher binding energy of -5.6 kJ/mol against Nsp14-ExoN domain compared to -5.5kJ/mol of GTP which further solidifies its position as promising drug candidate. When the binding affinity of a nucleoside mimetic is higher for Nsp14-ExoN domain compared to GTP, it could potentially inhibit viral replication by competing with GTP for binding to the ExoN domain (Malet *et. al.,* 2020). In one study, it was reported that "the high affinity of 2'-C-methylcytidine for Nsp14-ExoN domain suggests that it could be an effective inhibitor of viral replication by preventing the proofreading activity of Nsp14-ExoN" (Gordon et al., 2020). Thus, a nucleoside mimetic with a higher binding affinity for SARS-CoV-2 Nsp14-ExoN domain and RdRp compared to GTP has the potential to inhibit viral replication by preventing proofreading activity and competing with GTP for binding to the RNA polymerase. furthermore, the PI value of this compound is highest among the potential hits which suggests higher membrane permeation, absorption and longer residence time in target protein.

From the above results three compound namely compound 17, compound 1016 and LAS34154490 were chosen as potential leads. Therefore, these compounds were further studied for drug-protein interaction analysis.

Analysis of protein-ligand interaction:

Receptor-ligand complex visualization and Prediction of amino acids responsible for receptor-ligand interaction:

Three lead compounds screened from docking and PI analysis were further studied for understanding the interaction with the target receptor. Amino acids residue present at the active site are mainly involved in ligand-receptor interactions (Bugg, 2012). The flexibility inherent in biological macromolecules can modify the energy involved in identifying ligands by altering the configuration of the binding site (Di Cera, 2020). Thus, in order to predict the probable amino acids responsible for binding to the receptor, amino acid residue with in the 5 angstroms of ligands were determined using PyMol software. PyMol, a molecular graphic tool, offers a range of molecular modeling components, including tools for visualization and analysis enhancement, protein-ligand modeling, molecular simulations, and drug screening (Yuan *et. al.,* 2017). The prediction studies done via PyMol are depicted further in images below.

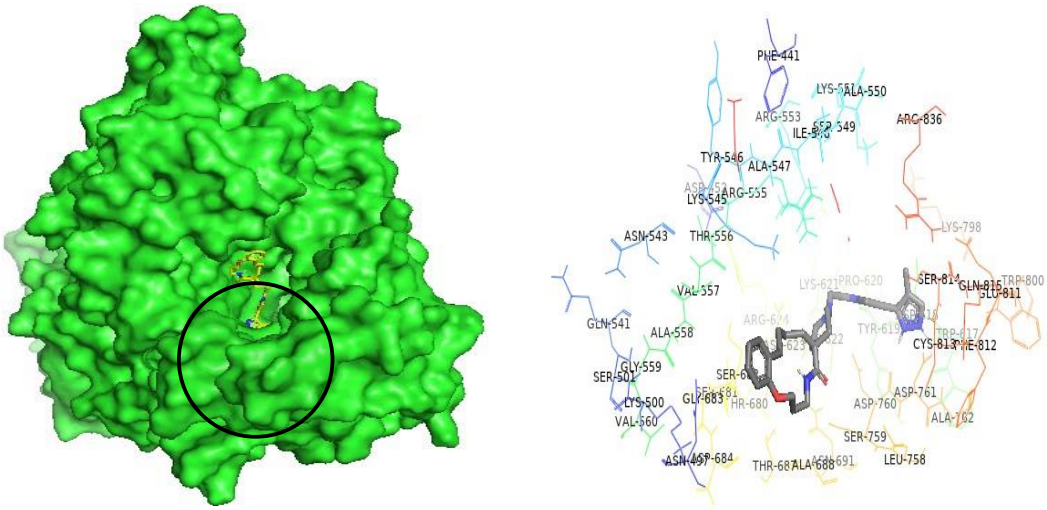


Fig 11: Receptor ligand complex of comp 17 shown as sticks with SARS-CoV-2 RdRp depicted by surface and amino acids residue of RdRp shown as lines within 5 Å of compound 17

Table 6: Amino acid residues within 5 Å of compound 17

Ligand	Amino acid residues within 5Å of ligand
Compound 17	PHE541, ASP542, ASN497, LYS500, SER501, GLN541, ASN543, LYS545, TYR546, ALA547, ILE548, SER549, ALA550, LYS551, ARG553, ARG555, THR556, VAL557, ALA558, GLY559, VAL560, TRP617, ASP618, TYR619, PRO620, LYS621, CYS622, ASP623, ARG624, THR680, GLY683, ASP684, THR687, ALA688, ASN691, LEU758, SER759, ASP760, ASP761, ALA762, LYS798, TRP800, GLU811, PHE812, CYS813, SER814, GLN815, ARG836, ASP845, ARG858

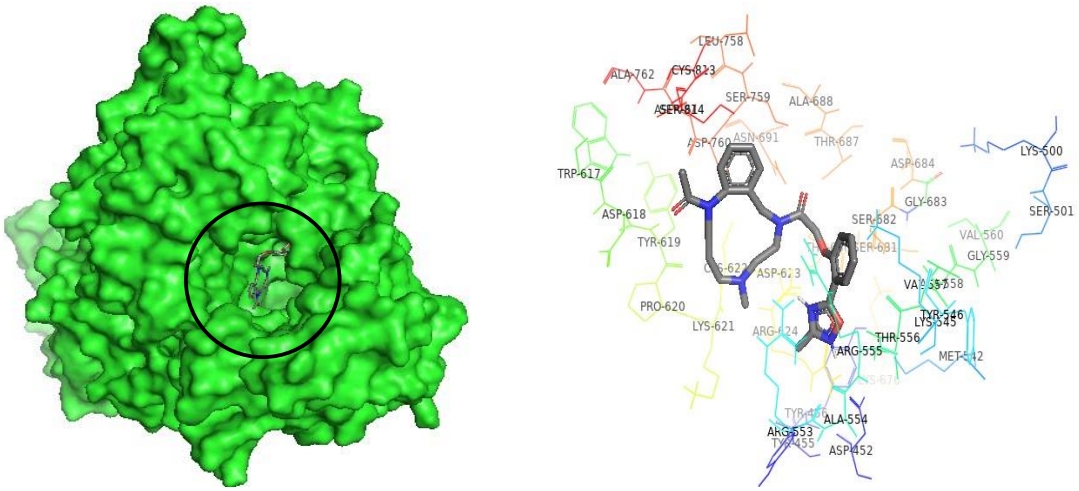


Fig 12: Receptor ligand complex of comp 1016 shown as sticks with SARS-CoV-2 RdRp depicted by surface and amino acids residue of RdRp shown as lines within 5 Å of compound 1016

Table 7: Amino acid residues within 5 Å of compound 1016

Ligand	**Amino acid residues within 5Å of ligand**
Compound 1016	ASP452, TYR455, TYR456, LYS500, SER501, ASP542, MET542, TYR546, ARG553, ALA554, ARG555, THR556, VAL557, ALA558, GLY559, VAL560, TRP617, ASP618, TYR619, PRO620, LYS621, CYS622, ASP623, ARG624, LYS676, CYS676, SER681, SER682, GLY683, ASP684, ALA688, ASN691, LEU758, SER759, ASP760, ASP761, ALA762, CYS813, SER814

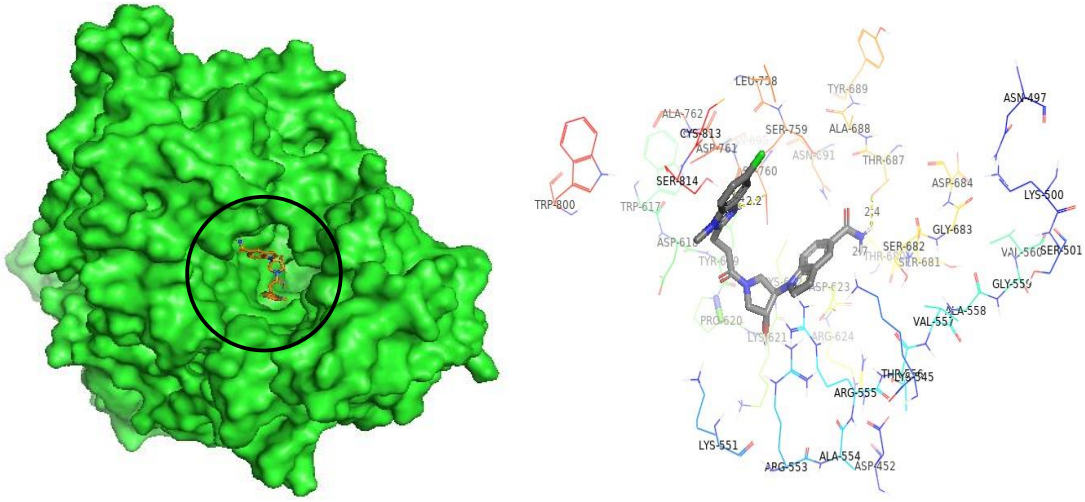


Fig 13: Receptor ligand complex of comp LAS34154490 shown as sticks with SARS-CoV-2 RdRp depicted by surface and amino acids residue of RdRp shown as lines within 5 Å of compound LAS34154490

Table 8: Amino acid residues within 5 Å of compound LAS34154490

Ligand	**Amino acid residues within 5Å of ligand**
LAS34154490	LYS438, HIS439, PHE440, PHE442, ILE548, SER549, ASP452, ASN497, LYS500, SER501, LYS545, ALA550, LYS551, ASN552, ARG553, ALA554, ARG555, THR556, VAL557, ALA558, GLY559, TRP617, ASP618, TYR619, PRO620, LYS621, CYS622, ASP623, ARG624, THR680, SER681, SER682, GLY683, ASP84, THR687, ALA688, TYR689, ASN691, LEU758, SER759, ASP760, ASP761, ALA62, TRP800, CYS813, SER814

The above results depicted that all the ligands were appropriately positioned in the binding cavity of the target receptor. The prediction of amino acid residues within 5Å

of ligands will act as a basis for validating interaction between ligand and amino acid otherwise not incorporated within gridbox prepared in molecular dynamic simulation. Thus, the evidence presented clearly indicates that the calculated binding energies between the ligand and SARS-CoV-2 RdRp correspond to the binding of the ligand within the amino acid residues of the enzyme's active site.

Analysis of interactions responsible for receptor/ligand binding:

In order to understand the mechanism responsible for higher affinities receptor/ligand binding, different interactions leading to bond formation and bond distance were analyzed. Discovery studio visualizer was used to analyze hydrogen bond, electrostatic and hydrophobic interactions (Studio, D., 2008). The hydrophobic interaction and other interaction were both generated 3D representations. The hydrophobic interactions were indicated by hydrophobic surfaces coded by colors. Brown surfaces indicated hydrophobic surfaces whereas blue surfaces indicated hydrophilic surfaces.

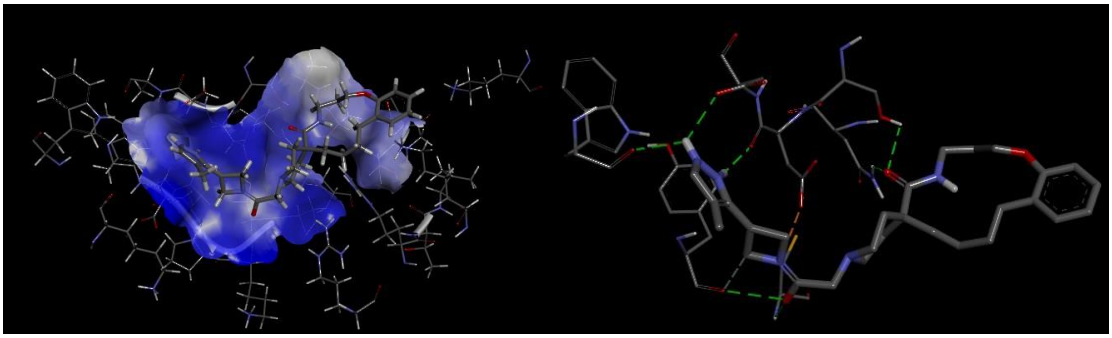


Fig 14: Receptor ligand interaction and hydrophobic interaction of compound 17 with SARS-CoV-2 RdRp imaged from Discovery Studio Visualizer

The hydrophobic interaction analysis of comp17 and SARS-CoV2 RdRp showed less hydrophobic areas. Hydrophobic regions are represented in brown color as per default colors spectrum of Discovery studio visualizer. However, no brown regions were observed while analysing interaction between comp17 and SARS-CoV2 RdRp. This result was further corroborated by data obtained on types of interaction, which showed hydrogen bond interaction as constitutive force to stabilize ligand/receptor binding.

Table 9: Types of interaction of compound 17 with active site amino residues of SARS-CoV-2 RdRp and bond distance datas obtained from Biovia Discovery studio visualizer

Chain	Amino acid	Interactions	Bond with	Bond length(Å)
A	TRP617	H-bond	Amino group of	2.45
A	TYR619	H-bond	Hydroxyl group	3.04
A	TYR619	Carbon hydrogen bond	Carbon in ring4	3.09
A	CYS622	H-bond	Hydroxyl group	2.38
A	ASN691	H-bond	Carbonyl group of ring 2	2.59
A	SER759	H-bond	Carbonyl group of ring 2	2.26
A	ASP760	H-bond	Amino group in substituted pyrazole	2.07
A	ASP760	Attractive charge	Tertiary Amine group in ring4	3.47
A	ASP761	H-bond	Amino group in substituted pyrazole	2.44

Seven hydrogen bonds were formed between compound 17 and TRP617, TYR619, CYS622, ASN691, SER759, ASP760 and ASP761 with corresponding bond length of 2.45Å, 3.04 Å, 2.38 Å, 2.59 Å, 2.26 Å, 2.07 Å, and 2.44 Å. The hydrogen bond length

below 3.3 Å suggests strong interaction between two moieties involved in the process (McRee, 1999). Additionally attractive interaction was observed between negative charge of ASP760 and positive charge present in tertiary amine structure of compound 17 with bond length of 3.47 Å.

Thus, from the result obtained upon analyzing different types of interaction and their corresponding bond length between active site amino acid residues of SARS-CoV-2 RdRp and different structural group of compound 17, it can be said that strong hydrogen bond and attractive forces were responsible for higher binding energy of -8.4 kJ/mol between compound 17 and SARS-CoV-2 RdRp.

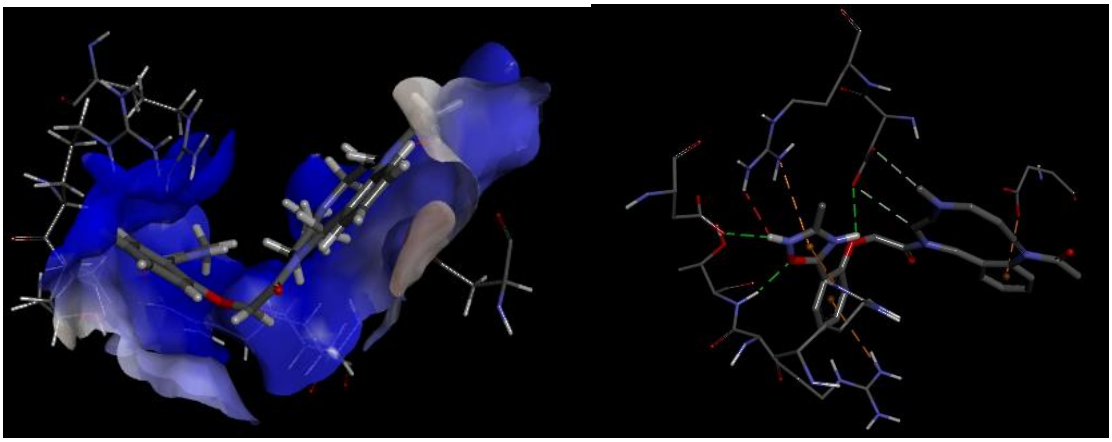


Fig 15: Receptor ligand interaction and hydrophobic interaction of compound 1016 with SARS-CoV-2 RdRp imaged from Discovery Studio Visualizer

Table 10: types of interaction of compound 1016 with active site amino residues of SARS-CoV-2 RdRp and bond distance datas obtained from Biovia Discovery studio visualizer

Chain	Amino acid	Interactions	Bond with	Bond length(Å)
A	ASP452	H-bond	Amino group in ring of substituted oxadiazolidine	2.44
A	ARG553	Pi-cation	Substituted oxadiazolidine	3.38

A	ARG555	Pi-cation	Substituted Benzene	4.40
A	THR556	H-bond	Oxo group of substituted oxadiazolidine	2.12
A	ASP623	H-bond	Amino group in ring of substituted oxadiazolidine	2.66
A	ASP623	Carbon Hydrogen bond	Amino group in ring 2	3.28
A	ASP623	Carbon Hydrogen bond	Methyl group in ring 2	3.53
A	ARG624	Pi-cation	substituted oxadiazolidine	4.18
A	ARG624	Unfavorable donor-donor	Amino group in ring of substituted oxadiazolidine	2.39
A	ASP760	Pi-anion	Substituted benzene	3.2

The 3d representation of Hydrophobic interaction between compound 1016 and SARS-CoV2 RdRp as shown in figure showed brownish white spots suggesting weak hydrophobic region which might play a role in stabilizing the protein/ligand complex. Additionally, bonding pattern of compound 1016 with SARS-CoV-2 showed hydrogen bond, pi-cation, pi-anion, carbon hydrogen bond and unfavorable donor-donor interactions as shown in table above.

Three hydrogen bond interactions were observed between compound 1016 with ASP452, THR556, ASP623 residues of SARS-CoV-2. The corresponding bond lengths

were found to be 2.44 Å, 2.12 Å, 2.66 Å. As bond length was smaller than 3.3 Å, it can be concluded that these hydrogen bond will result in tight binding of ligand with the target protein. Furthermore, one pi-anion interaction was seen between compound 1016 and ASP760 residue of SARS-CoV-2 RdRp with bond length of 3.2 Å. Anion-pi interactions refer to the advantageous non-covalent contacts established between an anion and an electron deficient (π-acidic) aromatic system (Schottel *et. al.,* 2008). Although anion-π bonds are often observed competing with other noncovalent interactions, they remain important due to their ability to significantly impact the stability, reactivity, and selectivity of molecular systems (Anstöter *et. al.,* 2019) and usually pi-anion interaction below 3.6 Å are of significance (Wang and Wang, 2013). From the data of pi-anion interaction, the interaction of compound 1016 with ASP 760 residue of SARS-CoV-2 RdRp is of high significance. In addition to that, three pi-cation interaction were formed between compound 1016 and ARG553, ARG555, ARG624 residues of SARS-CoV-2 RdRp with respective bond length of 3.38 Å, 4.40 Å, 4.18 Å. The cation-π interactions were of electrostatic nature and arises from the attraction between the positively charged cation and the negatively charged electron cloud of the π system, making it one of the strongest noncovalent interactions (Mahadevi and Sastry, 2013). The bond distance threshold for energetically significant cation-π interactions varies depending on the molecules involved, with proposed ranges from less than 3.7 Å (Levitt and Perutz, 1988) to 6 Å (Davis and Dougherty, 2015). Thus, the bond length data for pi-cation interaction between compound 1016 and SARS-CoV-2 RdRp residue suggests that this strong force which will further strengthen the binding of ligand and protein. However, an unfavorable donor donor interaction with bond length 2.39 Å between compound 1016 and ARG624 might destabilize the binding of protein and ligand.

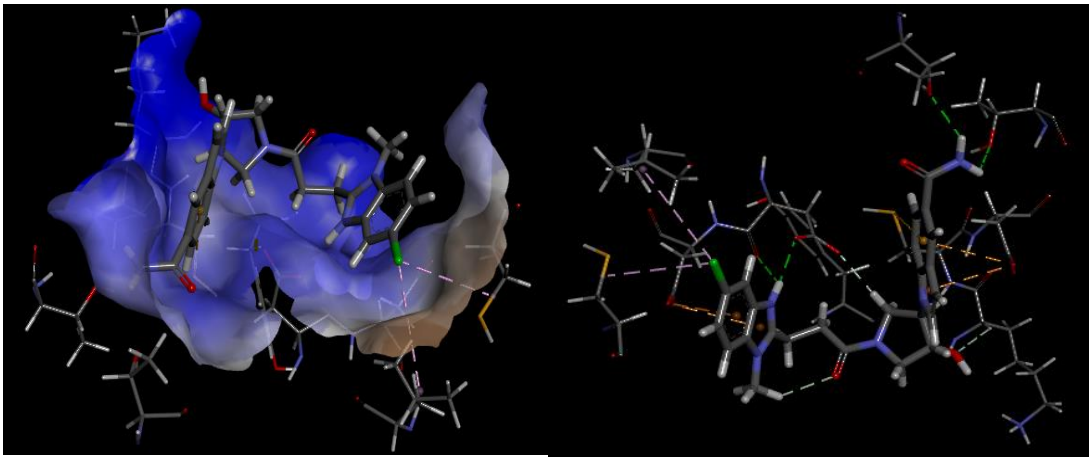


Fig 16: Receptor ligand interaction and hydrophobic interaction of LAS34154490 *with SARS-CoV-2 RdRp imaged from Discovery Studio Visualizer*

Table 11: Types of interaction of LAS34154490 with active site amino residues of SARS-CoV-2 RdRp and bond distance datas obtained from Biovia Discovery studio visualizer

Chain	**Amino acid**	**Interaction**	**Bond with**	**Bond length(Å)**
A	LYS621	H-bond	hydroxyl group	2.88
A	CYS622	Pi-Sulfur	Substituted benzene	5.04
A	ASP623	Pi-Anion	Substituted benzene	3.72
A	ASP623	Pi-Anion	Substituted Pyrrole	3.57
A	THR687	H-bond	Amido group	2.41
A	LEU758	Alkyl	Methyl group	4.87
A	ASP760	H-bond	Nitrogen of Imidazole	2.37

A	ASP760	H-bond	Hydrogen of imidazole	2.17
A	ASP761	Pi-Anion	Benzene of benzimidazole	3.23
A	ASP761	Pi-Anion	Imidazole ring	3.81
A	CYS813	Alkyl	Methyl group	4.31

The The 3d representation of Hydrophobic interaction between LAS34154490 and SARS-CoV2 RdRp as shown in figure showed brown spots suggesting favorable hydrophobic region which will pivotal role in stabilizing the protein/ligand complex. The hydrophobic interactions were depicted as two alkyl interaction as shown in figure. The presence of alkyl alkyl interaction is crucial in smectic phases formation which will result in microsegregation of the alkyl groups (Giese and Albrecht, 2020). Thus, this interaction will play major role in stabilizing of protein/ligand complex of LAS34154490 and SARS-CoV-2 RdRp.

Upon analyzing the bonding pattern of LAS34154490 molecule, four major interactions, namely Hydrogen bond, pi-anion, pi-sulphur and alkyl interactions were seen. Four hydrogen bond interaction between LAS34154490 and LYS621, THR687, ASP760, ASP760 residues of SARS-CoV-2 were found with respective bond length of 2.88 Å, 2.41 Å, 2.37 Å, 2.17 Å. The bond length of all these hydrogen bonds were lesser than 3.3 Å, which suggest tight bonding of LAS34154490 with SARS-CoV-2 RdRp. Furthermore, ASP623, ASP623, ASP761, ASP761 residues of SARS-CoV-2 were bonded with LAS34154490 using pi-anion interaction with respective bond length of 3.72 Å, 3.57 Å, 3.23 Å, 3.81 Å. However, aside from interaction of ASP761 with LAS34154490, other pi-anion interactions were weaker as their bond length were either comparable or greater than 3.6 Å which was upper limit value for strong pi-anion interaction.

Unique to interaction formed by other leads, pi-sulphur and alkyl interactions were seen between LAS34154490 and SARS-CoV-2 RdRp. The sulfur atom has an exceptional capability to interact with π systems uniformly throughout the entire

range (Motherwell *et. al.,* 2018). sulfur-aromatic interactions typically occur at distances below 7 Å, with a peak distance of approximately 5 Å (Gómez-Tamayo et al., 2016; Valley et al., 2012). Pi-sulphur interaction was seen between LAS34154490 and CYS622 residue of SARS-CoV-2 RdRp with bond length of 5.04 Å. This suggests that tight bonding will occur between LAS34154490 and RdRp due to pi-sulphur interactions. Additionally, hydrophobic interaction in the form of alky- alkyl interaction was seen between LAS34154490 and LEU758, CYS813 residues of SARS-CoV-2 RdRp with respective bond length of 4.87 Å and 4.32 Å. Hydrophobic interactions impact various water-related processes such as complexation, surfactant aggregation, and coagulation, and are important for stabilizing proteins (Bogunia and Markowski, 2020).

Interaction with mutated RdRp (P323L):

The lead compounds along with native ligand GTP were docked with P323L substituted RdRp of SARS-CoV-2, which was the most commonly found mutation existing in various RdRp variants. The binding energy was calculated using PyRX platform which are tabulated below.

Table 12: Binding energy against P323L RdRp

Ligands	**Binding energy (kJ/mol) against P323L RdRp**
Compound 17	-8.3
Compound 1016	-8.2
LAS34154490	-8.0
GTP	-7.9

Although the binding energy decreased for all lead compounds, however, it was found that all the molecules had higher binding energy when compared to native ligand against mutated P323L RdRp of SARS-CoV-2.

Interaction with human protein hMAT1A:

Three lead compounds were docked against hMAT1A with SAM as a native ligand to evaluate the cross reactivity of these lead compounds with human protein.

Table 13: Binding energy of lead compounds against hMAT1A

Ligands	Binding energy (kJ/mol) against hMAT1A
Compound 17	-9.8
Compound 1016	-10.3
LAS34154490	-7.9
SAM	-7.9

The result showed that compound 17 and compound 1016 showed higher binding energy of -9.8 kJ/mol and -10.3 kJ/mol against hMAT1A compared to -7.9 kJ/mol of SAM suggesting they might pose problem in SAM biosynthesis by competing against GTP to bind the enzyme. However, LAS34154490 showed equal binding energy of -7.9 kJ/mol against hMAT1A, thus suggesting less likeliness of this compound to compete for active site when compared to another molecule. Thus, although binding energy against SARS-CoV-2 RdRp is high for compound 17 and compound 1016, the additional properties of LAS34154490 including higher binding energy of -5.6 kJ/mol against ExoN domain of Nsp14 protein of SARS-CoV-2, high performance index of value 3 and lower cross reactivity against hMAT1A, infers the likeliness of this compound to be used for treatment against Covid-19 disease.

SUMMARY:

Covid-19 infection poses global threat due to its fatality and nature of spread. The rapid spread of coronavirus has caused global pandemic and impacted the globe negatively. In less than half a year the virus has spread nearly worldwide and several countries are facing outbreaks with emergence of new variants. Thus, this exhibits necessity of additional therapeutic interventions to control disease spread and speeding patients' recovery. Computer-aided drug design (CADD) is a new tool to ease discovery of new molecular entities and reduce later phase failures in drug discovery via early assessment of activity, selectivity and pharmacokinetic properties. In order to cope with the emerging threats of Covid-19 infection, virtual screening methods and molecular docking were employed to screen compound from different ligand databases. In this study, 21,376 compounds were filtered from large drug library of ZINC, ASINEX and UORSY database and further subjected to molecular docking against SARS-CoV-2 RNA dependent RNA polymerase (RdRp). Upon performing molecular docking along with the native ligand of RdRp, 56 compounds with binding affinity higher than that of GTP was obtained. These molecules were again redocked and further screening were done, which resulted to yield of three lead compounds. Among these three lead compounds, compound 17 and LAS34154490 was considered to be best. Compound 17 showed highest binding affinity against RdRp with ligand/receptor complexes forming strong interaction between them. Whereas, LAS34154490 showed binding energy higher than that of GTP against SARS-CoV-2 RdRp and Nsp14 Exonuclease domain. In addition to that, it also showed low hepatotoxicity when compared to other lead compound. Thus, with the use of computational tools three molecules with appropriate pharmacokinetic properties were determined, which might act as potential therapeutic option to combat Covid-19 infection.

CONCLUSION:

In conclusion, computer-aided drug design has provided an efficient and effective approach to identifying potential inhibitors of RdRp in SARS-CoV-2. Through the use of molecular docking, molecular dynamics simulations, and other computational tools, three compounds have been proposed as potential candidates for further development. These promising compounds have shown encouraging results in terms of their binding affinity, stability, and specificity for RdRp. However, further experimental studies, such as enzymatic assays and cell-based assays, are needed to verify the efficacy and safety of these compounds. Overall, the success of computer-aided drug design in identifying potential inhibitors of RdRp in SARS-CoV-2 highlights the importance of computational approaches in drug discovery and development, particularly in response to emerging infectious diseases such as COVID-19.

RECOMMENDATIONS:

It is advisable to validate the interaction between the SARS-CoV-2 RdRp and the top leads (compound 17 and LAS34154490) by performing molecular dynamics (MD) simulations. This compound shows promise as a potential drug candidate for further development. To solidify its candidacy as a suitable option for therapeutic use against SARS-CoV-2 RdRp, additional steps such as enzyme inhibition kinetic studies, animal testing, and toxicity testing can be pursued.

REFERENCES:

- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., & Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet. Infectious diseases*, *20*(5), 553–558. https://doi.org/10.1016/S1473-3099(20)30144-4
- Sharif, A., Aloui, C., & Yarovaya, L. (2020). COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach. *International Review of Financial Analysis*, *70*, 101496.
- Guan, L., Prieur, C., Zhang, L., Prieur, C., Georges, D., & Bellemain, P. (2020). Transport effect of COVID-19 pandemic in France. *Annual reviews in control*.
- Leroux-Roels, G., Bonanni, P., Tantawichien, T., & Zepp, F. (2011). Vaccine development. *Perspectives in Vaccinology*, *1*(1), 115-150.
- Plotkin, S., Robinson, J. M., Cunningham, G., Iqbal, R., & Larsen, S. (2017). The complexity and cost of vaccine manufacturing–an overview. *Vaccine*, *35*(33), 4064-4071.
- Nordström, P., Ballin, M., & Nordström, A. (2022). Risk of infection, hospitalisation, and death up to 9 months after a second dose of COVID-19 vaccine: a retrospective, total population cohort study in Sweden. *The Lancet*, *399*(10327), 814-823.
- Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C., & Greyson, D. (2011). The cost of drug development: a systematic review. *Health policy*, *100*(1), 4-17.
- Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological reviews*, *66*(1), 334-395.
- Lin, X., Li, X., & Lin, X. (2020). A review on applications of computational methods in drug screening and design. *Molecules*, *25*(6), 1375.

- Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, *24*(13), i232-i240.
- Bakheet, T. M., & Doig, A. J. (2009). Properties and identification of human protein drug targets. *Bioinformatics*, *25*(4), 451-457.
- Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., & Tramontano, A. (2018). Critical assessment of methods of protein structure prediction (CASP)-Round XII. Proteins 86 Suppl 1: 7–15.
- Ayton, G. S., Noid, W. G., & Voth, G. A. (2007). Multiscale modeling of biomolecular systems: in serial and in parallel. *Current opinion in structural biology*, *17*(2), 192-198.
- Handing, K. B., Niedzialkowska, E., Shabalin, I. G., Kuhn, M. L., Zheng, H., & Minor, W. (2018). Characterizing metal-binding sites in proteins with X-ray crystallography. *Nature protocols*, *13*(5), 1062-1090.
- Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862-865.
- Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature protocols*, *11*(5), 905-919.
- Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews Drug discovery*, *3*(8), 673-683.
- Baby, K., Maity, S., Mehta, C. H., Suresh, A., Nayak, U. Y., & Nayak, Y. (2020). Targeting SARS-CoV-2 RNA-dependent RNA polymerase: An *in silico* drug repurposing for COVID-19. *F1000Research*, *9*, 1166.
- Ullrich, S., & Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorganic & medicinal chemistry letters*, *30*(17), 127377.
- Peng, Y., Du, N., Lei, Y., Dorje, S., Qi, J., Luo, T., Gao, G. F., & Song, H. (2020). Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *The EMBO journal*, *39*(20), e105938. https://doi.org/10.15252/embj.2020105938

- Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., Zheng, M., Chen, L., & Li, H. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta pharmaceutica Sinica. B, 10*(5), 766–788. https://doi.org/10.1016/j.apsb.2020.02.008
- Malone, B., Urakova, N., Snijder, E. J., & Campbell, E. A. (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*, *23*(1), 21-39.
- Elfiky, A. A. (2021). SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: An in silico perspective. *Journal of Biomolecular Structure and Dynamics*, *39*(9), 3204-3212.
- Singh, S., Sk, M. F., Sonawane, A., Kar, P., & Sadhukhan, S. (2021). Plant-derived natural polyphenols as potential antiviral drugs against SARS-CoV-2 via RNA-dependent RNA polymerase (RdRp) inhibition: an in-silico analysis. *Journal of Biomolecular Structure and Dynamics*, *39*(16), 6249-6264.
- Shannon, A., Le, N. T. T., Selisko, B., Eydoux, C., Alvarez, K., Guillemot, J. C., Decroly, E., Peersen, O., Ferron, F., & Canard, B. (2020). Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *Antiviral research*, *178*, 104793. https://doi.org/10.1016/j.antiviral.2020.104793
- Wahl, A., Gralinski, L. E., Johnson, C. E., Yao, W., Kovarova, M., Dinnon III, K. H., Liu, H., Madden, V. J., Krzystek, H. M., De, C., White, K. K., Gully, K., Schafer, A., Zaman, T., Leist, S. R., Grant, P. O., Bluemling, G. R., Kolykhalov, A. A., Natchus M. G., Askin, F. B., Painter, G., Browne E. P., Jones, C. D., Pickles, R. J., Baric, R. S., & Garcia, J. V. (2021). SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature*, *591*(7850), 451-457. https://doi.org/10.1038/s41586-021-03312-w
- Singh, N. A., Kumar, P., & Kumar, N. (2021). Spices and herbs: potential antiviral preventives and immunity boosters during COVID-19. *Phytotherapy Research*, *35*(5), 2745-2757.

- Hall, M. D., Anderson, J. M., Anderson, A., Baker, D., Bradner, J., Brimacombe, K. R., Campbell, E. A., Corbett, K. S., Cate, K., Cherry, S., Chiang, L., Cihlar T., de Wit, E., Denison, M., Disney, M., Fletcher, C. V., Ford-Scheimer, S. L., Götte, M., Grossman, A. C., Hayden, F., Hazuda, D. J., Lanteri, C. A., Marston, H., Mesecar, A. D., Moore, S., Nwankwo, J. O., O'Rear, J., Painter, G., Saikatendu, K. S., Schiffer, C. A., Sheahan, T. P., Shi, P. Y., Smyth, H. D., Sofia, M. J., Weetall, M., Weller, S. K., Whitley, R., Fauci, A. S., Austin, C. P., Collins, F. S., Conley, A. J., & Davis, M. I. (2021). Report of the national institutes of health SARS-CoV-2 antiviral therapeutics summit. *The Journal of Infectious Diseases*, *224*(Supplement_1), S1-S21. https://doi.org/10.1093/infdis/jiab305
- Zumla, A., Chan, J. F., Azhar, E. I., Hui, D. S., & Yuen, K. Y. (2016). Coronaviruses—drug discovery and therapeutic options. *Nature reviews Drug discovery*, *15*(5), 327-347.
- of the International, C. S. G. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature microbiology*, *5*(4), 536.
- Fehr, A. R., & Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses*, 1-23.
- Chan, J. F. W., Lau, S. K. P., & Woo, P. C. Y. (2013). The emerging novel Middle East respiratory syndrome coronavirus: the "knowns" and "unknowns". *Journal of the Formosan Medical Association*, *112*(7), 372-381.
- Channappanavar, R., Zhao, J., & Perlman, S. (2014). T cell-mediated immune response to respiratory coronaviruses. *Immunologic research*, *59*(1), 118-128.
- Cheng, V. C., Lau, S. K., Woo, P. C., & Yuen, K. Y. (2007). Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical microbiology reviews*, *20*(4), 660-694.
- Chan, Jasper FW, Susanna KP Lau, Kelvin KW To, Vincent CC Cheng, Patrick CY Woo, and Kwok-Yung Yuen. "Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease." *Clinical microbiology reviews* 28, no. 2 (2015): 465-522.

- Gretebeck, L. M., & Subbarao, K. (2015). Animal models for SARS and MERS coronaviruses. *Current opinion in virology*, *13*, 123-129.
- He, F., Deng, Y., & Li, W. (2020). Coronavirus disease 2019: What we know?. *Journal of medical virology*, *92*(7), 719-725.
- Enjuanes L, Brian D, Cavanagh D, Holmes K, Lai MMC, Laude H, Masters P, Rottier PJM, Siddell SG, Spaan WJM, Taguchi F, Talbot P (2000) Coronaviridae. In: Virus Taxonomy, Seventh Report of the International Committee on Taxonomy of Viruses (MHV van Regenmortel, CM Fauquet, DHL Bishop, EB Carstens, MK Estes, SM Lemon, J Maniloff, MA Mayo, DJ McGeoch, CR Pringle, RB Wickner, eds) Academic Press, San Diego. pp 835–849
- Morris, D. R., & Geballe, A. P. (2000). Upstream open reading frames as regulators of mRNA translation. *Molecular and cellular biology*, *20*(23), 8635-8642.
- Brian, D. A., & Baric, R. S. (2005). Coronavirus genome structure and replication. *Coronavirus replication and reverse genetics*, 1-30.
- Brown, T. D. K., & Brierley, I. (1995). The coronavirus nonstructural proteins. In *The Coronaviridae* (pp. 191-217). Springer, Boston, MA.
- Zhao, L., Jha, B. K., Wu, A., Elliott, R., Ziebuhr, J., Gorbalenya, A. E., Silverman, R. H., & Weiss, S. R. (2012). Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. *Cell host & microbe*, *11*(6), 607–616. https://doi.org/10.1016/j.chom.2012.04.011
- 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://www.ncbi.nlm.nih.gov/pmc/articles/PMC4369385/figure/Fig1/
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., Zhao, K., Chen, Q. J., Deng, F., Liu, L. L., Yan, B., Zhan, F. X., Wang, Y. Y., Xiao, G. F., & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat

origin. *Nature*, *579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7

- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., Tan, W., & China Novel Coronavirus Investigating and Research Team (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England journal of medicine*, *382*(8), 727–733. https://doi.org/10.1056/NEJMoa2001017
- Kim, J. M., Chung, Y. S., Jo, H. J., Lee, N. J., Kim, M. S., Woo, S. H., Park, S., Kim, J. W., Kim, H. M., & Han, M. G. (2020). Identification of Coronavirus Isolated from a Patient in Korea with COVID-19. *Osong public health and research perspectives*, *11*(1), 3–7. https://doi.org/10.24171/j.phrp.2020.11.1.02
- Menachery, V. D., Graham, R. L., & Baric, R. S. (2017). Jumping species—a mechanism for coronavirus persistence and survival. *Current opinion in virology*, *23*, 1-7.
- Li, X., Song, Y., Wong, G., & Cui, J. (2020). Bat origin of a new human coronavirus: there and back again. *Science China. Life Sciences*, *63*(3), 461.
- Park, W. B., Kwon, N. J., Choi, S. J., Kang, C. K., Choe, P. G., Kim, J. Y., Yun, J., Lee, G. W., Seong, M. W., Kim, N. J., Seo, J. S., & Oh, M. D. (2020). Virus Isolation from the First Patient with SARS-CoV-2 in Korea. *Journal of Korean medical science*, *35*(7), e84. https://doi.org/10.3346/jkms.2020.35.e84
- Finlay, B. B., See, R. H., & Brunham, R. C. (2004). Rapid response research to emerging infectious diseases: lessons from SARS. *Nature Reviews Microbiology*, *2*(7), 602-607.
- Beniac, D. R., Andonov, A., Grudeski, E., & Booth, T. F. (2006). Architecture of the SARS coronavirus prefusion spike. *Nature structural & molecular biology*, *13*(8), 751-752.
- Collins, A. R., Knobler, R. L., Powell, H., & Buchmeier, M. J. (1982). Monoclonal antibodies to murine hepatitis virus-4 (strain JHM) define the viral glycoprotein responsible for attachment and cell-cell fusion. *Virology*, *119*(2), 358-371.

- Abraham, S., Kienzle, T. E., Lapps, W., & Brian, D. A. (1990). Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. *Virology*, *176*(1), 296-301.
- De Groot, R. J., Luytjes, W., Horzinek, M. C., Van der Zeijst, B. A. M., Spaan, W. J. M., & Lenstra, J. A. (1987). Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *Journal of molecular biology*, *196*(4), 963-966.
- Armstrong, J., Niemann, H., Smeekens, S., Rottier, P., & Warren, G. (1984). Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature*, *308*(5961), 751-752.
- Nal, B., Chan, C., Kien, F., Siu, L., Tse, J., Chu, K., Kam, J., Staropoli, I., Crescenzo-Chaigne, B., Escriou, N., van der Werf, S., Yuen, K. Y., & Altmeyer, R. (2005). Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *The Journal of general virology*, *86*(Pt 5), 1423–1434. https://doi.org/10.1099/vir.0.80671-0
- 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
- Godet, M., L'Haridon, R., Vautherot, J. F., & Laude, H. (1992). TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology*, *188*(2), 666-675.
- Nieto-Torres, J. L., DeDiego, M. L., Verdiá-Báguena, C., Jimenez-Guardeño, J. M., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodriguez, C., Alcaraz, A., Torres, J., Aguilella, V. M., & Enjuanes, L. (2014). Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS pathogens*, *10*(5), e1004077. https://doi.org/10.1371/journal.ppat.1004077
- Mukherjee, S., Bhattacharyya, D., & Bhunia, A. (2020). Host-membrane interacting interface of the SARS coronavirus envelope protein: Immense functional potential of C-terminal domain. *Biophysical chemistry*, *266*, 106452.

- Chang, C. K., Sue, S. C., Yu, T. H., Hsieh, C. M., Tsai, C. K., Chiang, Y. C., Lee, S. J., Hsiao, H. H., Wu, W. J., Chang, W. L., Lin, C. H., & Huang, T. H. (2006). Modular organization of SARS coronavirus nucleocapsid protein. *Journal of biomedical science*, *13*(1), 59–72. https://doi.org/10.1007/s11373-005-9035-9
- Hurst, K. R., Koetzner, C. A., & Masters, P. S. (2009). Identification of in vivo-interacting domains of the murine coronavirus nucleocapsid protein. *Journal of virology*, *83*(14), 7221-7234.
- Stohlman, S. A., & Lai, M. M. (1979). Phosphoproteins of murine hepatitis viruses. *Journal of virology*, *32*(2), 672-675.
- Stohlman, S. A., Baric, R. S., Nelson, G. N., Soe, L. H., Welter, L. M., & Deans, R. J. (1988). Specific interaction between coronavirus leader RNA and nucleocapsid protein. *Journal of virology*, *62*(11), 4288-4295.
- Kuo, L., & Masters, P. S. (2013). Functional analysis of the murine coronavirus genomic RNA packaging signal. *Journal of virology*, *87*(9), 5182-5192.
- Hurst, K. R., Koetzner, C. A., & Masters, P. S. (2013). Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *Journal of virology*, *87*(16), 9159-9172.
- Sturman, L. S., Holmes, K. V., & Behnke, J. (1980). Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *Journal of virology*, *33*(1), 449-462.
- Kumar, S., Nyodu, R., Maurya, V. K., & Saxena, S. K. (2020). Morphology, genome organization, replication, and pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In *Coronavirus Disease 2019 (COVID-19)* (pp. 23-31). Springer, Singapore.
- Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., Tan, K. S., Wang, D. Y., & Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Military Medical Research*, *7*(1), 11. https://doi.org/10.1186/s40779-020-00240-0

- Chan, J. F. W., Kok, K. H., Zhu, Z., Chu, H., To, K. K. W., Yuan, S., & Yuen, K. Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging microbes & infections*, *9*(1), 221-236.
- Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B. T., & Chaillon, A. (2020). Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *Journal of medical virology*, *92*(6), 602-611.
- 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., Meng, Y., Wang, J., Yang, L., Yuan, J., Xie, Z., Ma, J., Liu, W. J., Wang, D., Xu, W., Holmes, E. C., Gao, G. F., Wu, G., Chen, W., Shi, W., & 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
- Malone, B., Urakova, N., Snijder, E. J., & Campbell, E. A. (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*, *23*(1), 21-39.
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, *19*(3), 155-170.
- Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nature Reviews Microbiology*, *7*(3), 226-236.
- Lu, G., Wang, Q., & Gao, G. F. (2015). Bat-to-human: spike features determining 'host jump'of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends in microbiology*, *23*(8), 468-478.
- Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annual review of virology*, *3*, 237-261.

- He, Y., Zhou, Y., Liu, S., Kou, Z., Li, W., Farzan, M., & Jiang, S. (2004). Receptor-binding domain of SARS-CoV spike protein induces highly potent neutralizing antibodies: implication for developing subunit vaccine. *Biochemical and biophysical research communications*, *324*(2), 773-781.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, *181*(2), 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052
- 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
- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., & Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral research*, *176*, 104742.
- Matsuyama, S., & Taguchi, F. (2009). Two-step conformational changes in a coronavirus envelope glycoprotein mediated by receptor binding and proteolysis. *Journal of virology*, *83*(21), 11133-11141.
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3
- Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., ... & Stern-Ginossar, N. (2021). The coding capacity of SARS-CoV-2. *Nature*, *589*(7840), 125-130.
- Ziebuhr, J., Snijder, E. J., & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology*, *81*(4), 853-879.

- 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
- Malik, Y. A. (2020). Properties of coronavirus and SARS-CoV-2. *The Malaysian journal of pathology*, *42*(1), 3-11.
- Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., Cheng, J., Straub, J. H., Stürzel, C. M., Fröhlich, T., Berninghausen, O., Becker, T., Kirchhoff, F., Sparrer, K. M. J., & Beckmann, R. (2020). Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science (New York, N.Y.)*, *369*(6508), 1249–1255. https://doi.org/10.1126/science.abc8665
- Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L. A., Leibundgut, M., Thiel, V., Mühlemann, O., & Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nature structural & molecular biology*, *27*(10), 959–966. https://doi.org/10.1038/s41594-020-0511-8
- Cortese, M., Lee, J. Y., Cerikan, B., Neufeldt, C. J., Oorschot, V. M. J., Köhrer, S., Hennies, J., Schieber, N. L., Ronchi, P., Mizzon, G., Romero-Brey, I., Santarella-Mellwig, R., Schorb, M., Boermel, M., Mocaer, K., Beckwith, M. S., Templin, R. M., Gross, V., Pape, C., Tischer, C., Frankish, J., Horvat, N. K., Laketa, V., Stanifer, M., Boulant, S., Ruggieri, A., Chatel-Chaix, L., Schwab, Y., & Bartenschlager, R. (2020). Integrative Imaging Reveals SARS-CoV-2-Induced Reshaping of Subcellular Morphologies. *Cell host & microbe*, *28*(6), 853–866.e5. https://doi.org/10.1016/j.chom.2020.11.003
- Snijder, E. J., Limpens, R. W. A. L., de Wilde, A. H., de Jong, A. W. M., Zevenhoven-Dobbe, J. C., Maier, H. J., Faas, F. F. G. A., Koster, A. J., & Bárcena, M. (2020). A unifying structural and functional model of the coronavirus

replication organelle: Tracking down RNA synthesis. *PLoS biology*, *18*(6), e3000715. https://doi.org/10.1371/journal.pbio.3000715

- Ziebuhr, J. (2005). The coronavirus replicase. *Coronavirus replication and reverse genetics*, 57-94.
- Subissi, L., Posthuma, C. C., Collet, A., Zevenhoven-Dobbe, J. C., Gorbalenya, A. E., Decroly, E., Snijder, E. J., Canard, B., & Imbert, I. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(37), E3900–E3909. https://doi.org/10.1073/pnas.1323705111
- Kirchdoerfer, R. N., & Ward, A. B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature communications*, *10*(1), 1-9.
- Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., Ge, J., Zheng, L., Zhang, Y., Wang, H., Zhu, Y., Zhu, C., Hu, T., Hua, T., Zhang, B., Yang, X., Li, J., Yang, H., Liu, Z., Xu, W., Guddat, L. W., Wang, Q., Lou, Z., & Rao, Z. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science (New York, N.Y.)*, *368*(6492), 779–782. https://doi.org/10.1126/science.abb7498
- Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X., Wang, Q., Sun, Y., Fan, Z., Qi, J., Gao, G. F., & Shi, Y. (2020). Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. *Cell reports*, *31*(11), 107774. https://doi.org/10.1016/j.celrep.2020.107774
- Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., Chang, S., Xie, Y. C., Tian, G., Jiang, H. W., Tao, S. C., Shen, J., Jiang, Y., Jiang, H., Xu, Y., Zhang, S., Zhang, Y., & Xu, H. E. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science (New York, N.Y.)*, *368*(6498), 1499–1504. https://doi.org/10.1126/science.abc1560

- Hillen, H. S. (2021). Structure and function of SARS-CoV-2 polymerase. *Current Opinion in Virology*, *48*, 82-90.
- Krijnse-Locker J, Ericsson M, Rottier PJM et al (1994) Characterization of the budding compartment of mouse hepatitis virus: evidence that transport from the RER to the Golgi complex requires only one vesicular transport step. J Cell Biol 124:55–70
- Tooze J, Tooze S, Warren G (1984) Replication of coronavirus MHV-A59 in saccells: determination of the fi rst site of budding of progeny virions. Eur J Cell Biol 33:281–293
- Cohen, J. R., Lin, L. D., & Machamer, C. E. (2011). Identification of a Golgi complex-targeting signal in the cytoplasmic tail of the severe acute respiratory syndrome coronavirus envelope protein. *Journal of virology*, *85*(12), 5794-5803.
- Perrier, A., Bonnin, A., Desmarets, L., Danneels, A., Goffard, A., Rouillé, Y., Dubuisson, J., & Belouzard, S. (2019). The C-terminal domain of the MERS coronavirus M protein contains a *trans*-Golgi network localization signal. *The Journal of biological chemistry*, *294*(39), 14406–14421. https://doi.org/10.1074/jbc.RA119.008964
- de Haan CA, Rottier PJ (2005) Molecular interactions in the assembly of coronaviruses. Adv Virus Res 64:165–230
- Fung, T. S., & Liu, D. X. (2018). Post-translational modifications of coronavirus proteins: roles and function. *Future virology*, *13*(6), 405-430.
- Ghosh, S., Dellibovi-Ragheb, T. A., Kerviel, A., Pak, E., Qiu, Q., Fisher, M., Takvorian, P. M., Bleck, C., Hsu, V. W., Fehr, A. R., Perlman, S., Achar, S. R., Straus, M. R., Whittaker, G. R., de Haan, C. A. M., Kehrl, J., Altan-Bonnet, G., & Altan-Bonnet, N. (2020). β-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell*, *183*(6), 1520–1535.e14. https://doi.org/10.1016/j.cell.2020.10.039
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*, *20*(5), 533-534.

- 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. *The Lancet*, *395*(10225), 689-697.
- Xu, S., & Li, Y. (2020). Beware of the second wave of COVID-19. *The Lancet*, *395*(10233), 1321-1322.
- Gorbalenya, A. E., Krupovic, M., Mushegian, A., Kropinski, A. M., Siddell, S. G., Varsani, A., Adams, M., Davison, A., Dutilh, B., Harrach, B., Harrison, R., Junglen, S., King, A., Knowles, N., Lefkowitz, E., Nibert, M., Rubino, L., Sabanadzovic, S., Sanfaçon, H., Simmonds, P., Walker, P., Zerbini, F., & Kuhn, J. H. (2020). The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology*, *5*(5), 668-674. https://doi.org/10.1038/s41564-020-0709-x
- World Health Organization (2020a) Coronavirus disease (COVID-2019) situation reports. https://www.who.int/emergencies/diseases/novelcoronavirus-2019/situation-reports/
- 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.
- Tao, C., Guang, C., Wei, G., Min, X., Marco, L. Y., & Suhua, C. F. L. (2020). A quick guide to the diagnosis and treatment of pneumonia for novel coronavirus infections. *Herald Med*, *39*, 305-307.
- Gandhi, M., Yokoe, D. S., & Havlir, D. V. (2020). Asymptomatic transmission, the Achilles' heel of current strategies to control Covid-19. *New England Journal of Medicine*, *382*(22), 2158-2160.
- Meselson, M. (2020). Droplets and aerosols in the transmission of SARS-CoV-2. *New England Journal of Medicine*, *382*(21), 2063-2063.
- Morawska, L., & Cao, J. (2020). Airborne transmission of SARS-CoV-2: The world should face the reality. *Environment international*, *139*, 105730.

- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T. Y., Wu J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., & Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England journal of medicine*, *382*(13), 1199–1207. https://doi.org/10.1056/NEJMoa2001316
- Rubens, J. H., Karakousis, P. C., & Jain, S. K. (2020). Stability and viability of SARS-CoV-2. *N Engl J Med*, *382*(20), 1962-3.
- Van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., ... & Munster, V. J. (2020). Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England journal of medicine*, *382*(16), 1564-1567.
- Machhi, J., Herskovitz, J., Senan, A. M., Dutta, D., Nath, B., Oleynikov, M. D., Blomberg, W. R., Meigs, D. D., Hasan, M., Patel, M., Kline, P., Chang, R. C., Chang, L., Gendelman, H. E., & Kevadiya, B. D. (2020). The Natural History, Pathobiology, and Clinical Manifestations of SARS-CoV-2 Infections. *Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology*, *15*(3), 359–386. https://doi.org/10.1007/s11481-020-09944-5
- Moore, J. B., & June, C. H. (2020). Cytokine release syndrome in severe COVID-19. *Science*, *368*(6490), 473-474.
- Grifoni, A., Weiskopf, D., Ramirez, S. I., Mateus, J., Dan, J. M., Moderbacher, C. R., Rawlings, S. A., Sutherland, A., Premkumar, L., Jadi, R. S., Marrama, D., de Silva, A. M., Frazier, A., Carlin, A. F., Greenbaum, J. A., Peters, B., Krammer, F., Smith, D. M., Crotty, S., & Sette, A. (2020). Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell*, *181*(7), 1489–1501.e15. https://doi.org/10.1016/j.cell.2020.05.015

- Park, M. D. (2020). Macrophages: a Trojan horse in COVID-19?. *Nature Reviews Immunology*, *20*(6), 351-351.
- Yuki, K., Fujiogi, M., & Koutsogiannaki, S. (2020). COVID-19 pathophysiology: A review. *Clinical immunology*, *215*, 108427.
- Zhang, B., Zhou, X., Zhu, C., Song, Y., Feng, F., Qiu, Y., Feng, J., Jia, Q., Song, Q., Zhu, B., & Wang, J. (2020). Immune Phenotyping Based on the Neutrophil-to-Lymphocyte Ratio and IgG Level Predicts Disease Severity and Outcome for Patients With COVID-19. *Frontiers in molecular biosciences*, *7*, 157. https://doi.org/10.3389/fmolb.2020.00157
- Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R. C., Zella, D., & Ippodrino, R. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of translational medicine*, *18*(1), 179. https://doi.org/10.1186/s12967-020-02344-6
- Song, C. M., Lim, S. J., & Tong, J. C. (2009). Recent advances in computer-aided drug design. *Briefings in bioinformatics*, *10*(5), 579-591.
- Myers, S., & Baker, A. (2001). Drug discovery—an operating model for a new era. *Nature biotechnology*, *19*(8), 727-730.
- DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of health economics*, *22*(2), 151-185.
- DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of health economics*, *47*, 20-33.
- Vyas, P., & Vohora, D. (2018). Pharmaceutical Regulations for complementary Medicine. In *Pharmaceutical Medicine and Translational Clinical Research* (pp. 233-264). Academic Press.
- Zhong, F., Xing, J., Li, X., Liu, X., Fu, Z., Xiong, Z., Lu, D., Wu, X., Zhao, J., Tan, X., Li, F., Luo, X., Li, Z., Chen, K., Zheng, M., & Jiang, H. (2018). Artificial intelligence in drug design. *Science China. Life sciences*, *61*(10), 1191–1204. https://doi.org/10.1007/s11427-018-9342-2

- Hou, T., & Xu, X. (2004). Recent development and application of virtual screening in drug discovery: an overview. *Current pharmaceutical design*, *10*(9), 1011-1033.
- Klebe, G. (2006). Virtual ligand screening: strategies, perspectives and limitations. *Drug discovery today*, *11*(13-14), 580-594.
- Blundell, T. L., Dodson, G. G., Mercola, D., & Hodgkin, D. C. (1972). The structure, chemistry and biological activity of insulin. *AdvProteinChem*, *26*, 279-402.
- Beddell, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S., & Wootton, R. (1976). Compounds designed to fit a site of known structure in human haemoglobin. *British journal of pharmacology*, *57*(2), 201.
- Baig, M. H., Ahmad, K., Roy, S., Ashraf, J. M., Adil, M., Siddiqui, M. H., Khan, S., Kamal, M. A., Provazník, I., & Choi, I. (2016). Computer Aided Drug Design: Success and Limitations. *Current pharmaceutical design*, *22*(5), 572–581. https://doi.org/10.2174/1381612822666151125000550
- Veselovsky, A. V., & Ivanov, A. S. (2003). Strategy of computer-aided drug design. *Current drug targets. Infectious disorders*, *3*(1), 33–40.
- Surabhi, S., & Singh, B. K. (2018). Computer aided drug design: an overview. *Journal of Drug delivery and Therapeutics*, *8*(5), 504-509.
- Sun H. (2008). Pharmacophore-based virtual screening. *Current medicinal chemistry*, *15*(10), 1018–1024.
- Shoichet B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862–865
- 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 and Peptide Science*, *8*(4), 329-351.
- Mestres, J., & Knegtel, R. (2000). Similarity versus docking in 3D virtual screening. In *Virtual Screening: An Alternative or Complement to High Throughput Screening?* (pp. 191-207). Springer, Dordrecht.
- Mason, J. S., Good, A. C., & Martin, E. J. (2001). 3-D pharmacophores in drug discovery. *Current pharmaceutical design*, *7*(7), 567-597.

- Srinivasan, J., Castellino, A., Bradley, E. K., Eksterowicz, J. E., Grootenhuis, P. D., Putta, S., & Stanton, R. V. (2002). Evaluation of a novel shape-based computational filter for lead evolution: Application to thrombin inhibitors. *Journal of medicinal chemistry*, *45*(12), 2494-2500.
- Walters, W. P., Stahl, M. T., & Murcko, M. A. (1998). Virtual screening—an overview. *Drug discovery today*, *3*(4), 160-178.
- 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.
- Lengauer, T., Lemmen, C., Rarey, M., & Zimmermann, M. (2004). Novel technologies for virtual screening. *Drug discovery today*, *9*(1), 27–34.
- 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
- Sager, G., Ørvoll, E. Ø., Lysaa, R. A., Kufareva, I., Abagyan, R., & Ravna, A. W. (2012). Novel cGMP efflux inhibitors identified by virtual ligand screening (VLS) and confirmed by experimental studies. *Journal of medicinal chemistry*, *55*(7), 3049-3057.
- Sharma, R., Lawrenson, A. S., Fisher, N. E., Warman, A. J., Shone, A. E., Hill, A., Mbekeani, A., Pidathala, C., Amewu, R. K., Leung, S., Gibbons, P., Hong, D. W., Stocks, P., Nixon, G. L., Chadwick, J., Shearer, J., Gowers, I., Cronk, D., Parel, S. P., O'Neill, P. M., Ward, S. A., Biagini, G. A., & Berry, N. G. (2012). Identification of novel antimalarial chemotypes via chemoinformatic compound selection methods for a high-throughput screening program against the novel malarial target, PfNDH2: increasing hit rate via virtual screening methods. *Journal of medicinal chemistry*, *55*(7), 3144–3154. https://doi.org/10.1021/jm3001482
- Villoutreix, B. O., Eudes, R., & Miteva, M. A. (2009). Structure-based virtual ligand screening: recent success stories. *Combinatorial chemistry & high throughput screening*, *12*(10), 1000-1016.

- Anderson, A. C. (2003). The process of structure-based drug design. *Chemistry & biology*, *10*(9), 787-797.
- Li, Q., & Shah, S. (2017). Structure-based virtual screening. In *Protein Bioinformatics* (pp. 111-124). Humana Press, New York, NY.
- Wang, X., Song, K., Li, L., & Chen, L. (2018). Structure-based drug design strategies and challenges. *Current Topics in Medicinal Chemistry*, *18*(12), 998-1006.
- Lee, J., Freddolino, P. L., & Zhang, Y. (2017). Ab initio protein structure prediction. In *From protein structure to function with bioinformatics* (pp. 3-35). Springer, Dordrecht.
- Lemer, C. M. R., Rooman, M. J., & Wodak, S. J. (1995). Protein structure prediction by threading methods: evaluation of current techniques. *Proteins: Structure, Function, and Bioinformatics*, *23*(3), 337-355.
- Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology modeling a fast tool for drug discovery: current perspectives. *Indian journal of pharmaceutical sciences*, *74*(1), 1.
- Pitt, W. R., Calmiano, M. D., Kroeplien, B., Taylor, R. D., Turner, J. P., & King, M. A. (2013). Structure-based virtual screening for novel ligands. In *Protein-ligand interactions* (pp. 501-519). Humana Press, Totowa, NJ.
- Madhavi Sastry, G., 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.
- Batool, M., Ahmad, B., & Choi, S. (2019). A structure-based drug discovery paradigm. *International journal of molecular sciences*, *20*(11), 2783.
- Pan, L., Gardner, C. L., Pagliai, F. A., Gonzalez, C. F., & Lorca, G. L. (2017). Identification of the tolfenamic acid binding pocket in PrbP from Liberibacter asiaticus. *Frontiers in microbiology*, *8*, 1591.
- R Laurie, A. T., & Jackson, R. M. (2006). Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Current Protein and Peptide Science*, *7*(5), 395-406.

- Binkowski, T. A., Naghibzadeh, S., & Liang, J. (2003). CASTp: computed atlas of surface topography of proteins. *Nucleic acids research*, *31*(13), 3352-3355.
- Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, *28*(15), 2074-2075.
- Sun, J., & Chen, K. (2017). NSiteMatch: prediction of binding sites of nucleotides by identifying the structure similarity of local surface patches. *Computational and Mathematical Methods in Medicine*, *2017*.
- Tan, K. P., Varadarajan, R., & Madhusudhan, M. S. (2011). DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic acids research*, *39*(suppl_2), W242-W248.
- Zhu, H., & Pisabarro, M. T. (2011). MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics*, *27*(3), 351-358.
- Huang, B. (2009). MetaPocket: a meta approach to improve protein ligand binding site prediction. *OMICS A Journal of Integrative Biology*, *13*(4), 325-330.
- Laurie, A. T., & Jackson, R. M. (2005). Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics*, *21*(9), 1908-1916.
- Hergenrother, P. J. (2006). Obtaining and screening compound collections: a user's guide and a call to chemists. *Current opinion in chemical biology*, *10*(3), 213-218.
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, *40*(Database issue), D1100–D1107. https://doi.org/10.1093/nar/gkr777
- Irwin, J. J., & Shoichet, B. K. (2005). ZINC− a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, *45*(1), 177-182.

- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., & Bryant, S. H. (2016). PubChem Substance and Compound databases. *Nucleic acids research*, *44*(D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951
- Pence, H. E., & Williams, A. (2010). ChemSpider: an online chemical information resource.
- Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., & Hassanali, M. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, *36*(Database issue), D901–D906. https://doi.org/10.1093/nar/gkm958
- Chandrasekaran, B., Abed, S. N., Al-Attraqchi, O., Kuche, K., & Tekade, R. K. (2018). Computer-aided prediction of pharmacokinetic (ADMET) properties. In *Dosage form design parameters* (pp. 731-755). Academic Press.
- 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.
- 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
- Blagg, J. (2006). Structure–activity relationships for in vitro and in vivo toxicity. *Annual Reports in medicinal chemistry*, *41*, 353-368.
- López-López, E., Naveja, J. J., & Medina-Franco, J. L. (2019). DataWarrior: An evaluation of the open-source drug discovery tool. *Expert Opinion on Drug Discovery*, *14*(4), 335-341.
- Hung, C. L., & Chen, C. C. (2014). Computational approaches for drug discovery. *Drug development research*, *75*(6), 412-418.
- 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.

- 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*, 411-428.
- Bursulaya, B. D., Totrov, M., Abagyan, R., & Brooks, C. L., 3rd (2003). Comparative study of several algorithms for flexible ligand docking. *Journal of computer-aided molecular design*, *17*(11), 755–763.
- 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
- Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, *267*(3), 727-748.
- Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein–ligand docking: current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, *65*(1), 15-26.
- Oshiro, C. M., Kuntz, I. D., & Dixon, J. S. (1995). Flexible ligand docking using a genetic algorithm. *Journal of computer-aided molecular design*, *9*, 113-130.
- Hart, T. N., & Read, R. J. (1992). A multiple-start Monte Carlo docking method. *Proteins: Structure, Function, and Bioinformatics*, *13*(3), 206-222.
- Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, A. C. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British journal of pharmacology*, *153*(S1), S7-S26.
- Huang, S. Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics*, *12*(40), 12899-12908.

- Guedes, I. A., Pereira, F. S., & Dardenne, L. E. (2018). Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. *Frontiers in pharmacology*, *9*, 1089.
- Muegge, I. Pmf scoring revisited. J. Med. Chem. 2006, 49, 5895–5902
- Li, H., Peng, J., Leung, Y., Leung, K. S., Wong, M. H., Lu, G., & Ballester, P. J. (2018). The impact of protein structure and sequence similarity on the accuracy of machine-learning scoring functions for binding affinity prediction. *Biomolecules*, *8*(1), 12.
- Hecht, D., & Fogel, G. B. (2009). Computational intelligence methods for docking scores. *Current Computer-Aided Drug Design*, *5*(1), 56-68.
- 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*, *28*(22), 3002-3003.
- Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., & Chong, J. (2016). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic acids research*, *44*(D1), D1045-D1053.
- Anighoro, A., Bajorath, J., & Rastelli, G. (2014). Polypharmacology: challenges and opportunities in drug discovery: miniperspective. *Journal of medicinal chemistry*, *57*(19), 7874-7887.
- Kabinger, F., Stiller, C., Schmitzová, J., Dienemann, C., Kokic, G., Hillen, H. S., Höbartner, C., & Cramer, P. (2021). Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nature structural & molecular biology*, *28*(9), 740–746. https://doi.org/10.1038/s41594-021-00651-0
- Ahn, D. G., Shin, H. J., Kim, M. H., Lee, S., Kim, H. S., Myoung, J., Kim, B. T., & Kim, S. J. (2020). Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). *Journal of microbiology and biotechnology*, *30*(3), 313–324. https://doi.org/10.4014/jmb.2003.03011

- Fox, J. A., McMillan, S., & Ouellette, B. F. (2006). A compilation of molecular biology web servers: 2006 update on the Bioinformatics Links Directory. *Nucleic Acids Research*, *34*(suppl_2), W3-W5.
- Berman, H. M., Burley, S. K., Chiu, W., Sali, A., Adzhubei, A., Bourne, P. E., Bryant, S. H., Dunbrack, R. L., Jr, Fidelis, K., Frank, J., Godzik, A., Henrick, K., Joachimiak, A., Heymann, B., Jones, D., Markley, J. L., Moult, J., Montelione, G. T., Orengo, C., Rossmann, M. G., Rost, B., Saibil, H., Schwede, T., Standley, D. M., & Westbrook, J. D. (2006). Outcome of a workshop on archiving structural models of biological macromolecules. *Structure (London, England : 1993)*, *14*(8), 1211–1217. https://doi.org/10.1016/j.str.2006.06.005
- Vicenti, I., Zazzi, M., & Saladini, F. (2021). SARS-CoV-2 RNA-dependent RNA polymerase as a therapeutic target for COVID-19. *Expert opinion on therapeutic patents*, *31*(4), 325-337.
- Hashemian, S. M. R., Pourhanifeh, M. H., Hamblin, M. R., Shahrzad, M. K., & Mirzaei, H. (2022). RdRp inhibitors and COVID-19: Is molnupiravir a good option?. *Biomedicine & Pharmacotherapy*, *146*, 112517.
- Avila, M. A., Berasain, C., Prieto, J., Mato, J. M., Garcia-Trevijano, E. R., & Corrales, F. J. (2005). Influence of impaired liver methionine metabolism on the development of vascular disease and inflammation. *Current Medicinal Chemistry-Cardiovascular & Hematological Agents*, *3*(3), 267-281.
- Grimble, R. F., & Grimble, G. K. (1998). Immunonutrition: role of sulfur amino acids, related amino acids, and polyamines. *Nutrition*, *14*(7-8), 605-610.
- Panmanee, J., Antonyuk, S. V., & Hasnain, S. S. (2020). Structural basis of the dominant inheritance of hypermethioninemia associated with the Arg264His mutation in the MAT1A gene. *Acta Crystallographica Section D: Structural Biology*, *76*(6), 594-607.
- Ogando, N. S., Zevenhoven-Dobbe, J. C., van der Meer, Y., Bredenbeek, P. J., Posthuma, C. C., & Snijder, E. J. (2020). The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *Journal of virology*, *94*(23), e01246-20.

- Lin, S., Chen, H., Chen, Z., Yang, F., Ye, F., Zheng, Y., Yang, J., Lin, X., Sun, H., Wang, L., Wen, A., Dong, H., Xiao, Q., Deng, D., Cao, Y., & Lu, G. (2021). Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic acids research*, *49*(9), 5382–5392. https://doi.org/10.1093/nar/gkab320
- Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, *17*(4), 355-362.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic acids research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(suppl_2), W407-W410.
- Sippl, M. J. (1995). Knowledge-based potentials for proteins. *Current opinion in structural biology*, *5*(2), 229-235.
- Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, *7*(1), 95-99.
- Ramakrishnan, C., Lakshmi, B., Kurien, A., Devipriya, D., & Srinivasan, N. (2007). Structural compromise of disallowed conformations in peptide and protein structures. *Protein and Peptide Letters*, *14*(7), 672-682.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, *26*(2), 283-291.
- Yuan, S., Chan, H. S., & Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *7*(2), e1298.
- Tahir, M. (2021). Coronavirus genomic nsp14-ExoN, structure, role, mechanism, and potential application as a drug target. *Journal of Medical Virology*, *93*(7), 4258-4264.

- 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.
- Rizvi, S. M. D., Shakil, S., & Haneef, M. (2013). A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians. *EXCLI journal*, *12*, 831.
- Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Chemical biology: methods and protocols*, 243-250.
- Sander, T., Freyss, J., von Korff, M., & Rufener, C. (2015). DataWarrior: an open-source program for chemistry aware data visualization and analysis. *Journal of chemical information and modeling*, *55*(2), 460-473. https://doi.org/10.1021/ci500588j
- Schneider, G. (2013). Prediction of drug-like properties. In *Madame Curie Bioscience Database [Internet]*. Landes Bioscience.
- El-Kattan, A., & Varma, M. (2012). Oral absorption, intestinal metabolism and human oral bioavailability–Topics on drug metabolism, Dr. James Paxton. *InTech, Croatia, EU*.
- Williams, H. D., Trevaskis, N. L., Charman, S. A., Shanker, R. M., Charman, W. N., Pouton, C. W., & Porter, C. J. (2013). Strategies to address low drug solubility in discovery and development. *Pharmacological reviews*, *65*(1), 315-499.
- Delaney, J. S. (2005). Predicting aqueous solubility from structure. *Drug discovery today*, *10*(4), 289-295.
- Chandrasekaran, B., Abed, S. N., Al-Attraqchi, O., Kuche, K., & Tekade, R. K. (2018). Computer-aided prediction of pharmacokinetic (ADMET) properties. In *Dosage form design parameters* (pp. 731-755). Academic Press.
- 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.

- El Aissouq, A., Chedadi, O., Bouachrine, M., & Ouammou, A. (2021). Identification of novel SARS-CoV-2 inhibitors: A structure-based virtual screening approach. *Journal of Chemistry*, *2021*, 1-7.
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., Hohmann, E., Chu, H. Y., Luetkemeyer, A., Kline, S., Lopez de Castilla, D., Finberg, R. W., Dierberg, K., Tapson, V., Hsieh, L., Patterson, T. F., Paredes, R., Sweeney, D. A., Short, W. R., Touloumi, G., Lye, D. C., Ohmagari, N., Oh, M. D., Ruiz-Palacios, G. M., Benfield, T., Fätkenheuer, G., Kortepeter, M. G., Atmar, R. L., Creech C. B., Lundgren, J., Babiker, A. G., Pett, S., Neaton, J. D., Burgess, T. H., Bonnett, T., Green, M., Makowski, M., Osinusi, A., Nayak, S., Lane, H. C., & ACTT-1 Study Group Members (2020). Remdesivir for the Treatment of Covid-19 - Final Report. *The New England journal of medicine*, *383*(19), 1813–1826. https://doi.org/10.1056/NEJMoa2007764
- Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuman, B. W., & Canard, B. (2021). The crystal structure of the nsp14-N60 exonuclease domain from the COVID-19 virus in complex with nucleoside analogs reveals a conserved mechanism among coronaviruses and a potential therapeutic target. *Journal of Virology*, *95*(15), e0039421.
- Gordon, C. J., Tchesnokov, E. P., Woolner, E., Kocinkova, D., Perry, J. K., Feng, J. Y., & Gotte, M. (2020). Remdesivir- and 2'-C-methylcytidine-induced inhibition of the SARS-CoV-2 polymerase will suppress viral replication. *Nature Communications, 11*(1), 1-11.
- Bugg, T. D. (2012). *Introduction to enzyme and coenzyme chemistry*. John Wiley & Sons.
- Di Cera, E. (2020). Mechanisms of ligand binding. *Biophysics Reviews*, *1*(1), 011303.
- Studio, D. (2008). Discovery studio. *Accelrys [2.1]*.
- McRee, D. E. (1999). *Practical protein crystallography*. Elsevier.
- Schottel, B. L., Chifotides, H. T., & Dunbar, K. R. (2008). Anion-π interactions. Chemical Society Reviews, 37(1), 68-83.

- Anstöter, C. S., Rogers, J. P., & Verlet, J. R. (2019). Spectroscopic determination of an anion− π bond strength. *Journal of the American Chemical Society*, *141*(15), 6132-6135.
- Wang, D. X., & Wang, M. X. (2013). Anion− π interactions: generality, binding strength, and structure. *Journal of the American Chemical Society*, *135*(2), 892-897.
- Mahadevi, A. S., & Sastry, G. N. (2013). Cation− π interaction: Its role and relevance in chemistry, biology, and material science. *Chemical reviews*, *113*(3), 2100-2138.
- Levitt, M., & Perutz, M. F. (1988). Aromatic rings act as hydrogen bond acceptors. *Journal of molecular biology*, *201*(4), 751-754.
- Davis, M. R., & Dougherty, D. A. (2015). Cation–π interactions: computational analyses of the aromatic box motif and the fluorination strategy for experimental evaluation. *Physical Chemistry Chemical Physics*, *17*(43), 29262-29270.
- Giese, M., & Albrecht, M. (2020). Alkyl-Alkyl Interactions in the Periphery of Supramolecular Entities: From the Evaluation of Weak Forces to Applications. *ChemPlusChem*, *85*(4), 715-724.
- Motherwell, W. B., Moreno, R. B., Pavlakos, I., Arendorf, J. R. T., Arif, T., Tizzard, G. J., Coles, S. J., & Aliev, A. E. (2018). Noncovalent Interactions of π Systems with Sulfur: The Atomic Chameleon of Molecular Recognition. *Angewandte Chemie (International ed. in English)*, *57*(5), 1193–1198. https://doi.org/10.1002/anie.201708485
- Gómez-Tamayo, J. C., Cordomí, A., Olivella, M., Mayol, E., Fourmy, D., & Pardo, L. (2016). Analysis of the interactions of sulfur-containing amino acids in membrane proteins. *Protein Science*, *25*(8), 1517-1524.
- Valley, C. C., Cembran, A., Perlmutter, J. D., Lewis, A. K., Labello, N. P., Gao, J., & Sachs, J. N. (2012). The methionine-aromatic motif plays a unique role in stabilizing protein structure. *Journal of Biological Chemistry*, *287*(42), 34979-34991.

- Bogunia, M., & Makowski, M. (2020). Influence of ionic strength on hydrophobic interactions in water: dependence on solute size and shape. *The Journal of Physical Chemistry B*, *124*(46), 10326-10336.

iThenticate

Insilico drug development agains

Insilico drug development against RNA dependent RNA polymerase of SARS-CoV-2

M.Sc. Thesis

2023

Submitted to

Central Department of Biotechnology

Tribhuvan University

Kirtipur, Kathmandu, Nepal

For partial fulfillment of M.Sc degree in Biotechnology

By

Siddartha Gautam

Registration No: 5-3-28-300-2017

Quotes Excluded
Bibliography Excluded

6%
SIMILAR

Match Overview

1	**Internet** 373 words crawled on 11-Apr-2022 pubannotation.org	2%
2	**Internet** 194 words crawled on 02-Jan-2023 www.mdpi.com	1%
3	**Internet** 144 words crawled on 03-May-2020 www.mjpath.org.my	1%
4	**Internet** 107 words crawled on 14-Dec-2020 link.springer.com	1%
5	**Internet** 107 words crawled on 23-Aug-2022 www.hindawi.com	1%
6	**Internet** 82 words crawled on 14-Dec-2022 www.bms.kr	1%