Inhibitor design for drug targets in Zika virus and SARS CoV-2 using computational studies

A Thesis

Submitted for the Degree of

DOCTOR OF PHILOSOPHY

By

Durgam Laxman

(Reg. No. 16CHPH06)

University of Hyderabad

INSTITUTION OF EMINENCE

School of Chemistry

University of Hyderabad

Hyderabad – 500046

INDIA

December 2022

DEDICATION

Words will not be enough to express my gratitude towards my role model my Mother Late **Durgam Jamalamma** who have struggled throughout her life and supported me. Because of her love, affection, hard work, and sacrifices, I am who I am today, and I dedicate this thesis to her soul.

University of Hyderabad
School of chemistry
Hyderabad-500046

STATEMENT

I hereby declaring that the matter embodied in this thesis is the results of investigations carried out by me in the school of chemistry, University of Hyderabad, Hyderabad under the supervision of **Prof. Lalitha Guruprasad**

In keeping with general practice of reporting scientific observations, due to acknowledgements have been made whenever the work described is based on the finding of other investigators

Hyderabad
December 2022

Durgam Laxman
16CHPH06

University of Hyderabad
School of chemistry
Hyderabad-500046

DECLARATION

I, Durgam Laxman declare that the thesis entitled "**Inhibitor design for drug targets in Zika virus and SARS CoV-2 using computational studies**" submitted by me under the supervision of Prof. Lalitha Guruprasad is a Bonafide research work which is free from plagiarisms. I also declare that it has not been submitted previously in part or full to this university or any other university or Institution for the award of any degree or diploma. I hereby agree that my thesis can be deposited in shodhganga/INFLIBNET.

A report on plagiarism from the university library enclosed.

Prof. Lalitha Guruprasad

(Thesis Supervisor)

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.

Durgam Laxman

Signature

Reg. No: 16CHPH06

University of Hyderabad
School of chemistry
Hyderabad-500046

CERTIFICATE

Theis is to certify that the thesis entitled "**Inhibitor design for drug targets in Zika virus and SARS CoV-2 using computational studies**" submitted by **Mr. Durgam Laxman** holding registration number **16CHPH06** in partial fulfilment of the requirements for award of Doctor of Philosophy in the School of Chemistry is a bonafide work carried out by him under my supervision and Guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other University or Institutions for award of any degree or diploma.

Parts of thesis have been:

A. Published in the following publications

1. Durgam, L., & Guruprasad, L. Molecular mechanism of ATP and RNA binding to Zika virus NS3 helicase and identification of repurposed drugs using molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 2021, 1-18.
2. Durgam, L., & Guruprasad, L. Computational studies on the design of NCI natural products as inhibitors to SARS-CoV-2 main protease. *J. Biomol. Struct. Dyn.* 2022, 1-11.

B. Presented and participated in the following conferences

1. ChemFest-2017, 2018, 2019 and 2021

Further the student has passed the fallowing courses towards fulfilment of course work requirement for Ph.D.

Course code	Name	Credits	Pass/Fail
CY-801	Research Proposal	3	pass
CY-802	Chemistry pedagogy	3	pass
CY-805	Instrumental methods-A	3	pass
CY-551	Biological chemistry	3	pass

30/12/2022

Dean
school of chemistry

Dean
SCHOOL OF CHEMISTRY
University of Hyderabad
Hyderabad-500 046

Prof. Lalitha Guruprasad
(Thesis supervisor)

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.

ACKNOWLEDGEMENTS

It gives me pleasure to acknowledge those who supported me during my Ph.D journey, be it my friends, family and faculties. I feel that this dissertation would not have been possible without their constant encouragement and belief. Therefore, I would like to thank all of them as they are equally responsible for completion of this work either directly or indirectly.

First of all, I want to thank my almighty god Jesus for being with me in the entire journey. I also want to thank the Department of Chemistry and management of University of Hyderabad for allowing me to do a Ph.D. course here.

Next, I want to thank my Guru, my ideal supervisor Prof. Lalitha Guruprasad for her perspective, instructions, and great encouragement without which, this dissertation would not have reached conclusion. I am obliged to her for reposing faith in me and making me realize that 'I am capable'. Her patience to listen and comprehend my ideas and thoughts was always admiring. I can say undoubtedly that I owe outstanding debt to her for the amount of guidance and extreme help she provided both through one-to-one or online discussions, throughout the study. She had been very kind to me for any administrative hiccups whenever I approached her. In fact, it had been a very educative journey to work with Prof. Lalitha Guruprasad, who is pleasant and accessible, her continuous support and patient explanations will always remain indispensable for me.

I owe my sincere thanks to the University of Hyderabad, from where I gained not only knowledge but also became self-reliant. I also feel very lucky having been selected at University of Hyderabad which is a reputed institution in academics as well as research for doing a PhD program in Chemistry. Therefore, I want to express my gratitude to the Vice-Chancellor of the University, the doctoral committee, Prof.

Susatha Mahapatra, and Prof. Muralitharan for their support and encouragement. I would also like to thank the Dean of the School of Chemistry and the staff for whatever help they provided whenever I approached them for administrative purposes. I hope and wish that the University will continue to achieve excellence and growth.

I am grateful to my friends namely Balaji Manasa Rao (Balu), Bairu Nagaraju, Naidu, Jaya Prakash, Mallesh, Tharun, Shankar Mallepally and my labmates Indra Neela, Shalini Saxena, Divya, Mageed, Naresh, Jishu, Priyanka, Adreja for providing motivation and inspiring me to do better. I thank them for being like my family and showing belief on me. I cannot forget the discussions and conversations spent with them in this journey which provided me strength to keep moving forward even in difficult times. I also thank my friends at our Laboratory of Chemical Informatics and Bioinformatics, School of Chemistry, University of Hyderabad. The beautiful days I spent with you is a bliss of solitude which I will always cherish.

I am also grateful to my friends from my hometown including my neighbors, my childhood friends of the ZPSS School Kuppenakuntla and my classmates at kavitha degree college.

I owe a lot to my family for their sincere love, support, and encouragement. I thank all my family members who stood beside me my father Krishnaiah, aunt Savitri, Venkamma, my brothers Venkateshwara Rao, Nageswara Rao, Ramu, my sister Ester rani and brother-in-law Vijay Kumar and my sister-in-law's Vijayamma, Krishnaveni and Slagini.

Lastly, I am eternally grateful to all those I met during this journey even if I missed mentioning them here their presence in my life and work holds great importance. Any errors or mistakes in this dissertation are mine alone.

ABBREVIATIONS

Ala	Alanine
Asp	Aspartic
Arg	Arginine
Asn	Asparagine
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic
Gly	Glycine
His	Histidine
Leu	Leucine
Lys	Lysine
Ile	Isoleucine
Pro	Proline
Phe	Phenylalanine
Tyr	Tyrosine
Trp	Tryptophan

Thr	Threonine
Ser	Serine
BlAST	Basic Local Alignment Search Tool
PDB	Protein Data Bank
sdf	structure-data file
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
vdW	van der Waals
NMA	Normal mode analysis
CADD	Computer-aided drug design
SBDD	Structure-based drug design
LBDD	Ligand-based drug design
DS	Discovery studio
3-D	Three-dimension
PLP1/PLP2	Piecewise Linear Potential 1 and 2
PMF	Potential of mean force
MD	Molecular Dynamic
Amber	Assisted model building with energy refinement
GROMACS	GROningen MAchine for Chemical Simulations
CHARMM	Chemistry at Harvard Macromolecular Mechanics

ACPYPE	Antechamber PYthon Parser interface
VMD	Visual molecular dynamics
ADME	Absorption, Distribution, Metabolism, Excretion
DFT	Density functional theory
HOMO	Highest occupied molecular orbital
LUMO	Lowest unoccupied molecular orbital
PCM	Polarizable continuum model
ESP	Electrostatic potential
SAM	S-adenosyl-L-methionine
SAH	S-adenosyl-L-homocysteine

CONTENTS

CHAPTER-1

Introduction

1.1 Introduction

A virus is a pathogenic microorganism composed of a protein-coated nucleic acid segment (either DNA or RNA) (Wu KJ, 2020). A virus cannot multiply by itself; it has to infect a living host in order to utilize the host cell's replication machinery components to make copies of itself. A host cell that has been infected is frequently compelled to quickly make several thousands of copies of the infected virus. Viruses can exist as independent particles or virions, when they are not inside an infected cell or in the process of infecting a cell. These virions are made up of i) the genetic material, which are long molecules of DNA or RNA, that encode the structure of the proteins the virus uses to replicate; (ii) a protein coat, the capsid, which surrounds and protects the genetic material; and (iii) in some cases, an external envelope of lipids.

Viruses can spread in a wide variety of ways. One technique of transmission involves the employment of disease-carrying vehicles known as vectors. Insects that feed on blood could also carry viruses that impact mammals. For instance, aphids, which feed on living plants, are often used to spread viruses from one plant to another. The two most common ways for viruses to spread in the air are through coughing and sneezing, including influenza virus (McCullers, 2006), SARS-CoV-2 (Wang et al., 2020), chickenpox (Tugwell et al., 2004), smallpox (Moore et al., 2006), and measles (Tatsuo et al., 2000). Norovirus and rotavirus are known to be the main causes of viral gastroenteritis, and they spread from person to person through hand-to-mouth contact, contaminated food, or water. Less than 100 infectious norovirus particles are needed to cause an infection in humans (Robilotti et al., 2015). Several viruses, including human immunodeficiency virus (HIV) and Zika virus, can be spread through body fluids and by contact with blood that has

been contaminated. The term "host range" describes the range of host cells that a virus can infect. A virus may have a limited range of species it may infect, or it may have a wide range and be able to infect many different hosts.

1.1.1 Structure of virus

Viruses exhibit a vast variety of morphologies, size and shapes (Louten, 2016). The majority of virus species have virions that are one hundredth of the size of most bacteria and are therefore too small to be seen with an optical microscope. Viruses are typically much smaller than bacteria, therefore one can assume that it is possible to fit more than a thousand bacteriophage viruses inside the cell of an *Escherichia coli* bacterium. Many known viruses are spherical, with a diameter ranging from 20 to 300 nanometers (nm). Although the diameter of some filoviruses, which are filaments, is only about 80 nm, their total length can reach 1400 nm (Britannica, 2017). Scanning and transmission electron microscopes were used to view viruses because a majority of them cannot be viewed with an optical microscope (Goldsmith & Miller, 2009). Atomic force microscopy allows for mechanical (physical) examination of the capsid and the complete virus structure (Kuznetsov et al., 2001). Electron-dense "stains" were employed to make viruses stand out better against the background. These are solutions of heavy metal salts, such as tungsten that scatter electrons from stained areas. Positive staining, the coating of virions with stain, obscures the fine details. Negative staining solves this issue by simply staining the backdrop.

A whole virus particle or virion, is made up of protein-coated nucleic acid that is encased in a protective protein shell called the “capsid”. The protective shell is created from protein building blocks known as capsomeres (Sevvana et al., 2021). Viruses have an outer lipid "envelope" that is assembled from the lipid membrane of the host cell. The morphological differentiation of viruses is based on the form of the capsid that is composed of proteins encoded

by the viral genome (Caspar & Klug., 1962; Crick & Watson, 1956). In general, the virus genome must be present for the virally-coded protein components to self-assemble into a capsid. Complex viruses encode proteins that help build their capsids. Nucleoproteins and nucleocapsids are terms used to describe proteins that are linked to nucleic acids. There are generally five primary categories of morphological viruses: icosahedral, prolate, helical, enveloped, and complex. Most often, viruses have RNA genomes, but some of the viruses also have single-stranded DNA genomes. Depending on whether the single-strand is complementary to the viral messenger RNA (mRNA), they are either positive-sense (referred to as "plus-strand") or negative-sense (referred to as "minus-strand"). Since the positive-sense viral RNA has the same sense as the viral mRNA, the mRNA genome can be translated to synthesize the protein products right away by the host cell. The viral protein RNA-dependent RNA polymerase (RdRp) must convert negative-sense viral RNA to positive-sense RNA before the translation occurs because it is complementary to mRNA.

Zika virus is one of the mosquito-borne families of Flaviviridae, genus flavivirus. In the Zika forest near Lake Victoria in Uganda, febrile sentinel rhesus monkeys were first found to have the Zika virus in 1947 (Dick et al., 1952). It has been known to occur since 1950 in a small equatorial belt extending from Africa to Asia. The Zika virus epidemic of 2015–2016 was caused by the virus spread eastwards from 2007–2016 across the Pacific Ocean to the Americas. The virus is transmitted by aedes mosquitoes, primarily daytime-active female *Aedes aegypti* which is commonly found throughout the tropical and subtropical Americas and it can also be spread by the *Aedes albopictus* (Asian tiger) mosquito which is distributed upto the Great Lakes region in North America (Kraemer et al., 2015). People infected with Zika virus can transmit the virus to their sexual partners (Centers for Disease Control and Prevention, 2017). The infection caused by Zika virus is Zika disease or Zika fever that is accompanied by fever, headache, arthralgia,

myalgia, maculopapular rashes (Ioos et al., 2014). Zika infection of male adult mice can cause testicular and epididymal damage, resulting in cell death and destruction of the somniferous tubules (Govero et al., 2016). When a pregnant woman is infected with Zika virus, it can be transmitted from mother to the fetus, that can manifest in microcephaly and related acute brain irregularities in the infant (Rasmussen et al., 2016). In addition to several Pacific American islands, the epidemic also spread to other regions of South and North America. In 29 countries, there have been nearly 3000 cases of microcephaly-related birth defects, Brazil has been hardest hit, with 2,366 babies being born to Zika-infected women and their families (Lancet, 2017). As of November 2018, 157 cases have been identified in India out of these 63 were pregnant women https://www.who.int/emergencies/diseases/zika/india-november-2018/en/. The Zika virus outbreak was declared as a Public Health Emergency of International Concern by WHO during February 2016, as there was growing evidence that infection of Zika virus can lead to birth defects as well as neurological problems (Sikka et al., 2016).

Zika virus belongs to the family of flaviviruses that includes several important viral pathogens of human, such as dengue, west Nile, yellow fever, Murray valley encephalitis, Japanese encephalitis, tick-borne encephalitis, Kunjin and Kokobera viruses. Like other flaviviruses, Zika virus also has icosahedral structure and it comprises ~11-12 kb non-segmented, single-stranded RNA genome that is capped at the 5' end. The RNA genome of Zika virus is translated in the cytoplasm of the infected cells into a polyprotein, which is further proteolyzed by either host cell or viral protease into three structural proteins, called envelope protein, precursor membrane protein and capsid protein, and seven non-structural (NS) proteins; NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5, these NS proteins of Zika virus perform essential roles in genome replication. A multi-protein replication complex comprising of both NS proteins and host cofactors are assembled on

the endoplasmic reticulum derived membranes (Welsch et al, 2009), that forms a location for the RNA replication to occur. Among all the NS proteins, NS5 is a largest enzyme and the most conserved protein of the replication complex. The Zika virus NS5 protein exhibits two enzymatic activities, RdRp and methyltransferase (MTase) which are mostly required for viral replication. The MTase protein is located at the N-terminus and the RdRp is located at the C-terminus of the NS5 protein.

1.1.2 The life cycle of positive-sense single-stranded RNA viruses

In the flaviviruses, the positive-sense single-stranded RNA can act as mRNA for translation towards protein synthesis. The replication of viruses occurs in two ways, the ribosome in the host cell can directly translate into the viral proteins. The other replication process of the positive-sense RNA genome proceeds through double-stranded RNA intermediates. The positive-strand RNA encodes an RdRp, which helps to synthesize a complementary negative-strand RNA template to create a new positive-sense RNA genome during replication. There is a double-stranded RNA formed in the intermediates. RdRp differs from RNA polymerase as it works to catalyze the synthesis of an RNA strand complementary to a given RNA template, the RNA replication process is a four-step mechanism. The first step is to bind nucleoside triphosphate (NTP). Initially, the protein RdRp has a free active site in which an NTP binds that is complementary to the corresponding nucleotide on the template strand. Correct NTP binding induces RdRp to change its conformation. The conformational changes caused by the correct NTP binding result in active site access restriction and the formation of a catalytically competent state in the second step. In the third step, a phosphodiester bond formation takes place; with two Mg^{2+} ions located in the catalytically active state that can rearrange themselves in such a way that around the newly synthesized RNA chain the substrate NTP can undergo a phosphatidyl transfer and form a

phosphodiester bond with the newly synthesized chain. Using these Mg^{2+} ions, the active site is no longer catalytically stable, and the RdRp complex changes to an open conformation. The fourth step is translocation. Once the active site is open, the RNA template strand can move by one position through the RdRp protein complex and continue the chain elongation by binding a new NTP, unless otherwise specified by the template.

Due to the complementary nature of negative-sense viral RNA to mRNA, a RdRp is required to convert it to positive-sense RNA prior to translation. Positive-sense virus RNA can directly infect cells even though it may not be as contagious as the entire virus particle. The purified RNA from a negative-sense virus, in contrast, is not contagious on its own. It has to be converted into positive-sense RNA, and each virion can be converted into a number of positive-sense RNAs. Ambisense RNA viruses resemble negative-sense RNA viruses, except that they translate genes from both their negative and positive strands. The newly produced viral genome is ready to infect a new cell. Zika virus replication mechanism occurs in this manner and is shown in **Figure 1.1**. In the Zika virus replication, the NS proteins, NS3 helicase, NS5 RdRp and MTase (shown in **Figure 1.2)** play a crucial role in viral genome replication.

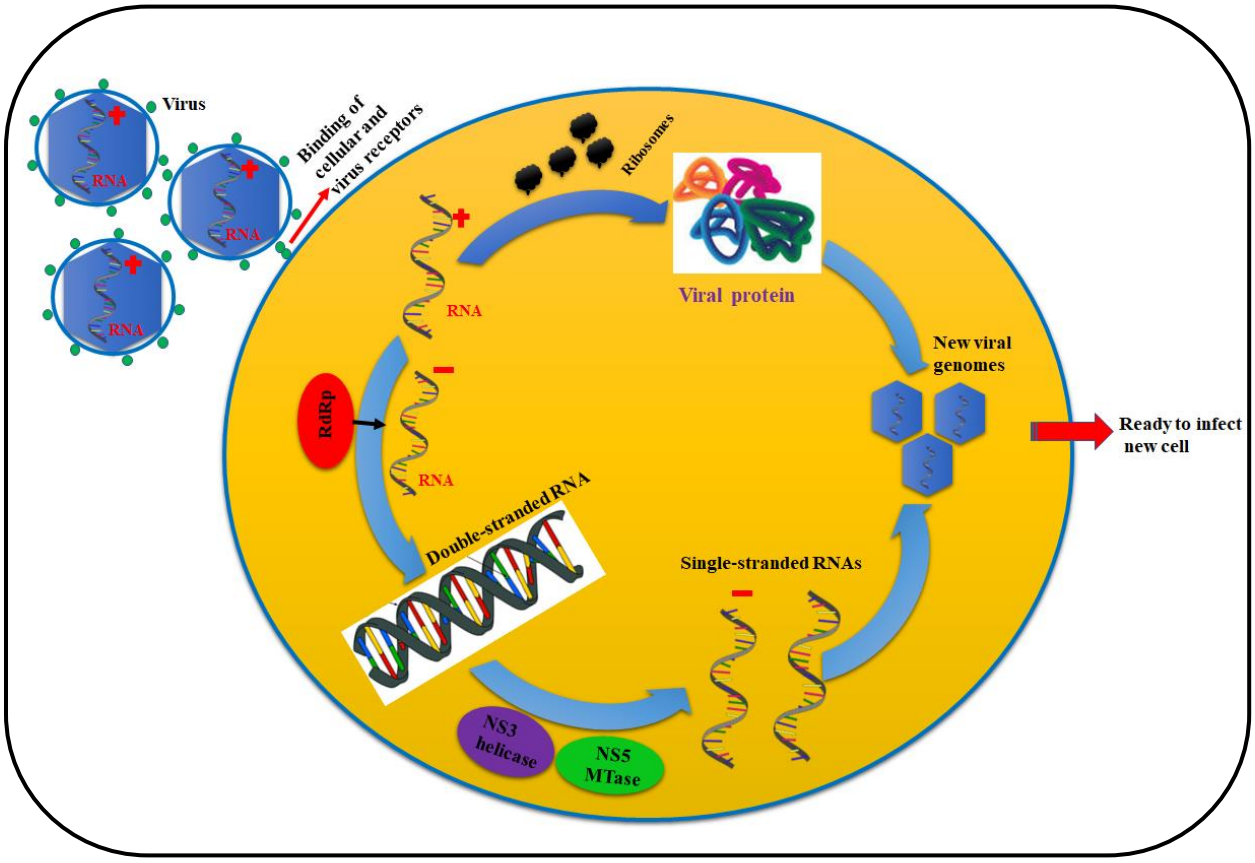


Figure 1.1: Life cycle of single-stranded RNA flaviviruses.

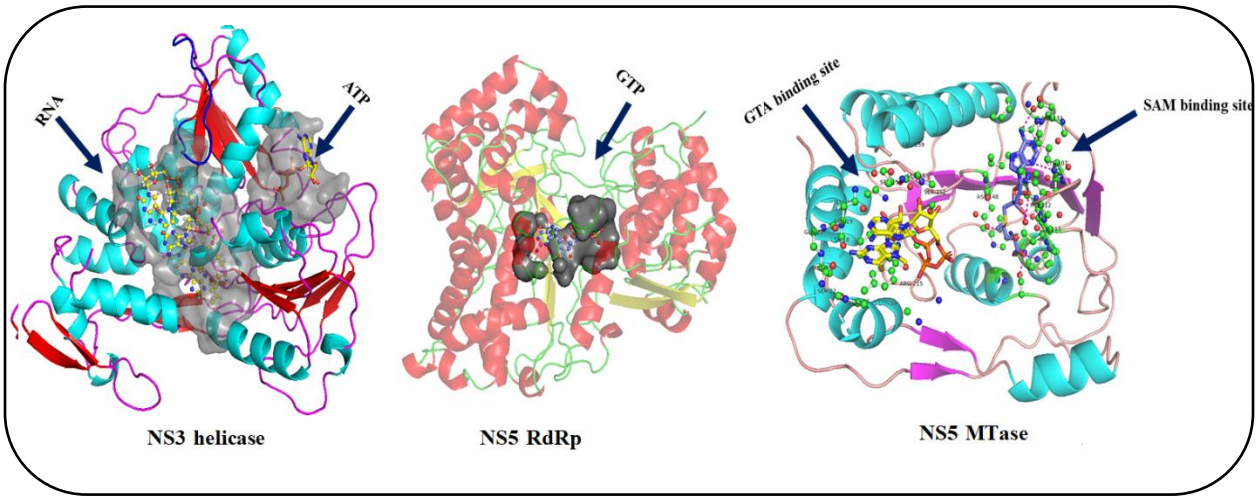


Figure 1.2: Structures of Zika virus NS3 helicase, NS5 RdRp and NS5 MTase.

1.1.3 Zika virus NS3 helicase

The N-terminus of the Zika NS3 helicase contains a protease domain and the C-terminus contains a helicase domain. The tertiary structure of the Zika virus NS3 helicase is composed of three domains, each of which contains 130-160 amino acid residues. Despite their low sequence homology, domains I (residues 192-332) and II (333-481) share a similar structural fold. Domain III (482-617) is made up of a four-helical bundle that is extended by two anti-parallel strands that are partially exposed to the solvent. The protease domain is represented by the amino acid residues 1-174. Two α-helices from domain I interact with a parallel α-helical bundle from domain III; domain II connects to domain III via a long β-hairpin, stabilizing interdomain interactions (Tian et al., 2016), and all three domains have clearly defined binding clefts. The cofactor ATP is present between domains I and II in Zika virus NS3 helicase; it is stabilized by the Mn^{2+} coordinated in octahedral geometry by interactions with side-chain of residues Thr201 and Glu286, two oxygen atoms of β- and γ-phosphate groups of ATP and two water molecules. The binding site of ATP is formed by the residues Gly197, Ala198, Gly199, Lys200, Thr201, Arg202 and Arg203 (P-loop), and Glu286, Ala317, Asn330, Gly415, Asn417, Gln455, Gly458, Arg459, Arg462 and Asn463. A single-stranded RNA binds the Zika virus NS3 helicase within the region formed by the amino acid residues; Pro224, Thr225, Arg226, Val227, Met244, Thr245, Thr246, Cys262, Ala264, Thr265, Phe289, Asp291, Pro292 (domain I), Pro364, Ser365, Val366, Arg367, Ser387, Arg388, Thr409, Asp410, Ile411, Leu430, Lys431, Pro432, Leu442 (domain II), His486, Lys537, Asp540, Arg598 and Ser601 (domain III). The single-stranded RNA makes hydrogen bonding interactions with Arg226, Thr245, Thr265, Asp291, Val366, Arg388, Thr409,

Asp410 and Lys431. RNA helicases play an essential role in the RNA metabolism and viral infection facilitation for RNA viruses.

1.1.4 NS5 RNA dependent RNA polymerase and NS5 methyltransferase

The Zika virus NS5 RdRp produces a negative-sense RNA by using the positive-sense viral RNA as a template for the formation of the intermediate stage double-stranded RNA. The Zika virus NS3 helicase unwound the double-stranded RNA to separate it into negative-sense and positive-sense single-stranded RNA molecules. The negative-sense RNA serves as a new template for the production of positive-sense genomic RNA (Xu et al., 2019). The replication of RNA occurs in a multi-protein replication complex comprising of NS proteins and cofactors, which assemble on endoplasmic reticulum derived membranes (Welsch et al, 2009). The NS5 protein comprises a MTase domain at the N-terminus and the C-terminal domain contains a RdRp (Zhao et al., 2017), The three-dimensional (3-D) structure of NS5 MTase is reported in Protein Data Bank (PDB), it comprises three domains referred to as finger (amino acid residues 321-488), palm (481-541 and 609-714) and thumb (715-903). The RdRp synthesizes the genome of RNA in the absence of a primer strand, in a *de novo* mechanism wherein single-nucleotide triphosphate provides a primer for nucleotide polymerization. The flavivirus RdRp contains a functional nuclear localization sequence, an important key region for viral and host proteins interactions. NS5 interacts with the NS3 protease-helicase and several host proteins (Yap et al., 2007; Tay et al., 2015). The Zika virus NS5 protein acts as an antagonist due to its enzymatic functions. That is because the NS5 RdRp plays a crucial role in viral genome replication. MTase enzyme is responsible for adding the methyl group to the 5' cap RNA structure, using cofactor S-adenosyl-L-methionine (SAM) as a methyl donor to form S-adenosyl-L-homocysteine (SAH) as a by-

product (Zhao et al., 2015) to facilitate the translation of polyprotein. These two enzymes RdRp and MTase are important for viral replication.

1.2 Methods

1.2.1 Databases

The purpose of building a variety of databases is to separate the knowledge information, create organized data, and make value addition through its annotation. Such databases have been made accessible for the scientific purposes with payment of a subscription fee or for free of charge.

1.2.1.1 Protein sequence database

The biological structure and function of a protein is determined by its amino acid sequence. The nucleotide sequences of the gene coding regions are obtained from the complete genome nucleotide sequencing projects at the organism level, these nucleotide regions are translated to protein amino acid sequences and is referred to as the primary protein structure (Xu & Xu, 2004). The protein primary structure information is stored in protein sequence databases. The National Center for Biotechnology Information (NCBI) available at http://www.ncbi.nlm.nih.gov, hosts a variety of information including both nucleotide and protein sequences. The NCBI Reference Sequence (NCBI RefSeq) database is a collection of nonredundant sequences of genomes, transcripts and proteins that are curated. A stable reference for genome annotation, gene identification and characterization, mutation and polymorphism analysis, and genetic testing are included with these entries (O'Leary et al., 2016). There is an issue with explosion of protein sequence information due to the sequencing of complete genomes. This problem of redundancy or duplication in protein sequence information has been resolved by creating the prokaryotic RefSeq protein dataset. This database has grown in size significantly due to the bacterial genome sequence submissions from individual isolates and their closely related bacterial strains, therefore, the other

kind of nonredundant RefSeq protein database has already been produced. The Universal Protein Resource (UniProt) is a comprehensive database of receptor sequences, with annotation data available online at http://www.uniprot.org/. The protein sequence databases available at NCBI and UniProt are most frequently used by the researchers. The nucleotide sequence database (http://www.ebi.ac.uk/embl/index.html) of the European Molecular Biology Laboratory (EMBL) is a complete set of primary nucleotide sequences maintained at the European Bioinformatics Institute (EBI). The other useful databases are GenBank (Benson et ai., 2003) PROSITE (Falquet et al., 2002), PRINTS (Attwood TK et al., 2003), Pfam (Bateman et al., 2003), ProDom (Corpet et al., 2000) SMART (Letunic et al., 2002), TIGRFAMS (Haft et al., 2003), PIR SuperFamilies (Huang et al., 2003) and SUPERFAMILY (Gough et al., 2001).

1.2.1.2 Protein structure database

The 3-D proteins structures along with nucleotides, heteromeric complexes, and complexes with cofactors/inhibitors/substrates are solved using X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy methods. The high-resolution 3-D structures of macromolecules is determined using these methods. A publicly accessible database of biomolecular structures is maintained by Worldwide Protein Data Bank (wwPDB), this is a freely accessible PDB archive to the research community (Berman et al., 2003). The Research Collaboratory for Structural Bioinformatics Protein Database (RCSB PDB) is one of four organisation members (PDBe, PDBj, RCSB, and BMRB) for the retrieval of protein 3-D structures (Velankar et al., 2010; Kinjo et al., 2012; Markley et al., 2008).

1.2.1.3 Small molecule chemical libraries

The 3-D structures of organic molecules and their physical properties are stored in small molecule databases. The biological activity of small molecules, drug targets and literature citations

are also available in some of the databases (Bento et al., 2014). While some of these chemical library databases are freely accessible and can be downloaded in a variety of file formats, other databases are proprietary and must be procured from the vendors. The accessibility of virtual libraries of ligand molecules supports the *in silico* screening, discovery and design of small molecule inhibitors for a selected receptor target. The virtual libraries of chemical compounds could hold billions of molecules, allowing a researcher to find hit molecules via virtual database screening using computational methods. Some of the most commonly used databases in computer-aided drug design (CADD) are; BindingDB contains data for 1,100,000 compounds and 8,900 targets (Gilson et al., 2016), ChEMBL contains 2,331,700 molecules for 15,072 targets (Gaulton et al., 2012), ChemSpider contains 115,000,000 chemical structures (Williams et al., 2010), Cambridge Structural database contains 1,100,000 chemical structures from experimental sources (Groom et al., 2016), DrugBank comprises over 500,000 molecules (Wishart et al., 2018), MCULE full database comprise 40,075,205 molecules (Kiss et al., 2012), PubChem database comprises 112,000,000 compounds (Kim et al., 2016), SciFinder comprises more than 182,000,000 compounds (Wagner, 2006), ZINC database comprises over 230 million compounds (Irwin & Shoichet, 2005), MolPort (20,000,000, https://www.molport.com), Asinex database comprises 575,302 compounds (http://www.asinex.com/), ChemBridge comprises over 1.3 million diverse and target-focused small molecule screening compounds (https://www.chembridge.com/), Chemical Diversity, ChemDiv comprises more than 180,000 active compounds (https://www.chemdiv.com/catalog/), AsisChem (2,109,738, http://www.asischem.com/), Enamine comprises 2,790,127 molecules with a range of molecular weights (Shivanyuk et al., 2007), SPECS (350,000, http://www.specs.net), and the National Cancer Institute (NCI) natural compounds set database (II, III, IV and V)

(https://wiki.nci.nih.gov/display/ncidtpdata/compound+sets), FDA approved drugs (https://www.bindingdb.org/bind/ByFDAdrugs.jsp).

1.2.2 Basic local alignment search tool

A protein sequence of interest (query) can be compared or searched with all of the entries in a protein sequence database using the heuristic technique known as the Basic Local Alignment Search Tool Protein (BLASTp). Above a predetermined threshold, it determines the protein sequences in the database that most closely resemble the query protein. The alignments are extended from these initial "hot spots" when short matches are made between two sequences. In addition to executing pairwise sequence alignments, it also gives statistical data regarding an alignment, such as the "expect" value, database length of the protein sequence, percentage identity, query coverage and matching score (Altschul et al., 2005).

1.2.3 Multiple sequence alignment

Sequence alignment is the arrangement of three or more protein amino acid sequences to identify regions of similarity. The similarity regions between the sequences could be the result of functional, structural and evolutionary relationships (Edgar & Batzoglou, 2006). The alignments produced would be examined to find the regions of amino acids insertions, mutations, or deletions among the sequences of interest (Needleman & Wunsch, 1970; Smith & Waterman, 1981; Lipman et al., 1989). The output format is used to create phylogenetic trees, which can be used to calculate evolutionary distances between the protein sequences and search for conserved domains that are essential for function (Sievers & Higgins, 2014). Multiple sequence alignments can be generated at both the global and local levels. The global multiple sequence alignment algorithms generate an alignment that cover the entire sequences and fills in any gaps. While the local multiple sequence

alignment algorithm aligns only the most similar regions of amino acid residues. The ultimate goal by both techniques would be to align longer sequence regions of amino acids with higher matching among the target proteins under consideration. Clustal Omega (Sievers & Higgins, 2014) and Tcoffee (Notredame et al., 2000) are some of the commonly used software tools to generate sequence alignment. Clustal omega produces multiple sequence alignments of nucleotide or protein sequences by selecting seeded guide trees and by applying hidden Markov model profile-profile techniques (Söding, 2005). The dynamic programming technique is used to identify the globally optimal alignment solution in a direct method for producing a multiple sequence alignment. For proteins, this approach typically uses two sets of parameters: a gap penalty and a substitution matrix that assigns scores or probabilities to the alignment of each potential pair of amino acids depending on the chemical characteristics of the amino acids and the likelihood that a mutation will occur over a time period. Similar gap penalties are applied to nucleotide sequences, but typically the substitution matrix is much more straightforward, taking only identical matches and mismatches into account. While the substitution matrix results for a global alignment could be all positive or a mix of positive and negative, they should be both positive and negative for a local alignment (Larkin et al., 2007; Thompson et al., 1994).

1.2.4 Structural comparison of proteins

Using servers such as DALI, the 3-D structure of a protein receptor could be evaluated by comparing with the already known proteins structures (Holm & Laakso, 2016). These searches on model protein structures support in the identification of protein structures that share comparable folds, binding sites for ligand, cofactor, and inhibitors binding. This structure comparison provides clues for the identification of proteins function and further in the drug design studies. Based on the amino acid side-chains that contribute to the 3-D active site space required for protein function,

the binding pocket of receptor identified to be similar via structure comparison searches (for example, based on a high Z score in DALI method) are most likely to have similar functions.

1.2.5 Bioinformatics

Bioinformatics is an emerging multidisciplinary area which has evolved owing to the high preference for computer assisted data analysis of the large biological data collected over the last 30 years (Luscombe et al., 2001). Bioinformatics incorporates the principles of physics and chemistry on biological systems by employing computational tools (Searls, 2010). Bioinformatics studies use biological data such as nucleotide and protein sequences to analyse and compare life forms and their evolutionary aspects. As a result, it is critical for data management in modern medical science and biology (Baxevanis et al., 2020). Bioinformatics is a resource for predicting and recognizing regulatory networks connecting genes, analysing genetic variations and expression, predicting gene coding regions, prediction of protein structure and function, modelling protein networks and their dynamics, simulating environment in conditions closer to living cells, and analyzing metabolic pathways to understand implications in disease.

1.2.6 Chemobioinformatics

Chemobioinformatics is an interdisciplinary field of chemistry, biology, mathematics, and physics that uses computer modelling and simulation to analyse the structure and characteristics of molecules and components, which include empirical and *ab initio* approaches (Martinez-Mayorga et al., 2020). Chemobioinformatics is a fast-emerging research area with methodologies for applications in CADD. The availability of high-speed computers with large data storage capability has accelerated the success in applying chemobioinformatics (Yu & MacKerell, 2017). To calculate the structure and properties of molecules, it employs theoretical chemistry

methodologies that are embedded into user friendly computer programmes. Computational chemistry methods and techniques typically vary from nearly approximate to precise (for large molecules) to greaterly accurate (for small molecules). The *ab initio* methods and techniques are based on quantum mechanics. Additional empirical characteristics are used in semi-empirical or empirical methods (MacKerell Jr et al., 1998). Computational tools are useful in molecular modelling ultimately leading up to its experimental synthesis in a laboratory, allowing unsuitable molecules to be ruled out. Furthermore, some of the molecular properties are more easily obtained computationally than experimentally.

1.2.7 Molecular animations and visualization

The molecular structures in chemical and biological systems can be visualized on computer graphics, on the screen in an interactive mode by the use of a variety of systems designed to display molecules in a virtual mode (Martinez et al., 2019). Designing such technologies for visualization of molecules necessitates knowledge in chemical as well as computer sciences. The software used for visualization of 3-D molecular structures in graphics are RasMol (Sayle & Milner-White, 1995), PyMOL (Schrodinger, 2010) (DeLano, 2002), UCSF Chimera (Pettersen et al., 2004), Jmol (Hanson, 2010), Deep View (Guex et al., 2009), Discovery Studio (DS) visualizer. The verification of 3-D model structures of proteins and ligand molecules, the manipulation and analysis of their molecular properties is carried out using these graphics visualizers. Visual molecular dynamics (VMD) is a freely downloadable 3-D graphics programme for displaying and animating large biomolecular systems. It is used to view 3-D molecular structures as well as plotting and analyzing the larger trajectory data files generated by traditional molecular dynamics (MD) simulations (Humphrey et al., 1996).

1.2.8 Artificial intelligence in drug discovery

Artificial intelligence is a method that combines computer simulations and intelligence gathering processing. Obtaining information, continuing to develop rules for using the information obtained, and reaching reasonable conclusion, and self-correction are all the steps in this process (McCarthy, 1987; Nilsson & Nilsson, 1998). To extract meaningful information from a large dataset, artificial intelligence employs complex algorithms and machine learning (Batool et al., 2019). For example, it aids in the identification of molecules which may bind to 'undruggable targets,' targets such as proteins with unidentified structures. A predictive set of molecules can be identified in a relatively shorter time scale by sequential simulations of intermolecular interactions in a variety of ligands with different regions in a protein (Hessler & Baringhaus, 2018). In less than a decade, artificial intelligence is being applied in novel drug discovery, this employs large data that is generated from gene expression data or libraries of small molecules and on an ensemble of protein structures. Over the last decade, artificial intelligence methods have employed deep learning and machine learning in new drug discovery and development. Automated software is available online that is offered via webservers to carry out studies in quantitative structure activity relationship (QSAR), pharmacophore generation, protein 3-D fold assessment, *in silico* screening of ligands, protein-protein interactions and protein-ligand, drug repurposing, drug design using *de novo* methods, evaluation of absorption, distribution, excretion, metabolism and toxicity (ADMET) properties of drugs (Zhong et al., 2018). Using artificial intelligence in the different stages of a drug development will decrease the project time and cost while driving the drug development in becoming more efficient.

1.2.9 Homology modeling of protein structures

A proteins structure is divided into four stages primary, secondary, tertiary, and quaternary. The protein 3-D structure is determined by the spatial distribution of all atoms on the main-chain and side-chains (Lüthy et al., 1992). PDB contains the protein 3-D structures defined through experimental studies (Berman et al., 2007). Knowledge of the protein 3-D structure provides useful information about the molecular mechanisms of their activities. Employing experimental techniques to determine structures of protein takes time and may not yield an advantage for proteins that tend to denature by aggregation under *in vitro* conditions and therefore precipitate out of solution. The unavailability of the 3-D structure of a protein has stymied efforts to determine the location of a ligand to bind the binding pockets in a protein. Under these conditions, one of the most reliable methods for obtaining structural information for a protein of interest is to build a model protein structure based on the protein 3-D structure that shares similar amino acid sequence (Cavasotto & Phatak, 2009). Some 3-D structures can be modelled to use homology modelling, *ab initio,* hybrid, and fold prediction methods (Hardin et al., 2002). Comparative modelling of protein structure, also known as homology modelling (Kopp & Schwede, 2004) is based on the observation that when primary sequences of two protein have a greater degree of homology, then their respective structures would have a high degree of similarity as well. The protein sequence of study with an unknown 3-D structure is referred to as the query sequence. The protein sequence similarity search method, i.e., BlastP searches against PDB (Altschul et al., 2005) can be employed to obtain the reference or a homologous structure on which the homology model can be constructed. The known 3-D structures that have the highest BlastP score, with fewer insertions and deletions, structures solved at higher resolution, and with none or fewer missing residues are obtained and are denoted as template structures. The homology modelling protocol necessitates a

pair-wise or multiple sequence alignment of the query and template protein sequences based on the availability of single or multiple template 3-D structures, respectively (Holm & Laakso, 2016). Comparative modelling constructs the 3-D structure of a query protein based on template protein sequence alignment. MODELLER (Sali & Blundell, 1993) is a popular method in comparative protein structure modelling methods that is based on the conditions of the satisfaction of spatial restraints. For building protein models, this software is available in a downloadable form and installation on local computers. The homology modelling process consists of four steps: identification of protein target of known 3-D structure, alignment of protein sequences, construction of model structure and refinement of structure. PRIMO (Hatherley et al., 2016), I-TASSER (Zhang, 2008), Phyre2 (Kelley et al., 2015) and SWISS-MODEL (Schwede et al., 2003) are some of the online servers available for homology based modelling of protein structures.

1.2.9.1 AlphaFold

Predicting the 3-D structure that a protein will adopt is purely based on its amino acid sequence, and this has been an important research problem for more than 50 years (Anfinsen, 1973). Despite recent advances (Senior et al., 2020; Wang et al., 2017; Zheng et al., 2019; Abriata et al., 2019) current methods fall far short of atomic precision, particularly when no identical structure is available. AlphaFold is the first computational technique that can predict protein structures with atomic precision even when no comparable structure exists. In the 14th Critical Assessment of Protein Structure Prediction (CASP14) (Kryshtafovych et al., 2021), an entirely newly designed version of a neural network-based model, AlphaFold, provided structures with high accuracy challenging the experimental structures in the majority of cases and outperformed other methods significantly. AlphaFold is an innovative machine learning method that incorporates biological and physical knowledge about structure of protein into the design of the deep neural

network algorithm by leveraging multi-sequence alignments. The CASP assessment employs recently remedied structures that have not been placed in the PDB or made publicly available, resulting in a blind test. CASP has long been the gold standard for determining the accuracy of structure prediction methods (Moult et al., 1995; Kryshtafovych et al., 2019). The AlphaFold network predicts the 3-D coordinates of all heavy atoms for a given protein using the primary amino acid sequence and aligned sequences of homologues as inputs. There are two stages to the network. The network's trunk first processes the inputs through repeated layers of a novel neural network block called Evoformer to generate a N_{seq} x N_{res} array (N_{seq}, number of sequences, N_{res}, number of residues) representing a processed multiple sequence alignment and a N_{res} x N_{res} array representing residue pairs.

1.2.9.2 Model validation methods

To ensure proper stereochemistry and protein folding, the protein 3-D structure identified using modelling methods should be validated. This structure evaluation process is critical in computational methods because the model structure will be employed for structural comparison of proteins, docking studies for molecular design, and MD simulations to analyse their transitions in conformations. The Ramachandran plot (Ramachandran et al., 1963) is used to divide amino acid residues into three different regions to verify the stereochemical specifications of the structure of a protein based on torsion angles: these are the preferred regions; allowed regions, disallowed regions, and the outlier regions. Web servers such as SAVES (https://saves.mbi.ucla.edu) and PROCHECK (Laskowski et al., 1993) provide additional details like main-chains, side-chains, bond lengths, bond angles, bonding and nonbonding interactions, ring planarity, and disulfide bonds. A better model structure has more amino acid residues in the allowed region and fewer amino acids in the disallowed regions of the Ramachandran plot. The VERIFY 3D server is

employed to compare the position and surroundings of a generated model 3-D structure with defined structures and to evaluate its secondary structure, area of buried residues, and side-chains covered by polar atoms (Lüthy et al., 1992). The model with the highest score is believed to be the best. The nonbonded interactions in protein structures are analysed by ERRAT, and a top scoring model validates the structures backbone conformations (Colovos & Yeates, 1993).

1.2.10 Computer aided drug design

CADD methodology incorporates chembioinformatics and computational biology methods (Zheng et al., 2013). These methods further help to generate structure-activity relationship (SAR) at the atomic level to support in the progression of drug development, limiting cost and time investment (Van De Waterbeemd, 2003). CADD primary role in drug discovery which is to divide large databases of compounds into smaller groups and correlate small compounds depends on their activity, allowing for the optimization and discovery of hit compounds by improving biological activity; such as ADMET and binding efficiency (Hassan et al., 2016). CADD is categorized into two approaches: ligand-based drug design and structure-based drug design, both of which convert features into models based on pharmacophore studies and QSAR (Mercader et al., 2016). CADD employs previous knowledge available on the 3-D structure of target protein to analyze the extent of intermolecular interactions between the receptor and ligand in the study. To differentiate between the known active and inactive molecules, ligand-based CADD employs the chemical similarity criteria and predict QSAR models that are generated from the molecules. QSAR modelling teaches about the impact of structure factors on bioactive compounds and how to construct ligands with enhanced and improved biological activities (Yu and MacKerell, 2017). The pharmacophore and QSAR models are employed in virtual screening to identify new hit molecules or ligands from commercial and open source chemical libraries in order to screen

molecules that make a large number of intermolecular interactions and high binding affinities. As a result, the CADD approach is important in the process of design and optimization of likely hit molecules to bind the receptor. It also has a wide range of applications in the different stages of drug discovery and development, such as drug target identification, validation, molecular design, improvisation of hit to lead molecule and interactions of hit or lead or drug molecules with protein targets of interest.

1.2.10.1 Structure-based drug design

Structure-based drug design methodology makes use of understanding of the 3-D structure target protein in complex with a hit compound and optimise the hit compound or a series of derivatives of a parent compound. It requires understanding of the protein-hit molecule intermolecular interactions in the complex formation. The experimental structures or homology models can be used to obtain structural information (Lounnas et al., 2013; Leach, 1994). This is a computational approach for locating possible hit compounds which aim to bind to a drug receptor implicated in a disease. This method involves swiftly searching through vast virtual libraries of chemical compounds, then docking the successful molecules into a receptor binding pocket of interest, which might be an active site or an allosteric binding pocket. The scoring function of a docking study is used to evaluate the potential of the hit molecules to bind with the receptor target in order to quantify the binding of these two molecules. The *in silico* screening of chemical databases by implementation of docking protocols or pharmacophore model based *in silico* screening are designed based on important residues that are also present in the template proteins and occupy similar regions is one of the techniques in structure-based drug design (Yang, 2010). The *de novo* molecule design is second category, which involves creating a molecule from fragments that bind to the active site and joining them together with a linker (Scott et al., 2012).

In the third category, a molecule can be optimised chemically so that the new analogues are more potent molecules with higher binding affinity to the receptor target (Pennington et al., 2020). The noteworthy gain of this exercise is that it enhances the speed of identification of hit compounds by substantially decreasing the number of hit compounds that need to be evaluated experimentally for their biological function, which tends to increase the success rate of the experiments from *in vitro* and *in vivo* studies.

1.2.10.2 Ligand-based drug design

In the absence of the 3-D structure of protein, ligand-based drug design is a useful methodology that relies on experimental data from ligands which are known to bind to the protein receptor target under consideration. The 3-D QSAR and pharmacophore modelling (Dixon et al., 2006; Lin, 2000) are important methodologies in ligand-based drug design. In pharmacophore and QSAR studies, knowledge on both active and inactive molecules evaluated *in vitro* or *in vivo* is represented as a source of data. This analysis results in the development of a model that can be employed in virtual screening, such as the SwisSimilarity server, to find new hit molecules (Zoete et al., 2016). These analyses can also provide extrapolative models that can be used to identify and optimise lead molecules. Analysing compounds that bind to the drug target of interest in a disease is a useful technique for advancing and improving pharmacologically active molecules. A QSAR is defined as the relationship that is created between both the hit compounds estimated biological features and their empirically determined bioactivities. Observations and findings from QSAR correlations are frequently utilized to identify the biological activity of newly designed and experimentally validated structural analogue molecules.

1.2.11 Pharmacophore modeling

A set of electrostatic and steric features that are necessary in order to confirm the interactions of a ligand with a receptor binding site of a specific biological target that is significant in disease so as to invoke a biological response is defined as a pharmacophore. The goal of these interactions is to elicit a biological response from the organism. The CADD technique that recognizes essential components necessary for function and recognition between macromolecule and ligand is the pharmacophore modelling (Wolber & Langer, 2005). A combination of inactive and active components, in addition to intermolecular interactions between ligand and receptor, make up a pharmacophore model. Hydrogen bond acceptors or donors, anions, hydrophobic centroids, aromatic rings, cations, metal contacts, and aromatic stacking or charge transfer interactions are examples of pharmacophore characteristics. A perfect pharmacophore model is constructed on a set of compounds 3-D characteristics and contains not more than seven desired features, with the majority of these features interacting with the receptor non-covalently. Pharmacophore models can be constructed manually or automatically, beginning with the structures of known active compounds (ligand-based approach) or the 3-D structure of the protein target (structure-based approach) (Leach et al., 2010). One of the numerous uses for pharmacophore models would be as a query to quickly screen vast chemical libraries in a virtual manner (Seidel et al., 2010). The primary objective is to find novel hit compounds that have a certain set of pharmacophore characteristics that are regarded to be essential for their bioactivity against a particular target of interest in disease. Screening for pharmacophores consistently yields hit molecules with a high structural variation. Furthermore, the ease with which pharmacophore features can be represented allows for rapid virtual screening of large small molecule chemical databases comprising billions of small organic compounds. Depending on the selection of the

necessary pharmacophore features, the application of precise matching requirements, and the volume of the small molecules database, a regular pharmacophore screening can typically yield tens to thousands or more hit molecules. Pharmit (Sunseri & Koes, 2016) is a dedicated web server for assigning preferred pharmacophore features and performing virtual screening; others are built into commercial software such as DS version 3.5 and Maestro, Schrodinger.

1.2.12 Molecular docking

CADD relies heavily on molecular docking, which is one of the important *in silico* structure-based rational drug development technique. The primary goal of receptor-ligand docking is to identify a ligand's predominant binding mode in the binding cavity of a known structure of protein (Pinzi & Rastelli, 2019). Docking would be the algorithmic plotting of bioactive molecules; the likely space occupied by all potential hit compounds that will be optimised in the later steps. The goal of docking is to attain the best conformation and spatial arrangement of a ligand within a protein's active site (Morris & Lim-Wilby, 2008). Docking protocols are designed in two components to achieve the highest success rate for the computational algorithm: docking alignment and scoring function. The molecular docking is a malleable method wherein the conformation of a ligand or protein can be altered during the docking (Leach, 1994). The docking techniques are defined into three categories based on the flexibility of the specific protein and hit compound. The primary objective of docking studies is to evaluate whether the protein or hit molecule is highly flexible (ligand). Flexible ligand docking, keeps the protein receptor conformation rigid and ligand is treated as flexible, while in rigid body docking, both the protein and the ligand are considered to be rigid. In the flexible receptor docking, both the receptor and ligand remain flexible. Flexible docking is one of most computationally intensive technique. Most docking processes regard the small ligand molecule as flexible while the receptor as rigid. This is

incorporated in molecular docking tools such as CDOCKER, AutoDock, FlexX and AutoDock Vina (Kramer et al., 1999; Morris et al., 2009; Wu et al., 2003; Trott & Olson, 2010). Some other docking techniques such as, GOLD, DOCK, LeDOCK, Glide (Verdonk et al., 2003; Lang et al., 2009; Friesner et al., 2004; Zhao & Caflisch, 2013) also recognize both the inhibitor and the receptor binding cavity as flexible orientations during docking so that the receptor and inhibitor molecule fit to one another in a complementary manner. Methods like RDOCK, ZDOCK and H-DOCK (Chen et al., 2003; Li et al., 2003) maintain the receptor and ligand molecules rigid throughout the docking. Forces of attraction which are van der Waals, hydrogen bonding, hydrophobic and electrostatic interactions facilitate intermolecular interactions among both small molecule and receptor. The complex crystal structure serves as a reference for comparing molecular docking results and predicting the suitable pose of compounds within the protein binding cavity (Chen et al., 2006). If the crystal structure for reference is not available, a better compound that binds with more interactions could be selected, and the compounds can be evaluated based on binding energy or binding scores. Homology models for receptors with unknown 3-D structures can be generated for docking (Sali & Blundell, 1993). Protein binding site prediction could be conducted in the absence of active site information for a protein using programmes such as sitemap (Halgren, 2009), Q-SiteFinder and CASTp (Laurie & Jackson, 2005; Tian et al., 2018). Molecular docking may also be applied to a single compound or on millions of compounds from a chemical database. Docking could be performed within binding pocket of a receptor for guided docking or in the whole receptor for blind docking (Hetényi & van der Spoel, 2006). Docking techniques that work to correctly prioritize docked compounds, search all binding pockets successfully and then use scoring functions to order the docked compounds (Kitchen et al., 2004). Docking could be used to conduct high-throughput screening on compound databases,

prioritize docked conformations, evaluate docked conformation binding to receptors via nonbonded interactions, and suggest structural hypotheses of how hit compounds inhibit the target, each of which is a valuable information for lead compound optimization. GOLD, DOCK, FlexX and ICM (Verdonk et al., 2003; Neves et al., 2012) are commonly used high-throughput docking tools. MD simulations can verify the stability of the hit chemical compound when bound to the specific receptor. Ligand scoring is a method for determining how well small molecules bind to protein binding cavities. The scores are determined by approximating mathematical functions that evaluate binding energy. Each and every docking process has its own scoring function for determining the stability of conformations. The popularly used scoring functions are the piecewise linear potential 1 and 2 (PLP1, PLP2) functions (Gelhaar et al., 1999; Gehlhaar et al., 1995), with PLP1 having a grid-based functional form and PLP2 having hydrogen bonding interactions that have an angular dependency. The potential of mean force (PMF, PMF04) scoring functions (Muegge, 2006; Muegge & Martin, 1999) score complexes by averaging the pairwise interaction terms of the receptor-hit molecule complexes across all interatomic pairs.

1.2.13 Drug repurposing

The process of FDA approved drugs being repositioned, repurposed, or re-tasked to be used in multiple diseases, i.e., outside of the purview of the original medical indication, is known as drug reprofiling or drug repurposing. The traditional drug development pipeline is time-consuming and expensive in terms of resources. Because of the time and money required to develop new drugs, several research organizations have focused their efforts on molecules that have been approved for human use in a disease (Ashburn & Thor, 2004). Some drugs, such as thalidomide derivatives (Sampaio et al., 1991), antibiotics (Konreddy et al., 2019), and antivirals, have been repurposed (Mercorelli et al., 2018). These drugs have achieved therapeutic success in

diseases other than those for which they were originally approved. Drug repurposing approaches are frequently classified as either drug-based or disease-based. When drug characteristics such as chemical, physical, and biochemical characteristics are accessible, drug-based techniques are most common and favoured. The availability of enormous computational capability in terms of space and cost, along with the advancement of high-throughput molecular, clinical, and molecular genetics methods, has opened up a new and suitable potential for rationale repurposing of existing drugs through computational methodologies rather than relying on chance for drug development. Following initial computational drug reprofiling discoveries, the findings are conformed using *in vitro* molecular screening, clinical and structure-based (biophysical) methods. In various rounds of clinical studies involving patient groups, these medications are even more verified. One can explore current PDBs and their 3-D amino acid patterns as well as the drug molecule binding interface on certain web servers, like Drug ReposER (Ghani et al., 2019).

1.2.14 Absorption, distribution, metabolism, excretion and toxicology

ADMET refers to the physical features that a drug-like compound must have in order to possess optimal pharmacokinetic properties while also having required pharmacological characteristics (Lipinski et al., 1997). The *in vitro* and *in vivo* ADMET properties collected from experimental studies indicate a thorough consideration of the pharmacokinetic properties of the chosen molecules. To evaluate the possible potential of the drug-like candidates in the initial phases of drug development, its effectiveness and safety parameters are critical. Before the molecule is studied in the first phase of clinical trials, it is necessary to record the safety profile of the drug-like molecule by evaluating its ADMET properties. Web servers are accessible for analysing the candidate compounds' physicochemical and drug-like characteristics, as well as their synthetic accessibility (Tian et al., 2015; Ertl & Schuffenhauer, 2009; Daina et al., 2017). These

theoretically calculated constraints based on their chemical structures will save time and funds while also accelerating the design of drug-like compounds that are likely to develop into drugs with a greater rate of success.

1.2.14.1 SwissADME sever

SwissADME (http://www.swissadme.ch/index.php) server is an online programme that calculates physicochemical characteristics like ADME properties as well as the physicochemical characteristics like solubility, lipophilicity and pharmacokinetic properties. Lipinski's rule of five (Lipinski et al., 1997; Lipinski et al., 2012; Lipinski, 2004) is a crucial requirement for ensuring a drug-like profile in orally administered drugs.

1.2.15 Molecular dynamics simulations

More than 60 years have passed since the first 3-D protein structure was determined by X-ray crystallography (Gutte, 1975), and the first MD simulations of proteins were performed in 1970, at the dawn of the modern computer (Levitt & Lifson, 1969; McCammon et al., 1977). Studies on protein structure engineering and the relation between sequence, structure and function were scarce at that time. Protein crystal structures that have been submitted to the PDB are regarded as static structures since they show the protein's configuration at a certain point in time. Protein structures are known to be dynamic instead, as the linkages between the single bonds move around, causing changes in conformation and occasionally even the function. The most popular approach, particularly for biological macromolecules like proteins, is to perform MD simulations to determine the conformational space filled by the molecules. One method to mimic their mobility is to use traditional MD simulations of protein structures derived from experiments, as well as computer models created using homology modelling (John & Sali, 2003; Dahiyat & Mayo, 1997). The position of the atom in 3-D space is investigated using MD simulations. This method replaces a

single-point model with a dynamic model that propels the nuclear system into motion. The traditional Newtonian dynamic equations are numerically solved to simulate the motion (Pace et al., 1996). The Newton's second law, also known as the equation of motion, F=ma, in which "F" stands for the force acting on the particle, "m" for its mass, and "a" for its acceleration, forms the foundation of the MD simulations technique. It is feasible to calculate each atoms acceleration in the system by knowing the force acting on it. The trajectory that results from the integration of the motion equations represents the locations, accelerations and velocities of the atoms as they change over a time scale. The average values of the attributes can be calculated from this trajectory. The process is deterministic, allowing for the prediction of the systems state at any point in the past or future once the locations and velocities of each atom are known. The contributions from interactions between bonded and nonbonded atoms are taken into account; nonbonded forces result from van der Waals interactions and are described using the Lennard-Jones potential, while charged (electrostatic) interactions are treated using Coulombs equation (Childers & Daggett, 2017; Geng et al., 2019). MD simulations can be time-consuming and expensive to run on computers (Shaw et al., 2008). Multiple software packages can be used to mimic protein flexibility through MD simulations of solvated proteins. CHARMM (Jo et al., 2008), AMBER (Case et al., 2005), CHARMm, DL_POLY (Smith et al., 2002), GROMOS (van Gunsteren & Berendsen, 1987), GROMACS (Lindahl et al., 2001), LAMMPS (Grindon et al., 2004), NAMD (Nelson et al., 1996) are some of the popular software capable of carrying out MD simulations.

1.2.15.1 Force fields

The "force field" refers to a computational methodology for a mathematical formula used to arrive at a set of parameters and calculate a proteins energy as a function of its all atom coordinates. A force field is a mathematical statement that describes how a system's energy

depends on the 3-D coordinates of its constituent components. The term "force field" refers to an equation and is divided into two terms: the first characterizes bonding interactions, which represents atoms connected by covalent bonds, and their mathematical form deals with bond lengths, bond angles, and torsion angles. The second term of the equation describes the contribution from nonbonding interactions caused by van der Waals and ionic interactions, and it is calculated using the Lennard-Jones potential and Coulombs law, respectively. It is defined in an analytical form to indicate the interatomic potential energy, and a set of parameters entering into the form.

The parameters of force field are often determined either through fitting to experimental data from techniques including X-ray and NMR, electron diffraction, infrared, Raman and neutron spectroscopy or through *ab initio* or semi-empirical quantum mechanical calculations (Weiner & Kollman, 1981; Chen & Yip, 2017). The force field substitutes a simplified model appropriate to the region being simulated for the real potential, which is simply defined as a collection of atoms bound together by simple elastic (harmonic) forces. It should ideally be basic enough to be evaluated quickly yet detailed enough to be able to recreate the characteristics of the systems being studied. A number of different types of force fields are available in the literature to describe compounds with varying levels of complexity and capable of handling different types of systems. The widely used force fields that include parameters for each and every atom in the periodic table includes Dreiding and Universal (UFF) force fields (Rappé et al., 1992). CHARMM (Brooks et al., 1983), GROMOS, AMBER, OPLS (Jorgensen et al., 1996), and COMPASS (Sun et al., 1998). For example, CHARMM19, CHARMM22, CHARMM27; GROMOS45A3, GROMOS53A5, GROMOS96, GROMOS53A6; AMBER02, AMBER91, AMBER94, AMBER96, AMBER99; etc. are just a few examples of these force fields that are constantly changing and have several

versions available (Maolepsza et al., 2010). Both the protein and the small chemical molecules under study must be compatible with the force field that is being used.

1.2.15.2 MD trajectory analysis

The trajectory of MD simulations is used to examine how atomic-level changes in biological and chemical structures occur over a period of time. Because knowledge about the proteins dynamical structure can only be achieved through MD simulations to construct an ensemble of structures, crystal structure is regarded as a snapshot of that structure. This is important for drug discovery research because it shows alternate protein conformations and, as a result, the allosteric sites present in the protein structure become available.

1.2.15.2.1 Root mean square deviation

The root mean square deviation (RMSD) is among the most frequently used quantitative metrics for the structural similarity of two or more molecules by the superposition of their 3-D atomic coordinates. The RMSD values are calculated for any kind or a subset of atoms in a molecules, such as all of the Cα-atoms in the protein as a whole or all of the atoms in a particular subset, or all of the carbon atoms in a protein that are involved in ligand binding. The secondary structural elements are stable and the loops are more flexible, which causes the RMSD of the system to increase. It has been shown that a system with stability will display a lower RMSD. The following equation can be utilized to determine RMSD.

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \qquad (1.1)$$

δ^2_i is the squared difference between the positions of the atom at index i and that same atom in the reference structure, and N is the number of atoms being counted.

1.2.15.2.2 Root mean square fluctuation

The term root mean square fluctuation (RMSF) can be employed to describe fluctuations around well-defined average positions in dynamical molecular systems. Data related to the temperature stability and regions of flexibility in the structure are provided by RMSF. The distance between the position of a atom and a reference position is measured by the RMSF.

$$RMSF = \sqrt{\frac{1}{N}\sum_{j}^{N}\left(x_{i(j)} - \langle x_i \rangle\right)^2} \qquad (1.2)$$

Where x_i represents the position at time i and $<x_i>$ denoted average value.

1.2.15.2.3 Hydrogen bonds

The nonbonding interactions that contribute to the stability of the molecular systems are categorized as ionic contacts, hydrogen bonds, hydrophobic and van der Waals interactions. These interactions contribute to the stability of the molecule in the form of intramolecular interactions and intermolecular interactions that stabilize the receptor-ligand interactions for sustenance in biological systems. An electronegative atom (donor) that has a covalent link with a hydrogen atom (acceptor) is attracted towards the lone pairs of electrons on an additional electronegative atom (acceptor), either within the same molecule (intramolecular hydrogen bond), or outside of it (intermolecular hydrogen bond). An average hydrogen bond contains 5–10% covalent bond character. Protein-ligand complexes are stabilized via hydrogen bonding. The directed connections provided by hydrogen bonds support protein structure and selectivity to molecular recognition through intermolecular interactions. A hydrogen bond must have a distance between hydrogen donor and acceptor of less than 3.2 Å and a D-H-A angle of 180° ± 30° in order to be considered to exist. According to Van Der Spoel et al. (2005), the GROMACS command "gmx hbond" can

compute the number of hydrogen bonds in a protein complex as well as the distance between the acceptor and donor atoms participating in hydrogen bonds. The number of intermolecular hydrogen bonds that are present in the docked conformation and those that are preserved during MD simulations are analysed by indexing the number of intermolecular hydrogen bonds between particular atoms.

1.2.15.2.4 Normal mode analysis

A quick and easy method for determining protein flexibility and vibrational modes is normal mode analysis (NMA) (Bahar et al., 2010). The atoms in NMA are modelled as point masses connected by springs, which stand for the interatomic force fields, and are occasionally restricted to Cα atoms exclusively. In order to disclose the dynamic properties of proteins, NMA have been created (Velázquez-Muriel et al., 2009; Bakan et al., 2011). The NMA is employed for analyses of protein structures and to study the slow dynamics and large-scale motion in biomolecules. One specific kind of NMA is the elastic network model. In this model, only the atom pairs that are close to a cutoff distance are taken into account, and all of the springs linking each node to its neighbors are of similar strength.

1.2.15.2.5 Mechanical stiffness

The pulling or stretching forces needed to start protein unfolding differ significantly depending on where the pressures are applied, which causes a different reaction in each protein to uniaxial strain. Anisotropic Network Model (ANM), a set of software tools included in Prodynamics, is used to create mechanical stiffness charts for all molecular systems in response to all potential pulling directions (Eyal et al., 2015).

1.2.15.2.6 Principal component analysis

In all the simulated systems, PCA was used to analyse the overall mobility of the residues in the protein structure. By lining up the structures from MD simulations trajectories, the MODE-TASK programme (Ross, C., et al. 2018,) was used to eliminate the translational and rotational motions of the protein Cα atoms. Following the creation of the 3N x 3N covariance matrix in Cartesian coordinates, the covariance matrix diagonalization was used to produce the eigenvectors. The PCA was generated using MD simulations trajectory data.

1.2.15.3 Binding free energy

For determining binding free energies in biological macromolecular systems like proteins, a variety of techniques with varying degrees of complexity have been used. Simplified scoring systems are used to achieve the necessary proficiency in screening vast chemical databases of small molecules to find a hit molecule that may potentially develop into a lead and therapeutic molecule (Parenti & Rastelli, 2012). The binding free energy can be estimated using quadratic fluctuations and continuum solvent approximation around a specific configuration of the molecular system (Kollman et al., 2000). A widely used methodology is Molecular Mechanics Poisson-Boltzmann and Surface Area (MM-PBSA) relies on a mixed scheme which combines free energy estimators based on an implicit continuum solvent model with configurations sampled from MD simulations with explicit solvent (Srinivasan et al., 1998; Hou et al., 2011). The MM-PBSA method is employed to determine the polar, non-polar and binding free energies of biomolecules (Gilson & Honig, 1988; Sitkoff et al., 1994). The command ‘g mmpbsa’ from a GROMACS trajectory output determines the binding free energy to a protein ligand complex (Kumari et al., 2014), and this tool includes a variety of non-polar solvation models, such as those based on the solvent accessible surface area (SASA), solvent accessible volume (SAV), and a model that covers

repulsive (SASA SAV). It also estimates the energy contribution of individual amino acid residues to the binding free energy. The binding free energy is calculated using the following formulae. When a protein is complexed with an inhibitor in a solvent such as water, the binding free energy is given as

$$\Delta G_{bind} = G_{complex} - G_{free\text{-}protein} - G_{free\text{-}inhibitor} \quad (1.3)$$

where, $G_{complex}$ is the total free energy of the protein-inhibitor complex and $G_{free\text{-}protein}$ and $G_{free\text{-}inhibitor}$ are total free energies of the isolated protein and inhibitor in the solvent, respectively.

The free energy of each individual entity "G" indicated above is represented by

$$G = E_{MM} - TS + G_{solvation} \quad (1.4)$$

TS denotes the entropic contribution to free energy in vacuum, where T and S represent temperature and entropy, respectively. $G_{solvation}$ refers to the free energy of solvation, which is the amount of energy required to move a solute from vacuum into a solvent. This is expressed as the total of G_{polar} and $G_{non\text{-}polar}$, the electrostatic and non-electrostatic contributions to the solvation free energy, respectively.

$$G_{solvation} = G_{polar} + G_{non\text{-}polar} \quad (1.5)$$

E_{MM}, which is determined using the parameters of the molecular mechanics (MM) force field, is the average molecular mechanics potential energy in vacuum, which contains the energy of both bonded and nonbonded interactions.

$$E_{MM} = E_{bonded} + E_{nonbonded} = E_{bonded} + (E_{vdW} + E_{elec}) \quad (1.6)$$

where E_{bonded} is bonded interactions consisting of the bond length, angle, dihedral and improper interactions. The $E_{nonbonded}$ is the nonbonded interactions that include both electrostatic and van der

Waals interactions and is modeled using Coulomb and Lennard-Jones (LJ) potential functions, respectively.

1.2.15.3.1 Linear interaction energy

The ligand-binding free energy (ΔG_{LIE}) (Almlöf et al., 2004;Brandsdal et al., 2003) was computed as the mean of the inhibitors Coulomb and van der Waals interaction differences with its neighboring atoms upon incorporation, that is, the individual ligand in the solvent (unbound state denoted as subscript u) and the inhibitor in the binding mode with protein (bound state denoted as subscript b) approach using gmx lie and gmx energy were calculated for the protein-ligand molecule complexes from the output trajectories of MD simulations.

$$\Delta G_{LIE} = \alpha \left(\langle V_{l\text{-}s}^{vdW} \rangle b - \langle V_{l\text{-}s}^{vdW} \rangle u + \beta \langle V_{l\text{-}s}^{cou} \rangle b - \langle V_{l\text{-}s}^{cou} \rangle u\right) + \gamma \qquad (1.7)$$

The above equation was used for calculating LIE. The coefficient γ, a constant, is associated with the alteration of the hydrophobic nature of the binding cleft conceding to various species of inhibitors, whereas the coefficients α and β are rating parameters for nonpolar and polar interactions, respectively. The values taken for α, β and γ are 0.181, 0.5 and 0, respectively.

1.2.16 Density functional theory calculations

DFT is a computational quantum mechanical modelling technique used to study the electronic structure (or nuclear structure) (primarily the ground state) of many-body systems such as atoms, molecules, and condensed phases in physics, chemistry, and materials science. This hypothesis can be used to determine the characteristics of a many-electron system. *Ab initio* DFT calculations, from the standpoint of computational materials science, allow for the prediction and

estimation of material behavior that depends on quantum mechanical factors without the need for higher-order parameters such as fundamental material properties. The electronic structure of a system is examined using a potential acting on its electrons in modern DFT techniques. This DFT potential is constructed by adding external potentials. V_{ext} is determined solely by the structure and elemental composition of the system, and V_{eff} is an effective potential that represents interelectronic interactions.

Geometry optimizations on all molecules were performed at the B3LYP (Becke, 1993; Lee et al., 1988) level of theory using the 6-311+G(2d,p) basis set (Petersson et al., 1988; Petersson et al., 1991). Furthermore, Polarizable Continuum Model (PCM) (Miertuš et al., 1981, Tomasi et al., 1994) optimizations with water as an implicit solvent were performed at the B3LYP/6-311+G(2d,p) level for all molecules to determine the effect of solvent presence on total energy of the molecule. The frequency calculations were then used to determine the characteristics of the stationary point. Molecular orbital analysis was performed at the B3LYP/6-311+G(2d,p) level of theory to find the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) (Fukuli et al., 1952) for the molecules studied through MD simulations. The purpose of this analysis was to identify the electron-rich zone in the molecules under consideration. To accurately analyse the charge distribution around these atoms, the electrostatic potential (ESP) energy values were determined (Vetrivel et al., 1996) and mapped over an isodensity surface equivalent to 0.0004 a.u. This ESP incorporates the van der Waals volumes of the molecules individual atoms, providing an accurate representation of the reactive regions surrounding the inhibitors. Gauss View was used to visualize the molecular structures. All calculations were carried out using the Gaussian 09 (Frisch et al., 2009) programme suite.

CHAPTER-2

Structure stability and drug repurposing of Zika virus NS3 helicase

2.1 Introduction

Most viral illnesses are brought on when a virus enters the body of a host and exploits that system to proliferate. Infections result when viruses multiply and spread to other cells because the host immune system is unable to prevent them. The following are significant signs of viral illnesses: Fever, exhaustion, sore throat, headache, irritability, rashes, malaise, sneezing, swollen tonsils, and excessive weight loss. The Zika virus is an arbovirus that is transmitted by aedes mosquitos via blood and other bodily fluids. The Zika virus caused mild illness and rash at first, but it was later discovered to cause Zika virus disease, which is characterized by fever, headache, arthralgia, myalgia, and maculopapular rashes (Loos et al., 2014). Because there are no specific drugs available to prevent or treat Zika virus infection, new drugs for the treatment of this critical disease are required.

New drug discovery is time-consuming and expensive, with a low success rate most of the time. Drug repurposing, also known as drug repositioning or reprofiling, is a strategy for identifying new uses for previously approved drugs (Ashburn & Thor, 2004). It is useful for identifying drugs that can function as multi-disease inhibitors. These drug molecules have already been tested in humans, and comprehensive information on their pharmacology, formulation, dosing, and potential toxicity is available (Ashburn & Thor, 2004, Pushpakom et al., 2019). As a result, this method has advantages over developing new drugs for a disease; specifically, the risk of drug failure is lower because the repurposed drug has already been proven to be sufficiently safe, resulting in a shorter timeframe for drug identification and a lower budget investment. For example, zidovudine was originally developed to treat cancer but was later redeveloped to treat acquired immune deficiency syndrome (AIDS) that is caused by HIV (Marchbanks, et al 1995).

The antimalarial drug hydroxychloroquine is an autophagy inhibitor; *in vitro* studies showed that it inhibited dengue virus (Al-Bari, 2015), and it is now being used to treat SARS-CoV-2 disease (Chen et al., 2020). Ivermectin (Mastrangelo et al., 2012), suramin (Basavannacharya et al., 2014), and aurintricarboxylic acids (Shadrick et al., 2013) have been identified as drugs that inhibit flavivirus helicase. Amodiaquine, an antimalarial drug, inhibits Zika virus pathogenicity by blocking autophagy (Han et al., 2018). A patent review of repositioning and investigational drugs for Zika virus disease identified ribavirin, sofosbuvir, α-interferons, lopinavir-ritonavir combination, and niclosamide as Zika virus inhibitors in clinical trials (Rosa et al., 2020). In human clinical trials, some Zika virus vaccine candidates based on nucleic acid vaccines, inactivated vaccines, viral-vectored vaccines, and attenuated vaccines have shown significant promise (Pattnaik et al., 2020). However, a systematic investigation for specific targets to discover efficient and swift-acting drugs are needed for the Zika virus because the virus spreads very fast when it infects the human host.

Flavinoids and their derivatives have been shown to inhibit the activity of the Zika virus proteins NS1 (Ahmad et al 2020) and NS2B-NS3 (Yadav et al 2021). The NS3 helicase is essential in the replication of viral genomes. RNA helicases are essential for RNA virus RNA metabolism and viral infection facilitation. The Zika virus NS5 RdRp generates a negative-sense RNA by using the positive-sense RNA as a template for the formation of the intermediate stage double-stranded RNA. The Zika virus NS3 helicase unwinds the double-stranded RNA to separate it into negative- and positive-sense single-stranded RNA molecules. Negative-sense RNA serves as a new template for the production of positive-sense genomic RNA (Xu et al., 2019). Because of its critical role in viral RNA replication, NS3 helicase has been identified as an important target for antiviral drug discovery. X-ray diffraction methods were used to report the 3-D structures of helicase from

dengue, hepatitis C, Zika, and coronaviruses, and their 3-D coordinates are deposited in the PDB (Abola et al., 1984). The RNA binding domain of the Zika virus NS3 helicase exhibits ATPase activity, which provides chemical energy that is then converted into mechanical energy required for viral RNA genome unwinding during viral replication (Tian et al., 2016). Helicase, in collaboration with the enzyme RdRp, aids in viral genome replication. Zika virus NS3 helicase is an significant target for drug development against Zika infections due to its critical role in genome replication. Using computational studies, the 1,4-benzothiazine derivatives were shown to be Zika virus NS3 inhibitors (Badshah et al., 2019). The crystal structures of Zika virus NS3 helicase complexed with ATP or RNA explain how the virus identifies its substrates during replication and provide structural insights for rational Zika virus drug design. However, the mechanisms by which Zika virus helicase distinguishes between the binding of nucleoside triphosphate (ATP) and viral RNA remain unknown, making the development of antiviral drugs difficult. These structural stability studies would aid in determining the molecular basis for cofactor/substrate binding to Zika virus NS3 helicase activity. Using molecular docking, MD simulations, and post-MD analyses, structural changes in the Zika virus NS3 helicase when bound to the substrate (RNA), cofactor (ATP), simultaneously binding to both ATP-RNA, and repurposed drugs are reported in this chapter.

2.2 Materials and Methods

2.2.1 Dataset

The crystal structures of the apo Zika virus NS3 helicase (PDB id: 5JMT) and when bound to ATP (5GJC) and RNA (5GJB) are available at 1.8, 2.2, and 1.7 Å resolutions, respectively. Crystal waters were removed from the structures, and the missing residues in the structure (5GJC) were built using the loop modelling method in MODELLER (Sali & Blundell, 1993). The valency

of all atoms in the structures were satisfied by the addition of hydrogen atoms, and the geometry of the structures was optimised using UCSF Chimera 2.1.1 (Petersen et al., 2004).

2.2.2 Drug repurposing and molecular docking studies

It has been proposed that some drugs not only interact with their therapeutic protein targets but also inhibit other proteins (Sanseau et al., 2011). **Figure 2.1** depicts a conceptual diagram of target-based drug repositioning and the outcome of our workflow. The first steps in drug repositioning include a high-throughput *in silico* virtual screening of FDA approved drugs obtained from the BindingDB. This is followed by virtual screening based on molecular docking and subsequent validation steps. Molecular docking entails creating different binding poses for a ligand within the target active site for conformational sampling, as well as evaluating the binding strength of each protein-ligand complex based on the extent of nonbonding interactions for scoring the pose (Meng et al., 2011). Both docking methods used in this study, AutoDock (Morris et al., 2009) and CDOCKER (Gagnon et al., 2016), were validated by docking the cofactor ATP into the binding site of the NS3 helicase.

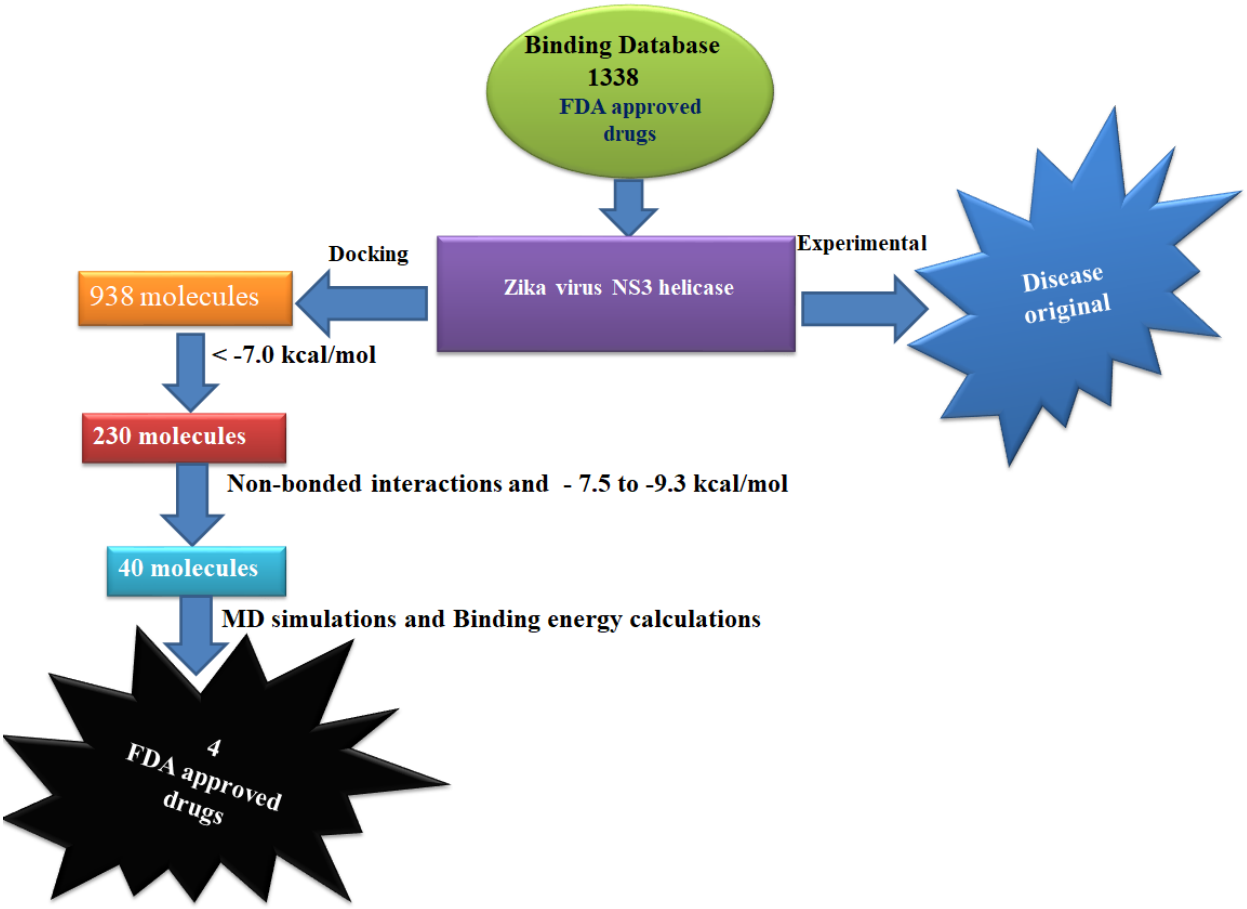


Figure 2.1. Repurposing of FDA approved drugs for Zika virus NS3 helicase.

The BindingDB (Liu et al., 2007; Gilson et al., 2016) is an exhaustive and freely accessible web resource that contains detailed drug, drug-target, drug action, and drug interaction information for FDA approved drugs as well as experimental drugs in the FDA approval process. The structures of 1,338 FDA approved drugs have been deposited in BindingDB (https://www.bindingdb.org/). These molecules 3-D coordinates were downloaded in .sdf format from BindingDB. The active site of Zika virus NS3 helicase was identified using key amino acid residues that interact with the cofactor, ATP. Virtual screening of FDA approved drugs using AutoDock 4.2 tools was employed to dock the compounds within the 5 Å cavities defined around the ATP binding pocket. This

revealed their predicted binding affinity, binding orientation within the active site of the Zika virus NS3 helicase, and best binding conformations in order to shortlist the FDA approved drugs. The protein structure was loaded into AutoDock tools (Morris et al., 2009) to generate the ten best conformations for each molecule. The molecules were initially loaded; torsions were determined and saved in PDBQT format. All the heteroatoms and crystal waters were deleted from the loop modeled structure of Zika virus NS3 helicase (5GJC) prior to molecular docking and saved in PDBQT format. The Lamarckian Genetic algorithm was used to perform all calculations for protein-ligand flexible docking (Wiley, et al 2006). A grid box with the dimensions X: 22.987 Å, Y: 19.338 Å and Z: 49.646 Å was used, with a grid spacing of 0.492 Å as the default. The best conformation with the highest binding affinity and greater number of hydrogen bonding interactions between the protein and the docked pose of an FDA approved drug was manually analysed on graphics.

The top-ranked molecules were then subjected to a second round of docking studies with CDOCKER (Gagnon et al., 2016), which is available in DS 3.5. To define the proteins active site, a sphere with a radius of 5 Å was generated around ATP. Twenty docking poses were generated for each molecule, and the conformations of molecules generated in the active site of the protein were shortlisted. The system was heated to 700 K for 2,000 steps as part of the docking protocol, then cooled to 300 K for 5,000 steps. The binding conformations of the molecules in Zika virus NS3 helicase were analysed using the "scoring ligand poses" implemented in the receptor-ligand interactions protocol in DS 3.5, and the docking poses were analysed using the scoring functions PLP1, PLP2 and PMF (Gehlhaar et al., 1995; Parrill et al., 1999; Muegge et al., 1999; Muegge, 2006). Top scores and intermolecular interactions with the Zika virus NS3 helicase were used to

select the best docking pose. The top-ranked molecules from the CDOCKER docking method were studied further using MD simulations to assess complex stability.

2.2.3 Molecular dynamics simulations

MD simulations are computational methodologies used to examine the physical and conformational changes of proteins as well as their interactions with other molecular species in a variety of environments (Beveridge & Ravinshanker, 1994; Cheatham et al., 1995; Cheatham et al., 1997; Young & Beveridge, 1998; Reyes & Kollmann, 1999; Tang & Nilsson, 1999). GROMACS 5.1.4 (Hess et al., 2008; Van Der Spoel et al., 2005) was used to study the MD simulations of all molecular systems for 150 ns. There were eight various types of molecular systems studied: apo Zika virus NS3 helicase, Zika virus NS3 helicase complexed with ATP, RNA, ATP-RNA, and four FDA approved drugs located within the proteins ATP binding site. Hydrogen atoms were added to all molecular systems, and the coordinates were saved in .mol2 format. AMBER03 force fields were applied by UCSF Chimera using AM1-BCC charges and saved in .mol2 format. ACPYPE script was used to generate topology of the compounds (Da Silva et al., 2012). MD simulations were run for 150 ns for each molecular system to investigate the mechanism of ATP-RNA binding to Zika virus NS3 helicase and the stability of repurposed drugs within the protein active site in order to validate them.

The molecular systems were solvated in a cubic box with water as the solvent, and the systems total charge was neutralised by the addition of Na^+ and Cl^- ions (Berendsen et al., 1981). To optimise the geometry of the systems, the steepest descent method was used for 1,000 steps of energy minimization. The distribution of water molecules was achieved during the position restraint phase, and the systems were equilibrated under NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature)

conditions for 100 ps each using the leap-frog integrator method. For long-range electrostatics, the particle mesh Ewald method was used, temperature coupling was set using V-rescale (Bussi et al., 2007), modified Berendsen thermostat (Berendsen et al., 1984) at 300 K, and pressure was maintained at 1 atm using the Parrinello-Rahman (Parrinello & Rahman, 1981) method. The MD simulations of all systems were executed using the periodic boundary conditions with a cutoff of 1.0 nm, the Lincs algorithm (Hess et al., 1997) was used to constrain the hydrogen bond parameters, and the final MD simulations in the production phase were run for 150 ns. The MD trajectories were analysed with the GROMACS utilities gmx rms, gmx rmsf, gmx gyrate, and gmx cluster (Van Der Spoel et al., 2005) to obtain the RMSD, RMSF, Rg and cluster analyses. VMD was used to visualise the MD trajectory file analyses (Humphrey et al., 1996); Chimera and Pymol were used to generate cartoon image representations. The RMSD, RMSF and Rg graphs were plotted using the Xmgrace software (http://plasma-gate.weizmann.ac.il/grace/). The secondary structural changes in the MD simulations as a function of simulations time were examined using GROMACS gmx do dssp command.

2.2.4 Normal Mode Analysis

Protein flexibility is a major challenge in the accurate prediction of protein-ligand docking and dynamics. NMA can provide a quick and systematic investigation of protein dynamics. An elastic network model-based NMA using dihedral angels as independent variables for all molecular systems was developed using the software suite of programs available in Prodynamics (Uyar et al., 2011 and Atilgan et al., 2001).

2.2.5 Mechanical Stiffness

It has been demonstrated that the pulling/stretching forces required to initiate protein unfolding vary considerably depending on the location of the application of the forces, thus

unraveling specific response of protein to uniaxial tension. Mechanical stiffness plots of all molecular systems in response to all possible pulling directions were constructed by using ANM using the software suite of programs in Prodynamics (Eyal et al., 2015).

2.2.6 Principal component analysis

PCA has been used to examine the overall motion of R-loop residues in the Zika virus NS3 helicase in all simulated systems. MODE-TASK software (Ross et al., 2018) was used to eliminate the translational and rotational motions of the protein Cα-atoms by aligning the structures from MD simulations trajectories. The 3N/3N covariance matrix was then constructed using Cartesian coordinates, followed by the construction of eigenvectors by diagonalization of the covariance matrix. The PCA was computed using MD simulations trajectories ranging from 0 to 150 ns.

2.2.7 Binding free energy calculations

The g mmpbsa tools (Kumari et al., 2014) were employed to evaluate the strength of ATP and FDA approved molecules binding to Zika virus NS3 helicase. These tools were designed to work with GROMACS output trajectories in order to calculate the binding free energy of each molecular system. Based on the RMSD results, binding free energy calculations were run on the stable region of MD simulations trajectories (140-150 ns). These binding free energy calculations predict the effective participation of amino acid residues in the Zika virus NS3 helicase in binding to ligands, as well as the contribution from van der Waals, electrostatic and polar and apolar solvation energy terms.

2.3 Results and Discussion

2.3.1 Zika virus NS3 helicase complexed with ATP, RNA and ATP-RNA

The missing residues in the ATP-bound Zika virus NS3 helicase (5GJC) structure were constructed by using MODELLER 9.17 software. Among the 5 generated models, the model with

the highest negative distinct optimized protein energy (DOPE) score was used for further studies. The tertiary structure of the Zika virus NS3 helicase is composed of three domains, each of which contains 130-160 amino acid residues. Despite their low sequence homology, domains I (residues 192-332) and II (333-481) have a similar structural fold. Domain III (482-617) is made up of a four-helical bundle that is extended by two anti-parallel β-strands that are partially exposed to the solvent. The protease domain is represented by the amino acid residues 1-174. Two α-helices from domain I interact with a parallel α-helical bundle from domain III; domain II connects to domain III via a long β-hairpin, stabilising interdomain interactions (Tian et al., 2016), and all the three domains have clearly defined binding clefts. The cofactor ATP is located between the domains I and II in Zika virus NS3 helicase; it is stabilized by the Mn^{2+} coordinated in octahedral geometry by interactions with side-chain of residues Thr201 and Glu286, two oxygen atoms of β and γ phosphate groups of ATP and two water molecules. The binding site of ATP is formed by the residues Gly197, Ala198, Gly199, Lys200, Thr201, Arg202 and Arg203 (P-loop), and Glu286, Ala317, Asn330, Gly415, Asn417, Gln455, Gly458, Arg459, Arg462 and Asn463. The single-stranded RNA binds the Zika virus NS3 helicase within the region formed by the amino acid residues; Pro224, Thr225, Arg226, Val227, Met244, Thr245, Thr246, Cys262, Ala264, Thr265, Phe289, Asp291, Pro292 (domain I), Pro364, Ser365, Val366, Arg367, Ser387, Arg388, Thr409, Asp410, Ile411, Leu430, Lys431, Pro432, Leu442 (domain II), His486, Lys537, Asp540, Arg598 and Ser601 (domain III). The single-stranded RNA makes hydrogen bonding interactions with Arg226, Thr245, Thr265, Asp291, Val366, Arg388, Thr409, Asp410 and Lys431. The ATP and single-stranded RNA binding sites in Zika virus NS3 helicase are shown in **Figure 2.2A**.

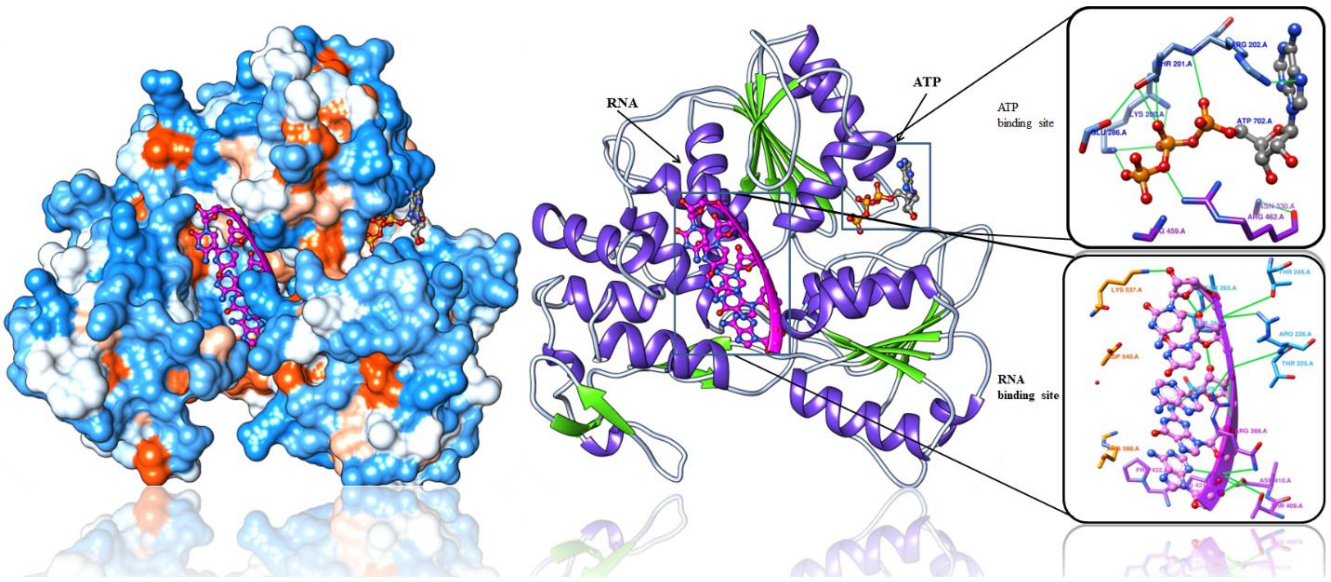


Figure 2.2A: Zika virus NS3 helicase bound to ATP-RNA, amino acid residues in hydrogen bonding interactions are indicated as green lines.

2.3.2 Virtual screening and molecular docking

To validate the docking methods, the loop-modeled Zika virus NS3 helicase (5GJC) was used to redock the cofactor ATP into its binding site using AutoDock 4.2 and CDOCKER. The cofactor docked into the active site pocket. It is located in the cavity formed by the residues Leu194, His195, Pro196, Gly197, Ala198, Gly199, Lys200, Thr201, Arg202, Arg203, Glu286, Ala317, Asn330, Gly415, Asn417, Gln455, Gly458, Arg459, Arg462 and Asn463 and forms hydrogen bonding interactions with the amino acids, Gly197, Gly199, Lys200, Thr201, Arg202, and Arg462. The crystal structures ATP superposition and molecular docking is shown in **Figure 2.2B**.

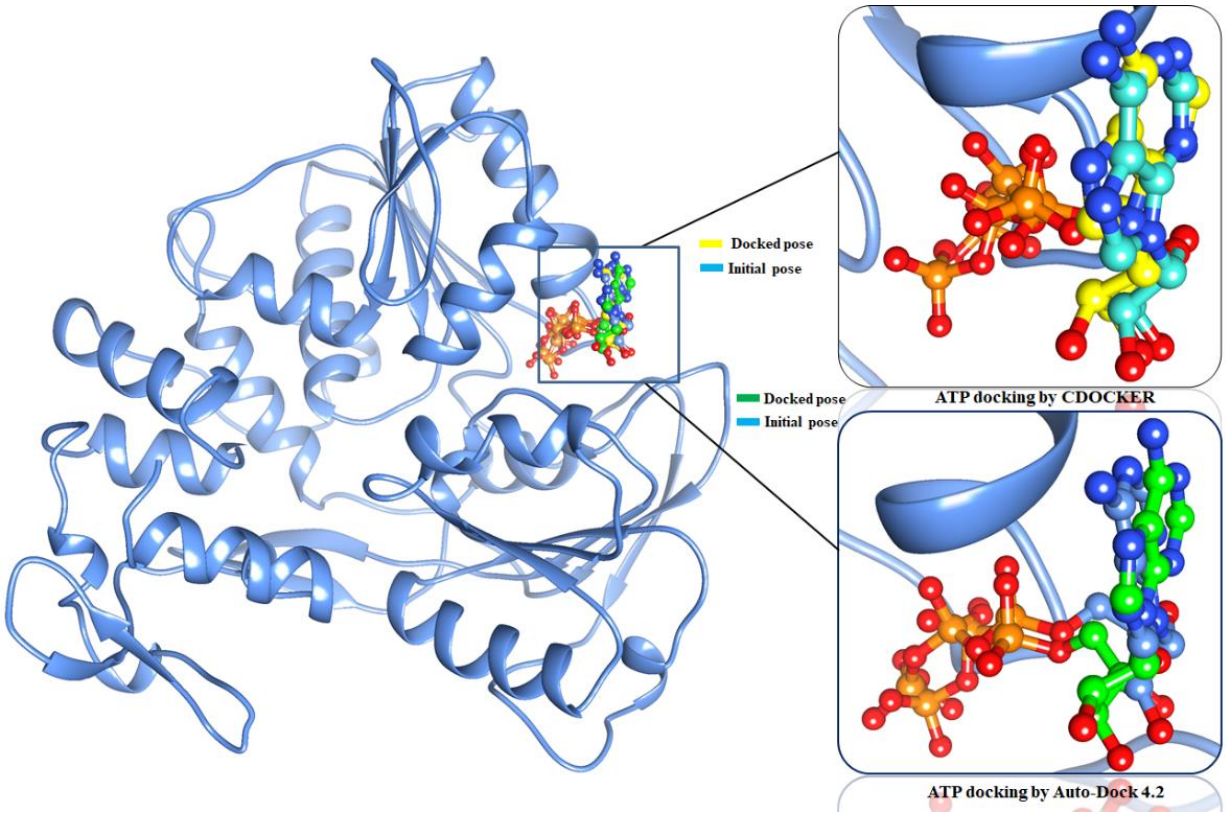


Figure 2.2B: docking of ATP to Zika virus NS3 helicas CDOCKER and AutoDock.

The virtual screening of 1,338 FDA approved drug molecules to bind the cofactor ATP binding site of Zika virus NS3 helicase was studied by using AutoDock 4.2 software. Out of these, only 938 molecules could be docked into the ATP binding site of NS3 helicase. From these docked compounds, 230 molecules were predicted to have binding affinity better than -7.0 kcal/mol. The docking conformers of each molecule was analysed and selected the molecules based on the docking score, docking conformation, and hydrogen bonding interactions mediated by key residues in the ATP binding site. Of these, the best 40 compounds were selected that had a docking score between (-9.3 to -7.5 kcal/mol). These molecules were further confirmed by docking into the ATP binding pocket of Zika virus NS3 helicase using CDOCKER. The docking scores and amino acid residues mediating interactions in the binding site using AutoDock and CDOCKER methods are shown in **Table 2.1**.

Table 2.1: The docking scores of the best selected molecules using AutoDock and CDOCKER and the interacting amino acids in the NS3 helicase.

S. No.	Compound Name	AutoDock score -kcal/mol	CDOCKER score			Active site amino acid residues
			-PLP	-PLP2	-PMF04	
1.	Lapatinib GW57201	9.3	82.85	73.09	55.12	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, **Thr201**, **Arg202**, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
2.	Dicumarol DB00266	9.0	73.48	52.49	69.41	Leu194, His195, Pro196, **Gly197**, Ala198, **Gly199**, **Lys200**, **Thr201**, **Arg202**, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, **Asn417**, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
3.	Sorafenib DB16673	8.8	70.48	49.36	61.32	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, Lys200, Thr201, Arg202, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
4.	Votrient DB26474	8.6	69.48	48.24	60.29	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200,** Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419,

						Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
5.	Fexofenadine DB22874	8.5	75.36	51.36	68.32	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, **Thr201**, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
6.	Beroccapn DB50367343	8.5	74.31	63.48	77.96	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, **Arg202**, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
7.	Pemetrexed DB18796	8.3	64.56	46.14	43.62	Leu194, His195, Pro196, Gly197, Ala198, Gly199, Lys200, **Thr201**, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
8.	Raloxifene DB1944	8.2	78.86	72.38	55.42	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
9.	Raltegravir DB25351	8.2	72.16	53.86	64.92	Leu194, His195, Pro196, Gly197, Ala198, Gly199, **Lys200**, Thr201, **Arg202**, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455,

						Gly458, Arg459, Arg462, Asn463, Pro464
10.	Amiodarone DB01118	8.2	76.51	65.81	77.38	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
11.	Povan DB78435	8.2	65.54	49.18	46.91	Leu194, His195, Pro196, **Gly197**, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
12.	Bicalutamide DB18678	8.1	79.68	71.48	56.21	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
13.	Linezolid DB00601	8.1	77.21	66.42	75.98	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**,Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
14.	Respiridone DB50001885	8.1	71.11	62.68	76.96	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419,

						Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
15.	Erivedge DB50249522	8.1	68.24	44.19	48.72	Leu194, His195, Pro196, Gly197, Ala198, Gly199, **Lys200**, **Thr201**, **Arg202**, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, **Arg459**, Arg462, Asn463, Pro464
16.	Rondomycin DB50368780	8.1	73.89	72.56	58.21	Leu194, His195, Pro196, **Gly197**, Ala198, **Gly199**, **Lys200**, Thr201, **Arg202**, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
17.	Lopinavir DB578	8.0	73.86	54.16	63.96	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
18.	Roxyzine pamoate DB38549	8.0	69.51	63.81	76.37	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
19.	Indapamide DB25901	7.9	67.64	47.16	49.96	Leu194, His195, Pro196, **Gly197**, Ala198, **Gly199**, **Lys200**, **Thr201**, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455,

						Gly458, Arg459, Arg462, Asn463, Pro464
21.	Chlorthalidone DB25900	7.9	70.68	68.48	58.21	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
22.	Diulo DB25899	7.9	65.91	61.83	73.32	Leu194, His195, Pro196, Gly197, Ala198, Gly199, **Lys200**, **Thr201**, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
23.	Tadalafil DB14777	7.9	67.59	47.27	46.95	Leu194, His195, Pro196, **Gly197**, Ala198, **Gly199**, Lys200, **Thr201**, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
24.	Vardenafil DB14476	7.8	72.88	70.48	56.21	Leu194, His195, Pro196, **Gly197**, Ala198, **Gly199**, Lys200, Thr201, Arg202, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
25.	Fenofibrate DB28700	7.8	73.21	64.43	72.99	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419,

						Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
26.	Azactam DB50240480	7.8	70.19	61.67	75.97	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
27.	Buspar DB50001859	7.8	67.54	45.18	49.71	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
28.	Terconazole DB31769	7.7	71.81	72.51	56.25	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
29.	Azulfidine DB50143010	7.7	67.55	46.17	49.78	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
30.	Floxuridine DB00322	7.6	65.58	44.11	43.35	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, **Glu231**, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419,

						Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
31.	Celecoxib DB11639	7.6	65.57	41.19	59.72	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
32.	Iressa DB00317	7.6	72.89	67.51	56.47	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, Lys200, Thr201, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
33.	Belinostat DB05015	7.5	79.85	73.19	55.22	Leu194, His195, Pro196, Gly197, Ala198, Gly199, **Lys200**, **Thr201**, **Arg202**, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, **Asn417**, Lys419, Gln455, Gly458, Arg459, **Arg462**, Asn463, Pro464
34.	Roflumailast DB14774	7.5	76.51	46.28	53.69	Leu194, His195, Pro196, Gly197, Ala198, **Gly199**, Lys200, Thr201, Arg202, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
35.	Dasatinib DB82071	7.5	69.87	73.57	59.29	Leu194, His195, Pro196, Gly197, Ala198, Gly199, **Lys200**, Thr201, Arg202, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455,

						Gly458, Arg459, **Arg462**, Asn463, Pro464
36.	Brominate dm DB50366613	7.5	67.15	47.43	57.18	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, Lys200, Thr201, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, **Gly415**, **Asn417**, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
37.	Tazobactam DB50157692	7.5	65.58	44.11	43.35	Leu194, His195, Pro196, **Gly197**, Ala198, **Gly199**, **Lys200**, **Thr201**, **Arg202**, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
38.	Tindal DB82475	7.5	69.51	46.21	61.71	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, **Lys200**, Thr201, **Arg202**, Arg203, Ala235, **Glu286**, His288, , Asn317, Asn330, Gly415, Asn417, Lys419, Gln455, Gly458, Arg459, Arg462, Asn463, Pro464
39.	Furadantin DB57045	7.5	71.91	69.61	57.17	Leu194, His195, Pro196, Gly197, Ala198, Gly199, Lys200, Thr201, Arg202, Arg203, Ala235, **Glu286**, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Lys419, Gly458, Arg459, **Arg462**, Asn463, Pro464
40.	Arestin DB6602603	7.5	65.58	44.11	43.35	Leu194, His195, Pro196, **Gly197**, Ala198, Gly199, Lys200, Thr201, Arg202, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, Asn417, Arg459, **Arg462**, Asn463,

The best four molecules identified from both the docking methods and MD simulations were dicumarol (DB00266), linezolid (DB00601), floxuridine (DB00322) and belinostat (DB015015), that have a range of docking scores (-9.0, -8.1 -7.6 and -7.5. kcal/mol), respectively as shown in **Figure 2.3** These molecules make hydrogen bonding interactions with active site residues Gly197, Gly199, Lys200, Thr201, Arg202, Glu286, Asn417, Arg459 and Arg462. These best docking pose of Zika virus NS3 helicase complexed with four repurposed drugs obtained using CDOCKER docking were further studied by using MD simulations studies.

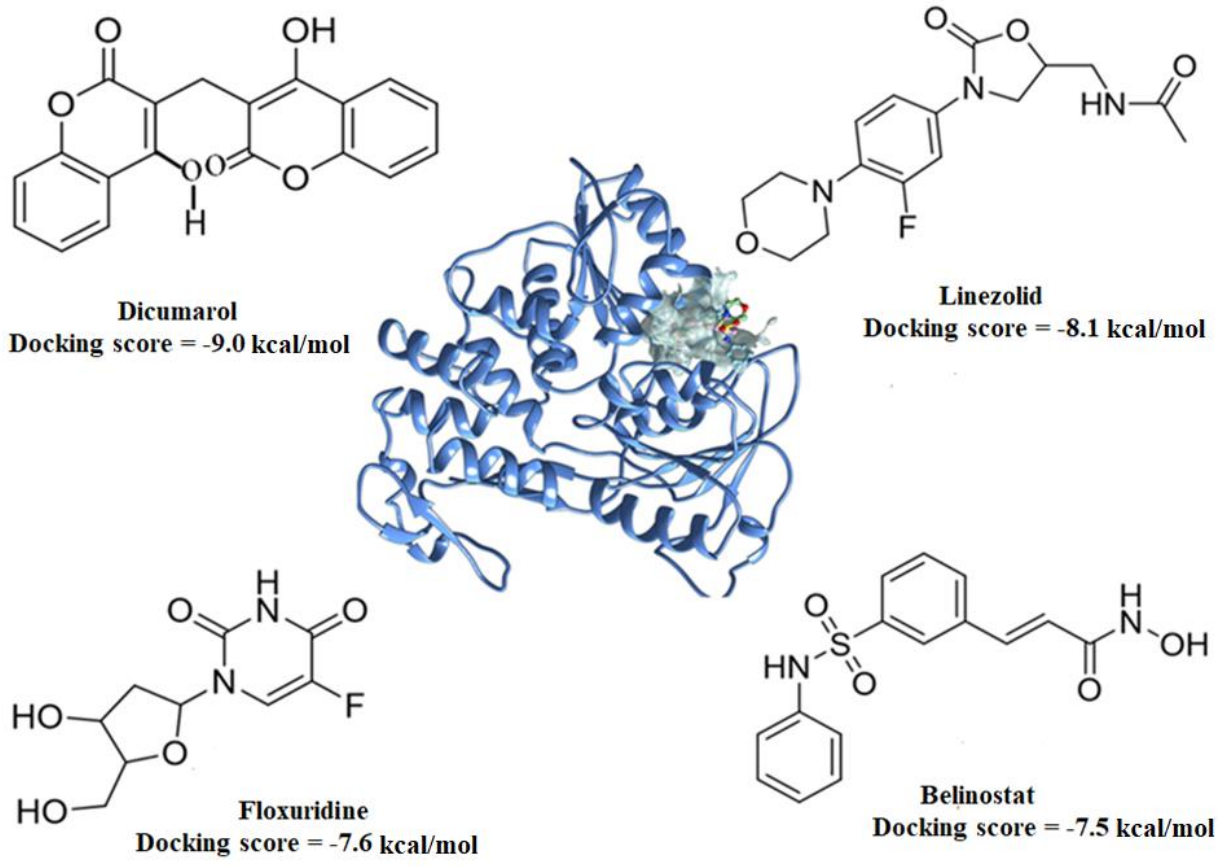


Figure 2.3: The selected docked structures of repurposed compounds binding to Zika virus NS3 helicase along with AutoDock score.

2.3.3 Molecular dynamics simulations

All molecular systems, apo Zika virus NS3 helicase and when complexed with ATP, RNA, ATP-RNA, and repurposed drugs were subjected to 150 ns MD simulations using GROMACS. Rg, RMSD, and RMSF plots were created to evaluate the stability of these molecular systems. The average structure of each molecular system was calculated using the data from 150 ns MD simulations. The structural changes in the protein before and after MD simulations were observed by superposing the initial and average structures of the respective molecular systems. The RMSD

of protein Cα-atoms is an important parameter for analysing the trajectories; these are plotted as a function of time to understand the stability of each molecular system throughout the MD simulations. The RMSD plots, as shown in **Figure 2.4A**, explain the range of deviations of all Cα-atom positions and indicate that the structures have reached stability after 10 ns of MD simulations. All complexes show good stability with low RMSD. The RMSD of RNA bound to Zika virus NS3 helicase reached 3.2 nm, which could be attributed to motor residue movement in domain II. Domain II moves away from the RNA binding groove during MD simulations. Based on these findings, it is proposed that the presence of both ATP and RNA improves protein structural stability.

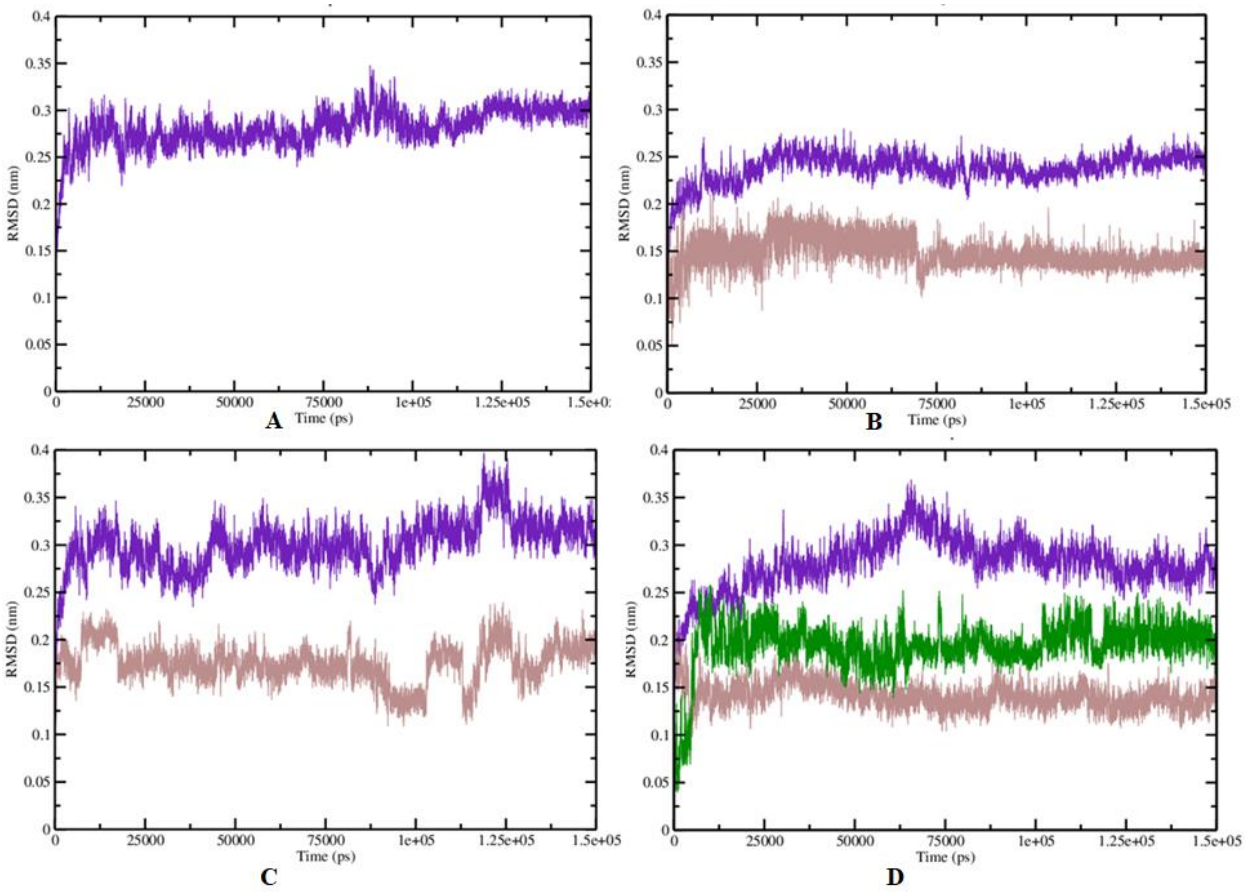


Figure 2.4: RMSD of Zika virus NS3 helicase **A)** apo protein (violet) **B**) NS3 helicase (violet) with ATP (brown) **C**) NS3 helicase (violet) with RNA (brown) and **D**) NS3 helicase (violet) with ATP (brown) and RNA (green).

To gain a better understanding of the structural changes that occur in the Zika virus NS3 helicase upon ligand binding, the RMSF of all molecular systems was calculated. The fluctuations of each Cα-atom in relation to its average position during MD simulations were observed using the RMSF plots of the protein **Figure 2.4B**. When NS3 helicase was bound to ATP, two regions fluctuated significantly: Met244-Ser253, which forms the RNA binding loop (R-loop), and Pro327-Ile333. The Pro327-Ile333 region is close to the ATP binding pocket and fluctuates throughout MD simulations.

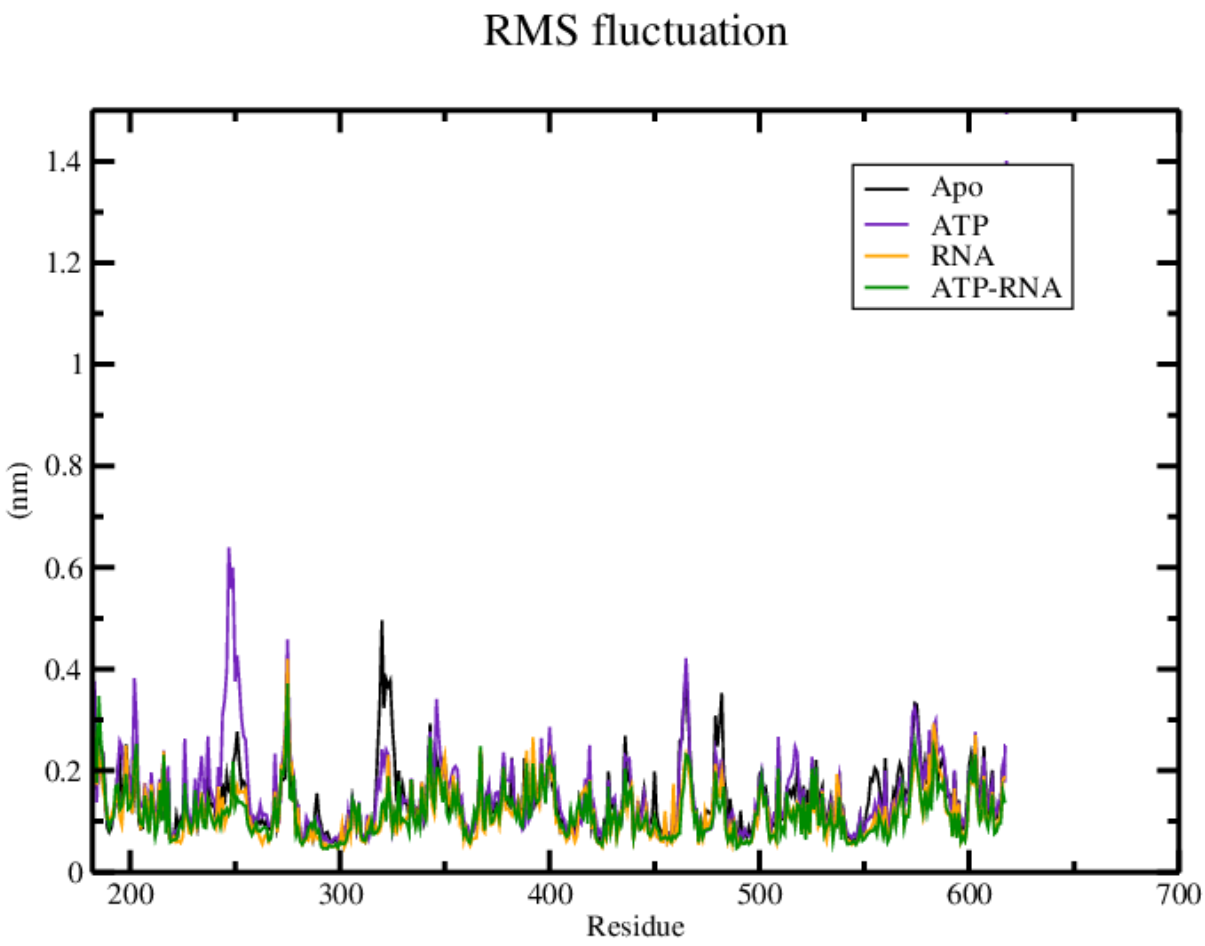


Figure 2.4B: RMSF plots of Zika virus NS3 helicase apo (monomer), and complexed with ATP (dimer), RNA (dimer), ATP-RNA (trimer).

In the Zika virus NS3 helicase when complexed with ATP-RNA, the nucleotide base cytosine; C5 from RNA forms hydrogen bonding interactions with Met244 and Thr245 in the R-loop. In the absence of RNA, this loop displays significant fluctuations as observed from the RMSF plots. When the conformations of these R-loop residues from MD simulations of apo and complexed molecular systems were compared, it was discovered that there was a significant difference in the flexibility of the R-loop in the presence and absence of single-stranded RNA. These findings are consistent with findings from MD simulations of Hepatitis C virus NS3 helicase complexed with ATP-RNA (Pérez-Villa et al., 2015). In the absence of RNA, the Zika virus NS3

helicase bound to ATP had a higher RMSF (0.6 nm) in the R-loop. These loop residues have lower fluctuations (0.25 nm) in the presence of RNA in all molecular systems, including the trimeric complex of Zika virus NS3 helicase bound to ATP-RNA, which has lower fluctuations in the R-loop when compared to the ATP bound protein.

2.3.4 Conformational changes in apo, ATP, RNA, and ATP-RNA bound Zika virus NS3 helicase molecular systems

The triphosphate moiety of ATP is stabilised in the Zika virus NS3 helicase complexed with ATP by interactions with the Walker A motif residues (195-204), which form the phosphate binding loop (P-loop residues); Gly197, Gly199, Lys200, and Arg202, and arginine finger residues (456-463); Arg459 and Arg462. The 3' end of RNA binds to domain I and the 5' end binds to domain II in single-stranded RNA binding. The R-loop residues interact with RNA and exhibit enhanced stability in MD simulations of all molecular systems. The superpositions of the starting and average structures from MD simulations of each molecular system are used to look into the structural changes that happen as a result of simulations time.

The comparison of protein conformational changes across all molecular systems investigated revealed greater changes in the domain I and mobility of the R-loop when bound to ATP. The results revealed that the R-loop underwent distinct conformational changes (open and closed) during the 150 ns MD simulations. The R-loop residues in the Zika virus NS3 helicase - RNA complex have a closed conformation, whereas the R-loop residues in the ATP bound Zika virus NS3 helicase have an open conformation during MD simulations, as shown in **Figure 2.5A**. In domain I of the initial structure of the Zika virus NS3 helicase complexed with ATP, the residues Ala230, Val242 Met244, Ala247, Val248, His252 and Tyr243 form hydrophobic interactions. During MD simulations, the ATP bound protein loses hydrophobic interactions with the surface

of domain I **Figure 2.5B**, resulting in higher fluctuations in the R-loop of the ATP bound molecular system compared to the other systems. As a result, in the Zika virus NS3 helicase-ATP complex, the R-loop residues move away from the RNA binding groove, whereas in the Zika virus NS3 helicase-RNA complex, the R-loop residues move closer to the RNA, such that HG1 of Thr246 interacts with O1P of C5, the residues Met244, Thr245 and Thr246 form hydrogen bonds, and other nonbonding interactions with RNA. These interactions are retained in the average structures obtained after MD simulations and are responsible for stabilizing the R-loop in Zika virus NS3 helicase - RNA complexes. These conformational changes are in correlation with the RMSF plots.

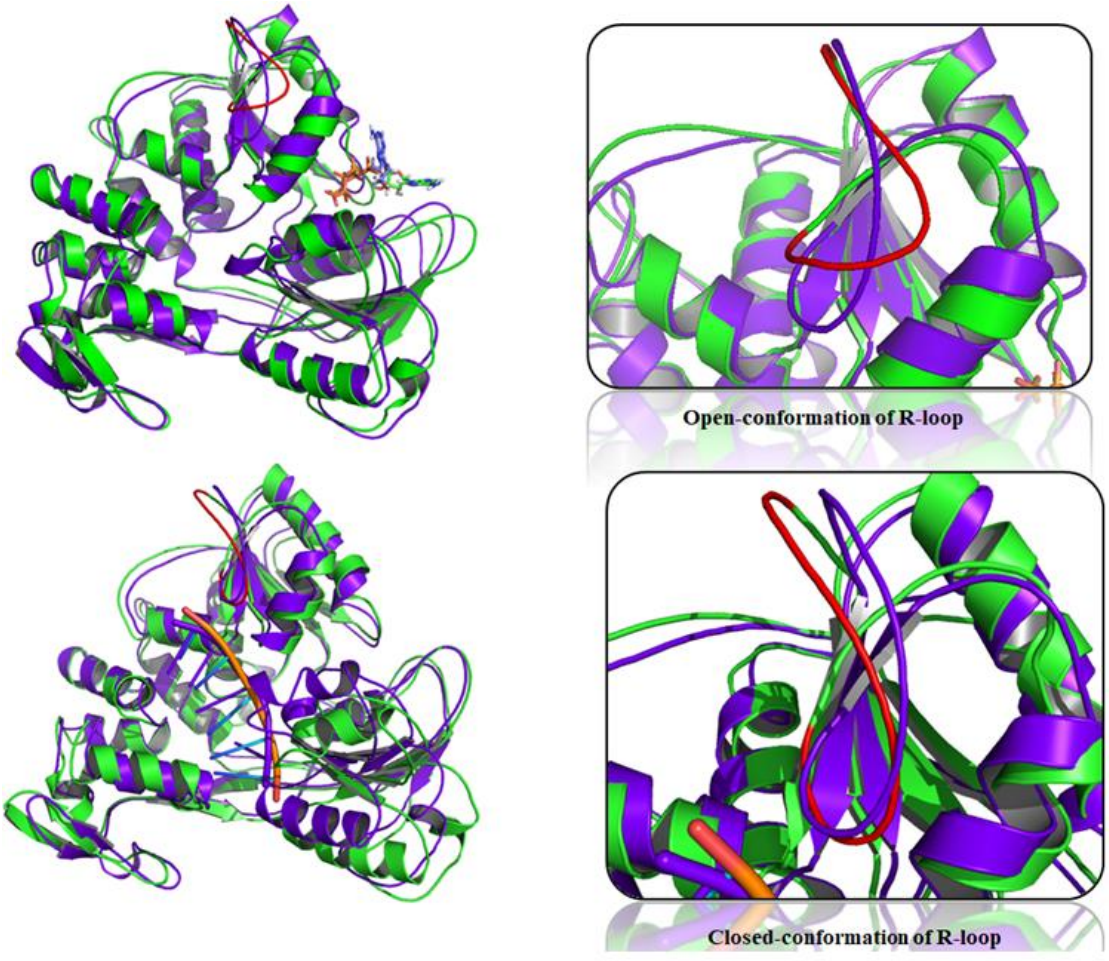


Figure 2.5A: Open and closed conformations of R-loop before (blue) and after (red) MD simulations and Zika virus NS3 helicase complexed with ATP and RNA. Initial (green) and after (blue) MD simulations.

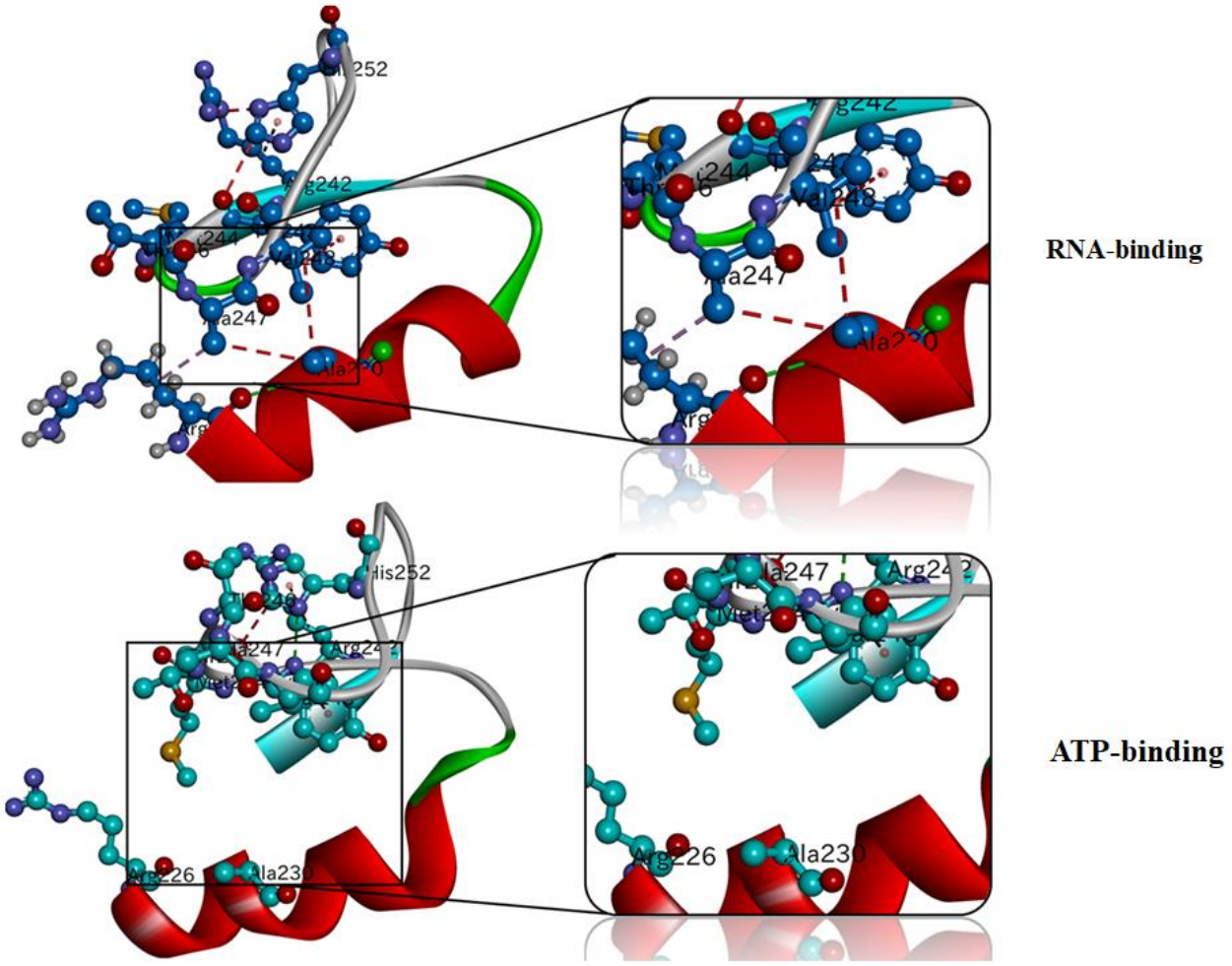


Figure 2.5B: Hydrophobic interactions of R-loop residues after MD simulations in Zika virus NS3 helicase binding with RNA and ATP.

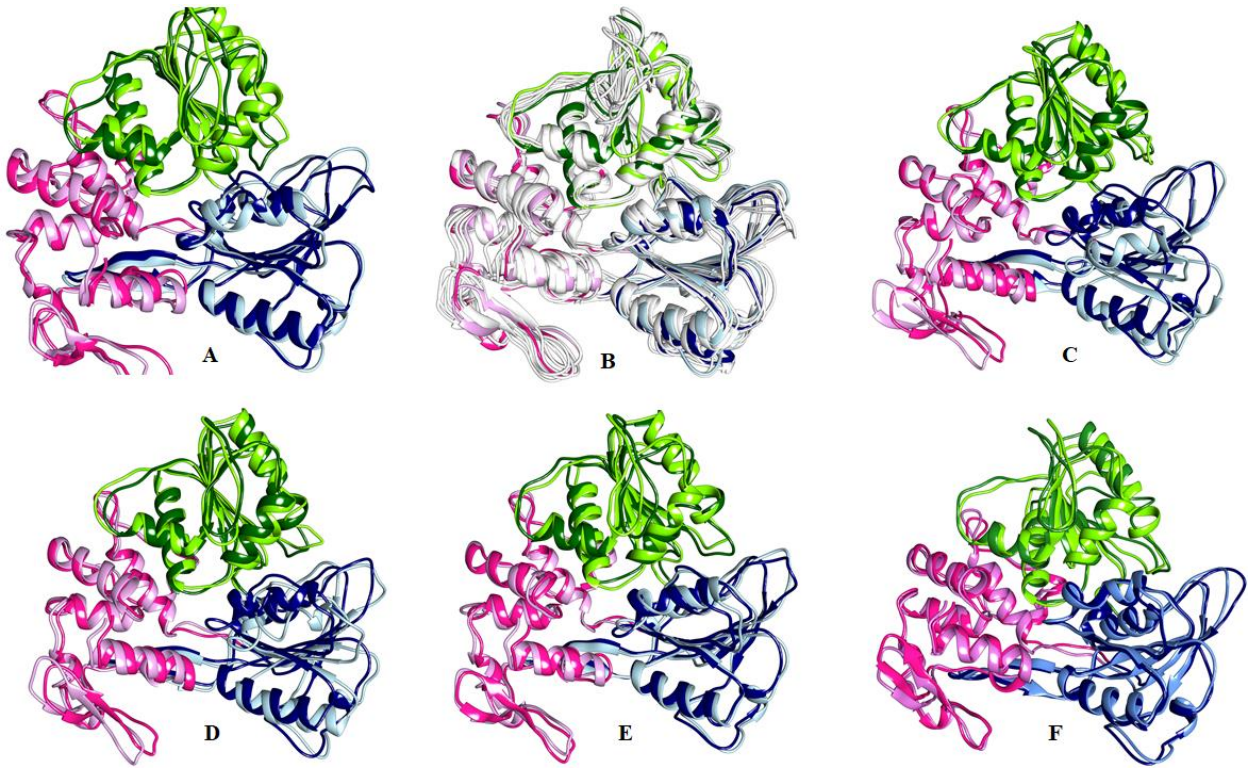
The amino acid residues region of domain I, Pro240-Tyr243, is initially present as a β-strand; however, during MD simulations in the ATP bound system, these residues are converted to a loop structure with an open conformation throughout the MD simulations, which is caused by

the loss of hydrophobic interactions between the R-loop and α-helix 2 (Arg226-Glu234) on the protein surface, which explains the change in the secondary structure of protein during MD simulations. Secondary structural analysis confirms this observation, demonstrating that in the ATP bound molecular system, Pro240-Tyr243 changes conformation from β-sheet to loop structure. From the initial to 150 ns MD simulations of each molecular system, distinct structural deviations were observed. The ATP, RNA, ATP-RNA bound Zika virus NS3 helicase structures exhibit more deviations from its apo form, during MD simulations. The distance between domains I and II in the apo and ATP bound forms is smaller, indicating that the protein is in a closed conformation. The distance between domain I and domain II increased in the Zika virus NS3 helicase when bound to RNA, ATP-RNA and greater movement was observed in RNA binding compared to ATP-RNA, indicating that these molecular systems are in an open state. This is due to the fact that when RNA binds to the protein, domain II moves away from the RNA binding groove. Previous research found that when dengue NS3 helicase was complexed with ATP and single-stranded RNA, it changed its conformation (Davidson et al., 2018). The MD simulations of the hepatitis C virus NS3 helicase revealed significant movement (open and closed conformations) and hydrogen bonding interactions in domains I and II (Pérez-Villa et al., 2015). In the case of the Zika virus NS3 helicase, the residues Arg226, Val227 (domain I) and Glu392 (domain II) exhibit hydrogen bonding interactions in apo and ATP bound forms, and these interactions are maintained during MD simulations. Because of the increased distance between the two domains in RNA and ATP-RNA bound Zika virus NS3 helicase, hydrogen bonding interactions are not observed. Based on these findings, it is now proposed that domains I and II maintain an open conformation in RNA and a closed conformation in ATP-RNA bound molecular systems. RNA binding to the protein results in domain II moving away from the RNA-binding groove, thus increasing the distance

between the two domains I and II, resulting in an open conformation. ATP binding between the domains I and II holds the two domains throughout MD simulations, resulting in a closed conformation.

2.3.5 Clustering analysis

For clustering analysis of open and closed conformations of Zika virus NS3 helicase, clustering analysis tool of GROMACS (gmx cluster) was used to explore the conformational heterogeneity in the ensemble of protein structures generated by computer simulations. The structurally similar clusters were determined using the GROMOS clustering algorithm (Daura et al., 1999) with a Cα-RMSD cut-off. The ATP bound molecular system has eight clusters, while the other molecular systems have two clusters that have distinct structures as shown in **Figure 2.6A**.



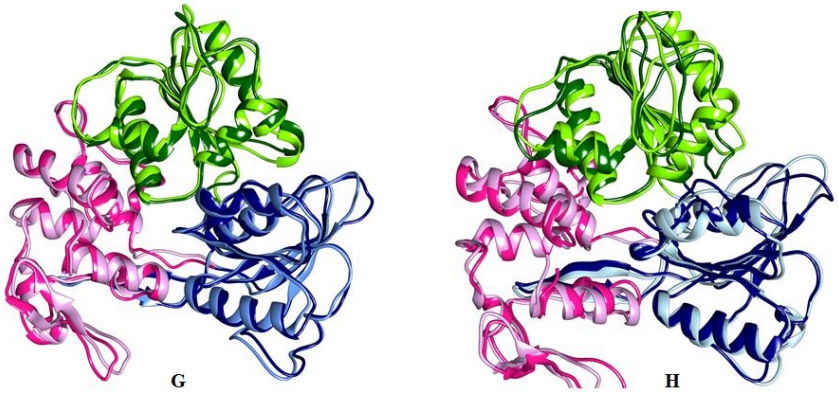


Figure 2.6A: Apo **(A)**, bound to ATP **(B)**, RNA **(C)**, ATP-RNA **(D)**, floxuridine **(E)**, belinostat **(F)**, dicumarol **(G)** and linezolid **(H)**.

The structures obtained from cluster analyses revealed that the protein changes its conformation to open state when RNA is bound and in the other molecular systems its exhibits closed conformation. Whereas in the ATP bound molecular system conformational changes are observed in the R-loop of all eight structures. The distribution of RMSD in each molecular systems were also plotted, that showed the RMSD distribution is greater in ATP bound molecular system (0.28 nm) and the other molecular systems exhibit < 0.25 nm as shown in **Figure 2.6B.**

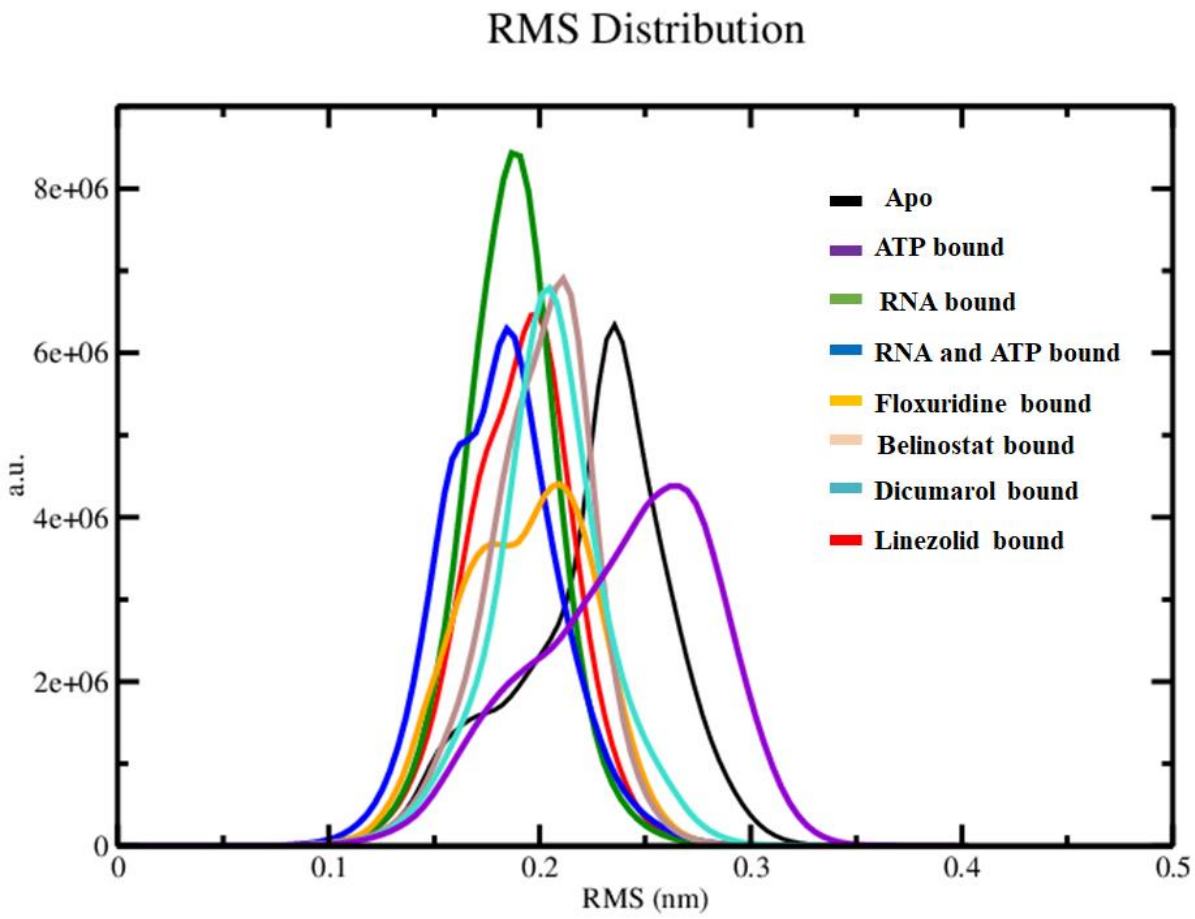


Figure 2.6B: RMS distribution plot of Apo **(A)**, bound to ATP **(B)**, RNA **(C)**, ATP-RNA **(D)**, floxuridine **(E)**, belinostat **(F)**, dicumarol **(G)** and linezolid **(H)**.

Molecular docking on the Zika virus NS3 helicase with various molecules at the ATP binding site were studied (Badshah et al., 2019; Kumar et al., 2020). However, the structural changes of single-stranded RNA when the Zika virus NS3 helicase is complexed with both ATP and RNA remain unknown till now. Significant conformational changes in the RNA bases are observed during MD simulations of both systems (RNA and ATP-RNA). The crystal structure of the Zika virus NS3 helicase complexed with RNA (Tian et al., 2016) shows hydrophobic and hydrogen bonding interactions between the RNA bases and the proteins amino acid residues. The

R-loop forms hydrogen bonding interactions with the C5 on single-stranded RNA. Several MD simulations studies on RNA have focused on conformational changes and torsional angle deviations between crystal and MD simulated structures (Hermann et al., 1998; Zacharias et al., 1999; Zacharias & Martin, 2000; Nifosi et al., 2000; Williams et al., 2000; Koplin et al., 2005). Protein-RNA interactions were studied structurally and biochemically to determine how a protein recognises a specific RNA site, the effect it has on RNA structure, and how their interactions promote a specific function (De Groot et al., 2019). When protein-RNA complex is formed, both the protein and the RNA undergo significant conformational changes (Draper & David, 1995; Flores et al., 2018). In this study, for Zika virus NS3 helicase complexed with RNA and ATP-RNA, significant conformational changes are observed in RNA structure during the course of MD simulations, the nucleoside bases in single-stranded RNA exhibit movements in opposite directions (clockwise and anti-clockwise) with respect to the neighboring nucleoside base as illustrated in **Figure 2.7A**. The nucleoside bases show alterations in structure throughout MD simulations. These alterations were analysed by calculating the torsion angles (α, β, γ, δ, ε, ζ and χ) of each base by using wDSSR web server (Lu et al., 2015). When the Zika virus NS3 helicase is bound to RNA and ATP-RNA, the nucleoside bases in RNA show deviations in torsion angles throughout the MD simulations compared with crystal structure of RNA shown in **Table 2.2**. This deviation in torsion angles provides an explanation for the structural alternations in RNA bases, compared with ATP-RNA molecular system, greater deviations in torsion angles of RNA were observed in RNA bound molecular system, because of the domain II that moves away from the RNA binding groove in RNA-bound system. Whereas for ATP-RNA bound Zika virus NS3 helicase, lower deviations in the torsion angles of RNA are observed because of the closed conformation of domain I and domain II.

Because of these structural alterations of RNA in Zika virus NS3 helicase bound to both RNA and ATP-RNA, the interactions between C5 and R-loop increased as a function of simulations time. The distance between RNA base of C5 and the amino acid residues in the vicinity were measured. These results show that in the initial structures, RNA base C5 close to R-loop makes interactions with Arg226, Thr245 and Thr246 in domain I; however, during MD simulations the RNA base C5 remains close to Thr246 and Thr245, and moves away from Arg226 in Zika virus NS3 helicase complexed with RNA and ATP-RNA. This deviation is more in ATP-RNA bound molecular system compared with RNA bound molecular system **Figure 2.7B**. This indicates that the nucleosides in RNA change their conformation throughout MD simulations and these conformational changes proceed to explain the mechanism of the intermediate state of double-stranded RNA converted into two single-strands, with positive and negative sense of direction.

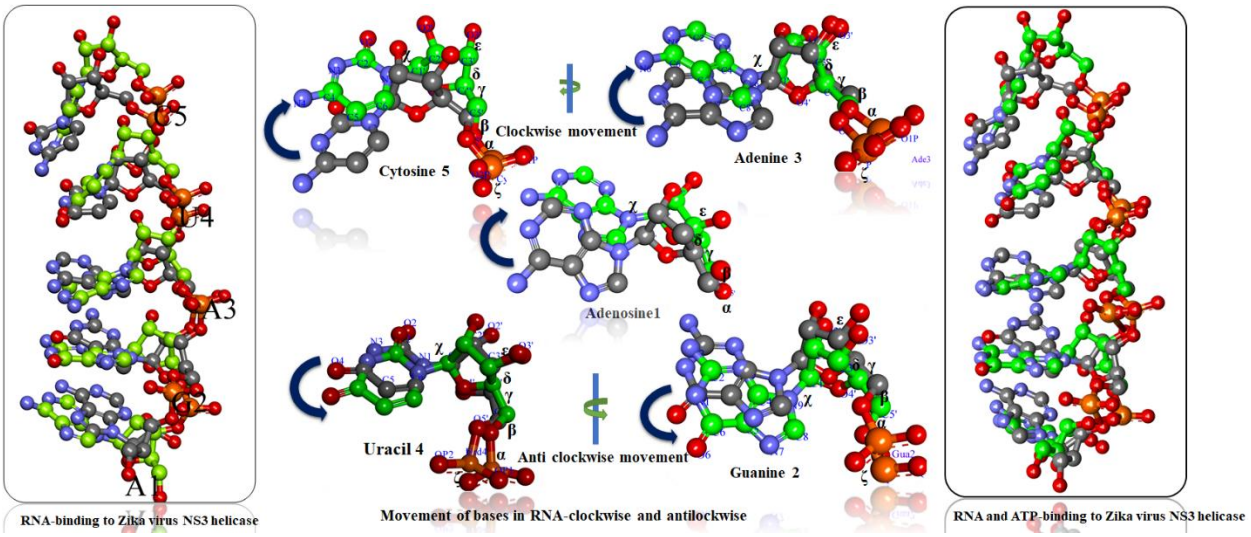


Figure 2.7A: Movements of each base in RNA initial (Grey) and after (Green) MD simulations for Zika virus NS3 helicase bound with RNA and ATP-RNA.

Table 2.2: Torsion angles in RNA from X-ray structure (PDB_id: 5GJB), comparison with MD simulations of RNA bound, and RNA-ATP bound Zika virus NS3 helicase.

Torsion angle	**A1**	**G2**	**A3**	**U4**	**C5**
α-X-ray		-60.697	-65.678	-45.725	-65.443
α-RNA		-61.937	-94.561	-78.09	**-93.16**
α-RNA-ATP		-67.539	-85.072	-49.318	**-100.089**
β-X-ray		163.33	-176.765	154.303	-169.231
β-RNA		167.611	**-154.281**	**176.831**	-168.978
β-RNA-ATP		170.009	-176.233	165.961	-171.833
γ-X-ray	-163.299	60.022	44.539	56.512	51.134
γ-RNA	-174.237	45.709	57.277	67.747	59.781
γ-RNA-ATP	**-135.791**	54.445	85.359	66.846	50.501
δ-X-ray	81.808	81.286	75.828	80.451	79.677
δ-RNA	78.273	82.791	**101.034**	84.173	90.001
δ-RNA-ATP	79.628	81.225	70.387	89.457	83.455
ε-X-ray	-131.079	-153.468	-151.614	-150.520	
ε-RNA	-147.364	**-178.551**	-172.925	-175.998	
ε-RNA-ATP	-134.217	**-176.369**	-143.248	-170.563	
ζ-X-ray	-72.074	-62.401	-99.933	-76.718	
ζ-RNA	-75.576	-75.83	-94.731	-77.229	
ζ-RNA-ATP	-68.757	-77.937	-93.544	-78.777	
χ-X-ray	-177.239	-169.297	-159.474	-145.65	-128.1
χ-RNA	**-156.352**	**-122.174**	**-109.521**	**-95.139**	**-81.279**
χ-RNA-ATP	-177.134	-177.512	-164.265	-133.777	**-140.977**

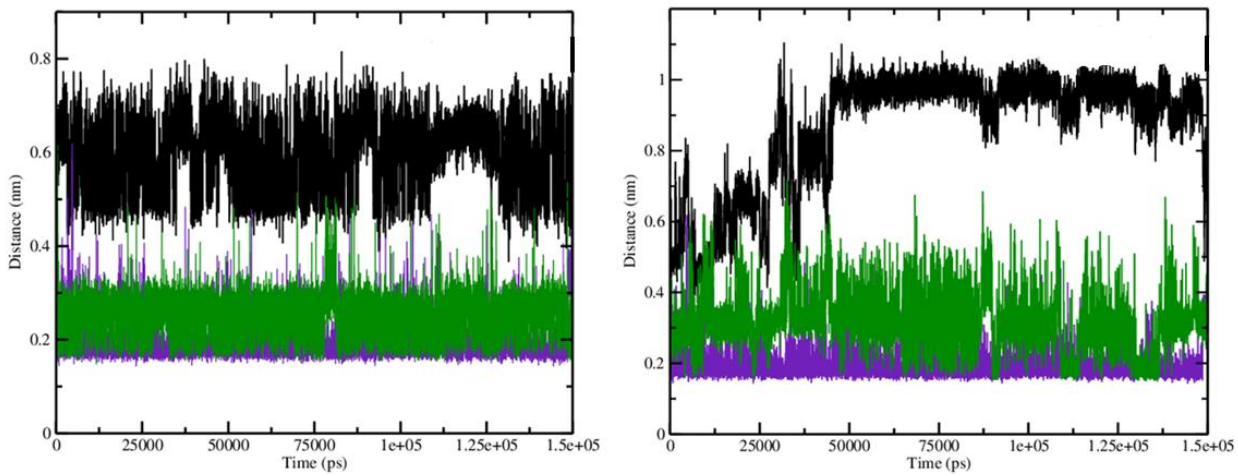


Figure 2.7B

RNA bound Zika virus NS3 helicase:

i)) Distance between NH1 of Arg226 and O1P of Cytosine (black)

ii) Distance between HG1 of Thr245 and O1P of Cytosine (purple)

iii) Distance between HG1- Thr246 and O1P Cytosine (green)

RNA and ATP bound Zika virus NS3 helicase:

i) Distance between NH1 of Arg226 and O1P of Cytosine (black)

ii) Distance between HG1 of Thr245 and O1P of Cytosine (purple)

iii) Distance between HG1- Thr246 and O1P Cytosine (green)

Based on the structural superposition and torsion angles, conformational changes in RNA when complexed to NS3 helicase in both RNA and ATP-RNA bound molecular systems were observed. Greater changes in RNA torsions were observed in RNA bound Zika virus NS3 helicase compared to ATP-RNA bound NS3 helicase. During MD simulations domain II moves away from domain I and exhibits open conformation, resulting in the expansion of the RNA binding groove leading to greater conformational changes in RNA. Whereas in ATP-RNA bound molecular system, ATP is coordinated between domain I and domain II, and the movement of domain II decreased throughout MD simulations, because ATP tightly holds domains I and II which exhibit closed conformation, and there are no changes in RNA binding groove and distance between the two domains is decreased which indicated that the interactions between domain II and RNA are

maintained throughout MD simulations, because of this conformational changes in RNA have decreased when protein is complexed with ATP-RNA. These observations lead to propose a mechanism for the generation of single-stranded RNA from double-stranded RNA of Zika virus. The hydrogen bonding interactions between RNA and protein in RNA bound, ATP-RNA bound molecular systems were monitored. The RNA binding complex exhibits hydrogen bonding interactions with Pro224, Arg226, Thr245, Thr265, Arg388, Thr409, Lys537 that are present throughout MD simulations, whereas the trimeric complex shows along with these interactions some additional hydrogen bonding interactions formed with Cys262, Thr246, Met536 and Asp540 throughout 150 ns MD simulations. The hydrogen bonding interactions between protein and RNA are higher in trimeric complex throughout 150 ns MD simulations. Based on hydrogen bonding interactions and structural alterations of RNA, the unwinding behavior of RNA was observed in both molecular systems. The alterations in the conformation from cofactor and substrate binding, and the mechanism of their recognition in Zika virus NS3 helicase are described in this work.

2.3.6 Molecular dynamics simulations of Zika virus NS3 helicase complexed with FDA approved drugs

Zika virus NS3 helicase complexed with four FDA approved compounds floxuridine, belinostat, dicoumarol and linezolid identified from molecular docking studies were proceeded for 150 ns MD simulations studies. The Cα-backbone RMSD of protein was monitored throughout the 150 ns MD simulations for all molecular systems that attained stability after 10 ns (RMSD < 3 Å) as shown in **Figure 2.8A**. Similarly, the repurposed drugs displayed low RMSD (~ 2 Å) as shown in **Figure 2.8B,** indicating that these drugs stabilize the Zika virus NS3 helicase protein and that a stable complex is formed. The RMSF plots **Figure 2.8C** indicated the stable complexes formed between Zika virus NS3 helicase and FDA approved drugs.

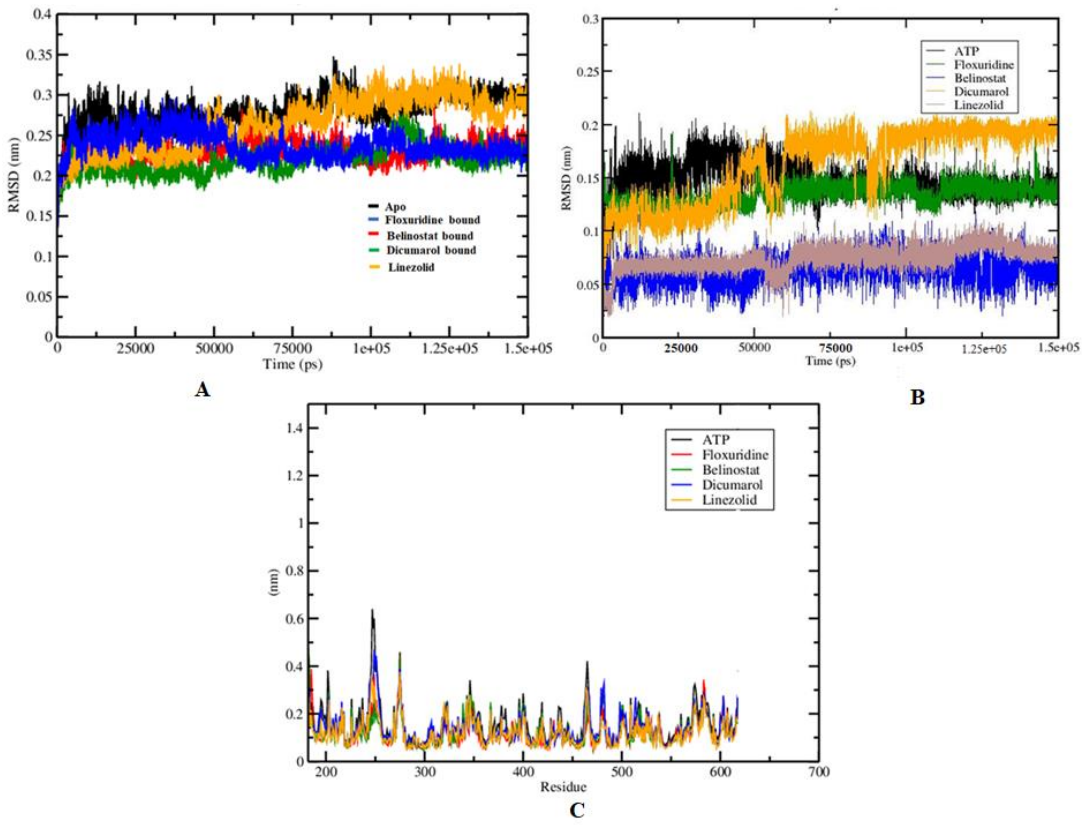
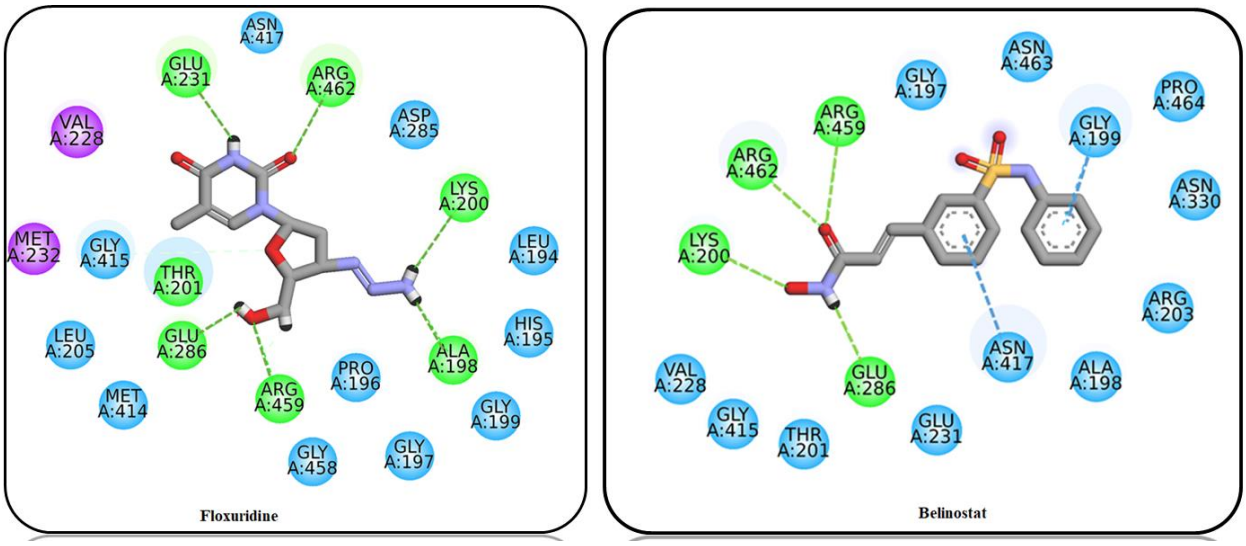


Figure 2.8: A) RMSD plots of Zika virus NS3 helicase complexes with floxuridine, belinostat, dicumarol, and linezolid. **B)** RMSD plots of repurposed drugs floxuridine, belinostat, dicumarol and linezolid. **C)** RMSF plots of Zika virus NS3 helicase complexed with floxuridine, belinostat, dicumarol and linezolid.

The amino acid residues in the P-loop (Ala198-Arg203) form hydrophobic and hydrogen bonding interactions with the drug molecules at the active site and show stability during the MD simulations. The R-loop in domain I that showed high fluctuations when complexed with ATP (~ 6 Å) is stabilized when Zika virus NS3 helicase is complexed with FDA approved drug molecules (< 4 Å). In the domain II, at the Ala325-Val338 region, fluctuations are observed when complexed with floxuridine similar to ATP bound complex, the fluctuations in the region around (Gln272-Tyr279) are observed in all molecular systems. The hydrogen bonding analysis of the four FDA approved drugs was carried out in order to study their hydrogen bonding network at the active site

based on MD simulations. Comparison of the initial and average structures of all these complexes showed that they maintain important interactions throughout MD simulations. **Figure 2.9** represents the 5 Å residue interaction plots of floxuridine, belinostat, dicumarol and linezolid within the active site of Zika virus NS3 helicase. The residues that line active site pocket for the four compounds are Gly197, Ala198, Gly199, Lys200, Thr201, Arg202, Glu288, Gly415, Asn417, Arg459 and Arg462. The residues Gly199, Lys200, Thr201, Arg202, Glu286, Gly415, Asn417 Arg459 and Arg462 formed stable hydrogen bonds with electronegative atoms present in the four compounds that were retained during MD simulations. These results confirmed that the four FDA approved drugs maintained the hydrogen bonding interactions throughout MD simulations. This indicates that these nonbonding interactions stabilize the drug molecules within the active site cavity of Zika virus NS3 helicase.



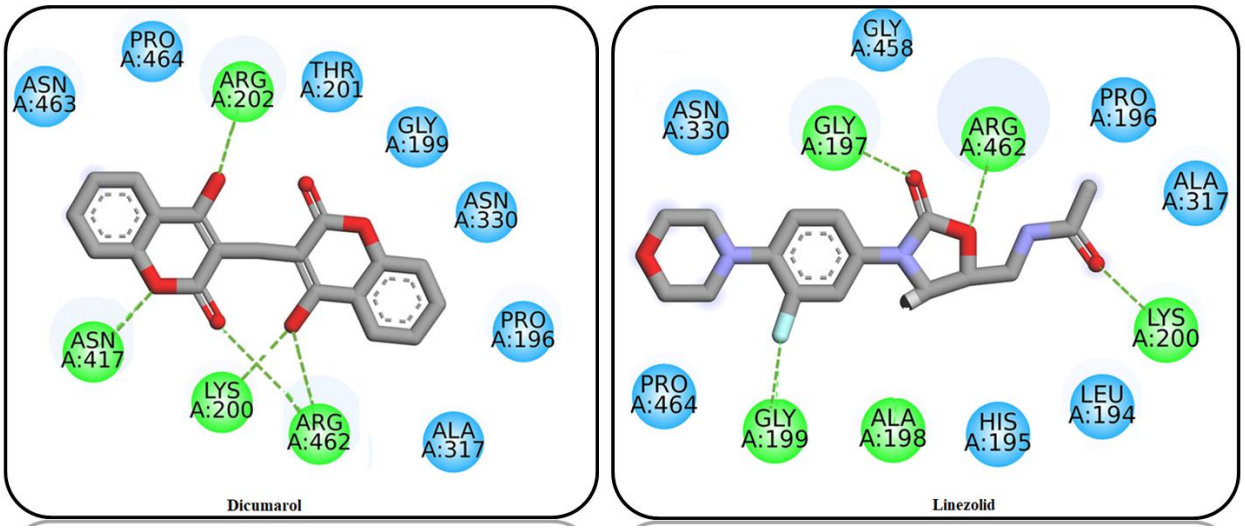


Figure 2.9: Interactions of FDA approved drugs floxuridine, belinostat, dicumarol and linezolid within 5 Å in Zika virus NS3 helicase after MD simulations.

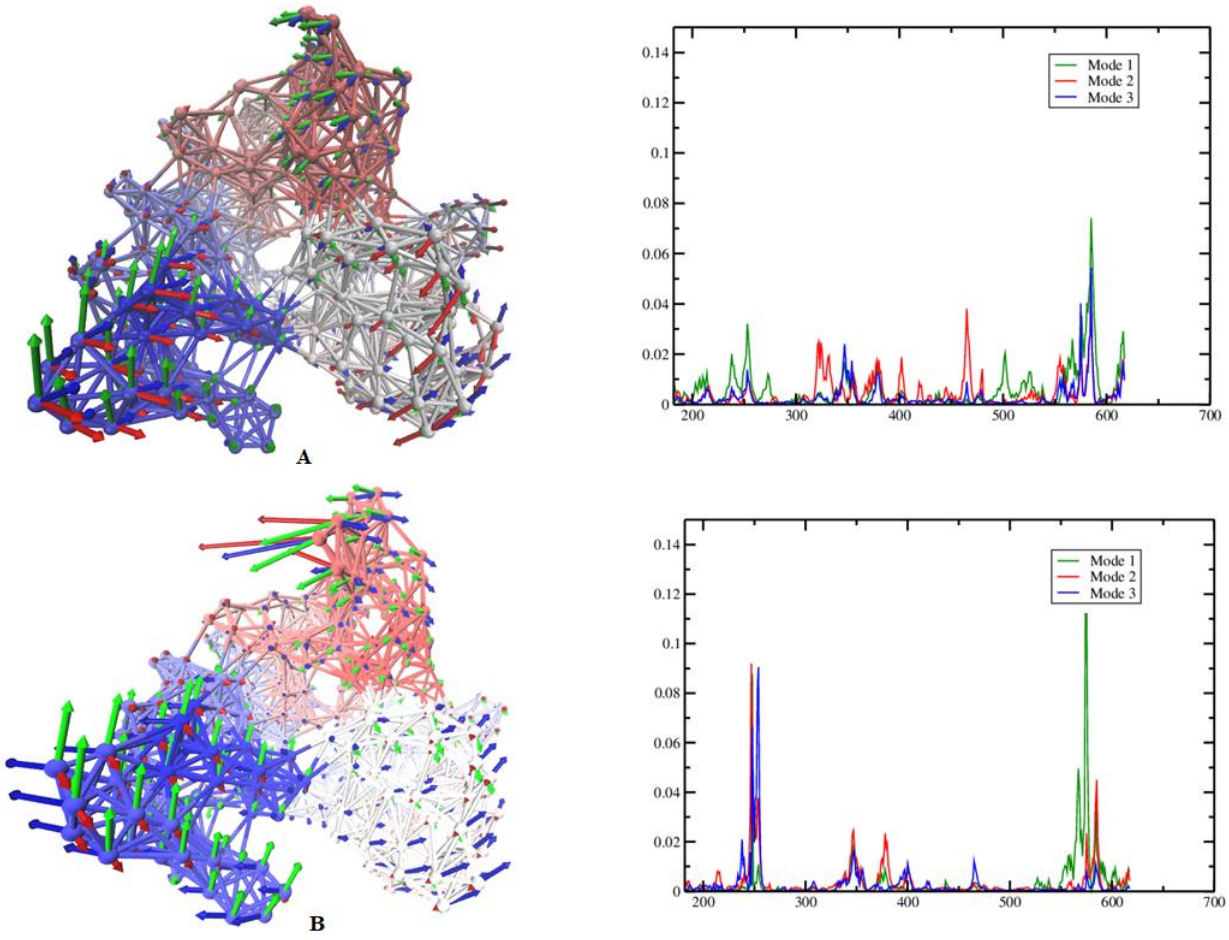
2.3.7 Radius of gyration of Zika virus NS3 helicase binding to ATP, RNA, ATP-RNA and repurposed drugs

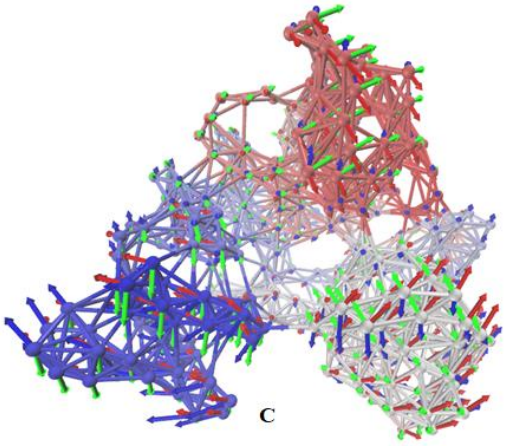
The Rg provides information about the compactness of the molecular systems under study during the course of MD simulations. The Rg of all molecular systems showed that the repurposed drugs have lower Rg as seen during the 150 ns MD simulations. When Zika virus NS3 helicase is bound to RNA the Rg is found to be higher indicating that the compactness of protein decreased because of the influence of RNA binding, and changes in domain II throughout MD simulations. When the protein binds with ATP-RNA, the Rg value slightly decreased compared with RNA bound protein, because of the decreased conformational changes in domain II. The Rg is found to be lowest when the protein is complexed with linezolid and dicumarol indicating the greater stability of these complexes.

2.3.8 Normal mode analysis

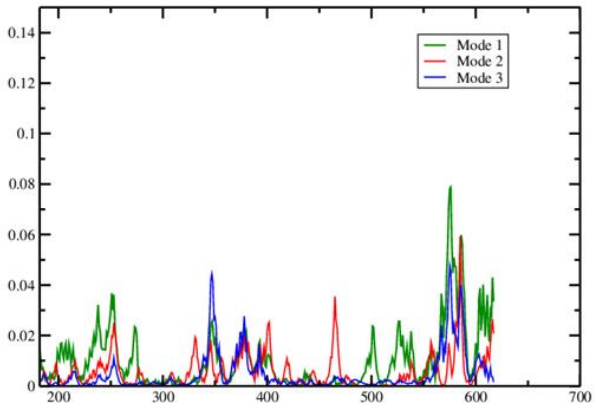
The NMA is a fast and simple method to calculate vibrational modes and protein flexibility (Alexandrov et al., 2005). It reveals the structural variations in the protein obtained from MD simulation studies. The NMA calculations involve atomic fluctuations, two types of atomic fluctuations are calculated by using this analysis; displacement vectors of individual normal modes of atoms and their average of all the normal modes and time. The displacement vectors of atoms for individual normal modes provides significant information on protein dynamics. The individual normal modes display characteristic motions that differ from mode to mode, the descriptions of slow-large amplitude motions are important which are analyzed by NMA. This study helps to explain the conformational changes of apo Zika virus NS3 helicase and its complexes with ATP, RNA, ATP-RNA and the FDA approved drug molecules. By using normal mode wizard which is available in the Prodynamics software, ten normal modes were generated for each system from 150 ns MD simulations trajectories and selected the first three modes for each system and analyzed the structural variations occurring during the MD simulations and compared the movement of each mode with the apo structure of Zika virus NS3 helicase. These NMA mobility plots indicated that the residues Val366-Ala379 and Ser387-Gln401 from two helices present in the motor domain (domain II) exhibit mobility in all molecular systems as shown in **Figure 2.10A-H.** However, this mobility is greater in RNA binding molecular systems compared with others, because of the movement of domain II in the presence of RNA. In Zika virus NS3 helicase bound to ATP (**Figure 2.10B)**, the R-loop region displays three modes with high intensity. The movements of these NMAs have decreased when RNA is bound to Zika virus NS3 helicase and the R-loop residues, as shown in **Figures 2.10C and D**. These results indicate that when the protein is complexed with RNA the conformational changes of R-loop residues have decreased because of decreasing movements of modes of the residues present in the R-loop. The Zika virus NS3 helicase complexed

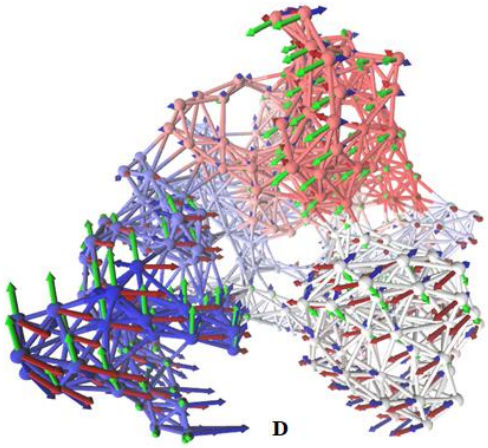
with FDA approved drugs showed lower movements of NMAs compared to the ATP bound molecular systems. In all the systems, movements in the amino acid residues of motor domain (Cys562-Thr582) in domain III were observed. These results indicate that the FDA approved drugs binding to Zika virus NS3 helicase form stable complexes throughout MD simulations.



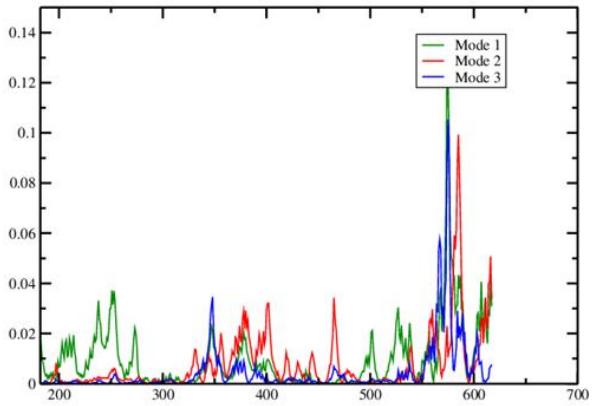


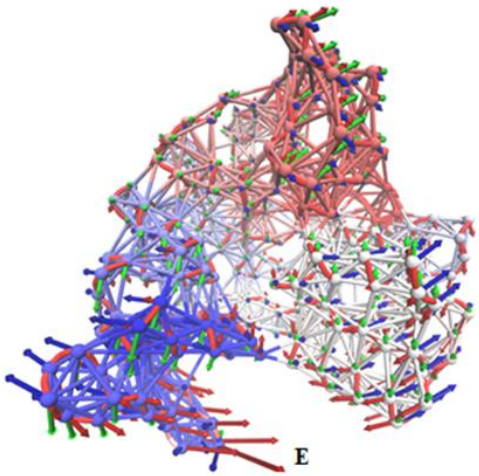
C


Mode 1
Mode 2
Mode 3
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
200
300
400
500
600
700

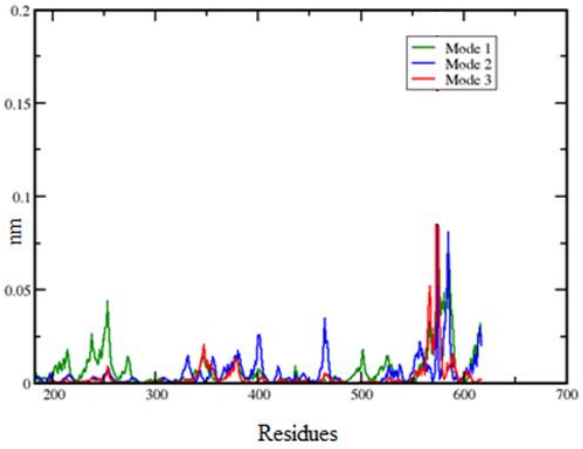


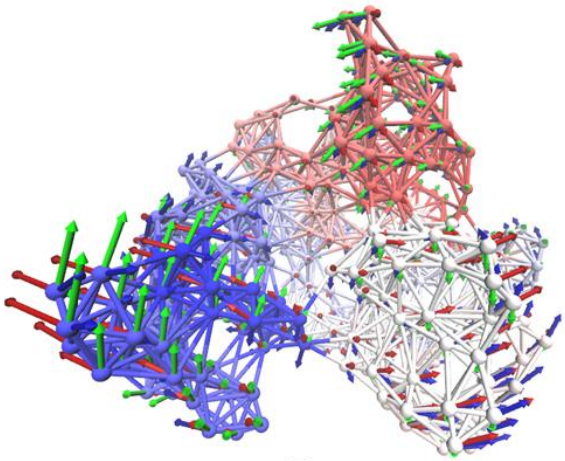
D


Mode 1
Mode 2
Mode 3
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
200
300
400
500
600
700

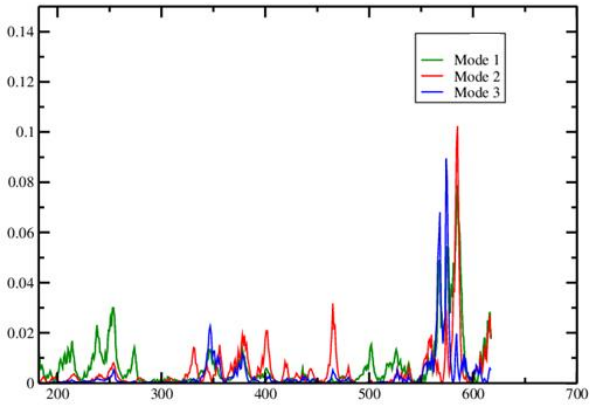


E





F



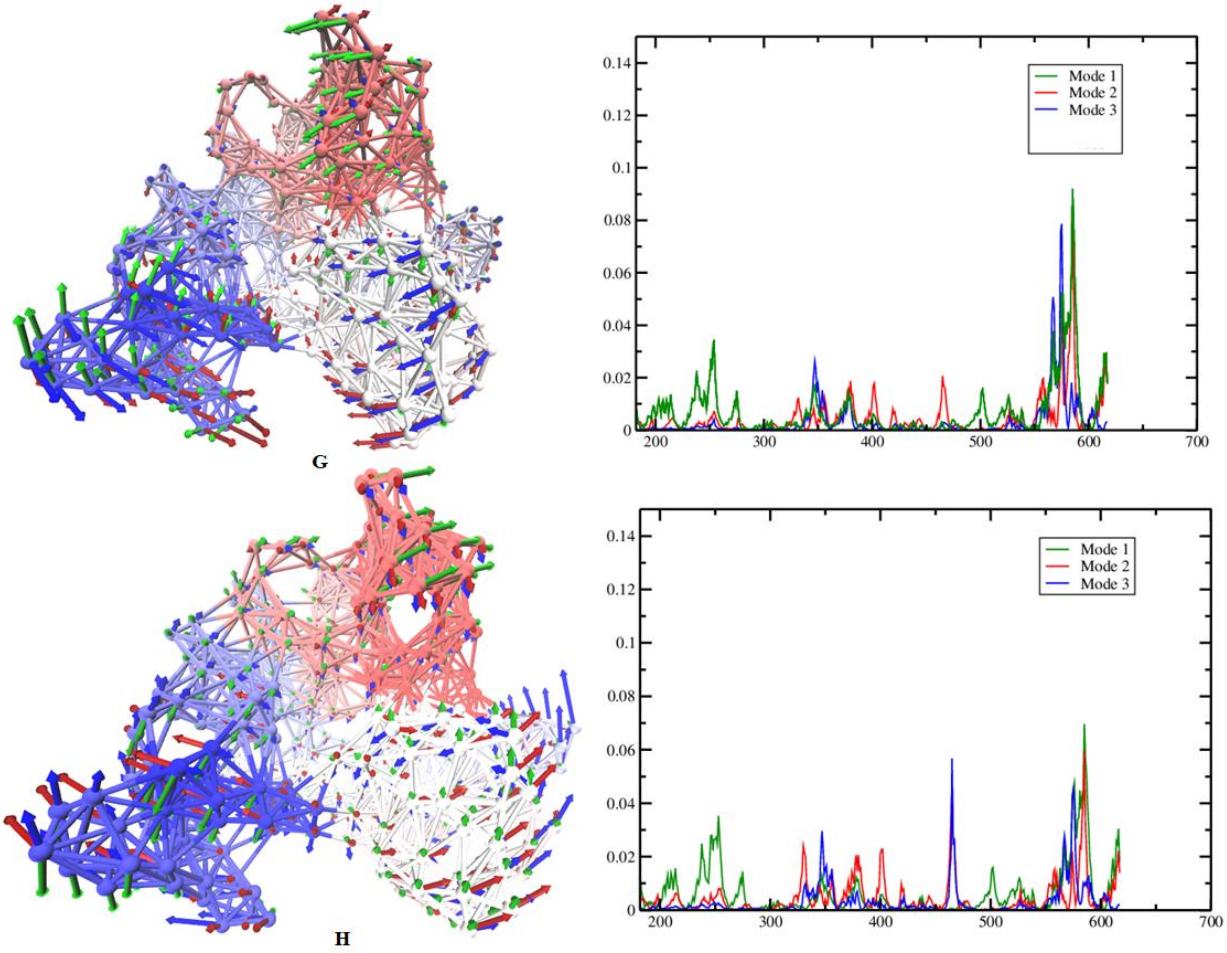


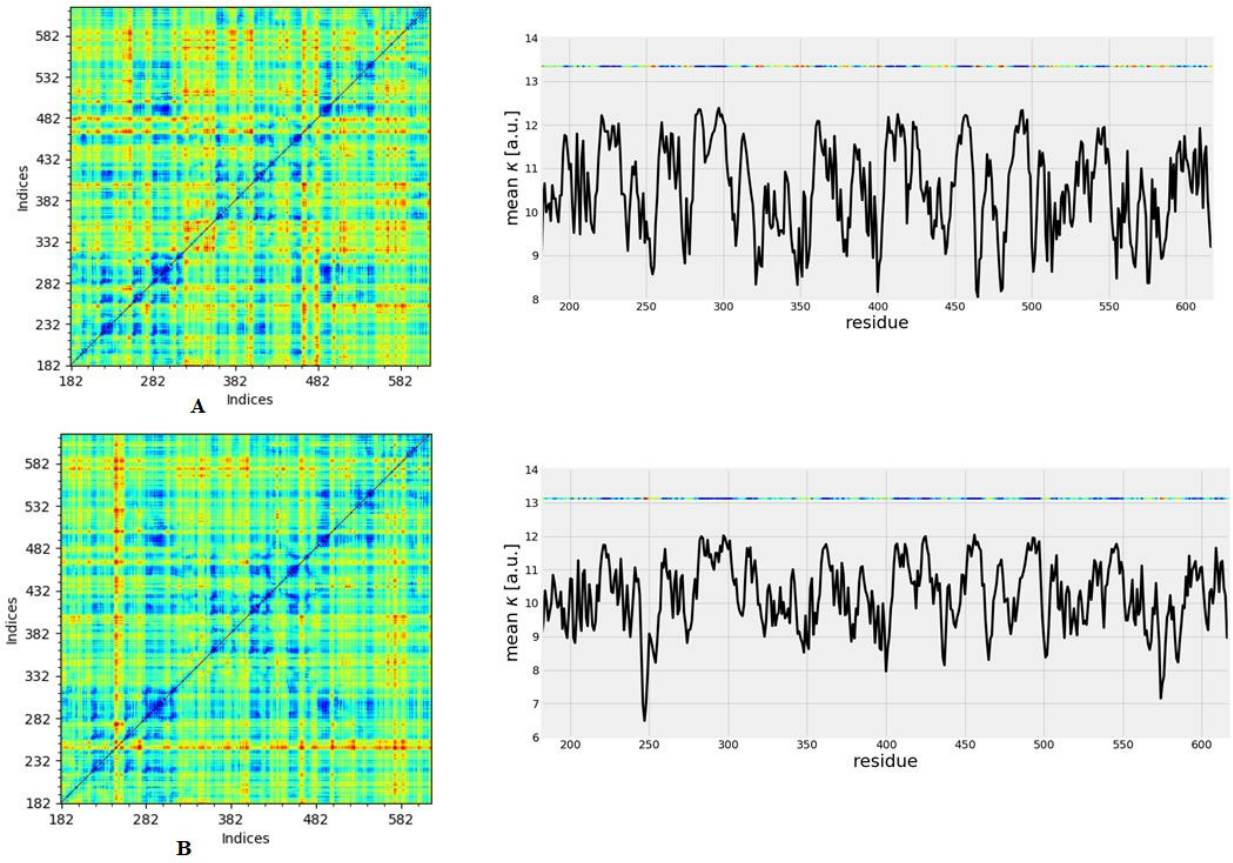
Figure 2.10: Network and mobility plots of Zika virus NS3 helicase apo and complexed with ATP, RNA, ATP-RNA, and FDA approved drugs **(A)** Apo protein, **(B)** ATP bound protein, **(C)** RNA bound protein, **(D)** ATP-RNA bound protein, **(E)** floxuridine, **(F)** belinostat, **(G)** dicumarol, and **(H)** linezolid.

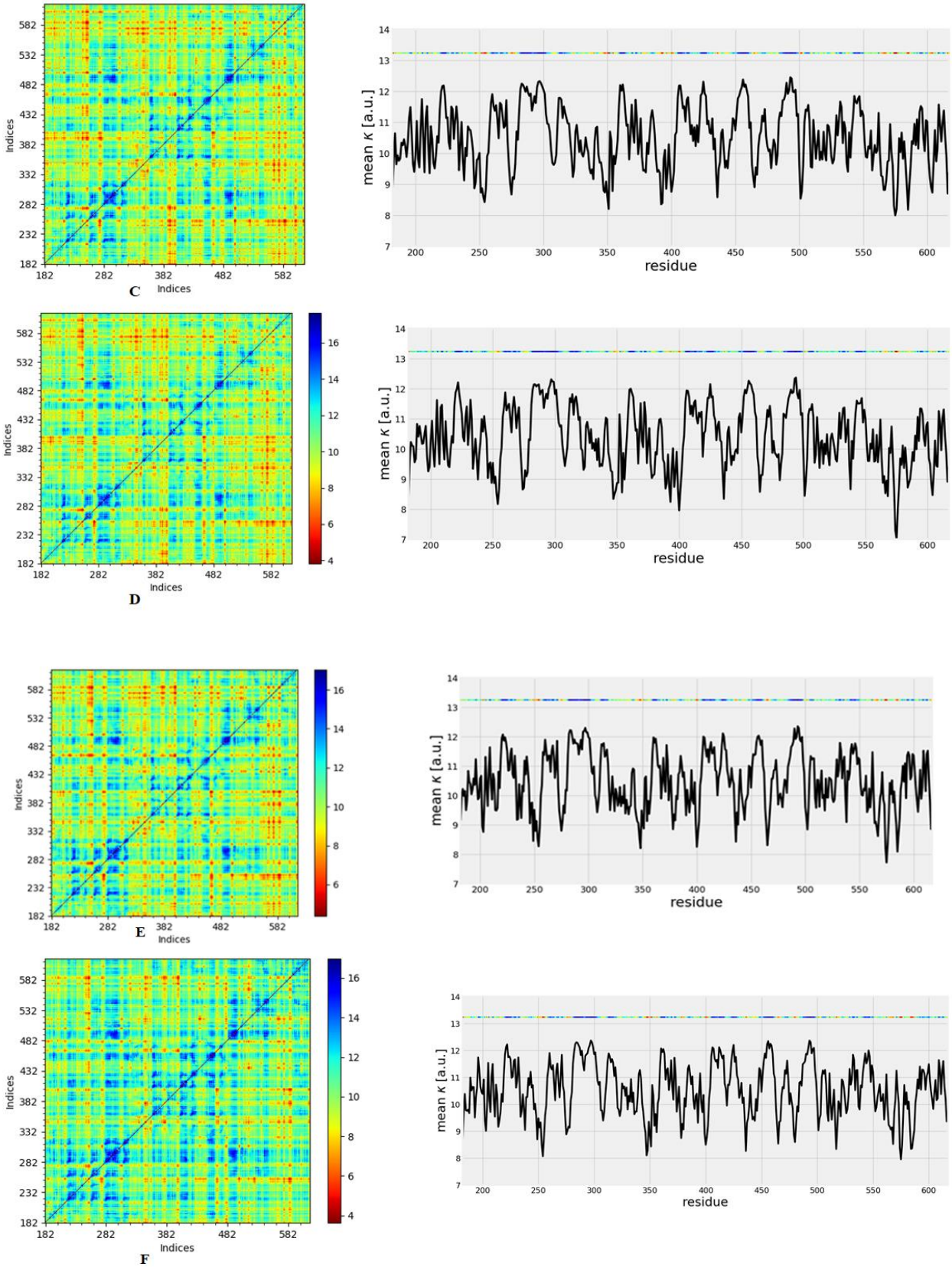
2.3.9 Mechanical stiffness

The mechanical stiffness calculations help to explain the mechanical and elastic nature of a protein, elastic proteins exhibit lower mechanical stiffness (Gosline et al., 2001). These plots are useful to identify the anisotropic response of the structure to external perturbations, determination of weak and strong pairs of interactions depending on the direction of the external force (Eyal &

Bahar, 2008) and the residues that belong to secondary structural elements display relatively strong resistance to alterations because of their inherent stabilization due to hydrogen bonds. The mean plot displays the results averaged over all pairs of residues, which provides an estimate of mechanical resistance of individual residues to structural alterations. Mechanical stiffness was calculated for 150 ns MD simulations structures of apo and complexes of Zika virus NS3 helicase (ATP, RNA, ATP-RNA, and FDA approved drug molecules) in all molecular systems and generated maps with effective force constant. Lower mechanical stiffness is indicative of the weak regions and higher mechanical stiffness is indicative of strong and stable regions during MD simulations. In all the molecular systems when compared with apo Zika virus NS3 helicase, the structural alterations were noted. From these plots it was observed that the R-loop residues exhibit lower effective stiffness which is decreased in ATP bound protein, whereas in RNA bound protein the mechanical stiffness of R-loop remains unchanged compared with the apo form. In the domain II, the residues from Ser387-Gln401 and Val366-Ala379 form α-helices and Arg343-Gly357 are present as loop-helix-loop, all of these regions show greater fluctuations and mechanically weak behavior in the presence of RNA binding. This is caused by the binding of RNA to NS3 helicase domain II that exhibits greater structural changes and mechanically weak behavior throughout MD simulations. In domain III, residues (Cys562-Thr582) also exhibit lower stiffness throughout the MD simulations in all molecular systems **Figure 2.11A-H**. In the mean plots of mechanical stiffness, the effective spring constant value of R-loop is less than 7 k (a.u) in the ATP bound system; whereas in other complexes it is greater than 8 k (a.u). These values indicated that the elastic nature of R-loop residues (Met244-Ser253) is higher in the ATP bound protein compared with other MD simulated structures of Zika virus NS3 helicase. From the results of mechanical stiffness calculations, it is now proposed that Zika virus NS3 helicase interacts with ATP, RNA,

ATP-RNA and FDA approved drugs. Some regions in the protein are mechanically weak or strong, but large conformational deviations and elastic nature is observed in the R-loop of ATP bound protein and motor domain residues (Cys562-Thr582) in domain III among all molecular systems.




Indices
Indices
mean κ [a.u.]
residue
C
D
E
F

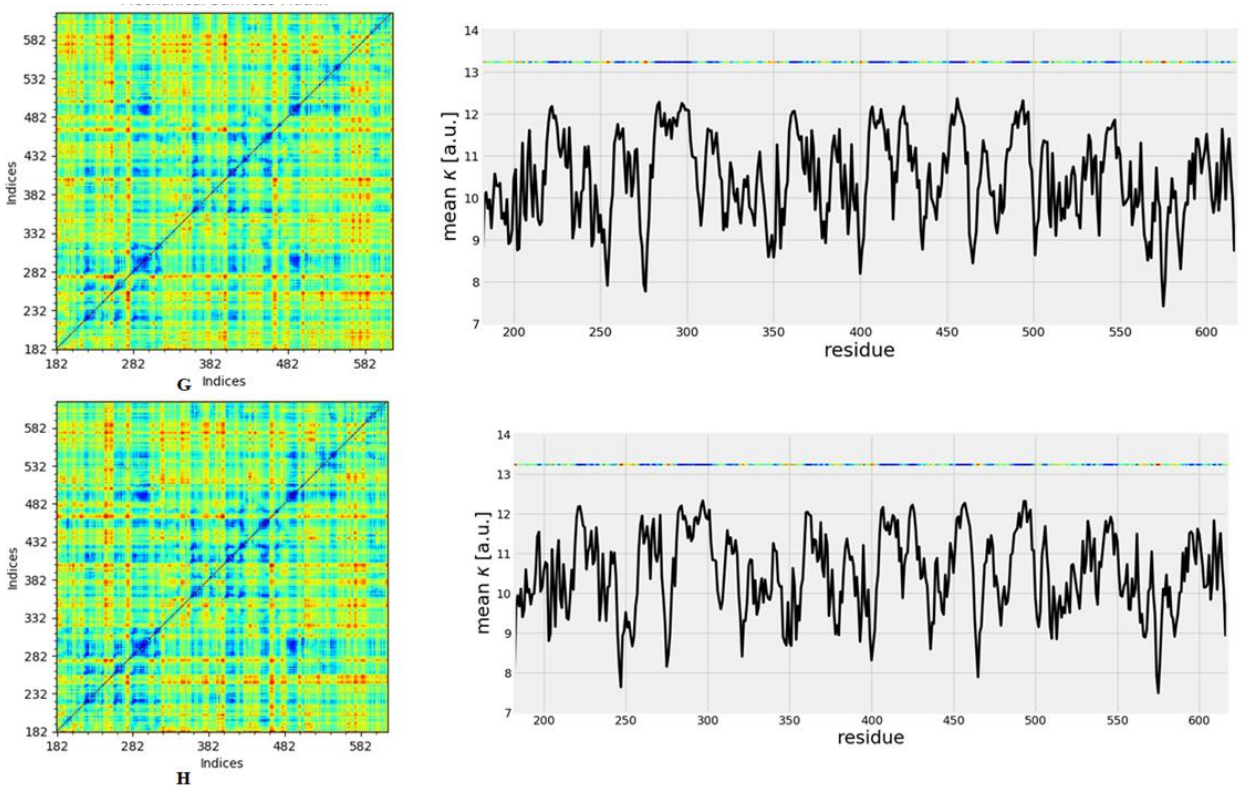
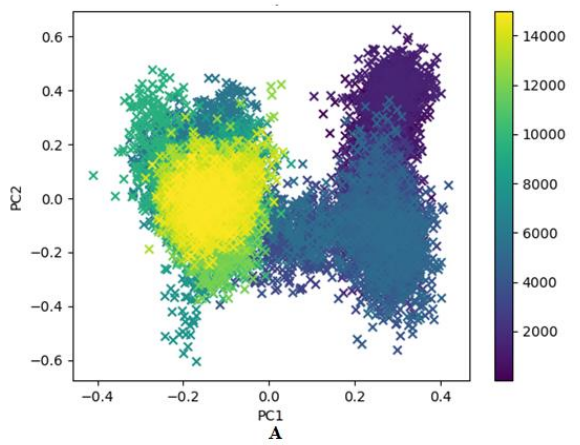


Figure 2.11: Mechanical stiffness and mean plots of Zika virus NS3 helicase Apo **(A)**, bound to ATP **(B)**, RNA **(C)**, ATP-RNA **(D)**, floxuridine **(E)**, belinostat **(F)**, dicumarol **(G)** and linezolid **(H)**.

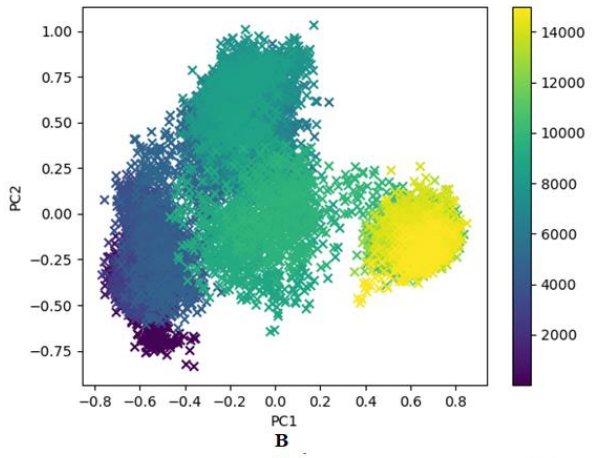
2.3.10 Principal component analysis

PCA was performed on 150 ns MD simulations trajectories to understand the alterations in the R-loop (Met244-Ser256) residues for open and closed conformations of Zika virus NS3 helicase. PCA is commonly employed on ensembles of protein structures obtained from MD simulations to reduce dimensionality by clustering the structures into different conformations and identifying conformational transitions. The resulting clusters allow for comparisons to be made between conformational changes that are generated during the MD simulations. For each molecular system 15,000 frames were generated and the motion of Cα-atoms in each molecular

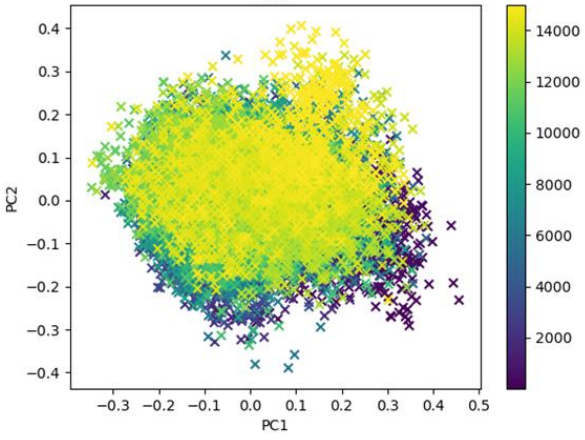
system of Zika virus NS3 helicase was monitored. It was observed that the first two principal components (PC1 and PC2) capture the majority of the variance in the original distribution of conformational ensembles in the molecular systems. The distribution of Cα-atoms in the ATP bound molecular system is greater than the other molecular systems. The conformational ensembles of the R-loop residues (Met244-Ser256) in all molecular systems under study were analyzed by projecting the trajectories of PC1 and PC2 into a two-dimensional (2-D) space. When these are mapped onto each other, the structures with a high degree of similarity cluster together. Therefore, each cluster represents a different R-loop conformational state in the protein. The observed conformational changes of the R-loop in apo, ATP bound, RNA bound, ATP-RNA bound, and repurposed drugs bound molecular systems were monitored and shown in **Figure 2.12**. Projecting the trajectory snapshots onto the plane formed by the PC1 and PC2 reveals a semicircle relationship; such a pattern probably indicates random diffusion during the simulations and is interpretable as motion along a shallow free-energy landscape. Among all the systems studied, the PCs show high-frequency motion in ATP bound Zika virus NS3 helicase as can be seen from the PCA scatter plots **Figure 2.12B**, indicating greater conformational changes of R-loop in the ATP bound molecular system. This demonstrated that the conformational distributions of R-loop residues in Zika virus NS3 helicase bound with ATP were remarkably different from other molecular systems. The frequencies of PCA scatter plots were quantified, and the highest frequency is observed in ATP bound Zika virus NS3 helicase and also in the linezolid bound system. These results indicated that Zika virus NS3 helicase bound with ATP and linezolid display more significant R-loop conformational changes compared to other molecular systems.


PC2
PC1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.4
−0.2
0.0
0.2
0.4
14000
12000
10000
8000
6000
4000
2000

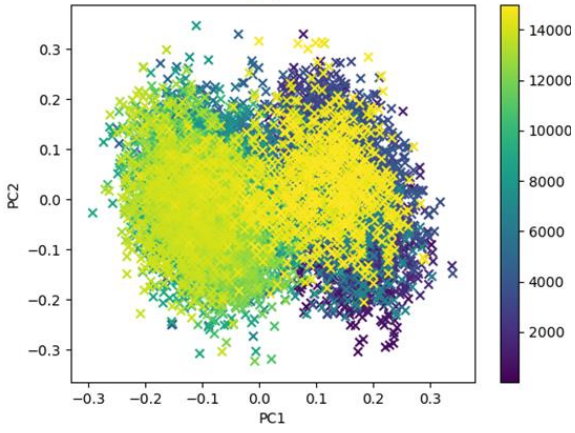
A


PC2
PC1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
14000
12000
10000
8000
6000
4000
2000

B


PC2
PC1
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
0.5
14000
12000
10000
8000
6000
4000
2000

C


PC2
PC1
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
14000
12000
10000
8000
6000
4000
2000

D

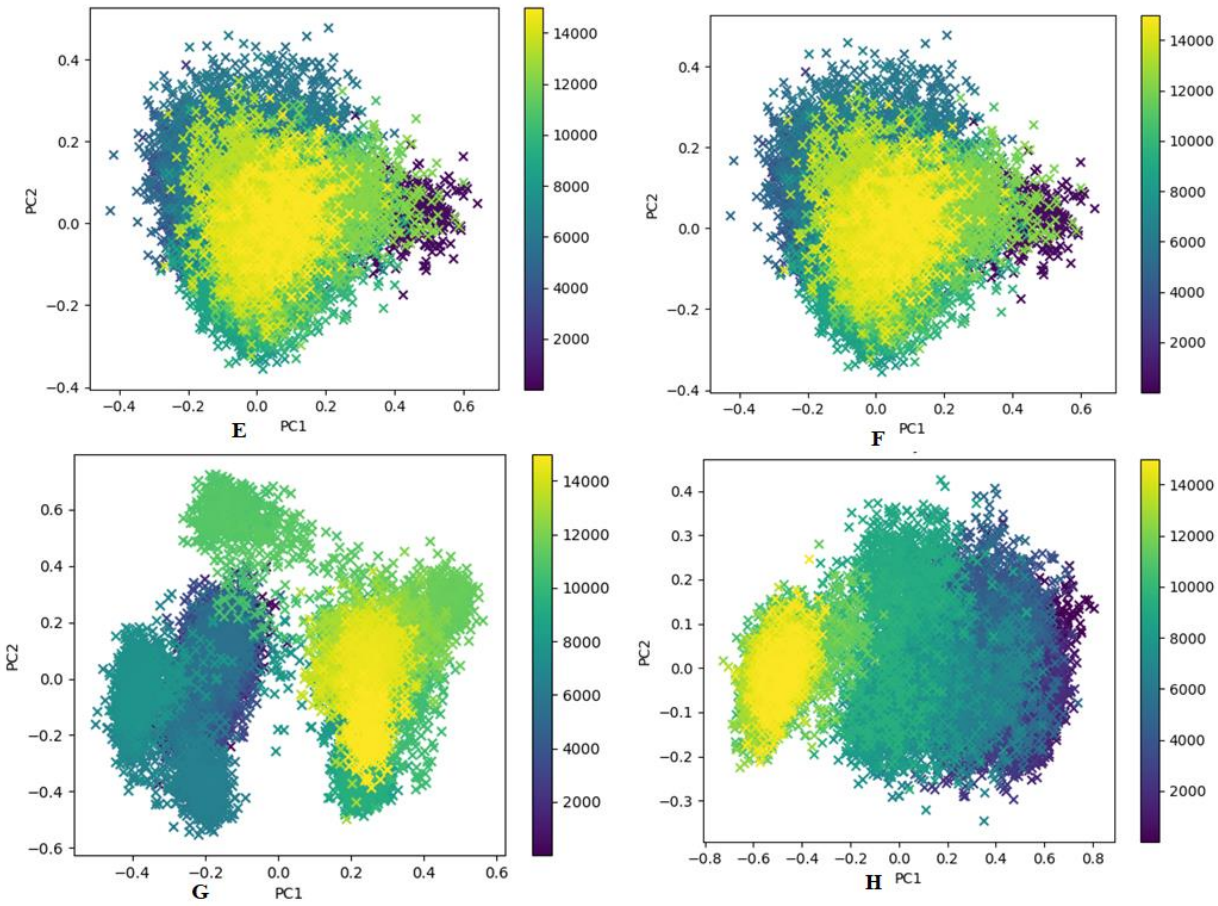


Figure 2.12: PCA Scatter plots of Zika virus NS3 helicase Apo **(A)** bound to ATP **(B)** RNA **(C)** ATP-RNA **(D)** floxuridine **(E)** belinostat **(F)** dicumarol **(G)** and linezolid **(H)**.

2.3.11 Binding free energies

The binding free energies of the Zika virus NS3 helicase binding to ATP and FDA approved drug molecules was calculated using MM-PBSA, which has been proved to be an important complement to the evaluation of protein-ligand interactions (Chen et al., 2018; Sun et al., 2014; Weng et al., 2019). The predicted binding free energies for Zika virus NS3 helicase bound to ATP, dicumarol, linezolid, floxuridine and belinostat are summarized in **Table 2.3**. The binding free energies for Zika virus NS3 helicase bound with ATP (-186.01± 0.00 kJ/mol), dicumarol (−95.007 ± 0.52 kJ/mol), floxuridine (−92.91 ± 0.62 kJ/mol), linezolid (−85.25 ± 0.99 kJ/mol) and belinostat (−68.03 ± 0.82 kJ/mol) are observed. The FDA approved drugs show

reasonably good binding affinity with Zika virus NS3 helicase. The contribution from electrostatic energy in ATP binding is higher than the FDA approved drugs. The binding free energy contribution from active site amino acids in Zika virus NS3 helicase is measured to explain the importance of active site residues in binding the ligand. The amino acid residues contributing to the binding free energies of the complexes is shown in the **Table 2.4** and **Figure 2.13.** The negative values suggested favorable nonbonding interactions with residues in the ligand binding pocket of Zika virus NS3 helicase. The residues Lys200, Asn417 and Arg462, show the highest binding free energy contribution, which may be attributed to the hydrophobic and hydrogen bonding interactions when bound to ATP and FDA approved drugs. These studies indicate that the FDA approved drugs may have good inhibitory activity against Zika virus NS3 helicase.

Table 2.3: Binding free energies of Zika virus NS3 helicase binding to ATP and FDA drug molecules.

Compound name	**van der Waals energy (kJ/mol)**	**Electrostatic energy (kJ/mol)**	**Polar solvation (kJ/mol)**	**SASA (kJ/mol)**	**∆G (kJ/mol)**
ATP	-188.611+/-0.00	-88.425 +/- 0.00	107.156 +/- 0.00	-17.136+/-0.00	-186.016+/- 0.00
Dicumarol	-155.663+/-0.35	-6.390 +/- 0.17	83.511 +/- 0.55	-16.472+/- 0.03	-95.007 +/- 0.52
Floxuridine	-142.676+/-0.55	-44.149 +/- 0.43	107.753+/- 0.86	-13.805+/- 0.05	-92.910 +/- 0.63
Linezolid	-148.542+/-0.52	-8.287 +/- 0.33	87.110 +/- 1.34	-15.536+/- 0.04	-85.250 +/- 0.99
Belinostat	-138.846+/-0.56	-19.299 +/- 0.69	104.173 +/- 1.15	-14.040+/- 0.04	-68.036 +/- 0.82

Table 2.4: Residue-wise contribution energy (kJ/mol) of Zika virus NS3 helicase binding with ATP and FDA approved molecules.

Residues	**ATP**	**Dicumarol**	**Floxuridine**	**Linezolid**	**Belinostat**
Gly199	-12.02	-9.11	-8.33	-12.17	-9.82
Lys200	-58.50	-25.77	-22.41	-23.68	-18.77
Thr201	-6.28	-5.27	-2.31	-12.13	-4.19
Arg202	-66.20	-10.50	-7.62	-9.93	-9.78
Glu286	-10.21	-2.63	-1.24	-0.27	-10.64
Ala317	-2.04	-4.14	-2.53	-8.74	-6.03
Asn417	-16.43	-20.63	-51.38	-20.00	-19.39
Arg428	-16.58	-2.36	-4.63	-2.01	-1.32
Arg458	-13.33	-10.21	-4.51	-3.66	-4.19
Arg459	-28.33	-5.17	-6.27	-7.54	-5.59
Arg462	-71.42	-45.94	-33.14	-20.36	-37.07

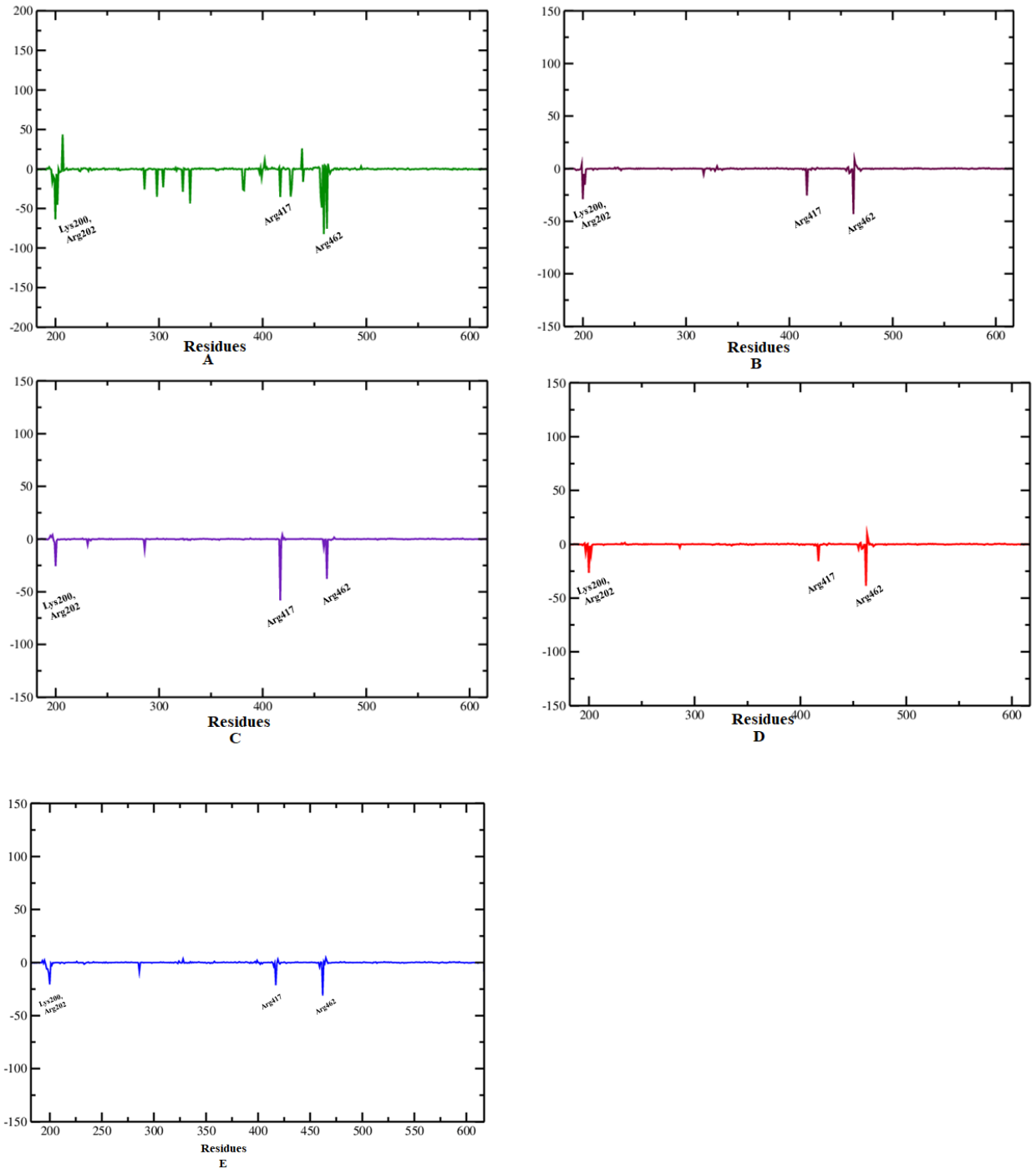


Figure 2.13: Contribution from amino acid residues in the Zika virus NS3 helicase to the binding of ATP **(A)**, dicumarol **(B)**, floxuridine **(C)**, linezolid **(D)** and belinostat **(E)**.

2.4 Conclusions

The MD simulations studies on apo Zika virus NS3 helicase and when complexed with ATP, RNA, ATP-RNA and FDA approved drugs was carried out. Conformational changes were observed in the protein among various molecular systems and significant conformational changes in single-stranded RNA. From the RMSD and RMSF plots, NMA, mechanical stiffness, PCA and cluster analysis, it was observed that the RNA binding loop in Zika virus NS3 helicase shows stable behavior when it binds with RNA and only minor deviation in its conformation in apo form and when complexed with FDA approved drugs. While the Zika virus NS3 helicase bound to ATP shows relatively higher fluctuations in the R-loop (Met244-Ser253) residues which leads to their increased mobility. This R-loop region exhibits mechanically weak behavior and is therefore more elastic in the presence of ATP. From the PCA, it was observed that the motion of R-loop residues is higher in ATP binding protein. The conformational changes in the single-stranded RNA nucleoside bases suggest the replicative mechanism of the conversion of double-stranded RNA to single-strand RNA molecules. The docking and MD simulations studies were applied to understand the stability of repurposed drugs within the ATP binding site of Zika virus NS3 helicase, the four FDA approved drugs floxuridine, belinostat, dicumarol and linezolid show significant stability and good binding affinity within the active site of protein and make hydrogen bonding interactions with protein amino acid residues which indicates that these drugs may also possess anti-viral inhibition for Zika virus NS3 helicase. The repurposing studies are helpful to develop novel anti-viral drugs against the Zika disease.

CHAPTER-3

Mutational analyses, pharmacophore-based inhibitor design and *in silico* validation for Zika virus NS3 helicase

3.1 Introduction

Sequence and structure analyses of NS3 helicases from flaviviruses were studied to understand the similarities and differences in this present study. The sequences of Zika virus NS3 helicases were analyzed for substitution mutations across the evolution of Zika virus. The availability of a proteins 3-D structures has aided in the rational design and discovery of pharmaceuticals (Lounnas et al., 2013), and CADD holds great promise for the development of novel inhibitors. Virtual screening is one of these strategies that is quick and reliable for finding novel ligands based on biological structures. It uses high-throughput screening techniques so that virtual libraries comprising millions of structures could be searched quickly and affordably (Walters et al., 1998). Virtual screening can be done using either ligand-based or structure-based methods, which are both widely acknowledged methodologies. In ligand-based virtual screening, vast databases of compounds are compared to known active molecules using 2-D or 3-D similarity searches. A variety of modelling tools are used in the structure-based virtual screening to simulate the binding interactions between a ligand and its biomolecular receptor (Merz et al., 2010). The modelling based on pharmacophore and QSAR generation, molecular docking are examples of virtual screening techniques used to find novel molecules with certain properties that reflect the bioactivity of the designed molecules (Cherkasov et al., 2014). The SBDD performs processes such as target receptor selection, its 3-D structure elucidation, development of compound libraries of small molecules, *in silico* screening of compound libraries, molecular docking, analyses of the results, and prioritizing the results using a standard scoring system. By using these techniques, the binding mechanism is clarified, the SAR is revealed, and the best compounds are chosen in comparison to those that have already been described (Irwin, 2008; Rella et al., 2006). Millions of

small compounds are represented in chemical libraries and the goal is to choose a small number of molecules with the highest anticipated inhibitory action. In this case, using the pharmacophore-based technique, potent inhibitors are used to generate a pharmacophore model for *in silico* screening of virtual libraries of chemical compounds. The combination of all electronic and steric properties which are important for binding and to make nonbonding interactions with a definite receptor in order to stimulate or prevent the biological response is defined as a "pharmacophore" (Wermuth et al., 1998). Additionally, there are two types of pharmacophore modelling: structure-based and ligand-based pharmacophore models. In the structure-based pharmacophore models, residues present in the binding pocket which are crucial to protein-ligand interactions are constructed. A pharmacophore model is created using a sequence of molecules with a SAR and a range of different inhibitory actions in the ligand-based pharmacophore model (Kandakatla & Ramakrishnan, 2014). The created pharmacophore models can be used as queries in a virtual screening protocol to look for new hit compounds that have the necessary pharmacophore features. These models were created using ligand-based and structure-based approaches. Molecular docking is used to validate the virtual screening of hit chemical molecules as potential inhibitors to a specific target receptor (Yang, 2010). The identification of new hit molecules from searching the big data is being carried out using the protocols in machine learning combined with virtual screening methodologies (Lavecchia, 2015). Such techniques are extremely advantageous because they can search millions of compounds in a short period of time.

Zika virus NS3 helicase comprises of a cofactor ATP bound at the active site and this site is considered as an active site for developing new inhibitors using the CADD approach. Based on the essential features of ATP cofactor, ZINC database was screened by applying various filters in order to find best molecules for molecular docking studies. CDOCKER was used for docking, the

best molecules at the receptor active site in all possible conformations. The molecules were validated using MD simulations and binding free energy analysis. The molecules obtained through MD simulations were used to perform DFT calculations for solvent and gas phase energies and observed the location of HOMO and LUMO in the molecules located inside the active site of the Zika virus NS3 helicase.

3.2 Materials and methods

3.2.1 Sequence analyses of Zika virus NS3 helicases and structure analyses of flavivirus NS3 helicases

The representative 3-D crystal structures of flavivirus NS3 helicases were downloaded from PDB (Berman et al. 2000; Burley et al., 2021). The molecular visualization and structure superposition was carried out using UCSF Chimera (Pettersen et al., 2004). The sequences of Zika virus NS3 helicases were retrieved from NCBI using the BLAST searches (Johnson et al., 2008). The proteins multiple sequence alignment was accomplished with the help of the NGPhylogeny server (http://www.NGPhylogeny.fr). The Zika virus sequence (NCBI accession code: YP_009428568.1) was used as a reference protein to examine the amino acid mutations. The software suite developed by ABREAST (https://www.abreast.in) was used to identify mutations in the Zika virus NS3 helicases.

3.2.2 Protein and ligand preparation

The crystal structures of apo Zika virus NS3 helicase (PDB_id: 5JMT) (Jain et al., 2016) and when bound to ATP (5GJC)) (Tian et al., 2016) were determined at 1.8 and 2.2 Å resolutions, respectively. Crystal waters were removed from the structures and the missing residues in 5GJC were built by the loop modeling method in MODELLER (Šali & Blundell, 1993). Hydrogen atoms were added and their geometry optimization was performed by using UCSF Chimera.

3.2.3 Pharmacophore generation

The ligand-based pharmacophore models were generated by using the Pharmit server (Sunseri & Koes, 2016). Pharmit is an online server, which provides an interactive environment for the virtual screening of libraries of small molecule databases. Pharmit provides both pharmacophore and molecular shape-based search on the pre-inbuilt libraries of small molecules. The pharmacophore models were constructed based on the cofactor ATP complexed with Zika virus NS3 helicase based on the structure 5GJC obtained from the PDB. To design the pharmacophore model for ATP, the features of hydrogen bond donor, hydrogen bond acceptor, negative ionizable, hydrophobic, and aromatic ring that are required for interactions with the receptor binding pocket were considered. The virtual screening of the ZINC database was carried out using the generated pharmacophore model comprising 12,996,897 molecules (Irwin et al., 2012). The molecules retrieved from the ZINC database were filtered and ranked based on lower RMSD and molecular weight. The hit compounds were downloaded in .sdf format.

3.2.4 Molecular docking

Molecular docking predicts the optimal pose of the ligand within the receptor binding site. The CDOCKER (Gagnon et al., 2016) molecular docking methodology accessible in DS 3.5 was employed to find the best orientation of the molecules identified from the pharmacophore based virtual screening of ZINC database when binding to the receptor. A sphere of 5 Å radius was generated around ATP to define the active site of the protein. Twenty docking poses were generated for each molecule and shortlisted the conformations of molecules defined in the protein active site. As a part of the docking protocol, the molecular system was heated up to 700 K for 2,000 steps and then cooled to 300 K for 5,000 steps. The binding conformations of the molecules in Zika virus NS3 helicase were analyzed using “scoring ligand poses” implemented in receptor-

ligand interactions protocol in DS 3.5, and the scoring functions PLP1, PLP2 and PMF (Gehlhaar et al., 1995; Parrill & Reddy, 1999; Muegge & Martin 1999; Muegge, 2006) were used to analyze the docking poses. The best docking pose was selected based on top scores and intermolecular interactions within the Zika virus NS3 helicase active site. The top-ranked molecules from the molecular docking results were used further for assessing their drug-like properties.

3.2.5 Drug-like properties

The capability of a drug to permeate, distribute, metabolize, and retain within the host system for a reasonable period of time are described as the drug-like properties. These are the significant parameters to be deciphered for hit molecules selection. The SwissADME server supports the calculation of the ADMET properties (Lipinski et al., 1997; Daina et al., 2017 and Daina et al., 2014) and drug-like properties of the molecules. The Lipinski's rule of five indicated that a drug molecule must exhibit, ≤5 LogP, ≤ 500 molecular weight, ≤10 hydrogen bond acceptors and ≤5 hydrogen bond donors (Lipinski, 2004). The molecules that displayed drug-like properties were studied using the MD simulations in order to understand their ability to bind the protein active site, and the nonbonding interactions that stabilize the protein-ligand complexes and to quantify their binding free energies.

3.2.6 Molecular dynamics simulations

MD simulations of the reference molecule ATP and the screened molecules from ZINC database in complex with Zika virus NS3 helicase were carried out for 150 ns using GROMACS 5.1.4 (Hess et al., 2008; Van Der Spoel et al., 2005). Amber99sb force field (Hornak et al., 2006) was applied to both the receptor and small organic compounds; force fields were assigned to small molecules in Antechamber using the ACPYPE script (Da Silva et al., 2012) with AM1-BCC

charges (Wang et al., 2006). All of the molecular systems were immersed in a cubic box, a three-point model (SPC) was used to solvate the systems. The system was neutralized by the addition of Na^+ and Cl^- ions (Berendsen et al., 1981). The steric stress of the system was eliminated and the system was allowed to relax, the energy minimization was carried out using steepest descent algorithm. The maximum force was adjusted at less than 1000 kJ/mol/nm, and a scaling factor of 0.01 was used to limit the number of steps to 50,000. After that, position restraint dynamics was employed to stabilise the system and retain the solvent and ions in their proper locations around the protein. Before being equilibrated at 1 atm and 300 K for 1000 ps until it achieved the right density, this system was heated to 300 K for 100 ps. With a time, step of 0.002 ps, the final MD simulations were performed at 300 K for 150 ns. The pressure was controlled using the Parrinello-Rahman method (Parrinello & Rahman, 1981), and the V-rescale thermostat (Bussi et al., 2007) was used to maintain the temperature. The particle mesh Ewald technique (Darden et al., 1993; Essmann et al., 1995) was used to manage the long-range electrostatics, with a real-space cutoff of 10 Å, particle mesh Ewald order of 4, and a relative tolerance among both long and short-range energies of 10^{-6}. While Lennard-Jones interactions and real-space electrostatic interactions were terminated at 9 Å, short-range interactions have been analyzed using a neighbour list of 10 Å, which was updated after every ten steps. Using the LINCS algorithm, the hydrogen bonds were restrained (Hess et al., 1997). After structural stabilization, the final models for each system were created by averaging the MD simulations generated trajectories. To study the conformational variations in protein-inhibitor complexes, the Cα-atoms RMSD in relation to their initial structures was computed using gmx rms. The convergence of MD simulations was examined using RMSD plots. To investigate the stability of trajectories relative to the initial structure, the RMSF of the Cα-atoms was calculated using gmx rmsf.

3.2.7 Binding free energies

The binding free energies on the last ten ns of MD simulations trajectories were calculated to determine the strength of small molecules binding to Zika virus NS3 helicase. The GROMACS output trajectory compatible g mmpbsa (Kumari et al., 2014) tools were used to determine the binding free energy of each complex. To understand the significant contribution of reference molecule ATP and screened molecules identified in this work, total energies were calculated using MM-PBSA (Homeyer & Gohlke, 2012; Miller et al., 2012). The computations of the binding free energy, and the contribution of binding free energies from van der Waals, electrostatic, polar, and apolar solvation energy components were estimated.

The LIE technique is an end point approach for determining binding affinities. This technique incorporates explicit conformational sampling (of protein-bound and unbound-ligand states) with accuracy in determining the protein-ligand binding free energy ΔG_{bind}. LIE is also defined by calculating differences in average nonbonded (i.e., van der Waals) interaction energies between the ligand and its surroundings in either the Zika virus NS3 helicase bound and unbound states (denoted as subscript b) or unbound state (denoted as subscript u) (Åqvist et al., 1994). To calculate the mean of van der Waals (vdW) and coulomb (cou) interaction energy variation of the ligand with its neighboring atoms, ΔG_{bind} from simulations of the ligand bound to free protein or in solvent, the obtained average van der Waals and coulomb interaction energies of the ligand with its surroundings are scaled by LIE parameters. The LIE calculations were carried out using gmx energy and gmx lie for Zika virus bound with ATP and hit molecules from their MD simulation trajectories.

3.2.8 Density functional theory calculations

In order to investigate the strength of the final five screened molecules obtained through MD simulations, quantum chemical calculations have been performed. Geometry optimizations have been performed on all molecules at B3LYP (Axel, 1993; Lee et al., 1988) level of theory using 6-311+G(2d,p) basis set (Petersson et al., 1988; Petersson et al., 1991). Water was used as an implicit solvent in PCM optimization (Miertuš et al., 1981; Tomasi & Persico, 1994) at the B3LYP/6-311+G(2d,p) level for all the substances to obtain an indication of the impact of solvent presence on total energy of the compound. The nature of the stationary point was then determined using frequency calculations. Molecular orbital analysis was performed at the B3LYP/6-311+G(2d,p) level of theory, to evaluate the HOMO and the LUMO (Fukuli et al., 1952) for all the molecules obtained through MD simulations. This analysis was performed to identify the electron-rich zone in the compounds under consideration. The calculated values for the ESP energy were projected onto an isodensity surface that corresponded to 0.0004 a.u. to accurately analyse the charge distribution around these atoms (Vetrivel et al., 1996). This ESP accurately depicts the reactive zones surrounding the inhibitors since it takes into account the van der Waals volumes of each individual atom within the molecule. The Gaussian 09 programme suite was used to perform all calculations (Frisch et al., 2009).

3.3 Results and discussion

The tertiary structure of Zika virus NS3 helicase has three domains, domain I (residues 182–332) and domain II (333–481) share a similar structural fold, despite low sequence homology shared between them. Domain III (482–617) comprises a four α-helical bundle extended by two anti-parallel β-strands partially exposed to the solvent. Two α-helices from domain I interact with the parallel α-helical bundle from domain III; domain II connects with domain III via a long β-

hairpin, thus stabilizing the interdomain interactions, and all the three domains have clearly defined binding clefts. The cofactor ATP is located between domains-I and II in Zika virus NS3 helicase; it is stabilized by the Mn^{2+} coordinated in octahedral geometry by interactions with side-chain of residues Thr201 and Glu286, two oxygen atoms of β and γ phosphate groups of ATP and two water molecules. The binding site of ATP is formed by the residues Gly197, Ala198, Gly199, Lys200, Thr201, Arg202 and Arg203 (P-loop), and Glu286, Ala317, Asn330, Gly415, Asn417, Gln455, Gly458, Arg459, Arg462 and Asn463 as shown in the **Figure 3.1**.

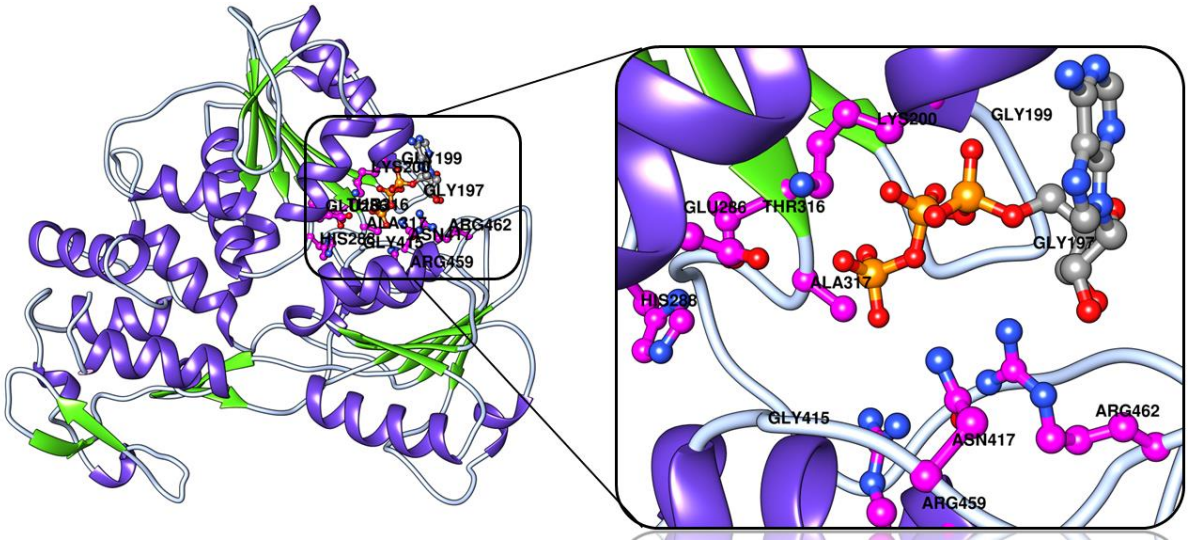
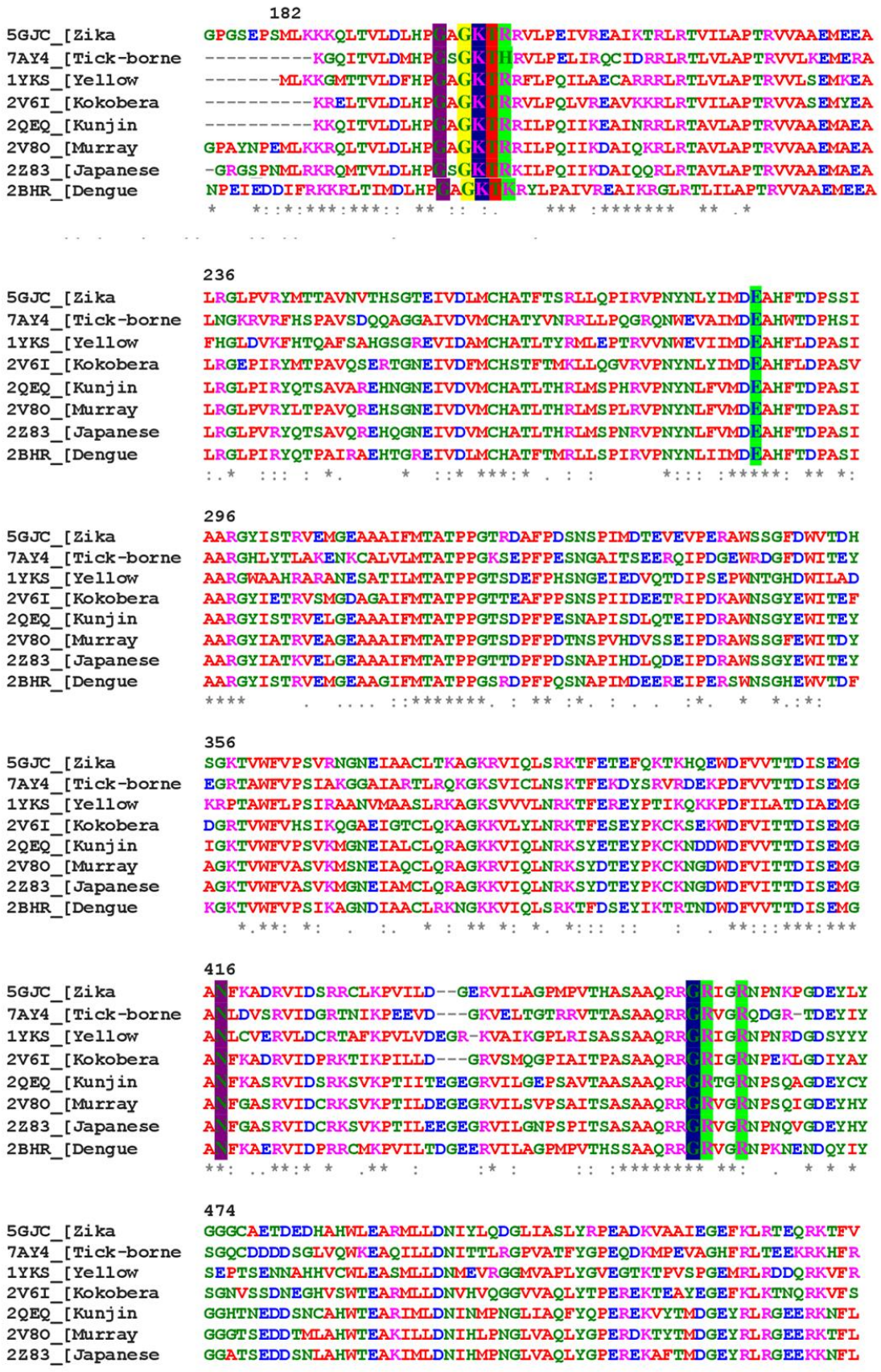


Figure 3.1: ATP binding site in Zika virus NS3 helicase.

3.3.1 Structure analyses of flavivirus and sequence analyses of Zika virus NS3 helicases

The multiple sequence alignment of flavivirus NS3 helicases reveals that all the amino acid residues within 4 Å from ATP (5GJC) and RNA (5GJB) binding site are highly conserved among all flaviviruses. The structures are highly superposable with low RMSD indicating their high structure conservation as shown in **Figure 3.2A**.


182
5GJC_[Zika GPGSEPSMLKKKQLTVLDLHPGAGKTKRVLPEIVREAIKTRLRTVILAPTRVVAAEMEEA
7AY4_[Tick-borne ----------KGQITVLDMHPGSGKTHRVLPELIRQCIDRRLRTLVLAPTRVVLKEMERA
1YKS_[Yellow -------MLKKGMTTVLDFHPGAGKTKRFLPQILAECARRRLRTLVLAPTRVVLSEMKEA
2V6I_[Kokobera ----------KRELTVLDLHPGAGKTKRVLPQLVREAVKKRLRTVILAPTRVVASEMYEA
2QEQ_[Kunjin ----------KKQITVLDLHPGAGKTKRILPQIIKEAINRRLRTAVLAPTRVVAAEMAEA
2V8O_[Murray GPAYNPEMLKKRQLTVLDLHPGAGKTKRILPQIIKDAIQKRLRTAVLAPTRVVAAEMAEA
2Z83_[Japanese -GRGSPNMLRKRQMTVLDLHPGSGKTKKILPQIIKDAIQQRLRTAVLAPTRVVAAEMAEA
2BHR_[Dengue NPEIEDDIFRKKRLTIMDLHPGAGKTKRYLPAIVREAIKRGLRTLILAPTRVVAAEMEEA
236
5GJC_[Zika LRGLPVRYMTTAVNVTHSGTEIVDLMCHATFTSRLLQPIRVPNYNLYIMDEAHFTDPSSI
7AY4_[Tick-borne LNGKRVRFHSPAVSDQQAGGAIVDVMCHATYVNRRLLPQGRQNWEVAIMDEAHWTDPHSI
1YKS_[Yellow FHGLDVKFHTQAFSAHGSGREVIDAMCHATLTYRMLEPTRVVNWEVIIMDEAHFLDPASI
2V6I_[Kokobera LRGEPIRYMTPAVQSERTGNEIVDFMCHSTFTMKLLQGVRVPNYNLYIMDEAHFLDPASV
2QEQ_[Kunjin LRGLPIRYQTSAVAREHNGNEIVDVMCHATLTHRLMSPHRVPNYNLFVMDEAHFTDPASI
2V8O_[Murray LRGLPVRYLTPAVQREHSGNEIVDVMCHATLTHRLMSPLRVPNYNLFVMDEAHFTDPASI
2Z83_[Japanese LRGLPVRYQTSAVQREHQGNEIVDVMCHATLTHRLMSPNRVPNYNLFVMDEAHFTDPASI
2BHR_[Dengue LRGLPIRYQTPAIRAEHTGREIVDLMCHATFTMRLLSPIRVPNYNLIIMDEAHFTDPASI
296
5GJC_[Zika AARGYISTRVEMGEAAAIFMTATPPGTRDAFPDSNSPIMDTEVEVPERAWSSGFDWVTDH
7AY4_[Tick-borne AARGHLYTLAKENKCALVLMTATPPGKSEPFPESNGAITSEERQIPDGEWRDGFDWITEY
1YKS_[Yellow AARGWAAHRARANESATILMTATPPGTSDEFPHSNGEIEDVQTDIPSEPWNTGHDWILAD
2V6I_[Kokobera AARGYIETRVSMGDAGAIFMTATPPGTTEAFPPSNSPIIDEETRIPDKAWNSGYEWITEF
2QEQ_[Kunjin AARGYISTRVELGEAAAIFMTATPPGTSDPFPESNAPISDLQTEIPDRAWNSGYEWITEY
2V8O_[Murray AARGYIATRVEAGEAAAIFMTATPPGTSDPFPDTNSPVHDVSSEIPDRAWSSGFEWITDY
2Z83_[Japanese AARGYIATKVELGEAAAIFMTATPPGTTDPFPDSNAPIHDLQDEIPDRAWSSGYEWITEY
2BHR_[Dengue AARGYISTRVEMGEAAGIFMTATPPGSRDPFPQSNAPIMDEEREIPERSWNSGHEWVTDF
356
5GJC_[Zika SGKTVWFVPSVRNGNEIAACLTKAGKRVIQLSRKTFETEFQKTKHQEWDFVVTTDISEMG
7AY4_[Tick-borne EGRTAWFVPSIAKGGAIARTLRQKGKSVICLNSKTFEKDYSRVRDEKPDFVVTTDISEMG
1YKS_[Yellow KRPTAWFLPSIRAANVMAASLRKAGKSVVVLNRKTFEREYPTIKQKKPDFILATDIAEMG
2V6I_[Kokobera DGRTVWFVHSIKQGAEIGTCLQKAGKKVLYLNRKTFESEYPKCKSEKWDFVITTDISEMG
2QEQ_[Kunjin IGKTVWFVPSVKMGNEIALCLQRAGKKVIQLNRKSYETEYPKCKNDDWDFVVTTDISEMG
2V8O_[Murray AGKTVWFVASVKMSNEIAQCLQRAGKRVIQLNRKSYDTEYPKCKNGDWDFVITTDISEMG
2Z83_[Japanese AGKTVWFVASVKMGNEIAMCLQRAGKKVIQLNRKSYDTEYPKCKNGDWDFVITTDISEMG
2BHR_[Dengue KGKTVWFVPSIKAGNDIAACLRKNGKKVIQLSRKTFDSEYIKTRTNDWDFVVTTDISEMG
416
5GJC_[Zika AMFKADRVIDSRRCLKPVILD--GERVILAGPMPVTHASAAQRRGRIGRNPNKPGDEYLY
7AY4_[Tick-borne AMLDVSRVIDGRTNIKPEEVD---GKVELTGTRRVTTASAAQRRGRVGRQDGR-TDEYIY
1YKS_[Yellow AMLCVERVLDCRTAFKPVLVDEGR-KVAIKGPLRISASSAAQRRGRIGRNPNRDGDSYYY
2V6I_[Kokobera AMFKADRVIDPRKTIKPILLD---GRVSMQGPIAITPASAAQRRGRIGRNPEKLGDIYAY
2QEQ_[Kunjin AMFKASRVIDSRKSVKPTIITEGEGRVILGEPSAVTAASAAQRRGRTGRNPSQAGDEYCY
2V8O_[Murray AMFGASRVIDCRKSVKPTILDEGEGRVILSVPSAITSASAAQRRGRVGRNPSQIGDEYHY
2Z83_[Japanese AMFGASRVIDCRKSVKPTILEEGEGRVILGNPSPITSASAAQRRGRVGRNPNQVGDEYHY
2BHR_[Dengue AMFKAERVIDPRRCMKPVILTDGEERVILAGPMPVTHSSAAQRRGRVGRNPKNENDQYIY
474
5GJC_[Zika GGGCAETDEDHAHWLEARMLLDNIYLQDGLIASLYRPEADKVAAIEGEFKLRTEQRKTFV
7AY4_[Tick-borne SGQCDDDDSGLVQWKEAQILLDNITTLRGPVATFYGPEQDKMPEVAGHFRLTEEKRKHFR
1YKS_[Yellow SEPTSENNAHHVCWLEASMLLDNMEVRGGMVAPLYGVEGTKTPVSPGEMRLRDDQRKVFR
2V6I_[Kokobera SGNVSSDNEGHVSWTEARMLLDNVHVQGGVVAQLYTPEREKTEAYEGEFKLKTNQRKVFS
2QEQ_[Kunjin GGHTNEDDSNCAHWTEARIMLDNINMPNGLIAQFYQPEREKVYTMDGEYRLRGEERKNFL
2V8O_[Murray GGGTSEDDTMLAHWTEAKILLDNIHLPNGLVAQLYGPERDKTYTMDGEYRLRGEERKTFL
2Z83_[Japanese GGATSEDDSNLAHWTEAKIMLDNIHMPNGLVAQLYGPEREKAFTMDGEYRLRGEEKKNFL

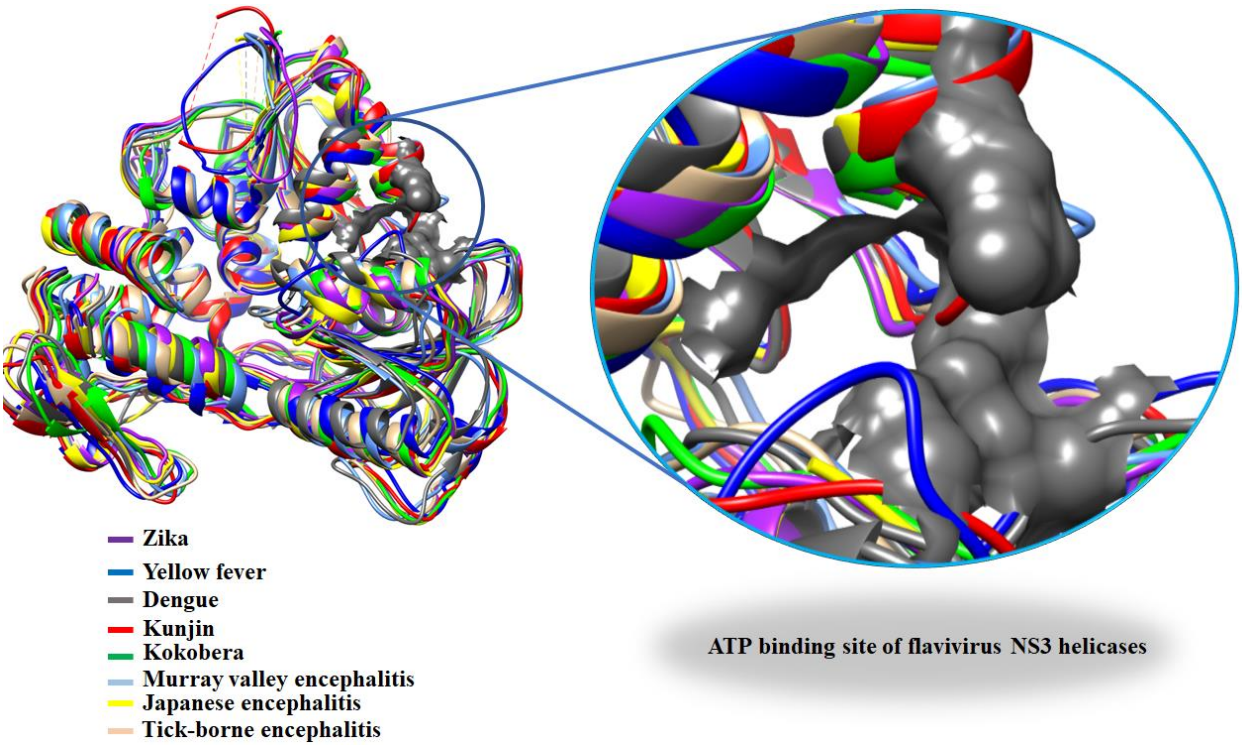


Figure 3.2A: sequence and structural alignment of Zika virus other flavivirus helicases.

From the NCBI protein sequence database, the primary structures of 1,083 Zika virus NS3 helicases were retrieved. Mutational analyses were carried out on the multiple sequence alignment generated by NGphylogeny server that revealed 115 unique mutation positions in a protein of 436 amino acid residues (182-617 positions)**.** Also, it was observed that only two mutations lie within 4 Å from the ATP binding sites. Gly458 is mutated to Arg only in one instance (NCBI id: QKF93433.1), and it was observed that the mutation of Gly458Arg increases the possibility of its interaction with ATP. The Asn463 is mutated to Asp in NCBI id: ATG29292.1. This Asn463Asp mutation does not make any significant alterations in the ability of the protein to bind ATP. This infers that despite the presence of mutations, the function of the protein would be retained. In the RNA binding pocket, mutations were observed at certain locations; Arg226Pro (AHL43503.1), Met244Gln (AHL43503.1), Asp291Gly (AZS35408.1), Ser365Arg (AMK79469.1), Lys537Arg

(AQS26826.1, AOO54270.1, APH11492.1, QLJ57653.1, ARB07949.1, AND01116.2, AOY08529.2, AQS26833.1, AMQ48981.1). The side-chain of Arg226 and Lys537 make hydrogen bonding interactions with the phosphate oxygen and 2' OH of the ribose sugar respectively from the C5 of RNA. High frequency mutations are observed at the amino acid positions, 185 (23 occurrences), 215 (132), 355 (77), 360 (42), 400 (132), 407 (124), 472 (128), 483 (97), 572 (232), 583 (122), 584 (253) and it was observed that these mutations are located in all the three domains. The β-sheet regions in domains I and II accommodate the mutations and also in the periphery of the protein **Figure 3.2B**. Interestingly, none of the mutations are located at the ATP, RNA binding sites and the interdomain interface indicating the unaltered NS3 helicase function despite a large number of mutations accumulated in the protein over a short period of its evolution.

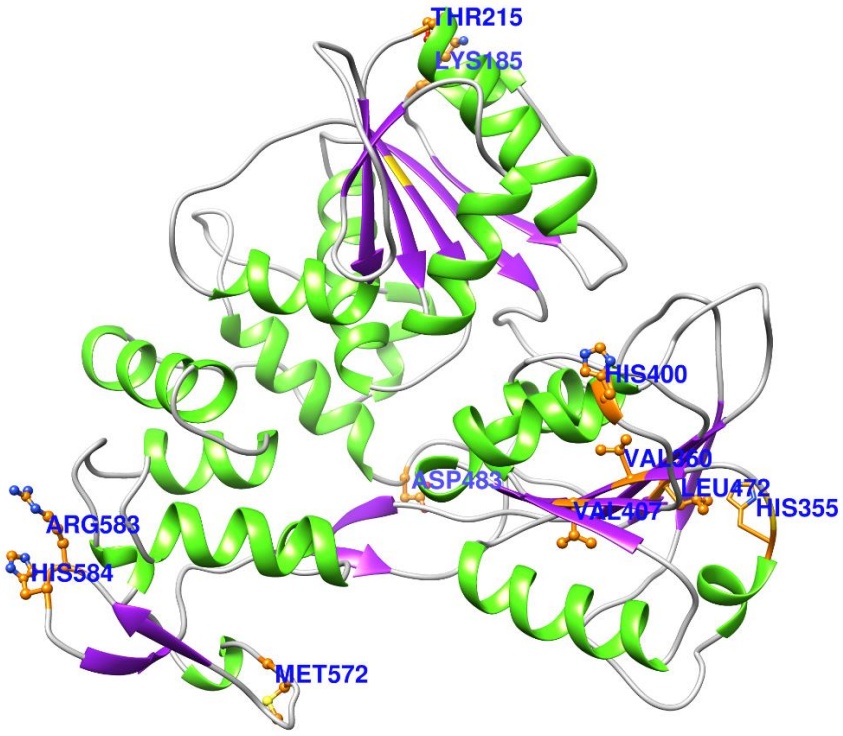


Figure 3.2B: Mutation of residues in Zika virus NS3 helicase Thr215Lys, Tyr355His, Val360Ile, His400Met, Val407Ile, Leu472Met, Asp483Gly, Met572Leu, Arg583Lys, His584Tyr.

3.3.2 Pharmacophore generation and virtual screening

Pharmacophore modeling using the Pharmit server identified the features of ATP that are required for interactions with the amino acid residues present in its binding site of Zika virus NS3 helicase. The five electronic and steric pharmacophore features, selected in ATP are shown in **Figure 3.3**. The aromatic feature that shows Pi-cation interaction with Arg202, four-hydrogen bond acceptor features; β-phosphate O1β, α-phosphate O3α, α-phosphate O1α and N7 at the imidazole ring of adenine form hydrogen bonding with Lys200, Gly199, Arg462 and Arg202, respectively, were selected as pharmacophore features for screening. The pharmacophore model was validated on 1000 molecules decoy set (inactive molecules) using Pharmit server. The pharmacophore model could not predict a single decoy molecule, which indicated the pharmacophore model to successfully qualify the validation process.

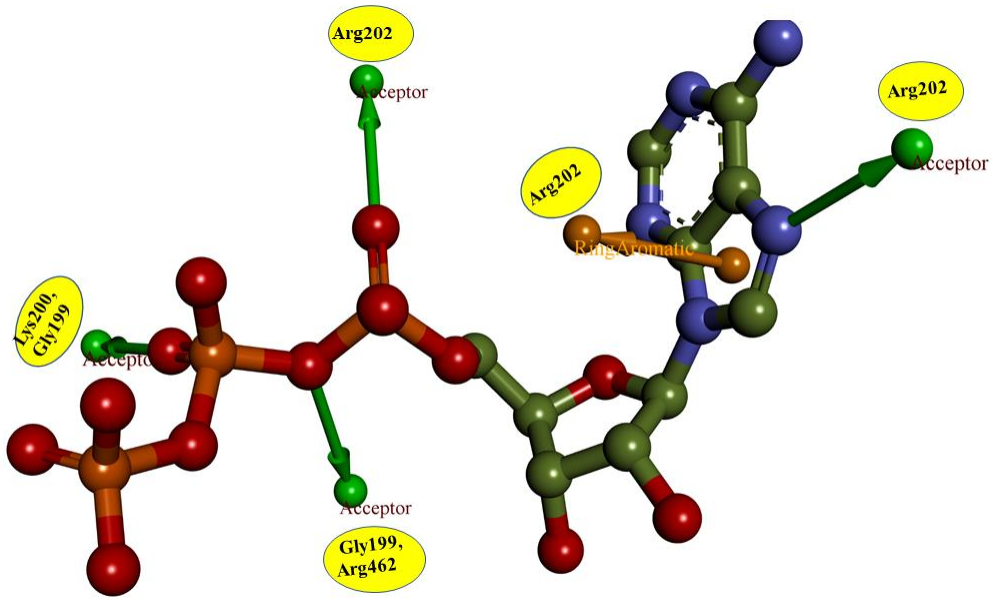


Figure 3.3: Pharmacophore features of ATP used for virtual screening of ZINC database.

The validated pharmacophore model was employed for *in silico* screening of the ZINC database of small molecules (Irwin et al., 2012). The 12,996,897 molecules in the ZINC database have 164,282,714 confirmations. The five featured pharmacophore model that was used as a 3-D query retrieved 7,526 hit molecules from the ZINC database. These hits were filtered and ranked based on RMSD < 0.25 Å and molecular weight < 500 Da; based on this criteria, 471 molecules were selected and downloaded for further molecular docking studies.

3.3.3 Molecular docking

The 471 hit molecules identified through pharmacophore based virtual screening were further analysed by molecular docking methodology, CDOCKER, to verify the binding orientation and binding affinity to Zika virus NS3 helicase. The binding site of protein was defined as a sphere generated 5 Å around ATP. For each molecule 20 docking poses were generated, out of the 20 docking poses, the conformers with the high CDOCKER docking scores were visualized on graphics for nonbonding interactions such as hydrogen bonds, hydrophobic and ionic interactions with protein active site residues. The results of molecular docking of the best ten hit molecules and ATP are shown in **Table 3.1.** These selected molecules bind at the location of the ATP and form hydrogen bonds with amino acids Gly199, Lys200, Thr201, Arg202, Asn417 and Arg462.

Table 3.1: Scoring of best docking structure of screened hit molecules.

S. No	Compound Name	-PLP1	-PLP2	-PMF	-PMF04	Active site interactions
1.	ATP	98.57	98.61	106.27	-8.4	Leu194, His195, Pro196, Gly197, Ala198, **Gly199**, **Lys200**, **Thr201**, **Arg202**, Arg203, Ala235, Glu286, His288, Thr316, Ala317, Asn317, Asn330, Gly415, **Asn417**, Lys419, Gln455,

						Gly458, Arg459, **Arg462**, Asn463, Pro464
2.	ZINC000032856030	74.53	70.12	58.87	21.45	Pro196, Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286, Ala317, Asn330, Gly415, **Asn417**, Phe418, Arg459, **Arg462**, Asn463, Pro464
3.	ZINC000299571648	52.32	47.55	45.09	12.88	Gly197, **Ala198**, **Gly199**, **Lys200**, Thr201, **Arg202**, Arg203, Asp285, Glu286, Asn330, Gly415, Asn417, Phe418, Arg459, **Arg462**, Asn463, Pro464
4.	ZINC000058178136	87.1	72.7	57.66	20.26	**Gly197**, Ala198, **Gly199**, **Lys200**, **Thr201**, Arg202, Arg203, Asp285, Glu286, Gly415, Asn417, Phe418, Arg459, **Arg462**, Asn463, Pro464
5.	ZINC000408750650	88.77	88.48	58.45	5.06	Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286, Gly415, **Asn417**, Phe418, Arg459, Arg462, Asn463, Pro464
6.	ZINC000015675488	79.07	71.34	51.92	36.26	Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286, Gly415, **Asn417**, Phe418, Arg459, **Arg462**, Asn463, Pro464
7.	ZINC000008731686	72.18	66.08	53.77	12.95	Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286,

						Gly415, **Asn417**, Phe418, Arg459, **Arg462**, Asn463, Pro464
8.	ZINC000071404039	89.6	94.00	47.14	8.62	Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286, Gly415, Asn417, Phe418, Arg459, **Arg462**, Asn463, Pro464
9.	ZINC000888101786	61.83	53.48	54.14	20.24	Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286, **Asn330**, Gly415, Asn417, Phe418, Arg459, **Arg462**, Asn463, Pro464
10.	ZINC000021168927	108.73	102.03	85.91	22.54	Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286, **Asn330**, Gly415, **Asn417**, Phe418, Arg459, **Arg462**, Asn463, Pro464
11.	ZINC000670457317	91.93	80.19	62.9	14.3	Gly197, Ala198, **Gly199**, **Lys200**, Thr201, Arg202, Arg203, Asp285, Glu286, Gly415, **Asn417**, Phe418, Arg459, Arg462, Asn463, Pro464

3.3.4 Drug-like properties

The drug-like features of the identified hit molecules were studied using SwissADME server and Lipinski's rule of five. The synthetic ease of access of the proposed molecules was graded on a scale of one to ten on the basis of complexity of the compounds, the total count of stereocenters, and other factors. The more synthetically accessible the compound should have a

lower value. The synthetic accessibility values in this work were reported to be less than 4.75, demonstrating the simplicity of their synthesis. The topological polar surface area (TPSA) is much less than 140 $Å^2$, the solubility in water (given as Log *S*) indicates that the majority of compounds are either slightly or fairly soluble in water, and the lipophilicity (represented as cLogP) is less than 5. The likelihood of skin permeation is indicated by the skin permeation probability represented as Log Kp, which is also suitable. All of these molecules exhibit LogP ≤5, molecular weight ≤ 500, no of hydrogen bond acceptors ≤10 and no of hydrogen bond donors ≤5 (Lipinski, 2004). This implies that all the ten selected hit molecules from molecular docking studies qualify to possess drug-like properties. The ADME properties are shown in **Table 3.2**. The compound identities (IDs) were represented in the main text by taking the last four digits of ZINC ID, for example ZINC000032856030 as ZINC6030. The ten selected hit molecules in complex with Zika virus NS3 helicase were subjected to MD simulations studies. For the sake of comparison, apo and ATP complexed Zika virus NS3 helicase were also subjected to MD simulations studies.

Table 3.2: Drug-like properties of screened ZINC compounds.

S. No	Compound name	TPSA $(A)^2$	Consensus po/w	Log kp (cm/s)	Synthetic Accessibility
1	ZINC6030	68.44	1.93	-6.3	3.50
2	ZINC1648	68.96	1.20	-7.32	3.50
3	ZINC8136	97.04	2.80	-6.73	2.21
4	ZINC0650	132.83	4.03	-5.22	2.97
5	ZINC5488	137.14	0.06	-8.48	4.67
6	ZINC1686	102.54	2.44	-6.38	3.01
7	ZINC4039	75.97	2.86	-6.19	2.76
8	ZINC1786	110.76	3.27	-6.68	4.25

9	ZINC6897	122.62	2.61	-6.39	3.75
10.	ZINC7317	72.70	2.99	-5.96	3.25

3.3.5 Molecular dynamics simulations

MD simulations offer information on the binding of screened hit molecules to the Zika virus NS3 helicase at the residue level. Protein-ligand bound complexes were studied in a dynamic environment to assess the stability of interactions over a period of time. In this analysis, apo-protein and Zika virus NS3 helicase complexed with ATP and five screened hit molecules ZINC6030, ZINC1648, ZINC8136, ZINC1786 and ZINC6897 were found to be stable during the MD simulations**.** The average structures were generated from 150 ns MD simulations trajectories for structure comparison. The superposition of the initial and average structures demonstrates that the molecules have good stability and low RMSD, which was confirmed by conserved hydrogen bonding interactions with active site residues, denoting that all these molecules are suitable as Zika virus NS3 helicase inhibitors. The overall conformational changes were analyzed through the Cα-atoms RMSD as a function of MD simulations time **Figure 3.4A**. From the plots it can be seen that the apo-protein gained stability after the initial ten ns of simulation time with an average RMSD of 2 Å. From the comparison of RMSD trajectories, it is evident that ZINC6030, ZINC8136, and ZINC1786 possess better binding and acquire stable interactions at the ATP binding pocket of NS3 helicase throughout 150 ns of MD simulations. The screened hit molecules show lower than 0.25 nm RMSD, as shown in **Figure 3.4B**, and the molecule ZINC6030 offers the highest RMSD value ~0.22 nm compared with the other molecules and reference ATP.

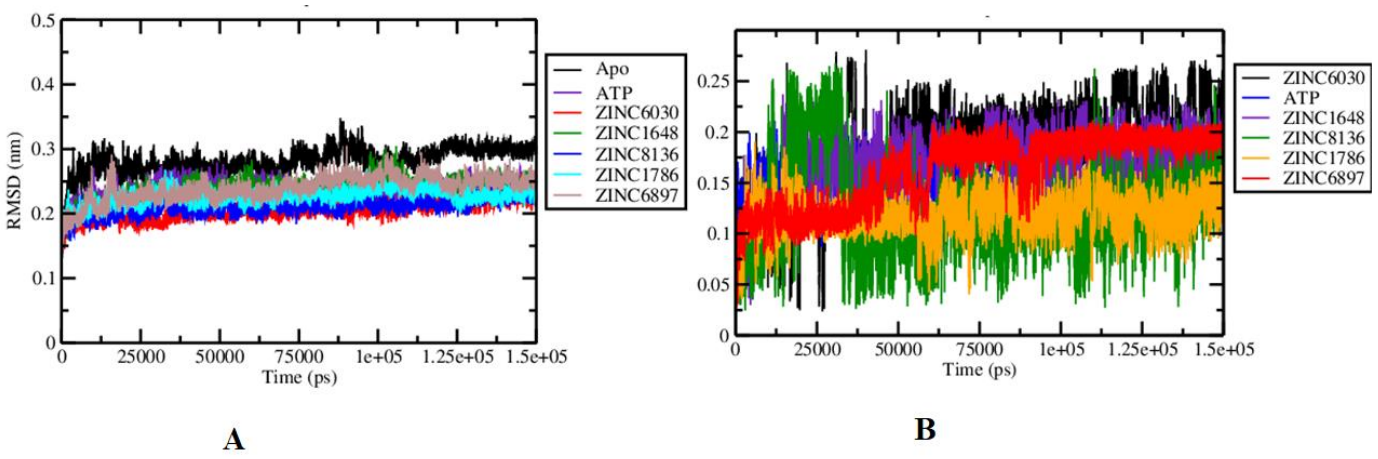


Figure 3.4: A) RMSD plot of Zika virus NS3 helicase apo and bound with screened molecules

B) RMSD plot of screened molecules bounds with Zika virus NS3 helicase.

From the RMSF plots for the Cα-backbone atoms **Figure 3.5**, it can be seen that most amino acid residues in the protein have lower fluctuations (< 0.2 nm), and only the residues that form a loop from Met244 to Ser256 display higher fluctuations and reach until 0.5 nm, in the ATP bound NS3 helicase as well as in complexes with ZINC1648 and ZINC6897 (0.23 nm). This is in correspondence with our previous studies that reported the mobile nature of the RNA binding loop (Met244-Sert253) reported in chapter 2. The amino acid residues important for inhibitor binding are, (Met244-Ser253; RNA-binding loop), (Pro273- Ala278; near to ATP binding site), (Pro319-Ala325; downward to the ATPase site) and (Pro464-Glu470; adjacent to ATP site) that show significant fluctuations. The overall RMSF results show that protein attains greater stability throughout MD simulations when it is complexed with screened hit molecules than the ATP bound molecular system.

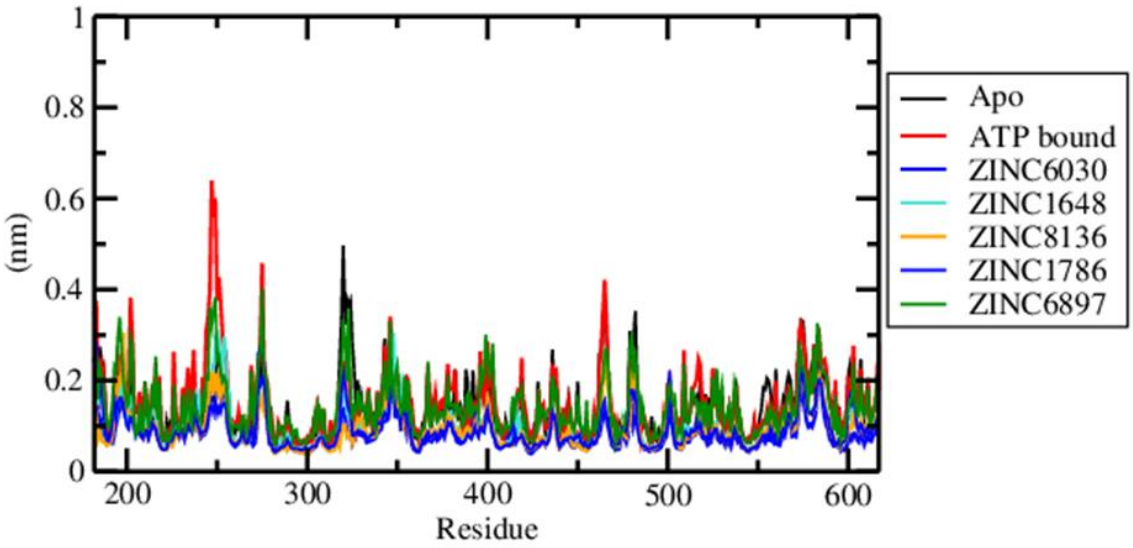
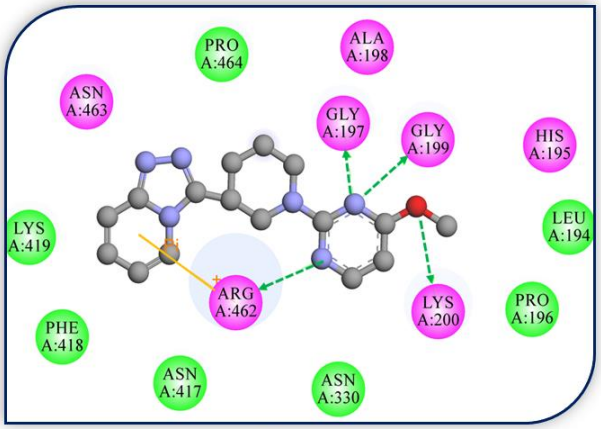
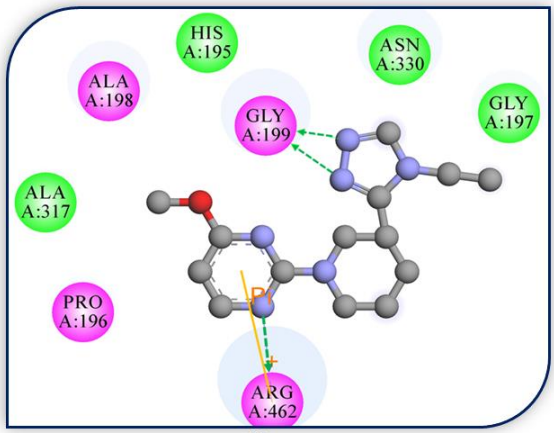


Figure 3.5: RMSF plots of Zika virus NS3 helicase apo, ATP and screened hit molecule complexes.

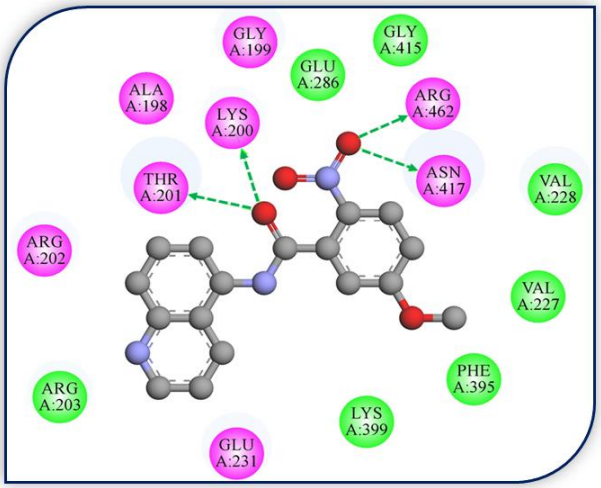
All hit molecules make hydrogen bonds with the active site amino acid residues Gly199, Lys200, Thr201, Arg202, Glu286, Asn417, and Arg462. Hydrogen bonding and hydrophobic interactions stabilize the Zika virus NS3 helicase screened inhibitor complexes. The protein-ligand interactions were generated throughout MD simulations using DS 3.5. ATP makes hydrogen bonding interactions with Gly199, Lys200, Thr201, Arg202, Glu286, Asn417, and Arg462; these interactions are retained throughout MD simulations. The screened hit molecules show hydrogen bonding and pi-cation interactions with active site residues; these interactions tightly hold the molecules inside the active site of Zika virus NS3 helicase. The molecule ZINC6030 shows Gly199, Lys200, Asn417, Arg462, and pi-cation Arg462, whereas in ZINC1648 shows hydrogen bonds Gly199, Arg462 and Arg462 involves Pi-cation, ZINC8136 form Lys200, Thr201, Asn417, Arg462, whereas ZINC1786 exhibits Lys200, Arg202, Arg462 and pi-cation with Arg459, Arg462 and ZINC6897 shows Ala198, Lys200, Arg203 and Arg462 as shown in **Figure 3.6**.


PRO A:464
ALA A:198
ASN A:463
GLY A:197
GLY A:199
HIS A:195
LYS A:419
LEU A:194
ARG A:462
LYS A:200
PRO A:196
PHE A:418
ASN A:417
ASN A:330

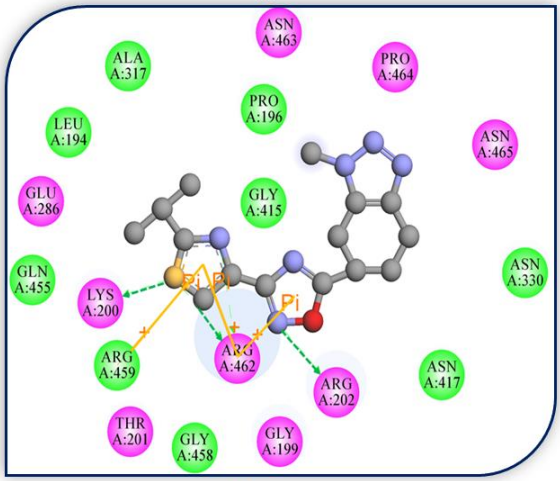
A


HIS A:195
ASN A:330
ALA A:198
GLY A:199
GLY A:197
ALA A:317
PRO A:196
ARG A:462

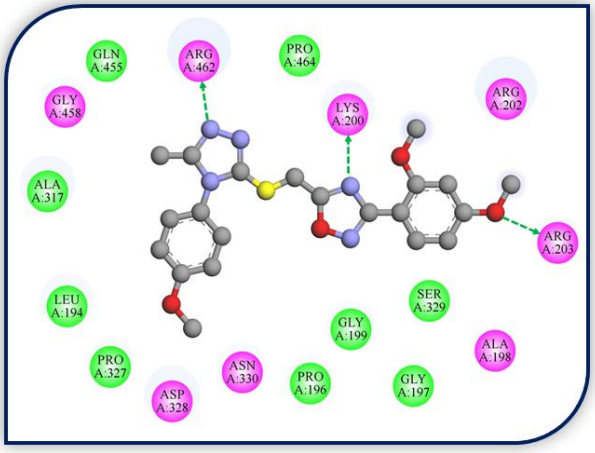
B


GLY A:199
GLY A:415
GLU A:286
ALA A:198
ARG A:462
LYS A:200
THR A:201
ASN A:417
VAL A:228
ARG A:202
VAL A:227
PHE A:395
ARG A:203
LYS A:399
GLU A:231

C


ASN A:463
ALA A:317
PRO A:464
PRO A:196
LEU A:194
ASN A:465
GLU A:286
GLY A:415
GLN A:455
ASN A:330
LYS A:200
ARG A:459
ARG A:462
ARG A:202
ASN A:417
THR A:201
GLY A:458
GLY A:199

D


GLN A:455
ARG A:462
PRO A:464
GLY A:458
LYS A:200
ARG A:202
ALA A:317
ARG A:203
LEU A:194
SER A:329
GLY A:199
ALA A:198
PRO A:327
ASN A:330
PRO A:196
GLY A:197
ASP A:328

E

Figure 3.6: Interactions of screened molecules ZINC6030 (**A**), ZINC8136 (**B**), ZINC1648 (**C**), ZINC1786 (**D**) and ZINC6897 (**E**) with Zika virus NS3 helicase.

The Rg demonstrated the compactness of the molecular systems under study during MD simulations. The Rg of all molecular systems showed that the screened hits have lower Rg, during the 150 ns MD simulations. When Zika virus NS3 helicase is bound to ATP and screened hits, the Rg is found to be below ~0.23 nm, indicating that all prepared molecular systems attain stability throughout MD simulations.

3.3.6 Binding free energy and residue-wise contribution energy

The binding free energies of the Zika virus NS3 helicase in complex with ATP and screened-in hit molecules were calculated using MM-PBSA and LIE methods using GROMACS as shown in **Table 3.3**. Reference compound and screened-in hit molecules can be compatible with one another due to the energy contributions from electrostatic, van der Waals, and polar solvation. The binding free energy for ATP Zika virus NS3 helicase complex was observed to be -186.016 +/- 0.00 kJ/mol, while the binding free energies for the screened molecules were ZINC6030 (-92.91 +/- 0.63 kJ/mol), ZINC1648 (-96.19 +/- 0.67 kJ/mol), ZINC8136 (-52.95 +/- 0.71 kJ/mol), ZINC1786 (-95.01 +/- 0.52 kJ/mol) and ZINC6897 (-131.46 +/- 0.71 kJ/mol). The binding site residues Gly199, Lys200, Thr201, Arg202, Glu286, Ala317, Asn417, Arg428, Gly458, Arg459, and Arg462 contribute to the highest binding energies in all the complexes studied, as shown in **Figure 3.7**.

Table 3.3. Binding free energy calculations of ATP and screened hit molecules.

S.No	Compound	van der Waals	Electrostatic energy	Polar solvation energy	SASA (kJ/mol)	ΔG (kJ/mol)	ΔG_{LIE} (kJ/mol)

	Name	energy (kJ/mol)	(kJ/mol)	(kJ/mol)			
1.	ATP	-188.611 +/- 0.0	-88.425+/- 0.0	107.156 +/- 0.0	-17.136 +/- 0.0	-186.016+/- 0.0	-168.016
2.	ZINC0000 32856030	-142.68+/- 0.55	-44.15+/- 0.43	107.75+/-0.87	-13.81+/-0.05	-92.91+/-0.63	-123.843
3.	ZINC0002 99571648	-143.63+/- 0.58	-48.61+/- 0.42	109.97+/-0.65	-13.89+/-0.05	-96.19+/-0.67	-92.096
4.	ZINC0000 58178136	-154.89+/- 0.47	-51.09+/- 0.45	168.61+/-1.02	-15.54+/-0.04	-52.95+/-0.71	-47.453
5.	ZINC0008 88101786	-155.66+/- 0.35	-6.39+/- 0.17	83.51+/-0.55	-16.47+/-0.03	-95.01+/-0.52	-98.501
6.	ZINC0000 2116897	-238.69+/- 0.48	-26.69+/- 0.33	154.35+/-0.70	-20.49+/-0.04	-131.46+/-0.71	-90.155

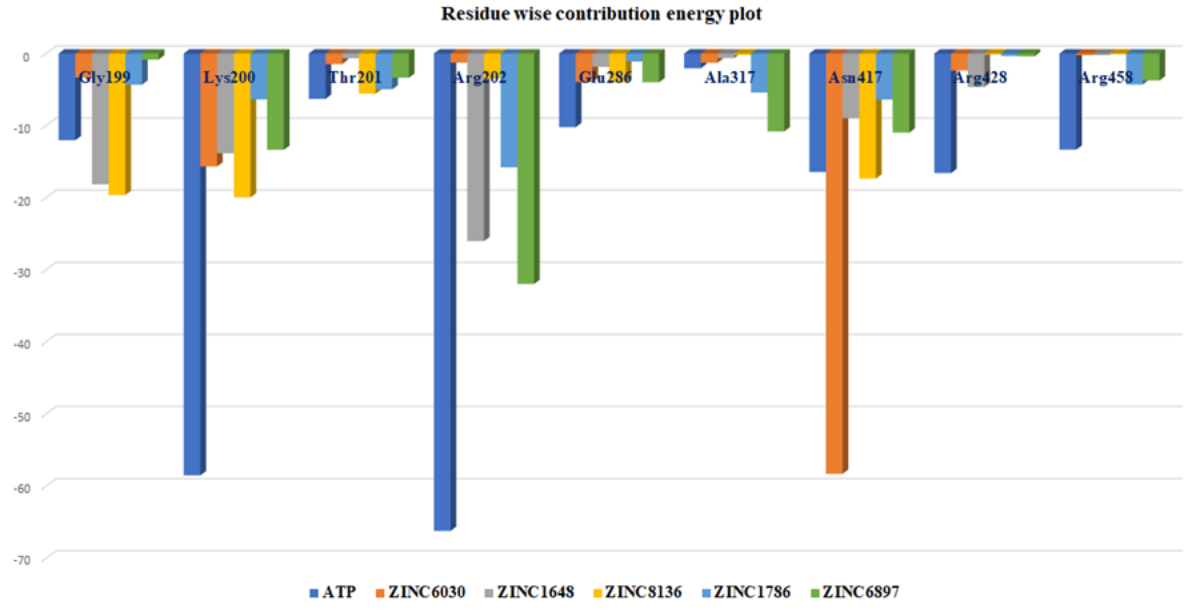


Figure 3.7: Residue-wise contribution (kJ/mol) of Zika virus NS3 helicase binds with cofactor ATP and screened molecules.

The Zika virus NS3 helicase bound with ATP shows LIE of –168.016 kJ/mol and the screened molecule ZINC6030 shows highest energy -123.843 kJ/mol and the molecule ZINC8136 shows lower energy (-47.453 kJ/mol), the other complexes ZINC1648, ZINC1786 and ZINC6897 show reasonably good binding free energies. The binding free energy of ATP is much lower, as it is a cofactor to the protein and makes several hydrogen bonding and other nonbonding interactions with the protein compared to the hit molecules. From the binding energy results it was noticed that the screened molecules attain stability when bound to Zika virus NS3 helicase at the ATP binding site.

3.4 Density functional theory calculations

The hit molecules ZINC6030, ZINC8136, ZINC1648, ZINC1786 and ZINC6897 validated through MD simulations, were further explored through DFT calculations. Quantum chemical calculations have been performed, in order to investigate the strength and nature on potential energy surface of the five screened molecules.

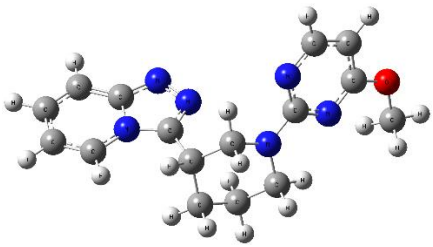
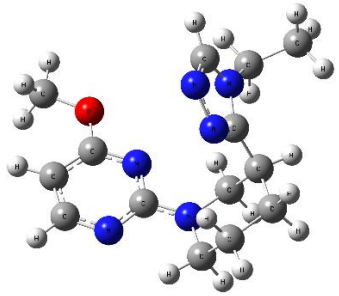
3.4.1 Geometry optimization

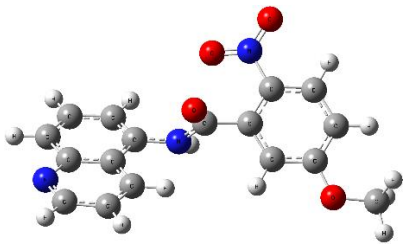
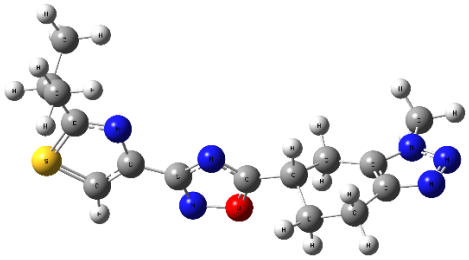
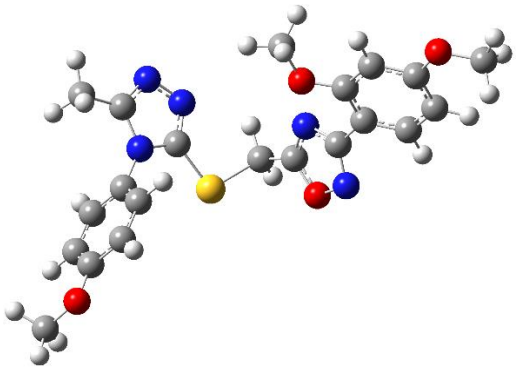
The Geometry optimization and frequency analysis were carried out at B3LYP/6-311+G(2d,p) level of theory upon the five ZINC6030, ZINC8136, ZINC1648, ZINC1786 and ZINC6897 molecules validated through MD simulations, to understand the nature and energy of these molecules. The total energy (in hartrees) in the considered molecules follows the trend ZINC6897 > ZINC1786 > ZINC1648 > ZINC6030 > ZINC8136. For the current study, all five conformations were local minima on the PES.

3.4.2 Solvent effect

To assess the effect of the solvents existence on the molecules total energy, PCM enhancement at the B3LYP/6-311+G(2d,p) level of theory was performed with water as the implicit solvent. Though the overall energy (in hartrees) in the solvated molecules is more compared to gas phase geometries, their trends remain the same ZINC6897 > ZINC1786 > ZINC1648 > ZINC6030 > ZINC8136. There is a difference of ~ -54 kJ/mol for the molecule ZINC6030, ~ -58 kJ/mol for the molecule ZINC8136, ~ -61 kJ/mol for the molecule ZINC1648, ~ -55 kJ/mol for the molecule ZINC1786 and ~ -72 kJ/mol for the molecule ZINC6897 respectively, thus suggesting that the molecules are more stable in solvent medium compared to gas phase as shown in **Table 3.4.**

Table 3.4: Total energy (in hartrees), point group and frequency of five Zika virus ligands determined using B3LYP/6–311G+(2d,p). (PCM total energy (in hartrees) in Bold with water as solvent).

S. No.	Name	Geometry	Total Energy	PG	NImag
1	ZINC6030		-1024.550965 **-1024.571623**	C_1	0
2	ZINC1648		-949.531433 **-949.553711**	C_1	0

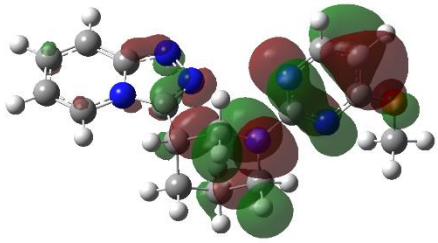
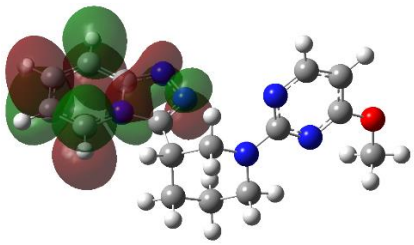
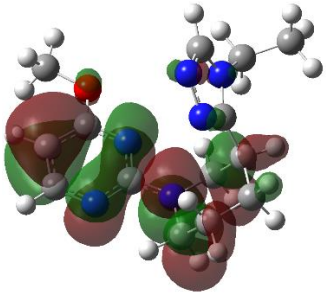
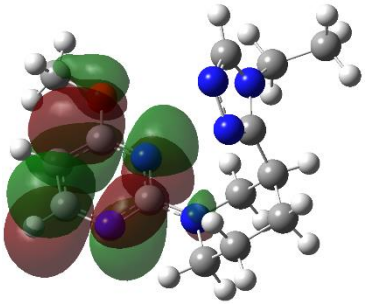
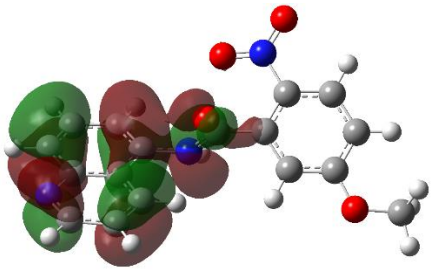
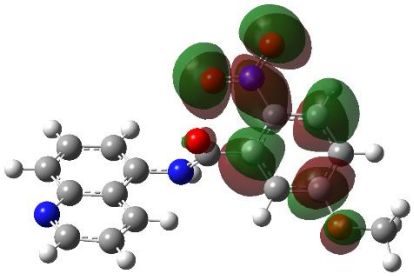
3	ZINC1648		-1121.028984 **-1121.052370**	C_1	0
4	ZINC1786		-1384.613938 **-1384.635012**	C_1	0
5	ZINC6897		-1786.069328 **-1786.096775**	C_1	0

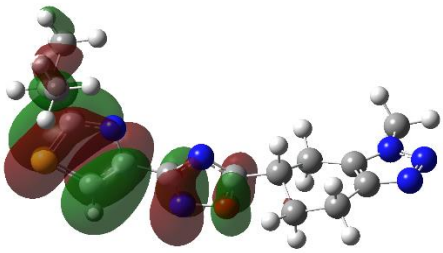
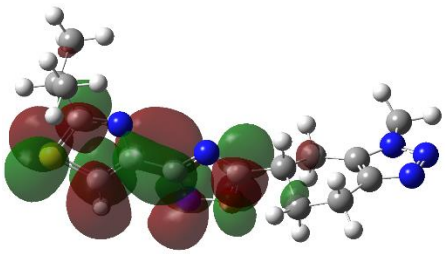
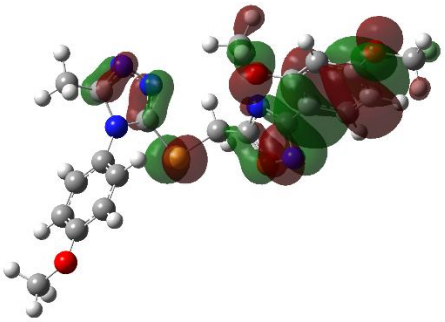
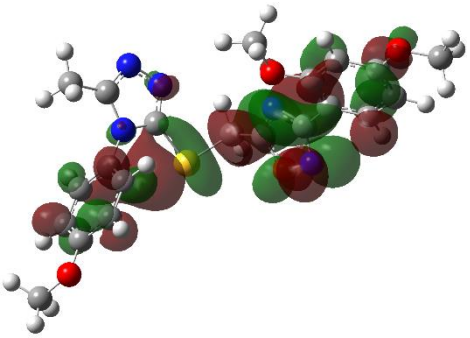
3.4.3 Molecular orbital analysis

The HOMO and LUMO were analysed to determine the electron-rich (donor) and electron-poor (acceptor) zones in each of the obtained molecules. In the case of molecule, ZINC6030, the HOMO is observed on the 1,2,4 triazole, piperidine and pyrimidine rings whereas LUMO is observed on the only 1,2,4 triazole ring of the inhibitor. For molecule ZINC8136, the HOMO is observed in piperidine and pyrimidine rings whereas LUMO located at the pyrimidine of the compound. In the case of the compound ZINC1648, HOMO is located on the quinoline ring and LUMO is observed on 4-nitro anisoline group of the molecule. For the molecule ZINC1786, HOMO and LUMO are both located at 1,2,4 oxadiazole and thiazole groups. In the case of

molecule ZINC6897, HOMO is observed at 1,2,4 oxadiazole, 1,2,4 triazole and pyrimidine rings whereas LUMO is located on two pyrimidine rings of the molecule, as shown in **Table 3.5**. Through this analysis, it was observed that the HOMO exists close to the protein active site amino acid residues Gly197, Gly199, Lys200 and Thr201 respectively.

Table 3.5: HOMO, LUMO of five Zika virus NS3 helicase screened in molecules determined at B3LYP/6–311G+(2d,p) level theory.

S. No.	Name	HOMO	LUMO
1	ZINC6030		
2	ZINC1648		
3	ZINC8136		

4	ZINC1786		
5	ZINC6897		

3.4.4 Electrostatic potential maps

For the molecules obtained through MD simulations, ESP maps were generated, which support the characterization of non-covalent interactions, including hydrogen bonding interactions, as well as the comprehension of the sites of nucleophilic and electrophilic attacks. The red area (electron-rich, negative) represents the electrophilic assault, whereas the blue area (electron-poor, positive) represents the nucleophilic attack. ESP maps are essentially related to the electron density of the molecule. The region between -3.560e-2 and +3.560e-2 a.u. is where the negative and positive electron densities are found. These sections make it easier to see the many charged regions of the molecules that were studied. The red colored region is observed close to active site residues of all the molecules in the protein which indicates that these regions play a major role for the binding of molecule at active site. The negative polarization effect was visible on the oxygen, nitrogen and sulfur atoms of the molecules as shown in **Figure 3.8**, thus representing them as the active regions in making interactions with the surrounding amino acids.

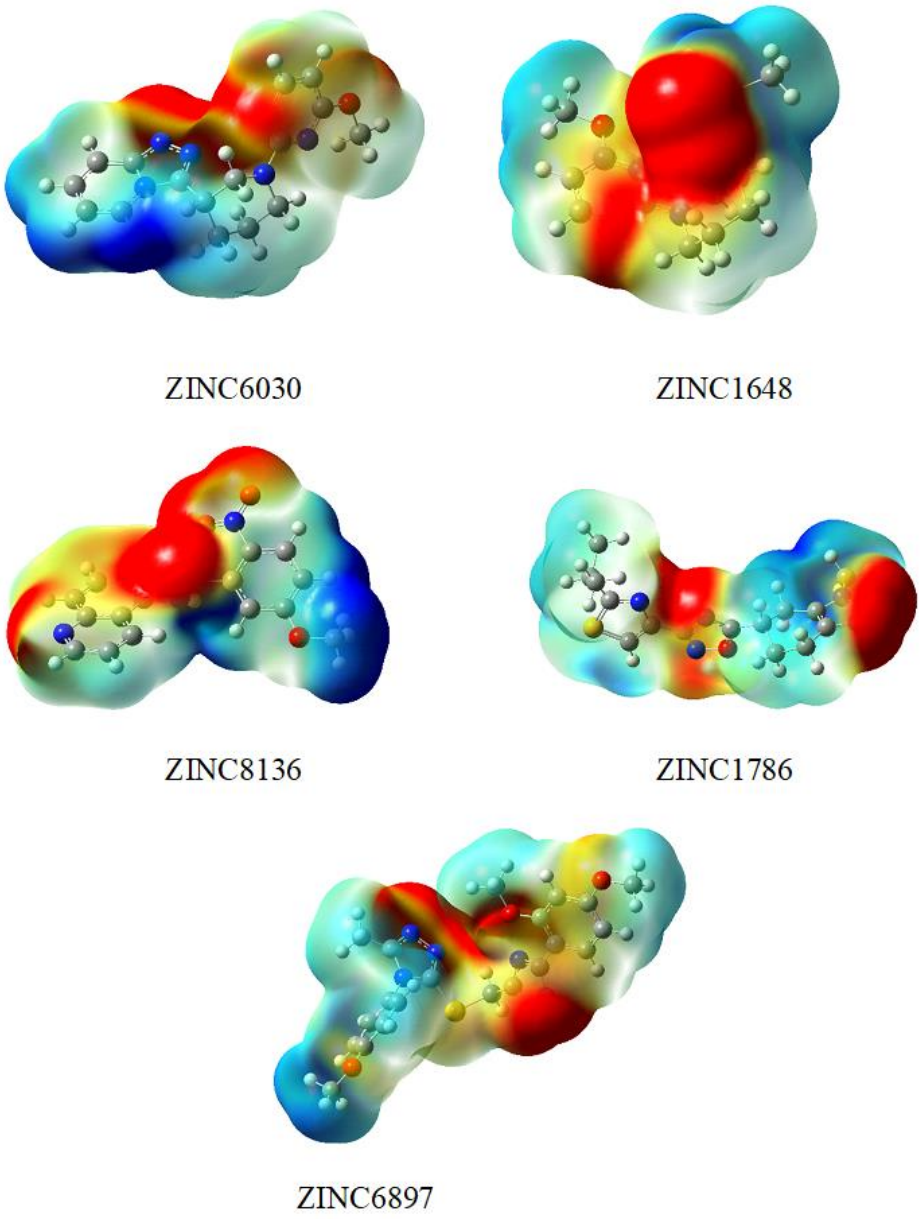


Figure 3.8: Molecular electrostatic potential (MESP) maps of five Zika virus NS3 helicase screened in molecules determined at B3LYP/6–311G+(2d, p) level of theory.

3.5 Conclusions

Sequence and structure analysis of Zika virus and flavivirus NS3 helicases, respectively revealed that the ATP and RNA binding sites are highly conserved and that the structures are highly superposable. The crystal structure of Zika virus NS3 helicase bound with ATP was explored using multiple features in pharmacophore model based on Pharmit for ZINC database screening to yield diverse leads after employing various filters that included molecular weight, RMSD and number of rotatable bonds. Molecular docking of the hit molecules and their analyses based on docking scores, intermolecular hydrogen bonding, ADME properties were applied to prioritize the molecules. The stability of Zika virus NS3 helicase bound to five hit compounds (ZINC6030, ZINC1648, ZINC8136, ZINC1786, and ZINC6897) was verified using MD simulations, and all the screened molecules show reasonable binding affinity throughout MD simulations. The DFT calculations performed on the molecules obtained through MD simulations reveal that they are stable on PES in solvent phase as compared to their gas phase energies. In light of these findings, it was concluded that creating new Zika virus NS3 helicase inhibitors to treat Zika sickness will benefit from a deeper understanding of the fundamental structural and binding properties revealed by the pharmacophore-based virtual screening.

CHAPTER-4

Zika virus NS5 methyltransferase and RNA dependent RNA polymerase inhibitor design by pharmacophore-based virtual screening, molecular docking and molecular dynamics simulations

4.1 Introduction

Zika virus NS5 MTase enzyme is responsible for adding a methyl group to the 5' cap RNA structure, using the cofactor SAM as a methyl donor to form SAH as a by-product (Zhao et al., 2015) to facilitate the translation of polyprotein. CADD is a broad and rapidly expanding research field that plays an important role in the early stages of new drug discovery. It includes analyses of the similarities between the protein sequences and structures, 3-D structure modelling of proteins, computational screening, docking of molecules, scoring of hit molecules, MD simulations and binding free energy calculations.

The 3-D structure of Zika virus NS5 MTase is reported in PDB as a complex with SAM and 7-methylguanosine-P3-adenosine-5',5'-triphosphate (GTA) (RNA analog) (5WZ2). In this chapter, analysis of primary sequences and 3-D structures of MTase enzyme from flaviviruses was carried out by sequence alignments, structure-based sequence alignments and structural comparison. The NS5 MTase comprises of a cofactor SAM bound at the active site and this site is considered as a druggable site for developing new inhibitors using the CADD approach. Based on the essential pharmacophore features of SAM, the screened molecules from the ZINC database were shortlisted by applying various filters, to be used for molecular docking studies in order to find best molecules. AutoDock 4.2 was used for docking the best molecules at the receptor active site in all possible conformations. MD simulations and binding free energy calculations were used to validate the molecules. These analyses provide an efficient method for designing novel hit

compounds and identifying key amino acid residues for intermolecular interactions in Zika virus NS5 MTase, which will be useful in future drug discovery studies.

4.2 Methods

4.2.1 Sequence and structure analyses of flavivirus MTases

PDB was used to download representative 3-D crystal structures of flavivirus NS5 MTase (Berman et al. 2000; Burley et al., 2021). UCSF Chimera was used for molecular visualization and structure superposition (Pettersen et al., 2004). BLAST searches were used to retrieve the sequences of Zika virus NS5 MTase from the NCBI (Johnson et al., 2008). The proteins multiple sequence alignment was accomplished using the ClustalW software.

4.2.2 Zika virus NS5 MTase structure and preparation

Zika virus NS5 MTase protein crystal structure complexed with SAM and GTA, a cap structure of RNA (5WZ2) deposited at 1.8 Å in PDB was used. For the preparation of the protein structure, crystal waters were deleted and hydrogen atoms were added using UCSF Chimera 2.1.1 (Petersen et al., 2004) and the coordinates were saved in PDBQT format. The cofactor SAM located in the active site was used for pharmacophore generation and to screen the molecules for further studies.

4.2.3 Structure-based pharmacophore generation

To identify novel hit molecules from *in silico* screening, the structure of SAM in Zika virus NS5 MTase was used to create the pharmacophore model for the structure-based pharmacophore generation. To generate the pharmacophore model for screening, initially, the Pharmit (Sunseri and Koes, 2016) server was used to map the possible ligand features. The probable interactions

from the binding cavity of Zika virus NS5 MTase protein with the cofactor SAM was used as a reference compound for developing interactions, and the pharmacophore features on SAM participating in nonbonding interactions with the receptor were selected. The list of five set of features of SAM were used to screen the molecules. These are hydrogen bond acceptor, hydrogen bond donor, hydrophobic, negative ionizable, and aromatic ring. The modeled pharmacophore hypothesis was used as a 3-D query for the screening of ZINC database comprising 12,996,897 molecules (Irwin & Shoichet, 2005). The database consisted of 164,282,714 exclusive conformations and were subjected to pharmacophore screening using the best generated pharmacophore model. This screening aims to retrieve the potential hit molecules suitable for further development. The successfully screened hit compounds were filtered based on RMSD (lower than 2.3 Å), molecular weight (< 500). The selected hit compounds were downloaded in .sdf format. The Lipinski's rule of five (Lipinski et al., 1997; Lipinski et al., 2012; Lipinski, 2004) are crucial factors in determining whether orally administered drugs will have a drug-like profile. The Lipinski's rule of five directs that a drug molecule exhibits LogP ≤5, molecular weight ≤ 500, number of hydrogen bond acceptors ≤10 and number of hydrogen bond donors ≤5 (Lipinski, 2004). The hit compounds which obey the above properties were selected for docking studies.

4.2.4 Molecular docking

The active site of Zika virus NS5 MTase is identified based on the key amino acid residues that interact with the cofactor SAM. The screened molecules were prepared by DS 3.5. The docking study was carried out by using AutoDock 4.2 (Morris et al., 2009). Initially, the molecules were loaded; torsions were set and saved in PDBQT format and the protein was also saved in PDBQT format. The protein structure with all the compounds was loaded into AutoDock 4.2 to generate 20 best conformations for each molecule. The Lamarckian Genetic algorithm was used

to accomplish all calculations for protein-ligand flexible docking (Wellock & Ross, 2001). A grid box with the dimensions of X: 22.987 Å, Y: 19.338 Å, and Z: 49.646 Å, with a grid spacing of 0.492 Å was used. These grid box dimensions were used to dock the filtered compounds from ZINC database in order to predict their binding affinity, binding orientation with the target protein and to analyze the best conformations. The best conformation with higher binding affinity was chosen, and hydrogen bonding interactions existing between the docked pose and protein were manually analyzed. The top-ranked molecules from docking method were further studied using ADMET calculations for their drug-likeness.

4.2.5 ADMET properties

The best docked molecular structures drug-like properties were examined by analyzing the pharmacokinetics profile with the SwissADME server (http://www.swissadme.ch/index.php). This is a software tool to calculate ADME and physicochemical properties such as solubility, lipophilicity and pharmacokinetics. MD simulations were used to study the hit molecules that meet the criteria of ADME properties.

4.2.6 Molecular dynamics simulations

MD simulations of apo and complexes of Zika virus NS5 MTase was carried out for 100 ns, using GROMACS 5.1.4 (Hess et al., 2008; Van Der Spoel et al., 2005) software. Four types of molecular systems were prepared; apo Zika virus NS5 MTase, Zika virus NS5 MTase complexed with SAM, Zika virus NS5 MTase complexed with SAM and GTA, Zika virus NS5 MTase complexed with hit molecules. During the preparation of molecular systems, water molecules and all other hetero atoms were removed from the protein, hydrogens were added and Amber99sb force field was applied to all the molecular systems. The applicable charges were added to SAM and hit

molecules using UCSF Chimera-2.1.1, and the coordinates were saved in .mol2 format to generate force fields using AM1-BCC charge in Antechamber using ACPYPE script (Da Silva et al., 2012). SPC water models were used to solvate the molecular systems, by adding Na^+ and Cl^- ions, such that the total charge of the systems was neutralized. Energy minimization was accomplished in 1000 steps that use the steepest descent algorithm to remove steric stress and enable the system to relax. Water molecules and ions were distributed during the position restraint phase. The systems were heated for 100 ps at 300 K before being equilibrated at 1 atm and 300 K for 1000 ps until the proper density was achieved. The temperature was kept constant using a V-rescale thermostat (Bussi et al., 2007), and the pressure was controlled using the Parrinello Rahman technique (Parrinello & Rahman, 1981). The particle mesh Ewald method was used to deal with long-range electrostatic interactions (Darden et al., 1993; Essmann et al., 1995). For 1000 ps, molecular systems were equilibrated using the NVT and NPT ensembles. The Lennard-Jones interactions and the real-space electrostatic interactions were truncated at 9 Å. LINCS algorithm was used to restrict hydrogen bonds (Hess et al., 1997). After every 2 ps, the coordinates from the production MD trajectories were generated and saved. Upon structure stabilization, the final models in all systems were obtained by averaging the snapshots from the output trajectories of MD simulations. The GROMACS utilities gmx rms, gmx rmsf, gmx hbond, and gmx gyrate were used to analyse the MD trajectories (Van Der spoel, et al., 2005). The RMSD of the Cα-atoms in relation to their starting structures was calculated in GROMACS. VMD was used to display the MD trajectory files (Humphrey et al., 1996). The Xmgrace software was employed for plotting the data obtained from MD simulations analysis, and UCSF Chimera (Pettersen et al., 2004) was employed for structure superposition and cartoon image generation.

4.2.7 Binding free energy

The molecular mechanics levels of energy combined with the MM-PBSA methods are widely used to estimate the free energy of small ligand molecules binding to biological macromolecules. The MM-PBSA method (Kumari et al., 2014) was used to calculate the binding free energy of protein-ligand complexes using g_mmpbsa tools. The output trajectories of MD simulations were used to calculate the binding free energy of protein-ligand complexes. The output results obtained by this g_mmbsa method are van der Waals, electrostatic, polar solvation, SASA, and binding free energies.

4.3 Results and discussions

4.3.1 Sequence and structure analyses

The multiple sequence alignment of flavivirus NS5 MTase (**Figure 4.1A**) reveals that all the amino acid residues within 5 Å from SAM (5WZ2) binding site are highly conserved among all flaviviruses. The structures are highly superposable with a low RMSD indicating their high structure conservation as shown in **Figure 4.1B.**

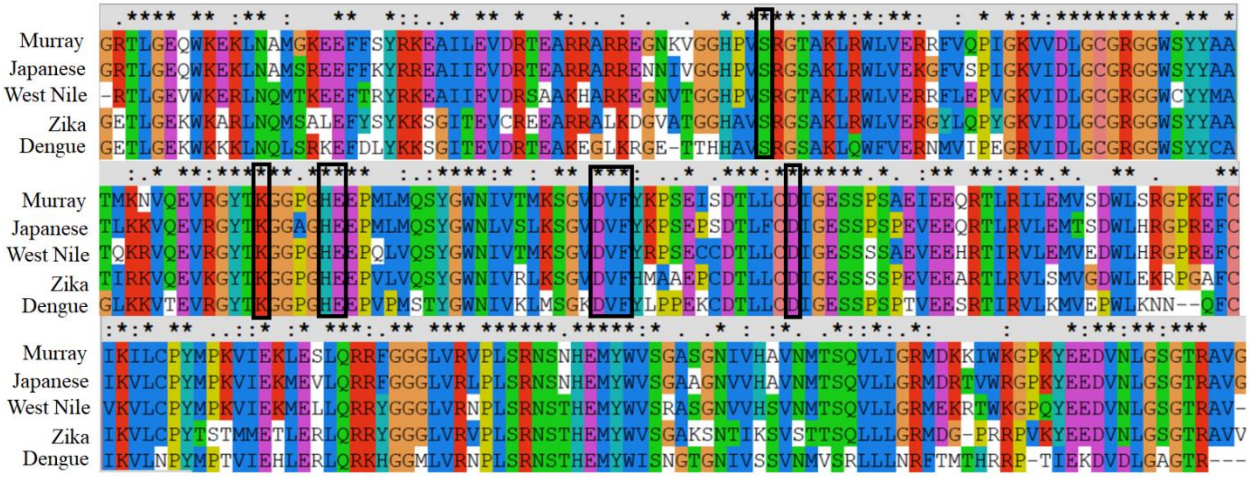


Figure 4.1A: Multiple sequence alignment of Zika and other flavivirus NS5 MTase proteins.

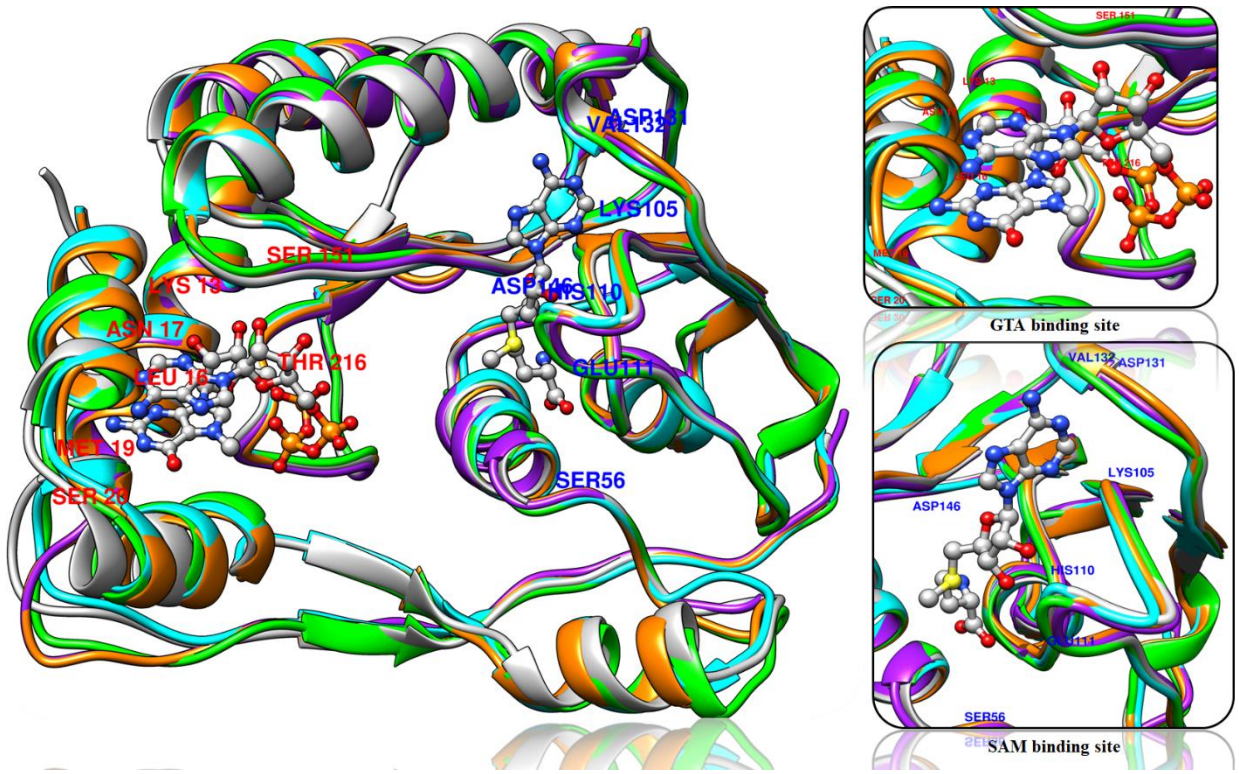


Figure 4.1B: Structural alignment of Zika and other flavivirus NS5 MTase proteins Zika (grey), Dengue (cyan), Japanese encephalitis (orange) and Murray valley encephalitis (purple).

In the Zika virus NS5 MTase crystal structure (5WZ2), complexed with cofactor SAM and GTA, the cofactor binding cavity is considered as the active site of the protein. For validation of the docking methodology, SAM was redocked into its binding site using the AutoDock 4.2. Twenty conformers were generated, out of these, the final docked poses were selected based on the binding affinity and intermolecular interactions. The essential interactions were conserved with residues within the active site of the protein. The redocked SAM was identified to be in the vicinity of crucial amino acid residues Ser58, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, Glu113, Val132, Asp133, Val134, Phe135, Asp148,

Ile149, Gly150, Glu151, Lys184 and Tyr222 and showed RMSD of 0.16 Å with the crystal structure conformation. The hydrogen bonding interactions were formed with the residues Ser58, Lys107, His112, Glu113, Asp133, Val134, Asp148. The SAM docked into the active site showed binding affinity -8.4 kcal/mol.

4.3.2 Pharmacophore generation and virtual screening

The structure-based pharmacophore model was generated to understand the key features necessary for biological functions. The intermolecular interactions between the hit molecule and the catalytically important active site amino acid residues provide appropriate information as inputs for the design of the structure-based pharmacophore model for the macromolecular receptor Zika virus NS5 MTase. Pharmacophore modeling using the Pharmit server identified the features of SAM required for nonbonded interactions with the amino acid residues available in its binding site. The five pharmacophore features selected in SAM for virtual screening are; hydrogen bonding interaction of the N6 atom of NH_2 on adenine ring with Asp133, N1 atom on adenine ring with Val134, O2' atom of OH on 5-memberd ring moiety with Glu113, the N atom on methionine NH_2 with Asp148 and the O atom from methionine carboxylate group with Ser58, shown in **Figure 4.2**. The pharmacophore model was validated on 1000 molecules decoy set (inactive molecules) using Pharmit server. The pharmacophore model could not predict a single decoy molecule, thus indicating that the pharmacophore model successfully qualified the validation process. The validated pharmacophore model was employed for virtual screening of small molecules in the ZINC database. The five featured pharmacophore model that was used as a 3D query retrieved 609 hit molecules from 12,996,897 molecules in the ZINC database and were downloaded as .sdf file. These hit molecules were filtered and ranked based on lower RMSD (< 0.23 Å) and the Lipinski's

rule was applied in prefiltering process. This retrieved 92 molecules that were used for molecular docking studies by AutoDock 4.2.

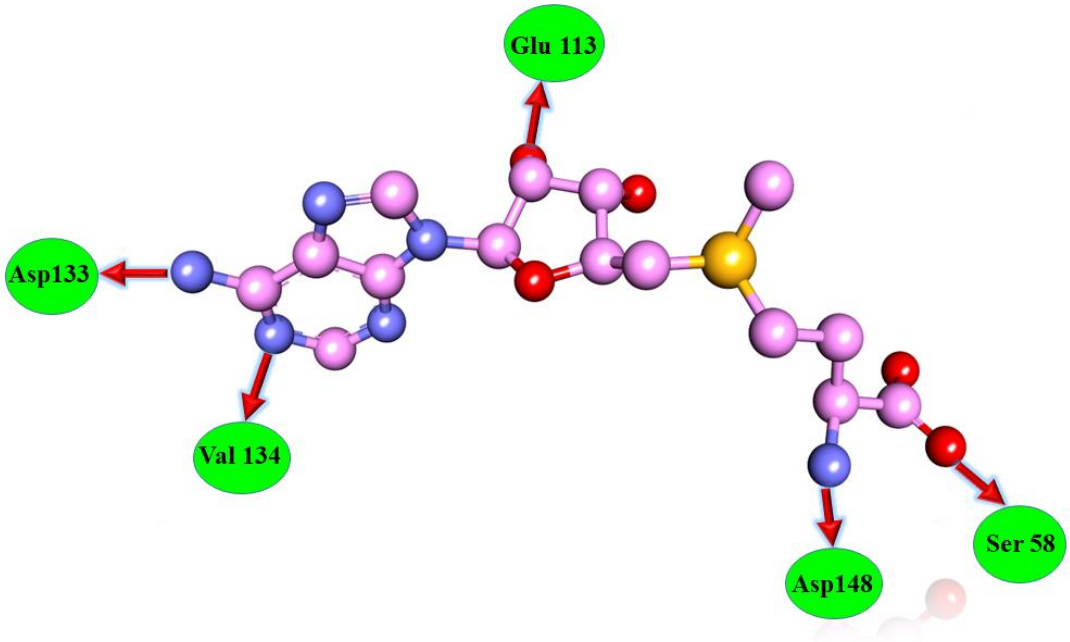
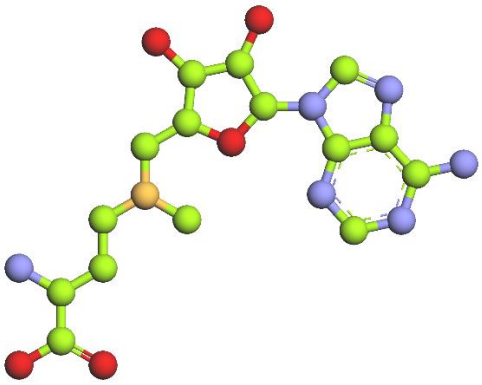
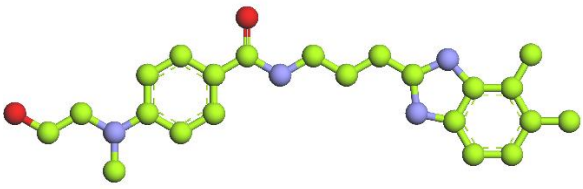


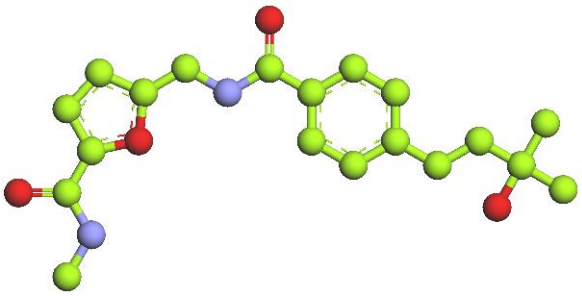
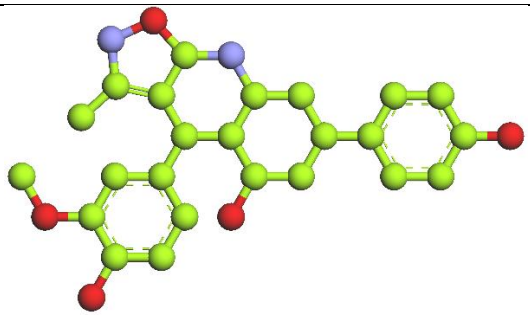
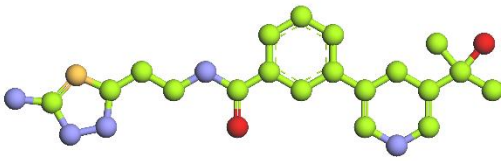
Figure 4.2: Pharmacophore features of SAM used for virtual screening of ZINC database.

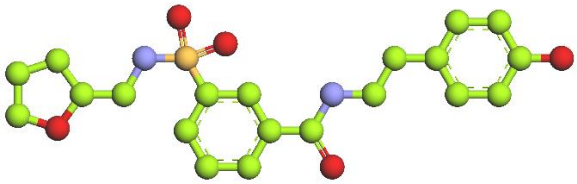
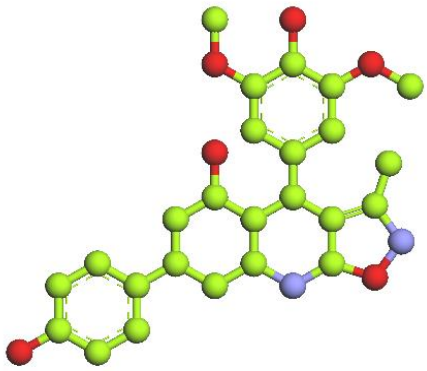
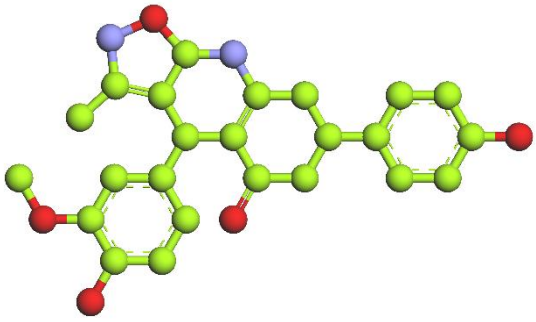
4.3.3 Molecular docking

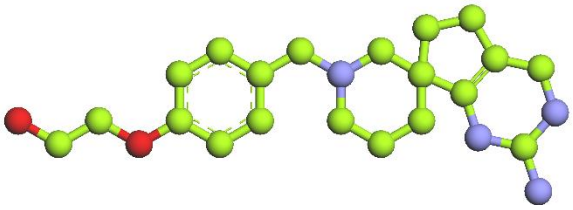
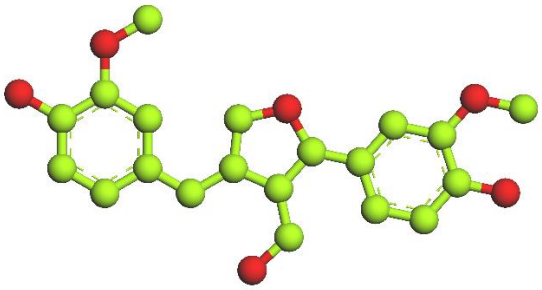
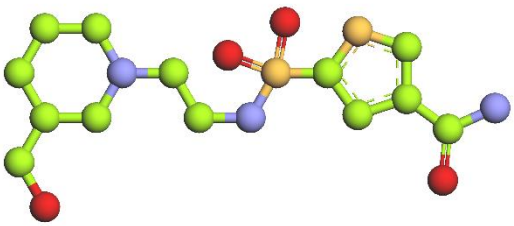
The 92 hit molecules identified from virtual screening were docked into the SAM binding cavity of Zika virus NS5 MTase. For each molecule, 20 docking poses were generated and the best docking pose was selected based on the binding affinity and hydrogen bonding interactions maintained by key amino acid residues in the SAM binding cavity. The selected 92 hit molecules were docked with a binding affinity within the range of -9.0 to - 6.8 kcal/mol. Of these, 50 hit molecules had a binding affinity higher than -7.8 kcal/mol and 15 hit molecules had a binding affinity between -9.0 to -8.0 kcal/mol. The binding affinities and the amino acid residues in the binding pocket are shown in **Table 4.1**.

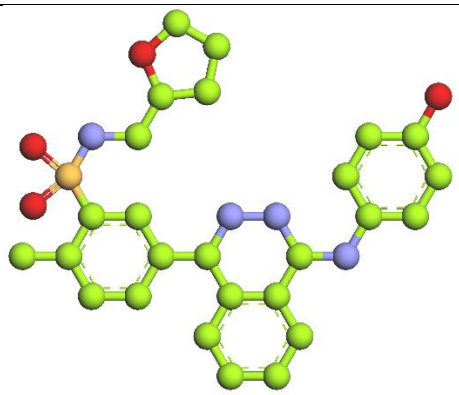
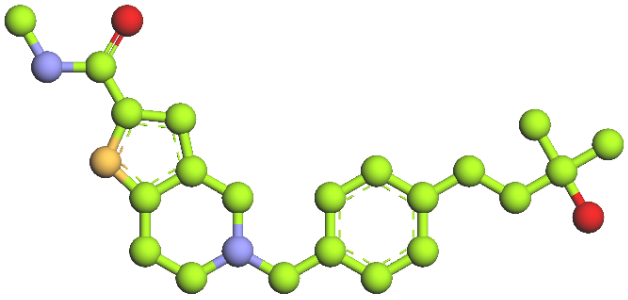
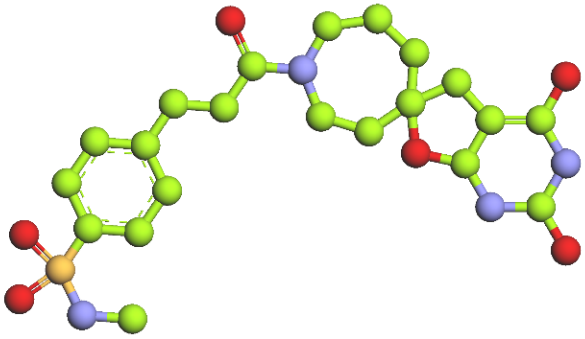
Table 4.1. Docking scores of SAM, screened hit molecules along with the interacting residues in the Zika virus NS5 MTase.

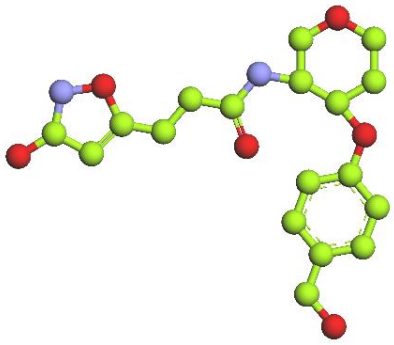
S. No	Structure and Name	Docking score (kJ/mol)	Binding site residues
1.	 **SAM** S-adenosyl-L-methionine	-8.4	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, **Lys107**, Gly108, **His112**, **Glu113**, Val132, **Asp133**, **Val134**, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222
2.	 **hit-1** N-(3-(4,5-dimethyl-1H-benzo[d]imidazol-2-yl)propyl)-4-((2-hydroxyethyl)(methyl)amino)benzamide	-8.5	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, Val132, Asp133, Val134, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222

3.	 **hit-2** 5-((4-(3-hydroxy-3-methylbutyl)benzamido) methyl)-N-methylfuran-2-carboxamide	-8.4	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, Val132, Asp133, **Val134**, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222
4.	 **hit-3** (4S,7R)-4-(4-hydroxy-3-methoxyphenyl)-7-(4-hydroxyphenyl)-3-methyl-4,6,7,8-tetrahydroisoxazolo[5,4-b]quinolin-5-ol	-8.2	Ser58, Arg59, Gly60, **Lys63**, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, Val132, Asp133, **Val134**, Phe135, Asp148, Ile149, Gly150, Glu151, **Lys184** and Tyr222
5.	 **hit-4** N-(2-(5-amino-1,3,4-thiadiazol-2-yl)ethyl)-3-(5-(2-hydroxypropan-2-yl)pyridin-3-yl)benzamide	-8.5	Ser58, Arg59, Gly60, Lys63, Leu82, Gly83, **Cys84**, Gly85, Arg86, **Gly87**, Thr106, Lys107, Gly108, His112, Glu113, **Val132**, **Asp133**, Val134, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222

6.	 **hit-5** (R)-N-(4-hydroxyphenethyl)-3-(N-((tetrahydrofuran-2-yl)methyl)sulfamoyl)benzamide	-9.0	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, **Val132**, Asp133, Val134, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222
7.	 **hit-6** (4S,7S)-4-(4-hydroxy-3,5-dimethoxyphenyl)-7-(4-hydroxyphenyl)-3-methyl-4,7,8,9-tetrahydroisoxazolo[5,4-b]quinolin-5(6H)-one	-8.3	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, Glu113, Val132, Asp133, **Val134**, Phe135, Asp148, **Ile149**, Gly150, Glu151, Lys184 and Tyr222
8.	 **hit-7** (4S,7S)-4-(4-hydroxy-3-methoxyphenyl)-7-(4-hydroxyphenyl)-3-methyl-4,7,8,9-tetrahydroisoxazolo[5,4-b]quinolin-5(6H)-one	-8.4	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, **Lys107**, Gly108, His112, Glu113, Val132, **Asp133**, Val134, Phe135, Asp148, Ile149, Gly150, Glu151, Lys184 and Tyr222

9.	 **hit-8** (S)-2-(4-((2-amino-5,6-dihydrospiro[cyclopenta[d]pyrimidine-7,3'-piperidin]-1'-yl)methyl)phenoxy)ethan-1-ol	-8.2	**Ser58**, Arg59, Gly60, **Lys63**, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, **Lys107**, Gly108, His112, Glu113, Val132, Asp133, Val134, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222
10.	 **hit-9** 4-((2S,3R,4S)-4-(4-hydroxy-3-methoxybenzyl)-3-(hydroxymethyl) tetrahydrofuran-2-yl)-2-methoxyphenol	-8.3	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, Val132, Asp133, **Val134**, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222
11.	 **hit-10** (S)-5-(N-(2-(3-(hydroxymethyl)piperidin-1-yl)ethyl)sulfamoyl)thiophene-3-carboxamide	-8.0	Ser58, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, **Lys107**, Gly108, His112, **Glu113**, Val132, Asp133, **Val134**, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222

12.	 **hit-11** (S)-5-(4-((4-hydroxyphenyl)amino)phthalazin-1-yl)-2-methyl-N-((tetrahydrofuran-2-yl)methyl)benzenesulfonamide	-8.6	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, Val132, **Asp133**, **Val134**, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222
13.	 **hit-12** 5-(4-(3-hydroxy-3-methylbutyl)benzyl)-N-methyl-4,5,6,7-tetrahydrothieno[3,2-c]pyridine-2-carboxamide	-8.5	Ser58, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, **Lys107**, Gly108, His112, Glu113, Val132, Asp133, Val134, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222
14.	 **hit-13**	-8.1	**Ser58**, Arg59, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, Val132, Asp133, **Val134**, Phe135, Asp148, Ile149, Gly150, Glu151, Lys184 and Tyr222

	(S)-4-(3-(2',4'-dihydroxy-5'H-spiro[azepane-4,6'-furo[2,3-d]pyrimidin]-1-yl)-3-oxopropyl)-N-methylbenzenesulfonamide		
15.	 **hit-14** 3-(3-hydroxyisoxazol-5-yl)-N-((3R,4R)-4-(4-(hydroxymethyl)phenoxy)tetrahydro-2H-pyran-3-yl)propanamide	-8.0	Ser58, **Arg59**, Gly60, Lys63, Leu82, Gly83, Cys84, Gly85, Arg86, Gly87, Thr106, Lys107, Gly108, His112, **Glu113**, Val132, **Asp133**, Val134, Phe135, **Asp148**, Ile149, Gly150, Glu151, Lys184 and Tyr222

4.3.4 Drug-like properties

The molecules selected from molecular docking were subjected to the next stage of ADME calculations using the SwissADME server and the results are shown in **Table 4.2**, and selected the molecules based on acceptable values within the range. The synthetic ease of access of the suggested compounds is graded on a scale of one to ten, on the basis of the complexity of the molecules, the number of stereocenters, and other factors. A lower value is indicative of greater synthetic accessibility of the molecule and the synthetic accessibility value was found to be less than 5.1, which indicates the ease of their synthesis. The TPSA was lower than 140 Å, the extent of solubility in water, stated as Log *S* indicated that most of the hit molecules are soluble or reasonably soluble in water; and lipophilicity expressed as Log K_P was lower than 3.5. The

possibility of skin permeation expressed as Log Kp was also reasonable. Out of the 15 selected molecules from docking studies, 12 molecules successfully qualified all the ADME properties within the range of acceptable values. These 12 molecules were further studied by MD simulations. The best Zika virus NS5 MTase hit molecule complexes were used as initial conformations for MD simulations, and complex with SAM and apo protein were also studied for the sake of comparison.

Table 4.2. Drug-like properties of screened hit compounds.

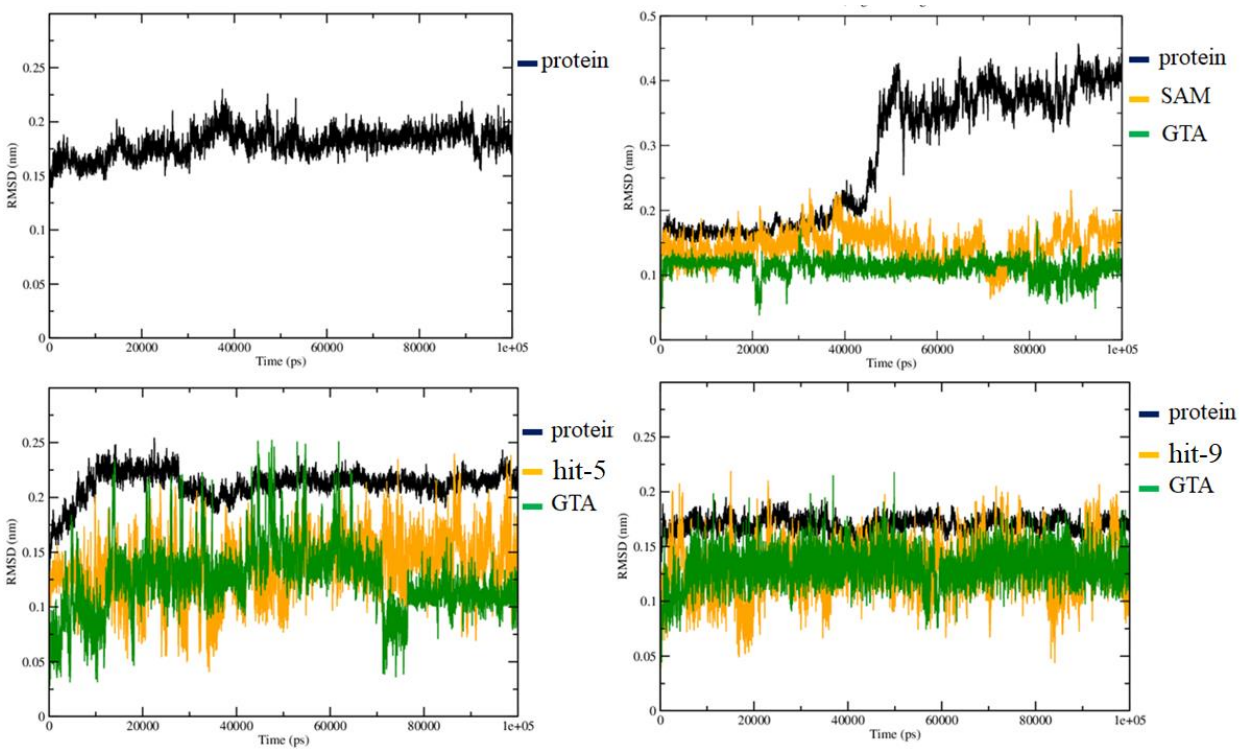
S. No	**Compound**	**TPSA ($Å^2$)**	**Consensus Log $P_{o/w}$**	**Log *K*p**	**Synthetic accessibility**
1.	hit-1	96.45	1.64	-7.85	4.88
2.	hit-2	122.70	2.24	-7.10	5.08
3.	hit-3	109.36	2.81	-5.53	4.04
4.	hit4	88.09	1.86	-7.33	3.27
5.	hit-5	87.00	1.81	-7.30	4.04
6.	hit-6	109.28	0.57	-8.75	4.19

7.	hit-7	109.28	2.34	-7.37	5.94
8.	hit-8	130.08	0.16	-8.75	4.83
9.	hit-9	88.38	2.38	-6.79	3.94
10.	hit-10	104.73	1.68	-7.17	2.27
11.	hit-11	124.32	3.46	-6.46	4.44
12.	hit-12	121.67	1.74	-7.36	3.46

4.3.5 Molecular dynamics simulations

The stability and dynamic characteristics of the protein-hit molecule complexes were determined using atomistic MD simulations with explicit solvent. MD simulations reveal detailed information about protein-hit molecule interactions as a course of the simulations time that contribute to their stable binding state, as well as the changes in the conformational states of protein upon binding to the hit molecule. The apo and Zika virus NS5 MTase complexed with all the 12 selected hit molecules were carried out for 100 ns MD simulations. Out of these 12 hit molecules, three hit molecules (hit-5, hit-9 and hit-11) obtained stability throughout MD simulations. Each system within the initial and average structures were compared and then analyzed the conformational changes of protein and the stability of hit molecules that bind to the protein. The RMSD plots describe the extent of deviation of all atom positions in the apo and protein-hit

complexes throughout MD simulations that demonstrated the protein stability throughout MD simulations. It was observed that the screened hit molecules from ZINC database show good stability when binding with Zika virus NS5 MTase protein and GTA molecule shows significant stability throughout MD simulations. Based on RMSD plots, it was observed that the RMSD of apo is 1.8 Å; when complexed with SAM and hit5 (~ 2.1 Å) and complexes with hit9 and hit11 (< 1.8 Å) as shown in **Figure 4.3A**. The hit molecules and SAM also show lower than 2 Å RMSD. From RMSF plots, it is seen that the residues from loop regions (Arg43-Thr52, Arg177-Phe181 and Val253-Glu256) show greater fluctuations during MD simulations (**Figure 4.3B)**.



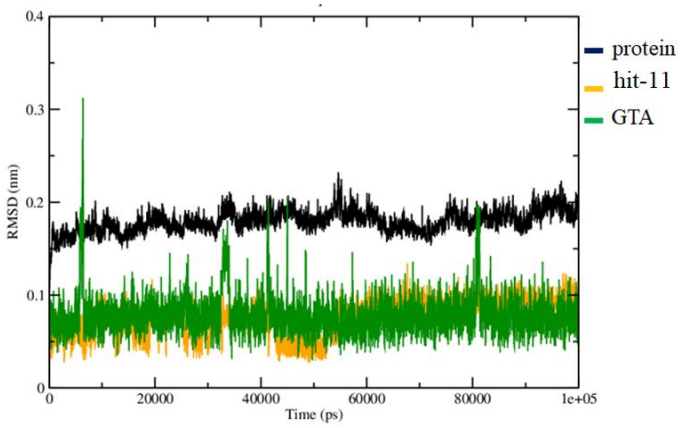


Figure 4.3A: RMSD plots of Zika virus NS5 MTase bound with SAM, GTA and screened hit molecules.

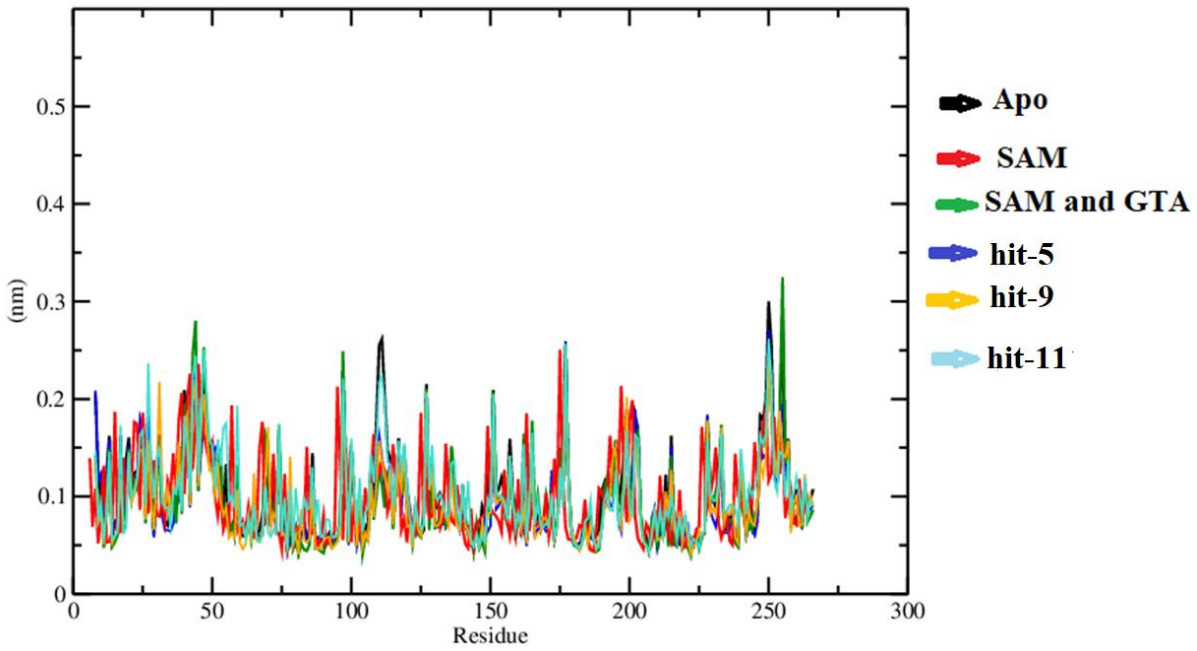


Figure 4.3B: RMSF plot of Zika virus NS5 MTase bound with SAM, GTA and screened hit molecules.

4.3.6 Hydrogen bonding interactions

One of the essential criteria for stable protein-ligand complex formation is hydrogen bonding interactions, which hold the molecule tightly in the protein active site. Intermolecular interactions in protein reference and screened hit molecules structures before and after MD simulations showed that SAM makes hydrogen bonding interactions with Ser56, Glu111, Val132 and Asp131 throughout MD simulations. The three hit molecules (hit-5, hit-9 and hit-11) also make hydrogen bonding interactions with the active site residues (Val57, Ser58, Glu113, Val134 and Asp133). In hit-5 the oxygen atom on phenol group shows hydrogen bonding with Asp133 and Val134, the oxygen atom on C=O of amide group shows hydrogen bonding with Glu133 and O atom on sulfate group make hydrogen bonding with Arg86. In hit-9 OH of methoxy phenol make hydrogen bonding with Ser86, the O atom of methoxy group forms hydrogen bonding with Arg86 and His112, the O atom of furan exhibits hydrogen bonding with Glu113. In hit-11 the O and N atoms of benzene sulfonamide show hydrogen bonding interactions with Ser58, Arg86 and the N of phthalazine shows hydrogen bonding with Glu113, the O atom of phenyl group shows hydrogen bonding with Asp133 and Val132 as shown in **Figure 4.4**, indicating that these interactions stabilize the complex formation.

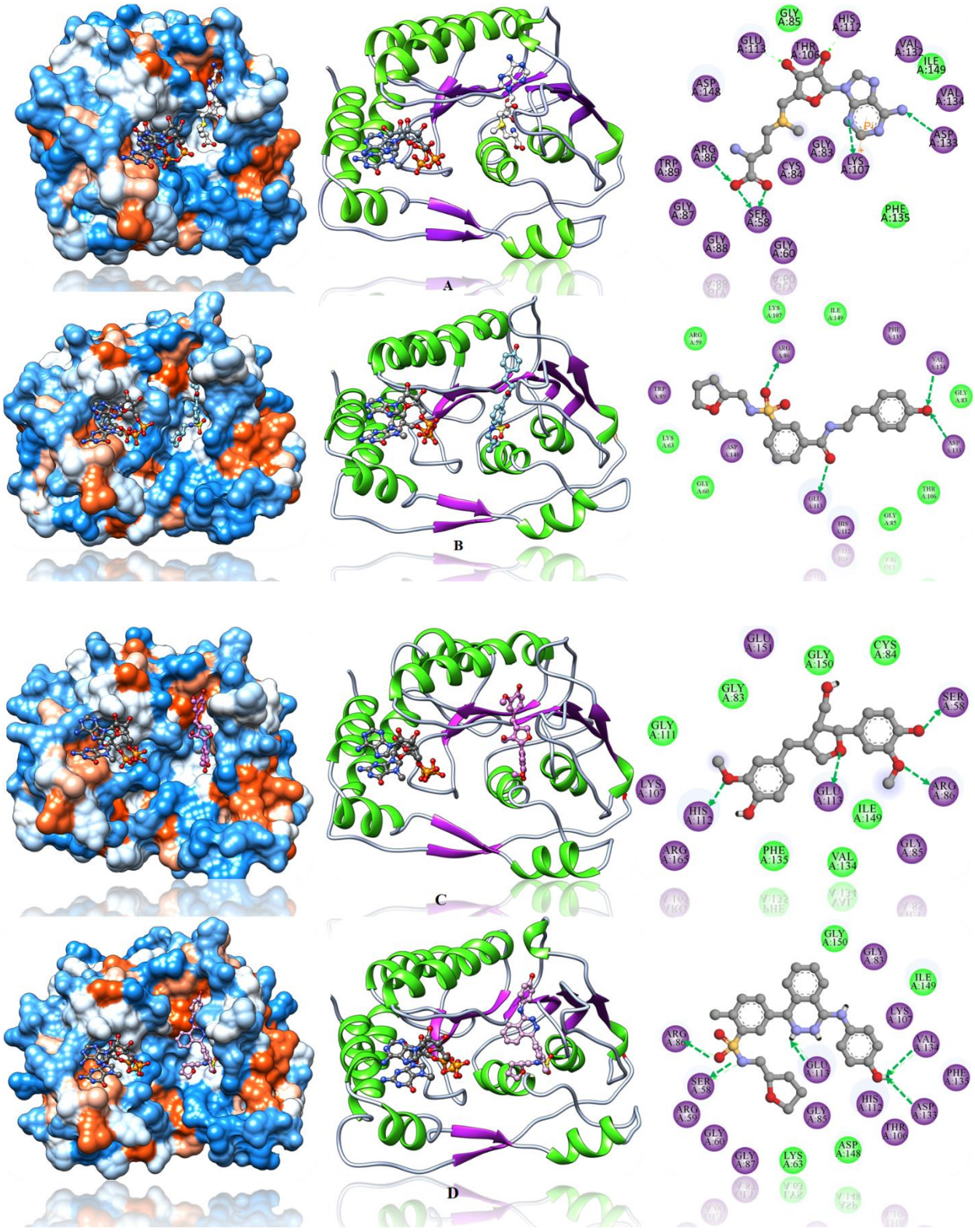

A
B
C
D

Figure 4.4: Interaction of residues Zika virus NS5 MTase complexed with hit molecules throughout MD simulations **A**) NS5 MTase with SAM, **B**) NS5 MTase with hit-5, **C**) NS5 MTase with hit-9 and **D**) NS5 MTase with hit-11.

4.3.7 Binding free energy calculations

The binding free energies of reference and screened hit compounds in this work were analysed from MM-PBSA calculations as shown in **Table 4.3**. The contribution from electrostatic, van der Waals and polar solvation energies indicate agreement with the reference molecule. The binding free energies were observed to be SAM (-70.66 kJ/mol), hit-5 (-67.41 kJ/mol), hit-9 (-70.57 kJ/mol) and hit-11 (-62.19 kJ/mol).

Compound Name	**van der Waals energy (kJ/mol)**	**Electrostatic energy (kJ/mol)**	**Polar solvation (kJ/mol)**	**SASA (kJ/mol)**	**ΔG (kJ/mol)**
SAM	-192.22+/0.700	-46.310+/0.418	185.74+/0.849	-17.89+/0.044	-70.66+/0.719
hit-5	-135.195+/0.849	-24.110+/1.012	107.495+/1.407	-15.59+/0.068	-67.41+/1.249
hit-9	-168.16+/0.516	-33.36+/0.664	147.534+/0.724	-16.59+/0.050	-70.58+/0.561
hit-11	-146.26+/0.647	-56.01+/0.548	156.882+/1.158	-16.790+/0.061	-62.198 +/- 0.77

In order to explain the importance of active site amino acid residues in binding to the ligands, their contribution to the binding energy was measured. The binding site residues Ser58, Glu113, Asp133 and Val134 show the highest binding energy contribution, that is attributed to the hydrogen bonding interactions observed in the residues while interacting with SAM and screened hit molecules as shown in **Table 4.4**. From this analysis of binding energy calculations, it is

suggested that hit molecules show good binding affinity with NS5 MTase and are therefore its likely inhibitors.

Table 4.4: Contribution of amino acid residues in the Zika virus NS5 MTase protein to the binding of SAM and screened hit molecules.

S. No	**Amino acid residues**	**SAM (kJ/mol)**	**hit-5 (kJ/mol)**	**hit-9 (kJ/mol)**	**hit-11 (kJ/mol)**
1	Ser 56	-4.84	-2.45	-2.38	-10.32
2	Arg57	-1.06	-3.35	-1.50	-1.06
3	Glu111	-37.21	-4.29	-0.18	-3.96
4	Asp131	-36.99	-4.44	-0.18	-9.94
5	Val132	-5.88	-4.97	-6.52	-3.74
6	Phe133	-1.93	-3.12	-8.39	-3.38
7	Asp146	-14.51	-1.81	-2.96	-14.32
8	Ile147	-28.24	-19.41	-15.43	-22.36

4.4 Conclusions

CADD techniques can be employed effectively to accelerate the development of therapeutic agents for Zika virus disease. The cofactor, SAM was used for pharmacophore based virtual screening of ZINC database. To investigate the binding interactions among protein and hit

molecule complexes, molecular docking and MD simulations were used. The stability of the Zika virus NS5 MTase bound to three hit molecules (hit-5, hit-9, and hit-11) was verified by using MD simulations followed by binding free energy calculations. Overall, the results indicated that the pharmacophore-based screening provided significant information towards comprehension of the essential structural binding features when developing new Zika virus inhibitors to treat Zika infection.

4.5 Zika virus NS5 RNA dependent RNA polymerase

4.5.1 Introduction

The RdRp synthesizes the genome of RNA in the absence of primer strand called as *de novo* mechanism where in single nucleoside triphosphate is provided as a primer for nucleoside polymerization to form a double-stranded RNA. The flavivirus RdRp incorporates a functional nuclear localization sequence region, which is important for viral and host protein interactions. NS5 interacts with the NS3 protease helicase as well as a number of host proteins (Yap et al., 2007; Tay et al., 2015). The Zika virus NS5 protein serves as an antagonist due to its enzymatic functions. That is because the NS proteins NS5 RdRp plays an important role in viral genome replication. The 3-D structure of Zika virus NS5 RdRp is already reported in PDB. This protein has three domains, called as finger (residues 321-488), palm (residues 481-541 and 609-714) and thumb (residues715-903) (**Figure 4.5)**. The structure of Zika virus NS5 RdRp has prominent similarities with the Japanese encephalitis (JEV) NS5 RdRp, with the RMSD of 0.55 Å, which is binding with cofactor GTP at the active site. The missing residues in the crystal structure of Zika virus NS5 RdRp were built on the JEV NS5 RdRp structure. The active site of protein was defined by

template structure JEV NS5 RdRp (PDB 4HDG) (Surana et al., 2014), extracted the GTP molecule into Zika virus NS5 RdRp and this site was considered as an active site for developing new inhibitors using the CADD approach. Based on the essential features of GTP, molecules were screened from the Asinex database by applying various filters in order to find best molecules for molecular docking studies. The CDOCKER docking methodology was used for docking the molecules at the receptor active site in all possible conformations. The molecules were validated using MD simulations and binding free energy calculations.

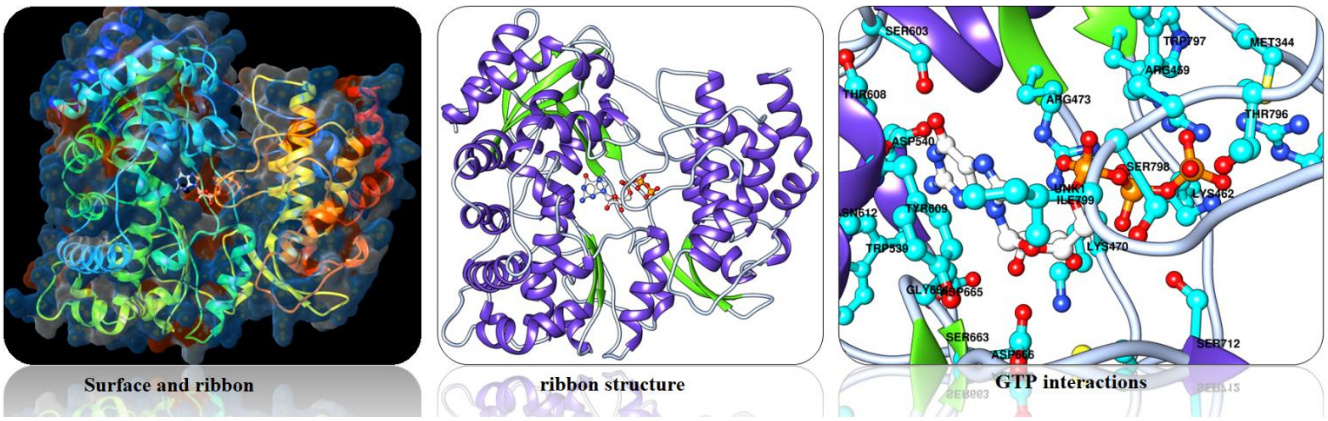


Figure 4.5: Structure of Zika virus NS5 RdRp binding with GTP.

4.5.2 Materials and methods

4.5.2.1 NS5 RNA dependent RNA polymerase structure and preparation

Zika virus NS5 RdRp protein crystal structure is already reported at 1.8 Å (Duan et al., 2017). During preparation of protein the protein were subjected to addition of missing residues and hydrogens. The water molecules and hetero atoms were removed by using UCSF Chimer 2.1.1 (Peterson et al., 2004). The 3-D structure of Zika virus NS5 RdRp was submitted to the Dali server to find the template structure to build the missing residues. Dali was used to compare the residues one-to-one and the method removes the amino acids with a mismatch between the two protein

structures under comparison. This server performs three types of structure associations: (i) The PDB search to compare one query structure with the members in the PDB to produce a list of proteins that share similar structure. (ii) A pair-wise comparison to compare the protein structure of interest with a set of 3-D structures that are provided by the user; (iii) Comparison of all structures against all to generate a structural similarity matrix, a dendrogram, and a multi-dimensional scaling projection of the set of structures defined by the user. The Dali server was also used to perform PDB search comparison to find the protein structures similar to the query (Zika virus NS5 RdRp) protein structure where the active site of a protein is known and provides the crystal structure of various flaviviruses which are closely related to Zika virus NS5 RdRp protein. The JEV NS5 RdRp (4HDG) was selected as the template structure for building the missing residues in Zika virus NS5 RdRp.

4.5.2.2 Building of missing residues in 3-D structure of Zika virus NS5 RNA dependent RNA polymerase

The missing residues in Zika virus NS5 RdRp were constructing by using template structure JEV NS5 RdRp. By using the sequence alignment, a 3-D model of Zika virus NS5 RdRp with the complete model was constructed by using MODELLER 9.17 (Šali & Blundell, 1993) based on 4HDG. The generated models were evaluated based on their DOPE scores, with the highest decreasing trend of DOPE score being ranked first. The template structure JEV NS5 RdRp (4HDG) has cofactor GTP present in the active site. The two structures Zika virus NS5 RdRp (5WZ3) and JEV NS5 RdRp (4HDG) were superposed to extract the GTP coordinates into modeled structure of Zika virus NS5 RdRp protein.

4.5.2.3 Structure based drug design

To increase the effectiveness of *in silico* screening in identifying active and novel hits, the redocked cofactor GTP into Zika virus NS5 RdRP protein was used for the structure-based pharmacophore generation. In this study, the Asinex database with 500,000 exclusive structure records was screened using a set of pharmacophore features.

4.5.2.4 Pharmacophore generation

After docking the GTP into active site of Zika virus NS5 RdRp protein, this structure was used for the pharmacophore generation to identify the novel inhibitors development for Zika virus NS5 RdRp. To develop the pharmacophore model for screening, initially, probable interactions from the active site of Zika virus NS5 RdRp protein were generated by employing the interaction generation protocol available in DS 3.5. The cofactor GTP, was used as a reference molecule to generate intermolecular interactions. The parameters for both density of lipophilic sites and polar sites were defined. Feature mapping tool available in DS 3.5 pharmacophores protocol was used to map the feasible ligand features.

4.5.2.5 Library screening

The Asinex database was used in this study which comprises 500,000 exclusive structure records were transferred to “Screen Library” protocol of DS 3.5 (Kabsch, 1976). In order to identify the hit molecules, Asinex database was screened by using the best pharmacophore model with preferred pharmacophore features and scored them on the basis of their fit value. In the Screen Library protocol DS 3.5, a minimum of two and a maximum of six pharmacophore features were chosen using a rigid fitting method of the pharmacophore and the small molecule. Based on their pharmacophore fit value, the molecules were selected.

4.5.2.6 Molecular docking

Molecular docking is an important application to identify the mechanism of protein-inhibitor binding. Once the active site of a protein based on the key amino acid residues is identified different ligand molecules can be docked into the active site of the protein such as Zika virus NS5 RdRp. DS 3.5 was used to prepare the ligand molecules, hydrogens are added, and the molecules were energy minimised using the CHARMm force field (Brooks et al., 1983). The new inhibitors were docked using CDOCKER. The temperature of the system heated up to 700 K for 2,000 steps and then cooled to 300 K for 5,000 steps. The protein active site was described based on GTP binding position and extended to the residues within its 5Å cut-off distance, with a total of 20 docking poses. To test the reliability of the CDOCKER docking protocol, the cofactor GTP was docked into the active site of Zika virus NS5 RdRp. The docking results were analysed using the PMF04 ligand scoring method (Muegge, 2006). Following molecular docking, the molecules were further analysed using MD simulations.

4.5.2.7 Molecular dynamics simulations

Protein 3-D structures have an inherent conformational change. The binding conformations of the designed compounds, their ability to stay bound to the active site, and the method of functional control cannot be determined from the docking pose of a single frame of a protein bound to a ligand. As a result, MD simulations of Zika virus NS5 RdRp bound to the hit compounds and GTP were performed using GROMACS 5.1.4 software (Hess et al., 2008; Van Der Spoel et al., 2005). The force fields for protein was generated using Amberff99sb (Da Silva et al., 2012) and the AM1-BCC (Wang et al., 2006) charges were added to hit molecules and GTP and saved in .mol2 format. The force fields for hit molecules and GTP was generated in Antechamber with

ACPYPE scripts. Each complex was solvated with SPC water models, and then Cl^- and Na^+ ions were added for neutralization. The unit cell was adjusted to cubic box dimensions of 1.0 nm (Berendsen et al., 1981). Treatment of long-range electrostatic interactions was done using the particle mesh Ewald approach (Darden et al., 1993; Essmann et al., 1995). Using the LINCS algorithm, the hydrogen bonds were restrained (Hess et al., 1997). The goal of the first stage of MD simulations is to remove steric stress and enable the system to relax. To achieve this goal, 50,000 steps were run until the systems maximal force was less than 1000 kJ/mol/nm^2. In order for the solvents and ions to equilibrate around the protein, the system is switched to NVT, equilibrated, and minimised for 100 ps. Once the system had the correct density, the equilibration has been set to NPT 1 atm pressure and 300 K for 1000 ps. The approaches used for temperature and pressure couplings are V-rescale and Parrinello-Rahman, respectively (Bussi et al., 2007; Parrinello & Rahman, 1981). The final stage of MD simulations on apo Zika virus NS5 RdRp and hit molecule complexes were carried up to 50 ns. The stability and conformational variations of the complexes of apo, and protein-hit molecules, GTP complexes are shown by the RMSD and RMSF graphs.

4.5.3 Results and discussions

The Zika virus NS5 RdRp crystal structure is already reported in PDB (5WZ3), it comprises missing residues in crystal structure (Pro313-Leu321, Ile342-Thr347 and Ser406-Glu425). The template protein structure was identified by submitting Zika virus NS5 RdRp to Dali server. This server compares the 3-D structure of target protein to provide the crystal structure PDBs which are structurally similar with Zika virus NS5 RdRp. Dali server identified the template structure of JEV NS5 RdRp (4HDG). The template protein comprises GTP cofactor and shows sequence identity of 67.34%. The sequence alignment between template JEV NS5 RdRp (4HDG)

and target Zika virus NS5 RdRp (5WZ3) proteins was performed by ClustalW. These missing residues in crystal structure of Zika virus NS5 RdRp protein were constructed by MODELLER9.17. The modeled structure with GTP and sequence alignment shown in **Figure 4.6A** and **4.6B**.

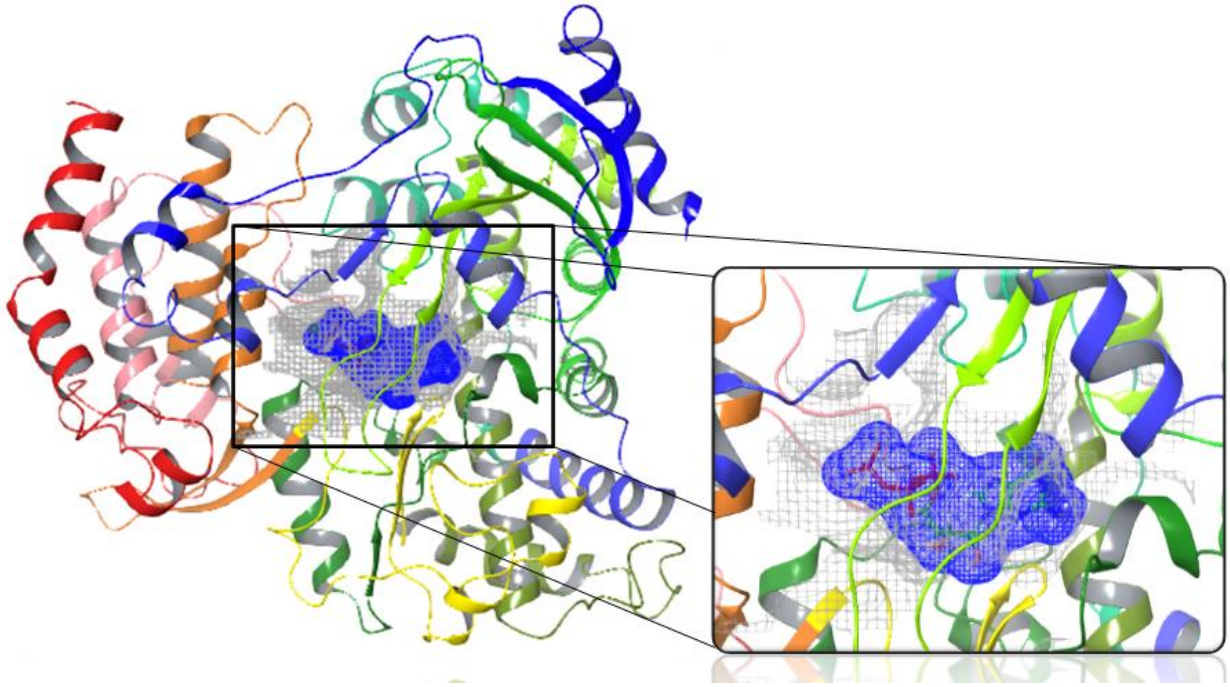


Figure 4.6A: The modeled structure of Zika virus NS5 RdRp complexed with GTP.

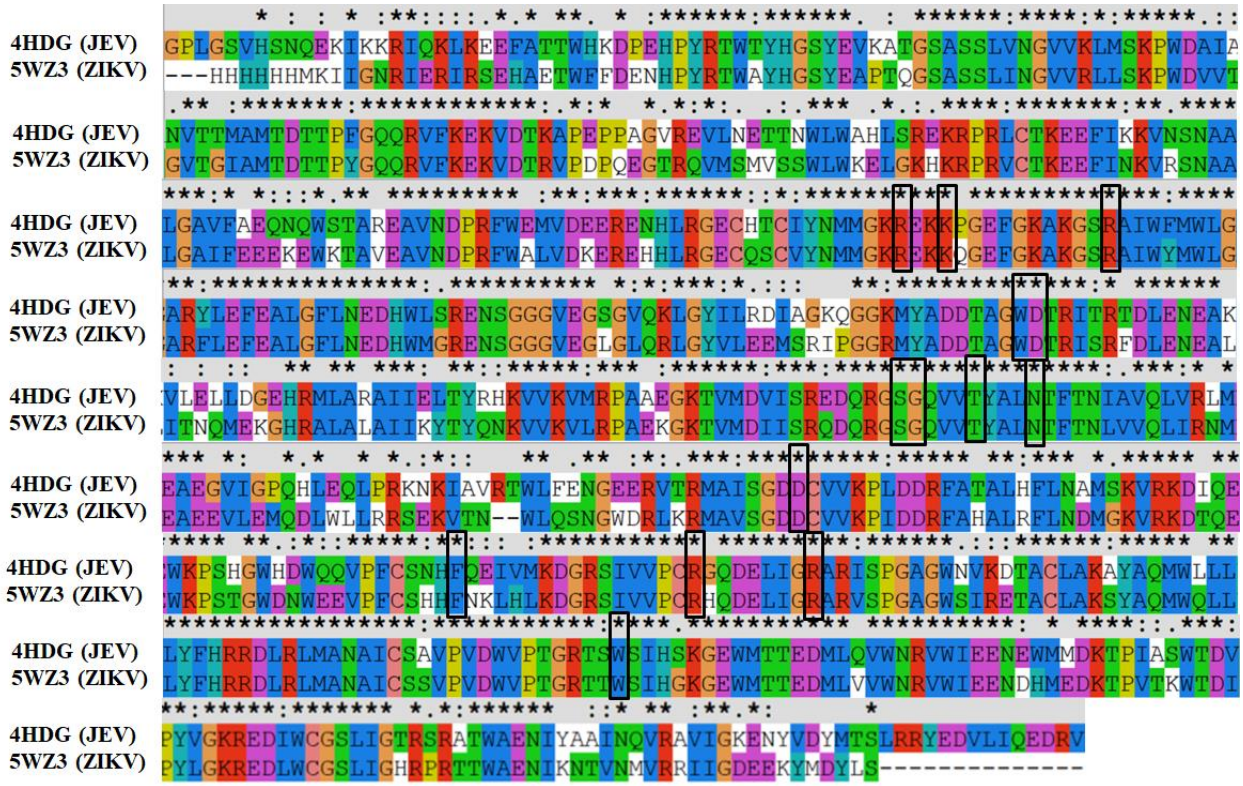


Figure 4.6B: Sequence alignment and conserved active site residues involved hydrogen bonding interactions of Zika virus NS5 RdRp (5WZ3) and JEV NS5 RdRp (4HDG).

4.5.3.1 Identification of active site and validation

Dali was used for validation of active site, as there is no crystal structure available for Zika virus NS5 RdRp protein binding to cofactor or any other small molecule. The JEV NS5 RdRp (4HDG) which is complexed with cofactor GTP in the active site was predicted. For the authentication of binding site and docking method, the substrate that was bound to the protein was removed and redocked the cofactor GTP into the defined active site cavity of Zika virus NS5 RdRp by using CDOCKER protocol from DS 3.5. Selected the final docked pose based on the CDOCKER score (PMF04 -146.84) and important interactions within the binding cavity residues of Zika virus NS5 RdRp. The redocked ligand was located in the proximity of important amino

acids such as, Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739, Arg741 and Trp797. Thus, the redocking results demonstrated that the ligand interacted well with the modelled Zika virus NS5 RdRp protein as shown in **Figure 4.7**.

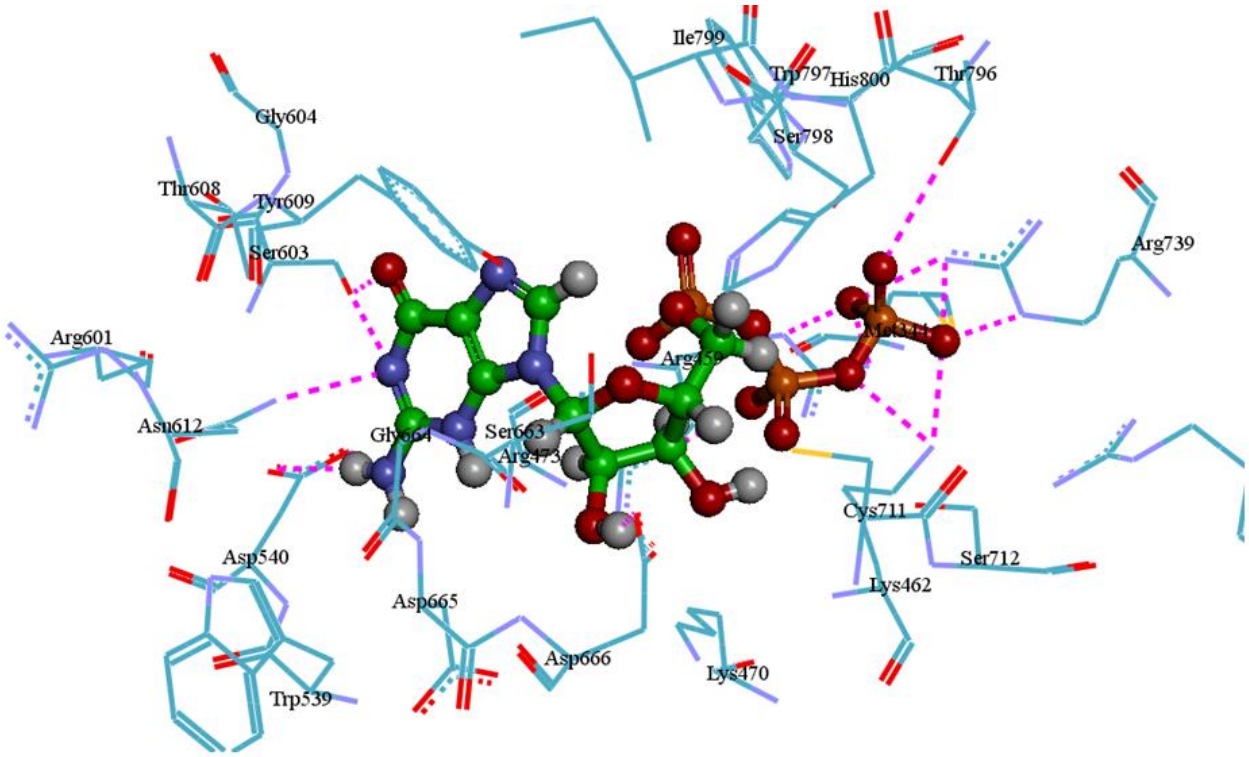


Figure 4.7: Interaction of binding site residues of Zika virus NS5 RdRp with GTP from molecular docking.

4.5.3.2 Feature mapping and pharmacophore generation

Structure-based pharmacophore modeling was used to understand the features which are important for biological function of protein. The interactions of GTP with the amino acid residues in the active center of the target protein Zika virus NS5 RdRp were an appropriate input to design a structure-based pharmacophore model. There are two methods used to generate pharmacophores.

The generation of interactions was analysed first by considering the docked model within certain specific dimensions of the interacting region, and then the features were reduced by performing feature mapping of GTP. The GTP features were mapped, yielding all possible pharmacophore features of the ligand from the selected features. The pharmacophore of GTP was mapped using the selected features, hydrogen bond acceptors and hydrogen bond donors. The feature mapping server generated 66 features on GTP, out of these seven features were considered based on the essential interactions between the receptor and ligand. Based on various interacting points on the active site with GTP, four hydrogen bonding acceptor features, two hydrogen bonding donor features and one negative ionizable feature were selected. The O6 atom of carbonyl group from the nucleoside base forms hydrogen bond with Ser603. The O2 of γ-phosphate forms hydrogen bonding with Lys462 and Arg739. The O3 of γ-phosphate with Arg459 in Zika virus NS5 RdRp. The hydrogen bond donor features are; N2 atom of NH_2 group on guanine ring forms hydrogen bonding with Asp540. The O2' atom on ribose sugar moiety forms hydrogen bonding with Asp666 and one negative ionizable ion O2 atom of β-phosphate forms hydrogen bonding interaction with Arg473. The selected pharmacophore features are shown in **Figure 4.8**. The pharmacophore model was validated on 1000 molecules decoy set (inactive molecules) using DS 3.5. The pharmacophore model successfully qualified the validation process. The seven featured validated pharmacophore model was used for virtual screening of the Asinex database of small molecules.

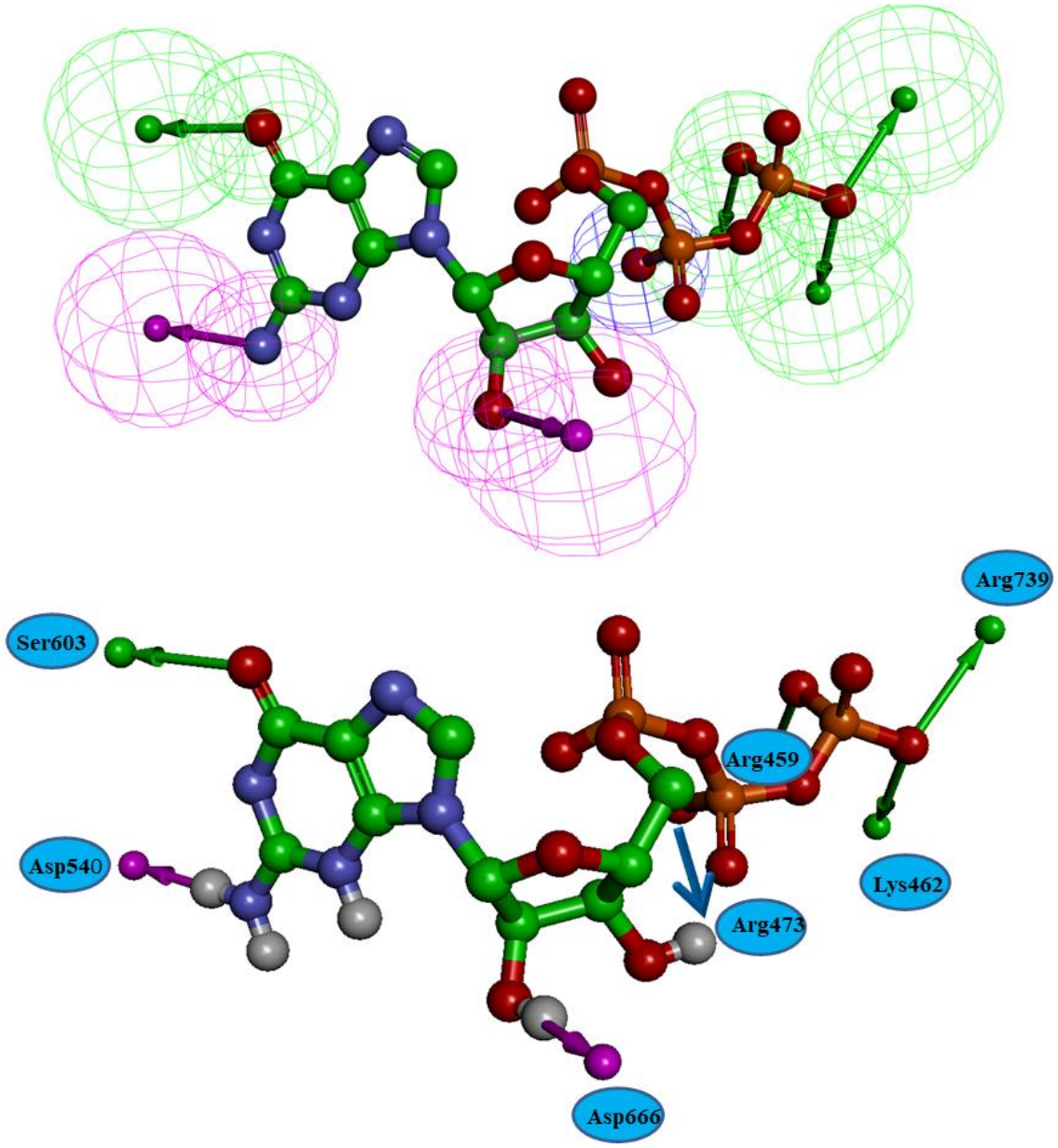


Figure 4.8: Pharmacophore features used for virtual screening of Asinex database.

4.5.3.3 Virtual screening and molecular docking

The virtual screening protocol relies on the use of filters to limit the number of compounds. The finalised pharmacophore was employed to screen the Asinex database with 500,000 compounds, followed by molecular docking studies of the selected hit molecules. The finalised pharmacophore model was used for virutral screeing of Asinex database based on the generated

structure based pharmacophore model using screen library protocol avalible in DS 3.5 with the best conformation generation and flexible fitting methods. This virtual screening retrieved 29,427 molecules from the database. These hit molecules were filtered and ranked based on lower RMSD (< 0.3 Å) and the Lipinski's rule was applied in prefiltering process. The top 59 hit molecules recognised through virtual screening were then docked using the CDOCKER docking program to confirm their binding conformation and affinity to Zika virus NS5 RdRp (5WZ3). The best-scoring molecules were represented graphically for nonbonding interactions such as hydrogen bonds and pi-stacking. PMF04 scores were used to assess the binding efficiency of protein-hit molecule complexes and GTP. The scores of the newly identified hit molecules ranged from -128.08 to -42.17. The greater docking score was identified for GTP and hit-1 (-146.84 and -128.08). These 59 molecules are docked into 5 Å cavity present at GTP binding site in the Zika virus NS5 RdRp protein. Twenty docking conformations were generated for each molecule by CDOCKER to predict the best conformer based on CDOCKER score and nonbonding interaction with protein. Out of the 59 molecules, 12 best molecules were selected for MD simulations studies as shown in **Table 4.5**.

Table 4.5: Fit values, docking results of GTP, and hit molecules screened from Asinex database.

S. No	Compound	Fit value	Docking Score PMF04	Hydrogen bonding interactions	Active site interactions
1	GTP	5.21	-146.84	Arg459, Lys462, Ser603, Arg473, Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
2.	hit-1	3.45	-128.08	Arg459, Lys462, Ser603, Arg473 and Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
3.	hit-2	3.155	-81.59	Arg459, Lys462, Arg473, Asn612, Asp666 and Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
4.	hit-3	2.91083	-76.52	Arg459, Lys462, Ser603, Arg473, Arg739, Trp797, Ser798 and His8 00	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
5	hit-4	3.73602	-76.49	Arg459, Lys462, Ser603, Arg473, Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
6.	hit-5	2.49697	-76.25	Arg459, Lys462 and Arg473	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
7	hit-6	2.6132	-76.13	Arg459, Lys462, Ser603, Arg473, Arg739,Trp797, Ser798 and His8 00	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741

8	hit-7	3.50634	-75.11	Arg459, Lys462 and Ser603	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
9	hit-8	2.98349	-74.39	Arg459, Lys462, Ser603, Arg473, Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
10	hit-9	3.155	-74.35	Arg459, Lys462 and Ser603	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
11	hit-10	3.07037	-73.61	Arg459, Lys462, Ser603, Arg473, Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
12	hit-11	3.42	-56.13	Arg459, Arg473 and Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741
13	hit-12	2.64	-42.17	Arg459, Lys462, Arg473, Arg739	Arg460, Lys462, Lys471, Arg473, Trp539, Asp540, Ser603, Thr608, Asn612, Asp666, Asp669, Ser715, Arg739 and Arg741

4.5.3.4 Molecular dynamics simulations

The stability and dynamic characteristics of the protein-hit molecule complexes were determined using atomistic MD simulations with explicit solvent. The details into the insights of intermolecular interactions between the receptor-ligand in motion is provided by MD simulations studies, which contribute to the protein stable bound conformation, as well as the influence of

ligand binding on protein conformational changes. Protein ligand-bound complexes were studied in a dynamic environment to determine the stability of interactions over time. In this study, apo protein and Zika virus NS5 RdRp complexed with GTP and screened hit molecules were used for 50 ns MD simulations. Out of the twelve hit compounds, four compounds obtained stability throughout MD simulations. The average structures were generated from 50 ns of MD simulations trajectories for comparison with the initial structures of each system to analyse the conformational changes of protein and stability of hit molecules binding to protein. The superposition of the initial and average structures demonstrate that the molecules have good stability and lower RMSD, which is confirmed by conserved hydrogen bonding interactions with active site residues, indicating that these molecules are suitable as Zika virus NS5 RdRp inhibitors. The extent to which all atom positions in the apo and protein-hit complexes deviate during MD simulations is assessed by the RMSD plots. The overall conformational changes were analyzed through the Cα-atom RMSD as a function of MD simulations time. From the plots it can be seen that the apo protein gained stability after the initial ten ns with an average RMSD of 0.33 nm throughout MD simulations and when binding with hit molecules it shows lower RMSD (< 0.3 nm). The hit molecules also show less than 0.2 nm RMSD (as shown in **Figure 4.9)**. From these results it is evident that GTP and hit molecules possess better binding and acquire stable interactions at the binding pocket of NS5 RdRp throughout the 50 ns of MD simulations.

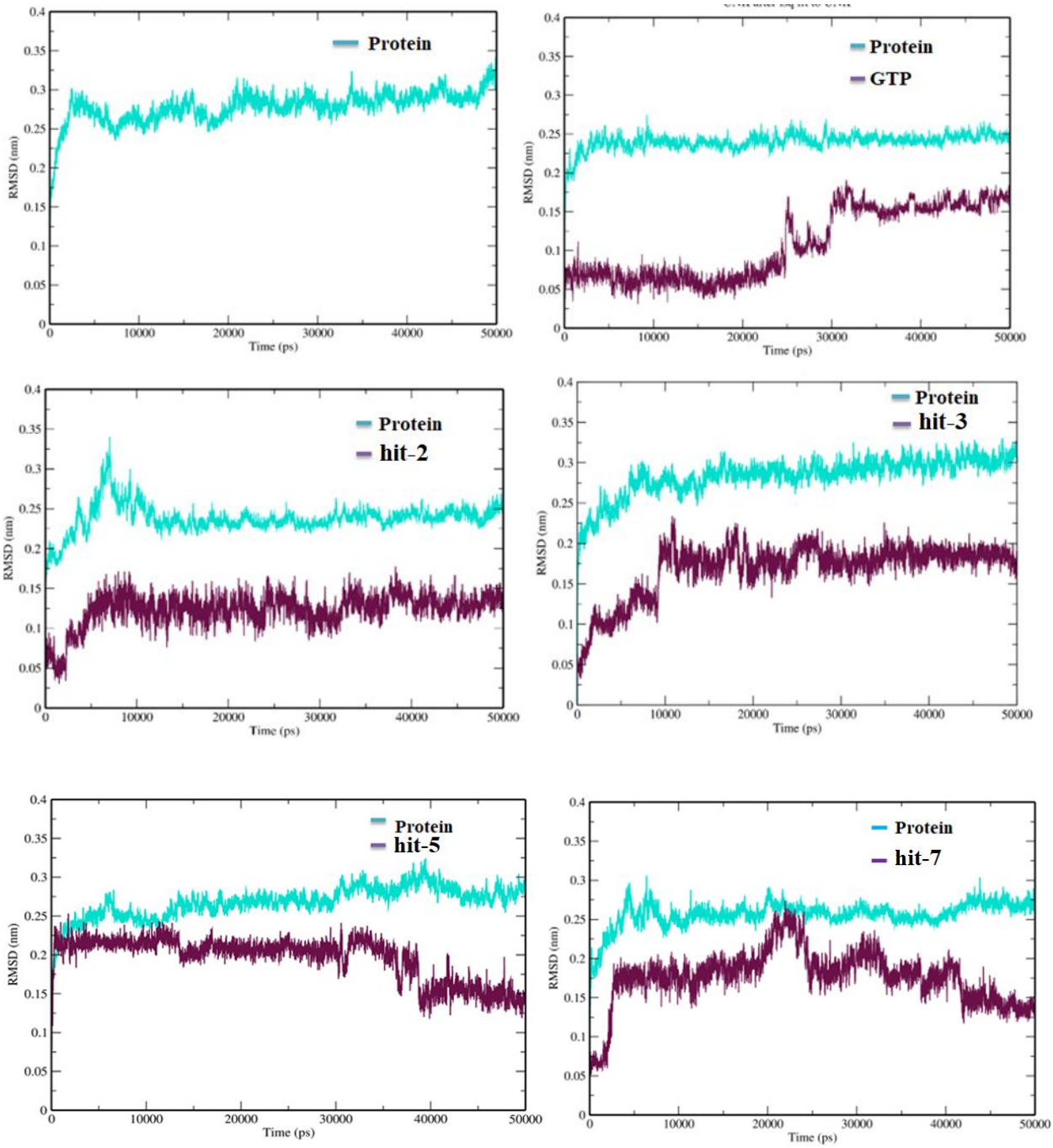


Figure 4.9: RMSD plots of Zika virus NS5 RdRp apo and bound with GTP, screened hit molecules.

From the RMSF plots it was identified that the Cα-backbone atoms **Figure 4.10** for majority of the protein amino acid residues have lower fluctuations (< 0.2 nm), and only the residues that form a loop display higher fluctuations and reaches until 0.5 nm. The amino acid residues in Zika virus NS5 RdRp protein from Gln315-Leu323, is the loop residues, Thr348-Gln352 in helices and Trp748-Thr753 amino acids adjacent to the priming loop show significant fluctuations during MD simulations in the apo Zika virus NS5 RdRp, and when complexed with hit molecules. The residues Glu465-Phe487 present in loop-β-sheet-helix shows fluctuations in all complexes. By the comparison of the RMSF plots from apo, GTP and hit molecules complexed protein, it was observed that the complexed protein showed greater stability compared with apo Zika virus NS5 RdRp protein.

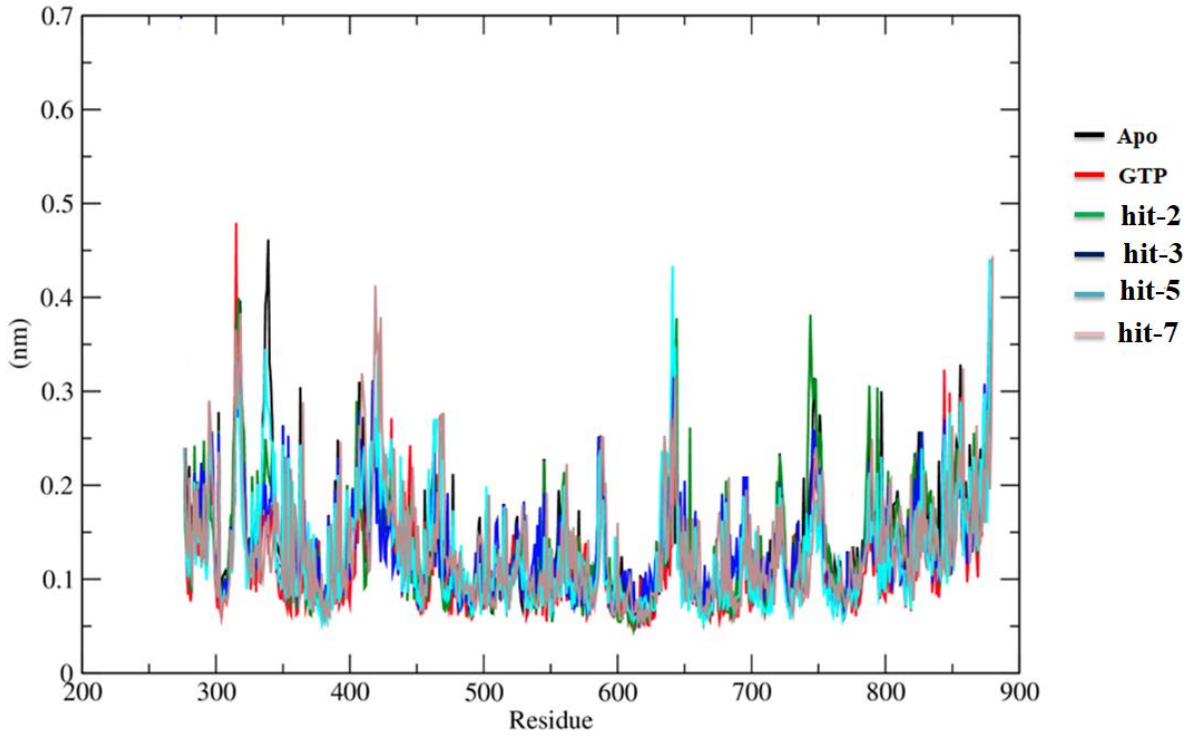
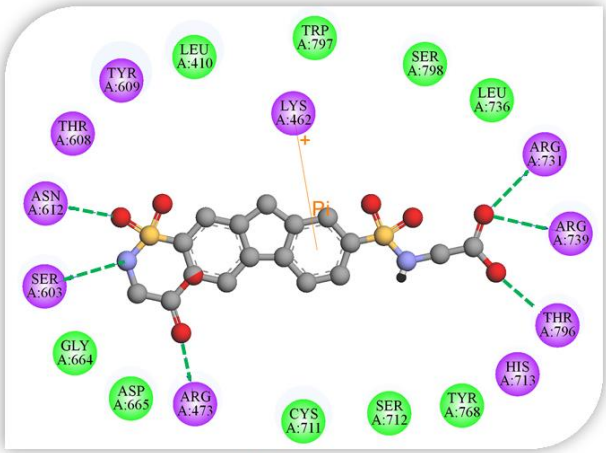


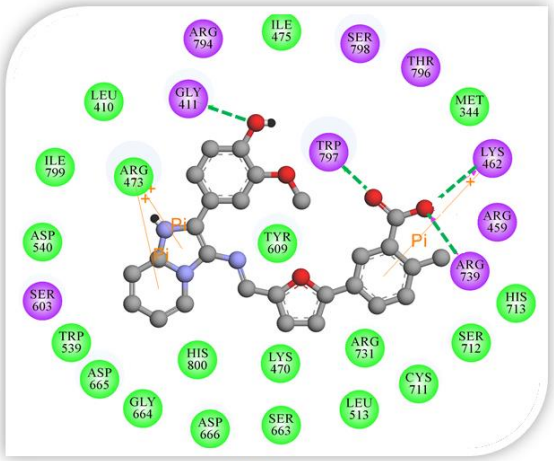
Figure 4.10: RMSF plot of Zika virus NS5 RdRp apo, bound with GTP and screened hit molecules.

4.5.3.5 Hydrogen bonding interactions

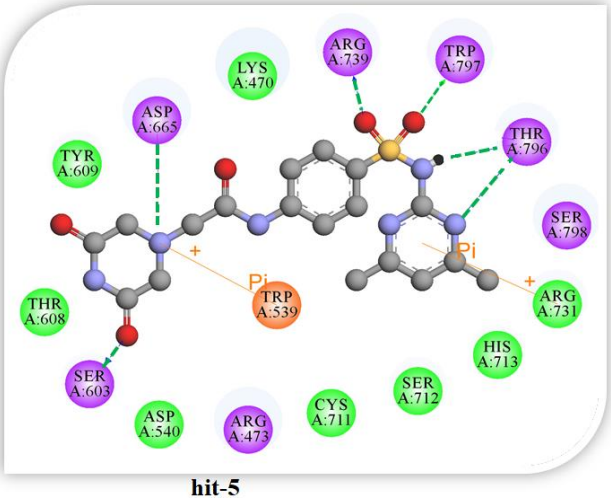
For the majority of receptor-ligand complexes, the formation of hydrogen bonds represents one of the crucial interactions required to keep the ligand molecule firmly in the receptor binding cavity. The hydrogen bonding interpretation of the four hit molecule complexes was performed to investigate their hydrogen bonding network at the binding site based on MD simulations. Based on a comparison of the initial input and average structures of hit molecules when bound to Zika virus NS5 RdRp, it was revealed that ligands formed several nonbonding interactions with the binding pocket residues (Lys462, Arg473, Arg459, Asp540, Ser603, Arg742, Arg739 and Trp800). Apart from these interactions the characteristic hydrogen bonds with Lys462, Arg473, Arg459, Ser603, Arg739, Thr796 and Trp797 were retained in all complexes throughout MD simulations, indicating these interactions stabilized the complexes as shown in **Figure 4.11**.



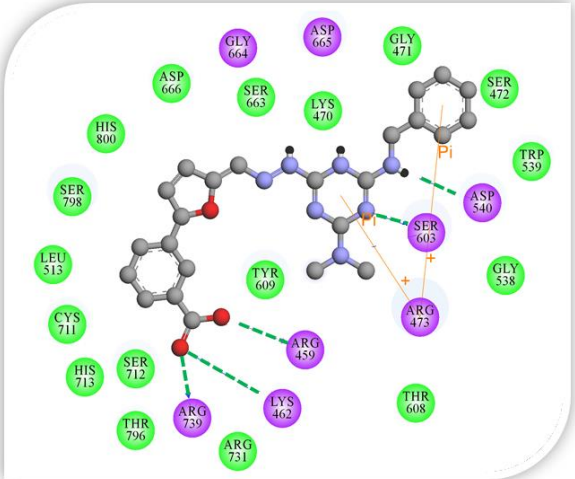
hit-2



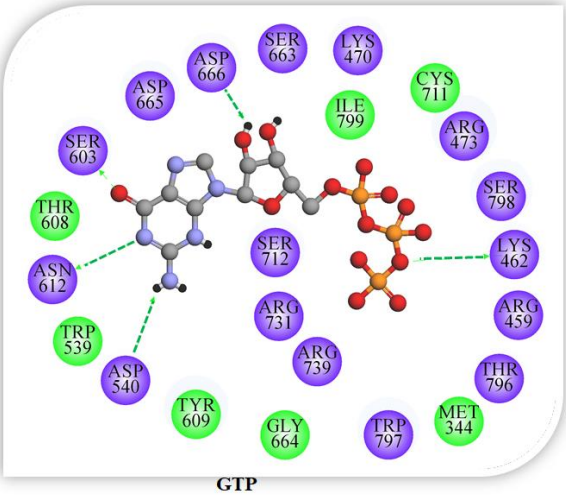
hit-3



hit-5



hit-7



GTP

Figure 4.11: Hydrogen bonding interactions of GTP and screened hits with Zika virus NS5 RdRp MD simulations.

4.5.3.6 Binding free energy analysis

The binding free energies of GTP and screened hit molecules in this work were estimated using MM-PBSA and indicated in **Table 4.6.** The contribution from electrostatic, van der Waals and polar solvation energies show compatibility with the reference (GTP) molecule. The binding energies were observed to be GTP (-204.66 kJ/mol), hit-2 (-96.41 kJ/mol), hit-3 (-95.01 kJ/mol) hit-5 (-88.91 kJ/mol) and hit-7 (-74.58 kJ/mol).

Table 4.6: Binding energy of Zika virus NS5 RdRp bound with GTP and screened hit molecules.

Compound Name	**van der Waals energy (kJ/mol)**	**Electrostatic energy (kJ/mol)**	**Polar solvation (kJ/mol)**	**SASA (kJ/mol)**	**ΔG (kJ/mol)**
GTP	-192.22+/0.700	-96.425+/- 0.00	102.156 +/- 0.00	-18.136+/- 0.00	-204.625+/0.719
hit-2	-135.195+/0.849	-24.110+/1.012	107.495+/1.407	-15.59+/0.068	-96.41+/1.249
hit-3	-155.663+/- 0.349	-6.390 +/- 0.17	83.511 +/- 0.55	-16.472+/- 0.03	-95.007 +/- 0.52
hit-5	-144.68+/-0.55	-46.15+/-0.43	115.75+/-0.87	-13.81+/-0.05	-88.91+/-0.63
hit-7	-158.16+/0.516	-38.36+/0.664	138.534+/0.724	-16.59+/0.050	-74.58+/0.561

In order to explain the importance of active site amino acid residues in binding the ligands, their contribution to the binding energy was measured. The contribution based on energy decomposition analysis of majorly participating residues in Zika virus NS5 RdRp to the

interactions of hit molecules are Arg459, Lys462, Arg473, Trp539, Asp540, Ser603, Gly604, Arg731, Arg739, Thr796 and Trp797. The energy contribution of GTP and screened hit compounds are shown in **Table 4.7**. GTP showed greater contribution to the amino acid residues in the binding pocket, compared to the other hit compounds. The screened hit molecule bound to NS5 RdRp showed high contribution from the residues Arg459, Lys462, Asp540 and Arg739 and the residues Trp539 and Gly604 show lower contribution. **Figure 4.12** represents the contribution from binding cavity amino acid residues to the formation of a complex. The compounds discovered through the virtual screening have nearly the same and greater contributions in comparison with the GTP. By this analysis of binding energy calculations, it was suggested that hit molecules show good binding affinity with Zika virus NS5 RdRp.

Table 4.7: Residue-wise contribution (kJ/mol) of Zika virus NS5 RdRp with GTP and screened hit molecules.

Residues	**GTP**	**hit-2**	**hit-3**	**hit-5**	**hit-7**
Arg459	-67.44	-46.60	-9.07	-9.44	-62.77
Lys462	-42.28	-18.96	-16.76	-22.26	-33.33
Arg473	-18.77	-7.05	-11.99	-3.04	-4.84
Trp539	-8.19	-6.40	-4.39	-2.98	-6.23
Asp540	-62.15	-41.30	-19.15	-57.60	-14.35
Ser603	-28.23	-5.94	-15.15	-19.76	-7.09
Gly604	-9.24	-4.64	-0.31	-0.65	-3.07
Arg731	-36.20	-18.97	-21.51	-19.76	-9.14
Arg739	-78.04	-32.74	-69.87	-3.43	-56.07
Thr796	-32.32	-26.94	-28.33	-21.50	-14.19
Trp797	-26.18	-10.15	-19.02	-24.97	-18.31

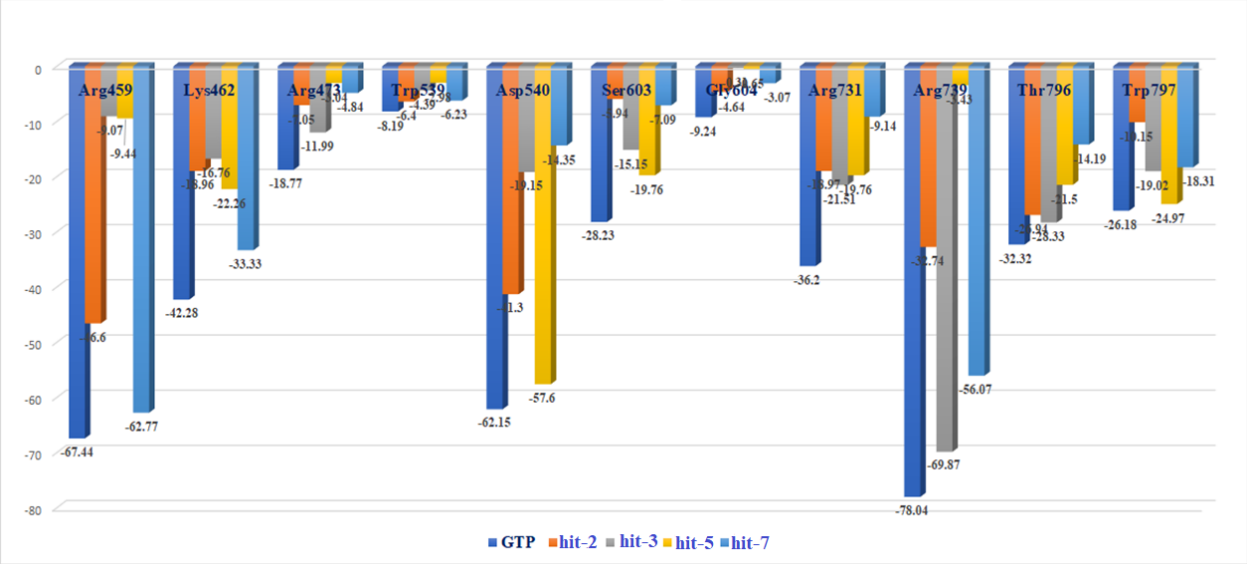


Figure 4.12: Residue-wise contribution energy (kJ/mol) of Zika virus NS5 RdRp binding with GTP and screened hit molecules.

4.5.4 Conclusions

CADD methods can be used successfully to accelerate the development of therapeutic agents for Zika virus disease. The pharmacophore-based virtual screening method combines modelling based on ligands and modelling based on structures. Compared to a straightforward ligand-based pharmacophore search, structure-based pharmacophore appears to be more effective at locating potent hits with structural variety. Four hit molecules (hit-2, hit-3, hit-5, and hit-7) that were bound to the Zika virus NS5 RdRp were found to be stable using MD simulations. Overall examination of these findings indicated that the design of new Zika virus inhibitors for the treatment of Zika infection was made possible by the structure-based pharmacophore modelling, which supplied the necessary information for understanding the crucial structural binding properties.

CHAPTER-5

Computer aided design of NCI natural products as inhibitors to SARS-CoV-2 main protease

5.1 Introduction

Acute respiratory illness caused by the SARS-CoV-2 was initially identified as the COVID-19 in Wuhan, China, in December 2019. Nearly 30K nucleotides from the SARS-CoV-2 genome are translated into the overlapping pp1a and pp1b. Two proteases, 3C-like and papain-like, are present in the pp1a polypeptide. Leu-Gln↓(Ser/Ala/Gly) is the particular recognition sequence motif for cleavage (↓) by the main protease. Because it cleaves polyproteins upon translation into the host cell cytosol, the protease enzyme is essential in viral protein growth and development for many viruses. Drug targets for viral proteases are common. Slowing the formation of mature viral particles can be slowed by inhibiting viral protease. To date, numerous antiviral drugs have been developed to combat viral infections by targeting proteases. Among the FDA approved drugs are the HIV-1 protease inhibitors (tipranavir, darunavir, amprenavir, lopinavir, saquinavir, atazanavir, indinavir, ritonavir, and nelfinavir) (Lv et al., 2015) and hepatitis C virus (HCV) NS3/4A protease inhibitors (boceprevir). As a result, developing antiviral drugs that inhibit the main protease of SARS-CoV-2 could have clinical applications. The main protease of coronaviruses is a potential drug target because it is responsible for its own maturation (Ziebuhr et al., 2000).

The crystal structure of SARS-CoV-2 main protease (PDB id: 6LU7) is a homodimer. Each protomer contains three domains I, II and III (Zhang et al., 2020). The crystal structures of the protein binding with the peptide inhibitor N3 (Jin et al., 2020) and α-ketoamide inhibitor (6Y2G) (Zhang et al., 2020) shown in **Figure 5.1A** and **B** are reported. Another inhibitor GC376, a broad-spectrum dipeptidyl inhibitor bound at the active site of main protease (7CB7) (Wang et al., 2020)

is also reported. So far, research on dengue viruses has shown that the molecules in the NCI Diversity Database have good inhibitory action (Abduraman et al., 2018). These positive outcomes led to the virtual screening of compounds from natural product datasets in the NCI database using docking at the main protease active site. Following this, estimates of their binding free energies to determine their binding affinity when complexed with SARS-CoV-2 main protease, molecular docking and MD simulations were used to examine the binding interactions of the screened compounds.

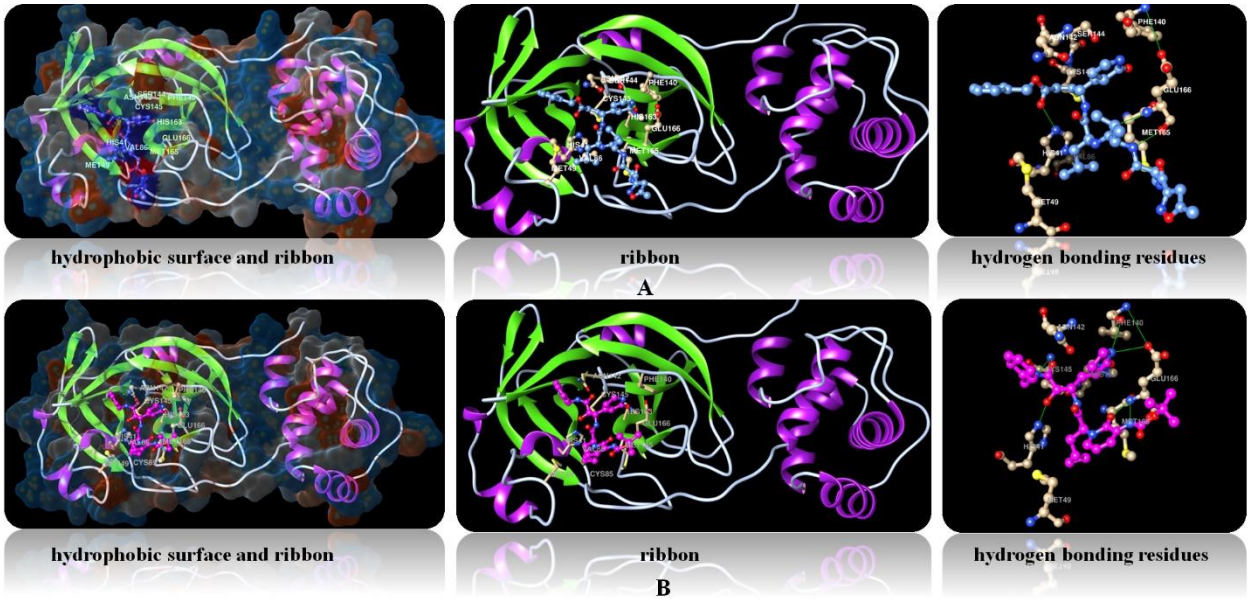


Figure 5.1 A) SARS-CoV-2 main protease binding with N3-inhibitor (PDB id: 6Y2G)

B) SARS-CoV-2 main protease binding with α-ketoamide inhibitor (PDB id: 6LU7).

5.2 Materials and Methods

5.2.1 Protein preparation

The crystal structures of SARS-CoV-2 main protease complexed with inhibitor N3 (PDB id: 6LU7) (Jin et al., 2020) and α-ketoamide inhibitor (PDB id:6Y2G) (Zhang et al., 2020) deposited in PDB were used for this study. The heteroatoms and water molecules were deleted from the protein crystal structures, hydrogen atoms were added in order to prepare the protein for screening the molecules and the 3-D structure coordinates were saved in .pdbqt format for virtual screening using PyRx server (Dallakyan & Olson, 2015).

5.2.2 Ligand preparation

The coordinates of N3 and α-ketoamide inhibitors were extracted from the crystal structures of main protease. Library of compounds from the NCI natural compounds set database (II, III, IV and V) (https://wiki.nci.nih.gov/display/ncidtpdata/compound+sets) comprising 1,046 molecules were downloaded in .sdf format. Hydrogen atoms were added at pH 7.0 and the coordinates of the compounds were saved in .pdbqt format.

5.2.3 Virtual screening and molecular docking

The docking-based virtual screening using PyRx software was performed as the initial step to identify potential main protease inhibitors. The 1,046 natural compounds were screened by docking into the active site at the α-ketoamide inhibitor binding location in the PDB id: 6Y2G. The compounds were ranked according to their binding mode and scoring analysis. The best molecules obtained from virtual screening were docked using AutoDock Vina (Trott & Olson, 2010; Morris et al., 2009) that employs protein-ligand flexible docking using the Broyden-

Fletcher-Goldfarb-Shanno method. The protein structure with all the compounds was loaded and ten conformations were generated for each ligand molecule by AutoDock Vina, the grid box was centered at 30.71, 50.48, 4.10 Å in x, y, z coordinates, respectively, with a grid spacing; 0.492 Å, box size of 25 x 25 x 25 points and exhaustiveness was set to 10. Initially, the molecules were loaded; torsions were set and saved in .pdbqt format. The screened-in molecules were docked within a 5 Å cavity defined around the α-ketoamide binding pocket in the SARS-CoV-2 main protease. The best conformer selected based on binding affinity and the number of hydrogen bonding interactions between the docked pose of natural product and protein were manually visualized. The virtual screening and molecular docking methods were validated by redocking the crystal ligands N3 and α-ketoamide inhibitors into the receptor active site.

5.2.4 Validation of molecular docking

The top-ranked molecules from AutoDock Vina were further proceeded for another round of docking studies using CDOCKER (Gagnon et al., 2016) available in DS 3.5. A sphere of 5 Å radius was generated around α-ketoamide inhibitor to define the active site of protein. Ten docking poses were generated for each molecule in the protein active site. The binding conformations of the molecules in SARS-CoV-2 main protease were analyzed using "scoring ligand poses" implemented in receptor-ligand interactions protocol in DS 3.5. The scoring functions PLP1, PLP2 and PMF (Gehlhaar et al., 1995; Parrill et al., 1999; Muegge et al., 1999; Muegge 2006) were used to assess the docking poses. The selection of docking pose was based on top scores and intermolecular interactions with SARS-CoV-2 main protease. The best hit molecules chosen from both AutoDock Vina and CDOCKER docking methods were subjected to study their drug-like properties.

5.2.5 Lipinski rules and ADME properties

The drug-like properties of the best docked compounds were studied by analyzing the pharmacokinetics profile using the SwissADME server. This is a software tool to calculate molecular properties such as ADMET and physicochemical properties such as solubility, lipophilicity and pharmacokinetics. The Lipinski's rule of five (Lipinski et al., 1997; Lipinski et al., 2012; Lipinski, 2004) are an essential criterion to ensure a drug-like profile for orally administered drugs. The hit molecules that qualify the ADME properties were studied by MD simulations studies in complex with SARS-CoV-2 main protease.

5.2.6 Molecular dynamics simulations

MD simulations of the apo and SARS-CoV-2 main protease in complex with hit molecules was carried out using GROMACS-5.1.4 for 150 ns. These studies reveal the stability of protein-ligand complexes during MD simulations. The Amber99sb force field (Hornak et al. 2006) was applied to the protein, force fields were assigned to the small molecules using ACPYPE script (Da Silva et al., 2012) with AM1-BCC charges in Antechamber (Wang et al., 2006). The molecular systems were immersed in a cubic box, SPC waters were added to the system, Na^+ and Cl^- ions were added to neutralize (Berendsen et al., 1981) the systems and periodic boundary conditions were applied. Energy minimization was carried out with a tolerance of 1000 kJ/mol/nm^2. The systems were heated until 300 K for 100 ps; in the subsequent step, the system was equilibrated at 1 atm and 300 K for 1000 ps until it reaches proper density. The temperature was maintained using a V-rescale thermostat (Bussi et al., 2007) and Parrinello–Rahman method was used to control the pressure (Parrinello & Rahman 1981). The long-range electrostatics were handled using the particle mesh Ewald method (Darden et al., 1993 and Essmann et al., 1995). The equilibration of

molecular systems was performed under NVT and NPT ensembles for 1000 ps. The Lennard Jones interactions and the real-space electrostatic interactions were truncated at 9 Å. Hydrogen bonds were constrained using the LINCS algorithm (Hess et al., 1997). The coordinates from production MD trajectories were generated and saved for every 2 ps. The final models in all the systems were obtained by averaging the snapshots from the trajectories generated by MD simulations after the structure stabilization was achieved. The RMSD of the Cα-atoms concerning their starting structures was calculated using gmx rms, and the RMSF were calculated using gmx rmsf commands in GROMACS. The xmgrace software was used to plot the data, UCSF Chimera (Pettersen et al., 2004) was used for structure superposition and PyMOL was used for cartoon image generation. For the sake of comparison; apo, N3 and α-ketoamide bound SARS-CoV-2 main protease were also studied by MD simulations.

5.2.7 Binding free energies of protein-ligand complexes

The protein-ligand binding affinities describe the extent of intermolecular recognition. The ligand binding free energies were calculated based on MM-PBSA approach (Kumari et al., 2014) using g_mmpbsa tools.

The LIE (Almlöf et al., 2004 and Brandsdal et al., 2003) was computed as the mean of van der Waals (vdW) and coulomb (cou) interaction energy differences of the inhibitor with its neighboring atoms upon incorporation, that is, the individual ligand in the solvent (unbound state denoted as subscript u) and the inhibitor in the binding mode with SARS-CoV-2 main protease (bound state denoted as subscript b). The ligand binding free energies were calculated using gmx energy and gmx lie for the SARS-CoV-2 main protease-hit molecule complexes from the output trajectories of MD simulations.

5.2.8 Normal mode analysis and mechanical stiffness

NMA can provide a quick and systematic investigation of protein dynamics. Elastic network model-based NMA was developed using dihedral angels as independent variables for all molecular systems using the software suite of programs in Prodynamics (Uyar et al., 2011; Atilgan et al., 2001). Mechanical stiffness plots of all molecular systems in response to all possible pulling directions were constructed by using ANM using the software suite of programs in Prodynamics (Eyal et al., 2015).

5.2.9 Principal component analysis

PCA was performed to study the overall motion of SARS-CoV-2 main protease in all the simulated systems using MODE-TASK (Ross et al., 2018). A 3N × 3N covariance matrix was created using Cartesian coordinates, followed by the construction of eigenvectors by diagonalization of the covariance matrix. The PCA was calculated from 0 to 150 ns MD simulations trajectories.

5.3 Results and discussion

The crystal structure of SARS-CoV-2 main protease has three domains: domain I (1-99 amino acid residues), domain II (100-182 residues) and domain III (199-307 residues). Domains I and II are each made up of a six-stranded β-barrel, with the substrate binding site located at the intersection of the two domains. The binding site is made up of subsites S1, S2, S3, S4, and S1', which are

represented in **Figure 5.2** based on the position of the substrate (Jin, Zhenming, et al., 2020).

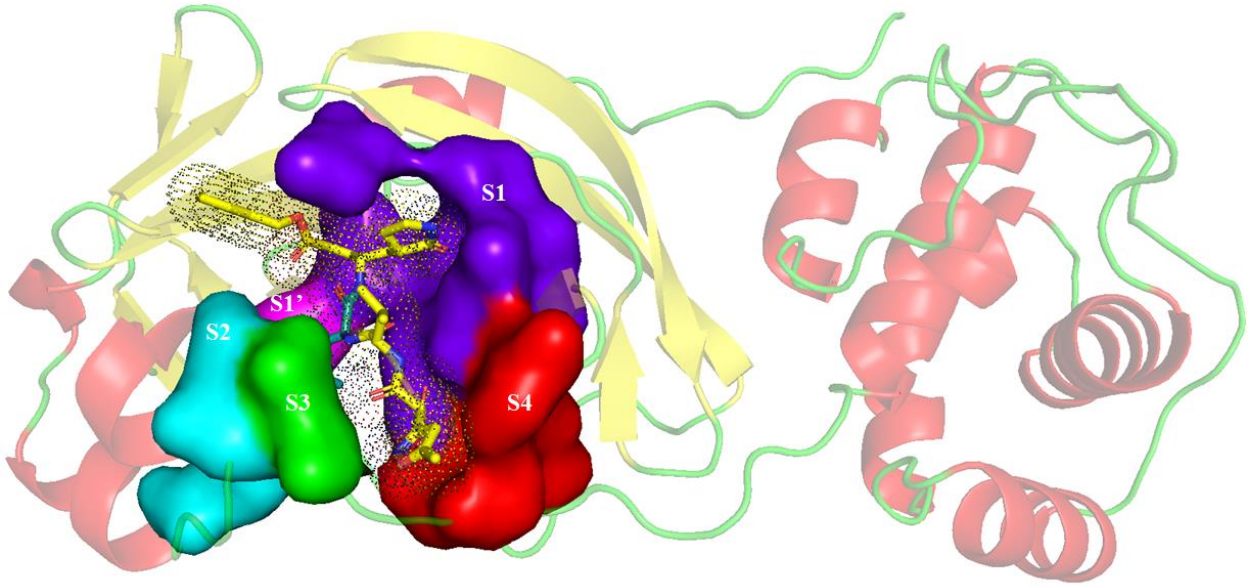


Figure 5.2: Subsites of SARS-CoV-2 main protease N3 and α-ketoamide binding site.

The domains I and II, which are linked to hinge region residues (182-198), assist in forming the S3 and S4 subsites. Based on the crystal structure of the α-ketoamide bound protein, the lactam ring at the P1 position of the inhibitor is in the S1 subsite formed by the side-chains of Phe140, Asn142, Glu166, His163 and His172. The lactam nitrogen at position P1 forms hydrogen bonds with the Phe140 main-chain carbonyl oxygen. The cyclopropyl methyl group at P2 is incorporated into the S2 subsite constituted by His41, Met49, Tyr54, Met165 and Asp187 side-chains. The inhibitors carbonyl oxygen close to the lactam ring forms a hydrogen bond with the His41 side-chain. The amide nitrogen located between the lactam and cyclopropyl methyl groups forms hydrogen bonds with His164 main-chain carbonyl oxygen. The carbonyl oxygen on pyridone forms hydrogen bonds with the main-chain NH of Glu166. The inhibitors OH functional group on imine carbon forms hydrogen bonds with the main-chain NH of Ser144 and Cys145 in the S1' subsite. The N3 inhibitor covalently bonded with Cys145 and hydrogen bonded with Phe140,

Gly143, His164, Glu166 (S1 subsite), Gln189, and Thr190 (S4 subsite) (Jin, Z. et al., 2020). Both crystal structures are highly superimposable, with a low RMSD of 0.69 Å.

5.3.1 Virtual screening and molecular docking

The N3 and α-ketoamide inhibitors binding site was considered as the active site of SARS-CoV-2 main protease. PyRx server based virtual screening of natural products (1,046 molecules) into the active site of main protease successfully screened 736 potential hit molecules. These screened-in molecules were docked into the SARS-CoV-2 main protease active site using AutoDock Vina. The N3 and α-ketoamide inhibitors docked into the protein active site with binding affinity -7.8 kcal/mol and -9.6 kcal/mol, respectively, and formed hydrogen bonding interactions similar to the crystal structure. Thirty natural product molecules were retrieved with AutoDock Vina score lower than -7.0 kcal/mol that also contribute hydrogen bonding interactions similar to the reference molecules. These 30 molecules were studied for another round of docking by CDOCKER using receptor ligand interaction protocols available in DS 3.5. The docking protocols validated by redocking the reference molecules N3 and α-ketoamide at the active site of SARS-CoV-2 main protease is shown in **Figure 5.3**. Eight best compounds **Table 5.1** were selected based on intermolecular hydrogen bonding interactions with SARS-CoV-2 main protease active site and the highest docking scores from AutoDock Vina and CDOCKER.

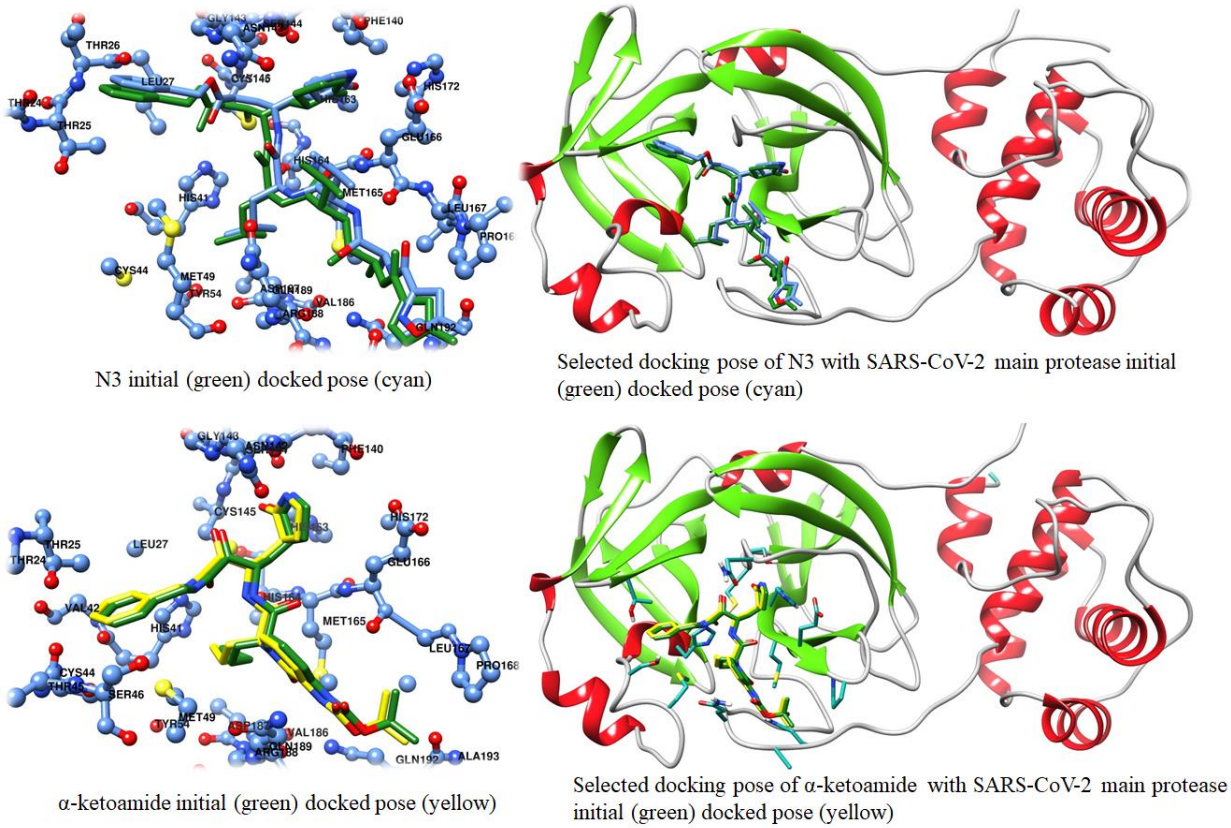


Figure 5.3: Redocking of N3 and α-ketoamide at active site of SARS-CoV-2 main protease.

Table 5.1. Docking scores of N3, α-ketoamide and screened molecules along with the interacting residues in the SARS-CoV-2 main protease binding site.

S No	Compound name	AutoDock Vina score (kcal/mol)	CDOCKER score			Hydrogen bonding forming residues	Active site residues
			-PLP1	-PLP2	-PMF		
1.	N3	-7.8	119.02	71.73	44.22	Phe140, Gly143, Cys145, His164, Glu166, Gln189 Thr190	Leu27, His41, Leu50, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164 Met165, Glu166, Leu167, Pro168, Gln189, Thr190
							Leu27, His41, Leu50, Phe140, Leu141,

2.	α-ketoamide	-9.6	78.98	38.17	46.39	His41, Phe140, Ser144, Cys145, His164, Glu166	Asn142, Gly143, Ser144, Cys145, His163,His164 Met165, Glu166, Leu167, Pro168, Gln189
3.	NSC36398	-8.1	95.76	97.49	18.22	Phe140, Leu141, Ser144, Met165, Glu166, Gln189	Leu27, His41, Leu50, Phe140, Leu141, Gly143, Ser144, Cys145, His163, His164 Met165, Glu166, Leu167, Pro168 and Gln189
4.	NSC281245	-8.0	109.47	94.08	54.88	His41, Phe140 Ser144, Glu166	Leu27, His41, Leu50, Phe140, Leu141, Gly143, Ser144, Cys145, His163, His164 Met165, Glu166, Leu167, Pro168 and Gln189
5.	NSC11926	-7.8	86.04	40.27	41.69	Phe140 Gly143, Ser144, Cys145, Glu166	Leu27, His41, Leu50, Phe140, Asn142, Gly143, Ser144, Cys145, His163, His164 Met165, Glu166, Leu167, Pro168 and Gln189
6.	NSC44175	-7.8	98.62	67.12	42.11	His41, Gly143, Cys145, Glu166	Leu27, His41, Leu50, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, Leu167, Pro168, Gln189
7.	NSC5113	-7.6	75.13	61.32	22.13	Leu141, Gly143, Ser144, Cys145, Glu166	Leu27, His41, Leu50, Phe140, Leu141, Gly143, Ser144, Cys145, His163, His164 Met165, Glu166, Leu167, Pro168, Gln189
8.	NSC107067	-7.6	102.59	74.27	10.25	Leu141, Gly142, Ser144	Leu27, His41, Leu50, Phe140, Leu141, Gly143, Ser144, Cys145, His163,

							His164 Met165, Glu166, Leu167, Pro168, Gln189
9.	NSC22842	-7.5	74.3	70.76	11.68	Ser144, His163, Glu166	Leu27, His41, Leu50, Phe140, Leu141, Gly143, Ser144, Cys145, His163, His164 Met165, Glu166, Leu167, Pro168, Gln189
10.	NSC94600	-7.3	78.97	62.31	8.2	Ser144, Asn142, Glu166	Leu27, His41, Leu50, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, Leu167, Pro168, Gln189

5.3.2 Drug-like properties

The eight molecules selected from both docking methods were assessed for Lipinski's rule of five and ADME properties. The results shown in **Table 5.2** reveal that the selected molecules were within the acceptable range of synthetic accessibility (score less than 5.93), TPSA was between 20 and 140 Å, lipophilicity; expressed as cLogP was less than 4.4, and water solubility expressed as Log *S* shows that most molecules are soluble or moderately soluble in water. The skin permeation possibility expressed as Log Kp was also reasonable, indicating the possibility of skin permeation. The ADME properties also lie within the range of acceptable values. Based on these results, all the eight hit molecules were selected for *in silico* validation using MD simulations.

Table 5.2. Drug-like properties of reference and screened NCI natural compounds.

S. No	Compound	TPSA $Å^2$	$LogP_{o/w}$	Log K_p (cm/s)	Synthetic accessibility
1	N3	193.39	2.48	-8.40	6.79
2	α-ketoamide	172.80	1.72	-8.15	5.54
3	NSC5113	100.13	4.40	-4.90	4.23
4	NSC107067	124.54	2.55	-5.28	5.93
5	NSC11926	110.81	2.30	-5.85	2.63
6	NSC22842	83.81	2.46	-5.98	3.16
7	NSC36398	127.45	0.52	-5.79	3.52
8	NSC44175	80.92	3.16	-5.43	2.21
9	NSC281245	72.83	3.78	-5.94	5.42
10	NSC94600	88.14	2.71	-5.86	3.41

Standard values for reference (Ertl P et al., 2000; Daina et al., 2014; Daina et al., 2017; Potts et al., 1992) TPSA = 0 to 140 $Å^2$, Log $P_{o/w}$ = -4.0 to 5.6, Log K_p= -6.1 to -0.19 cm/s, Synthetic accessibility scale = 1 to 10

5.3.3 Molecular dynamics simulations

Classical MD simulations of all the selected molecular systems; apo SARS-CoV-2 main protease, complexes with inhibitors N3, α-ketoamide and the screened-in molecules was performed using GROMACS 5.1.4 for 150 ns. Out of eight screened-in hit molecules, four molecules (NSC36398, NSC281245, NSC44175 and NSC11926) showed stability at the active site of SARS-CoV-2 main protease as shown in **Figure 5.4**. The covalent bond between the N3 inhibitor and active site residue Cys145 was not observed during the MD simulations because the Amber99sb force filed cannot account for a covalent bond formation.

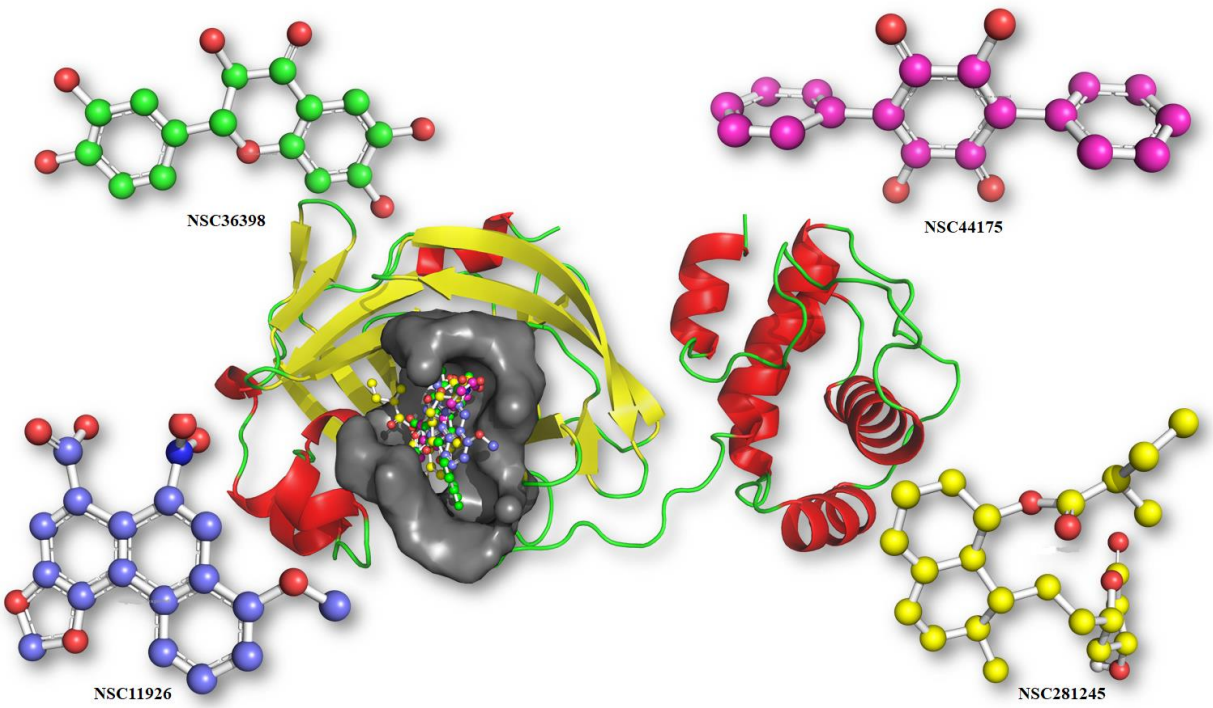


Figure 5.4: Location of the stabilized hit molecules in the SARS-CoV-2 main protease active site throughout MD simulations.

Studies on natural compounds-like bioactive molecules are reported as inhibitors of SARS-CoV-2 drug targets. For example, assafoetidnol A, conferol, farnesiferol B, sesamin, sesaminol, sesamolin show potential activity in targeting main protease, spike protein, and human ACE-2 receptors (Natesh, Jagadish, et al. 2021). Plant based natural compounds such as apigenin, coriandrin, curcumin, glabridin and oleanolic acid (Sampangi-Ramaiah et al., 2020, Verma et al., 2020) have been reported as main protease inhibitors, and some of the spice molecules (piperine, capsaicin, gingerol and terpinen-4-ol) (Rout et al., 2020) have been shown to bind spike protein and main protease. All these studies describe those natural compounds showing good inhibitory activity on SARS-CoV-2 proteins. The inhibition of these target proteins may lead to either

attenuation of viral replication or reduce the infectivity of this virus. In this work it was observed that the screened molecules from natural compounds NCI database show good binding affinity and nonbonding interactions with SARS-CoV-2 main protease. The results demonstrated that protein attains stability when it binds with screened-in hit molecules and maintain the hydrogen bonding interactions with important amino acid residues compared with the reference molecules (N3 and α-ketoamide). The amino acid Cys145 shows covalent interactions with reference molecules, which also maintains distance with screened-in hit molecules throughout the MD simulations. Cys145 interaction is most significant in inhibition of the drug inside the active site of SARS-CoV-2 main protease. The RMSD plots in **Figure 5.5A** revealed that the structures attained stability within the first 10 ns of MD simulations. The main protease when complexed with NSC36398 showed greater RMSD (~ 0.3 nm) among all the systems studied. The N3, α-ketoamide, NSC44175, NSC281245 bound main protease displayed lower RMSD (~ 0.22 nm) indicating greater structural stability of these four complexes. The screened-in hit molecules NSC11926, NSC281245 and NSC44175 exhibit lower RMSD (lower than 0.1 nm) whereas the reference molecules N3 and α-ketoamide showed relatively higher RMSD as shown in **Figure 5.5B**. The RMSF plots analyzed the residual fluctuations of protein during MD simulations. Higher fluctuations are observed in the regions; Asp153-Val157 and Asn221-Thr225 that are away from the active site and dimer interface of the SARS-CoV-2 main protease. The region, Leu50-Arg60 that contributes to the S2 subsite undergoes fluctuations up to 0.25 nm, the region Glu270-Gly283 located at the intersubunit interface shows fluctuations between 0.2 to 0.3 nm in all the molecular systems studied. The residues located in the region Cys117-Pro122 show greater fluctuations in N3 and α-ketoamide complexed proteins, Ser139-Cys145 (S1’ subsite) region also shows higher

fluctuations in the N3 binding protein. These regions of fluctuations in the apo and complexed SARS-CoV-2 main protease are shown in the RMSF plots, **Figure 5.5C**.

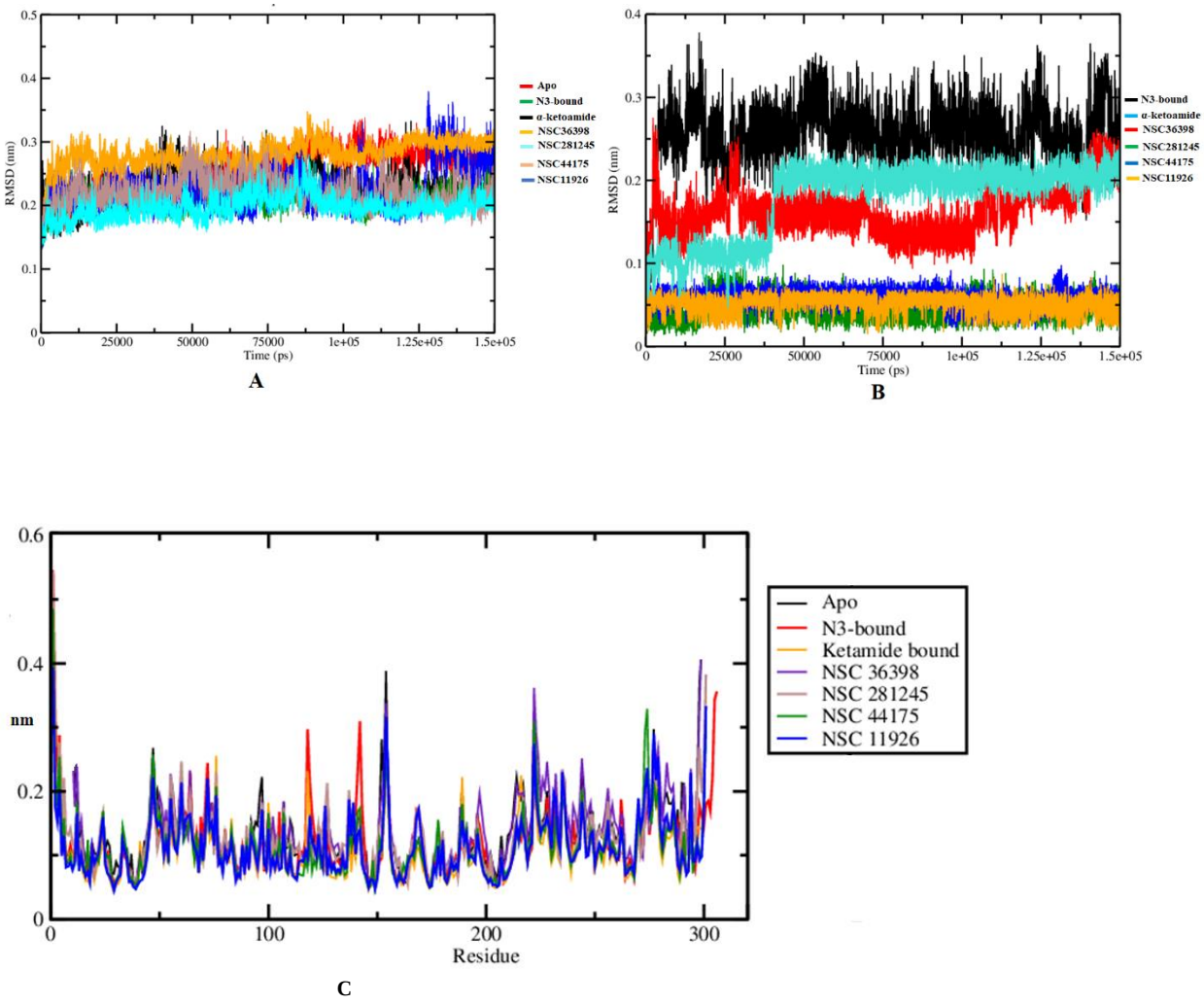
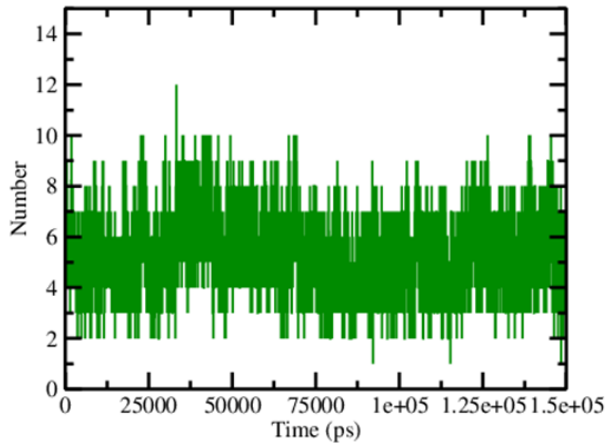


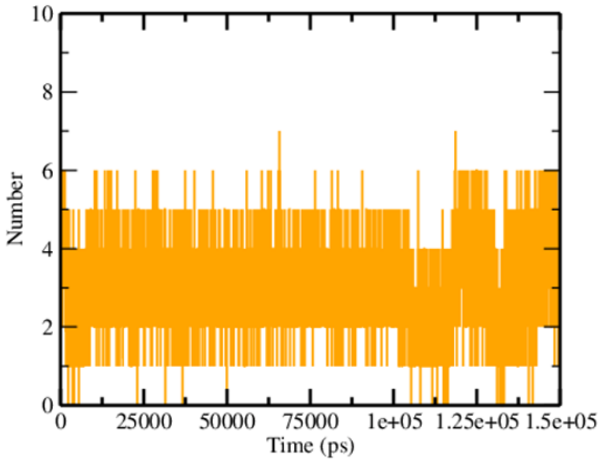
Figure 5.5: A) RMSD of apo SARS-CoV-2 main protease and when complexed with N3, α-ketoamide, NSC36398, NSC281245, NSC44175 and NSC11926. **B)** RMSD of N3, α-ketoamide, NSC36398, NSC281245, NSC44175 and NSC11926. **C)** RMSF plot of SARS-CoV-2 main protease apo, and when complexed with N3, α-ketoamide and screened-in molecules.

The Rg provides information about the compactness of protein throughout the MD simulations. It was observed that the apo and NSC281245 complexed SARS-CoV-2 main protease

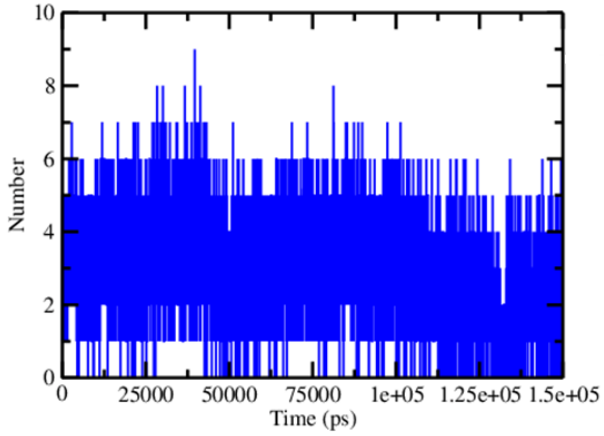
have a relatively higher Rg among all the systems studied. The stability of reference and screened molecules in the protein active site was analysed by comparing the initial and average structures, different types of nonbonding interactions were measured within 5 Å around the ligand. The intermolecular interactions in the structures between the protein-reference and screened-in molecules before and after MD simulations showed that the N3-inhibitor maintains interactions with Gly143, Ser144, Cys145, Glu166, Glu189, Gln192; and α-ketoamide also has interactions with Leu141, Ser144, Glu166, His164 and Gln189 throughout MD simulations. The protein complexed with NSC36398 made hydrogen bonding interactions with Leu141, Ser144, His163, Glu166, Arg188, and Gln189; NSC281245 with Ser46, Ser144, Cys145, and Glu166; NSC44175 with Gly143, Cys145, and Glu166 and NSC281245 with Ser144, Cys145, and Glu166 are retained during the MD simulations. The hydrogen bonding interactions of all molecular systems throughout MD simulations are shown in **Figure 5.6**



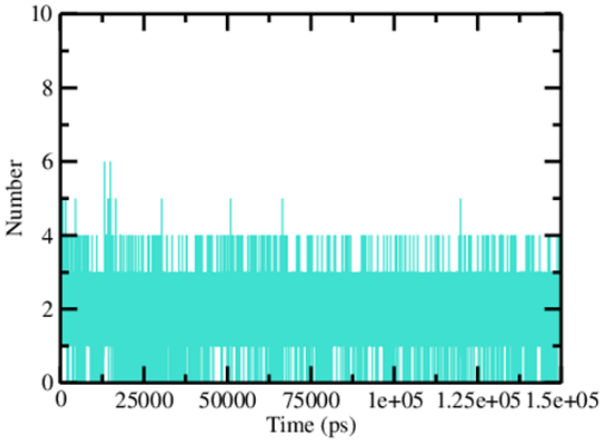
N3-inhibitor



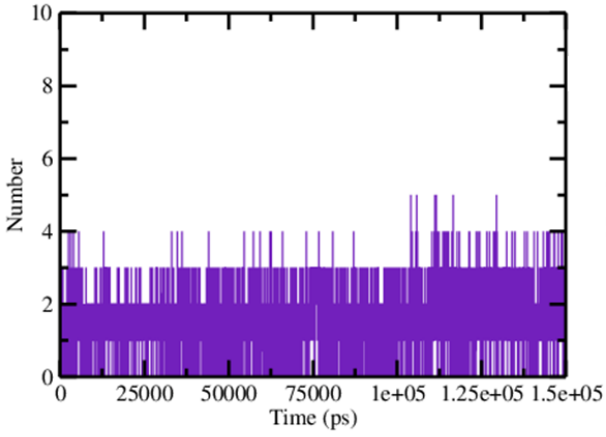
α- ketoamide



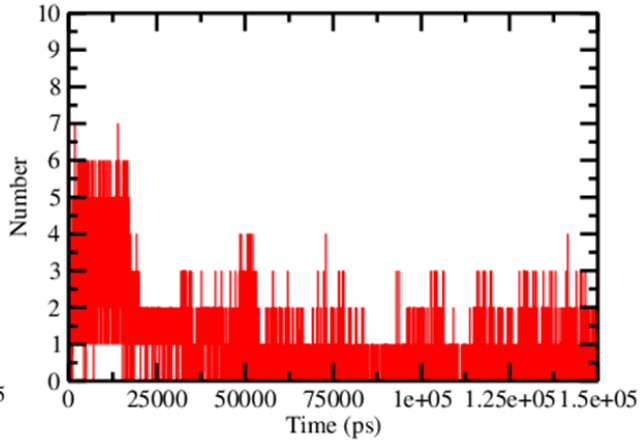
NSC36398



NSC281245



NSC44175



NSC11926

Figure 5.6: Hydrogen bonding interaction plot of SARS-CoV-2 main protease with reference and screened-in hit molecules.

5.3.4 Normal mode and mechanical stiffness analysis

The NMA is a fast and simple method to calculate protein flexibility (Alexandrov et al., 2005) involving atomic fluctuations. It reveals the structural variations and mobility in protein which are a collection of micro-ensemble states fluctuating about thermodynamically stable states. The RMSF plots revealed certain flexible regions in SARS-CoV-2 main protease during the MD simulations. To further confirm this observation, the NMA for all molecular systems were performed. Ten normal modes were obtained for each system from MD simulations trajectories, the first mode was selected and the structural variations were compared with apo structure of main protease. The regions that displayed higher RMSF (Leu50-Arg60, Asn221-Thr225, Glu240-Asp245 and Glu270-Gly283), also display normal modes with higher mobility in the presence of screened molecules as shown in **Figure 5.7**.

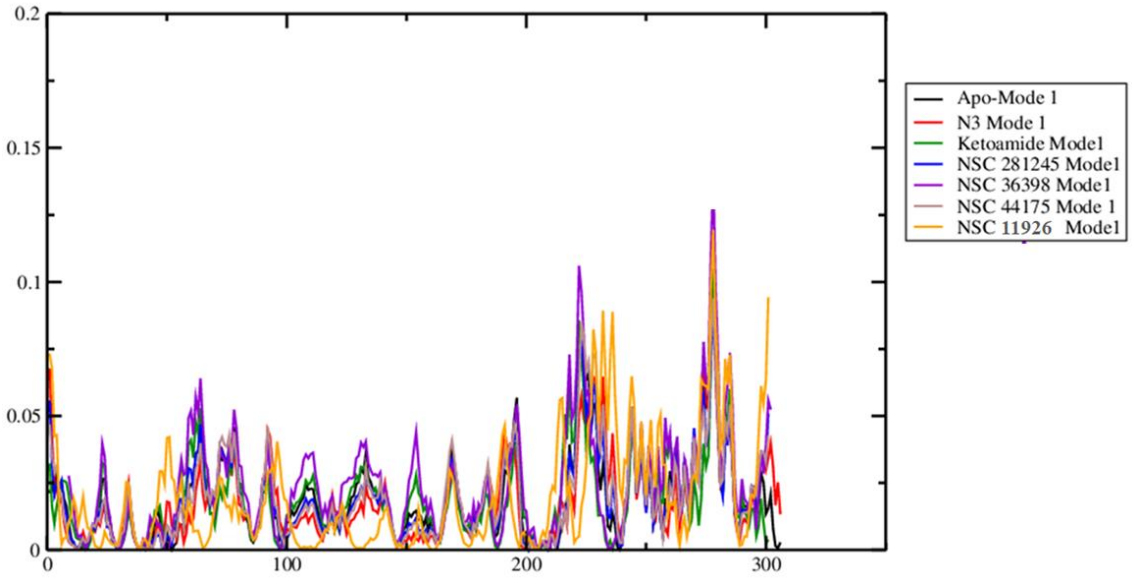


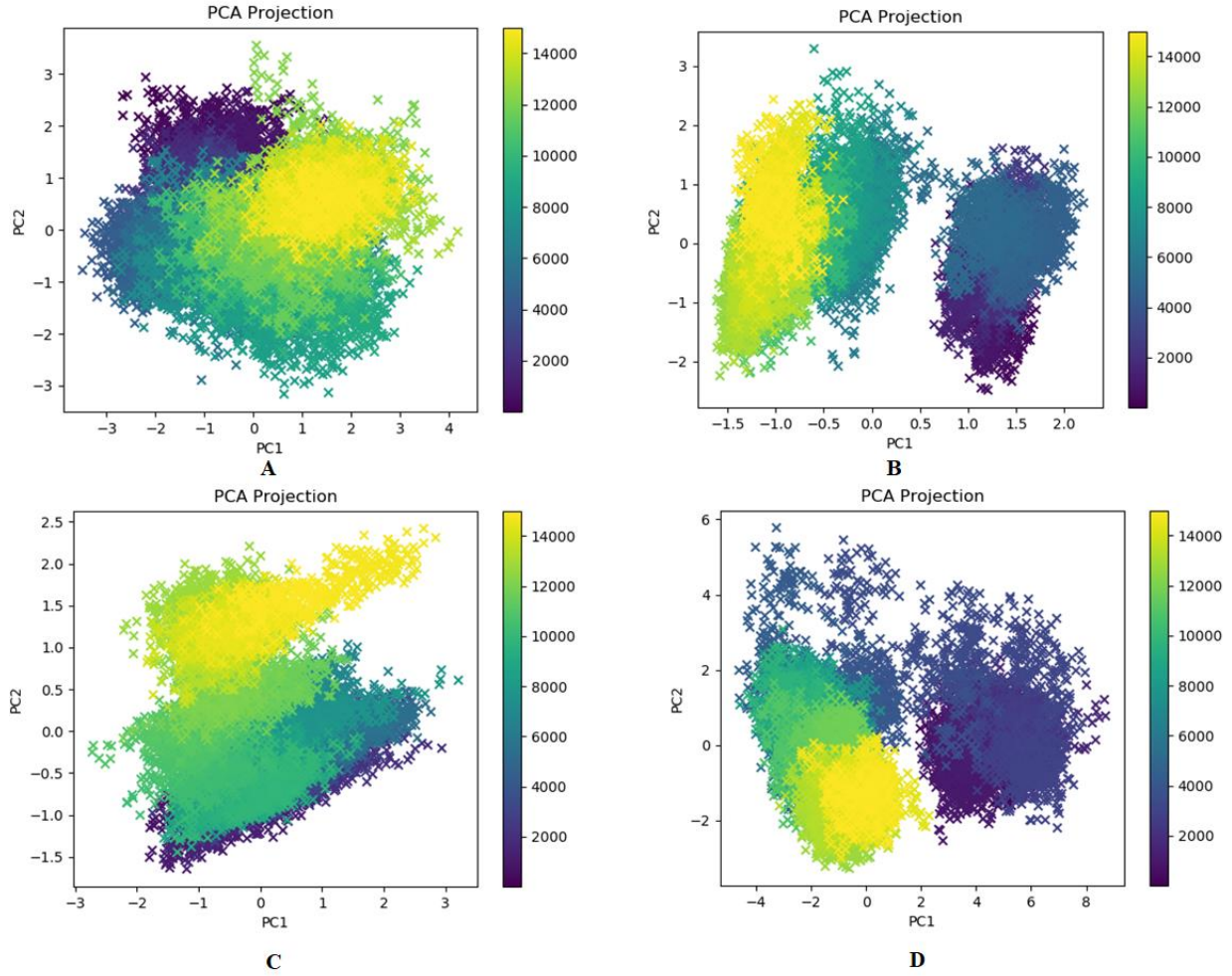
Figure 5.7: Mobility plot of NMAs in SARS-CoV-2 main protease apo and binding with N3, α-ketoamide and screened-in hit molecules.

This study explains the conformational changes in apo and ligand bound complexes of main protease and indicate relatively higher flexibility in domain III. The mechanical stiffness plots are useful to identify the anisotropic response of the protein structure to external perturbations, and the determination of weak and strong pairs of interactions that depend on the direction of the external force (Eyal & Bahar, 2008). Lower mechanical stiffness is indicative of the weak regions and higher mechanical stiffness is indicative of strong regions. In all the molecular systems studied, structural deformations were noted when compared with apo protein. From these plots, it was observed that the regions (Leu50-Arg60, Asn221-Thr225, Glu240-Asp245 and Glu270-Gly283) exhibit lower effective stiffness in all molecular systems. In the mean plots of mechanical stiffness, the effective spring constant value for fluctuating regions of residues was greater than 8 k (a.u) and was larger than 12 k (a.u) for the stable regions in proteins. These values indicated that the elastic nature of regions Leu50-Arg60, Asn221-Thr225, Glu240-Asp245 and Glu270-Gly283 is higher in the all molecular systems throughout MD simulations. From the results of mechanical stiffness and NMA it was proposed that the regions of residues Leu50-Arg60, Asn221-Thr225, Glu240-Asp245 and Glu270-Gly283 in the protein exhibit mechanically weak behavior. These large deviations of conformational changes indicated the elastic nature of protein in all systems studied.

5.3.5 Principal component analysis

PCA deciphers the conformational changes in a protein as a function of time from the MD simulations trajectories. The PCA scatter plots of all molecular systems studied is shown in **Figure 5.8**. The conformational changes of the SARS-CoV-2 main protease in apo form, N3, α-ketoamide

and natural products bound molecular systems were monitored. The Cα-atoms distribution is greater in NSC36398 bound molecular system which indicates that greater conformational changes of protein are observed. This demonstrated that the conformational distributions of main protease bound with NSC36398 was remarkably different from other molecular systems. The frequencies of PCA scatter plots were quantified and the highest-frequency is observed in NSC36398 bound main protease. These results indicated that SARS-CoV-2 main protease bound with NSC36398 displayed higher protein conformational changes compared to other molecular systems.



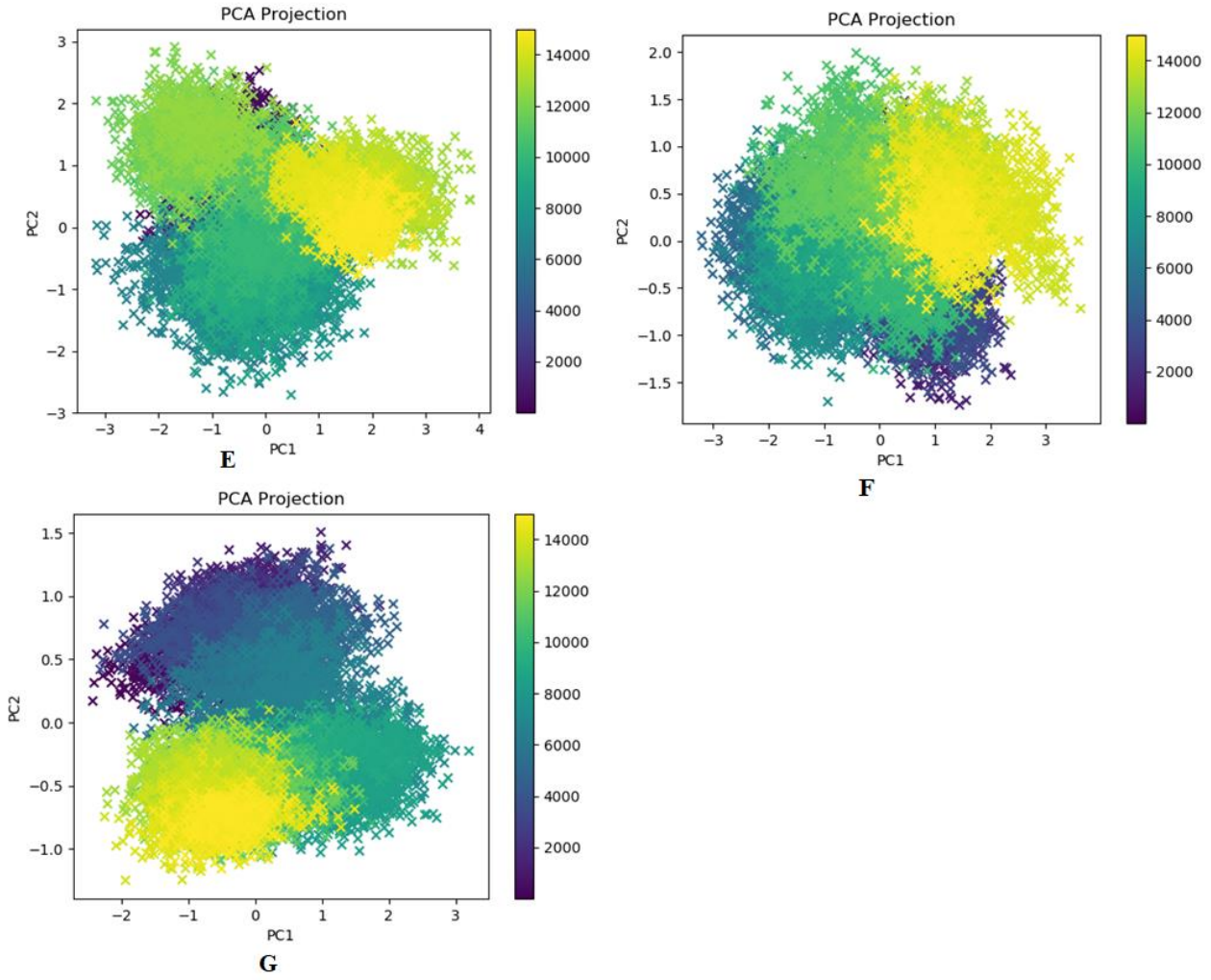


Figure 5.8. PCA scatterplots of SARS-CoV-2 main protease apo (A), complex with N3 (B), α-ketoamide (C), NSC36398 (D), NSC281245 (E), NSC44175 (F) and NSC11926 (G).

5.3.6 Binding free energy and residue-wise contribution analysis

The binding free energies of the reference and screened natural products calculated using MM-PBSA and LIE methods are shown in **Table 5.3**. The contributions from van der Waals, electrostatic and polar solvation energies for MM-PBSA binding free energies show compatibility with each other and reference molecules already reported. The binding free energies for N3 and α-ketoamide inhibitors in complex with SARS-CoV-2 main protease were -150.06 kJ/mol and -90.11

kJ/mol, respectively. The binding energies observed from AutoDock Vina and the MM-PBSA scores observed in this work are in correspondence with previous reports (Keretsu et al., 2020). The binding free energies for the natural products selected along with reference were NSC281245 (-133.79 kJ/mol), NSC11926 (-93.22 kJ/mol), NSC44175 (-81.97 kJ/mol) and NSC36398 (-70.75 kJ/mol) as shown in **Figure 5.9**.

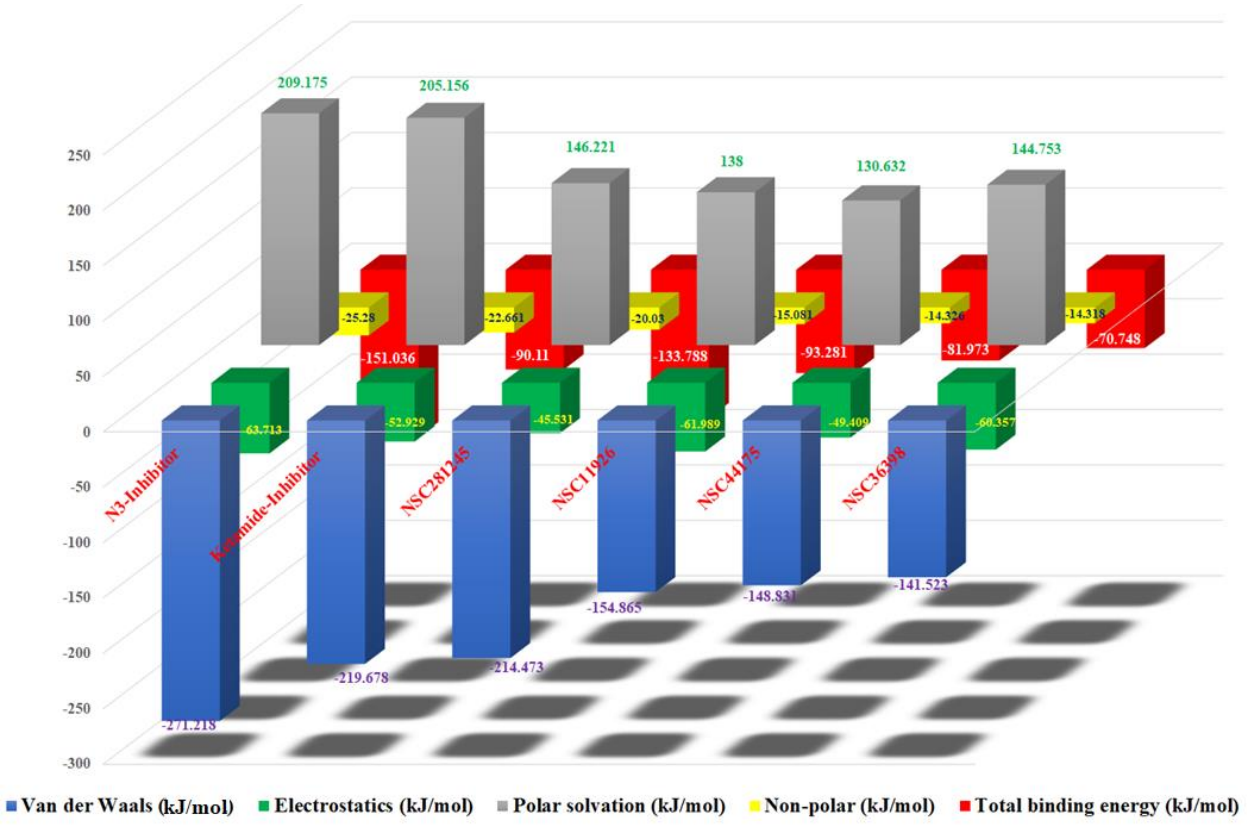


Figure 5.9: Binding free energy plot of SARS-CoV-2 main protease with screened-in hit molecules.

The binding site residues Leu27, His41, Gly143, Ser144, Cys145, His164, Met165 and Glu166 contribute to the highest binding free energies in all the complexes studied. Pro168 contributes to binding free energy in both the reference molecules as shown in **Figure 5.10**. The LIE values show that the binding free energies of reference and screened-in hit molecules with

SARS-CoV-2 main protease N3 shows (-140.69 kJ/mol), α-ketoamide (-143.78 kJ/mol), NSC281245 (-117.63 kJ/mol), NSC36398 (-83.71 kJ/mol), NSC11926 (-78.65 kJ/mol), and NSC44175 (-73.82 kJ/mol).

Table 5.3: Comparison of docking scores and binding free energy (MM-PBSA and LIE) of reference and screened-in molecules binds with SARS-CoV-2 main protease.

S. No	Name of the compound	AutoDock Vina score (kcal/mol)	CDOCKER score			Δ_{LIE} (kJ/mol)	$\Delta G_{(MM-PBSA)}$ (kJ/mol)
			-PLP	-PLP2	-PMF04		
1.	N3	-7.8	119.02	71.73	44.22	-140.64+/0.01	-151.04 +/- 0.86
2.	α-ketoamide	-9.6	78.98	38.17	46.39	-143.76+/0.02	-90.11 +/- 0.55
4.	NSC281245	-8.0	109.47	94.08	54.88	-117.627+/0.01	-133.79 +/- 0.55
5.	NSC11926	-7.8	86.04	40.27	41.69	-78.65+/0.01	-93.28 +/- 0.47
3.	NSC36398	-8.1	95.76	97.49	18.22	-83.71+/0.01	-70.75 +/- 0.39
6.	NSC44175	-7.8	98.62	67.12	42.11	-73.82+/0.00	-81.97 +/- 0.49

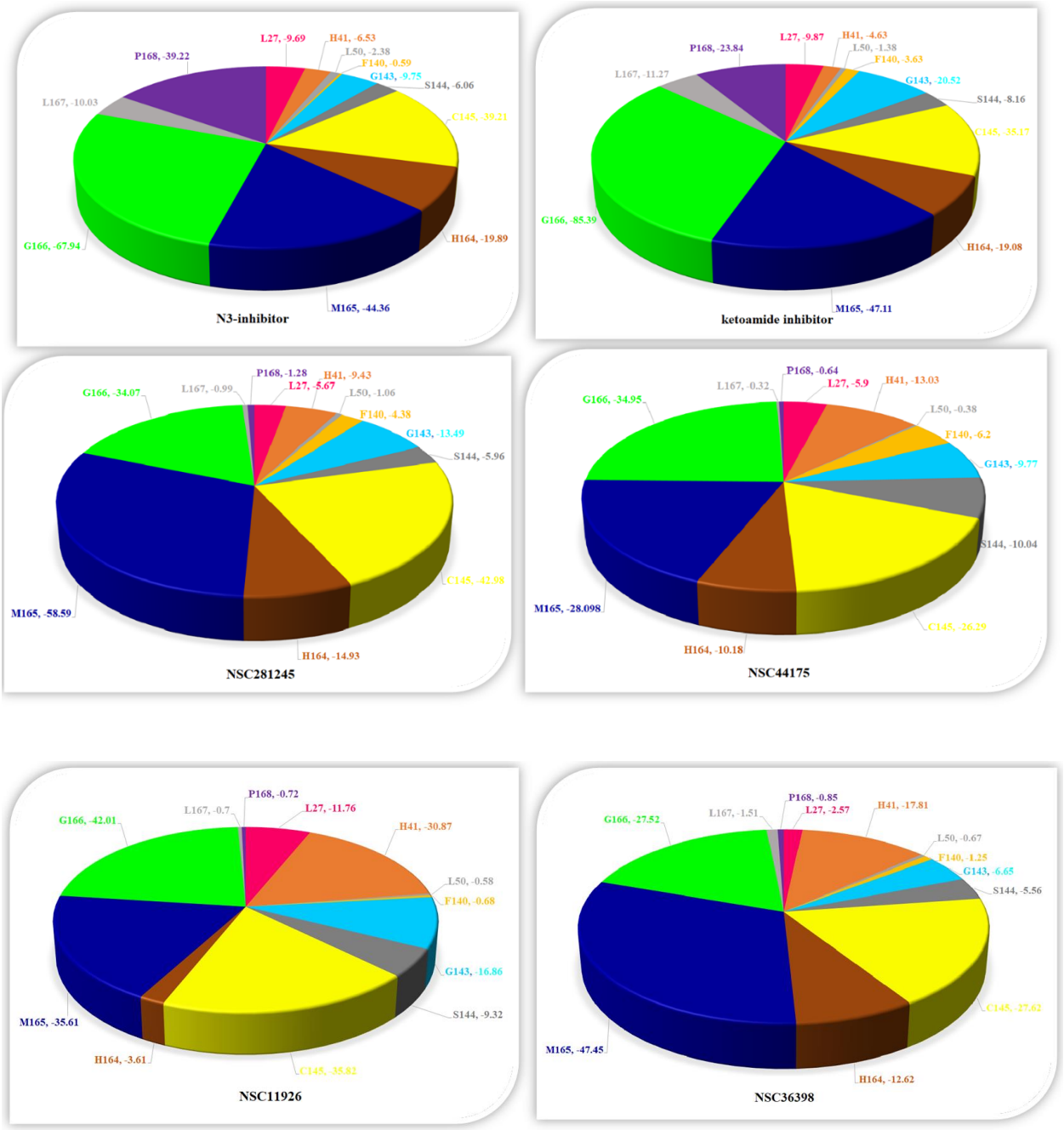


Figure 5.10: Residual contribution (kJ/mol) of SARS-CoV-2 main protease with reference and screened hit molecules.

5.4 Conclusions

CADD methodologies can be used effectively to speed-up the process of developing therapeutic agents for the treatment of COVID-19 disease. In this study, the docking based virtual screening of NCI diversity set of natural compounds were used to identify the potential hits for SARS-CoV-2 main protease. Molecular docking and MD simulations were carried out to study the binding interactions between protein and ligand molecules. Binding free energies were calculated to identify the potential hit molecules for SARS-CoV-2 main protease. Four compounds that showed good binding affinity and stability in the protein active site throughout 150 ns MD simulations were identified in this study. The amino acid residues Cys145, Met165 and Glu166 have high contribution to the binding free energies of all the molecules studied. In all molecular systems studied, certain regions in SARS-CoV-2 main protease domain III showed greater flexibility and NSC36398 bound protein displayed higher protein conformational changes revealing the molecular mechanisms of protein-NCI natural products interactions. All natural compounds studied also displayed drug-like properties indicating their suitability as probable inhibitors for SARS-CoV-2 main protease.

References

Abduraman, M. A.; Hariono, M.; Yusof, R.; Rahman, N. A.; Wahab, H. A.; Tan, M. L. Development of a NS2B/NS3 Protease Inhibition Assay Using AlphaScreen® Beads for Screening of Anti-Dengue Activities. *Heliyon* **2018**, *4*, e01023.

Abola, E. E.; Bernstein, F. C.; Koetzle, T. FThe Protein Data Bank. In *Neutrons in Biology*; Springer: Boston, MA. **1984,** pp 441–441.

Abriata, L. A.; Tamò, G. E.; & Dal Peraro, M. A further leap of improvement in tertiary structure prediction in CASP13 prompts new routes for future assessments. *Proteins: Struct. Funct. Genet.* **2019**, *87*, 1100-1112.

Ahmad, N.; Badshah, S. L.; Junaid, M.; Ur Rehman, A.; Muhammad, A.; Khan, K. Structural Insights into the Zika Virus NS1 Protein Inhibition Using a Computational Approach. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3004–3011.

Al-Bari, M. A. A. Chloroquine Analogues in Drug Discovery: New Directions of Uses, Mechanisms of Actions and Toxic Manifestations from Malaria to Multifarious Diseases. *J. Antimicrob. Chemother.* **2015**, *70*, 1608–1621.

Alexandrov, V.; Lehnert, U.; Echols, N.; Milburn, D.; Engelman, D.; Gerstein, M. Normal Modes for Predicting Protein Motions: A Comprehensive Database Assessment and Associated Web Tool. *Protein Sci.* **2005**, *14*, 633–643.

Almlöf, M.; Brandsdal, B. O.; Aqvist, J. Binding Affinity Prediction with Different Force Fields: Examination of the Linear Interaction Energy Method: Binding Affinity Prediction with Different Force Fields. *J. Comput. Chem.* **2004**, *25*, 1242–1254.

Altschul, S. F.; Wootton, J. C.; Gertz, E. M.; Agarwala, R.; Morgulis, A.; Schäffer, A. A.; Yu, Y.-K. Protein Database Searches Using Compositionally Adjusted Substitution Matrices. *FEBS J.* **2005**, *272*, 5101–5109.

Anfasa, F.; Siegers, J. Y.; Van Der Kroeg, M.; Mumtaz, N.; Stalin Raj, V.; De Vrij, F. M. & van Riel, D. Phenotypic Differences between Asian and African Lineage Zika Viruses in Human Neural Progenitor Cells. *MSphere* **2017**, *2*, e00292-309.

Anfinsen, C. B. Principles that govern the folding of protein chains. *Science*. **1973**, *181*, 223-230.

Åqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng. Des. Sel.* **1994**, *7*, 385–391.

Ashburn, T. T.; Thor, K. B. Drug Repositioning: Identifying and Developing New Uses for Existing Drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673–683.

Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **2001**, *80*, 505–515.

Attwood, T. K.; Bradley, P.; Flower, D. R.; Gaulton, A.; Maudling, N.; Mitchell, A. L.; Moulton, G.; Nordle, A.; Paine, K.; Taylor, P.; Uddin, A.; Zygouri, C. PRINTS and Its Automatic Supplement, PrePRINTS. *Nucleic Acids Res.* **2003**, *31*, 400–402.

Axel, D. B. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *The Journal of chemical physics* **1993**, *98*, 5648–5652.

Badshah, S. L.; Ahmad, N.; Ur Rehman, A.; Khan, K.; Ullah, A.; Alsayari, A. & N Mabkhot, Y. Molecular Docking and Simulation of Zika Virus NS3 Helicase. *BMC chemistry* **2019**, *13*, 1–8.

Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chem. Rev.* **2010**, *110*, 1463–1497.

Bakan, A.; Meireles, L.M.; Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*. **2011**, 27, 1575-1577.

Banerjee, R.; Perera, L.; Tillekeratne, L. M. V. Potential SARS-CoV-2 Main Protease Inhibitors. *Drug Discov. Today* **2021**, *26*, 804–816.

Barrows, N. J.; Campos, R. K.; Powell, S. T.; Prasanth, K. R.; Schott-Lerner, G.; Soto-Acosta, R.; Galarza-Muñoz, G.; McGrath, E. L.; Urrabaz-Garza, R.; Gao, J.; Wu, P.; Menon, R.; Saade, G.; Fernandez-Salas, I.; Rossi, S. L.; Vasilakis, N.; Routh, A.; Bradrick, S. S.; Garcia-Blanco, M. A. A Screen of FDA-Approved Drugs for Inhibitors of Zika Virus Infection. *Cell Host Microbe* **2016**, *20*, 259–270.

Basavannacharya, C.; Vasudevan, S. G. Suramin Inhibits Helicase Activity of NS3 Protein of Dengue Virus in a Fluorescence-Based High Throughput Assay Format. *Biochem. Biophys. Res. Commun.* **2014**, *453*, 539–544.

Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Eddy, S. R.; Griffiths-Jones, S.; Howe, K. L.; Marshall, M.; Sonnhammer, E. Sonnhammer ELL:The Pfam Protein Families Database. *Nucleic Acids Res* **2002**, *30*, 276–280.

Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci.* **2019**, *20*, 2783.

Baxevanis, A.D.; Bader, G.D.; Wishart, D.S. Bioinformatics. John Wiley & Sons. **2020**.

Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Wheeler, D. L. Wheeler DL:GenBank. *GenBank. Nucleic Acids Res* **2003**, *31*, 23–27.

Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083-90.

Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*; Springer Netherlands: Dordrecht, **1981**; pp 331–342.

Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. The protein data bank. *Nucleic acids research*. **2000**, 28, 235-242.

Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The Worldwide Protein Data Bank (WwPDB): Ensuring a Single, Uniform Archive of PDB Data. *Nucleic Acids Res.* **2007**, *35* (Database issue), D301-3.

Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980.

Beveridge, D. L.; Ravishanker, G. Molecular Dynamics Studies of DNA. *Curr. Opin. Struct. Biol.* **1994**, *4*, 246–255.

Bowen, J. R., Quicke, K. M., Maddur, M. S., O'Neal, J. T., McDonald, C. E., Fedorova, N. B., ... & Suthar, M. S. Zika virus antagonizes type I interferon responses during infection of human dendritic cells. *PLoS pathogens*. **2017**, 13, e1006164.

Brandsdal, B. O.; Osterberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J. Free Energy Calculations and Ligand Binding. *Adv. Protein Chem.* **2003**, *66*, 123–158.

Brasil, P.; Pereira, J. P.; Moreira, M. E.; Ribeiro Nogueira, R. M.; Damasceno, L.; Wakimoto, M. & Nielsen-Saines, K. Zika Virus Infection in Pregnant Women in Rio de Janeiro. *N. Engl. J. Med.* **2016**, *375*, 2321–2334.

Britannica.com. https://www.britannica.com/science/filovirus. (accessed 2022-12-30).

Burley, S. K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G. V. & Zhuravleva, M. RCSB Protein Data Bank: Powerful New Tools for Exploring 3D Structures of Biological Macromolecules for Basic and Applied Research and Education in Fundamental Biology,

Biomedicine, Biotechnology, Bioengineering and Energy Sciences. *Nucleic acids research* **2021**, *49*, D437–D451.

Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.A.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem*. **1983**, 4, 187-217.

Burley, S. K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G. V.; ... & Zhuravleva, M. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic acids research*. **2021**, 49, D437-D451

Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

Cao-Lormeau, V.-M.; Musso, D. Emerging Arboviruses in the Pacific. *Lancet* **2014**, *384*, 1571–1572.

Cao-Lormeau, V. M.; Roche, C.; Teissier, A.; Robin, E.; Berry, A. L.; Mallet, H. P. *& Musso, D. Zika Virus, French Polynesia, South Pacific*; 2013.

Carteaux, G.; Maquart, M.; Bedet, A.; Contou, D.; Brugières, P.; Fourati, S. & Mekontso Dessap, A. Zika Virus Associated with Meningoencephalitis. *New England Journal of Medicine* **2016**, *374*, 1595–1596.

Caspar, D. L.; Klug, A. Physical Principles in the Construction of Regular Viruses. *Cold Spring Harb. Symp. Quant. Biol.* **1962**, *27*, 1–24.

Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discov. Today* **2009**, *14*, 676–683.

Clinical Guidance for Healthcare Providers for Prevention of Sexual Transmission of Zika Virus. *Centers for Disease Control and Prevention* **2017**.

Cheatham, T. E., Iii; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Am. Chem. Soc.* **1995**, *117*, 4193–4194.

Chen, H.; Lyne, P.D.; Giordanetto, F.; Lovell, T; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model*. **2006**, 46, 401-415.

Chen J.; Liu D.; Liu L.; Liu P.; Xu Q.; Xia L.; Ling Y.; Huang D.; Song S.; Zhang D.; Qian Z.; Li T.; Shen Y.; Lu H. A pilot study of hydroxychloroquine in treatment of patients with moderate COVID-19. *Zhejiang Da Xue Xue Bao Yi Xue Ban.* **2020**, *49*, 215–219.

Chen, R.; Tong, W.; Mintseris, J.; Li, L.; Weng, Z. ZDOCK predictions for the CAPRI challenge. Proteins: *Struct., Funct., Bioinf.* **2003**, 52, 68-73.

Chen, S. H. *Spectroscopy in Biology and Chemistry: Neutron, X-Ray, Laser*; Laser. Academic Press, **2017**.

Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M. & Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of medicinal chemistry.* **2014**, *57*, 4977–5010.

Childers, M.C.; Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* **2017**, *2*, 9-33.

Colovos, C.; Yeates, T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**, *2*, 1511–1519.

Corpet, F.; Servant, F.; Gouzy, J.; & Kahn, D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic acids research.* **2000**, *28*, 267-269.

Crick, F. H.; & Watson, J. D. Structure of small viruses. *Nature.* **1956**, *177*, 473-475.

Da Silva, A.W.S.; Vranken, W.F. ACPYPE-Antechamber python parser interface. *BMC Res. Notes.* **2012**, *5*, 1-8.

Dahiyat, B.I.; Mayo, S.L. De novo protein design: fully automated sequence selection. *Science*. **1997**, *278*, 82-87.

Daina, A., Michielin, O., & Zoete, V. iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Model.* **2014**, *54*, 3284-3301.

Daina, A., Michielin, O., & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep*. **2017**, *7*, 1-13.

Dallakyan, S.; Olson, A. J. Small-Molecule Library Screening by Docking with PyRx. *Methods Mol. Biol.* **2015**, *1263*, 243–250.

Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N· Log (N) Method for Ewald Sums in Large Systems. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

Davidson, A.; Slavinski, S.; Komoto, K.; Rakeman, J.; Weiss, D. Suspected Female-to-Male Sexual Transmission of Zika Virus - New York City, 2016. *MMWR Morb. Mortal. Wkly. Rep.* **2016**, *65*, 716–717.

Davidson, R. B.; Hendrix, J.; Geiss, B. J.; McCullagh, M. Allostery in the Dengue Virus NS3 Helicase: Insights into the NTPase Cycle from Molecular Simulations. *PLoS Comput. Biol.* **2018**, *14*, e1006103.

De Groot, N. S.; Armaos, A.; Graña-Montes, R.; Alriquet, M.; Calloni, G.; Vabulas, R. M.; Tartaglia, G. G. RNA Structure Drives Interaction with Proteins. *Nature communications* **2019**, *10*, 1–13.

Deckard, D. T.; Chung, W. M.; Brooks, J. T.; Smith, J. C.; Woldai, S.; Hennessey, M.; Mead. Male-to-Male Sexual Transmission of Zika Virus-Texas. *Morbidity and Mortality Weekly Report* **2016**, *65*, 372–374.

Delano, W. L. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter on protein crystallography* **2002**, *40*, 82–92.

Dick, G. W. A.; Kitchen, S. F.; Haddow, A. J. Zika Virus (I). Isolations and Serological Specificity. *Trans. R. Soc. Trop. Med. Hyg.* **1952**, *46*, 509–520.

Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J. Comput. Aided Mol. Des.* **2006**, *20*, 647–671.

Draper, D. E. Protein-RNA recognition. *Annu. Rev. Biochem.* **1995**, *64*, 593-620.

Driggers, R. W.; Ho, C. Y.; Korhonen, E. M.; Kuivanen, S.; Jääskeläinen, A. J.; Smura, T. & Vapalahti, O. Zika Virus Infection with Prolonged Maternal Viremia and Fetal Brain Abnormalities. *N. Engl. J. Med.* **2016**, *374*, 2142–2151.

Duan, W.; Song, H.; Wang, H.; Chai, Y.; Su, C.; Qi, J.; Shi, Y.; Gao, G. F. The Crystal Structure of Zika Virus NS 5 Reveals Conserved Drug Targets. *EMBO J.* **2017**, *36* (7), 919–933.

Duffy, M. R.; Chen, T. H.; Hancock, W. T.; Powers, A. M.; Kool, J. L.; Lanciotti, R. S. & Hayes, E. B. Zika Virus Outbreak on Yap Island, Federated States of Micronesia. *N. Engl. J. Med.* **2009**, *360*, 2536–2543.

Eddy, S.R. Where did the BLOSUM62 alignment score matrix come from? *Nat. Biotechnol.* **2004**, *22*, 1035-1036.

Edgar, R.C.; Batzoglou, S. Multiple sequence alignment. *Curr Opin Struct Biol*. **2006**, *16*, 368-373.

Ekins, S.; Mietchen, D.; Coffee, M.; Stratton, T. P.; Freundlich, J. S.; Freitas-Junior, L.; ... & Andrade, C. Open drug discovery for the Zika virus. *F1000Research*. **2016**, 5.

Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.

Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminform*. **2009**, *1*, 1-11.

Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

Eyal, E.; Bahar, I. Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models. *Biophys. J.* **2008**, *94*, 3424–3435.

Eyal, E.; Lum, G.; & Bahar, I. The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics.* **2015**, *31*, 1487-1489.

Falquet, L.; Pagni, M.; Bucher, P.; Hulo, N.; Sigrist, C. J. A.; Hofmann, K.; Bairoch, A. The PROSITE Database, Its Status in 2002. *Nucleic Acids Res.* **2002**, *30*, 235–238.

Faria, N. R.; Azevedo, R. D. S. D. S.; Kraemer, M. U.; Souza, R.; Cunha, M. S.; Hill, S. C.; ... & Vasconcelos, P. F. Zika virus in the Americas: early epidemiological and genetic findings. *Science.* **2016**, *352*, 345-349.

Flores, J. K.; Ataide, S. F. Structural Changes of RNA in Complex with Proteins in the SRP. *Front. Mol. Biosci.* **2018**, *5*, 7.

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. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749.

Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. Gaussian 09, Revision A.02. **2009**.

Fukuli, K.; Yonezawa, T.; Shingu, H. A molecular orbital theory of reactivity in aromatic hydrocarbons. *J. Phys. Chem.* **1952**, *20*, 722–725.

Gabriel, E.; Ramani, A.; Karow, U.; Gottardo, M.; Natarajan, K.; Gooi, L. M.; Goranci-Buzhala, G.; Krut, O.; Peters, F.; Nikolic, M.; Kuivanen, S.; Korhonen, E.; Smura, T.; Vapalahti, O.; Papantonis, A.; Schmidt-Chanasit, J.; Riparbelli, M.; Callaini, G.; Krönke, M.; Utermöhlen, O.;

Gopalakrishnan, J. Recent Zika Virus Isolates Induce Premature Differentiation of Neural Progenitors in Human Brain Organoids. *Cell Stem Cell.* **2017**, *20*, 397-406.e5.

Gagnon, J. K.; Law, S. M.; Brooks, C. L., 3rd. Flexible CDOCKER: Development and Application of a Pseudo-Explicit Structure-Based Docking Method within CHARMM: Adding Receptor Flexibility Improves Protein-Ligand Docking Within CDOCKER. *J. Comput. Chem.* **2016**, *37*, 753–762.

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. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100-D1107.

Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol.* **1995**, *2*, 317–324.

Geng, H.; Chen, F.; Ye, J.; Jiang, F. Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 1162–1170.

Ghani, A.; Ramlan, N. S.; Firdaus-Raih, E. I. Drug ReposER: A Web Server for Predicting Similar Amino Acid Arrangements to Known Drug Binding Interfaces for Potential Drug Repositioning. *Nucleic Acids Res* **2019**, *47*, W350–W356.

Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; & Chong, J. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic acids research.* **2016**, *44*, D1045-D1053.

Gilson, M.K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins: Struct., *Funct., Bioinf.* **1988**, *4*, 7-18.

Goldsmith, C. S.; Miller, S. E. Modern Uses of Electron Microscopy for Detection of Viruses. *Clin. Microbiol. Rev.* **2009**, *22*, 552–563.

Gosline, J.; Lillie, M.; Carrington, E.; Guerette, P.; Ortlepp, C.; Savage, K. Elastic Proteins: Biological Roles and Mechanical Properties. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2002**, *357*, 121–132.

Gough, J.; Karplus, K.; Hughey, R.; Chothia, C. Assignment of Homology to Genome Sequences Using a Library of Hidden Markov Models That Represent All Proteins of Known Structure. *J. Mol. Biol.* **2001**, *313*, 903–919.

Govero, J.; Esakky, P.; Scheaffer, S. M., Fernandez, E., Drury, A., Platt, D. J., ... & Diamond, M. S. Zika virus infection damages the testes in mice. *Nature.* **2016**, *540*, 438-442.

Grindon, C.; Harris, S.; Evans, T.; Novik, K.; Coveney, P.; Laughton, C. Large-Scale Molecular Dynamics Simulation of DNA: Implementation and Validation of the AMBER98 Force Field in LAMMPS. *Philos. Trans. A Math. Phys. Eng. Sci.* **2004**, *362*, 1373–1386.

Groom, C.R.; Bruno, I.J.; Lightfoot, M.P.; Ward, S.C. The Cambridge structural database. Acta Crystallogr., *Sect. B: Struct. Sci., Cryst. Eng. Mater*. **2016**, *72*, 171-179.

Guex, N.; Peitsch, M.C.; Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis, **2009**, *30*, S162-S173.

Gutte, B. A Synthetic 70-Amino Acid Residue Analog of Ribonuclease S-Protein with Enzymic Activity. *J. Biol. Chem.* **1975**, *250*, 889–904.

Haft, D. H.; Selengut, J. D.; White, O. White O: The TIGRFAMs Database of Protein Families. *Nucleic Acids Res* **2003**, *31*, 371–373.

Halgren, T.A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377-389.

Hamel, R.; Ferraris, P.; Wichit, S.; Diop, F.; Talignani, L.; Pompon, J. & Missé, D. African and Asian Zika Virus Strains Differentially Induce Early Antiviral Responses in Primary Human Astrocytes. *Genetics and Evolution* **2017**, *49*, 134–137.

Hanson, R.M. Jmol–a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.* **2010**, *43*, 1250-1260.

Hardin, C.; Pogorelov, T.V.; Luthey-Schulten, Z. Ab initio protein structure prediction. *Curr. Opin. Struct. Biol*. **2002**, *12*, 176-181.

Baig, M. H.; Ahmad, K.; Roy, S.; Ashraf, J. M.; Adil, M.; Siddiqui, M. H.; Khan, S.; Kamal, M. A.; Provazník, I.; Choi, I. Computer Aided Drug Design: Success and Limitations. *Curr. Pharm. Des.* **2016**, *22*, 572–581.

Hatherley, R.; Brown, D.K.; Glenister, M.; Tastan Bishop, Ö. PRIMO: an interactive homology modeling pipeline. *PLoS One*. **2016**, *11*, e0166698.

He, D.; Gao, D.; Lou, Y.; Zhao, S.; & Ruan, S. A comparison study of Zika virus outbreaks in French Polynesia, Colombia and the State of Bahia in Brazil. *Sci. Rep*. **2017**, *7*, 1-6.

He, D.; Gao, D.; Lou, Y.; Zhao, S.; Ruan, S. A Comparison Study of Zika Virus Outbreaks in French Polynesia, Colombia and the State of Bahia in Brazil. *Sci. Rep.* **2017**, *7* (1), 273.

Hermann, T.; Westhof, E. Aminoglycoside Binding to the Hammerhead Ribozyme: A General Model for the Interaction of Cationic Antibiotics with RNA. *J. Mol. Biol.* **1998**, *276* (5), 903–912.

Hess, B.; Bekker, H.; Berendsen, H.J.; Fraaije, J.G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463-1472.

Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

Hessler, G.; Baringhaus, K.H. Artificial intelligence in drug design. *Molecules.* **2018**, 23, 2520.

Hetényi, C.; van der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* **2006**, *580*, 1447-1450.

Holm, L.; Laakso, L. M. Dali Server Update. *Nucleic Acids Res.* **2016**, *44*, W351-5.

Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson− Boltzmann Surface Area Method. *Molecular informatics* **2012**, *31*, 114–122.

Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725.

Huang, H.; Barker, W. C.; Chen, Y.; Wu, C. H. Wu CH: IProClass: An Integrated Database of Protein Family, Function and Structure Information. *Nucleic Acids Res* **2003**, *31*, 390–392.

Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33- 38.

Ioos, S.; Mallet, H.-P.; Leparc Goffart, I.; Gauthier, V.; Cardoso, T.; Herida, M. Current Zika Virus Epidemiology and Recent Epidemics. *Med. Mal. Infect.* **2014**, *44*, 302–307.
Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; & Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757-1768.

Irwin, J. J. Community Benchmarks for Virtual Screening. *J. Comput. Aided Mol. Des.* **2008**, *22*, 193–199.

Islami, F.; Goding Sauer, A.; Miller, K. D.; Siegel, R. L.; Fedewa, S. A.; Jacobs, E. J.; McCullough, M. L.; Patel, A. V.; Ma, J.; Soerjomataram, I.; Flanders, W. D.; Brawley, O. W.; Gapstur, S. M.; Jemal, A. Proportion and Number of Cancer Cases and Deaths Attributable to Potentially Modifiable Risk Factors in the United States: Potentially Preventable Cancers in US. *CA Cancer J. Clin.* **2018**, *68*, 31–54.

Jácome, R.; Campillo-Balderas, J. A.; Ponce de León, S.; Becerra, A.; Lazcano, A. Sofosbuvir as a Potential Alternative to Treat the SARS-CoV-2 Epidemic. *Sci. Rep.* **2020**, *10*, 9294.

Jain, R.; Coloma, J.; García-Sastre, A.; Aggarwal, A. K. Structure of the NS3 Helicase from Zika Virus. *Nat. Struct. Mol. Biol.* **2016**, *23*, 752–754.

Javorac, D.; Grahovac, L.; Manić, L.; Stojilković, N.; Anđelković, M.; Bulat, Z. & Djordjevic, A. B. An Overview of Safety Assessment of the Medicines Currently Used in the Treatment of COVID-19 Disease. *Food and Chemical Toxicology.* **2020**.

Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289–293.

Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859-1865.

John, B.; Sali, A. Comparative Protein Structure Modeling by Iterative Alignment, Model Building and Model Assessment. *Nucleic Acids Res.* **2003**, *31*, 3982–3992.

Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; & Madden, T. L. NCBI BLAST: a better web interface. *Nucleic acids research*. **2008**, *36*, W5-W9.

Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

Kandakatla, N.; Ramakrishnan, G. Ligand Based Pharmacophore Modeling and Virtual Screening Studies to Design Novel HDAC2 Inhibitors. *Adv. Bioinformatics.* **2014**, *2014*, 812148.

Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858.

Keretsu, S.; Bhujbal, S. P.; Cho, S. J. Rational Approach toward COVID-19 Main Protease Inhibitors via Molecular Docking, Molecular Dynamics Simulation and Free Energy Calculation. *Sci. Rep.* **2020**, *10*, 17716.

Kiernan, P.; Jelmayer, R. Zika Fears Imperil Brazil's Tourism Push. *The Wall Street Journal* **2016**, 1–8.

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. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202-13.

Kinjo, A. R.; Suzuki, H.; Yamashita, R.; Ikegawa, Y.; Kudou, T.; Igarashi, R.; Kengaku, Y.; Cho, H.; Standley, D. M.; Nakagawa, A.; Nakamura, H. Protein Data Bank Japan (PDBj):

Maintaining a Structural Data Archive and Resource Description Framework Format. *Nucleic Acids Res.* **2012**, *40*, D453-60.

Kiss, R.; Sandor, M.; Szalai, F.A. http://Mcule. com: a public web service for drug discovery. *J. cheminformatics*. **2012**, *4*, 1-1.

Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949.

Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

Konreddy, A. K.; Rani, G. U.; Lee, K.; Choi, Y. Recent Drug-Repurposing-Driven Advances in the Discovery of Novel Antibiotics. *Curr. Med. Chem.* **2019**, *26*, 5363–5388.

Koplin, J.; Mu, Y.; Richter, C.; Schwalbe, H.; & Stock, G. Structure and dynamics of an RNA tetraloop: a joint molecular dynamics and NMR study. *Structure*. **2005**, *13*, 1255-1267.

Kopp, J., & Schwede, T. Automated protein structure homology modeling: a progress report. *Pharmacogenomics.* **2004**, *5*, 405-416.

Kraemer, M. U.; Sinka, M. E.; Duda, K. A.; Mylne, A. Q.; Shearer, F. M.; Barker, C. M. & Hay, S. I. *The Global Distribution of the Arbovirus Vectors Aedes Aegypti and Ae. Albopictus.* **2015**.

Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins: *Struct., Funct., Bioinf.* **1999**, *37*, 228-241.

Krupanidhi, S.; Abraham Peele, K.; Venkateswarulu, T. C.; Ayyagari, V. S.; Bobby, M.; John Babu, D. & Aishwarya, G. Screening of Phytochemical Compounds of Tinospora Cordifolia for Their Inhibitory Activity on SARS-CoV-2: An in Silico Study. *J. Biomol. Struct. Dyn.* **2021**, *39*, 5799–5803.

Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)-Round XIII. *Proteins* **2019**, *87*, 1011–1020.

Kryshtafovych, A.; Schwede, T., Topf, M., Fidelis, K., & Moult, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function, and Bioinformatics.* **2021**, *89*, 1607-1617.

Kumar, D.; Aarthy, M.; Kumar, P.; Singh, S. K.; Uversky, V. N.; Giri, R. Targeting the NTPase Site of Zika Virus NS3 Helicase for Inhibitor Discovery. *J. Biomol. Struct. Dyn.* **2020**, *38*, 4827–4837.

Kumar, D.; Sharma, N.; Aarthy, M.; Singh, S. K.; Giri, R. Mechanistic Insights into Zika Virus NS3 Helicase Inhibition by Epigallocatechin-3-Gallate. *ACS Omega.* **2020**, *5*, 11217–11226.

Kumari, R.; Kumar, R. Open-Source Drug Discovery Consortium, & Lynn, A. G_mmpbsa A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962.

Kuznetsov, Y. G.; Malkin, A. J.; Lucas, R. W.; Plomp, M.; McPherson, A. Imaging of Viruses by Atomic Force Microscopy. *J. Gen. Virol.* **2001**, *82*, 2025–2034.

Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; Mcgettigan, P. A.; Mcwilliam, H. & Higgins, D. G. Clustal W and Clustal X Version 2.0. *J. Bioinform.* **2007**, *23*, 2947–2948.

Lancet, T. Artificial Intelligence in Health Care: Within Touching Distance. *Lancet* **2017**.
Leach, A. R. Ligand Docking to Proteins with Discrete Side-Chain Flexibility. *J. Mol. Biol.* **1994**, *235*, 345–356.

Lang, P.T.; Brozell, S.R.; Mukherjee, S.; Pettersen, E.F.; Meng, E.C.; Thomas, V.; Rizzo, R.C.; Case, D.A.; James, T.L.; Kuntz, I.D. DOCK 6: Combining techniques to model RNA–small molecule complexes. *Rna*. **2009**, *15*, 1219-1230.

Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*. **1993**, *26*, 283-291.

Laurie, A.T.; Jackson, R.M. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics.* **2005**, *21*, 1908-1916.

Lavecchia, A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discov. Today.* **2015**, *20*, 318-331.

Leach, A.R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **1994**, *235*, 345- 356.

Leach, A.R.; Gillet, V.J.; Lewis, R.A.; Taylor, R. Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* **2010**, *53*, 539-558.

Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B Condens. Matter* **1988**, *37*, 785–789.

Letunic, I.; Goodstadt, L.; Dickens, N. J.; Doerks, T.; Schultz, J.; Mott, R.; Ciccarelli, F.; Copley, R. R.; Ponting, C. P.; Bork, P. Recent Improvements to the SMART Domain-Based Sequenceannotation Resource. *Nucleic Acids Res* **2002**, *30*, 242–244.

Levitt, M.; Lifson, S. Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure. *J. Mol. Biol.* **1969**, *46* (2), 269–279.

Li, L.; Chen, R.; Weng, Z. RDOCK: refinement of rigid-body protein docking predictions. *Proteins: Struct. Funct. Bioinf.* **2003**, *53*, 693-707.

Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.; Zhou, L.; Lin, Y.; Wu, D.; Zhang, L.; Liu, H.; Xu, X.; Yu, K.; Zhang, Y.; Cui, J.; Zhan, C.-G.; Wang, X.; Luo, H.-B. Identify Potent SARS-CoV-2 Main Protease Inhibitors via Accelerated Free Energy Perturbation-Based Virtual Screening of Existing Drugs. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 27381–27387.

Lindahl, E.; Hess, B. Van Der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model* **2001**, *7*, 306–317.

Lindahl, T.; Barnes, D. E. Repair of Endogenous DNA Damage. *Cold Spring Harb. Symp. Quant. Biol.* **2000**, *65*, 127–133.

Lipinski, C. A. Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337-341.

Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2012**, *64*, 4–17.

Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.

Lipman, D. J.; Altschul, S. F.; Kececioglu, J. D. A Tool for Multiple Sequence Alignment. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 4412–4415.

Liu, Z.; Zhou, T.; Lai, Z.; Zhang, Z.; Jia, Z.; Zhou, G.; Williams, T.; Xu, J.; Gu, J.; Zhou, X.; Lin, L.; Yan, G.; Chen, X.-G. Competence of Aedes Aegypti, Ae. Albopictus, and Culex Quinquefasciatus Mosquitoes as Zika Virus Vectors, China. *Emerg. Infect. Dis.* **2017**, *23*, 1085–1091.

Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. Current Progress in Structure-Based Rational Drug Design Marks a New Mindset in Drug Discovery. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302011.

Louten, J. Virus structure and classification. *Essential human virology*. **2016**, 19.

Lu, X. J., Bussemaker, H. J., & Olson, W. K. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic acids research*. **2015**, *43*, e142-e142.

Luscombe, N. M.; Greenbaum, D.; Gerstein, M. What Is Bioinformatics? A Proposed Definition and Overview of the Field. *Methods Inf. Med.* **2001**, *40*, 346–358.

Lüthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature.* **1992**, *356*, 83-85.

MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.;

Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

Majumder, R., & Mandal, M. Screening of plant-based natural compounds as a potential COVID-19 main protease inhibitor: an in silico docking and molecular dynamics simulation approach. *J. Biomol. Struct. Dyn.* **2022**, *40*, 696-711.

Markley, J. L.; Ulrich, E. L.; Berman, H. M.; Henrick, K.; Nakamura, H.; Akutsu, H. BioMagResBank (BMRB) as a Partner in the Worldwide Protein Data Bank (WwPDB): New Policies Affecting Biomolecular NMR Depositions. *J. Biomol. NMR.* **2008**, *40* (3), 153–155.

Martinez, X.; Krone, M.; Alharbi, N.; Rose, A.S.; Laramee, R.S.; O'Donoghue, S.; Baaden, M.; Chavent, M. Molecular graphics: bridging structural biologists and computer scientists. *Structure*. **2019**, *27*, 1617- 1623.

Martinez-Mayorga, K.; Madariaga-Mazon, A.; Medina-Franco, J. L.; Maggiora, G. The Impact of Chemoinformatics on Drug Discovery in the Pharmaceutical Industry. *Expert Opin. Drug Discov.* **2020**, *15*, 293–306.

Mastrangelo, E.; Pezzullo, M.; De Burghgraeve, T.; Kaptein, S.; Pastorino, B.; Dallmeier, K.; de Lamballerie, X.; Neyts, J.; Hanson, A. M.; Frick, D. N.; Bolognesi, M.; Milani, M. Ivermectin Is a Potent Inhibitor of Flavivirus Replication Specifically Targeting NS3 Helicase Activity: New Prospects for an Old Drug. *J. Antimicrob. Chemother.* **2012**, *67*, 1884–1894.

McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nature.* **1977**, *267*, 585-590.

McCarthy, J. Generality in artificial intelligence. *Commun. ACM*. **1987**, *30*, 1030-1035.

McConkey, B. J.; Sobolev, V.; & Edelman, M. The performance of current methods in ligand–protein docking. *Current Science*. **2002**, 845-856.

McCullers, J. A. Insights into the interaction between influenza virus and pneumococcus. *Clin. Microbiol. Rev.* **2006**, *19*, 571-582.

Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157.

Mercorelli, B.; Palù, G.; Loregian, A. Drug Repurposing for Viral Infectious Diseases: How Far Are We? *Trends Microbiol.* **2018**, *26*, 865–876.

Merz, B.; Kreibich, H.; Schwarze, R.; Thieken, A. Review Article" Assessment of Economic Flood Damage". *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 1697–1724.

Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilizaion of AB Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117–129.

Miller, I.; Mcgee, B. R.; Swails, T. D.; Homeyer, J. M.; Gohlke, N.; Roitberg, H.; Mmpbsa, A. E. Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput* **2012**, *8*, 3314–3321.

Mlakar, J.; Korva, M.; Tul, N.; Popović, M.; Poljšak-Prijatelj, M.; Mraz, J.; Kolenc, M.; Resman Rus, K.; Vesnaver Vipotnik, T.; Fabjan Vodušek, V.; Vizjak, A.; Pižem, J.; Petrovec, M.; Avšič Županc, T. Zika Virus Associated with Microcephaly. *N. Engl. J. Med.* **2016**, *374*, 951–958.

Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

Moore, Z. S., Seward, J. F., & Lane, J. M. Smallpox. *The Lancet*. **2006**, *367*, 425-435.

Morris, G. M.; Lim-Wilby, M. Molecular Docking. *Methods Mol. Biol.* **2008**, *443*, 365–382.

Moulin, E.; Selby, K.; Cherpillod, P.; Kaiser, L.; Boillat-Blanco, N. Simultaneous Outbreaks of Dengue, Chikungunya and Zika Virus Infections: Diagnosis Challenge in a Returning Traveller with Nonspecific Febrile Illness. *New Microbes New Infect.* **2016**, *11*, 6–7.

Moult, J.; Pedersen, J. T.; Judson, R.; & Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins: Struct. Funct. Genet*. **1995**, *23*, ii-iv.

Muegge, I.; Martin, Y.C. A general and fast scoring function for protein− ligand interactions: a simplified potential approach. *J. Med. Chem*.**1999**, *42*, 791-804.

Muegge, I. PMF scoring revisited. *J. Med. Chem*. **2006**, *49*, 5895-5902.

Musso, D.; Ko, A. I.; & Baud, D. Zika virus infection—after the pandemic. *N. Engl. J. Med.* **2019**, *381*, 1444-1457.

Natesh, J.; Mondal, P.; Penta, D.; Salam, A. A. A.; & Meeran, S. M. Culinary spice bioactives as potential therapeutics against SARS-CoV-2: Computational investigation. *Comput. Biol. Med*. **2021**, *128*, 104102.

Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443-453.

Nelson, M.T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kalé, L.V.; Skeel, R.D.; Schulten, K. NAMD: a parallel, object-oriented molecular dynamics program. *Int. j. supercomput. appl. high perform. Comput*. **1996**, *10*, 251-268.

Neves, M.A.; Totrov, M.; Abagyan, R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput. Aided Mol. Des.* **2012**, *26*, 675-686.

Nifosì, R.; Reyes, C. M.; & Kollman, P. A. Molecular dynamics studies of the HIV-1 TAR and its complex with argininamide. *Nucleic Acids Research*. **2000**, *28*, 4944-4955.

Nilsson, N.J.; Nilsson, N.J. Artificial intelligence: a new synthesis. Morgan Kaufmann. **1998**.

Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205-217.

O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. **2016**, *44*, D733- D745.

Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. Forces Contributing to the Conformational Stability of Proteins. *FASEB J.* **1996**, *10*, 75–83.

Parenti, M. D.; Rastelli, G. Advances and Applications of Binding Affinity Prediction Methods in Drug Discovery. *Biotechnol. Adv.* **2012**, *30*, 244–250.

Parrill, A. L. Rational Drug Design: Novel Methodology and Practical Applications. *J. Am. Chem. Soc* **1999**.

Pattnaik, A.; Sahoo, B. R.; Pattnaik, A. K. Current Status of Zika Virus Vaccines: Successes and Challenges. *Vaccines (Basel)* **2020**, *8*, 266.

Peele, K. A.; Potla Durthi, C.; Srihansa, T.; Krupanidhi, S.; Ayyagari, V. S.; Babu, D. J.; Indira, M.; Reddy, A. R.; Venkateswarulu, T. C. Molecular Docking and Dynamic Simulations for Antiviral Compounds against SARS-CoV-2: A Computational Study. *Inform. Med. Unlocked* **2020**, *19*, 100345.

Pennington, L. D.; Aquila, B. M.; Choi, Y.; Valiulin, R. A.; Muegge, I. Positional Analogue Scanning: An Effective Strategy for Multiparameter Optimization in Drug Design. *J. Med. Chem.* **2020**, *63*, 8956–8976.

Pérez-Villa, A.; Darvas, M.; & Bussi, G. ATP dependent NS3 helicase interaction with RNA: insights from molecular simulations. *Nucleic acids research*. **2015**, *43*, 8725-8734.

Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Open–shell systems and the total energies of the first–row atoms. *J. Chem. Phys*. **1991**, *94*, 6081–6090.

Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A. Complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first–row atoms. *J. Chem. Phys*. **1988**, *89*, 2193–2198.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

Pierson, T. C.; Diamond, M. S. *Flaviviruses. In Fields Virology: Sixth Edition*; ESP, 2013.
Pinzi, L.; Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int. J. Mol. Sci.* **2019**, *20*, 4331.

Pitici, F.; Beveridge, D. L.; Baranger, A. M. Molecular Dynamics Simulation Studies of Induced Fit and Conformational Capture in U1A-RNA Binding: Do Molecular Substates Code for Specificity? *Biopolymers* **2002**, *65* (6), 424–435.

Pompon, J.; Morales-Vargas, R.; Manuel, M.; Huat Tan, C.; Vial, T.; Hao Tan, J.; ... & Missé, D. A Zika virus from America is more efficiently transmitted than an Asian virus by Aedes aegypti mosquitoes from Asia. *Sci. Rep.* **2017**, *7*, 1-8.

Potts, R. O.; Guy, R. H. Predicting Skin Permeability. *Pharm. Res.* **1992**, *9*, 663–669.

Pushpakom, S.; Iorio, F.; Eyers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; ... & Norris, A. Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **2019**, *18*, 41-58.

Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*, 95–99.

Rappé, A.K.; Casewit, C.J.; Colwell, K.S.; Goddard III, W.A.; Skiff, W.M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992,** *114*, 10024-10035.

Rasmussen, S. A.; Jamieson, D. J.; Honein, M. A.; & Petersen, L. R. Zika virus and birth defects—reviewing the evidence for causality. *N. Engl. J. Med.* **2016**, *374*, 1981-1987.

Rella, M.; Rushworth, C. A.; Guy, J. L.; Turner, A. J.; Langer, T.; & Jackson, R. M. Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *J. Chem. Inf. Model*. **2006**, *46*, 708-716.

Ren, Z.; Yan, L.; Zhang, N.; Guo, Y.; Yang, C.; Lou, Z.; & Rao, Z. The newly emerged SARS-like coronavirus HCoV-EMC also has an" Achilles' heel": current effective inhibitor targeting a 3C-like protease. *Protein Cell*. **2013**, *4*, 248.

Rosa, G.; Fierro, S.; Santos, I. M. Repositioning and Investigational Drugs for Zika Virus Infection Treatment: A Patent Review. *Expert Opin. Ther. Pat.* **2020**, *30*, 847–862.

REYES, C. M.; & KOLLMAN, P. A. Molecular dynamics studies of U1A-RNA complexes. *Rna*. **1999**, *5*, 235-244.

Robilotti, E.; Deresinski, S.; & Pinsky, B. A. Norovirus. Clin. *Microbiol. Rev.* **2015**, *28*, 134-164.

Ross, C.; Nizami, B.; Glenister, M.; Sheik Amamuddy, O.; Atilgan, A. R.; Atilgan, C.; & Tastan Bishop, Ö. MODE-TASK: large-scale protein motion tools. *Bioinformatics*. **2018**, *34*, 3759-3763.

Rout, J.; Swain, B. C.; Tripathy, U. In Silico Investigation of Spice Molecules as Potent Inhibitor of SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2022**, *40*, 860–874.

Ruchusatsawat, K.; Wongjaroen, P.; Posanacharoen, A.; Rodriguez-Barraquer, I.; Sangkitporn, S.; Cummings, D. A.; & Salje, H. Long-term circulation of Zika virus in Thailand: an observational study. *Lancet Infect. Dis*. **2019**, *19*, 439-446.

Šali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol*. **1993**, *234*, 779-815.

Sampaio, E.P.; Sarno, E.N.; Galilly, R.; Cohn, Z.A.; Kaplan, G. Thalidomide selectively inhibits tumor necrosis factor alpha production by stimulated human monocytes. *J. Exp. Med.* **1991**, *173*, 699-703.

Sampangi-Ramaiah, M. H.; Vishwakarma, R.; &Shaanker, R. U. Molecular docking analysis of selected natural products from plants for inhibition of SARS-CoV-2 main protease. *Curr. Sci.* **2020**, *118*, 1087-1092.

Sanseau, P.; Koehler, J. Editorial: Computational Methods for Drug Repurposing. *Brief. Bioinform.* **2011**, *12*, 301–302.

Satija, N.; Lal, S. K. The Molecular Biology of SARS Coronavirus. *Ann. N. Y. Acad. Sci.* **2007**, *1102*, 26–38.

Saxena, S. K.; Kumar, S.; Sharma, R.; Maurya, V. K.; Dandu, H. R.; Bhatt, M. L. Zika Virus Disease in India-Update October 2018. *Travel Med Infect Dis.* **2018**, *27*, 121–122.

Sayle, R. A.; & Milner-White, E. J. RASMOL: biomolecular graphics for all. Trends *Biochem. Sci.* **1995**, *20*, 374-376.

Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: An Automated Protein Homologymodeling Server. *Nucleic Acids Res.* **2003**, *31*, 3381–3385.

Scott, D.E.; Coyne, A.G.; Hudson, S.A.; Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry*. **2012**, *51*, 4990-5003.

Searls, D.B. The roots of bioinformatics. *PLoS Comput Biol*. **2010**, *6*, e1000809.

Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D Pharmacophore-Based Virtual Screening. *Drug Discov. Today Technol.* **2010**, *7*, e203-70.

Sevvana, M.; Klose, T.; Rossmann, M. G. Principles of Virus Structure. In Encyclopedia of Virology; *Elsevier.* **2021,** pp 257–277.

Shadrick, W. R.; Mukherjee, S.; Hanson, A. M.; Sweeney, N. L.; & Frick, D. N. Aurintricarboxylic acid modulates the affinity of hepatitis C virus NS3 helicase for both nucleic acid and ATP. *Biochemistry.* **2013**, *52*, 6151-6159.

Shaw, D.E.; Deneroff, M.M.; Dror, R.O.; Kuskin, J.S.; Larson, R.H.; Salmon, J.K.; Young, C.; Batson, B.; Bowers, K.J.; Chao, J.C.; Eastwood, M.P. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM.* **2008**, *51*, 91-97.

Sheridan, M. A.; Yunusov, D.; Balaraman, V.; Alexenko, A. P.; Yabe, S.; Verjovski-Almeida, S.; Schust, D. J.; Franz, A. W.; Sadovsky, Y.; Ezashi, T.; Roberts, R. M. Vulnerability of Primitive Human Placental Trophoblast to Zika Virus. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E1587–E1596.

Sievers, F.; Higgins, D. G. Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. *Methods Mol. Biol.* **2014**, *1079*, 105–116.

Sikka, V.; Chattu, V. K.; Popli, R. K.; Galwankar, S. C.; Kelkar, D.; Sawicki, S. G.; Stawicki, S. P.; Papadimos, T. J. The Emergence of Zika Virus as a Global Health Security Threat: A Review and a Consensus Statement of the INDUSEM Joint Working Group (JWG). *J. Glob. Infect. Dis.* **2016**, *8*, 3–15.

Sitkoff, D.; Sharp, K.A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978-1988.

Smith, T.F.; Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol*. **1981**, *147*, 195-197.

Smith, W.; Yong, C.W.; Rodger, P.M. DL_POLY: Application to molecular simulation. *Mol. Simul.* **2002**, *28*, 385-471.

Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics.* **2005**, *21*, 951-960.

Srinivasan, J.; Cheatham, T.E.; Cieplak, P.; Kollman, P.A.; Case, D.A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate− DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401-9409.

Styczynski, A. R.; Malta, J. M. A. S.; Krow-Lucal, E. R.; Percio, J.; Nóbrega, M. E.; Vargas, A.; Lanzieri, T. M.; Leite, P. L.; Staples, J. E.; Fischer, M. X.; Powers, A. M.; Chang, G.-J. J.; Burns, P. L.; Borland, E. M.; Ledermann, J. P.; Mossel, E. C.; Schonberger, L. B.; Belay, E. B.;

Salinas, J. L.; Badaro, R. D.; Sejvar, J. J.; Coelho, G. E. Increased Rates of Guillain-Barré Syndrome Associated with Zika Virus Outbreak in the Salvador Metropolitan Area, Brazil. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005869.

Sun, H.; Ren, P.; Fried, J. R. The COMPASS Force Field: Parameterization and Validation for Phosphazenes. *Comput. Theor. Polym. Sci.* **1998**, *8* (1–2), 229–246.

Sunseri, J.; Koes, D.R. Pharmit: interactive exploration of chemical space. *Nucleic Acids Res.* **2016**, *44*, W442-W448.

Surana, P.; Satchidanandam, V.; & Nair, D. T. RNA-dependent RNA polymerase of Japanese encephalitis virus binds the initiator nucleotide GTP to form a mechanistically important pre-initiation state. *Nucleic acids research*. **2014**, *42*, 2758-2773.

Tang, Y.; & Nilsson, L. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J*. **1999**, *77*, 1284-1305.

Tatsuo, H.; Ono, N.; Tanaka, K.; & Yanagi, Y. SLAM (CDw150) is a cellular receptor for measles virus. *Nature.* **2000**, *406*, 893-897.

Tay, M. Y. F.; Vasudevan, S. G. The Transactions of NS3 and NS5 in Flaviviral RNA Replication. *Adv. Exp. Med. Biol.* **2018**, *1062*, 147–163.

Team, E. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19)-China, 2020. China CDC Weekly. **2020**, *2*.

Tian, H.; Ji, X.; Yang, X.; Xie, W.; Yang, K.; Chen, C. & Yang, H. The Crystal Structure of Zika Virus Helicase: Basis for Antiviral Drug Design. *Protein & cell* **2016**, *7*, 450–454.

Tian, H.; Ji, X.; Yang, X.; Zhang, Z.; Lu, Z.; Yang, K. & Xie, W. Structural Basis of Zika Virus Helicase in Recognizing Its Substrates. *Protein Cell.* **2016**, *7*, 562–570.

Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*. **2018**, *46*, W363-W367.

Tomasi, J.; Persico, M. Molecular interactions insolution: An overview of methods based on continuous distributions of the solvent. *Chem. ReV*. **1994**, *94*, 2027–2094.

Touret, F.; Gilles, M.; Barral, K.; Nougairède, A.; van Helden, J.; Decroly, E.; de Lamballerie, X.; Coutard, B. In Vitro Screening of a FDA Approved Chemical Library Reveals Potential Inhibitors of SARS-CoV-2 Replication. *Sci. Rep.* **2020**, *10*, 13093.

Trott, O., & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455-461.

Tugwell, B. D., Lee, L. E., Gillette, H., Lorber, E. M., Hedberg, K., & Cieslak, P. R. Chickenpox outbreak in a highly vaccinated school population. *Pediatrics.* **2004**, *113*, 455-459.

Uyar, A.; Kurkcuoglu, O.; Nilsson, L.; Doruker, P. The Elastic Network Model Reveals a Consistent Picture on Intrinsic Functional Dynamics of Type II Restriction Endonucleases. *Phys. Biol.* **2011**, *8*, 056001.

Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; & Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem*. **2005**, *26*, 1701-1718.

van Gunsteren, W.F.; Berendsen, H.J. Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J. Comput. Aided Mol. Des*. **1987**, *1*, 171- 176.

Velankar, S.; Best, C.; Beuth, B.; Boutselakis, C.H.; Cobley, N.; Sousa Da Silva, A.W.; Dimitropoulos, D.; Golovin, A.; Hirshberg, M.; John, M.; Krissinel, E.B. PDBe: protein data bank in Europe. *Nucleic Acids Res*. **2010**, *38*, D308-D317.

Van De Waterbeemd, H.; Gifford, E. ADMET in silico modelling: towards prediction paradise?. *Nat. Rev. Drug Discov*. **2003**, *2*, 192-204.

Velázquez-Muriel, J.A.; Rueda, M.; Cuesta, I.; Pascual-Montano, A.; Orozco, M.; Carazo, J.M. Comparison of molecular dynamics and superfamily spaces of protein domain deformation. *BMC Struct. Biol*. **2009**, *9*, 1-14.

Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved protein–ligand docking using GOLD. *Proteins: Struct. Funct. Bioinf.* **2003**, *52*, 609-623.

Verma, S., Twilley, D., Esmear, T., Oosthuizen, C. B., Reid, A. M., Nel, M., & Lall, N. Anti-SARS-CoV natural products with the potential to inhibit SARS-CoV-2 (COVID-19). *Front. Pharmacol.* **2020**, 1514.

Vetrivel, R.; Deka, R. C.; Chatterjee, A.; Kubo, M.; Broclawik, E.; Miyamoto, A. Studies on the molecular electrostatic potential inside the microporous material and its relevance to their catalytic activity. *Theor. Comput. Chem.* **1996**, *3*, 509–541.

Viveiros-Rosa, S. G., Regis, E. G., & Santos, W. C. Vector competence of Culex mosquitoes (Diptera: Culicidae) in Zika virus transmission: an integrative review. *Rev. Panam. Salud Publica*. **2020**, *44*, e7.

Wagner, A. B. SciFinder Scholar 2006: An Empirical Analysis of Research Topic Query Processing. *J. Chem. Inf. Model.* **2006**, *46*, 767–774.

Walters, W. P.; Stahl, M. T.; & Murcko, M. A. Virtual screening—an overview. *Drug Discov. Today*. **1998**, *3*, 160-178.

Wang, F.; Chen, C.; Tan, W.; Yang, K.; & Yang, H. Structure of main protease from human coronavirus NL63: insights for wide spectrum anti-coronavirus drug design. *Sci. Rep*. **2016**, *6*, 1-12.

Wang, H. Y.; Li, X. L.; Yan, Z. R.; Sun, X. P.; Han, J.; & Zhang, B. W. Potential neurological symptoms of COVID-19. *Ther. Adv. Neurol. Disord*. **2020**, *13*, 1756286420917830.

Wang, J.; Wang, W.; Kollman, P. A.; & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model*. **2006**, 25, 247-260.

Wang, L.; Valderramos, S. G.; Wu, A.; Ouyang, S.; Li, C.; Brasil, P.; Bonaldo, M.; Coates, T.; Nielsen-Saines, K.; Jiang, T.; Aliyari, R.; Cheng, G. From Mosquitos to Humans: Genetic Evolution of Zika Virus. *Cell Host Microbe* **2016**, *19*, 561–565.

Wang, S.; Sun, S.; Li, Z.; Zhang, R.; & Xu, J. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput. Biol.* **2017**, *13*, e1005324.

Wang, Y. C.; Yang, W. H.; Yang, C. S.; Hou, M. H.; Tsai, C. L.; Chou, Y. Z.; ... & Chen, Y. Structural basis of SARS-CoV-2 main protease inhibition by a broad-spectrum anti-coronaviral drug. *Am. J. Cancer Res*. **2020**, *10*, 2535.

Watts, A. G.; Huber, C.; Bogoch, I. I.; Brady, O. J.; Kraemer, M. U.; & Khan, K. Potential Zika virus spread within and beyond India. *J Travel Med.* **2019**, *26*, tay132.

Weger-Lucarelli, J.; Rückert, C.; Chotiwan, N.; Nguyen, C.; Garcia Luna, S. M.; Fauver, J. R.; ... & Ebel, G. D. Vector competence of American mosquitoes for three strains of Zika virus. PLoS Negl Trop Dis. **2016**, *10*, e0005101.

Weiner, P.K.; Kollman, P.A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem*. **1981**, 2, 287-303.

Welsch, S.; Miller, S.; Romero-Brey, I.; Merz, A.; Bleck, C. K.; Walther, P. & Bartenschlager, R. Composition and Three-Dimensional Architecture of the Dengue Virus Replication and Assembly Sites. *Cell Host Microbe.* **2009**, *5*, 365–375.

Wermuth, C. G., Ganellin, C. R., Lindberg, P., & Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, 70, 1129-1143.

Wikan, N., & Smith, D. R. Zika virus: history of a newly emerging arbovirus. *Lancet Infect. Dis.* **2016**, *16*, e119-e126.

Wiley, E. A., MacDonald, M., Lambropoulos, A., Harriman, D. J., & Deslongchamps, G. LGA-dock/EM-dock exploring Lamarckian genetic algorithms and energy-based local search for ligand receptor docking. *Can. J. Chem.* **2006**, *84*, 384-391.

Williams, A. J., Tkachenko, V., Golotvin, S., Kidd, R., & McCann, G. ChemSpider-building a foundation for the semantic web by hosting a crowd sourced databasing platform for chemistry. *J. Cheminformatics.* **2010**, *2*, 1-1.

Williams, D. J., & Hall, K. B. Experimental and computational studies of the G [UUCG] C RNA tetraloop. *J. Mol. Biol.* **2000**, *297*, 1045-1061.

Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074-D1082.

Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45*, 160-169.

Wu, G.; Robertson, D.H.; Brooks III, C.L; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549-1562.

Wu, K. J. There Are More Viruses than Stars in the Universe. Why Do Only Some Infect Us? *National Geographic.* **2020**.

Xu, D.; Xu, Y. Protein databases on the internet. Curr. Protoc. *Mol. Biol*. **2004**, *68*, 19-4.

Xu, S., Ci, Y., Wang, L., Yang, Y., Zhang, L., Xu, C., ... & Shi, L. Zika virus NS3 is a canonical RNA helicase stimulated by NS5 RNA polymerase. *Nucleic acids research*. 2019, 47, 8693-8707.

Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., ... & Rao, Z. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol*. **2008**, *82*, 2515-2527.

Yadav, P. D., Malhotra, B., Sapkal, G., Nyayanit, D. A., Deshpande, G., Gupta, N., ... & Mourya, D. T. Zika virus outbreak in Rajasthan, India in 2018 was caused by a virus endemic to Asia. *Infect. Genet. Evol.* **2019**, *69*, 199-202.

Yadav, R., Selvaraj, C., Aarthy, M., Kumar, P., Kumar, A., Singh, S. K., & Giri, R. Investigating into the molecular interactions of flavonoids targeting NS2B-NS3 protease from ZIKA virus through in-silico approaches. *J. Biomol. Struct. Dyn*. **2021**, *39*, 272-284.

Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., ... & Rao, Z. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **2005**, *3*, e324.

Yang, S.Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today*. **2010**, *15*, 444-450.

Yap, T. L., Xu, T., Chen, Y. L., Malet, H., Egloff, M. P., Canard, B., ... & Lescar, J. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J. Virol*. **2007**, *81*, 4753-4765.

Young, M. A., & Beveridge, D. L. Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. J. Mol. Biol. 1998, 281, 675-687.

Yu, W.; MacKerell, A. D., Jr. Computer-Aided Drug Design Methods. *Methods Mol. Biol*. **2017**, *1520*, 85–106.

Zacharias, M. Simulation of the Structure and Dynamics of Nonhelical RNA Motifs. *Curr. Opin. Struct. Biol*. **2000**, *10*, 311–317.

Zacharias, M.; Sklenar, H. Conformational Analysis of Single-Base Bulges in A-Form DNA and RNA Using a Hierarchical Approach and Energetic Evaluation with a Continuum Solvent Model. *J. Mol. Biol*. **1999**, *289*, 261–275.

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., ... & Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science.* **2020**, *368*, 409-412.

Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics*. **2008**, *9*, 40.

Zhao, B.; Yi, G.; Du, F.; Chuang, Y.-C.; Vaughan, R. C.; Sankaran, B.; Kao, C. C.; Li, P. Structure and Function of the Zika Virus Full-Length NS5 Protein. *Nat. Commun.* **2017**, *8*, 14762.

Zhao, H.; Caflisch, A. Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. *Bioorg. Med. Chem. Lett*. **2013**, *23*, 5721-5726.

Zheng, M.; Liu, X.; Xu, Y.; Li, H.; Luo, C.; Jiang, H. Computational Methods for Drug Design and Discovery: Focus on China. *Trends Pharmacol. Sci*. **2013**, *34*, 549–559.

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. Artificial Intelligence in Drug Design. Sci. *China Life Sci*. **2018**, *61*, 1191–1204.

Ziebuhr, J.; Snijder, E. J.; Gorbalenya, A. E. Virus-Encoded Proteinases and Proteolytic Processing in the Nidovirales. *J. Gen. Virol*. **2000**, *81*, 853–879.

Zoete, V.; Daina, A.; Bovigny, C.; Michielin, O. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand- Based Virtual Screening. *J. Chem. Inf. Model*. **2016**, *56*, 1399–1404.

List of publications

1. **Durgam, L.,** & Guruprasad, L. Molecular mechanism of ATP and RNA binding to Zika virus NS3 helicase and identification of repurposed drugs using molecular dynamics simulations. *J. Biomol. Struct. Dyn*. **2021**, 1-18.

2. **Durgam, L.,** & Guruprasad, L. (2022). Computational studies on the design of NCI natural products as inhibitors to SARS-CoV-2 main protease. *J. Biomol. Struct. Dyn*. **2022**, 1-11.

3. Saxena, S., **Durgam, L.,** & Guruprasad, L. Multiple e-pharmacophore modelling pooled with high-throughput virtual screening, docking and molecular dynamics simulations to discover potential inhibitors of Plasmodium falciparum lactate dehydrogenase (PfLDH). *J. Biomol. Struct. Dyn*. **2019**, 37, 1783-1799.

4. Andola, P., Pagag, J., **Laxman, D.,** & Guruprasad, L. Fragment-based inhibitor design for SARS-CoV2 main protease. *Struct Chem*. **2022**, *33*, 1467-1487.

Inhibitor design for drug targets in Zika virus and SARS CoV-2 using computational studies

by Durgam Laxman

Inhibitor design for drug targets in Zika virus and SARS CoV-2 using computational studies

ORIGINALITY REPORT

lalithaal

43%	14%	41%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Laxman Durgam, Lalitha Guruprasad. "Molecular mechanism of ATP and RNA binding to Zika virus NS3 helicase and identification of repurposed drugs using molecular dynamics simulations", Journal of Biomolecular Structure and Dynamics, 2021 Publication	22%
2	Laxman Durgam, Lalitha Guruprasad. "Computational studies on the design of NCI natural products as inhibitors to SARS-CoV-2 main protease", Journal of Biomolecular Structure and Dynamics, 2022 Publication	10%
3	Submitted to University of Hyderabad, Hyderabad Student Paper	2%
4	link.springer.com Internet Source	1%
5	www.embopress.org Internet Source	<1%

lalithaal

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.

lalithaal

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.

6	Saroj Kumar Panda, Shalini Saxena, Lalitha Guruprasad. " Homology modeling, docking and structure-based virtual screening for new inhibitor identification of heptosyltransferase-III ", Journal of Biomolecular Structure and Dynamics, 2019 Publication	<1 %
7	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
8	Maaged Abdullah, Lalitha Guruprasad. "Computational basis for the design of PLK-2 inhibitors", Structural Chemistry, 2019 Publication	<1 %
9	en.wikipedia.org Internet Source	<1 %
10	Submitted to National Institute of Technology, Raipur Student Paper	<1 %
11	www.frontiersin.org Internet Source	<1 %
12	www.nature.com Internet Source	<1 %
13	www.mdpi.com Internet Source	<1 %
14	www.tandfonline.com	

	Internet Source	<1%
15	Submitted to Sir George Monoux College Student Paper	<1%
16	www.ncbi.nlm.nih.gov Internet Source	<1%
17	"Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design", Springer Science and Business Media LLC, 2021 Publication	<1%
18	Submitted to Key 2 Learning College Student Paper	<1%
19	wikimili.com Internet Source	<1%
20	Submitted to University of KwaZulu-Natal Student Paper	<1%
21	John Jumper, Richard Evans, Alexander Pritzel, Tim Green et al. "Highly accurate protein structure prediction with AlphaFold", Nature, 2021 Publication	<1%
22	hollywoodsubliminals.wordpress.com Internet Source	<1%
23	peerj.com Internet Source	

<1%

24 Son Tung Ngo, Nguyen Minh Tam, Minh Quan Pham, Trung Hai Nguyen. "Benchmark of Popular Free Energy Approaches Revealing the Inhibitors Binding to SARS-CoV-2 Mpro", Journal of Chemical Information and Modeling, 2021
Publication
<1%

25 bio.gnu.ac.kr
Internet Source
<1%

26 Tejeswara R. Allaka, Jaya S. Anireddy. "Novel 7-substituted Fluoroquinolone Citrate Conjugates as Powerful Antibacterial and Anticancer Agents: Synthesis and Molecular Docking Studies", Current Organic Chemistry, 2019
Publication
<1%

27 www.arkat-usa.org
Internet Source
<1%

28 Son Tung Ngo, Nguyen Minh Tam, Pham Minh Quan, Trung Hai Nguyen. "A Benchmark of Popular Free Energy Approaches Revealing the Inhibitors Binding to SARS-CoV2 Mpro", American Chemical Society (ACS), 2020
Publication
<1%

Submitted to University of South Florida

29	Student Paper	<1%
30	Submitted to Queen's University of Belfast Student Paper	<1%
31	d-nb.info Internet Source	<1%
32	Melina Mottin, Rodolpho C. Braga, Roosevelt A. da Silva, Joao H. Martins da Silva et al. "Molecular dynamics simulations of Zika virus NS3 helicase: Insights into RNA binding site activity", Biochemical and Biophysical Research Communications, 2017 Publication	<1%
33	Submitted to York St John University Student Paper	<1%
34	escholarship.org Internet Source	<1%
35	Submitted to University of Leicester Student Paper	<1%
36	Submitted to Lipscomb Academy Student Paper	<1%
37	Submitted to Jawaharlal Nehru Technological University Student Paper	<1%

38	Rashmi Kumari, Rajendra Kumar, Andrew Lynn. " —A GROMACS Tool for High-Throughput MM-PBSA Calculations ", Journal of Chemical Information and Modeling, 2014 Publication	<1 %
39	Ravi Kant, Prakash Jha, Daman Saluja, Madhu Chopra. " Identification of novel inhibitors of MurI using homology modeling, structure-based pharmacophore, molecular docking, and molecular dynamics simulation-based approach ", Journal of Biomolecular Structure and Dynamics, 2022 Publication	<1 %
40	Bala Divya M, Maaged Abdullah, Shalini Saxena, Lalitha Guruprasad. " Inhibitor Binding Studies of MraY (Rv2156c): Insights from Molecular Modeling, Docking and Simulation Studies ", Journal of Biomolecular Structure and Dynamics, 2018 Publication	<1 %
41	Submitted to Indian Institute of Technology Student Paper	<1 %
42	scholar.harvard.edu Internet Source	<1 %
43	Submitted to Middlesex University Student Paper	<1 %

44	Birendra Kumar, P. Parasuraman, Thirupathihalli Pandurangappa Krishna Murthy, Manikanta Murahari, Vivek Chandramohan. " screening of therapeutic potentials from against the dimeric main protease (M) structure of SARS-CoV-2 ", Journal of Biomolecular Structure and Dynamics, 2021 Publication	<1 %
45	Haoqi Wang, Nirmitee Mulgaonkar, Lisa M. Pérez, Sandun Fernando. "ELIXIR-A: An Interactive Visualization Tool for Multi-Target Pharmacophore Refinement", ACS Omega, 2022 Publication	<1 %
46	Marco Wiltgen. "Algorithms for Structure Comparison and Analysis: Homology Modelling of Proteins", Elsevier BV, 2018 Publication	<1 %
47	worldwidescience.org Internet Source	<1 %
48	www.greenpharmacy.info Internet Source	<1 %
49	www.imrpress.com Internet Source	<1 %
50	Mohsen Sisakht, Aida Solhjoo, Amir Mahmoodzadeh, Mohammad Fathalipour,	<1 %

Maryam Kabiri, Amirhossein Sakhteman. "Potential inhibitors of the main protease of SARS-CoV-2 and modulators of arachidonic acid pathway: Non-steroidal anti-inflammatory drugs against COVID-19", Computers in Biology and Medicine, 2021
Publication

51	scholarspace.manoa.hawaii.edu Internet Source	<1%
52	Priyanka Andola, Jishu Pagag, Durgam Laxman, Lalitha Guruprasad. "Fragment-based inhibitor design for SARS-CoV2 main protease", Structural Chemistry, 2022 Publication	<1%
53	"Human Viruses: Diseases, Treatments and Vaccines", Springer Science and Business Media LLC, 2021 Publication	<1%
54	Manoj G. Damale, Shahebaaz K. Pathan, Rajesh B. Patil, Jaiprakash N. Sangshetti. " Pharmacoinformatics approaches to identify potential hits against tetraacyldisaccharide 4'-kinase (LpxK) of ", RSC Advances, 2020 Publication	<1%
55	Submitted to S.D. Asfendiyarov Kazakh National Medical University Student Paper	<1%

lalitha

Lalitha Guruprasad
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046.

56	Submitted to University of Hong Kong Student Paper	<1%
57	researchspace.ukzn.ac.za Internet Source	<1%
58	www.medtextpublications.com Internet Source	<1%
59	Seketoulie Keretsu, Suparna Ghosh, Seung Joo Cho. "Computer aided designing of novel pyrrolopyridine derivatives as JAK1 inhibitors", Scientific Reports, 2021 Publication	<1%
60	openaccess.sgul.ac.uk Internet Source	<1%
61	repository-tnmgrmu.ac.in Internet Source	<1%
62	udspace.udel.edu Internet Source	<1%
63	Amena Ali, Abuzer Ali, Musarrat Husain Warsi, Mohammad Akhlaquer Rahman, Mohamed Jawed Ahsan, Faizul Azam. "Toward the Discovery of a Novel Class of Leads for High Altitude Disorders by Virtual Screening and Molecular Dynamics Approaches Targeting Carbonic Anhydrase", International Journal of Molecular Sciences, 2022 Publication	<1%

64	Maaged Abdullah, Lalitha Guruprasad. "Computational fragment-based design of Wee1 kinase inhibitors with tricyclic core scaffolds", Structural Chemistry, 2018 Publication	<1 %
65	Raj Kumar, Vikas Kumar, Keun Woo Lee. "A computational drug repurposing approach in identifying the cephalosporin antibiotic and anti-hepatitis C drug derivatives for COVID-19 treatment", Computers in Biology and Medicine, 2020 Publication	<1 %
66	dbtindia.gov.in Internet Source	<1 %
67	hdl.handle.net Internet Source	<1 %
68	Abhishek Thakur, Gaurav Sharma, Vishnu Nayak Badavath, Venkatesan Jayaprakash, Kenneth M. Merz, Galia Blum, Orlando Acevedo. " Primer for Designing Main Protease (M) Inhibitors of SARS-CoV-2 ", The Journal of Physical Chemistry Letters, 2022 Publication	<1 %
69	Pushpendra Mani Mishra, Vladimir N. Uversky, Rajanish Giri. "Molecular Recognition Features in Zika Virus Proteome", Journal of Molecular Biology, 2018 Publication	<1 %

70	Shiv Bharadwaj, Akhilesh Kumar Rao, Vivek Dhar Dwivedi, Sarad Kumar Mishra, Umesh Yadava. "Structure-based screening and validation of bioactive compounds as Zika virus methyltransferase (MTase) inhibitors through first-principle density functional theory, classical molecular simulation and QM/MM affinity estimation", Journal of Biomolecular Structure and Dynamics, 2020 Publication	<1 %
71	kclpure.kcl.ac.uk Internet Source	<1 %
72	prr.hec.gov.pk Internet Source	<1 %
73	Jitender Dev G, Y. Poornachandra, K. Ratnakar Reddy, R. Naresh Kumar et al. "Synthesis of novel pyrazolo[3,4- b]quinolinyl acetamide analogs, their evaluation for antimicrobial and anticancer activities, validation by molecular modeling and CoMFA analysis", European Journal of Medicinal Chemistry, 2017 Publication	<1 %
74	Madeleine I. Ezeh, Onyinyechi E. Okonkwo, Innocent N. Okpoli, Chima E. Orji et al. "Chemoinformatic Design and Profiling of Derivatives of Dasabuvir, Efavirenz, and Tipranavir as Potential Inhibitors of Zika Virus	<1 %

RNA-Dependent RNA Polymerase and Methyltransferase", ACS Omega, 2022
Publication

75	Molecular Virology, 2013. Publication	<1 %
76	pubs.rsc.org Internet Source	<1 %
77	www.researchsquare.com Internet Source	<1 %
78	A.K. Saxena, S. Saxena, S.S. Chaudhaery. "Molecular modelling and docking studies on heat shock protein 90 (Hsp90) inhibitors", SAR and QSAR in Environmental Research, 2010 Publication	<1 %
79	Ernesto Estrada. "COVID-19 and SARS-CoV-2. Modeling the present, looking at the future", Physics Reports, 2020 Publication	<1 %
80	Io Antonopoulou, Eleftheria Sapountzaki, Ulrika Rova, Paul Christakopoulos. "Inhibition of the main protease of SARS-CoV-2 (Mpro) by repurposing/designing drug-like substances and utilizing nature's toolbox of bioactive compounds", Computational and Structural Biotechnology Journal, 2022 Publication	<1 %

81	Nabarun Chandra Das, Parth Sarthi Sen Gupta, Satyaranjan Biswal, Ritwik Patra, Malay Kumar Rana, Suprabhat Mukherjee. " evidences on filarial cystatin as a putative ligand of human TLR4 ", Journal of Biomolecular Structure and Dynamics, 2021 Publication	<1 %
82	Rajesh Ghosh, Ayon Chakraborty, Ashis Biswas, Snehasis Chowdhuri. " Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors – an docking and molecular dynamics simulation study ", Journal of Biomolecular Structure and Dynamics, 2020 Publication	<1 %
83	coek.info Internet Source	<1 %
84	www.cus.cam.ac.uk Internet Source	<1 %
85	Amira A. Alakhdar, Amr H. Saleh, Reem K. Arafa. " Targeting homologous recombination (HR) repair mechanism for cancer treatment: discovery of new potential UCHL-3 inhibitors virtual screening, molecular dynamics and binding mode analysis ", Journal of Biomolecular Structure and Dynamics, 2020 Publication	<1 %

86	Rahman Abdizadeh, Farzin Hadizadeh, Tooba Abdizadeh. "Evaluation of apigenin-based biflavonoid derivatives as potential therapeutic agents against viral protease (3CLpro) of SARS-CoV-2 via molecular docking, molecular dynamics and quantum mechanics studies", Journal of Biomolecular Structure and Dynamics, 2022 Publication	<1 %
87	dspace.cc.tut.fi Internet Source	<1 %
88	"Progress in the Chemistry of Organic Natural Products 110", Springer Science and Business Media LLC, 2019 Publication	<1 %
89	Andre S. Godoy, Gustavo M. A. Lima, Ketllyn I. Z. Oliveira, Naiara U. Torres et al. "Crystal structure of Zika virus NS5 RNA-dependent RNA polymerase", Nature Communications, 2017 Publication	<1 %
90	Submitted to Aston University Student Paper	<1 %
91	Sugunadevi Sakkiah, Keun Woo Lee. "Pharmacophore-based virtual screening and density functional theory approach to	<1 %

identifying novel butyrylcholinesterase inhibitors", Acta Pharmacologica Sinica, 2012
Publication

92	Submitted to University of Bristol Student Paper	<1%
93	chemicalsearch.info Internet Source	<1%
94	www.biorxiv.org Internet Source	<1%
95	www.chemdiv.com Internet Source	<1%
96	www.dovepress.com Internet Source	<1%
97	www.future-science.com Internet Source	<1%

Exclude quotes On
Exclude bibliography On
Exclude matches < 14 words