Computer-aided Drug Designing for Dengue targeting NS3 Helicase

M.Sc. Thesis

by

Mohini
(Roll No. 2203171012)

DEPARTMENT OF BIOSCIENCES & BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE

May 2024

Computer-aided Drug Designing for Dengue targeting NS3 Helicase

A THESIS

Submitted in partial fulfillment of the
requirements for the award of the degree
of
Master of Science

by
Mohini

DEPARTMENT OF BIOSCIENCES & BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
May 2024

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**Computer-aided Drug Designing for Dengue targeting NS3 Helicase"** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF BIOSCIENCES and BIOMEDICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2023 to May 2024 under the supervision of Dr. Parimal Kar, Associate Professor, Department of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Mohini

Signature of the student
Mohini

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Parimal Kar

Signature of the Supervisor

Dr. Parimal Kar

Mohini has successfully given her M.Sc. Oral Examination held on **9th May 2024.**

P.V. Kodgire

Signature of Supervisor of MSc thesis
Date:

DPGC
Convenor
Date:

Dedicated to my parents

and brother Harsh

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, **Dr. Parimal Kar**, for his invaluable guidance, support, and encouragement throughout the course of this research. His expertise, mentorship, and constructive feedback have been instrumental in shaping this thesis and my academic journey.

I am also thankful to the Head of the Department, **Dr. Amit Kumar**, the DPGC Convenor, **Dr. Prashant Kodgire**, and the **faculty members** at the Department of Biosciences and Biomedical Engineering at the Indian Institute of Technology Indore for their academic insights and continuous support.

I am grateful to all my colleagues at the Computational Biophysics Lab, **Mr. Sayan Poddar**, **Ms. Subhasmita Mahapatra**, **Mr. Suman Koirala**, **Mr. Kapil Ursal**, **Mr. Sunanda Samanta** and **Ms. Trupti Rathod**. With their constant understanding and support, the first tryst with research was fruitful and intellectually stimulating.

My heartfelt appreciation goes to my parents, **Mr. Rajbir Singh** and **Ms. Jyoti**, and my brother **Harsh** for their unwavering love, understanding, and encouragement. Their encouragement and belief in my abilities have been a constant source of strength and motivation.

I would like to extend my gratitude to my **friends** for their companionship, encouragement, and occasional distractions during moments of stress. Finally, I am grateful to **IIT Indore** for providing the necessary resources, facilities, and conducive environment for conducting this research.

Thank you to everyone who has contributed in any way to this endeavor. Your support and encouragement have been deeply appreciated.

Yours truly,

Mohini

ABSTRACT

The Dengue fever is considered as a “neglected tropical disease” as despite the staggering number of cases every year and significant loss of human and economic resources, a specific antiviral drug is not yet available for humankind. With the advent of global warming, increase in migration and urbanisation worldwide, dengue threat looms large over humanity, highlighting the dire need of an antiviral against the infection. In the present study titled as the “Computer-aided Drug Designing for Dengue targeting NS3 Helicase”, the main objective is to predict inhibitors against the target protein, that is, the Non-Structural protein 3 (NS3) of the Dengue virus. This protein is a multifunctional entity, performing NTPase, protease as well as 5’ RNA-triphosphatase activities. Its major function, however, is to perform unwinding of the single stranded RNA genome of the virus so that it becomes more accessible to the replication machinery, which it performs by deriving energy from ATP. We have performed thorough in-silico research and analysis, ranging from conformational analysis of the target protein, virtual screening, ADMET studies, Molecular Dynamics simulations and post-simulation analysis for structural stability as well as binding energetics, for the target protein in order to predict potential inhibitors. In the results of the various experiments, we have obtained positive findings and discovered three potential compounds against our chosen target protein of the Dengue virus. Two of the compounds are from the Natural Product Activity and Species Source Database, that is, they are of natural origin, while the third molecule comes from DrugBank, a repository of drug molecules approved by the FDA and can be repurposed as a drug against Dengue infection. All the three molecules have been subjected to extensive in-silico research and analysis from which we have produced positive findings that demonstrate their potential as inhibitors against the NS3 Helicase of Dengue virus. The present study is a small effort from our side to the larger cause of combating the dengue infection.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ADE	Antibody-dependent Enhancement
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
AI	Artificial Intelligence
AMBER	Assisted Model Building with Energy Refinement
CADD	Computer-aided Drug Designing
DCCM	Dynamic Cross Correlation Matrix
DENV	Dengue Virus
DFT	Density Functional Theory
DHF	Dengue Haemorrhagic Fever
DSS	Dengue Shock Syndrome
FDA	Food and Drug Administration
FEP	Free Energy Perturbation
GAFF2	General AMBER Force Field 2
HMDB	Human Metabolome Database
HTVS	High-throughput Virtual Screening
ICAMS	Intercellular Adhesion molecules
IVM	Integrated Vector Management
KDE	Kernel Density Estimation
LBDD	Ligand-based Drug Design

LBVS	Ligand-based Virtual Screening
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MM-GBSA	Molecular Mechanics with Generalised Born and Surface Area Solvation
MM-PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
NPASS	Natural Products Activity and Species Source Database
NSAIDs	Nonsteroidal Anti-Inflammatory Drugs
NSP	Non-structural Proteins
OPLS	Optimized Potentials for Liquid Simulations
PBC	Periodic Boundary Conditions
PCA	Principal Component Analysis
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
QM	Quantum mechanics
QSAR	Quantitative Structure-Activity Relationship
ROG	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
sander	Simulated Annealing with NMR-Derived Energy Restraints
SASA	Solvent-accessible Surface Area

SBDD	Structure-based Drug Design
SBVS	Structure-based Virtual Screening
SDF	Structured Data File
SMILES	Simplified Molecular Input Line Entry System
SP	Standard Precision Screening
TIP3P	Transferable Intermolecular Potential with 3 points
VS	Virtual Screening
WHO	World Health Organization
XP	Extra Precision Screening

Chapter 1
Dengue

Dengue is a viral disease inflicted by the **Dengue Virus (DENV).** It is also called **Breakbone fever**. This infection is prevalent usually in the sub-tropical and tropical regions of the planet, however, with the advent of global warming and climate change, infections are now spreading to non-endemic areas as well, raising major concerns regarding public health all over the globe. This hike in the count of dengue cases is also due to upward trends in overseas travel, growth of population and rapid urbanization, dearth of sanitation, inefficient vector (mosquito) control, and increases in the scrutiny and formal coverage of dengue cases. Data from the Centre for Disease Control and Prevention shows that each year, up to **400 million** humans are infected with dengue. Close to **100 million** people catch the infection, and **40,000** die from severe form of the disease. About **4 billion** people currently face the risk of contracting the disease. [1]

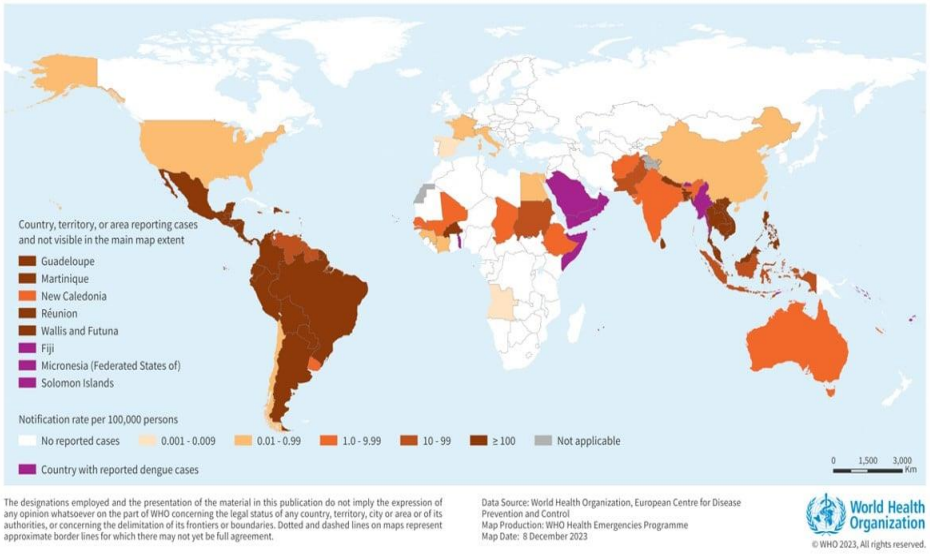


Fig.1: Countries/territories/areas reporting the number of dengue cases from November 2022-November 2023. [2]

Dengue is considered as a **neglected tropical disease** as despite the staggering number of infections and mortality rates, **no specific anti-viral drug** for dengue is yet available for humankind. Infection

management is majorly done via vector control, that is, limiting the spread as well as breeding of the major mosquito vector ***Aedes aegypti***.

1.1 Dengue Virus

The Dengue virus belongs to the family *Flaviviridae* and the genus *Flavivirus*. Dengue virus exists in four serologically and genetically different seroforms known as the Dengue virus 1, 2, 3 and 4. Recently, a fifth serotype was also discovered by Mustafa et. al [3] which highlights the rapidly evolving dynamics of the virus and presents further difficulties in combating the associated disease. As per the analysis of recent outbreaks, the serotype 2 has been shown to be the most prevalent in the Indian sub-continent. [4]

The virus is **enveloped** and contains a **single positive strand RNA genome of size 11 kb**. The genetic material encodes for **three structural**: Capsid (C), pre-membrane (prM) and Envelope (E) and **seven non-structural** proteins: NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. The envelope, capsid and membrane proteins make up the biological assembly of the virus. Out of these, the E protein helps the virus to enter into the host cells by binding to host cellular receptors (Eg. ICAMS-grabbing non-integrin, mannose receptor, Rab5, CD209, GRP78 etc). The other non-structural proteins are engaged in forming the **replication complex** that amplifies the viral genetic material and interacting with host proteins.

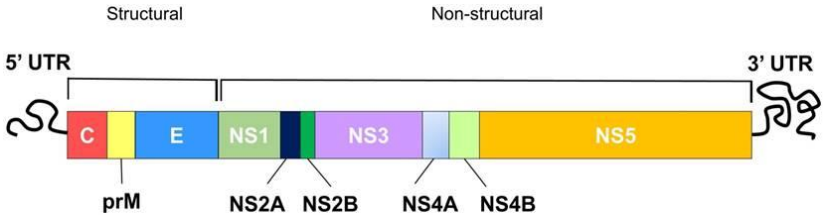


Fig.2: Representative image of the Dengue virus genome showing structural and non-structural polyproteins that are encoded by the viral genome. [5]

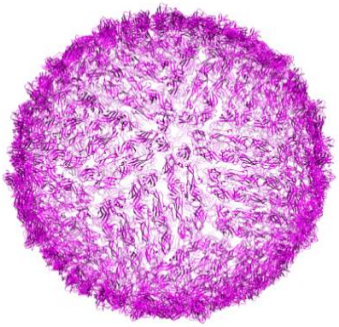


Fig.3: Dengue Virus (immature) [PDB ID: **1K4R]**

Table 1: Constituent proteins of DENV with molecular weight and amino acid residues.

Protein	**No. of Amino acids**	**Molecular weight (kDa)**
Capsid (C)	100	12
Membrane (M)	75	8.2-8.5
Envelope (E)	495	-
NS1	350	45
NS2 A/B	218	22
NS3	618	69
NS4 A/B	127/248	16/27
NS5	900	104

1.2 Life cycle of the Dengue virus

The Dengue virus enters inside the human body via the bite of an infected female *Aedes* mosquito. Upon entering the blood stream, the virus enters the White Blood Cells to reproduce. The envelope (E) protein of virus binds to host membrane receptors. Dengue virus then undergoes the phenomenon of endocytosis. The acidification of endosome occurs causing structural changes in the E protein and the

fusion of the endosomal membrane with the envelope. This causes the release of virion particles in the endoplasm of the cell. Inside the endoplasm, the uncoating of viral particle takes place and viral ssRNA becomes accessible to the host translational machinery. The host ribosomes translate (+) ss RNA into a single polyprotein. Following translation, the host and the viral enzymes together cleave the polyprotein into a total of 10 proteins (3 structural & 7 non-structural). These proteins include the RNA polymerase as well which synthesizes and further amplifies the viral genome, thus making the cell a "viral factory". The assembly happens on the intracellular membrane, which buds into the Endoplasmic reticulum forming the viral envelope. The Post translational modifications (glycosylation and pH-based transformational arrangements) occur within the Golgi apparatus. After successful assembly of a fully functional viral particle, it egresses via exocytosis, on its way to infect other healthy cells and increasing the viral load of the host. Meanwhile the process of replication, the WBCs produce cytokines and interferons which are responsible for causing pain and fever in the host. The main target of the virus are cells with binding receptors like DC-SIGN, TIM-TAM, Heparin sulfate and Heparin. The events of the viral life cycle and the increasing viral cycle correspond to the symptoms of Dengue fever in the host such as high vascular permeability, thrombocytopenia, cytokine storms etc.

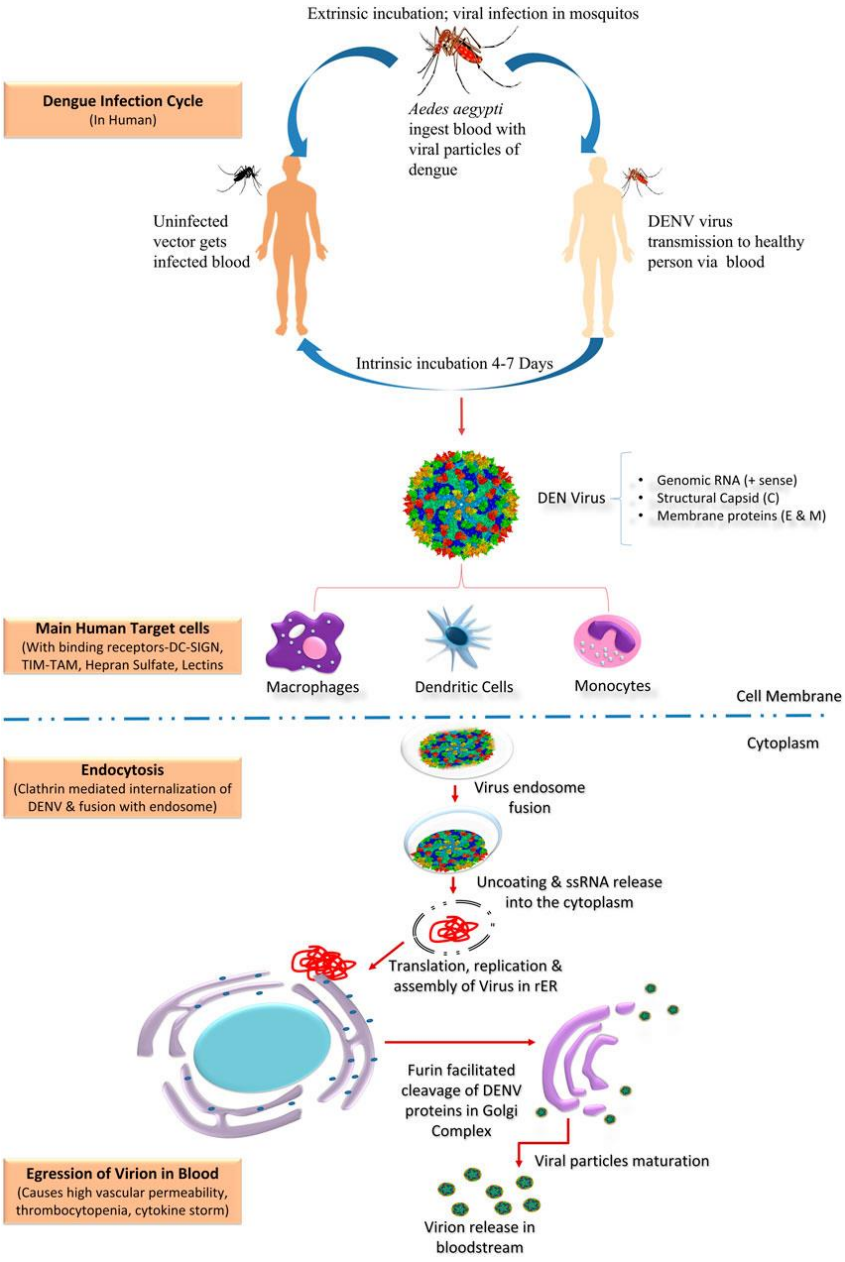


Fig.4: Processes of entry of DENV into the host cell and its life cycle. [6]

1.3 Transmission of DENV

The female mosquito *Aedes aegypti* carrying the Dengue virus is the main vector for the virus's transfer from human to mosquito to human. The tiny, dark-coloured *Aedes aegypti* mosquito has a silver-white pattern of scales on its body that help identify it, as do the white bands on its legs. This species of mosquito inhabits tropical and subtropical

regions of the world where humans are in close proximity to them. Because they need a warm environment to survive, *Aedes* mosquitoes typically do not reside above 1000 meters, where the temperature is lower. After feeding on the blood of an infected person, the female mosquito develops into a dengue vector. For the remainder of their lives, infected mosquitoes carry on spreading dengue with every blood meal they consume. Other species responsible for transmission includes *Aedes albopictus*, *Aedes polynesiensis* and *Aedes scuellaris*. However, their contribution is minimal as compared to *Aedes aegypti*. Non-vector transmission may happen through (a) Transfusion of blood, (b) Transplantation of organs or (c) needle stick injuries. Vertical transmission from mother to baby is common if she is viraemic at the time of delivery. Sexual transmission of the virus is rare and unusual, but one case was reported in Spain. [7] No other cases of sexual transmission, however, have been reported so far. Dengue poses the biggest threat in thickly inhabited areas with rainy seasons where there are large populations of the vector and a high degree of contact between the vector and the host.

1.4 Clinical manifestations of dengue fever

The spectrum of clinical symptoms of dengue fever is broad. The period of incubation for the virus is 4-7 days after which symptoms start appearing. The major symptoms of Dengue fever are high fever and chills, headache, retro-orbital pain, bone and muscle pain, nauseousness, vomiting, and change in sensation of taste and smell. In severe cases, plasma leakage may cause dengue to progress and manifest as **Dengue Shock Syndrome (DSS)** or **Dengue Haemorrhagic Fever (DHF)** which may prove to be fatal for the patient. [8]

1.5 Diagnosis of Dengue

Diagnosis of dengue involves the detection of the virus, genetic material of the virus, associated antigens or antibodies, or an aggregate of these substances. The choice of test depends on the day after onset

of illness. Laboratory diagnosis of NS1 antigen can help in early diagnosis in pyretic patients. For initial 4-5 days following the initiation of the infection, the virus can be detected in serum, plasma, blood cells, and other body organs. Antigen detection, examination of nucleic acid, and virus extraction can all be used to diagnose the disease during the acute phase of the infection. Serology is the recommended technique for diagnosis when the acute phase of the infection is coming to an end. In lieu of virus isolation or PCR (Polymerase Chain Reaction), serological assays that identify IgM or IgG antibodies against the dengue virus are very accessible and can assist in the diagnosis of dengue fever. IgM responses towards dengue virus infections are generally more potent and specific after the initial (primary) infection; in contrast, IgG responses are more potent but the IgM response is weaker in later (secondary) infections. The requirement to assess the specificity of commercially available assays is highlighted by these variable IgM response patterns to infection, particularly for the detection of secondary dengue infections. The Torniquet test or the Capillary Fragility Test is also frequently employed for diagnosing dengue. It determines the haemorrhagic tendency or capillary fragility and thromobocytopenia, that is, decrease in platelet count. The Torniquet test is recommended by the WHO for diagnosis of dengue but it is not very sensitive.

1.6 Infection management

Currently, **no specific antiviral drug exists** for dengue and management is usually done by **vector control.** Few drugs are in development, major approach has been to repurpose already existing drugs, Eg. Balapiravir, Chloroquine, Lovastatin, Ivermectin etc. Balapiravir is Hepatitis C virus NS5 polymerase protein inhibitor which progressed to clinical trials as a treatment for Dengue but was stopped in phase II due to no efficacy. [9] Janssen Pharmaceuticals has recently developed an NS4B inhibitor which has progressed to phase II clinical trials and shows good potential. The mechanism of action of the compound is to disrupt the interaction between two viral proteins

(NS3 and NS4B), hence disrupting the ability of the virus to replicate. [10] Treatment is usually by providing supportive care. Acetaminophen (paracetamol) is used for fever and pain. Aspirin, other salicylates, and nonsteroidal anti-inflammatory drugs (NSAIDs) are usually avoided as they may aggravate the risk of bleeding in patient.

The major focus is on Vector control. These days **Integrated Vector Management (IVM)** is practiced wherein multiple approaches such as biological, chemical and environmental are used for vector control.

- Biological methods: *Bacillus thuringiensis*, larva-eating fish and copepods are used for limiting mosquito larvae population.
- Release of *Wolbachia* infected mosquitoes: Bacteria *Wolbachia* renders mosquito resistant to arboviruses and can spread in population.
- Chemical methods: Insecticides like Temephos or Pyriproxyfen are used to control larvae of the mosquito.
- Environmental methods: This involves decreasing sites for mosquito breeding by manual intervention.

1.7 Vaccines

Currently two vaccines have been approved:

1. **Dengvaxia (CYD-TDV)**

This vaccine has been developed by **Sanofi Pasteur** (2015). It is a recombinant, live-attenuated and tetravalent vaccine developed on the 17D backbone of the yellow fever virus. It has been registered in 20 countries till now. The Clinical trials of the vaccine has shown complex performance. Its efficacy depends on serotype, age and baseline serostatistics. It increases susceptibility to severe complications of dengue fever due to phenomenon of antibody-dependent enhancement. [11]

2. **Qdenga (DENVax or TAK-003)**

This vaccine has been developed by **Takeda**. It is a recombinant chimeric attenuated vaccine. It is constituted by proteins of DENV 1, 3 and 4 etched onto DENV2 backbone. [12]

1.8 The Phenomenon of Antibody-dependent Enhancement (ADE) in the infection of Dengue virus

The phenomenon of Antibody-dependent enhancement in dengue infection was first postulated by **Halstead and colleagues**. [13]

The 4 serotypes of Dengue virus share some structural antigens. Infection in a person with one serotype induces two types of immunity in the host: type-specific, which lasts longer and cross-reactive, which is usually short-lived (2 months to 2 years). Immunization of an individual with one serotype of DENV enhances the risk of developing severe dengue on occurrence of subsequent infection with another serotype. This happens because cross reactive antibody binds heterologous virus to assist viral entry through Fc receptors of host cells. Also, virus and host interaction during ADE enables virus to escape host immune responses that would otherwise decrease infection. Hence, ADE increases infection burden, this causes imbalance of pro- and anti-inflammatory responses. This imbalance is responsible for inducing vascular leakage. This may result in potential hypovalaemic shock, that is, Dengue Shock Syndrome (DSS). DSS is fatal in most cases. However, the risk of developing ADE is not universal among all secondary Dengue infections, but requires an appropriate ratio of antibody-to-virus.

1.9 DENV NS3 Helicase

Out of the seven non-functional proteins translated from the DENV ssRNA, our chosen target for drug discovery is the NS3 Helicase. The Dengue virus NS3 Helicase is a large **multifunctional** protein having **618 residues** and molecular weight **69 kDa**. The structure of this ATP-dependent Helicase protein consists of two domains, surface charge distribution is not uniform for the unit and a tunnel-like cavity is present which can house single-stranded RNA of about six nucleotides but not a duplex. In order to take in a double-stranded substrate in this cavity, large changes in conformation, like interdomain hinge motions, would be required.

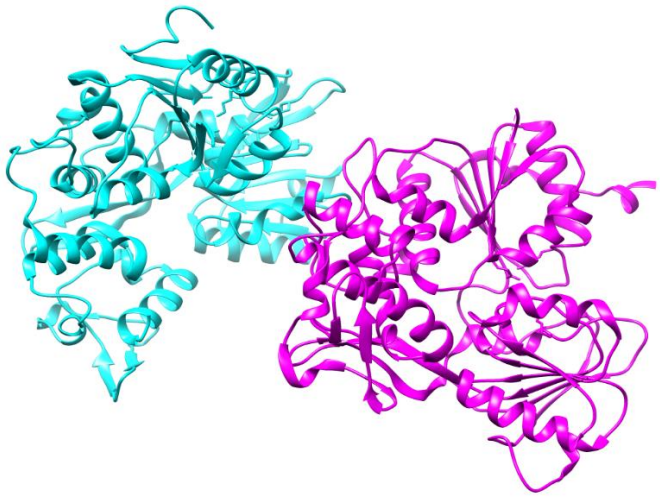


Fig.5: Dengue Virus NS3 Helicase protein (**PDB ID: 2BHR**) at resolution of 2.4 Å; Blue colored is the Chain A (1-180) (Serine protease domain) while the pink colored is the Chain B (180-618) (Helicase domain). [14]

This cavity is encrusted with several basic residues. The protein exhibits **two separate binding pockets**: one for binding ATP and the other tunnel-like cavity for binding of the RNA. ATP is the source of energy for the enzyme which unwinds the single-stranded genetic material of the Dengue virus and makes it accessible to replication and translational machinery.

The first **180 residues (N-terminal) of the enzyme** function as a **Serine protease**. The serine protease activity requires NS2B protein of the virus as co-factor. The next residues **180-618 (C-terminal)** show **RNA helicase** and **RTPase/NTPase** activities. This domain is a 3-lobed flattened structure containing a substantial number of loop regions and the dimensions are **60 × 60 × 35 Å**.

The helicase region, that is, from sequence 180 to 618 of the viral NS3 comprise of two motifs called "**Walker A, GK(S/T)",** and "**Walker B, DEx(D/H)"**. These motifs relate the helicase to a large group of nucleotide binding proteins that contribute to several cellular functions by coupling the breakdown of NTP with unwinding of nucleic acid duplex processing of genetic material and repair.

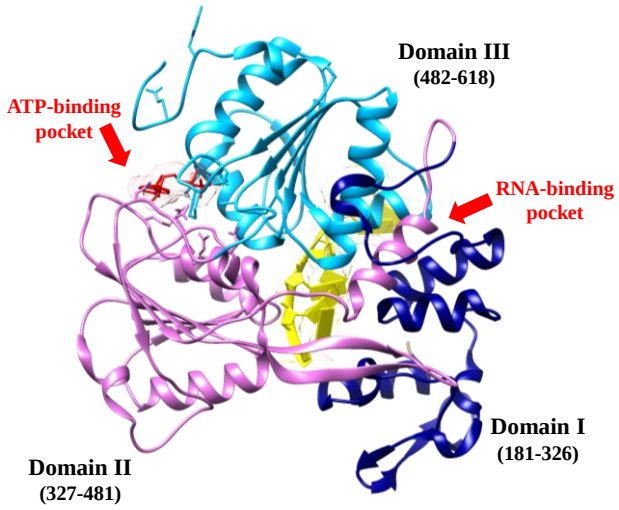


Fig.6: The NS3 Helicase protein of Dengue virus depicting the three domains and the two binding pockets: ATP-binding pocket and RNA-binding pocket.

NS3 Helicase plays a significant part in replication of viruses and exemplifies a fascinating target for the buildup of specific dengue virus inhibitors due to its diverse functions.

Chapter 2

Computer-aided Drug Designing

The process of discovering a drug is lengthy, extravagant and arduous that takes years to complete and costs billions of dollars. This process includes the identification of biological targets, discovering lead compounds, optimizing the discovered leads and finally, preclinical testing. In spite of such large ventures in terms of both capital and time, the success percentage of clinical testing of novel drugs is below 15%.[15] The reason for failure of approximately 50 per cent of novel molecules in the clinical testing is because of inferior pharmacokinetic properties, that is, ADMET properties ("Absorption, Distribution, Metabolism, Excretion and Toxicity"). However, nowadays, the pace as well as successful incidences of drug discovery have been dramatically enhanced with the advent of computational methods. Computational approaches and algorithms have had considerable impact on shortening time span as well as capital investments in finding novel drug candidates.

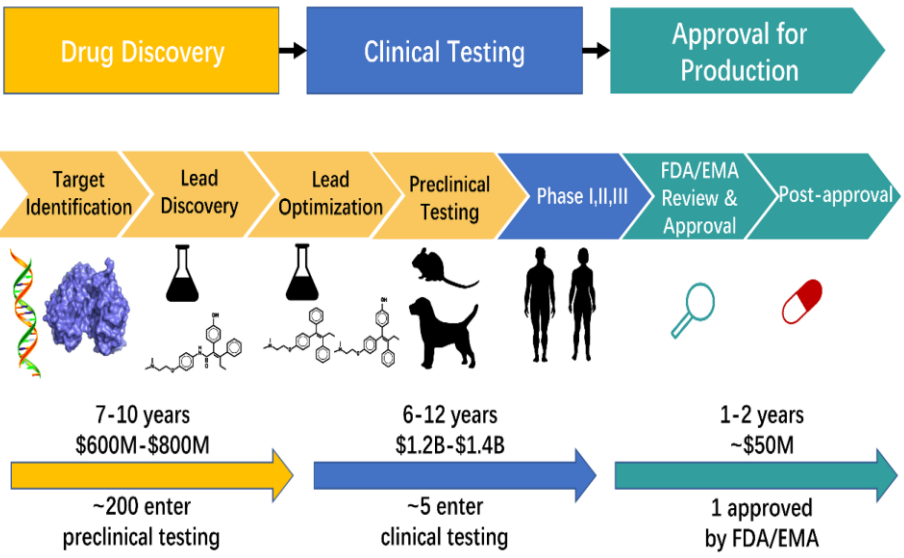


Fig.7: The arduous journey of a drug from the compound database to a chemist shop. [16]

In the process of discovering a drug, the input of computational science has been in the characterization of molecular mechanism underlying

the binding of the ligand and the biological target, discovery of novel binding or active sites and the refinement of binding positions of the ligand towards the target. Computational techniques can also assist in mapping the specific residues of these binding pockets. This information proves useful in assisting modification as well as optimization of the initial hit molecule and create new optimal interactions between the target and the ligand. In certain instances, pathologic activity may not be determined entirely by assessing the active site only. Cases where mutations in areas away from the binding site, changes in conformation, resistance towards drug and expression levels can induce or influence pathogenesis, need to be investigated holistically and computational approaches play an indispensable role in this. Computational biology, especially MD simulations, comes across as an influential and robust method for reporting the molecular mechanism of the biological target and contributing new contexts for drug designing.

MD simulations have been extensively deployed in the drug discovery process. According to Newton Mechanics, MD simulations map the position as well as the motion of every atom in the system under study. This assessment reveals the specifics of binding, unbinding, and changes in the conformation of the target protein. In addition to this, MD simulations also predict the kinetics, thermodynamics and the free energy profiles of the ligand-receptor interactions. Besides, due to the more trustworthy prediction of binding affinity by MD simulations, they can be used to legitimize the accuracy of predicted docking scores. Furthermore, quantum mechanics (QM) techniques, like the Density Functional Theory (DFT) and ab initio calculation approaches can be utilized to explore the interactions at the atomic as well as electronic levels between the ligand and the target during Virtual Screening. However, the Quantum Mechanics approaches are computationally very expensive and not applicable every time to the process of Virtual Screening.

Drug candidates and active compounds with comparable biological features are often found using computer-aided drug design, which also helps to speed up drug discovery and lower failure rates and costs. Most CADD-discovered medication candidates come from libraries of small molecules. Numerous techniques, like Virtual Screening, Pharmacophore modelling, Molecular docking, and Quantitative Structure–Activity Relationship (QSAR), are used to make these discoveries. Among these methods, VS is the main one used to filter novel compounds from big chemical databases that meet the necessary criteria. There are two types of VS: Ligand-based VS (LBVS) and Structure-based VS (SBVS). VS is used to reduce the number of compounds that need to be evaluated in the in-vivo testing and speed up drug discovery. Additionally, VS is crucial for drug repurposing and repositioning since it expedites the process of designing and developing new drugs by rapidly optimizing the drug candidates.

Recently, artificial intelligence has been touted as a potentially effective method for learning from and uncovering pharmacological large data in drug discovery, which has increased the chances of success rate of drug identification. Artificial intelligence is able to learn and generate new models for converting data into useful information by using the vast datasets from biomedical research. Prominent pharmaceutical corporations have used AI to improve the efficacy of their medication prospects, which has saved money and time by reducing the need for needless testing and synthesis. Machine learning and deep learning have also been integrated into the VS process. The methods aid in the creation of new chemicals as well.

2.1 Molecular Mechanics in Drug Design

By treating molecules according to the principles of classical mechanics, molecular mechanics (MM) avoids the need for the computer power needed for quantum calculations. The MM potential energy function, which is the total of various energy components

known as force fields, is responsible for doing this. MM potential energy functions find application in multiple sampling techniques, including Monte Carlo (MC) and MD. One of the most widely used algorithms for sampling is MD. It applies different integration techniques to the interpretation of classical Newton's equations of motion in order to study the trajectories, motions, and interactions in the biomolecular system under consideration. These algorithms include Verlet's Algorithm, Leap-frog Algorithm, and Beeman's Algorithm. MD provides time-dependent properties. A biomacromolecule within a solvent environment, like a protein or an enzyme, often makes up the system. Physical investigations are used to determine the basic protein structure for this protein or system. Afterwards, other techniques could be used to model the structure. After that, the prepared model is used to begin simulations. One experimental technique for determining the 3-D protein structure is X-ray crystallography. Unfortunately, the formation of stable crystals for X-ray imaging depends on the quality of the crystal, which restricts the ability to generate high-quality protein structures, particularly for membrane proteins. In the event that an experimental structure is not available, the structure must be predicted or modelled. DeepMind's AlphaFold and homology modelling are the recommended methods for obtaining the first protein structure. The dynamic properties of the system are provided by interactions between the system's atoms and molecules during the predetermined time in the molecular dynamics simulation. Newton's equations of motion often determine the paths of atoms. To compute the energy of the system, different force fields and molecular mechanics techniques are used. When compared to tests, molecular dynamics (MD) simulations can offer comprehensive insights into the folding process of the target protein and explain how variations in temperature, pH, and residue mutations affect the protein's structure. They can also provide extensive energetic information about these changes. Currently, a lot of research has been done using MD simulations to examine the target protein's molecular mechanisms in order to help in medication creation.

2.2 Molecular Docking

Compounds are docked in a specific spot in molecular docking based on space complementarity and energy match. Following that, a score system is used to rank and rate the docking stances. For the purpose of drug development, VS has proven invaluable based on molecular docking. With VS, drug discovery expenses and time are reduced, and different molecular scaffolds are obtained quickly. Pre-processing of compound libraries, molecular docking, and pretest compound selection make up the entire VS process. Generally speaking, VS success is mostly determined by the “enrichment factor”. As a method of validation, the enrichment factor calculates the ratio of active compounds among the total tested molecules in the original library, thereby assessing the efficacy of VS. Various techniques are applied to increase the enrichment factor for every VS phase. The precision of binding affinity and the rationale of target-ligand docking poses determine the VS outcomes. The binding affinity of the ligand with the receptor is evaluated using MD simulations in association with binding free energy estimates, in inclusion to the optimization of the conformations of the target and ligand. There are two general methods for calculating binding free energy: MM-PBSA and MM-GBSA. Van der Waals energy, solvation energy, and electrostatic energy are computed depending on the trajectories from MD simulations. Normal mode analysis can be used to determine the entropy change. It is therefore possible to obtain the binding free energy.

The calculations of binding free energy are excellent and helpful in increasing the precision of the binding affinity of docking conformations and enhancing the enrichment factor. However, these expensive sampling computations are frequently applied to a far smaller group of plausible hits.

2.3 Lead Optimization

For the purpose of comprehending the interactions between the receptor and the ligand and directing the modification of the screened molecules, the ideal binding mechanism and precise binding affinity are essential. Experiments can be used to obtain the ligand-target thermodynamical data, which are used to differentiate between molecules that are active and inactive. Examples of these data are entropy change (ΔS) and free energy change (ΔG). Nevertheless, additional structural alterations to the compounds are restricted due to the incomplete understanding of target-ligand interactions. One effective method for accurately assessing the ligand-target binding modes is molecular dynamics modelling. It has the ability to identify the precise ligand-target interactions as well as the free energy contribution of each binding site residue. The data can offer direction for optimizing leads.

2.4 Computer-Aided Drug Design (CADD)

Up till now, CADD has been responsible for the creation of nearly 70 authorized medications, ranging from Remdesivir in 2021 to Captopril in 1981. Structure-based drug design (SBDD) and Ligand-based drug design (LBDD) are two significant subcategories of CADD. To ascertain ligand-target interactions, SBDD is dependent on the target's and the active sites' three-dimensional structures. However, in situations where the target's three-dimensional structure is unknown, LBDD is employed. It starts with single or combination of compounds that are efficient against the target and is based on the relationship between structure and activity.

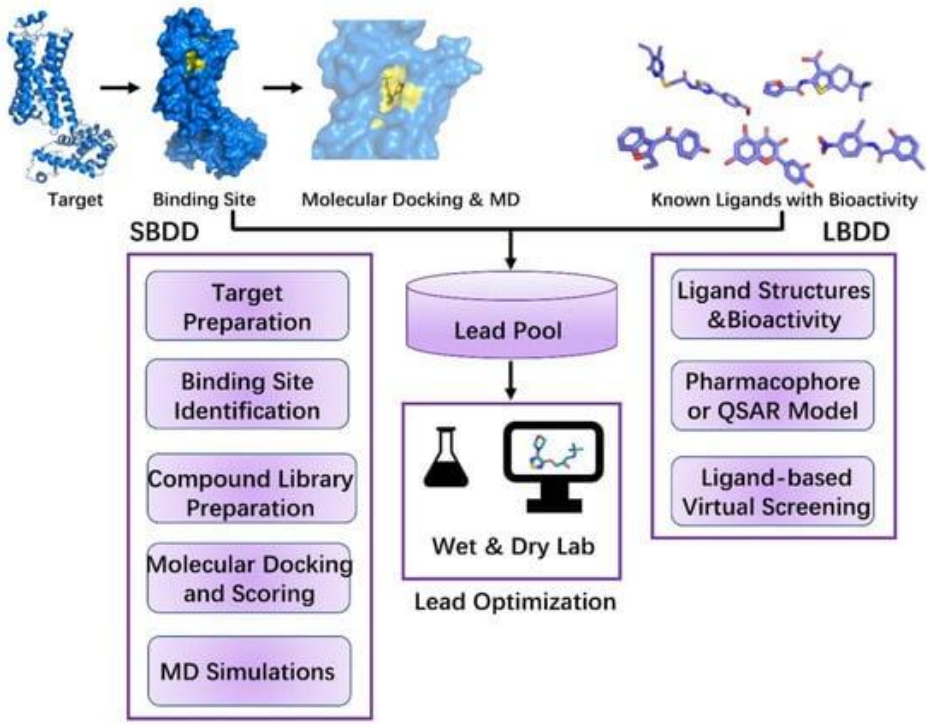


Fig.8: The process of designing drugs using both structure-based and ligand-based approaches. [16]

The first step in SBDD is target identification. Next, the target's binding site must be located, and a compound library must be assembled. Dock every compound in the library into the designated binding site after that, then calculate the score. MD simulations can be used in molecular docking to restore the docking process and produce a more flexible target. Furthermore, lead optimization through ligand-target interactions can be accomplished with MD simulations. It is mostly through these processes that leads are obtained. First, ligands with known bioactivity are used in LBDD. After that, create a pharmacophore or QSAR model by removing these ligands' chemical characteristics. Next, leads are screened using LBVS in the compound library based on information about known ligands (such as ligand similarity). Wet and dry labs are used to further improve these leads.

2.4.1 Structure-Based Drug Design

SBDD is a productive method for optimizing and finding leads. The three most popular SBDD techniques—molecular docking, SBVS, and MD simulations—are used to assess binding affinity, ligand–target interactions, and investigate conformational changes in the target. Several approved medications were found using SBDD, including Imatinib, Indinavir, Nilotinib, and Lifitegrast. Target preparation, binding site detection, compound database preparation, molecular docking, and Molecular Dynamics simulations are the primary components of SBDD.

2.4.2 Preparation of Target

As structural biology progresses, more target protein structures become available and are added to the PDB. Certain goal structures remain unattainable due to the constraints of experimental methodologies. Target structures can be predicted using computational techniques like homology modelling, AlphaFold, and ab initio protein structure prediction based on their sequences. In order to build the target structure, homology modelling chooses a suitable template structure. DeepMind's AlphaFold artificial intelligence technology uses amino acid sequences to predict three-dimensional protein structures with an approximate accuracy of experimentation. When the template structure is not accessible in the PDB, the ab initio i.e. from scratch protein structure prediction is thought to be appropriate.

Taking into account the primary structure of the particular target, this technique can assist in determining the tertiary structure with the least amount of energy.

2.4.3 Identification of Binding Site

Determining the binding location is a necessary first step in the molecular docking process. Site-directed mutation and the syn-crystallized complexes of ligand with protein provide knowledge about the binding sites of proteins under study. Blind docking is necessary to forecast the binding locations when any previous information about the

binding pocket is not known. To determine the most likely binding mode using the blind docking method, docking must be done throughout the whole protein surface. Several trials and energy evaluations are required throughout the entire process in order to determine the ideal ligand-target complex posture. Though blind docking is less dependable and more constrained than normal docking because of insufficient sampling at the docking space, blind docking is significant in identifying novel interactions that might be present in the undiscovered binding modes.

Several tools, such as DeepSite, Fpocket, RaptorX-Binding Site, COACH, and PocketDepth, are designed to predict target protein binding sites using blind docking.

2.4.4 Compound Library Preparation

ZINC, PubChem, DrugBank, ChEMBL, and ChemDB et cetera are some of the compound libraries from which compounds used for VS are chosen. The Lipinski's Rule of Five, the Veber criteria, ADMET, and additional particular characteristics (such as carcinogenicity, immunogenicity and hepatotoxicity) were used to filter the compounds. According to Lipinski's "Rule of Five" and the Veber criteria, a compound is considered orally bioactive if its molecular weight is less than 500 Da, its hydrogen bond donors (HBD) and acceptors (HBA) are less than 5, its octanol-water partition coefficient logP is less than 5, its rotatable bonds (RotB) are less than 10, and its topological polar surface area is less than 140. [17]

In addition, the compounds' synthetic accessibility needs to be taken into account. The optimal three-dimensional structure of the ligand should be modelled after ligands have been filtered out of libraries.

2.4.5 Molecular Docking and Scoring

Currently, in the presence of a 3D biological structure, the search procedure is made simpler by combining molecular docking with visual search. It is employed to evaluate interactions of ligand with the target at the atomic level and use scoring systems to rank the ligands in

order of binding affinity. The molecular docking tools that are most commonly used are SwissDock, Autodock, AutoDock Vina, and GLIDE. In terms of ligand and target flexibility, molecular docking approaches comprise three types of docking: (1) rigid docking, in which the ligand and target structures are both rigid; (2) semi-flexible docking, the most popular type of docking, in which the ligand structure is kept flexible but the target is in rigid confirmation; and (3) flexible docking, in which the ligand and target conformations are both flexible.

Scoring functions, which are used to ascertain binding affinities and ligand-target binding modalities as well as suggest possible therapeutic candidates, are what make molecular docking accurate. Scoring functions based on physics, empirical data, expertise, and machine learning are accessible.

2.4.6 MD Simulations

SBDD has made substantial use of MD simulations. MD simulations can help target conformations with well-defined binding sites as well as flexibility for molecular docking by increasing the target protein's flexibility. Additionally, lead tuning and docking scoring can be accomplished with MD simulations. When used in conjunction with free energy computations, MD simulations can precisely determine binding affinity and enhance the precision of compound ranking. MD simulations can be used in lead optimization on small groups of compounds to identify ligand-target interactions, which can then be used to guide ligand synthesis.

Molecular Dynamics (MD) simulation is a computational technique used to study the conduct of atoms and molecules over time. Following are the detailed steps involved in an MD simulation:

System Setup: This involves specifying the molecular or atomic composition of the system under study. It includes determining the types and numbers of molecules or atoms and their initial spatial arrangement.

Choosing a force field: A force field is a set of mathematical equations that expresses the interactions between atoms or molecules within the system. These equations typically account for bonded and non-bonded interactions such as stretching of bonds, distortion of angles, torsional rotations, and van der Waals forces.

Specifying boundary conditions: This step involves defining the size and shape of the simulation box in which the system resides. It also determines whether the simulation box has periodic boundary conditions, allowing for the simulation of an infinite system by replicating the contents of the box periodically.

Solvation: Solvent molecules need to be added around the system under study to mimic the biological environment conditions accurately.

Initialization: Initial velocities are assigned to each atom or molecule according to the desired temperature, typically following a Maxwell-Boltzmann distribution.

Calculating initial forces: Forces acting on each atom are computed based on their positions and the force field parameters. These forces determine how atoms will move in the subsequent steps of the simulation.

Integration of Equations of Motion: An integration scheme dictates how positions and velocities are updated at each time step. Popular schemes include Verlet, Leapfrog, and Velocity Verlet.

Integrating Newton's equations of motion: Using the chosen integration scheme, the positions and velocities of atoms are updated iteratively over small time steps (Δt), typically on the order of femtoseconds (10^{-15} seconds), based on the forces acting on them.

Simulation Run: The integration of equations of motion is repeated for a specified number of time steps or until a desired simulation time is reached.

Output trajectory data: At regular intervals, the positions, velocities, and energies of atoms are recorded to generate a trajectory file, allowing for the analysis of the system's evolution over time.

Analysis: Structural analysis: This involves analyzing changes in the molecular structure, such as bond lengths, angles, and dihedral angles, to understand the conformational changes within the system.

Dynamic properties: Properties like diffusion coefficients, radial distribution functions, and vibrational spectra are calculated to gain insights into the system's dynamic behavior.

Thermodynamic properties: Thermodynamic quantities such as temperature, pressure, and free energy are computed using statistical mechanics methods to characterize the system's thermodynamic state.

Visualization: Trajectory data is visualized using software tools to aid in the interpretation and understanding of the simulation results.

The specific processes involved in an MD simulation are:

Minimization: Minimization is the process of optimizing the initial configuration of the system to alleviate steric clashes and reduce potential energy. It involves iteratively adjusting atomic positions to minimize the potential energy of the system while satisfying force and energy convergence criteria. Minimization is typically performed using optimization algorithms such as steepest descent or conjugate gradient until a local minimum in the potential energy surface is reached.

Heating: Heating involves increasing the temperature of the system from an initial low temperature to the desired simulation temperature. This process is often necessary to initiate molecular motion and overcome energy barriers, allowing the system to explore its conformational space more effectively. Heating can be achieved by scaling the velocities of atoms according to a temperature schedule or by applying a thermostat that regulates the system's temperature.

Equilibration: Equilibration is the phase where the system is allowed to reach a stable thermodynamic state under the desired conditions (e.g., temperature, pressure). During equilibration, fluctuations in thermodynamic properties such as temperature, pressure, and density are monitored to ensure that the system has reached equilibrium. Various ensemble-based algorithms, such as NVT (constant Number of particles, Volume, and Temperature) or NPT (constant Number of particles, Pressure, and Temperature), are used to maintain the desired conditions during equilibration.

Production Run: The production run is the main phase of the MD simulation where data for analysis is collected. It involves running the simulation for an extended period, allowing the system to explore its phase space and sample various configurations. Trajectory data generated during the production run is used for subsequent analysis of system properties and behaviors. Each of these steps is crucial in ensuring the accuracy, reliability, and relevance of MD simulations in exploring the dynamics and properties of molecular systems.

2.5 Conclusions and Perspectives

Drug design has been greatly aided by the widespread application of computational biology techniques in target discovery, mechanism analysis, virtual screening, and optimization of leads. These methods are theoretically well-founded. Research on thermodynamic and kinetic properties, as well as investigations of molecular mechanisms, continue to rely heavily on Molecular Dynamics simulations, including force field-based and ab initio simulations. MD simulations are still necessary for precise determination of the free energy change of the ligand-target as well as for capturing the structural and dynamical characteristics of targets in drug design. While molecular force field-based simulations are more accurate than the ab initio method, they can be used at greater scales. This flaw is compensated for by the QM/MM approach, which is progressively used in drug discovery.

However, the extensive applicability of ab initio MD simulation is limited by the sheer quantity of computational chores that impede MD simulation's growth to greater sizes. All things considered, drug design has expedited due to the introduction and widespread use of CADD, molecular docking, virtual screening, and QSAR. Lately, scientists have made significant strides in expediting the conventional drug discovery paradigm through the application of AI techniques. Generative models on the basis of molecular graphs or strings, like SMILES, have gained popularity in the field of molecular generation because of their superior performance in a range of molecular optimization tasks. While AI techniques and computational modelling of intricate protein machines have proven to be more effective in drug creation, there are still a number of issues with the current framework that need to be resolved.

Chapter 3

Objectives

- **To study the NS3 Helicase protein of Dengue virus using Molecular Dynamics simulations.**

  This objective involves performing detailed analysis of our target structure derived from the Protein Data Bank, including identification of missing residues and their construction using appropriate tools and techniques and studying the dynamics of the target at a molecular level by performing MD simulations. This objective shall assist in identification of suitable sites for inhibitor binding.

- **To study the structural dynamics of the inhibitory action of Suramin against NS3 Helicase using Molecular Dynamics Studies.**

  Upon survey of literature, we have identified Suramin as a potential Control molecule for the present study. Suramin is a medicine used for treating diseases African sleeping sickness and river blindness. The molecular formula of suramin is $C_{51}H_{40}N_6O_{23}S_6$.

- **To identify novel inhibitors against the target protein using in-silico approaches.**

  This objective involves screening compounds from different databases against our target molecule in order to identify potential inhibitors. Here, we are employing Structure-based Virtual Screening (SBVS). Several software for conducting Virtual Screening are available viz. Glide module of Schrodinger, MolSoft, PharmScreen, PyRx etc. Recently, AI and Deep Learning based VS software have also been

developed which have enhanced the speed as well as accuracy of VS.

Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) are the pharmacokinetic properties of a compound which are responsible for the failure of majority of novel molecules in clinical trials. It is hence of utmost importance to assess the ADMET properties of the hits carefully in order to minimize chances of rejection. Several webservers are available, for example, ADMET Lab 2.0, ProTox-II, SwissADME etc, which can be utilized for the purpose.

- **To explore the structural dynamics and energetics of binding of the discovered molecules to DENV NS3 Helicase using MD simulations and free energy calculations respectively.**

  Molecular Dynamics simulations are an indispensable tool in Computer-aided Drug discovery. They provide crucial insights into the molecular level structural and binding changes of our receptor-ligand complex. The binding free energy calculations provide information regarding the quantitative energy changes associated with the formation of the complex. This information can be used to score the different hits and identify the best ligand.

  The potential lead molecules are subjected to extended MD analysis wherein the time scale of simulation is prolonged. This assists us in deciphering the long-term interactions as well as stability and activity of our lead molecules with the target protein. Under this objective, we also plan to compare the lead molecules from the two different pockets investigating which pocket shows more potential as a drug binding site.

Chapter 4

Methodology

4.1 System preparation

Receptor Preparation: The protein structure was derived from the database "Protein Data Bank" (PDB ID: 2BHR). The structure has a resolution of 2.8Å. [18] The structure was analysed and the two respective binding pockets were mapped using molecular visualisation software package Chimera. [19] We found that the structure has missing residues from 243 to 254. The missing residues (loop) of the structure were added using ModLoop software. [20]

Ligand Preparation: The choice of control ligand (Suramin) was made after conducting literature survey. The ligand structure was derived from the database PubChem (ID: 8514). [21] The molecular formula of the compound is $C_{51}H_{34}N_6Na_6O_{23}S_6$ while the molecular weight is 1429.2 g/mol.

(a) (b)

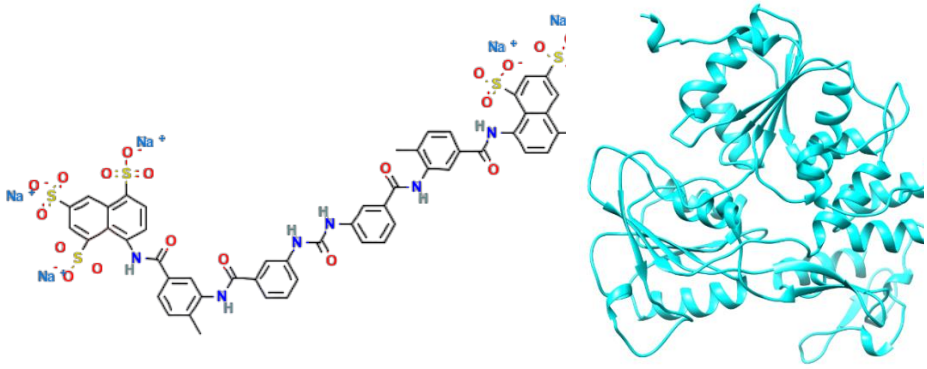


Fig.9: (a) Structure of Ligand molecule Suramin and (b) Apo structure of DENV NS3 Helicase (2BHR)

4.2 Virtual Screening of databases against NS3 Helicase

4.2.1 Database selection and Ligand preparation

For realising our objective of finding a suitable inhibitor against DENV NS3 Helicase, we have seleected three databases to screen i.e. the **Human Metabolome Database (HMDB)**, [22] the **Natural Product Activity and Species Source Database (NPASS)** [23] and the **DrugBank** [24]. Our focus was on discovering a natural inhibitor and hence the choice of first two databases. Our objective was also to repurpose already existing drugs and hence the choice of DrugBank. In total, 12122 compounds from HMDB, 30926 compounds from NPASS and 11586 compounds from Drugbank, were screened twice, once against the ATP-binding pocket of the protein and once against the RNA-binding pocket, making the total number of compounds screened to be **54634×2=1,09,268**.

All the compounds were imported in the Structured Data File (SDF) format onto the LigPrep module of the Schrodinger suite. [25] All the conformer states as well as the tautomeric states of the compounds were produced in the Maestro portal. Finally, all the compounds underwent minimization and optimization using the force field OPLS3. [26]

4.2.2 Virtual Screening

All the 109268 compounds were screened against the two different pockets of 2BHR using the virtual screening function of the **Glide module of the Schrodinger suite**. In the screening process, we have employed the QikProp module and Lipinski's rule of five to screen out the compounds for each database in the course of screening. However, Lipinski filter was disabled during screening against RNA-binding pocket as discussed in later sections. We have employed flexible molecular docking during the screening process. The selected method of docking was semi-flexible, wherein our receptor is inflexible but the ligand is flexible to move inside of the binding pocket. The virtual

screening was executed progressively in three levels: **High-throughput Virtual Screening (HTVS), Standard Precision Screening (SP)** and **Extra Precision Screening (XP)**. The top scoring molecules from XP were taken and used for further investigation.

4.2.3 Pharmacological profiles and toxicity investigation of compounds

Following screening, the resultant molecules underwent various computational based pharmacological profiles and toxicity assessment. We have primarily used the ADMET Lab 2.0 server [27] to assess the ADME (Absorption, Distribution, Metabolism, and Excretion) parameters of molecules, representing drug-like tendency. We have also verified the results using the SwissADME server [28]. Another crucial aspect in development of a drug is to evaluate the toxicity characteristics of compounds. Therefore, hepatotoxicity, carcinogenicity, immunotoxicity, mutagenicity, and cytotoxicity of the chosen compounds were estimated using the ProTox-II web server [29]. After doing the ADMET analysis of our Glide-XP screened molecules, the top lead systems i.e. with highest docking score values, were taken for analysis of structural stability using the MD simulation studies and MM-PBSA free energy calculations.

The complete workflow has been summarised in the diagram below.

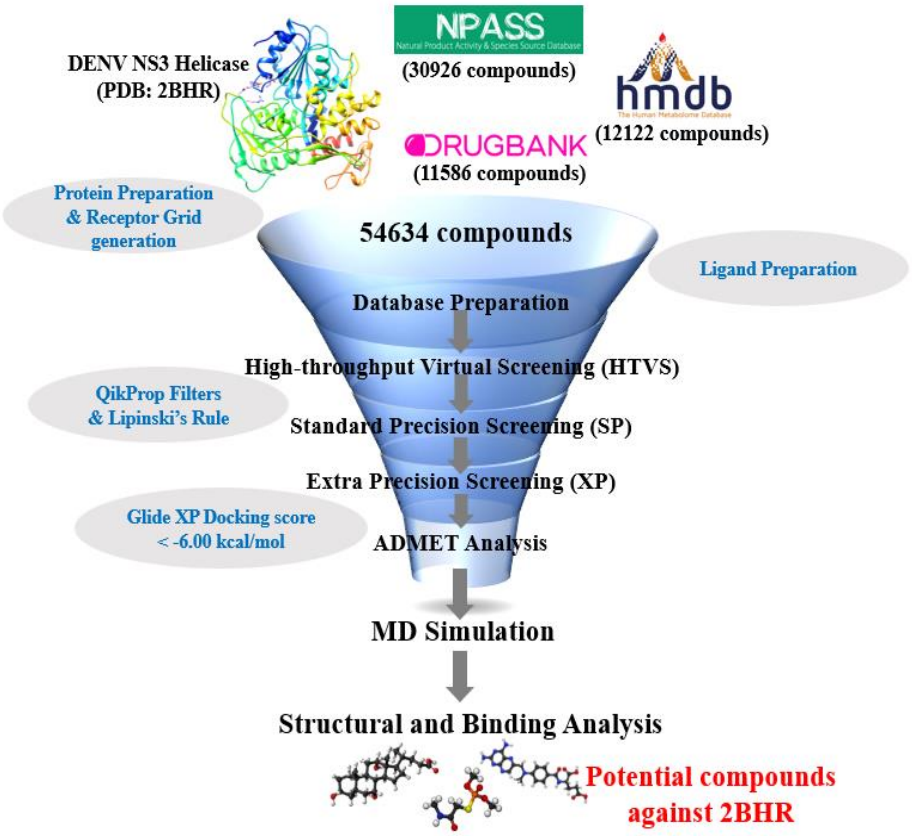


Fig.10: Schematic diagram representing the workflow of virtual screening and post-screening analysis of the hit molecules.

4.3 Molecular Dynamics Simulation Protocol

MD simulations are based on the Newtonian laws of motion, where the movement of each atom is analysed computationally over a period of time. The coordinates and corresponding velocity of all atoms are calculated, which helps to visualize the dynamics of the molecule.

All the simulations have been executed using the **AMBER18** software package. [30] The **LEaP module** from **AMBERTools19** has been used to add the required hydrogen atoms. The systems were neutralized with suitable counterions (Na^+ and Cl^-). For solvation of the system, we have employed the **TIP3P** [31] **octahedron** water box at distance of 10 Å. For receptor, the Amber **ff14SB** [32] forcefield was used while **GAFF2** [30] was used for the ligands. **2.0 fs** was the

time step. Restraining was done using the **SHAKE** [34] algorithm. The **PDB 2BHR** was used for apo protein simulation. All the water molecules and ions present on the PDB were deleted and then it was further prepared by using the **pdb4amber** module.

The first step in performing the simulation was **minimization** wherein lowest energy conformations of the assembly is obtained and other stereochemistry defects corrected. Subsequently, **heating** of the system is done slowly from 0 to 300 K. The temperature and pressure were sustained using the **Langevin thermostat** [35] and the **Barendsen barostat** [36] respectively. After expelling all restraints, we performed **equilibration** in the NPT ensemble and then **production** run was performed for 200 ns (×2).

Post MD Simulation, analysis has been done using the **cpptraj module** [37] of Amber18 for the following parameters:

1. ***Root Mean Square Deviations (RMSD)***

   RMSD, which can indicate conformational similarity between the two sets of values (atomic coordinates), is performed by aligning the structures by least square method which rotates one structure around the centre to match the other structure's orientation. RMSD essentially measures the deviation of the target coordinate from the reference coordinate.

$$\text{RMSD} = \sqrt{\frac{\sum_{i=0}^{N} m_i (X_i - Y_i)^2}{M}} \qquad (5.1)$$

   In this equation, N is the atom number, mi is the mass of the atom i, Xi the target coordinate of atom i, Yi is the reference coordinate of atom i, and M is the total mass.

2. ***Solvent-Accessible Surface Area (SASA)***

SASA is used to get electrostatic interactions of molecules in ionic solvent where the equation is

$$\nabla . [\varepsilon(r)\nabla\varphi(r)] - k' \sinh[\varphi(r)] = -4\pi\rho(r) \quad (5.2)$$

3. *Root Mean Square Fluctuation (RMSF)*

RMSF is defined as the root mean square of the distance from an atom in a structure to its average position over time. Hence it represents the fluctuation of the said atom from its average position.

$$\text{RMSF} = \sqrt{\frac{\sum_{t_j=1}^{T}\left[x_{i(t_j)} - (x_i)\right]^2}{T}} \quad (5.3)$$

In contrast to RMSD where time specific value is obtained by taking average over particles, RMSF averages the value for each particle over time and reflects the positional deviation of the entire structure over time. Flexibility of different segments of a protein in a simulation can be understood from RMSF.

4. *Radius of Gyration (R_g)*

It is used to measure protein compactness as per the formula:

$$R_g = \left(\frac{\Sigma_i |r_i|^2 m_i}{\Sigma_i m_i}\right)^2 \quad (5.4)$$

Here, the mass of the atom i is m_i and the position of atom i with respect to the centre of mass of the molecule is represented by r_i.

5. *Binding Free Energy calculation (MM-PBSA Scheme)*

This is a post-processing technique that computes the free energy shift between two states (usually the bound and free states of a receptor and ligand) using representative snapshots from a wide collection of conformations. [38] The calculation of differences in free energy involves the integration of both the solvation free energy components (polar and non-polar) derived from an implicit solvent model for each species and the gas phase energy contributions that are independent of the selected solvent model. As an additional refinement, entropy contributions to the total free energy are added. A mmpbsa_py_nabnmode [39] application created in the nab programming language is used to calculate entropy in the gas phase. Based on the force field used to construct the topology files, the gas phase free energy contributions are computed using either sander from the Amber software suite or mmpbsa_py_energy from the AmberTools package. It is possible to further break down the solvation free energy contributions into hydrophobic and electrostatic components. The Poisson Boltzmann (PB) equation, the Generalized Born technique, or the Reference Interaction Site Model (RISM) are used to compute the electrostatic component. The LCPO technique used in Sanders approximates the hydrophobic contribution.

The binding free energy has been calculated as per following formulas:

$$\Delta G_{\text{binding}} = \Delta H - T\Delta S \approx \Delta E_{\text{internal}} + \Delta G_{\text{sol}} - T\Delta S \quad (5.5)$$

$$\Delta E_{\text{internal}} = \Delta E_{\text{covalent}} + \Delta E_{\text{elec}} + \Delta E_{\text{vdw}} \quad (5.6)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}} \quad (5.7)$$

6. MM-PBSA per residue

By adding up each residue's interactions over the entire system, per-residue breakdown determines the energy contribution of individual residues. Amber offers multiple methods for breaking the computed free energy into distinct residue contributions by utilizing the GB or PB implicit solvent models. Per-residue decomposition is a method of breaking down interactions for each residue by only including those in which at least one of the residue's atoms is involved. Alternatively, interactions can be pairwise dissected by designating certain residue pairs and only include interactions in which one atom from each of the residues under analysis is involved. These breakdown techniques can offer helpful insights into significant interactions in computations of free energy.

7. DCCM

Quantifying the correlation coefficients of motions between atoms has been done a lot using the dynamic cross correlation (DCC) methodology. The following formula, where ri(t) indicates the vector of the ith atom's coordinates as a function of time t, indicates the temporal ensemble average, and defines the DCC between the ith and jth atoms. The cross corelation coefficient of the i t™ and j t™ alpha carbons that correspond to the particular residues is given in the matrix, and the DCC generates a N x N heatmap, where N is the number of alpha carbons in the protein. In this case, the correlation coefficient ranges from -1 to +1, where +1 denotes correlation and -1 anticorrelation.

$$\mathrm{DCC} = \frac{\langle \Delta r_i(t) \cdot \Delta r_j(t) \rangle_t}{\sqrt{\langle \|\Delta r_i(t)\|^2 \rangle_t} \sqrt{\left\langle \|\Delta r_j(t)\|^2 \right\rangle_t}} \tag{5.8}$$

where $\boldsymbol{r_i(t)}$ denotes the vector of the *i*th atom's coordinates as a function of time *t*.

The DCC creates a heatmap of N x N, where N is the number of alpha carbons in the protein and the cross-correlation coefficient of the ith and jth alpha carbons that corresponds to the specific residues is given in the matrix. Here the correlation coefficient varies between -1 and +1 where -1 indicates anti correlation and +1 indicates correlation. This function helps us understand the activities of residues by providing information on their correlation.

8. Principal Component Analysis

Principal Component Analysis or PCA is a method of dimension reduction that calculates the principal components (eigenvectors with corresponding high eigenvalues) using the atomic coordinates from the MD trajectories. The eigenvectors represent the motion direction, and the corresponding eigenvalues describe the motion amplitude. In PCA analysis from MD trajectories, the number of dimensions of each atom is calculated and to eliminate the translation and rotational motions, the mean value is set to zero by subtracting the average from each dimension. Next, a covariance matrix is calculated where the Cij element is represented by $Cij = < xi >)$ $(xj- < xj >)$ where the $i\ th$ and $j\ th$ atom coordinates are represented by xi and xj respectively and the mean average of $i\ th$ and $j\ th$ atom coordinates are given by $< xi >)$ and $< xj >$ respectively. Covariance matrix of 3N x 3N is generated for N number of atoms for the three dimensions and is diagonalized to get the eigenvalues. $A\ TC\ A = \gamma$. Here, A is the eigenvector and γ is the eigenvalue. The eigenvectors are arranged in decreasing order of eigenvalue and the eigenvector corresponding to the highest eigenvalue gives the first principal component. The first few principal components, that together

can describe the system with sufficient accuracy, are generally used in analysis. [40]

9. Hydrogen Bond analysis

Hydrogen bonds arise from electrostatic interactions between hydrogen donor and acceptor groups, facilitated by the partial positive charge of hydrogen and the electronegative atoms (e.g., oxygen or nitrogen) of the receptor. The geometry and strength of hydrogen bonds modulate ligand-receptor interactions, influencing binding kinetics and thermodynamics. Specificity and Selectivity: Hydrogen bonds confer specificity and selectivity to ligand-receptor interactions by enabling precise geometric complementarity and recognition between complementary binding partners. Structural analyses reveal the intricate network of hydrogen bonds orchestrating molecular recognition events with exquisite specificity. Affinity and Binding Kinetics: The formation of hydrogen bonds contributes to the overall binding affinity of ligand-receptor complexes, influencing association and dissociation kinetics. Molecular dynamics simulations and kinetic experiments elucidate the dynamic nature of hydrogen bond-mediated interactions and their impact on binding energetics.

10. Potential Mean Force (PMF)

In molecular dynamics (MD) simulations, the potential mean force (PMF) is a thermodynamic quantity that characterizes the free energy landscape along a reaction coordinate or a specific degree of freedom within a system. It provides valuable insights into the energetics and kinetics of molecular processes, such as ligand binding, protein conformational changes, or chemical reactions. [41]

The PMF represents the reversible work done on a system as it undergoes a transition along a chosen reaction coordinate. It quantifies the energy barriers and minima encountered along

the reaction pathway, allowing researchers to elucidate the driving forces and mechanisms governing the process of interest.

The calculation of PMF in MD simulations typically involves employing enhanced sampling techniques, such as umbrella sampling, metadynamics, or adaptive biasing force methods. These methods bias the simulation to explore specific regions of the configurational space corresponding to different values of the reaction coordinate, thereby reconstructing the free energy profile.

11. Network Analysis

In a protein structure network, the amino acids are represented by the nodes, and the relationships between them are shown by the edges. A node's degree is determined by how many direct links it forms, and hubs are higher-degree nodes. Hubs are defined as residues that make four or more connections with other residues. If a residue appears as a hub in at least 50% of the MD simulation snapshots, it is considered to be dynamically stable. The variations in protein kinase's structural connection between its ligand bound and unbound states are reflected in alterations to the quantity and location of the dynamically stable hub residues. They are referred to as hot spots, which is essential for preserving both structural stability and information transmission. However, a subset of residues can be defined as intra-linked nodes in which each residue is related to every other residue. In the context of protein structures, these indicate areas of structural rigidity by signifying higher-order connectivity in a network. Communities are defined as an assemblage of intra-linked nodes through common edges and interactions. Communities are the means by which structural rigidity percolates across the network of protein structures. Comparing nodes and

communities together reveals minute conformational alterations that modify the rigidity and flexibility of protein structural organization. We have performed Network analysis of our systems using the webPSN server. [42]

Chapter 5

Results and Discussion

5.1 Screening and structural analysis results for leads of the ATP-binding pocket

5.1.1 Virtual Screening

Table 2: List of top hit compounds obtained from virtual screening of the databases against the **ATP-binding pocket** of our target protein. Molecular weight is given in **Daltons** while Dock score is given in **kcal/mol**. The given compounds also showed positive results in ADMET analysis and exhibit no toxicity.

Database	Ligand ID	Mol. weight	Dock score
HMDB	HMDB0028851	176.17	-11.67
HMDB	HMDB0028847	206.26	-11.26
HMDB	HMDB0029043	218.25	-10.89
HMDB	HMDB0028854	174.10	-10.87
HMDB	HMDB0028682	203.20	-10.38
DrugBank	DB08045	359.81	-7.80
DrugBank	DB11924	301.52	-7.44
DrugBank	DB02829	252.25	-7.30
DrugBank	DB12937	382.22	-7.23
NPASS	NPC329495	192.10	-11.33
NPASS	NPC317147	188.12	-11.27
NPASS	NPC18223	155.07	-10.46
NPASS	NPC68974	162.10	-10.29
NPASS	NPC327542	119.06	-10.02
NPASS	NPC60672	86.04	-9.97

We have performed the Virtual screening in steps: HTVS, SP, followed by XP screening. Based on favourable docking scores, we have filtered out compounds at each stage. Finally, we selected compounds with docking score below -4.52 kcal/mol (which is the docking score of our control drug Suramin). The selected compounds had a dock score in the range of -12.98 to -7.23 kcal/mol. The compounds were then subjected to ADMET analysis, and the best compounds were subjected for structural analysis by Molecular simulation investigation and Free energy estimation. Following all the steps, we concluded with the four best leads, whose structural analysis is as follows.

Table 3: ADMET Properties of the top hit compounds for ATP-binding pocket.

Compound	**MW**	**logP**	**logS**	**Caco-2**	**MDCK**	**TPSA**
Suramin	1429.20	0.21	-4.68	-5.99	8.5e-06	500.73
HMDB0028851	176.17	-2.34	-0.53	-6.3	0.005	112.65
HMDB0028847	206.26	-1.46	-0.99	-6.21	4.7e-06	92.42
HMDB0029043	218.25	-1.20	-0.49	-6.12	0.006	112.65
HMDB0028854	174.10	-1.22	-0.06	-6.19	0.0022	92.42
HMDB0028682	203.20	-2.88	-0.50	-6.34	0.003	135.51
DB08045	359.81	3.35	-4.03	-5.14	4e-06	104.39
DB11924	301.52	5.12	-4.55	-5.15	2.6e-05	66.48
DB02829	252.25	-0.54	-2.57	-6.28	6.9e-06	123.14
DB12937	382.22	0.95	-3.41	-5.86	8.6e-06	177.89
NPC329495	192.10	-2.71	-0.44	-6.43	0.018	132.88
NPC317147	188.12	-2.02	-0.39	-5.29	0.002	95.91
NPC18223	155.07	-3.25	-1.08	-5.50	1.4e-05	92.00
NPC68974	162.10	-3.21	-0.01	-6.14	0.004	109.57
NPC327542	119.06	-2.59	0.39	-6.09	0.006	83.55
NPC60672	86.04	-3.44	-1.04	-6.29	0.003	134.45

Parameters for ADMET:

1. **logP**: Predicted octanol/water partition coefficient (acceptable range: 2.0 to 6.5).
2. **logS**: Predicted aqueous solubility in mol/L (acceptable range: 6.5 to 0.5).
3. Predicted **Caco-2 cell permeability** in nm/s (acceptable range: < 25 is poor and > 500 is optimal).
4. Predicted apparent **MDCK cell permeability** in nm/s (acceptable range: < 25 is poor > 500 is optimal).
5. **Topological Polar Surface Area (TPSA)** (optimal range from 60-160 $Å^2$).

5.1.2 Structural and flexibility analysis: RMSD

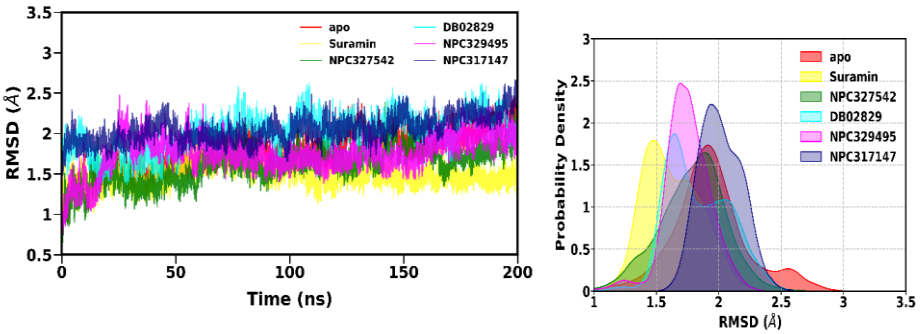


Fig.11: Backbone RMSD analysis of the top lead compounds in comparison to the apo and control, with associated probability density plot.

To check the simulation convergency, structural deviation and stability, we have calculated the RMSD of backbone atoms of protein with respect to their respective initial coordinates for all systems. To gather insights into the stability of the complete protein backbone, we have employed the Kernel Density Estimation (KDE) method to plot the probability density of RMSD for the entire course of the simulation. It is observed that all the simulated systems achieve convergence and RMSD values range between 1.0 to 2.5 and are

comparable to the control. The peak RMSD of NPC317147, however, shows a slightly higher value than apo.

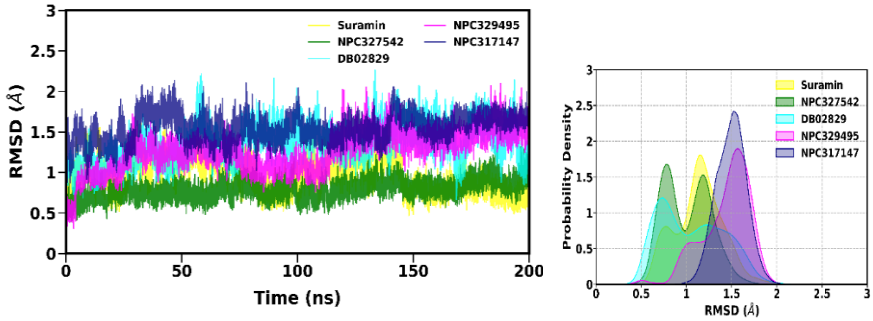


Fig.12: RMSD analysis of the binding pocket, with associated probability density plot.

The RMSD results of the binding pocket of our receptor in different systems depicts convergence within a window of 0.5 to 2. In all systems, binding pocket shows considerable stability.

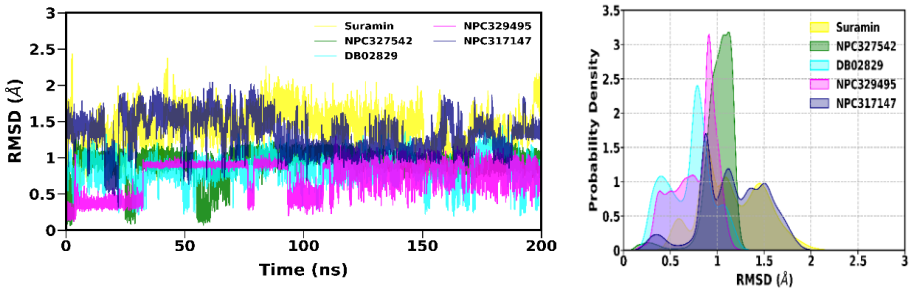


Fig.13: RMSD analysis of the ligands, with associated probability density plot.

In the RMSD analysis of ligands, it is observed that despite initial fluctuations, the system converges at the end within a window of 2.0 RMSD value.

5.1.3 Ligand-Receptor distance analysis

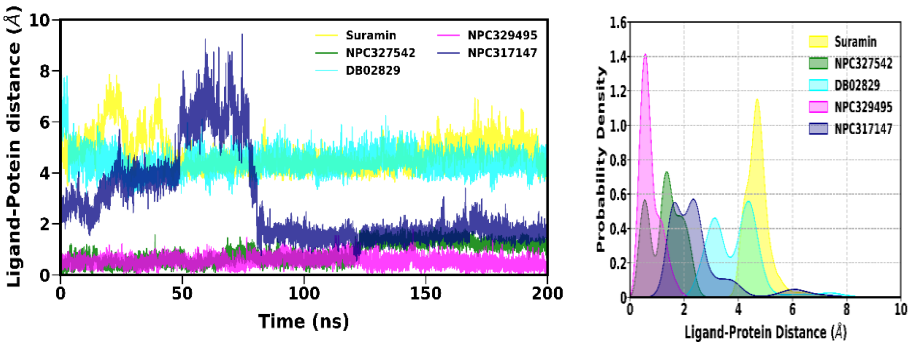


Fig.14: The Ligand-Protein distance analysis, with associated probability density plot.

Upon analyzing the distance between the ligand and the receptor over the course of simulation, we observe that DB02829, NPC329495 and NPC327542 show a constant position with respect to the protein. However, our control shows fluctuation for the initial 50ns but attains stability later on. Similarly, NPC317147 shows high fluctuation for the first 80ns but stabilizes for the rest simulation course.

5.1.4 Protein compactness and solvent accessibility analysis

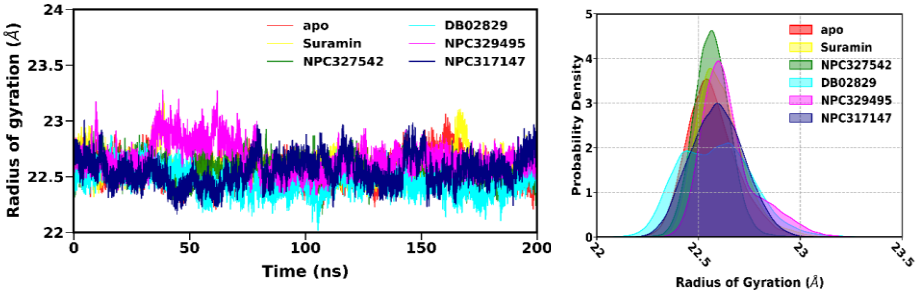


Fig.15: The Radius of Gyration analysis of the lead compounds, with associated probability density plot.

The Radius of Gyration patterns tell us about the compactness of the structure. The plot shows comparable Radius of Gyration values for all the systems.

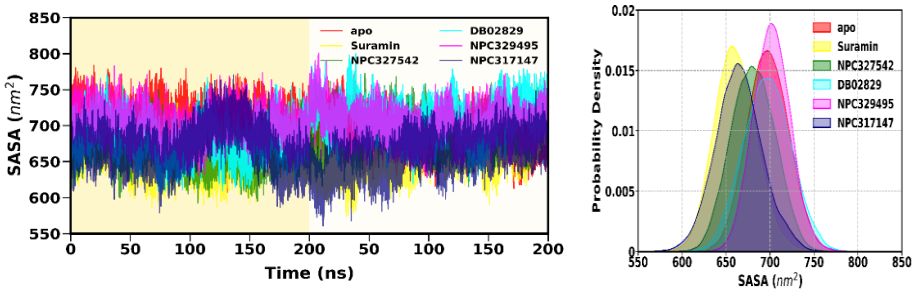


Fig.16: The Solvent-Accessible Surface Area Analysis (SASA) analysis, with associated probability density plot.

The SASA values for systems Suramin, NPC317147 and NPC327542 are less in comparison to the apo structure. The systems NPC329495 and DB02829 show comparable values of SASA to the apo protein structure.

5.1.5 Residual Stability analysis

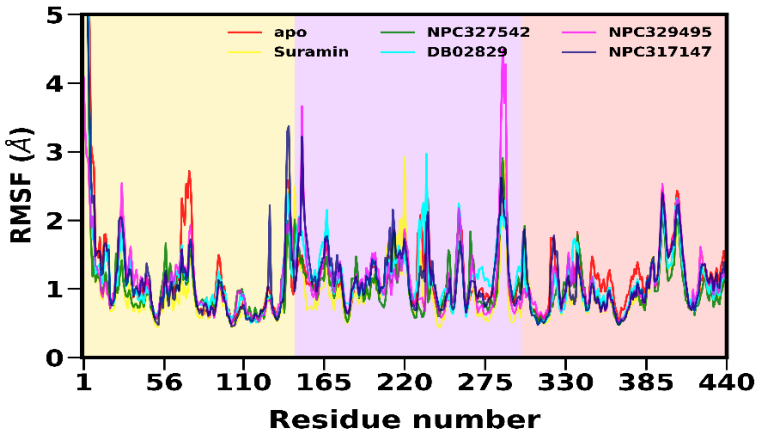


Fig.17: The Root Mean Square Fluctuation analysis of the lead compounds.

The per-residue fluctuation plot for our systems shows comparable values except for certain instances. Residues in the range 65-80 show considerable fluctuation in apo structure while they are much stabilized in complexes. The residues in in range 279-285 show considerable fluctuation in complex NPC329495, the phenomenon can be an area of further investigation.

5.1.6 Binding Free Energy Analysis

All the four systems show good binding affinity towards the target protein in comparison to the control. We have not calculated the entropy contribution because of the high computation cost. It is observed that van der Waals and electrostatic interactions contribute towards the binding of the complex. The contribution of electrostatic interactions is however, very significant. This reflects the involvement of charged molecules in the binding. The polar interactions are shown to oppose the binding of the complex.

Table 4: The Binding Free Energy of the various ligand-protein complexes with their respective components.

Ligand	**ΔG_{vdW}**	**ΔG_{elec}**	**ΔG_{polar}**	**$\Delta G_{nonpolar}$**	**ΔG_{bind} (kcal/mol)**
Suramin	-41.05	-31.49	55.05	-3.57	-21.06
NPC327542	-11.62	-178.09	140.49	-1.84	-51.07
DB02829	-23.35	-207.96	183.35	-2.45	-50.41
NPC329495	-19.53	-228.25	200.04	-2.60	-50.34
NPC317147	-34.78	-56.44	57.49	2.68	-36.42

5.2 Screening and structural analysis results for leads of the RNA-binding pocket

5.2.1 Virtual Screening

Table 5: List of top hit compounds obtained from virtual screening of the databases against the **RNA-binding pocket** of our target protein. Molecular weight is given in **Daltons** while Dock score is given in **kcal/mol**. The given compounds also showed positive results in ADMET analysis and exhibit no toxicity.

Database	Ligand ID	Molecular weight	Dock score
HMDB	HMDB0029794	938.75	-16.31
HMDB	HMDB0012253	990.86	-15.05
HMDB	HMDB0006602	1072.96	-14.81
HMDB	HMDB0006593	998.88	-14.15
HMDB	HMDB0006696	853.77	-13.24
DrugBank	DB03198	1039.06	-16.29
DrugBank	DB03971	969.89	-15.98
DrugBank	DB01719	876.92	-14.40
DrugBank	DB03208	940.68	-12.90
DrugBank	DB04453	807.74	-12.74
DrugBank	DB11938	679.86	-11.74
NPASS	NPC60982	788.26	-14.08
NPASS	NPC106944	356.13	-13.64
NPASS	NPC57751	596.20	-13.29
NPASS	NPC141455	756.25	-12.98
NPASS	NPC205195	640.2	-12.93

We have performed the Virtual screening in steps: HTVS, SP, followed by XP screening. Based on favourable docking scores, we have filtered out compounds at each stage. Finally, we selected compounds with docking score below -8,70 kcal/mol (which is the docking score of our control drug Suramin). The selected compounds had a dock score in the range of -16.31 to -10.45 kcal/mol. The compounds were then subjected to ADMET analysis and the best compounds were subjected for structural analysis by Molecular simulation studies and Free energy calculations.

Table 6: ADMET Properties of the top hit compounds for RNA-binding pocket

Compound	MW	logP	logS	Caco2	MDCK	TPSA
Suramin	1429.20	0.21	-4.67	-5.99	8.5e-06	500.73
HMDB0029794	938.75	-0.61	-3.20	-6.12	0.001	442.66
HMDB0012253	990.86	-4.29	-1.19	-6.68	0.002	506.13
HMDB0006602	1072.96	-4.57	-0.49	-6.70	0.001	531.71
HMDB0006593	998.88	-4.06	0.08	-7.05	0.001	510.09
HMDB0006696	853.77	-4.05	0.34	-6.65	0.001	423.46
DB03198	1039.06	-3.18	-1.08	-6.51	0.001	478.44
DB03971	969.89	-3.89	-0.86	-6.51	0.002	479.47
DB01719	876.92	-2.49	-1.25	-6.36	0.001	388.29
DB03208	940.68	2.39	-2.31	-7.15	1.3e-05	444.18
DB04453	807.74	-3.38	-0.69	-6.51	0.001	400.32
DB11938	679.86	0.25	-2.09	-6.30	8.8e-06	222.97
NPC60982	788.26	-3.02	-0.90	-6.72	0.001	366.29
NPC106944	356.13	-0.49	-2.69	-6.52	2.8e-05	259.45
NPC57751	596.20	-2.54	-0.98	-6.65	0.001	277.91
NPC141455	756.25	-0.22	-2.90	-6.53	2.1e-05	304.21
NPC205195	640.20	-0.41	-2.43	-6.45	1.9e-05	265.52

Following all the steps, we concluded with the four best leads, whose structural analysis is as follows.

5.2.2 Structural and flexibility analysis: RMSD

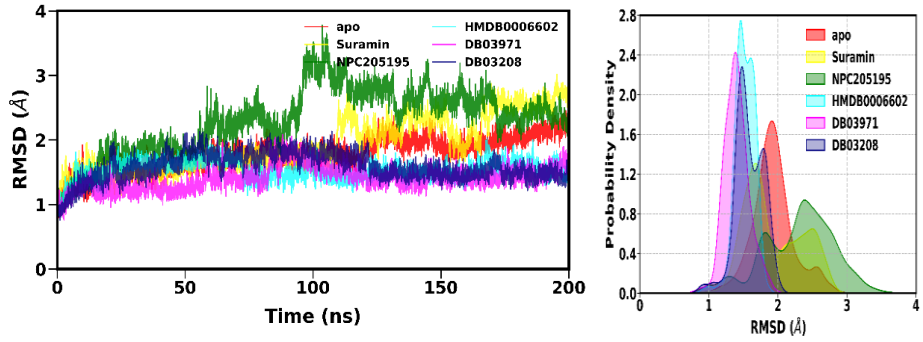


Fig.18: Backbone RMSD analysis of the top lead compounds in comparison to the apo and control, with associated probability density plot.

In the backbone RMSD analysis, HMDB0006602, DB03971 and DB03208 show less RMSD values than the apo structure, indicating structural stability. For NPC205195, however, fluctuations between 80 to 120 ns are seen but it ultimately converges. For our control structure, we observe high variance after 100 ns.

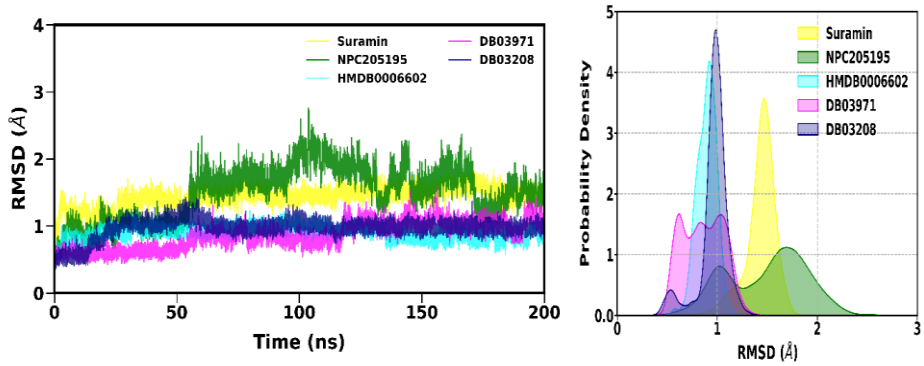


Fig.19: RMSD analysis of the binding pocket, with associated probability density plot.

The RMSD results of the binding pocket of all the systems show considerable stability and convergence within a window of 0.5 to 2 RMSD values. NPC205195 complex shows fluctuation from 50 to 170 ns but converges ultimately.

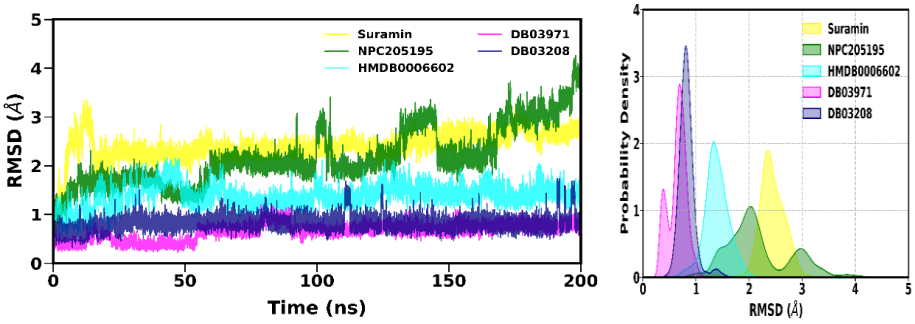


Fig.20: RMSD analysis of the ligands, with associated probability density plot.

The RMSD analysis of ligand for DB03971, DB03208 and HMDB0006602 show high structural stability at RMSD values less than 2 Å. Our control complex shows high RMSD value, while NPC205195 ligand shows high fluctuation, especially after160 ns.

5.2.3 Ligand-Receptor distance analysis

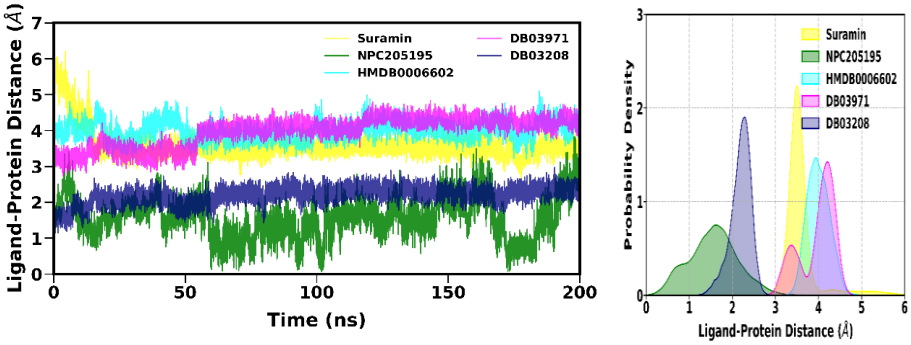


Fig.21: The Ligand-Protein distance analysis, with associated probability density plot.

The analysis of the distance between the ligand-protein shows uniform constant respective values for each system as seen in the graph.

5.2.4 Protein compactness and solvent accessibility analyses

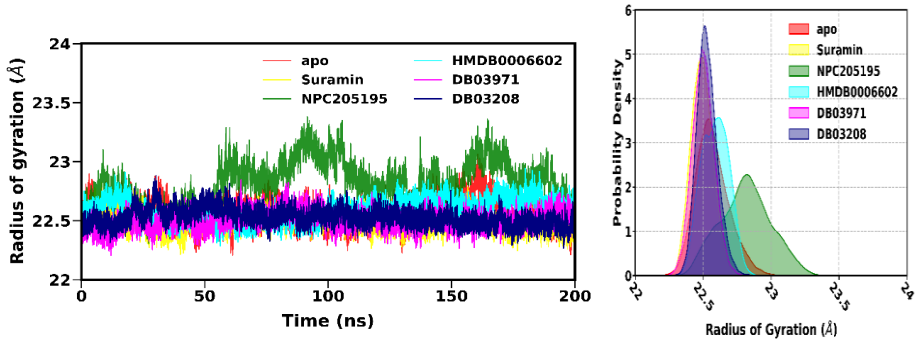


Fig.22: The Radius of Gyration analysis of the lead compounds, with associated probability density plot.

The Radius of Gyration values for the system are assessed to estimate the structure compactness. Here, we observe that all the systems, except NPC205195, show comparable values of Radius of gyration to apo structure. The ROG values for NPC205195 fluctuate over the course of simulation but are seen to converge at the end.

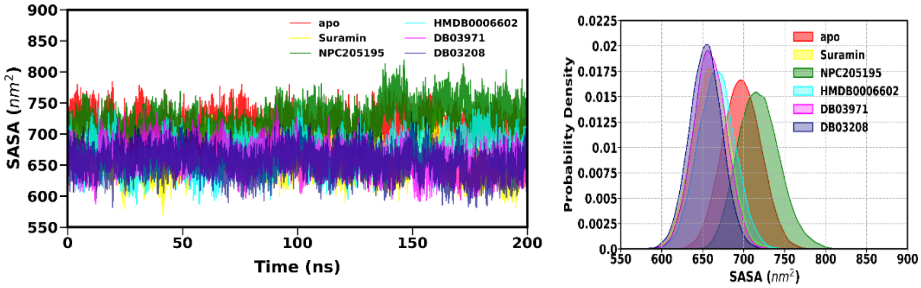


Fig.23: The Solvent-Accessible Surface Area Analysis (SASA) analysis, with associated probability density plot.

The SASA values for all systems, except NPC205195, are less than apo structure. This is because upon binding of the inhibitor, the surface area earlier accessible to the solvent decreases. In the case of NPC205195, however, we notice that the SASA value is higher than the apo value.

5.2.5 Residual Stability analysis

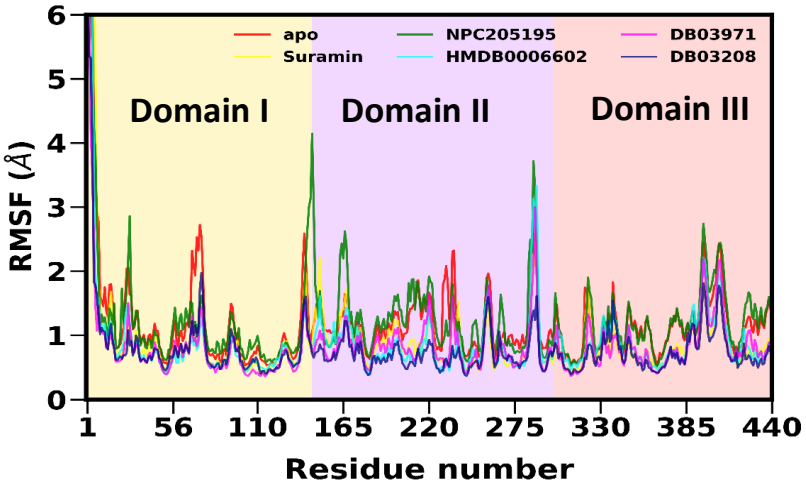


Fig.24: The Root Mean Square Fluctuation analysis of the lead compounds. Also highlighted in the figure are the respective Domains of DENV NS3 Helicase.

The per-residue fluctuation analysis plot shows the overall stability in complexes is higher than the apo structure. However, in the case of NPC205195, we observe high fluctuations in residues 25-30, 130-135 and 163-165.

5.2.6 Binding Free Energy Analysis

Table 7: The Binding Free Energy of the various ligand-protein complexes with their respective components.

Ligand	ΔG_{vdW}	ΔG_{elec}	ΔG_{polar}	$\Delta G_{nonpolar}$	ΔG_{bind} (kcal/mol)
Suramin	-107.54	-68.14	150.04	-8.77	-34.40
NPC205195	-51.86	-86.40	103.66	-6.38	-40.98
HMDB0006602	-75.70	-128.79	167.97	-8.56	-45.09
DB03971	-66.65	-134.96	158.16	-7.44	-50.90
DB03208	-55.28	-118.39	138.62	-6.64	-41.70

All the four systems show good binding affinity towards the target protein in comparison to the control. We have not calculated the entropy contribution because of the high computation cost. It is observed that van der Waals and electrostatic interactions contribute towards the binding of the complex. The contribution of electrostatic interactions is however, slightly higher. This reflects the involvement of charged molecules in the binding. The polar interactions are shown to oppose the binding of the ligand to the receptor.

5.3 Selected compounds and their potential as DENV NS3 Helicase inhibitors

After obtaining the enthalpy results for the compounds, we have included the entropy contributions also to the overall binding free energy and obtained the following results:

1) For the ATP-pocket leads

Table 8: Hit molecules for the ATP-pocket and their respective Binding free energy values. The contribution from various interactions is also depicted. All values are in kcal/mol.

Ligand	ΔG_{vdW}	ΔG_{elec}	ΔG_{polar}	ΔG_{np}	ΔH	$T\Delta S$	ΔG_{bind}
Suramin	-41.05	-31.49	55.05	-3.57	**-21.06**	**-18.22**	**-2.84**
NPC327542	-11.62	-178.09	140.49	-1.84	**-51.07**	**-18.61**	**-32.46**
DB02829	-23.35	-207.96	183.35	-2.45	**-50.41**	--	--
NPC329495	-19.54	-228.25	200.04	-2.60	**-50.34**	**-22.19**	**-28.15**
NPC317147	-34.78	-56.44	57.49	2.68	**-36.42**	**-20.88**	**-15.54**

2) For the RNA-pocket leads

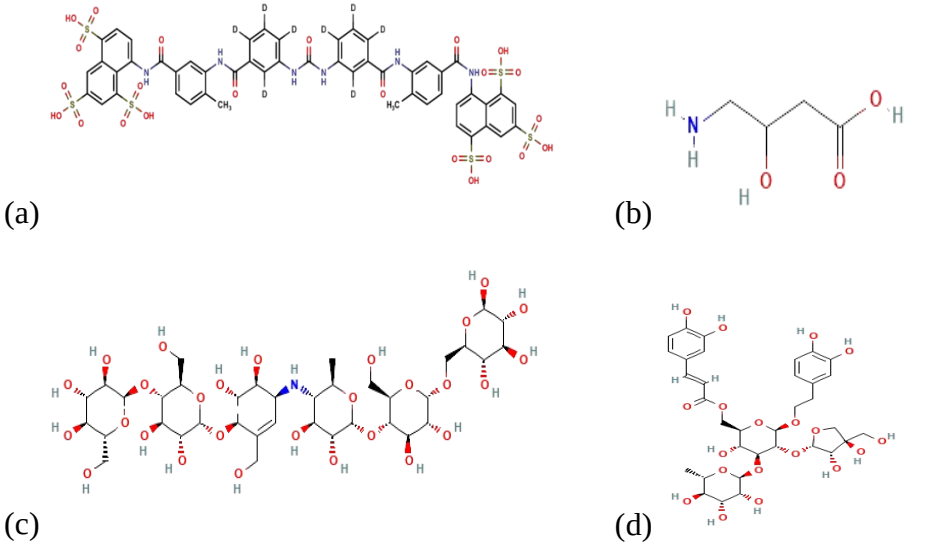
Table 9: Hit molecules for the RNA-pocket and their respective Binding free energy values. The contribution from various interactions is also depicted. All values are in kcal/mol.

Ligand	ΔG_{vdW}	ΔG_{elec}	ΔG_{polar}	ΔG_{np}	ΔH	$T\Delta S$	ΔG_{bind}
Suramin	-107.5	-68.14	150.04	-8.77	**-34.40**	**-34.68**	**0.28**
HMDB0029794	-66.89	-91.69	119.13	-8.14	**-47.49**	**-42.38**	**-5.11**
HMDB0006602	-75.70	-128.79	167.97	-8.56	**-45.09**	**-42.37**	**-2.72**
DB03971	-66.65	-134.96	158.16	-7.44	**-50.90**	**-40.77**	**-10.13**
NPC141455	-78.39	-15.09	50.44	-6.29	**-49.33**	**-28.73**	**-20.6**
DB03208	-55.28	-118.39	138.61	-6.64	**-41.70**	**-40.38**	**-1.32**

After accounting for the entropy contributions towards the binding free energy, it was observed that few of the leads which were earlier showing good enthalpy results were not showing optimal overall binding free energy values. Therefore, we have chosen to proceed with ligands with higher binding free energy values viz. **NPC327542** for the **ATP-binding pocket** and **DB03971** and **NPC141455** for the **RNA-binding pocket.**

Subsequent to this, we have subjected the chosen compounds to extended MD simulation of 1 μs duration. (The runs were performed in triplicates, making the total simulation length to be 1 μs × 3 = 3 μs).

The extended simulations were then again subjected to extensive structural analysis, the results for which have been provided below:



(a) (b)

(c) (d)

Fig.25: (a) 2D Structure of Suramin (Control drug), (b) 2D structure of NPC141455, (c) 2D structure of DB03971 and (d) 2D structure of NPC327542

5.3.1 Structural and flexibility analysis: RMSD

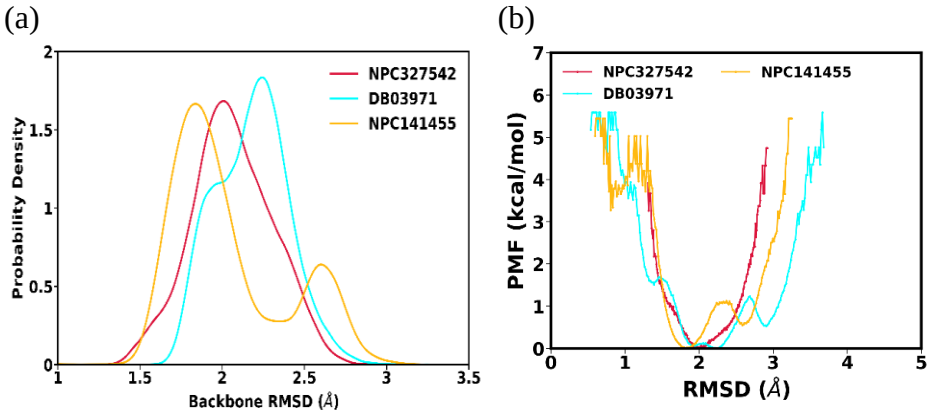


Fig.26: (a) The backbone RMSD analysis of the complexes with the (b) associated PMF plot.

From the kernel distribution plot of the backbone RMSD of the three complexes, it is seen that all of them show prominent peaks under the range of 2.5 Å RMSD. These results assert the stability of the protein in bound state with the three ligands as compared to free (apo) state. Corresponding to the KDE plot, we have also plotted the Potential Mean Force graph of the backbone RMSD. PMF quantifies the free

energy changes associated with transitions between different states of a system. Herein, it is observed in the PMF plot that NPC327542 shows minima at 2 Å RMSD while in the cases of other two ligands, shoulder (side) peaks have also been observed. However, the energy difference between the two is very less.

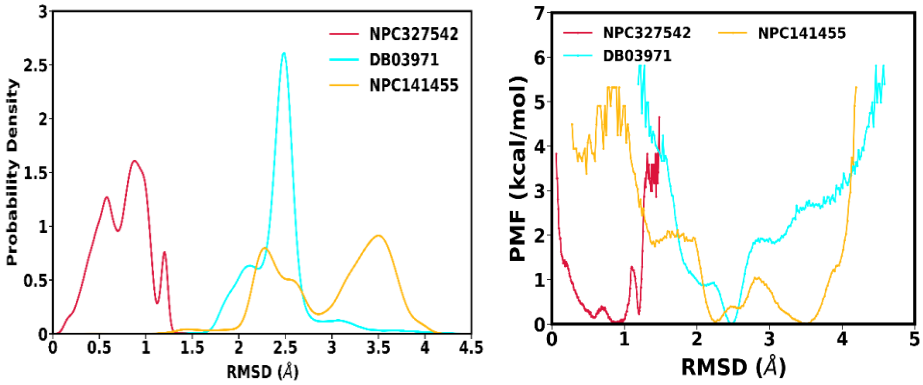


Fig.27: RMSD analysis of the ligands and the associated PMF plot.

Ligand DB03971 shows single peak at 2.5 Å while the other two are observed to have scattered peaks over the range of 0-4 Å. This can be accounted for by the small size of the ligand molecules. In the associated PMF plot, it is observed that DB03971 is showing single peak at 2.5 Å, which is in agreement to the RMSD values. The ligand NPC141455 is showing two minima at 2.2 and 3.5 Å respectively while in the case of NPC327542, peak region is observed at 1 Å.

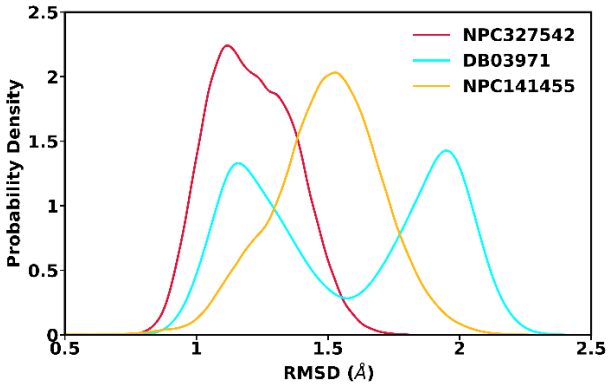


Fig.28: RMSD analysis of the Binding pocket of the three complexes.

The time evolution plot of the RMSD values of all the complexes show stable values, all under the range of 2.4 Å. In the case of only DB03971, two comparable peaks are seen, one at 1.2 Å and the other at 2 Å RMSD.

5.3.2 Ligand-Receptor distance analysis

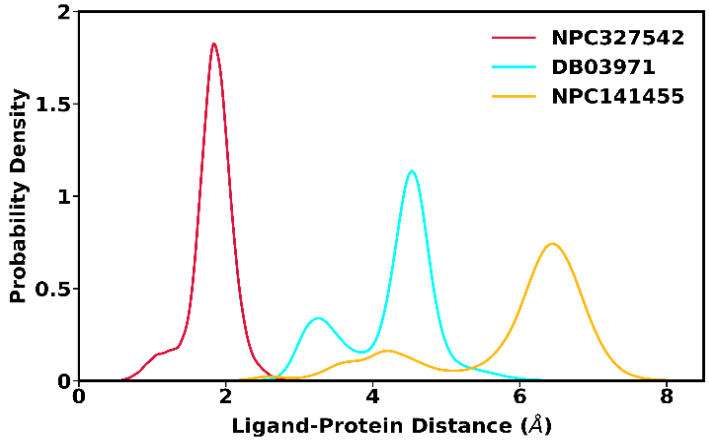


Fig.29: Analysis of distance between the ligand and the protein over the course of simulation.

Upon observing the distance between the ligand and the protein over the course of simulation, it is inferred that ligand NPC327542 shows a stable distance of 2 Å, while in DB03971, prominent peak is seen at 4.3 Å and a shoulder peak is also observed. In the case of NPC141455 as well, we have observed that the distance between the ligand and the protein over the course of simulation has been majorly 6.5 Å while a minor peak at 2.8 Å is also seen.

5.3.3 Protein compactness and solvent accessibility analysis

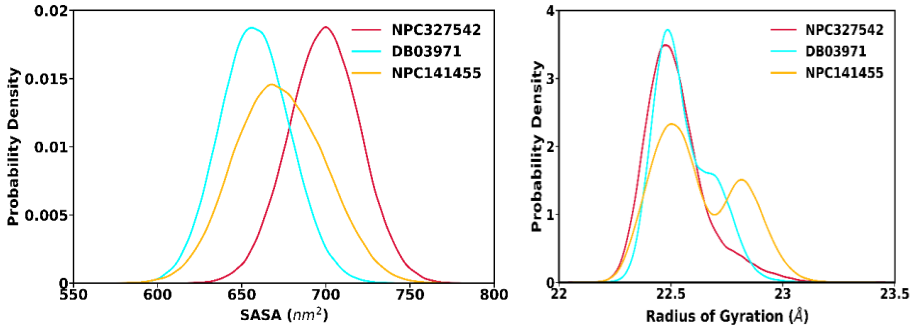


Fig.30: Plots depicting the Solvent-accessible surface area (SASA) and the radius of Gyration values of the three complexes over the course of simulation.

Upon inspection of the SASA values, it is observed that all the three complexes show prominent peaks as per follows: NPC327542 at 710 nm^2, DB03971 at 654 nm^2 and NPC141455 at 660 nm^2. The SASA value for the ATP-pocket lead i.e. NPC327542 is seen to be greater than the other two RNA-pocket leads. This is because of the peripheral location of the ATP pocket making it more accessible to outer solvent environment while in the case of the RNA pocket, central location surrounded by residues reduces the solvent-accessible surface area.

From the Radius of Gyration plot, we see that NPC327542 and DB03971 show prominent peaks at 22.5 Å while the third ligand NPC141455 shows one major peak at 22.5 Å and a shoulder peak at 22.7 Å. The radius of gyration values gives us an indication about the compactness of the protein over the course of simulation and we can infer that all the three complexes show a comparable value of 22.5 Å, with a slight deviation in the case of NPC141455.

5.3.4 Residual Stability analysis

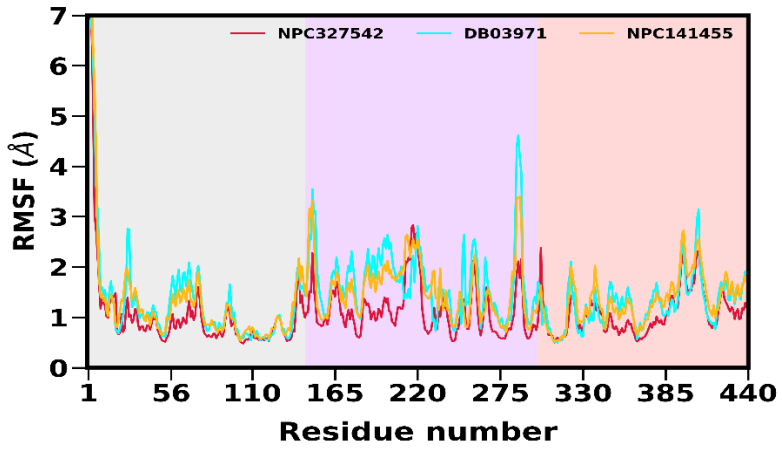


Fig.31: Plot depicting the per residue flexibility of the three complexes over the course of MD simulation.

From the RMSF plot, it is observed that NPC327542 is showing the least residual flexibility followed by NPC141455 and DB03971 respectively. Overall, domain II is shown to have the highest number of flexible residues. This may be accounted for by the high number of loop regions present in the domain II.

5.3.5 Binding Free Energy Analysis

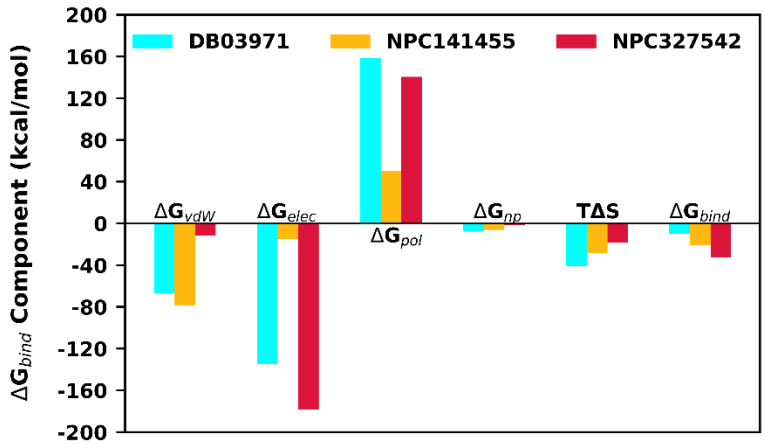


Fig.32: The binding free energy of the three hit molecules along with the contribution of respective forces.

In the case of NPC327542, it is observed that electrostatic interactions contribute significantly towards the binding of the ligand to the receptor. In complex DB03971 as well, electrostatic interactions play the major towards formation of the complex. In NPC141455, we observe that the major contribution towards complex formation is from van der waals forces. In all the three complexes, we see that polar interactions are opposing the binding. The entropy contributions are the highest for DB03971, followed by NPC14155 and NPC327542 respectively, which can be accounted for by the size of the respective molecules as entropy values tend to increase with increase in size of the molecules due to higher degrees of freedom and flexibility.

5.3.6 Per residue energy decomposition

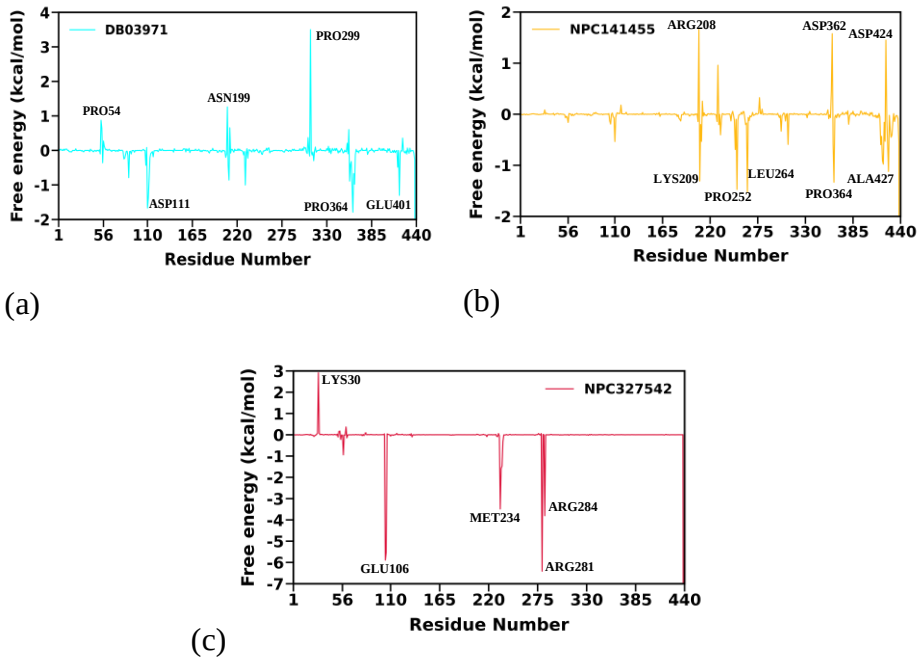


Fig.33: The per residue energy decomposition for the three complexes highlighting major contributing residues.

The per residue binding free energy decomposition indicates about the individual contribution of residues towards the binding of the complexes. In all the three cases, we can observe residues which show

negative free energy implying that they are contributing positively towards the binding while other residues which show positive free energy values oppose the binding of the complex.

DB03971: Asp111, Pro364 and Glu401 contribute positively towards binding while Pro54, Asn199 and Pro299 are shown to oppose complex formation.

NPC141455: Lys209, Pro252, Leu264, Pro364 and Ala427 contribute positively towards formation of complex and Arg208, Asp362 and Asp424 oppose the binding.

NPC327542: In this case, Glu106, Met234, Arg281 and Arg284 are favoring binding while Lys30 residue is opposing it.

The per residue energy contribution to complex formation values comes handy wherein we are studying cases with mutations. If it is observed that mutation has been caused in one of the prominent residues, we may encounter changes in complex formation or a slightly reduced affinity.

5.3.7 Ligand Interaction studies

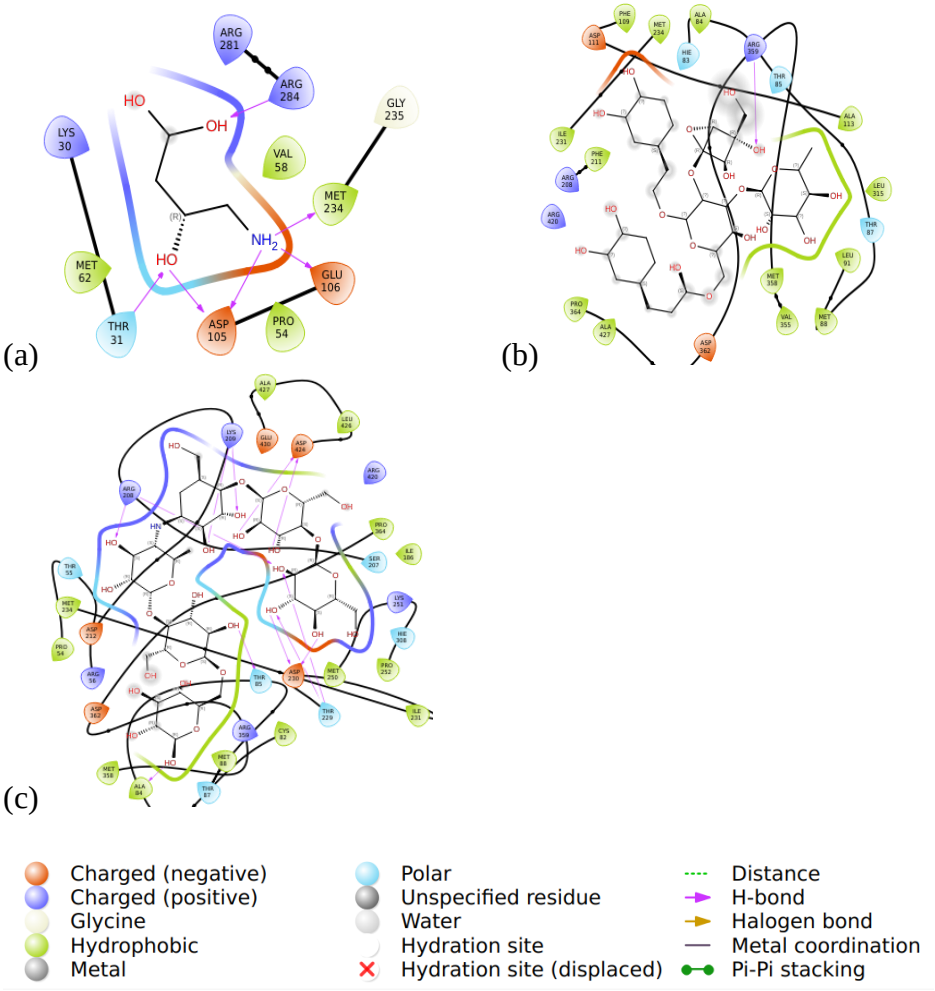


Fig.34: 2D Maestro maps showing the types of interactions between the ligand and the receptor for the three complexes. The symbols and representations of the various types of interactions have been provided in the legend.

In the case of NPC327542, we observe from the ligand interaction diagram that the major interactions formed are Hydrogen bonds. Due to the specificity and strength of hydrogen bonds, they impart greater stability to the complex and consequently a high binding affinity. In the case of NPC141455, it has been investigated that major interactions are van der waals and hydrophobic forces. In DB03971, significant number of hydrogen bond interactions are seen from the ligand

interaction 2D Maestro map. The determination of the type of interactions responsible for binding help us assess the dynamics of the formation of the complex and provides a scope for lead optimisation as well.

5.3.8 Hydrogen Bond Analysis

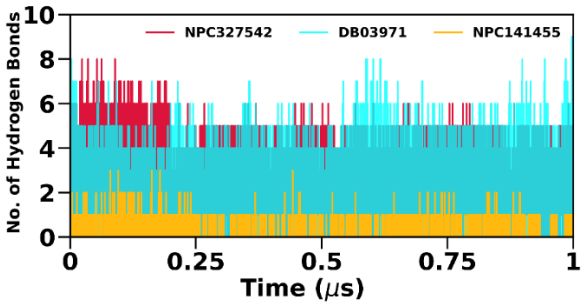


Fig.35: Plot representing the number of hydrogen bonds formed in the three complexes over the course of simulation.

Table 10: Table showing the hydrogen bond interaction of complex DB03971 mentioning the donor, acceptor, distance (in Å) and occupancy (in %).

Binding couples		Molecular dynamics	
Acceptor	**Donor**	**Distance (Å)**	**Occupancy (%)**
Lig_440@021	THR_229@H1	2.81	35.91
Lig_440@O26	ARG_420@H11	2.89	21.66
ARG_208@O	Lig_440@H47	2.73	46.11
ASP_230@O1	Lig_440@H56	2.64	40.70
ASP_230@O1	Lig_440@H58	2.66	39.31
LEU_363@O	Lig_440@H42	2.78	34.67
ASP_111@O2	Lig_440@H57	2.66	34.26
ASP_111@O2	Lig_440@H63	2.71	33.47

Table 11: Table showing the hydrogen bond interaction of complex NPC327542 mentioning the donor, acceptor, distance (in Å) and occupancy (in %).

Binding couples		Molecular dynamics	
Acceptor	**Donor**	**Distance (Å)**	**Occupancy (%)**
Lig_440@01	THR_31@H1	2.74	61.50
Lig_440@O2	ARG_284@H12	2.84	43.35
Lig_440@O3	ARG_284@H12	2.84	39.76
Lig_440@O2	ARG_284@H22	2.83	28.69
ASP_105@O2	Lig_440@H9	2.62	77.63
GLU_106@O2	Lig_440@H8	2.67	41.64
ASP_105@O2	Lig_440@H6	2.69	41.23
MET_234@O	Lig_440@H7	2.76	37.16
ASP_105@O2	Lig_440@H7	2.69	28.24

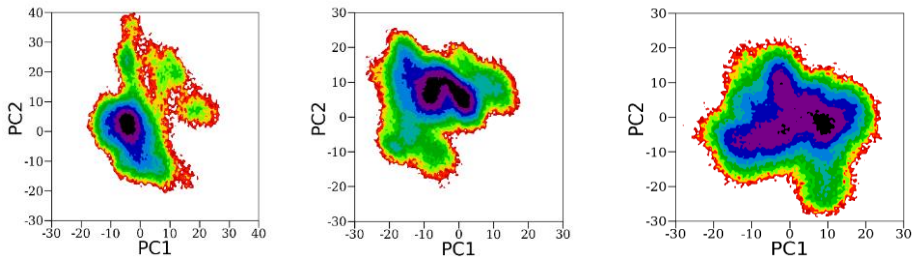
Table 12: Table showing the hydrogen bond interaction of complex NPC141455 mentioning the donor, acceptor, distance (in Å) and occupancy (in %).

Binding couples		Molecular dynamics	
Acceptor	**Donor**	**Distance (Å)**	**Occupancy (%)**
Lig_440@O6	ARG_359@H11	2.89	0.24
Lig_440@O5	THR_85@H1	2.89	0.21

Hydrogen bonds, due to their specificity and strong interaction strength, significantly influence the binding of a biological protein-

ligand complex. It is hence very significant to study the contribution of hydrogen bonds. Here, we have analysed the number of hydrogen bonds formed in our three complexes over the complete duration of the simulation run. It has been observed that NPC141455 shows the least number of hydrogen bond interactions. This is because the major interactions for this complex is van der waals and other non-polar interactions. In the case of DB03971, we observe a good number of hydrogen bonds over time, ranging from 2 to 8. In complex NPC327542, the average number of hydrogen bonds formed over the course of simulation ranges from 4 to 8. We have also investigated the actual donor and acceptor residues forming the hydrogen bonds in case of all the complexes, the details for which are provided in the tables. Asp, Leu, Glu and Met are the major residues involved in formation of hydrogen bonds, the other significant participant being the ligands.

5.3.9 Principal Component Analysis (PCA)



NPC327542 (ATP-pocket) | DB03971 (RNA-pocket) | NPC141455 (RNA-pocket)

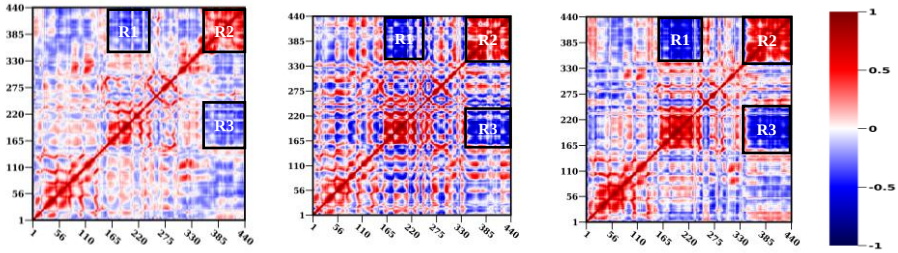
Fig.36: The PCA analysis for the three complexes.

The Principal Component Analysis or PCA is a great tool for analyzing and reducing the complexity of multi-dimensional data. PCA involves computing the covariance matrix of the aligned trajectory data. The covariance matrix describes the correlations between different pairs of atomic positions or structural variables over the course of the simulation. The covariance matrix is symmetric and positive semi-definite. The next step is to perform eigenvalue decomposition (or singular value decomposition) of the covariance matrix. This yields a set of eigenvectors and eigenvalues. The eigenvectors represent the principal components (PCs) or modes of motion in the system, while

the corresponding eigenvalues indicate the magnitude of the variance along each PC. The eigenvectors obtained from PCA can be used to project the original trajectory data onto a reduced-dimensional space spanned by the principal components. This allows for the visualization and analysis of the dominant modes of motion in the system. This facilitates the visualization and analysis of complex molecular dynamics by focusing on the most relevant collective motions.

In our PCA analysis for the three complexes, we observe that complex NPC327542 shows a prominent minimum. This point corresponds to the most stable conformation of the complex over the course of the whole simulation. In the case of DB03971, we observe minima point which is a bit scattered meaning that the stable conformations span over a range of free energy. In NPC141455, single minima are observed, however, another very small region of minimum free energy is also seen. However, the energy barrier between the two is negligible.

5.3.10 Dynamic Cross-Correlation Matrix (DCCM)



NPC327542 (ATP-pocket) DB03971 (RNA-pocket) NPC141455 (RNA-pocket)

Fig.37: Plots depicting the Dynamic cross-correlation matrix for the three systems.

The DCCM is a method used to analyze the correlated motions of atoms or groups of atoms within a molecular system over time. The DCCM provides valuable insights into the correlated motions of atoms within the molecular system. Positive values in the DCCM indicate correlated motions, suggesting that the fluctuations of the two atoms are positively correlated over time. Conversely, negative values

indicate anti-correlated motions, where the fluctuations of the two atoms are negatively correlated. In the case of our three ligands, we have highlighted the three regions wherein significant correlation is observed highlighted as R1, R2 and R3 respectively. The red and blue lines correspond to correlated and anti-correlated motions, respectively.

The region R1 is showing anti-correlation which is the least in NPC327542, followed by DB03971 and highest in NPC141455. The region corresponds to residues of the Domain II and this anti-correlation between the residues can be attributed to the significant number loop regions present in the concerned domain.

The region R2 depicts positive correlation among the three systems and coincides with Domain III residues 330 to 440 of the protein complex. This domain interacts with both the pockets: ATP-binding as well as RNA-binding and is hence showing positively correlated motions. Structurally as well, the domain comprises mostly of alpha-helix and beta-sheets with very less number of loop regions imparting significant stability.

The region R3 also coincides with Domain II (residues 150-230) and Domain III (residues 330-440). Herein we are observing anti-correlation motion between the three systems. This can be accounted for the by interactions of the two domains and the coinciding of the two binding pockets over the course of simulation.

5.3.11 Network Analysis

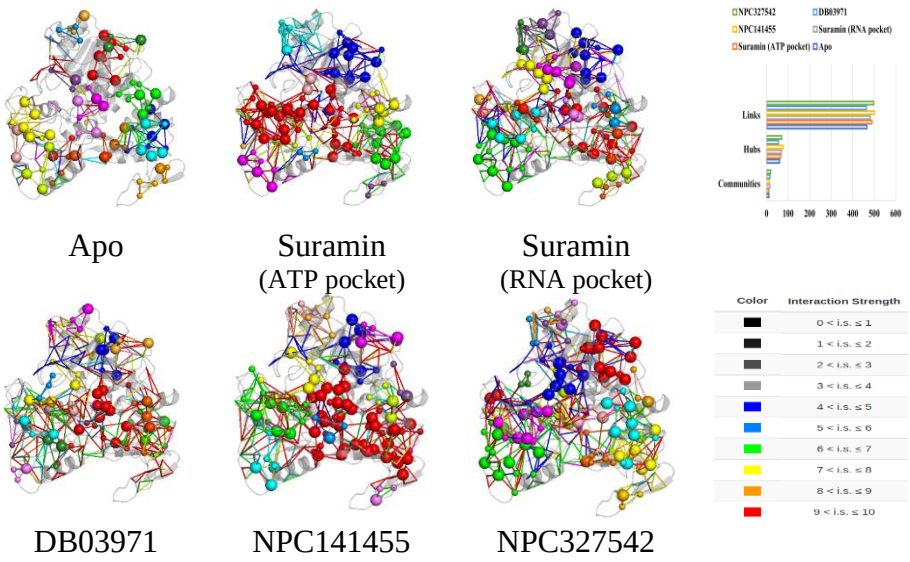


Fig.38: Network analysis for apo, control and the three complexes depicting hubs, links and communities.

Network analysis in molecular dynamics (MD) simulation involves the characterization and study of molecular systems as networks of interconnected nodes (atoms, residues, or functional groups) linked by edges (interactions or relationships). In our results for Network analysis, we have compared the three systems with apo structure (unbound) as well as control Suramin for both the pockets. It is observed that in comparison to the apo and the control structures, the three systems are showing higher numbers of hubs, links as well as communities. This depicts the enhanced stability of the structure NS3 Helicase in the bound state with the three ligands in comparison to a free state.

Chapter 6

Conclusion and Scope of Future work

In the present study, extensive computational approaches have been employed in order to predict potential compounds which can act as inhibitors against the NS3 Helicase protein of the Dengue virus. Three novel compounds have been discovered, two of them are of natural origin, derived from the Natural Product Activity and Species Source database [NPC327542: Gamma-Amino-Beta-Hydroxybutyric Acid and NPC141455: Luteoside B (It is a phenylpropanoid glycoside from the medicinal plant *Markhamia lutea*)], while the third compound is an existing FDA-approved drug from DrugBank [DB03971: Acarbose Derived Hexasaccharide (Acarbose is a complex oligosaccharide that acts as a competitive, reversible inhibitor of pancreatic alpha-amylase and membrane-bound intestinal alpha-glucoside hydrolase. It is used for the treatment of Diabetes mellitus.)] that can be repurposed against dengue. All the three compounds have undergone extensive in-silico testing and experimentation and we have obtained positive results that showcase their potent activity against the dengue virus. For further improvement in the potency of the compounds, lead optimization can be performed which involves modifying and introducing changes in the structure of the compound in order to enhance its potency against the target. Several computational techniques exist which can be employed to realise the objective. Also, in-silico approaches represent only the first step in the long road to the discovery of a drug. Hence, although the three novel compounds have performed well in computational studies, it is equally important to perform in-vivo experimentation on the discovered compounds to further establish their authenticity as dengue inhibitors and moving one step ahead in the global war of humanity against the dengue infection.

REFERENCES

1. Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases, Division of Vector-Borne Diseases
2. World Health Organization. Disease Outbreak News; Dengue – Global situation
3. Discovery of 5th serotype of dengue virus: A new public health dilemma in dengue control; Lt Col M.S. Mustafa et. Al; *Medical Journal Armed Forces India*; 2015.
4. Current perspectives on the spread of dengue in India; Ekta Gupta &Neha Ballani; *Infection and Drug Resistance*; 2015
5. Chew, M.F., Poh, K.S., Poh, C.L. Peptides as Therapeutic Agents for Dengue Virus. *International Journal of Medical Sciences*, 2017.
6. Dhiman, Mamta & Sharma, Lakshika & Dadhich, Abhishek & Dhawan, Poonam & Sharma, M. (2022). Traditional Knowledge to Contemporary Medication in the Treatment of Infectious Disease Dengue: A Review. *Frontiers in Pharmacology*, 2022
7. Chee H Liew; The first case of sexual transmission of dengue in Spain; *Journal of Travel Medicine*; 2020
8. Nanaware N et. Al; Dengue Virus Infection: A Tale of Viral Exploitations and Host Responses. *Viruses*. 2021
9. Nguyen NM et. Al; A randomized, double-blind placebo-controlled trial of balapiravir, a polymerase inhibitor, in adult dengue patients. *J Infect Dis*. 2013
10. Safety, Tolerability, and Pharmacokinetics of JNJ-1802, a Pan-serotype Dengue Direct Antiviral Small Molecule, in a Phase 1, Double-Blind, Randomized, Dose-Escalation Study in Healthy Volunteers; Oliver Ackaert et. Al; *Clinical Infectious Diseases*

11. Tully D, Griffiths CL. Dengvaxia: the world's first vaccine for prevention of secondary dengue. *Ther Adv Vaccines Immunother*. 2021
12. Qdenga® - A promising dengue fever vaccine; can it be recommended to non-immune travellers? Martin Angelin et.al, *Travel Medicine and Infectious Disease*.
13. Halstead SB. Dengue Antibody-Dependent Enhancement: Knowns and Unknowns. *Microbiol Spectr*. 2014
14. Structure of the Dengue Virus Helicase/Nucleoside Triphosphatase Catalytic Domain at a Resolution of 2.4 Å; Ting Xu et. al; ASM Journals, *Journal of Virology*, 2004
15. Tsaioun, K., Bottlaender, M., Mabondzo, A. *et al*. ADDME – Avoiding Drug Development Mistakes Early: central nervous system drug discovery perspective. *BMC Neurol* (Suppl 1), S1, 2009.
16. Honarparvar, B.; Govender, T.; Maguire, G.E.; Soliman, M.E.; Kruger, H.G. Integrated approach to structure-based enzymatic drug design: Molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev*. 2014, 114, 493–537.
17. Christopher A. Lipinski, Drug-like properties and the causes of poor solubility and poor permeability, *Journal of Pharmacological and Toxicological Methods*, 2000.
18. Xu TSampath AChao A, Wen D, Nanao M, Chene P, Vasudevan SG, Lescar J.Structure of the Dengue Virus Helicase/Nucleoside Triphosphatase Catalytic Domain at a Resolution of 2.4 Å. *J Virol*, 2005
19. UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, et. al, *J Comput Chem.* 2004
20. Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics.* 2003
21. PubChem [Internet]. Bethesda (MD): National Library of Medicine (US), National Centre for Biotechnology Information; 2004. PubChem Compound Summary for CID 8514, Suramin sodium.

22. Wishart DS, Tzur D, Knox C, et al., HMDB: the Human Metabolome Database. *Nucleic Acids Res*. 2007.
23. Hui Zhao, Yuan Yang, Shuaiqi Wang, et al. NPASS database update 2023: quantitative natural product activity and species source database for biomedical research. *Nucleic Acids Research*, 2022.
24. Wishart DS, Feunang YD, Guo AC, et. al, DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2017.
25. Halgren, T. A.; Murphy, R. B.; Friesner, R. A. et. al, "Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening", *J. Med. Chem*., 2004
26. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins; Edward Harder, et. al., J*ournal of Chemical Theory and Computation* 2016
27. Xiong G, Wu Z, et. al, ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 2021.
28. 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* 7, 2017.
29. Banerjee P, Eckert AO, Schrey AK, Preissner R. ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res*. 2018.
30. Case, D. A. et. al; AMBER 2018 Univ. Calif. San Franc. 2018
31. A modified TIP3P water potential for simulation with Ewald summation; Price, Daniel J.; Brooks, Charles L., III; *Journal of Chemical Physics* 2004
32. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB; Maier et. al; *Journal of Chemical Theory and Computation*, 2015
33. Development and testing of a general Amber force field; Wang, Junmei et. al; *Journal of Computational Chemistry* 2004

34. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations; Krautler, Vincent et. al; *Journal of Computational Chemistry* 2001
35. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide; Loncharich, Richard J. et. al; *Biopolymers* 1992.
36. Molecular dynamics with coupling to an external bath: Berendsen, H. J. C. et. al; *Journal of Chemical Physics* 1984.
37. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data; Roe, Daniel R. et. al; *Journal of Chemical Theory and Computation* 2013.
38. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models; Kollman et. al; *Accounts of Chemical Research* 2000
39. *MMPBSA.py*: An Efficient Program for End-State Free Energy Calculations Bill R. Miller III, T. Dwight McGee Jr., Jason M. Swails, Nadine Homeyer, Holger Gohlke, and Adrian E. Roitberg *Journal of Chemical Theory and Computation* 2012
40. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016
41. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? Wanli You, Zhiye Tang, and Chia-en A. Chang *Journal of Chemical Theory and Computation* 2019
42. Seeber M, Felline A, Raimondi F, Mariani S, Fanelli F webPSN: a web server for high-throughput investigation of structural communication in biomacromolecules. *Bioinformatics*, 2015