COMPUTER AIDED DRUG DESIGNING FOR DENGUE NS5 METHYLTRANSFERASE

M.Tech. Thesis

By

PRADEEP KUMAR

DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2025

COMPUTER AIDED DRUG DESIGNING FOR DENGUE NS5 METHYLTRANSFERASE

A THESIS

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

by

PRADEEP KUMAR

DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
MAY 2025

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 NS5 METHYLTRANSFERASE** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINERRING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from JULY, 2024 to MAY 2025 under the supervision of **DR PARIMAL KAR**.

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.

05/05/2025

Signature of the student with date
(PRADEEP KUMAR)

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

Signature of the Supervisor of M.Tech. thesis (with date) 05.05.2025

(DR PARIMAL KAR)

PRADEEP KUMAR has successfully given his M.Tech. Oral Examination held on **5 MAY 2025**.

Signature of Supervisor of M.Tech. thesis
Date: 05.05.2025

P.V. Kodgire
Convener, DPGC
Date: 05/05/2025

ACKNOWLEDGEMENTS

The first steps into the world of scientific research, wobbly and unsteady and frequently tripping over. The steps can not be taken in solitude, without any helping hands extended towards you. It is my great fortune, having been surrounded by great people who came to my assist during the times of struggles, and helped me take the steps slow and steady and taught me how to walk.

First, I would like to extend my deepest gratitude to my project supervisor **Dr. Parimal Kar** for his continuous support and guidance throughout this journey. It was his courses that properly introduced me to the field of computational biophysics. From then till now, his continuous guidance and helpful advice have helped me immensely.

My warmest thanks to **Ms. Subhasmita Mahapatra**, a great mentor and friend, who not only taught me various aspects of the computational field but also showed me how to handle a research work properly.**. Sayan Poddar, Mr. Suman Koirala, Mr. Kapil D. Ursal, Mr. Sunanda Samanta, Ms. Dibyanka Dalai and Ms. Ahana Chakraborty** were always there in times of struggles and lent their help unconditionally. It has been a wonderful experience to be a part of the computational biophysics group.

The acknowledgements would not be complete without thanking my family, my parents **Mr. Satyabhan Singh**, **Mrs. Saroj** my sister **Ms. Pragati** and my brother **Mr. Avesh kumar** Constant source of inspiration, love and encouragement, they made this journey possible.

Finally, I thank **Indian Institute of Technology Indore** for providing me this valuable learning opportunity.

---PRADEEP KUMAR

Dedicated to my parents

ABSTRACT

Dengue fever, caused by the dengue fever virus, poses a significant public health threat worldwide. A crucial enzyme in the viral replication process is the NS5 methyltransferase (MTase). This enzyme functions to limit viral RNA and protect it from the host's immune response. Targeting NS5 MTase presents a promising strategy for developing treatments for dengue virus disease. Computer-aided drug discovery has emerged as an effective method to identify potential inhibitors of NS5 MTase by leveraging computational techniques like molecular docking, virtual screening, and molecular dynamics simulations. These approaches enable the identification of the enzyme's active site and small molecules that can specifically bind to and inhibit its functioning, thereby impeding virus replication. In this study, we explore the application of CADD to identify and optimize potential inhibitors of NS5 MTase, providing insights into their binding affinity, stability, and mechanism of action. The efficacy of CADD in accelerating the discovery of novel antiviral compounds highlights its potential as a cost-effective and time-efficient strategy for dengue drug development. Further experimental investigation and optimization of these inhibitors may facilitate the development of effective treatments to prevent dengue virus infection.

Virtual screening of the NPASS database identified three promising lead molecules with potential therapeutic relevance. These molecules underwent comprehensive **ADMET analysis** to evaluate their pharmacokinetic and toxicity profiles, followed by **molecular dynamics (MD) simulations** to assess their stability and interactions within the target binding site. The results of these analyses were systematically compared with **Sinefungin**, the control molecule in the **computer-aided drug design (CADD)** process. Comparative insights from MD simulations, binding affinities, and structural dynamics highlighted the potential of the identified leads as viable candidates for further drug development. The findings contribute to a deeper understanding of ligand-protein interactions, reinforcing the role of computational methodologies in accelerating drug discovery.

Contents

List of Figures

List of Tables

Abbreviations

TIP3P	Transferable Intermolecular Potential with 3 points
VS	Virtual Screening
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
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
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

Chapter 1

Introduction

Dengue fever, a significant public health issue in India, is attributed to the dengue virus (DENV), which comprises four distinct serotypes: **DENV-1, DENV-2, DENV-3, and DENV-4**. The predominance of these serotypes are influenced by climatic factors. Notably, Denv-2 and Denv-3 have exhibited varying levels of circulation during outbreaks, with Denv-2 accounting for approximately **66.40% of cases in North India** during specific outbreaks, whereas Denv-3 became more prevalent in subsequent years, particularly in 2020 and 2021, when it constituted 72.08% of cases [1]. The dynamic nature of DENV serotype distribution in India presents challenges for effective monitoring and public health intervention. Over time, the emergence of diverse serotypes has heightened concerns regarding severe dengue manifestations and the potential for increased transmission rates, especially in densely populated and urbanized areas. Research has not only identified geographical disparities in serotype circulation but has also highlighted the underlying environmental and social determinants that elevate the risk of dengue outbreaks, such as housing conditions, hygiene practices, and climate change [2][3]. To develop effective dengue control strategies, it is crucial to comprehend these serotype fluctuations. Public health officials have emphasized the necessity for continuous monitoring and adaptive responses to manage the ongoing risk of dengue outbreaks [2][4]. Furthermore, the evolution of DENV serotypes and their respective genotypes in India underscores

the need for robust monitoring systems and targeted interventions to mitigate the impact of this mosquito-borne disease on the affected population.

Dengue fever is a transmitted by the bite of ***Aedes aegypti*** mosquito and ***Aedes albopictus***. Globalization and climate change are likely to expand the distribution of the *Aedes aegypti* mosquito, potentially increasing health risks. Dengue fever infections worldwide are estimated at half a million per year, leading to serious symptoms [5].

At present, numerous antiviral drugs are under investigation in clinical trials. However, many have proven unsuccessful due to the four serotypes of dengue fever, which are responsible for toxicity concerns, limited duration of treatment, and efficacy issues. Only one Sanofi **Dengvaxia vaccine** has been associated with adverse effects in children and may exacerbate symptoms in individuals who have not previously contracted dengue fever. Consequently, there is an urgent need for effective antiviral treatment.

FIGURE 1.1: Dengue mosquito.

1.1 Structural analysis of NS5 methyltransferase

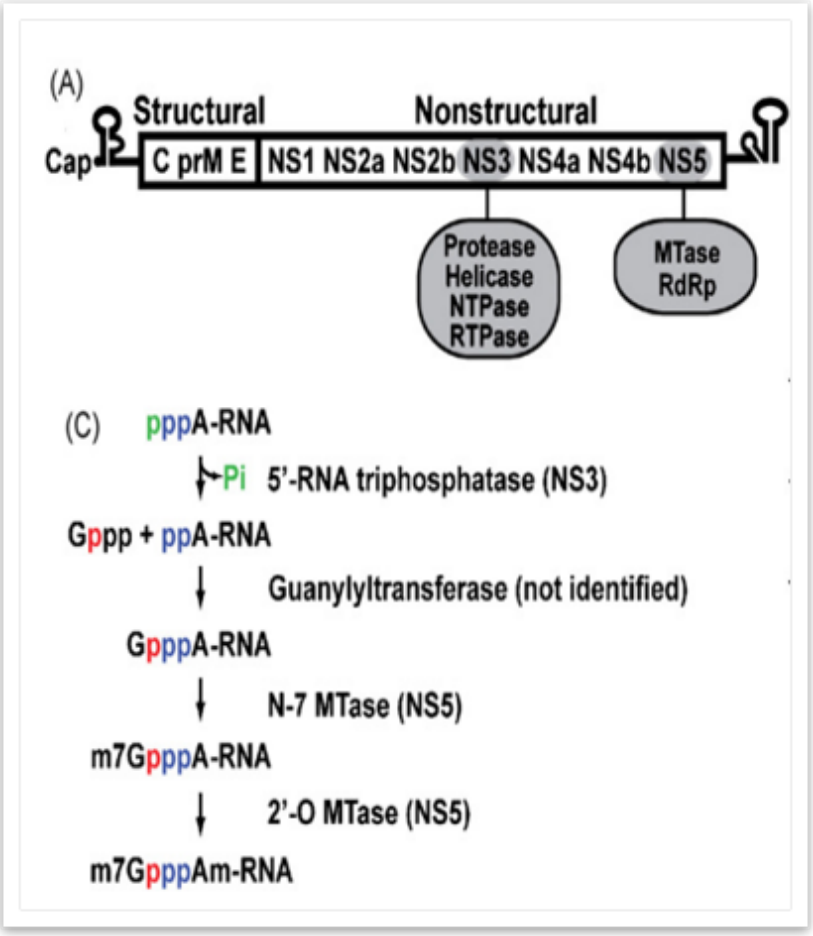

FIGURE 1.2: Dengue virus RNA structure.

The NS5 protein of dengue fever virus (DENV) is a multifunctional enzyme that consists of two main functional domains: RDRP domain and the methyltransferase (MTase) domain. A methyl group is added to the 5' cap of viral RNA for RNA stability, translation, and host immune response. And the most important thing is theft. The enzyme steps catalyse two key reactions: the addition of a guanosine cap (m7'G) to the viral RNA and the subsequent methylation of the 7'N position of the guanine base [6]. These activities are required for DENV replication. This makes NS5 MTase an attractive target for antiviral drug development [7]. Understanding the structural features and mechanism of action of NS5

MTase in DENV is critical for the development of new antiviral treatments aimed at inhibiting its activity. This report presents an in-depth structural analysis of DENV NS5 MTase, including its catalytic mechanism and interaction with potential inhibitors, using high concentrations of inhibitors.

The structure of DENV NS5 MTase reveals a highly conserved SAM-dependent methyltransferase cluster [7]. It provides a stable structure for the enzyme to function. The main structural features identified in DENV NS5 MTase include:

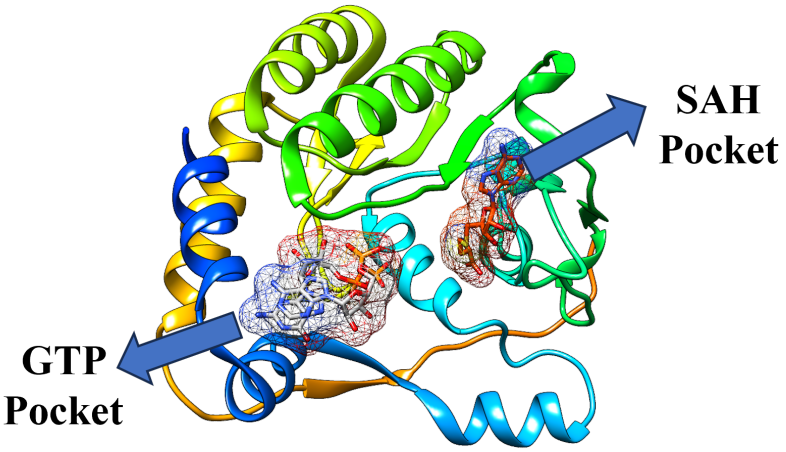

FIGURE 1.3: NS5 Methyltransferase Protein structure PDB ID (2P3L).

The SAM-binding pocket: It consists of several conserved residues, including Phe241, Tyr211, and Glu236. These residues facilitate the interaction between the adenine ring of SAM and the enzyme while the Methionine is transferring a methyl group to the RNA substrate. The SAM molecule is tightly bound to a hydrophobic vesicle surrounded by stabilizing polar residues [8].

RNA Binding Site: The RNA substrate binding site is adjacent to the binding pocket of the SAM. The enzyme specifically recognizes the 5' cap structure of viral RNA. Important residues such as Asp307 is crucial for the coordination of the guanine bases of the RNA precursor to ensure

proper alignment for methylation. The RNA-binding region undergoes a conformational change upon binding to the substrate.

Catalytic Mechanism: The catalytic cycle of NS5 MTase follows a mechanism that is dependent on the methyl group from SAM being transferred to the 7' N position of the guanine base in the RNA cap structure. This methylation is facilitated by the interaction of key residues, such as Asp307, and activation of water molecules, which initiates the transfer. Conformational changes during the reaction cycle are important for efficient methyl group transfer [6]. Molecular dynamics simulations show that NS5 MTase undergoes significant structural rearrangements during the methylation reaction. The enzyme switches between closed conformations (when bound to SAM and RNA substrates) and open conformation (In the absence of a substrate), a closed framework is essential for proper orientation of the substrate for methyl transfer [9]. An open framework is required for post-reaction separation of substrates. Enzyme flexibility is particularly evident in the RNA-binding domain, where loops and helices adapt to fit the RNA substrate [7]. These dynamic changes are important for structure recognition and processing the 5'-cap of RNA.

1.2 Role of methyltransferase

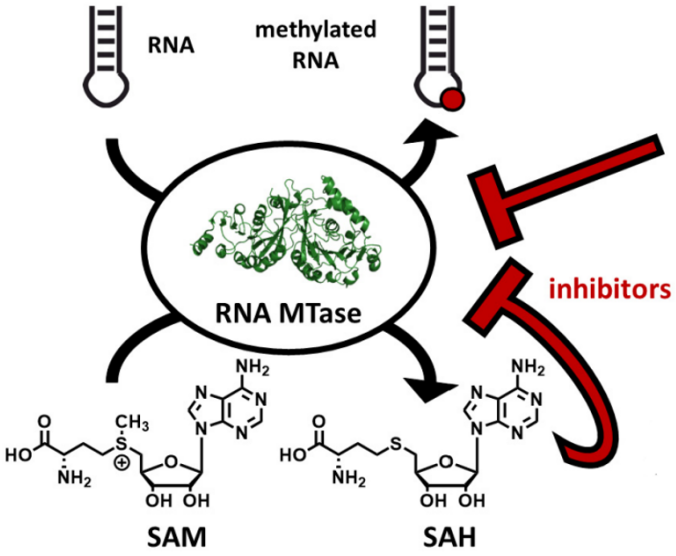

FIGURE 1.4: Working mechanism of NS5 methyltransferase.

RNA Capping: The main function of NS5 MTase is to add methyl groups to the 5' cap structure of viral RNA, an important modification for mRNA stability and efficient translation. The enzyme catalyses two methylation reactions: N7-methylation: Addition of a methyl group to the 7th nitrogen of the RNA cap structure's guanine base (m7G). 2'-O-methylation: The addition of a methyl group to the 2' hydroxyl group of the ribose sugar in the first nucleotide of RNA transcript, stabilizes the RNA and increases translation efficiency.

Evasion of Host Immune System: By adding a cap structure to its RNA, DENV can mimic host mRNA. Most importantly, he avoids host's innate immune response. Therefore, efficient replication occurs in the absence of antiviral pathways.

Substrate Specificity and Catalysis: The NS5 MTase domain binds S-adenosylmethionine (SAM) as a cofactor, which acts as a methyl donor in the given maximum reaction. The enzyme has a conserved SAM-binding pocket and an RNA-binding groove through which it interacts

with viral RNA at the 5' end [6]. The structural configuration of the MTase domain allows for its orientation. The methyl groups are precisely assigned to the guanine base and ribose sugar of the RNA, which are required for proper cap formation and methylation [10].

Replication and translation: Methylated 5' cap and adjacent 2'-O-methylation promote efficient translation by acting as a protective structure that prevents viral RNA degradation by exonuclease [9]. Additionally, the cap structure facilitates translation. Recruiting host translation machines. This leads to more efficient protein synthesis and viral replication.

1.3 Motivation

The motivation behind CADD for Dengue NS5 Methyltransferase is driven by the urgent need for effective antiviral treatments. Dengue virus is responsible for millions of infections annually, and the lack of specific antiviral drugs makes its management challenging [11]. NS5 methyltransferase is a critical enzyme in the dengue virus replication cycle, as it is involved in the capping of viral RNA, which is essential for the virus's stability and immune evasion. Targeting this enzyme with selective inhibitors could disrupt the replication process and provide a novel therapeutic approach for dengue. CADD offers a powerful, cost-effective tool to accelerate the identification and design of potential inhibitors for dengue NS5 methyltransferase. By simulating the binding interactions and assessing the pharmacokinetics and toxicity profiles of these compounds in silico, CADD reduces the time and resources typically required for experimental drug discovery. Moreover, CADD enables the optimization of drug-like properties, such as bioavailability and specificity, before synthesizing and testing them in the lab. With the increasing burden of dengue infections and the lack of targeted treatments, the use of CADD in discovering effective inhibitors of NS5 methyltransferase holds great promise for developing advanced antiviral therapies,

ultimately improving public health outcomes and reducing the global impact of dengue.

Chapter 2

Literature Review

The Dengue Virus, a member of the Flaviviridae family, poses a significant global health threat, infecting approximately 390 million individuals annually. Among the various non-structural proteins encoded by the DENV genome, NS5 plays a crucial role in viral replication and immune evasion. The N-terminal domain of NS5, (SAM) dependent methyltransferase (MTase), exhibits methyltransferase activity, which is essential for the methylation of the viral RNA cap at the N'7 and 2-O positions [6]. This methylation is critical for RNA stability, efficient translation, and evasion of the host immune response. Due to its vital function and conserved structure across serotypes, the NS5 MTase is considered a promising target for antiviral drug development. The application of CADD has significantly extrapolated the discovery of inhibitors targeting viral enzymes, including NS5 MTase. CADD combines structural biology, computational chemistry, and molecular modeling to identify potential inhibitors in a cost-effective manner. A key aspect of CADD is (SBDD), which has been facilitated by the availability of high-resolution crystal structures of the NS5 MTase domain for all four dengue serotypes, such as PDB entries 1L9K [12] and 2P3L [6]. These structures have enabled molecular docking, pharmacophore modeling, and virtual screening efforts aimed at identifying inhibitors that can bind to the SAM-binding site or RNA-binding grooves.

Numerous in silico studies have demonstrated the efficacy of CADD in identifying potential inhibitors of DENV NS5 MTase. For instance, in research conducted for virtual screening and molecular docking of extensive chemical libraries, small molecules capable of binding to the SAM-binding pocket [13]. Similarly, [14] employed molecular dynamics (MD) simulations and free energy calculations to assess the binding affinity and stability of candidate inhibitors [15]. These studies often incorporate ADMET predictions to efficiently filter pharmacological compounds in the drug development pipeline. The natural product database has also been explored using CADD for DENV MTase inhibition. Phytochemicals from medicinal plants such as ***Andrographis paniculata***, ***Panax ginseng***, and ***Azadirachta indica*** have shown promising docking scores and favorable binding interactions with the NS5 MTase catalytic site [16]. For example, flavonoids and polyphenolic compounds have demonstrated strong hydrogen bonding with key **active-site residues such as Lys61, Asp146, and Glu149** [13].

Machine learning techniques and (QSAR) models are increasingly integrated with traditional (CADD) workflows to enhance accuracy and facilitate scaffold hopping. Recent advancements in deep learning have further enabled the generation of novel molecular scaffolds with desirable drug-like properties, particularly those conforming to the NS5 methyltransferase (MTase) pocket [17]. Despite these advancements, the translation of in silico findings to effective in vitro and in vivo inhibitors remains a significant challenge. Issues such as off-target effects, poor bioavailability, and resistance mutations must be addressed through iterative adaptation and experimental validation. In summary, CADD for NS5 has emerged as a powerful approach for the rational design of MTase inhibitors. By leveraging molecular docking, dynamics, pharmacophore modeling, and virtual screening, researchers have established a robust foundation for the development of novel anti-dengue therapeutics. The continuous integration of artificial intelligence, cheminformatics, and high-throughput screening holds promise for further enhancing the identification of lead compounds targeting this critical viral enzyme.

Chapter 3

Objectives

- To study the structural dynamics of NS5 Methyltransferase of Dengue virus.
- To analyze the inhibitory action of Sinefungin against NS5 Methyltransferase using MD Simulation.
- To perform Virtual Screening of the selected database and predict inhibitors to NS5 Methyltransferase.
- To perform ADMET analysis of the screened hit molecules.
- To explore the energetics of binding of the hit molecules to DENV NS5 Methyltransferase using MD simulations and free energy calculations.

Chapter 4

Methodology

4.1 Target protein selection and preparation

Preparation of the target protein is a pivotal step in virtual screening, as the structural integrity of the target protein can substantially impact the results of computational analyses. Proper preparation encompasses several essential steps to ensure that the protein structure utilized in docking simulations is both accurate and suitable for binding studies. Initially, the crystal structures of the Dengue NS5 methyltransferase (2P3L) protein are acquired from the Protein Data Bank (PDB) [6]. It is imperative to address common issues associated with experimental structures, such as missing hydrogen atoms and residues [13]. Determining the correct protonation states of the protein is crucial, particularly as it can influence binding interactions. Protonation states are computed across a pH range of 7 ± 2, aligning the proteins with physiological conditions. This step is vital for accurate molecular interactions during virtual screening. Identifying the binding site is another critical step in preparing the target protein, which involves selecting all residues located within 5 Å of the ligand in the crystal structure of the target. Following this selection, visual inspection may refine the list by adding residues beyond the 5 Å threshold if they are deemed necessary for maintaining the continuity of the binding cavity. This ensures that all relevant residues are included

for potential interactions with the ligand. After proper preparation of the protein structure, docking is performed against reference receptors, typically within a specified region surrounding the co-crystallized ligands [7]. For consistent and unbiased results across multiple virtual screening programs, a common docking method known as rigid docking is employed. In this approach, the target atoms remain fixed while small molecules are flexible, allowing for the evaluation of various conformations during docking simulations.

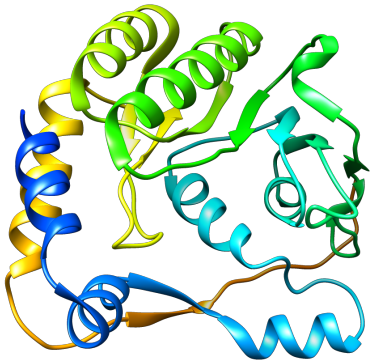

FIGURE 4.1: Dengue virus protein structure PDB (2P3L).

4.2 Ligand library preparation

In the virtual screening campaign aimed at the NS5 MTase of the dengue virus, a chemically diverse and biologically pertinent compound library was curated from four esteemed databases: NPASS [18], Enamine [19], ChemSpace [20], and Asinex [21]. These databases were selected to ensure comprehensive coverage of both synthetic and natural product chemical spaces. The aggregated dataset comprised 545,337 small molecules, approximately distributed as follows: NPASS included natural products and derivatives with known or predicted bioactivity; Enamine provided commercially available, drug-like, and lead-like synthetic compounds;

ChemSpace offered a diverse library encompassing fragment-like, drug-like, and scaffold-based molecules; and Asinex focused on pharmacologically relevant and novel scaffolds. The final ligand library, consisting of 545,337 drug-like molecules, was compiled into a unified database, indexed, and prepared for high-throughput virtual screening utilizing molecular docking tools.

4.3 Virtual screening is conducted using molsoft software

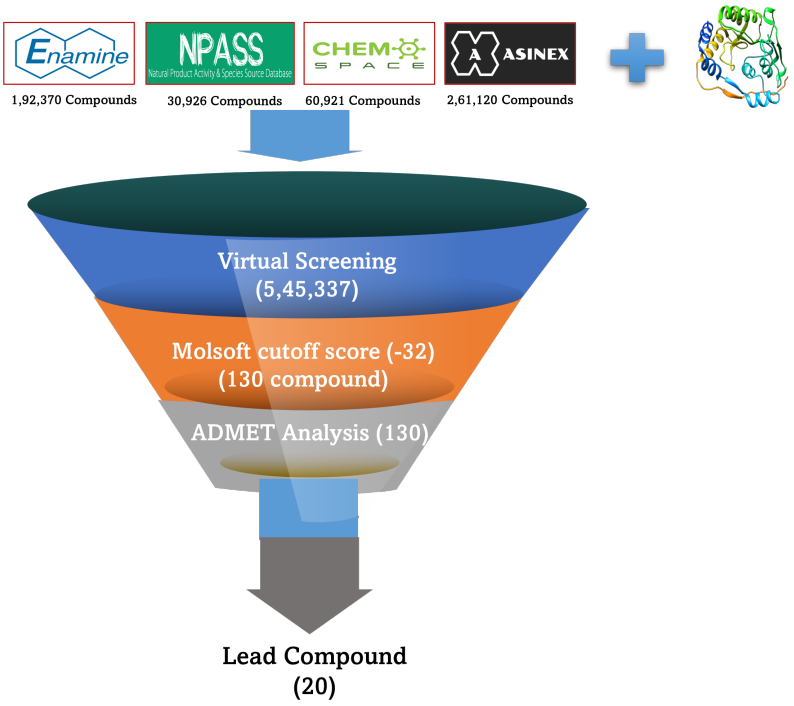

FIGURE 4.2: Virtual screening of 545,337 molecules to get lead molecules.

In a comprehensive virtual screening study utilizing MolSoft software, we analyzed the binding interaction between the protein 2P3L (Dengue protein) and the ligand (Sinefungin). The process commenced with the preparation of the 2P3L protein structure, which involved optimizing its structure through the addition of hydrogen atoms, removal of water

molecules, and correction of residues or missing atoms. Subsequently, the 3D structure of Sinefungin was obtained and optimized to ensure proper protonation and molecular charge distribution for accuracy. The docking procedure was performed using MolSoft ICM (internal coordinate mechanics) molecular docking module. Sinefungin was docked at the specified binding site of 2P3L [22]. Multiple docking poses were generated, and the software's scoring function, incorporating parameters such as ligand-protein interaction energy and binding affinity, was employed to rank the poses. Data visualization was conducted. Additionally, the binding energy for each docked complex was calculated to evaluate the stability of ligand-protein complexes. The highest-scoring poses indicated a favourable binding interaction between Sinefungin and 2P3L, with specific interactions involving the polar and non-polar regions of the protein. This result suggested that Sinefungin is a promising ligand for 2P3L, and the top-scoring complex was selected for further investigation, including potential experimental validation to confirm binding affinity and biological relevance [23].

4.4 ADMET analysis of selected molecule

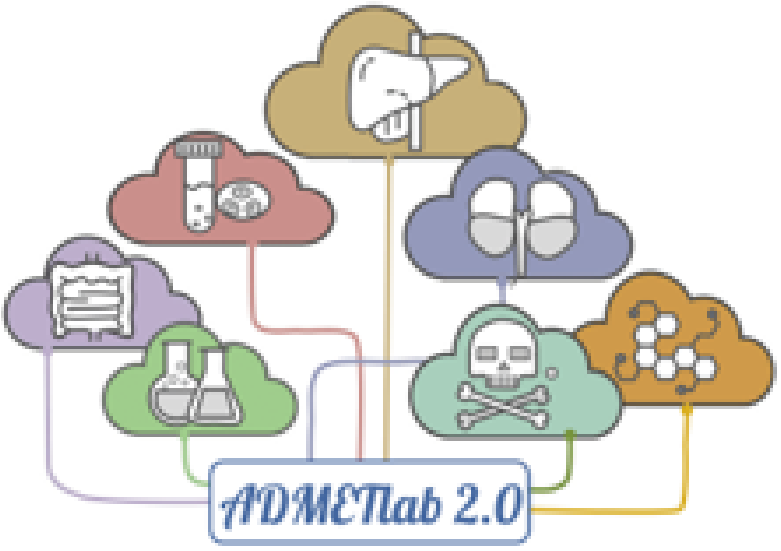

FIGURE 4.3: ADMET performed using web server.

ADMETLab 2.0 [24] is a sophisticated computational tool utilized for analysing the ADMET properties of compounds to facilitate the prediction of pharmacokinetic and toxic behaviour [6]. Upon input of a compound's structure, typically in SMILES notation or as a 3D file, the software predicts critical properties such as intestinal absorption, blood-brain barrier penetration, binding to plasma proteins, Cytochrome P450 enzyme interactions, estimated renal clearance, metabolic stability, potential for hepatotoxicity, and other toxicity concerns including mutagenicity and carcinogenicity [24]. ADMETLab 2.0 provides comprehensive data on a compound's potential as a biotoxin agent, enabling researchers to evaluate whether a drug candidate possesses favourable pharmacokinetic properties.

4.5 MD simulation protocol

First, using AMBER 24's pmemd.cuda module, [25] conventional Molecular Dynamic simulations were performed. Force field variables were generated in AMBER before simulations [22]. using the Leap Module and the force field used for protein is ff19SB, and for ligand (Sinefungin) is GAFF2 [26]. The AMBER antechamber module was employed to ascertain the ligand atomic charges. Each system was solvated in an octahedral box, with the solute and the box separated by a 10 Å gap using the TIP3P water model [27]. To neutralize the system, the right amount of Na^+ and Cl^- ions was added. The SHAKE method was used to restrict the lengths of all bonds, including the hydrogen bond.The method used to manage non-bonded electrostatic interactions was Particle Mesh Ewald (PME), with a threshold set at 10 Å [28]. A timestep of 2 fs was kept constant during the simulation. A clear methodology comprising the sequential processes of minimization, heating, and equilibration was carefully followed before starting the production simulation. For the solvated complexes, two stages of energy minimization were carried out. A weak harmonic constraint of 2 kcal $\text{mol}^{-1}\text{Å}^2$.

Binding energy calculation plays a critical role in dengue drug discovery, particularly in understanding the interaction between ligands and target proteins, such as NS5 methyltransferase. The binding energy ($_{\text{binding}}$) quantifies the free energy change associated with the formation of a stable protein-ligand complex. It can be calculated using Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) [29] or Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods, represented by the equation:

$$\Delta G_{\text{binding}} = E_{\text{MM}} + \Delta G_{\text{solvation}} - T\Delta S \tag{4.1}$$

Where: (E_{MM}) represents the molecular mechanics energy, including: [$E_{\text{MM}} = E_{\text{vdW}} + E_{\text{electrostatic}}$] ($\Delta G_{\text{solvation}}$) consists of polar ((ΔG_{polar})) and non-polar (($\Delta G_{\text{nonpolar}}$)) contributions: [$\Delta G_{\text{solvation}} = \Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}}$

] ($T\Delta S$) is the entropic contribution, where (T) is the temperature and () is the entropy change.

These equations are typically applied in MM-PBSA or MM-GBSA methods to compute binding free energy in drug discovery studies [30].

Docking simulations are used to predict the binding pose and calculate preliminary binding scores, which guide the selection of promising drug candidates. Molecular dynamics simulations refine these calculations [31], allowing for dynamic analysis of the complex stability. These simulations consider conformational flexibility and the solvent environment, providing realistic insights into protein-ligand interactions. Calculated binding energies help identify ligands with high affinity, enabling further optimization for improved efficacy against the dengue virus.

4.6 Trajectory analysis of lead molecule

4.6.1 Root mean square deviation

RMSD is also integral to molecular dynamics (MD) simulation, where it is used to monitor the stability and transformative changes of biomolecules over time. By calculating the RMSD of a molecular trajectory against a reference structure, researchers can identify deviations that may indicate structural deviation, folding events, or corresponding flexibility [32]. This application is important to understand dynamic biological processes and the functional implications of molecular motion.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \tag{4.2}$$

where (N) is the number of points being compared, (x_i) represents the coordinates of the first structure, and (y_i) represents the coordinates of the second structure

4.6.2 Radius of gyration ($\mathbf{R}_g$)

$$k = \sqrt{\frac{r_1^2 + r_2^2 + r_3^2 + \ldots + r_n^2}{n}} \tag{4.3}$$

where $(r_1, r_2, r_3, \ldots, r_n)$ are the perpendicular distances from the axis of rotation to each particle of mass (m) constituting the body, and (n) represents the number of particles

4.6.3 Solvent accessibility of surface area

SASA is particularly important in the context of protein folding, where the burial of hydrophobic residues within the protein structure and their exposure to the solvent are critical for stability [33].

The degree to which hydrophobic residues are buried correlates with the protein's overall stability and folding kinetics. In computational studies, SASA is used to predict the folding behavior of proteins by examining how changes in surface area can impact the energetic landscape of folding [33].

$$SASA = \sum_i A_i \tag{4.4}$$

where (A_i) represents the accessible surface area of atom (i)

4.6.4 Root mean square fluctuation

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.

$$RMSF = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \langle x \rangle)^2} \tag{4.5}$$

where (x_i) represents the coordinates of particle (i) and $(\langle x \rangle)$ denotes the ensemble average position of that particle.

In contrast to RMSD where time specific value is obtained by taking the 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.

4.6.5 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. The differences in free energy involves integration of both polar and non-polar solvation energy components. As an additional refinement, entropy Contributions to the total free energy are added. A MMPBSA.py nabnmode 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, is used to compute the electrostatic component. The LCPO technique used in Sanders approximates the hydrophobic contribution.

$$\Delta G_{\text{binding}} = \Delta H - T\Delta S \approx \Delta E_{\text{internal}} + \Delta G_{\text{sol}} - T\Delta S \tag{4.6}$$

$$\Delta E_{\text{internal}} = \Delta E_{\text{covalent}} + \Delta E_{\text{elec}} + \Delta E_{\text{vdw}} \tag{4.7}$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}} \tag{4.8}$$

4.6.6 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 including 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.

4.6.7 Hydrogen bond analysis

Hydrogen bonds arise from electrostatic interactions between hydrogen donors 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.

4.6.8 Dynamic cross-correlation matrix analysis

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-correlation coefficient of the i tTM and j tTM alpha carbons that correspond to the particular residues is given in the matrix, and the DCC generates an 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.

$$DCC = \frac{\langle \Delta \mathbf{r}_i(t) \cdot \Delta \mathbf{r}_j(t) \rangle_t}{\sqrt{\langle \|\Delta \mathbf{r}_i(t)\|^2 \rangle_t} \sqrt{\langle \|\Delta \mathbf{r}_j(t)\|^2 \rangle_t}} \tag{4.9}$$

The DCC creates a heatmap of N x N matrix, 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.

4.6.9 Principle component analysis

Principal component analysis 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 is given by $< xi >)$ and $< xj >$ respectively. Covariance matrix of 3N x 3N is generated for N atoms in three dimensions and is diagonalized to get the eigenvalues. A TC A = . Here, A is the eigenvector and 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, which together can describe the system with sufficient accuracy, are generally used in analysis [34].

4.6.10 Network analysis

To visualize protein structures other than secondary structure and fold arrangements is to represent the interactions between residues in a network. trajectories. The network can be built based on C, C, atom pairs, centroid networks, or interaction-strength networks. Here, the node is created based on the C atom of an amino acid residue, and an unweighted edge is constructed if the paired residue C-C distance lies within the threshold distance (RC) of 7 Å. Overall network view supports various analyses for identifying functional residue, predicting coevolving residues, understanding the mechanism of protein-protein interactions or domain-domain for understanding the communications between them. We have performed Network analysis of our systems using the webPSN server [35].

Chapter 5

Result and Discussion

5.1 Virtual screening of selected dataset

A virtual screening study was carried out using Molsoft ICM-VLS-Pro software [36] to identify potential drug candidates against the dengue virus. The study involved a chemical library of 545,337 ligands, which were evaluated for their binding affinity to the active site of the target protein, likely Pocket (NS5 methyltransferase) [6]. Molsoft's scoring function was used to rank the ligands, with a cutoff score of < -32 applied to filter out low-affinity candidates. The cutoff represents the potential inhibitor.

From the initial 545,337 ligands screened, 130 molecules achieved binding scores better than or equal to the cutoff score, highlighting their strong interaction potential with the target protein. These 130 molecules represent the most promising drug candidates and serve as the basis for further investigation.

The selected ligands are now prioritized for advanced computational and experimental evaluations. Molecular dynamics simulations will refine the insights by analyzing the stability of the protein-ligand complexes over time, accounting for conformational changes and solvent effects. Furthermore, MM-PBSA or MM-GBSA methods will be used to refine

the binding energy values and provide more accurate predictions of ligand binding efficacy.

Experimental validation, including in vitro and in vivo assays, will help confirm the inhibitory activity and pharmacokinetic properties of these molecules [24]. Structural optimization may also be pursued to enhance their drug-likeness and efficacy.

This study underscores the capability of computational tools, particularly Molsoft ICM-VLS-Pro, to streamline the drug discovery process. By efficiently screening a large library of ligands, the study successfully identified 130 potential inhibitors, marking a significant step toward the development of effective treatments for dengue virus infection.

TABLE 5.1: Virtual screening result.

Database	Ligand ID	Molecular Weight (g/mol)	Score
NPASS	NPC262282	320.05	-48.30
NPASS	NPC476622	320.25	-47.43
NPASS	NPC278652	600.11	-46.87
NPASS	NPC474895	340.08	-44.06
NPASS	NPC313754	268.08	-43.98
NPASS	NPC84362	478.11	-38.5
ASINEX	LMK 19839623	346.34	-50.49
ASINEX	LMK 19841893	431.44	-45.40
ASINEX	SYN 19838209	365.41	-44.33
ASINEX	LMK 19840884	391.4	-43.05
ASINEX	ADM 22811797	378.39	-42.82
CHEMSPACE	CSCS00102252427	301.32	-48.38
CHEMSPACE	CSCS00020661893	285.68	-45.87
CHEMSPACE	CSCS06256037270	350.76	-45.66
CHEMSPACE	CSCS02125539956	249.31	-44.31
CHEMSPACE	CSCS06623409447	210	-44.13
ENAMINE	LMK 19839623	239.15	-49.03
ENAMINE	LMK 19841893	390	-48.18
ENAMINE	SYN 19838209	450	-47.34
ENAMINE	LMK 19840884	310	-45.42
ENAMINE	ADM 22811797	440	-43.83

5.2 Performing ADMET analysis

A systematic ADMET analysis was performed using ADMETlab 2.0 to evaluate the pharmacokinetic and toxicity profiles of 130 molecules, aiming to identify 20 promising candidates. Initially, the molecular dataset was prepared with the SMILES strings or structural files of the compounds, ensuring compatibility with the software's batch evaluation module. The molecules were then analyzed for key ADMET properties, including absorption, distribution, metabolism, excretion, and toxicity. ADMETlab 2.0's predictive models generated 17 physicochemical properties, 13 medicinal chemistry metrics, 23 ADME endpoints, and 27 toxicity endpoints, providing a comprehensive evaluation of the compounds' profiles [24]. Thresholds were applied to filter the dataset based on essential properties like solubility, permeability, metabolic stability, and toxicity. Molecules with poor performance were excluded, while those with excellent or medium property ratings were retained. Following the scoring process, the molecules were ranked according to their overall ADMET profiles. Out of the initial 130 molecules, the top 20 were selected for further consideration. Detailed profiles of these candidates were reviewed to ensure they met the criteria for drug-likeness and safety. The resulting compounds from this analysis provide a strong foundation for subsequent stages of drug discovery. These include experimental validation and optimization to enhance the therapeutic potential of the selected molecules. By systematically narrowing down the candidates, this protocol demonstrates the efficacy of ADMETlab 2.0 in accelerating the identification of viable drug candidates. Let me know if you'd like additional details or further assistance!

TABLE 5.2: ADMET Analysis of selected virtual screening molecules.

Ligand ID	Ligand Database	Molecular Weight (g/mol)	HBA	HBD	Lipinski rule	Toxicity
NPC84362	NPASS	478.11	12	7	NO	CLEAR
NPC474895	NPASS	340.08	9	5	YES	CLEAR
NPC48863	NPASS	406.15	11	7	NO	CLEAR
LMK 13948248	ASINEX	374.11	7	2	YES	CLEAR
LMK 13948248	ASINEX	374.11	7	2	YES	CLEAR
LMK 13958241	ASINEX	382.11	9	2	YES	CLEAR
ART19553705	EMANINE	292.12	7	2	YES	CLEAR
ART10537076	EMANINE	182.04	3	1	YES	CLEAR
ART19553705	EMANINE	292.12	7	2	YES	CLEAR
CSC0002360457	CHIMERSPACE	224.16	4	3	YES	OI FAIR
CSC0002360253	CHIMERSPACE	240	8	5	YES	OI FAIR

5.3 Structural stability and solvent accessibility analysis

First, we want to oberserve the dynamic behaviour and stablity of the apo and complex of the system throughout the simulation. The simulation was 600ns for apo, complex, and lead molecule.

5.3.1 RMSD backbone

This (Figure 5.1) showcases a probability density plot of the RMSD values for backbone atoms in a protein structure under different conditions. The X-axis represents RMSD values (in Angstroms), while the

Y-axis denotes the corresponding probability density. Specific curves illustrate distinct conditions: APO (red), sinefungin (green), NPC84362 (blue), NPC474895 (orange), and NPC48863 (yellow). These curves indicate variations in structural mobility in response to different systems. Under the APO condition, the red curve exhibits peaks around 1.5 Å, suggesting minimal structural deviations in the absence of a ligand. In contrast, the sinefungin condition displays a peak at 1.8 Å, indicating more pronounced structural changes in the protein backbone. Other lead molecules demonstrate unique peaks in their respective curves, reflecting varying degrees of structural fluctuations induced by these molecules. This analysis highlights the impact of ligands and inhibitors on protein stability. Higher RMSD values are associated with greater structural changes, while the comparative evaluation of the curves reveals the stabilizing or destabilizing effects of each molecule. Such insights are invaluable for understanding protein-ligand interactions and guiding the design of therapeutic agents, particularly in drug discovery and development.

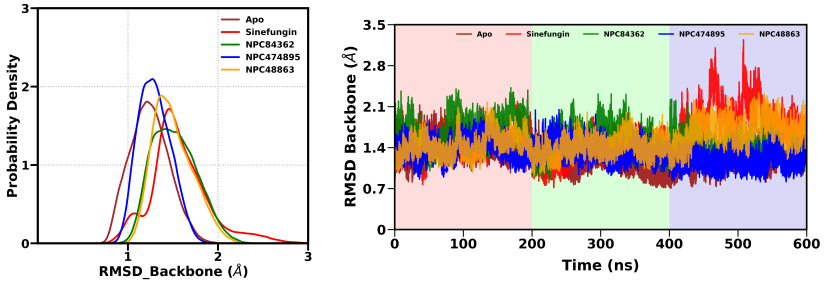

FIGURE 5.1: Probability distribution of the Apo and Complex Root Mean Square Deviation Curve of Backbone.

5.3.2 Radius of gyration

This (Figure 5.2) showcases a probability density plot of the ROG values for protein compactness throughout the simulation. The peaks in the plot provide insights into the conformational preferences of the molecule. The ROG distributions for Apo and Sinefungin conditions show peaks around 18.2 Å and 18.4 Å, respectively, indicating minor conformational variations. In contrast, the peaks for NPC84362, NPC474895,

and NPC48863 are closer to 18 Å, suggesting more compact molecular conformations under these conditions. The variations in probability densities indicate the degree of structural fluctuation, with higher values reflecting the dominance of certain conformations. This analysis is critical for understanding the impact of different ligands and inhibitors on the molecule's structural behavior. The probability density plot serves as a valuable tool for comparing molecular conformations, offering insights into ligand-induced changes that can inform further research and drug development efforts.

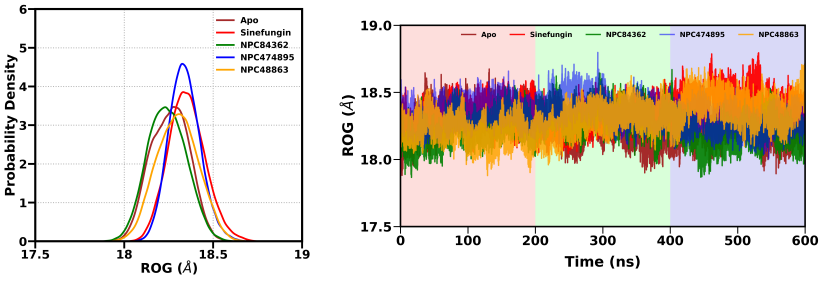

FIGURE 5.2: Probability distribution of the Apo and Complex Radius of gyration.

5.3.3 Solvent accessibility surface analysis

This (Figure 5.3) showcases a probability density plot of the SASA values for the protein Surface accessible with Protein throughout the simulation For the apo condition peak was observed around 440 nm², n contrast, the peaks for NPC84362, NPC474895, and NPC48863 also show a sharp peak in a similar range compared to sinefungin. In a time-series plot of SASA, also in the range of around 360-500 nm², rather than sinefungin, after 200ns goes down to 350 nm². These variations suggest that different ligands influence the molecule's solvent exposure in unique ways, altering its interaction with the surrounding environment. After this analysis, we interpret the result and say that all novel molecules have **similar conformational changes** upon binding.

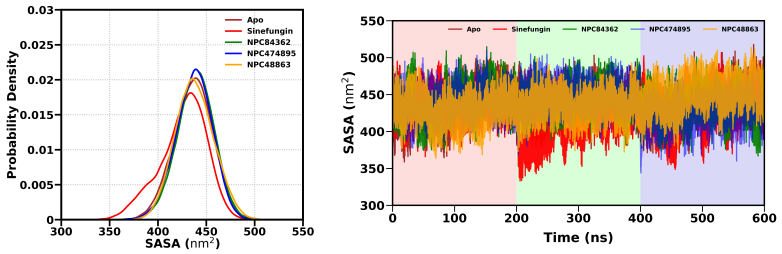

FIGURE 5.3: Probability distribution of solvent accessibility of surface analysis.

5.4 Binding pocket analysis

5.4.1 RMSD Pocket

This (Fig. 5.4) refers to a possibility density plot of RMSD Pocket value (in Angstrom) for various experimental conditions, which provides insight into the structural variations of the pocket area throughout the simulation.

The peaks in the curves represent the most potential RMSD pocket values under their respective conditions. For the APO position, the brown curve shows a major peak, suggests minimal structural variation in the pocket area when no ligand is bound. In comparison, red curves show different distributions for NPC inhibitors and red curves for green, blue, and orange curves, indicating that these inspire different levels of changes in ligand pockets.

This plot is important to analyze ligand-inspired flexibility or hardness of the pocket area, which can affect ligand binding and efficacy. By comparing RMSD distribution, it is possible to evaluate the stabilization or destabilizing effect of various molecules on the pocket structure, offering valuable guidance for drug design and molecular study.

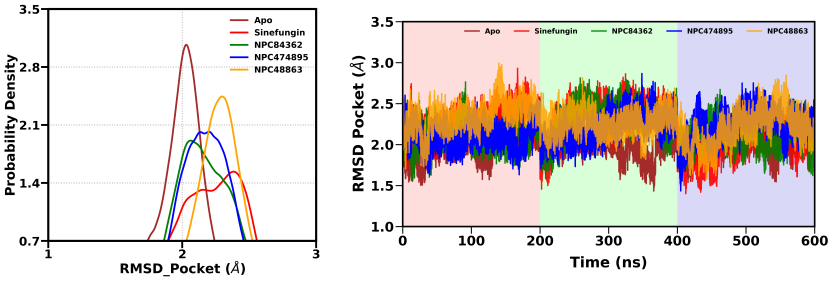

FIGURE 5.4: Probability distribution of the Apo and Complex Root Mean Square Deviation curve of Pocket.

5.5 Root mean square fluctuation Analysis

The (figure 5.5) represents a probability density plot showing the RMSF values for different residues in a protein structure across five experimental conditions: Apo (brown), Sinefungin (red), NPC84362 (green), NPC474895 (blue), and NPC48863 (orange). The x-axis corresponds to the residue numbers, while the y-axis represents the RMSF values in angstroms (Å), which indicate the flexibility of individual residues. Peaks in the graph highlight regions of the protein with higher flexibility, where the atoms exhibit significant motion. In contrast, troughs represent rigid regions with minimal atomic displacement. For example, specific peaks in the apo condition suggest that certain residues are more flexible in the absence of ligands, while the other conditions display shifts in peak positions and heights, reflecting the influence of ligands on protein dynamics. The analysis of RMSF values is essential for understanding how ligands and inhibitors affect the conformational flexibility of protein residues. Such flexibility is crucial for biological functions like ligand binding and enzymatic activity. Comparing the conditions provides insight into the stabilizing or destabilizing effects of different molecules, aiding in the rational design of drugs and the study of molecular mechanisms.

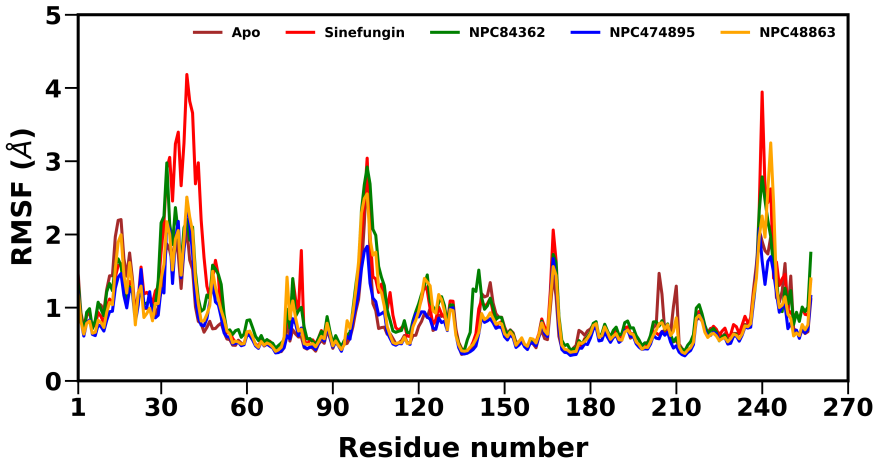

FIGURE 5.5: Root Mean Square Fluctuation Analysis of Apo, Control, and Lead Molecules.

5.6 Hydrogen bond occupancy per compound Analysis

Hydrogen bonds play a crucial role in stabilizing protein–ligand complexes due to their directional nature and relatively strong interaction strength. Therefore, analyzing their contribution is essential. In this study, we examined the number of hydrogen bonds formed across the entire simulation period for all three complexes. Among them, NPC474895 exhibited the fewest hydrogen bond interactions, primarily relying on van der Waals and other non-polar interactions for binding. In contrast, NPC48863 maintained a moderate number of hydrogen bonds throughout the simulation, typically ranging from 2 to 8. NPC834362 showed a slightly higher average, with hydrogen bonds fluctuating between 4 and 8 during the simulation. We also identified the specific donor and acceptor residues involved in hydrogen bonding for each complex, with key residues including Asp and Leu alongside significant contributions from the ligands themselves.

TABLE 5.3: Hydrogen Bond Occupancy Per Compound.

Binding couples		**Molecular dynamics**	
Acceptor	**Donor**	**Distance(Å)**	**Occupancy(%)**
Sinefungine			
LEU_81@O	LIG_266@O5	2.76	58.2
ASP_80@OD2	LIG_266@O4	2.66	48.4
ASP_132@OD1	LIG_266@O11	2.59	41.3
ASP_132@OD2	LIG_266@O4	2.60	40.32
ASP_80@OD2	LIG_266@O4	2.60	40.32
LIG_266@O3	GLY_76@N	2.87	16.4
LIG_266@O4	TRP_82@N	2.90	12.8
NPC834362			
ASP_80@OD1	LIG_266@O5	2.64	74.2
ASP_132@OD1	LIG_266@O12	2.63	46.4
ASP_80@OD2	LIG_266@O4	2.73	46.3
LIG_266@O3	GLY_87@N	2.89	21.1
LIG_266@O1	ARG_85@N	2.89	16.9
LIG_266@O8	GLY_149@N	2.92	10.7
NPC474895			
LEU_81@O	LIG_266@O7	2.71	62.5
ASP_80@OD1	LIG_266@O9	2.64	56.6
ASP_80@OD2	LIG_266@O9	2.65	41.9
LIG_266@O7	GLY_87@N	2.89	38.01
NPC48863			
ASP_80@OD1	LIG_266@O5	2.62	99.7
ASP_80@OD2	LIG_266@O12	2.76	78.6
LIG_266@O9	GLY_87@N	2.87	47.1
LIG_266@O10	GLY_87@N	2.91	34.8
ASP_80@OD8	LIG_266@N	2.90	31.3
LIG_266@H19	GLY_86@N	2.83	24.4

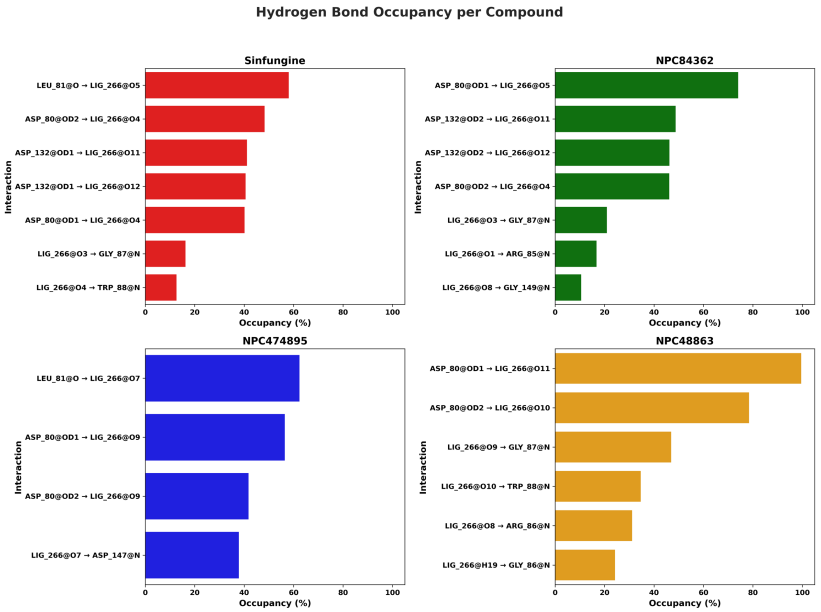

FIGURE 5.6: Hydrogen Bond Analysis of Control and Lead molecules.

5.7 Binding free energy analysis (MM-PBSA)

The (figure 5.7) summarizes thermodynamic parameters for four ligands—Sinefungin, NPC84362, NPC474895, and NPC48863—highlighting their interactions with the target protein. The parameters include van der Waals energy (ΔE_{vdW}), electrostatic energy (ΔE_{elec}), polar solvation energy (ΔG_{polar}), non-polar solvation energy (ΔG_{polar}), enthalpy change ΔH, entropy contribution $T\Delta S$, and Gibbs free energy ΔG, all measured in kcal/mol. Sinefungin shows strong van der Waals and electrostatic interactions (ΔE_{vdW}) = -40.16, ΔE_{elec} = -85.48) and a relatively low ΔG = (-14.35), indicating a stable binding. NPC84362 exhibits the highest ΔE_{elec}(- 101.13) and a ΔG of -19.77, suggesting that it forms the most stable complex among the ligands. NPC474895 has weaker van der Waals and electrostatic interactions (ΔE_{vdW}) = -27.87, ΔE_{elec} = -83.69) and the least negative ΔG (-7.77), indicating a less stable binding. NPC48863 shows moderate stability with ΔE_{vdW} = -34.35 and ΔE_{elec} = -71.34, leading to a ΔG of -8.63.

This analysis underscores the role of different energy components in ligand binding. A comparison of ΔG values provides insights into ligand binding affinities, which are essential for drug development and optimizing ligand design. The small standard deviations indicate reliable measurements.

TABLE 5.4: Binding free energy analysis (MM-PBSA) of sinefungin and lead molecules.

Ligand	ΔE_{vdw} **(kcal/mol)**	ΔE_{ele} **(kcal/mol)**	ΔG_{polar} **(kcal/mol)**	$\Delta G_{\text{nonpolar}}$ **(kcal/mol)**	ΔH **(kcal/mol)**	ΔG_{bind} **(kcal/mol)**
Sinefungin	-40.16 (0.05)	-85.48 (0.14)	8.24 (0.09)	-4.49 (0.002)	-42.9 (0.06)	-14.5 (0.64)
NPC48863	-37.97 (0.05)	-101.13 (0.15)	92.88 (0.09)	-4.59 (0.002)	-50.75 (0.07)	-19.77 (0.75)
NPC474895	-27.87 (0.05)	-83.69 (0.14)	80.41 (0.1)	-3.65 (0.001)	-34.81 (0.05)	-7.77 (0.68)
NPC84362	-34.35 (0.05)	-68.76 (0.15)	71.34 (0.12)	-3.83 (0.002)	-35.60 (0.07)	-8.63 (0.79)

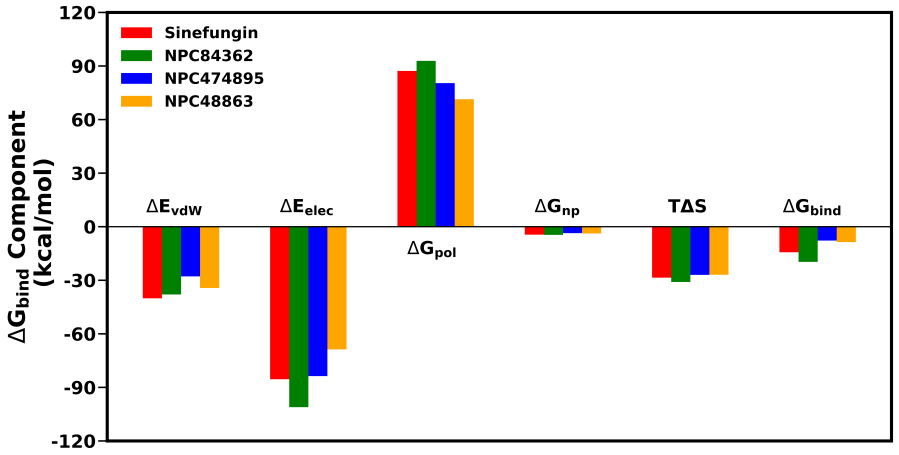

FIGURE 5.7: Graphical representation of binding free energy analysis (MM-PBSA).

5.7.1 Per residue calculation

In Fig. (5.8) the per-residue binding free energy decomposition reveals the specific contributions of individual amino acid residues to the overall binding of the complexes. In all three cases, some residues exhibit negative free energy values, indicating a favorable contribution to binding, whereas others display positive values, suggesting they hinder or oppose the binding interaction.

Sinefungin: ASP124 and ILE140 contributed favour binding while ASP139 disfavouring in complex formation.

NPC84362: All residue in ASP72, ASP124, and ILE140 contributed favour binding in complex formation.

NPC474895: ASP134 and ILE140 residue contributing in complex formation, while ASP72 disfavouring binding in complex formation.

NPC48863: All binding residue favouring ASP77, TRP80, and ASP139 contribute to binding complex formation.

The per-residue energy contributions to complex formation are particularly useful when analyzing cases involving mutations. If a mutation occurs in a key residue that significantly contributes to binding, it may lead to alterations in complex stability or result in a reduced binding affinity.

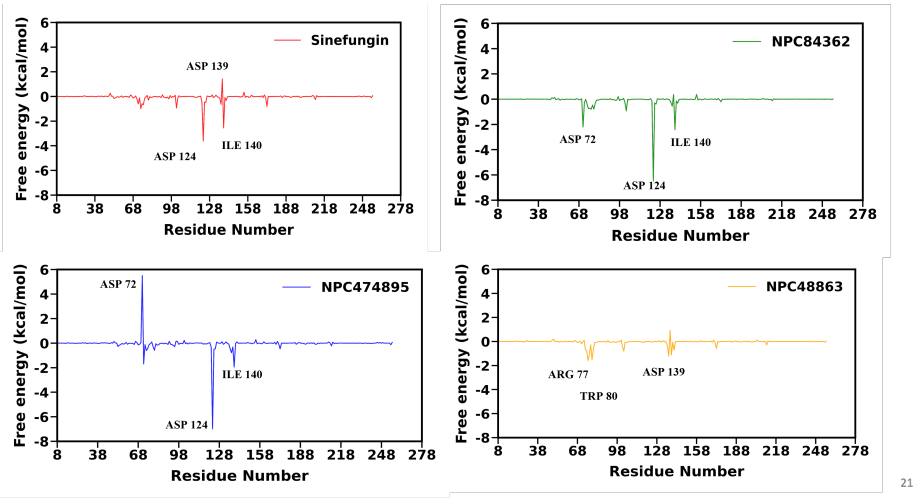

FIGURE 5.8: Per residue calculation of sinefungin and Leads molecule.

5.8 Ligand interaction study

In the case of NPC84362, 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 NP4888635, it has been investigated that major interactions are van der Waals and hydrophobic forces. In Sinefungin, a 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.

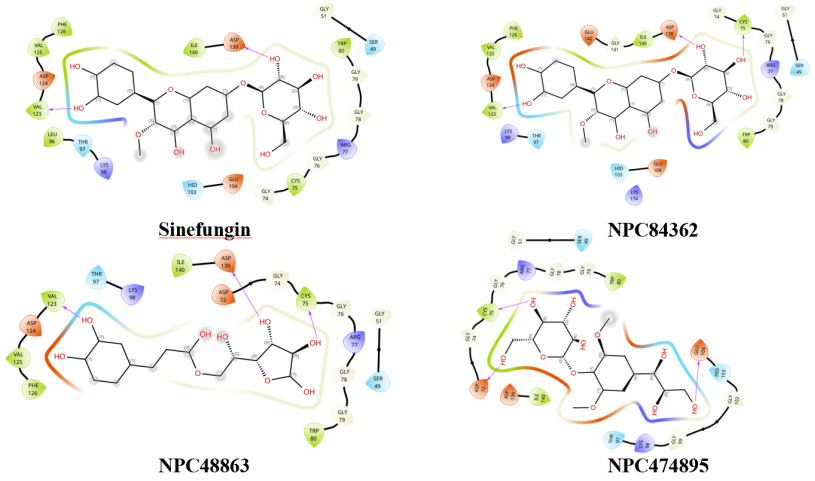

FIGURE 5.9: Ligand-Interaction study of the Apo and Complex.

5.9 Dynamics cross-correlation matrix analysis

Dynamic Cross-Correlation Matrix (DCCM) analysis investigates the correlated and anticorrelated motions of protein residues during molecular dynamics simulations. The results are presented in heatmaps, where the x-axis and y-axis represent residue numbers, and the color scale ranges from -1 to 1. Positive correlations, shown in red, indicate residues moving in a coordinated manner, often reflecting shared structural or functional roles. Conversely, negative correlations, depicted in blue, reveal residues with anticorrelated motions, which often occur in flexible regions such as loops or binding pockets. These anticorrelations can signify allosteric coupling, where the movement of one region triggers a compensatory shift in another. Analyzing these patterns under different conditions, such as ligand binding, helps identify structural dynamics and functional hotspots in the protein. This information is valuable for understanding protein mechanisms, guiding drug design, and exploring potential therapeutic targets.

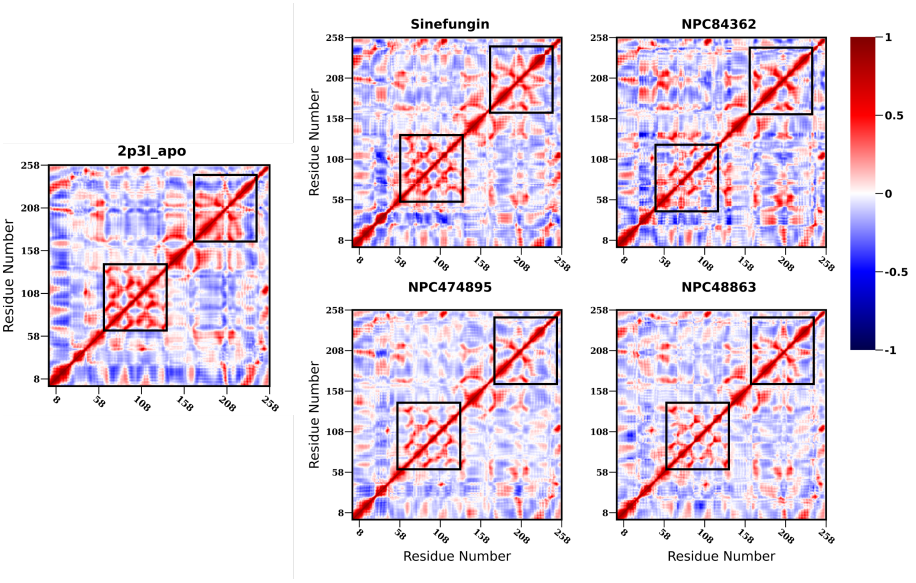

FIGURE 5.10: Dynamics Cross-correlation Matrix Analysis of Apo, Sinefungin(Control) and Leads molecule.

5.10 Hydrogen bond analysis

The (figure 5.11) displays a graph of the number of hydrogen bonds formed over time, measured in nanoseconds (ns), for four different compounds. The plot is divided into three sections, with the first half (0–200 ns) shaded pink, the second half (200–400 ns) shaded green, and the thired (400-600 ns) shaded blue. This division emphasizes the temporal changes in hydrogen bonding patterns, showing fluctuations across the three simulation phases. For instance, Sinefungin and NPC84364 demonstrates a relatively stable number of hydrogen bonds all interval, while NPC48863 shows significant variation. This analysis reveals the dynamic behavior of hydrogen bonding interactions, which are crucial for assessing the stability and efficacy of ligand-protein complexes. By comparing the compounds, it becomes possible to identify molecules with stronger or more stable interactions, providing valuable insights for molecular studies and drug design.

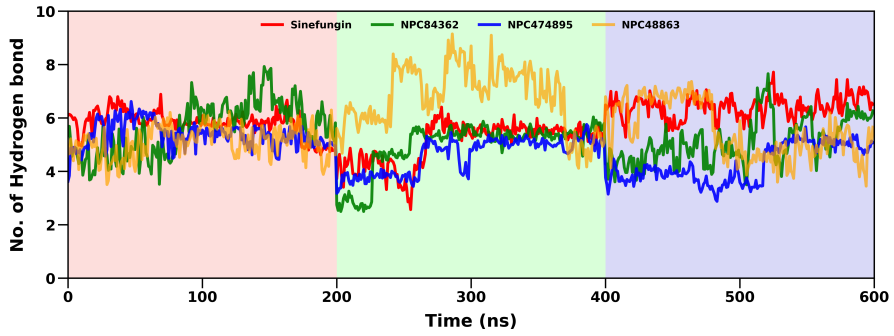

FIGURE 5.11: H-Bond Analysis of Control and Lead molecules.

5.11 Principle component analysis

Principal Component Analysis (PCA) is an effective method for simplifying and interpreting high-dimensional data, particularly in molecular dynamics studies. It begins with the calculation of a covariance matrix from the aligned trajectory data, which captures the correlations between atomic positions or structural variables throughout the simulation. This matrix is symmetric and positive semi-definite. By performing eigenvalue decomposition (or singular value decomposition) on the covariance matrix, A set of eigenvectors and corresponding eigenvalues is obtained. The eigenvectors represent the principal components (PCs)—the main modes of motion—while the eigenvalues quantify the variance along each mode.

These principal components allow projection of the original trajectory into a lower-dimensional space, making it easier to visualize and analyze the dominant motions driving the system's dynamics. This reduction helps in identifying key conformational changes and stable states.

In our PCA analysis of the four complexes, complex NPC84362 exhibits a distinct global minima, indicating a highly stable conformation during the simulation. Sinefungin shows two distinct global minima separated by a high energy barrier, indicating impossible interconversion between them. For NPC48863, a single global minimum is observed, indicating

a stable confirmation. In NPC474895, a prominent global minima is observed along with the presence of a local minima separated by a low energy barrier. It indicates that the structures are interconvertible.

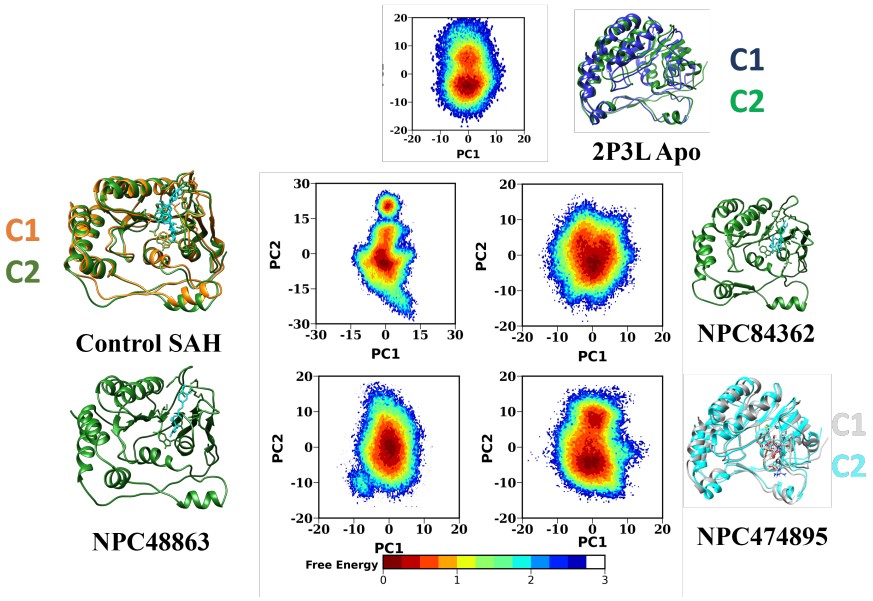

FIGURE 5.12: PCA Analysis of Apo, Sinefungin and Leads molecule.

5.12 Network analysis

TABLE 5.5: Comparison of network parameters across different conditions.

Parameter	Apo	Control	NPC84362	NPC474895	NPC48863
$I_{\min}$	3.76	4	3	3.23	3.01
Number of Isolated Nodes	243	247	241	243	246
Number of Links	270	271	307	268	274
Number of Hubs	34	45	49	31	38
Number of Links Initiated by Hubs	122	154	128	127	145
Number of Communities	8	7	10	8	9
Number of Nodes Involved in Communities	44	47	93	36	45
Number of Links Involved in Communities	57	65	119	44	55

From Table (5.4), we can see the number of nodes, links, hubs, and communities present in the systems. NPC84362 shows a high number of links, hubs, and communities as compared to the control and other lead molecules. Thus, it is evident that the residues present in NPC84362 are highly interacting with each other.

Network analysis in molecular dynamics (MD) simulations provides insight into the structural organization and stability of molecular systems by representing them as networks, where nodes (such as atoms, residues, or functional groups) are connected by edges representing interactions. In our analysis, we compared the four ligand-bound complexes with both the apo (unbound) structure and the control compound Sinefungin across both binding pockets. The results show that the ligand-bound systems exhibit a greater number of hubs, links, and communities compared to the apo and control structures. This increase in network complexity suggests enhanced structural stability of the NS5 methyltransferase when bound

to the four ligands, highlighting their stabilizing effect relative to the unbound state. From Fig(5.13) it can be concluded that NPC84362 has the highest number of populous communities like red, green, and blue as compared to apo, sinefungin, and other leads.

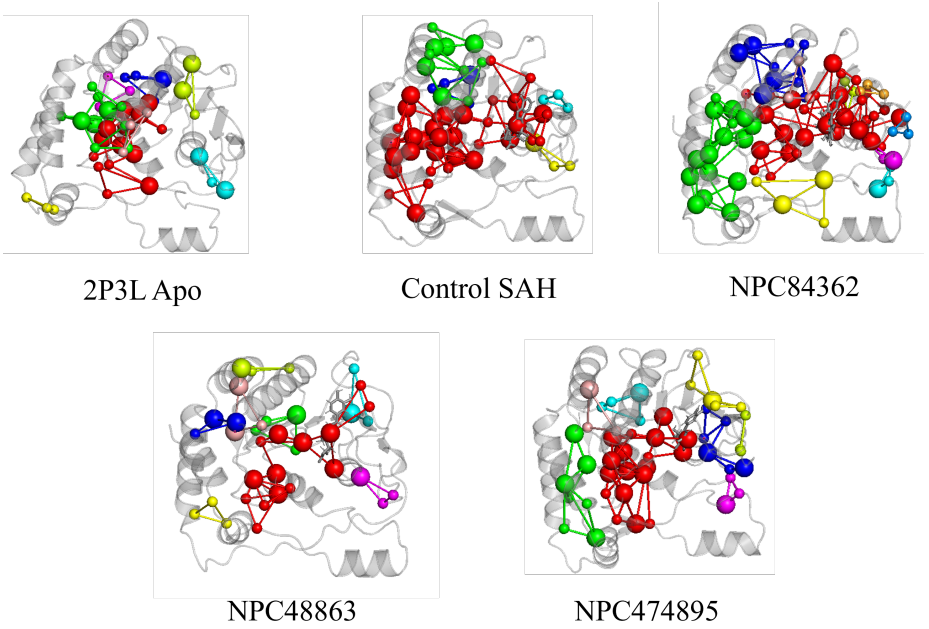

FIGURE 5.13: Network Analysis of Apo, Sinefungin and Leads molecule.

Chapter 6

Conclusion

In conclusion, CADD is a promising approach for identifying potential therapeutic agents that target the dengue virus NS5 methyltransferase. By utilizing molecular docking, virtual screening, and MD simulations. The comprehensive analysis of molecular dynamics simulations reveals distinct effects of different ligands on the structural and functional properties of the target protein. RMSD distributions showed the Apo condition had the lowest peak at 1.5 Å, while Sinefungin induced greater backbone deviation with a peak at 1.8 Å, and NPC inhibitors exhibited distinct ranges, indicating ligand-specific stabilization. The SASA data highlighted variations in solvent accessibility, with Apo and NPC ligands peaking near 400–450 nm², while Sinefungin caused a higher solvent exposure at 18.4 Å radius of gyration. RMSF analysis identified flexible regions critical for functional dynamics, with Apo showing localized peaks compared to ligand-bound conditions. Hydrogen bonding analysis demonstrated stable interactions for Sinefungin (average 6 bonds) over time, while NPC inhibitors exhibited fluctuating binding patterns. In MMPB-SA analysis NPC84362 shows higher gibbs free energy compare to Sinefungun which was -19.77 kcal/mol. while other lead molecules have lower binding energy as well as enthalpy compare to sinefungin. In Per-residue calculation for complexes ASP124, ILE140 and ASP72 play an important role in favouring the complex formation.

in NPC84362 ASP72, ASP124 and ILE140 all residue favour the complex formation while Sinefungin ASP139 disfavouring the complex formation. Ligand-protein distances highlighted closer binding for Sinefungin (6–7 Å), contrasting with NPC ligands that exhibited slightly higher distances and spatial variability. The DCCM analysis revealed correlated motions within secondary structures and anticorrelated dynamics in flexible loops and binding pockets, indicating ligand-induced allosteric effects.

Bibliography

[1] Zeeshan Mustafa, Haris Manzoor Khan, Mohd Azam, Hiba Sami, Syed Ghazanfar Ali, Islam Ahmad, Adil Raza, and Mohammad Azam Khan. Insight into the seroepidemiology and dynamics of circulating serotypes of dengue virus over a 4 year period in western uttar pradesh, india. *Access Microbiology*, 5(6):000567–v4, 2023.

[2] Nitin Singh, Amresh Kumar Singh, and Ankur Kumar. Dengue outbreak update in india: 2022. *Indian Journal of Public Health*, 67(1):181–183, 2023.

[3] Nitish Mondal. The resurgence of dengue epidemic and climate change in india. *The Lancet*, 401(10378):727–728, 2023.

[4] Gaurav Badoni, Pratima Gupta, Manju O Pai, Neelam Kaistha, Radhakanta Ratho, and Nusayha Sokeechand. dengue burden and circulation of dengue-2 serotype among children along with clinical profiling in uttarakhand, india: A cross-sectional study from 2018 to 2020. *Cureus*, 15(1), 2023.

[5] Usman Sumo Friend Tambunan, Mochammad Arfin Fardiansyah Nasution, Fauziah Azhima, Arli Aditya Parikesit, Erwin Prasetya Toepak, Syarifuddin Idrus, and Djati Kerami. Modification of s-adenosyl-l-homocysteine as inhibitor of nonstructural protein 5 methyltransferase dengue virus through molecular docking and molecular dynamics simulation. *Drug Target Insights*, 11:1177392817701726, 2017.

[6] Marie-Pierre Egloff, Etienne Decroly, Hélène Malet, Barbara Selisko, Delphine Benarroch, François Ferron, and Bruno Canard. Structural and functional analysis of methylation and 5-rna sequence requirements of short capped rnas by the methyltransferase domain of dengue virus ns5. *Journal of molecular biology*, 372(3):723–736, 2007.

[7] Mohd Imran, Abida, Nawaf M Alotaibi, Hamdy Khamees Thabet, Jamal Alhameedi Alruwaili, Lina Eltaib, Ahmed Alshehri, and Mehnaz Kamal. Investigation of natural compounds as methyltransferase inhibitors against dengue virus: an in silico approach. *Journal of Biomolecular Structure and Dynamics*, pages 1–16, 2024.

[8] Viwan Jarerattanachat, Chompunuch Boonarkart, Supa Hannongbua, Prasert Auewarakul, and Ruchuta Ardkhean. In silico and in vitro studies of potential inhibitors against dengue viral protein ns5 methyl transferase from ginseng and notoginseng. *Journal of Traditional and Complementary Medicine*, 13(1):1–10, 2023.

[9] Chandrabose Selvaraj, Dhurvas Chandrasekaran Dinesh, Umesh Panwar, Rajaram Abhirami, Evzen Boura, and Sanjeev Kumar Singh. Structure-based virtual screening and molecular dynamics simulation of sars-cov-2 guanine-n7 methyltransferase (nsp14) for identifying antiviral inhibitors against covid-19. *Journal of Biomolecular Structure and Dynamics*, 39(13):4582–4593, 2021.

[10] Brian J Geiss, Aaron A Thompson, Andrew J Andrews, Robert L Sons, Hamid H Gari, Susan M Keenan, and Olve B Peersen. Analysis of flavivirus ns5 methyltransferase cap binding. *Journal of molecular biology*, 385(5):1643–1654, 2009.

[11] Andrea Scarpino, Gyorgy G Ferenczy, and Gyorgy M Keseru. Comparative evaluation of covalent docking tools. *Journal of Chemical Information and Modeling*, 58(7):1441–1458, 2018.

[12] Marie-Pierre Egloff, Delphine Benarroch, Barbara Selisko, Jean-Louis Romette, and Bruno Canard. An rna cap (nucleoside-2-o-)-methyltransferase in the flavivirus rna polymerase ns5: crystal structure and functional characterization. *The EMBO journal*, 2002.

[13] See Ven Lim, Mohd Basyaruddin A Rahman, and Bimo A Tejo. Structure-based and ligand-based virtual screening of novel methyltransferase inhibitors of the dengue virus. In *BMC bioinformatics*, volume 12, pages 1–12. Springer, 2011.

[14] Amandeep Kaur, Anupamjeet Kaur, Deepti Goyal, and Bhupesh Goyal. How does the mono-triazole derivative modulate aβ42 aggregation and disrupt a protofibril structure: insights from molecular dynamics simulations. *ACS omega*, 5(25):15606–15619, 2020.

[15] Fatiha Benmansour, Iuni Trist, Bruno Coutard, Etienne Decroly, Gilles Querat, Andrea Brancale, and Karine Barral. Discovery of novel dengue virus ns5 methyltransferase non-nucleoside inhibitors by fragment-based drug design. *European journal of medicinal chemistry*, 125:865–880, 2017.

[16] Li Jian Yap, Dahai Luo, Ka Yan Chung, Siew Pheng Lim, Christophe Bodenreider, Christian Noble, Pei-Yong Shi, and Julien Lescar. Crystal structure of the dengue virus methyltransferase bound to a 5-capped octameric rna. *PloS one*, 5(9):e12836, 2010.

[17] Yongqian Zhao, Tingjin Sherryl Soh, Jie Zheng, Kitti Wing Ki Chan, Wint Wint Phoo, Chin Chin Lee, Moon YF Tay, Kunchithapadam Swaminathan, Tobias C Cornvik, Siew Pheng Lim, et al. A crystal structure of the dengue virus ns5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS pathogens*, 11(3):e1004682, 2015.

[18] Xian Zeng, Peng Zhang, Weidong He, Chu Qin, Shangying Chen, Lin Tao, Yali Wang, Ying Tan, Dan Gao, Bohua Wang, et al. Npass: natural product activity and species source database for natural

product research, discovery and tool development. *Nucleic acids research*, 46(D1):D1217–D1222, 2018.

[19] Manikandan Jayaraman, Lakshmanan Loganathan, Karthikeyan Muthusamy, and Krishna Ramadas. Virtual screening assisted discovery of novel natural products to inhibit the catalytic mechanism of mycobacterium tuberculosis inha. *Journal of Molecular Liquids*, 335:116204, 2021.

[20] Jose L Medina-Franco, Karina Martinez-Mayorga, Marc A Giulianotti, Richard A Houghten, and Clemencia Pinilla. Visualization of the chemical space in drug discovery. *Current Computer-Aided Drug Design*, 4(4):322–333, 2008.

[21] Tejas M Dhameliya, Prinsa R Nagar, and Normi D Gajjar. Systematic virtual screening in search of sars cov-2 inhibitors against spike glycoprotein: pharmacophore screening, molecular docking, admet analysis and md simulations. *Molecular diversity*, 26(5):2775–2792, 2022.

[22] Andreas W Gotz, Mark J Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 1. generalized born. *Journal of chemical theory and computation*, 8(5):1542–1555, 2012.

[23] Hideaki Shimizu, Akatsuki Saito, Junko Mikuni, Emi E Nakayama, Hiroo Koyama, Teruki Honma, Mikako Shirouzu, Shun-ichi Sekine, and Tatsuo Shioda. Discovery of a small molecule inhibitor targeting dengue virus ns5 rna-dependent rna polymerase. *PLoS neglected tropical diseases*, 13(11):e0007894, 2019.

[24] Guoli Xiong, Zhenxing Wu, Jiacai Yi, Li Fu, Zhijiang Yang, Changyu Hsieh, Mingzhu Yin, Xiangxiang Zeng, Chengkun Wu, Aiping Lu, et al. Admetlab 2.0: an integrated online platform for accurate and comprehensive predictions of admet properties. *Nucleic acids research*, 49(W1):W5–W14, 2021.

[25] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.

[26] Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling*, 25(2):247–260, 2006.

[27] Wesam S Qayed, Rafaela S Ferreira, and José Rogério A Silva. In silico study towards repositioning of fda-approved drug candidates for anticoronaviral therapy: Molecular docking, molecular dynamics and binding free energy calculations. *Molecules*, 27(18):5988, 2022.

[28] David A Case, Hasan Metin Aktulga, Kellon Belfon, David S Cerutti, G Andrés Cisneros, Vinícius Wilian D Cruzeiro, Negin Forouzesh, Timothy J Giese, Andreas W Gotz, Holger Gohlke, et al. Ambertools. *Journal of chemical information and modeling*, 63(20):6183–6191, 2023.

[29] Bill R Miller III, T Dwight McGee Jr, Jason M Swails, Nadine Homeyer, Holger Gohlke, and Adrian E Roitberg. Mmpbsa. py: an efficient program for end-state free energy calculations. *Journal of chemical theory and computation*, 8(9):3314–3321, 2012.

[30] Daniel R Roe and Thomas E Cheatham III. Ptraj and cpptraj: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, 9(7):3084–3095, 2013.

[31] Daniel R Roe and Thomas E Cheatham III. Parallelization of cpptraj enables large scale analysis of molecular dynamics trajectory data, 2018.

[32] Oliviero Carugo and Sándor Pongor. A normalized root-mean-spuare distance for comparing protein three-dimensional structures. *Protein science*, 10(7):1470–1473, 2001.

[33] Shennon Lu and Amy S Wagaman. On methods for determining solvent accessible surface area for proteins in their unfolded state. *BMC research notes*, 7:1–7, 2014.

[34] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[35] Michele Seeber, Angelo Felline, Francesco Raimondi, Simona Mariani, and Francesca Fanelli. Webpsn: a web server for high-throughput investigation of structural communication in biomacromolecules. *Bioinformatics*, 31(5):779–781, 2015.

[36] Polo C-H Lam, Ruben Abagyan, and Maxim Totrov. Hybrid receptor structure/ligand-based docking and activity prediction in icm: development and evaluation in d3r grand challenge 3. *Journal of Computer-Aided Molecular Design*, 33:35–46, 2019.

[37] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

[38] Chuan Tian, Koushik Kasavajhala, Kellon AA Belfon, Lauren Raguette, He Huang, Angela N Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, et al. ff19sb: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of chemical theory and computation*, 16(1):528–552, 2019.

[39] Lloyd K Daniels. Rapid in-office and in-vivo desensitization of an injection phobia utilizing hypnosis. *American Journal of Clinical Hypnosis*, 18(3):200–203, 1976.

[40] Romelia Salomon-Ferrer, Andreas W Gotz, Duncan Poole, Scott Le Grand, and Ross C Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald. *Journal of chemical theory and computation*, 9(9):3878–3888, 2013.