Pharmacophore Modeling and 3D QSAR Combined with Virtual Screening to Identify Novel Inhibitors of Protein Targets of Infectious Diseases

THESIS

Submitted in partial fulfilment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

J. THERESE PATRISHA

ID No 2010PHXF809H

Under the Supervision of

P. YOGEESWARI

BITS Pilani

Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2014

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

CERTIFICATE

This is to certify that the thesis entitled **"Pharmacophore Modeling and 3D QSAR Combined with Virtual Screening to Identify Novel Inhibitors of Protein Targets of Infectious Diseases"** and submitted by **J. THERESE PATRISHA** ID No **2010PHXF809H** for award of Ph.D. of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor _____

Name in capital block letters: **P. YOGEESWARI**

Designation: **Professor**

Date:

ACKNOWLEDGEMENTS

*First and foremost I would like to thank **God Almighty** who has given me this opportunity.*

*It is a great pleasure and privilege in expressing my deep gratitude towards my research supervisor, **Prof. P. Yogeeswari,** Professor and Associate Dean (Sponsored Research and Consultancy Division), Department of Pharmacy, BITS-Pilani, Hyderabad Campus for her continuous guidance, suggestions and support. She was always an inspiration to me for my work.*

*I deeply acknowledge and heartfelt thanks to **Prof. D. Sriram,** Professor, Department of Pharmacy, BITS-Pilani, Hyderabad Campus for his valuable suggestions, guidance offered to me during this period.*

*I am thankful to acknowledge my DAC member **Prof. Sajeli Begum** for her support and encouragement during this period.*

*I am grateful to **Prof. Bijendra N. Jain,** Vice-Chancellor, BITS-Pilani and **Prof. V. S. Rao,** Director, BITS-Pilani, Hyderabad Campus for allowing me to carry out my doctoral research work in the institute.*

*I am thankful to **Prof. M. M. S. Anand,** Registrar and **Prof. S. K. Verma,** Dean, Academic Research (Ph.D. Programme), BITS Pilani for their support to do my research work.*

*I would like to express my sincere thanks to **Prof. M.B. Srinivas,** Dean, Administration and **Prof. Vidya Rajesh,** Associate Dean, Academic Research (Ph.D. Programme), BITS-Pilani, Hyderabad campus for their continuous support and encouragement during my research work.*

*I would like to express my gratitude to **Prof. Shrikant Yashwant Charde,** Associate Professor & Head, Department of Pharmacy, BITS-Pilani, Hyderabad campus for providing me with all the necessary laboratory facilities and for having helped me at various stages of my research work.*

I sincerely acknowledge the help rendered by ***Prof. Punna Rao*** *and other* ***faculty members****, Department of Pharmacy, BITS-Pilani, Hyderabad campus.*

I am grateful to express my sincere thanks and acknowledgements to my supervisor's collaborators ***Dr. Neerja Koushik Basu,*** *New Jersy Medical College, New Jersy, USA,* ***Dr. Vinay Nandhikori,*** *NII, India and* ***Dr. Hanna,*** *NIRT, India for doing the activity studies.*

I am very much grateful to all my friends ***Pritesh, Ram, Madhu, Mallika, Manoj, Monika, Jean, Rukaiyya, Shalini, Reshma, Srikant, Gangadhar, Saketh, Brindha, Koushik, Renuka, Priyankas, Praveen, Mahibalan, Santhose, Brahmam, Ganeshs, Bobesh, Sreedevi, Hasitha, Joseph, Suman, Peter, Abik, Poorna, Jose, Omkar, Preethi, Radhika,*** *late* ***Thimmappa*** *and late* ***Srividya*** *for the time they had spent for me and helped me to complete my work.*

I express my thanks to our laboratory attendants and demonstrators ***Ms. Saritha, Ms. Rekha, Mr. Venkat, Mr. Rajesh and Mr. Seenu*** *for all their support.*

I would like to thank my parents ***Mr. A.F. Joseph*** *and* ***Ms. Agnes*** *and my brothers* ***Fr. Rajan, Robert, Vincent*** *and* ***Tilak*** *and their family members and my aunt late* ***Ms. Esther*** *for all their support and prayers.*

Last but most important I would like to thank and deeply appreciate my husband ***Prof. Michael Alphonse*** *for his valuable advice and support and my dear daughters* ***Binu Valancia*** *and* ***Juno Felecia*** *for their deep patience and support.*

Date: **J. Therese Patrisha**

ABSTRACT

In this study we evaluated few novel leads for few protein targets of infectious diseases like hepatitis C virus non-structural 5B polymerase, human immunodeficiency virus protease and N-acetylglucosamine-1-phosphate uridyltransferase of MTB.

We employed computer-aided drug design tools like structure based drug design and ligand based drug design to identify new leads for some of the infectious disease targets like HCV NS5B, HIV protease and MTB GlmU. Strategically two major modeling procedures were employed that included structure based pharmacophore and ligand based 3D QSAR modeling. For the structure based approach we generated energy-based pharmacophore hypotheses based on the crystal structures of HCV NS5B and HIV protease bound to inhibitors and validated using enchrichment calculations. Simultaneously ligand based pharmacophore hypotheses generated and validated and employed for 3D QSAR modeling. Further virtual high-throughput screening and docking was performed to identify new inhibitors for the above targets.

In the present work we selected 5 crystal structures of HCV NS5B polymerase bound with inhibitors and obtained 5 e-phamacophores. In ligand based approach 1568 NS5B inhibitors with wide range of activity 0.04 nM – 500 μM were used to develop best 3D QSAR model with PLS factor five and employed 5 e-pharmacophores and one validated ligand based pharmacophore model and structural diversity for HCV NS5B inhibitors for screening the commercially available Asinex database. Based on docking score, fitness, number of H-bonds, ADME properties, interaction diagram and visual inspection we selected top 10 promising leads for HCV NS5B polymerase. These compounds were procured from Asinex compound database and performed anti-HCV activity and cytotoxicity studies for our 10 leads. The promising two leads **H-5** and **H-6** showed IC_{50} values as **28.8 μM** and **47.3 μM** respectively with inhibition of HCV NS5B RdRp activity as nearly as 67% and 50%. With regard to HIV protease inhibitor design, we selected 8

crystal structures and obtained 12 e-pharmacophores. In the ligand based approach we employed 1535 compounds in the activity range of 0.026 nM to 316 μM and developed a 3D QSAR model with PLS factor five the best statistical model. We employed all 13 pharmacophore for screening the Asinex database and obtained 13 lead compounds. We performed anti-HIV protease inhibition studies for all our 13 leads and found that all compounds showed more than 60% inhibition at 25 μM concentration except the compound **L-4**.

For MTB GlmU, crystal structure with bound inhibitor was not available. Hence we employed ligand based pharmacophore approach with 27 GlmU inhibitors from PubChem Bioassay AID-1376. AAD was chosen as the best pharmacophore and 3D QSAR model was developed and validated for its external predictivity. It showed good r^2_m (LOO), R^2 and r^2_{cv} values which showed that the developed 3D QSAR model was reliable. The highest active reported compound **1** used for 3D QSAR study was then optimized based on contour maps and we designed few compounds. The compounds selection was based on docking them in the acetyl transferase substrate binding pocket and those that exhibited good docking score, fitness and interaction with important amino acid residues and were tested for GlmU acetyltransferase activity. The inhibitors **R, S, T, V** and **W** showed inhibition at 1 μM concentration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LISTS OF ABBREVATIONS

A	Hydrogen bond acceptor
ADME	Adsorption, distribution, metabolism and excretion
ADP	Adenosine di phosphate
AG	Arabinogalactan
AIDS	Acquired immune deficiency syndrome
AP	Allosteric binding pockets
ATP	Adenosine tri phosphate
BEDROC	Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative molecular similarity indices analysis
CYP enzyme	Cytochrome P450
D	Hydrogen bond donor
DABCYL	4-((4-(dimethylamino)phenyl)azo)benzoic acid
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxy ribonucleic acid
DOTS	Directly Observed Treatment, Short-course
DTT	Dithiothreitol
EDANS	(5-((2-Aminoethyl)amino)naphthalene-1-sulfonic acid

EF	Enrichment factor
EID	Emerging infectious diseases
FBS	Fetal bovine serum
FDA	Food and Drug Administration
FRET	Forster resonance energy transfer
FFD	Fractional factorial design
FLDs	First-line-anti-TB drugs
GOLPE	General Optimal Linear PLS Estimation
GH	Goodness of fit
GlmU	N-acetylglucosamine-1-phosphate uridyltransferase
GOLD	Genetically Optimized Ligand Docking
gp	glycoprotein
H	Hydrophobic
HAART	Highly active antiretroviral therapy
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HOMO	Highest occupied molecular orbital
HTVS	High throughput virtual screening
IC_{50}	Half maximal inhibitory concentration
IFN-α	Interferon-α
IPTG	Isopropyl β -D-1-thiogalactopyranoside

IRES	Internal ribosome entry site
kNN	k-nearest neighbors
kNN-MFA	k-nearest neighbors molecular field analysis
LM	Lipomannan
LOO	Leave-one-out
ManLAM	Mannose-capped lipoarabinomannan
MDR-TB	Multi-drug resistant tuberculosis
MDR-HIV	Multi-drug resistant-Human immunodeficiency virus
MLR	Multiple linear regression
m-RNA	Messenger RNA
MTB	*Mycobacterium tuberculosis*
MW	Molecular weight
N	Negative ionizable
NANBH	Non-A, non-B hepatitis
NI	Nucleoside inhibitors
Ni-NTA	Nickel-nitriloacetic acid
NNI	Non-nucleoside inhibitors
NNRTIs	Non-nucleoside reverse transcriptase inhibitors
NRTs	Nucleoside reverse transcriptase inhibitors
NS5B	Non-structural 5B proteins
NTP	Nitriloacetic acid

OECD	Organisation for economic Co-operation and development
OPLS	Optimized potentials for liquid simulations
P	Positive ionizable
PAS	p-Amino salicylic acid
PCA	Principal Component Analysis
PCR	Principal Component Regression
PDB	Protein data bank
PEG-IFNα+RBV	Peginterferon α and ribavirin
PG	Peptidoglycan
PGL	Phenolic glycolipid
PIs	Protease inhibitors
PIMs	Phosphatidyl-myo-inositol mannosides
PKnB	Protein kinase B
PKR	RNA–activated protein kinase
PLS-QSAR	Partial least squares quantitative structure–activity relation
PMSF	Phenylmethylsulfonyl fluoride
QSAR	Quantitative structure activity relationship
R	Aromatic ring
RBV	Ribavirin
RdRp	RNA dependent RNA polymerase

REACH	Registration, Evaluation, and Authorization of Chemicals
RMSD	Root mean square deviation
RMSE	Root mean square error
RIE	Robust Initial Enhancement
RIG	Retinoic-acid
RNA	Ribonucleic acid
ROC	Receiver Operating Characteristic curve value
SAR	Structure-Activity Relationship
SBDD	Structure based drug design
SDEP	Secondary dual educator program
SP	Standard precision
SL	Sulfolipid
SLDs	Second-line-anti-TB drugs
TB	Tuberculosis
TDM	Trehalose dimycolate
TDR-TB	Totally drug resistant tuberculosis
TGs	Triglycerides
TLC	Thin layer chromatography
TLR3	Toll-like receptor 3
TSRI	The Scripps Research Institute
UTP	Glucose-1-phosphate uridyltransferase

VLP	Virus-like particle
WHO	World health organization
XDR-TB	Extremely drug resistant tuberculosis
XP	Extra precision

CHAPTER 1

INTRODUCTION

It is one of the most infamous quotes in the history of biomedicine: "It is time to close the book on infectious diseases, and declare the war against pestilence won." Long attributed to the United States Surgeon General, Dr. William H. Stewart (1965-1969), the statement is frequently used as a foil by scientific and lay authors to underscore the ever-increasing problems of antibiotic-resistant and emerging infections. However, the primary source for the quote has never been identified.

Infectious Disease Poverty. 2013; 2: 3.

Published online Feb 18, 2013.

1.1. INFECTIOUS DISEASES

Infectious diseases are caused by pathogenic microorganisms, such as bacteria, viruses, parasites or fungi; the diseases can be spread, directly or indirectly, from one person to another. Infectious diseases kill more people worldwide than any other single cause. Infectious diseases have for centuries ranked with wars and famine as major challenges to human progress and survival. The last decade has seen a renaissance of advanced low-cost molecular diagnostics to detect and monitor the spread of deadly infectious agents, including drug-resistant variants. A single organism can now be reliably identified from small amounts of human blood, tissue and respiratory specimens, as well as from environmental sources like air and water.

The ongoing explosion of antibiotic-resistant infections continues to plague global health care. Meanwhile, an equally alarming decline has occurred in the research and development of new antibiotics to deal with the threat [1]. Emerging infectious diseases are a significant burden on global economies and public health. Their emergence is thought to be driven largely by socio-economic, environmental and ecological factors [2]. Emerging infectious disease (EID) events have risen significantly over time after controlling for reporting bias, with their peak incidence (in the 1980s) concomitant with the HIV pandemic. EID events are dominated by zoonoses (60.3% of EIDs): the majority of these (71.8%) originate in wildlife (for example, severe acute respiratory virus, Ebola virus), and are increasing significantly over time. Against a constant background of established infections, epidemics of new and old infectious diseases periodically emerge, greatly magnifying the global burden of infections. Studies of these emerging infections reveal the evolutionary properties of pathogenic microorganisms and the dynamic relationships between microorganisms, their hosts and the environment [3]. The need for new antimicrobial agents is greater than ever because of the emergence of multidrug resistance in common pathogens, the

rapid emergence of new infections, and the potential for use of multidrug-resistant agents in bioweapons. Paradoxically, some pharmaceutical companies have indicated that they are curtailing anti-infective research programs. Among the most common infections, we have attempted to study novel protein targets of various infectious diseases and the details of these are summarized in further sections.

1.2. HEPATITIS C VIRAL (HCV) INFECTION

Hepatitis is an inflammatory condition of the liver. The name has its Greek origin. 'Hepat' means liver and 'itis' means inflammation [4]. There are two stages of hepatitis, acute and chronic. The generation of scar tissues on the liver is known as liver cirrhosis and further damage to the liver leads to liver cancer as shown in **Figure 1**.

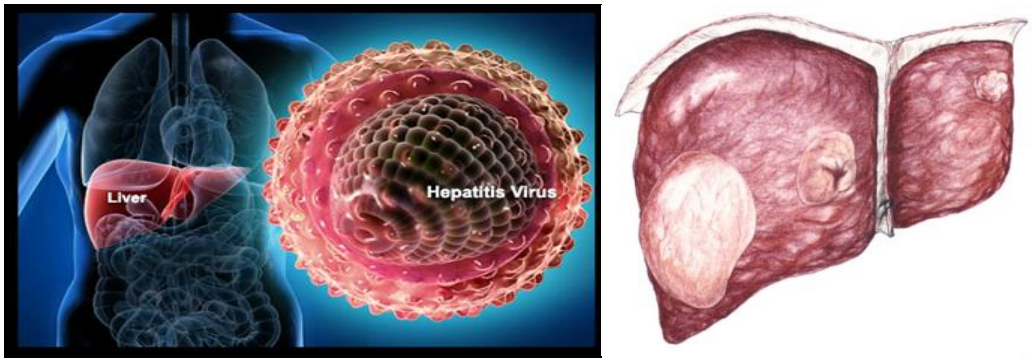


Figure 1: The structure of Hepatitis C Virus and the HCV affected liver [5]

Hepatitis C is a liver disease caused by the hepatitis C virus. The disease can range in severity from a mild illness lasting a few weeks to a serious, lifelong condition that can lead to cirrhosis of the liver or liver cancer. The hepatitis C virus is transmitted through contact with the blood of an infected person. According to WHO report [6] about 150 million people are chronically infected with hepatitis C virus, and more than 350 000 people die every year from hepatitis C-related liver diseases.

Every year around 1 lakh people die due to hepatitis. 1.8-2.5% of total population gets infected by hepatitis. Around 50-60% of drug users are also affected [6]. HCV infection is ten times more contagious than HIV and co-infection with HIV could be lethal. One vial of interferon α costs Rs.23, 100 [7]. The full course of the treatment goes around 48 weeks. It is very difficult for a common man to afford the expenses for the treatment.

1.2.1. *Hepatitis C virus (HCV)*

HCV is a positive, single-stranded RNA virus that contains a 9.6 kb genome that encodes several structural and nonstructural proteins. HCV is a blood-borne pathogen belonging to the Flaviviridae 1 family of viruses; this also includes the West Nile, Yellow Fever, and Dengue viruses. HCV infection is also one of the most significant cause for liver cirrhosis and hepatocellular carcinoma [8] leading to liver failure and as such is a growing medical problem that affects an estimated 200 million individuals worldwide. HCV is a positive strand RNA virus, and its genome comprises of 9600 base pairs [9, 10].

This dreadful pathogen was identified by a team of scientists in 1987, namely Michael Houghton, Qui-Lim Choo and George Kuo at Chiron Corporation in collaboration with Dr. D.W. Bradley. They developed a diagonistic kit and employed a novel cloning technique. This was confirmed by Alter in the year 1988, by confirming its presence in a panel of non-A, non-B hepatitis (NANBH) specimens. In 1989, this work was published in the journal Science in two articles [11, 12]. In 2000, Drs. Alter and Houghton received the Lasker Award for Clinical Medical Research for their discovery.

HCV a positive-strand RNA virus has become a significant global pathogen causing acute hepatitis and chronic liver disease, including cirrhosis and liver cancer. About 200 million people are chronically infected with HCV, of which ~350,000 die every year from hepatitis C-related

liver diseases. Every year, 3-4 million people are newly infected with this virus. In some countries chronic infection rate is more than 5%. The main mode of transmission is through the usage of contaminated unsafe injections [13].

1.2.2. *Life cycle of HCV and interventions*

The life cycle of HCV starts with the entry into the host cell and then there is uncoating of HCV particle nucleocapsid as shown in **Figure 2.**

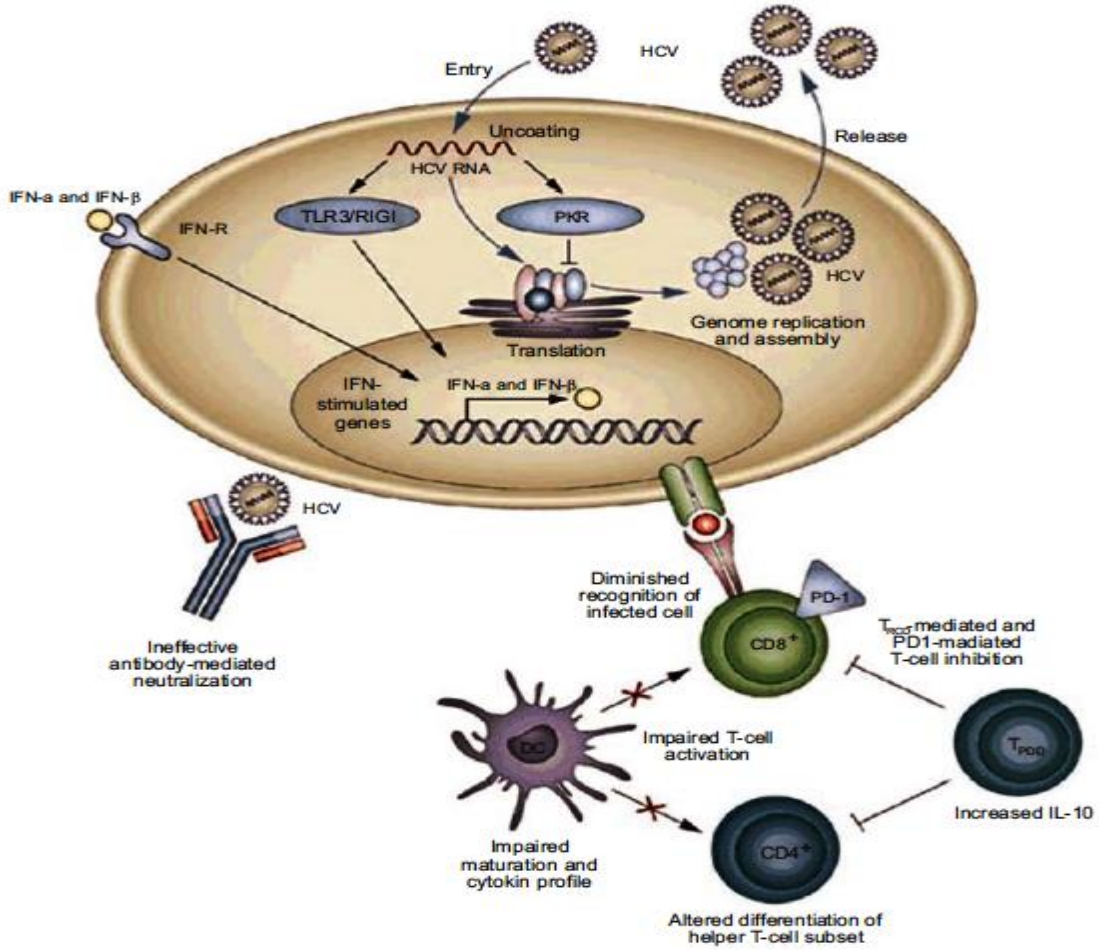


Figure 2: The various stages of HCV [14]

The factors which are present in the host cell that is important for the HCV entry include the tetraspanin, CD81, human scavenger receptor class B type I (SR-BI) and the tight junction proteins, claudin 1 (ClDn1), and occluding [15]. HCV E2 envelope glycoprotein also binds to dendritic-cell-specific and liver-cell-specific intercellular adhesion molecule-3-grabbing

nonintegrins (CD209 and CD209L) [16]. The glycoproteins of HCV envelope, E1 and E2 are crucial for the entry into host cells and they have become an attractive target for neutralizing antibodies [17].

Once the virus enters into the host cell cytoplasm, it undergoes a process of uncoating of HCV particle nucleocapsid and then enters to translation step. The presence of viral, double-stranded RNA in the cytoplasm initiates several antiviral mechanism in the host, like the activation of double-stranded RNA–activated protein kinase (PKR), Toll-like receptor 3 and retinoic-acid-inducible protein 1 (TLR3 and RIG-1), which helps in the release of Interferonα (IFN-α) and Interferon β (IFN- β). It uses internal ribosome entry site (IRES)-mediated translation mechanism to access the viral translation. Host cells would develop few mechanisms to stop the use of its own protein-translation mechanism by the viral particle. A single ~3,000 amino-acid polyprotein produced by the translation of HCV genome; gets further cleaved into at least 10 different protein products that include the structural proteins, the viral particle (core and envelope proteins E1 and E2) and non-structural proteins like P7, NS3, NS4A, NS4B, NS4A and NS5B [18] as shown in **Figure 3.** In association with intracellular membranes, replication of viral RNA happens. But the exact mechanism of HCV-RNA replication is not yet very clear. But the important protein which is responsible for viral RNA replication is the HCV-NS5B protein, which is a catalytic subunit of the replication complex and possesses the RNA dependent RNA polymerase (RdRp) activity [19].

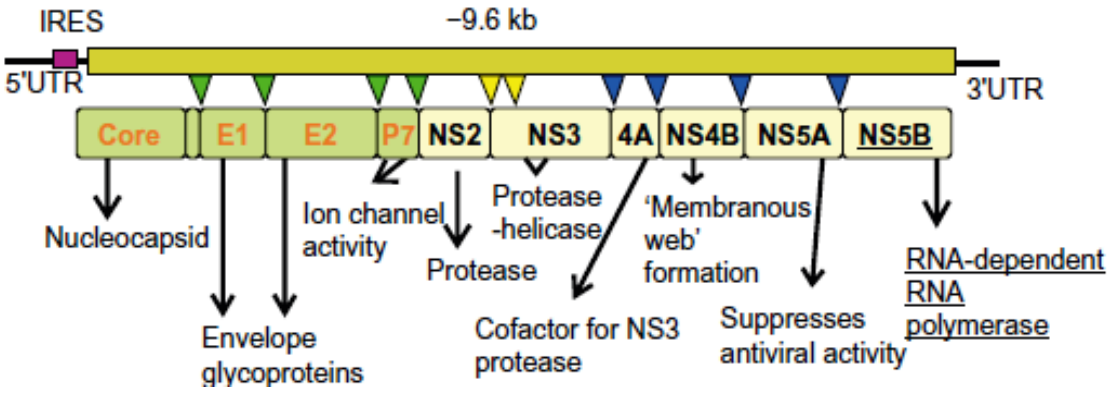


Figure 3: The organization of the HCV genome and the scheme for HCV polyprotein cleavage [20]

Proof reading function is absent in HCV RdRp due to this it generates genetically diversed genome population within the infected person known as a quasispecies. The replication rate of HCV is quite high in the range of 10^{10} to 10^{12} virions per day and their predicted half-life is 2 to 3 hours. Secreted HCV particles are low in density, which combine with lipoproteins for the viral release. This combination prevents the viral particle from the attack of host immune system [18]. There are few promising drug development targets for HCVs that include NS2-NS3 autoprotease, NS3 protease, NS3 helicase, NS4A cofactor, NS5B polymerase and envelope proteins as seen in **Figure 3.**

1.2.3. *Treatment of HCV infection*

There are 6 genotypes of the hepatitis C virus and they may respond differently to treatment. Careful screening is necessary before starting the treatment to determine the most appropriate approach for the patient.

The available standard combination therapy for HCV infection is with peginterferon α and β (PEG-IFN plus RBV). More than 50% of patients with genotype 1 are non-responder to standard combination therapy [21]. Treatment with interferon alone has effectiveness in 10% to 20% patients but combination therapy shows effectiveness in 30% to 50% patients. Hence uses of RBV

alone are not so effective. Non-responders to PEG-IFN plus RBV are a major challenge because these patients develop lot of complications that are lethal [22].

Combination antiviral therapy with interferon and ribavirin has been the main stay of hepatitis C treatment. Unfortunately, interferon is not widely available globally, it is not always well tolerated, some virus genotypes respond better to interferon than others, and many people who take interferon do not finish their treatment. This means that while hepatitis C is generally considered to be a curable disease, for many people this is not a reality [23].

Ribavirin and its liver targeting prodrug viramidine show favorable pharmacokinetic properties and viramidine [24] showed very good antiviral activity. Scientific advances have led to the development of new antiviral drugs for hepatitis C, which may be more effective and better, tolerated than existing therapies. Two new therapeutic agents, telaprevir and boceprevir, have recently been licensed in some countries. Much needs to be done to ensure that these advances lead to greater access and treatment globally.

Telaprevir, is a pharmaceutical drug for the treatment of hepatitis C co-developed by Vertex Pharmaceuticals and Johnson & Johnson. It is a member of a class of antiviral drugs known as protease inhibitors [25]. Specifically, telaprevir inhibits the hepatitis C viral enzyme NS4A serine protease. Telaprevir is only indicated for use against hepatitis C genotype 1 viral infections and has not been proven to have an effect on or being safe when used for other genotypes of the virus. The standard therapy of pegylated interferon and ribavirin is less effective on genotype 1. The cost of telaprevir-based triple therapy for hepatitis C is $189,000 per sustained viral response. The most common adverse effect is rash. Grade 3 adverse events (mainly anemia and leukopenia/neutropenia) were more frequent in the telaprevir groups than in the control group

(37% vs. 22%). Boceprevir is a protease inhibitor used as a treatment for hepatitis C genotype 1. It binds to HCV nonstructural 3NS3 (HCV) active site [26].

1.2.4. *Hepatitis C vaccine*

Although vaccines exist for hepatitis A and hepatitis B, development of a hepatitis C vaccine has presented challenges. Over the last decade numerous HCV vaccine approaches have been assessed in mice and primates. Only a small fraction of animal HCV vaccine studies have progressed to human trials. Most vaccines work through inducing an antibody response that targets the outer surfaces of viruses. However the Hepatitis C virus is highly variable among strains and fast mutated, making an effective vaccine very difficult. The detailed structure of E2 envelope glycoprotein, believed to be the key protein the virus uses to invade liver cells, is elucidated by scientists at The Scripps Research Institute (TSRI) in November 2013 [27].

In recent years many vaccine strategies have shown promising results during clinical trials. Four important strategies studied include recombinant protein, peptide, DNA, and vector vaccines. Some of these technologies have generated robust antiviral immunity in healthy volunteers and infected patients. Novel future vaccine approaches include virus-like particle (VLP)-based vaccines that have been successfully employed for viral infections such as hepatitis B. Additional strategies include molecules that induce innate immune responses, with secondary effects on adaptive responses (such as TLR-9 ligands) that are either encoded within a vaccine construct or used as a vaccine adjuvant [28]. Recently vaccines called IC41, Chron Vac-C, and Isomatrix have reached the final stages of human clinical trial.

1.3. HUMAN IMMUNODEFICIENCY VIRAL (HIV) INFECTION

Human immunodeficiency virus (HIV) is a lentivirus (a member of the retrovirus family) that causes acquired immunodeficiency syndrome (AIDS) [29, 30] a condition in humans in which

progressive failure of the immune system allows life-threatening opportunistic infections and cancers to thrive. Infection with HIV occurs by transfer of blood, semen, vaginal fluid or breast milk. Within these bodily fluids, HIV is present as both free virus particles and virus within infected immune cells. HIV infection in humans is considered pandemic by the World Health Organization (WHO). Nevertheless, complacency about HIV may play a key role in HIV risk. From its discovery in 1981 till 2006, AIDS has killed more than 25 million people. For every one person around the world getting treatment, there are 2 new reported cases of HIV infection records and currently there are 16 million orphans in the world those who have lost either one or both parents because of AIDS. Due to global AIDS awareness programs and increase in scientific research and drug treatment programs, the number of new HIV infections dropped 20% over the past decade [31].

1.3.1. *Human immunodeficiency virus (HIV)*

HIV, the virus being a member of Lentiviruses, use the most effective replication method and transmit their genetic material into host cell through their deoxyribonucleic acid (DNA) and replicate in non-dividing cells. The peculiar character of reteroviruses is that they can transform ribonucleic acid (RNA) into DNA which is unnatural of the process as RNA is from DNA. Reverse transcriptase, a DNA polymerase of reteroviruses is responsible for the transcription of single-stranded RNA into single-stranded DNA and hence HIV is also termed as a retrovirus. [32, 33].

HIV is of 2 types, HIV-1 and HIV-2 with HIV-1 being the major cause for worldwide infections. There are few cases in or from West Africa with HIV-2 infections. HIV-1 can be divided into 4 viral groups: M, N, O and P. HIV-1 from group M plays a major role in global pandemic among the various subtypes and recombinants [34].

HIV consists of two single chains of RNA and infects mainly the CD4+ lymphocytes (T cells). These infected cells lose their original function in human immuno system and start replicating HIV cells. The HIV viral particle is spherical in shape with a diameter of 1/10,000 mm as shown in **Figure 4.** Viral envelope consists of two lipid layers and proteins. The outer spikes consists of

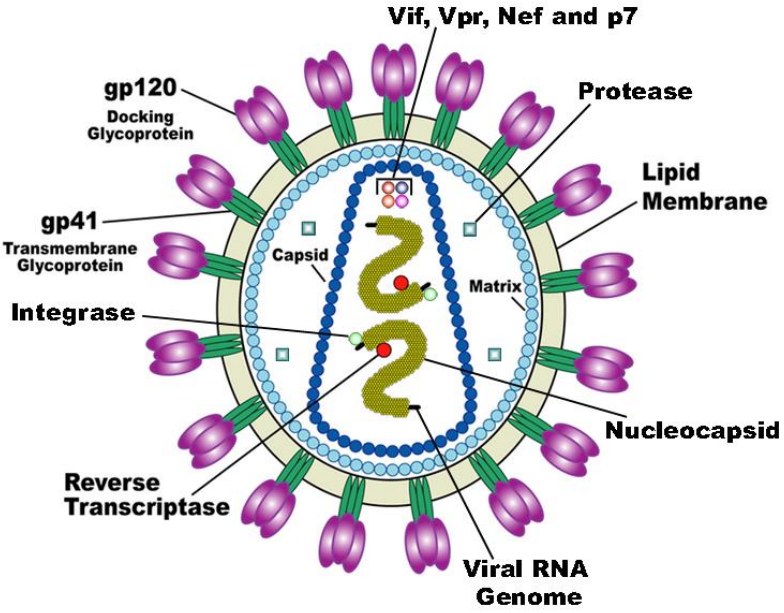


Figure 4: Structure of Human Immunodeficiency Virus [35]

glycoprotein (gp) 120 and the transmembrane has gp41. The HIV matrix proteins consist p17 proteins present between the envelope and the core. The viral core consists of viral capsule protein p24 that sorrounds two single-stranded HIV RNA and the enzymes needed for viral replication. The glycoproteins 120 and 41 (gp120 and gp41) on the surface plays vital role in fusion with the immune cells of host cells to form glycoprotein complex, which is the initial stage of infection. Protective protein sheath or capsid, which is present inside the virus with viral RNA core and enzymes is surrounded by the viral envelope composed of phospholipids [36].

1.3.2. *Life cycle of HIV*

There are eight stages of HIV life cycle as shown in **Figure 5** and it is a complex process.

I. Attachment: HIV life cycle begins with conformational change of gp120 envelope protein, which allows binding with the co-receptors on the T cell's surface to obtain entry and expose gp41 molecules for fusion of the lipid bilayers of the virus and the T cells [37].

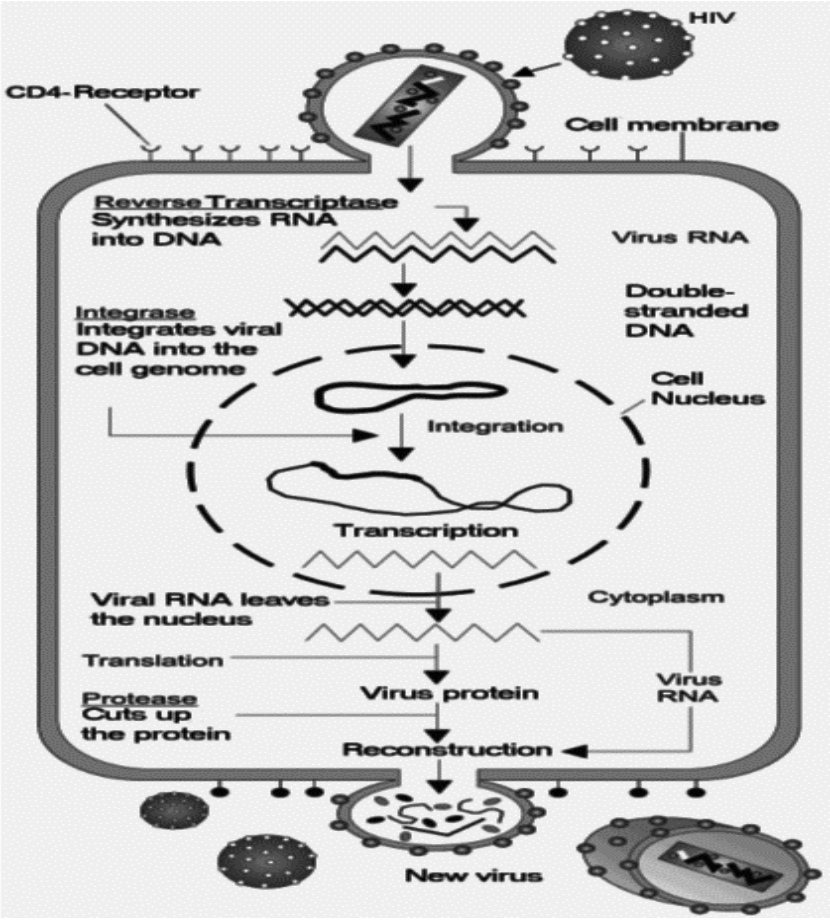


Figure 5: A representation of HIV life cycle [38]

II. Viral entry and Reverse Transcription: HIV penetration into the T cells occurs and the matrix with capsid protein is mixed together which allows the release of viral enzymes and RNA into the cytoplasm of the cell. With the help of host nucleotides, the HIV enzyme

reverse transcriptase transforms the viral RNA into single-stranded DNA which then converts single-stranded DNA into double-stranded DNA [39].

III. Integration: The HIV enzyme integrase grasps the viral double-stranded DNA and carries it through the nuclear pore into host cell's nucleus. Then it integrates with the host chromosome to insert the HIV DNA [40].

IV. Transcription: Viral RNA is produced during this process and acts as a template for new viruses to be produced *via* ribosomes (mRNA). The enzyme RNA polymerase produces mRNA that encodes for various viral proteins like envelope proteins. The viral proteins embedded in the cell membrane are merged with other newly formed envelope proteins to form cluster on the surface of the infected cell. These proteins exist as multi-protein chains and transformed to the surface with a strand of RNA and some of the enzymes [41].

V. HIV morphogenesis: HIV DNA insertion leads to the production of HIV protease enzyme which plays a vital role in later stages of HIV morphogenesis. Protease enzymes cut the HIV protein chain into individual small chains and a new working virus is produced [42].

There are three final stages of HIV life cycle; that consists of assembly, budding and maturation. These are carried out by gag polyproteins in combination with proteolytic maturation products. During assembly, the viral proteins collected at the plasma membrane are transported to the cell's surface as an irregular form of virus and it involves not only viral and cellular proteins but also lipids and RNAs. Finally breakage of immature virus on the surface of the host cell leads to budding and mature virions [43].

1.3.3. *Treatment*

There are six major types of drugs based on their activity towards HIV/AIDS. As HIV life cycle has many stages, each stage of its life can be inhibited by drugs.

The replicative cycle of HIV comprises a number of steps, which could be considered as adequate targets for chemotherapeutic intervention. The important targets are viral adsorption, virus-cell fusion, virion uncoating, reverse transcription (RNA $\rightarrow$ ds DNA) [RT enzyme], proviral DNA integration, viral transcription (DNA $\rightarrow$ RNA), viral translation (mRNA $\rightarrow$ Protein), viral budding (assembly/release) and maturation (protease and glucosidase enzymes). Most of the substances which have been identified as anti-HIV agents can be allocated to one of the ten classes of HIV inhibitors [44], according to the stage at which they interfere with the HIV replicative cycle (**Table 1**).

Table 1: Anti-HIV agents

S.No.	Types	Inhibitors
1.	Adsorption Inhibitors	a) Polysulfate- Dextran sulfate, Curdlan sulfate, Pentosan polysulfate
		b) Polysulfonates- Suramin, Evans blue
		c) Polycarboxylate- Aurin tricarboxylic acid
		d) Glycyrrhizin
2.	Fusion Inhibitors	Betulinic acid, Mannose-specific plant lectinoylsulfate
3.	Virus uncoating Inhibitors	Bicyclam derivatives
4.	Reverse transcription Inhibitors	a) Nucleoside derivatives- Zidovudine, Stavudine, Lamivudine- Zalcitabine, Didanosine, Abacavir
		b) Non-nucleoside derivatives- Nevirapine, Delavirdine, Efavirenz Loviride, Trovirdine, Emivirine
5.	Integration Inhibitors	Curcumin, L-chicoric acid
6.	DNA replication Inhibitors	Antisense constructs
7.	Transcription Inhibitors	1,4-Benzodiazepine and Fluoroquinolone derivatives
8.	Translation Inhibitors	Ribozymes, Trichosanthin
9.	Maturation Inhibitors	a) Protease inhibitors – Saquinavir, Indinavir, Ritonavir, Nelfinavir, Lopinavir, Amprenavir, Atazanavir, Tipranavir, Darunavir, Amprenavir, Lopinavir
		b) Glucosidase inhibitors – Castanospermine
10.	Budding Inhibitors	Interferon, Hypericin

HAART Therapy:

HAART is nothing but the combinations of two or more of the different class of inhibitors (drugs) into a single product to prevent or stop HIV functions in host cells and is also known as multi-class combination products. The treatment for the patients those who are resistant to one type of

antiretroviral drug can take a combination of antiretroviral drugs to get highly effective therapy [45]. The combinations of drug commonly employed are given in **Table 2**.

Table 2: Multi-class combination products for HAART therapy

Brand Name	**Generic Name**	**Manufacturer Name**
Atripla	Efavirenz, Emtricitabine and Tenofovir disoproxil fumarate	Bristol-Myers Squibb and Gilead Sciences
Complera	Emtricitabine, Rilpivirine, and Tenofovir disoproxil fumarate	Gilead Sciences
Stribild	Elvitegravir, Cobicistat, Emtricitabine, Tenofovir disoproxil fumarate	Gilead Sciences

1.4. TUBERCULOSIS DISEASE

Tuberculosis (TB) is a lethal, infectious disease caused by various strains of mycobacteria, majorly *Mycobacterium tuberculosis* (MTB). Most infections are asymptomatic and latent, but about one in ten latent infections eventually progresses to active form [46]. One third of the world's population is thought to have been infected with MTB, with new infections cropping in about 1% of the population each year. The distribution of tuberculosis is not uniform across the globe; about 80% of the population in many Asian and African countries tests positive in tuberculin tests, while only 5–10% of the United States is affected. The risk of TB disease and death is highly increased by the usage of tobacco [47].

Tuberculosis (TB) is the second greatest killer globally due to a single infectious agent. In 2012, 8.6 million people were affected by this infectious agent, among which about 1.3 million people died. More than 95% of TB deaths occur in low and middle income countries and in these countries, it is one of the top three causes of deaths among women between ages 15 to 44. Ninety-five per cent of all cases and 99 % of deaths occurred in developing countries, because of compromised immunity, largely due to high rates of HIV infection and the corresponding development of AIDS and with the greatest burden in sub-Saharan Africa and South East Asia. In 2013, totally 530,000 children (0-14 years) were affected by TB and 74,000 children with HIV-negative died because of TB [47].

People getting affected every year are on a decrease which shows that the world is on track to achieve the Millennium Development Goal to reverse the spread of TB by 2015. The TB death rate reduced to 45% between 1990 and 2012. Nearly 22 million lives were saved through the use of DOTS and because of the strategy Stop TB recommended by WHO. In 2013, about 60% of new TB cases came from the region of Asia. But sub-Saharan Africa showed the highest proportion of 255 new cases for TB per population in the year 2012. In 2013, 80% of the TB cases reported from 22 countries as shown in **Figure 6**. Among these 22 countries, Brazil and China that showed a sustained decline in TB cases over past 20 years [48].

Multi-drug-resistant TB (MDR-TB) is a form of TB caused by the bacteria which is resistant to the most of the available first line anti-TB medicines like isoniazid and rifampicin. This is because of the improper treatment, inappropriate use of anti-TB drugs, improper knowledge about the disease and low quality of medicines.

They are treatable only with the second-line anti-TB drugs like fluroquinolones, aminoglycosides, etc. Extensive chemotherapy for more than two years give a complete cure to the patients, but the treatment is bit costly and lead to lots of harmful side effects [50].

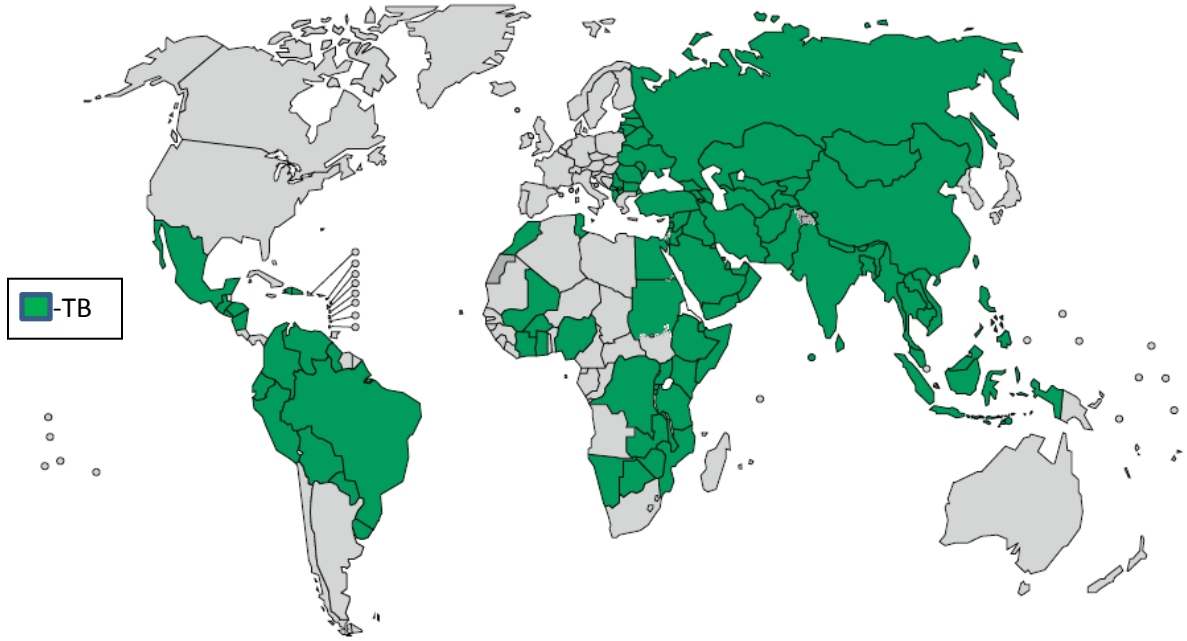


Figure 6: Countries with TB disease burden 2008-2013 [49]

More severe drug resistance lead to extensively drug- resistant TB (XDR-TB), a form of multi-drug resistant tuberculosis that responds to very less number of available the most effective second-line anti-TB drugs. Nearly 450,000 people were identified with MDR-TB worldwide in 2012, among these 50% were reported from India, China and Russian Federation (**Figure 7).** About 9.6 % of all MDR-TB cases had XDR-TB.

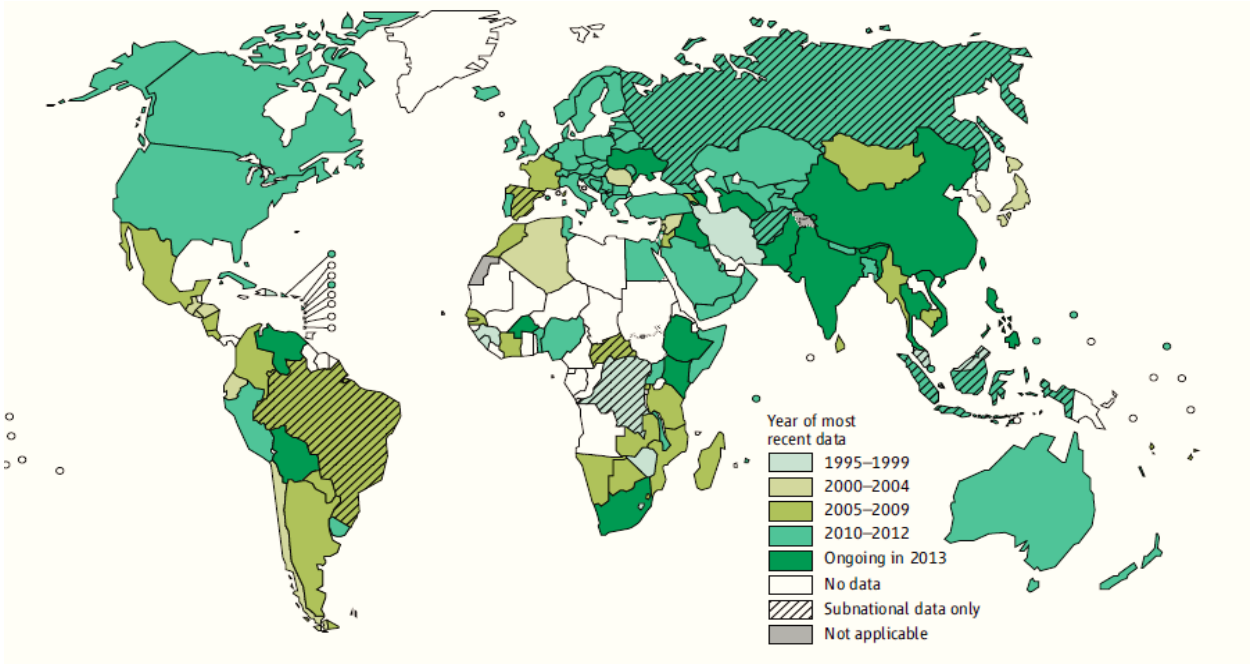


Figure 7: Increasing number of drug resistance cases globally, 1994-2013 [51]

Earlier patients were reported to be totally drug resistant (TDR-TB) and they had resistance to all first-line-anti-TB drugs (FLD) and the second-line-anti-TB drugs (SLD) for which they were tested. Furthur 15 more people were reported for the TDR-TB who showed resistance to all the available anti-TB drugs tested. Later 4 patients with TDR-TB were reported from India with subsequent media reports of further 8 patients. This term TDR-TB is not yet recognized by WHO. The term “totally drug resistant” is not clearly defined till date [52].

1.4.1. *Mycobacteria*

MTB then known as the tubercle bacillus, was first discovered on 24 March 1882 by Robert Koch, who subsequently received the Nobel Prize in physiology or medicine for this discovery in 1905; the bacterium is also known as Koch's bacillus [53]. MTB is pathogenic bacterial species in the genus Mycobacterium and the causative agent of most cases of tuberculosis. MTB has an unusual, waxy coating on the cell surface (primarily mycolic acid), which makes the cells impervious to Gram staining so acid-fast detection techniques are used instead. The physiology of

MTB is highly aerobic and requires high levels of oxygen. Primarily a pathogen of the mammalian respiratory system, MTB infects the lungs and is the causative agent of tuberculosis. The most frequently used diagnostic methods for TB are the tuberculin skin test, acid-fast stain, and chest radiographs.

The life cycle of MTB infection is depicted as in the **Figure 8**. The infection starts when MTB bacilli, present in droplets are inhaled and are phagocytosed by resident alveolar macrophages. This results in proinflammatory response and invasion of epithelium which leads to the recruitment of monocytes from the circulation, as well as extensive neovascularization of the infection site. The macrophages in the granulomas then gets differentiated to epithelioid cells, multinucleate giant cells, and foam cells which are filled with lipid droplets now called as granuloma. Finally the virulent bacilli are released into the airways when the granuloma collapses into the lungs.

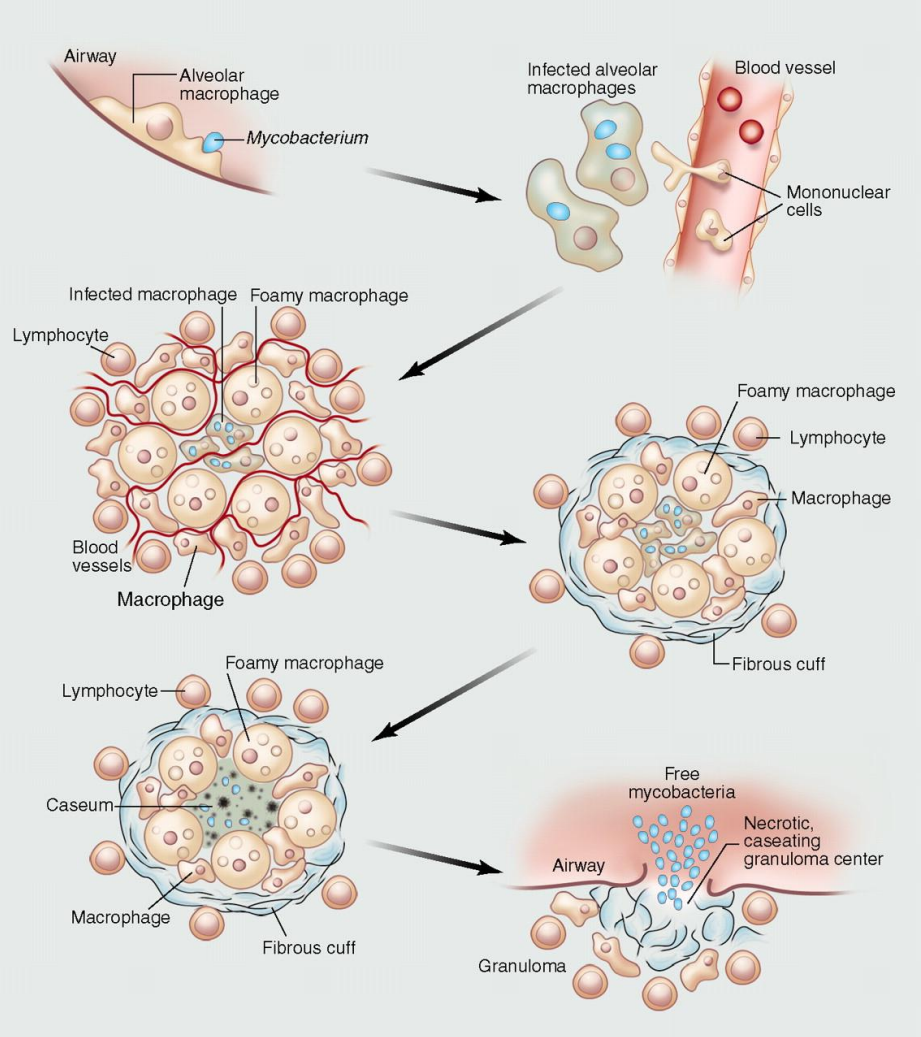


Figure 8: Life cycle of MTB [54]

1.4.1.1. *The mannosylated cell envelope components of MTB*

The MTB cell envelope is characterized by the presence of a variety of unique complex lipids, constituting 60% of the bacillus total weight. This lipid-rich low permeability matrix contributes to the difficulty in combating mycobacterial diseases by endowing the organism with innate resistance to therapeutic agents and host defenses. The complex MTB cell envelope can be divided into two major structures, the cell wall and the capsule-like outermost structures (**Figure 9)**. The outermost components are solvent-extractable non-covalently bound free lipids, carbohydrates and proteins associated with the mycolyl-arabinogalactanepeptidoglycan complex

(cell wall core). These surface components may be prone to release, shedding, and/ or cleavage upon contact with the host cell or within an appropriate intracellular environment of the cell [55]. The surface of MTB is particularly rich in mannose-containing biomolecules, including mannose-capped lipoarabinomannan (ManLAM), the related lipomannan (LM), phosphatidyl-myo-inositol mannosides (PIMs), arabinomannan, mannan and manno-glycoproteins PIMs, LM and ManLAM are incorporated into the plasmamembrane and also exposed on the MTB cell surface. They act as ligands for host cell receptors and contribute to the pathogenesis of MTB.

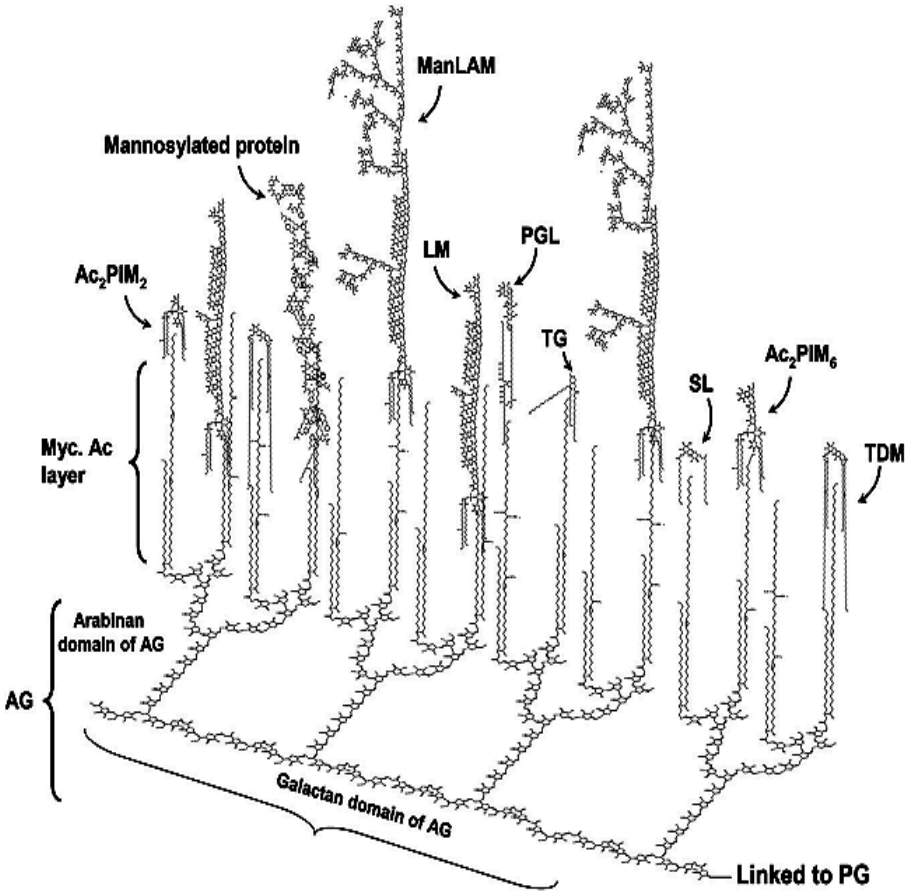


Figure 9: The cell envelope of MTB with an emphasis on exposed mannosylated cell envelope components. This scheme depicts the cell envelope "skeleton or core" determinants (mycolyl-arabinogalactanepeptidoglycan complex) and emphasizes the distribution of intercalated major mannosylated cell envelope components that are exposed on the MTB surface. AG is covalently linked to PG via the galactan chain and the arabinan chain is in turn linked to the mycolic acids

(Myc Ac) which are shown perpendicular to the plasma membrane. The polar groups (i.e. carbohydrate domains) of several mannosylated cell envelope components are exposed on the cell surface and their lipid domains are intercalated with the Myc Ac acid layer. These envelope components include ManLAM, LM, higher- and lower-order PIMs, and lipomannoproteins. Other known virulence factors described for MTB that interact with the Myc Ac layer [i.e. TDM, SL; and TGs and PGL, the latter on some MTB strains)] are also depicted. Not all Myc Ac are depicted interacting with cell surface components. Not shown are capsule-like components (i.e. arabinomannan, glucan, mannan, and xylan). In order to maintain simplicity, molecular quantities depicted (relative number of molecules) do not accurately reflect experimental data. AG (arabinogalactan); PG (peptidoglycan); Myc Ac (mycolicacids); ManLAM (mannose-capped lipoarabinomannan); LM (lipomannan); PIMs (phosphatidyl-myo-inositol mannosides); TDM (trehalose dimycolate); SL (sulfolipid); TGs (triglycerides); PGL (phenolic glycolipid). [56]

Unlike some fast growing mycobacteria and other actinomycetales, MTB is rarely pleomorphic; it does not elongate into filaments, and does not branch in chains when observed in clinical specimens or culture. When numerous and actively multiplying, the bacilli are strongly acid fast and show an evident and distinctive tendency to form hydrophobic bundles. Free bacilli can also be seen, though, especially at the border of the swarms. In unlysed host tissue, the bacilli are more numerous within the phagocytic cells. Once the disease has been controlled, dying bacilli become sparser, often faintly and unevenly colored, due to partial loss of the internal contents. Of course, irregular staining may also be the consequence of technical defectiveness of dyes or staining procedures.

1.4.2. *TB treatment*

The standard "short" course treatment for TB is isoniazid, rifampicin (also known as rifampin in the United States), pyrazinamide, and ethambutol for two months, then isoniazid and rifampicin alone for a further four months. The patient is considered cured at six months (although there is still a relapse rate of 2 to 3%). For latent tuberculosis, the standard treatment is six to nine months of isoniazid alone.

If the organism is known to be fully sensitive, then treatment is with isoniazid, rifampicin, and pyrazinamide for two months, followed by isoniazid and rifampicin for four months. Ethambutol need not be used.

First Line: All first-line anti-tuberculous drug names have a standard three-letter and a single-letter abbreviation as shown in **Table 3**.

Second Line: There are six classes of second-line drugs (SLDs) used for the treatment of TB. A drug may be classed as second-line instead of first-line for one of three possible reasons: it may be less effective than the first-line drugs (e.g., p-Aminosalicylic acid); or, it may have toxic side-effects (e.g., cycloserine); or it may be unavailable in many developing countries (e.g., fluoroquinolones).

The treatment for MDR-TB patients includes an intensive phase of 8 months treatment and for patient newly diagnosed with MDR-TB, total treatment duration of 20 months is recommended for most of the patients and the duration may be modified based on the patient's response rate. For HIV-positive patients with drug-resistance TB, antiretroviral therapy is highly recommended along with second-line-anti TB drugs, irrespective of their CD4 cell-count, within the first 8 weeks followed by anti-TB treatment as shown in **Table 3**.

Third Line: Other drugs that may be useful, but are not on the WHO list of SLDs. These drugs may be considered "third-line drugs" and are listed here in **Table 3** either because they are not very effective (e.g., clarithromycin) or because their efficacy has not been proven and their toxicity and cost are high. Rifabutin is effective, but is not included on the WHO list because for most developing countries, it is impractically expensive.

Adjuvant therapy helps the patients who are getting intensive treatment with SLDs and other third line drugs to avoid the serious side effects.

Table 3: Available anti-TB drugs

TB treatment	Drugs	
First Line	a) Ethambutol EMB or E,	
	b) Isoniazid INH or H,	
	c) Pyrazinamide PZA or Z,	
	d) Rifampicin RMP or R,	
	e) Streptomycin STM or S.	
Second Line	Second-line parenteral agent: Injectable anti-TB drugs) Aminoglycosides:	amikacin (Amk), kanamycin (Km), capreomycin (Cm);
	Fluoroquinolones:	levofloxacin (lfx), moxifloxacin (Mxf) gatifloxacin (Gfx) ofloxacin (Ofx)
	Oral bacteriostatic SLDs: thioamides:	ethionamide (Eto) prothionamide (Pto) cycloserine (Cs) terizidone (Trd) p-aminosalicylic acid (PAS)
	Group 5 drugs (Agents with unclear role in DR-TB treatment (not recommended by WHO for routine use in DR-TB patients)	clofazimine (Cfz) linezoloid (Lzd) amoxicillin/clavulanate (Amx/Clv) thioacetazone (Thz) clarithromycin (Clr) imipenem (Ipm)
Third Line	a) Rifabutin	b) Vitamin D
	c) Macrolides: e.g., clarithromycin (CLR)	d) Imipenen/cilstatin
	e) Linezolid (LZD)	f) High-dose isoniazid
	g) Thioacetazone (T)	h) Arginine
	i) Thioridazine	j) R207910

Vitamin B6 (pyridoxine): It is used to prevent neurological side effects for the patients who are using cycloserine. The standard dose is 50mg for every 250mg of cycloserine. Multivitamins and

minerals supplements (zinc, iron, calcium, etc.) should be given and they should be avoided when fluoroquinolones are used because they can interfere with its absorbtion. The standard dose is 50mg for every 250mg of cycloserine. Multivitamins and minerals supplements (zinc, iron, calcium, etc.) should be given and they should be avoided when fluoroquinolones are used because they can interfere with its absorbtion.

Gastric protection with H_2 blockers (e.g. Ranitidine 150 mg once daily); proton pump inhibitors (e.g. omeprazole 20mg once daily), it should be avoided for patients prescribed with antitubercular therapy due to drug-drug interaction. Corticosteroids can be beneficial in conditions such as severe respiratory insufficiency and central nervous system or pericardial involvement. To get immediate response it is advisable to use injectable corticosteroids [50-52].

1.5. STRATEGIES OF DRUG DESIGN

Based on biological target knowledge the inventions of new therapeutics begin and become the crucial starting point for every drug design. A drug is nothing but a small organic compound which either activates or inhibits the activity of macromolecule like protein resulting in therapeutic advantage to the patients. Drug design depends on computational modelling tools known as computer-aided drug design. There are two strategies in computer drug design:

1.5.1. *Structure based drug design:*

Based on the knowledge of the interacting pattern of the bound inhibitor in the crystal structure of protein, the drugs are designed [57]. If the crystal structures are not available, using homology modeling tools we can design the crystal structures. Molecular dynamics studies are used to predict the drug with high affinity and high selectivity. Docking studies used to screen large compound libraries and databases to identify the potent lead. The structure based strategy has proved to be successful over random screening of existing chemical libraries.

The first success of structure based drug design was the identification of peptide-based HIV protease inhibitor [58]. As in the recent years, there is a rise in the structural information on protein targets; structure based drug design methods have potential advantages over the conventional drug discovery methods. Identification of target active site plays an important role in structure based design protocols. The ligand binding site can be the active site, as in an enzyme, an assembly site with another macromolecule, or a communication site necessary in mechanism of the molecule.Various strategies involved in structure based drug design include the following but not limited to,

i. Fragment-based lead discovery
ii. Virtual screening
iii. Receptor-based pharmacophore modeling
iv. De nova design
v. Homology modeling and molecular dynamics

Over the 35 years that have followed the first published work describing SBDD in 1976, computer aided molecular design and SBDD has surmounted several hurdles and has played a key role in the development of several marketed drugs [59].

1.5.2. *Ligand based drug design:*

Based on the knowledge of already reported drugs for that particular biological target new drugs are designed. The already reported molecules can be used to design the pharmacophore. Pharmacophore is nothing but a model which represents the minimum necessary features of the molecule should possess to bind to the target. This pharmacophore model is used to screen large database compounds to find the potent lead [60].

Pharmacophore approaches is vital tool for accelerating discovery efforts when more extensive data are available by providing a means of superimposing structures for 3D quantitative structure/activity relationship (QSAR) development, or by acting as a rapid prefilter on real or virtual libraries that are too large for routine treatment with more expensive structure-based techniques, such as docking. Ligand-based design generally utilizes pharmacophore modeling for virtual screening of libraries irrespective of the availability of the target structure. Pharmacophore models define important functional groups involved in binding, and the relative positions in 3-dimensional space which implicates the conformational requirement of a ligand. In the following sections, the application of pharmacophore in 3D-QSAR modeling has been described in detail.

1.5.2.1. *Quantitative structure-activity relationship (QSAR)*

Quantitative structure-activity relationships (QSAR) are statistically derived models that can be used to predict the physicochemical and biological (including toxicological) properties of molecules from the knowledge of chemical structure.

The description of QSAR models has been a topic for scientific research for more than 40 years and a successful topic within the framework of regulatory for more than 20 years. The main aim of QSAR field is to examine the relationships using mathematical models which further validate and predict the model statistically. QSARs have many applications in diverged fields like drug discovery and lead optimization, risk assessment and toxicity prediction, regulatory decisions and agrochemicals. One of the major applications of QSAR model is to predict the biological activity of untested compounds from their molecular structures. The estimation of accuracy of prediction is a critical problem in QSAR modeling. With the advent of molecular modelling, three-dimensional (3D) descriptors have replaced the traditional physicochemical and bi-dimensional descriptors [61]. Rigorous analysis and fine-tuning of independent variables has led to an

expansion in development of molecular and atom-based descriptors, as well as descriptors derived from quantum chemical calculations and spectroscopy.

Based on the statistical method which we employ to correlate the structure with activity, QSAR techniques can be divided into two types:

Linear method: Linear QSAR methods include analysis of linear regression, multiple linear regression (MLR), partial-least square (PLS), principal component analysis (PCA) and principal component regression (PCR).

Non-Linear method: Non-linear methods include *k*-nearest neighbors (*k*NN), artificial neural networks, Bayesian neural nets.

1.5.2.2. *3D QSAR*

3D QSAR techniques are the most powerful computational means to support the chemistry side of drug design projects. The primary aim of these techniques is to establish a correlation of the biological activities of a group of structurally and biologically characterized compound with a spatial finger prints of numerous field properties of each molecule, such as steric, lipophilicity and electrostatic interactions. Especially, 3D QSAR study allow to identify the pharmacophoric arrangement of molecular fragments in space and provides guidelines for the design of the next generation of compounds with enhanced biological potencies [61].

1.5.2.3. *CoMFA method*

The CoMFA (Comparative Molecular Field Analysis) method was developed as a tool to study 3D QSAR. A CoMFA analysis begins with a traditional pharmacophore modeling to suggest a bioactive conformation for each molecule and to superimpose the molecules under study. Steric and electrostatic fields provide the information about the biological properties for a set of compounds under study. The steric potential, expressed in a Leonard-Jones function and

electrostatic potential expressed in a simple Coulomb function are the two potentials used in CoMFA study. The standard application of CoMFA provides only enthalpic contributions of the free energy of binding; however, one should find the importance of other contributions from hydrophobic and entropic interactions and to judge whether CoMFA will be able find results under these extraordinary conditions [62].

1.5.2.3. *CoMSIA method*

CoMSIA (Comparative Molecular Similarity Indices Analysis) model overcomes the problems associated with the functional form of the Lennard-Jones potentials used in most of the CoMFA methods. This is based on similarity indices similar to CoMFA and was developed by Klebe *et al* [63]. This method adopted Gaussian type functions instead of traditional CoMFA potentials. The descriptors used in this approach are spatial similarity and dissimilarity of the molecules. Three different indices related to steric, electrostatic and hydrophobic potentials were used in their study of the classical Tripos steroid benchmark dataset. Models of comparable statistical quality with respect to both internal cross-validations of the training set and predictivity of the test set were derived using the CoMSIA method. The CoMSIA contour maps are easily interpretable compared to CoMFA maps. The CoMSIA approach also avoids the cutoff values for the potential functions and it also included the hydrogen bond descriptors to evaluate hydrogen bonds [63].

1.5.2.3. *GRID and GOLPE method*

It is an alternative to the original COMFA method and used by many researches to calculate the interaction fields. The reduced number of potential functions (6-4) compared to Lennard-Jones potential (6-12) in COMFA, for the calculation of interaction energies at the grid points. A statistical method GOLPE (General Optimal Linear PLS Estimation) was developed by Baroni *et al.* to improve the predictivity of QSAR models. GRID force field in combination with GOLPE

program yielded very good statistical results. The dataset was particularly interesting because the X-ray structures of all protein-ligand complexes were solved. The concept of variable selection and reduction was used in the refinement of original CoMFA method. Approaches for separating the useful variables from the less useful ones were needed. In GOLPE program several variable selection methods, such as D-optimal design and fractional factorial design (FFD) were implemented. The predictivity of each variable was determined by generating a large number of 3D QSAR models, and by calculating the SDEP. After the completion of an FFD run, each variable was evaluated and classified into one of three categories: helpful for predictivity, detrimental for predictivity or uncertain. By applying this variable selection method, QSAR models with higher cross-validated Q^2 values were derived compared to conventional CoMFA method [64].

1.5.2.4. *QSAR validation*

In recent years, QSAR validation received more attention that there were four tools to validate a QSAR models 1. randomization of the response data 2. cross-validation 3. bootstrapping 4. external validation by splitting the total data set into test and training set. REACH (Registration, Evaluation and Authorization of Chemicals) legislation enforced in the European Unoin agreed that QSAR models should be validated scientifically and also regulatory bodies should take decisions based on sound scientific background. Many ideas for validating QSAR models were proposed at an International workshop held in Setubal (Portugal), that were reformed in 2004 by OECD work programme on QSARs. There is an urgent need of a novel parameter to validate the QSAR model to overcome the traditional validation parameters. There is always argument between the users of internal *versus* external validation. External validation supporters considered that internal validation alone is not a parameter to check the robustness of the models and external

validation also must be done. Supporters of internal validation suggested that cross validation was able to measure the fitness of the model and also to cross check the datas used for prediction were not involved in the QSAR development. There were few inconsistency reported in some QSAR models internal and external validations. That is high internal predictivity may have very low external predictivity and *vice versa.* Recently it has been proved that predictive R^2 (R^2_{pred}) may not be a suitable to measure external predictability because of its dependence on training set mean. An alternative measure r_m^2 was suggested to measure external predictivity. r_m^2 depends on the observed and predicted data of the test set compounds, on the training set in LOO method to correlate the observed and the predicted values and also depends on the whole test and training set in LOO method to calculate the overall correlation. For an acceptable QSAR model, one more measure the average correlation coefficient (R_r) of randomized models should be less than the correlation coefficient (R) of the non-randomized model. But there is no proper definition for these two measures in the literatures. But a parameter which penalizes the model R^2 for the difference between these two squared {(R_r^2) & (R^2) values [65].

CHAPTER 2

LITERATURE REVIEW

2.1. HCV NS5B Polymerase

2.1.1. *Structure of HCV NS5B RdRp Polymerase*

This non-structural protein 5B (NS5B) is a RNA-dependent RNA polymerase (RdRp) become very important as drug target because of its ability to play a crucial role in HCV replication and viral genome. Its inhibition doesn't cause any side effect because of its absence in mammalian cells. The crystal structures of NS5B have revealed a unique "right hand" topology of polymerase family, with thumb, palm, and fingers subdomains [66]. There were many available high resolution 3D crystal structures of NS5B complexes either with nucleoside or non-nucleoside inhibitors which revealed that the presence of five distinct allosteric binding pockets as shown in **Figure 10**. Allosteric binding pockets AP-1, AP-2 and AP-3 are presented in thumb and palm site. A unique feature of the HCV RdRp and that of related viral RdRps is that the thumb and the fingers domains are bridged by two loops called the loop 1 and loop 2. These two loops are responsible for the closed conformation of the enzyme that results in a complete encircling of the active site from the front side (front side as viewed with the thumb domain at the left and fingers domain at the right side of the viewer in (**Figure 10**). The back side is covered by another loop called the β loop that forms part of the template channel. This loop extends toward the metal coordinating residues in the active site [20].

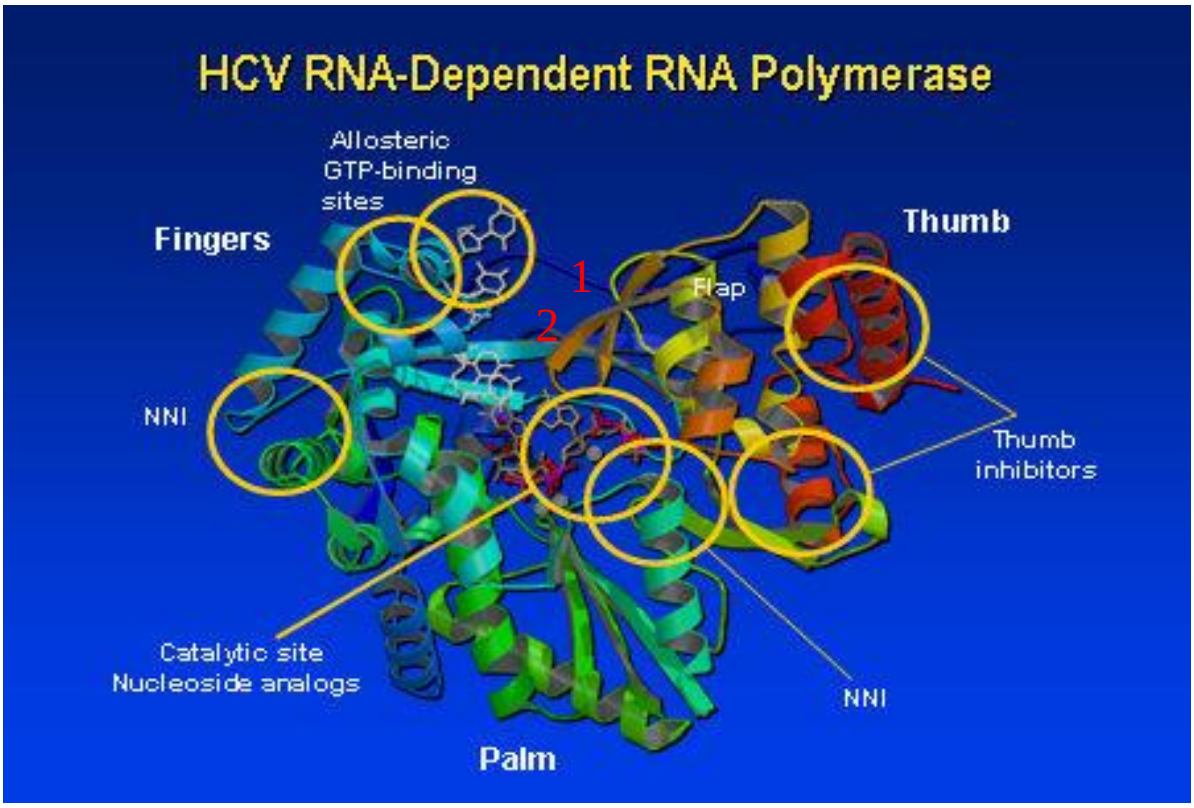


Figure 10: The protein structure of HCV NS5B with the fingers (F), thumb (T) and palm (P) domains [67]

Structural and functional studies of the HCV RdRp have delineated the role of crucial residues in the enzymes however the NS5B bound to ternary complex (template/primer/nucleotide) is yet to be solved. During virus replication, HCV NS5B forms a replication complex in which, NS5B acts as a key component in copying (+) strand HCV RNA into (−) strand HCV RNA. This newly copied (-) strand RNA is then used as the template to produce a large number of progeny (+) strand RNA. It was shown that in both of these critical replication steps, NS5B is involved as a key component and mainly responsible for replicating viral RNA and hence it is a key target for drug development. Also, the catalytic subunit of the replicase complex is the HCV encoded NS5B, which contains all the sequence motifs highly conserved among all the known RdRps. Analogous to the studies from the human immunodeficiency virus (HIV), where the reverse

transcriptase is a primary target for effective antivirals, the HCV RdRp is considered an important target for drug development. Using the right-hand analogy for polymerases, the HCV RdRp has discernable fingers, palm and thumb subdomains. An unusual feature of this polymerase is that, due to the extensive interactions between the finger and thumb subdomains, the HCV RdRp has an encircled active site. Similar to other known RdRps, the HCV NS5B also contains six conserved motifs designated as A-F and more than hundred crystal structures have been determined independently by several groups [68]. Many inhibitors bound with HCV NS5B crystal structures are also available in the protein databank (PDB).

2.1.2. *HCV NS5B inhibitors*

Based on the structure and site of binding with HCV NS5B polymerase, the inhibitors are classified into two types: Nucleoside or nucleotide Inhibitors which during RNA synthesis, they act as the competitors of NTPs while non-nucleoside inhibitors (NNI), are the second major class of inhibitors which inhibits the starting stage of RNA synthesis.

2.1.2.1. *Nucleoside or nucleotide inhibitors:*

The only FDA approved drug for HCV treatment is the purine analog ribavirin which showed pleiotropic effects on both cellular and viral enzymes. Several sugar moieties were reported as promising inhibitors of the replication of HCV, among these 2′-and 3′-substituted analogs of ribonucleosides showed very good potency. The important nucleosides (**Figure 11)** were 2'-C-methyladenosine (a), 2'-O-methylcytidine (b) 2′-C-methylguanosine (c), 2′-deoxy-2′-flouro-2′-C-methylcytidine (d) and 3′-deoxyribonucleosides derivative like 3′-deoxycytidine (e).

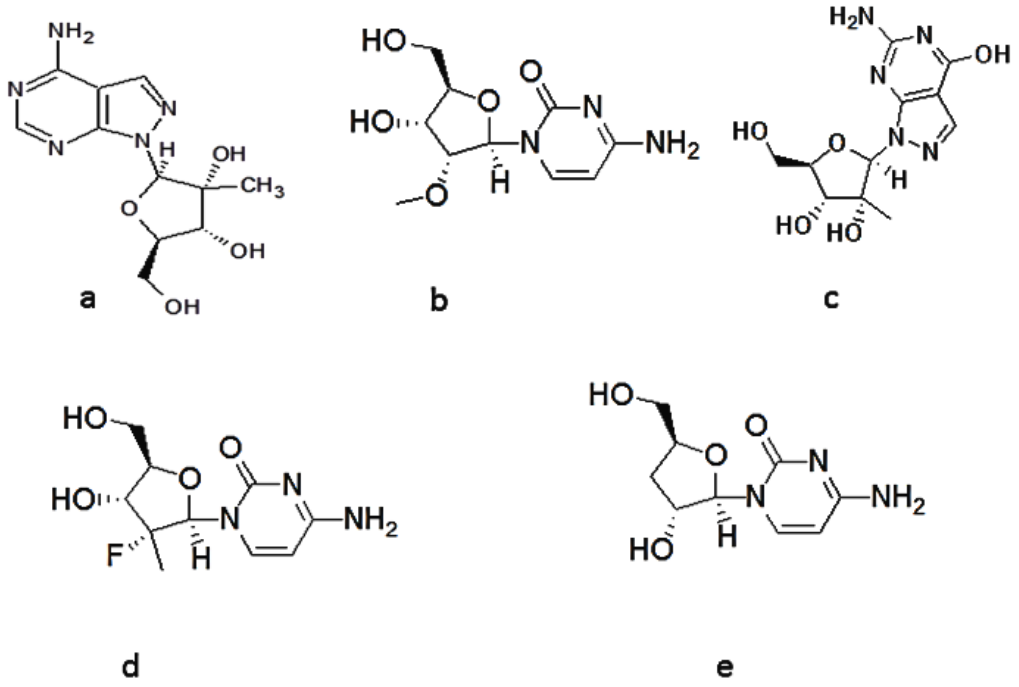


Figure 11: Nucleoside or nucleotide inhibitors as HCV NS5B inhibitors

These nucleosides were good competitors for the substrate and behaved as chain terminators [69]. 2′-Deoxy-2′-flouro-2′-C-methylcytidine (**Figure 11**) showed an EC_{90} of 5.4 μM, comparable to that of 2′-deoxy-2′-fluorocytidine (d) and it was not cytotoxic upto 100 μM. The compound 3′-deoxycytidine (e) showed submicromolar activity in a biochemical assay [70].

Recently, Roche showed interim results from a multiple ascending dose study of R1626, a novel nucleoside analog targeting HCV polymerase in chronic HCV-infected individuals. R1626 a pro-drug of R1479 exhibited potent anti-HCV activity. After oral administration, R1626 was efficiently converted to R1479 and was well tolerated in native chronic patients. Moreover, dose dependent antiviral activity was observed with a mean serum viral RNA reduction of 1.2 log10 from baseline after treatment with 1500 mg of R1626 for 14 days [71].

In a recent communication, Varaprasad *et al.* [72] reported the synthesis of some pyrrolo[2, 3-*d*]pyrimidine nucleoside derivatives (**Figure 12**) as potential anti-HCV agents.

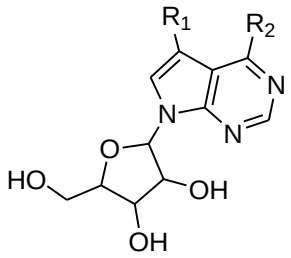


Figure 12: Pyrrolo[2, 3-*d*]pyrimidine nucleoside derivatives as HCV NS5B inhibitors

Larger size of alkyl group at R_2 has been found to improve activity along with increased cytotoxicity. Further modifications at C4 and C5 positions were also found to increase the activity but resulting compounds were relatively toxic. The carboxamide oximes and carboxamidine derivatives revealed good selectivity. These results will lead to further development of potent combination therapy with other type of inhibitors of HCV NS5B.

2.1.2.2. *Non-nucleoside inhibitors (NNI):*

Till date, five distinct inhibitor binding pockets have been identified, where NNI were found to bind. Allosteric binding pockets AP-1, AP-2 and AP-3 are present in thumb and palm site.

AP-1 is situated on the surface of thumb domain adjacent to the allosteric GTP-binding site. Indole, benzimidazole, quinoxaline, coumestan and thieno[3,2-b]pyyrole derivatives were the identified as inhibitors for this site. Thiazolone, N, N-disubstituted phenylalanine, thiophene-2-carboxylic acid, pyranoindole, dihydropyranone and thiazolidin-4 one derivatives were reported as inhibitors for the AP-2 site located in the thumb domain adjacent to AP-1. While, benzothiadiazine, benzylidene, proline sulfonamide, anthranilic acid, acrylic acid and pyrrolidine derivatives were inhibitors identified for AP-3 site located adjacent to the polymerase active site in palm site [14]. There were so many NNIs reported for HCV NS5B. Here we have discussed only the recent developments.

In 2005, Ding et al. reported few pteridine derivatives with IC_{50} of 15 μM and 1.6 μM for two compounds respectively as shown in **Figure 13** and further also reported 5-cyano-6 aryl-2-thiourail compound with IC_{50} of 27 μM concentration and later did SAR studies and developed few deriatives [73].

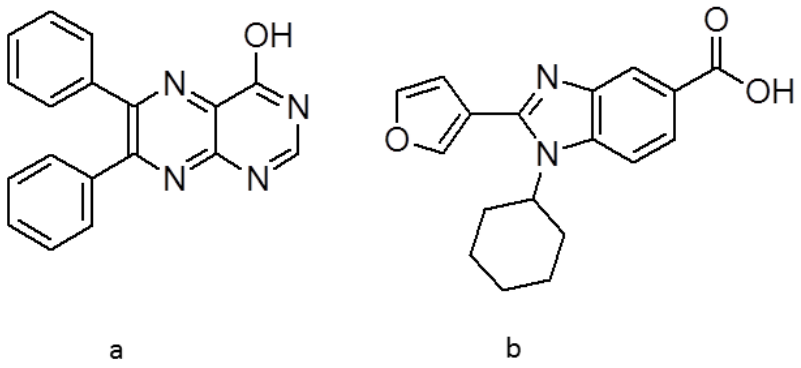


Figure 13: HCV NS5B inhibitors reported by Ding *et al* [73]

In 2006, Yan et al. [74] modified the already reported thiazolone derivatives as shown in **Figure 14** and few of the newly synthesized compounds showed IC_{50} 5 μM against HCV.

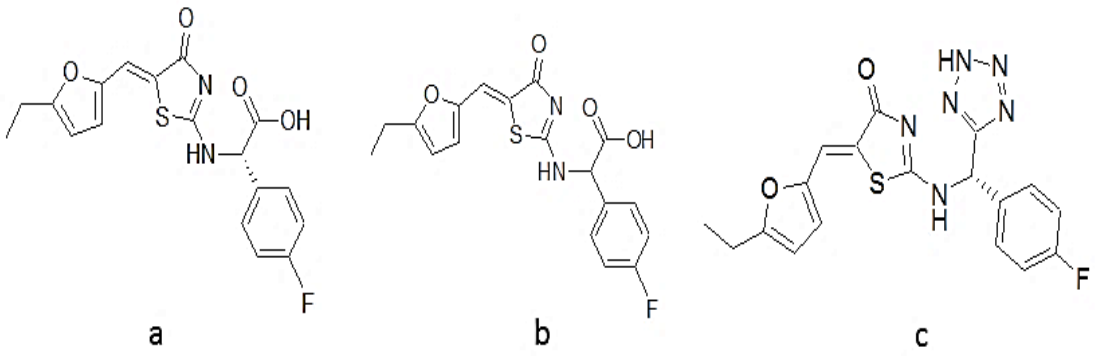


Figure 14: Thiazolone derivatives as HCV NS5B inhibitors

In 2007, Puerstiger *et al.* [75] reported 5-benzyl-2-phenyl-5H imidazolo[4,5-c]pyridines as potent lead with antiviral properties and further synthesized based on SAR that showed good EC_{50} (1.0 μM) values. Sriram *et al.* [76] reported some aminopyrimidinimino isatin analogues with good

inhibitory activity against HCV (**Figure 15),** and the most active compound showed 100% inhibition and was non-toxic upto 50µg/mL. Parhenolide identified as inhibitor of HCV replication by Hwang *et al.* [77].

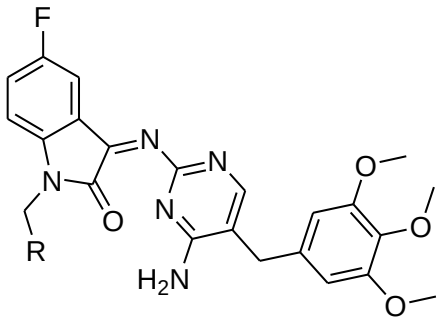


Figure 15: Isatin analogues as HCV NS5B inhibitors

In 2009, Hwu *et al.* [78] reported the synthesis and evalution of benzimidazole-coumarine derivatives with a methylenethio linker and the corresponding N-glucosides. The substituents –F, -Cl, and -Me on benzimidazole ring as in **Figure 16** showed very little inhibition compared to –Br, -H, and -OMe groups at the coumarine ring that showed very good inhibition. In the year 2009, Kim *et al.* [79] reported few novel aryl diketoacid analogues, in which the p-chloro analogues were found to show more efficiency. Optimization of a pyrrolidine-based template was reported by Slater *et al.* [80] and the SAR findings concluded the importance of the bulkier

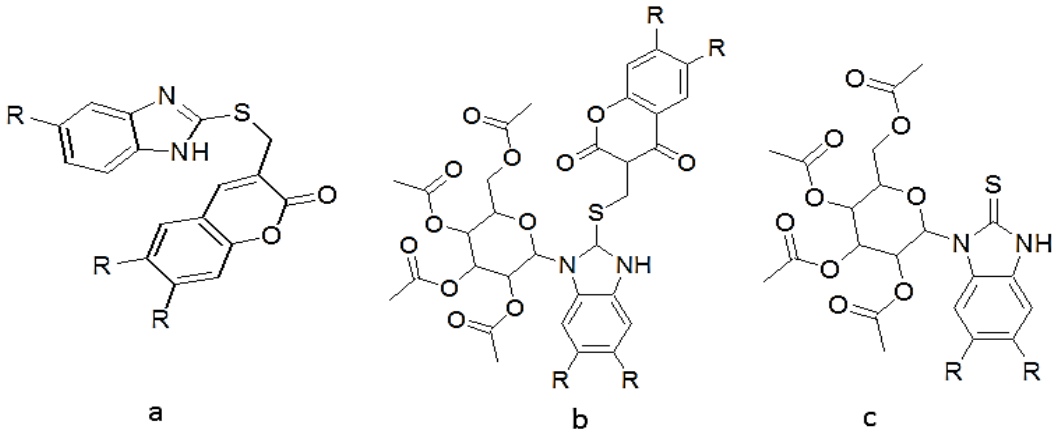


Figure 16: Benzimidazole-coumarine derivatives as HCV NS5B inhibitors

4-*tert*-Butyl in the benzamide and 2-thiazole at C5 and a primary amide at C4 showed good polymerase activity. Lee *et al.* [81] reported that *n*-butanol-methanol extract obtained from *Acacia confuse* plant, showed inhibition of HCV replication. The EC_{50} value and CC_{50}/EC_{50} selectivity index (SI) were found to be 5 ± 0.3 g/ml and >100 with, respectively. The extract also showed antiviral synergism in combination with IFN-α and Telaprevir; VX-950 and 2'-*C* methylcytidine; and NM-107. The extracts were also found to significantly suppress COX-2 expression in HCV replicon cells.

2.1.3. *Molecular modelling studies done on HCV NS5B polymerase target*

Using computer aided drug design tool with high-throughput virtual screening of different libraries, some proline derivatives were reported by Gopalsamy et al. [184] and derivatized for SAR analysis. Another study on a series of indole-N-acetamide was reported as inhibitors of replication of subgenomic HCV RNA in HUH-7 cells in which the role of H-bonding with NS5B enzyme was indicated [82].

In 2007, Yan *et al.* [83] reported few thiazolone-acylsulfonamides as potent allosteric inhibitors of HCV NS5B polymerase using structure based drug design tools and X-ray crystallographic tools. Melagraki *et al.* [84] performed a QSAR study on a set of 98 genotype 1 HCV polymerase inhibitors and developed a MLR QSAR model using lipophilicity, HOMO energy, Kier and Hall index order 2 (Ki2) and Kier and Hall information indices (KiInf0, KiInf3).

In 2010, Talele *et al.* [85] reported 23 inhibitors with rhodanine scaffold with IC_{50} range from 7.7 - 68.0 μM by screening ChemBridge database of 260,000 compounds against the tetracyclic indole inhibitor binding allosteric pocket (AP-1) of NS5B to identify novel inhibitors through a combined use of virtual screening, SAR analysis, synthesis and biological evaluation. In 2010, a detailed QSAR study was done by Patil *et al.* [86] on pyrrolo[2,3,-d]pyrimidine nucleoside

derivatives. QSAR study was done by Varaprasad *et al.* [72] in 2007 on benzimidazole-coumarin conjugates and also by Hwu *et al.* [78].

In 2010, using kNN-MFA approach, 3D QSAR model was reported on benzimidazole [87] and thiouracil [88] derivatives by Gupta *et al.* They validated the model with q^2 and predictive r^2 values. Combined molecular docking studies and 3D QSAR (COMFA and CoMSIA) studies on benzimidazole derivatives by Patel *et al.* [89] explained the inhibitory activity based on the CoMFA steric and electrostatic contour maps. Few naphthyridine compounds which were identified through molecular modelling studies of benzimidazoles are given in **Figure 17**.

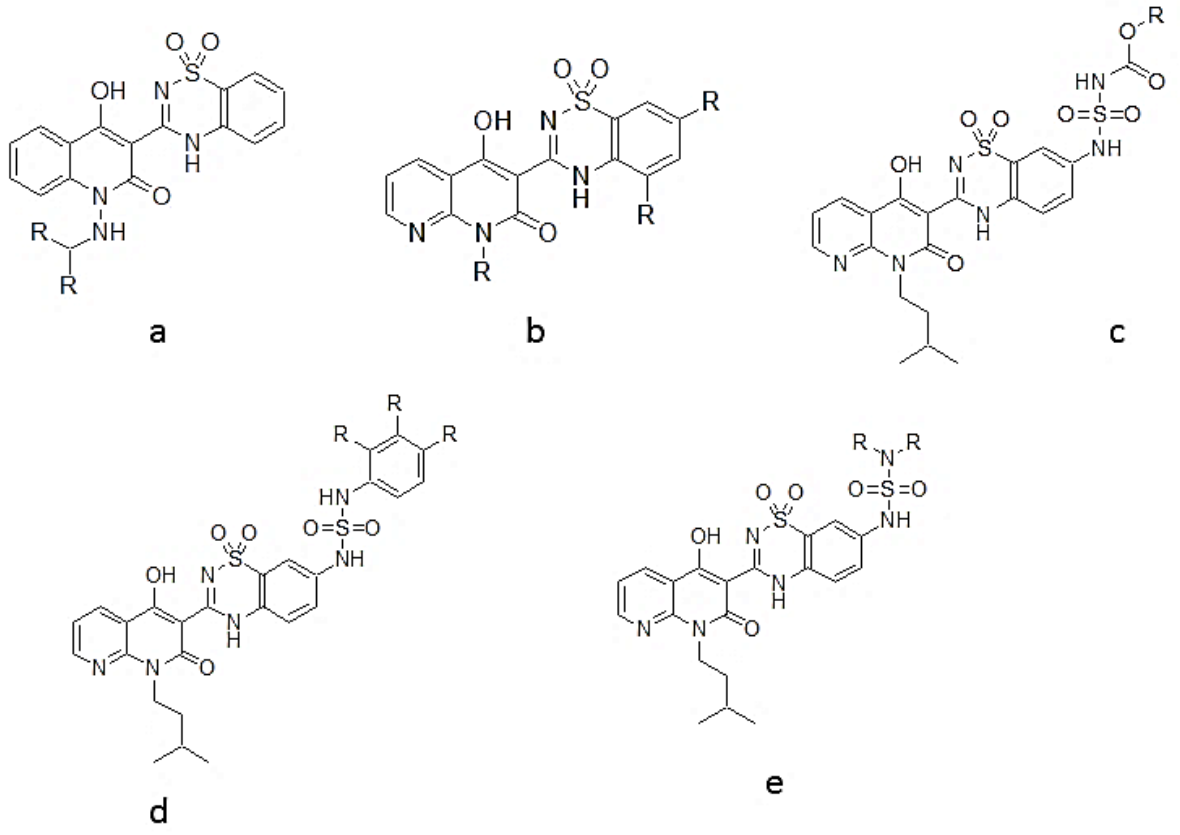


Figure 17: Leads identified through molecular modeling studies as HCV NS5B inhibitors

The site specific HCV inhibitors and their mode of activity are presented in **Table 4**. There were many inhibitors reported till date, but very few have entered into clinical trials. HCV polymerase inhibitors that reached clinical trials of certain phases are shown in **Table 5**.

Table 4: The site specific important HCV inhibitors and their mode of action [20]

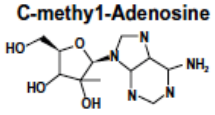
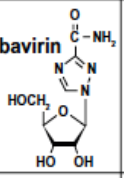
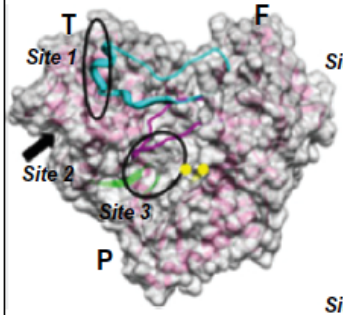
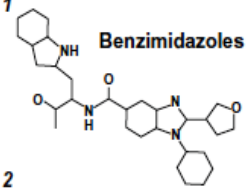
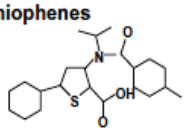
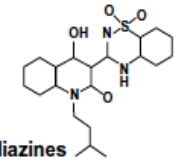
Binding sites	Compounds	MOI	Comments
NI Active site	C-methy1-Adenosine  Ribavirin 	Incorporation into nascent RNA; inhibits both initiation and elongation	Non-obligate chain terminators; Ribavirin acts by multiple mechanisms
NNI  Site 1 Site 2 Site 3 Benzothiadiazines	Site 1 Benzimidazoles  Site 2 Thiophenes  Site 3 	Inhibits initiation Not clear; alters NS5B conformation Inhibits initiation	Referred to as 'finger-loop' inhibitors; cocrystalized with NS5B MOA is not clear; cocrystalized with NS5B Binds different drugs; benzofurans cocrystalized with NS5B

Table 5: The HCV polymerase inhibitors that reached the clinical trials of certain phases

Polymerase Inhibitors	**Company**	**Current clinical phase**
RG-7128 Mericitabine	Roche	II
GS-7977 Sofosbuvir	Gilead	III
NNI-Site 1 BI207127	Bohringer Ingelheim	III
NNI-Site 1 BMS 791325	BMS	II
NNI-Site 2 Filibuvir	Pfizer	II
NNI-Site 2 VX-222	Vertex	II
NNI-Site 3 Setrobuvir	Anadys	I
NNI-Site 3 ABT 333	Abbott	III
NNI-Site 3 ABT-072	Abbott	II
NNI-Site 4	Gilead	II

Musmuca *et al.* [90] reported a detailed study on the combined alignment of on 3D QSAR (ligand based) and structure based procedures and used this alignment to identify potent leads for HCV NS5B polymerase.

Wang *et al.* [91] reported few benzothiadiazine derivatives with the help of 3D QSAR, docking studies and molecular dynamics. Also, Zhang *et al.* [92] reported the requirments needed for the benzothiadiazine derivatives as inhibitors of HCV NS5B polymerase using computational tools like 3D QSAR, docking and dynamics studies.

Recently Yu *et al.* [93] reported in 2013 few 5-hydroxy-2H-pyridazine-3-one derivatives as HCV NS5B polymerase inhibitor by employing drug design tools like 3D QSAR, molecular docking, molecular dynamic simulation and binding free energy calculation studies.

Hucke *et al.* [94] reported a potent lead with picomolar cellular potency by using molecular dynamic study and structure based drug design tools in 2013.

Early 2014, Jin *et al.* [95] reported few indole derivatives by using virtual screening techniques for their in-house library with 6000 compounds and using the identified hits they found the presence of indole moiety in most of the hits. Furthur synthesis of few indole derivatives and HCV NS5B replication studies yielded a potent lead with IC_{50} of 292 nM.

Hence, HCV NS5B is an important target for antiviral therapies, which has both nucleoside and non-nucleoside polymerase inhibitors. R7128 is a prodrug of cytidine analog, PSI-6130, which shows positive, results in phase I clinical trials and shows promising reports in phase II. HCV NS5B nucleoside inhibitor R1626, showed hematological and ocular toxic effects during its clinical trials has been halted. GS9190, filibuvir, VCH-222 are non-nucleoside inhibitors shown potent antiviral activity in clinical trials.

2.2. Human immunodeficiency virus (HIV) protease

2.2.1. *Structure of HIV protease*

HIV protease plays an important role in the life cycle of the virus by processing the viral gag and gag-pol polyproteins into smaller structural and functional proteins which are crucial for the viral maturation. Inhibition of HIV protease leads to the generation of non-infectious virus particles and acts as a promising drug target for antiviral drug design. HIV protease inhibitors are the most potent anti-AIDS drugs reported to date and are crucial components of most active antiretroviral therapy (HAART) [96]. Due to HIV mutants that are resistant to current drug regimens it leads to the clinical failure of antiviral therapy. Currently FDA approved protease inhibitors must face the challenge of the emergence multi-drug-resistant (MDR) protease variants to prove their efficacy. There is a need of next generation protease inhibitors which should overcome this MDR virus, with improved pharmacological properties and good activity profile. The crystal structure of HIV protease was first reported by Navia and colleagues from Merck Laboratories in 1989 [97]. Various crystal structures of HIV protease revealed a vast difference in the binding mode of the substrate and different inhibitors. Substrate form mostly conserved H-bond with backbone not with side chains. Inhibitors which show this type of conserved H-bond with the backbone would show more effectiveness. Eg: Amprenavir and Darunavir. They form H-bond with backbone Asp29 and Asp30 residues. The crystal structure of protease is well characterized in terms of its function, substrate specificity and its inhibitor binding nature. It is a homodimer of two identical polyprotein chains with only one active site. The interface of the dimer consisted of four stranded beta sheet with beta strands of both amino and carboxyl terminals. The outer part of the dimer interface is formed by amino terminal beta strand ‘a’ (residues 1-4). It continues through a loop to strand ‘b’ (residues 9-15). The strand ‘c’ follows strand ‘b’ ends in the active site trid (three

residues 25-27). The beta strand‘d’ follows the active site loop (residues 30-35) followed by a broad loop (residues 36-42). The second half monomer topology resembles that of the first half topology by pseudo-dyad and is represented in the diagram **Figure 18**.

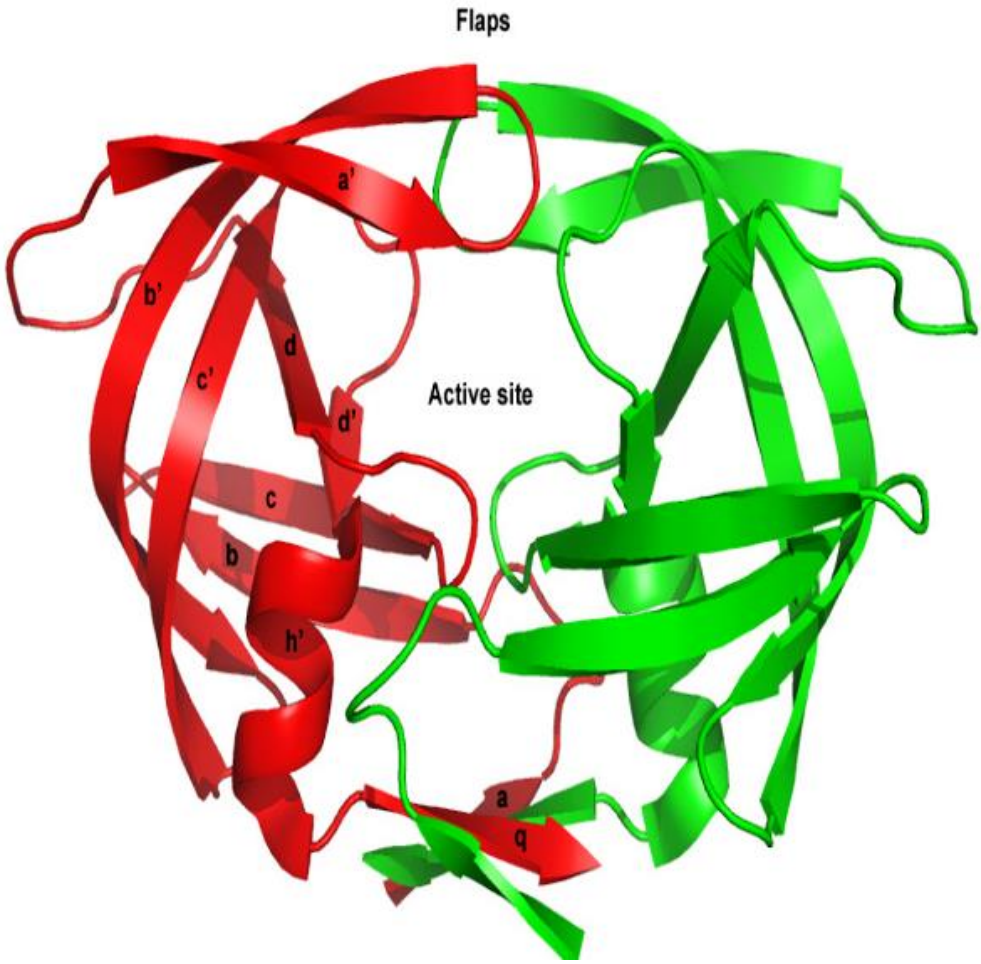


Figure 18: The active site pocket of HIV protease [98]

The residues 43-49 form a beta strand a’which forms a part of flap with residues 52-58 to form the other part of flap and part of strand b’ (residues 52-66). Strand c’ with residues 68-78 continues through a loop (residues 79-82) to chain d’ (residues 83-85) which follows a small helix h’ (residue 86-94). A carboxyl terminus beta strand q’ (residues 95-99) is followed by helix h’ which form the inner core of the dimer interface of the four stranded beta sheet and gives psi shaped sheet which is a characteristic feature of aspartic proteases [98].

The important role of HIV protease is to cleave the polyprotein chain into structural proteins like p17, p24, p7, p6, p2, p1 and functional proteins like protease p11, reverse transcriptase p66/p51, and integrase p32. Later maturation will occur followed by the infection of virions. Number of low molecular weight inhibitors of HIV-1 protease (MW < 1000 Da) are now available including saquinavir, ritonavir, indinavir, nelfinavir and amprenavir. These are among the first successful examples of receptor/structure based designer drugs and were developed using structures of compounds bound in the active site of HIV-1 protease.

The active site of HIV protease is situated at the dimer interface. The beta ribbons which form the roof of the active site contain many glycine moieties and are conformationally highly flexible. Compared to free HIV protease crystal structures with Inhibitor bound structures have 7-15 Å movement of flap tips around the residues 50/50'. The specificity of the enzyme is confirmed by the hydrogen bond interaction of the substrate in its extended conformation with the respective amino acid residues [99].

The residues which line up the S1/S1' sub-sites include Arg8, Leu23, Asp25, Gly27, Gly48, Gly49, Ile50, Thr80, Pro81 and Val82. The S2/S2' sub-sites are mostly hydrophobic (Ala28/28', Leu23/23', Ile47/47', Gly49/49', Ile50/50', Leu76/76' and Ile84/84') except Asp29, Asp29', Asp30 and Asp30' [100].

2.2.2. *FDA approved protease inhibitors*

US FDA approved, at present 25 compounds for clinical trial for the treatment of AIDS which includes 10 protease inhibitors. The 10 FDA-approved protease inhibitors: saquinavir (1), ritonavir (2), darunavir (3), indinavir (4), tipranavir (5), fosamprenavir (6), nelfinavir (7), atazanavir (8), lopinavir/ritronavir (9), and amprenavir (10) (**Figure 19** & **Table 6).** There are three generations of PIs which aimed to improve the efficacy of drug, to increase the patients

survival, their quality of live, immunological response and decrease the rate of infection. Except tipranavir (has coumarine scaffold) all PIs have hydroxyl ethylene scaffold which mimics the normal peptide linkage [101].

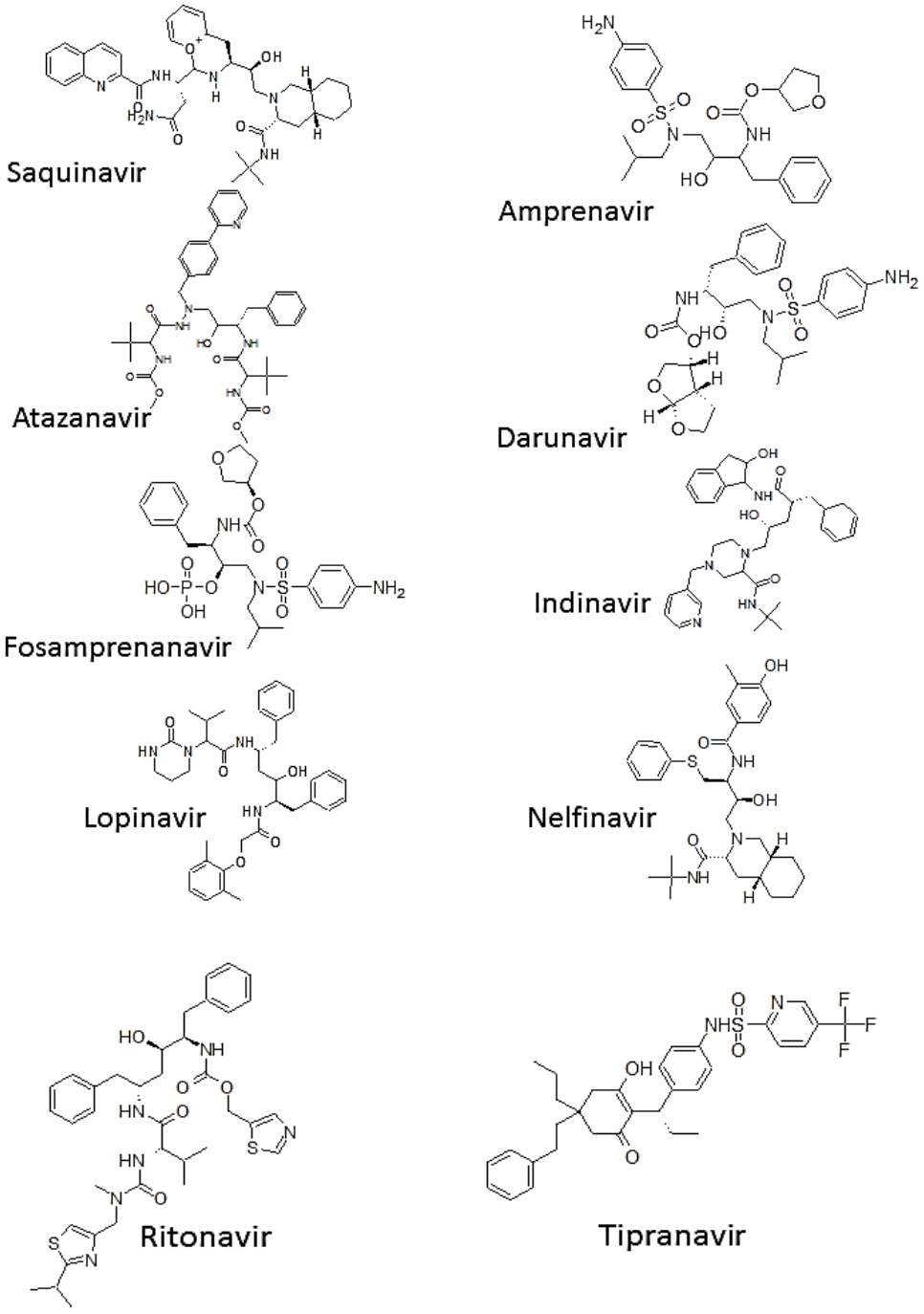


Figure 19: FDA approved HIV protease Inhibitors

Table 6: FDA approved protease inhibitors (PIs) for HAART therapy

Brand Name	Generic Name	Manufacturer Name	Approval Date	Time to Approval
Agenerase	amprenavir, APV (no longer marketed)	GlaxoSmithKline	15-Apr-99	6 months
Aptivus	tipranavir, TPV	Boehringer Ingelheim	22-Jun-05	6 months
Crixivan	indinavir, IDV,	Merck	13-Mar-96	1.4 months
Fortovase	saquinavir (no longer marketed)	Hoffmann-La Roche	7-Nov-97	5.9 months
Invirase	saquinavir mesylate, SQV	Hoffmann-La Roche	6-Dec-95	3.2 months
Kaletra	lopinavir and ritonavir, LPV/RTV	Abbott Laboratories	15-Sep-00	3.5 months
Lexiva	Fosamprenavir Calcium, FOS-APV	GlaxoSmithKline	20-Oct-03	10 months
Norvir	ritonavir, RTV	Abbott Laboratories	1-Mar-96	2.3 months
Prezista	Darunavir	Tibotec, Inc.	23-Jun-06	6 months
Reyataz	atazanavir sulfate, ATV	Bristol-Myers Squibb	20-Jun-03	6 months
Viracept	nelfinavir mesylate, NFV	Agouron Pharmaceuticals	14-Mar-97	2.6 months

First generation PIs: Saquinavir (Invirase and Fortovase) was the first FDA introduced PIs in the year 1996 and followed by ritonavir (Norvir), indinavir (Crixivan) and in 1997 the nelfinavir (Crixivan). In 1999, amprenavir (Agenerase and Prozei) was introduced. These were the first generation PIs and were the widely available drugs till the end of 1990s [102].

Saquinavir has low bioavailability and gets metabolised through cytochrome P450. It inhibits both HIV-1 and HIV-2. It is well tolerable drug with mild gastrointestinal symptoms [103]. Ritonavir has very high oral absorption and is not affected by food and also metabolized by CYP3D6 [104]. Indinavir has a major drawback in drug accumulation and has no inductive effect on hepatic CYP enzymes which affects its own metabolism and also evident of nephrolithiasis and gastrointestinal complaints. Though it has good antiretroviral activity, due to its unfavorable pharmacological

reports prevented its use for chronic treatment. This was overcome by the co-administration of low-dose ritonavir which improved its pharmacokinetic parametres [105].

Nelfinavir is a less effective inhibitor compared to other PIs and resistance to a unique mutation of the gene. The adverse event in patients using nelfinavir based combination is diarrhea, rashes, nausea, head-ache and asthenia. Its prolonged viral suppression, good tolerability and a unique resistance report made it as a well known drug for the treatment of adults, pregnant women, paediatric patients with HIV infection and patients those who are unable to tolerate other PIs [106, 107].

Amprenavir is a sulfonamide based drug with low toxicity [108]. Due to low bioavailabilty, high pill burdens of first generation PIs led to less adherence and limited long-term viral inhibition [109]. The second generations PIs were introduced in early 2000, due to resistance to previously reported drugs. These includes Lopinavir (Aluviran and Koletra) in 2000 and in 2003, fosamprenavir (Lexiva and Telzir). Due to increase in number of potent antiretroviral agents, that resulted in complete suppression of viral replication, in 2005, FDA approved third generation PIs tipranavir (Aptivus) and in 2008 darunavir (Prezista) which showed high potency and effective barrier for resistance than all other reported PIs [110].

2.2.3. *Recent molecular modelling studies on HIV protease as target*

Boutton *et al.* [111] in 2005 did a genotype dependent QSAR study for HIV protease inhibitors. In this study he used computational structure based approach to predict the resistance of HIV strains to amprenavir drug by calculating the interaction energy of the drug and HIV protease.

Surleraux *et al.* [112] in 2005 reported few fused heteroaromatic sulfonamides and new classes of compounds as presented in **Figure 20** for P2' region with improved activity (pEC_{50} 7.5-8) and good pharmacokinetic properties by employing the molecular modeling tools.

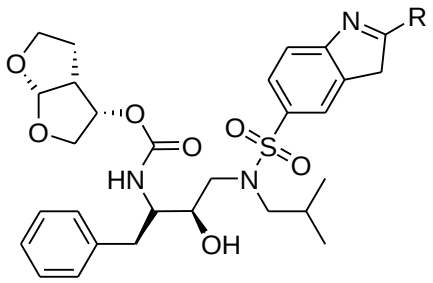


Figure 20: Derivatives of benzoxazole and benzothiazole as HIV protease inhibitors

Ghosh *et al.* [113] specifically designed few compounds using structure based drug design tools to interact with backbone of HIV protease to combat resistance and found that one compound showed an IC_{50} 1.8 nM (**Figure 21).**

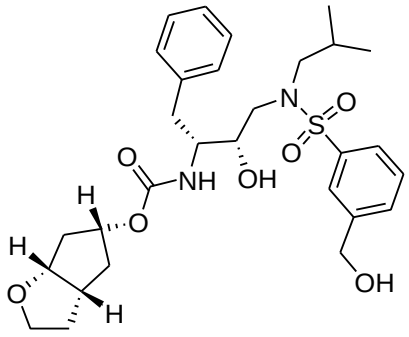


Figure 21: 5-Hexahydrocyclopenta[*b*]furanyl urethane as HIV protease inhibitors

In 2006 Ali *et al.* [114] designed few phenyloxazolidinone derivatives (**Figure 22)** by incorporating N-phenyloxazolidinone-5-carboxamide into (hydroxyethylamino)sulfonamide scaffold using structure based drug design tools and the most potent compound showed K_i of 108 pM against wild type HIV and against MDR variants it showed picomolar to nanomolar K_i**.**

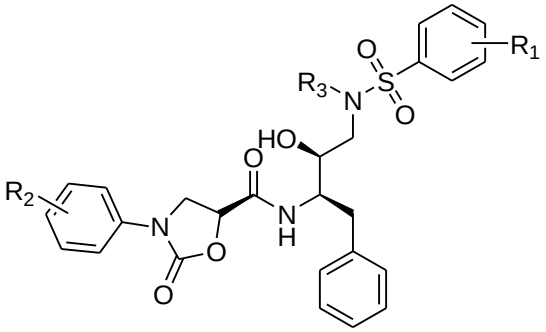
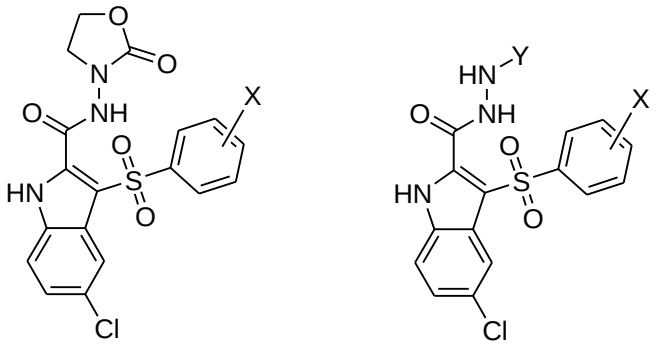


Figure 22: Phenyloxazolidinone derivatives as HIV protease inhibitors

Ragno *et al.* [115] reported few indolyl aryl sulfones derivatives using molecular modelling and 3D QSAR studies as shown in **Figure 23**. Two compounds showed very good activity with X-Me_2 and Y-i-Pr against HIV.



a b

Figure 23: Indolyl aryl sulfones derivatives as HIV protease inhibitors

Further Volarath *et al.* [116] in 2007 discussed about the various structure based drug design tools which included from molecular modeling to cheminformatics. Ghosh *et al.* [117] in 2009 reported few non-peptidic macrocyclic HIV protease inhibitors and found that cyclic inhibitors were more potent than their acyclic and unsaturated analogs than saturated. The compounds (**Figure 24)** showed good enzyme inhibitory, antiviral and potent activity against MDR HIV-1 variants.

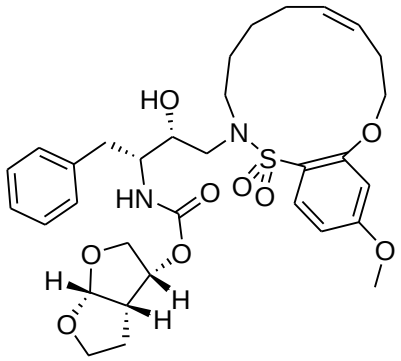


Figure 24: Non-peptidic macrocyclic derivatives as HIV protease inhibitors

Ali *et al.* [118] reported few substituted phenyloxazolidinones (**Figure 25)** by incorporating phenyloxazolidinones into hydroxyethylamine core based on SAR studies.

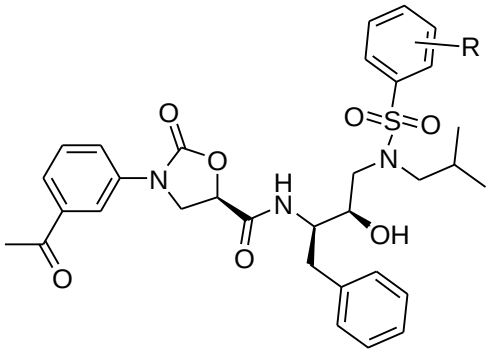


Figure 25: Substituted phenyloxazolidinones as HIV protease inhibitors

In 2011 on the basis of pharmacophore modeling of known PIs, conformationally restricted sulfonamides (Figure 26) were designed and synthesized by Ganguly *et al* [119].

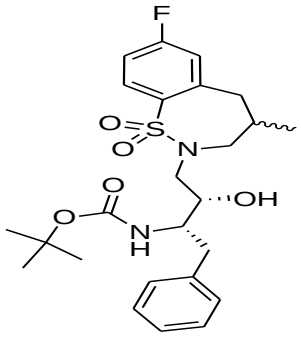


Figure 26: Sulfonamides derivatives as HIV protease inhibitors

In 2012 Yadav *et al.* [120] found few hit molecules from Maybridge library using combined structure based and ligand based virtual screening methods. They developed a pharmacophore model based on structure based tool and compared that pharmacophore with the ligand based pharmacophore and screened the Maybridge database. The final four hit molecules obtained are shown in **Figure 27.**

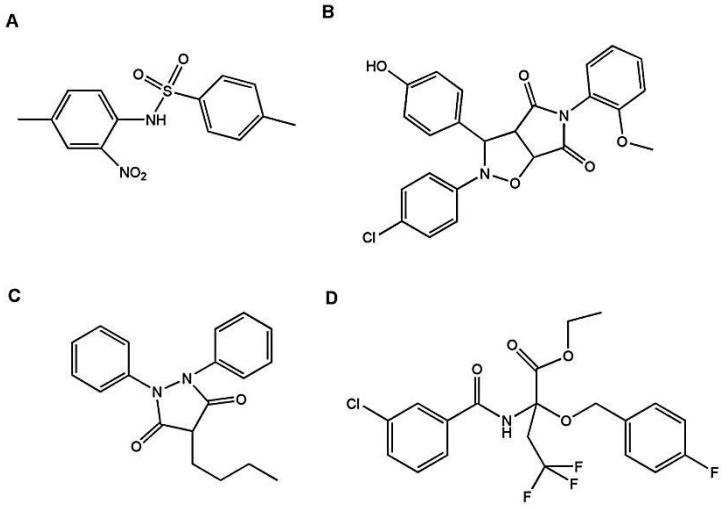


Figure 27: Four Maybridge hit compounds as HIV protease inhibitors

Joshi *et al.* in 2013 [121] reported three 14 and 15-member macrocyclic inhibitors (**Figure 28)** by using drug design tools and also evaluated ten novel linear PIs. They found that the macrocyclic systems showed higher activity than the linear ones having K_i=3.1 nM and EC_{50}= 0.37 μM.

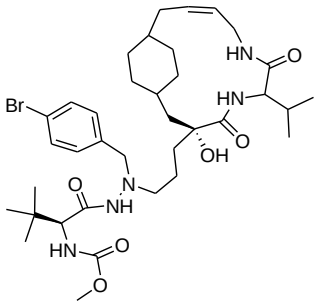


Figure 28: Macrocyclic compounds as HIV protease inhibitors

Recently in 2014 Tiefenbrunn *et al.* [122] performed a crystallographic fragment-based drug discovery to understand the surface sites in soaking experiments. They used 68 brominated fragments and identified two new compounds that bound to two known surface sites of HIV protease.

Furthur, 2014 Pang X *et al.* [123] reported some advances of non-peptidometric inhibitors for HIV protease. In this article they broadly discussed about the advantages of the molecules like low molecular weight, good bioavailabilty, high stability through *in vivo*, low resistance and their low cost of production.

Antunes *et al.* [124] in 2014 used *in silico* methods like molecular docking and molecular dynamic simulation studies to evaluate the unusual mutations which caused the major drug resistance to nelfinavir drug. And Yedidi *et al.* [125] showed that a compound (**Figure 29)** with para-fluoro phenyl group had shown very good binding effect and enhanced activity against MDR HIV protease using structural based virtual screening of extended lopinavir analogs.

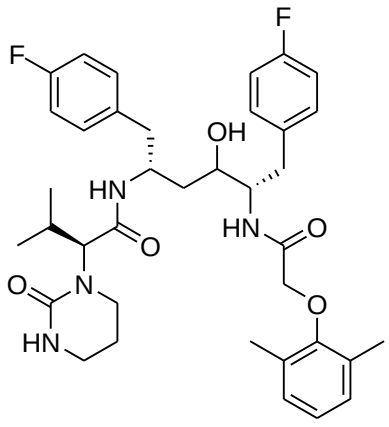


Figure 29: p-Fluoro phenyl substituted lopinavir as HIV protease inhibitors

2.3. MTB N-Acetylglucosamine 1-phosphate uridyltransferase (GlmU)

2.3.1. *Structure of GlmU*

GlmU (N-Acetylglucosamine-1-phosphate uridyltransferase**)**, a bifunctional enzyme, produces UDP-GlcNAc, an essential precursor for the biosynthesis of peptidoglycan and lipopolysaccharide

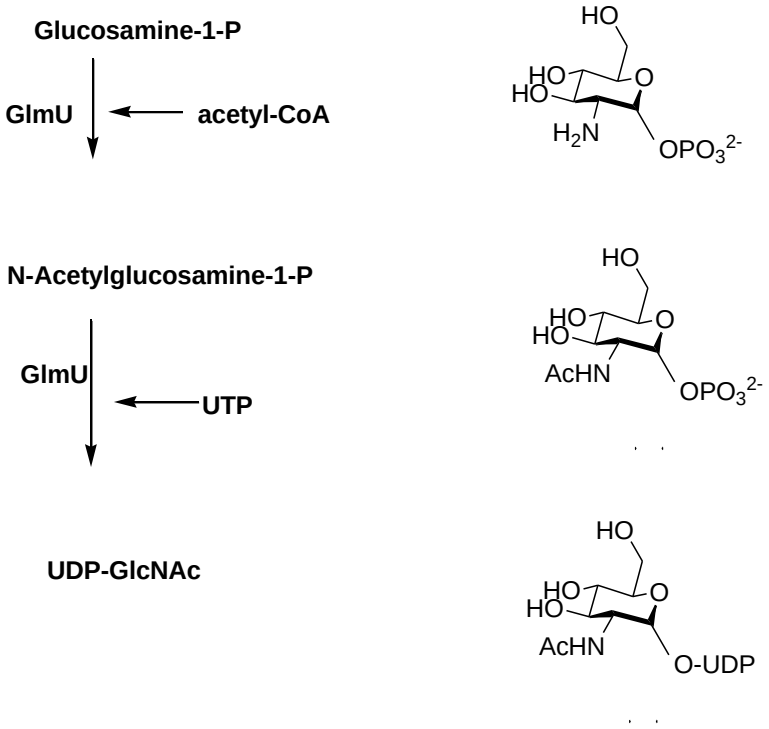


Figure 30: Biosynthetic pathway of bifunctional enzyme GlmU

Shock mutagenesis experiments revealed *MTB GlmU* to be an essential gene. In line with this, recent work demonstrated the inability of *M. smegmatis* to grow in the absence of GlmU. GlmU carries out two important biochemical activities: a C-terminal domain catalyses the transfer of acetyl group from acetyl coenzyme A (acetyl-CoA) to glucosamine-1-phosphate (GlcN-1-P) to produce N-acetylglucosamine- 1-phosphate (GlcNAc-1-P), which is converted to UDP-N-acetylglucosamine (UDP-GlcNAc) by the transfer of UMP (from UTP), a reaction catalyzed by the N-terminal domain [126] (**Figure 30)**. Though the second step is present in prokaryotes as

well as in humans, the first step is present only in prokaryotes. The absence of the first step in human makes it suitable for designing non-toxic inhibitors.

GlmU from MTB displays two-domain architecture – an N-terminal domain (residues 6 – 241) with α/β like fold and a C-terminal domain (residues 264 to 473) that forms a left-handed parallel-β-helix structure (LβH) **(Figure 31)**. A long extended C-terminal tail (residues 450-495) with little secondary structure accounts for the remaining C-terminal region [127].

The two domains are connected by a 33Å long α-helix (residues 241 to 263) that forms a hinge. The N and C-terminal domains are responsible for uridyltransferase and acetyltransferase activities, respectively. The formation of the C-terminal active site requires a trimeric arrangement [128]. The C terminal tail residues contributes to the acetyltransferase active site by making several stabilizing contacts such that its extreme end contributes to the active site formed by the other two monomers. In the case of MTB GlmU too, the tail adopts a similar conformation. However, a major difference is a 30 amino acid extension of the tail. Structural and biochemical analyses showed the significance of a variable C-terminal tail in regulating acetyltransferase activity **(Figure 31)**.

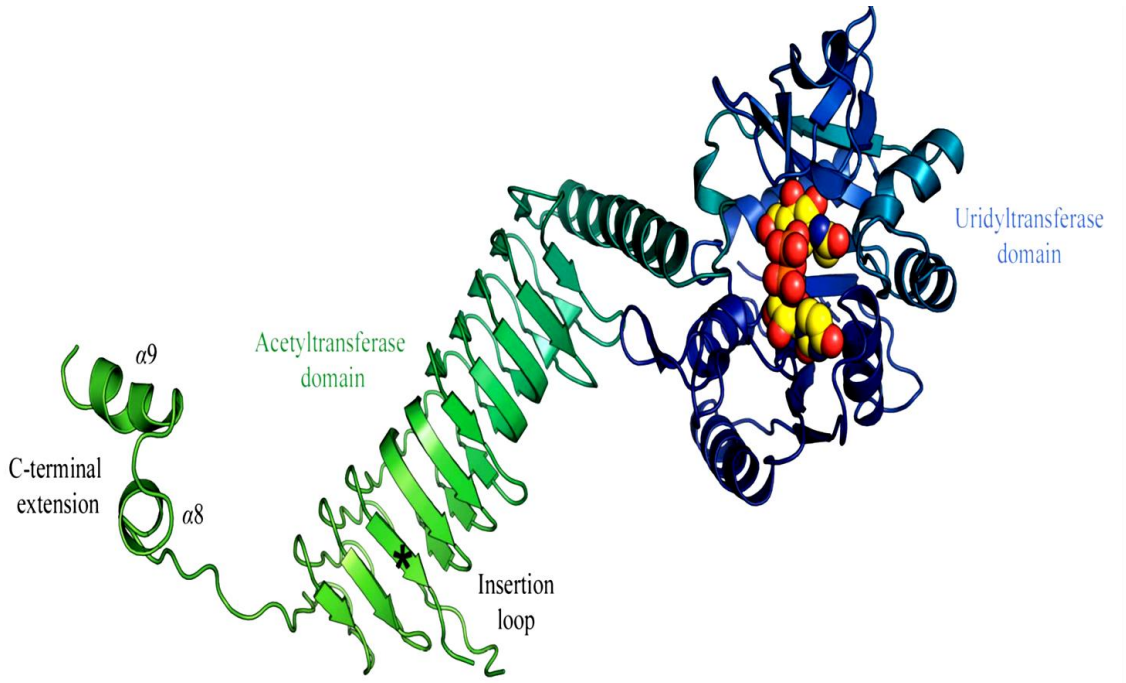


Figure 31: Structure of GlmU (N-Acetylglucosamine-1-phosphate uridyltransferase) [129]

GlmU is a substrate for eukaryotic like serine/threonine kinase PknB. Kinase assays with PknB using the N and C-terminal domains of GlmU as substrates illustrated that PknB phosphorylates GlmU in the C-terminal domain. Interestingly, it was demonstrated that PknB mediated phosphorylation of GlmU does not affect its uridyltransferase activity, but significantly modulates the acetyltransferase activity.

2.3.2. *Recent research on GlmU inhibitors*

Recent studies on the mycobacterial proteome using *in silico* analysis suggested GlmU to be a potential drug target. To date, however, no inhibitors have been reported for the MTB enzyme.

In the year 1993 Mengin-Lecreulx D. and van Heijenoort J. identified the gene *GluM* which encodes N-acetylglucosamine-1-phosphate uridyl-transferase in *E coli*. and in 1994, they identified the bifunctions of GlmU [130].

Brown *et al.* [126] in 1999 first reported the crystal structures of GlmU. Based on the catalytic activity of GlmU, the bio-synthetic pathway of UDP-GlcNAc in prokaryotes and eukaryotes were

different. This was reported by Milewski *et al.* [131] in 2006. The main functions of GlmU were reported by Zhang *et al.* [129] in 2009 they reported the important residues in the binding site pocket.

Further in 2009, Pereira *et al.* [132] reported 63 identified hits using virtual screening technique, among which 37 were acetyltransferase active specified hits but they didn't reveal the structures of the hits. In 2011, Li *et al.* [133] designed and synthesized few cell wall inhibitors for both GlmM and GlmU shown in **Figure 32**. Only compound c showed inhibitory activity against GlmU and other three were inactive in both the cases.

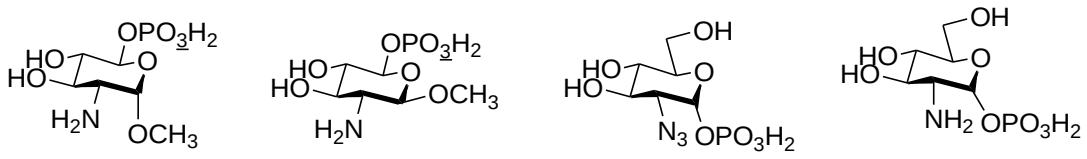


Figure 32: Cell wall inhibitors for both GlmM and GlmU

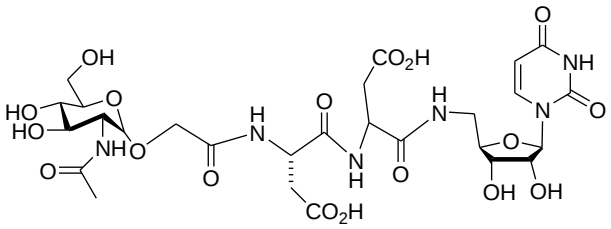
In 2011, Buurman *et al.* [134] designed some GlmU inhibitors for *Haemophilus influenza* and the sulfonamide derivatives showed some antimicrobial properties and they reported in the same year the *in vitro* studies of acetyltransferase activity of GlmU on *H influenza*. They proved the antibacterial activity of GlmU inhibitors. In this study GlmU acetyltransferase domain was found to act as molecular target. X-ray co-crystal structure and this study revealed the binding mode affinity of the inhibitors and their lack of potency against Gram-positive GlmU isozymes. This was the first report made on the growth inhibitory effects of antimicrobial compounds particularly through GlmU.

Zhou *et al.* [135] reported kinetic studies done on GlmU. Two colorimetric assays for the bifunctional GlmU enzyme activities were performed and its kinetic properties resulted for further studies like high-throughput screening of its inhibitors.

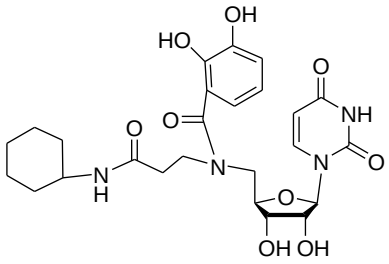
In 2011, using QSAR and docking studies Singla *et al.* [136] reported few hit compounds for GlmU. For this study they employed Pubchem Bioassay (AID 1376) results of 84 diverse compounds and performed docking studies using AutoDock software tool.

In 2013, Jagtap *et al.* [137] reported the crystal structures of MTB GlmU bound with substrate/products at the acetyl transferase active site and also found the uniqueness of GlmU from MTB possessing a 20 residue extension at the C terminus.

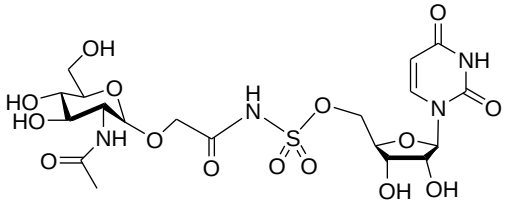
In 2013 Tran *et al.* [138] reported few inhibitors for MTB GlmU; which was the first inhibition study against MTB GlmU. To design the inhibitors they employed the substrate and putative transition state of uridyl transferase reaction as the beginning point. Two transition state mimics were designed but they showed very weak inhibition against the enzyme (10% and 60% inhibition *in vitro* at 2 mM (**Figure 33)**. In this compound 5and 6 were already reported by Larsen *et al.* [139] from Astra Zeneca for *H. influenza*. These compounds were tested for MTB GlmU uridyltransferase activity and they showed promising activity i.e. compound 5 showed 44% and compound 6 showed 36% inhibition at 50 μM concentration.



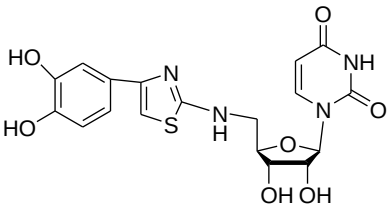
1. 60% Inhibition at 2 mM



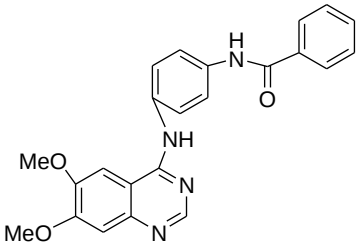
2. 10% Inhibition at 2 mM



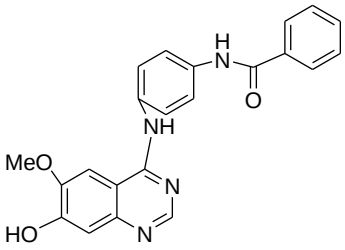
3. 55% Inhibition at 2 mM



4. 37% Inhibition at 2 mM



5. 44% Inhibition at 50 uM



6. 36% Inhibition at 50uM

Figure 33: First generation MTB uridyltransferase inhibitors

These two compounds 4 & 5 were used for SAR studies and based on these two compounds, few compounds were designed and synthesized as shown in **Figure 34.** The most potent compound showed the IC_{50} of 74 μM against GlmU uridyltransferase activity and this is the first reported inhibitor for MTB GlmU uridyltransferase activity.

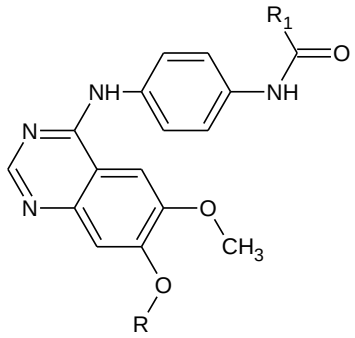


Figure 34: Modified structure scaffold using compounds 4 & 5

CHAPTER 3

OBJECTIVES AND WORK PLAN

From thorough literature survey it was very clear that there is a crucial need for new medication or new regimen for microbial targets like TB, HCV and HIV. Against these three targets, drugs available now are unable to sustain the growing inherent and emerging challenges of treatment. Due to the development of genome biology and medicinal chemistry, there are few identification of new targets and pathways for drug discovery.

Current HCV therapy faces important problems that are inadequate for sustained viral response, increasing effect of drug resistance and also lot of side effect which led to discontinuation of therapy. There are large numbers of available peptidic inhibitors with lack of non-peptidic especially small molecule inhibitors which are easy to synthesize and also may have fewer side effects. Current HIV therapy suffers from the rapid viral mutation and resistance to available drugs. This MDR-HIV is the recent challenge for HAART therapy. Among all type of inhibitors, HIV protease inhibitors play very potent action in inhibiting the HIV virus. Non-peptidic and small molecules show promising effect on virus which led to the development of new small molecule inhibitors for HIV protease. Current TB therapy also suffers from two main constrains like MDR-TB and TDR-TB. The causative organism can be killed by effective means of targetting and inhibiting the vital MTB enzymes. There are many new targets available with very few leads or no lead. A bi-functional enzyme MTB GlmU (N-Acetylglucosamine-1-phosphate uridyltransferase) is a new drug target for tuberculosis as evident from extensive literature survey.

An effective study involved in these targets could result in introduction of new type of drugs in the market which can overcome the problem of resistance, side effects and course of treatment. So

to identify new inhibitors there are two strategies structure based and ligand (3D QSAR) based drug design methods that were followed to screen the commercial and in-house databases to identify the potent leads.

Objectives

HCV NS5B polymerase & HIV protease inhibitors

- To develop multiple e-pharmacophores by employing different available crystal structures which are downloaded from PDB.
- Constructions of ligand based pharmacophores using available already reported HCV polymerase & HIV protease inhibitors.
- Validation of both e-pharmacophores and ligand based pharmacophores using enrichment calculations.
- Development of 3D QSAR model and its external validation.
- Combined virtual screening of databases using the selected pharmacophores and employing different stages of docking studies to identify the potent leads.
- Anti-HCV & anti-HIV activity studies.

MTB GlmU inhibitors

- Generation of ligand based pharmacophore using compounds from PubChem assay result.
- Validation of the pharmacophores.
- 3D QSAR model development and its validation.
- SAR studies and design of new leads based on the best active compound.
- Enzymatic and biochemical analysis of the synthesized compounds in MTB GlmU.

CHAPTER 4

MATERIALS & METHODS

4.1. Computational details

All computations were carried out on an Intel Core 2 Duo E7400 2.80 GHz capacity processor with a memory of 2GB RAM running with the RHEL 5.2 operating system. GLIDE 5.8 module of Schrödinger suite [140, 141] used to screen the public database like Asinex and in-house database. PHASE 3.4 [146-148] implemented in the Maestro 9.2 software package (Schrödinger, LLC) was used to generate pharmacophore and 3D QSAR models. Canvas 1.5 chemoinformatics package [151] of Schrödinger suite was employed to cluster the molecules.

4.2. Structure (e-pharmacophore) based approach

4.2.1. *Preparation of protein*

For protein targets whose crystal structures were available with their crystal ligands from PDB with good resolution were employed for the studies. Protein preparation wizard in Maestro which is part of the Maestro software package (Maestro, 9.3, Schrodinger, LLC, NY) was used to prepare the proteins [140, 141]. If the protein structure is a multimer with duplicate binding sites, the redundant site was removed by picking molecules or chains. If the binding interaction required both identical units to form active site pocket none were removed. Water molecules were removed. A brief relaxation was performed using an all-atom constrained minimization carried out with the Impact Refinement module (Impref) [142] using the OPLS-2005 force field to alleviate steric clashes that may exist in the original PDB structures. The minimization was terminated when the energy converged or the rmsd reached a maximum cutoff of 0.30 Å.

4.2.2. *Energy-optimized structure based pharmacophores: Hypothesis generation*

Receptor grid generation tool in Maestro software package was used to generate energy grids for all prepared protein structures. We docked the crystal ligands using the XP (extra precision) docking in Glide 5.8. [143-145]. Default settings were employed to minimize and optimize the structure. Based on the XP descriptor information pharmacophore features were generated using PHASE 3.4. PHASE [146-148] provided a built-in set of six pharmacophore features, hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively ionizable (N), positively ionizable (P), and aromatic ring (R). Hydrogen bond acceptor sites were represented as vectors along the hydrogen bond axis in accordance with the hybridization of the acceptor atom. Hydrogen bond donors were represented as projected points, located at the corresponding hydrogen bond acceptor positions in the binding site. Projected points allow the possibility for structurally dissimilar active compounds to form hydrogen bonds to the same location, regardless of their point of origin and directionality. Each pharmacophore feature site is first assigned an energetic value equal to the sum of the Glide XP contributions of the atoms comprising the site [149], allowing sites to be quantified and ranked on the basis of the energetic terms. Glide XP descriptors included terms for hydrophobic enclosure, hydrophobically packed correlated hydrogen bonds, electrostatic rewards, π-π stacking, π-cation, and other interactions. ChemScore, hydrogen bonding and lipophilic atom pair interaction terms were included when the Glide XP terms for hydrogen binding and hydrophobic enclosure were zero.

4.2.3. *Pharmacophore validation*

To prove the specificity and selectivity of a pharmacophore hypothesis we validated the best hypotheses with enrichment factor calculation. Ligand decoy sets were available for download (http://www.Schrödinger.com/glide_decoy_set). We generated a decoy database by using

Generate Phase Database sub application window from PHASE application. Decoy set consists of 1000 decoys and few known active molecules of interested target inhibitors. Decoy set was used to check how well the hypothesis was able to discriminate the active inhibitor compounds from other molecules [144], based on parameters such as total number of compounds in the hit list (H_t), number of active percent of yields (%Y), percent ratio of actives in the hit list (%A), Enrichment factor (EF) and goodness of fit (GH) were calculated using the pharmacophore mapping protocol. Equations (1) and (2) were used to calculate the EF and GH.

$$EF = \frac{(H_a \times D)}{(H_t \times A)} \quad \text{---------- (1)}$$

$$GH = \left(\left(\frac{H_a}{4H_tA}\right) \times (3A + H_t)\right) \times \left(1 - \left(\frac{H_t - H_a}{D - A}\right)\right) \quad \text{---------- (2)}$$

Where in 'H_t' represented total number of compounds in the hit list, 'H_a' was the total number of actives molecules in the hit list, 'A' was the total number of actives in the decoy set and 'D' was the total number of molecules in the decoy set. In addition to EF, parameters like RIE, ROC and BEDROC were in-built in the software and were also generated for reascertaining the pharmacophore validation as reported earlier [183].

4.3. 3D QSAR (Ligand) based approach

4.3.1. *Generation of datasets*

We chose few already reported inhibitors from the literature with known IC_{50} values for each of our targets. These inhibitors showed a wide range of activity (few nM–μM) and structural diversity. The LigPrep 2.5 [150] application from Schrödinger software package was utilized to build and energetically minimize structures and to add hydrogens and generate stereoisomers at neutral pH 7 using ionizer subprogram. Canvas 1.5 chemoinformatics package [151] was employed to cluster molecules based on tanimoto similarities between a set of linear fingerprint

descriptors to determine the structural diversity among the compounds. Finally a set of compounds were selected from clusters as representative molecules. Clustered molecules with structural diversity were utilized for 3D QSAR development. Set of already reported inhibitors were used as actives to validate the pharmacophores which were also prepared using LigPrep 2.5. The IC_{50} values were converted to pIC_{50} to get the linear relationship in the QSAR equation, using the following formula:

$$pIC_{50} = -\log IC_{50} \quad \text{---------- (3)}$$

where IC_{50} was the concentration of the compound producing 50% inhibition of that particular enzyme. The dataset consisted of some highly active and inactive molecules with few molecules as moderately active.

4.3.2. ***Molecular alignment and development of pharmacophore model for 3D QSAR approach***

To develop 3D QSAR model, pharmacophore models and statistical analyses were performed using PHASE. Conformers were generated using a Macromodel torsion angle search approach followed by minimization of each generated structure using OPLS-2005 as force field with implicit distance dependent dielectric solvent model. A maximum of 1000 conformers were generated per structure using MacroModel search method (ConfGen) [152] in the preprocess minimization of 100 steps, including post process minimization of 50 steps. Each generated conformers were further filtered using a relative energy scale of 10.0 kCal/mol and a RMSD of 1 Å. After conformers generation step there was an essential step of creating pharmacophore sites on each ligand structure responsible to facilitate non-covalent binding interaction between the ligand and the receptor. The threshold range of the active and inactive pIC_{50} was selected. Structural alignment is one of the most important parameter in 3D QSAR analyses. The accuracy of the prediction of 3D QSAR models and the reliability of the contour models depend strongly

on the structural alignment. In the development of 3D QSAR approach, pharmacophore models and statistical analyses were performed using PHASE [146-148]. After conformers generation step there was an essential step of creating pharmacophore sites on each ligand structure responsible to facilitate non-covalent binding interaction between the ligand and the receptor. Next, 'find common pharmacophore' step generated identical sets of features with similar spatial arrangements which group to form a common pharmacophore hypothesis with requirement that all active must match. Pharmacophore features were created using the clean minimized structure. Last step in pharmacophore generation was 'scoring hypothesis' in which hypotheses were ranked to make rational choices among the hypotheses and the most appropriate one for further exploration. Common pharmacophores were examined by a scoring protocol to identify the pharmacophore from each surviving n-dimensional box that yielded the best alignment of the active set ligands. The inactive molecules were scored to observe the alignment of these molecules with respect to the pharmacophore hypothesis to enable selection of the hypothesis. Larger the difference between the scores of active and inactives better was the hypothesis in distinguishing the actives from inactives. The final selected ligands were aligned with the best pharmacophore template of compound with high active score. These EF and GH based validated pharmacophores were further validated by building a 3D QSAR model and by external statistical validation.

4.3.3. *3D QSAR modeling*

PHASE QSAR models may be either atom-based or pharmacophore-based, the difference being whether all atoms were taken into account, or merely the pharmacophore sites that can be matched to the hypothesis. We used the selected promising compounds to develop our 3D QSAR model. We divided randomly few compounds for training set and for test set by using the method

"Automated Random Selection" option present in the PHASE module. PHASE provided the means to build QSAR models using the activities of the ligands that match a given hypothesis. PHASE QSAR models were based on PLS regression, applied to a large set of binary valued variables which were individually derived from a regular grid of cubic volume elements, with each cubic element is represented by a set of bit values (0 or 1) to account for the different type of pharmacophore features in the training set [153]. The independent variables in the QSAR model were derived from a regular grid of cubic volume elements that span the space occupied by the training set ligands. The ligand set used were diverse structures and hence pharmacophore-based QSAR models were generated for the best pharmacophore hypothesis using the selected molecules of training set and a grid spacing of 1 Å. QSAR models containing few PLS factors were generated. A model with a PLS factor showing high R^2 and Q^2 (> 0.5), less standard deviation value (< 0.3), less RMSE value and high F-value was considered as the best statistical model. This model was validated by predicting activities of test set molecules [154].

4.3.4. *PLS analysis and external statistical validation of QSAR models:*

All 3D QSAR models were generated by using significant statistical method of partial least square analysis. The cross validation analysis was performed using the leave one out (LOO) method which evaluated the predictive ability of QSAR model [155]. The cross validated coefficient, r^2_{cv} (also called as LOO-q^2) was calculated using the following equation:

Formula:

$$r^2_{cv} = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - Y_{mean})^2} \qquad \text{---------- (4)}$$

Where Ypred, Yobs and Ymean were the predicted, observed and mean values of the target property (pIC_{50}) respectively. $(Y_{obs}\text{-}Y_{mean})^2$ was the predictive residual sum of squares (PRESS). The predictive correlation coefficient (r^2_{pred}), based on molecules of test set and was defined by,

$$r^2pred = \frac{SD-PRESS}{SD}$$ ----------(5)

Where SD was the sum of the squared deviations between the biological activities of the test set and mean activities of the training set molecules. PRESS represented the sum of squared deviation between predicted and actual activity values for every molecule in the test set.

Based on the reported external validation methods of 3D QSAR, we evaluated the true predictive abilities of the generated models of target of interest inhibitors. To validate the true predictivity of the established models, it was crucial to perform external validation. According to literature [155], 3D QSAR models were accepted if they satisfied all of these following conditions:

$r_{cv}^2 > 0.5, r^2 > 0.6,\ R_0^2$ or $R_0'^2$ close to r^2, i. e$[[(r2 - R_0^2)/ r2]] < 0.1$, or $[(r2 - R_0'^2)/ r2] < 0.1,\ 0.85 \leq k \leq 1.15, and\ r^2m > 0.5$

The r^2 value was calculated using the following formula:

$$R = \frac{\sum(Y_i-\bar{Y}o)(\tilde{Y}_i-\bar{Y}p)}{\sqrt{\sum(Y_i-\bar{Y}o)^2 \sum(\tilde{Y}_i-\bar{Y}p)^2}}$$ ---------- (6)

In these above equation, y_i and $\tilde{y}_i$ were the observed and predicted activity, $\bar{Y}o$ and $\bar{Y}p$ were the average values of the observed and predicted pIC_{50} values of the test set molecules.

If we plot y (observed activity) versus $\bar{y}$ (predicted activity) for the ideal QSAR model, the regression line will bisect the angle formed by positive directions of the orthogonal axes $\bar{y}_i$ and y_i. The regression line is expressed by $y^r = a\bar{y}+b$,

Where

$$a = \frac{\sum(y_i-\bar{y})(\tilde{y}_i-\bar{\tilde{y}})}{\sum(\tilde{y}_i-\bar{\tilde{y}})^2}$$ ---------- (7)

and

$$b = (\bar{y} - a\bar{\tilde{y}})$$ ---------- (8)

In the above equation (7) and (8), $\bar{y}$ and $\bar{\tilde{y}}$are the average values of observed and predicted activities respectively, and the summation indicates the overall n compounds in the test set.

The slopes k and k' [156] were calculated using the following equations:

$$k=\frac{\in y_i\tilde{y}_i}{\in \tilde{y}^2} \quad \text{----------(9)}$$

$$k' = \frac{\in y_i\tilde{y}_i}{\in y_i^2} \quad \text{----------(10)}$$

Regression lines which were passing through the origin defined by $y^{ro}= k\tilde{y}$ and $\tilde{y}^{ro}= k'y$ which were not close to the optimum regression lines $y^r= a\bar{y}+b$ and by $\tilde{y}^r= a'\bar{y}+b'$. Correlation coefficients for these lines were R_0^2 and $R'_0{}^2$ have different values which can be calculated using the following formulae:

$$R_0{}^2 = 1 - \frac{\sum(\tilde{y}_i - y^{ro})^2}{\sum(\tilde{y}_i - \bar{\tilde{y}})^2} \quad \text{---------- (11)}$$

$$R'_0{}^2 = 1 - \frac{\sum(y_i - \tilde{y}^{ro})^2}{\sum(y_i - \bar{y})^2} \quad \text{---------- (12)}$$

The summations indicated for overall n compounds in the test set.

For better external predictive potential of the model, a parameter of modified r^2 [$r_m^2(LOO)$] was an very important factor , which can be used for the whole set considering LOO-predicted values for the training set and predicted values of the test set compounds. The r_m^2 (LOO) statistic equation for overall test and training set values was generally used for selection of the best predictive models from among comparable models [157, 158]. Substantiation of the particular QSAR models with $r_m^2(LOO)$ value can be defined by the following equation:

$$r_m^2(loo) = r^2\left(1 - \sqrt{|r^2 - R_0^2|}\right) \quad \text{---------- (13)}$$

3D QSAR model with good predictive ability and its respective pharmacophore were selected among the top pharmacophore models and were used to find newer leads.

Further the integrity of the model was predicted by r^2_{pred} for test set with the value of >0.5.The accepted leave-one-out (LOO) -cross validated value of training set (R^2) should be greater than 0.5, LOO cross validated value for test set (Q^2) [159] should show a value greater than 0.5 to attain good predictive capacity, and standard deviation (SD) below 0.3, with minimum root mean square error (RMSE), and high value of variance ratio (F) to provide conventional QSAR validation limits. And the predictive correlation coefficient (r^2_{pred}) value generated based on molecules of test set should demonstrate real predictive capacity and robustness of the QSAR model [160-162].

4.4. High-Throughput Virtual screening (HTVS) and Molecular Docking

For e-pharmacophore approach, sites with score above -1.0 kcal mol^{-1} were energetically favoured. The minimum three of site points should be matched. We set 2.0Å as intersite distance matching tolerance for both strong and weak alignment. Pharmacophore screening hits were ranked based on their fitness score which gave the knowledge of how well each conformer aligned to the pharmacophore hypothesis calculated from the knowledge of vector score, volume score, rmsd, no of site matched, energy. Fitness score value range from 0 to 3 as given in the default pharmacophore screening in Phase was empolyed. Database ligands were docked into the binding sites of the crystal structures in the active site regions with Glide 5.8 [143] utilizing HTVS and scoring function to estimate protein–ligand binding affinities. Glide HTVS was faster and more tolerant to suboptimal fits than Glide XP, making it better for comparison in this work. The center of the Glide grid was defined by the position of the co-crystallized ligand. Default settings were used for both the grid generation and docking. Post docking minimization was

implemented to optimize the ligand geometries. Compounds with best docking and Glide scores were then subjected to Glide SP and XP screening [163].

The energy-based and ligand based pharmacophore models were validated using enrichment analysis, and all selected models were further employed for the high-throughput virtual screening and docking. The filtered hits on the basis of the selected pharmacophore models were subjected to molecular docking and application of filtering criteria pertaining to their interaction with amino acids in the active site pocket and 3D QSAR predictions. The hit compounds were finally short-listed for testing against selected target activity.

4.5. ADME predictions

The lead compounds were evaluated for the pharmaceutically relevant properties to check drug likeness and predictions for drug's pharmacokinetics in the human body including its ADME. QikProp module [164] was used for evaluation of drug-like behavior through analysis of pharmacokinetic parameters required for absorption, distribution, metabolism and excretion (ADME).

4.6. Compounds Details

Few lead compounds were procured from Asinex (ASINEX, ASINEX Platinum Collection, ASINECorp, Winston-Salem, NC, USA. http://www.asinex.com) and few lead compounds were procured from commercial outsourcing the synthesis and screened for biological assay. The purity of the compounds was found to be ≥95%.

4.7. Biological assays

4.7.1. HCV NS5B inhibitory activity

All compounds were made soluble in DMSO at 10 mM stock and further diluted just prior to evaluations. Huh7/Rep-Feo1b cells were grown at 37°C with 5% CO_2 in Dulbecco's Modified

Eagle's Medium (DMEM) containing 10% FBS, 1% penicillin/streptomycin, 1% glutamine, and 500 µg/ml of G418. IC_{50} values were calculated using CalcuSyn software (Biosoft).

4.7.1.1. Anti-HCV activity and cytotoxicity in cell culture

Anti-HCV activity of compounds in cell culture was evaluated in the Huh7/Rep-Feo1b cells replicating the subgenomic HCV genotype 1b replicon RNA carrying the firefly luciferase reporter. The evaluation of compounds was performed similarly as described earlier [165]. In brief, the cells were grown in 98 well plates and treated with compounds or equal amount of DMSO (control). Final concentration of DMSO was 1%. After 48h of treatment, cells were washed with phosphate buffered saline and inhibitory effect of compounds on HCV replication was measured as the level of firefly luciferase activity using luciferase assay system (promega).

To evaluate the cytotoxic effects of compounds, Huh7/Rep-Feo1b cells were treated with compounds or DMSO (control) in 96 well plates similarly as above. After 48 h post treatment, the cells were supplemented with 20 µl CellTiter 96 AQ_{ueous} One solution and cell viability was measured by following manufacturer's protocol (Promega, USA) [166].

4.7.1.2. NS5B RdRp assay

HCV NS5B carrying a deletion of 21 hydrophobic amino acids from C terminal and bearing hexa-histidine tag was recombinantly purified using Ni-NTA column chromatography as described earlier [167]. RdRp assay was performed similarly as described previously in a final volume of 20 µl [168]. In brief, the assay containing 20 mM Tris–HCl (pH 7.0), 100 mM NaCl, 100 mM sodium glutamate, 0.5 mM DTT, 0.01% BSA, 0.01% Tween-20, 5% glycerol, 0.5 µM of pre-annealed template/primer (poly rA/U_{12}), 25 µM UTP, 1–2 µCi [α-^{32}P]UTP and ~300 ng of NS5B was incubated with compounds or DMSO (control) on ice for 5 minute and reactions were started by adding 0.5 mM $MnCl_2$. The reactions were incubated at 30°C for 1h and reactions were

terminated by adding ice cold 5% TCA solution. The precipitated acid insoluble materials were filtered on glass fiber filter and washed successively with 5% TCA, water and ethanol. The filters were air dried and placed in a vial containing EcoLite scintillation cocktail and the amount of incorporated radioactive nucleotides were counted using scintillation counter. NS5B activity in the presence of DMSO was set at 100% and that in the presence of the compounds was determined relative to this control.

4.7.2. Biological assay for HIV protease:

In first step, assay buffer (2X) 5ml, 1M DTT (1000X) 10µL and deionized water 5mL made the total volume 10mL. HIV-1 protease substrate (50X) of volume 100µL and assay buffer 4.9mL were mixed these reagents and every time prepared as fresh substrate solution. Protease was diluted as necessary step. The recommended volume for HIV-1 protease diluent was 40µL/well (96-well late). Test compounds were diluted with deionized water with the volume of 10µL/assay (96-well plate). 20µL of DMSO was added into one vial of saquinavir to get 2mM concentration by vortexing to dissolve completely and added this to 2µM assay buffer. 10µL (96-well plate) per well of 2µM saquinavir was prepared. Saquinavir solution was highly unstable either it should store in -20^{o}C or freshly prepared [169].

4.7.2.1. HIV protease enzymatic reaction

The positive control contained HIV-1 protease diluent without the test compound. Inhibitor control contained of protease diluent and known inhibitor Saquinavir. In vehicle control we use DMSO as vehicle which was used to dissolve test compound. Pre-incubation of the plate at 37°C for 10-15 min was done and HIV-1 protease substrate was also incubated for 37°C. To initiate the enzymatic reaction 50 µL per well of HIV-1 protease substrate solution was added and the reagents were mixed well by shaking the plate gently for 30-60 sec. Then the fluorescence

intensity was measured immediately at Ex/Em= 340±30 nm and recorded continuously for every 5 min for 30-60 min. To get end point reading, the reaction mixture was incubated at room temperature for 30-60 min and the plate was kept way from direct light. The intensity was measured at Ex/Em= 340±30 nm [170, 171].

4.7.3. GlmU acetyltransferase assay

Acetyltransferase activity assays were carried out in a 8 µl reaction volume containing 25 mM HEPES buffer (pH 7.6), 10 mM MgCl2, 1 mM DTT, 0.1 mM GlcN-1-P, 0.2 mM acetyl-CoA, 0.002 µCi of [14C] acetyl-CoA, and GlmU. Reactions were terminated by the addition of 2 µl of 2 M formic acid, and the samples were resolved on polyethylenimine cellulose TLC plates. Radioactive spots corresponding to [14C] GlcNAc-1-P and unincorporated [14C] acetyl-CoA was visualized by phosphorimaging.

4.7.3.1. Expression and purification of MTB GlmU

M.tuberculosis GlmU has 495 amino acids long. The gene encoding GlmU from *E.coli* (DH10B) and *M. tuberculosis (H37Rv)* was amplified using respective genomic DNA as a template. Amplicon obtained was digested with NdeI and HindIII cloned into corresponding sites in pQE2 (Qiagen) expression vector. Expression plasmids were freshly transformed into *E. coli* BL21 (DE3) CODON PLUS cells. Fresh transformants were grown in LB broth at 37ºC with constant aeration, in the presence of 100 µg/ml ampicillin. Exponentially growing cultures (A600 of ~0.6) were induced with 0.2 mM IPTG and further grown for 12-16 h at 18ºC. Cells were harvested and lysed by sonication in lysis buffer. The cell lysate containing His6-fusion proteins were mixed with equilibrated Ni-NTA agarose affinity resins. His tagged proteins were eluted with buffer containing 25 mM Tris-HCl (pH 8.0), 140 mM NaCl, 5% glycerol, 1 mM DTT, 100 mM PMSF

and 250 mM imidazole. Purified dialyzed protein was estimated by Bradford protein estimation method [172, 173].

CHAPTER 5

DESIGN OF HCV NS5B INHIBITORS: RESULTS & DISCUSSIONS

5.1. Background

Nonstructural protein 5B (NS5B) is the RNA-dependent RNA-polymerase from Hepatitis C Virus which is responsible for HCV replication. This NS5B RNA dependent RNA polymerase synthesizes RNA using a RNA template, which is absent in mammalian cells and hence becomes an important and attractive target to develop drugs for the treatment of this lethal disease. The current HCV therapy suffers from inadequate sustained viral response rate, rapid emergence of drug resistance, in particular for patients infected with genotype 1 HCV, along with significant side effects, resulting in discontinuation of therapy [174, 175]. Given the absolute requirement of NS5B to synthesize nascent HCV RNA, NS5B represents an attractive target for the development of anti-HCV inhibitors. Therefore, an inhibitor that blocks NS5B RdRp activity should, in theory, have minimal or no effect on host biological processes. HCV NS5B is a 66 kDa protein with a hydrophobic C-terminal membrane anchorage domain, which poses a challenge for full-length NS5B purification. Consequently, structure function investigations on NS5B have been achieved employing recombinant NS5B with 21 to 55 amino acid C-terminal truncations, which does not compromise its enzymatic activity. Crystal structures of NS5B have revealed a classical "right-hand" topology of the polymerase family, with discernable fingers, palm and thumb subdomains. An unusual feature of NS5B is the extensive interactions between its finger and thumb subdomains, resulting in an encircled active site [66]. Several crystal structures of NS5B either alone or in complex with diverse ligands ranging from ions and small molecules to nucleotides and non-nucleoside analogue inhibitors (NNI), have been elucidated [176-179]. The NNIs are

reported to bind to five distinct allosteric pockets on NS5B located on its thumb or palm domain [180, 181]. These structures offer an excellent platform for the identification and development of new NS5B inhibitors.

5.2. Structure (e-pharmacophore) based drug design (Work flow is presented in Figure 38)

The HCV NS5B polymerase crystal structures with low resolution and with high active bound inhibitors were retrieved from PDB. Here, we used a total of five crystal structures with a resolution between 1.65 Å - 2.31 Å and activity of inhibitors (IC_{50}) ranging from ≤10 nM to 291 nM to enhance the chances of retrieving similarly active and diverse hits from the database screening. The crystal ligand structure with resolution and the affinities are presented in **Figure 35.** Three of the crystal structures (3CVK [177], 3CO9 [178], and 2GIQ [182] represented in complex with inhibitors in the palm region and the other two (2GIR [182] and 2D3Z [179]) represented inhibitors bound to the thumb region. The bound crystal ligand of 3CVK was pyridazinone derivative and showed H-bond intaractions with Asn291, Asp318 and Tyr448. The bound crystal ligand of 3CO9 was also pyridazinone derivative and showed similar H-bond interaction like 3CVK. 2GIQ crystal ligand was quinolone derivative having H-bond with Tyr448. The bound inhibitor of 2D3Z and 2GIR of thumb region was non-nucleosidic thiophene derivative with H-bond interactions with Ser476 and Tyr477. This showed that Asn291, Asp318 and Tyr448 were the important H-bond interactions for palm region and for thumb region Ser476 and Tyr477. When we shortlisted our hits we would prefer the compounds which were showing these interactions as prior condition.

We prepared these proteins using protein preparation wizard by adding hydrogens, and energy minimization by converging the heavy atoms to RMSD 0.30 Å using OPLS2005 as force field, removed water molecules and finally refined the protein structure shown in **Figures 36 & 37**. All

the crystal ligands were redocked onto the prepared protein to generate e-pharmacophore. The root mean square deviation (rmsd) was less than 1Å for all the 5 crystal ligands. Ligand interaction analysis showed that the important residues in palm region were Asp318, Asn291 and Tyr448 and for thumb region Ser476 and Tyr477.

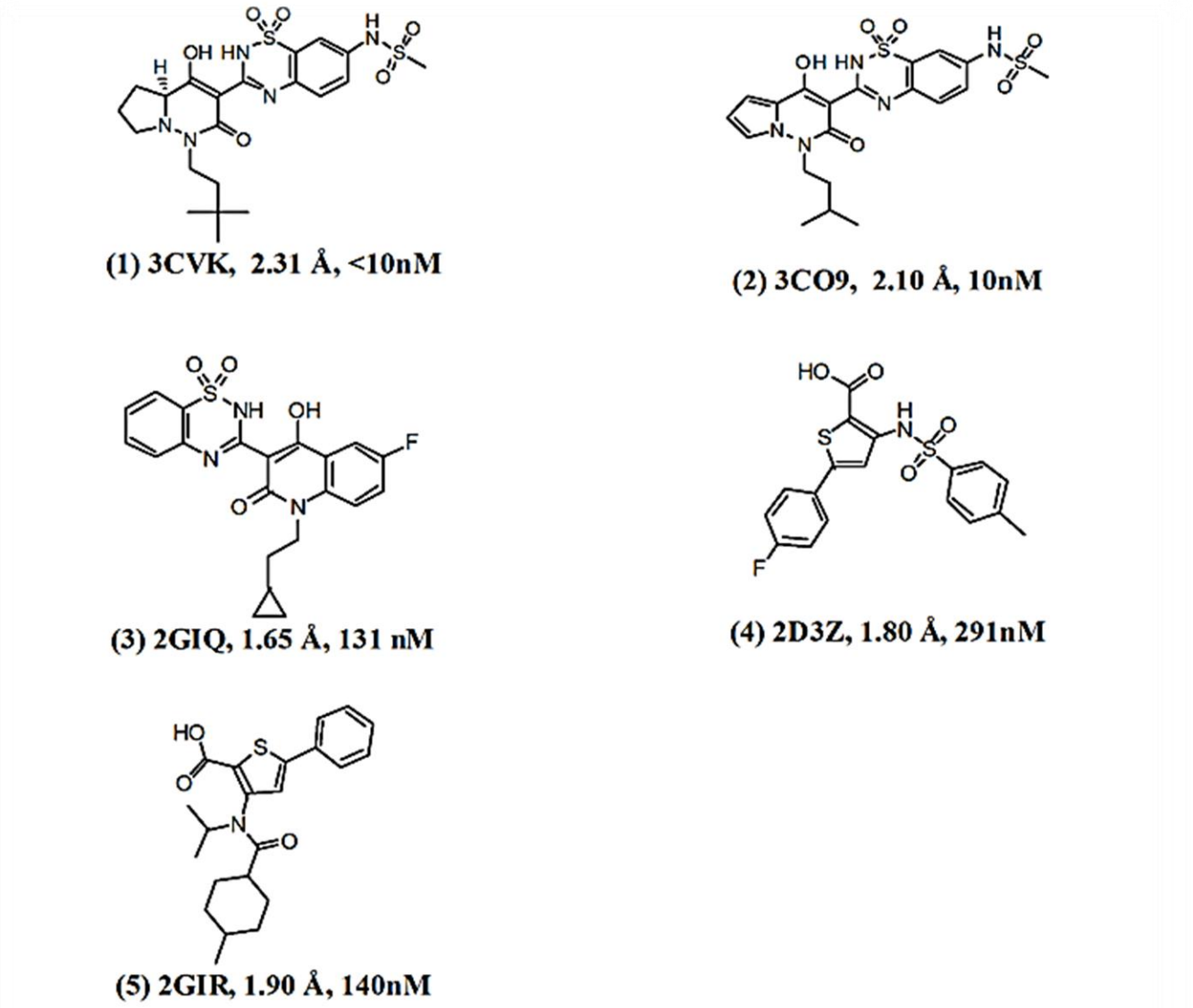


Figure 35: Structures of bound inhibitors listed with their respective PDB IDs, IC_{50} values and resolutions for HCV NS5B target

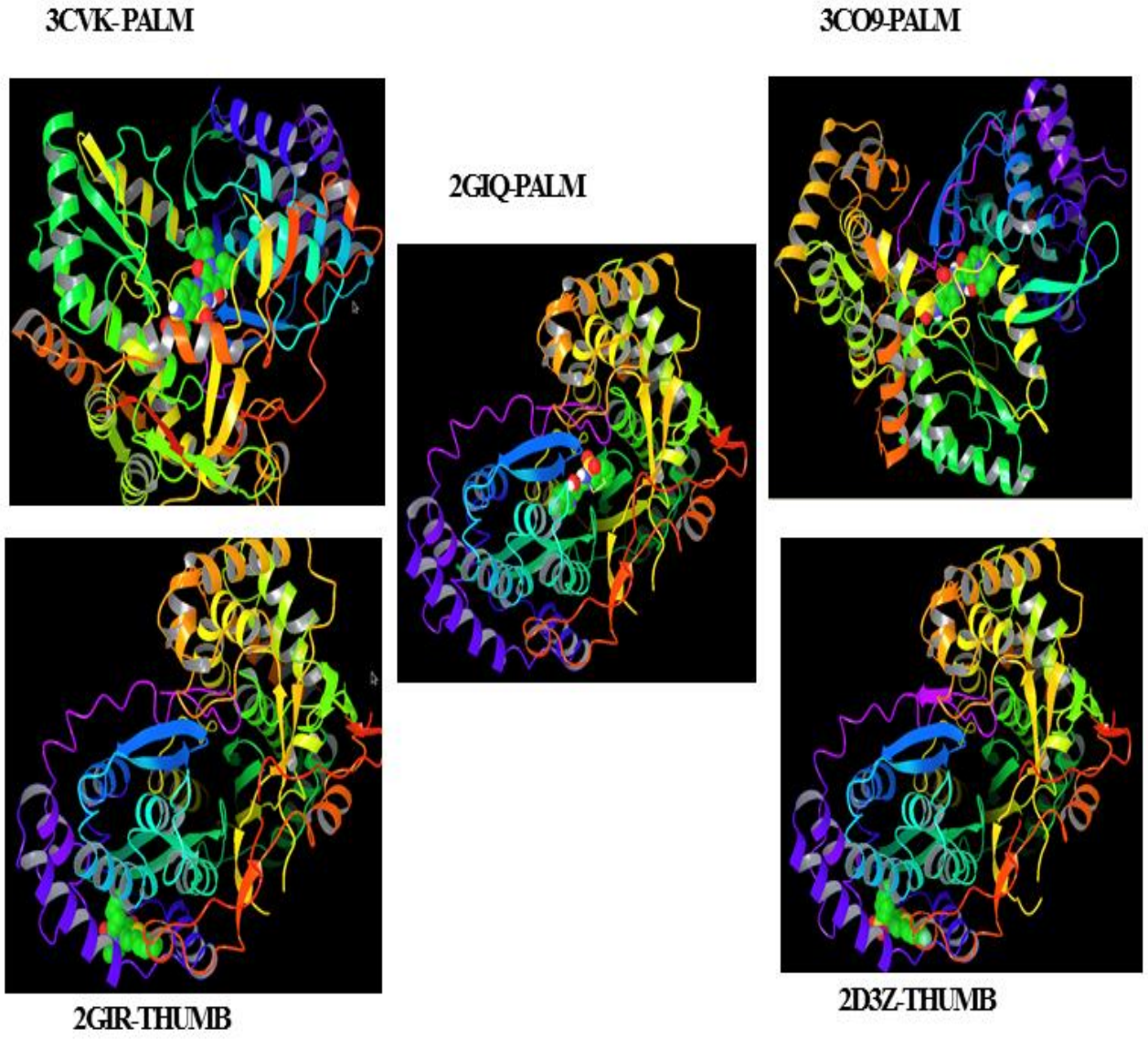


Figure 36: The five selected PDB structures with their respective IDs and their bound crystal ligand for HCV NS5B polymerase

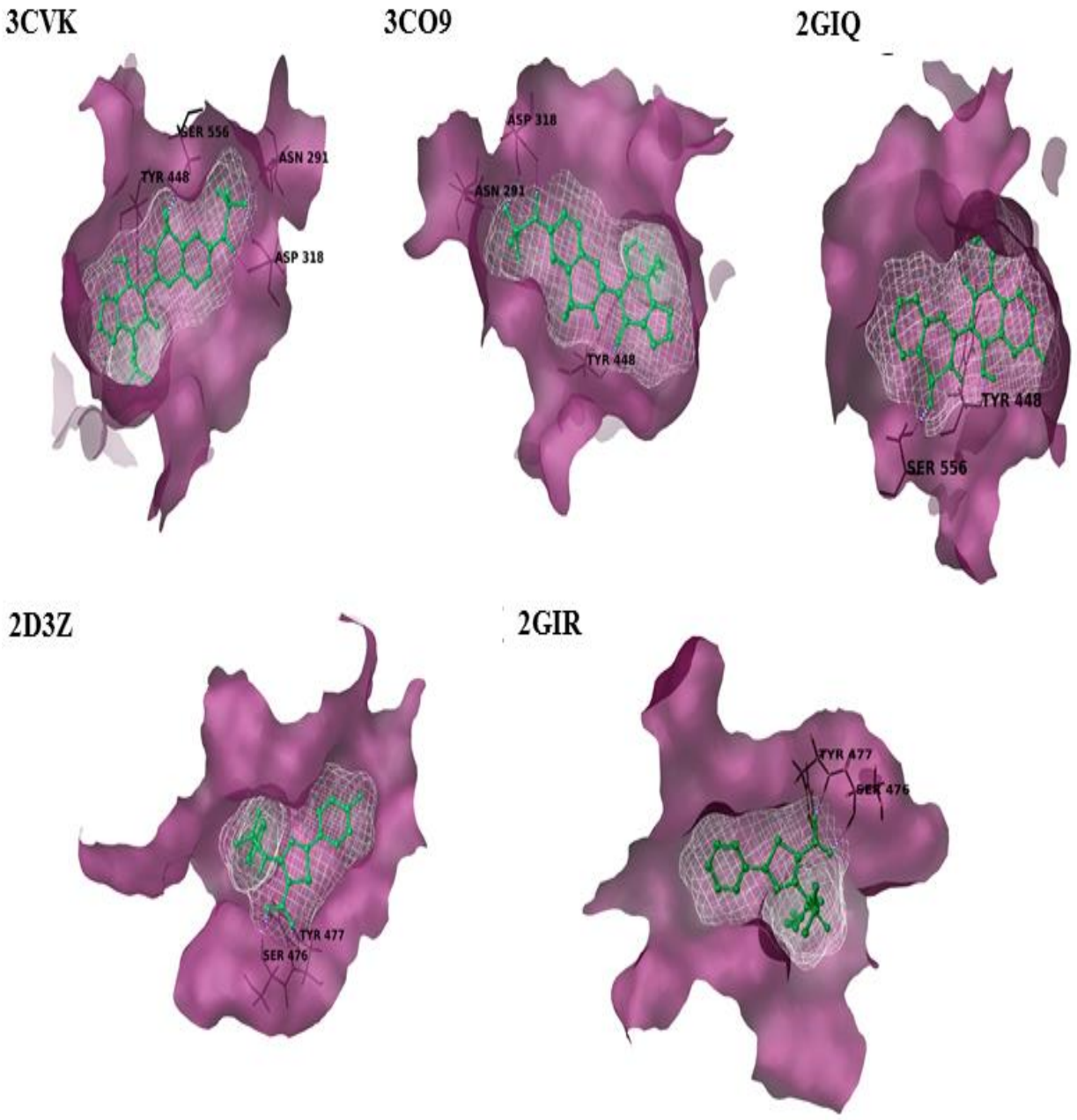


Figure 37: The binding site pocket of the five indicated crystal structures, demonstrated in white and magenta color created using 'Create binding site surfaces' module of Schrodinger suite for HCV NS5B polymerase

Receptor Grid Generation tool in Maestro software package was used to generate energy grids for all prepared protein structures. The Grid table is given in **Table 7.**

Table 7: Grid information along with their PDB IDs employed for docking studies

S.No	PDB code	X-Centre	Y-Centre	Z-Centre
1.	3CVK	96.2294	9.309	54.8521
2.	3CO9	33.5356	44.5081	53.9472
3.	2D3Z	9.325	33.7615	74.9164
4.	2GIQ	9.177	-8.3279	-12.7974
5.	2GIR	-33.8597	-18.8336	33.3814

The pharmacophores were generated, validated and used for further studies like virtual screening and multiple steps docking studies as given in the work flow for structure based drug design in the **Figure 38**.

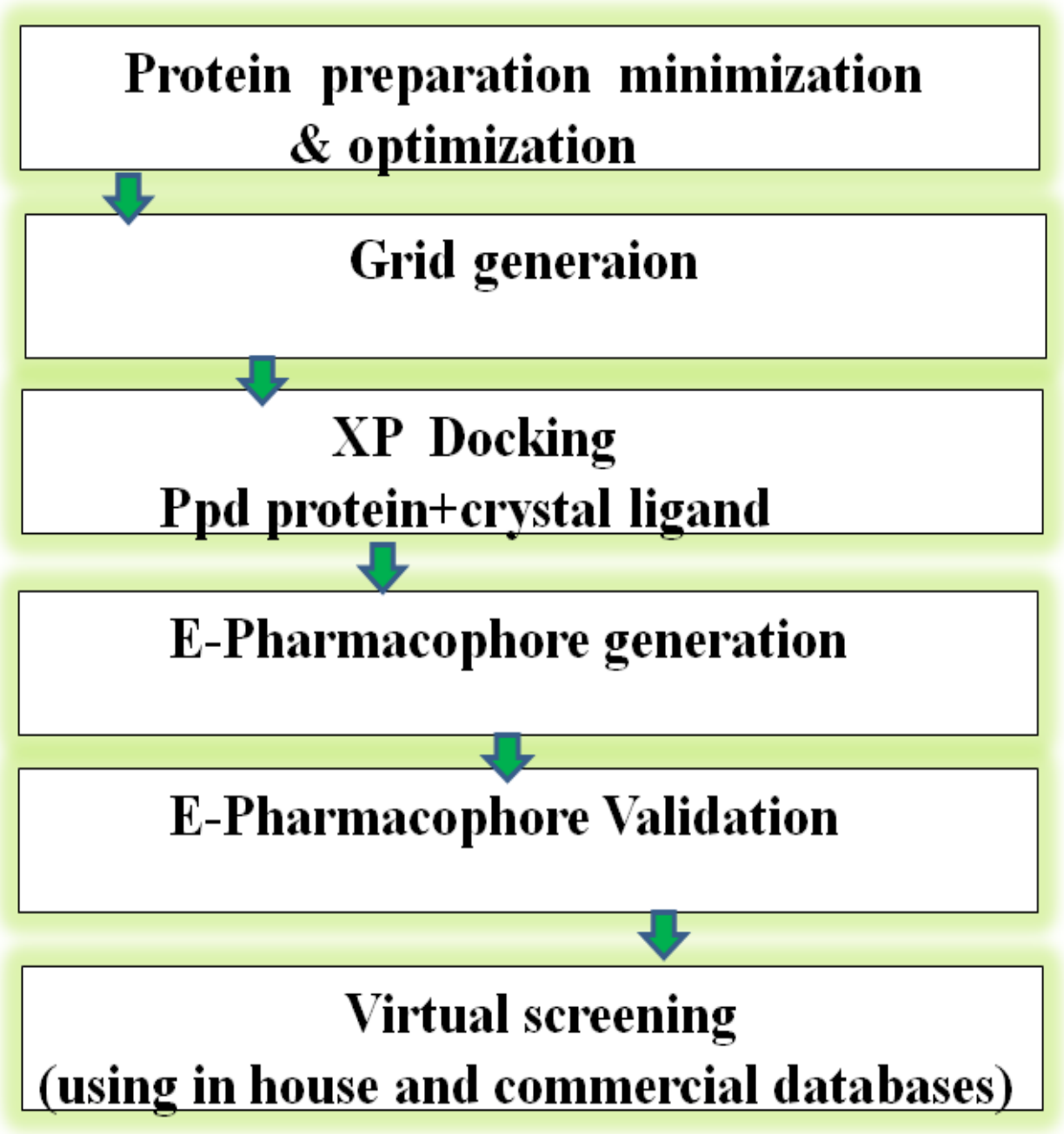


Figure 38: Work flow for structure based approach

5.2.1. Energy-based Pharmacophore Generation and its validation

We docked the crystal ligand using XP None refine option in Glide module. Using this XP descriptor information pharmacophore hypotheses were generated. Each feature of

pharmacophore hypotheses was assigned with energy values; lower the energy value higher the stablility of the feature as shown in **Table 8**. We kept 10 as maximum number of features. The e-pharmacophore method combined the aspects of structure based and ligand based techniques and was explored for 5 crystal structures of HCV NS5B polymerase employed in this study.

Table 8: The possible features and their energy score for each PDB IDs

PDB	Features*	Score
3CVK	A7	-0.66
	D10	-0.33
	R15	-0.59
3CO9	A7	-0.66
	D9	-0.45
	R14	-0.62
2GIQ	A5	-0.23
	R11	-0.82
	R13	-0.59
2D3Z	R9	-1.27
	N6	-0.69
	R8	-1.07
	R7	-0.60
2GIR	N5	-0.70
	H2	-0.22
	R7	-0.96

*****A-Acceptor, D-Donar, R-Ring aromatic, N-Negative ionic and H-Hydrophobic

The pharmacophore hypotheses were developed by mapping Glide XP energetic terms onto pharmacophore sites which were calculated based on the structural and energy information between the protein and the ligand **Table 8**. The optimized sites for hypothesis generation for the 5 crystal structures of NS5B are represented in **Figure 39**.

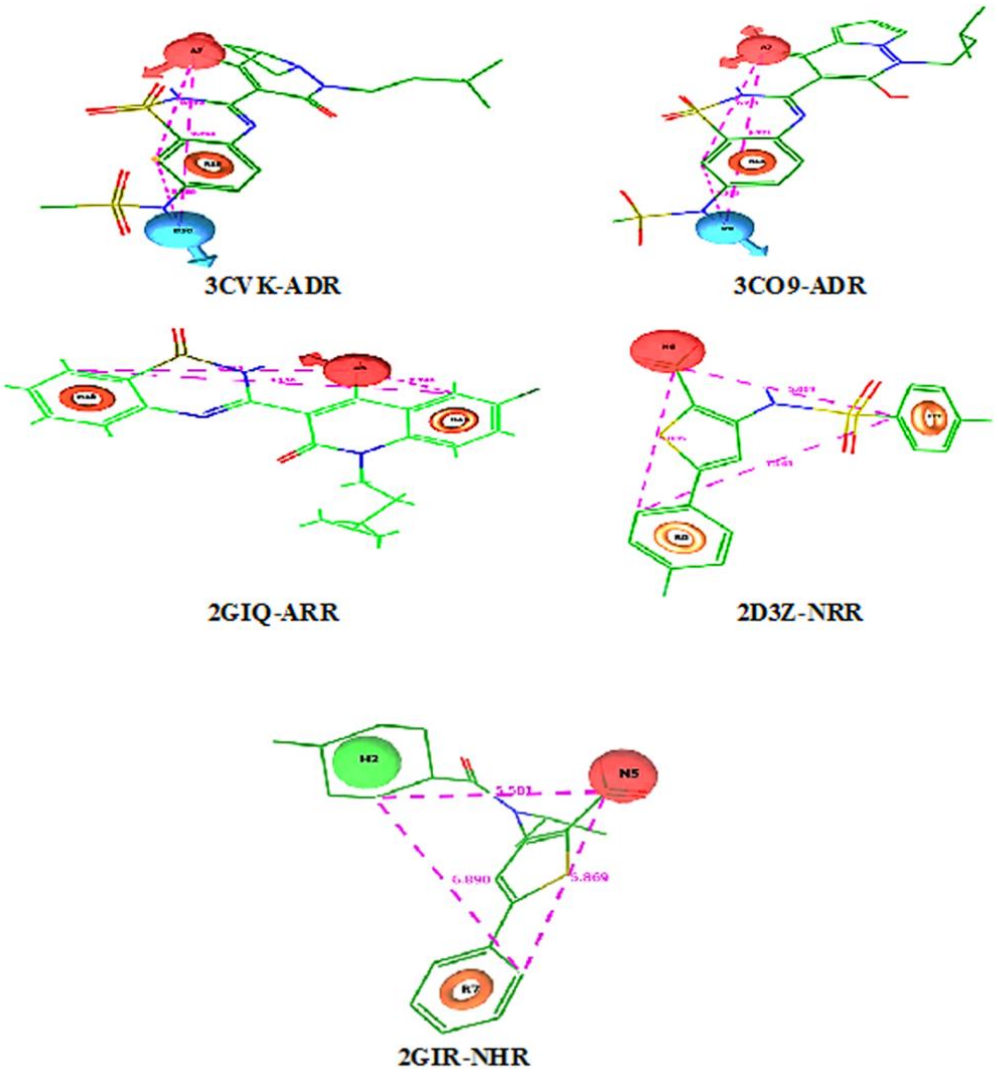


Figure 39: Finally selected five e-pharmacophores. Pink sphere with arrow-hydrogen bond acceptor (A), yellow open circle-aromatic ring (R), blue sphere with arrow-hydrogen bond donor (D), green sphere-hydrophobic (H), pink sphere- negatively ionizable (N) for HCV NS5B polymerase

Maximum number of features available for each crystal structures and the selected features are given in **Table 9.**

Table 9: The possible number of available pharmacophore sites from each PDB structure, the final selected sites and with their final selected hypotheses

S.No	PDB code	No of possible sites	No of selected sites	Hypothesis*
1.	3CVK	3	3	ADR
2.	3CO9	3	3	ADR
3.	2D3Z	4	3	ARR
4.	2GIQ	3	3	NRR
5.	2GIR	3	3	NHR

*A-Acceptor, D-Donar, R-Ring aromatic, N-Negative ionic and H-Hydrophobic.

The pharmacophoric features for the palm region were ADR (3CVK and 3CO9) and ARR (2GIQ), for thumb region NRR (2D3Z) and NHR (2GIR). One Ring (R) feature was common in all the hypotheses. All the crystal structures yielded three feature pharmacophore. The maximum distance between the features was 8.489 Ǻ and the minimum distance was 2.341 Ǻ. The distance mapping among the features is presented in **Table 10**. Acceptor (A) and Ring (R) features were common among the pharmacophoric sites derived from palm domain crystal structures (3CVK, 3CO9 and 2D3Z), where in thumb domain crystal structures (2GIQand 2GIR) Ring (R) and negatively ionizable (N) feature were common.

Table 10: Distance between the features of e-pharmacophores

S.NO	PDB CODE	A-R (Ǻ)	R-D (Ǻ)	D-A (Ǻ)	N-R (Ǻ)	H-R (Ǻ)	R-R (Ǻ)	R-A (Ǻ)	N-H (Ǻ)	R-N (Ǻ)
1.	3CVK	3.751	4.754	8.489	NA	NA	NA	3.751	NA	NA
2.	3CO9	3.73	4.727	7.924	NA	NA	NA	3.73	NA	NA
3.	2GIQ	5.31	NA	NA	NA	NA	6.757	2.341	NA	NA
4.	2D3Z	NA	NA	NA	5.033	NA	6.683	NA	NA	5.313
5.	2GIR	NA	NA	NA	5.324	5.447	NA	NA	4.184	5.324

Distance shown in Å, A-Acceptor, D-Donar, R-Ring aromatic, N-Negatively ionizable and H-Hydrophobic. NA-not applicable.

The distance between A and R when compared among the palm region structures, the ligands of 3CVK and 3CO9 were similar (3.73 Å) while 2GIQ (**Table 10)** generated two R features and the

distances with A feature were 5.31 Å and 2.34 respectively. This depicted that the e-pharmacophore derived from 2GIQ was very much different from other e-pharmacophores of the palm region and hence had the possibility of yielding diverse structure libraries. We also compared the 1% EF, RIE, ROC, BEDROC (α=20) and (α=160). Enrichment factor indicated the fraction of actives retrieved from a database with both actives and inactives (Decoys) [144] after screening. High EF 1% meant top 1% of actives retrieved in the top 1% decoys which gave hope that pharmacophore model could screen good number of active molecule from the database we used. One more enrichment metrics we used was the Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC). BEDROC was a generalization of the receiver operating characteristic (ROC) that gave the knowledge of the "early scoring problem" by Boltzmann weighing the hits based on how early they were retrieved. We had estimated both α=20.0 and α=160.9 for comparisons. To evaluate virtual screening results, α=20.0 was the best choice because it accounted the 80% results in top 1% of the database. The enrichment results for all 5 crystal ligands using the e–pharmacophore method were compared for the enrichment factor (EF1%), BEDROC (α=160.9 and α=20.4) [183], based on recovery rate of actives against the ranked decoy database that consisted of 1090 compounds in which 90 were known inhibitors. The results are as shown in **Table 11** where 3CO9 and 2GIR pharmacophores showed the highest enrichment at 1% (EF1%) and BEDROC values. And the overall enrichment factor (EF) and goodness of fit (GH) was better for 2GIR and 2D3Z. Finally 5 potential e- pharmacophore hypotheses were selected for virtual screening of compound databases.

Table 11: Validation of e-pharmacophore hypotheses

PDB code	e-Pharmacophore features	EF 1%	RIE	ROC	BEDROC (α-20)	EF	GH
3CVK	ADR	2.2	1.01	2.33	0.274	0.501	0.045
3CO9	ADR	7.7	1.02	1.55	0.668	0.513	0.047
2GIQ	ARR	6.6	0.96	2.46	0.507	0.760	0.060
2D3Z	NRR	3.3	1.04	1.97	0.190	1.568	0.129
2GIR	NHR	7.7	1.02	4.4	0.591	2.663	0.240

EF- Enrichment factor at 1% of the decoy dataset
RIE- Robust initial enhancement
ROC-Receiver operating characteristic curve value
BEDROC- Boltzmann-enhanced discrimination of receiver operating characteristic
EF- Overall enrichment factor
GH- Goodness of fit

5.3. Ligand based drug design

The workflow for ligand based design is indicated in **Figure 40.**

We retrieved 1568 NS5B inhibitors from various literatures [176-179, 184-229] with known IC_{50} values. These inhibitors showed a wide range of activity (0.04 nM – 500 μM) and structural diversity. All ligands were prepared, energetically minimized and optimized. Finally 132 compounds were selected from 151 clusters as representative molecules [176-179, 184-229]. The compounds with their observed activity, predicted activity, residual error and their QSAR set are given in **Table 12**. The threshold range of the active and inactive pIC_{50} was 9.398 and 3.301 respectively. Selected 132 promising compounds were used to develop our 3D QSAR model. The work flow for ligand based drug design is shown in the **Figure 40**. The structures of the compounds employed for the study are given in **Figure 41**.

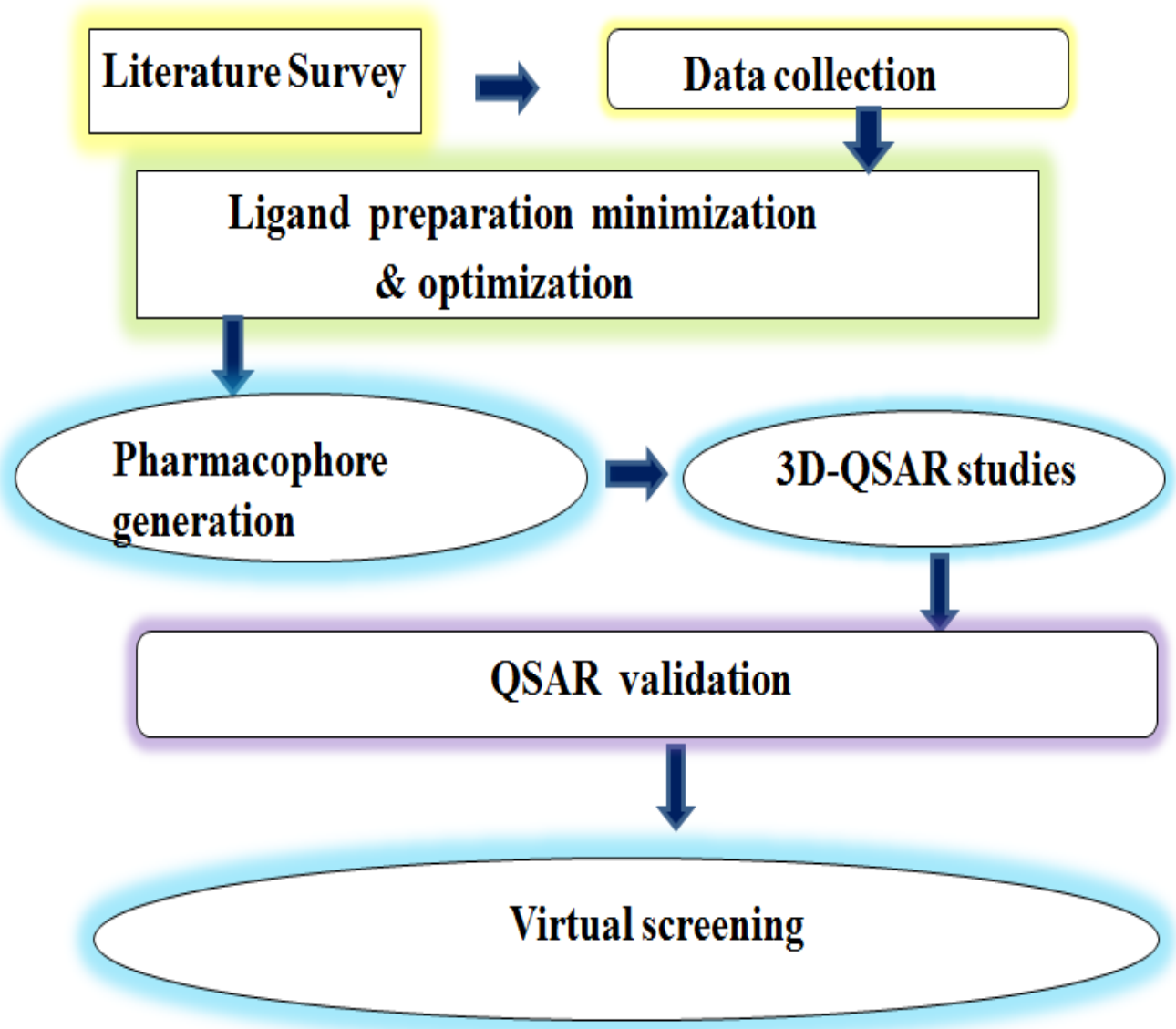
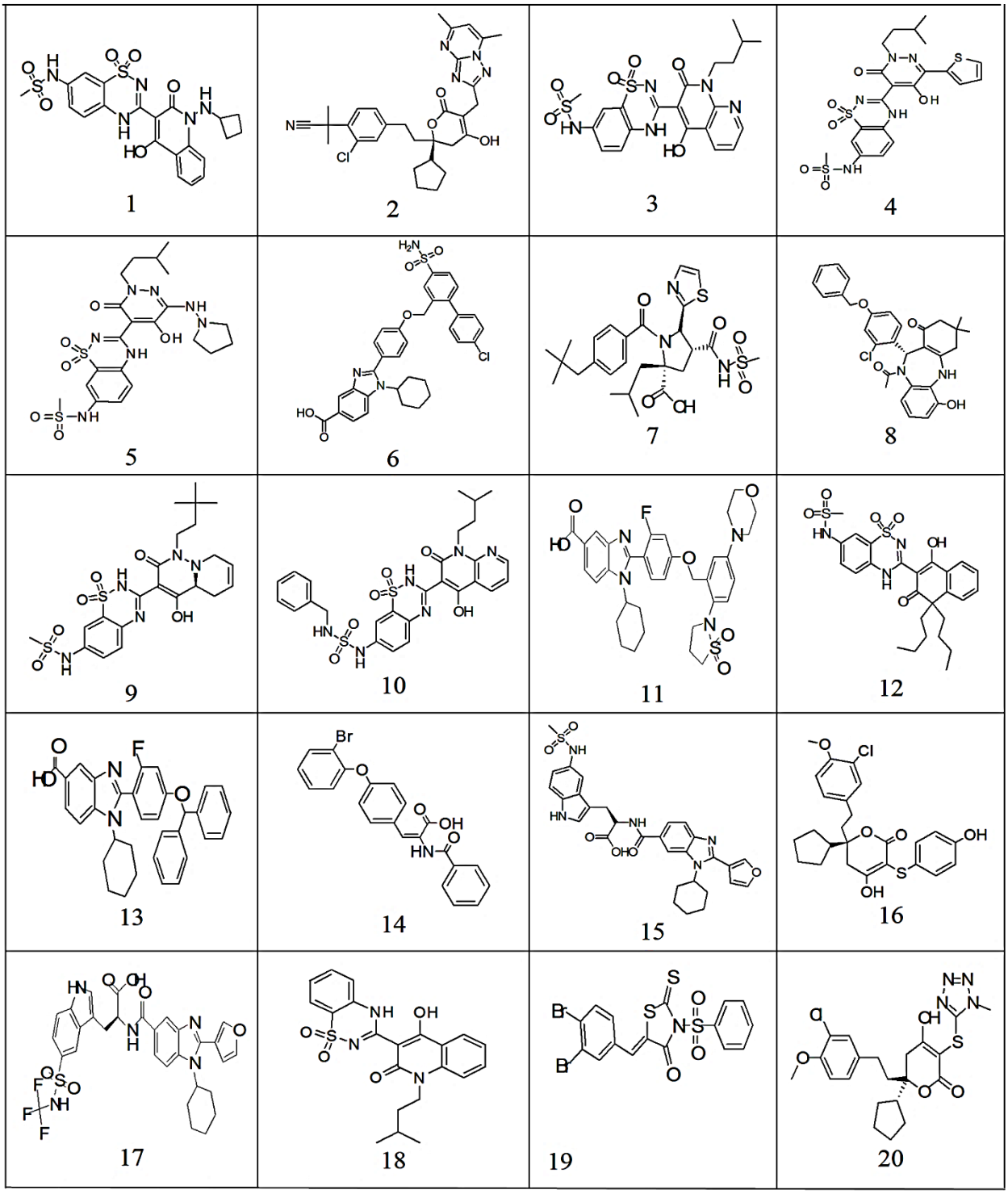
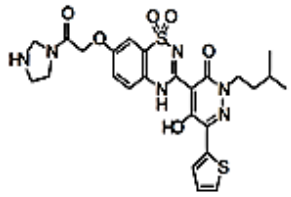
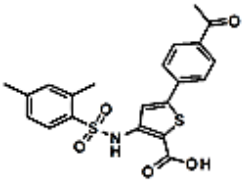
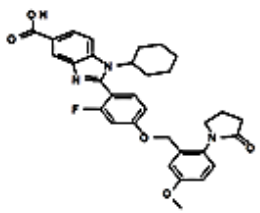
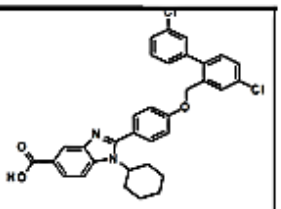
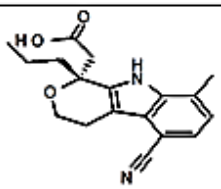
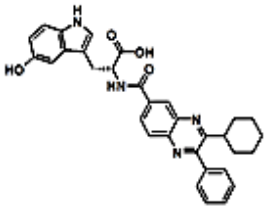
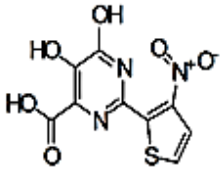
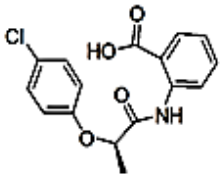
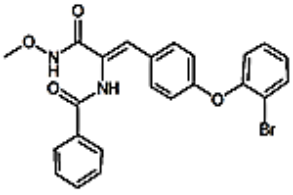
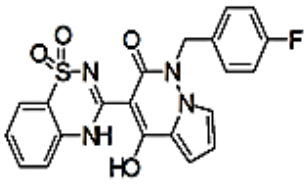
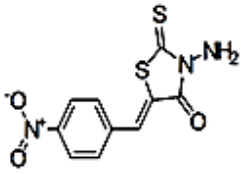
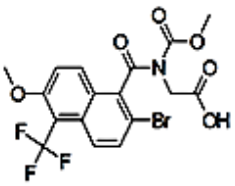
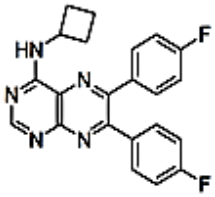
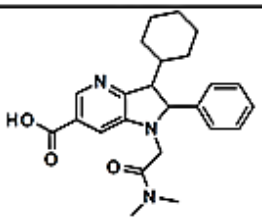
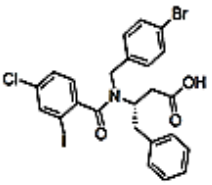
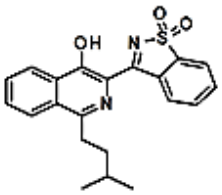
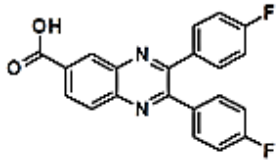
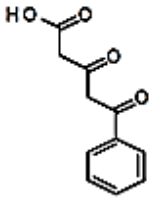
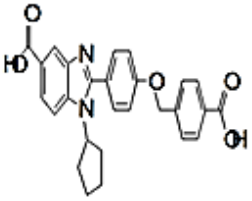
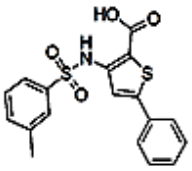


Figure 40: Work flow for 3D QSAR (ligand) based approach

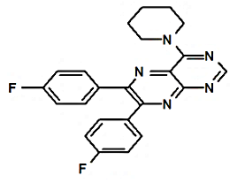
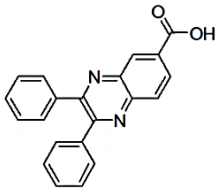
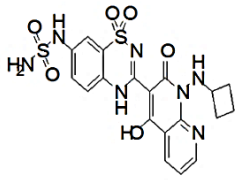
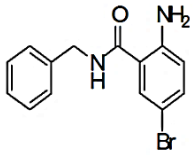
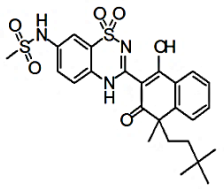
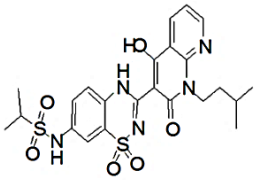
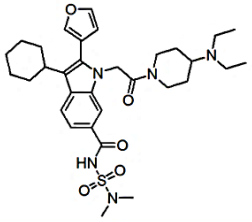
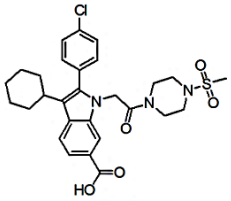
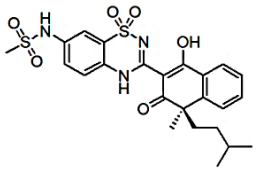
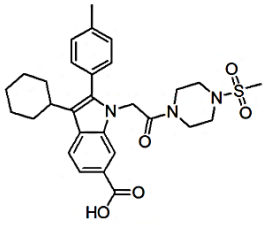
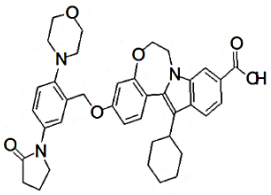
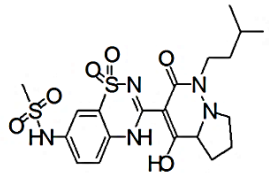
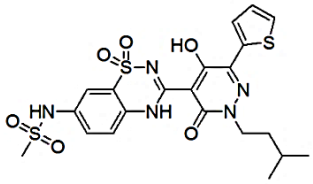
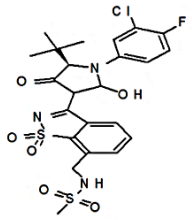
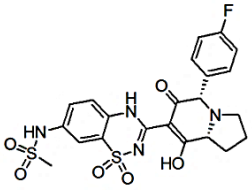
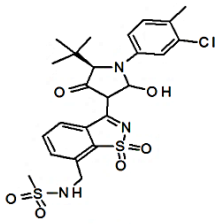
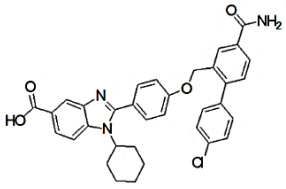
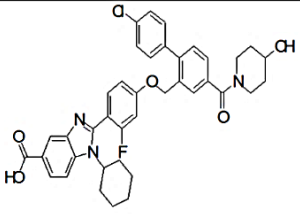
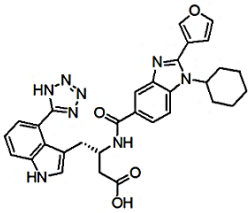
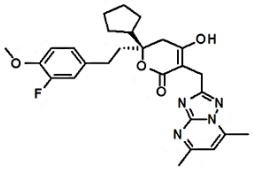
Figure 41: Compounds utilized for ligand based pharmacophore and 3D QSAR model for HCV NS5B polymerase



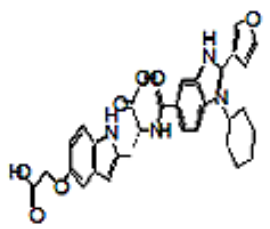
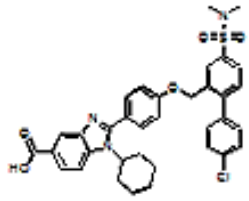
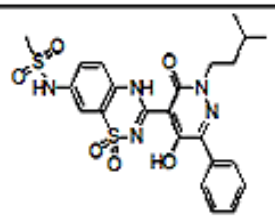
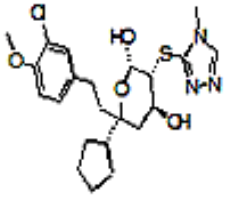
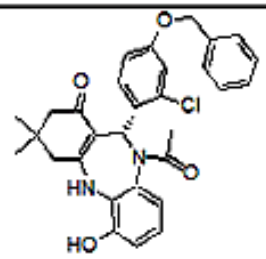
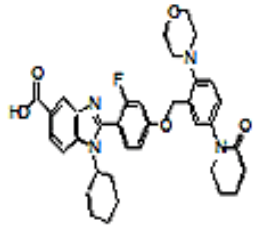
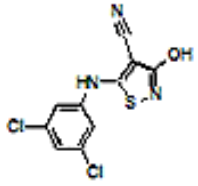
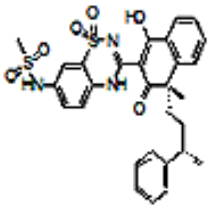
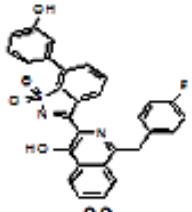
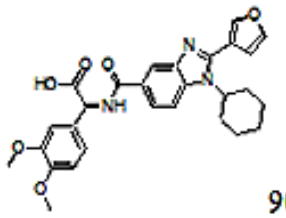
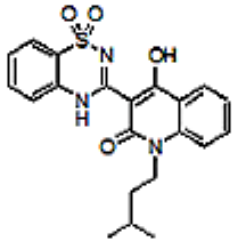
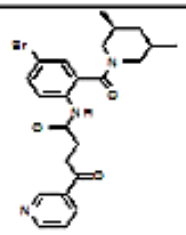
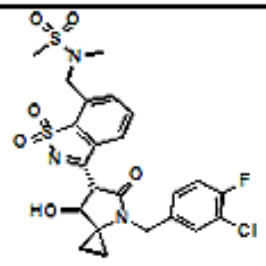
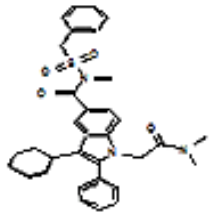
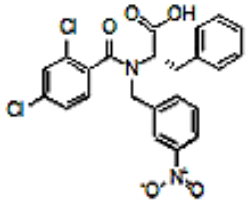
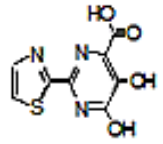
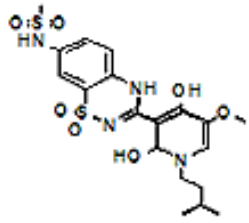
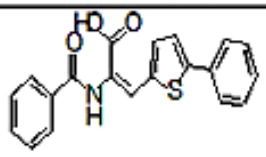
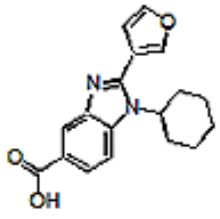
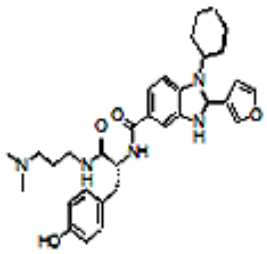
Contd..

 21	 22	 23	 24
 25	 26	 27	 28
 29	 30	 31	 32
 33	 34	 35	 36
 37	 38	 39	 40

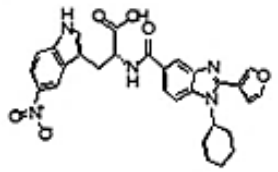
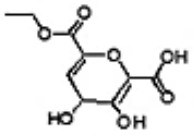
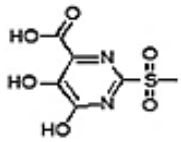
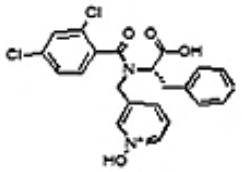
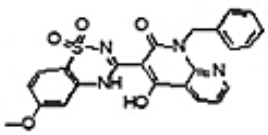
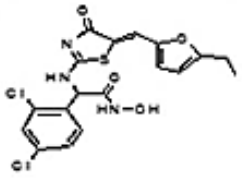
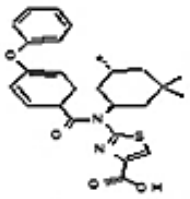
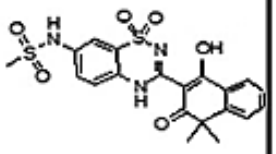
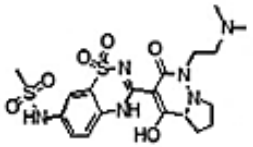
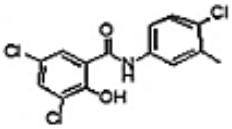
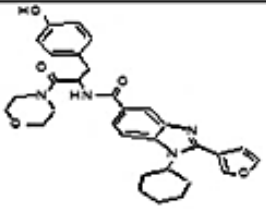
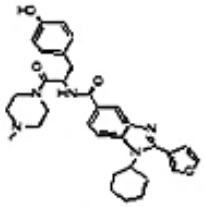
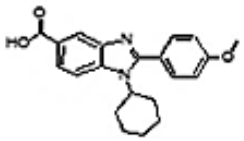
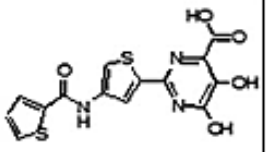
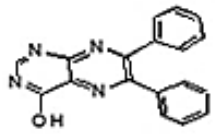
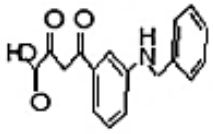
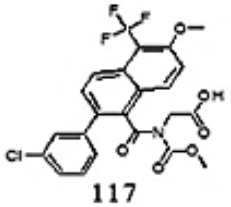
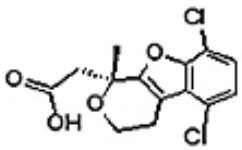
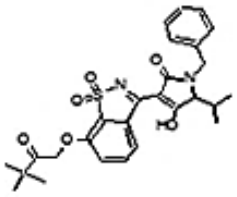
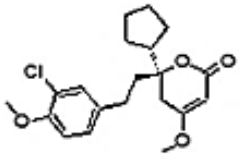
Contd..

 61	 62	 63	 64
 65	 66	 67	 68
 69	 70	 71	 72
 73	 74	 75	 76
 77	 78	 79	 80

Contd..

 81	 82	 83	 84
 85	 86	 87	 88
 89	 90	 91	 92
 93	 94	 95	 96
 97	 98	 99	 100

Contd..

 101	 102	 103	 104
 105	 106	 107	 108
 109	 110	 111	 112
 113	 114	 115	 116
 117	 118	 119	 120

Contd..

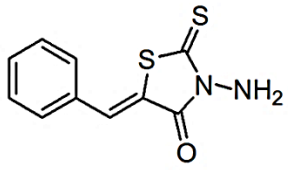
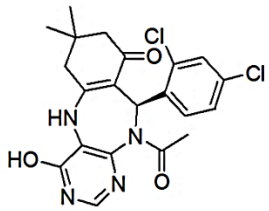
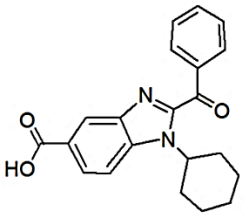
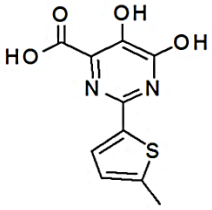
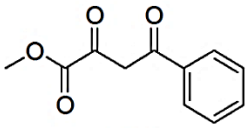
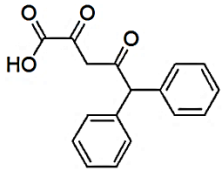
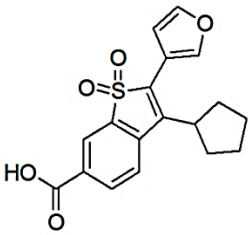
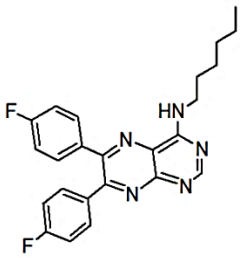
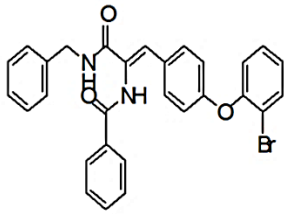
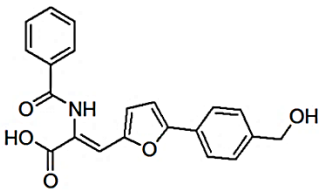
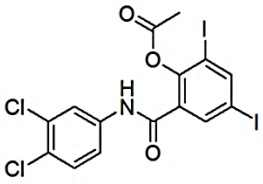
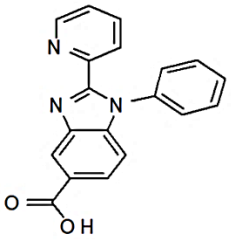
 121	 122	 123	 124
 125	 126	 127	 128
 129	 130	 131	 132

Table 12: 3D QSAR data set of 132 compounds with their fitness score, experimental activity, predicted activity and the difference between their predicted and actual activity

ID	Fitness	Experimental activity	phase predicted activity	Error	Phase qsar set
1.	2.967	8.796	8.170	0.626	Test
2.	1.485	8.699	6.890	1.809	Test
3.	2.924	8.155	8.160	-0.005	Test
4.	2.213	8.00	7.230	0.770	Test
5.	2.371	7.854	6.550	1.304	Test
6.	1.307	7.796	7.120	0.676	Test
7.	1.366	7.398	6.020	1.378	Test
8.	1.053	7.398	7.010	0.388	Test
9.	2.342	7.387	6.400	0.987	Test
10.	2.784	7.357	8.360	-1.003	Test
11.	1.398	7.149	7.050	0.099	Test
12.	2.745	7.137	7.070	0.067	Test
13.	1.299	7.018	6.270	0.748	Test
14.	1.282	7.000	6.140	0.860	Test
15.	1.564	6.921	6.490	0.431	Test
16.	1.044	6.886	6.060	0.826	Test
17.	1.481	6.824	7.470	-0.646	Test
18.	2.363	6.699	6.300	0.399	Test
19.	1.525	6.699	5.380	1.319	Test
20.	1.237	6.553	6.290	0.263	Test
21.	1.972	6.469	5.510	0.959	Test
22.	1.619	6.409	6.200	0.209	Test
23.	1.339	6.387	6.940	-0.553	Test
24.	1.404	6.319	6.000	0.319	Test
25.	1.558	6.244	5.190	1.054	Test
26.	1.220	6.222	5.950	0.272	Test
27.	1.513	6.167	4.650	1.517	Test
28.	1.457	6.119	5.330	0.789	Test
29.	1.490	6.097	5.960	0.137	Test
30.	2.252	6.071	5.430	0.641	Test
31.	1.107	6.000	4.410	1.590	Test
32.	1.446	5.854	5.150	0.704	Test
33.	1.612	5.678	5.240	0.438	Test
34.	1.327	5.658	4.940	0.718	Test
35.	1.611	5.569	5.340	0.229	Test
36.	1.744	5.347	6.290	-0.943	Test
37.	0.984	5.260	4.680	0.580	Test
38.	1.445	5.244	4.590	0.654	Test

Contd..

39.	1.281	5.174	6.280	-1.106	Test
40.	1.653	5.131	5.360	-0.229	Test
41.	1.546	4.959	5.020	-0.061	Test
42.	1.599	4.886	5.380	-0.494	Test
43.	1.534	4.854	4.140	0.714	Test
44.	1.581	4.839	5.110	-0.271	Test
45.	1.598	4.796	5.790	-0.994	Test
46.	1.467	4.699	5.220	-0.521	Test
47.	1.296	4.699	5.800	-1.101	Test
48.	1.434	4.638	4.370	0.268	Test
49.	1.578	4.569	5.600	-1.031	Test
50.	1.582	4.538	5.440	-0.902	Test
51.	1.556	4.523	5.330	-0.807	Test
52.	1.099	4.523	4.490	0.033	Test
53.	1.585	4.509	5.700	-1.191	Test
54.	0.926	4.495	4.670	-0.175	Test
55.	1.621	4.357	4.970	-0.613	Test
56.	0.832	4.357	5.320	-0.963	Test
57.	1.459	4.31	4.670	-0.360	Test
58.	1.664	4.301	4.570	-0.269	Test
59.	1.631	4.301	5.370	-1.069	Test
60.	1.265	4.301	5.780	-1.479	Test
61.	1.574	4.167	5.200	-1.033	Test
62.	1.255	4.102	4.500	-0.398	Test
63.	3.000	9.398	8.850	0.548	Training
64.	2.739	9.222	8.150	1.072	Training
65.	2.899	8.523	8.490	0.033	Training
66.	1.524	8.398	8.280	0.118	Training
67.	0.658	8.222	8.110	0.112	Training
68.	2.664	8.155	7.870	0.285	Training
69.	1.325	8.155	8.590	-0.435	Training
70.	1.587	8.143	8.170	-0.027	Training
71.	2.359	8.000	7.770	0.230	Training
72.	2.197	8.000	7.720	0.280	Training
73.	1.743	8.000	8.350	-0.350	Training
74.	1.643	8.000	8.150	-0.150	Training
75.	1.439	8.000	7.840	0.160	Training
76.	1.228	7.886	7.820	0.066	Training
77.	1.464	7.854	7.900	-0.046	Training
78.	1.485	7.824	7.490	0.334	Training
79.	1.368	7.824	7.740	0.084	Training
80.	1.476	7.721	7.550	0.171	Training
81.	1.403	7.638	7.750	-0.112	Training
82.	2.099	7.509	7.600	-0.091	Training
83.	0.979	7.420	7.710	-0.290	Training

Contd..

84.	1.044	7.301	7.270	0.031	Training
85.	1.351	7.046	6.950	0.096	Training
86.	1.513	6.699	6.730	-0.031	Training
87.	2.658	6.553	6.830	-0.277	Training
88.	1.336	6.456	6.490	-0.034	Training
89.	1.311	6.398	6.440	-0.042	Training
90.	2.364	6.377	6.300	0.077	Training
91.	1.818	6.229	6.500	-0.271	Training
92.	1.320	6.208	6.340	-0.132	Training
93.	1.806	6.194	6.360	-0.166	Training
94.	1.637	6.155	6.180	-0.025	Training
95.	1.579	6.119	6.400	-0.281	Training
96.	2.442	6.097	6.250	-0.153	Training
97.	1.595	5.824	6.040	-0.216	Training
98.	1.445	5.796	5.860	-0.064	Training
99.	1.161	5.721	5.840	-0.119	Training
100.	1.604	5.699	5.710	-0.011	Training
101.	1.385	5.648	5.710	-0.062	Training
102.	1.666	5.62	5.590	0.030	Training
103.	0.851	5.569	5.400	0.169	Training
104.	2.328	5.409	5.650	-0.241	Training
105.	1.173	5.26	5.440	-0.180	Training
106.	1.066	5.26	4.930	0.330	Training
107.	2.820	5.203	6.530	-1.327	Training
108.	2.409	5.167	5.690	-0.523	Training
109.	1.405	5.046	5.150	-0.104	Training
110.	1.354	5.022	5.020	0.002	Training
111.	1.440	4.959	4.880	0.079	Training
112.	1.326	4.951	5.040	-0.089	Training
113.	1.720	4.86	4.950	-0.090	Training
114.	1.425	4.824	4.440	0.384	Training
115.	1.602	4.777	4.820	-0.043	Training
116.	1.370	4.699	4.810	-0.111	Training
117.	1.295	4.699	4.760	-0.061	Training
118.	1.568	4.678	4.480	0.198	Training
119.	1.288	4.538	4.500	0.038	Training
120.	1.103	4.523	4.470	0.053	Training
121.	0.848	4.495	4.270	0.225	Training
122.	1.363	4.357	4.290	0.067	Training
123.	1.509	4.313	4.640	-0.327	Training
124.	1.624	4.301	4.110	0.191	Training
125.	1.466	4.301	4.430	-0.129	Training
126.	1.708	4.000	3.750	0.250	Training
127.	1.546	4.000	4.070	-0.070	Training
128.	1.531	4.000	4.030	-0.030	Training

Contd..

129.	0.659	4.000	3.830	0.170	Training
130.	1.442	3.699	3.640	0.059	Training
131.	1.261	3.444	3.310	0.134	Training
132.	1.337	3.301	3.160	0.141	Training

5.3.1. Ligand based pharmacophore modeling and its validation

PHASE QSAR models may be either atom-based or pharmacophore-based, the difference being whether all atoms were taken into account, or merely the pharmacophore sites that can be matched to the hypothesis. To build the QSAR model for HCV NS5B inhibitors, the compounds with pIC_{50} values more than 8.5 were labelled as actives and less than 4.0 were labelled as inactives. We divided our data set into actives, inactives and moderately actives. In our pharmaset we had three actives which had pIC_{50} more than 8.5, seven inactives and remaining were moderately actives. We used this pharmaset to bulid our 3D QSAR model. We kept 5 maximum number of sites to be generated for common pharmacophore and 4 for minimum number of sites and the common pharmacophore generated was constrained to match with all the three actives. There were 34 different combinations of common pharmacophores generated. These were AAADP with 347 maximum hypotheses, AADDR 415, DDHPR 6, ADDHP 19, AAPRR 98, AHPRR 70, AAADH 479, AADDP 140, AADRR 320, AAAPR 260, AAAAP 125, ADHPR 288, ADDHR 39, AADHP 339, AAADR 885, DDHRR 5, AADPR 527, AAAAH 332, ADDRR 145, AAAHR 705, DHPRR 26, AADHR 838, ADDPR 168, DDPRR 26, AAAHP 359, AAAAD 377, ADHRR 168, AAHPR 462, AAAAR 384, AAARR 189, AAHRR 223, AADDH 44, ADPRR 109 and AAADD 145. There were 498 clusters of seven hundred and twenty one hypotheses were produced after all scoring functions were calculated (survival active, survival inactive, post-hoc). We then selected top three pharmacophore hypotheses based on good survival activity, vector, volume, energy scores, best active alignment and number of matches (**Table 13 & Figure 42**).

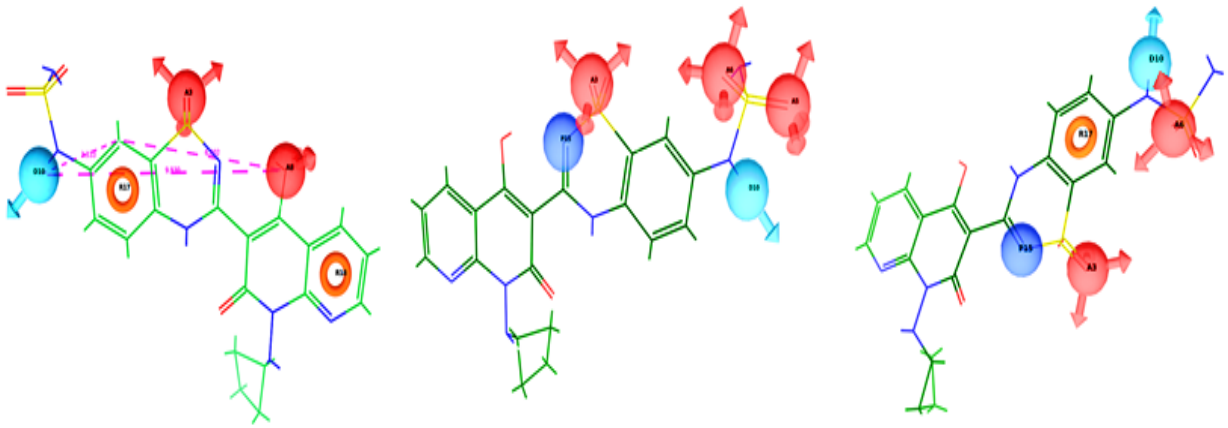
In ligand based approach, clustering the scored hypotheses, we got few numbers of clusters out of these we selected top three hypotheses based on their survival actives, inactives, vector, volume scores and their energy values.

Table 13: Ligand based pharmacophore hypotheses with their scores

S.NO	Hypothesis*	Survival Score	Survival-inactive Score	Vector Score	Volume Score	Site Score	Energy
1.	AADRR	4.029	2.936	0.998	0.837	0.98	0.000
2.	AAADP	3.959	3.108	0.999	0.810	0.94	0.087
3.	AADPR	3.986	3.072	1.000	0.801	0.98	0.933

*A-Acceptor, D-Donar, R-Ring aromatic, N-Negative ionic and H-Hydrophobic.

The hypothesis 1 (AADRR) showed 2 hydrogen bond acceptors, 1 hydrogen bond donor and two aromatic rings, hypothesis 2 (AAADP) showed 3 hydrogen bond acceptors, 1 hydrogen bond donar and 1 positively ionizable, and the hypothesis 3 (AADPR) showed 2 hydrogen bond acceptors, 1 hydrogen bond donor, 1 positively ionizable and 1 aromatic ring as features as shown in **Figure 42**. Among these hypotheses, model 1 (AADRR) showed the highest survival score (**Table 13**). The distance between the pharmacophoric features are given in the **Table 14.**



AADRR **AAADP** **AADPR**

Figure 42: Top three pharmacophores selected from ligand based approach for HCV NS5B polymerase

Table 14: Distance between the features of 3D QSAR based pharmacophores

*Hypothesis	A-R1 (Å)	A-D (Å)	A-A1 (Å)	A-P (Å)	A-R2 (Å)	D-R1 (Å)	A-A2 (Å)	A2-P (Å)	A1-P (Å)
AADRR	7.325	5.545	4.624	NA	3.252	11.646 ^{max}	NA	NA	NA
AAADP	NA	5.497	4.922	2.568	NA	NA	7.201	4.231	8.703
AADPR	3.225	6.404	5.759	2.566 ^{min}	NA	3.228	NA	NA	7.229

Distance shown in Å, *A-Acceptor, D-Donar, R-Ring aromatic, N-Negatively ionizable, H-Hydrophobic and P-Positively ionizable. NA- not applicable

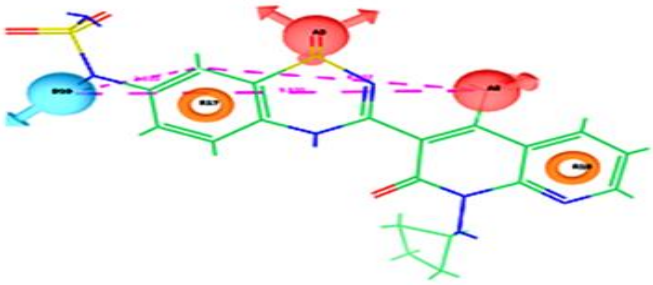
To further validate these pharmacophore models for virtual screening, we evaluated the enrichment factor similar to that performed for e-pharmacophores. The decoy set consisted of 1082 compounds in which 82 compounds were known inhibitors which were not utilized in the pharmacophore hypotheses building. Enrichment factor (EF1%), BEDROC (α=160.9 and α=20.4), RIE, ROC and also using the formulae 1 and 2 we calculated the overall enrichment factor (EF), goodness of fit (GH) values we selected the best pharmacophore hypothesis. Using the Find matches in Phase module of Schrödinger suite, with the total number of molecules in the database (D), 68 compounds were obtained as hits (H_t) for the hypothesis 1, in which 8.5% were active yields (%Y), 82.93% ratio of actives were retrieved in the hit lists (%A), and the values of EF 1% 11, RIE 2.63, ROC 2.33, high BEDROC [183] values, EF (1.122) and GH (0.048) indicated a good sign of the high efficiency of hypothesis 1. Though the hypothesis 3 AADPR had high RIE value of 2.7 compared to hypothesis 1 but other parameters were lesser than hypothesis 1(**Table 15)**.

Table 15: Validation of ligand based pharmacophore hypotheses

Pharmacophore features	EF 1%	RIE	ROC	BEDROC (α-160.9)	BEDROC (α-20)	EF	GH
AADRR	11	2.63	2.33	0.737	0.256	1.122	0.048
AAADP	8.4	1.81	1.55	0.647	0.176	0.766	0.037
AADPR	9.6	2.70	1.97	0.678	0.263	0.882	0.023

EF- Enrichment Factor at 1% of the decoy dataset
RIE- Robust Initial Enhancement
ROC-Receiver Operating Characteristic curve value
BEDROC- Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic
EF- Overall Enrichment Factor
GH- Goodness of Fit

The results as shown in **Table 15** rated the hypothesis 1 (AADRR) as the best one as indicated by the highest EF1%, BEDROC and GH values is given in **Figure 43.**



3D QSAR- AADRR

Figure 43: Finally selected ligand based pharmacophore for HCV NS5B polymerase. Pink sphere with arrow-hydrogen bond acceptor (A), yellow open circle-aromatic ring (R), blue sphere with arrow-hydrogen bond donor (D)

5.3.2. 3D QSAR and PLS Analysis

3D QSAR models were then developed for the pharmacophore hypothesis using the training set structures that match the pharmacophore on three sites. However, we utilized the three models for the 3D QSAR studies by generating pharmaophore based 3D QSAR models and PLS analysis.

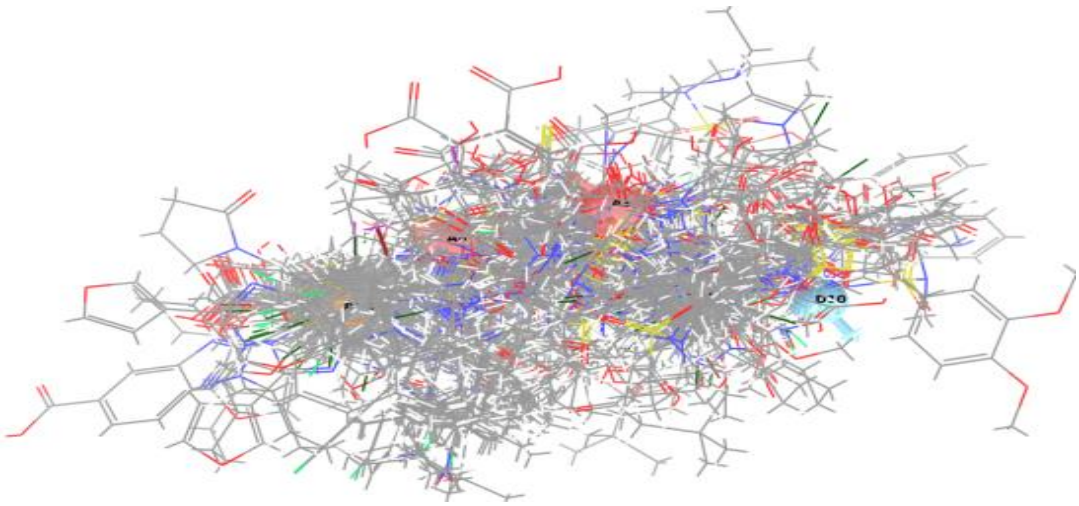


Figure 44: All 132 compounds were aligned with the best pharmacophore selected (AADRR) for 3D QSAR study for HCV NS5B polymerase

Hypotheses 1 (AADRR) showed the highest survival score, all selected 132 molecules for 3D QSAR study were aligned with that selected pharmacophore as shown in **Figure 44**. To develop superlative 3D QSAR models which were meant to exhibit reliable predictions it necessitated internal and external statistical validation. Models which were capable of fulfilling statistical validation parameter boundaries were expected to display more reliable predictions. Randomly we chose compounds in the training set and in the test to develop 3D QSAR model. Important parameters obtained based on LOO method, favored the internal statistical validation by PLS analysis. Among the models, the best hypothesis should show good external predictive ability for each combination as compared to others. Randomly we chose 70 compounds in the training set and 62 compounds in the test to develop 3D QSAR model (**Table 12**), by using the method "Automated Random Selection" option present in the PHASE module. Important parameters obtained based on LOO method, (**Table 16**) favored the internal statistical validation by PLS analysis. A model with PLS factor five was considered as the best statistical model [154]. This model was validated by predicting activities of test set, training set and overall molecules. The

best pharmacophore selected was used to screen the database compounds and the hits were further used for the docking studies. Among the three models, hypothesis 1 (AADRR) showed good external predictive ability for each combination as compared to others. Hypothesis 1 showed a good R^2 value for the training set of 0.9666, good predictive power with Q^2 of 0.5810 for the test sets, with SD of 0.2936, and F value of 381.8. Further the integrity of the model was predicted by r^2_{pred} for test set which yielded a value of 0.5955 [156] (**Table 16)**

Table 16. PHASE 3D QSAR and PLS statistics for the internal validation of the dataset

Statistical parameters	**AADRR**	**AAADP**	**AADPR**
Number of molecules in Training set	70	70	70
Number of molecules in Test set	62	62	62
R^2	0.9666	0.9766	0.9841
Q^2	0.5810	0.3381	0.3238
SD	0.2936	0.2459	0.2045
F-value	381.8	451.3	781.3
Pearson-R	0.7673	0.5847	0.6161
RMSE	0.8086	1.0513	1.0216
r^2_{pred}	0.5920	0.5097	0.1156

SD - Standard deviation of the regression
R^2 - for the regression.
F - variance ratio.
r^2_{pred} - predictive correlation coefficient value.
RMSE - root mean square error.
Q - squared (Q^2)value of Q^2 for the predicted activities.
Pearson R - correlation between the predicted and observed activity for the test set.

In the present study the best predictive ability of the model was characterized by correlation coefficient R = 0.7679 (r2 = 0.5895) [158]. High slope of regression lines through the origin k value of 1.018 and k′ value of 0.9651 gave the substantial values for R_0^2 value 0.9883 and the R'^2_0 value 0.9722, which were obtained by calculating correlation coefficient of regression lines of the scatter plot obtained by means of actual activity versus predicted activity (**Figure 45**). The

relation between r^2, R_0^2 and R'^2_0 gave (r2 $-R_0^2$ /r2) values of -0.6765 and second relation (r2 $-R'^2_0$ /r2) value -0.6492 showed optimum values within the statistical limits (**Table 17**).

Table 17: External statistical validation results of quantitative structure activity relationship (QSAR) result for the hypothesis 1 AADRR

External validation	**Parameter calculated**	**Limitations**
r_{cv}^2	0.5920	$r_{cv}^2 > 0.5$
R	0.7679	Must close to 1
r^2	0.5895	$r^2 > 0.5$
k value	1.018	$0.85 \leq k \leq 1.15$
k′value	0.9651	$0.85 \leq k' \leq 1.15$
R_0^2	0.9883	R_0^2 or R'^2_0close to r^2
R'^2_0	0.9722	R_0^2 or R'^2_0close to r^2
$[(r^2 - R_0^2) / r^2]$	-0.6765	$[(r^2 - R_0^2 / r^2] < 0.1$
$[(r^2 - R'^2_0) / r^2]$	-0.6492	$[(r^2 - R'^2_0 / r^2] < 0.1$
r_m^2 (LOO)	0.8242	r_m^2(LOO)> 0.5

r_{cv}^2 - cross validated coefficient
R (or r^2) - correlation coefficient between the actual and predicted activities
k and k′ - slope values of regression lines
R_0^2 and R'^2_0 - correlation coefficients for the regression lines through the origin
$[(r^2 - R_0^2) / r^2]$ and
$[(r^2 - R'^2_0) / r^2]$ - to calculate relation between r^2, R_0^2 and R'^2_0
r_m^2 (LOO) - modified squared correlation coefficient for Leave One Out method.

Yet, our established QSAR model from hypothesis 1 (finalized after PLS analysis), gave r_{cv}^2 value of 0.592. A parameter of modified r^2 [r_m^2 (LOO)] [65] was considered as a better external predictive potential for the whole set of compounds which was of 0.8242 (>0.5) defined through scatter plot best fit line values (**Figure 45**).

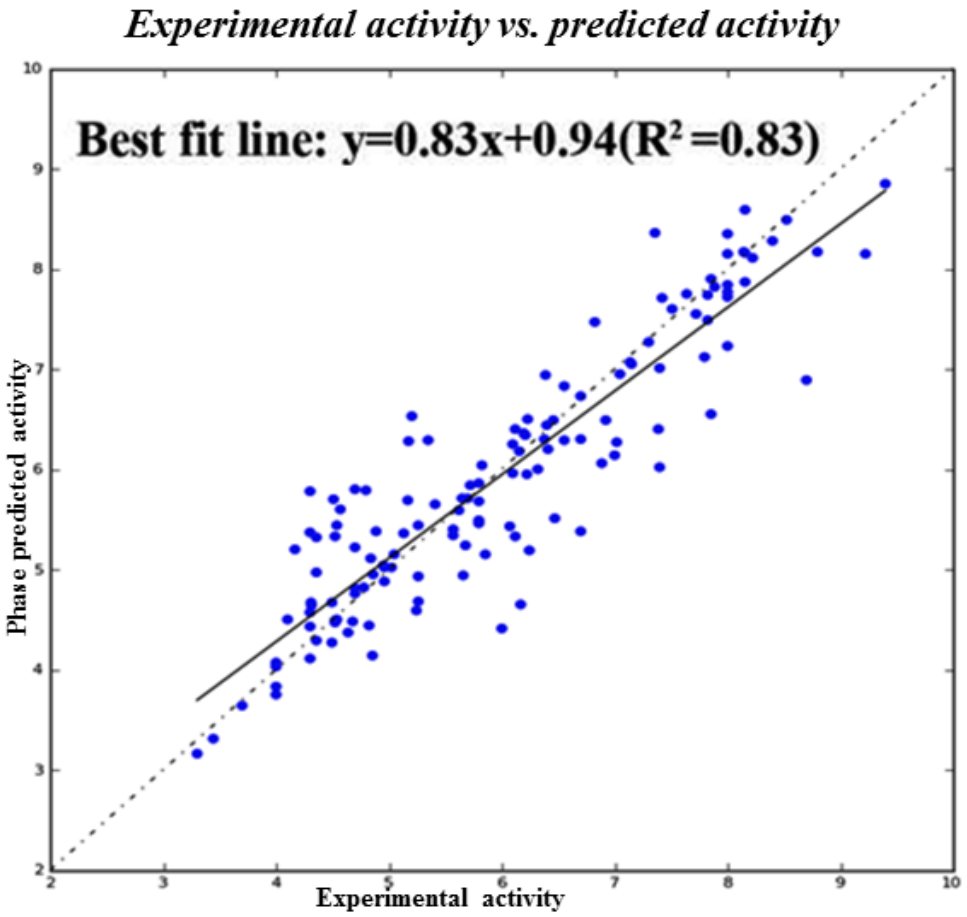


Figure 45: Scatter plot of the experimental vs. predicted activity of NS5B inhibitors generated by the best model for HCV NS5B polymerase

This appeared to be truly predictive by fulfilling the requirements of every parameter in the external validation **(Table 17**). Truly, we considered this model as statistically significant model. Besides, we resumed further steps to predict the activities of new leads from the compound libraries by using hypothesis 1.

5.3.3. Contour maps

The final validated hypothesis 1 obtained from 3D QSAR was used to generate contour maps. These contour maps were important to identify the positions of the substitutions or replacements of atoms to enhance bioactivity. Inhibitory activity can be gained by visualizing and understanding the maps against most active (**1**) and least active (**131**) compounds. This could help

in discovering novel scaffolds with good biological activity. The most and least active ligand contour maps were generated and are shown in **Figure 46**. Contour maps indicated H-bond donor effect on the most active ligand (**1**) and least active ligand (**131**) (**Figures 46 A and 46B**), the hydrophobic effect of the ligands **(Figures 46C and 46D**) and the electron withdrawing nature (**Figures 46E and 46F**) of both ligands represented in the figure were discussed further.

The hydrogen bond donor nature for the most active compound **1** and the least active compound **131** when compared showed their most favorable region blue color and unfavorable regions red color. Hydrogen bond donor mapping revealed that favorable regions lied near the nitrogens of naphthyridine and in benzothiadiazine indicating their importance for activity compared to the least active compound **131.** Therefore the presence of cyclic rings with hydrogen donor group in the scaffold backbone was very much needed for the activity.

Figure 46C and 46D when compared for their hydrophobic nature for the most active compound **1** and least active compound **131** revealed that favored green color region around the naphthyridin, cyclobutyl and benzothiadiazine rings that revealed that the terminal hydrophobic rings were very much needed for the activity of the compound and unfavorable region (yellow color) on aromatic ring hydrogen moiety revealed that increase in the carbon chain could increase the activity.

In **Figure 46E,** the favored red color regions were observed near hydrogen bond acceptors along with respective acceptor hypothesis features of most active compound which indicated that these features were crucial for the activity and these groups should be unsubstituted when further lead modifications indicated. In the least active compound as in **Figure 46F** the unfavorable region (blue color) surrounded the napthyridin ring moiety which indicated that a decrease in the ring size could increase the biological activity of the compound.

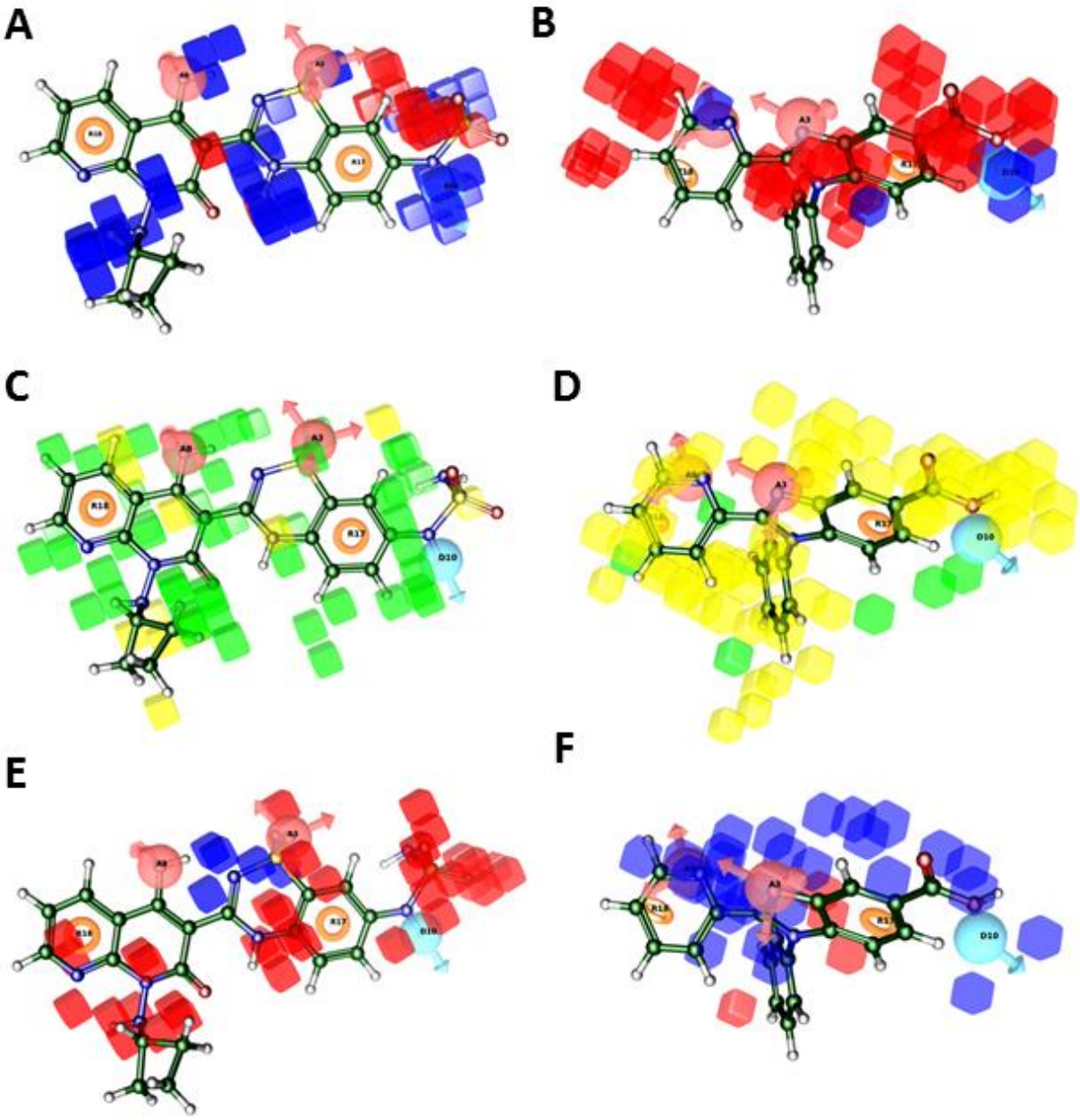


Figure 46: Contour map for HCV NS5B polymerase:
A- H-bond donor effect: Most active; **B**- Least active (Blue- favorable, Red unfavorable); **C**- Hydrophobic effect: Most active; **D**- Least active (Green- favorable, Yellow-unfavorable); **E**- Electon with-drawing effect: Most active; **F**- Least active (Red-favorable, Blue -unfavorable)

5.4. Multiple pharmacophore models based virtual screening and docking

Virtual screening studies of the commercial database were fruitful resource for initial lead identification. Fit value indicated a measure of how well the ligand fit the pharmacophore. Therefore, the hits with a high fit value were probably very active. In the present study, we employed the best e-pharmacophores and in second method, pharmacophore based on the 3D QSAR were utilized to find important features for the inhibition of NS5B, to help in designing of lead molecules. The best validated five e-pharmacophore model and the pharmacophore from 3D QSAR (hypothesis 1) were used to screen against the databases of 500000 compounds (Asinex) as presented in the flow chart (**Figure 47)**. The numbers of hits derived from pharmacophore filtering are shown in the **Table 18** and in **Figure 48A.**

For the e-pharmacophore obtained from ADR of 3CVK, fitness more than 1.5 was taken as cut-off for the HTVS (High throughput virtual screening), which yielded 13226 compounds and we clustered these compounds. We obtained 1812 clusters and the 13226 ligands were docked to palm region of 3CVK, and the docking score of the crystal ligand was -7.289 kcal/mole. During shortlisting the hits we kept the docking score of the crysta ligand as one of the criterion. 7257 ligand molecules as hits from HTVS with docking score above -6 kcal/mole, fitness above 1.6 and the number of H-bonds above 2 which belonged to 1026 clusters were selected. These hits were further docked using Glide SP (standard precision) docking module and 1549 ligand molecules were selected based on the docking score above -7 kcal/mole, fitness above 1.7, above 2 H-bond and visual inspection of occupancy of ligand into the pocket. Finally, we subjected the Glide SP 1549 filtered ligands to Glide XP (extra precision) docking simulation.

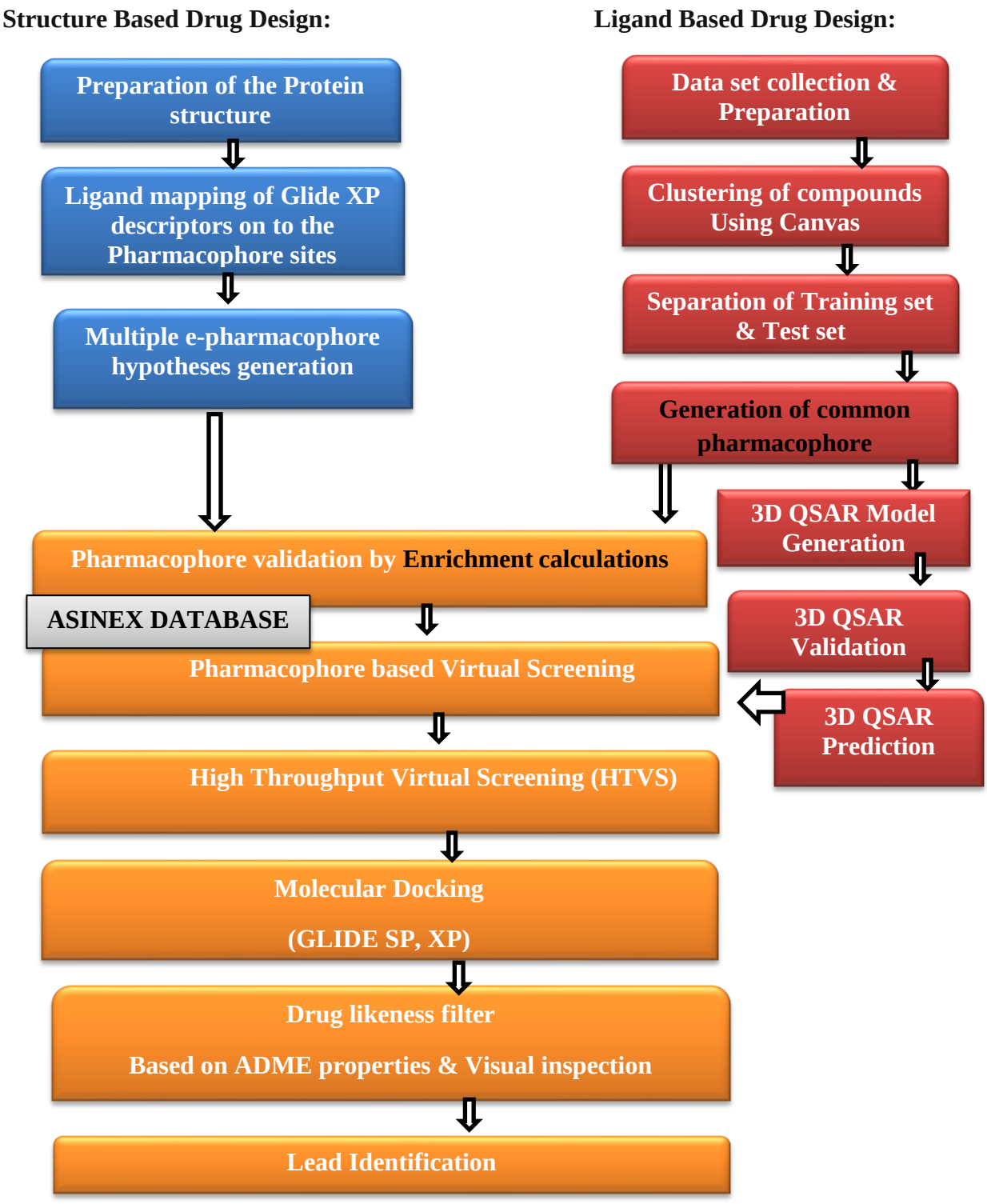


Figure 47: Work flow for combined virtual screening workflow for both structure based and ligand based Approaches

Top 58 ligand molecules more than a docking score of -7.00 kcal/mole were visually inspected for the pose and important binding residues. Top 58 ligands were found to belong to 18 diverged structural scaffolds and those matching all the three e-pharmacophoric features were selected.

For the e-pharmacophore ADR for 3CO9, fitness above 1.5 (7576 ligands of 541 clusters) were selected for HTVS docking. The docking score of the crystal ligand was -5.171 kcal/mole. For SP docking we shortlisted 2092 (356 clusters), with docking score above -5 kcal/mole, fitness above 1.6 and more than 2 H-bonds. For XP docking we shortlisted 1513 (306 clusters), with docking score above -6 kcal/mole, fitness above 1.7 and 2 H-bonds. Top 391 ligands were found clustered into 98 diverged structural scaffolds and those ligands with docking score above -7 kcal/mole, fitness above 1.8 and 2 H-bond and fit with all the three e-pharmacophoric features were selected.

For the e-pharmacophore NRR of 2D3Z, fitness above 1.35 (28687 ligands of 1923 clusters) were selected for HTVS docking. The docking score of the crystal ligand was -4.948 kcal/mole. For SP docking we shortlisted 2035 (364 clusters), with docking score above -4 kcal/mole, fitness above 1.4 and 2 H-bond. For XP docking we shortlisted 961 (183 clusters), with docking score above -6 kcal/mole, fitness above 1.5 and 2 H-bonds. Top 482 ligands resulted into 98 diverse structural scaffolds and those with docking score above -6 kcal/mole, fitness above 1.8 and 2 H-bond and matching with all the three e-pharmacophoric features were selected

For the e-pharmacophore ARR of 2GIQ, fitness above 1.5 (13367 ligands of 1646 clusters) were selected for HTVS docking. The docking score of the crystal ligand was -5.771 kcal/mole. For SP docking we shortlisted 10000 (1345 clusters), with docking score above -4 kcal/mole, fitness above 1.5 and 2 H-bond. For XP docking we shortlisted 2110 (430 clusters), with docking score above -6 kcal/mole, fitness above 1.6 and 2 H-bonds. Top 78 ligands were found to belong to 8

diverged structural scaffolds those with docking score above -7 kcal/mole, fitness above 1.7 and 2 H-bonds and matching all the three e-pharmacophoric features were selected.

For the e-pharmacophore NHR from 2GIR, fitness above 1.4 (11256 ligands of 978 clusters) were selected for HTVS docking. The docking score of the crystal ligand was -6.419 kcal/mole. For SP docking we shortlisted 6334 (478 clusters), with docking score above -5 kcal/mole, fitness above 1.5 and 2 H-bonds. For XP docking we shortlisted 3843 (334 clusters), with docking score above -6 kcal/mole, fitness above 1.6 and 2 H-bonds. Top 205 ligands were found to belong to 28 diverse structural scaffolds and those with docking score above -7 kcal/mole, fitness above 1.8 and 2 H-bond and matching all the three e-pharmacophoric features were selected.

Similarly, the best pharmacophore based on the ligand approach (ADDRR) was also employed in database screening using phase findmatches. There were 50000 ligands above fitness score 1.4 which were selected for docking studies. We chose 2GIQ crystal structure and default grid because of its very low resolution and docked our top 50000 compounds in the palm active site pocket. Fitness above 1.4 (50000 ligands of 3458 clusters) were selected for HTVS docking. For SP docking we shortlisted 4231 compounds (352 clusters), with docking score above -5 kcal/mole, fitness above 1.5 and 2 H-bonds. For XP docking we shortlisted 1552 (269 clusters), with docking score above -6 kcal/mole, fitness above 1.7 and 2 H-bonds.

Top 108 ligands belonging to 19 diverse structural scaffolds were further selected based on docking score above -6.5 kcal/mole, fitness above 2.0 and 2 H-bond and while matching at least all the three e-pharmacophoric features. At each stage of screening we found common hits between two pharmacophore models and the overall overlap of number of retrieved compounds were analyzed with pie graph which is presented as **Figure 48.**

GOLD docking programe was employed to cross check our finally selected XP hits. There were two simultaneous runs performed and compounds with high GOLD score were selected.

Figures 48A-48D depicts similar analyses at every stage of the design cycle from pharmacophore filtering, HTVS, Glide SP and XP steps and there were little overlap in later stages indicating that possibility of more diverse ligand generation. Analyses of the hits based on peptidic vs non-petidic leads revealed that most of the leads generated were peptidic ligands (**Table 18)**. Shortlisting was performed by comparing among non-peptidic leads ranked in more than one pharmacophore models.

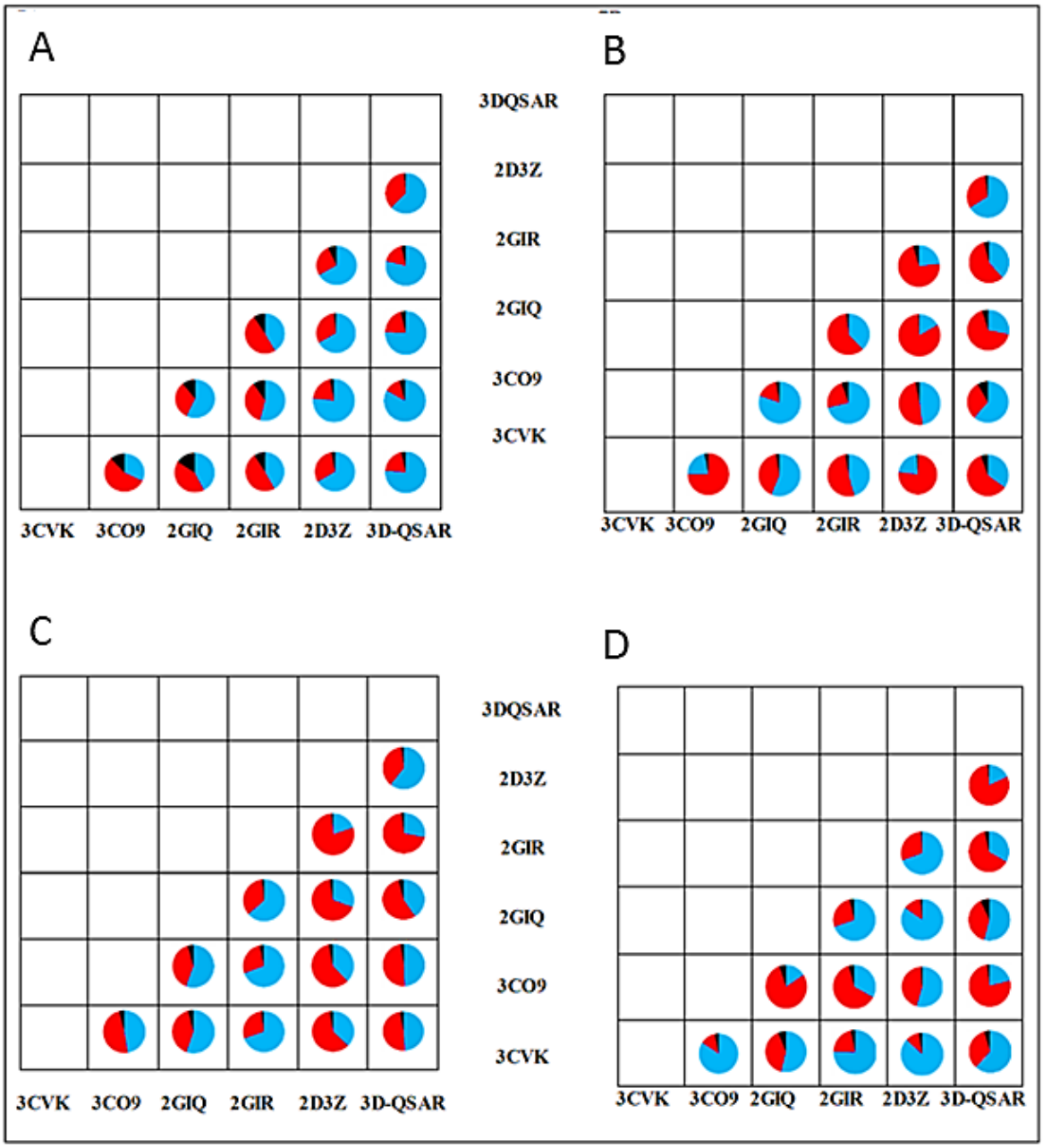


Figure 48: Pie graph showing the extent of overlapping of compounds at each stage of screening for all six pharmacophore models. A- Phase find match hits, B- HTVS docking hits, C- SP docking hits, D-XP docking hits. Red and blue colored regions represent the number of screened hit compounds from any two pharmacophore models and the black color region represents the overall overlap of the screened compounds for HCV NS5B polymerase

Structural diversity is an important index for the quality of the hits by an *in-silico* approach.

Accordingly, we compared the structural diversity of the hits retrieved from the Asinex database.

The compounds retrieved from each of the 5 crystal structure and 3D QSAR model after docking results were clustered using Canvas [151] clustering algorithm from Schrodinger. The total number of clusters from each structure as represented in **Table 18** was analyzed for structural diversity among the compounds generated.

Table 18: Number of compounds (Hits) retrieved at each stage of screening

PDB CODE	**Phase find matches hits**[a]	**Clusters**	**HTVS Hits**[b]	**Clusters**	**SP Hits**[c]	**Clusters**	**XP Hits**[d]	**Clusters**	**Non-peptidic**	**Peptidic**
3CVK	13226	1812	7257	1026	1549	168	58	18	11	47
3CO9	7576	541	2092	356	1513	306	391	98	59	332
2D3Z	28687	1923	2035	364	961	183	482	126	32	450
2GIQ	13367	1646	9987	1345	2110	430	78	8	24	54
2GIR	11256	978	6334	478	3843	334	205	28	61	144
3D QSAR (2GIQ)	50000	3458	4231	352	1552	269	108	19	22	86

[a]Total number of hits from Phase findmatches.
[b]Total number of hits from HTVS docking.
[c]Total number of hits from SP docking.
[d]Total number of hits from XP docking

The hits retrieved from 3CO9 were found to be very restrictive and retrieved very limited hits from Asinex with only 7576 hits demonstrating that different pharmacophore models may have quite different performance in screening a chemical database. Therefore, multiple pharmacophore models can be used to improve the overall screening efficacy. Similar evaluations were performed for each pharmacophoric hypotheses as explained in pie graph. All the final hits were analysed for their ranking and the final hit list were prepared based on their ranking in more than one pharmacophore models. Predicted activities, Glide scores, fitness, H-bond data and Gold Scores are presented in the **Table 19.**

When compared among the similar pharmacophore models of 3CVK, 3CO9 and 2GIQ (**Figure 49**) based on the similar features of A and R, there were more common hits as compared to other pharmacophores which were diverse. The distance between common features in the e-pharmacophores were further mapped along with ligand based models and were found to show rmsd values within 2 Å as shown in **Figure 49.**

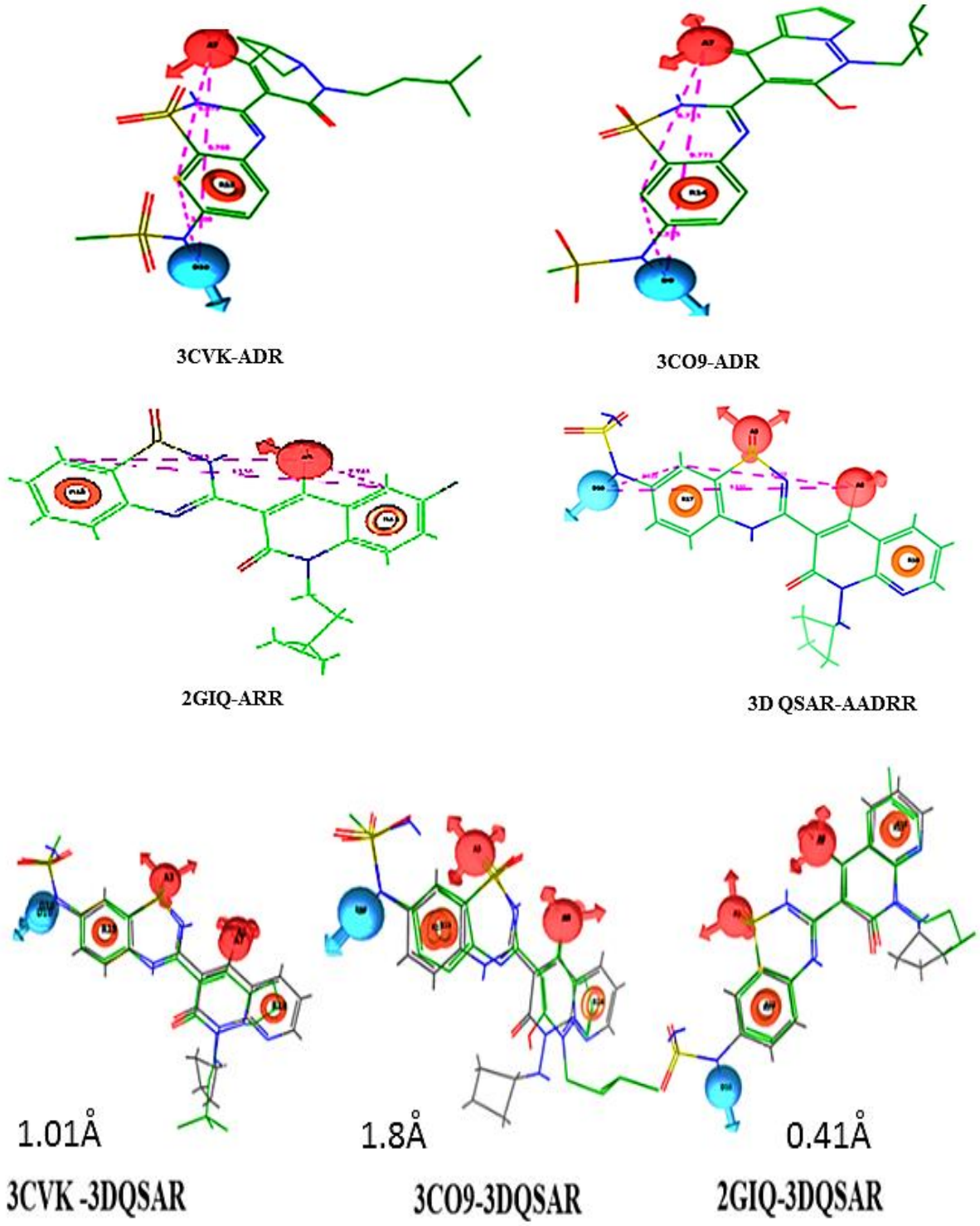


Figure 49: Comparison of the distance between the common features in the pharmacophores obtained from structure based (ADR, ARR) and ligand (3D QSAR) based (AADRR). The e-pharmacophores were mapped with the best ligand based pharmacophore and given very less rmsd values

The top 1% hits retrieved by each pharmacophore hypothesis were aligned with their respective pharmacophore shown in **Figure 50.**

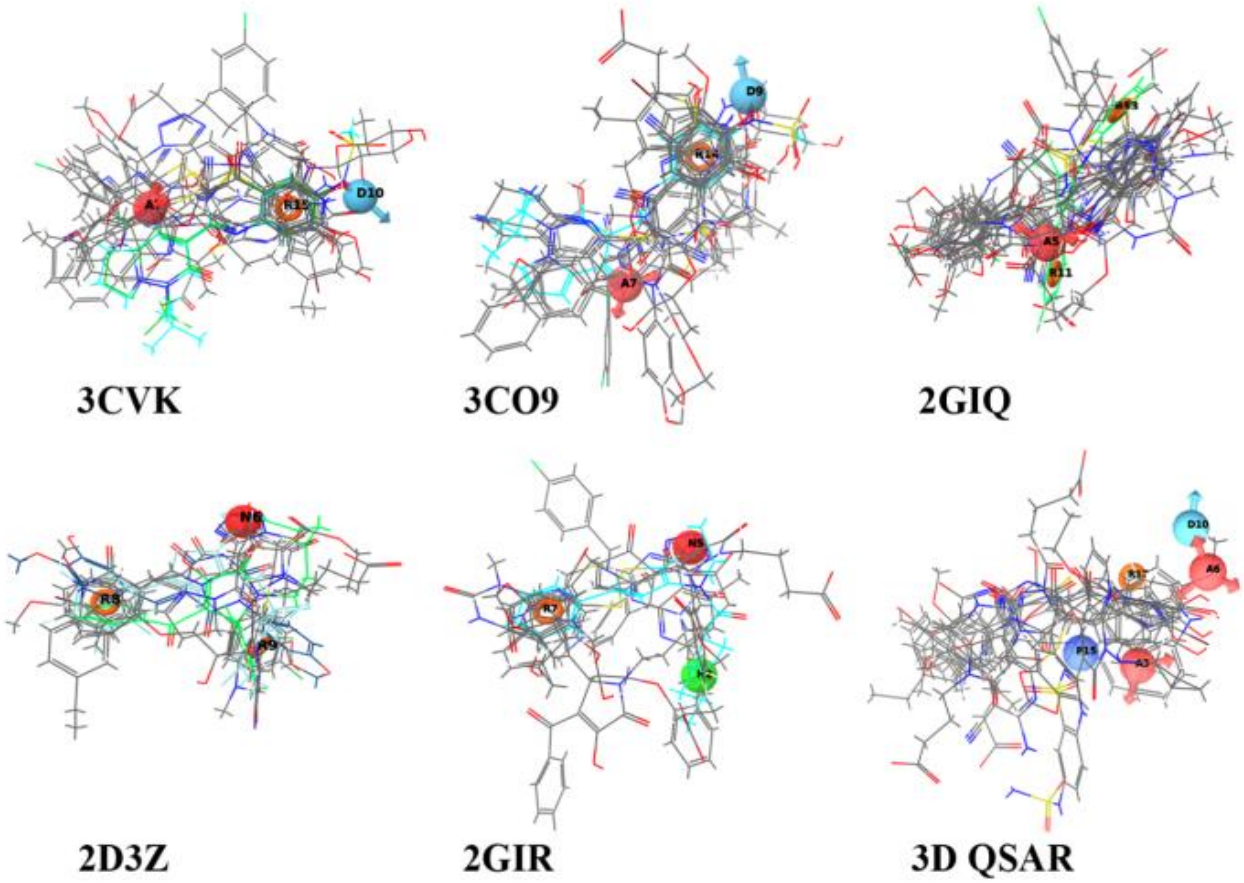


Figure 50: Top 1% of the selected compounds aligned with their pharmacophore with their respective PDB codes for HCV NS5B polymerase

The structures of all top 10 hits with their docking score, fitness, number of H-bonds, important amino acid interaction and gold score are presented in **Table 19 and Figure 51.**

Table 19: Lead compounds with their fitness, docking score, number of H-bond, interaction with important amino acids and their two respective gold scores

S.No.	Fitness	Docking Score	H-bond	Predicted Activity	Ligand Interactions	Gold Score1	Gold Score2
H-1	1.105	-9.959	5	5.077	Tyr477, Ser476 (2), Arg501, Lys533	68	63
H-2	1.707	-8.665	4	6.181	Asp318 (2),Tyr415, Ser556	75	70
H-3	1.569	-8.221	2	5.199	Ser407, Tyr448	49	45
H-4	1.579	-8.134	5	5.917	Asn291, Asp318, Gln446, Tyr448, Ser556	63	62
H-5	2.058	-6.561	2	5.075	Asp318 (2)	61	62
H-6	2.018	-7.471	4	6.215	Asp318, Met414, Tyr448, Ser556	72	71
H-7	1.476	-7.818	4	6.051	Ser476, Trp528, Lys533, Arg422	58	59
H-8	1.804	-7.746	3	5.008	Asn291, Tyr448, Ser556	67	61
H-9	1.413	-7.120	2	5.689	Tyr448, Gln446	42	43
H-10	1.863	-6.518	2	5.917	Asp318, Ser288	67	74

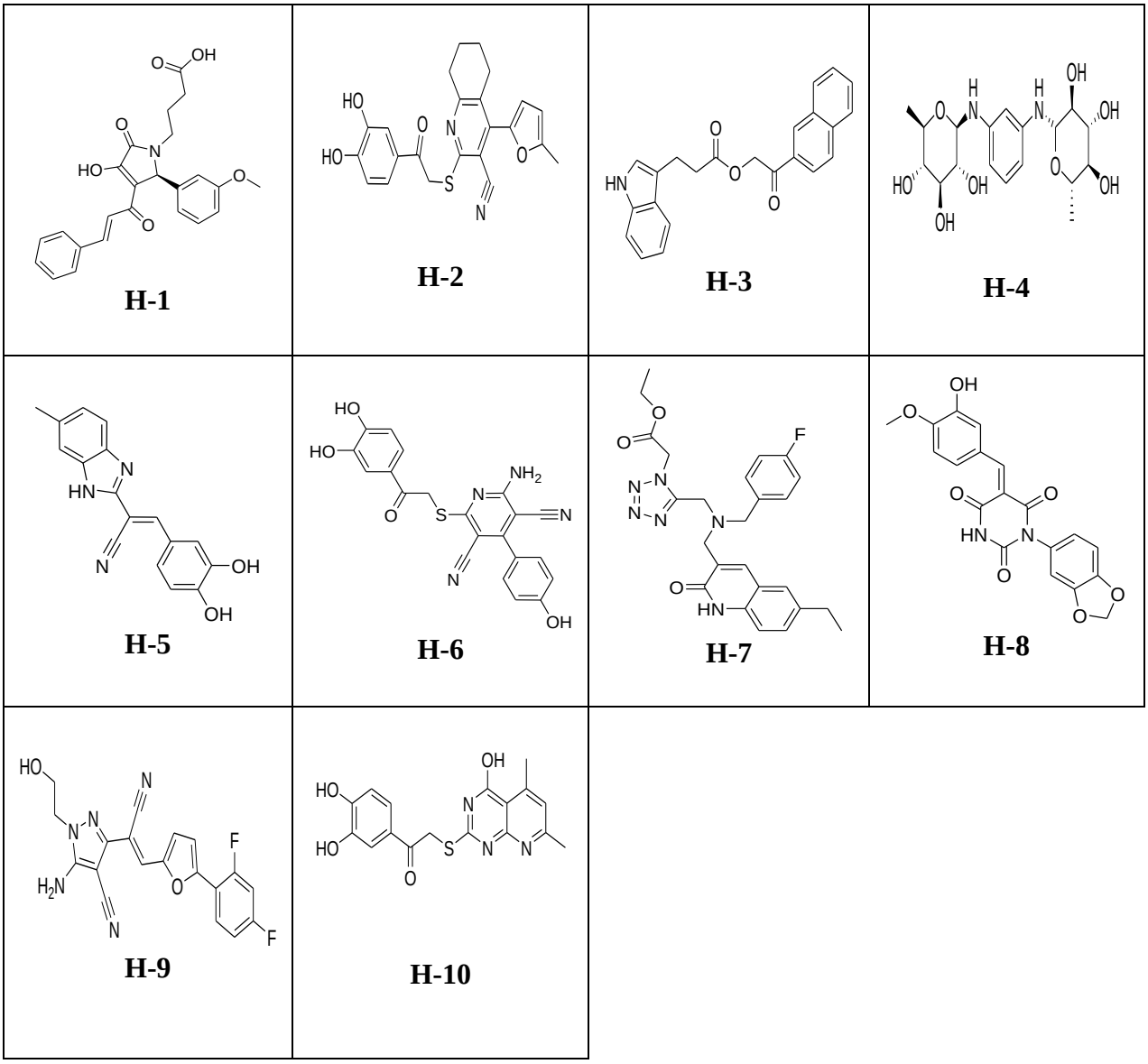


Figure 51: Top 1% of the selected compounds for HCV NS5B polymerase

5.5. ADME predictions

We finally evaluated the 10 lead compounds for the pharmaceutically relevant properties to check drug-likeness and predictions for drug's pharmacokinetics in the human body including its ADME. QikProp module of Schrödinger suite [164] was used for evaluation of drug-like behavior through analysis of pharmacokinetic parameters required for absorption, distribution,

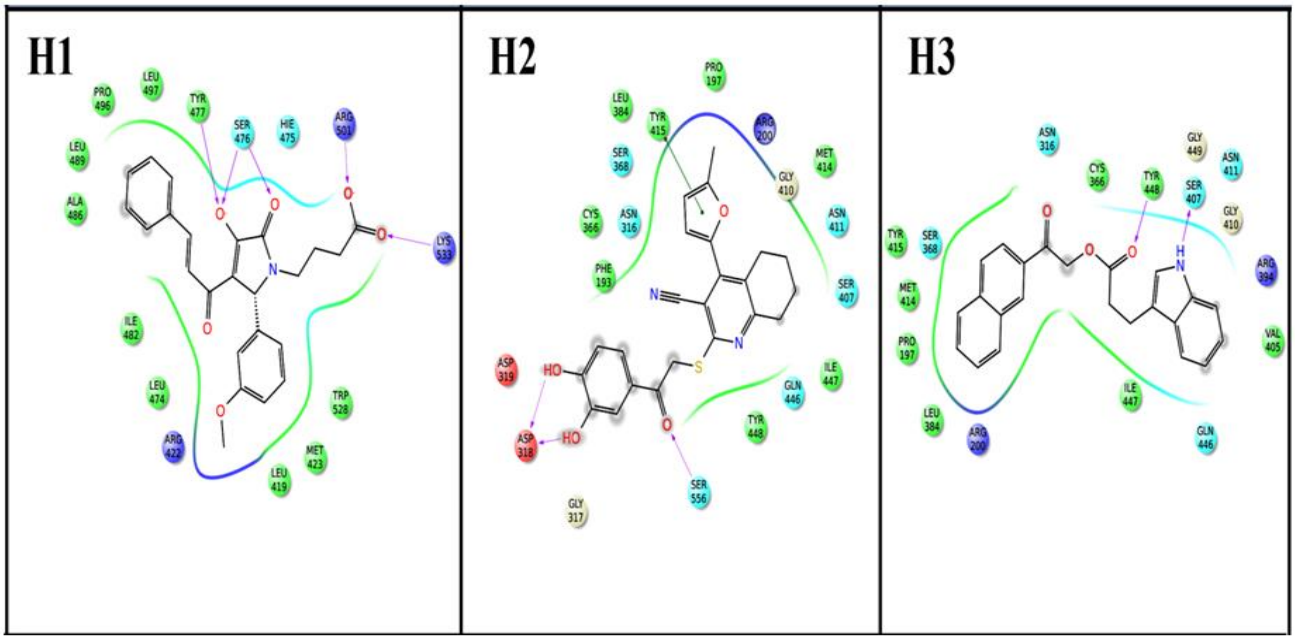
metabolism and excretion (ADME). All the ten lead compounds showed good partition coefficient (QPlogPo/w) values which were critical for understanding of absorption and distribution of drugs, to range from -1.535 to 4.448. Factor QPPCaco indicating permeability of the 10 lead compounds ranged from 2.966 to 777.368, where QPPCaco was a predicted apparent Caco-2 cell permeability in nm/sec value a key factor for estimation of cell permeability in biological membranes and its metabolism. All the lead compounds passed the entire pharmacokinetic requirement for a drug-like compound and were within the acceptable range defined for human use. Additional parameters such as molecular weight, H-bond donors, H-bond acceptors, and human oral absorption according to Lipinski's rule of 5 etc. were also evaluated for their drug-like behavior and are represented in **Table 20**.

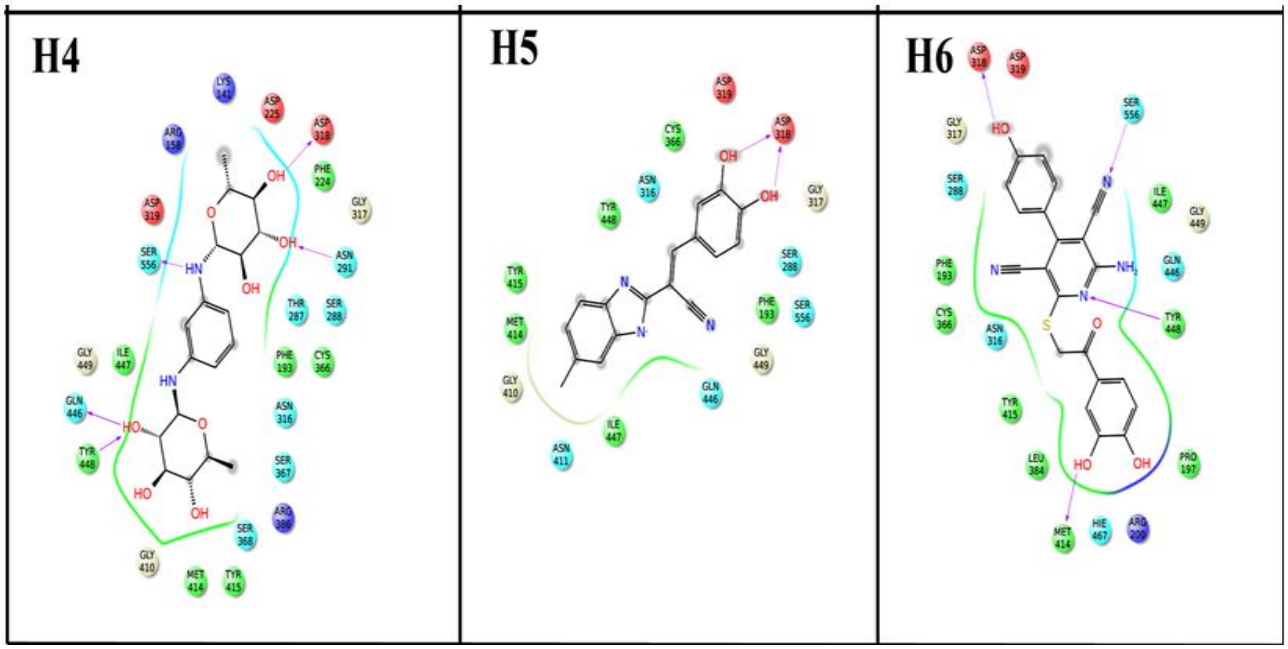
Table 20: Lead compounds with their Qikprop properties

S.NO	MW[a]	QPlogPo/w[b]	QPPCaco[c]	QPlogHERG[d]	Percent Human oral absorption[e]	Rule of five[f]
H-1	421.449	3.515	16.562	-4.263	69.346	0
H-2	420.482	3.427	139.168	-5.640	85.378	0
H-3	357.408	4.429	777.368	-6.348	100.000	0
H-4	400.428	-1.535	52.237	-4.873	35.750	1
H-5	291.309	2.010	130.085	-5.733	76.554	0
H-6	418.426	0.586	2.966	-6.000	38.826	0
H-7	478.525	3.025	27.360	-6.945	70.381	0
H-8	382.329	1.884	205.604	-4.951	79.376	0
H-9	381.341	2.498	112.270	-5.971	78.267	0
H-10	357.383	1.781	91.850	-4.881	72.509	0

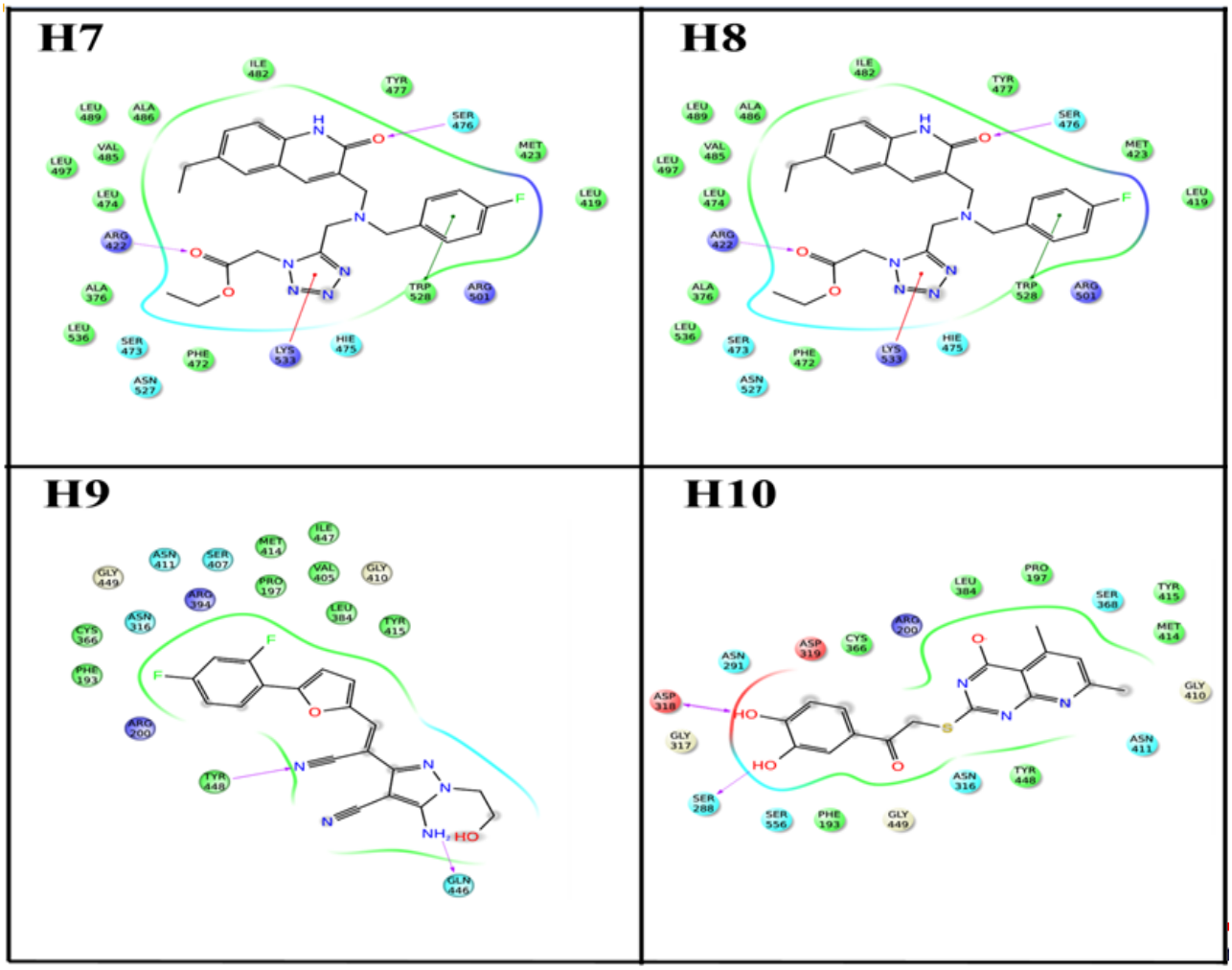
[a] Molecular weight (acceptable range < 500)
[b] Predicted octanol/water partition co-efficient log p (acceptable range from −2.0 to 6.5).
[c] Predicted Caco-2 cell permeability in nm/s (acceptable range: <25 is poor and >500 is great).
[d] Predicted value for blockage of HERG K+ channels (concern below −6.5).
[e] Percentage of human oral absorption (<25% is poor and >80% is high).
[f] Rule of five (no. of violations of Lipinski's rule of five: 0 is good and 4 is bad)

Thus, compounds with predicted interaction and good predicted pharmacokinetic properties were finalized. The shortlisted hits belonged to diverse scaffolds like, pyridine, pyrrole, indole pyrano, benzimidazole, pyrazole, quinoline, pyrimidinotrione, pyrimidopyridine, etc. The hits **H1-10** when analyzed for their binding interactions in the active site pockets of palm and thumb regions, was found to be (**H-1, H-2, H-4-6, H-8-10**) retrieved from palm region docking while hits **H-3** and **H-7** were retrieved from thumb region binding. Among these hits, **H-2, H-6** and **H-7** showed higher predicted bioactivity based on 3D QSAR model (**Table 19**). The binding pose of all these hit compounds is presented in **Figures 52 & 53**. It is clear that important amino acid residues reported for palm or thumb region has been involved in the interaction.





Figures 52: Ligand interaction diagram of hits (**H-1- H-6**) with their important amino acid residues for HCV NS5B polymerase



Figures 53: Ligand interaction diagram of hits (**H-7- H-10**) with their important amino acid residues for HCV NS5B polymerase

5.6. Anti-HCV activities and cellular cytotoxicity

Having shortlisted the lead compounds from the design protocol, we hypothesized that the compounds may possess the ability to bind with HCV NS5B and as a result could inactivate the RdRp activity of NS5B. To test this hypothesis, the compounds were preliminarily screened at 50 µM concentrations in standard *in vitro* RdRp assays and the potency of anti-NS5B RdRp activity was evaluated. Among the tested compounds, **H-5** and **H-6** were able to demonstrate more than 50% inhibition of HCV NS5B RdRp activity as nearly 67% and 50% inhibition,

respectively were noted in the preliminary screening. Furthermore, compounds **H-5** and **H-6** were evaluated in a dose-dependent manner to estimate the potency against NS5B polymerase and their IC_{50} values were found to be 28.8 µM and 47.3 µM, respectively. Among the other compounds, **H-10** showed moderately poor inhibition in RdRp assay as nearly 27% inhibition was noted in preliminary screening. On other hand, no other compounds showed significant inhibition in preliminary screening against RdRp activity (**Table 21**).

To further evaluate the anti-HCV activity and cellular cytotoxicity of hit compounds, we measured the ability of these compounds to inhibit HCV replication in cells carrying HCV subgenomic replicons in Huh7/Rep-Feo1b cells. Simultaneously, the effect of these compounds on cellular cytotoxicity was also measured. As expected, approximately 96% and 86% inhibition in HCV replication was demonstrated by compounds **H-5** and **H-6** at 50 µM, respectively. Compounds **H-5** and **H-6** also did not show any cytotoxicity as nearly 114% and 90% cells were found to be viable after the treatment of both of these compounds at 50 µM, respectively. To our surprise, compound **H-2**, **H-7**, **H-9** and **H-10** showed <80% inhibition of HCV replication at 50 µM concentration with cell viability of 64%, 68%, 77% and 113%, respectively. The inhibition obtained by these compounds could be due to either the inhibition of viral proteins other than HCV NS5B or due to inhibition of cellular factors facilitating HCV replication. Compound **H-3** showed approximately 50% inhibition of HCV replication at 50 µM concentration whereas compounds **H-4** and **H-8** did not show considerable inhibition but were found to have no cellular cytotoxicity at the tested concentration.

Table 21: Anti-HCV potency and cytotoxicity of hit compounds

No.	Feo1b replicon Inhibition[a] (%)	Viability[b] (%)	NS5B RdRp inhibition[c] (%)	IC_{50}[d] (µM)
H-1	NI	88 ± 1	5 ± 2	ND
H-2	97 ± 1	64 ± 2	9 ± 1	ND
H-3	51 ± 5	88 ± 1	1 ± 1	ND
H-4	16 ± 4	109 ± 3	NI	ND
H-5	97 ± 1	114 ± 11	67 ± 4	28.8 ± 2.0
H-6	86 ± 1	90 ± 4	50 ± 3	47.3 ± 5.0
H-7	91 ± 2	68 ± 9	8 ± 3	ND
H-8	25 ± 8	106 ± 5	NI	ND
H-9	80 ± 4	77 ± 6	13 ± 5	ND
H-10	81 ± 1	113 ± 12	27 ± 1	ND

[a]The anti-HCV activity of the compounds in Feo1b cells replicating HCV subgenomic replicons reporters at 50 µM concentration. Cells incubated in the presence of equal amounts of DMSO served as control. The data represents an average of three independent experiments. [b]Cell viability was determined at 50 µM concentrations using CellTiter 96® AQueous One Solution Cell Proliferation kit (Promega) in Feo1b cells. Cells treated with equal amounts of DMSO was considered at 100% viable and served as control. [c]Inhibition of NS5B RdRp activity at 50 µM concentrations of indicated compounds on homopolymeric template-primer (polyrA-U_{12}) and $MnCl_2$ as divalent metal ion as described in Materials and Methods. The inhibition of RdRp activity in the present of DMSO was considered as control. The results shown are an average of three independent set of experiments in triplicate. [d] Ten quarter log dilutions of compounds were employed in RdRp assay. NI, no inhibition. ND, not determined

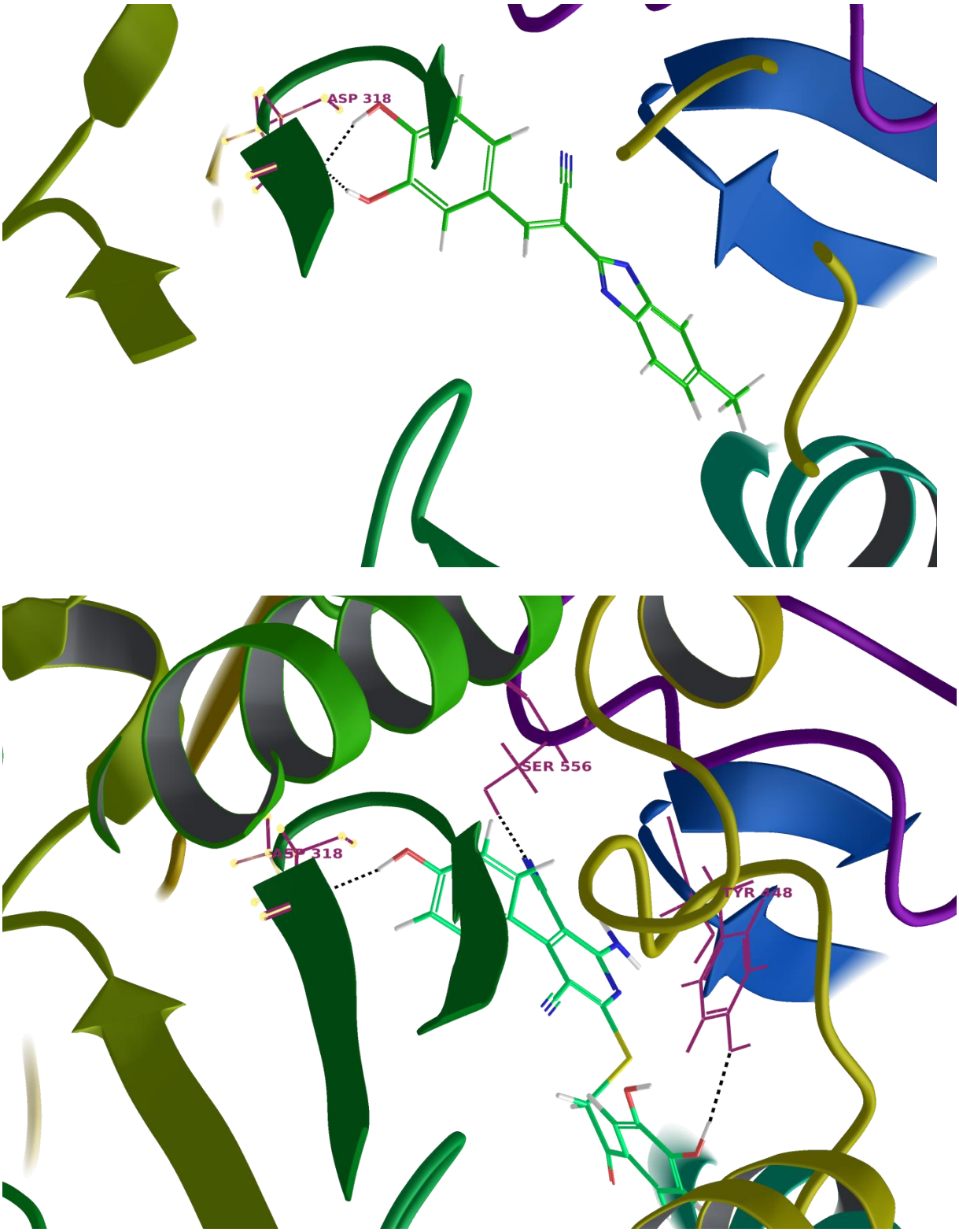
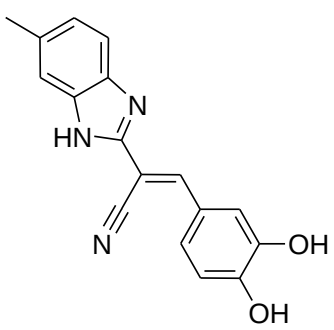


Figure 54: Interaction picture of the most potent leads **H-5** and **H-6**

The interaction picture (**Figure 54)** of the two most active compounds were shown that the same interactions of crystal ligands of palm region. The compound **H-5** showed two H-bonds with Asp318 which was an important interaction responsible for the activity shown by all bound inhibitors and showed that the binding site of **H-5** was similar to that of the crystal ligands belong to 3CVK and 3CO9. The compound **H-6** showed three H-bond interaction with Asp318, Tyr448 and Ser556. All these three were important interactions shown by all three crystal ligands. These interactions showed that the compound **H-6** was binding to the same binding site of the crystal ligands of 3CVK, 3CO9 and 2GIQ. This was the evidence that H-bond interaction pattern was an important criterion to finalise the hit compounds.

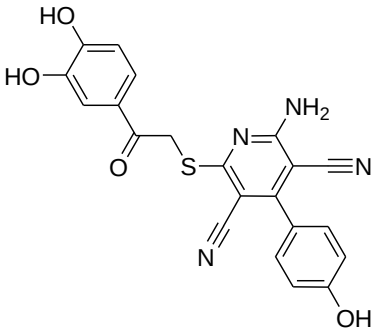
5.7. CONCLUSION

The available information on HCV NS5B crystal structures bound with inhibitors in the two sites namely palm and thumb regions along with the ligand with known inhibitory potential were explored for a comparative pharmacophore analyses followed by a high-throughput virtual screening and docking to identify diverse non-peptidic leads. It was evident from this study that the hits retrieved by combining multiple pharmacophore hypothesis and 3D QSAR predictions yielded 10 hit compounds which were diverse. The most promising hits identified in the biological screening were **H-5** and **H-6** which showed IC_{50}s of 28.8 µM and 47.3 µM respectively against NS5B polymerase and anti-HCV inhibition of **96%** and **86%** respectively. The lead compounds identified were benzimidazole (**H-5**) and pyridine (**H-6**) scaffolds which could be considered as protypical leads for further drug development.



H-5

IC_{50} = 28.8 µM



H-6

IC_{50} = 47.3 µM

CHAPTER 6

DESIGN OF HIV PROTEASE INHIBITORS: RESULTS & DISCUSSIONS

6.1. Background

The protease of the HIV-1 is an attractive drug target because of its crucial role in the replication process of HIV. The crystal structure of HIV protease was first reported by Navia and colleagues from Merck Laboratories in 1989 [97]. It is a noncovalent homodimer, with each subunit possessing 99 amino acids. The active site pocket lies between the two identical subunits. It resembles aspartic proteases by having the sequence Asp25-Thr26-Gly27. The active site triad Asp25-Thr26-Gly27 is stabilized by a group of hydrogen bond network between the amino acid residues in the active site known as 'Fireman's grip' shown in **Figure 55**. The roof of the active site is covered by the two double stranded beta sheets from each subunit also called as flaps which contains glycine residues and are highly flexible in nature [98].

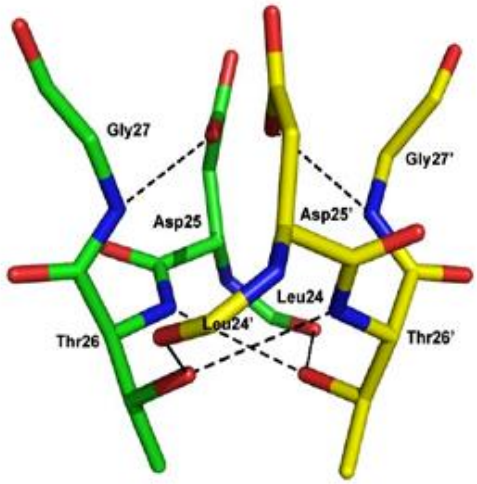


Figure 55: Fireman's grip: The H-bond networks in the active site pocket [98]

HIV protease inhibitors play a very crucial role in retarding the viral replication and the formation of infectious virions. These drugs were found to be very useful in reducing the viral load and improving the CD4 cell counts in AIDS patients. However, rapid emergence of drug resistance has been reported for almost all protease inhibitors currently in clinical use due to site-specific mutations in the enzyme [230]. The bioavailability and toxicity profiles of protease inhibitors are also of importance; thus, there exists an urgent need to discover a new generation of protease inhibitors that are more potent against these mutant forms of the virus and, at the same time, that exhibit low toxicity and high bioavailability [231, 232].

6.2. Structure (e-Pharmacophore) based design (Work flow as given in Figure 38)

There were more than 180 crystal structures available in PDB. We selected 8 crystal structures based on low resolution and high inhibitory activity of the bound crystal ligand as shown in **Figure 56**. They were 3OXC, 2Q5K, 3DJK, 3H5B, 3I6O, 3NDX, 2AID and 3T11 [233-238]. The bound crystal ligand of 3OXC was isoquinoline derivative (drug name-Saquinavir) with H-bond interaction with Asp25, Asp30, Gly48 and Asp125. The bound inhibitor of 2Q5K was pyrimidine derivative (drug name–Lopinavir) with H-bond interaction with Asp25, Gly27, Asp29 and Asp125. 3DJK had dioxepan derivative as its bound inhibitor and its H-bond interaction with Asp25, Asp125 and Gly127. 3H5B had furan derivative as its inhibitor and H-bonds with Asp25, Asp29 and Asp130. 3I6O had furan derivative and H-bonds with Asp25, Asp29, Asp30, Asp125, Asp130 and Gly27. 3NDX had a thiazole derivative (drug name-Ritonavir) as its inhibitor and H-bonds with Asp25, Asp29 and Gly27, Gly48. 2AID had a non-peptidic inhibitor of piperidine derivative and H-bonds with Asp25. 3T11 had a diazepin derivative as its bound inhibitor and its interactions were not yet revealed.

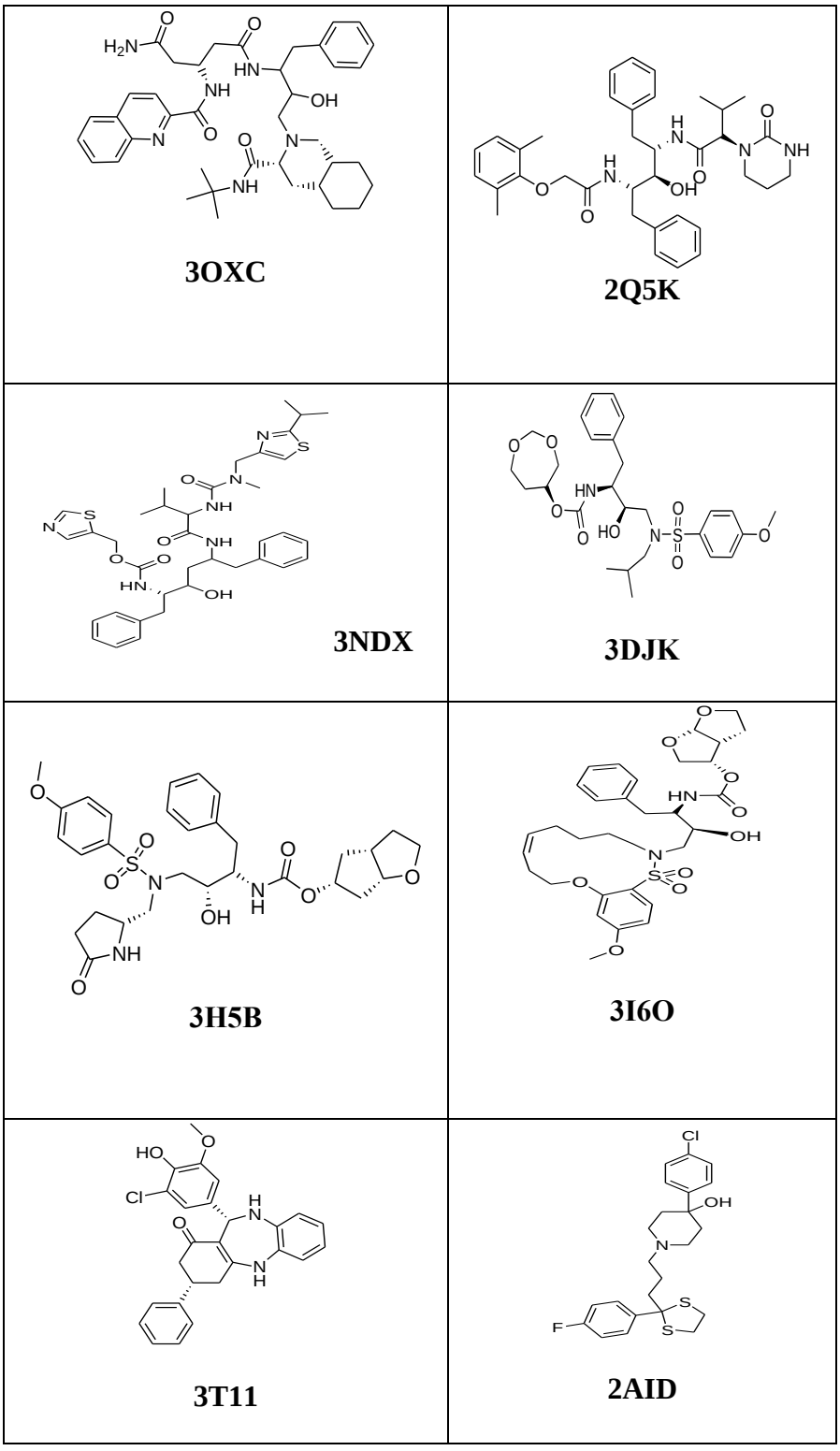


Figure 56: Structures of bound inhibitors from indicated co-crystals for HIV protease

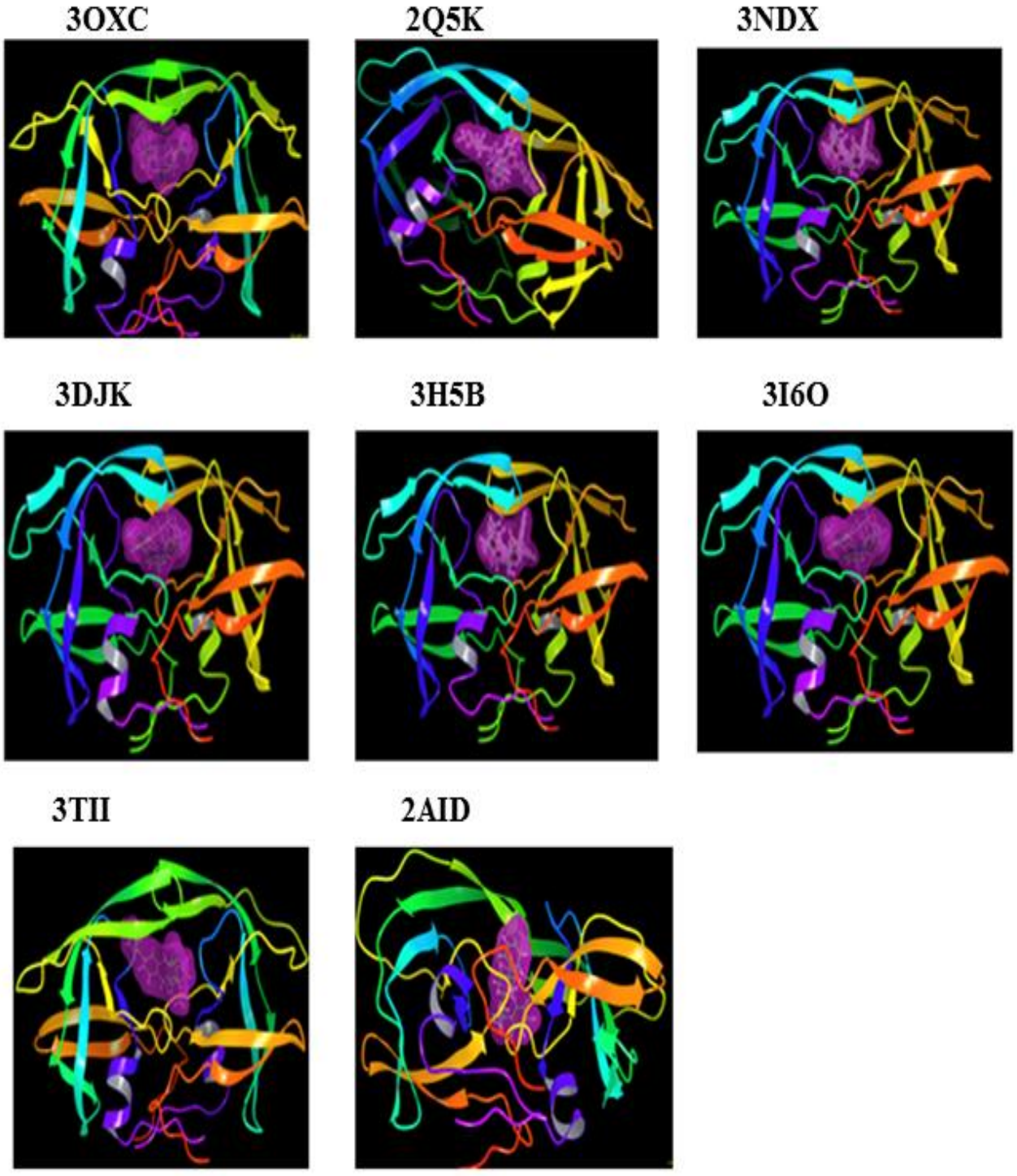


Figure 57: The eight PDB structures with their respective IDs and their active site region and the binding site pocket of the eight indicated crystal structures, demonstrated in purple color createdusing 'Create binding site surfaces' module of Schrodinger suite for HIV protease target

From the interaction patterns of these crystal ligands, the H-bond interactions found to be important were with Asp25, Asp29, Asp30, Asp125, Gly27, Gly48, Gly127 and Asp130.These crystal structures showed good resolution (1.00Å - 2.40Å) and their ligand activity ranged from 0.2 nM to 270 nM. 3OXC, 2Q5K, 3DJK, 3H5B, 3I6O and 3NDX were bound to peptidic inhibitors and 2AID and 3T11 were to bound non-peptidic inhibitors (**Figure 57).** The PDB structure resolutions and their activity values are given in the **Table 22.**

Table 22: Selected Crystal structures with their PDB IDs, resolution and their bound inhibitor's IC_{50} values for HIV protease target

S.No.	PDB-ID	Resolution Å	IC_{50} nM
1.	3OXC	1.16	0.2- 270
2.	2Q5K	1.95	35
3.	3NDX	1.03	67
4.	3DJK	1	4.9
5.	3H5B	1.29	26
6.	3I6O	1.17	7
7.	2AID	1.90	Ki=15000nM
8.	3T11	2.4	NA

We downloaded all these protein structures from PDB. We prepared all structures using protein preparation wizard tool in Maestro. We prepare energetically optimized and minimized protein structures for further use. Receptor Grid Generation tool in Maestro software package was used to generate energy grids for all prepared protein structures. The Grid table is given in **Table 23.**

Table 23: Grid information along with their PDB IDs employed for docking studies for HIV protease target

S.No	PDB code	X-Centre	Y-Centre	Z-Centre
1.	30XC	5.1111	-2.6919	14.8325
2.	2Q5K	20.3871	29.8650	12.8521
3.	3NDX	31.1662	0.1820	12.2442
4.	3DJK	16.2781	22.8492	17.2581
5.	3H5B	16.0505	22.7720	17.0518
6.	3I6O	16.9537	-20.3752	-5.6856
7.	3T11	1.5613	1.2062	19.4417
8.	2AID	5.4462	-0.0196	13.045

The protocol followed was similar to that of HCV NS5B **Figure 38**. Each feature pharmacophore hypotheses was assigned with energy values; lower the energy value higher the stablility of the feature as shown in **Table 24**. We kept 10 as maximum number of features. We used all possible feature combination and got 28 different combinations.

6.2.1. Energy-based pharmacophore generation and their validation

The e–pharmacophore method that of structure based and ligand based techniques was explored for 8 crystal structures of HIV proteases. Using the defalut grids (**Table 23)** the crystal ligands were docked and based on the XP descriptor information, the e-paharmacophores were generated. The pharmacophore hypotheses were developed by mapping Glide XP energetic terms onto pharmacophore sites which were calculated based on the structural and energy information between the protein and the ligand. The initial number of pharmacophore sites was set up to ten for all the crystal structures [233-238]. Generated e-pharmacophores with their all possible features and their energy values are given in **Table 24.**

Table 24: Possible e-pharmacophoric features with their score for each crystal structure for HIV protease target

PDB	Features	Score	Features
3OXC	A_5	-0.70	H-bond Acceptor
	A_2	-0.64	H-bond Acceptor
	D_7	-0.43	H-bond Donor
	A_3	-0.37	H-bond Acceptor
	D_9	-0.35	H-bond Donor
	D_8	-0.21	H-bond Donor
	R_{19}	-0.84	Aromatic Ring
2Q5K	A_3	-0.53	H-bond Acceptor
	D_6	-0.33	H-bond Donor
	D_8	-0.32	H-bond Donor
	A_2	-0.31	H-bond Acceptor
	D_7	-0.26	H-bond Donor
	R_{15}	-0.93	Aromatic Ring
	R_{16}	-1.02	Aromatic Ring
	R_{17}	-1.27	Aromatic Ring
3NDX	A_7	-0.63	H-bond Acceptor
	A_4	-0.57	H-bond Acceptor
	D_9	-0.32	H-bond Donor
	R_{19}	-0.70	Aromatic Ring
	R_{20}	-0.89	Aromatic Ring
3DJK	D_9	-0.47	H-bond Donor
	D_{10}	-0.65	H-bond Donor
	R_{14}	-0.99	Aromatic Ring
	R_{15}	-0.98	Aromatic Ring
3H5B	A_6	-0.63	H-bond Acceptor
	D_9	-0.41	H-bond Donor
	A_4	-0.28	H-bond Acceptor
	A_3	-0.09	H-bond Acceptor
	R_{16}	-1.00	Aromatic Ring
	R_{17}	-0.89	Aromatic Ring
3I6O	A_6	-0.63	H-bond Acceptor
	D_{10}	-0.39	H-bond Donor
	A_3	-0.28	H-bond Acceptor
	A_7	-0.13	H-bond Acceptor
	R_{16}	-1.07	Aromatic Ring
	R_{17}	-1.17	Aromatic Ring
3T11	D_6	-0.47	H-bond Donor
	R_{11}	-0.58	Aromatic Ring
	R_{12}	-1.62	Aromatic Ring
2AID	A_1	-0.63	H-bond Acceptor
	D_4	-0.33	H-bond Donor
	R_{13}	-0.59	Aromatic Ring

We preferred the pharmacophoric features with high negative values that were lower energy and were more stable. There were 28 e-pharmacophores with all possible combinations of features. The crystal structure from 3OXC gave an e-pharmacophore with seven features which was the maximum number of features compared to all other crystal structures. The crystal structures 3DJK, 3H5B, 3I6O and 2Q5K gave six maximum number of features; 3NDX gave 5 as maximum whereas 3T11 and 2AID gave only three features as maximum number of features. Using the decoy sets we validated all these possible e-pharmacophores. Using phase findmatches in Schrödinger suite we did pharmacophore screening for all these 28 e-pharmacophores with the data set collection of 1000 decoys and 64 known reported HIV inhibitors to validate the pharmacophores. Though few e-pharmacophores like 3NDX (AADRR) showed very good EF and 3OXC (AAADDDR), 3DJK (ADRR), 3H5B (ADRR) had EF values as shown in **Table 25**, we couldn't select them because they had very less number of actives in the pharmacophore screening output. The available number of pharmacophore sites and selected sites for each crystal structures are given in **Table 26**.

Table 25: Enrichment calculations for HIV protease target

S.N0	PDB	No of Features	Features	Total Hits[a]	Actives[b]	EF[c] 1%	Active[d] 1%	%Ya[e]	% A[f]	F(-ve)[g]	F(+ve)[h]
1.	3OXC	7	AAADDDR	20	10	11	9.4	50	16.6	54	10
2.	3OXC	6	AAADDR	125	20	4.5	4.7	16	31.3	44	105
3.	3OXC	5	AAADR	373	40	7.6	7.8	10.7	62.5	24	333
4.	3OXC	4	AADR	681	51	9.1	9.4	7.5	80	13	630
5.	3OXC	3	AAR	703	54	6	6.2	7.7	84	10	649
6.	3DJK	6	AADDRR	7	5	7.6	7.8	71.4	7.8	59	2
7.	3DJK	5	AADRR	25	7	7.6	7.8	28	10.9	57	18
8.	3DJK	4	ADRR	168	26	11	10.9	15.5	40.6	38	104
9.	3DJK	3	DRR	652	46	6	4.7	7.1	71.9	18	606
10.	3H5B	6	AAADRR	18	7	9.1	7.8	38.9	10.9	57	11
11.	3H5B	5	AADDR	28	7	7.6	7.8	25	10.9	57	21
12.	3H5B	4	ADRR	196	26	11	9.4	13.3	40.6	38	170
13.	3H5B	3	ARR	568	40	4.5	4.7	7.1	62.5	24	528
14.	3NDX	5	AADRR	155	24	14	14.1	15.5	37.5	40	131
15.	3NDX	4	AARR	325	36	7.6	7.8	11.1	56.3	28	289
16.	3NDX	3	ARR	645	45	9.1	9.4	7	70.3	19	600
17.	3I6O	6	AAADRRR	8	3	4.5	4.7	37.5	37.5	61	5
18.	3I6O	5	AAARR	28	5	3	3.1	17.9	7.8	59	23
19.	3I6O	4	AARR	112	11	4.5	4.7	9.8	17.2	53	101
20.	3I6O	3	ARR	530	41	3	3.1	7.7	64.1	23	489
21.	2Q5K	6	ADDRRR	8	4	6	6.2	50	6.3	60	4
22.	2Q5K	5	ADRRR	8	4	6	6.2	50	6.3	60	4
23.	2Q5K	4	ARRRR	59	13	11	9.4	22	20	51	46
24.	2Q5K	3	ARR	727	48	9.1	9.4	6.6	75	16	679

Contd..

25.	2Q5K	3	ARR	246	29	11	9.4	11.8	45	35	217
26.	2Q5K	3	ARR	474	40	11	9.4	8.5	62.5	24	434
27.	3T11	3	DRR	784	47	0	0	6	73	17	737
28.	2AID	3	ADR	711	52	1.5	1.6	7.3	81.3	12	659

[a] Total number of hit molecules from the database

[b] Total number of active molecules in hit list

[c] Enrichment factor using formula

[d] Top 1% Actives in the hit.

[e] Yield of actives = [$(H_a/H_t) \times 100$]

[f] Ratio of actives = [$(H_a/A) \times 100$]

[g] False negatives = [$A - H_a$]

[h] False Positives = [$H_t - H_a$]

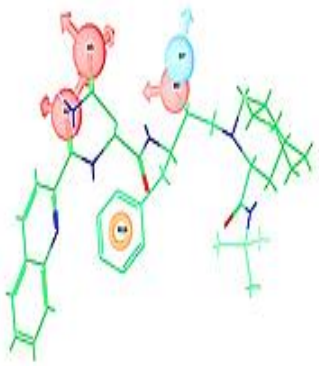
Table 26: The possible number of available pharmacophore sites from each PDB structure, the final selected sites and their final selected hypotheses for HIV protease target

S.No	PDB code	Maximum no of possible sites	No of selected sites	Hypothesis*
1.	3OXC	7	5	AAADR
			4	AADR
			3	AAR
2.	2Q5K	8	3	ARR
			3	ARR
3.	3NDX	5	4	AARR
			3	ARR
4.	3DJK	6	3	DRR
5.	3H5B	6	3	ARR
6.	3I6O	7	3	ARR
7.	3T11	3	3	DRR
8.	2AID	3	3	ADR

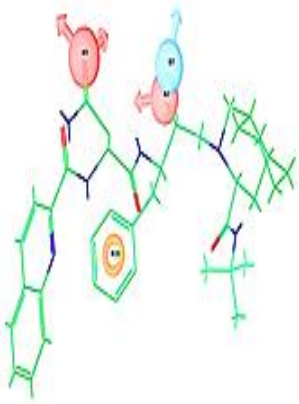
*A-Acceptor, D-Donor, R-Ring aromatic, N-Negatively ionizable and H-Hydrophobic

Among these 28 e-pharmacophores we selected 12 e-pharmacophores (**Table 26)** based on their good EF 1%, % of actives, BEDROC, RIE. The selected e-pharmacophores are shown in **Figure 58** and their respective EF values are given in **Tables 27 & 28**.

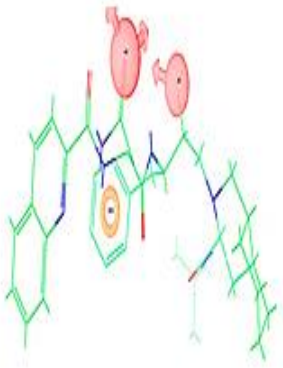
The e-pharmacophores 3OXC (AAADR) and 3NDX (AARR) showed high calculated EF and GH values. The e-pharmacophores 3OXC (AADR), 3OXC (AAR) and 2AID (ADR) retrieved more than 80% of the actives and other selected e-pharmacophores retrieved more than 60% and 70% actives except the e-pharmacophore 3NDX (AARR) which retrieved only 56.3% but showed very good calculated EF and GH values. This indicated that these e-pharmacophores were eligible to retrieve more number of actives during pharmacophore screening.



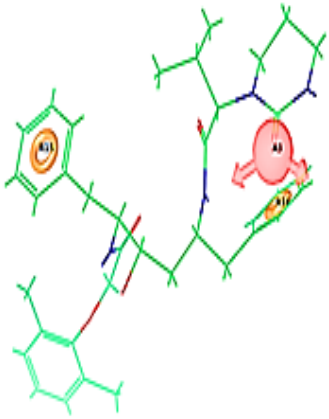
3OXC-5F



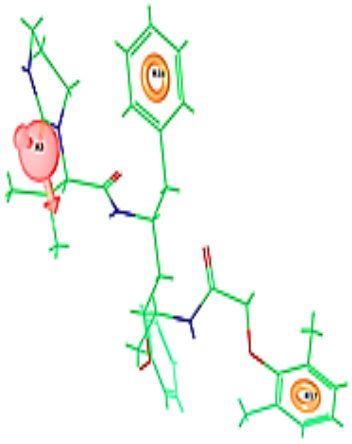
3OXC-4F



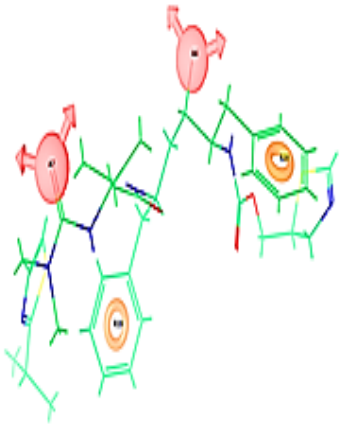
3OXC-3F



2Q5K-3F



2Q5K-3F



3NDX-4F

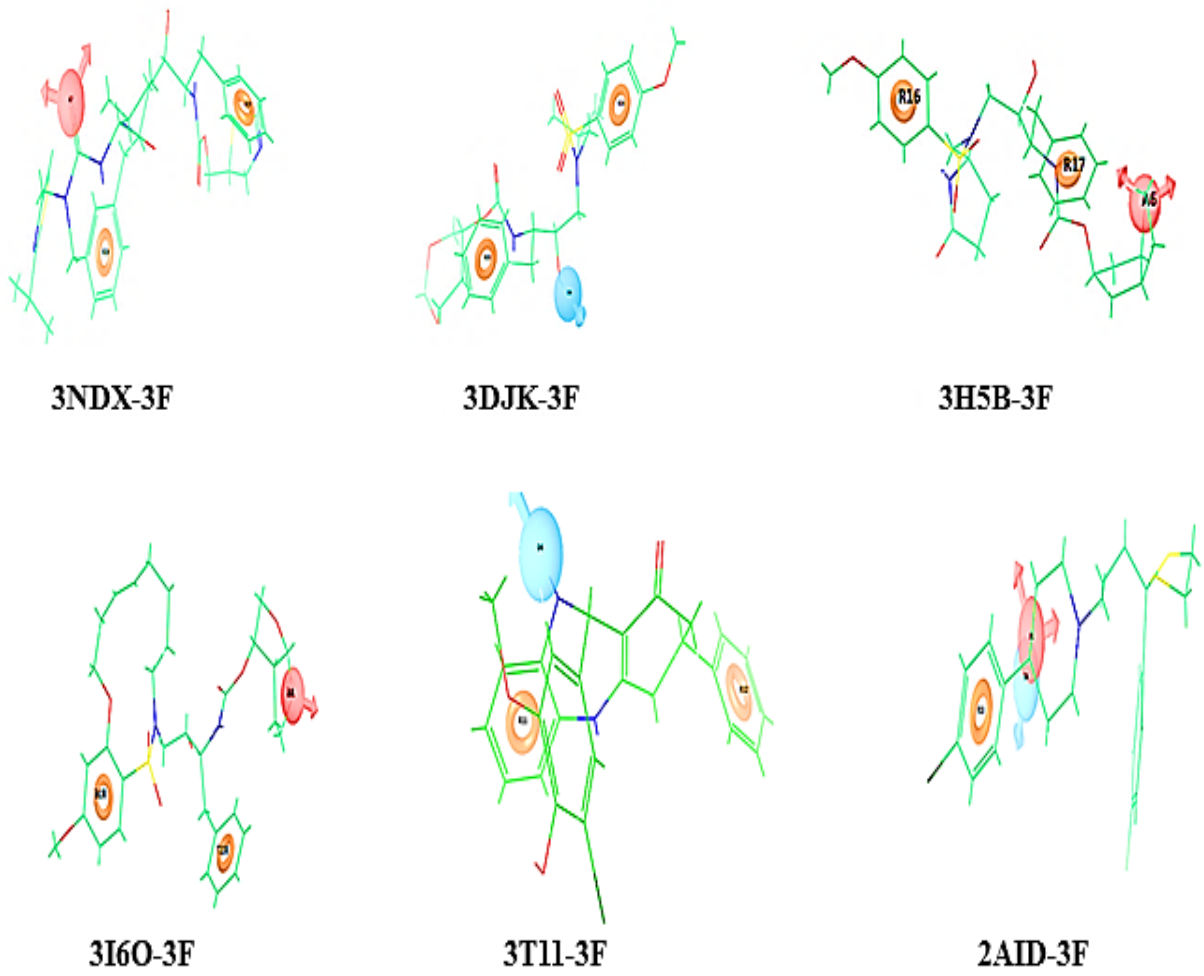


Figure 58: The finally selected 12 e-pharmacophores for HIV protease target

Table 27: Selected 12 e-pharmacophores with their calculated overall EF, goodness of fit (GH), % of actives, yield of actives, false negative and false positive values for HIV protease target

S.NO	PDB	No of Features	Features	T.Hits[a]	Actives[b]	EF[c]	GH[d]	%Ya[e]	%A[f]	F(-ve)[g]	F(+ve)[h]
1.	3OXC	5	AAADR	373	40	1.68	0.153	10.7	62.5	24	333
2.		4	AADR	681	51	1.17	0.083	7.5	80	13	630
3.		3	AAR	703	54	1.20	0.082	7.7	84	10	649
4.	2Q5K	3	ARR	727	48	1.03	0.065	6.6	75	16	679
5.		3	ARR	474	40	1.32	0.118	8.5	62.5	24	434
6.	3NDX	4	AARR	325	36	1.73	0.155	11.1	56.3	28	289
7.		3	ARR	645	45	1.09	0.082	7	70.3	19	600
8.	3DJK	3	DRR	652	46	1.10	0.082	7.1	71.9	18	606
9.	3H5B	3	ARR	568	40	1.10	0.091	7.1	62.5	24	528
10.	3I6O	3	ARR	530	41	1.21	0.104	7.7	64.1	23	489
11.	3T11	3	DRR	784	47	0.94	0.049	6	73	17	737
12.	2AID	3	ADR	711	52	1.14	0.076	7.3	81.3	12	659

[a] Total number of hit molecules from the database

[b] Total number of active molecules in hit list

[c] Enrichment factor using formula

[d] Goodness of Fit using formula

[e] Yield of actives = [$(H_a/H_t) \times 100$]

[f] Ratio of actives = [$(H_a/A) \times 100$]

[g] False Negatives = [$A - H_a$]

[h] False Positives = [$H_t - H_a$]

Table 28: The selected 12 e-pharmacophores with their number of features, their various combinations of features, enrichment factor at top 1%, RIE, ROC, BEDROC (α-20) and (α-160.9) for HIV protease target

S.NO	PDB	No of Features	features	EF 1%	RIE	ROC	BEDROC (α =20)	BEDROC (α =160)
1.	3OXC	5	AAADR	7.6	5.14	0.99	0.442	0.522
2.		4	AADR	9.1	5.19	0.94	0.446	0.557
3.		3	AAR	6	4.84	0.90	0.416	0.431
4.	2Q5K	3	ARR	9.1	4.66	0.94	0.401	0.442
5.		3	ARR	11	5.21	0.99	0.448	0.535
6.	3NDX	4	AARR	7.6	5.12	1.00	0.440	0.478
7.		3	ARR	9.1	3.90	0.94	0.335	0.440
8.	3DJK	3	DRR	6	3.11	0.91	0.267	0.346
9.	3H5B	3	ARR	4.5	3.87	0.97	0.333	0.313
10.	3I6O	3	ARR	3	3.04	0.96	0.261	0.263
11.	3T11	3	DRR	0	0.42	0.82	0.036	0.000
12.	2AID	3	ADR	1.5	3.38	0.86	0.290	0.068

EF- Enrichment Factor at 1% of the decoy dataset
RIE- Robust Initial Enhancement
ROC-Receiver Operating Characteristic curve value
BEDROC- Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic

The distance mapping among the features is presented in **Table 29**. Ring (R) feature was common among all the pharmacophoric sites. Acceptor (A) feature was common among many pharmacophores except in 3DJK (DRR) and 3T11 (DRR). There were only one e-pharmacophore with five features that was 3OXC (AAADR) and two with four features 3OXC (AADR) and 3NDX (AARR) while remaining pharmacophores had only 3 features. Among the crystal structure 3OXC e-pharmacophores there were three e-pharmacophores selected (AADRR, AADR and AAR).

Table 29: Distance between the features of e-pharmacophores for HIV protease target

PDB	Features*	A1-R1 (Å)	R1-D (Å)	D-A1 (Å)	D-R2 (Å)	R1-R2 (Å)	R2-A1 (Å)	R1-A2 (Å)	A1-A2 (Å)	A2-D1 (Å)	A3-R1 (Å)	A2-A3 (Å)
3OXC	AAADR	4.167	5.101	0.953	NA	NA	NA	7.336	6.845	6.880	7.336	3.584
	AADR	4.167	5.101	0.953	NA	NA	NA	7.336	6.845	6.880	7.336	3.584
	AAR	4.167	NA	NA	NA	NA	NA	7.336	6.845	NA	NA	NA
2Q5K	ARR	9.199	NA	NA	NA	8.203	4.765	NA	NA	NA	NA	NA
	ARR	4.765	NA	NA	NA	8.942	12.070^{max}	NA	NA	NA	NA	NA
3NDX	AARR	5.799	NA	NA	NA	8.102	9.386	4.311	7.828	NA	NA	NA
	ARR	5.799	NA	NA	NA	8.102	9.386	4.311	NA	NA	NA	NA
3DJK	DRR	NA	4.309	NA	6.185	6.530	NA	NA	NA	NA	NA	NA
3H5B	ARR	11.140	NA	NA	NA	6.312	6.302	NA	NA	NA	NA	NA
3I6O	ARR	6.524	NA	NA	NA	6.419	11.760	NA	NA	NA	NA	NA
3T11	DRR	NA	2.419	NA	5.925	7.020	NA	NA	NA	NA	NA	NA
2AID	ADR	3.538	3.831	0.951^{min}	NA	NA	NA	NA	NA	NA	NA	NA

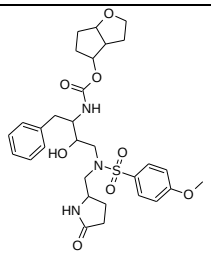
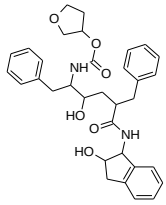
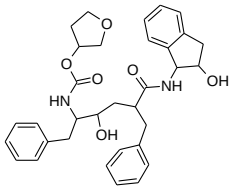
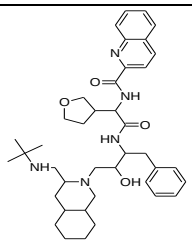
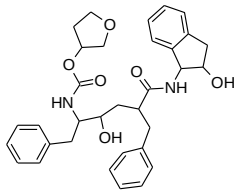
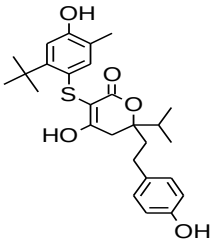
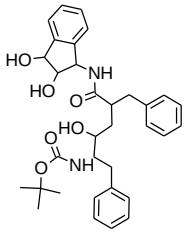
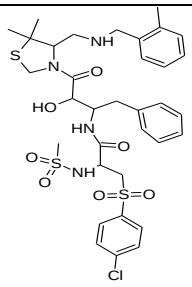
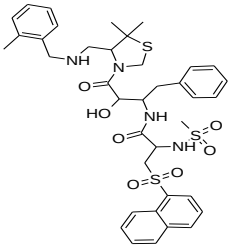
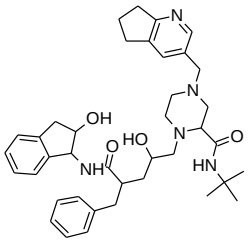
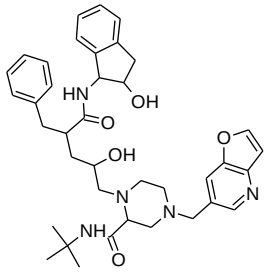
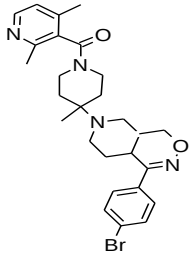
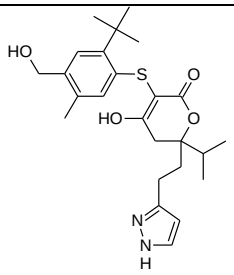
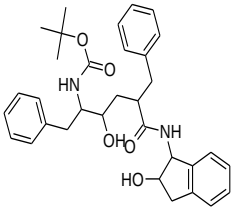
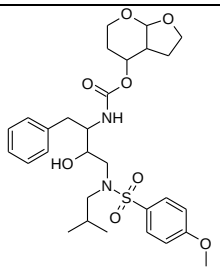
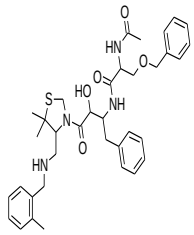
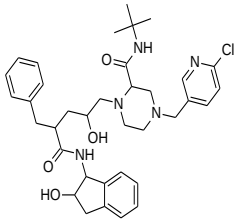
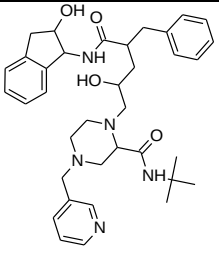
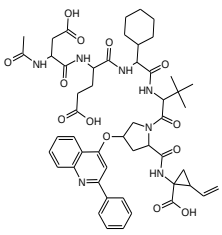
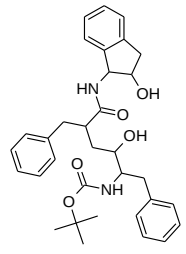
Distance shown in Å, *A-Acceptor, D-Donor, R-Ring aromatic, N-Negatively ionizable and H-Hydrophobic. NA-not applicable

While screening the database, we kept the option of must match five sites, four sites and three sites respectively to get diverse compounds though all these three were derived from the same crystal structure. In four feature e-pharmacophore models there were two acceptor (A) features and one ring (R) were found to be common. But when we compared the distance between these common features like A1-A2 in 3OXC (AADR) which was 11.503 Å and in 3NDX (AARR) was 7.828 Å and A-R distances were 4.167 Å and 7.336 Å 3OXC and 3NDX 4.311 Å and 5.799 Å. Though they had similar features, the distances were different and hence utilized further to retrieve more diversed compounds. There were nine pharmacophores with three features, all of them had atleast one ring (R) feature as common, and this showed that the new inhibitor which we would design could possess atleast one ring feature. Among these nine 3 feature pharmacophores, two had DRR, one AAR, one ADR and five ARR. When we compared the distance between the features of DRR, the R1-R2 distance was little closer (6.530 Å and 6.296 Å), but the distance between D-R was quite different from each other 4.309Å and 6.185Å in 3DJK (DRR) and 3.340 Å and 7.481 Å in 3T11 (DRR). There were five e-pharmacophore with ARR features and when we compared the distance between the R1-R2 features, three showed little close values like 8.203 Å in 2Q5K (A3R15R16), 8.942 Å in 2Q5K (A3R16R17) and 8.102 Å in 3NDX (ARR) and also two showed 6.312Å in 3H5B (ARR) and 6.419Å in 3I6O (ARR). But for these e-pharmacophores A-R distances were different. The pharmacophore 2Q5K (A3R16R17) showed the maximum distance of 12.07 Å and the e-pharmacophore 2AID (ADR) had the minimum distance of 0.951Å. This showed that all twelve e-pharmacophores were unique and they must have the ability to screen more number of diverse ligands.

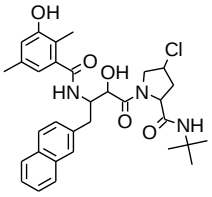
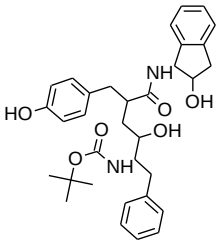
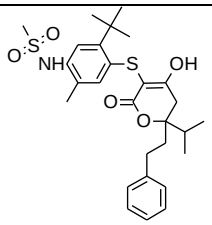
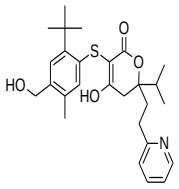
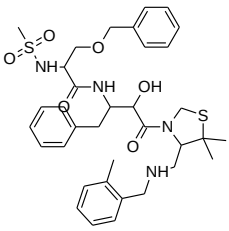
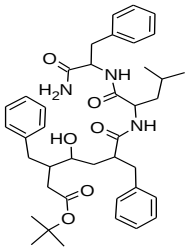
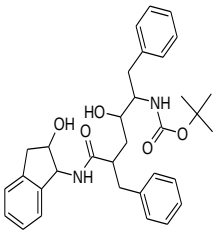
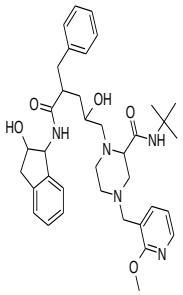
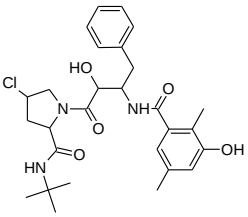
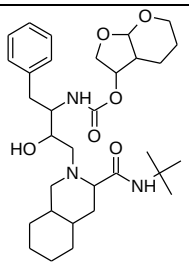
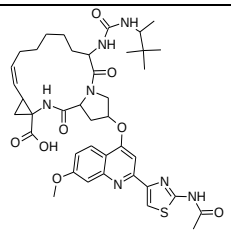
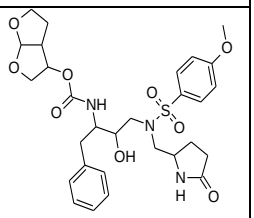
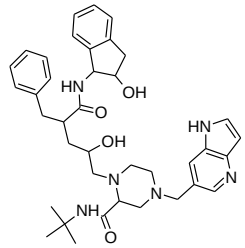
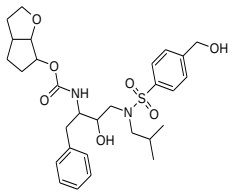
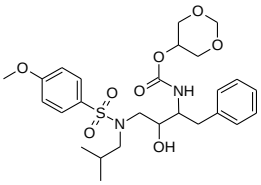
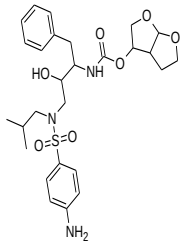
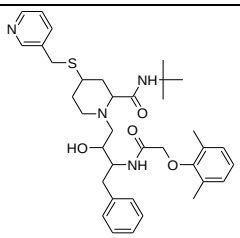
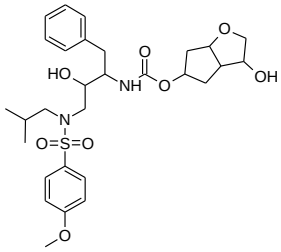
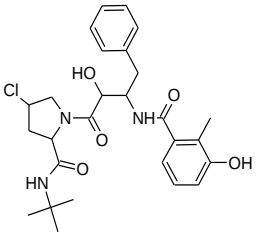
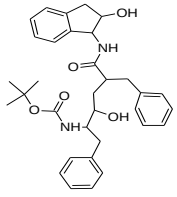
6.3. Ligand based design

We chose 1535 HIV protease inhibitors from various literatures with known IC_{50} values [234-269]. These inhibitors showed a wide range of activity from 0.026 nM to 316 µM with structural diversity. 134 compounds were selected from 168 clusters as representative molecules as shown in **Figure 59**. Clustered molecules with structural diversity were utilized for 3D QSAR development. Set of sixty four already reported inhibitors were used as actives to validate the pharmacophores and were also prepared using LigPrep 2.5 [150]. The threshold range of the active and inactive pIC_{50} was 10.585 and 3.500 respectively.

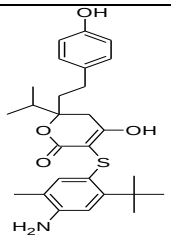
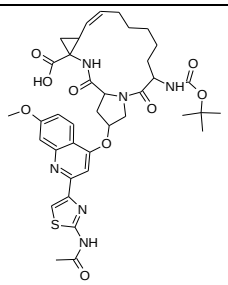
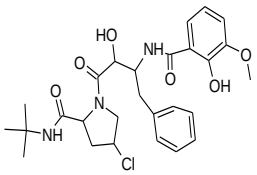
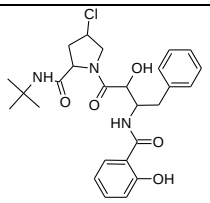
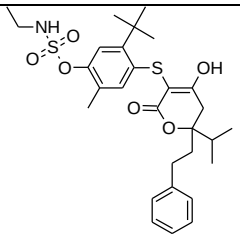
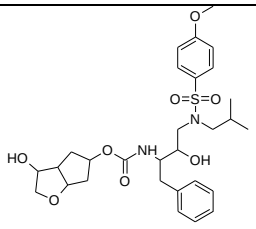
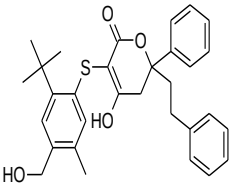
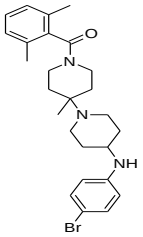
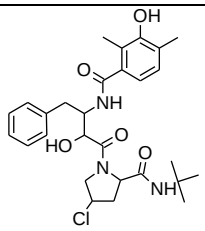
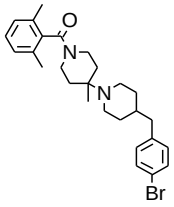
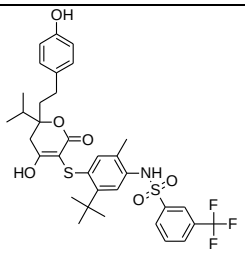
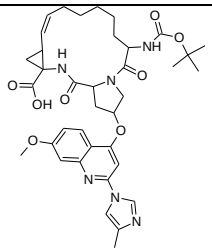
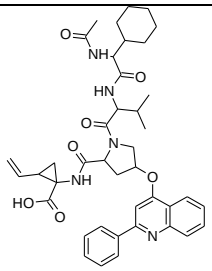
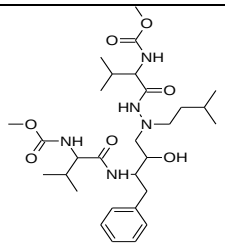
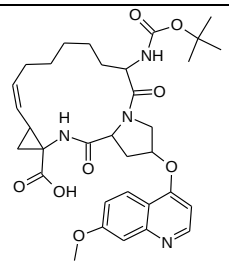
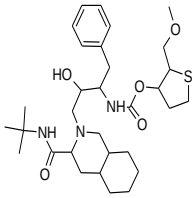
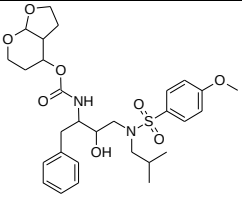
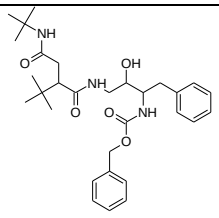
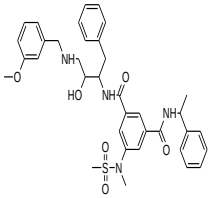
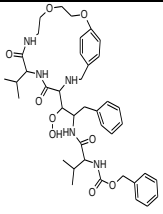
Figure 59: Compounds used for 3D QSAR (ligand) based pharmacophore development and 3D QSAR model for HIV protease target

1. 	2. 	3. 	4. 
5. 	6. 	7. 	8. 
9. 	10. 	11. 	12. 
13. 	14. 	15. 	16. 
17. 	18. 	19. 	20. 

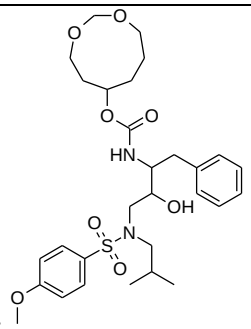
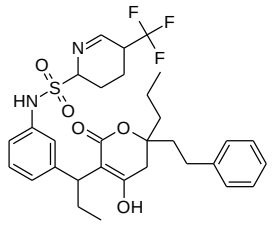
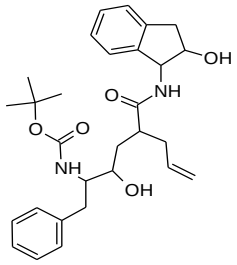
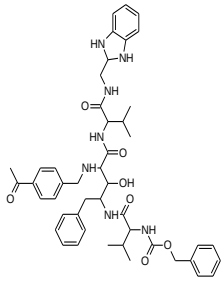
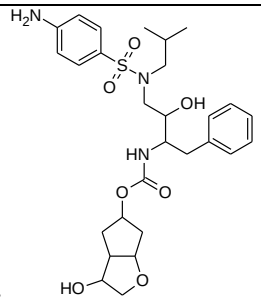
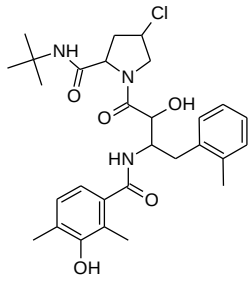
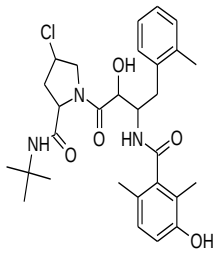
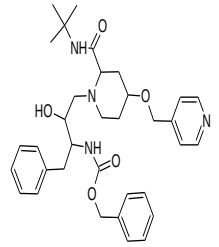
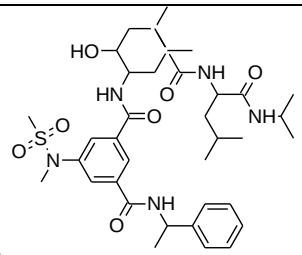
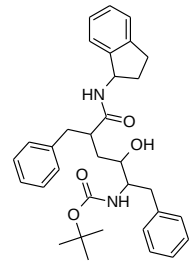
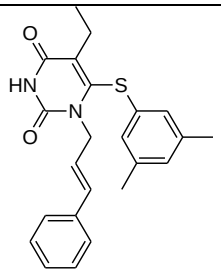
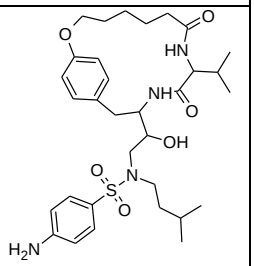
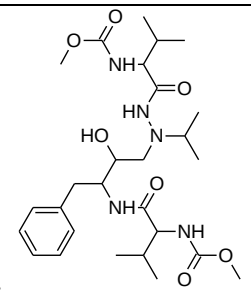
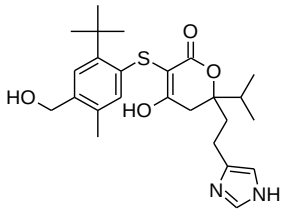
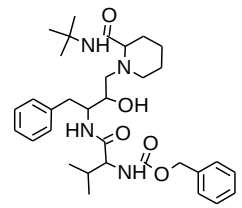
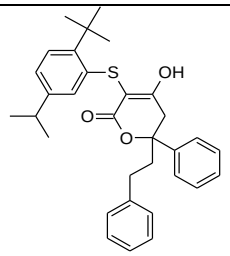
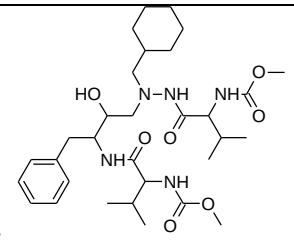
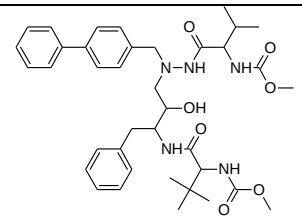
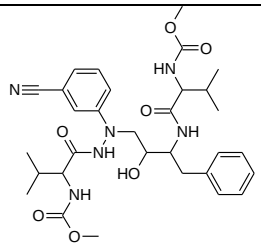
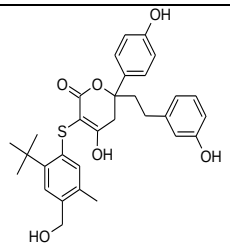
Contd..

21. 	22. 	23. 	24. 
25. 	26. 	27. 	28. 
29. 	30. 	31. 	32. 
33. 	34. 	35. 	36. 
37. 	38. 	39. 	40. 

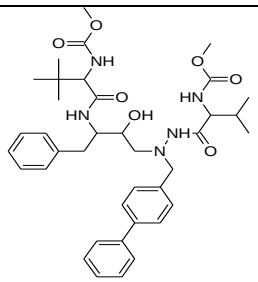
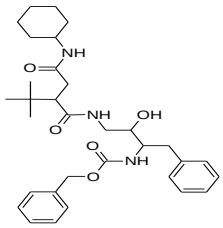
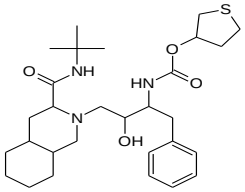
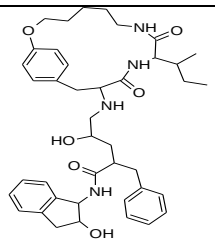
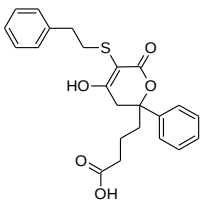
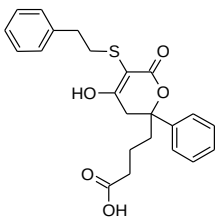
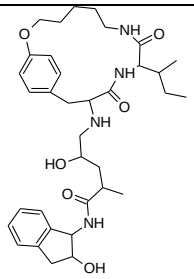
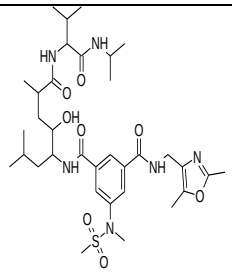
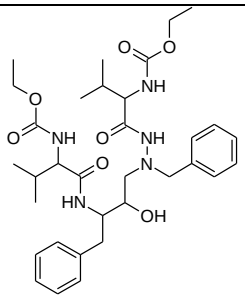
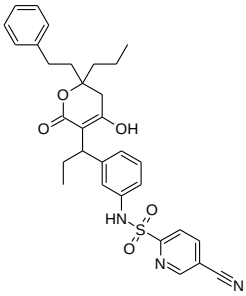
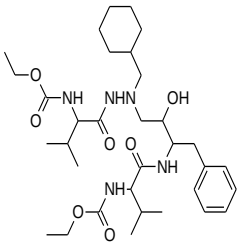
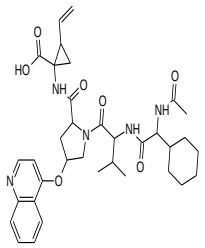
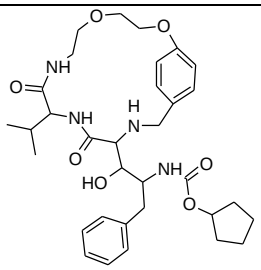
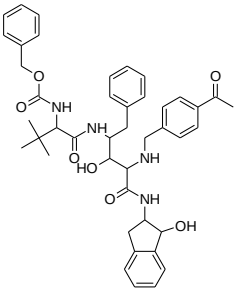
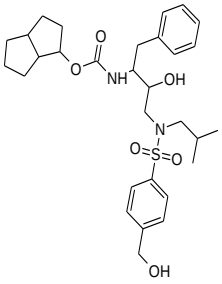
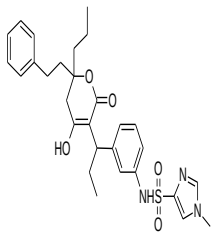
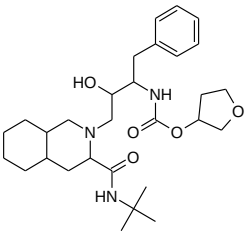
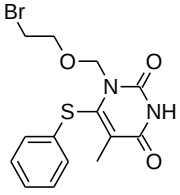
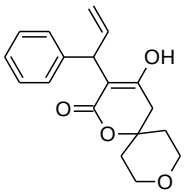
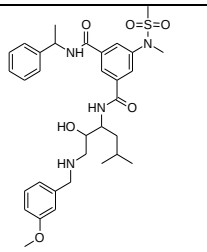
Contd..

41. 	42. 	43. 	44. 
45. 	46. 	47. 	48. 
49. 	50. 	51. 	52. 
53. 	54. 	55. 	56. 
57. 	58. 	59. 	60. 

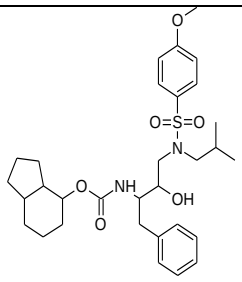
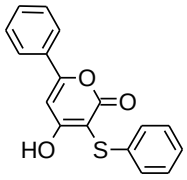
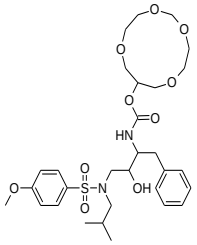
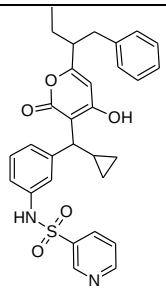
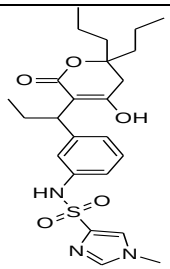
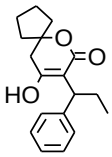
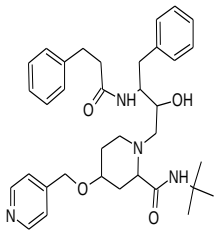
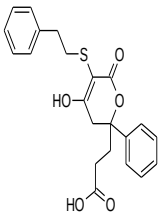
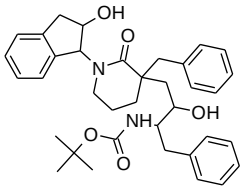
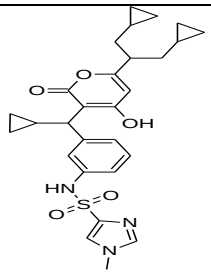
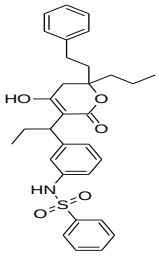
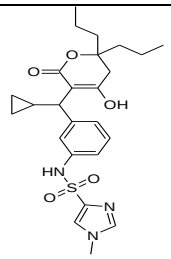
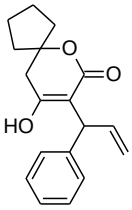
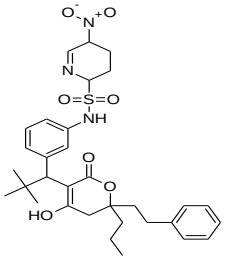
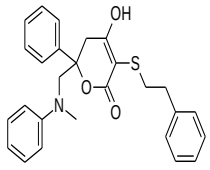
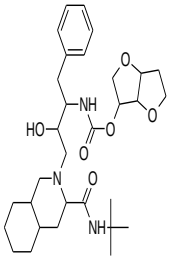
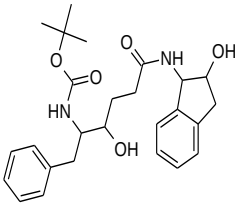
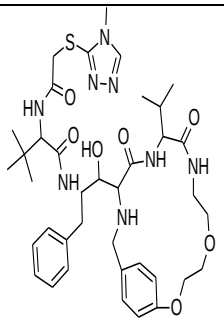
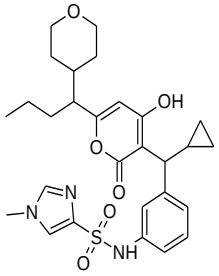
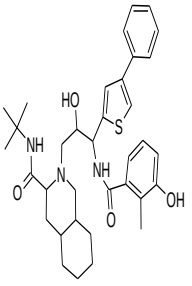
Contd..

61. 	62. 	63. 	64. 
65. 	66. 	67. 	68. 
69. 	70. 	71. 	72. 
73. 	74. 	75. 	76. 
77. 	78. 	79. 	80. 

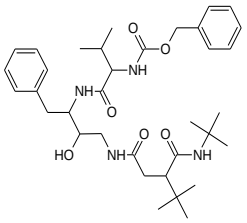
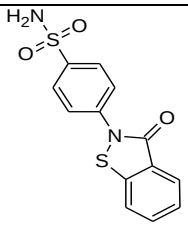
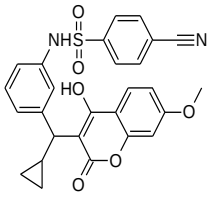
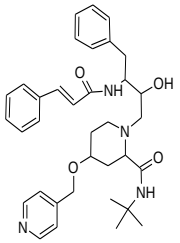
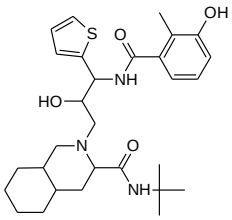
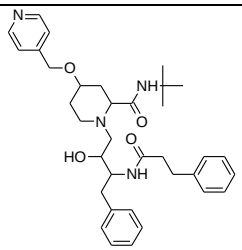
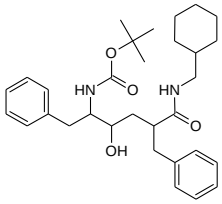
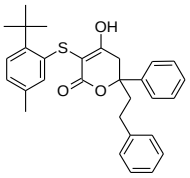
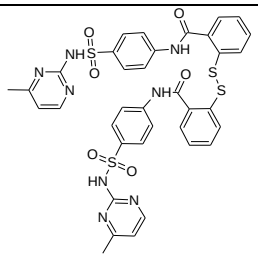
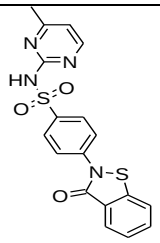
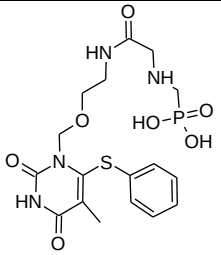
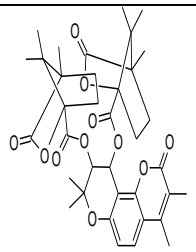
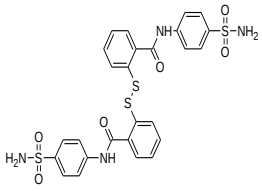
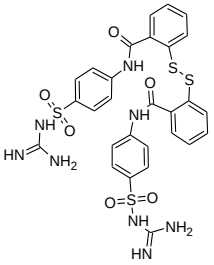
Contd..

81. 	82. 	83. 	84. 
85. 	86. 	87. 	88. 
89. 	90. 	91. 	92. 
93. 	94. 	95. 	96. 
97. 	98. 	99. 	100. 

Contd..

101. 	102. 	103. 	104. 
105. 	106. 	107. 	108. 
109. 	110. 	111. 	112. 
113. 	114. 	115. 	116. 
117. 	118. 	119. 	120. 

Contd..

121. 	122. 	123. 	124. 
125. 	126. 	127. 	128. 
129. 	130. 	131. 	132. 
133. 	134. 		

The 134 ligands were aligned with the pharmacophore template of compound with high active score compound. We selected 134 promising compounds to develop our 3D QSAR model. We were randomly as 103 for training set and 31 for test set (**Table 30**), by using the method "Automated Random Selection" option present in the PHASE module. A model with PLS factor five was considered as the best statistical model. This model was validated by predicting activities of test set, training set and overall molecules **Table 30**.

Table 30: Compounds for 3D QSAR study with their experimental, predicted activity and fitness score for HIV protease target.

S. No.	Fitness	Experimental activity	Predicted activity	Error	Phase QSAR set
1.	1.502	9.800	9.080	0.720	test
2.	2.051	9.600	9.130	0.470	test
3.	0.698	9.600	8.420	1.180	test
4.	2.053	9.319	8.800	0.519	test
5.	2.311	9.301	9.380	-0.079	test
6.	1.565	9.252	9.010	0.242	test
7.	1.539	9.222	9.010	0.212	test
8.	1.587	9.097	8.430	0.667	test
9.	1.582	8.921	9.010	-0.089	test
10.	1.409	8.886	8.280	0.606	test
11.	1.434	8.770	8.540	0.230	test
12.	1.036	8.770	8.750	0.020	test
13.	1.456	8.569	8.370	0.199	test
14.	1.589	8.260	8.150	0.110	test
15.	0.956	8.086	8.280	-0.194	test
16.	0.712	7.738	8.050	-0.312	test
17.	1.424	7.538	7.510	0.028	test
18.	1.378	7.409	7.440	-0.031	test
19.	1.593	7.328	7.140	0.188	test
20.	1.739	7.208	6.970	0.238	test
21.	1.453	6.886	6.700	0.186	test
22.	1.713	6.804	6.660	0.144	test
23.	1.045	6.770	6.670	0.100	test
24.	1.347	6.398	7.450	-1.052	test
25.	1.438	6.102	6.780	-0.678	test
26.	1.012	6.000	5.740	0.260	test
27.	1.415	5.745	6.750	-1.005	test
28.	1.186	5.553	6.130	-0.577	test
29.	1.362	5.000	6.050	-1.050	test
30.	1.385	4.790	5.250	-0.460	test
31.	1.029	4.542	5.940	-1.398	test
32.	3.000	10.585	10.690	-0.105	training
33.	1.528	10.523	10.150	0.373	training
34.	1.237	10.523	10.410	0.113	training

Contd..

35.	1.484	10.270	10.240	0.030	training
36.	2.106	10.220	10.340	-0.120	training
37.	1.536	10.220	10.190	0.030	training
38.	1.647	10.081	9.940	0.141	training
39.	1.809	10.000	10.030	-0.030	training
40.	1.663	9.959	9.970	-0.011	training
41.	1.061	9.854	9.920	-0.066	training
42.	1.064	9.770	9.910	-0.140	training
43.	1.272	9.699	9.660	0.039	training
44.	0.949	9.553	9.610	-0.057	training
45.	1.214	9.523	9.540	-0.017	training
46.	1.277	9.409	9.350	0.059	training
47.	0.986	9.310	9.340	-0.030	training
48.	1.162	9.301	9.140	0.161	training
49.	1.326	9.229	9.240	-0.011	training
50.	1.345	9.097	9.120	-0.023	training
51.	0.826	9.081	9.050	0.031	training
52.	1.269	8.959	8.900	0.059	training
53.	1.169	8.921	8.870	0.051	training
54.	1.550	8.890	8.930	-0.040	training
55.	1.483	8.886	8.880	0.006	training
56.	0.797	8.854	8.830	0.024	training
57.	2.279	8.745	8.640	0.105	training
58.	1.931	8.538	8.440	0.098	training
59.	1.116	8.523	8.600	-0.077	training
60.	1.425	8.469	8.640	-0.171	training
61.	0.596	8.319	8.300	0.019	training
62.	1.510	8.180	8.190	-0.010	training
63.	1.160	8.155	8.220	-0.065	training
64.	1.084	8.086	8.120	-0.034	training
65.	1.448	8.051	8.260	-0.209	training
66.	1.437	8.000	8.110	-0.110	training
67.	1.293	7.886	7.820	0.066	training
68.	1.038	7.854	7.900	-0.046	training
69.	1.738	7.796	7.790	0.006	training
70.	0.907	7.796	7.660	0.136	training
71.	1.925	7.721	7.740	-0.019	training
72.	1.182	7.721	7.740	-0.019	training
73.	1.198	7.706	7.760	-0.054	training

Contd..

74.	1.551	7.678	7.610	0.068	training
75.	1.229	7.620	7.780	-0.160	training
76.	1.241	7.602	7.510	0.092	training
77.	1.632	7.560	7.510	0.050	training
78.	1.568	7.523	7.570	-0.047	training
79.	0.595	7.523	7.550	-0.027	training
80.	0.813	7.470	7.430	0.040	training
81.	1.412	7.465	7.470	-0.005	training
82.	1.593	7.409	7.360	0.049	training
83.	1.897	7.387	7.340	0.047	training
84.	0.958	7.347	7.360	-0.013	training
85.	1.255	7.302	7.490	-0.188	training
86.	1.148	7.208	7.220	-0.012	training
87.	1.721	7.155	7.330	-0.175	training
88.	1.484	7.076	7.130	-0.054	training
89.	1.702	7.071	6.940	0.131	training
90.	1.422	7.071	7.010	0.061	training
91.	1.448	6.959	6.860	0.099	training
92.	1.479	6.921	6.960	-0.039	training
93.	1.245	6.879	6.870	0.009	training
94.	1.406	6.870	6.920	-0.050	training
95.	1.033	6.824	6.760	0.064	training
96.	1.664	6.752	6.690	0.062	training
97.	1.620	6.752	6.730	0.022	training
98.	1.818	6.616	6.580	0.036	training
99.	1.402	6.569	6.770	-0.201	training
100.	0.700	6.301	6.260	0.041	training
101.	1.518	6.159	6.170	-0.011	training
102.	1.659	6.108	5.910	0.198	training
103.	0.969	6.097	5.930	0.167	training
104.	1.430	6.038	6.060	-0.022	training
105.	2.075	6.000	6.280	-0.280	training
106.	1.477	6.000	6.110	-0.110	training
107.	1.415	6.000	6.050	-0.050	training
108.	1.125	6.000	6.040	-0.040	training
109.	1.227	5.921	5.880	0.041	training
110.	1.029	5.921	5.840	0.081	training
111.	1.381	5.900	5.870	0.030	training
112.	0.975	5.854	5.910	-0.056	training

113.	0.701	5.796	5.770	0.026	training
114.	1.306	5.642	5.590	0.052	training
115.	1.549	5.638	5.610	0.028	training
116.	1.508	5.530	5.440	0.090	training
117.	1.468	5.523	5.680	-0.157	training
118.	1.445	5.523	5.390	0.133	training
119.	1.409	5.523	5.520	0.003	training
120.	1.057	5.398	5.340	0.058	training
121.	1.315	5.377	5.380	-0.003	training
122.	1.663	5.314	5.300	0.014	training
123.	1.468	5.000	5.040	-0.040	training
124.	1.172	4.959	4.980	-0.021	training
125.	1.327	4.699	4.570	0.129	training
126.	1.227	4.523	4.850	-0.327	training
127.	1.167	4.520	4.570	-0.050	training
128.	1.365	4.456	4.560	-0.104	training
129.	1.302	4.246	4.310	-0.064	training
130.	1.135	4.241	4.170	0.071	training
131.	1.408	4.000	3.890	0.110	training
132.	1.598	3.812	3.750	0.062	training
133.	1.162	3.664	3.560	0.104	training
134.	1.231	3.500	3.550	-0.050	training

6.3.1. Ligand based pharmacophore generation and its validation

Pharmacophore dataset comprising of 134 molecules were divided randomly as 103 for training set and 31 for test set as described earlier. PHASE provided the means to build QSAR models using the activities of the ligands that match a given hypothesis. To bulid the QSAR model, the compounds with pIC_{50} values more than 10.00 were labelled as actives and less than 4.50 were labelled as inactives. The molecules present in the Pharmaset were utilized for the generation of the hypothesis. The Pharmaset consisted of 8 actives, 7 inactives and remaining 119 were moderately actives. There were 8 active molecules in the Pharmaset, which was utilized to develop the pharmacophore and the 7 inactives were used in the scoring function to identify the best common pharmacophore. The Find common hypotheses step was run with option to generate maximum five pharmacophoric sites and minimum four pharmacophore sites led to generation of total of 16 pharmacophore models with different combinations of features. 2131 pharmacophore hypotheses were generated with AAADH combination, AAAAH 2465, AAAHR 6465, AADRR 1121, AADHR 3687, AAARR 1490, AAADD 7, ADDHR 301, ADHRR 232, AADDH 239, AAADR 5754, ADDRR 27, AAHRR 268, AAAAR 8882, AADDR 458 and the combination AAAAD, with which 226 pharmacophore hypotheses were generated. These pharmacophores were scored according to their fitness to the inhibitors. Clustering the scored hypothesis resulted in 57 groups, out of these we selected top four hypotheses based on their survival actives, inactives, vector, volume scores and their energy values as shown in **Table 31** and **Figure 60**.

Table 31: 3D QSAR Hypotheses with their scores

S.NO	Hypothesis	Survival Score	Survival-inactive Score	Vector Score	Volume Score	Site Score	Energy
1.	AAAHR	2.583	1.560	0.787	0.377	0.42	1.232
2.	AAHRR	2.460	1.266	0.744	0.372	0.34	9.069
3.	AAAAR	2.567	1.543	0.775	0.312	0.48	9.815
4.	AAADH	2.555	1.386	0.858	0.311	0.37	8.885

*A-Acceptor, D-Donor, R-Ring aromatic, N-Negatively ionizable and H-Hydrophobic

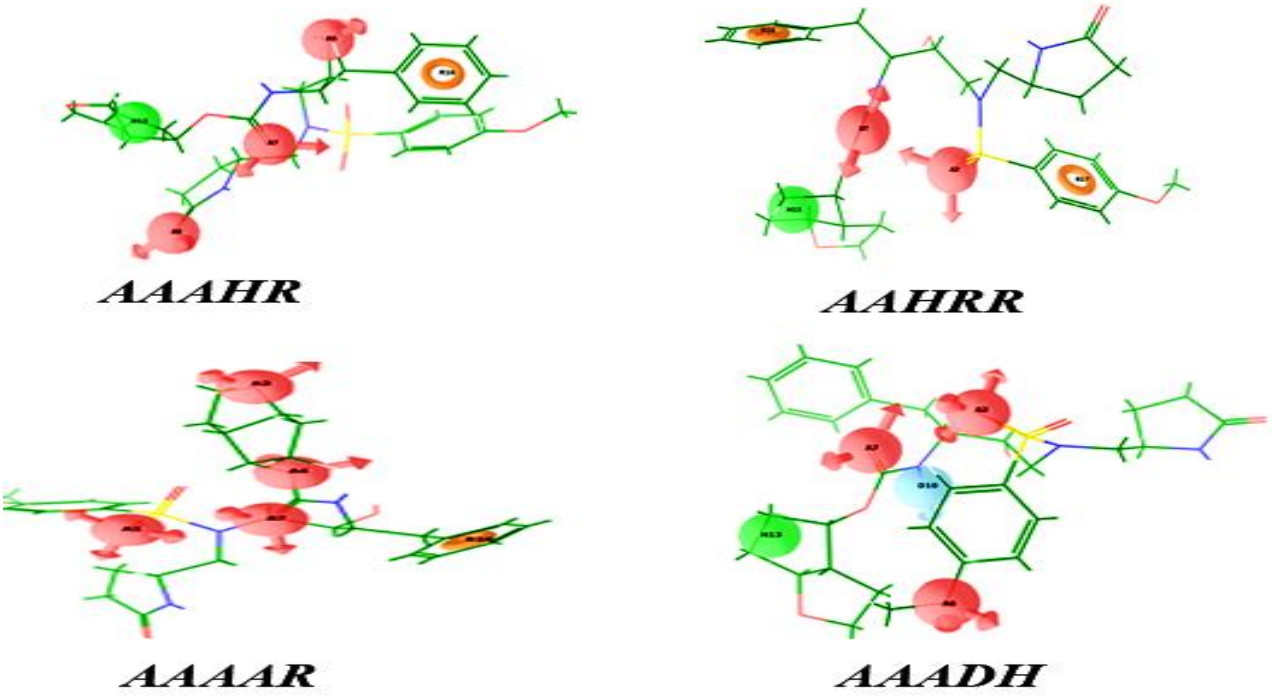


Figure 60: The finally selected top four pharmacophores from ligand based approach for HIV protease target

We did enrichment calculations for these pharmacophores, the hypothesis1 had EF 1% as 12, RIE 5.08, ROC 0.83 and EF 1.21 as shown in **Table 32**. The hypothesis 1 was selected as best hypothesis based on the survival active score of 2.583, survival inactive score of 1.560, vector score of 0.787, site score of 0.42 and volume score of 0.377 (**Table 31)** and it had good overall

calculated EF, GH, EF 1%, RIE, ROC and BEDROC [145] values compared to other three ligand based pharmacophores shown in **Table 33.** Though the hypothesis 1 (AAAHR) retrieved less number (750) of overall hits compared to other three hypotheses (1381, 1290, 1354) but the number actives retrieved was higher than other three hypotheses. Therefore the hypothesis 1 was selected as the best one as shown in **Figure 61.** The distance between the features were also shown in **Table 34**. Then we aligned all 134 compounds with the best pharmacophore hypothesis to build the QSAR model shown in **Figure 62**.

Table 32: Enrichment calculations for HIV protease target

S.No.	No. of Features	Features	Total Hits[a]	Actives[b]	EF[c] 1%	Active[d] 1%	%Ya[e]	% A[f]	F(-ve)[g]	F(+ve)[h]
1.	5	AAAHR	750	58	12	10.9	7.7	90.6	6	692
2.	5	AAHRR	1381	53	4.5	4.7	3.84	82.8	11	1328
3.	5	AAAAR	1290	48	11	10.9	3.72	75	16	1242
4.	5	AAADH	1354	53	11	10.9	3.91	82.8	11	1301

[a] Total number of hit molecules from the database

[b] Total number of active molecules in hit list

[c] Enrichment factor using formula

[d] Goodness of Fit using formula

[e] Yield of actives = [$(H_a/H_t) \times 100$]

[f] Ratio of actives = $[(H_a/A) \times 100]$

[g] False Negatives = $[A - H_a]$

[h] False Positives = $[H_t - H_a]$

Table 33: Selected pharmacophore with its calculated overall EF, goodness of fit (GH), % of actives, yield of actives, false negative and false positive values for HIV protease target

No of Features	Features	EF[a] 1%	EF[b]	GH[c]	RIE[d]	ROC[e]	BEDROC[f] α =20	BEDROC[f] α =160
5	AAAHR	12	1.21	0.074	5.08	0.83	0.437	0.629

[a] EF- Enrichment Factor at 1% of the decoy dataset

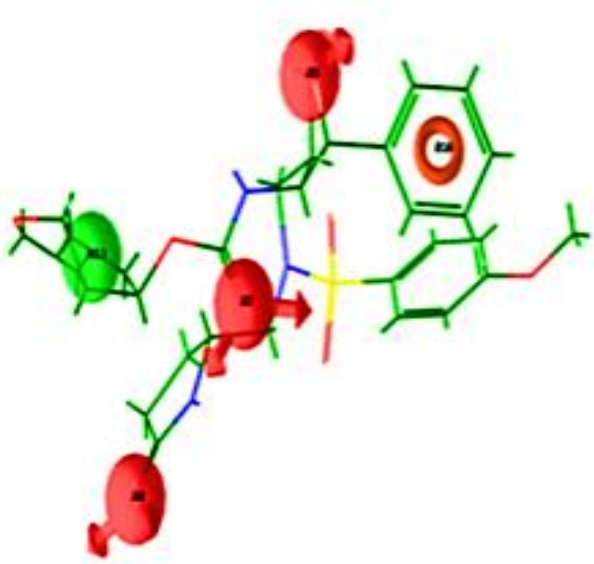
[b] Enrichment factor using formula

[c] Goodness of Fit using formula

[d] RIE- Robust Initial Enhancement

[e] ROC-Receiver Operating Characteristic curve value

[f] BEDROC- Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic



3D QSAR- AAAHR

Figure 61: The finally selected ligand based pharmacophore

Table 34: Distance between the features of 3D QSAR based pharmacophores for HIV protease target

Features*	**A1-H (Å)**	**H-A2 (Å)**	**A2-A3 (Å)**	**A3-R1 (Å)**	**R1-A1 (Å)**	**A3-D (Å)**	**D-A1 (Å)**	**A2-R2 (Å)**	**A1-A2 (Å)**	**A3-A4 (Å)**	**A4-R1 (Å)**	**R2-R1 (Å)**
AAAHR	8.123	6.548	5.359	4.632	3.425	NA	NA	NA	7.121	NA	NA	NA
AAHRR	3.540	5.240	NA	NA	3.764	NA	NA	2.617	3.125	NA	NA	8.176
AAAAR	NA	NA	4.162	6.725	8.107	NA	NA	NA	7.630	2.283	5.224	NA
AAADH	3.428	6.218	6.592	NA	NA	4.617	3.193	NA	3.011	NA	NA	NA

Distance shown in Å, *A-Acceptor, D-Donor, R-Ring aromatic and H-Hydrophobic. NA- non applicable

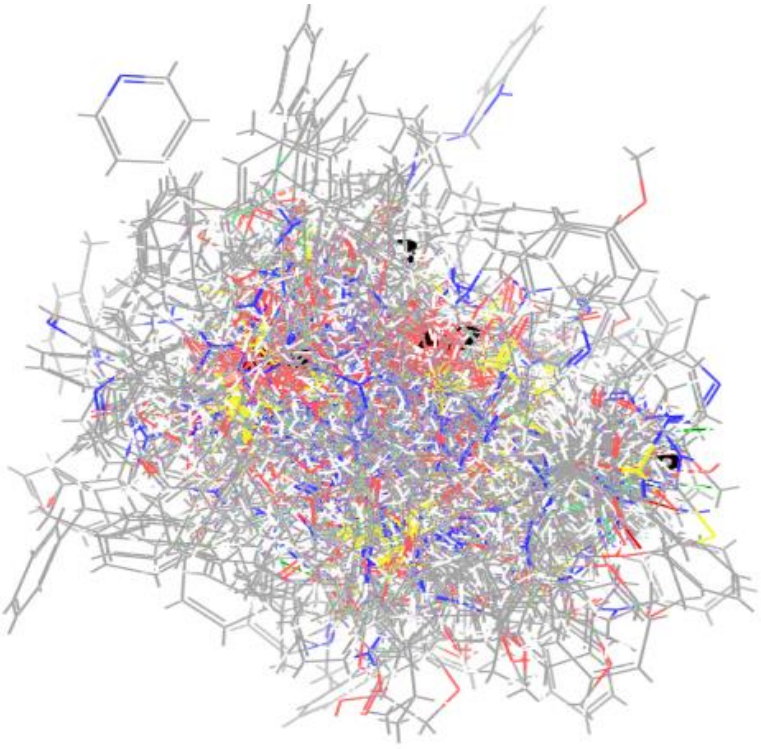


Figure 62: 134 selected compounds to build 3D QSAR were aligned with the best pharmacophore AAAHR for HIV protease target

6.3.2. 3D QSAR modelling and PLS analysis

The ligands used were of diverse structures and hence pharmacophore-based QSAR models were generated for AAAHR hypothesis using the 103-member training set and a grid spacing of 1.00 Å. For each run we kept on changing the training set and test set compounds to get good Q^2 and R^2 values. At 553 run we had 103 compounds in training set and 31 compounds in test set, QSAR models containing one to five PLS factors were generated. A model with PLS factor five was considered as the best statistical model and thorough analysis was done and their results are shown in **Table 35**. This model was validated by predicting activities of test set, training set and over all molecules. The best 3D QSAR model should possess good internal and external statistical validation, which can display more reliable predictions. Important parameters obtained based on LOO method, (**Table 35**) favored the internal statistical validation by PLS analysis.

Table 35: The PLS analysis results for all three selected 3D QSAR based pharmacophores for HIV protease target

Statistical parameters	**AAAHR**	**AAHRR**	**AAAAR**	**AAADH**
Number of molecules in Training set	103	103	103	103
Number of molecules in Test set	31	31	31	31
R^2	0.997	0.995	0.993	0.99
Q^2	0.865	0.008	-0.199	0.271
SD	0.103	0.129	0.154	0.186
F-value	6044	4159	2692	2002
RMSE	0.568	1.471	1.715	1.319
r^2_{pred}	0.8651	0.1005	0.0091	0.1156

PLS statistic parameters:

SD - Standard deviation of the regression.

R^2 - for the regression.

F - variance ratio.

r^2_{pred} - predictive correlation coefficient value.

RMSE - root mean square error.

Q - squared (Q^2)value of Q^2 for the predicted activities.

Pearson R -correlation between the predicted and observed activity for the test set.

Among these four models, hypothesis 1 (AAAHR) showed good external predictive ability as compared to others. Hypothesis 1 showed a good R^2 value for the training set of 0.997, good predictive power with Q^2 of 0.865 for the test sets, with SD of 0.103, and F value of 6044. Further the external validation of the model was predicted by r^2_{pred} for test set with the value of 0.865 (**Table 35**). The accepted LOO-cross validated value of training set (R^2) should be greater than 0.5 and **Table 36** LOO cross validated value for test set (Q^2) should show a value greater than 0.5 to attain good predictive capacity, and standard deviation (SD) below 0.3, with minimum root mean square error (RMSE), and high value of variance ratio (F) to provide conventional QSAR validation limits. And the predictive correlation coefficient (r^2_{pred}) value generated based on HIV-PR molecules of test set demonstrated real predictive capacity and robustness of the QSAR model. For a QSAR model, internal validation of LOO cross validated Q^2 was commonly used to assess predictive ability, where a high value of Q^2 is necessary and important but Q^2 alone is not sufficient condition for a model to have a high predictive power [159, 160]. A reliable model should also be characterized by a high correlation coefficient R (or r^2) between the predicted and observed activities of compounds from an external test set. In the present study the best predictive ability of the model was characterized by correlation coefficient $R = 0.952$ ($r^2 = 0.907$).

Table 36: External statistical validation results of quantitative structure activity relationship (QSAR) result for the hypothesis 1 (AAAHR) pharmacophore hypotheses for HIV protease target

External validation	**Parameters calculated**	**Limitations**
r^2_{cv}	0.865	r^2_{cv}> 0.5
R	0.952	Must close to 1
r^2	0.907	r^2> 0.5
k value	1.001	0.85 ≤ k ≤ 1.15
k′value	0.991	0.85 ≤ k′≤ 1.15
R^2_0	0.999	R^2_0 or R'^2_0close to r^2
R'^2_0	0.998	R^2_0 or R'^2_0close to r^2
$[(r^2 - R^2_0) / r^2]$	-0.101	$[(r^2 - R^2_0 / r^2]$ <0.1
$[(r^2 - R'^2_0) / r^2]$	-0.100	$[(r^2 - R'^2_0 / r^2]$ <0.1
r^2_m (LOO)	0.905	r^2_m(LOO)> 0.5
r^2_m(all)	0.953	$r^2_m(all)$> 0.5
r^2_m(test)	0.823	$r^2_m(test)$> 0.5

r^2_{cv} - cross validated coefficient
R (or r^2) - correlation coefficient between the actual and predicted activities
k and k′ - slope values of regression lines
R^2_0 and R'^2_0 - correlation coefficients for the regression lines through the origin
$[(r^2 - R^2_0) / r^2]$ and
$[(r^2 - R'^2_0) / r^2]$ - to calculate relation between r^2, R^2_0 and R'^2_0
r^2_m (LOO) - modified squared correlation coefficient for Leave One Out method.
r^2_m(all) - modified squared correlation coefficient for all test set and training set
r^2_m(test) - modified squared correlation coefficient for all test set

High slope of regression lines through the origin k value of 1.003 and k′ value of 0.991 (either k or k′ should be close to 1) gave substantial values of R_0^2 value 0.999 and the R'^2_0 value 0.998, which were obtained by calculating correlation coefficient of regression lines of the scatter plot obtained by means of actual activity versus predicted activity and predicted activity versus actual activity plots respectively (**Figure 63**). The relation between r^2, R_0^2 and R'^2_0 gave (r2 $-R_0^2$ /r2) values of -0.101 and second relation (r2 $-R'^2_0$ /r2) value -0.100 showed optimum values within the statistical limits (**Table 36**). Yet, our established QSAR model from hypothesis 1 (finalized after PLS analysis), gave r^2_{cv} value of 0.865. A parameter of modified r^2 [r^2_m (LOO)] [65] was considered as a better external predictive potential for the whole set of compounds which was of 0.8242 (>0.5) defined through scatter plot best fit line values.

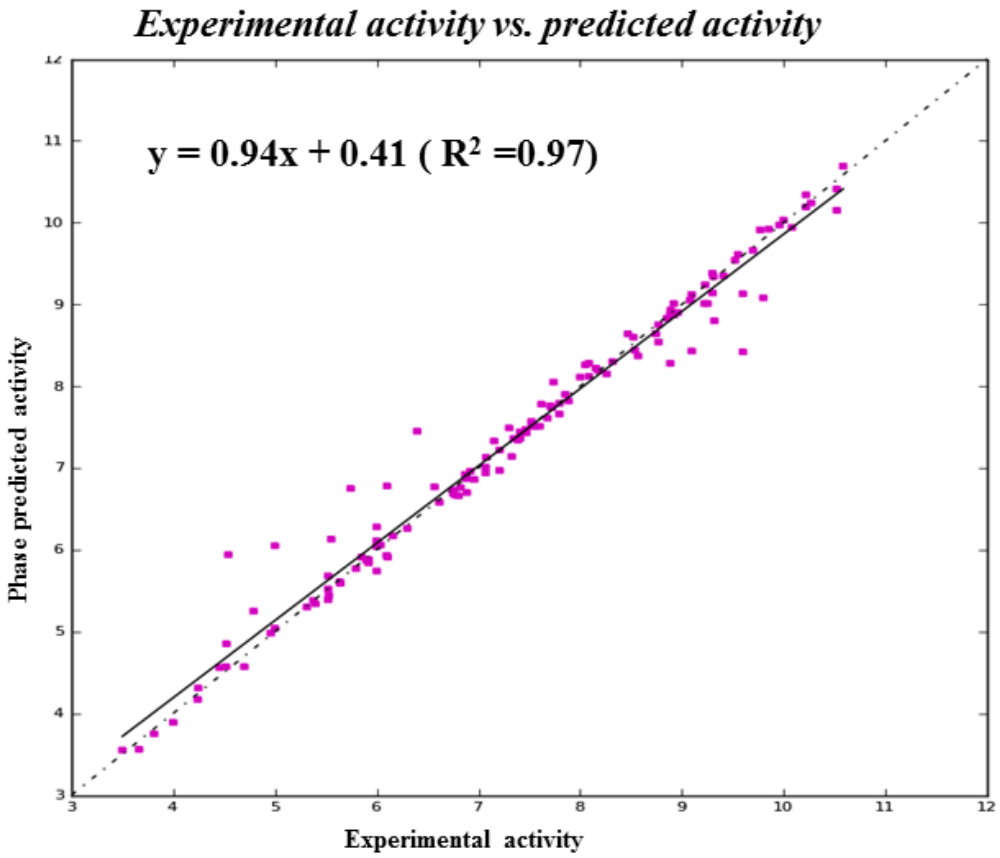


Figure 63: Best fit line for HIV Protease target

This showed to be truly predictive by fulfilling the requirements of every parameter in the external validation. We considered this model as statistically significant model. Besides, we resumed further steps to predict the activities of new leads from the compound libraries by using hypothesis 1.

6.3.3. Contour maps

The best pharmacophore hypothesis 1 selected from 3D QSAR was employed to generate the contour maps. These were useful to identify the important sites for substitutions or replacements of groups to increase the biological activity of the compound. The counter maps generated from the most active compound (**1**) and the least active compound (**134**) were used to analyze the important positions which enhanced the inhibitory activity and could be useful in designing novel scaffolds. The contour maps are given in **Figure 64.** In **Figure 64A** and **B,** the H-bond donor effect was seen with the most active (**1**) and least active (**134)** compounds, In **Figure 64C** and **D,** the hydrophobic effect and in **Figure 64E** and **F,** the electron withdrawing effect was displayed.

The hydrogen bond donor nature for the most active compound **1** and the least active compound **134** when compared showed their most favorable region blue color and unfavorable regions red color. In the pharmacophore model, the H-bond donor feature was absent. We discovered from the contour maps the H-bond donor effect in **Figures 64A** and **B,** the presence H-bond donor positive region present (blue colour) in most active compound **1** between the two H-bond acceptor features (A-A) and between H-bond acceptor and hydrophobic feature (A-H) enhance the activity and this was because of the presence of the nitrogen group which was absent in these region in the least active compound **134**. Pyrrolidinyl, benzamide substituted valinamide had more than two nitrogen atom in between the acceptor-acceptor and acceptor- hydrophobic

features and showed high inhibitory activity against HIV protease. Therefore it can be inferred that the presence of H-bond donor group was very much needed for the activity.

In **Figure 64C** and **D,** the hydrophobic nature of the most active compound **1** and the least active compound **134** revealed that the favored green colored regions were more around the aromatic rings of pyrrolidinyl, benzamide, furanyl, pyrrolinone, benzocycloalkyl amine, pyranone, substituted valinamide, pyranyl derivatives showed that terminal aromatic rings were very much needed for the activity of the compounds. But the least active compound **134** had the unfavored yellow region on the extended aliphatic chain which hindered the activity.

In **Figure 64E** and **F,** the favored red color regions were exactly on the acceptor features of the pharmacophore. This enhances the activity. When we compared the position of the acceptor features in both most active compound **1** and the least active compound **134 ,** the most active compound **1** had acceptor features in the main core of the compound but in the case of the least active compound **134** were in terminal position. This showed that the presence of electron with drawing group in the core of the compound enhances the activity.

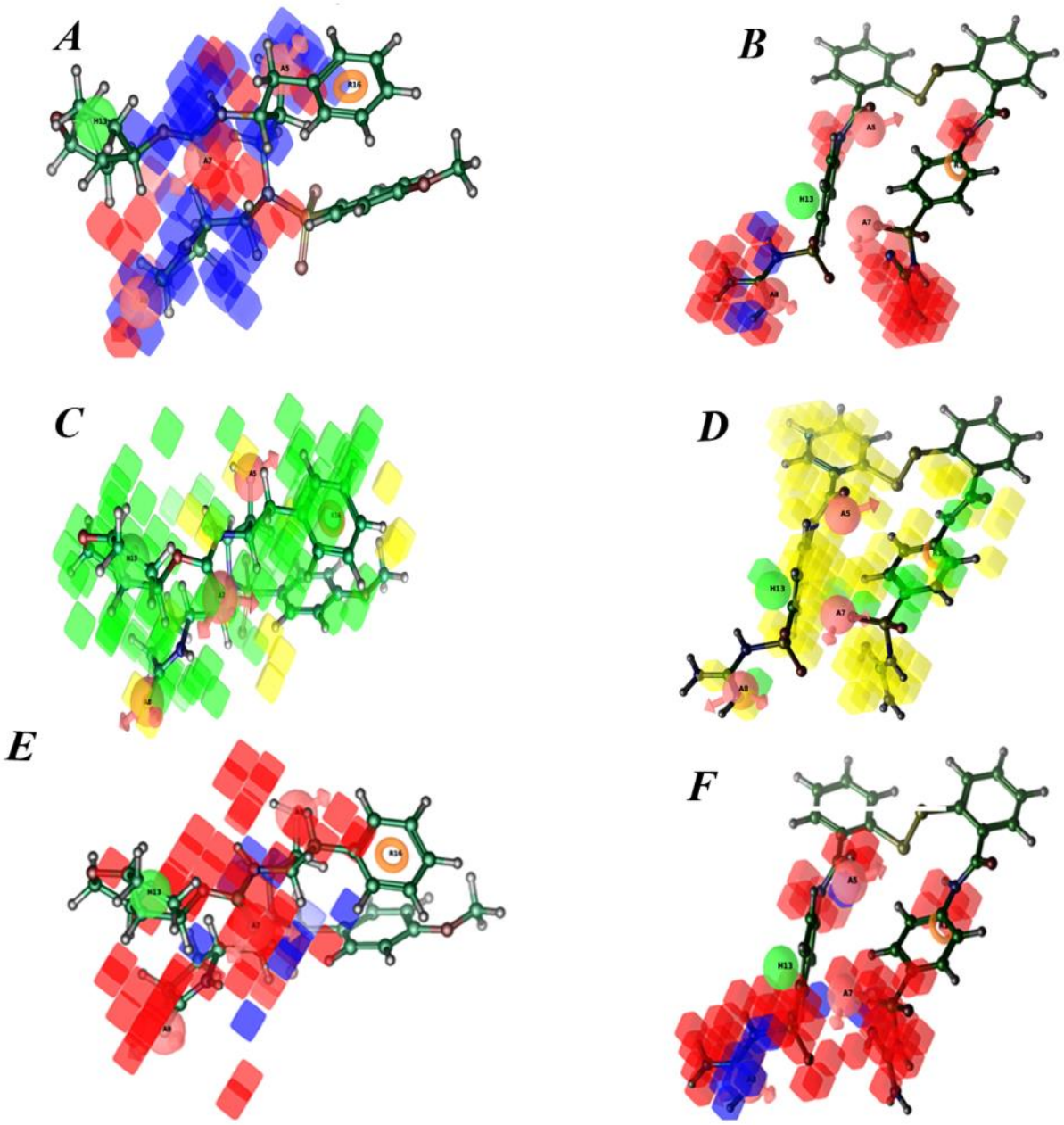


Figure 64: Contour maps for HIV protease target

A- H-bond donor effect: Most active; **B**- Least active (Blue- favorable, Red unfavorable); **C**- Hydrophobic effect: Most active; **D**- Least active (Green- favorable, Yellow-unfavorable); **E**- Electon with-drawing effect: Most active; **F**- Least active (Red-favorable, Blue -unfavorable).

6.4. Multiple pharmacophore models based Virtual screening and docking

Virtual screening of commercial data bases was the preliminary step in lead identification. We employed both approaches like structure based and ligand based to design new lead compounds for this target to identify small molecule inhibitors against HIV protease. In this study, we had 13 pharmacophores (12 e-pharmacophores and 1 3D QSAR pharmacophore). We followed similar protocol as that employed for HCV NS5B inhibitors wherein first we did pharmacophore screening of commercial database (Asinex Database). We had 5, 4 and 3 featured pharmacophores, so we kept the must match on at least 5, 4 and 3 site points and for ligand based pharmacophore we kept must match for at least 3 site points because the structure based approach had many possible combinations of features but in case of ligand based we had only one five featured pharmacophore. If we employed only the must match five sites, we would have missed few potent leads for any of three site points. The hits derived at each stage were validated based on their fitness score which indicated how well they fitted with the pharmacophore features. The compounds with high fitnesss scores were shortlisted as phase find matches hits and are shown in **Table 37.**

From the phase findmatches module output of the e-pharmacophore AAADR for 3OXC, we selected fitness more than 1.00 as limit for the HTVS and got 45630 compounds and with 2672 clusters. The bound crystal ligand was isoquinoline derivative and its docking score was -11.980 kcal/mole. These ligands were docked to the HIV protease active site of 3OXC using the grid which was already generated, and finally we selected 4536 ligand molecules as hits from HTVS with the criteria of docking score above -5 kcal/mole, fitness above 1.10 and the number of H-bond above 2. We got 443 diversed clusters. These hits were further docked using Glide SP docking module and 4006 ligand molecules were selected based on the docking score above -6

kcal/mole, fitness above 1.20, 2 H-bond and visual occupancy of ligand into the pocket. Finally, we subjected the Glide SP 4006 filtered ligands to Glide XP (extra precision) docking simulation. Top 1991 ligand molecules more than a docking score of -7.00 kcal/mole were visually inspected for the pose and important binding residues. These belonged to 175 diverse structural scaffolds, matching all the five e-pharmacophoric features and were selected. Though the docking score of the crystal ligand was very high but we selected compounds with docking score above -7.00 kcal/mole compounds because of their good fitness and number of H-bonds.

Table 37: The number of compounds (Hits) retrieved from each stage of screening results

PDB	Phase Find [a]	clusters	HTVS[b]	clusters	SP[c]	clusters	XP[d]	cluster	Peptidic	Non-peptidic
3OXC-5	45630	2672	4536	443	4006	401	1991	175	1864	127
3OXC-4	85000	6521	4688	515	2752	269	2306	258	2119	187
3OXC-3	166818	8135	4560	445	2900	423	2874	129	2676	198
2Q5K-3	186900	9045	8900	904	3870	625	793	217	649	144
2Q5K-3	126900	11586	4841	616	2143	349	847	187	693	154
3NDX-4	9900	952	1625	344	1271	129	774	109	693	81
3NDX-3	12035	1452	3605	597	2214	182	1246	163	1108	138
3H5B-3	147800	7870	4509	675	3834	577	1279	279	1111	168
3I6O-3	127800	8352	8500	1352	1927	171	1611	53	1572	39
3DJK-3	157800	8542	4502	466	1475	271	971	197	807	164
3T11-3	52421	1356	4836	360	3019	249	1541	181	1400	141
2AID-3	125569	9014	8932	510	3998	389	1998	287	1829	169
3DQSAR-5 (3DJK)	132545	3542	4921	569	1171	297	727	103	659	68

[a]Total number of hits from Phase findmatches.
[b]Total number of hits from HTVS docking.
[c]Total number of hits from SP docking.
[d]Total number of hits from XP docking

For the e-pharmacophore AADR from 3OXC, we got 85000 compounds as phase find matches output with fitness score above 1.5 and belonged to 6521 clusters. These hits were docked into the active site pocket of HIV protease of 3OXC by HTVS that yielded 4688 molecules of 515 clusters with docking score more than -5.00 kcal/mole, 2 H-bond and fitness above 1.6. We selected 2752 leads with 269 clusters with SP docking score above -5.00 kcal/mole, 2 H-bond and fitness above 1.6. Finally in XP docking we got 2306 with 258 clusters with XP docking score above -6.00 kcal/mole and above 1.7 fitness score. When we further grouped them into peptidic and non-peptidic we got 2119 peptidics and 187 non-peptidic leads.

Similar protocol for AAR from 3OXC resulted in 2874 (129 clusters) as hits with docking sore above -6.00 kcal/mole, 2 H-bonds and fitness above 2. From these XP hits, we got 2676 peptidic and 198 non-peptidic.

There were two e-pharmacophores obtained from 2Q5K and for the e-pharmacophore A3R15R16 186900 (9045 clusters) compounds were retrieved from phase find matches screening having fitness score above 1.6. The bound crystal ligand was pyrimidine derivative and its score was -6.25 kcal/mole kept the docking score of the crystal ligand as the cut-off value to select the hits. When we performed HTVS for these compounds in the active site of 2Q5K, we got 8900 leads (904 clusters) with docking score above -5.00 kcal/mole, 2 H-bonds and above 1.73 fitness. These were then subjected to SP docking and further XP docking to yield 793 leads of 217 clusters with docking score above -6.00 kcal/mole, 2 H-bond and fitness above 1.73 which belonged to 649 peptidic and 144 nonpeptidic.

Similarly for A3R16R17 e-pharmacophore from 2Q5K, it resulted in 154 non-peptidic and 649 peptidic from the XP hit file of 793 compounds of 217 clusters.

Utilizing 3NDX crystal ligand, two e-pharmacophore were obtained and for the e-pharmacophore AARR for, we got 9900 leads with pharmacophore filter and when subjected to The bound inhibitor was a thiazole derivative with docking score -5.334 kcal/mole. HTVS in the active site of 3NDX, we got 1625 leads (344 clusters) with docking score above -5.00 kcal/mol, 2 H-bonds and above 1.4 fitness score. We got 1271 compounds (129 clusters) from SP docking with score above -5.00 kcal/mol, 2 H-bonds and above 1.4 fitness. Finally 81 non-peptidic and 693 peptidic compounds were obtained from XP docking.

For the e-pharmacophore ARR from 3NDX, we finally obtained 1108 peptidic compounds and 138 non-peptidic compounds from XP docking output which about of 1246 hits (163 clusters).

For the e-pharmacophore DRR for 3H5B, the crystal protein had pyyrolidine derivative as its bound crystal ligand and its docking score was -6.140 kcal/mole. we got 147800 leads (7870 clusters) from pharmacophore screening with above 1.5 and were subjected to HTVS resulted in 4509 leads (675 clusters) of docking score above -5.00 kcal/mol, 2 H-bonds and above 1.5 fitness. 3834 (577 clusters) leads yielded from SP docking with docking score above -6.00 kcal/mol, 2 H-bonds and above 1.5 fitness. 168 non-peptidic and 1111 peptidic compounds from the final XP docking of 1279 compounds (279 clusters).

For the e-pharmacophore ARR from 3I6O, we obtained 127800 compounds of 8352 clusters having fitness above 1.8. 8500 compounds (1352 clusters) were obtained from HTVS output having docking score above -5.00 kcal/mol, 2 H-bonds and above 1.8 fitness. It had a furan derivative as bound ligand with the docking score of -6.557 kcal/mole. Then we performed SP docking, which yielded 1927 (171 clusters) leads with docking score above -7.00 kcal/mol, 2 H-bonds and above 1.8 fitness. From this e-pharmacophore we got only 39 non-peptidic and 1572 peptidic from XP docking of 1611 (53 clusters).

Similarly for the e-pharmacophore ARR from 3DJK, we obtained 157800 leads of 8542 clusters having fitness above 1.96. . The crystal protein had dioxepan derivative as its bound crystal ligand and its docking score was -6.854 kcal/mole. 4502 compounds emerged (466 clusters) from HTVS output having docking score above -5.00 kcal/mol, 2 H-bonds and above 1.99 fitnessThen in SP docking, we obtained 1475 (271 clusters) leads with docking score above -6.00 kcal/mol, 2 H-bonds and above 1.99 fitness. Finally only 164 non-peptidic and 807 peptidic from XP hits of 971 (197 clusters).

E-pharmacophore DRR from 3T11, the crystal protein had diazepin derivative as its bound crystal ligand and its docking score was -6.430 kcal/mole .yielded 52421 compounds (1356 clusters) from pharmacophore screening with fitness above 1.4 and then HTVS docking retrieved 4836 (360 clusters) compounds having docking score above -5.00 kcal/mol, 2 H-bonds and above 1.4 fitness SP hits obtained with docking score above -6.00 kcal/mol, 2 H-bond and above 1.4 fitness were 3019. Finally we retrieved 141 non-peptidic and 1400 peptidic compounds from the XP docking of 1541 (181 clusters) compounds.

For the e-pharmacophore ADR from 2AID, we retreived 125569 compounds (9014 clusters) pharmacophore filtering with fitness above 1.96 and then HTVS docking yielded 8932 (510 clusters) compounds having docking score above -5.00 kcal/mol, 2 H-bonds and above 1.99 fitness. The crystal ligand was simple piperidine derivative and its docking score was -5.130. Therefore there were more possibility of retrieving non-peptidic compounds and kept the docking score of the crystal ligand as the cut-off value. There were 3998 (389 clusters) SP hits with docking score above -6.00 kcal/mol, 2 H-bond and above 1.99 fitness. Finally we retrieved 169 non-peptidic and 1829 peptidic compounds from the XP docking of 1998 (287 clusters) compounds.

The ligand based pharmacophore AAAHR, when subjected to pharmacophore filtering and virtual screening with 3DJK with low resolution. We retrieved 4921 (569 clusters) compounds as HTVS hits with docking score above -5.00 kcal/mol, 2 H-bonds and above 1.4 fitness. Further SP docking yielded 1171 (297 clusters) compounds as SP hits with docking score above -6.00 kcal/mol, 2 H-bonds and above 1.4 fitness. Finally we retrieved 68 non-peptidic and 659 peptidic compounds from the final XP docking hits of 727 compounds (103 clusters).

Out of three e-pharmacophores AAADR, AADR and AAR from 3OXC, the three feature pharmacophore AAR retrieved more number of non-peptidic leads. When we compared two e-pharmacophores A3R15R16 and A3R16R17 from 2Q5K, the e-pharmacophore A3R15R16 was able to retrieve more number of diverse compounds in each stage of screening than A3R16R17 and the distance between the ring (R) features R15 and R16 could have played very crucial role. But e-pharmacophore A3R16R17 retrieved more number of non-peptidic hits than A3R15R16. Though the features A3 and R15 were found to aligne when mapped R16 and R17 were in opposite direction. This shows that these 2 pharmacophores were completely different from each other as shown in **Figure 65.**

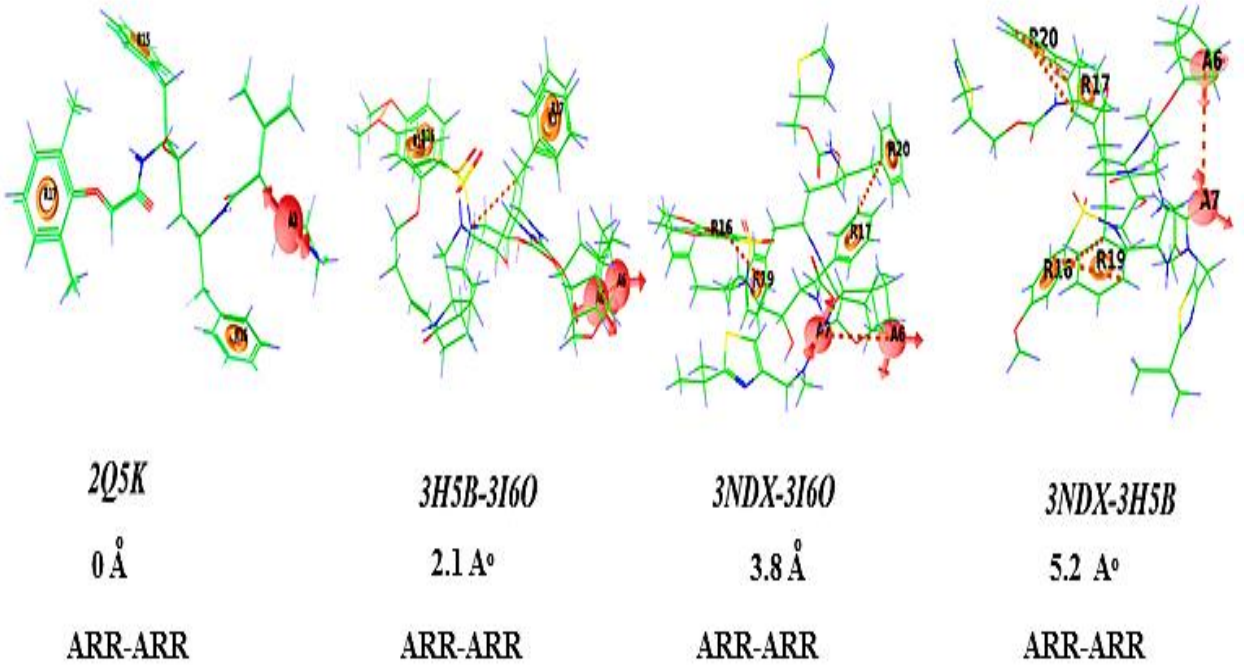


Figure 65: Comparison of pharmacophores with similar features for HIV protease target

Similar mapping of other pharmacophores like 3H5B to 3I6O, 3NDX to 3I6O and 3NDX to 3H5B with common features ARR were found not exactly aligning with each other but showed rmsd deviation from 0 Å to 5.2 Å. This shows that these pharmacophore though they have the same features but they were unique in nature because of their distance between the features and their deviation as shown in **Figure 65**. When we compared the compounds retrieved by both AARR and ARR from 3NDX, the three feature e-pharmacophore ARR (3NDX) retrieved more number of diverse compounds in each stage of screening process compared to AARR (3NDX) and also retrieved the most number of non-peptidic leads. When we compared all the three feature pharmacophores, AAR (3OXC), A3R15R16 (2Q5K), A3R16R17 (2Q5K), ARR (3NDX), DRR (3H5B), ARR (3I6O), ARR (3DJK), DRR (3T11) and ADR (2AID), the e-pharmacophore AAR (3OXC) retrieved the highest number (198) of non-peptidic compounds among all the 13 pharmacophores. E-pharmacophore ARR (3I6O) was the one which retrieved very less number

(36) of non-peptidic among all the 13 pharmacophores. The e-pharmacophores A3R15R16 (2Q5K) had 144, A3R16R17 (2Q5K) had 154, ARR (3NDX) had 138, DRR (3H5B) had 168, ARR (3DJK) had 164, DRR (3T11) had 141 and ADR (2AID) had 169 non-peptidic compounds. When we compared the five feature pharmacophores from structure based AAADR (3OXC) and ligand based AAAHR (3D QSAR), the e-pharmacophore AAADR (3OXC) retrieved more number of non-peptidic compounds. Among the four feature pharmacophores AADR (3OXC) and AARR (3NDX), the e-pharmacophore AADR (3OXC) retrieved more number (187) of non-peptidic compounds. When we compared all the 13 pharmacophores, the pharmacophore from 3NDX retrieved very less number of compounds from each stage of screening (**Table 37**). 3T11 and 2AID were bound to non-peptidic inhibitors and other had bound peptidic inhibitors and when compared with the screening data, surprisingly 3T11 and 2AID retrieved comparably good number of non-peptidic inhibitors.

We cross checked all our final XP hits with GOLD docking program [270]. We performed simultaneously two runs. We selected the compounds which had good scores. Finally we selected 13 compounds based on their docking score, fitness score, predicted activity, gold score, visual inspection and their interaction patterns. Most of the compounds were found to have more than two H-bonds and all were with reported H-bonds interactions which the crystal ligands showed. Predicted activities, Glide scores, fitness, H-bond data are presented in the **Table 38** and **Figure 66.**

Table 38: Lead compounds with their respective number of H-bonds, fitness score and docking score, predicted activity, interaction with important amino acids and their respective gold scores for HIV protease target

Title	Fitness	Docking score	H Bond	Predicted Activity	Ligand Interaction	Gold score 1	Gold score 2
L-1	1.95	-10.53	5	6.38	Asp29, Asp30, Asp125, Ile50, Ile150	52	53
L-2	1.49	-10.20	4	6.84	Asp25, Gly27, Asp125 (2)	49	50
L-3	1.51	-9.42	3	6.29	Asp25, Asp125 (2)	43	47
L-4	1.41	-9.28	3	6.55	Asp25, Asp30, Pro79	51	52
L-5	1.96	-9.15	2	7.33	Asp30, Gly48	44	47
L-6	1.38	-8.86	5	6.23	Asp25, Gly27, Asp129, Asp130 (2)	49	53
L-7	1.90	-8.79	5	6.27	Arg8 (2), Ile50, Asp129, Ile150	57	58
L-8	2.05	-7.93	6	6.52	Arg8 (2), Asp25 (4)	46	49
L-9	1.56	-7.93	4	5.95	Asp25, Asp30, Ile50 (2)	48	48
L-10	1.58	-7.49	2	7.50	Asp30 (2)	46	47
L-11	2.13	-7.44	3	5.73	Asp125, Asp129, Ile150	54	56
L-12	1.52	-7.37	4	6.64	Asp29, Asp30, Ile50, Ile150	50	53
L-13	1.45	-6.81	4	6.35	Arg8, Asp25, Asp125, Gly127	49	52

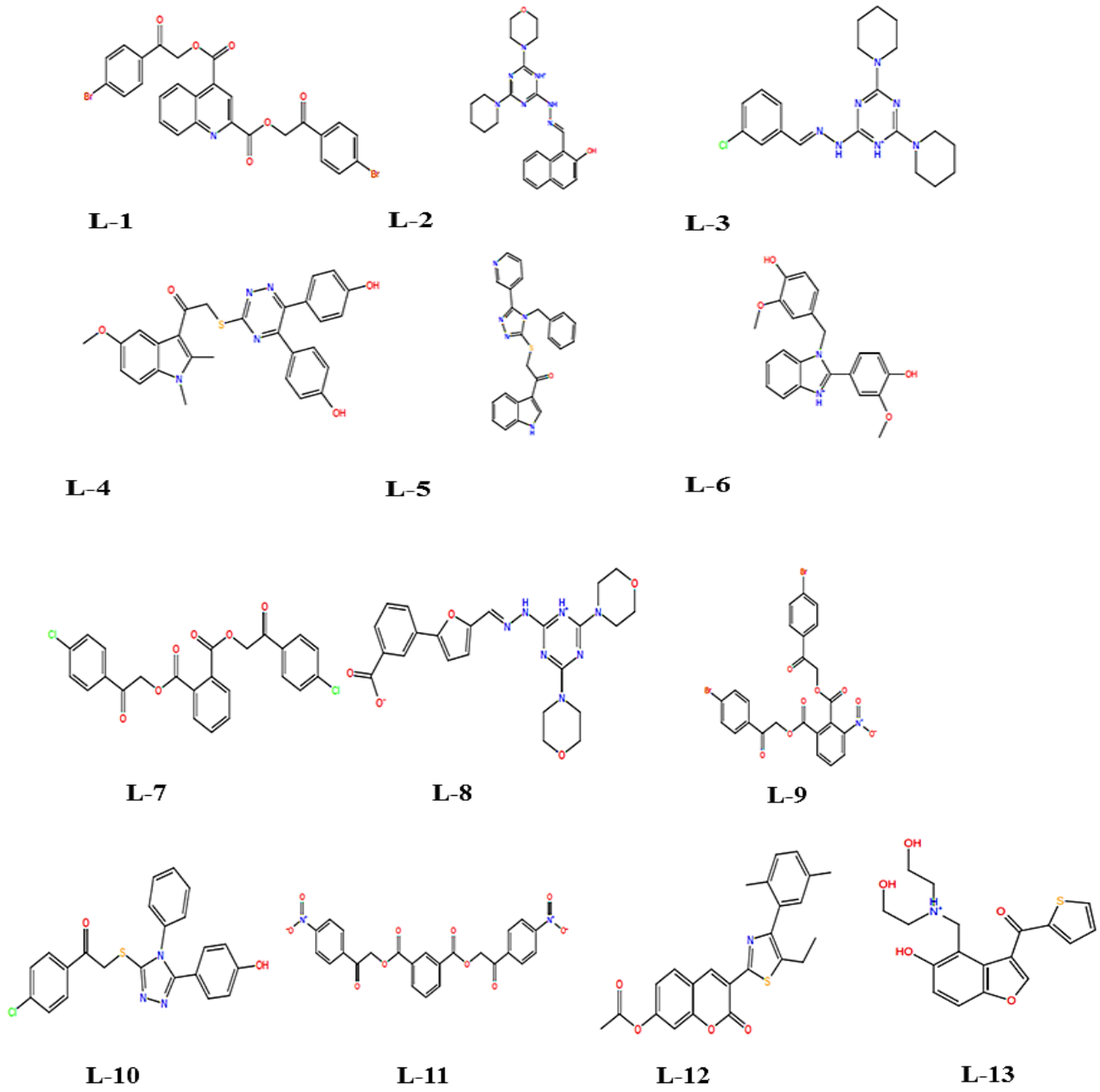


Figure 66: The final 13 hit compounds selected from the multiple pharmacophores screening and different stages of docking studies for HIV protease target

In each stage of screening we were interested in diverse structure search and analyzed the compound retrieved at each stage using pie graph as shown in **Figure 67**. In each stage of screening like pharmacophore screening, HTVS, SP and XP we compared the hits with any two pharmacophores and tried to identify the common hits. The peptidic and non-peptidic identification was also carried out for comparison. Top 1% hits obtained from each pharmacophore were aligned with their respective pharmacophores as shown in **Figure 68**.

The shortlisted final scaffolds as hits contained diverse scaffolds like quinolone dicarboxylate, piperidine, triazine morpholine, indole, benzimidazole, phthalate, thiazole, benzofuran derivatives. There were 11 hits having predicted pIC_{50} values more than 6 and remaining **L-9** and **L-11** having 5.95 and 5.73 respectively. Most of the hits having interaction with the imported aminoacids which are responsible for the activity of the crystal ligands. Interaction patterns revealed that all 13 hits showed more than two important H-bond interactions which included Asp25, Asp29, Asp30, Asp125, Gly27, Gly48 and Gly127. Out of the 13 top selected hits seven were having H-bond interaction with Asp25 which was shown by the non-peptidic crystal ligand 2AID. This showed that the non-peptidic compounds would be potent inhibitor for HIV protease. The interaction patterns of all the top 13 hits are presented in **Figure 69.**

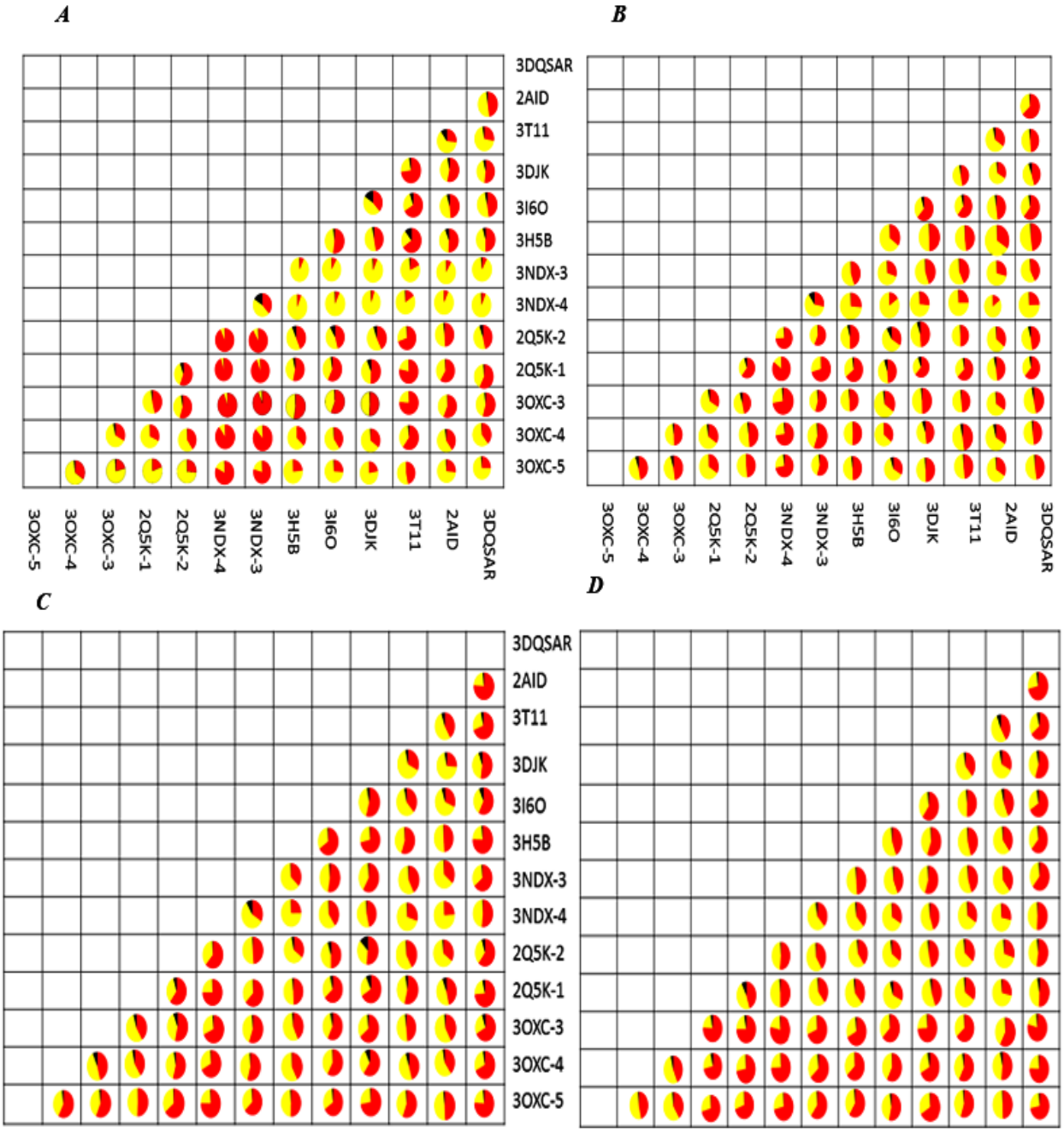


Figure 67: Pie graph shows the extent of overlapping of compounds at each stage of screening for all thirteen pharmacophore models. A- Phase find match hits, B- HTVS docking hits, C- SP docking hits, D-XP docking hits. Red and yellow colored regions represent the number of screened hit compounds from any two pharmacophore models and the black color region represents the overall overlap of the screened compounds for HIV protease

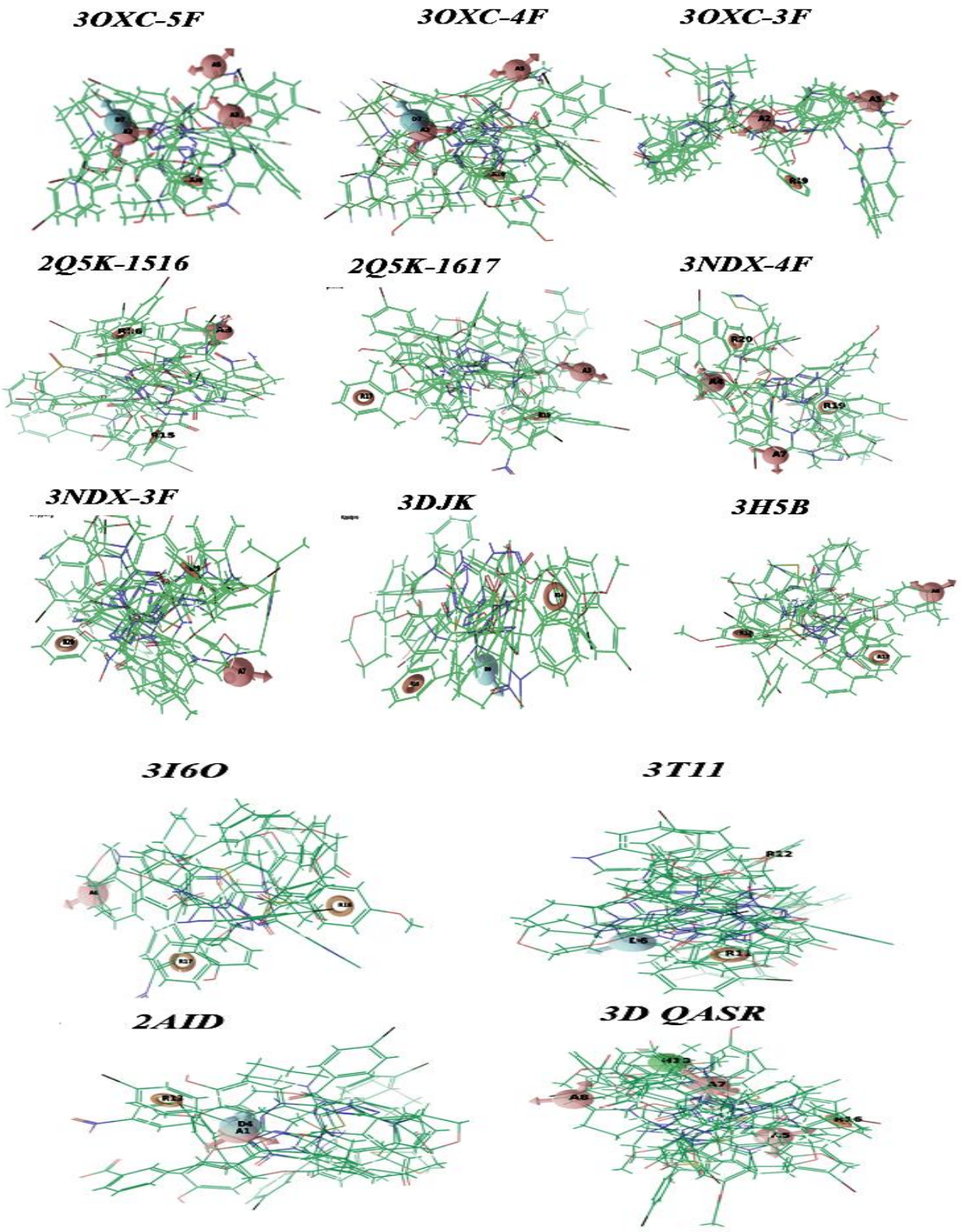
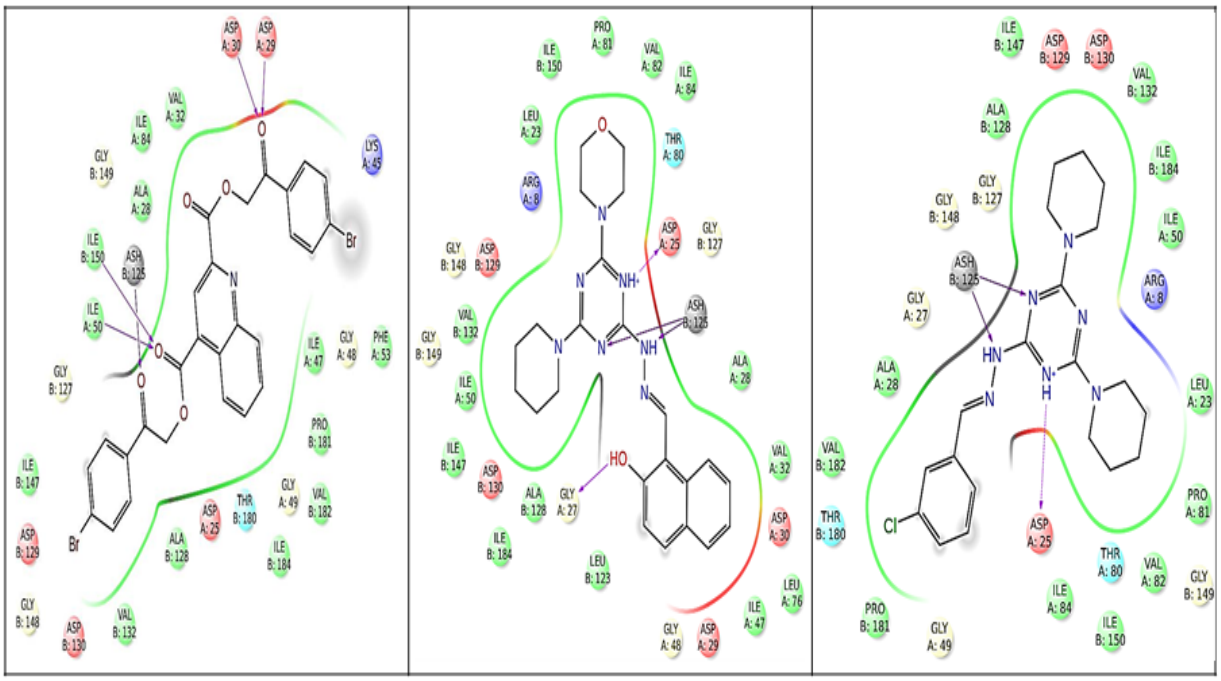
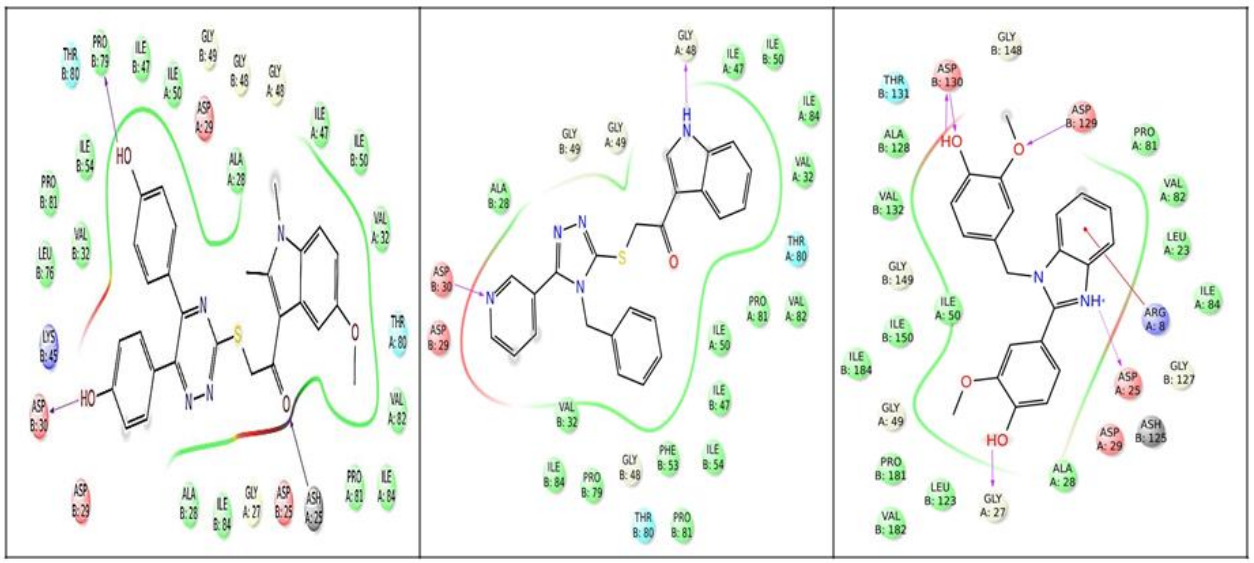


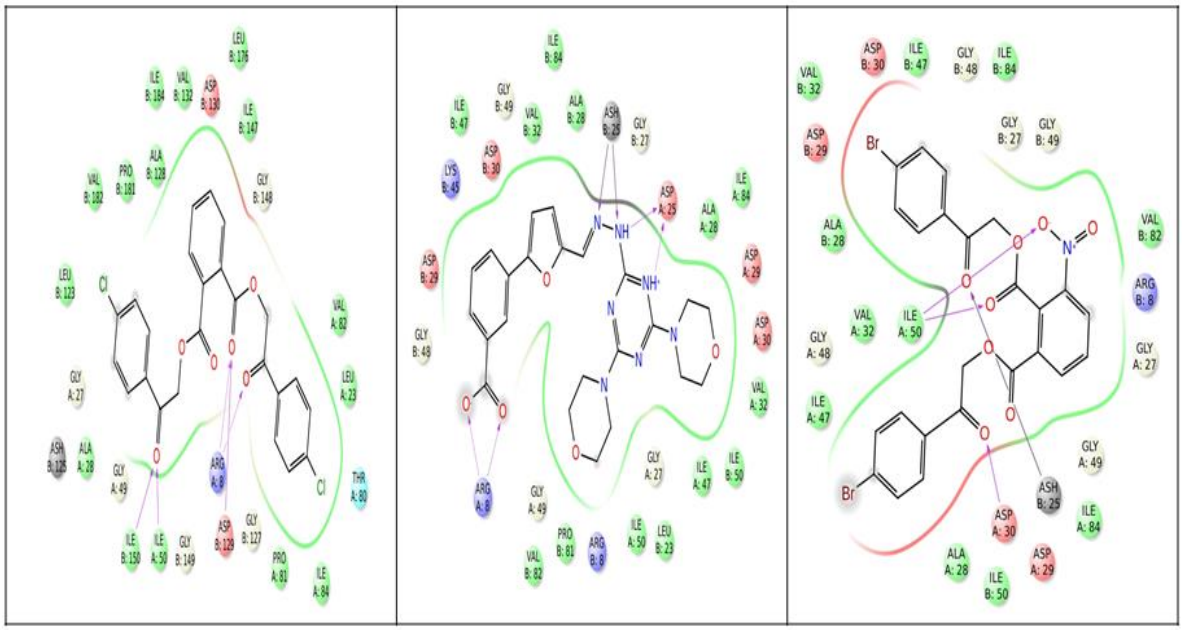
Figure 68: Top 1% hits aligned with their pharmacophore with their respective PDB codes for HIV protease target



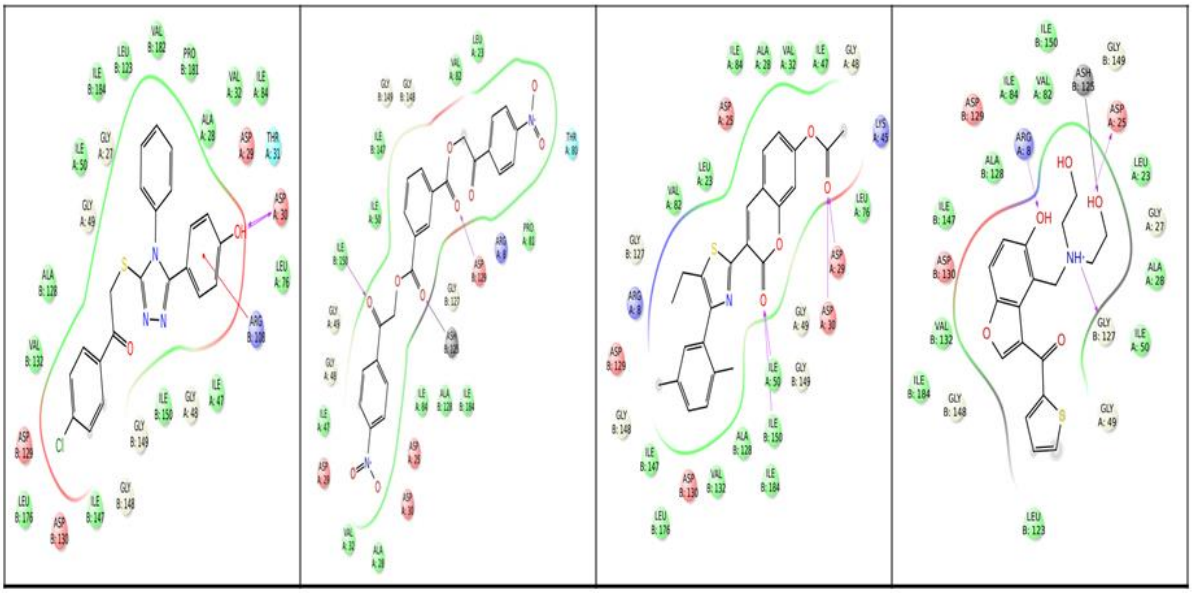
L-1 **L-2** **L-3**



L-4 **L-5** **L-6**



L-7 L-8 L-9



L-10 L-11 L-12 L-13

Figure 69: The interaction pictures of the finally selected 13 leads for HIV protease target

6.5. ADME predictions

We finally evaluated the 13 lead compounds for the pharmaceutically relevant properties to check drug-likeness and including its ADME. QikProp module [164] was used for evaluation of drug-like behavior through analysis of pharmacokinetic parameters required for absorption, distribution, metabolism and excretion. The 13 lead compounds showed good partition coefficient (QPlogPo/w) values which were critical for understanding of absorption and distribution of drugs. Factor QPPCaco for 13 lead compounds was predicted for apparent Caco-2 cell permeability in nm/sec value (Caco-2 cells are model for gut-blood barrier) a key factor for estimation of cell permeability in biological membranes and its metabolism.

Table 39: Lead compounds with their Qikprop properties for HIV protease target

Title	mol MW[a]	QPlogPo/w[b]	QPPCaco[c]	QPlogHERG[d]	% Human Oral Absorption[e]	Rule Of Five[f]
L-1	611.24	4.57	135.89	-7.59	78.93	1
L-2	433.51	3.75	1454.96	-6.25	100.00	0
L-3	399.93	4.44	2250.15	-5.47	100.00	0
L-4	512.58	5.01	234.16	-7.08	72.75	2
L-5	425.51	5.51	844.29	-7.50	100.00	1
L-6	376.41	4.86	999.66	-7.07	100.00	0
L-7	471.29	4.08	378.18	-6.56	96.97	0
L-8	479.49	2.60	57.31	-4.42	60.69	1
L-9	605.19	3.89	116.71	-6.56	73.75	1
L-10	421.90	5.26	734.00	-7.04	96.06	1
L-11	492.40	1.56	12.19	-5.94	42.54	1
L-12	419.49	4.32	1067.31	-5.98	100.00	0
L-13	361.41	1.01	135.30	-4.90	70.98	0

[a] Molecular weight
[b] Predicted octanol/water partition co-efficient log p (acceptable range from −2.0 to 6.5).
[c] Predicted Caco-2 cell permeability in nm/s (acceptable range: <25 is poor and >500 is great).
[d] Predicted value for blockage of HERG K+ channels (concern below −6.5).
[e] Percentage of human oral absorption (<25% is poor and >80% is high).
[f] Rule of five (no. of violations of Lipinski's rule of five: 0 is good and 4 is bad)

All the 13 lead compounds passed all the pharmacokinetic parameters and are within the acceptable range defined for human use (**Table 39**) with additional parameters such as molecular weight, H-bond donors, H-bond acceptors, and humoral absorption according to Lipinski's rule of 5. Thus, compounds with binding and pharmacokinetic properties were finalized as valuable leads for further biologic assays.

6.6. Anti HIV activity

The SensoLyte 490 HIV-1 protease assay kit provides a well convenient assay for high throughput screening of HIV-1 protease inhibitors and continuous quatification of HIV-1 protease activity using a fluorescence resonance energy transfer (FRET) peptide. Basically the sequence of FRET peptide is derived from the native p17/p24 cleavage site on Pr^{gag} for HIV-1 protease. In the FRET peptide, the fluorescence of EDANS is quenched by DABCYL until this peptide is cleaved at Tyr-Pro bond into two separate fragments by HIV-1 protease. While cleavage the fluorescence of EDANS is recovered and monitored at excitation/emission = 340 nm/490 nm and can be performed in a convenient 96 or 384-well microplate format.

The designed lead compounds for HIV protease as target, were procured and screened at single concentration of 25 μM and saquinavir (10 μM) was used as standard drug using the assay protocol as described in materials & method. Lead **L-8** showed the highest inhibition of 92% and **L-7** showed 82% inhibition. This showed that they were more potent compared to all other 11 leads. Among the other 11 leads, 10 of them showed more than 60% inhibition and **L-4** showed only 58% inhibition as shown in **Table 40.**

Table 40: Anti HIV activity of the selected hit compounds at 25 μM concentration

Compound code	% Inhibition at 25 μM
L-1	66
L-2	70
L-3	60
L-4	58
L-5	64
L-6	71
L-7	82
L-8	92
L-9	69
L-10	64
L-11	63
L-12	69
L-13	70
Saquinavir*	93

***At 10 μM cocentration

The interaction pictures (**Figure 70**) of these two most active compounds were shown that they had H-bond interactions with the amino acid moieties which were crucial for the activity of the crystal ligands. The most active compound **L-8** had six H-bonds, they were two with Arg8 and four with Asp25. The H-bond interaction with Asp25 was shown by six crystal ligands out of eight selected for these study. Therefore it showed that these interactions were responsible for the activity of this compound **L-8**. The compound **L-9** was shown four H-bonds interactions, one with Asp25, one with Asp30 and two with Ile50. Among these three, Asp25 and Asp30 were important amino acid interactions shown by selected six crystal ligands for these study. Therefore, these two interactions were responsible for the activity of **L-9**.

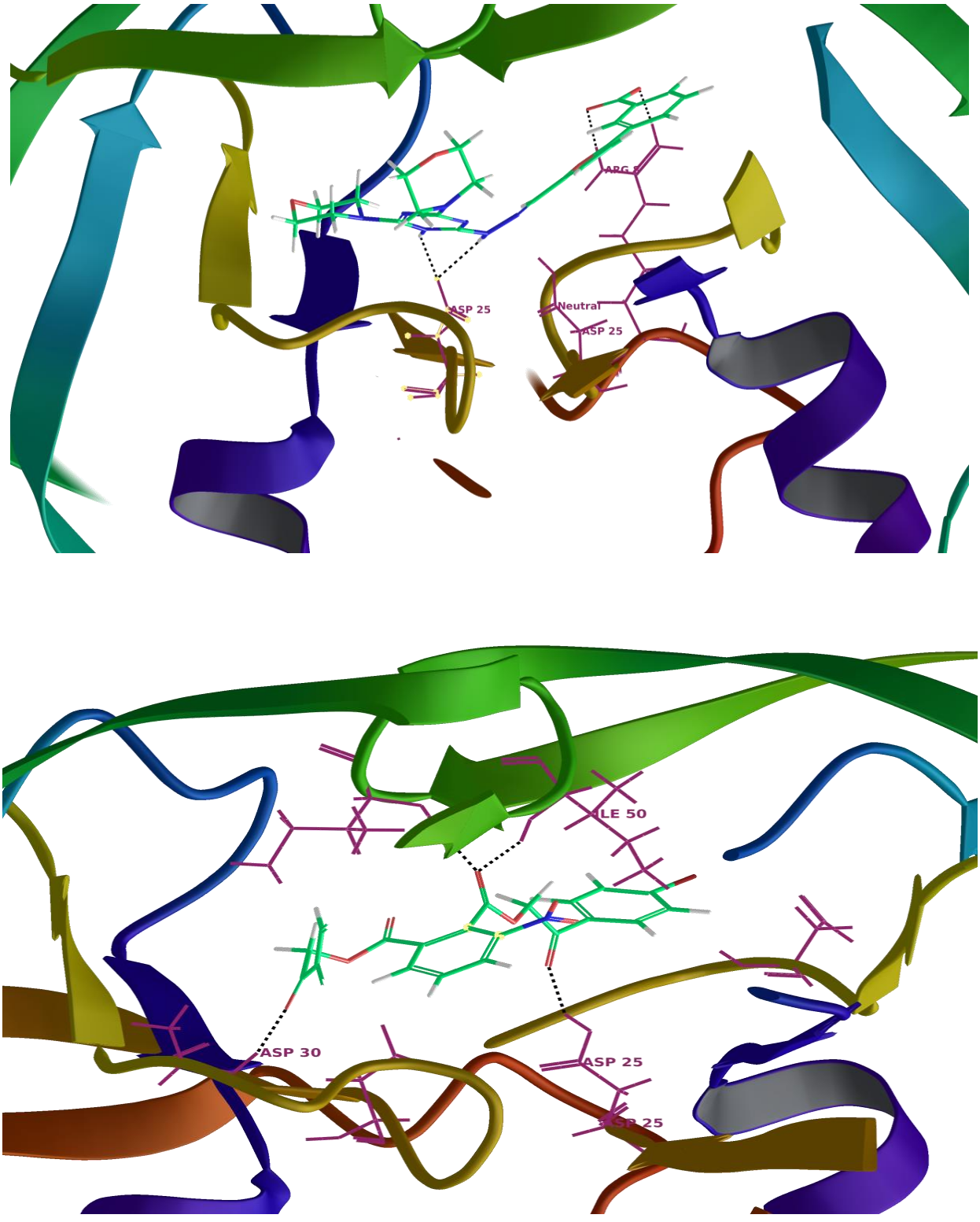
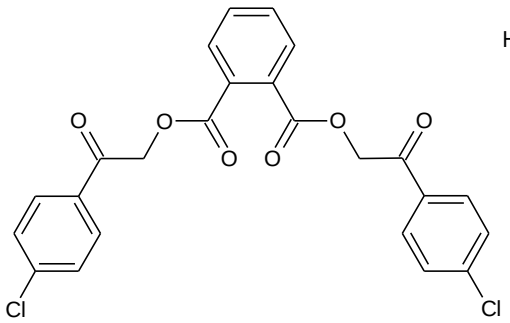


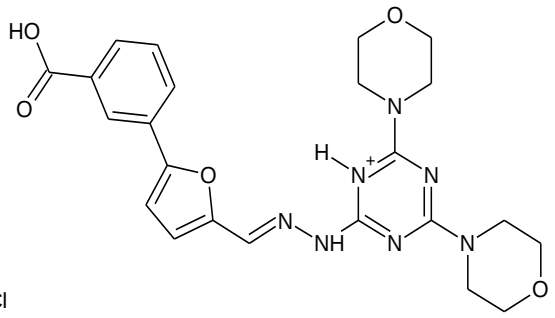
Figure 70: The interaction pictures of two most potent leads **L-8** & **L-9**

6.7. Conclusion

To find new inhibitors for HIV protease enzyme present work extended multiple pharmacophore and 3D QSAR studies using 9 crystal structures and 134 diverse HIV protease inhibitors. We used these reported inhibitors for our studies to generate pharmacophore modeling and 3D QSAR techniquies in PHASE. 3D QSAR model was developed using these 134 selected compounds and validated by both LOO and external validation methods. Using 3D QSAR counter map visualization, SAR studies were done that could help to give insights for further design of newer leads. Virtual screening of Asinex database using the best pharmacophores to yield hits which were filtered by three consecutive docking runs (HTVS, Glide SP, Glide XP and Gold) to finally identify top ranked hits. Furthermore, refinement based on the ADME predictions resulted in 13 valuable hits with good binding scores and pharmacokinetic properties. We performed biological screening for these 13 leads and found, phthalate scaffold **L-7** and morpholine scaffold **L-8** to show very good inhibition of 82% and 92% respectively. Thus the present work revealed two new diverse, benzene dicarboxylate and triazine morpholine derivatives as HIV protease inhibitors valuable leads for further drug development.



L-7 82%



L-8 92%

CHAPTER 7

DESIGN OF MTB GlmU INHIBITORS: RESULTS & DISCUSSION

7.1. Background

Drug resistance has become a major stumbling block to overcome diseases and thus there is always a need to find new drugs and new pathways. Now there are very few drugs for the treatment of *M. tuberculosis* disease due to drug resistant strains that resulted in reduced efficacy. It is necessary to find new drugs and new targets to kill the pathogen. There are some new targets and pathways discovered recently for tuberculosis due to resistance [271] or for dormat survival of the bacteria which were not much explored. GlmU is one such target which is essential for the survival of the pathogen [272]. Recent studies on the mycobacterial proteome using in-silico analysis suggested GlmU to be a potential drug target [273].

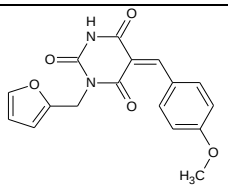
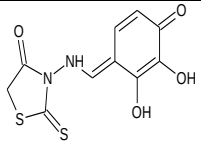
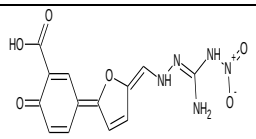
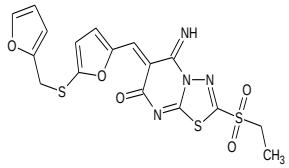
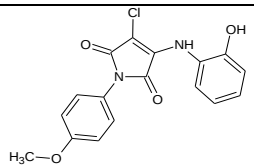
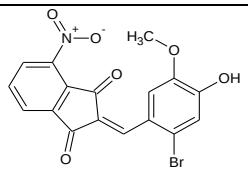
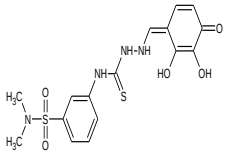
GlmU plays its bifunctional role to two functionally autonomous active sites: the acetyltransferase active site and the uridyltransferase active site. Kinetic and structural studies demonstrated that these two active sites were present on two different protein domains. Within the C-terminal (acetyl transferase) domain acetyltransferase reaction occured but the rate-limiting uridyl transferase reaction is in the N-terminal (uridyl transferase) domain. It has been previously shown that the uridyltransferase activity follows a sequential substrate-binding order with UTP binding first noncovalently to the GlmU enzyme [129, 130]. Till GlcNAc-1-P binds, the uridyltransferase active site will be in an open apo conformation which gives lots of conformational change. After GlcNAc-1-P binding, the phosphate oxygen of GlcNAc-1-P undergoes nucleophilic attack on phosphate of UTP.

The identification of inhibitors using wet lab techniques is an expensive and time consuming task. Thus, there is need to develop theoretical models for predicting inhibitors against a potential target. Ligand based approaches have been used to provide insights into the structural and binding environment based on the key interactions shown by reported PubChem inhibitors. Several other computational approaches have also been employed for the design and development of the GlmU inhibitors like 3D QSAR studies, pharmacophore map generation, docking studies and so on as discussed earlier.

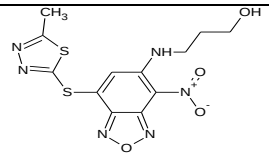
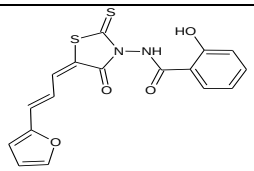
7.2. Drug design based on ligand based strategy

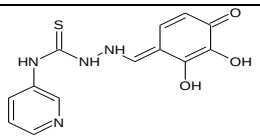
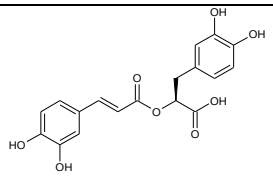
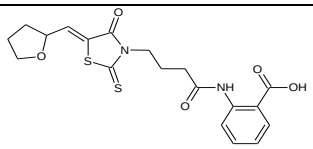
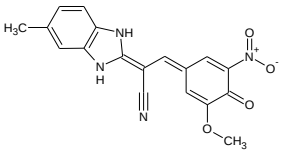
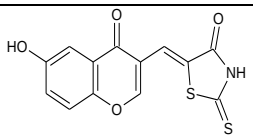
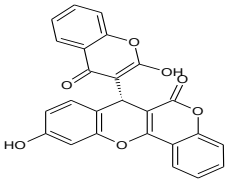
The flow of work accomplished for the design was as similar to that of **Figure 40**. We chose 27 GlmU inhibitors from PubChem Bioassay AID-1376 with known IC_{50} values against MTB GlmU [274, 275]. These inhibitors showed a wide range of activity (1.79 μM-1073 μM) and structural diversity. The IC_{50} values were converted to pIC_{50} where IC_{50} was the concentration of the compounds producing 50% inhibition of GlmU. The dataset consisted of some highly active and inactive molecules with few molecules moderatly active. A total of 27 **(Table 41**) molecules were randomly selected with pIC_{50} values, of which 22 molecules were selected for training set and 5 molecules were chosen as test set. All structures were drawn using 2D sketcher tab on Maestro graphical user interface.

Table 41: Compounds used for 3D QSAR study with their observed and predicted activity for MTB GlmU target

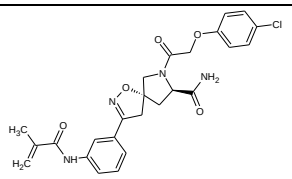
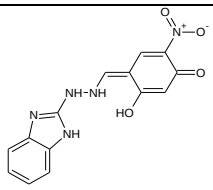
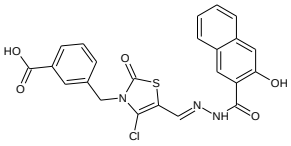
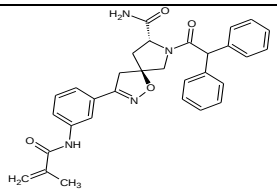
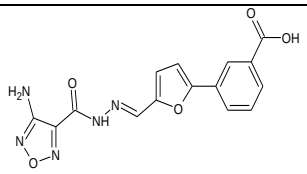
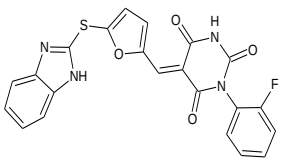
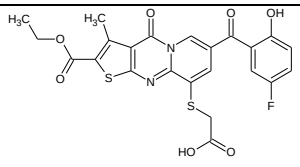
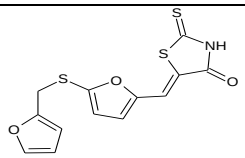
S.No.	Compounds	IC_{50} (µM)	Observed pIC_{50}	Predicted pIC_{50}	Residual error	Data set
1.		1.79	5.747	5.210	-0.537	Training
2.		7.79	5.108	4.850	-0.258	Training
3.		28.41	4.547	4.390	-0.157	Training
4.		49.4	4.306	4.390	0.084	Training
5.		53.6	4.271	4.010	-0.261	Training
6.		60.9	4.215	4.340	0.125	Training
7.		68.4	4.165	4.310	0.145	Training

Contd..

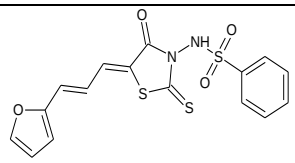
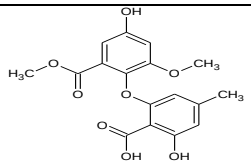
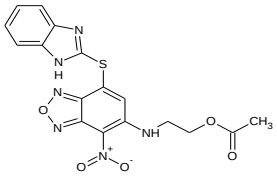
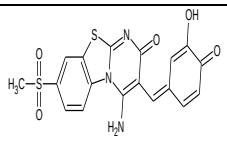
8.		77.98	4.108	3.990	-0.118	Training
9.		80.62	4.094	4.150	0.056	Training

10.		82.97	4.081	4.280	0.199	Training
11.		84.94	4.071	4.390	0.319	Training
12.		86.02	4.065	3.920	-0.145	Training
13.		94.27	4.026	3.990	-0.036	Training
14.		111.69	3.952	4.200	0.248	Training
15.		117.18	3.931	4.020	0.089	Training

Contd..

16.		118.06	3.928	4.010	0.082	Training
17.		124.94	3.903	3.970	0.067	Training
18.		134.17	3.872	3.910	0.038	Training
19.		166.77	3.778	3.550	-0.228	Training
20.		188.13	3.726	3.930	0.204	Training
21.		219.12	3.659	4.060	0.401	Training
22.		1073.41	2.969	2.640	-0.329	Training
23.		17.74	4.751	4.400	-0.351	Test

Contd..

24.		49.58	4.305	4.100	-0.205	Test
25.		82.7	4.082	4.190	0.108	Test
26.		116.16	3.935	4.170	0.235	Test
27.		159.09	3.798	3.860	0.062	Test

7.3. Determination of the best pharmacophore model and its validation

The training set molecules were sufficiently rigid and congeneric, and hence our 3D QSAR approach involved the generation of a common pharmacophore hypothesis built on the principle of identification and alignment of pharmacophoric features of the chemical structures. We divided our data set into 2 actives, 2 inactives and remaining 23 were found to be moderately actives. We kept the three maximum and minimum number of sites to be matched and the common pharmacophore generated should match with the two actives. Three different combinations of common pharmacophore were produced after all scoring functions were employed. They were AAH with maximum 3 hypotheses, AAD 23 hypotheses and AAA 17 hypotheses. We clustered these available hypotheses and we obtained 13 clusters. Based on the calculated survival active, survival inactive, post-hoc, vector score, energy and volume score we

selected three hypotheses (**Table 42**). The hypothesis 1 (AAD) showed 2 hydrogen bond acceptors and 1 hydrogen bond donor, hypothesis 2 (AAA) showed 3 hydrogen bond acceptors and hypothesis 3 (AAH) showed 2 hydrogen bond acceptors and 1 hydrophobic as features (**Figure 71)**.

Table 42: Top three hypotheses selected based on their scores for MTB GlmU

S.NO	Hypothesis	Survival Score	Survival-inactive Score	Vector Score	Volume Score	Site Score	Energy
1.	AAD	3.254	2.616	0.826	0.462	0.87	0.001
2.	AAA	3.031	2.211	0.815	0.442	0.63	0.042
3.	AAH	2.720	1.756	0.513	0.442	0.67	0.031

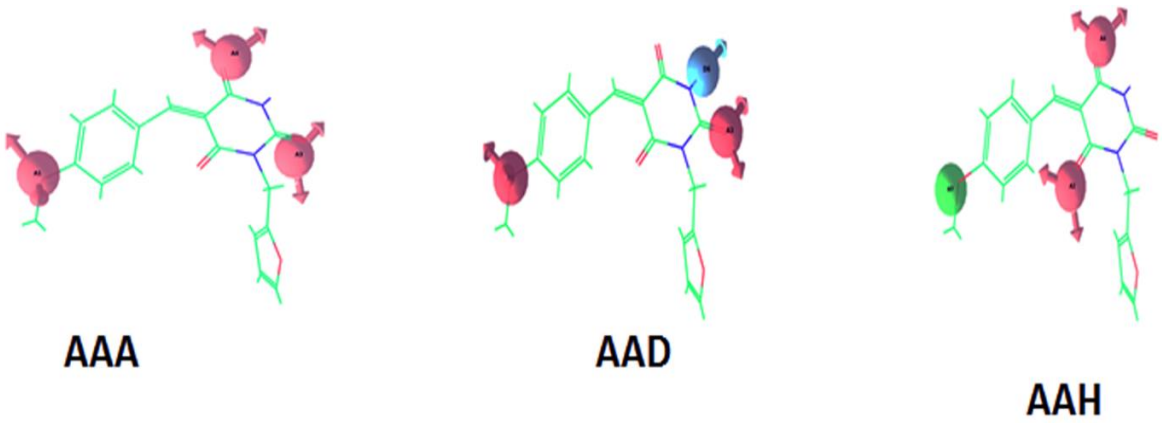


Figure 71: Pharmacophores selected from 3D QSAR based approach. Out of these three pharmacophore, AAD pharmacophore was the best pharmacophore based on the EF and GH calculations for MTB GlmU target

These top three hypotheses were validated using decoys set of 1027 compounds in which 27 were known actives. Using the Find matches in Phase module of Schrödinger suite, with the total number of molecules in the database (D) 1027, 45 compounds were obtained as hits (H_t) for the hypothesis 1, in which 46.67% were active yields (%Y), 77.78% ratio of actives were retrieved in the hit lists (%A), and the values of EF (8.64) and GH (0.52) indicated a good sign of the high

efficiency of hypothesis 1. The values of EF 1%, RIE, ROC and BEDROC values were also high for the hypothesis 1(**Table 43**).

Table 43: Pharmacophore validation parameters for the best three hypotheses for MTB GlmU target

Top 3 Hypotheses	**AAD**	**AAA**	**AAH**
[a] H_t	45	55	60
[b] H_a	21	19	13
[c] %Y	46.67	34.55	21.67
[d] % A	77.78	70.37	48.15
[e] EF	8.64[◊]	6.39	4.01
[f] Fn	6	8	14
[g] Fp	24	36	47
[h] GH	0.52[◊]	0.40	0.26
[i] EF 1%	12	7	8
[j] RIE	1.45	0.98	1.02
[k] ROC	2.41	1.20	1.11
[l] BEDROC (α-20)	0.54	0.11	0.24

[a] Total number of hit molecules from the database
[b] Total number of active molecules in hit list
[c] Yield of actives =[$(H_a/H_t) \times 100$]
[d] Ratio of actives =[$(H_a/A) \times 100$]
[e] Enrichment factor using formula
[f] False negatives =[$A - H_a$]
[g] False Positives = [$H_t - H_a$]
[h] Goodness of fit score using the formula
[◊] best EF and GH scores
[i] EF- Enrichment Factor at 1% of the decoy dataset
[j] RIE- Robust Initial Enhancement
[k] ROC-Receiver Operating Characteristic curve value
[l] BEDROC- Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic

When we compared the distance between the pharmacophoric features (**Table 44**), it was found to be 10.227 Å between the two H-bond acceptors as the maximum distance and the minimum distance was 2.4 Å between A2 and D. Though the pharmacophore hypothesis AAA showed

better EF and GH values, it was not utilized for further screening our compound library because of three non-diversity of pharmacophore features as there were only 3 acceptors in this model.

Table 44: The distance between the pharmacophoric features for the top three selected pharmacophores for MTB GlmU target

Features	A1-A2 (Å)	A2-D (Å)	D-A1 (Å)	A2-A3 (Å)	A3-A1 (Å)	A2-H (Å)	H-AI (Å)
AAD	10.227	2.400	9.969	NA	NA	NA	NA
AAA	10.227	NA	NA	4.436	8.488	NA	NA
AAH	4.746	NA	NA	NA	NA	8.488	5.821

Distance shown in Å, *A-Acceptor, D-Donor and H-Hydrophobic

We selected the pharmacophore hypothesis based on high EF and GH values and also they should possess different combinations of features. The hypothesis AAH had different combination, but the EF and GH values are very low. From the overall validations, we were assured that hypothesis 1(AAD) can predict most of the experimentally active molecules in the same scale compared to the remaining 2 hypotheses. Hence hypothesis 1, a three point model AAD, had two hydrogen bond acceptors (A) and one hydrogen bond donor (D) with good scores of EF, GH, % of actives and other parameters was selected for further studies (**Table 43).**

All 27 compounds were aligned to the best pharmacophore hypothesis 1 AAD as shown in the **Figure 72.** However, we utilized all the three models for the 3D QSAR studies and PLS analysis.

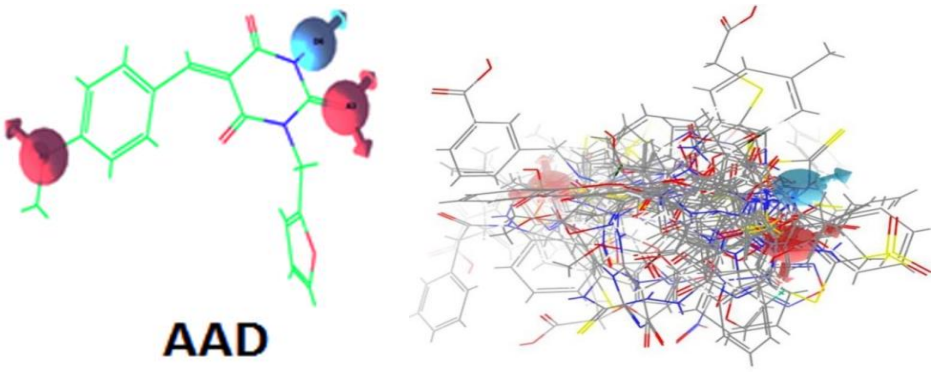


Figure 72: Selected pharmacophre AAD. All selected 27 compounds were aligned to the best selected ligand based pharmacophore AAD for MTB GlmU target

7.4. 3D QSAR models generation, PLS analysis and its external statistical validation.

To develop superlative 3D QSAR models which were meant to exhibit reliable predictions it necessitates internal and external statistical validation. Models with high statistical validation parameter could offer more reliable predictions which could get good optimized leads. Randomly chosen 22 compounds in the training set and 5 compounds in the test were utilized to develop 3D QSAR. Important parameters obtained based on LOO method, favored the internal statistical validation by PLS analysis. Among the best three pharmacophore models, hypothesis 1 showed good external predictive ability for each combination as compared to others. Hypothesis 1 showed a good R^2 value for the training set with 0.8101, good predictive power with Q^2 of 0.5701 for the test sets, with SD of 0.2369, and F value of 85.3. Further the integrity of the model was predicted by r^2_{pred} for test set with the value of 0.5893 **(Table 45).**

Table 45: PLS statistics results for 3D QSAR studies for MTB GlmU target

Statistical parameters	**AAD**	**AAA**	**AAH**
Number ofmolecules in Training set	22	22	22
Number of molecules in Test set	5	5	5
R^2	0.8101	0.7384	0.7543
Q^2	0.5701	- 1.143	-0.0095
SD	0.2369	0.2383	0.205
F-value	85.3	79	150.4
Pearson-R	0.8155	-0.7726	0.2232
RMSE	0.2189	0.5857	0.2941
r^2_{pred}	0.5893	0.5097	0.1156

PLS statistic parameters:

SD - Standard deviation of the regression.
R^2 - for the regression.
F - variance ratio.
r^2_{pred} - predictive correlation coefficient value.
RMSE - root mean square error.
Q - squared (Q^2)value of Q^2 for the predicted activities.
Pearson R -correlation between the predicted and observed activity for the test set.

The accepted LOO-cross validated value of training set (R^2) should be greater than 0.6, LOO cross validated value for test set (Q^2) should show a value greater than 0.55 to attain good predictive capacity, and standard deviation (SD) below 0.3, with minimum root mean square error (RMSE), and high value of variance ratio (F) to provide conventional QSAR validation limits. The predictive correlation coefficient (r^2_{pred}) value generated based on the molecules of test set demonstrated real predictive capacity and robustness of the QSAR model [158]. In the present study the best predictive ability of the model was characterized by correlation coefficient R = 0.8219 (r^2 = 0.6755). High slope of regression lines through the origin k value of 1.008 and k′ value of 0.9892 gave substantial values of R^2_0 value 0.9607 and the R'^2_0 value 0.9816. The calculated relation between r^2, R^2_0 and R'^2_0 gave (r2 $-R^2_0$ /r2) values of -0.4222 and second relation (r2 $-R'^2_0$ /r2) of value -0.4532 showed optimum values within the statistical limits. QSAR model from hypothesis 1 gave r^2_{cv} value of 0.5893. A parameter of modified r^2 [r^2_m] [65] was considered as a better external predictive potential for the whole set of compounds which was found to be 0.5100. This showed to be truly predictive by fulfilling the requirements of every parameter in the external validation **(Table 46)**. From these statistically significant values this model was found to be efficient for predicting efficient leads using Hypothesis 1. Plot of predicted vs. actual pIC_{50} for training and test set is given in **Figure 73.**

Table 46: External statistical validation results of quantitative structure activity relationship (QSAR) result for the Hypothesis 1 (AAD. 1) common pharmacophore hypotheses for MTB GlmU target

External validation	**Parameter calculated**	**Limitations**
r^2_{cv}	0.5893	r^2_{cv}> 0.5
R	0.8219	Must close to 1
r^2	0.6755	r^2> 0.6
k value	1.0083	0.85 ≤ k ≤ 1.15
k′value	0.9892	0.85 ≤ k′≤ 1.15
R^2_0	0.9607	R^2_0 or R'^2_0close to r^2
R'^2_0	0.9816	R^2_0 or R'^2_0close to r^2
$[(r^2 - R^2_0) / r^2]$	-0.4222	$[(r^2 - R^2_0 / r^2]$ <0.1
$[(r^2 - R'^2_0) / r^2]$	-0.4532	$[(r^2 - R'^2_0 / r^2]$ <0.1
r^2_m	0.5100	r^2_m> 0.5

r^2_{cv} - cross validated coefficient
R (or r^2) - correlation coefficient between the actual and predicted activities
k and k′ - slope values of regression lines
R^2_0 and R'^2_0 - correlation coefficients for the regression lines through the origin
$[(r^2 - R^2_0) / r^2]$ and
$[(r^2 - R'^2_0) / r^2]$ - to calculate relation between r^2, R^2_0 and R'^2_0
r^2_m - modified squared correlation coefficient

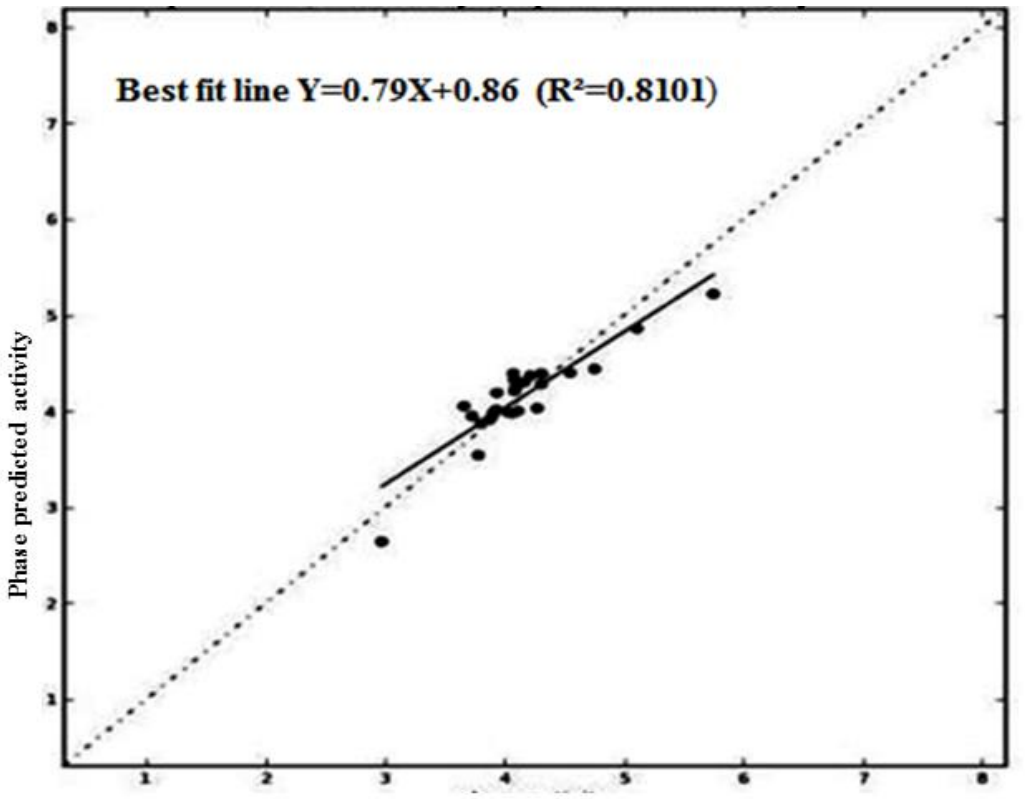


Figure 73: Scatter plot plotted between observed vs predicted activity of GlmU inhibition by the best model obtained using compounds 22 as the training set and validated using compounds 5 as the test set for MTB GlmU target

7.5. Contour maps

The final validated hypothesis 1 obtained from 3D QSAR was used to generate contour maps. These contour maps were important to identify the positions of the substitutions or replacements of atoms to enhance bioactivity. Inhibitory activity can be gained by visualizing and understanding the maps against most active (**1**) and least active (**22**) compounds. This could help in discovering novel scaffolds with good biological activity. The most and least active ligand contour maps were generated and are shown in **Figure 74**. Contour maps indicated H-bond donor effect on the most active ligand (**1**) and least active ligand (**22**) **(Figures 74A** and **74B),**

the hydrophobic effect of the ligands (**Figures 74C** and **74D**) and the electron withdrawing nature (**Figures 74E** and **74F**) of both ligands represented in the figure are discussed further.

The hydrogen bond donor nature for the most active compound **1** and the least active compound **22** when compared showed their most favorable region blue color and unfavorable regions red color (**Figures 74A** and **74B**). Hydrogen bond donor mapping revealed that favorable regions lied near the nitrogens and oxygen of pyrimidinetrione indicating their importance for activity compared to the least active compound **22.** Therefore the presence of pyrimidinetrione in the scaffold backbone is very much needed for the activity.

Figures 74C and **74D** when compared for their hydrophobic nature for the most active compound **1** and least active compound **22** revealed that favored green color region around the furyl, benzyl rings showed that the terminal hydrophobic rings were very much needed for the activity of the compound and unfavorable region yellow color on methoxy moiety revealed that increase in the carbon chain could increase the activity. In **Figure 74E**, the favored red color regions were observed near hydrogen bond acceptors along with respective acceptor hypothesis features of most active compound which indicated that these features were crucial for the activity and these groups should be unsubstituted when further lead modifications indicated. In the least active compound as in **Figure 74F** the unfavorable region blue color surrounded the napthyl ring moiety which indicated that a decrease in the ring size could increase the biological activity of the compound.

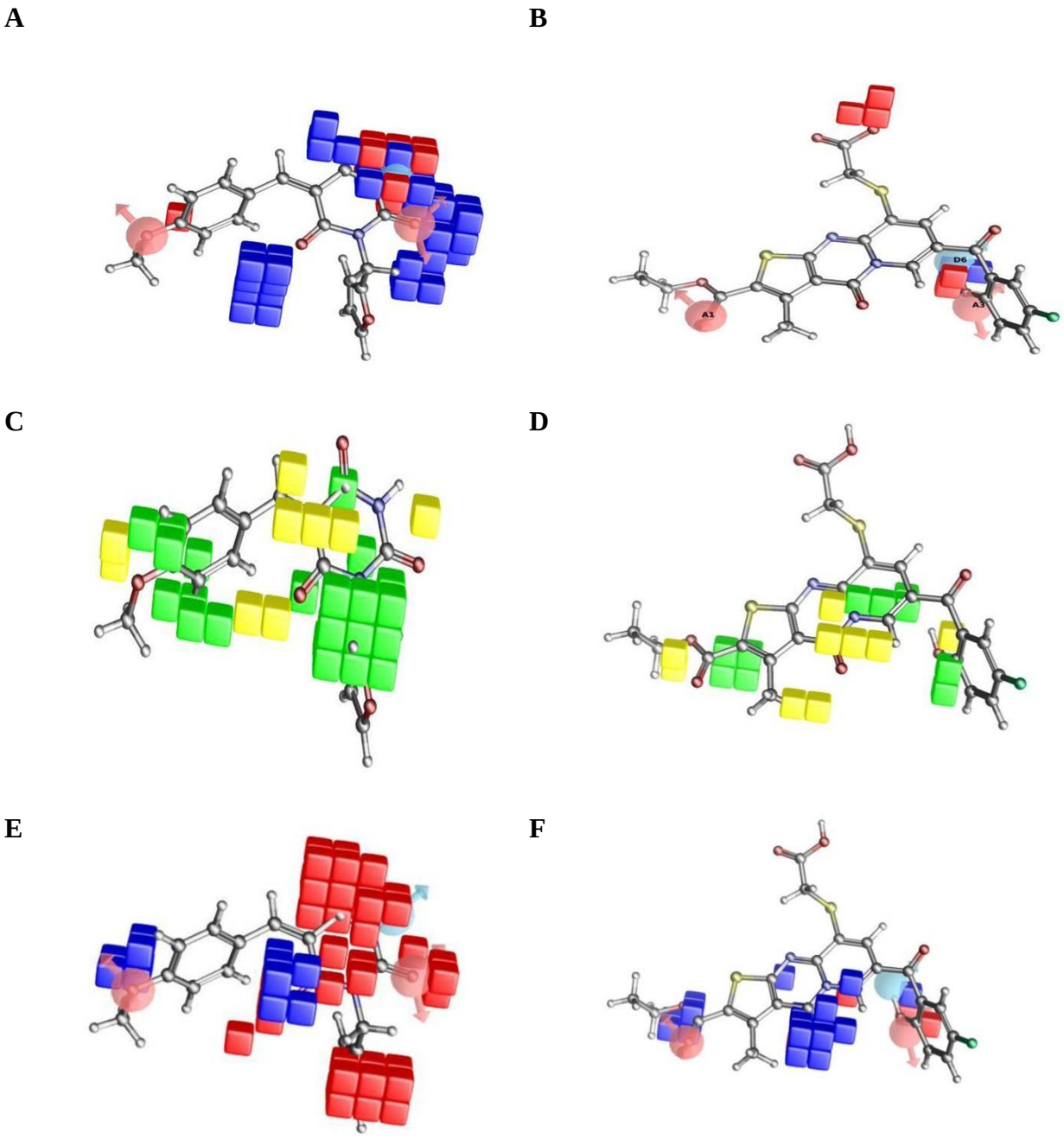
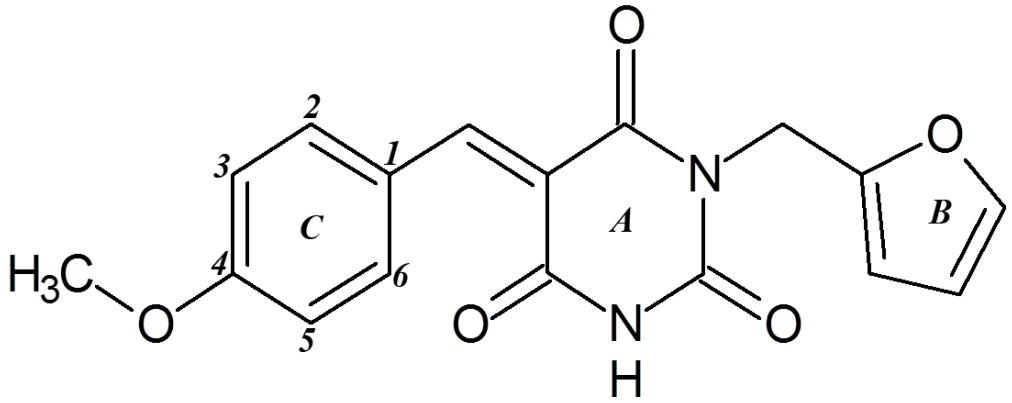


Figure 74: Contour map for MTB GlmU target:

Figure A: H-bond donor effect: Most active; **Figure B**: Least active (Blue- favorable, Red-unfavorable); **Figure C**: Hydrophobic effect: Most active; **Figure D**: Least active (Green-favorable,Yellow- unfavorable); **Figure E**: Electron with-drawing effect: Most active; **Figure F**: Least active (Red-favorable, Blue –unfavorable)

7.6. Lead modification and SAR

Based on the knowledge of 3D QSAR studies and contour maps, we optimized compound **1** **(Figure 75)** for our further studies.



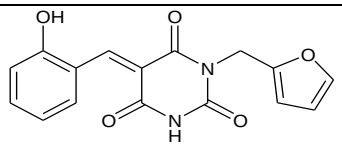
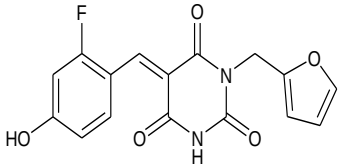
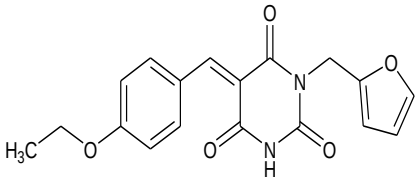
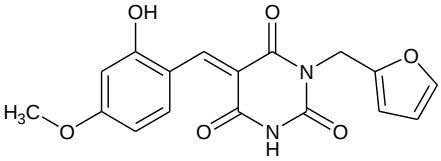
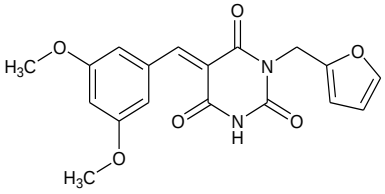
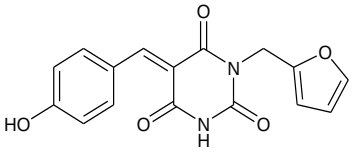
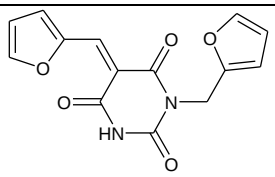
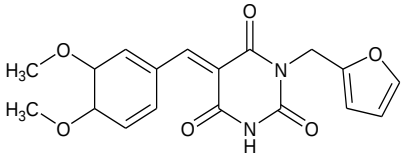
Compound 1

IC50 =1.79 (µM)

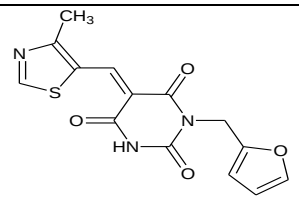
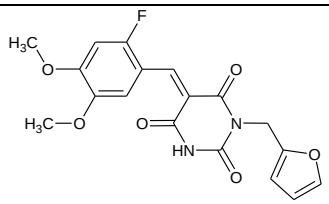
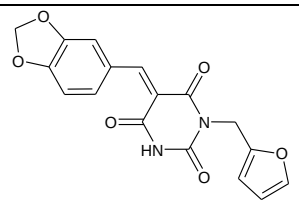
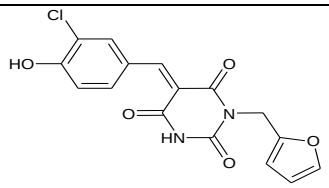
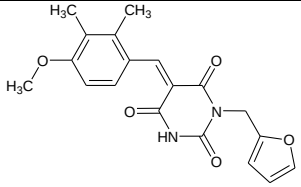
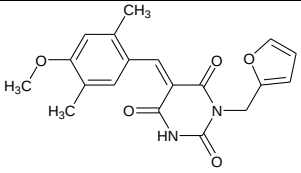
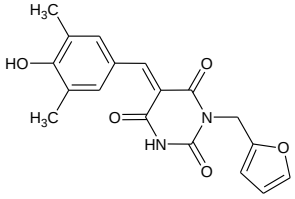
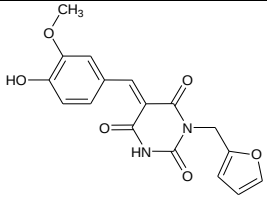
Figure 75: Structure of the highest active compound for MTB GlmU target

Based on 3D QSAR and its corresponding pharmacophore we utilized to find important features for the inhibition of GlmU, to help in designing of lead molecules. Based on contour map (**Figure 74A** & **74B**) the presence of pyrimidinetrione in the scaffold backbone was found to be needed for the activity. Therefore we did not modify the central backbone i.e A ring. The contour maps on hydrophobicity (**Figures 74C & 74D)** revealed that the extreme hydrophobic moiety was very much needed for the activity of the compound, so we keep the furan ring (B ring) as such. The contour map on electron withdrawing group (**Figure 74E & 74F)** revealed that the presence of hydrogen bond acceptor groups in the benzene ring system would increase the activity. Based on these ideas we tried to modify the lead based on substitution in the ring C (**Figure 75).** The derivatised compounds were screened with the pharmacophore hypothesis AAD and showed good fitness and predicted activity values as shown in the **Table 47.**

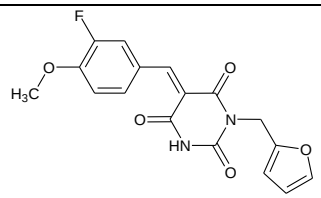
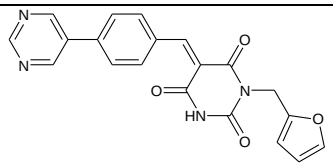
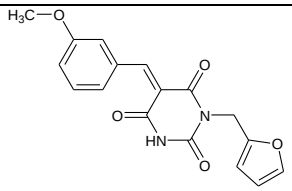
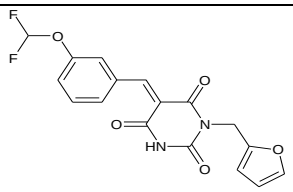
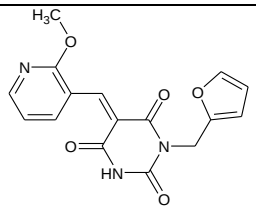
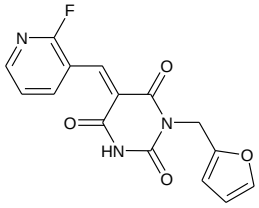
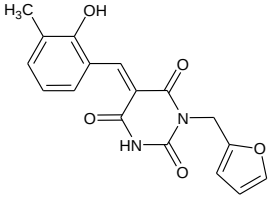
Table 47: The possible leads with their fitness and predicted activity values for MTB GlmU target

Compound ID	Structure	Fitness	Predicted activity
A		1.55	4.10
B		2.84	5.01
C		2.93	5.06
D		2.92	5.06
E		2.43	4.52
F		2.90	5.01
G		1.84	4.72
H		2.89	5.06

Contd..

I		1.82	4.82
J		2.81	5.04
K		2.82	4.99
L		2.85	4.98
M		2.77	4.99
N		2.63	2.80
O		2.82	4.99
P		2.81	4.98

Contd..

Q		2.88	5.07
R		2.27	4.69
S		2.14	4.53
T		2.87	5.04
U		2.21	4.44
V		1.92	4.84
W		1.53	4.06

7.7. Virtual screening and docking studies

We used 3SPT [276] protein with resolution 2.33 Å from Protein Databank (PDB) for our docking studies. We prepared our protein 3SPT using protein preparation wizard by adding hydrogens, optimization and impref minimization by converging the heavy atoms to RMSD 0.30Å using OPLS2005 as forcefield, and finally refined the protein structure. As this protein does not have any bound inhibitors, we used the substrate acetyl CoA binding pocket. The substrate binding pocket of acetyl transferase site was used to generate the grid. The grid center values were as 5.983, -11.513 and 85.085. The docking score of the substrate acetyl CoA was -3.389 kcal/mol. The important H-bond interactions shown by the substrate acetyl CoA were Ser416 and Ala434. To validate our docking protocol, RMSD between crystal structure and docked structure was calculated which was found to be 0.3229 Å (<2 Å) and was in an accepted range. The derivatised compounds were docked into the acetyl transferase substrate binding pocket. All 23 compounds showed good docking score and many showed interactions with important amino acid residues like Ser416 and Ala434.

All 23 leads with their docking score, the number of H-bond and the important amino acids with which they are interacting are given in the **Table 48**. The fitness of the leads was above 1.5 which indicated that all lead compounds were well aligned with the pharmacophore hypothesis AAD and also showed very good predicted activity which was above 4.0 except compound N which showed 2.8.

Table 48: Lead compounds with their docking score, number of H-bonds and important interacting amino acid residues for MTB GlmU target

Leads	Docking score	H-bond	Important interaction
A	-4.589	4	Asn388, Ser416 (2), Thr432
B	-4.044	4	Asn388, Ser416 (2), Thr414
C	-3.467	3	Asn388 (2), Ser416
D	-4.301	3	Thr414, Ser416 , Ala434
E	-4.289	2	Asn388, Ala434
F	-3.426	2	Thr418, Ala434
G	-3.107	3	Asn388 (2), Thr414
H	-3.758	2	Asn388, Ala434
I	-3.526	2	Ser416 (2)
J	-4.118	3	Asn388 (2), Ala434
K	-4.418	4	Asn388,Ser 416 (2), Ala434
L	-3.582	2	Ser416 , Ala434
M	-3.743	3	Asn388, Ser416, Ala434
N	-3.373	2	Asn388 (2)
O	-3.467	2	Asn388, Ser416
P	-3.671	2	Ser416
Q	-3.674	3	Ser416 (2), Ala434
R	-3.640	2	Ser416, Ala434
S	-4.047	3	Asn388 (2), Ala434
T	-3.476	2	Ser416, Ala434
U	-2.962	2	Ser416 (2)
V	-3.464	2	Asn388 (2)
W	-3.980	2	Ser416 (2)

Most of the leads showed interactions with the important amino acid residues like Ser416 and Ala434 except the leads G, N, and V.

7.8. ADME predictions

All 23 designed compounds showed good partition coefficient (QPlogPo/w) values which were critical for understanding of absorption and distribution of drugs, to range from 1.8 to 3.2. Factor QPPCaco indicated that permeability of the lead compounds ranged from 196 to 862, where QPPCaco was a predicted apparent Caco-2 cell permeability in nm/sec value. All the lead

compounds passed the entire pharmacokinetic requirements for drug-like compounds and were within the acceptable range defined for human use. Additional parameters such as molecular weight, H-bond donors, H-bond acceptors, and human oral absorption according to Lipinski's rule of 5 were also evaluated for their drug-like behavior and are represented in **Table 49**.

Table 49: Lead compounds with their Qikprop properties for MTB GlmU target

Title	mol MW [a]	QPlogPo/w [b]	QPlogHERG [c]	QPPCaco [d]	Human Oral Absorption [e]	Rule Of Five [f]
A	312.281	2.058	-5.235	371.623	3	0
B	330.272	2.126	-5.212	215.757	3	0
C	340.335	2.976	-5.396	683.470	3	0
D	342.307	2.053	-4.948	304.608	3	0
E	356.334	2.725	-5.143	693.043	3	0
F	312.281	1.924	-5.383	196.984	3	0
G	286.243	1.827	-4.639	654.473	3	0
H	358.35	2.097	-4.895	862.327	3	0
I	317.319	1.919	-4.756	417.168	3	0
J	374.325	2.931	-5.048	705.537	3	0
K	340.292	2.112	-4.511	747.769	3	0
L	346.726	2.405	-5.281	247.711	3	0
M	354.362	3.000	-4.641	820.511	3	0
N	354.362	3.248	-5.252	727.093	3	0
O	340.335	2.539	-4.993	331.144	3	0
P	342.307	2.005	-4.962	262.538	3	0
Q	344.298	2.815	-5.006	765.250	3	0
R	374.355	2.237	-5.94	220.853	3	0
S	326.308	2.639	-5.252	672.259	3	0
T	362.289	3.475	-5.274	757.674	3	0
U	327.296	2.579	-5.209	741.575	3	0
V	315.26	2.036	-4.964	384.853	3	0
W	326.308	2.308	-4.956	405.873	3	0

[a] Molecular weight (acceptable range < 500)
[b] Predicted octanol/water partition co-efficient log p (acceptable range from −2.0 to 6.5).
[c] Predicted value for blockage of HERG K+ channels (concern below −6.5).
[d] Predicted Caco-2 cell permeability in nm/s (acceptable range: <25 is poor and >500 is great).
[e] Percentage of human oral absorption (<25% is poor and >80% is high).
[f] Rule of five (no. of violations of Lipinski's rule of five: 0 is good and 4 is bad)

Thus, all lead compounds showed good binding interaction and good predicted pharmacokinetic properties. Substituted 1-((furan-2-yl)methyl)pyrimidine-2,4,6(1H,3H,5H)-trione derivatives were procured from commercial outsourcing the synthesis and screened for biological assay.

7.9. GlmU acetyltransferase activity

In order to screen for compounds that could inhibit the GlmU acetyltransferase activity, it was necessary to determine the concentration of GlmU where the activity was ~30%. Therefore, we performed acetyltransferase reactions in 8 µl containing 0.1mM of GlcNAc-1-P and 0.2 mM Acetyl CoA and 0.002 µCi [14C]acetyl-CoA and various concentration of enzyme. Based on the results obtained, we decided to use 0.4 pmoles of MTB GlmU and 0.06 pmoles of E. coli GlmU per reaction. Since, we were not sure of the inhibitory potential of synthesized compounds, the first set of experiments were performed with 100 µM inhibitor. Each time we repeated the assay with the change in inhibitor concentration and DMSO concentration.

Acetyltransferase titration assay (image obtained using Phosphoimager) with a range of enzyme concentration (pmole) along with negative control without enzyme. Spots corresponded to the product [14C] GlcNAc-1-P and unused substrate ([14C]acetyl-CoA) were indicated at 5µM inhibitor concentration, 0.02% DMSO, 1.5pmole of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P. The compounds R, S, T, V and W (**Figure 76)** showed good inhibition in the TLC of the assay when the concentration of the inhibitor was 20µM, 0.1% DMSO, 1.5pmole of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P as shown in the **Figure 77**.

Acetyltransferase titration assay (image obtained using Phosphoimager) with a range of enzyme concentration (pmole). C-ve is negative control without enzyme. Spots corresponds to the product [14C] GlcNAc-1-P and unused substrate ([14C]acetyl-CoA) are indicated at 5µM inhibitor concentration, 0.02% DMSO, 1.5 pmole of MTB GlmU and 100µM concentration of

AcCoA and GlcN-1-P. The compounds R, S, T, V and W showed good inhibition shows in the TLC of 5µM inhibitor concentration, 0.02% DMSO, 1.5pmole of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P as shown in the **Figure 78**.

Acetyltransferase titration assay (image obtained using Phosphoimager) with a range of enzyme concentration (pmole). C-ve is negative control without enzyme. Spots corresponds to the product [14C] GlcNAc-1-P and unused substrate ([14C] acetyl-CoA) are indicated at 1µM inhibitor concentration, 0.1% DMSO, 1.5pmole of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P shows in the TLC of 1µM inhibitor concentration, 0.1% DMSO, 1.5pmole of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P as shown in the **Figure 79**.

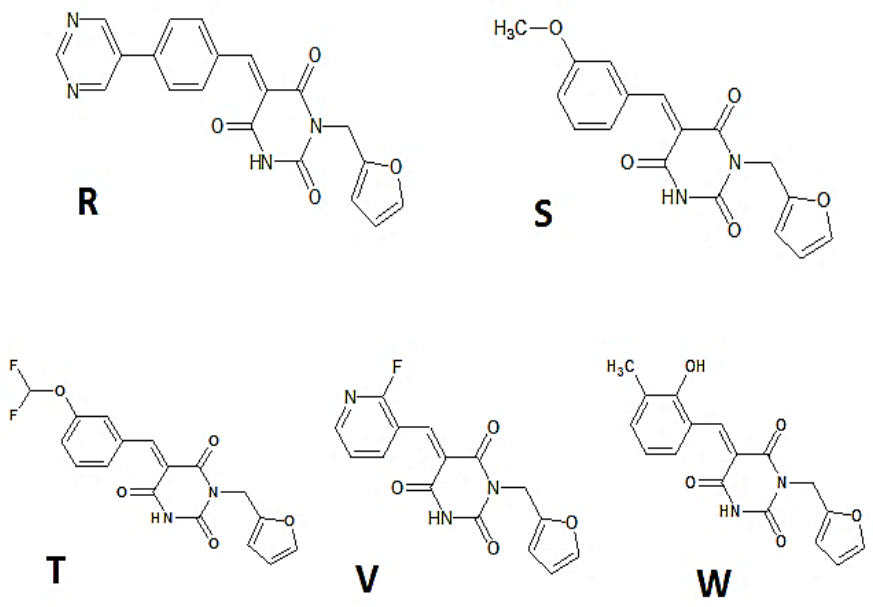


Figure 76: Inhibitors working at various concentrations for acetyltransferase titration assay for MTB GlmU target

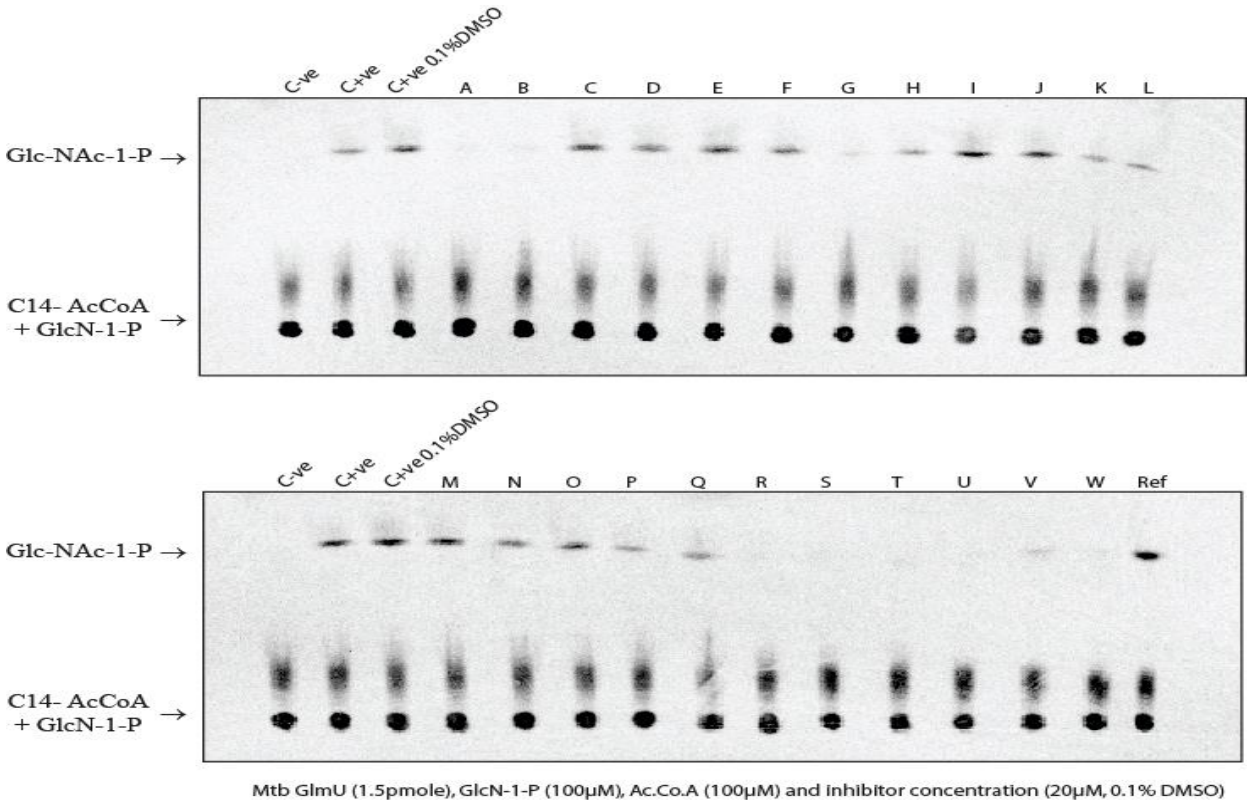


Figure 77: Acetyltransferase titration assay (image obtained using Phosphoimager) with a range of enzyme concentration (pmole). C-ve is negative control without enzyme. C+ve is positive control without DMSO. Spots corresponds to the product [14C] GlcNAc-1-P and unused substrate ([14C] acetyl-CoA) are indicated at 20µM inhibitor concentration, 0.1% DMSO, 1.5pM of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P. The compounds R, S, T, U, V and W showed good inhibition

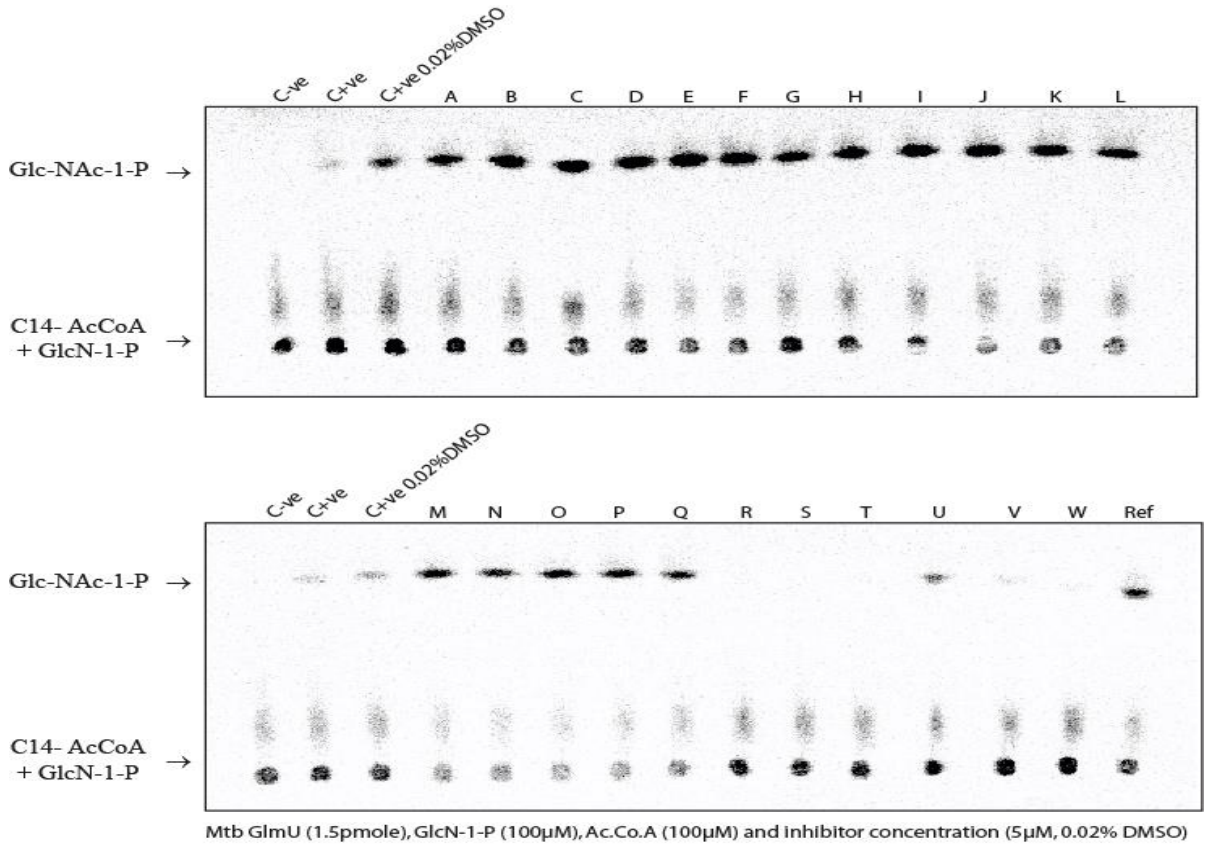


Figure 78: Acetyltransferase titration assay (image obtained using Phosphoimager) with a range of enzyme concentration (pmole). C-ve is negative control without enzyme. C+ve is positive control without DMSO. Spots corresponds to the product [14C] GlcNAc-1-P and unused substrate ([14C]acetyl-CoA) are indicated at 5µM inhibitor concentration, 0.02% DMSO, 1.5pM of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P. The compounds R, S, T, V and W showed good inhibition

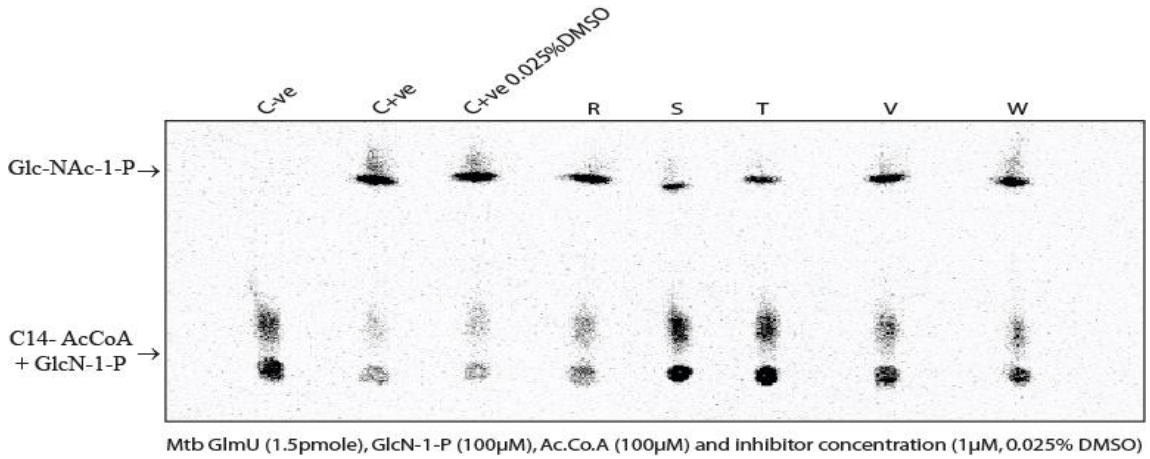
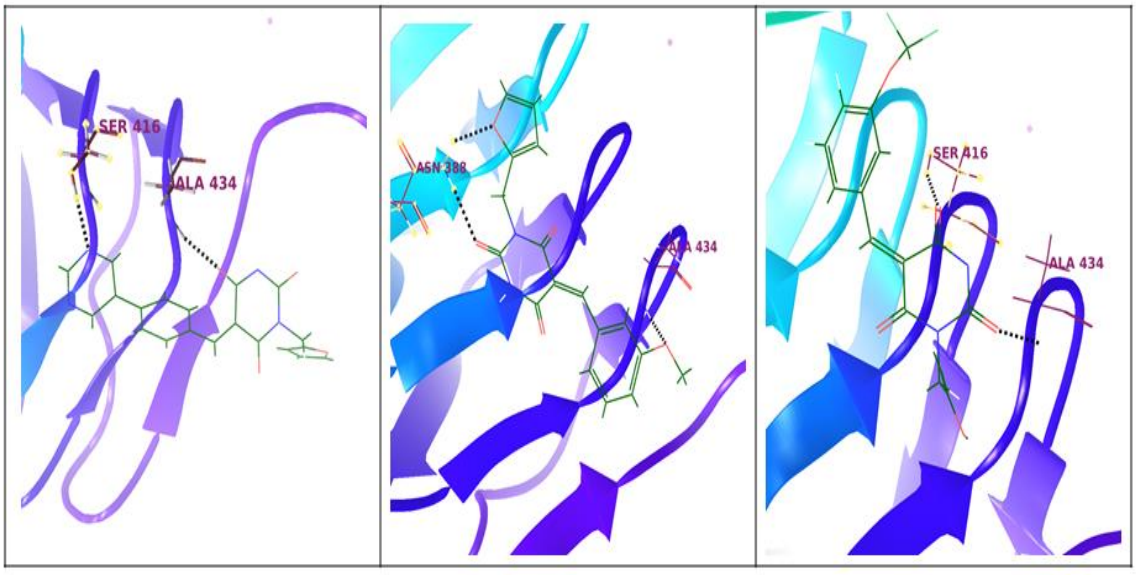
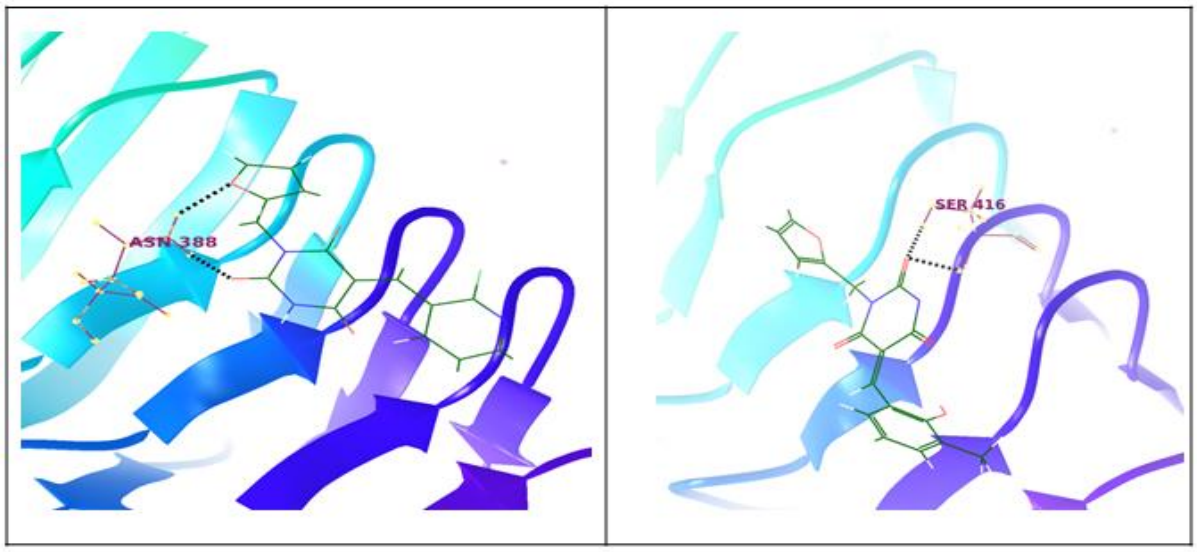


Figure 79: Acetyltransferase titration assay (image obtained using Phosphoimager) with a range of enzyme concentration (pmole). C-ve is negative control without enzyme. . C+ve is positive control without DMSO. Spots corresponds to the product [14C] GlcNAc-1-P and unused substrate ([14C] acetyl-CoA) are indicated at 1µM inhibitor concentration, 0.025% DMSO, 1.5pM of MTB GlmU and 100µM concentration of AcCoA and GlcN-1-P

Based on the above experiments, we concluded that reference compound used for the design was not much potent. However, number of compounds derived from the reference compound such as R, S, T, V and W were good inhibitors even at 1 µM concentration.



R **S** **T**



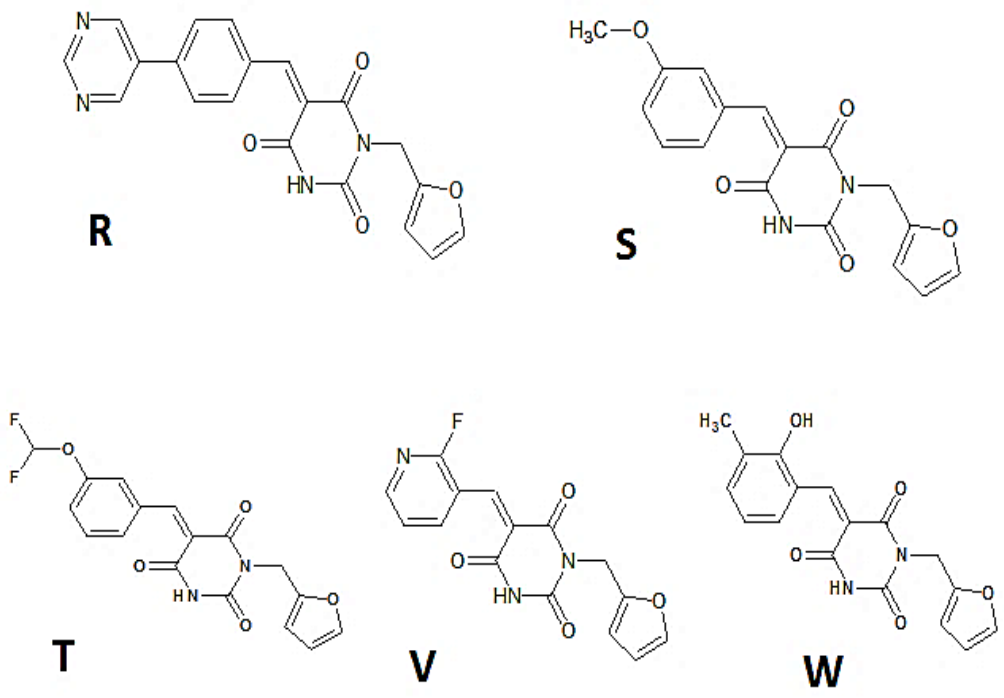
V W

Figure 80: The docked pose of the most potent lead compounds and their ligand interaction diagram for MTB GlmU target

7.10. Conclusion

New inhibitors were designed for GlmU as one of the potential new antimicrobial targets for combatting MTB. Pharmacophore and 3D QSAR studies using 27 diverse GlmU inhibitors were explored. 3D QSAR model was developed using these 27 selected compounds and validated by both LOO and external validation methods. A three point pharmacophore model with two hydrogen bond acceptors and one hydrogen bond donor was developed. Hypothesis 1 was selected as the best one based on its $R^2 = 0.8101$ and $Q^2 = 0.5701$. Using 3D QSAR counter map visualization, SAR studies were derived that could help to yield an insight for further design of newer leads. We employed SAR studies on the reported compound **1** as lead with high pIC_{50} (5.747) value. 23 designed leads were screened using the best pharmacophore AAD and docking studies were done using the PDB 3STP. All leads showed good docking score, fitness, predicted activity and interaction with important amino acid residues. ADME predictions results showed that all leads having good binding scores and pharmacokinetic properties. Thus the present work

revealed new diverse GlmU inhibitors as valuable leads for further biological assays. Based on these 23 compounds were procured and performed acetyltransferase activity studies. These studies revealed that the derivatised leads **R, S, T, V,** and **W** were more potent than the parent compounds. These results showed that using computer-aided drug design tools like 3D QSAR, docking studies, SAR studies, etc. we could get more potent leads.



CHAPTER 8

SUMMARY AND CONCLUSION

With the aim of developing new antimicrobial agents, we tried to explore three different enzymatic targets like HCV NS5B polymerase, HIV protease and MTB GlmU to find potential leads. We applied both structure based drug design and ligand based drug design methods to identify efficient hit molecules.

8.1. HCV NS5B Polymerase

- The HCV NS5B polymerase crystal structures with low resolution and high active bound inhibitors were retrieved from PDB.
- For ligand based approach, we chose 1568 reported compounds from available literatures and clustered them to 132 promising compounds from 151 clusters and divided them into actives, inactives and moderately actives to develop common pharmacophore hypothesis. We selected top three pharmacophores based on their survival active score, survival inactive score, vector score and volume score.
- Further we developed a 3D QSAR model by dividing our compounds into 70 training set and 62 test set compounds.
- We used these six pharmacophores 3CVK (ADR), 3CO9 (ADR), 2GIQ (ARR), 2D3Z (NRR), 2GIR (NHR) and 3D QSAR (AADRR) (5 e-pharmacophore and one ligand based pharmacophore) for virtual screening and docked.
- Based on the pharmacophore screening, 10 hit compounds were selected for biological assays.

- We performed anti HCV activity and cytotoxicity studies for our 10 leads. The leads **H-5** and **H-6**, IC_{50} values were found to be 28.8 µM and 47.3 µM with inhibition of HCV NS5B RdRp activity as nearly 67% and 50%.

8.2. HIV Protease

- 8 crystal structuresof HIV protease 3OXC, 2Q5K, 3DJK, 3H5B, 3I6O, 3NDX, 2AID and 3T11 with their crystal ligands from PDB based on their good resolution (1.00Å - 2.40Å) and their activity ranges from 0.2 nM to 270 nM were selected for the study.
- For the ligand based approach we selected 1535 compunds of range activity 0.026 nM to 316 µM and structural diversity from available literature. Out of which 134 compounds with 168 clusters were divided into actives, inactives and moderately actives to develop common pharmacophore.
- We developed a 3D QSAR model by dividing our 134 compounds into 103 training set and 31 test set.
- We employed all 13 pharmacophore for screening the Asinex database through HTVS, SP and XP dockings.
- Based on docking score, fitness, number of H-bonds, ADME properties, interaction diagram and visual inspection we selected top 13 promising leads for HIV PR.
- We performed anti HIV enzyme inhibition studies for all our 13 leads. All compounds showed more than 60% inhibition at 25 µM concentration except the compound **L-4**.

8.3. MTB GlmU

- We chose 27 GlmU inhibitors from PubChem Bioassay AID-1376 with known IC50 values against MTB GlmU (1.79 µM-1073 µM) and were divided them into actives, inactives and moderately active. Three pharmacophores were obtained and validated.

- We developed a 3D QSAR model by dividing our compounds into 22 training set and 5 test set and we validated our QSAR model for its external predictivity.
- The highest active reported compound **1** used for 3D QSAR study was optimized using contour maps and designed 23 compounds.
- Using the best pharmacophore AAD pharmacophore screening was done and showed good fitness score and predicted activity. The hits were then docked in the acetyl transferase domain's substrate binding pocket of the protein 3SPT.
- Designed leads showed good docking score, interaction with the important amino acid residues and allowed range of pharmacokinetic properties.
- The designed compounds were procured and were tested for GlmU acetyltransferase activity. Inhibitors showed % inhibition even at 1 μM.

CHAPTER 9

FUTURE PERSPECTIVES

In the present work we had attempted to design novel diverse inhibitors for three infectious diseases, HCV, HIV and TB. Initial hits identified had shown some potential to be further developed as candidate drugs.

For HCV polymerase target, we designed drugs by using palm and thumb region of the active site pocket and demonstrated that compounds **H-5** and **H-6** as potential lead compounds. Further optimization and synthesis could be evolved to optimize the lead structures. Also further consideration of other active sites like the finger region of HCV for drug design studies could give more insights on potential leads.

Although HIV protease has been the most studied target with regard to computer-based design, development of small molecules as non-peptidic ligands has been slow. The protypical molecules which are from diverse structural class need to further optimized based on the quantification of bioactivity and selectivity towards HIV. The lead optimization through medicinal chemistry approach could itself be a major development that takes some time to study.

With regard to MTB GlmU as target as there are no drugs available, this work could be a major leap in the discovery step. Wherein the compounds identified as inhibitors in this study could be further developed after considering their cytotoxicity.

REFERENCES:

1. Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H.W., Scheld, W.M., Bartlett, J.G., Edwards, J.Jr. The epidermic of antibiotic-resistant infections: a call to action for the medical community from the infectious diseases society of America. *Clin Infect Dis* 2008, 46: 155-164.
2. Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P. Global trends in emerging infectous diseases. *Nature* 2008, 451: 990-993.
3. Morens, D.M., Folkers, G.K., Fauci, A.S. The challenge of emerging and re-emerging infectious diseases. *Nature* 2004, 430: 242-249.
4. Online Etymolgy Dictionary, www.Etymonline.com browsed on 15/03/14.
5. http://www.google.cp.in/search?q=hcv+structures&espv=210&es_sm=1227source browsed on 15/03/14.
6. http://www.who.int/mediacentre/factsheets/fs164/en/ browsed on 21/03/14.
7. http://www.hemaware.org/story/costs-and-coverage-new-hepatitis-c-drugs browsed on 21/03/14.
8. Rice, C. Prespective: miles to go before we sleep. *Nature* 2011, 474: S8.
9. Hagedorn, C.H., van Beers, E.H., De Staercke, C. Hepatitis C Virus RNA-dependent RNA polymerase (NS5B polymerase). *Curr Top Microbiol Immunol* 2000, 242: 225-260.
10. Tomei, L., Vitale, R.L., Incitti, I., Serafini, S., Altamura, S., Vitelli, A., De Francesco, R. Biochemical characterization of a hepatitis C virus RNA-dependent RNA polymerase mutant lacking the C-terminal hydrophobic sequence. *J Gen Virol* 2000, 81: 759-767.
11. Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., Houghton, M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 1989, 244: 359-362.

12. Kuo, G., Choo, Q.L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J., Stevens, C.E., *et al.* An assay for circulating antibodies to amajor etiologic virus of human non-A, non-B hepatitis. *Science* 1989, 244: 362-364.
13. WHO, Hepatitis C. Factsheet no.164. 2011, accessed at www.who.int/mediacenter/fact sheets/f164/en/ browsed on 21/03/14.
14. Patil, V.M., Gupta, S.P., Samata, S., Masand, N. Current perspective of HCV NS5B inhibitors: a review. *Curr Med Chem* 2011, 18: 5564-5597.
15. Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, Cortese, R., Vitelli, A. The human scavenger receptor class B type 1 is a novel candidate receptorfor the hepatitis C virus. *EMBO J* 2002, 21: 5017-5025.
16. Acton, S., Rigotti, A., Landschulz, K.T., Xu, S., Hobbs, H.H., Krieger, M. Identification of scavenger receptor SR-B1 as a high density lipoprotein receptor. *Science* 1996, 271: 518-520.
17. Evans, M.J., von Hahn, T., Tscherne, D.M., Syder, A.J., Panis, M., Wolk, B., Hatziioannou, T., McKeating, J.A., Bieniasz, P.D., Rice, C.M. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* 2007, 446: 801-805.
18. Lindebach, B.D., Evans, M.J., Syder, A.J., Wolk, B., Tellighuisen, T.L., Liu, C.C., Maruyama, T., Hynes, R.O., Burton, D.R., McKeating, J.A., Rice, C.M. Complete replication of hepatitis C virus in cell culture. *Science* 2005, 309: 623-626.
19. Gastaminza, P., Cheng, G., Wieland, S., Zhong, J., Liao, W., Chisari, F.V. Cellular determinants of hepatitis C virus assembly, maturation, degradation, and secretion. *J Virol* 2008, 82: 2120-2129.

20. Chinnasamy, S., Cai, H., Kao, C. An update on small molecule inhibitors of the HCV NS5B polymerase: effects on RNA synthesis in vitro and in cultured cells, and potential resistance in viral quasispecies. *Virus Adapt Treat* 2010, 2: 73-89.
21. Rossignol, J.F., Elfert, A., El-Gohary, Y., Keeffe, E.B. Improved virologic response in chronic hepatitis C genotype 4 treated with nitazoxanide, peginterferon, and ribavirin. *Gastroenterology* 2009, 136: 856-862.
22. Rossignol, J.F., Elfert, A., Keeffe, E.B. Evaluation of a 4 week lead-in phase with nitazoxanide (NTZ) prior to peginterferon (PeGiFN) plus NTZ for treatment of chronic hepatitis C: final report. *J Hepatol* 2008, 48: 344A-345A.
23. Gorbakov, V.V., Kim, H., Oronsky, B., Lang, W. HCV RNA results from phase II, randomized, open-label study of omega interferon (IFN) with or without ribavirin in IFN-naïve genotype 1 chronic hepatitis c patients. *Hepatol* 2005, 42: 705-714.
24. Lin, C.C., Philips, L., Xu, C., Yeh, L.T. Pharmacokinetics and safety of viramidine, a prodrug of ribavirin, in healthy volunteers. *J Clin Pharmacol* 2004, 44: 265-275.
25. Revill, P., Serradell, N., Bolos, J., Rosa, E. Telaprevir. *Drugs of the future* 2007, 32:788.
26. Njoroge, F.G., Chen, K.X., Shih, N.Y., Piwinski, J.J. Challenges in modern drug discovery: a case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection. *Acc Chem Res* 2008, 41: 50-59.
27. Kong, L., Giang, E., Nieusma, T., Kadam, R.U., Cogburn, K.E., Hus, Y., Dai, X., Standfield, R.L., Burton, D.R., Ward, A.B., Wilson, I.A., Law, M. Hepatitis C virus E2 envelope glycoprotein core structure. *Science* 2013, 342: 1090-1094.
28. Halliday, J., Klenerman, P., Barnes, E. Vaccination for hepatitis C virus: closing in on an evasive target. *Expert Rev Vaccines* 2011, 10: 659-672.

29. Weiss, R.A. How does HIV cause AIDS? *Science* 1993, 260: 1273-1279.

30. Douek, D.C., Roederer, M., Koup, R.A. Emerging concepts in the immunopathogenesis of AIDS. *Annu Rev Med* 2009, 60: 471-484.

31. http://www.who.int/mediacentre/factsheets/fs360/en/ browsed on 21/03/14.

32. Hu, W.S., Temim, H.M. Retroviral recombination and reverse transcriptase. *Science* 1990, 250: 1227-1233.

33. Charpentier, C., Nora, T., Tenaillon, O., Clavel, F., Hance, A.J. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Viro* 2006, 80: 2472-2482.

34. Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 1999, 397: 436-441.

35. http://en.wikipedia.org/wiki/File:HIV Virion-en.png browsed on 24/03/14

36. Chan, D.C., Fass, D., Berger, J.M., Kim, P.S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 1997, 89: 263-273.

37. Goto, T., Nakai, M., Ikuta, K. The life cycle of human immunodeficiency virus type 1. *Micron* 1998, 29: 123-138.

38. http://en.wikipedia.org/wiki/File:HIV_gross_cycle_only.png browsed on 24/03/14.

39. Freed, E.O. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virol* 1998, 251: 1-15.

40. Englemann, A., Mizuuchi, K., Craigie, R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 1991, 67: 1211-1221.

41. Van Lint, C., Bouchat, S., Marcello, A. HIV-1 transcription and latency: an update. *Retrovirol* 2013, 10: 67.

42. Wapling, J., Srivastava, S., Shehu-Xhilaga, M., Tachedijian, G. Targeting human immunodeficiency virus type 1 assembly, maturation, and budding. *Drug Target Insights* 2007, 2: 159-182.

43. Ganser-Pornillos, B.K., Yeager, M., Sunquist, W.I. The structural biology of HIV assembly. *Curr Opin Struct Biol* 2008, 18: 203-217.

44. Yogeeswari, P., Sriram, D. "Medicinal Chemistry, (2nd ed.,) Pearson Education Singapore Ltd., 2010, 505-532.

45. Crum, N.F., Riffenburgh, R.H., Wegner, S. Comparisons of causes of death and mortality rates among HIV-infected persons: analysis of the pre-, early, and late HAART (highly active antiretroviral therapy) eras. *J Acquir Immune Defic Syndr* 2006, 41: 194-200.

46. Kumar, V., Abbas, A.K., Fausto, N., Mitchell, R.N. Robbins Basic Pathology (8th ed.), Saunhders Elsevier. 2007, 516-522.

47. http://www.who.int/mediacentre/factsheets/fs104/en/ browsed on 15/03/14.

48. http://www.who.int/mediacentre/news/en/ browsed on 15/03/14.

49. www.who.int/tb/publication/inventorystudies/en/index.html browsed on 15/03/14.

50. http://www.who.int/tb/challenges/mdr/en/ browsed on 15/03/14.

51. www.who.int/tb/publications/inventorystudies/en/index.html browsed on 15/03/14.

52. http://www.who.int/tdr/research/tb-hiv/en/ browsed on 15/03/14.

53. McCarthy, O.R. The key to the sanatoria. *J R Soc Med* 2001, 94:413-417.

54. Russell, D.G., Barry, C.E.3rd., Flynn, J.L. Tuberculosis: what we don't know can, and does, hurt us. *Science* 2010, 328: 852-856.

55. van Soolingen, D., Hoogenboezem, T., de Haas, P.E., Hermans, P.W., Koedam, M.A., Teppema, K.S., Brennan, P.J. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: Characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* 1997, 47: 1236-1245.

56. hhtp://wapedia.mobi/en/Tuberculosis_treatment browsed on 15/03/14.

57. Tollenaere, J.P. The role of structural-based ligand design and molecular modeling in drug discovery. *Pharm World Sci* 1996, 18: 56-62.

58. De Clercq, E. The history of antiretrovirals: key discoveries over the past 25 years. *Rev Med Virol* 2009, 19: 287-299.

59. Lounnas, V., Ritschel, T., Kelder, J., McGuire, R., Bywater, R.P., Foloppe, N. Current progress in structure based rational drug design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* 2013, 5:e201302011.

60. Selassie C.D., Abraham, J.D. "History of quantitative structure-activity relationships" In Burger's Medicinal Chemistry and Drug Discovery. (6^{th} ed.), Vol 1: Drug Discovery. John Wiley & Sons. 2003, 1-48.

61. Tropsha, A., Abraham, J.D. "Recent trends in quantitative structure-activity relationship" In: Burger's Medicinal Chemistry and Drug Discovery. (6^{th} ed.), Vol 1: Drug Discovery. John Wiley & Sons. 2003, 49-77.

62. Cramer, R.D. 3^{rd}, Patterson, D.E., Bunce, J.D. Recent advances in comparative molecular field analysis (CoMFA). *Prog Clin Biol Res* 1989, 291: 161-165.

63. Klebe, G., Abraham, U., Mietzer, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 1994, 37: 4130-4146.

64. Ragno, R., Simeoni, S., Valente, S., Massa, S., Mai, A. 3D QSAR studies on histone deacetylase inhibitors. A GOLPE/GRID approach on different series of compounds. *J Chem Inf Model* 2006, 46: 1420-1430.

65. Pratim Roy, P., Paul, S., Mitra, I., Roy, K. On two novel parameters for validation of predictive QSAR models. *Molecules* 2009, 14: 1660-1701.

66. Lesburg, C.A., Cable, M.B., Ferrari, E., Hong, Z., Mannarino, A.F., Weber, P.C. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol* 1999, 6: 937-943.

67. Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., Stuart, D.I. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* 2001, 410: 235-240.

68. Poch, O., Sauvaget, I., Delarue, M., Tordo, N. Identification of four conserved motifs among the RNA- dependent polymerase encoding elements. *EMBO J* 1989, 8: 3867-3874.

69. Carroll, S.S., Tomassini, J.E., Bosserman, M., Getty, K., Stahlhut, M.W., Eldrup, A.B., Bhat, B., Hall, D., Simcoe, A.L., LaFemina, R., Rutkowski, C.A., Wolanski, B., Yang, Z., Migliaccio, G., De Francesco, R., Kuo, L.C., MacCoss, M., Olsen, D.B. Inhibition of hepatitis C virus RNA replication by 2'-modified nucleoside analogs. *J Biol Chem* 2003, 278: 11979-11984.

70. Clark, J.L., Hollecker, L., Mason, J.C., Stuyver, L.J., Tharnish, P.M., Lostia, S., McBrayer, T.R., Schinazi, R.F., Watanabe, K.A., Otto, M.J., Furman, P.A., Stec, W.J., Patterson, S.E., Pankiewicz, K.W. Design, synthesis, and antiviral activity of 2'-deoxy-2'fluro-2'-C-methylcytidine, a potent inhibitor of hepatitis C virus replication. *J Med Chem* 2005, 48: 5504-5508.

71. Ismaili, H.M.A., Cheng, Y.X., Lavallee, J.F., Siddiqui, A., Storer, R. Method for the treatment or prevention of flavivirus infections using nucleoside analogues. *US Patent* 2002, 20020019363.

72. Varaprasad, C.V.N.S., Ramasamy, K.S., Girardet, J.L., Gunic, E., Lai, V., Zhong, W., An, H., Hong, Z. Synthesis of pyrrolo[2,3-d]pyrimidine nucleoside derivative as potential anti-HCV agents. *Bioorg Chem* 2007, 35: 25-34.

73. Ding, Y., Girardet, J.L., Smith, K.L., Larson, G., Prigaro, B., Wu, J.Z., Yao, N. Parrellel synthesis of 5-cyano-6-aryl-2- thiouracil derivatives as inhibitors for hepatitis c viral NS5B RNA-dependent RNA polymerase. *Bioorg Chem* 2006, 34: 26-38.

74. Yan, S., Appleby, T., Larson, G., Wu, J.Z., Hamatake, R., Hong, Z., Yao, N. Strutcure-based drug design of a novel thiazolone scaffold as HCV NS5B polymerase allosteric inhibitors. *Bioorg Med Chem Lett* 2006, 16: 5888-5891.

75. Puerstinger, G., Paeshuyse, J., Clercq, E.D., Neyts, J. Antiviral 2, 5-disubstituted imidazo[4,5-c]pyridines: anti-pestivirus to anti-hepatitis C virus activity. *Bioorg Med Chem lett* 2007, 17: 390-393.

76. Sriram, D., Bal, T.R., Yogeeswari, P. Design, synthesis and biological evalution of novel non-nucleoside HIV-1 reverse transcriptase inhibitors with broad-spectrum chemotherapeutic properties. *Bioorg Med Chem* 2004, 12: 5865-5873.

77. Hwang, D.R., Wu, Y.S., Chang, C.W., Lien, T.W., Chen, W.C., Tan, U.K., Hsu, J.T.A., Hsieh, H.P. Synthesis and anti-viral activity of a series of sesquiterpene lactones in the sub genomic HCV replicon system. *Bioorg Med Chem* 2006, 14: 83-91.

78. Hwu, J.R., Singha, R., Hong, S.C., Chang, Y.H., Das, A.R., Vliegen, I., De Clercq, E., Neyts, J. Synthesis of new benzimidazole-coumarin conjugates as anti-hepatitis C virus agents. *Antiviral Res* 2008, 77: 157-162.

79. Kim, J., Kim, K.S., Lee, H.S., Park, K.S., Park, S.Y., Kang, S.Y., Lee, S.J., Park, H.S., Kim, D.E., Chong, Y. Effects of the aryl linker and the aromatic substituent on the anti-HCV activities of aryl diketoacid (ADK) analogues. *Bioorg Med Chem Lett* 2008, 18: 4661-4665.

80. Slater, M.J., Amphlett, E.M., Andrews, D.M., Bravi, G. Optimization of novel acyl pyrrolidine inhibitors of hepatitis C virus RNA-Dependent RNA polymerase leading to a development candidate. *J Med Chem* 2007, 50: 897-900.

81. Lee, J.C., Chen, W.C., Wu, S.F., Tseng, C., Chiou, C.Y., Chang, F.R., Hsu, S., Wu, Y.C. Anti-hepatitis C virus activity of *Acacia Confusa* extract via suppressing cyclooxygenase-2. *Antiviral Res* 2011, 89: 35-42.

82. Harper, S., Pacini, B., Avolio, S., Filippo, M.D., Migliaccio, G., Laufer, R., Francesco, R.D., Rowley, M., Narjes, F. Development and preliminary optimization of indole-N-acetamide inhibitors of hepatitis C virus NS5B polymerase. *J Med Chem* 2005, 48: 1314-1317.

83. Yan, S., Appleby, T., Larson, G., Wu, J.Z., Hamatake, R.K., Hong, Z., Yao, N. Thiazolone-acylsulfonamides as novel HCV NS5B polymerase allosteric inhibitors: convergence of structure based drug design and X-ray crystallographic study. *Bioorg Med Chem Lett* 2007, 17: 1991-1995.

84. Melagraki, G., Afantitis, A., Sarimveis, H., Koutentis, P.A.,Markopoulos, J., Markopoulou, O.I. Identification of a series of novel derivatives as potent HCV inhibitors by a ligand based virtual screening optimized procedure. *Bioorg Med Chem* 2007, 15: 7237-7247.

85. Talele, T.T., Arora, P., Kulkarni, S.S., Patel, M.R., Singh, S., Chudayeu, M., Kaushik-Basu, N. Structure based virtual screening, synthesis and SAR of novel inhibitors of hepatitis C virus NS5B polymerase. *Bioorg Med Chem* 2010, 18: 4630-4638.

86. Patil, V.M., Gupta, S.P., Samanta, S. A QSAR study on some series of anti-hepatitis C virus (HCV) agents. *Lett Drug Design Disc* 2010, 7: 139-148.

87. Gupta, S.P., Samanta, S., Patil, V.M. A 3D QSAR study on a series of benzimidazole derivatives acting as hepatitis C virus inhibitors: Application of kNN-Molecular field analysis. *Med Chem* 2010, 6: 87-90.

88. Gupta, S.P., Samanta, S., Patil, V.M., Masand, N. 3D QSAR kNN-MFA studies on thiouracil derivatives as hepatitis c virus inhibitors. *Med Chem Res* 2011, 20: 1616-1621.

89. Patel, P.D., Patel, M.R., Kaushik-Basu, N., Talele, T.T. 3D QSAR and molecular docking studies of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors. *J Chem Inf Model* 2008, 48: 42-55.

90. Musmuca, I., Caroli, A., Mai, A., Kaushik-Basu, N., Ragno, R. Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *J Chem Inf Model* 2010, 50: 662-676.

91. Wang, X., Yang, W., Xu, X., Zhang, H., Li, Y., Wang, Y. Studies of benzothiadiazine derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D QSAR, molecular docking and molecular dynamics. *Curr Med Chem* 2010, 17: 2788-2803.

92. Zhang, H.X., Li, Y., Wang, X., Xiao, Z.T., Wang, Y.H. Insight into the structural requirements of benzothiadiazine scaffold-based derivatives as hepatitis C virus NS5b

polymerase inhibitors using 3D QSAR, molecular docking and molecular dynamics. *Curr Med Chem* 2011, 18: 4019-4028.

93. Yu, H., Fang, Y., Lu, X., Liu, Y., Zhang, H. Combined 3D QSAR, molecular docking, molecular dynamics simulation, and binding free energy calculation studies on the 5-hydroxy-2H-pyridazin-3-one derivatives as HCV NS5B polymerase inhibitors. *Chem Biol Drug Des* 2014, 83: 89-105.
94. Hucke, O., Coulombe, R., Bonneau, P., Bertrand-Laperle, M., Brochu, C., Gillard, J., Joly, M.A., Landry, S., Lepage, O., Llinas-Brunet, M., Pesant, M., Poirier, M., McKercher, G., Marquis, M., Kukolj, G., Beaulieu, P.L., Stammers, T.A. Molecular dynamics simulations and structure based rational design lead to allosteric HCV NS5B polymerase thumb pocket 2 inhibitor with picomolecular cellular replicon potency. *J Med Chem* 2014, 57: 1932-1943.
95. Jin, G., Lee, S., Chio, M., Son, S., Kim, G.W., Oh, J.W., Lee, C., Lee, K. Chemical genetics-based discovery of indole derivatives as HCV NS5b polymerase inhibitors. *Eur J Med Chem* 2014, 75: 413-425.
96. Barlett, J.A., Demasi, R., Quinn, J., Moxham, C., Rousseau, F. Overview of the effectiveness of triple combination therapy in antiretroviral-naïve HIV-1 infected adults. *AIDS* 2001, 15: 1369-1377.
97. Navia, M.A., Fitzgerald, P.M., McKeever, B.M., Leu, C.T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P. Three- dimensional structure of aspatyl protease from human immunodeficiency virus HIV-1. *Nature* 1989, 337: 615-620.
98. http://www.google.co.in/search?q=Vishal+Prashar+LIFE01200604006 browsed on 15/03/14.

99. Pettit, S.C., Moody, M.D., Wehbie, R.S., Kaplan, A.H., Nantermet, P.V., Klein, C.A., Swanstrom, R. The p2 domain of human immunodeficiency virus type 1 gag regulates sequential proteolytic processing and its required to produce fully infectious virions. *J Virol* 1994, 68: 8017-8027.

100. Brunger, A.T., Adams, P.D., Clorem, G.M., Delano, W.L., Gros, P., Gross-Kunstleve, R.W., Jiang, J-S., Kuszewski, J., Nilges, N., Pannu, N.S., Read, R.J., Rice, L.M. Simonson, T., Warren, G.L. Crystallography & NMR system (CNS), a new software suit for macromolecular structure determination. *Acta Crystalloogr D Biol Crystallogr* 1998, 54: 905-921.

101. De clercq, E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J Antimicrob Agents* 2009, 33: 307-320.

102. Piacenti, F.J. An update and review of antiretroviral therapy. *Pharmacother* 2006, 26: 1111-1133.

103. Vella, S., Floridia, M. Saquinavir clinical pharmacology and efficacy. *Clin Pharmacokinet* 1998, 34: 189-201.

104. Hsu, A., Granneman, G.R., Bertz, R.J. Ritonavir: clinical pharmacokinetics and interaction with other anti-HIV agents. *Clin Pharmacokinet* 1998, 35: 275-291.

105. Cressey, T.R., Plipat, N., Fregonese, F., Chokephaibulkit, K. Indinavir/ritonavir remains an important component of HAART for the treatment of HIV/AIDS, particularly in resource-limited settings. *Expert Opin Drug Metab Toxicol* 2007, 3: 347-361.

106. Olmo, M., Podzamczer, D. A review of nelfinavir for the treatment of HIV infection. Expert *Opin Drug Metab Toxicol* 2006, 2: 285-300.

107. Perry, C.M., Frampton, J.E., McCormack, P.L., Siddiqui, M.A., Cvetkovi, R.S. Nelfinavir: a review of its use in the management of HIV infection. *Drugs* 2005, 65: 2209-2244.

108. Noble, S., Goa, K.L. Amprenavir: a review of its clinical potential in patients with HIV infection. *Drugs* 2000, 60: 1383-1410.

109. Wensing, A.M., van Maarseveen, N.M., Nijhuis, M. Fifteen years of HIV protease inhibitors: raising the barrier to resistance. *Antiviral Res* 2010, 85: 59-74.

110. Fernandez-Montero, J.V., Barreiro, P., Soriano, V. HIV protease inhibitors: recent clinical trials and recommendations on use. *Expert Opin Pharmacother* 2009, 10: 1615-1629.

111. Boutton, C.W., De Bondt, H.L., De Jonge, M.R. Genotype-dependent QSAR for HIV-1 protease inhibition. *J Med Chem* 2005, 48: 2115-2120.

112. Surleraux, D.L., de Kock, H.A., Verschueren, W.G., Pille, G.M., Maes, L.J., Peeters, A., Vendeville, S., De Meyer, S., Azijn, H., Pauwels,R., De Betune, M.P., King, N.M., Prabu Jayabalan, M., Schiffer, C.A., Wigerinck, P.B. Design of HIV-1 protease inhibitors active on multidrug-resistant virus. *J Med Chem* 2005, 48: 1965-1973.

113. Ghosh, A.K., Sridhar, P.R., Leshchenko, S., Hussain, A.K., Li, j., Kovalvsky, A.Y., Walters, D.E., Wedekind, J.E., Grum-Tokars, V., Das, D., Koh, Y., Maeda, K., Gatanaga, H., Weber, I.T., Mitsuya, H. Structure based design of novel HIV-1 protease inhibitors to combat drug resistance. *J Med Chem* 2006, 49: 5252-5261.

114. Ali, A., Reddy, G.S., Cao, h., Anjum, S.G., Nalam, M.N., Schiffer, C.A., Rana, T.M. Discovery of HIV-1` protease inhibitors with picomolar affinities incorporating N-aryl-oxazolidinone-5-carboxamides as novel p2 ligands. *J Med Chem* 2006, 49: 7342-7356.

115. Ragno, R., Coluccia, A., La Regina, G., De Martino, G., Piscitelli, F., Lavecchia, A., Novellino, E., Bergamini, A., Ciaprini, C., Sinistro, A., maga, G., Crespan, E., Artico, M.,

Silvestri, R. Design, molecular modeling, synthesis, and anti-HIV-1 activity of new indolyl aryl sulfones novel derivatives of the indole-2-carboxamide. *J Med Chem* 2006, 49: 3172-3184.

116. Volrath, P., Harrison, R.W., Weber, I.T. Structure based drug design for HIV protease: From molecular modeling to cheminformatics. *Curr Top Med Chem* 2007, 7: 1030-1038.
117. Ghosh. A.K., Kulkarni, S., Anderson, D.D., Hong, L., Balbridge, A., Wang, Y.F., Chumanevich, A.A., Kovalevsky, A.Y., Tojo, Y., Amano, m., Koh, Y., Tang, J., Weber, I.T., Mitsuya, H. Design, synthesis, protein-ligand X-ray structures and biological evaluation of a series of novel macrocyclic HIV-1 protease inhibitors to combat drug resistance. *J Med Chem* 2009, 52: 7689-7705.
118. Ali, A., Reddy, G.S., Nalam, M.N., Anjum, S.G., Cao, H., Schiffer, C.A., Rana, T.M. Structure based design, synthesis, and structure-activity relationship studies of HIV-1 protease inhibitors incorporating phenyloxazolidinones. *J Med Chem* 2010, 53: 7699-7708.
119. Ganguly, A.K., Alluri, S.S., caroccia, D., Biswas, D., Wang, C.H., Kang, E., Zhang, Y., McPhail, A.T., Carroll, S.S., Burlein, C., Munshi, V., Orth, P., Strickland, C. Design, synthesis, and X-ray crystallographic analysis of a novel class of HIV-1 protease inhibitors. *J Med Chem* 2011, 54: 7176-7183.
120. Yadav, D., Paliwal, S., Yadavm, R., Pal, M., Pandey, A. Identification of novel HIV-1 protease inhibitors: application of ligand and structure based pharmacophore mapping and virtual screening. *PLoS ONE* 2012, 7: e48942.
121. Joshi, A., Veron, J.B., Unge, j., Rosenquist, A., Wallberg, h., Samuelsson B., Hallberg, A., Larhed, M. Design and synthesis of p1-p3 macrocyclic tertiary-alcohol-comprising HIV-1 protease inhibitors. *J Med Chem* 2013, 56: 8999-9007.

122. Tiefenbrunn, T., Forli, S., Happer, M., Gonzalex, A., Tsai, Y., Soltis, M., Elder, J.H., Olson, A.J., Stout, C.D. Crystallographic fragment-based drug discover: use of a brominated fragment library targeting HIV protease. *Chem Biol Drug Des* 2014, 83: 141-148.

123. Pang, X., Liu, Z., Zhai, G. Advances in non-peptidimimetric HIV protease inhibitors. *Curr Med chem* 2014, [Epup ahead of print]

124. Autunes, D.A., RIgo, M.M., Sinigagalia, M., de Medeiros, R.M., Jungueira, D.M., Almeida, S.E., Vieira, G.F. New insights into the in silico prediction of HIV protease resistance to nelfinavir. *PLoS ONE* 2014, 9: e87520.

125. Yedida, R.S., Liu, Z., Kovaria, I.A., Wosterb, P.M., Kovaria, L.C. P1 and P1'-para-fluro phenyl groups show enhanced binding and favorable predicted pharmacological properties: structure based virtual screening of extended lopinavir analogs against multi-drug resistant HIV-1 protease. *J Mol graph Model* 2014, 47: 18-24.

126. Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C., Bourne, Y. Crystal structure of the bifunctional N-acetylglucosamine 1-phosphate uridyltransferase from Escherichia coli: a paradigm for the related pyrophosphorylase super family. *EMBO J* 1999, 18: 4096-4107.

127. Kostrewa, D., D'Arcy, A., Takacs, B., Kamber, M. Crystal structures of Str eptococcus pnemoniae N-acetylglucosamine-1-phosphate uridyltransferase, GlmU, In Apo form at 2.33Å resolution and in complex with UDP-N-Acetylglucosamine and Mg2+ at 1.96Å resolution. *J Mol Biol* 2001, 305: 279-289.

128. Mochalkin, I., Lightle, S., Zhu, Y., Ohren, J.F., Spessard, C., Chirgadze, N.Y., Banotai, C., Melnick, M., Mcdowell, L. Characterization of substrate binding and catalysis in the

potential antibacterial target N-acetylglucosamine-1-phosphate uridyltransferase (GlmU). *Protein Sci* 2007, 16: 2657-2666.

129. Zhang, Z., Bulloch, E.M., Bunker, R.D., Baker, E.N., Squire, C.J. Structure and function of GlmU from Mycobacteriu tuberculosis. *Acta Crystallogr D Biol Crystallogr* 2009, 65: 275-283.

130. Mengin-Lecreulx, D., van Heijenoort, J. Idetification of the GlmU gene encoding N-acetylglucosamine-1-phosphate uridyltransferase in Escherichia Coli. *J Bacteriol* 1993, 175: 6150-6157.

131. Milewski, S.I., Gabriel, I., Olchowy, J. Enzymes of UDP-GlcNAc biosynthesis in yeast. *Yeast* 2006, 23: 1-14.

132. Pereira, M.P., Jan, J.E., Murphy, B.C., Roderick, S.L., Brown, E.D. High-throughput screening identifies novel inhibitors of the acetyltransferase activity of Escherichia Coli GlmU. *Antimicrob Agents Chemother* 2009, 53: 2306-2311.

133. Li, Y., Zhou, Y., Mac, Y., Li, X. Design and synthesis of novel cell wall inhibitors of Mycobacterium tuberculosis GlmM and GlmU. *Carbohydr Res* 2011, 346: 1714-1720.

134. Buurman, E.D., Andews, B., Gao, N., Hu, J., Keating, T.A., Lahiri, S., Otterbein, L.R., Patten, A.D., Stokes, S.S., Shapiro, A.B. In Vitro validation of acetyltransferase activity of GlmU as an antibacterial target in Haemophilus influenza. *J Biol Chem* 2011, 286: 40734-40742.

135. Zhou, Y., Xin, Y., Sha, S., Ma, Y. Kinetic properties of Mycobacterium tuberculosis bifunctional GlmU. *Arch Microbiol* 2011, 193: 751-757.

136. Singla, D., Anura, M., Dash, D., Raghava, G.P. A web server for predicting inhibitors against bacterial target GlmU protein. *BMC Pharmacol* 2011, 11: 5.

137. Jagtap, P.K., Verma, S.K., Vithani, N., Bais, V.S., Prakash, B. Crystal structures identify an aptical two-metal-ion mechanism for uridyltransfer in GlmU; its significance to sugar nucleotidyl transferases. *J Mol Biol* 2013, 425: 1745-1759.

138. Tran, A.T., Wen, D., West, N.P., Baker, E.N., Brittonc. W.J., Payne, R.J. Inhibition studies on Mycobacterium tuberculosis N-acetylglucosamine-1-phosphate uridyltransferase (GlmU). *Org Biomol Chem* 2013, 11: 8113-8126.

139. Larsen, N.A., Nash, T.J., Moringstar, M., Shapiro, A.B., Joubran, C., Blackett, C.J., Patten, A.D., Boriack-Sjodin, P.A., Doig, P. An aminoquinazoline inhibitors of the essential bacterial cell wall synthetic enzyme GlmU has a unique non-protein-kinase binding mode. *Biochem J* 2012, 446: 405-413.

140. Schrödinger, L.L.C., New York, www.schrodinger.com.

141. Maestro, version 9.2, 2011, Schrodinger, LLC, NewYork, NY.

142. Impact, version 5.8, 2011, Schrodinger, LLC, New York, NY

143. Glide, version 5.8, 2011, Schrodinger, LLC, New York, NY.

144. Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knol, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., Shenkin, P.S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004, 47: 1739–1749.

145. Kawatkar, S., Wang, H., Czerminski, R., Joseph-McCarthy, D. Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. *J Comput Aided Mol Des* 2009, 23: 527-539.

146. Phase 3.4, 2011, Schrodinger, LLC, New York, NY.

147. Dixon, S.L., Smondyrev, A.M., Knoll, E.H., Rao, S.N., Shaw, D.E., Friesner, R.A. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des* 2006, 20, 647 – 671.

148. Dixon, S.L., Smondyrev, A.M., Rao, S.N. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 2006, 67: 370-372.

149. Loving, K., Salam, N.K., Sherman, W. Energetic analysis of fragment docking and application to structure based pharmacophore hypothesis generation. *J Comput Aided Mol Des* 2009, 23: 541-554.

150. LigPrep, version 2.5, 2011, Schrodinger, LLC, New York, NY.

151. Canvas, version 1.4, 2011, Schrodinger, LLC, New York, NY.

152. MacroModel, version 9.9, 2011, Schrödinger, LLC, New York, NY.

153. Lewis, R.A. A general method for exploiting QSAR models in lead optimization. *J Med Chem* 2005, 48: 1638-1648.

154. Roy, P.P., Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 2008, 27: 302-313.

155. Battu, M.B., Chandra, A.M., Sriram, D., Yogeeswari, P. Pharmacophore-Based 3D QSAR and Molecular Docking Studies to Identify New Non-Peptidic Inhibitors of Cathepsin S. *Curr Med Chem* 2013, [Epub ahead of print]

156. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P. Random Forest: A classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 2003, 43: 1947-1958.

157. Stanton, D.T. On the physical interpretation of QSAR models. *J Chem Inf Comput Sci* 2003, 43: 1423-1433.

158. Schüürmann, G., Ebert, R.U., Chen, J., Wang, B., Kühne, R. External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs. training set activity mean. *J Chem Inf Model* 2008, 48: 2140-2145.

159. Consonni, V., Ballabio, D., Todeschini, R. Comments on the definition of the Q2 parameter for QSAR validation. *J Chem Inf Model* 2009, 49: 1669-1678.

160. Golbraikh, A., Tropsha, A. Beware of q2. *J Mol Graph Model* 2002, 20, 269–276.

161. Roy, K., Roy, P.P. Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *Eur J Med Chem* 2009, 44: 2913-2922.

162. Li, Y., Wang, Y., Zhang, F. Pharmacophore modeling and 3D QSAR analysis of phosphoinositide 3-kinase p110 α inhibitors. *J Mol Model* 2010, 16: 1449–1460.

163. Sakkiah, S., Meganathan, C., Sohn, Y.S., Namadevan, S., Lee, K.W. Identification of important chemical features of 11β-hydroxysteroid dehydrogenase type1 inhibitors: Application of ligand based virtual screening and density functional theory. *Int J Mol Sci* 2012, 13: 5138-5162.

164. QikProp, version 3.4, 2011, Schrodinger, LLC, New York, NY.

165. Kim, K., Kim, K.H., Kim, H.Y., Cho, H.K., Sakamoto, N., Cheong, J. Curcumin inhibits hepatitis C virus replication via suppressing the Akt-SREBP-1 pathway. *FEBS Lett* 2010, 584: 707-712.

166. Nichols, D.B., Fournet, G., Gurukumar, K.R., Basu, A., Lee, J.C., Sakamoto, N., Kozielski, F., Musmuca, I., Joseph, B., Ragno, R., Kaushik-Basu, N. Inhibition of hepatitis

C virus NS5B polymerase by S-trityl-L-cysteine derivatives. *Eur J Med Chem* 2012, 49: 191-199.

167. Kaushik-Basu, N., Bopda-Waffo, A., Talele, T.T., Basu, A., Costa, P.R., da Silva, A.J., Sarafianos, S.G., Noel, F. Identification and characterization of coumestans as novel HCV NS5B polymerase inhibitors. *Nucleic Acids Res* 2008, 36: 1482-1496.

168. Chen, Y., Bopda-Waffo, A., Basu, A., Krishnan, R., Silberstein, E., Taylor, D.R., Talele, T.T., Arora, P., Kaushik-Basu, N. Characterization of aurintricarboxylic acid as a potent hepatitis C virus replicase inhibitor. *Antivir Chem Chemother* 2009, 20: 19-36.

169. Porter, D.J., Hanlon, M.H., Carter, L.H.3rd, Danger, D.P., Furfine, E.S. Effectors of HIV-1 protease peptidolytic activity. *Biochem* 2001, 40: 11131-11139.

170. Schneider, J., Kent, S.B. Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. *Cell* 1988, 54: 363-368.

171. Gehringer, H., Von der Helm, K., Seelmeir, S., Weissbrich, B., Eberle, J., Nitschko, H. Development and evalution of a phenotypic assay monitoring resistance formation to protease inhibitors in HIV-1-infected patients. *J Virol Methods* 2003, 109: 143-152.

172. Gehring, A.M., Lees, W.J., Mindiola, D.J., Walsh, C.T., Brown, E.D. Acetyltransfer precedes uridyltransfer in the formation of UDP-N- acetylglucosamine in separable active sites of the bifunctional GlmU protein of Escherichia Coli. *Biochem* 1996, 35: 579-585.

173. Zhang, W., Jones, V.C., Scherman, M.S., Mahapatra, S., Crick, D., Bhamidi, S., Xin, Y., McNeil, M.R., Ma, Y. Expression, essentiality, and a microtiter plate assay for mycobacterial GlmU, the bifunctional glucosamine-1-phosphate acetyltransferase and N-acetylglucosamine-1-phosphateuridyltransferase. *Int J Biochem Cell Biol* 2008, 40: 2560-2571.

174. Manns, M.P., Foster, G.R., Rockstroh, J.K., Zeuzem, S., Zoulim, F., Houghton, M. The way forward in HCV treatment-finding the right path. *Nat Rev Drug Discov* 2007, 6: 991-1000.

175. Sarrazin, C., Zeuzem, S. Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterlogy* 2010, 138: 447-462.

176. Ruebsam, F., Sun, Z., Ayida, B.K., Webber, S.E., Zhou, Y., Zhao, Q., Kissinger, C.R., Showalter, R.E., Shah, A.M., Tsan, M., Patel, R., Lebrun, L.A., Kamran, R., Sergeeva, V., Bartkowski, D.M., Nolan, T.G., Norris, D.A., Kirkovsky, L. Hexahydropyyrolo-and hexa hydro-1H-pyrido[1,2-b]pyridazin-2-ones as potent inhibitors of HCV NS5B polymerase. *Bioorg Med Chem Lett* 2008, 18: 5002-5005.

177. Ruebsam, F., Webber, S.E., Tran, M.T., Tran, C.V., Murphy, D.E., Zhao, J., Dragovich, P.S., Kim, S.H., Li, L.S., Zhou, Y., Han, Q., Kissinger, C.R., Showalter, R.E., Lardy, M., Shah, A.M., Tsan, M., Patel, R., Lebrun, L.A., Kamran, R., Sergeeva, M.V., Barkowski, D.M., Nolan, T.G., Norris, D.A., Kirkovsky, L. Pyrrolo[1,2-b]pyridazin-2-ones as potent inhibitors of HCV NS5B polyme4rase. *Bioorg Med Chem Lett* 2008, 18: 3616-3621.

178. Maynard, A., Crosby, R.M., Ellis, B., Hamatake, R., Hong, Z., Johns, B.A., Kahler, K. M., Koble, C., Leivers, A.L., Leivers, M.R., Mathis, A., Peat, A.J., Pouliot, J.J., Roberts, C.D., Samano, V., Schmidt, R.M., Smith, G.K., Spaltenstein, A., Stewart, E.L., Thommes, P., Turner, E.M., Voitenleitner, C., Walker, J.T., Waitt, G., Weatherhead, J., Weaver, K.L., Williams, S., Wright, L., Xiong, Z.Z., Haigh, D., Shotwell, J.B. Discovery of a Potent boronic acid derived inhibitor of the HCV RNA-dependent RNA polymerase. *J Med Chem* 2013, 57: 1902-1913.

179. Biswal, B.K., Wang, M., Cherney, M.M., Chan, L., Yannopoulos, C.G., Bilimoria, D., Bedard, J., James, M.N. Non-nucleoside inhibitors binding to hepatitis C virus NS5B polymerase reveal a novel mechanism of inhibition. *J Mol Biol* 2006, 361: 33-45.

180. Bressanelli, S., Tomei, L., Rey, F. A., De Francesco, R. Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J Virol* 2002, 76: 3482-3492.

181. Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De Francesco, R., Rey, F.A. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc Natl Acad Sci* U. S. A. 1999, 96: 13034-13039.

182. Le Pogam, S., Kang, H., Harris, S.F., Leveque, V., Giannetti, A.M., Ali, S., Jiang, W., Rajyaguru ,S., Tavares, G., Oshiro, C., Hendricks, T., Klumpp, K., Symons, J., Browner, M.F., Cammack, N., Najera, I. Selection and characterization of replicon variants dually resistant to thumb-and palm-binding nonnucleoside polymerrase inhibitors of the hepatitis C virus. *J Virol* 2006, 80: 6146-6154.

183. Truchon, J.F., Bayly, C.I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J Chem Inf Model* 2007, 47: 488-508.

184. Gopalsamy, A., Chopra, R., Lim, K., Ciszewski, G., Shi, M., Curran, K.J., Sukits, S.F., Svenson, K., Bard J., Ellingboe, J.W., Agarwal, A., Krishnamurthy, G., Howe, A.Y., Field, B., O'Connell, J., Mansour, T.S. Discovery of proline sulfonamides as potent and selective hepatitis C virus NS5b polymerase inhibitors. Evidence for a new NS5b polymerase binding site. *J Med Chem* 2006, 49: 3052-3055.

185. Beaulieu, P.L., Bos, M., Bousquet, Y., DeRoy, P., Fazal, G., Gauthier, J., Gillard, J., Goulet, S., McKercher G., Poupart, M. Non-nucleoside inhibitors of the hepatitis C virus

NS5B polymerase: discovery of benzimidazole 5-carboxylic amide derivatives with low nanomolar potency. *Bioorg Med Chem Lett* 2004, 14: 967-971.

186. Beaulieu, P.L., Bos, M., Bousquet, Y., Fazal, G., Gauthier, J., Gillard, J., Goulet, S., LaPlante, S., Poupart, M., Lefebvre, S., McKercher, G., Pellerin, C., Austel, V., Kukolj, G. Non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase: discovery and preliminary SAR of benzimidazole derivatives. *Bioorg Med Chem Lett* 2004, 14: 119-124.

187. Burton, G., Ku, T.W., Carr, T.J., Kiesow, T., Sarisky, R.T., Lin-Goerke, J., Baker, A., Earnshaw, D.L., Hofmann G.A., Keenan, R.M., Dhanak, D. Identification of small molecule inhibitors of the hepatitis C virus RNA-dependent RNA polymerase from a pyrrolidine combinatorial mixture. *Bioorg Med Chem Lett* 2005, 15: 1553-1556.

188. Chan, L., Pereira, O., Reddy, T.J., Das, S.K., Poisson, C., Courchesne, M., Proulx, M., Siddiqui, A., Yannopoulos C.G., Nguyen-Ba, N. Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 2: tertiary amides. *Bioorg Med Chem Lett* 2004, 14: 797-800.

189. Chan, L., Reddy, T.J., Proulx, M., Das, S.K. Pereira, O., Wang, W., Siddiqui, A., Yannopoulos, C.G. Poisson, C., Turcotte, N. Identification of N,N-disubstituted phenylalanines as a novel class of inhibitors of hepatitis C NS5B polymerase. *J Med Chem* 2003, 46: 1283-1285.

190. Crescenzi, B., Poma, M., Ontoria, J., Marchetti, A., Nizi, E., Matassa, V., Gardelli, C. Phenyldihydroxypyrimidines as HCV NS5B RNA dependent RNA polymerase inhibitors. Part I: Amides and ureas. *Lett Drug Design Discov* 2005, 2: 451-455.

191. Ding, Y., Girardet, J., Smith, K.L., Larson, G., Prigaro, B., Lai, V.C., Zhong, W., Wu, J.Z. Parallel synthesis of pteridine derivatives as potent inhibitors for hepatitis C virus NS5B RNA-dependent RNA polymerase. *Bioorg Med Chem Lett* 2005, 15: 675-678.

192. Ding, Y., Smith, K.L., Varaprasad, C.V., Chang, E., Alexander J., Yao, V. Synthesis of thiazolone-based sulfonamides as inhibitors of HCV NS5B polymerase. *Bioorg Med Chem Lett* 2007, 17: 841-845.

193. Donner, P.L., Xie, Q., Pratt, J.K., Maring, C.J., Kati, W., Jiang, W., Liu, Y., Koev, G., Masse S., Montgomery, V., Molla, A., Kempf, D.J. Des-A-ring benzothiadiazines: Inhibitors of HCV genotype 1 NS5B RNA-dependent RNA polymerase. *Bioorg Med Chem Lett* 2008, 18: 2735-2738.

194. Eldrup, A.B., Allerson, C.R., Bennett, C.F., Bera, S., Bhat, B., Bhat, N., Bosserman, M.R., Brooks, J., Burlein, C., Carroll, S.S., Cook, P.D., Getty, K.L., MacCoss, M., McMasters, D.R., Olsen, D.B., Prakash, T.P., Prhavc, M., Song, Q., Tomassini, J.E., Xia, J. Structure-activity relationship of purine ribonucleosides for inhibition of hepatitis C virus RNA-dependent RNA polymerase. *J Med Chem* 2004, 47: 2283-2295.

195. Evans, K.A., Chai, D., Graybill, T.L., Burton, G., Sarisky, R.T., Lin-Goerke, J., Johnston V.K., Rivero, R.A. An efficient, asymmetric solid-phase synthesis of benzothiadiazine-substituted tetramic acids: Potent inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *Bioorg Med Chem Lett* 2006, 16: 2205-2208.

196. Gopalsamy, A., Aplasca, A., Ciszewski, G., Park, K., Ellingboe, J.W., Orlowski, M., Feld, B., Howe, A.Y. Design and synthesis of 3, 4-dihydro-1H-[1]-benzothieno[2, 3-c]pyran and 3,4-dihydro-1H-pyrano[3,4-b]benzofuran derivatives as non-nucleoside inhibitors of HCV NS5B RNA dependent RNA polymerase. *Bioorg Med Chem Lett* 2006, 16: 457-460.

197. Gopalsamy, A., Lim, K., Ciszewski, G., Park, K., Ellingboe, J.W., Bloom, J., Insaf, S., Upeslacis, J., Mansour T.S., Krishnamurthy, G., Damarla, M., Pyatski, Y., Ho, D., Howe, A.Y., Orlowski, M., Feld, B., O'Connell, J. Discovery of pyrano[3,4-b] indoles as potent and selective HCV NS5B polymerase inhibitors. *J Med Chem* 2004, 47: 6603-6608.

198. Gopalsamy, A., Lim, K., Ellingboe, J.W., Krishnamurthy, G., Orlowski, M., Feld, B., van Zeijl, M., Howe, A.Y. Identification of [(naphthalene-1-carbonyl)-amino]-acetic acid derivatives as nonnucleoside inhibitors of HCV NS5B RNA dependent RNA polymerase. *Bioorg Med Chem Lett* 2004, 14: 4221-4224.

199. Gopalsamy, A., Shi, M., Ciszewski, G., Park, K., Ellingboe, J.W., Orlowski, M., Feld B., Howe, A.Y. Design and synthesis of 2, 3, 4, 9-tetrahydro-1H-carbazole and 1, 2, 3, 4-tetrahydro-cyclopenta [b] indole derivatives as non-nucleoside inhibitors of hepatitis C virus NS5B RNA-dependent RNA polymerase. *Bioorg Med Chem Lett* 2006, 16: 2532-2534.

200. Hirashima, S., Oka, T., Ikegashira, K., Noji, S., Yamanaka, H., Hara, Y., Goto, H., Mizojiri, R., Niwa, Y., Noguchi, T., Ando, I., Ikeda, S., Hashimoto, H. Further studies on hepatitis C virus NS5B RNA-dependent RNA polymerase inhibitors toward improved replicon cell activities: benzimidazole and structurally related compounds bearing the 2-morpholinophenyl moiety. *Bioorg Med Chem Lett* 2007, 17: 3181-3186.

201. Hirashima, S., Suzuki, T., Ishida, T., Noji, S., Yata, S., Ando, I., Komatsu, M., Ikeda S., Hashimoto, H. Benzimidazole derivatives bearing substituted biphenyls as hepatitis C virus NS5B RNA-dependent RNA polymerase inhibitors: structure-activity relationship studies and identification of a potent and highly selective inhibitor JTK-109. *J Med Chem* 2006, 49: 4721-4736.

202. Ishida, T., Suzuki, T., Hirashima, S., Mizutani, K., Yoshida, A., Ando, I., Ikeda, S., Adachi T., Hashimoto, H. Benzimidazole inhibitors of hepatitis C virus NS5B polymerase: identification of 2-[(4-diarylmethoxy) phenyl]-benzimidazole. *Bioorg Med Chem Lett* 2006, 16: 1859-1863.

203. Kim, S.H., Tran, M.T., Ruebsam, F., Xiang, A.X., Ayida, B., McGuire, H., Ellis, D., Blazel, J., Tran C.V., Murphy, D.E., Webber, S.E., Zhou, Y., Shah, A.M., Tsan, M., Showalter, R.E., Patel, R., Gobbi, A., LeBrun, L.A., Bartkowski, D.M., Nolan, T.G., Norris, D.A., Sergeeva, M.V., Kirkovsky, L., Zhao, Q., Han, Q., Kissinger, C.R. Structure based design, synthesis, and biological evaluation of 1,1-dioxoisothiazole and benzo[b]thiophene-1,1-dioxide derivatives as novel inhibitors of hepatitis C virus NS5B polymerase. *Bioorg Med Chem Lett* 2008, 18: 4181-4185.

204. Koch, U., Attenni, B., Malancona, S., Colarusso, S., Conte, I., Di Filippo, M., Harper, S., Pacini, B., Giomini C., Thomas, S., Incitti, I., Tomei, L., De Francesco, R., Altamura, S., Matassa, V.G., Narjes, F. 2-(2-Thienyl)-5, 6-dihydroxy-4-carboxypyrimidines as inhibitors of the hepatitis C virus NS5B polymerase: discovery, SAR, modeling, and mutagenesis. *J Med Chem* 2006, 49: 1693-1705.

205. Krueger, A.C., Madigan, D.L., Green, B.E., Hutchinson, D.K., Jiang, W.W., Kati, W.M., Liu, Y., Maring, C.J., Masse S.V., McDaniel, K.F., Middleton, T.R., Mo, H., Molla, A., Montgomery, D.A., Ng, T.I., Kempf, D.J. Inhibitors of HCV NS5B polymerase: Synthesis and structure activity relationships of unsymmetrical 1-hydroxy-4,4-dialkyl-3-oxo-3,4-dihydronaphthalene benzothiadiazine derivatives. *Bioorg Med Chem Lett* 2007, 17: 2289-2292.

206. LaPorte, M., Lessen, T., Leister, L., Cebzanov, D., Amparo, E., Faust, C., Ortlip, D., Bailey, T., Nitz, T., Chunduru, S.K., Young, D.C., Burns, C.J. Tetrahydrobenzothiophene inhibitors of hepatitis C virus NS5B polymerase. *Bioorg Med Chem Lett* 2006, 16: 100-103.

207. Li, H., Tatlock, J., Linton, A., Gonzalez, J., Borchardt, A., Dragovich, P., Jewell, T., Prins, T., Zhou, R., Blazel, J., Parge, H., Love, R., Hickey, M., Doan, C., Shi, S., Duggal, R., Lewis, C., Fuhrman, S. Identification and structure based optimization of novel dihydropyrones as potent HCV RNA polymerase inhibitors. *Bioorg Med Chem Lett* 2006, 16: 4834-4838.

208. Li, H., Tatlock, J., Linton, A., Gonzalez, J., Jewell, T., Patel, L., Ludlum, S., Drowns, M., Rahavendran, S.V., Skor, H., Hunter, R., Shi, S.T., Herlihy, K.J., Parge, H., Hickey, M., Yu, X., Chau, F., Nonomiya, J., Lewis, C. Discovery of (R)-6-cyclopentyl-6-(2-(2,6-diethylpyridin-4-yl)ethyl)-3-((5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2yl)methyl)-4-hydroxy-5,6 dihydro pyran-2-one (PF-00868554) as a potent and orally available hepatitis C virus polymerase inhibitor. *J Med Chem* 2009, 52: 1255-1258.

209. Liu, Y., Donner, P.L., Pratt, J.K., Jiang, W.W., Ng, T., Gracias, V., Baumeister, S., Wiedeman, P.E., Traphagen L., Warrior, U., Maring, C., Kati, W.M., Djuric, S.W., Molla, A. Identification of halosalicylamide derivatives as a novel class of allosteric inhibitors of HCV NS5B polymerase. *Bioorg Med Chem Lett* 2008, 18: 3173-3177.

210. Louise-May, S., Yang, W., Nie, X., Liu, D., Deshpande, M.S., Phadke, A.S., Huang M., Agarwal, A. Discovery of novel dialkyl substituted thiophene inhibitors of HCV by in silico screening of the NS5B RdRp. *Bioorg Med Chem Lett* 2007, 17: 3905-3909.

211. McGowan, D., Nyanguile, O., Cummings, M.D., Vendeville, S., Vandyck, K., Van den Broeck, W., Boutton, C.W., De Bondt, H., Quirynen, L., Amssoms, K., Bonfanti, J.F., Last,

S., Rombausts, K., Tahri, A., Hu, L., Delouvroy, F., Vermeiren, K., Vandercruyssen, G., Van der Helm, L., Cleiren, E., Mostmans, W., Lory, P., Pille, G., Van Emelen, K., Fanning, G., Pauwels, F., Lin, T.I., Simmen, K., Raboisson, P. 1,5-Benzodiazepine inhibitors of HCV NS5B polymerase. *Bioorg Med Chem Lett* 2009, 19: 2492-2496.

212. Nittoli, T., Curran, K., Insaf, S., DiGrandi, M., Orlowski, M., Chopra, R., Agarwal, A., Howe, A.Y., Prashad A., Floyd, M.B., Johnson, B., Sutherland, A., Wheless, K., Feld, B., O'Connell, J., Mansour, T.S., Bloom, J. Identification of anthranilic acid derivatives as a novel class of allosteric inhibitors of hepatitis C NS5B polymerase. *J Med Chem* 2007, 50: 2108-2116.

213. Ontoria, J.M., Martin Hernando, J.I., Malancona, S., Attenni, B., Stansfield, I., Conte, I., Ercolani, C., Habermann, J., Ponzi, S., Di Filippo, M., Koch, U., Rowley M., Narjes, F. Identification of thieno[3,2-b]pyrroles as allosteric inhibitors of hepatitis C virus NS5B polymerase. *Bioorg Med Chem Lett* 2006, 16: 4026-4030.

214. Ontoria, J.M., Rydberg, E.H., Di Marco, S., Tomei, L., Attenni, B., Malancona, S., Martin Hernando, J.I., Gennari, N., Koch, U., Narjes, F., Rowley, M., Summa, V., Carroll, S.S., Olsen, D.B., De Francesco, R., Alamura, S., Migilaccio G., Carfi, A. Identification and Biological Evaluation of a Series of 1H-benzo[de]isoquinoline-1,3(2H)-diones as Hepatitis C Virus NS5B Polymerase Inhibitors. *J Med Chem* 2009, 52: 5217-5227.

215. Pfefferkorn, J.A., Greene, M.L., Nugent, R.A., Gross, R.J., Mitchell, M.A., Finzel, B.C., Harris, M.S., Wells, P.A., Shelly, J.A., Anstadt, R.A., Kilkuskie, R.E., Kopta L.A., Schwende, F.J. Inhibitors of HCV NS5B polymerase.Part1:Evaluation of the southern region of (2Z)-2-(benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg Med Chem Lett* 2005, 15: 2481-2486.

216. Pfefferkorn , J.A., Nugent, R., Gross, R.J., Greene, M., Mitchell, M.A., Reding, M.T., Funk, L.A., Anderson, R., Wells, P.A., Shelly, J.A., Anstadt, R., Finzel, B.C., Harris, M.S., Kulkuskie, R.E., Kopta, L.A., Schwende, F.J. Inhibitors of HCV NS5B polymerase. Part 2: Evaluation of the northern region of (2Z)-2-benzoylamino-3-(4-phenoxy-phenyl)-acrylic acid. *Bioorg Med Chem Lett* 2005, 15: 2812-2818.

217. Powers, J.P., Piper, D.E., Li, Y., Mayorga, V., Anzola, J., Chen, J. M., Jaen, J.C., Lee, G., Liu, J., Peterson, M.G., Tonn, G.R., Ye, Q., Walker, N.P., Wang, Z. SAR and mode of action of novel non-nucleoside inhibitors of hepatitis C NS5b RNA polymerase. *J Med Chem* 2006, 49: 1034-1046.

218. Pratt, J.K., Donner, P., McDaniel, K.F., Maring, C.J., Kati, W.M., Mo, H., Middleton, T., Liu, Y., Ng, T., Xie, Q., Zhang, R., Montgomery, D., Molla, A., Kempf, D.J., Kohlbrenner, W. Inhibitors of HCV NS5B polymerase: synthesis and structure activity relationships of N-1-heteroalkyl-4-hydroxyquinolon-3-yl-benzothiadiazines. *Bioorg Med Chem Lett* 2005, 15: 1577-1582.

219. Rawal, R.K., Katti, S., Kaushik-Basu, N., Arora, P., Pan, Z. Non-nucleoside inhibitors of the hepatitis C virus NS5B RNA-dependant RNA polymerase: 2-aryl-3-heteroaryl-1,3-thiazolidin-4-one derivatives. *Bioorg Med Chem Lett* 2008, 18: 6110-6114.

220. Rong, F., Chow, S., Yan, S., Larson, G., Hong, Z., Wu, J. Structure activity relationship (SAR) studies of quinoxalines as novel HCV NS5B RNA-dependent RNA polymerase inhibitors. *Bioorg Med Chem Lett* 2007, 17: 1663-1666.

221. Ruebsam, F., Tran, C.V., Li, L., Kim, S.H., Xiang, A.X., Zhou, Y., Blazel, J.K., Sun, Z., Dragovich, P.S., Zhao, J., McGuire, H.M., Murphy, D.E., Tran, M.T., Stankovic, N., Ellis, D.A., Gobbi, A., Showalter, R.E., Webber, S.E., Shah, A.M., Tsan, M., Patel, R.A., Lebrun,

L.A., Hou, H.J., Kamran, R., Sergeeva, M.V., Bartkowski, D.M., Nolan, T.G., Norris, D.A., Kirkovsky, L. 5,6-Dihydro-1H-pyridin-2-ones as potent inhibitors of HCV NS5B polymerase. *Bioorg Med Chem Lett* 2009, 19: 451-458.

222. Shipps G.W. Jr, Deng, Y., Wang, T., Popovici-Muller, J., Curran, P.J., Rosner, K.E., Cooper, A.B., Girijavallabhan, V., Butkiewicz, N., Cable, M. Aminothiazole inhibitors of HCV RNA polymerase. *Bioorg Med Chem Lett* 2005, 15: 115-119.

223. Stansfield, I., Avolio, S., Colarusso, S., Gennari, N., Narjes, F., Pacini, B., Ponzi, S., Harper, S. Active site inhibitors of HCV NS5B polymerase. The development and pharmacophore of 2-thienyl-5,6-dihydroxypyrimidine-4-carboxylic acid. *Bioorg Med Chem Lett* 2004, 14: 5085-5088.

224. Stansfield, I., Ercolani, C., Mackay, A., Conte, I., Pompei, M., Koch, U., Gennari, N., Giuliano, C., Rowley, M., Narjes. F. Tetracyclic indole inhibitors of hepatitis C virus NS5B-polymerase. *Bioorg Med Chem Lett* 2009, 19: 627-632.

225. Summa, V., Petrocchi, A., Matassa, V.G., Taliani, M., Laufer, R., De Francesco, R., Altamura, S., Pace, P. HCV NS5b RNA-dependent RNA polymerase inhibitors: from α,ϒ-diketoacids to 4,5-dihydroxypyrimidine -or 3-methyl-5-hydroxy pyrimidinone carboxylic acids. Design and synthesis. *J Med Chem* 2004, 47: 5336-5339.

226. Warner, R., Larson, D.P., Beno, D.W., Bosse, T.D., Darbyshire, J.F., Gao, Y., Gates, B.D., He, W., Henry, R.F., Hernandez, L.E., Hutchinson, D,K., Jiang, W.W., Kati, W.M., Klein, L.L., Koev, G., Liu, J., Kohlbrenner, W., Krueger, A.C., Liu, J., Long, M.A., Maring, C.J., Masse, S.V., Middleton, T., Prat, J.K., Montgomery, D.A., Stuart, P., Molla, A., Kempf, D.J. Inhibitors of hepatitis C virus polymerase: Synthesis and biological characterization of

unsymmetrical dialkyl-hydroxynaphthalenoyl-benzothiadiazines. *J Med Chem* 2009, 52: 1659-1669.

227. Zhou, Y., Li, L.S., Dragovich, P.S., Murphy, D.E., Tran, C.V., Ruebsam, F., Webber, S.E., Shah, A.M., Tsan, M., Averill, A., Showalter, R.E., Patel, R., Han, Q., Zhao, Q., Hermann, T., Kissinger, C.R., Lebrun, L., Sergeeva, M.V. Novel HCV NS5B polymerase inhibitors derived from 4-(1',1'-dioxo-1',4'-dihydro-1'lambda(6)-benzo[1',2',4']thiadiazin-3'-yl)-5-hydroxy-2H-pyridazin-3-ones. Part 2: Variation of the 2-and 6-pyridazinone substituents. *Bioorg & Med Chem Lett* 2008, 18: 1419-1424.

228. Chan, L., Das, S.K., Reddy, T.J., Poisson, C., Proulx, M., Pereira, O., Courchesne, M., Roy, C., Wang W., Siddiqui, A., Yannopoulos, C.G., Nguyen-Ba, N., Labrecgue, D., Bethell, R., Hamel, M., Courtemanche-Asselin, P., L'Heureux, L., David, M., Nicolas, O., Brunette, S., Bilimoria, D., Bedard, J. Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 1: Sulfonamides. *Bioorg Med Chem Lett* 2004, 14: 793-796.

229. Burton, G., Ku, T.W., Carr, T.J., Kiesow, T., Sarisky, R.T., Lin-Goerke, J., Hofmann, G.A., Slater, M.J., Haigh, D., Dhanak, D., Johnson, V.K., Parry, N.R., Thommes, P. Studies on acyl pyrrolidine inhibitors of HCV RNA-dependent RNA polymerase to identify a molecule with replicon antiviral activity. *Bioorg Med Chem Lett* 2007, 17: 1930-1933.

230. Boden, D., Markowitz, M. Resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob Agents Chemother* 1998, 42: 2775-2783.

231. Condra, J.H., Schleif, W.A., Blahy, O.M., Gadryelski, L.J., Graham, D.J., Quintero, J.C., Rhodes, A., Robbins, H.L., Roth, E., Shivaprahash, M., Titus, D., Yang, T., Teppler, H.,

Squires, K.E., Deutsch, P.J., Emini, E.A. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* 1995, 374: 569-571.

232. Ala, P.J., Huston, E.E, Klabe, R.M., McCabe, D.D., Duke, J.L., Rizzo C.J., Korant, B.D., DeLoskey, R.J., Lam, P.Y.S., Hodge, N.C., Chang, C.H. Molecular basis of HIV-1 protease drug resistance; structural analysis of mutant protease complexed with cyclic urea inhibitors. *Biochem* 1997, 36: 1573-1580.
233. Tie, Y., Kovalevsky, A.Y., Boross, P., Wang, Y.F., Ghosh, A.K., Tozser, J., Harrison, R.W., Weber, I.T. Atomic resolution crystal structures of HIV-1 protease and mutants V82A and 184V with saquinavir. *Proteins* 2007, 67: 232-242.
234. Reddy, G.S., Ali, A., Nalam, M.N., Anjum, S.G., Cao, H., Nathans, R.S., Schiffer, C.A., Rana, T.M. Design and synthesis of HIV-1 protease inhibitors incorporating oxazolidinones as P2/P2' ligands in pseudosymmetric dipeptide isosteres. *J Med Chem* 2007, 50: 4316-4328.
235. Olajuyigbe, F.M., Demitri, N., Geremia, S. Investigation of 2-fold disorder of inhibitors and relative potency by crystallizations of HIV-1 protease in ritonavir and saquinavir mixtures. *Cryst Growth Des* 2011, 11: 4378-4385.
236. Ghosh, A.K., Gemma, S., Baldridge, A., Wang, Y.F., Kovalevsky, A.Y., Koh, Y., Weber, I.T., Mitsuya, H. Flexible cyclic ethers/polyethers as novel P2-ligands for HIV-1 protease inhibitors: design, synthesis, biological avalution, and protein-ligand X-ray studies. *J Med Chem* 2008, 51:6021-6033.
237. Brynda, J., Rezacova, P., Saskova, G.K., Kozisek, M., Konvalinka, J. Dimeric inhibitor of HIV-1 protease. DOI: 10.2210/pdb3t11/pdb

238. Rutenber, E., Fauman, E.B., Keenan, R.J., Fong, S., Furth, P.S., Ortiz de Montellano, P.R., Meng, E., Kuntz, I.D., DeCamp, D.L., Salto, R. Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure based drug design. *J Biol Chem* 1993, 268:15343-15346.

239. Tyndall J.D., Reid R.C., Tyssen D.P., Jardine D.K., Todd B., Passmore M., March D.R., Pattenden L.K., Bergman D.A., Alewood D., Hu S.H., Alewood P.F., Birch C.j., Martin J.L., Fairlie D.P. Synthesis, stability, antiviral activity, and protease-bound structures of substrate-mimicking constrained macroctyclic inhibitors of HIV-1 protease. *J Med Chem* 2000, 43: 3495-3504.

240. Hwang Y.S., Chmielewsi J. Development of low molecular weight HIV-1 protease dimerization inhibitors. *J Med Chem*. 2005, 48: 2239-2242.

241. Rocheblave L., Bihel F., De Michelis C., Priem G., Courcambeck J., Bonnet B., Chermann J.C., Kraus J.L. Synthesis and antiviral activity of new anti-HIV amprenavir bioisosteres. *J Med Chem*. 2002, 45: 3321-3324.

242. Nakatani S., Hidaka K., Ami E., Nakahara K., Sato A., Nguyen J.T., Hamada Y., Hori Y., Ohnishi N., Nagai A., Kimura T., Hayashi Y., Kiso Y. Combination of non-natural D-amino acid derivatives and allophenylnorstatine-dimethylthioproline scaffold in HIV protease inhibitors have high efficacy in mutant HIV. *J Med Chem* 2008, 51: 2992–3004.

243. Ghosh A.K. Harnessing nature's insight: design of aspartyl protease inhibitors from treatment of drug-resistant HIV to Alzheimer's disease. *J Med Chem* 2009, 52: 2163-2176.

244. Ghosh A.K., Leshchenko-Yashchuk S., Anderson D.D., Baldridge A., Noetzel M., Miller H.B., Tie Y., Wang Y.F., Koh Y., Weber I.T., Mitsuya H. Design of HIV-1 protease inhibitors with pyrrolidinones and oxazolidinones as novel P1'-ligands to enhance

backbone-binding interactions with protease: synthesis, biological evalution, and protein-ligand X-ray studies. *J Med Chem* 2009, 52: 3902-3914.

245. Wang T., Yin T., Zhang Z., Bender J.A., Yang Z., Zadjura L.M., D'Arienzo C. J., DiGiugno Parker D., Gesenberg C., Yamanaka G.A., Gong Y. F., Ho H. T., Fang H., Zhou N., McAuliffe B.V., Eggers B.J., Fan L., Nowicka-sans B., Dicker I.B., Gao Q., Colonno R.J., Lin P.F., Meanwell N.A., Kadow J.F. Inhibitors of human immunodeficiency virus type 1 (HIV-1) attachment.5. An evolution from indole to azaindoles leading to the discovery of 1-(4-benzoylpiperazine-1-yl)-2-(4,7-dimethoxy-1H-pyrrolo[2,3-c]pyrindin-3-yl)ethane-1,2-dione (BMS-488043), a drug candidate that demonstrates antiviral activity in HIV-1-infected subjects. *J Med Chem* 2009, 52: 7778-7787.

246. Ettmayer P., Billich A., Hecht P., Rosenwirth B., Gstach H., Paracyclophanes: a novel class of water-soluble inhibitors of HIV proteinase. *J Med Chem* 1996, 39: 3291-3299.

247. Wang S., Milne G.W., Yan X., Posey I.J., Nicklaus M.C., Graham L., Rice W.G. Discovery of novel, non-peptide HIV-1 protease inhibitors by pharmacophore searching. *J Med Chem* 1996, 39: 2047-2054.

248. Thaisrivongs S., Janakiraman M.N., Chong K.T., Tomich P.K., Dolak L.A., Turner S.R., Strohback J.W., Lynn J.C., Horng M.M., Hinshaw R.R., Watenpaugh K.D. Structure based design of novel HIV protease inhibitors: sulfonamide-containing 4-hydroxycoumarins and 4-hydroxy-2-pyrones as potent non-peptidic inhibitors. *J Med Chem* 1996, 39: 2400-2410.

249. Thaisrivongs S., Romero D.L., Tommasi R.A., Janakiraman M.N., Strohbach J.W., Turner S.R.,Biles C., Morge R.R., Johnson P.D., Aristoff P.A., Tomich P.K., Lynn J.C., Horng M.M., Chong K.T., Hinshaw R.R., Howe W.J., Finzel B.C., Watenpaugh K.D. Structure

based design of HIV protease inhibitors:5,6-dihydro-4-hydroxy-2-pyrones as effective, nonpeptidic inhibitors. *J Med Chem* 1996, 39: 4630-4642.

250. Kim C.U., McGee L.R., Krawczyk S.H., Harwood E., Harada Y., Swaminathan S., Bischofberger N., Chen M.S., Cherrington J.M., Xiong S.F., Griffin L., Cundy K.C., Lee A., Yu B., Gulnik S., Erickson J.W. New series of potent, orally bioavailable, non-peptidic cyclic sulfones as HIV-1 protease inhibitors. *J Med Chem* 1996, 39: 3431-3434.

251. Thaisrivongs S., Turner S.R., Strohbach J.W., Tommasi R.A., Johnson P.D., Aristoff P.A., Hinshaw R.R., Chong K.T., Lynn J.C., Horng M.M., Tomich P.K., Howe W.J., Watenpaugh K.D. Structure based design of HIV protease inhibitors: sulfonamide-containing 5,6-dihydro-4-hydroxy-2-pyrones as non-peptidic inhibitors. *J Med Chem* 1996, 39: 4349-4353.

252. Jadhav P.K., Ala P., Woerner F.J., Chang C.H., Garber S.S., Anton E.D., Bacheler L.T. Cyclic urea amides: HIV-1 protease inhibitors with low nanomolar potency against both wild type and protease inhibitor resistant mutants of HIV. J Med Chem 1997, 40: 181-191.

253. Debnath A.K. Three-dimensional quantitative structure-activity relationship study on cyclic urea derivatives as HIV-1 protease inhibitors: application of comparative molecular field analysis. *J Med Chem* 1999, 42: 249-259.

254. Mimoto T., Kato R., Takaku H., Nojima S., Terashima K., Misawa S., Fukazawa T., Ueno T., Sato H., Shintani M., Kiso Y., Hayashi H. Structure-activity relationship of small-sized HIV protease inhibitors containing allophenylnorstatine. *J Med Chem* 1999, 42: 1789-1802.

255. Boyer F.E., Vara Prasad J.V., Domagala J.M., Ellsworth E.L., Gaida C., Hagen S.E., Markoski L.J., Tait B.D., Lunney E.A., Palovsky A., Ferguson D., Graham N., Holler T., Hupe D., Nouhan C., Tummino P.J., Urumov A., Zeikus E., Zeikus G., Gracheck S.J.,

Sanders J.M., VanderRoest S., Brodfueher J., Iyer K., Sinz M., Gulnik S.V. 5,6-Dihydropyran-2-ones possessing various sulfonyl functionalities: potent nonpeptidic inhibitors of HIV protease. *J Med Chem* 2000, 43: 843-858.

256. Beaulieu P.L., Wernic D., Abraham A., Anderson P.C., Bogri T., Bousguet Y., Croteau G., Guse I., Lamarre D., Liard F., Paris W., Thibeault D., Pav S., Tong L. Potent HIV protease inhibitors containing a novel (hydroxyetyl) amide isostere. *J Med Chem* 1997, 40: 2164-2176.
257. Bouras A., Boggetto N., Benatalah Z., de Rosny E., Sicsic S., Reboud-Ravaux M. Design, synthesis, and evalution of conformationally constrained tongs, new inhibitors of HIV-1 protease dimerization. *J Med Chem* 1999, 42: 957-962.
258. Alfonso Y., Monzote L. HIV protease inhibitors: Effect on the opportunistic protozoan parasites. *Open Med Chem J* 2011, 5: 40-50.
259. Jayatilleke P.R., Nair A.C., Zauhar R., Welsh W.J. Computational studies on HIV-protease inhibitors: influence of calculated inhibitor-enzyme binding affinities on the ststistical quality of 3D QSAR CoMFA models. *J Med Chem* 2000, 43: 4446-4451.
260. Artico, M., Silvestri, R., Pagnozzi, E., Bruno, B., Novellino, E., Greco, G., Massa, S., Ettorre, A., Loi, A.G., Scintu, F., La Colla, P. Structure based design, synthesis, and biological evaluation of novel pyrrolyl aryl sulfones: HIV-1 non-nucleoside reverse transcriptase inhibitors active at nanomolar concentrations, *J Med Chem* 2000, 43: 1886-1891.
261. Ohrui, H., Kohgo, S., Kitano, K., Sakata, S., Kodama, E., Yoshimura, K., Matsuoka, M., Shigeta, S., Mitsuya, H. Syntheses of 4'-C-ethynyl-$_{\beta}$-D-arabino- and 4'-C-ethynyl-2'-deoxy-

beta-D-ribo-pentofuranosylpyrimidines and -purines and evaluation of their anti-HIV activity. *J Med Chem* 2000, 43: 4516-4525.

262. Hagen, S.E., Domagala, J., Gajda, C., Lovdahl, M., Tait, B.D., Wise, E., Holler, T., Hupe, D., Nouhan,C., Urumov, A., Zeikus, G., Zeikus, E., Lunney, E.A., Pavlovsky, A., Gracheck, S.J., Saunders, J., VanderRoest, S., Brodfuehrer, J. 4-Hydroxy-5,6-dihydropyrones as inhibitors of HIV protease: The effect of heterocyclic substituents at C-6 on antiviral potency and pharmacokinetic parameters. *J Med Chem* 2001, 44: 2319-2332.

263. Breccia, P., Boggetto, N., Pe´rez-Ferna´ndez, R., Van Gool, M., Takahashi, M., Rene, L., Prados, P., Badet, B., Reboud-Ravaux, M., de Mendoza, J. Dimerization inhibitors of HIV-1 protease based on a bicyclic guanidinium subunit. *J Med Chem* 2003, 46: 5196-5207.

264. Ghosh, A.K., Chapsal, B.D., Parham, G.L., Steffey, M., Agniswamy, J., Wang, Y.F., Amano, M., Weber, I.T., Mitsuya, H. Design of HIV-1 protease inhibitors with C3-substituted hexahydrocyclopentafuranyl urethanes as P2-ligands: synthesis, biological evaluation, and protein-ligand X-ray crystal structure. *J Med Chem* 2011, 54: 5890–5901.

265. Smith, A.B., Cantin, L.D., Pasternak, A., Guise-Zawacki, L., Yao, W., Charnley, A.K., Barbosa, J., Sprengeler, P.A., Hirschmann, R., Munshi, S., Olsen, D.B., Schleif, W.A., Kuo, L.C. Design, synthesis, and biological evaluation of monopyrrolinone-based HIV-1 protease inhibitors *J Med Chem* 2003, 46: 1831-1844.

266. Bonini, C., Chiummiento, L., De Bonis, M., Di Blasio, N., Funicello, M., Lupattelli, P., Pandolfo, R., Tramutola, F., Berti, F. Synthesis of new thienyl ring containing HIV-1 protease inhibitors: promising preliminary pharmacological evaluation against recombinant HIV-1 proteases. *J Med Chem* 2010, 53: 1451–1457.

267. Ghosh, A.K., Kincaid, J.F., Walters, D.E., Chen, Y., Chaudhuri, N.C., Thompson, W.J., Culberson, C., Fitzgerald, P.M., Lee, H.Y., McKee, S.P., Munson, P.M., Duong, T.T., Darke, P.L., Zugay, J.A., Schleif, W.A., Axel, M.G., Lin, J., Huff, J.R. Nonpeptidal P2 ligands for HIV protease inhibitors: structure based design, synthesis, and biological evaluation. *J Med Chem* 1996, 39: 3278-3290.

268. Dorsey, B.D., McDonough, C., McDaniel, S.L., Levin, R.B., Newton, C.L., Hoffman, J.M., Darke, P.L., Zugay-Murphy, J.A., Emini, E.A., Schleif, W.A., Olsen, D.B., Stahlhut, M.W., Rutkowski, C.A., Kuo, L.C., Lin, J.H., Chen, I.W., Stuart R. Michelson, S.R., Holloway, M.K., Huff, J.R., Vacca, J.P. Identification of MK-944a: A second clinical candidate from the hydroxylaminepentanamide isostere series of HIV protease inhibitors. *J Med Chem* 2000, 43: 3386-3399.

269. Ghosh, A.K., Chapsal, B.D., Baldridge, A., Steffey, M.P., Walters, D.E., Koh, Y., Amano, M., Mitsuya, H. Design and synthesis of potent HIV-1 protease inhibitors incorporating hexahydrofuropyranol-derived high affinity P(2) ligands: structure-activity studies and biological evaluation. *J Med Chem* 2011, 54: 622–634.

270. Gold, version 4.1.2., Cambridge Crystalographic Data Center, Cambridge, UK.

271. Frieden, T.R., Munsiff, S.S. The DOTS strategy for controlling the global tuberculosis epidermic. *Clin Chest Med* 2005, 26: 197-205.

272. Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., Blanot, D. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev* 2008, 32: 168-207.

273. Anurag, M., Dash, D. Unraveling the potential of intrinsically disordered proteins as drug targets: application to Mycobacterium tuberculosis. *Mol Biosyst* 2009, 5: 1752-1757.

274. Wang, Y., Bolton, E., Dracheva, S., Karapetyan, K., Shoemaker, B.A., Suzek, T.O., Wang, J., Xiao, J., Zhang, J., Bryant, S.H. An overview of the PubChem BioAssay resource. *Nucleic Acids Res* 2010, 38: D255-266.

275. Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., Bryant, S.H. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 2009, 37: W623-633.

276. Jagtap, P.A., Verma, S.K., Prakash, B. Structure of Mycobacterium tuberculosis GlmU in complex with acetyl CoA. 2012, DOI: 10.2210/pdb3spt/pdb.

APPENDIX

LIST OF PUBLICATIONS

1. **Therese, P.J.,** Manvar, D., Kondepudi, S., Battu, M.B., Sriram, D., Basu, A., Yogeeswari, P., Kaushik-basu, N. Multiple e-pharmacophore modeling, 3D QSAR and high-throughput virtual screening of hepatitis C virus NS5B polymerase inhibitors. *J Chem Inf Model* 2014, 54: 539–552.
2. **Patrisha, J.T.,** Sriram, D., Yogeeswari, P. 3D QSAR studies combined with virtual screening to identify novel inhibitors of N-acetyl glucosamine 1-phosphate uridyltransferase from *Mycobacterium tuberculosis*. *Int J Drug Des Discov* 2013, 4: 1134-1148.
3. **Patrisha, J.T.,** Sridevi, K., Shruthi, K., Sriram, D., Yogeeswari, P. Discovery of new small molecule inhibitors of the hepatitis C virus RNA-dependent RNA polymerase: pharmacophore modeling, 3D QSAR studies and high throughput screening. *J Biotechnol Biomater* 2012, 2: 6.
4. Sriram, D., Yogeeswari, P., **Patrisha, J.T.,** Senthilkumar, P., Kurre, P.N., Prasad, Y.R. Exploring aryl thiazolidine carboxamides as a new class of antimycobacterials. *Pharmacology Online Newsletter* 2011, 1: 185-195.
5. **Patrisha, J.T.,** Sriram, D., Yogeeswari, P. 3D QSAR, pharmacophore modeling and docking studies based virtual screening to identify novel small molecule non-peptidic inhibitors of human immunodeficiency virus (HIV) protease. [To be communicated]

PAPERS PRESENTED AT NATIONAL/ INTERNATIONAL CONFERENCES

1. Yogeeswari, P., **Patrisha J.T.,** Sriram, D., Structure based drug design utilizing multiple crystal structure to design anti-viral agents, **3rd Science One Conference on Drug Discovery and Development (SCDDD),** 21-23 January, 2014, Dubai.
2. **Patrisha, J.T.,** Sriram, D., Yogeeswari, P. Discovery of novel small molecule inhibitors of Human immunodeficiency virus (HIV) protease: Pharmacophore Modeling, 3D QSAR studies and High-throughput screening, **2nd UK-India MedChem Congress,** 22-23 March, 2013, Hyderabad, India.
3. **Patrisha, J.T.,** Sridevi, K., Shruti, K., Sriram, D., Yogeeswari, P. Discovery of New Small Molecule Inhibitors of Hepatitis C Virus RNA-dependent RNA polymerase:Pharmacophore modelling, 3D QSAR studies and High-throughput screening, **3rdCongress on Biotechnology,** 13-15 Septemeber 2012, Hyderabad, India.
4. Jean Kumar V.U., Manoj C., Vijay S., **Patrisha J.T.,** Mallika A., Yogeeswari P., Nandicoori, V.K., Sriram D., Discovery of novel *Mycobacterium tuberculosis* GlmU inbibitors, **16th Annual Conference of the Swedish Structural Biology Network**, 15-18 June 2012, Talberg, Sweden.

BIOGRAPHY OF J. THERESE PATRISHA

Ms. Therese Patrisha J. completed her Bachelar of Science in Chemistry from Thiruvalluvar College, Manonmanium Sundaranar University and Masters of Science in Chemistry from Annamalai University, Chidambaram, Tamilnadu. She is pursuing Ph.D. from Department of Pharmacy, Birla Institute of Technology & Science-Pilani (BITS-Pilani), Hyderabad campus, under the guidance of Prof. P. Yogeeswari since March, 2010. During her doctoral studies she received Junior Research Fellowship (JRF) by CSIR-OSDD, DBT-CoE and BITS-Institute Fellowship. She has four publications and four conference publications in different national and international conferences.

BIOGRAPHY OF PROF. P. YOGEESWARI

Dr. P. Yogeeswari is presently working in the capacity of Professor and Associate Dean (Sponsored Research and Consultancy Division), Department of Pharmacy, Birla Institute of Technology and Science, Pilani, Hyderabad Campus. She received her Ph.D. degree in the year 2001 from Banaras Hindu University; Varanasi. She has been involved in research for the last 14 years and in teaching for 13 years. APTI honored her with YOUNG PHARMACY TEACHER AWARD for the year 2007. In 2010, ICMR honored her by awarding "Shakuntala Amir Chand Award" for her excellent biomedical research. She has also been selected for IASP 2014 award for "Excellence in Pain Research and Management in Developing Countries" under the basic science research category to be received at the "15th World Congress on Pain" at Argentina in October 2014. She has collaborations with various national and international organizations that include National Institute of Health, Bethesda, USA, National Institute of Mental Health and Neurosciences, Bangalore, Karolinska Institute, Stockholm, Sweden, National Institute of Immunology, New Delhi, India, Pasteur Institute, University of Lille, France, Bogomoletz Institute of Physiology National Academy of Science, Ukraine, and Faculty of Medicie of Porto, Porto, Portugal,. She has to her credit more than 190 research publications and one Indian Patent, Application No: 1138/CHE/2009. She is an expert reviewer of many international journals like Journal of Medicinal Chemistry (ACS), Journal of Chemical Information & Modeling (ACS, USA), Bioorganic Medicinal Chemistry (Elsevier), Recent Patents on CNS Drug Discovery (Bentham), etc. She has also co-authored a textbook on organic medicinal chemistry with Dr. D Sriram titled "Medicinal Chemistry" published by Pearson Education and one book chapter in in Jan 2013 by IGI Global. She is a lifetime member of Association of Pharmacy Teachers of India and Indian Pharmacological Society. She has successively completed many sponsored projects and currently on projects sponsored by DST, DBT INDO-BRAZIL, ICMR-INSERM, and CSIR. She has guided six Ph.D students and currently thirteen students are pursuing their Ph.D. work.