Computational Design, Validation and Delivery of HIV-1 Entry and Early replication inhibitors

DOCTOR OF PHILOSOPHY

By

Akhila Bommakanti

Department of Biotechnology and Bioinformatics
School of Life Sciences
University of Hyderabad
Hyderabad-46, Telangana., INDIA

Computational Design, Validation and Delivery of HIV-1 Entry and Early replication inhibitors

A thesis submitted to University of Hyderabad for the award of Ph.D. Degree in Biotechnology and Bioinformatics

By

Akhila Bommakanti

Department of Biotechnology and Bioinformatics
School of Life Sciences
University of Hyderabad
Hyderabad-46, Telangana., INDIA

Enrollment No: 14LTPH05
JULY 2019

UNIVERSITY OF HYDERABAD
(A central university established in 1974 by an act of Parliament)
Department of Biotechnology and Bioinformatics School of Life Sciences
University of Hyderabad, Hyderabad 500 046, India

CERTIFICATE

This is to certify that thesis entitled "**Computational Design, Validation and Delivery of HIV-1 Entry and Early replication inhibitors**" is a record of bonafide work done by Ms. Akhila Bommakanti, a research scholar for Ph.D. program in Department of Biotechnology and Bioinformatics, University of Hyderabad under my guidance and supervision. The thesis has not been submitted previously in part or full to this or any other university or institution for the award of any degree or diploma. I recommend her thesis for submission towards the partial fulfillment of Doctor of Philosophy degree in Biotechnology and Bioinformatics.

Supervisor	**Head** Department of Biotechnology and Bioinformatics	**Dean** School of Life Science

CERTIFICATE

This is to certify that parts of the thesis "**Computational Design, Validation and Delivery of HIV-1 Entry and Early replication inhibitors**" submitted by **Ms. Akhila Bommakanti,** bearing registration number **14LTPH05** in partial fulfilment of the requirement for award of Doctor of philosophy has been

A) Published in following publications:

- Kurumurthy K, Devaraya K, Bommakanti A, Kondapi A K. Development of Pyridine Dicoumarols as potent anti HIV-1 leads, targeting HIV-1 associated Topoisomerase IIβ Kinase, *Future Medicinal Chemistry* (2017) 9:14, 1597-1609.
- Ponraj K, Prabhakar M, Rathore R.S, Bommakanti A, Kondapi A.K. HIV-1 Associated Topoisomerase IIβ Kinase: A Potential Pharmacological Target for Viral Replication, *Current Pharmaceutical Design* (2013) 19: 4776.

B) Presented in following conferences:

- EMBO workshop DNA topoisomerases and DNA topology on "HIV-1 associated Topoisomerase II beta kinase: A distinctly charged active site environment" conducted from 17-21, September 2017 at Les Diablerets, Switzerland.
- Biology And Therapeutics of HIV & Associated Infections on "Prevalence of Drug Resistance in Reverse Transcriptase region of HIV-1" from 20-21, January 2019 at University of Hyderabad, Hyderabad.

Further, the student has passed the following courses towards fulfillment of coursework requirement for Ph.D .

S.No	Course Code	Course Name	Credits	Pass//Fail
1	BT801	Seminar	1	PASS
2	BT802	Research Ethics & Management	2	PASS
3	BT803	Biostatistics	2	PASS
4	BT804	Analytical Techniques	3	PASS
5	BT805	Lab Work	4	PASS

Supervisor

Head
Department of Biotechnology and Bioinformatics

Dean
School of Life Science

UNIVERSITY OF HYDERABAD
(A central university established in 1974 by an act of Parliament)
Department of Biotechnology and Bioinformatics School of Life Sciences
University of Hyderabad, Hyderabad 500 046, India

Declaration

I, **Akhila Bommakanti**, hereby declare that the work presented in this thesis, entitled "**Computational Design, Validation and Delivery of HIV-1 Entry and Early replication inhibitors**" has been carried out by me under the supervision of **Prof. Anand K. Kondapi**, Department of Biotechnology and Bioinformatics. To the best of my knowledge this work has not been submitted for the award of any degree or diploma at any other university or institution. I hereby agree that my thesis can be deposited in Shodganga/INFLIBNET. A report on plagiarism statistics from the University Librarian is enclosed.

Place: Hyderabad
Date:

Akhila Bommakanti
14LTPH05

ACKNOWLEDGEMENTS

PhD is not just a degree but a journey. A mixture of Eureka moments and moments of despair, my journey has been the most memorable. It allowed me to identify my strength, weakness, interest and passion and if not for this opportunity, I wouldn't have been exposed to new avenues and learnings, which I now consider my strengths. A lot of people have been instrumental in my journey, and I take this opportunity to thank them all.

I will always remain indebted to my supervisor Prof. Anand Kumar Kondapi who has been instrumental in gently introducing me to the world of research and for offering constant encouragement, freedom and support in designing and implementation of this thesis. His insights and knowledge are unmatchable, and it provided a direction to the work helping in the timely completion. I hope to always remain his student and gain from his experience and knowledge.

I would like to express my sincere thanks to present Dean School of Life Sciences, Prof. S. Dayananda and former Dean Prof. K.V.A Ramaiah, Prof. Reddanna, Prof. Aparna Dutta Gupta, Prof. R.P. Sharma, Prof. A. S. Raghavendra for permitting me to use necessary facilities to carry out my work.

I would also like to express my sincere regards to the present Head, Department of Biotechnology and Bioinformatics, Prof. K.P.M.S.V Padmasree and former Head of the Department Prof. J.S.S Prakash, Prof. Anand. K Kondapi, Dr. Niyaz Ahmed. I offer my sincere gratitude to my doctoral committee members Dr. Sunanda Bhattacharya and Dr.Prakash Prabhu for their valuable suggestions and guidance throughout my Ph.D. I sincerely thank all the faculty members of School of Life sciences.

I thank ICMR (SRF) for providing me fellowship during my Ph.D and the funding bodies ICMR, DBT and DST of Govt. of India for providing funds to our lab during my work. I thank Prof. Caroline Austin for giving the opportunity to present my work at the EMBO workshop conducted at Switzerland, an experience of my lifetime.

I thank Mr. Rajesh for his constant help at BIF and server facilities during my work. I would like to thank Dr. Balakrishna for patiently teaching and guiding me through the in-vitro work. I thank my previous lab mates Dr. Sarada, Dr. Farhan, Dr. Prashant, Dr. Pankaz, Dr. Lakshmi, Dr. Kurumurthy, Harikiran, Sonali, Ajay, Prashanthi, Lalith, Bhuvnesh.

My special thanks to present lab mates Kiran, Chuku, Satyajith, Neha, Pritikana, Vidya, Veena for being a wonderful, enthusiastic and positive group, always ready to help during my work.

I thank Mr. Murthy, and all non-teaching staff, especially administration of UOH for their help. I thank Sreenivas, Bhanu and Chandra for their dedicated work in the lab. I would like to express heartfelt thanks to my friends and my Ph.D batchmates and my seniors for their helping nature.

My special heartfelt thanks to Jagadeesh and Suresh for their support, company and help throughout my work. Their presence has been a blessing and is beyond the scope of this section of acknowledgement. Thanks to Veeru and Yashu for being the breath of fresh air during hectic workdays. My love to Snigdha Pola and Shaik Sabiha for holding me up during the bad times. Special thanks to Sagar, Varul and RAYS for constant support and encouragement.

I would finally like to mention my wonderful parents for giving me everything I have today and making me who I am. Their encouragement and support have been the guiding torch in my life, for which I will always be indebted. I thank my brother Aakarsh for all the arguments, fights and mind stimulating discussions we have. I thank my extended family, Aunts, Uncles and Cousins who have lived this journey through me, encouraging me throughout.

Finally, I thank the life force that lets me wake up every day with new hope, vigor and vitality.

-Akhila Bommakanti

Dedicated to my
Beloved Grandparents

Abbreviations:

HAART: Highly Active Anti-Retroviral Therapy

ARV: Anti Retro-viral

NRTI : Nucleotide Reverse Transcriptase Inhibitors

NNRTI: Non-Nucleotide Reverse Transcriptase Inhibitors

EPAP-1: Early Pregnancy Associated Protein-1

TopoIIBK$_{HIV}$: HIV-1 associated Topoisomerase II Beta Kinase

QSAR: Quantitative Structure Activity Relationship

COMFA: Comparative Molecular Field Analysis

TAM: Thymidine Analogue Mutations

AZT: Zidovudine

FTC: Emtricitabine

3TC:Lamivudine

TDF: Tenofovir

NVP: Nevirapine

EFV: Efavirenz

ETR: Etravirine

RPV: Rilpivirine

CurNPs: Curcumin Nanoparticle

INDEX

Chapter I: Introduction **1**

Chapter II: Materials **19**

Chapter III: Objective 1 **35**

Prevalence of HIV-1 Reverse Transcriptase drug resistance in India.

Chapter IV: Objective 2 **53**

Design, Development and Validation of HIV-1 Entry Inhibitors.

Chapter V: Objective 3 **92**

Insights into the binding pocket of HIV-1 Associated Topoisomerase II Beta Kinase using in-silico studies.

Chapter VI: Objective 4 **109**

Facilitating cellular localization of Topoisomerase II Beta Kinase inhibitors using Lactoferrin Nanoparticles.

Chapter VII: Summary & Conclusion **122**

CHAPTER I

INTRODUCTION

In 1981, an increase in unexplained opportunistic infections claiming lives of young homosexual men lead to the recognition of Acquired Immuno-Deficiency Syndrome (AIDS) as a new disease and Human Immuno-deficiency Virus as its causative agent (Sharp et al., 2011). Soon after, in 1986 India reported its first known case of HIV in Chennai, Tamil Nadu amongst female sex workers (Solomon et al., 2006). Since then, the disease claimed 35 million lives in total, making it one of the deadliest diseases that gripped mankind and affected populations around the world (WHO, Global Health Observatory Data). 36.7 million people were reported to be living with HIV in 2017 (Katz et al., 2018), among which 35.1 million were adults and 1.8 million were children (<15 years) (Iqbal et al., 2019). Middle and lower-income countries have been the most affected, with the majority in sub-Saharan Africa (Shao et al., 2012). According to a 2016 WHO report, HIV/AIDS stands to be the fifth major cause of death in low-income countries(Watkins et al., 2018).

India is ranked 3rd, worldwide, for the number of people living with HIV, preceded by Nigeria and South Africa. In the year 2016, 80,000 new HIV infections and 62, 000 AIDS-related deaths were recorded in India (UNAIDS) . By 2017, the states with a high prevalence of HIV included North Eastern states, Maharashtra, Telangana (India HIV Estimations, 2017). By 2024, the government plans to eradicate AIDS by working towards objectives to reduce the number of new infections by 80% and to eliminate Mother to child transmission of HIV (UNAIDS).

Origin and Subtypes

The origin of HIV has been explained by many theories, a popular one being, the origin from a primate virus, Simian Immuno-deficiency Virus, SIV found in Sub-Saharan Africa. The presence of SIV infections has been identified in over 40 primate species and is believed to have transmitted to human multiple times through at-least two different infected non-human primates (Sharp et al., 2011) (Hemelaar, 2012). The virus transmitted through Chimpanzees are designated HIV-1, and the one transmitted through Sooty Mangabeys is known as HIV-2 (Sharp et al., 2010). It is also believed that within these major classes, the transmission occurred multiple times, resulting in the evolution of various subgroups (Hills et al., 2000). Among the evolved subgroups, HIV-1, M group spread across continents and is considered the cause for the devastating epidemic, AIDS. HIV-1 O group is confined to West Africa and is not as dangerous as the other subgroup. Subgroups N and P have also been reported (Sharp et al., 2011).

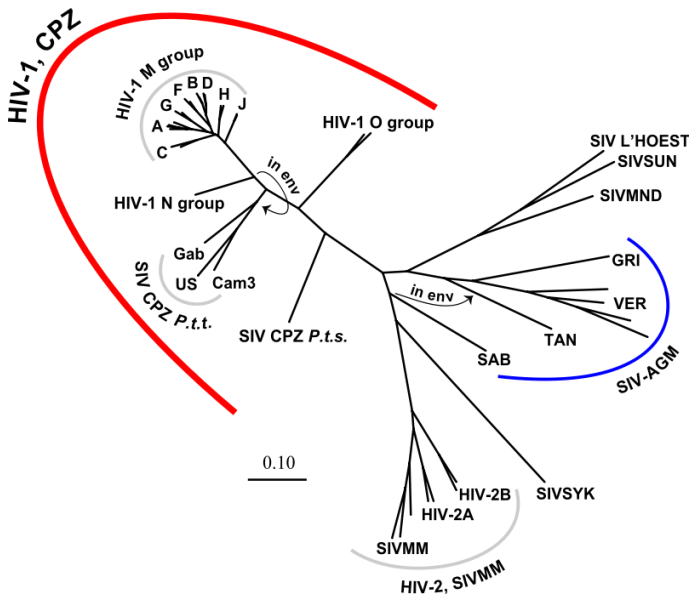


Fig1: The Phylogenetic Tree depicts the branching of SIV_{CPZ} and HIV-1 drawing a relation between the two viruses. A similar branching pattern is observed for SIV_{MM} and HIV-2. The tree was constructed by taking reference sequences from the different subtypes of group M. (Kuiken et al., 1999)

HIV-1 M group diversified into various subtypes, each dominating a different region in the world. The subtypes are

- Subtype A is widely spread across West Africa
- Subtype B is commonly observed in Europe, Japan and both of the American continents. It is well studies and documented among all the other subtypes. It is also seen circulating in Middle East and Australia.
- Subtype C is dominant in the lower income & developing nations like Southern Africa, Eastern Africa, India, Nepal, and parts of China. It is considered the most aggressive subtype and most prevalent, with 50% of people living with HIV (PLWH) infected with this subtype.
- Subtype D is commonly seen circulating in East and Central Africa.
- Subtype F is found commonly in central Africa, Eastern Europe and South America.
- Subtype G and a circulatory recombinant of A & G (CRF02_AG) are prevalent in Africa and central Europe.
- Subtype H is concentrated specifically in central Africa.

- Subtype J & K are mostly seen in North, Central and West Africa & Democratic Republic of Congo and Cameroon.

This diversity adds to the complexity of the virus, making the management and treatment difficult. Though this list throws some light on the various subtypes of HIV-1, it is not exhaustive as new circulatory recombinants are being discovered and added to the list (Taylor et al., 2009).

HIV Infection

Human Immuno-deficiency Virus is a lentivirus, belonging to the family Retroviridae, which affects the functioning of the host immune system. The transmission of this virus occurs through unprotected sexual contact, contact of broken skin/wound with infected blood or body fluids, use of HIV contaminated needles and vertical transmission from mother to child during pregnancy, birth or breastfeeding (Lange et al., 2008). People of certain backgrounds, habits and orientations are characterized as high-risk populations based on the probability of their exposure to HIV. The identification of this population helps to target these groups for various outreach programs to create awareness and to provide extra care. MSM, people in prisons, intravenous drug users, sex workers, infants of HIV+ mothers and transgender people are a part of this high-risk group, with MSM being at highest risk (Paranjape et al., 2016).

HIV, once in contact, preferentially targets the CD4+ T Cells hijacking them to replicate and produce infectious virions. This eventually results in the fall of CD4+ T Cells followed by immune-suppression and exhaustion leading to the manifestation of the condition, AIDS (Archin et al., 2014). According to the studies conducted on the closest SIV that infects macaques, the infection of CD4+ T Cells can be detected as early as two days post-viral contact (Deeks et al., 2017). Within two weeks, the virus reaches the distal lymph nodes and other tissues after affecting the mucosal layer and draining lymph nodes at the site of infection (eclipse phase) . A viral reservoir is established during this phase, and Gut-associated lymphoid destruction is seen (Deeks, 2017). Following this, the progression of the disease occurs through the following three stages namely.

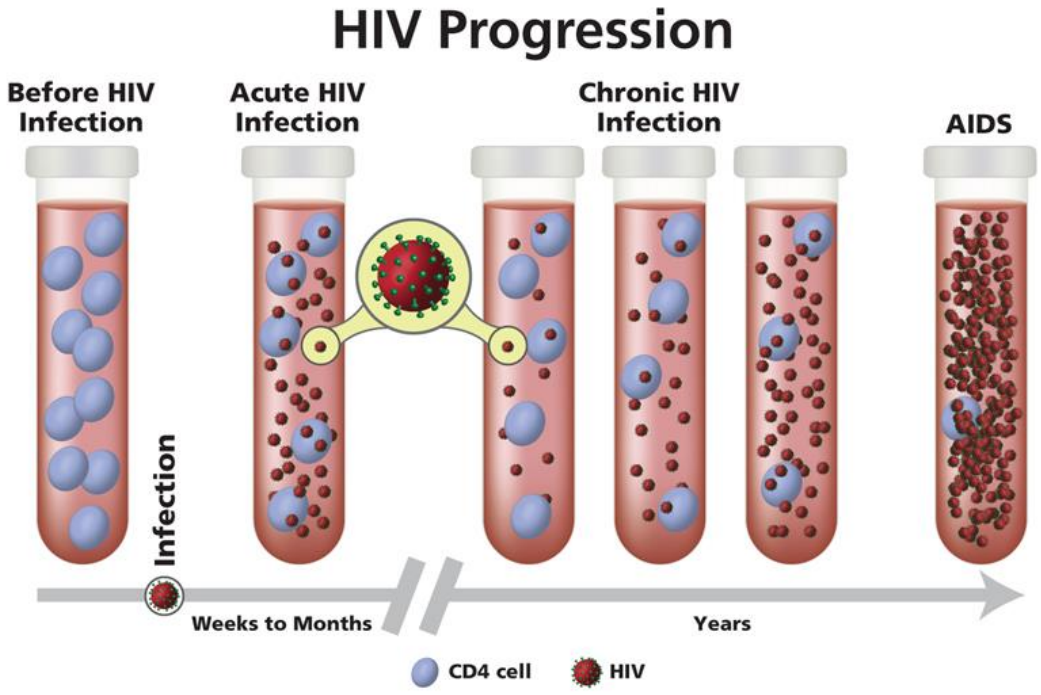


Fig 2: Depicts the changes in the CD4 cell count and HIV-1 viral load as the disease progresses.

The **acute infection**, which develops within 2-4 weeks of exposure, is usually characterized by flu-like symptoms in the patient following seroconversion. This is the phase, where HIV targets the CD4 cells, replicates in large numbers, and spreads through the body. The CD4 T-cell count drops due to HIV mediated destruction and the viral load spikes making it detectable in blood. This makes the patients highly infective to their sexual partners. CTL response can be noted.

Chronic infection is also known as the latent period as many patients remain asymptomatic during this phase for a long period. The viral set point is established. Progressive drop in CD4 T-Cell count leads to Chronic inflammation. A decrease in the level to lower than 350 cells/µl increases the risk of other opportunistic infections due to suppressed immunity. Further decrease in the CD4 count leads to the onset of AIDS.

AIDS is characterized by a CD4 count lower than 200 cells/µl. The US CDC defines AIDS based on HIV infection if the CD4 count falls below 200 cells/ µl or the presence of AIDS-defining opportunistic infection and malignancy (Deeks, 2017).

Structure of mature HIV virion

HIV is a spherically shaped enveloped retrovirus with an approximate diameter of 100nm. The outer layer of the virion is a lipid bilayer that encloses the core consisting of the genome and the proteins necessary for its functioning. The envelope contains the protein responsible for facilitating virus entry into the host cell. Underneath the viral envelope is the matrix (MA) made up of approximately 2000 units of p17 protein. The core of the virion is distinctly conical in shape. Enclosed by capsid (CA) made up of protein p24, the nucleocapsid protein (NC) that forms a layer underneath the capsid is made of the protein p7/p9.(Barré-Sinoussi et al., 1996) (Turner et al., 1999)The core consists of two copies of RNA that form the genome of HIV surrounded by proteins Reverse Transcriptase, Protease, Integrase and accessory proteins Nef, Vif and Vpu (Turner et al., 1999). The lipid bilayer consists of Env spikes that are exposed on the surface are highly glycosylated. Since the envelope spike is exposed, it is the only part of the virus that elicits an immune response in the host (Pantophlet et al., 2006).

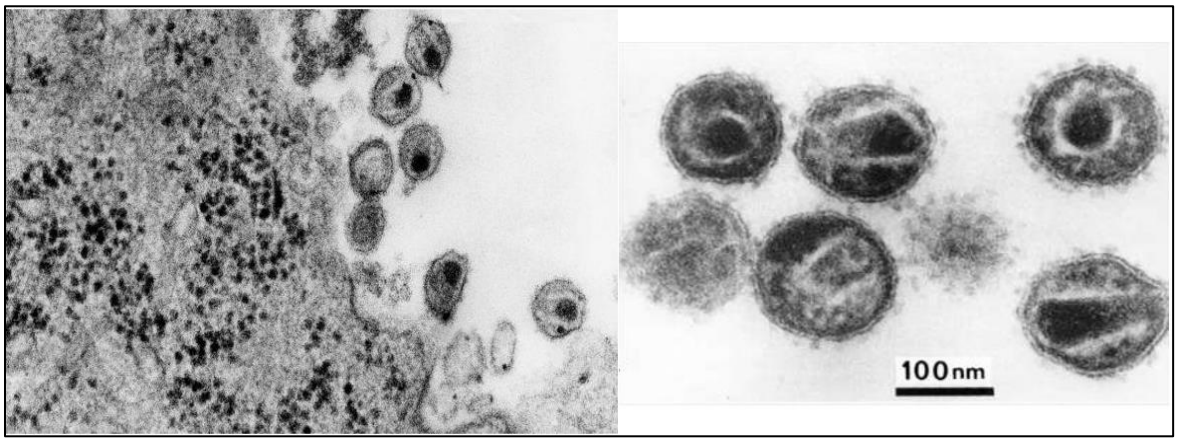


Fig 3: A transmission electron micrograph image of the HIV taken in 1983. The image produced under high magnification is of an infected tissue sample from AIDS patient. The diameter of the particle is approximately 100 nanometers.

HIV Therapy

In the year 1986, first drug, Zidovudine (ZDV), to treat HIV infection was approved. It inhibits Reverse Transcriptase (RT) a viral protein that facilitates translation of the viral single-stranded RNA to double-stranded DNA in the cytoplasm of the host cell (Vella et al., 2012) . This drug marked the beginning of HIV treatment with about 22 molecules and 21 combination drugs approved and many more in the clinical trials as of September 2018 (AIDS Info). Following ZDV, a larger number of drugs were approved in the coming years not only against RT but also other viral enzymes.

Despite the increase in the available options for treatment, the mode was mostly restricted to prescription of a single compound/drug, known as monotherapy (Arts et al., 2012). The effectiveness of this mode of treatment was limited and varied with individuals. Low effectiveness, along with a better understanding of the virus' functioning in the host and emergence of resistance to single drugs, lead to the shift from monotherapy to combination therapy (Johnson et al., 2009). Identification of new drug classes which target different stages in the HIV life cycle facilitated the development of combinations, that effectively control HIV infection. Treatment with a combination of drugs from different classes enables attacking the virus at different stages, thus disarming and preventing progression.

The use of a combination of drugs is known as Highly Active Anti-Retroviral Treatment. HAART became popular in 1997 and since then has been the mode of treatment for HIV infected patients. Strong clinical evidence suggesting the effectiveness of the combination of one or two classes of drugs in controlling infection was reported making HAART, the protocol for treatment (Lu et al., 2017). This combination has been shown to control viral titres in the blood, maintain CD4 count and improve overall health of the patient (O'Connell et al., 2009). Since the initiation of HAART, a decline in the number of deaths across the globe was reported (Brady et al., 2010).

WHO lays out guidelines for initiation of ART Therapy based on the clinical stage and CD4 count of the patient. However, WHO suggests not to delay initiation of treatment due to lack of CD4 results if the patient is eligible based on the clinical stage (Ford et al., 2018). Initiation of ART patients in Clinical stage I and II, is recommended if CD4 count falls below 350 cells/ml and for patients exhibiting Clinical stage III and IV, it can be initiated irrespective of the CD4 count (NACO Technical Guidelines, 2018).

ART is administered in regimens, usually known as first, second and sometimes third line regimens. The second- and third-line regimens are initiated upon failure of the first line treatment.(WHO guidelines 2018)

First Line Regimen:

This usually comprises of two Nucleoside Reverse Transcriptase Inhibitors (NRTIs) and one non-nucleoside Reverse Transcriptase inhibitor (NNRTI) .Failure of the first line compels the initiation of the second line treatment and deciding when to shift the treatment regimen is important as too early or too late will result in inefficient treatment outcome. Treatment failure can be defined in three ways, clinically, immunologically and virologically. Clinical failure is defined as recurrence of Stage IV condition after a minimum of 6 months of ART treatment. Immunological failure is characterized by the fall in CD4 counts to levels before the start of treatment or 50% of the level attained during treatment or persistent count of < 100 cells/mm3. Virological failure is when the viral load is > 5000 copies/ml after a minimum of 6 months of ART treatment (NACO 2013) .

Second Line Regimen:

Based on the cause of failure and a complete understanding of the resistance in the patients, a second line treatment is prescribed. Two NRTIs+ a ritonavir-boosted protease inhibitor (PI) combined as a fixed dose is the most preferred treatment option. Nucleoside cross-Resistance and drug interactions should be considered before prescribing any combination.

Third Line Regimen:

WHO urges nations to have strong policies for third-line regimens. Ideally, the third line should contain drugs from new and novel classes to avoid cross-resistance from drugs used in previous regimens. These include Integrase inhibitors, Entry inhibitors, second-generation Reverse transcriptase inhibitors and others. Patients with failure of both lines should be prescribed this at acceptable dosages (WHO guidelines 2018).

HIV-1 Inhibitors

Reverse Transcriptase inhibitors:

This class of inhibitors act by blocking the activity of Reverse Transcriptase. The enzyme Reverse Transcriptase is responsible for catalyzing the complex conversion of viral RNA to intermediate double stranded DNA which is later integrated into host genomic DNA. This enzyme in HIV-1 exhibits three-fold activities; RNA dependent DNA polymerase activity responsible for synthesis of (-) DNA from viral (+) RNA, DNA dependent DNA polymerase activity responsible for generation of (+) DNA and Ribonuclease H activity which acts on the RNA in the RNA-DNA intermediate. These inhibitors are classified as

Nucleoside Reverse Transcriptase Inhibitors (NRTI) analogs act as naturally occurring nucleotides when phosphorylated by host metabolizing enzymes, but lack 3'OH group. These act in two ways, one by preventing binding of the naturally occurring dNTPs by competitively binding to the enzyme active site during synthesis of (-)DNA strand. Secondly, the lack of 3'OH group inhibits the elongation process, this inhibition is known as chain termination. Nucleoside inhibitors, as previously mentioned, need to be phosphorylated by host system to act as inhibitors. Compounds developed to act directly as nucleotide analogs without requiring phosphorylation are known as Nucleotide Reverse Transcriptase Inhibitors (NtRTI). Despite the fool proof inhibition mechanism, NRTIs/NtRTIs are not highly effective, as HIV-1 can surpass the hurdles put forth by their incorporation during the copying. It is capable of distinguishing between naturally occurring nucleotides and the NRTI- Tri phosphate analogs. Further it can excise the NRTI-mono phosphate base responsible for chain termination using ATP as substrate.

However, it is still the preferred drug class to initiate therapy but due to acquired resistance is replaced by boosted protease inhibitor. Interestingly, recent clinical trial EARNEST, has suggested that no improved activity was observed when resistance acquired NRTIs (genotypically determined) were substituted by other class of drugs in the second line regimen. After discontinuing two drugs considering their long-term toxic effects, currently six NRTIs are available for use in ART therapy (Herman et al., 2012).

Non-nucleoside Reverse Transcriptase Inhibitors (NNRTI):

These compounds act by binding to an allosteric site on the RT enzyme, resulting in conformational change in the active site pocket thus affecting its activity. It is also suggested

that the binding of these compounds at the allosteric site restricts the flexibility of enzyme resulting in further change in activity. These along with the NRTIs are used as first line treatment due to their high potency and specificity. However, this class of drugs are the most affected by cross resistance from mutations developed in the binding site of NRTIs. Around four drugs have been approved till date for use in clinics and many are in clinical trials (Poongavanam et al., 2018).

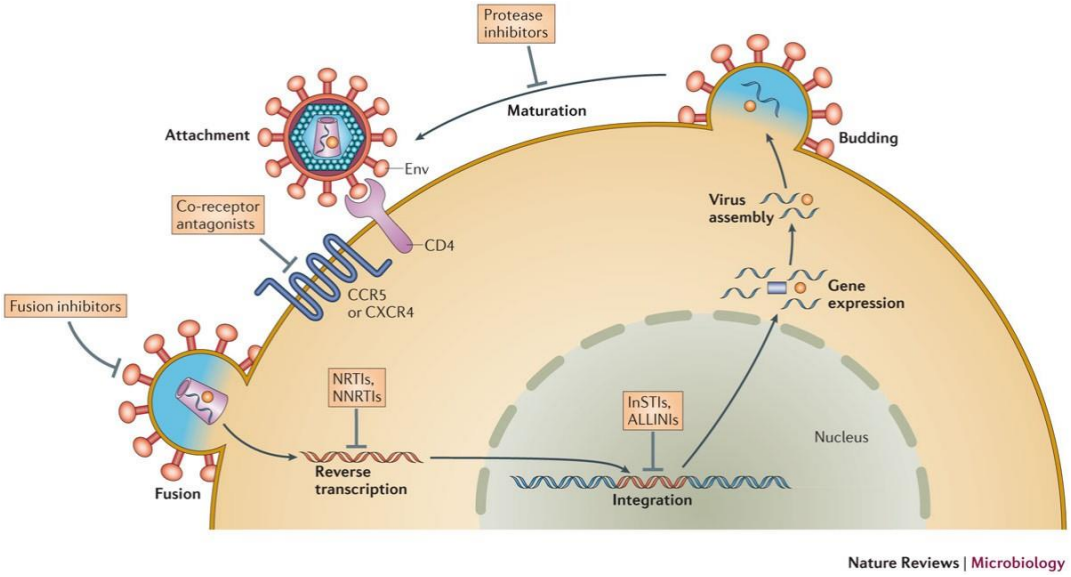


Fig 4: Indicates various classes of HIV-1 inhibitors targeting various stages of the virus life cycle.

Protease Inhibitors (PI):

An aspartyl-protease existing as a homodimer, HIV-1 Protease functions to cleave Gag and Gag-Pol precursors to yield mature, functional proteins. The active site of the protease, consisting of well conserved catalytic Aspartic acid residue, is usually inaccessible to the substrates due to the presence of two flexible β-hair pin loops. The inhibition of this enzyme occurs by the binding of inhibitors in the active site when it is exposed. There are 10 FDA approved PI and all the inhibitors possess similar structure and mode of binding. PIs are usually linked to high metabolic toxicities over long periods of treatment. These are usually given when the first line treatment fails (Chu et al., 2015).

Integrase Inhibitors (INSTI):

The enzyme Integrase works on the integration of double stranded proviral DNA into the host genomic DNA. Currently two drugs approved by FDA are being used for treatment that act by inhibiting the activity of Integrase and are known as Integrase strand transfer inhibitors (INSTIs). These are a fairly new class of drugs and are known for their high potency, greater clinical response, safety and tolerability (Blanco et al., 2015). The drawback of these inhibitors is the resistance and cross resistance that can be tackled by development of newer second-generation inhibitors (Yadav et al., 2018).

Entry Inhibitors:

HIV-1 entry is a multistage complex process involving receptor binding, coreceptor binding and eventually the membrane fusion. The entry inhibitors can target any of the stages to ultimately block the entry of virus into the host cell. Based on the stages, they are classified as fusion and post attachment inhibitors. Three drugs have been approved to target the entry and more are in trials. Approved drugs have different mode of action and have issues like tropism dependence and bioavailability. Many new strategies are being employed to design novel molecules to target viral entry (Tilton et al., 2010).

Computer Aided-Drug Discovery and anti-HIV leads:

The process of drug discovery; from identifying a target to marketing the drug takes about an average of 12 years or more and the cost to take the final product to market is as high as $2.6 billion (Mohs et al., 2017). Attrition rates in Drug Discovery are well known, and the failures are seen in different phases of Clinical Trials. The rates of failure at Phases I, II and III are 46%, 66% and 30% respectively and are usually due to efficacy and toxicity (Mignani et al., 2016). However, this process has been revolutionized owing to the advancement of interdisciplinary areas like genomics, bioinformatics, systems biology etc. Informatics techniques and other in-silico experimental methods used in support of in-vitro and in-vivo experiments increase the chance of obtaining a promising lead with desired properties (Pugazhendhi, 2013). Overall, the interference of computational techniques is a potentially useful guide for the process of drug development towards being efficient and cost-effective.

Improved experimental techniques that aide in High Throughput Screening (HTS) allow screening of thousands of molecules with limited manual intervention. These techniques made lead identification, an easier process. However, the low success rates and the costs involved necessitates the emergence of technologies that can tackle the problem (Leelananda et al., 2016). Computer-Aided Drug Discovery (CADD) techniques make use of a collection of informatics tools to considerably shrink the number of molecules that require screening. The molecules obtained after CADD techniques also have a better success at later stages of development (Hassan Baig et al., 2016). Drug Discovery pipelines of major pharmaceutical companies now have CADD as an essential part of their activities (Leelananda et al., 2016).

CADD methods have been instrumental in some note-worthy discoveries of drugs approved for the treatment of a wide array of diseases, one among them being HIV-AIDS (Santos et al., 2015). Increased implementation of computational methods is seen in HIV research to support the results of previous studies and to establish novel ones. The major targets on focus for computational drug biologists are HIV-1 Reverse Transcriptase, Protease, Integrase and co-receptors of HIV-1 entry (Phillips et al., 2018).

Drugs such as ritonavir (Norvir®), indinavir (Crixivan®) and saquinavir (Invirase®) belonging to PIs, raltegravir (Isentress®) of INSTIs, rilpivirine (RPV) (Edurant®) an RT inhibitor and enfuvirtide (Fuzeon®) a fusion inhibitor have all been designed and developed using computational methods (Santos et al., 2015).

Some other computational studies on HIV-1 inhibitors include, Molecular docking approaches applied to phenanthrene beta-diketo acid derivatives against HIV-1 integrase revealed two distinct binding modes that were different from the modes of elvitegravir and raltegravir, suggesting a novel region of binding in the Integrase active site (Gu et al., 2014). After the success of Raltegravir, a large number of analogues were developed, tested and promoted into clinical trials. However, the variations in the in vitro activity of these analogues were reported to be minor. Hence, researchers predicted the failure of these analogues in the later stages of the trials owing to their similarities. To elucidate this point, the interacting residues of the analogues were predicted and compared with the residues in the binding site of Raltegravir. Molecular Docking studies were employed for the purpose, and the experiments showed similar binding patterns in the analogues with the involvement of the same set of residues. Interestingly, the mutations that confer the virus to be resistant to

integrase drugs occur in the binding pocket, specifically on E29Q. Thus, the analogues would have the same resistance profile as the mother compound (Phillips et al., 2018).

In another study, virtual screening of 1,430 natural products and their derivatives resulted in the selection of 38 compounds which were later evaluated by toxicity screens and in-vitro assays, leading to the identification of one compound (NPD170) with high anti-viral activity (Gu et al., 2014).

In an attempt to identify compounds that disrupt the interaction between LEDGF/p75 with integrase, 26 known drugs identified from DrugBank were filtered using molecular docking. Eight potent compounds were identified from the study (Gu et al., 2014). Azaindole carboxylic acid derivatives were previously identified to possess potential as HIV-1 integrase inhibitors. A later study exploited these structures through 3D QSAR to explore the pharmacophore of the compounds. The binding modes were studied using Molecular Docking studies and based on the information obtained from both the experiments, new derivatives were designed, and their activities were predicted (Gu et al., 2014).

Among the approved HIV Protease inhibitors, Danuravir is known as the most potent inhibitor. Several researchers focused on developing and testing analogues of this drug, and these derivatives were evaluated by Docking and 3D-QSAR methods. Results obtained from both the experiments were in agreement with each other and also with the experimental results (Gu et al., 2014).

The most exploited target by computational drug discovery researchers is the HIV Reverse Transcriptase, and the credit partly goes to the high number of crystal structures available for the wild type as well as mutant protein. These structures enable researchers to dive deeper into the nuances of the RT binding pocket facilitating further improvements to existing drugs or discovery of new ones (Gu et al., 2014). The structure of the unliganded enzyme was used to screen derivatives of 4-thiazolidinone. Promising leads were identified in the experiment.

Another study screened a commercial library against two RT crystal structures. The drugs screened were labelled drug-like compounds based on Lipinski's rule and were ranked based on the scores. Seven hundred forty molecules were selected and tested experimentally and after ruling out low active molecules, 71 of the compounds showed inhibition of 85% or more. A total of 17 compounds with the highest activity were chosen for further evaluation. It was interesting to note that the selected compounds had elements similar to the phenylethylthiazolylthiourea(PETT) derivatives known to inhibit RT. Further, in the crystal

structure chosen to perform docking experiments, the native ligand is a PETT derivative, suggesting that these methods could be used to screen or retrieve molecules that bear a close resemblance to the native inhibitors in the structure chosen (Gu et al., 2014) (Santos et al., 2015) (Phillips et al., 2018).

Similar methods were implemented for identifying potential inhibitor candidates for blocking HIV entry into the host cell by targeting its co-receptors. A recent study reported a pharmacophore model that is highly specific for CXCR4 antagonists and could be exploited to develop molecules that could act as entry inhibitors. No approved drugs targeting CXCR4 exist until now, but many are in the preclinical and clinical trials. Unlike former, there are FDA approved drugs targeting CCR5 that act as entry inhibitors. It also has natural ligands, MIP-1β, MIP-1α and RANTES that bind and interact with it. The first approved drug in this category is Maraviroc and following this another small molecule Vicriviroc was developed and is in the Phase III clinical trials. The availability of crystal structures for the coreceptors in the last decade has enabled researchers to implement structure-based methods to identify potent compounds that could be developed into promising lead molecules (Phillips et al., 2018).

Limitations of HIV-1 Therapy:

Drug Resistance:

HIV incorporates a large number of mutations regularly because of the faulty proof-reading of Reverse Transcriptase and also its rapid replication rate (Pachamuthu et al., 2006). This acts in disguise for the virus to develop Drug-Resistant Mutations (DRM). When the virus mutates and reproduces, despite the presence of anti-retroviral drugs, the virus is said to be Drug Resistant (Mali et al., 2015). The consequences of drug resistance include a high rate of treatment failure, reduction in drug absorption and mainly the spread of resistant phenotypes among the population (Beyrer et al., 2017).

Primary Drug Resistance is resistance acquired by an individual when he is infected with resistant HIV-1 strains (Treatment Naïve), Secondary resistance is a resistance acquired by the individual due to long-term exposure to ART therapy (Treatment-Experienced). Pretreatment ARV is the resistance seen in drug naïve patients starting ART or restarting first-line treatment (Guha et al., 2012). This could be acquired or transmitted.

WHO has elucidated common causes for the development of Resistance as (Mali et al., 2015):

- Treatment with less than three drugs
- Prolonged treatment with a failed regimen
- Discontinuation of treatment
- Inappropriate drug selection
- Natural genetic variability
- Prescribing a single drug to a failing regimen
- Virus Fitness

All the drugs under the drug class of NRTIs select resistance mutations through two mechanisms. The first one is by the mutated reverse transcriptase enzyme selectively avoiding incorporating nucleotide analogues into the extending proviral DNA. The second one, usually seen in Thymidine analogues includes the phosphorolytic excision of incorporated nucleotide analogues on the 3'end, called primer unblocking. Some mutations that occur in the RT enzyme confer resistance to both classes of drugs (Rai et al., 2018).

The recently approved drug class INSTIs also leads to clinical failure due to acquired resistance mutations. The mutations conferring resistance to some of them are cross functional. One of the approved drug Dolutegravir is reported to have a high genetic barrier and has not yet been reported in treatment naïve patients. Primary mutations in Integrase usually occur in the active site and appear after initial exposure to INSTIs. Secondary mutations occur after prolonged exposure to drugs and appear to enhance the effects of primary mutations (Anstett et al., 2017).

Mutations that confer resistance to inhibitors of Protease occur both in active site as well as other allosteric sites. Some mutations have also been reported in gag cleavage sites. The ones that occur in the active site prevent binding of the PI thus enabling the cleavage of polyproteins resulting in the completion of viral lifecycle. These mutations are known as Primary mutations. The mutations that occur outside the active site are known as Secondary mutations and these alone do not confer resistance towards Protease Inhibitors. In presence of primary mutations or a cluster of secondary mutations lead to the decreased susceptibility to PIs. Secondary mutations usually occur to reverse the affect caused by the existence of primary mutations or to increase viral fitness. The third class of mutations that result in PI resistance occur outside the protein region and usually occur in combination with Primary and Secondary mutations. In cases such as these, the mutations occur to restore the Protease

Class	Inhibitors	Side Effects/ADR
NRTI	abacavir	Hypersensitivity, Cardiovascular Events
	emtricitabine	Hepatomegaly with steatosis, Lactic acidosis
	lamivudine	Pancreatitis, Hepatomegaly with steatosis, Lactic acidosis
	tenofovir	Renal impairment, Hepatomegaly, Drug Interactions
	zidovudine	Myopathy, Neutropenia, Anemia
NNRTI	doravirine	Immune Reconstitution Syndrome
	efavirenz	Hypersensitivity, Prolonged QTc interval, Psychiatric disorders
	nevirapine	Hepatotoxicity, Skin Disorders
	rilpivirine	Depression, Hypersensitivity, Hepatotoxicity
PI	atazanavir	Cardiac Toxicity, Hyperbilirubinemia, Renal impairment
	darunavir	Skin reactions, Hepatitis, Drug Interactions
	fosamprenavir	Hypersensitivity, Drug Interactions, Hyperglycemia
	ritonavir	Drug Interactions, Hepatotoxicity, Pancreatitis, Cardiac Issues.
	saquinavir	QT, PR elongation, Hyperglycemia, Hemophilia.
	tipranavir	Hepatotoxicity, Intracranial Hemorrhage
INSTI	dolutegravir	Hypersensitivity, Hepatotoxicity, Immune reconstitution syndrome
	raltegravir	Rhabdomyolysis, Myopathy
Fusion	enfuvirtide	Pneumonia, Hypersensitivity
CCR5	maraviroc	Cardiovascular events
Post-attachment	ibalizumab	Immune Reconstitution Inflammatory Syndrome

Tab 1: Lists the drugs of different classes approved for ART along with the ADR/ Side effects noted for each drug

activity and viral replicative capacity. However, these mutations have also been reported as the sole mechanism by which resistance to PIs have been conferred. These usually occur in the Gag cleavage sites thus being the only ones with resistance mechanism involving enzyme substrates (Baxter et al., 2016). The inhibitors that are approved to act against the HIV entry target the host co-receptor and gp41 subunit of the envelope protein. The resistance conferred to Co-receptor targeted inhibitors occurs via. One the following ways, i) Co-receptor switching from CCR5 to CXCR4 ii) Usage of inhibitor-bound co-receptor iii) Increased binding to co-receptor iv) Decreased time of HIV-Entry. Reports have shown that a stretch in the gp41 protein is prone to resistant mutations resulting in decreased affinity to entry inhibitor Enfuvirtide (Lobritz et al., 2010).

According to 2017 report by WHO, there is an increase in the prevalence of drug resistance in patients initiating first-line therapy mainly in African countries. In 2010 the prevalence in pretreatment ARV was at 6.8%, and the estimates in 2017 suggest the levels to be above 10% to the widely used first-line ART drugs (Drug Resistance Report 2017, WHO). The widespread use of thymidine analogues led to the development of resistance to these analogues known as Thymidine Associated Mutations (TAM), which also resulted in the development of cross-resistance among other NRTI drugs. Changes in regimen after 2012 resulted in the decreased prevalence of TAMs and the emergence of other resistance (Gregson et al., 2017).

Adverse Drug Reactions and Side Effects:

In the last couple of decades, the rollout of HAART has resulted in the decline of AIDS-related morbidity and mortality across the world. Despite the effectivity of ART, there still exist two major concerns, one being the circulation of multi-drug resistant strains, especially in the low resource settings and the prevalent side effects resulting from extended use of these drugs. The latter remains the major cause for non-adherence to treatment plans, which in turn results in the emergence of drug-resistant strains (Zhan et al., 2016). Some common Adverse reactions observed with the approved drugs are mentioned in Tab1.

Latent Reservoirs of Virus:

When HAART was initially rolled out, it sought after a complete cure for HIV (Zhan et al., 2016). However, it was later observed that even in patients surviving on treatment for more than 30 months, with higher CD_4+ count and negligible viral load, there existed reservoirs of CD_4+ cells that carried copies of competent virus (Persaud et al., 2000). These reservoirs

created during early stages of primary infection cannot be completely eradicated by the retroviral therapy (Pitman et al., 2015). These latently infected T-Cells have a half-life of 6-44 months and the activation of them can lead to replenishing of viral reservoirs (Finzi et al., 1997) . The presence of these latent reservoirs poses an obstacle in finding a complete cure for HIV.

Rationale & Objectives

As discussed, despite being effective in delaying the progression of AIDS and increasing the life span of individuals, the use of anti-retroviral agents is greatly obstructed by the emerging drug-resistant strains of HIV. Moreover, the necessity to continue the treatment for long term results in metabolic disorders and other toxicities. The resistance, toxicities and cost burden of the current treatment impact the treatment adherence. Hence there is a need to understand the prevalence of resistance in the country and also for better treatment and drugs. Based on this rationale, the following objectives have been framed.

Objective 1:

Prevalence of HIV-1 Reverse Transcriptase drug resistance in India.

Objective 2:

Design, Development and Validation of HIV-1 Entry Inhibitors.

Objective 3:

Insights into the binding pocket of HIV-1 Associated Topoisomerase II Beta Kinase using in-silico studies.

Objective 4:

Facilitating cellular localization of Topoisomerase II Beta Kinase inhibitors using Lactoferrin Nanoparticles.

CHAPTER II

MATERIALS

1. **Cell Lines:**

- SupT1 (Cat No: 100) Non-Hodgkin's T cell lymphoma, NIH AIDS Reference and Reagent program, USA.
- HL2/3 (Cat No: 1294) Cervical epithelial cell line, Hela modified to contain stably integrated copies of the HIV-1 molecular clone HXB2/3gpt obtained from NIH AIDS Reference and Reagent program, USA.
- U937 (ATCC® CRL1593.2™) Lymphocytic cell line derived from Histiocytic Lymphoma obtained from ATCC. These are pro-monocytic and can be differentiated in-vitro to macrophages.
- SK-N-SH (ATCC® HTB11™): Neuroblastoma cell line obtained from ATCC.

2. **Virus Strains:**

All virus strains used were laboratory adapted and the stocks were provided by NIH AIDS Reagent program, USA. The viruses used for the study include

- HIV-1 93RW024 (Cat No:2200) R5X4 tropic virus belonging to group M and subtype A of Rwanda origin.
- HIV-1 IIIB (Cat No:398) An R5X4 tropic virus belonging to group M, subtype B of US origin.
- HIV-1 85US_BA-L (Cat No:510) R5 tropic virus belonging to group M, subtype B of US origin.
- HIV-1 93IN101 (Cat No: 2900) A mixture of R5 and X4 tropic virus belonging to group M, subtype C of Indian origin.

3. **Cell Culture Materials:**

- Cell culture media: RPMI-1640 (Cat No: 31800-022), Dulbecco's modified Eagle medium (DMEM) (Cat No: 12500-062) and Minimal Essential Media (MEM) (Cat No: 61100-061) were purchased from Gibco, Thermo Fischer Scientific, USA
- Foetal bovine serum (FBS) (Cat No: 10270-106) was purchased from Gibco, Thermo Fischer Scientific, USA
- Trypsin EDTA 0.5% (Cat No: TCL034) was purchased from Himedia Laboratories Pvt Ltd. Non-Essential Amino acids (NEAA) (Cat No: 11140-035) and Sodium Pyruvate (Cat No: 11360-039) were purchased from Thermo Fischer Scientific, USA. Pencillin-Streptomycin antibiotic (Cat No: A018) was purchased from Himedia Laboratories Pvt Ltd.

- Cell culture flasks, dishes and tubes were purchased from Corning Inc. Life Sciences, MA, and USA and Axygen, Inc., CA, USA
- Multi-well cell culture plates and culture bottles were purchased from Corning Inc. Life Sciences, MA, and USA.

4. Reagents and Kits:

- MTT(3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) (Cat No: M6494) purchased from Invitrogen, USA.
- Calcein, AM, cell-permeant dye(Cat No: C3100MP) and Calcein Blue, AM dye (Cat No: C1429), DAPI Prolong Gold Antifade Reagent (Cat No: P36935) were procured from Thermo Fischer Scientific**,** USA .
- HIV-1 p24 Elisa Kit was procured from Advanced Bioscience Laboratories (ABL),Kensignton, MD, USA. (Cat No: 5447)
- PCR Reagents (Master mix, PCR grade water) were purchased from Invitrogen, USA.
- Culture grade DMSO (Cat No: C6295) was purchased from Sigma India(Merck).

5. Instrumentation:

- ELISA Plate reader supplied by HBiosciences Ltd. Model: SM600
- Fluorescent Microscope Leica
- TEM- Techni
- SEM- Carl ZEISS Ultra-55, Germany
- HPLC- Waters 2487
- CO_2 Incubator: Forma Scientific, Marietta, OH, USA
- Zetasizer : Malvern

6. Computational Tools & Software:

6.1 Sequence Retrieval, Subtyping, Alignment & Resistance Profiling:

- **ENTREZ DIRECT** (Edirect): NCBI's Edirect, to retrieve data from its repositories of publications, structures, sequences, gene, expression, variants, etc. It is accessible through UNIX terminal window by means of simple arguments (Kans, 2019) .
- **RIP v 3.0**: Recombinant Identification Program is a HIV subtyping tool hosted by HIV Los Almas National Laboratory (LANL) (SIEPEL et al., 1995).

- **REGA v 3.0**: REGA is a tool maintained by Stanford University under HIV Drug Resistance Database and developed in collaboration with the REGA institute at the Katholieke Universiteit Leuven, Leuven, Belgium and the HIV-1 Pathogenesis and Immunotherapeutics Program at University of Pretoria, South Africa. It is used in determination of HIV-1 subtypes based on sequences (Carlos, Luiz Alcantara et al., 2009).
- **BioEdit v7.0.5** : BioEdit is a user-friendly interface manual alignment editor. It is integrated with many internal as well as external tools for analysis of sequences/alignments, like ClustalW, BLAST, DNADist, etc. Originally developed at , North Carolina State University, many researchers contributed towards the development of further version (Hall, 1999).
- **HIV Drug Resistance Database(HIVdb) v 8.7:** HIVdb is developed and maintained by Stanford University as an important resource for clinicians, health care organizations and scientists across the world. HIVDB is a collection of drug resistance mutations that are identified by the research team at Stanford or user submitted mutations that have been validated (Rhee et al., 2003) (Shafer, 2006).
- **Calibrated Population Resistance (CPR) v 6.0**: CPR is a tool maintained by Stanford University HIV drug resistance database. It has been built to identify mutations in the Reverse Transcriptase, Integrase and Protease sequences that confer resistance to NRTI, NNRTI, INSTIs and PI (Liu et al., 2006).

6.2 Peptide structure determination, refinement and validation:

- **PEP-FOLD 3.0** is a tool that builds peptide structures based on input amino acid sequences based on a de novo approach. A Hidden Markov Model(HMM) derived, structural Alphabet(SA) of 27 patterns of 4 consecutive amino-acids is used by PEP-FOLD to describe proteins as overlapping fragments. To understand the process of organization of local conformations, the orientation of consecutive fragments in an alphabet are not independent on each other but follow a Markovian process(First order). PEP-FOLD learns the SA patterns profile from the given amino acid sequence. It is enabled as below,

Firstly, the SA patterns for each position are predicted based on the BLAST profiles obtained for the given amino acid sequence using PSI-BLAST (Altschul et al., 1997) against Uniref databank (Suzek et al., 2015). The program makes use of Support Vector Machine (SVM) that takes the PSI-BLAST profile of each four-residue amino-acid fragment with 2 amino-

acids extended on either side as input. Thus, a vector is generated to predict each SA pattern. The output of the SVM would be L-3*27 probabilities of each SA pattern fitting every fragment of the protein. The output of the SVM is used as an input to the HMM-forward-backward algorithm and HMM-SA model (Camproux et al., 2004), this forms the second step of the algorithm. Upon trying all the SA patterns at each position, the authors have found that 8 best SA patterns at each position were sufficient to represent protein's native conformation. To build the 3D structure of the protein from the obtained SA pattern profile, a greedy algorithm guided by modified coarse-grained force field, OPEP3.1 (Maupetit et al., 2007) is used. The greedy algorithms assemble structure starting at any random position and extending by 1 character on either side. After the 3D assembly of protein, the best model is refined using Monte-Carlo simulation. Post-refinement, each peptide is subjected to 50 greedy runs that start at different random positions, with randomness of extension at each position intact. Cluster Analysis is performed on the models thus obtained and the one with lowest energy centroid is returned (Thevenet et al., 2012).

With the new version of PEP-FOLD version 3.0, The 3D assembly from SA pattern profile is by using Taboo sampling algorithm that generates sub-optimal trajectory series and every such series leads to the generation of one conformation. The formation follows the same one residue at time generation and follows coarse grained force field, OPEP3.0. However, implementation of the new algorithm makes the tool ten times faster (Lamiable et al., 2016).

- **MODLOOP** is a tool developed to assist protein structure modelling by specifically assigning structure to loops of length less than or equal to 20 residues. As implemented by Modeller, prediction of loop conformation using optimization of objective function is also used in MODLOOP. This model uses the protocol consisting of minimizing with conjugate gradient and molecular dynamics with simulated annealing for optimization of all positions of non-hydrogen atoms of the input loop region. Bonds, few dihedral angles and bond angle terms in the pseudo-energy function are restrained based on the terms of CHARMM-22 function (Mackerell et al., 1998) . The dihedral angles corresponding to side chain and main chain along with the non-bonded atoms are based on the potentials mined from a large data set of known protein structures. This model generates structures that have been independently optimized starting from a randomly generated structure. The final reported structure would be the one with lowest energy score (Fiser et al., 2003).

- **MODREFINER** is a tool used in protein/peptide structure refinement. It follows a full-atomic energy minimization protocol for refinement of structures. This energy model consists of 9 terms

$$E_{full} = E_{restr} + W_1 E_{length} + W_2 E_{angle} + W_3 E_{rama} + W_4 E_{clash} + W_5 E_{hb} + W_6 E_{dfire} + W_7 E_{LJ} + W_8 E_{rot}$$

The terms $W_1 – W_8$ are weights of each energy term with E_{restr} being the base energy term. These are determined by weight optimizations done from a set of 262 nonhomologous globular proteins selected from PISCES list. The initial structures for some selected proteins (training set) is done using I-TASSER, for which homologous structures were not available in the template library. The refinement of these structures was done using ModRefiner program and the weights were determined from super-dimensional grid system. The model quality was maximized and the weights for each energy term was accordingly adjusted. The initial weights were set to zero and were continuously modified until the value of the coupled term (W_iEj) had no effect on the model quality. Finally, the optimized best weights were fixed for the equation.

The energy terms, E_{restr} as suggested previously is the base energy term that defined the absolute difference between C_α distance of consecutive amino acids in refined and reference models. The terms E_{length} and E_{angle} are the number of outliers in terms of bond length and angle as defined by Engh and Huber (Engh et al., 1991) ,E_{rama} is the number of outliers in terms of torsion angles of backbone pairs as per Ramachandran Plot, ,E_{clash} is the sum of clases between each pair of atoms. ,E_{hb} is the total number of H-bonds between main chain-side chain, main chain - main chain, side chain - side chain ,E_{dfire} is the potential calculated pair-wise according to DFIRE (Zhou et al., 2002).E_{LJ} is the pair-wise Leonard Jones 12-6 potential. E_{rot} is the side angle χ angle distributions calculated from PDB structures. In order to save CPU time, instead of calculating energy of the refined model from scratch the energy at each step is calculated based on the energy of the previous model. The energy of the previous model plus the energy change caused due to the movement of the model. After each step, only the atoms of the residues that have changed due to the refinement are checked and the energy difference between the models before and after the refinement step is calculated.

The total energy of the current model would be the sum of difference calculated and the total energy of the previous model (Xu et al., 2011).

- **RAMPAGE** is a tool used for validating protein structures that have been computationally determined or deduced by X-Ray crystallography or NMR analysis. It is based on a large set of protein structures derived from PDB using specific criteria, such as B factor, resolution etc. RAMPAGE works on the idea that φ, ψ torsion angles are not independent entities and hence need to be analysed together. The data obtained from the known structure is used to plot the Ramachandran Plot with contours for allowed and favourable regions. The data points on the plot are represented by percentage i.e 85% Contour suggests that 85% of datapoints fall in the contour area. Upon calculating the torsion angles as suggested by Lovell et.al. for the input, the validation for the structures is reported (Lovell et al., 2003).

6.3 Molecular sketching, energy minimization and Visualization

- **ChemSketch**: Managed by Advanced Chemistry Development, Inc Labs, is a package that enables drawing of chemical structures (organic, organo-mettalic, Markush etc). The package also has capabilities to 2D and 3D structure viewing, cleaning, labelling and logP prediction. The sketched molecules can be stored in numerous formats including mol, sdf etc.

- **Biovia Discovery Studio visualizer:** Managed and distributed by Dassault Systems, DS Visualizer is a tool developed for graphics visualization used for the purpose of 3D structure viewing, analysing and sharing of protein structures and sketching and cleaning of small molecules. It has capabilities to analyse protein-ligand complexes, intra and inter molecular interactions, bond distances, angles etc. It supports many file formats such as pdb, sybyl mol2 etc.

- **PyMol v2.0:** Maintained and distributed by Schrodinger**,** it is a popular visualization tool with backend python scripts is used commonly to visualize protein structures.

- **Tripos Sybyl-X v8.0:** Managed and distributed by Certara**,** Sybyl-X is a molecular modelling and simulation suite with integrated computer-aided drug design tools. It has

capabilities to perform energy minimization of designed small molecules based on various minimization algorithms and forcefields. It can also perform molecular dynamics simulations, Docking and QSAR methodologies.

6.4 Virtual Screening and Molecular Databases

- **PubChem:** Maintained by National Institute of Health-National Centre of Biotechnology Information (NCBI), Pubmed is a public database of ~ 97,00,000 chemical compounds, ~ 258,00,000 substances along with other data regarding Bioassays conducted for the compounds, Activities of the compounds, respective protein and gene targets. The search options for the database include structure search and Entrez search. The database can return structures similar to the query based on similarity measures and relevant biological activity data can be accessed for the respective structures (Kim et al., 2016).

- **Pep:MMs:MIMIC:** Developed by Matteo Floris and Stefano Moro, it is web-based tool that generates small molecule equivalents, in terms of 3D conformation, through conformational similarity search, given 3D structure of peptides and their interaction with proteins. Briefly, the algorithm starts with identifying key residues of interaction in a given peptide-protein pair. This is followed by reducing the complexity of peptide by replacing it by the pharmacophore model defined by a set of structural features derived from the interactions and key resides. All possible structural feature arrangements of the pharmacophore are considered, and the model is depicted as a bit string. Once this is generated it is used to screen multi conformational version of the database MMsINC (a database of 4 million energetically minimized 3D chemical compounds), the multi conformational version of the database (multi-confMMsINC) consists of 17 million low energy conformers generated from every entry of the MMsINC database. Scoring functions are used for refining the top mimetic compounds obtained (Floris et al., 2011).

The main steps are as follows

Conformer generation: MultI-conf MMsINC: Rotate ver1.0 is used to generate best five lowest energy conformers of each entry of MMsINC resulting in the generation of ensemble of conformers for the whole database.

Pharmacophore generation: Pharmacophoric features tyrosine side chain, tryptophan side chain, , histidine side chain; H-bond donor, H-bond acceptor, negatively ionisable, positively ionisable, hydrophobic and aromatic, are used for labelling all amino acid residues. The annotation ensembles of peptides and the conformers were described in terms of three-point pharmacophore, where each possible centroid is categorized based on feature distance. Two main criteria are considered for encoding the three-point pharmacophore into pep-MMs-MIMIC bitstring. These are atom type and centroid distance. Nine centroid types have been defined by pep-MMs-MIMIC and all possible triplet combinations can encoded by their respective bit strings. Centroids are continuous set of atoms that belong to the same pharmacophoric type i.e in six carbon atoms of Benzene can be binned into aromatic centroid with the centroid lying in the middle of the ring. In the second criterion, according to FuzCav method (ref), a distance cutoff was used and bins were defined for distanced between all centroids in a pharmacophoric triplet with each bin defining the vertices consisting of the three-point centroids and the relative distance between each triplet pair. The same criterions are adopted to represent both conformers as well as peptides.

USR metodology

USR is a rapid 3D search method that assumes that the structure of a molecule depends on the relative positioning of its atoms. According to this methodology, any molecule or an ensemble of atoms can be defined by , molecular centroid(ctd), closest atom to centroid (cst), farthest atom to centroid (fct), farthest atom to fct (ftf). Each of these can be defined by three vectors and thus a total of 12 shape descriptors can be used to define the molecules. Pep-MMs-MIMIC has implemented a USR method for similarity searching (Floris et al., 2011).

Scoring

A weighted similarity index (S_W) is used as a measure of similarity between fingerprints (bitstrings), defined as

$$S_w = c/(c + 2.5 \times m)$$

Where *c* represents the shared bits between the query peptide fingerprint and the fingerprint of the conformer and *m* is the total number of bits on the query. Along with S_w , three other

metrics are used for scoring. Shape score (ShS) based on USR; weighted score (S_w) based pharmacophoric similarity (PFS); combined ShS and PFS where ShS is used for filtering followed by sorting using PFS; A hybrid scoring function where weighted ShS and PFS are implemented based on S_w (Floris et al., 2011).

6.5 Molecular Docking and Protein- Peptide docking

- **HEX** : HEX is a Fourier Transform based docking algorithm developed for protein-protein, protein-peptide docking. It is a compilation of number of Shell scripts and Web pages that communicate via MySQL database. The algorithms implemented for the docking have been explained in great detail in the following publications (Ghoorah et al., 2013) (Macindoe et al., 2010).

- **GOLD (2015) :** Maintained and distributed by CCDC, GOLD is a well-accepted and widely used molecular docking tool that implements Genetic Algorithm. The program requires the user to specify the binding site in terms of its location and size. The location of the binding site is specified with ORIGIN and the size is specified with RADIUS. Flood-fill algorithm, described by Ho. et.al, is used to identify SAS area within the distance starting from ORIGIN across the RADIUS specified by the user. This is followed by a cavity detection algorithm developed by Delany et.al, is implemented to identify concave pocket where the ligand can interact with the protein. Hydrogen bond donor, acceptors in the cavity are identified by SYBYL atom-type characterization, given by Clark.et.al.

In the designated cavity area, the terminal hydrogen bond donor and acceptors connected to the protein by a single bond are treated as rotatable and thus are allowed to rotate into positions that can participate in binding with the ligand. A binding point at a distance of 2.9A^o is marked, from the terminal donor and acceptor atoms that are considered rigid. If this binding point lies within the cavity detected as per the RADIUS specified, it is considered available for the ligand binding. The ligand was minimized prior to the application of docking algorithm, lone pairs were added in appropriate positions and following geometry. All single, acyclic non-terminal bonds were considered rotatable and to avoid initial conformational bias, random rotations and translations are introduced into the rotatable bonds to generate conformers.

The conformations of the protein and the ligand were encoded into bitstrings which were designated as chromosomes. Each byte of the bitstring represented an angle of rotation across

the rotatable bond. Apart from the bitstrings, two integer strings were introduced to represent the mapping of hydrogen bond donor and acceptors between the protein and ligand pairs. The fitness of the chromosomes generated after each run are scored based on a function. This was evaluated in the following stages a) Firstly, the conformations of the protein's active site and the ligand was generated, b) A least square fitting was used to place the ligand within the detected active site, c) H bond energy: A bind energy corresponding to the Hydrogen bonding was calculated for the complex, d) Complex Energy: Energy of the complex corresponding to the pairwise steric interactions was obtained, e) Internal Energy :Molecular mechanics was used to come up with the energy term that measures ligand's internal energy, f) Final fitness score : The energy terms were summed to get final score.

The term H_Bond_energy was determined by summing individual bond energies of all possible combinations of H bond acceptors of the ligand and H bond donors of the protein active site and all possible combinations of H bond donors of ligand and H bond acceptors of active site. To replicate the interaction energy the following terms are used

$$E_{pair} = (E_{da} + E_{ww}) - (E_{dw} + E_{aw})$$

Where, *d* is the hydrogen bond donor, *a* is the bond acceptor, E_{da} is the energy of interaction when donor and acceptor form a bond while displacing water, *w*.

The term Complex_ Energy, describing the interaction energy, was determined by calculating 4-8 potential described by Jones et.al, after placing the ligand in the active site. The 4-8 potential is given by

$$E_{ij} = \frac{A}{d_{ij}^{8}} - \frac{B}{d_{ij}^{4}}$$

The term Internal_Energy is calculated as a sum of ligand's torsional and steric energies. 6-12 potential is used to calculate the steric energy and is of the form

$$E_{ij} = \frac{C}{d_{ij}^{12}} - \frac{D}{d_{ij}^{16}}$$

The terms A,B,C and D are calculated as given in (ref). The final fitness was determined by the sum of all the above terms as:

$$-H\_Bond\_Energy\ -(Complex\_Energy + Internal\_Energy)$$

The manipulation of the chromosomes happens in an island model, where the chromosomes are divided into several sub populations and the migration of chromosome happens between these populations. 5 populations with 100 individuals in each population were used. The Three genetic operators, cross over, migration and mutation are used to generate chromosomes in each generation. The operators are determined by roulette wheel method with mutation and cross over appearing with equal probability and migration happening 5% of the times. The algorithm is terminated after 100,000 genetic operations are applied, finally reporting highest scored docking.

- **MGL Tools 1.5.4:** A visualization and analysis software developed at Scripps Research Institute, Molecular Graphics Laboratory. The three main functions that can be performed with MGL Tools include, Docking(AutoDock Tools), Molecular structure visualization with varying capabilities(PyMol Veiwer) (Sanner, 2009), Visualization of data by building networks (Vision). The Autodock Tools integrated into MGL Tools, enables user to pre-process receptor and ligand, define binding site, select interacting residues, perform docking and visualize and analyze the docked results (Morris et al., 2010).

- **Autodock Vina:** A modified, and refined version of docking algorithm, developed from AUTODOCK 4.0. The new algorithm is approximately two times faster and more accurate than the previous version and enabling parallelization further increases the speed of the program.

AUTODOCK 4.0 (Morris et al., 2010) handles the problem of searching the conformational space of the protein available for the ligand by a grid-based method. This involves defined a grid around the protein target followed by a probe atom being introduced at each grid point and storing the interaction energies in the grid. The energies stored are used as a look up table when a ligand is introduced for docking. A hybrid Lamarckian Genetic Algorithm is

employed for searching through the conformational space. A global search is enabled by GA and local search by LA (Morris et al., 1998) which is adaptive in a way that the step size is adjusted based on the history of energies and other user-defines parameters. An initial population of conformations are created which mutate, exchange parameters and compete to survive like in Evolution. AUTODOCK follows approaches similar to other Genetic Algorithms but a Lamarckian aspect is introduced that allows individuals to explore their conformational space in each generation and pass on information to next generation. The chromosomes in AUTODOCK are not bitstrings but are represented by strings of real-value, three co-ordinate values, four variables for ligand orientations and one real-valued parameter for each torsion angle. Torsion angles are created by the user, through AUTOTORS, torsion tree that enables selection of rotatable bonds in the ligand.

The metric that is used to rank docked ligands and poses is Binding energy in Kcal/mol and is calculated using semi-empirical scoring method. The method is based on a thermodynamic model that incorporates intramolecular energies to predict energy of binding. A charge based desolvation model is also introduced. The energy function is given by

$$\Delta G = \Delta G_{vdW} \sum_{i,j} \left(\frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{i,j}}{r_{i,j}^{12}} - \frac{D_{i,j}}{r_{i,j}^{10}} \right) + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}}$$
$$+ \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i,j} \left(S_i V_j - S_j V_i \right) e^{\left(-r_{ij}^2 - 2\sigma^2 \right)}$$

Where, the terms ΔG on the R.H.S of the equation are derived by a regression model fit to a set of protein-ligand complexes with known constants. The other terms represent a 12-6 dispersion repulsion Lennard Jones term; a 12-10 directional hydrogen-bond term where E(t) is a directional weight term, and a electrostatic Coulumbic term. (Morris et al., 1998)

Vina derives its scoring function from a dataset obtained from PDBbind and is implemented in a manner similar to X-Score. The conformational dependent part of the scoring function is given as

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij})$$

while, the calculation is over all the pairs of atoms that move relative to each other, excluding the 1-4 interactions. Atom *i* is assigned to type t_i with pair-wise interaction function $f_{t_i} f_{t_j}$ at a distance r_{ij}. In other terms, it can be defined as

$$c = c_{inter} + c_{intra}$$

Vina ranks the conformations based on this equation and calculates the free energy of binding based on the intermolecular interactions of the lowest energy conformation, given by the below function, where, g is an arbitrary non-linear function

$$s_1 = g(c_1 - c_{intra1})$$

The advantages of the scoring function used in Vina is that it combines the goodness of knowledge-based and empirical scoring functions i.e it considers the preferences of known protein-ligand complexes taken from PDBbind and also takes experimental binding affinity values. Apart from the scoring function, Vina also employs a different optimization technique in comparison to AUTODOCK. In the method implemented, the algorithm used successive steps with each mutation followed by a local optimization and at each step, a acceptance criteria based on Metropolis principle is applied. Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is adapted for the local optimization. The advantage of BFGS method is that it takes gradient of each step along with the scoring function values to determine the next position, orientation and torsion values for the rotatable bond of the ligand and active site residues. This speed up the process of optimization.

AUTODOCK Vina can be executed through a command on LINUX and WINDOWS OS. It can be automated to perform docking on a set of ligands efficiently and in less time (Trott et al., 2011).

6.7 Quantitative Structure Activity Relations & Molecular Dynamics

- **Sybyl COMFA**: Comparative Molecular Field Analysis(COMFA) is a method to develop a 3D-Quantitative Structure Activity relation for given set of ligands and their experimental activities. The algorithm has pre-set rules that need to be followed before proceeding with the protocol. One such pre-condition is the requirement of the ligands to be interacting with a similar kind of receptor and in a similar mechanism i.e. a similar binding site with a relative geometry. The next step requires the user to define a sub-set of molecules as a training set to develop the COMFA model. The left-out molecules end up in the test set which would be used to check the validity of the model. The molecules used for the model-development and validation need to be processed to obtain low-energy conformations of the structures. Defining a scaffold for the selected set of molecules is another prerequisite that represents a common structure in the molecules.

The development of the model starts with the super positioning of the molecules to obtain a pharmacophore to represent the molecules in the dataset. Defining a grid distance, a large grid box is placed around the molecules to enable introduction of probe atoms. The probe atoms include, a Carbon atom, a positively charged atom, negatively charged atom, a hydrogen bond donor or acceptor atom and a lipophilic probe. These are used to calculate energies that the probe would experience at each grid pint based on the atoms on the molecules in the grid box. These values recorded, quantitatively represent the molecules. These values/ fields are recorded in columns that could be around 1000s in number and need to be correlated to the biological activity of the molecules.

Regression analysis is implemented to draw a correlation between the activity values and the derived field values. Partial Least Square(PLS) regression is widely used by COMFA protocol and cross-validation is used for confirming the accuracy of prediction of the model. The output of the model is a regression equation with thousands of variables which is represented in the form of a contour maps. These maps elucidate the favourable and unfavourable regions for steric substitutions on the common scaffold, similarly favourable and unfavourable regions for electropositive and electronegative substitutions are elucidated. These maps can be used as a guide to design and develop molecules with probable high activities. The prediction of activities for test set and for any newly designed molecules can be done using the developed PLS model (Kim, 1995).

- **GROMACS v 5.0:** GROningen MAchine for Chemical Simulations is a molecular dynamics suite that simulates Newtonian motion equations for a cluster of millions of particles. GROMACS offers high performance and is user-friendly. It can be easily executed from command line with input files and options to specify output files. GROMACS supports all possible algorithms for the implementation of MD. It supports energy minimization algorithms and has inbuilt topology builder for proteins. A basic trajectory analysis tool and a host of trajectory analysis tools make it possible to analyse the final out of the MD run. It has been regularly used to obtain a as an initial step in pre-processing of receptor molecules(proteins) before performing any operations on the structures, both X-Ray structures and computationally derived structures (James et al., 2015).

Programming and scripting

- **LINUX Shell scripting**

CHAPTER III

OBJECTIVE 1 : Prevalence of HIV-1 Reverse Transcriptase drug resistance in India.

Introduction

Despite being the most prevalent in US and the cause of 60% infections in Europe , subtype B only totals to a sum of 11% of the prevalent virus across the world (Pyne et al., 2013). It has been reported that several mutations leading to Drug resistance occur in specific subtypes only. This has been attributed to differences in codon usage (Chaplin et al., 2011). For example, an NNRTI mutation V106M has been reported to occur often in subtype C than B in patients experienced with NVP or EFV because the mutation requires a single change in the codon for C subtype whereas, in subtype B it requires two base-pair change (Wainberg et al., 2010). Many such examples have been reported, and these bring light onto the importance of subtype on the prevalence of drug resistance .

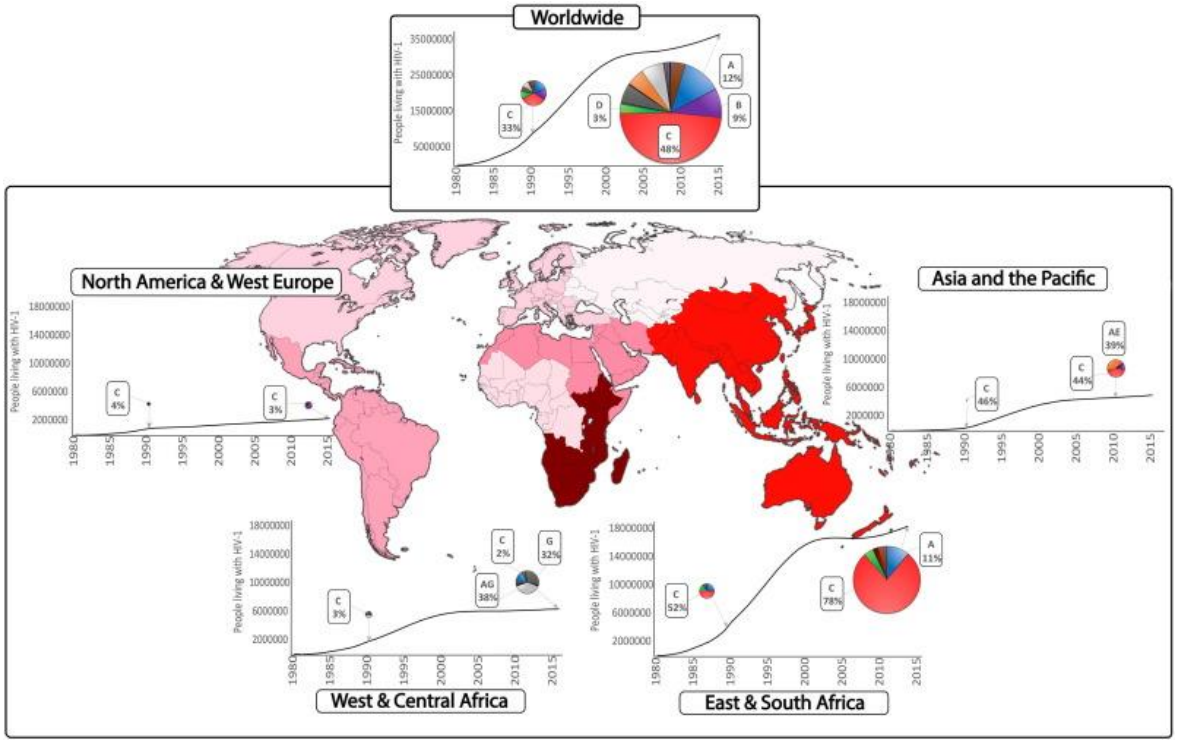


Fig. 5: Depicts the prevalence of subtypes in various parts of the world and the increasing trend of PLWH. Subtype C dominates the world-wide statistics with 48% of the PLWH being infected with this subtype.(Venner et al., 2016)

Despite Subtype C being predominant across the world, information about subtype specific Drug-Resistant mutations is not readily available(Guha et al., 2012). Studying DRMs in developing countries is of great importance as the chances of developing resistance is higher in these populations. The drop-out rates from the therapy and non-adherence, play a major role in developing resistance, and this is seen commonly in countries where the infections are

prevalent in lower income class group (Gupta et al., 2017). Same is the scenario in India where the awareness is a major drawback (Kumar et al., 2019) adding to which the discontinuation of therapy makes it difficult to control and monitor infection. In addition to this, the social stigma associated with the disease inhibits an infected individual to seek help and allow follow up (Kumar et al., 2019).

The limited resource settings in India forces implementation of a standardized approach to prescribing ART to facilitate rapid availability of treatment. Meanwhile, an individual patient management system is followed in developed countries that allows for follow up through routine monitoring (Dutta et al., 2017). Routine viral load monitoring is rare in Indian ART centers, and treatment failure is usually determined by immunological or clinical failure, and once that is determined, viral load is checked (Dutta et al., 2017)(Saravanan et al., 2017). The virological failure happens much before clinical/immunological failure, thus resulting in accumulation of DRMs while the failed regimen is still under use (Saravanan et al., 2017). Accumulation of these mutations compromises the effectiveness of treatment, use of subsequent regimens and carries a risk of transmission to newly infected patients resulting in health consequences to the public. WHO guidelines specified that the resistance that has been acquired by transmission (primary drug resistance) should be as low as 5% for the country to be declared free from resistance (Bennett et al., 2006).

To understand this prevalence in our country, an in-depth analysis of the DRMs should be made. Drug Resistance Genotyping studies for enrolled patients gives an insight into the mutations an individual developed for ART. This is difficult in India due to the high cost and loss in the follow-up of the patients. However, some isolated studies have been performed in the country, and the sequences have been uploaded in the GenBank Database, which is publicly available. The database also contains sequences that have not been published. We performed secondary analysis on this data to understand the prevalence of Reverse Transcriptase drug resistance in the country and its implications on ART treatment.

Methodology

1. Data Retrieval

NCBI search was used to locate HIV-1 data from Indian origin from NCBI linked databases. The search revealed records from various sources; Research articles from PubMed, Gene records, protein records, whole genome data etc. Filtering of data belonging to Nucleotide database listed a large number of sequences of HIV-1 including, gag, pol, env and whole length sequences. NCBI E-direct functionalities were used for accessing and retrieving sequences in bulk by setting appropriate filters. The following functionalities were applied to access and retrieve the sequences. E-search was used to perform an Entrez search using specific search terms. E-filter was used to define restrictions to the search terms to fine tune the results. E-fetch was used to download the records/files searched by Entrez in the designated format. E-link was used to retrieve records from databases related to the query database. The instructions were pipelined to construct a command.

The NCBI database was searched using the search term *HIV-1* and filters set were *India, Reverse Transcriptase* and years from *2005-2015*. These were piped with E-fetch to download the sequences

esearch -db nuccore -query "HIV-1 AND India" | *efilter -query "Reverse Transcriptase"* | *efetch -format fasta*

The output obtained was redirected to a text file with all the sequences in fasta format. A total of 7558 reverse transcriptase sequences were retrieved from India. After filtering based on preset criteria, E-link was used to obtain linked PubMed records for the retained sequences to understand the origin and the studies conducted on them. A Perl script was constructed to obtain sequence Ids and the respective PubMed records.

esearch -db "nucleotide" -query "ID retrieved from fasta sequence" | *elink -target pubmed* | *efetch -format Abstract*

2. Subtyping

Sequence-based subtyping was performed to understand the most prevalent HIV-1 subtype in the country. To avoid discrepancies and to increase confidence levels of the output, two different tools that analyze the sequences for determining subtype were used. Recombinant Identification Program 3.0 (RIP), developed by HIV LANL and REGA maintained by Stanford Drug resistance database, were used for the analysis.

2.1. RIP: The sequences to be analyzed were uploaded onto the RIP server in bulk. The default alignment of subtypes A-K was used in the background for the analysis. A window size of 100 and a confidence interval of 90% was selected. Any gaps in the data were stripped before analysis. Minimum length of 200 was considered when uploading sequences.

2.2. REGA: REGA uses phylogenetic analysis to categorize sequences into subtypes. The sequences were uploaded onto the server in bulk and were submitted for analysis. REGA considers pure subtypes as well as circulatory recombinant factors while constructing a master alignment and subsequent phylogenetic tree. Since the tool accepts only 2000 sequences per run, the data was analyzed in sets of three. The results were analyzed by coalescing outputs of the three runs.

3. Filtering and Trimming Sequences

The sequences were preprocessed before proceeding with the drug resistance mutation analysis to ensure uniformity in the data. The sequences reported to be of subtype C were selected for analysis and were imported into BioEdit Sequence Alignment Editor to visualize, edit and analyze sequences. The sequences were aligned with the Subtype C reference, 93IN101 and PMJ-4 using CLUSTAL-W v1.4. The matches were set at score 3, and a purine-purine or pyrimidine-pyrimidine substitution was set to 1. Any other substitution was set to 0. The k-tup value was set at default 4. Post-alignment, the region consisting of nucleotides from 2550-3870 (93IN101) was selected for analysis as the region was intact in most of the sequences. The region lying within the selected nucleotides belongs to Reverse Transcriptase region of the pol gene. The sequences were trimmed accordingly and those that were too short or too long were discarded from the study.

4. Metadata

The PubMed articles corresponding to the sequences, obtained through the Entrez system, were manually curated to extract relevant information. Microsoft Excel 6.0 was used to segregate and store data from each article for later use. Information about the aim, Number of patient samples, the year of study, Drug exposure status of the patients, patient sex and geographical location were extracted and recorded.

5. Distribution of Sequences

The data was sorted to understand the distribution of sequences based on year and location. The sequences collected were submitted in the span of ten years, from 2005 to 2015. They were segregated and plotted according to the year. Similarly, data was also segregated based on location, and after segregation, the source of the data could be divided into four geographical areas, East, West, North and South. For studies where the location was not available, the sequences were grouped under Unknown Region. To study the prevalence from 2015 to 2018, literature survey was conducted.

6. Identification of Drug Resistance Mutations

Calibrated Population Resistance (CPR) tool provided by Stanford Drug Resistance database was used to identify mutations in the sequences responsible for drug resistance. FASTA formatted pol (RT selected region) nucleotide sequences were submitted as a query. Maximum of 500 sequences were processed with each run. The list from the latest DRM surveillance, 2009 Surveillance of Drug Resistance Mutations (SDRM2009) was used to perform the analysis. The submitted sequences were first aligned to the Subtype B reference sequence using a local alignment algorithm. Upon setting the reading frame, the sequences are translated into protein sequences as per the codons. Upon confirming the quality of the sequences, the mutations were reported and categorized into five groups. a) Sequences containing at least one NRTI, NNRTI SDRMs b) Sequences containing at least one NRTI SDRMs c) Sequences containing at least one NNRTI SDRMs d) Sequences containing both NRTI and NNRTI SDRMs. The mutations reported were matched with the commonly used NRTI/NNRTI drugs used in the first and second line of treatments in India.

7. Identification of Polymorphism

The filtered and trimmed sequences aligned using CLUSTAL-W in BioEdit were imported to generate a Positional Amino Acid Numerical Summary matrix. The matrix reports the position of amino acid and the number of polymorphisms observed at the specified position. The positions with a higher number of repeated polymorphisms were retrieved from the matrix, identified and analyzed.

8. Data Representation

The analysis of sequences was presented as a percentage of the total number of sequences analyzed. The HIV-1 subtype, geographical and year wise distribution of sequences are represented through pie-charts. The region wise distribution of DRM observed is represented as bar charts. The DRMs were tabulated with the number of sequences observed for each class/type of mutations divided according to geographical location to explore region-based differences in the mutations. The mutations observed were analyzed discretely for drug naïve, and drug-experienced patients to identify mutations dominant in the two classes of patients. The results were analyzed to understand the drugs to which resistance is predominant and the analysis was tabulated.

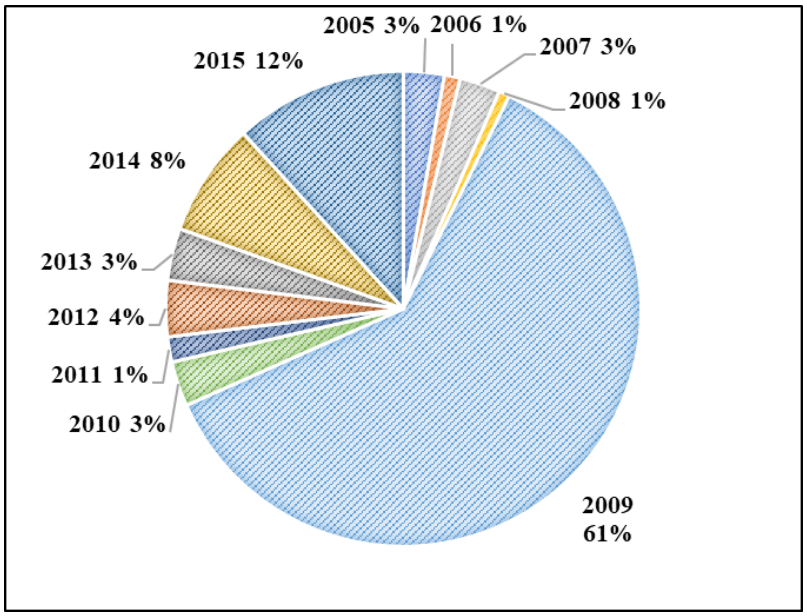


Fig 6: The graph depicts the year wise sequence collection/submission with highest being in 2009 (61%) followed by 2015 (12%) and 2014 (8%).

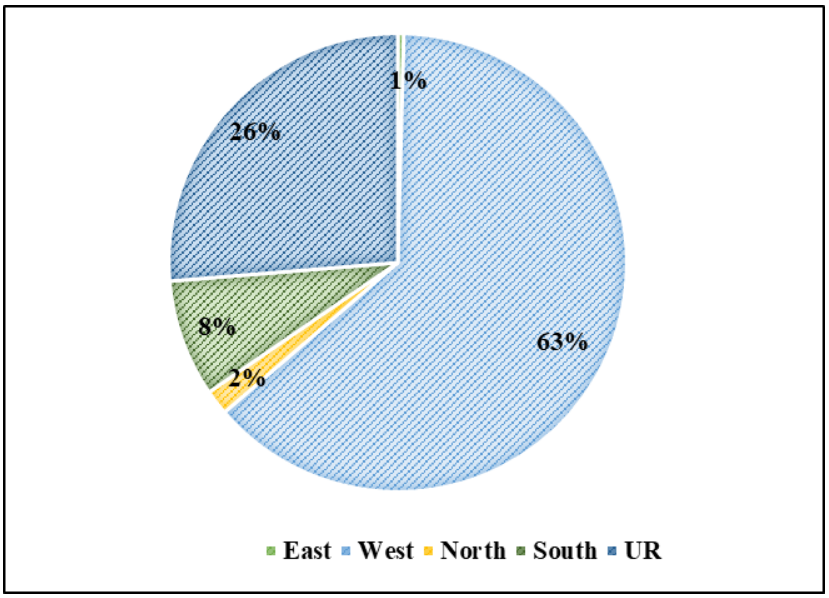


Fig. 7: The graph depicts the region wise sequence collection/submission with highest being from West (63%) followed by South (8%) and North (2%). The sequences with no information regarding the locations were grouped under Unknown region.

Results and Discussion

1. Distribution of sequences

The first step in analyzing the sequences is to understand the distribution of the data. The year-wise distribution of the curated data reported the highest number of sequence submissions (61%) in 2009. This was followed by the years 2014 (8%) and 2015 (12%). The results are depicted as a graph and represented in **Fig 6**. Location wise distribution of this dataset revealed the predominant contribution by the Western Region with a maximum number of sequences being uploaded and samples collected (63%), followed by the South (8%). Results depicted in **Fig 7.**

This pattern of distribution is seen, as the National AIDS Research Institute (NARI) with its large research contribution is situated in Pune, Maharashtra categorized in this study as West. Other contributors to the data from West are Sir JJ Hospital, Mumbai; Grant Medical College, Mumbai and Command Hospital, Pune. The number of metropolitan cities in the South with access to health care for patients and a large number of research centres specializing in HIV/AIDS is reflected in the number of sequences uploaded from this region. YRG Care centre, Chennai; Christian Medical College, Vellore; St. John's National Academy of Health Sciences, Bangalore and Tuberculosis Research Centre, Chennai are the contributors for sequences from the Southern Region of the country. The unequal distribution of sequences according to region shows the negligible work being conducted in parts of the country, other than West (NACO, NIIHAR institutes).

India and its socio-economic and cultural diversity influence the evolution of disease pathogens (Duff, 2014). To accommodate and to analyze the region-based differences in the DRMs, further analysis of the data was conducted region-wise.

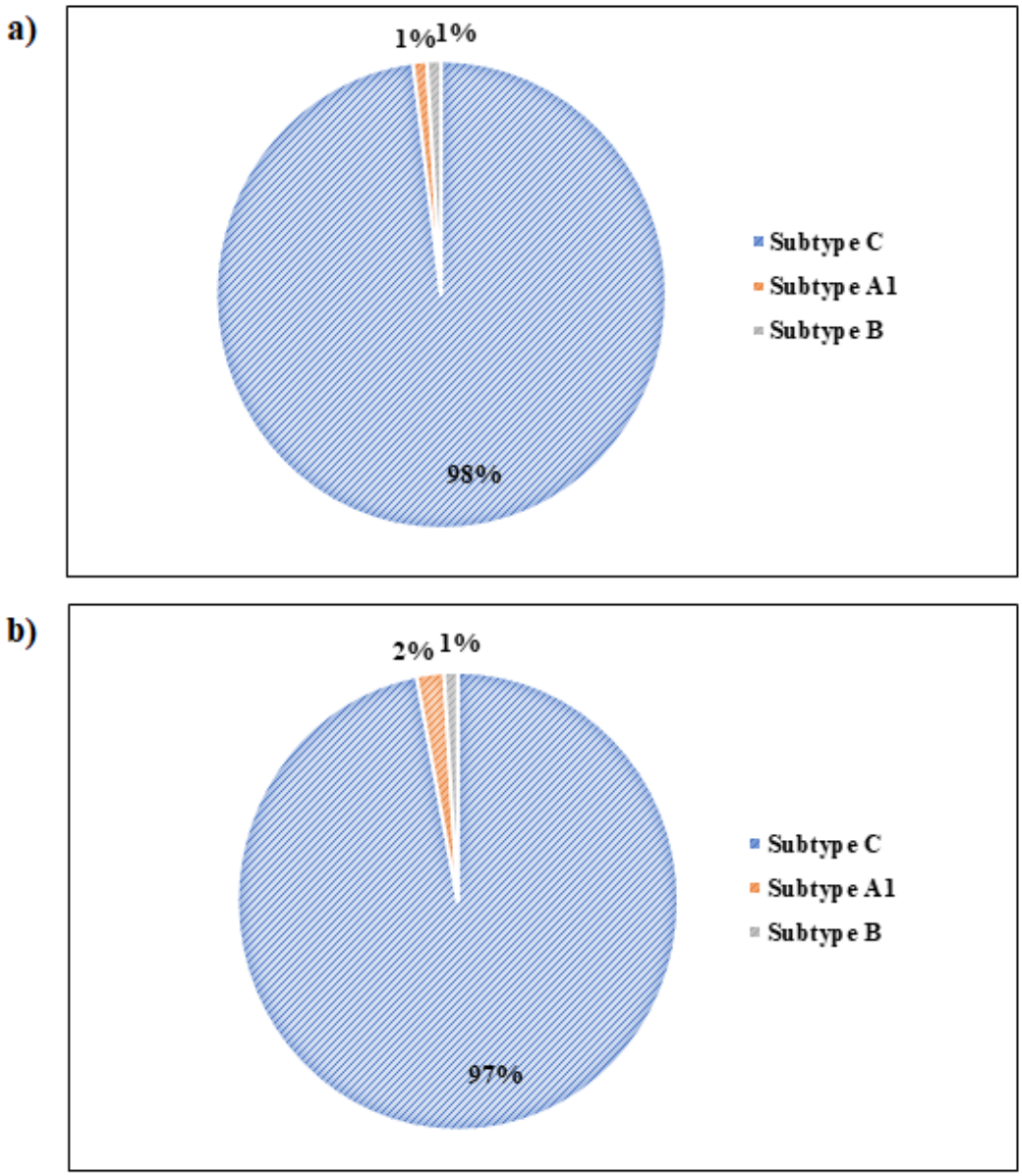


Fig. 8: The graphs depicts the reported subtypes of the sequences by programs a) Recombinant Identification Package (RIP) and b) REGA tool. 98% and 97% of the total sequences were reported to be of Subtype C by the tools respectively.

2. Subtyping

Subtyping analysis conducted on the sequences revealed a dominance of Subtype C with 98% by RIP (**Fig 8a**) and 97% by REGA (**Fig 8b**) reporting type C. Other subtypes observed included A1, B, D, H, CRF_01 AE etc. but in small number. A clear overlap between the results reported by both the tools was seen. This allowed discarding ambiguities while selecting sequences of a single subtype for further analysis. Despite Subtype C being prevalent in India, the sequences diverge from type C sequences seen in Africa and other Subtype C dominant countries. Results of the current study aligned with many research/review/clinical reports from India.

A study conducted by Neogi et al., 2012 at Karolinska Institute, Sweden studied the origin and spread of Subtype C variant of HIV-1 in India. The study analyzed sequences obtained from seven different states of the country to understand the divergence of the virus. The study reported the dominance of Subtype C strains when a single gene was examined, however when three genes pol, env and gag regions were analyzed, an increase in the percentage of recombinant strains was observed. Among the geographical regions studied, North East showed the highest prevalence of recombinant strains when more than one gene was analyzed. Other subtypes seen in the populations include types A1 and B. A study conducted by Rodriguez et al., 2009 examined sequences from different regions of the country to understand the prevalence of subtypes and to examine the sequences in relation to Subtype C sequences from other parts of the world. The study reported the dominance of Subtype C, followed by Subtypes B and A1. The group suggested that HIV-1 Subtype C virus is genetically distinct from those in Africa and South America. The study also showed increased viral fitness and higher transmissibility of Indian Subtype C in comparison to other C subtypes and other subtypes in general.

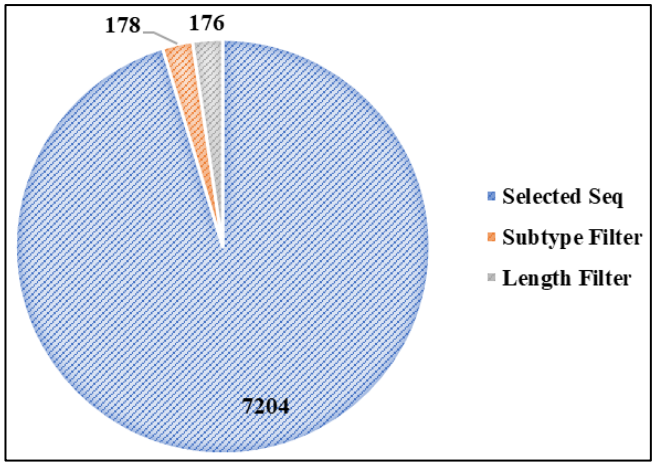


Fig 9: The pie chart depicts the selected sequences (7204) after discarding sequences from other subtypes (178) and sequences of length shorter than 50 nucleotides (176).

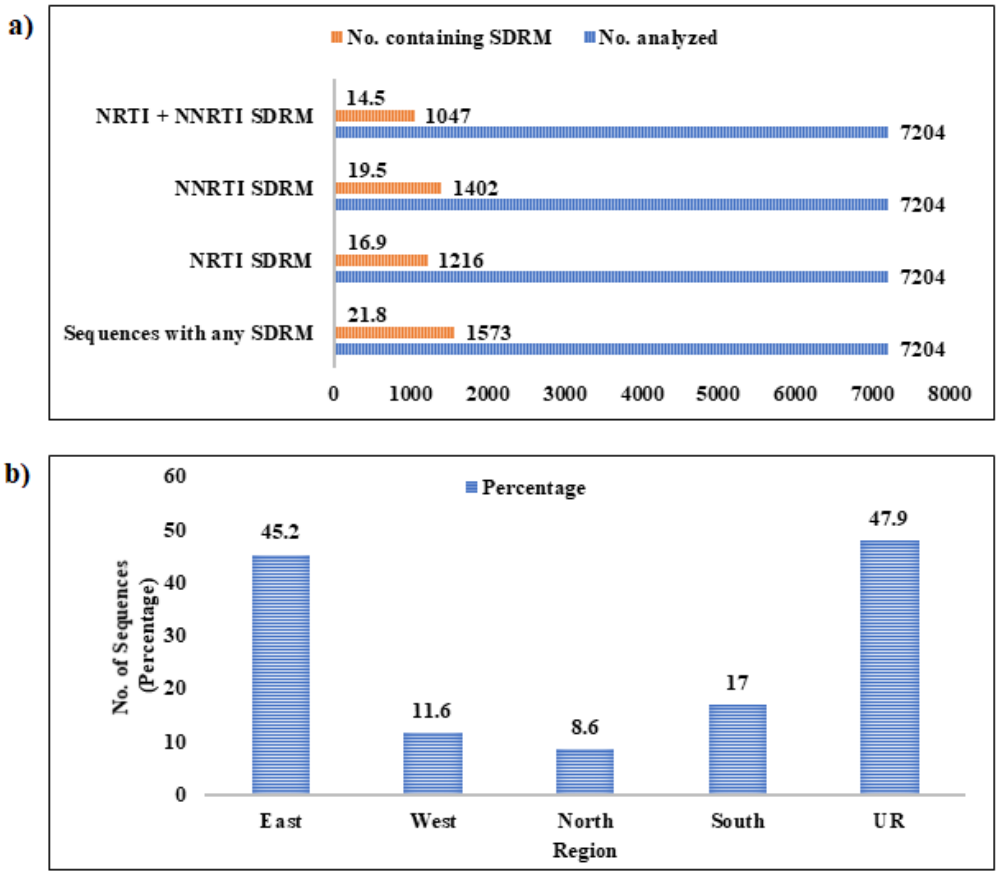


Fig.10: a): Represents the number of prevalent drug resistant mutations of RT observed in the population. **b)**Represents region-wise distribution of Drug Resistance Mutations with East (45.2%) and South (17%) having high prevalence.

3. Filtering and Trimming sequences

The subtype C sequences selected for the study were trimmed to obtain the desired nucleotide length. Sequences having length < 50 nucleotides were discarded. Twenty-one sequences were filtered for not satisfying the length criteria from Subtype C sequences (**Fig 9**).

4. Distribution of DRM (NRTI and NNRTI)

Seven thousand two hundred four sequences obtained after filtering were analyzed for drug resistance and the results showed 21.8% (1573) sequences containing DRM. 19.5% of analyzed sequences reported NNRTI DRMs (**Fig 10a**) thus being highly prevalent in India, in comparison to NRTI DRM and a combination of both (NRTI+NNRTI).

The prevalence of HIV-1 infection in India is dominated majorly by states in the North Eastern region and Southern regions dominate in new infections reported. Risk factors such as Low economic status, increased number of migrant workers and a large population of Intravenous Drug Users (IDU) form the basis of the emergent infections in these regions .On analyzing the distribution region-wise, the DRM data replicated the HIV-1 prevalence with East contributing majorly with 45.2% of the sequences showing at least one NNRTI/NNRTI resistant mutation. The DRM prevalence in the South was next to East with 17.0% sequences showing at least one NRTI/NNRTI followed by West region with 11.6% (**Fig 10b**).

The distribution of NRTI and NNRTI resistance is not regular in all the geographical areas. East and North dominated in harbouring mutations resistant to NRTI with East showing 85.7% sequences with NRTI and 64.3% with NNRTI; similarly, North showed 72.8% sequences with NNRTI resistant mutations and 54.5% NRTI mutations. West harboured mutations majorly resistant to NNRTI (84%) while South despite showing slight dominance in NNRTI resistant mutations, the prevalence was seen to be almost similar with 85.3% and 91.2% sequences with NRTI and NNRTI DRMs respectively. Sequences grouped as Unknown Region also showed increased NNRTI mutations (92.9%) than NRTI mutations (88.7%) but with little variation (**Fig.11**).

South region with a prevalence of 76.5% sequences was the region with the highest resistance for NRTI+NNRTI mutations. Following South, 50% sequences from East showed resistance

to both NRTI and NNRTI mutations while North (27.3%) seems to harbour the lowest. This unequal distribution triggers a need to further understand the kind of mutations, prevalent and their effect on drugs. This trend was reported in various articles published after 2015, sighting similar prevalence rates.

A study reported by Sharma et al., 2016 Manipur University, Manipur, on a sample size of 110 reported 37% and 29% individuals that harboured NNRTI and NRTI resistant mutations respectively. Manipur's geographical location at the Eastern border of India with International neighbours makes it vulnerable to various practices that result in increased HIV-1 diversity and possibility for the emergence of resistance. The study also reported the circulation of recombinants in the area at 11.8%. Among the risk groups, the study observed increased DRMs in DUs, followed by heterosexual individuals. A study reported by Dutta et al., 2017 on a population of 44 samples from all over East India, showed a 100% prevalence of any type of DRM (NRTI/NNRTI/PI) with 77% and 34.09% is the percentage of individuals with NRTI and NNRTI respectively. A study reported from South in Sivamalar et al., 2016 on 400 HIV-1 positive individuals failing first-line therapy showed a 94.7% prevalence of any DRMs. Another study reported in the same year Thirunavukarasu et al., 2016 on 23 isolates showed a prevalence of 87% of any DRMs and 91% showed NNRTI based resistant mutations and 87% showed NRTI based mutations. Karade et al., 2017 reported a study on 75 sequences derived from patients with Western Indian origin that showed a prevalence of 56.67% resistance. The patients showing NNRTI mutations were greater than patients with NRTI mutations.

These reiterated the results obtained by this study and showed an increase in resistance prevalence emphasizing the need for regular monitoring of resistance emergence in the patient population.

5. **Highly prevalent mutations**

Analyzing the results obtained from the CPR tool of Stanford Drug Resistance Database, it was observed that TAM and M184V/I mutations are most commonly observed in NNRTI. K70E and L74V/I was observed in some samples. TAM was observed at 21.4% frequency and M184V/I mutation was observed at 12.5% in the Eastern part of the country. However, in the West, both TAM and M184V/I were observed at ~5% frequency. TAM was seen at

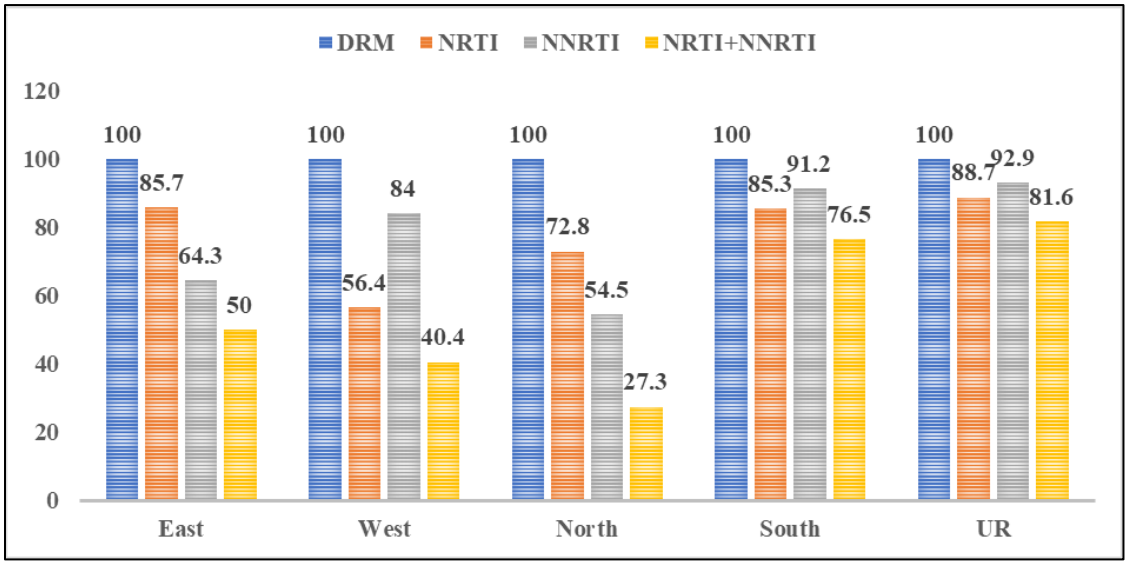


Fig. 11: The graph represents the NRTI, NNRTI and NRTI+NNRTI based mutations and their region-wise prevalence. East (85.7%) and North (72.8%) showed high prevalence of NRTI mutations while South (91.2%) and West (84%) showed high prevalence of NNRTI mutations.

NRTI/NNRTI Mutations Observed(region-wise)						
Drug Class	Mutation	East	West	North	South	UR
NRTI	TAM	12(21.4)	271(5.8)	8(6.2)	85(13.9)	789(41.4)
	M184VI	7(12.5)	241(5.2)	4(3.1)	75(12.3)	736(38.6)
	K65R	0	26(0.6)	0	3(0.5)	31(1.6)
	K70E	1(1.7)	84(1.8)	0	16(2.6)	215(11.3)
	L74VI	2(3.5)	30(0.6)	0	12(2.0)	82(4.3)
	Y115F	0	0	0	1(0.2)	18(0.9)
	T69ins	0	26(0.5)	0	3(0.5)	80(4.2)
	Q151	0	18(0.4)	0	2(0.3)	29(1.5)
Drug Class	Mutation	East	West	North	South	UR
NNRTI	L100I	0	1(0.02)	0	1(1.7)	15(0.8)
	K101E/P	0	49(1.0)	0	14(2.3)	172(9.0)
	K103N/S	4(7.1)	174(3.7)	3(2.3)	31(5.1)	351(18.4)
	V106A/M	0	35(0.7)	2(1.5)	17(2.8)	121(6.3)
	E138A/G/K/Q	0	0	0	0	0
	Y181C/I/V	4(7.1)	176(3.8)	1(0.8)	38(6.2)	323(17.0)
	Y188L/C/H	0	52(1.1)	0	7(1.1)	80(4.2)
	G190A/S/E	2(3.6)	115(2.5)	2(1.5)	26(4.2)	285(15.0)
	M230L	0	5	0	5(0.8)	18(1.0)

Tab. 2: Table depicts the prevalence of common NRTI and NNRTI mutations observed in the sequences of the study represented region-wise.

6.2% and 13.9% in North and South respectively, while M184V/I was at 3.1% and 12.3 % respectively. Mutation K65R was seen in West and South at ~ 0.5% but was not seen in other locations. This difference in the appearance of K65R could also be due to lack of sufficient data from East and North regions. A similar trend was seen in mutations T69ins and Q151.

The prevalence among NRTI showed a common pattern, with mutations K103N/S being observed in all geographical locations. 7.1% sequences from East harboured this mutation while 3.7% in the West, 2.3% in North and 5.1% in South showed K103N/S. Another mutation that showed high prevalence is Y181C/I/V with 7.1% sequences in East reflecting this mutation. 3.8% and 6.2% sequences from West and South respectively contained this mutation, but it did not reflect in the sequences from North where the percentage was as low as 0.8%. The mutations V106A/M and G190A/S/E dominated the population in North with 1.5% each **(Tab. 2)**. This is an interesting finding that needs a further follow up as Y181C is a commonly observed NNRTI mutation across populations exhibiting first-line treatment failure.

6. Status in Drug Naïve and Drug Experienced patients

To understand the level of resistance among patients diagnosed with HIV or patients initiating ARV treatment, the sequences in the study were divided into Drug Experience (n=6664) and Drug Naïve (n=540) based on the information provided in the scientific reports corresponding to the sequences. The percentage of patients harbouring primary resistance (Drug Naïve) were low in comparison to the Drug Experienced patients. Only 3.7% of the sequences obtained from DNs harboured resistant mutations, and the prevalence showed slight differences in the pattern when compared to DE. NRTI based mutation, M184V/I showed 68.0% prevalence, followed by TAMs with a prevalence of 51.2% in Drug Experienced patients. However, TAM showed the highest prevalence in Drug Naïve with 44.4% prevalence, followed by M184V/I with a prevalence of 16.7%. This pattern was previously reported in other studies and is suspected to be due to reverting of the virus to wild type in the absence of selective pressure of the ART (Sinha et al., 2018).

Among NNRTI mutations, K103N/S was observed in drug naïve patients at a prevalence of 22.2%, followed by V106A/M with a prevalence of 16.7%. K101E/P and Y181C/I/V were the other mutations seen at a prevalence of 5.6%. In Drug Experienced patients, K103N/S

and Y181C/I/V mutations were most commonly observed with 35.9 and 34.8% prevalence respectively. G190A/S/E was also observed in Drug Experienced with a prevalence rate of 28.1% **(Tab. 3)**.

A study reported at YRG CARE, Chennai by Saravanan et al., 2017 among children failing first-line NNRTI based ART treatment in South India showed that the highest prevalent NRTI mutation among the children was M184V/I with 94% of the resistant samples harbouring this mutation. T215F/Y, a mutation not observed in our study was reported to be the next highest with a frequency of 41%. TAMs were observed to be present in frequencies lower than T215F/Y followed by other mutations, Q151M, K65R, T69i were observed at frequencies similar to our study. Among the children harbouring at least one NNRTI mutation, Y181C/IV was seen to be the most common with a prevalence of 45% followed by K103N/S with frequency 37%, these results correlated with the overall mutation prevalence observations drawn by our study. G190A/S, an NNRTI mutation, was seen at a frequency of 32.5% followed by K101E/H/P and V106M. Despite this study being restricted to South Indian Children, the results correlated with the current study.

A study reported by Kannangai et al., 2015 conducted at Christian Medical College, Vellore in treatment naïve patients, showed the prevalence of K101E, Y181C and G190A NNRTI mutations . Similarly, a study conducted in North India by Sinha et al., 2012 at AIMS, Delhi reported a low prevalence of primary resistance at 2.9% with a prevalence of M184V. Another study reported by Deshpande et al., 2011 in Mumbai observed K103R and V106M among NNRTI mutations. Almost all reports suggested the prevalence of RT DRMs to be lower than 5% consistent with our study and according to the WHO guidelines.

NRTI/NNRTI Mutations Observed(DN&DE)			
Drug Class	Mutation	Drug Experienced	Drug Naive
NRTI	TAM	798(51.2)	8(44.4)
	M184VI	1060(68.0)	3(16.7)
	K65R	60(3.9)	0
	K70E	316(20.3)	0
	L74VI	126(8.1)	0
	Y115F	19(1.2)	0
	T69ins	109(7.0)	0
	Q151	49(3.1)	0
Drug Class	Mutation	Drug Experienced	Drug Naive
NNRTI	L100I	17(1.1)	0
	K101E/P	234(15.0)	1(5.6)
	K103N/S	559(35.9)	4(22.2)
	V106A/M	172(11.0)	3(16.7)
	E138A/G/K/Q	0	0
	Y181C/I/V	542(34.8)	1(5.6)
	Y188L/C/H	139(9.0)	0
	G190A/S/E	438(28.1)	0
	M230L	28(1.8)	0

Tab. 3: Observation table of NRTI and NNRTI mutation frequency in drug experienced and drug naïve sequences.

7. Implication of mutations

The common first-line treatment available in India as per NACO 2018 guidelines is Tenofovir (TDF) + Lamivudine(3TC) + Efavirenz (EFV); however, the alternate treatments include Zidovudine (AZT) + 3TC+ EFV or Abacavir (ABC) + 3TC+ EFV or ABC+EFV+ Nevirapine (NVP). The failure of the first line treatment necessitates the change to second-line treatment. The failure is diagnosed as Immunological or Virological failure and is usually a consequence of the development of resistance to the first line drugs. The current second-line treatment options suggested by NACO are Ritonavir-boosted Protease Inhibitor + AZT or TDF + 3TC. In cases where required, an Integrase Inhibitor can be added.

Some NRTI/NNRTI mutations developed under drug pressure confer cross-resistance to other drugs of the same family or its neighbour. For example, TAMs cause cross-resistance to some degree in all drugs of NRTI class, similarly M184V causes resistance to Abacavir

Highest resistance observed in Drug Naïve patients		
Drug Class	Mutation	Drug
NRTI	TAM	D4T, ZDV
	M184VI	3TC, FTC
NNRTI	K103N	NVP, EFV
	Y181C	NVP, ETR, RPV,EFV
	K101E/P	NVP,EFV,RPV
	V106A/M	NVP.EFV

Tab. 4: Table detects the drugs to which resistance mutations were prevalent in Drug Naïve patients and hence might require further surveillance and accordingly, prior testing before prescribing the drugs.

and ddI (IAS-USA Drug Resistance Mutations Group, 2007). As reported by this study, the most prevalent mutations observed in HIV-1 patients in the Indian scenario are TAMs and M184VI in NRTI class of drugs and G190A/S, K103N, Y181C/I/V in the NNRTI class of drugs (**Tab. 4**).

TAMs are Thymidine-analogs associated mutations seen in patients under treatment with AZT or Stavudine (d4T). The widespread use of thymidine analogues since the ART roll-out lead to the gradual accumulation of two classes of TAM mutations (TAM1 and TAM2). This is reflected in its high prevalence observed in all parts of the country. These confer 50-500 fold decreased susceptibility to AZT and cross-resistance to other NRTI drugs. Drug toxicities of AZT reduced its usage in the clinic as guided by NACO and hence is not commonly prescribed. However, previous reports and the current study recommend prior testing of patients with treatment failure before prescribing AZT or other analogues for the second line or as an alternative to other drugs in the first line treatment.

M184V causes high resistance to emtricitabine (FTC) and 3TC but has an advantage of decreasing viral fitness and increasing susceptibility to second class NRTIs like TDF. The emergence of K65R in West and South needs to be further studied and surveilled as it confers resistance to NRTI drugs like ABC, ddI, d4T etc. apart from resistance to TDF. Constant monitoring for the prevalence of this mutation in the population will help in understanding and managing the drugs in the second line treatment. Other NRTI mutations observed in the population include Q151M and T69i that are both well known for multi-resistance to all known NRTI drugs (IAS-USA Drug Resistance Mutations Group, 2007). The appearance of these mutations in drug naïve patients, especially, TAMs requires testing and monitoring of patients before initiating first-line treatment when Thymidine analogues are considered as an option.

Among NNRTI mutation, K103N/S is highly observed throughout all geographical regions, and these mutations cause resistance to Nevirapine (NVP) and Efavirenz (EFV) with K103N reducing their efficacy by 50 and 20-fold respectively. As previously discussed, Efavirenz is currently the most commonly used drug in the first line of treatment, and this could have lead to an increase in resistance. K103N/S mutation is generally seen in patients administered with a single dose of NVP, which is given as a prophylactic measure to pregnant women to inhibit mother to child transmission. This results in an increased rate of K103N/S mutation in drug

naïve children. This high prevalence of the mutation in drug naïve patients necessitates testing before initiating treatment in naïve patients.

Y181C/I/V is also seen with a high prevalence and confers resistance to NVP, Etravirine (ETR), Rilpivirine (RPV). The mutation Y181C confers reduced susceptibility of >50 fold to NVP, 5-fold reduction to ETR, three-fold to RPV and two-fold reduction to EFV. Despite the reduction in response to EFV by only 2-fold, it has been reported to suppress efficacy of EFV containing regimens. The mutation Y181I/V causes ten fold and 15 fold reduced susceptibility to ETR and RPV, respectively.

Another commonly observed mutation in the Drug Experienced population was G190A/S. This was seen to be selected by NVP and EFV, and G190A causes >50-fold reduced susceptibility to NVP and 5-10-fold reduction in the efficacy of EFV. G190S causes > 50-fold reduced susceptibility to both NVP and EFV. However, this mutation was not observed in naïve patients in our study.

K101E was observed in both drug experienced and naïve patients with considerable prevalence. It usually occurs in combination with other NNRTI mutations and causes decreased susceptibility to NVP by 3-10 fold; EFV by 1-5 fold. Similarly, K101P reduces the susceptibility of NVP, EFV and RPV by >50-fold.

V106A, a mutation commonly observed in subtype C virus in patients experienced with NVP. It causes 50-fold reduced susceptibility to NVP and a five-fold reduction in EFV. V106M is primarily selected by EFV and NVP and cause >30 fold reduced susceptibility to both the drugs. It was also shown to cause cross-resistance to NNRTI drugs as reported by a study conducted by (Brenner et al., 2003). This was also observed in drug naïve patients making it important to consider pre-testing for this mutation before initiating patients with EFV based treatments.

Summary

The necessity to understand the distribution of Drug Resistant Mutations in India is much more than it is perceived today. Standing as the third highest country with HIV-1 infection in the world, proper control of infection should be a goal to achieve. This requires continuous monitoring for changes in the population. With a virus as notorious as HIV-1 it is of utmost importance, due to less treatment options and lack of awareness. Studies conducted throughout, to understand DRMs reported results which without a complete picture, may stand baseless in understanding the trend of developing resistance. A study to understand the sequences in the population may turn out to clarify the trend to prescribe effective treatment. A study to explore the Drug Resistance to HIV-1 ART drugs in the context of Indian Subcontinent was performed.

The sequences obtained from Genbank for this objective belonged to diverse studies conducted in various research and ART centres across the country. A study conducted by Kurle et al., 2007. from NARI, Pune studied the emergence of NNRTI drug resistance as a consequence of single dosage administration of Nevirapine to prevent Mother- Child HIV transmission. Another study conducted by Kandathil et al., 2009 from Christian Medical College Vellore, analyzed sequences obtained from patient samples to compare three different genotypic resistance algorithms. Lall et al., 2008 from Armed Forces Medical College, Pune conducted a study to understand primary resistance in drug naïve patients. Rajesh et al., 2009 from Tuberculosis Research Centre (ICMR), Chennai studied the prevalence of resistant mutations at base-line and at the time of therapy failure in patients with HIV-1 coinfected with TB. A study conducted at Johns Hopkins Bloomberg School of Public Health, Baltimore collected samples from India to study NVP resistance, breast-milk transmission of HIV-1 in single and extended dosage while under NVP prophylaxis. A study in NARI, Pune by Thorat et al., 2011 was conducted to analyze drug resistance in HIV-1 positive woman from Kakinada, Andhra Pradesh. Some studies conducted in Bangalore, Delhi, Chennai explored the primary and secondary resistance in patients from different parts of the country. This objective was framed to understand the DRM circulating in India from these isolated studies to obtain a wholesome view.

Most prevalent Subtype in the population was identified as Subtype C. Large number of sequence submissions were seen in the West and South regions of India. Region wise study reported an increased circulation of resistant mutations in East and South regions. NRTI based

mutations were commonly observed in East and North .NNRTI based mutations were commonly observed in South and West. Most commonly observed mutations were TAMs and M184V in NRTI . K103N was Y181C/I/V commonly observed in NNRTI.

Some major polymorphisms were observed in the sequences, in well conserved regions across Subtype B and C. Some commonly observed are at positions I135, I178, I5, W229, K32, H221. The highest being I135, I178 and I5. W229, K32 and H221 though present in small number are reported first time in India. It has been observed in this study that the percentage of Drug-Resistant Mutations to Reverse Transcriptase was higher in Acquired Drug Resistance (Drug experienced) than Transmitted Resistance (Drug naïve). The percentages were observed at 27% and 3.7 % respectively. The percentage in Drug naïve is acceptable under WHO guidelines (<5%); however, from previous studies, the trend shows that resistance could spread at a higher rate. Added to this, the unavailability of sufficient data and a large number of unreported cases further increases the chance of the percentage to be higher than observed.

The TAM and M184V mutations make the virus resistant to ZDV, D4T and 3TC, FTC. It is most prevalent across the country. These drugs are most commonly prescribed as the first line regimen and are the ones most commonly acquire resistant to. This is the phenomenon observed worldwide. The resistance to NNRTI is most common to drugs NVP, EFV, ETR and RPV irrespective of region. NVP and EFV are most commonly prescribed drugs, NVP is prescribed as a single dose regimen for pregnant ladies to prevent the transfer of infection to the unborn child and also to the new-born as protection from transfer to breast milk.

This study has projected that resistance to drugs most commonly prescribed as the first line regimen in India might be on the rise. Thus, there might be a need for continuous monitoring. Further, algorithms and databases specific to this sub-continent need to be developed. Major sequencing projects should be undertaken periodically across the country to understand various trends of the infection and for effective treatment. A database and algorithm in line with Stanford Drug Resistance Database should be developed to assist health care providers and researchers to access the information on DRM specific to this sub-continent.

CHAPTER IV

OBJECTIVE 2: Design, Development and Validation of HIV-1 Entry Inhibitors.

Introduction

The first step in HIV entry is the adsorption of virus on the cell surface of the host and the further steps culminate into their final fusion in the subsequent release of the HIV-1 core into the host cell. The surface of HIV-1 particles possesses spikes of a 160 kDa glycoprotein envelope(Env). This Env facilitates fusion of the viral and host cell membranes, enabling the virus entry. gp160, coded by the Env gene, is the envelope precursor, which releases fragments, gp120 and gp41 upon proteolytic cleavage. The mature viral spike consists of Env glycoprotein trimer made of copies of heterodimers of gp120 and gp41. The former establishes contact with cellular surface receptors and latter facilitates the process of fusion between viral and target membranes. Each wild-type HIV-1 particle possess 7-14 spikes on the surface (Kijewski et al., 2015).

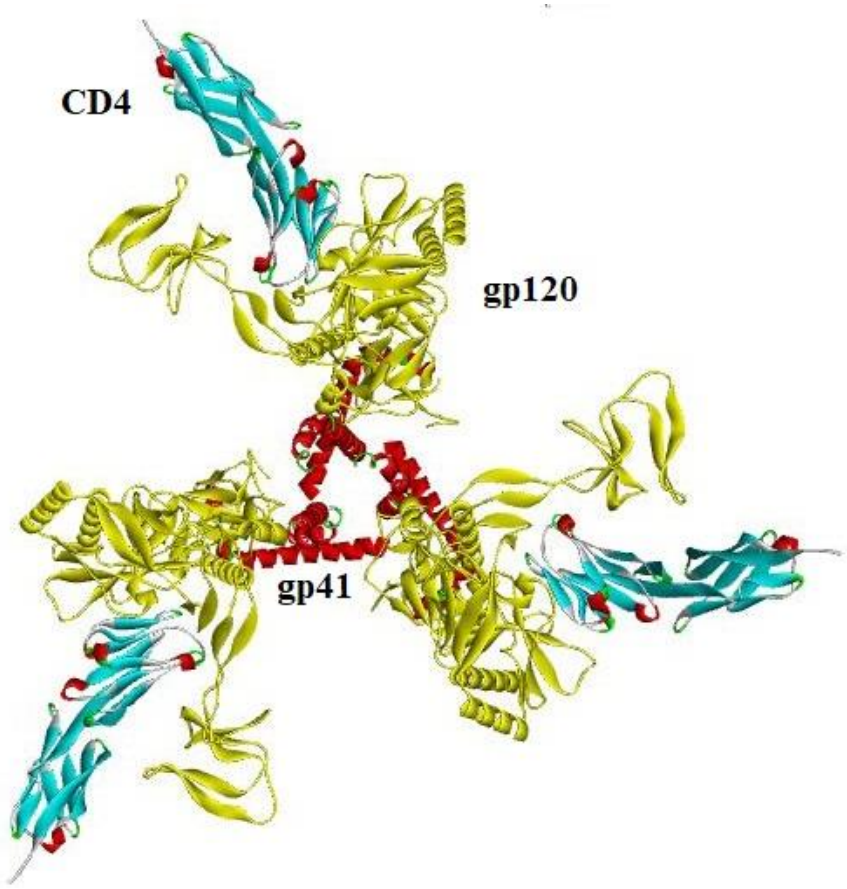


Fig. 12: HIV-1 Env Trimer. Source: Crustal structure 3J70 from RCSB PDB

HIV Entry

The fusion is either directly mediated by gp120-CD4 receptor interaction or initially via cell attachment factors such as DC-SIGN, negatively charged heparan sulfate proteoglycans, α4β7 integrin, Langerin, CD169 etc. These attachment factors involve in the initial interaction with the Env, mediating the transfer of virion particles into proximity of the host cell receptors (Wilen et al., 2012; Kijewski et al., 2015). HIV-1 Entry, mediated by the envelope gp120 is guided by structural rearrangements to achieve fusion of the virus and host, and hence gp120 is called the Type 1 fusion machine (Ward et al., 2015) .

Once in proximity, the gp120 subunit of Env trimer interacts with the host cell receptor CD4 and this is essentially the first and the most important in the entry process. Structural changes especially in the variable loops 1,2 and 3 are the consequence of the interaction of gp120 and the CD4 receptor. CD4 binding to the Env subunit also initiates the the bridging sheet (Wilen et al., 2012).

The next crucial step in the process of HIV entry is the engagement of coreceptor by gp120. The conformational changes in the V3 loop caused by the CD4 binding and the formation of the bridging sheet are critical in this engagement. The coreceptor binding causes a cascade of structural changes leading to the viral particles approaching the host membrane for productive fusion. Membrane fusion is mediated by the gp41 subunit of the Env, which is exposed as a consequence of coreceptor binding. The exposed gp41 hydrophobic fusion peptide bores into the host cell membrane and connects the host and viral membranes, resulting in the folding of the fusion peptide at a hinge region. This brings the amino-terminal region (HR-N) and a carboxy-terminal region (HR-C) together forming a six-helix bundle. This drives the movement of the opposing membranes into close contact, leading to the formation of a membrane pore known as fusion pore which serves as an entry for the transfer of viral core into the host (Wilen et al., 2012).

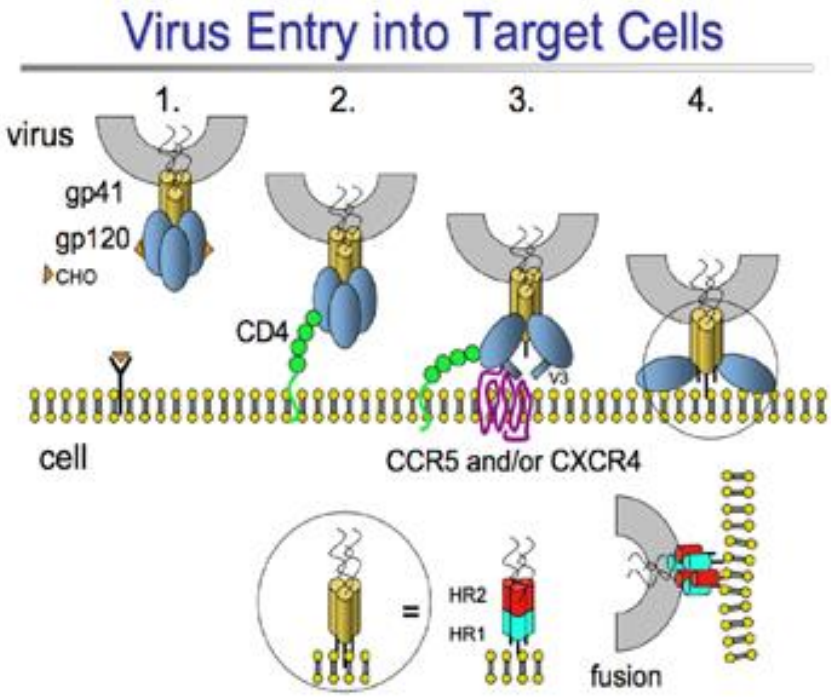


Fig. 13: HIV-1 Entry mechanism

Source:https://www.scripps.edu/mosier/images/donsresearchimage.png

gp120

The gp120 subunit of Env trimer is a highly glycosylated protein that has five conserved (C1-C5) regions interspersed with five variable regions (V1-V5), known for their genetic variability. The core of gp120 is composed of 5 α-helices, 25 β-strands, and ten defined loop segments. The gp120 polypeptide contains two major domain and certain extensions that protrude from this body (Kwong et al., 1998).

The glycans, decorated on the surface of the structure, once considered a shield of protection from the host immune system, have now been identified as a major factor of recognition for broadly neutralizing antibodies (Ward et al., 2015). The glycans are alternated by the variable loops that are surface exposed, thus protecting the conserved domains present in the core. High mutation rates, increased flexibility and high glycosylation, are characteristic traits of the variable loops. The flexibility of the variable loops makes it difficult to obtain their structure by X-Ray crystallography, and hence, most of the structures available do not include the variable regions. The variable region V1 projects towards the patch present on the gp120 outer domain, dominated by high mannose. Both V1 and V2 protrude away

from the apex of the trimer with V2 folded over the CD4 binding site. V3 loop is positioned to stabilize the trimeric structure with its tip making contact with the V2 base (Ward et al., 2015).

The CD4 binding site of gp120 is well conserved and comparatively free of glycans, but its location at the junction of promoters restricts its access. The CD4 receptor thus approaches at a specific angle to engage in interaction with gp120, while avoiding the V1, V2, V3 loops that form an apex over the binding site (Ward et al., 2015).

CD4 occupies a depression formed by the outer and inner domains with the bridging sheet. The interacting residues on CD4 span the amino acids 25-64, while the region in gp120 is spread over different segments of the protein (Kwong et al., 2017). Unlike the CD4 binding site that is partially exposed on the surface of the closed gp120 structure, the chemokine co-receptor binding region is well buried and only exposed upon CD4 binding (Bonsignori et al., 2017) and is well conserved (Ward et al., 2015). The binding of CD4 results in the V2 loop moment that overshadows V3 loop and CD4i (CD4 induced) epitopes. The binding not only exposes V3 loop but also enhances its susceptibility to proteolysis along with stabilizing the bridging sheet. Electrostatic interactions between the basic residues of the bridging sheet and the acidic surface of the co-receptor trigger a cascade of further steps for viral entry (Ward et al., 2015).

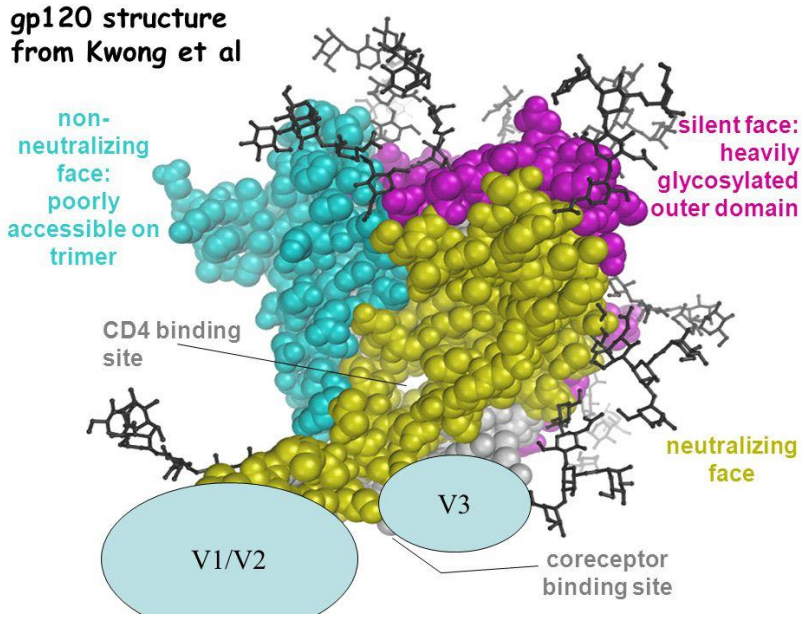


Fig. 14: gp120 structure represented with its various surfaces**.**
Source: https://player.slideplayer.com/11/3038957/data/images/img43.jpg

gp120-gp41 of the Env trimer, highly exposed on the surface of the virus particle is a potent vaccine target as it elicits immune response. Some regions of gp120: CD4 binding site, V1/V2 glycan region, the V3 loop region have been reported to be major targets for the broadly neutralizing antibodies (bNabs). Two classes of CD4 binding site antibodies have been defined based on the approach and antigen recognition. Some examples of the first class, CD4 mimicking antibodies are CH31, VRC01, CH235 and 8ANC131. The second class are known as variable heavy third complementarity determining region (CDR H3)-binder-class of CD4bs bnAbs. The examples include HJ16, CH98 and CH103. The broadly neutralizing antibodies of V1V2-glycan region is a new generation bnAb discovered. The examples of these include CAP256-VRC26 lineages, PG9, CH01 and PGT145. These are usually anionic, long, protruding and often tyrosine-sulfated that escape the glycans and reach for the apex of the spike and the epitope in the vicinity. Another group of antibodies that target the coreceptor-binding site (CoRbs) are also known as CD4-induced (CD4i) antibodies, as the exposure of this site is induced by the structural changes that occur upon binding to the CD4 receptor. Antibodies targeting this region include m36,17b, 48d, X5, and E51 (Sun et al., 2016).

Coreceptor (CCR5/CXCR4)

The coreceptors CCR5 and CXCR4 belong to a large family of G-protein coupled seven transmembrane receptors. The coreceptors consist of seven transmembrane helices, three extracellular loops (ECLs), three intracellular loops, a surface exposed N terminus, variable in length and a cytoplasmic C-terminus tail (Chan et al., 2018; Shaik et al., 2019). This family of proteins are involved with chemokine signalling and serve as the main coreceptors for HIV-1 entry. Based on the coreceptor usage, the HIV strains are classified as R5 tropic, utilize CCR5 as coreceptor or X4, utilize CXCR4 as coreceptor and virus strains that use both receptors are called dual-tropic. R5 tropic viruses are in abundance in the asymptomatic initial stages of the infection, while the X4 tropic or dual-tropic emerge along with the symptoms and loss of immunity (Dogo-isonagie et al., 2012).

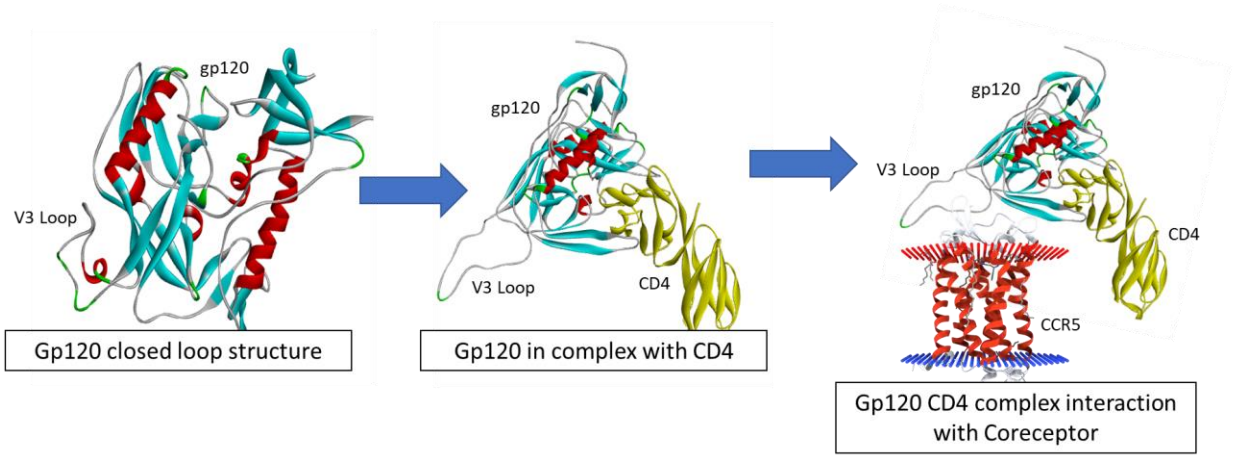


Fig. 14: Structural changes in the V3 loop of gp120 resulting from receptor and co-receptor binding.

A multitude of studies have been reported that aim at studying the interaction between the coreceptor and gp120-CD4 complex. The unavailability of crystal structures of the co-receptors until 2013 restricted the understanding of interactions. The studies conducted on the gp120 binding with small peptides or sulfonated antibodies that mimic CCR5 or other G-Protein coupled receptor family proteins along with some mutational studies identified the important regions for interaction, including the second extracellular loop, ECL2 and residues 2-15 of N-terminus (Dogo-isonagie et al., 2012). The angle at which the N-Terminus of the coreceptor approaches is critical in productively establishing contact with the gp-120. This angle of approach resembles the approach of CD4, resulting in the binding of the N-terminus of CCR5 at the bridging sheet-V3 loop intersection (Huang et al., 2005). This interaction requires a post-translational modification on some of the tyrosine residues, which involves its O-sulfonation, thus converting the tyrosine to sulfotyrosine (Farzan et al., 1999). This occurs at tyrosines located at positions 3, 10, 14 and 15 but the sulfotyrosines specifically at positions 10 and 14 facilitate engagement of gp120. Similarly, the studies elucidated the importance of N-Terminus, ECL2 and ECL3 in the interaction with CXCR4. The N-terminus of CXCR4 has not been shown to require O-sulfonation for its engagement with gp120.

V3 Loop

The V3 loop is a 35- residues long (Yokoyama et al., 2012) segment with the sequence 296 TRPNQNTRKSIHIGPGRAFYTTGEIIGDIRQAHC 331. The loop is held stable by a disulphide bond formed between the first and the thirty-fifth residue. The loop contains both

acidic and basic residues but is dominated by a positive charge. Many V3 loop structures have been resolved by X-ray and NMR studies in gp120 bound with CD4 and with many human monoclonal antibodies (mAbs) (Jiang et al., 2010). The structure can be divided into base, stem and tip. Located at the core of gp120, the base of the V3 loop is the most conserved region stabilized by disulphide bridges, the highly flexible stem protrudes out from the gp120 core. The crown of the loop consists of ~13 residues located in the centre of loop to form a hairpin β conformation (Jiang et al., 2010). The middle of the crown contains both hydrophobic and hydrophilic residues. But the net charge of the region is positive and can be in the range of +2 to +10, based on the phenotype of HIV-1. Generally, positive charges of the loop of CCR5 tropic HIV-1s are less than those of CXCR4 tropic HIV-1 (Yokoyama et al., 2012).

The crown can further be divided into the band, circlet and arch. The arch located in the centre of the loop is composed of a conserved motif, GPGR or GPGQ. The circlet located in the middle of the β-hairpin has two faces, the hydrophobic face, in its core contains Ile307 and Ile309 from the N-terminal and Phe317 from the C-terminal strands. The other face is known as the hydrophilic face. The band is characterized by a distinct positive charge due to the residues at the N-terminus. A conserved Tyr318 is present at the C-terminus of the crown (Chan et al., 2018).

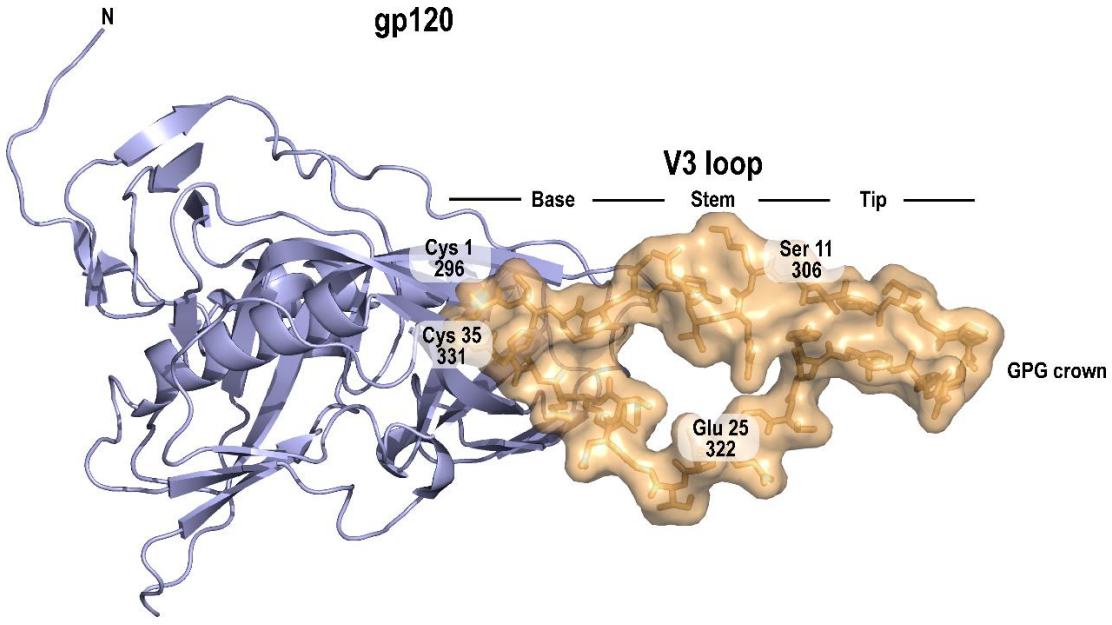


Fig. 15: Structure of V3 loop with depiction of crown , base, step and tip regions.
Source: doi: https://doi.org/10.1371/journal.pcbi.0030058.g001

V3 is known as the "principal neutralizing determinant" as it is highly immunogenic and antibodies against V3 are commonly found in the sera of all individuals infected with HIV-1. A large set of anti-V3 mAbs has been identified, explored and reported. Anti-V3 mAbs have been reported to neutralize up to ~50% of the viruses across various clades of virus (Jiang et al., 2010). Three types of V3 antibodies have been defined based on their interaction with gp120. The crown region elicits the highest immune response and the Abs raised against this region have been found in all infected individuals. This class of antibodies are called ladle-like glycan independent antibodies as the antibodies to this region are glycan-independent. The name "ladle-like" is given from the conformation of the V3-loop interaction with the Ab. The V3 tip of gp120 binds to the ladle like anti-V3 antibody and sits in the bowl of the ladle and the N-terminus strand contacts the handle of the ladle. 447-52D is an example of this type of antibody (Mayr et al., 2015).

Cradle type antibodies, exemplified by mAb 2557, are the second type of antibodies that are glycan independent and use an antigen-binding mode. In this class, the binding site on the co-receptor contains a groove holding the epitope, like a baby in a cradle. The binding site of V3 loop in this class of antibodies is the conserved hydrophobic core of the V3 crown, with residues 307, 309 and 317 making an interaction. The third class of antibodies are glycan-dependent and target the V3 base, which is usually poorly immunogenic. However, antibodies like PGT121 and PGT128 are highly potent and reactive. These require a glycan at N332 for interaction and hence are glycan-dependent (Mayr et al., 2015).

gp120-coreceptor interaction:

The X-Ray structure published by TanQ et al. in 2013, enabled a better understanding of the coreceptor engagement by gp120. A recent X-Ray crystallography study published in 2019 by Shaik. M. M et al (Shaik et al., 2019). explored the interactions between CCR5 and gp120. The first region to interact with CCR5 is the V3 loop. 309–316 residues of the V3 loop adopt a particular structure that includes a one-turn helix with a proline residue, Pro311, that penetrates the CRS2, and aligns against Trp86 and Tyr108 side chains of CCR5. Arg313 in V3 appears to be sandwiched between Tyr251 and Glu283 of CCR5. The CRS2 is made of two sub-pockets minor and major, of which minor sub-pocket is formed by the I-III and VII transmembrane helices and the major sub-pocket is formed by the III-VII helices. V3 loops occupy these sub-pockets and may also engage in water-mediated interactions with CCR5.

The ECL2 of CCR5 wraps around V3 loop and contacts stem and crown of the loop. The Glu172 residue of ECL2 and Arg304 of V3 appear to form a salt bridge. The ECL2 of CCR5 interacts with gp120 via three tyrosine residues Tyr10, Tyr14 and Tyr15 which can be sulphated and can intimately contact gp120. The aromatic ring of Tys10(sulfated Tyrosine) stacks with the Arg326 side chain of the V3 loop of gp120, assumed to be through a cation-π interaction; thus placing the putative sulfate group close to the gp120 Lys416 and Arg414 residues. Tys14 is seen to adjust between the V3 stem base and the base of the bridging sheet, positioning its sulfate group near the side chains of the Arg 298 and Lys 435 of gp120. The Tyr15 side chain lies close to residue Lys207 at the V1V2 stem base, where there is a possibility to form another cation-π interaction.

An interesting observation made by this study is the absence of any major structural changes in the gp120-CD4-CCR5 complex that might directly lead to the changes in gp41. However, gp120 dissociation was assumed to trigger an irreversible folding in gp41, finally leading to the fusion process. Structural rearrangements starting from the movement of the C-terminus of gp120 as a consequence of CD4 binding, lead to the formation of a pocket that is occupied by the fusion peptide stacking against the N-terminus. This is followed by structural dynamics disassociating the peptide from the pocket causing the loosening of gp41's contact with N-terminus allowing the terminus to bend back. These rearrangements in the terminus of gp120 prevent the filling of the pocket by the fusion peptide, thus weakening the bonding between gp120-gp41 leading to its disassociation. Despite, the negligible role of CCR5 in rearrangements of the termini and the gp41 release, the engagement of gp120 with the coreceptor brings the virion particles closer to the host cell membrane reducing the distance between the disassociated fusion peptide of gp41 and the membrane, facilitating fusion.

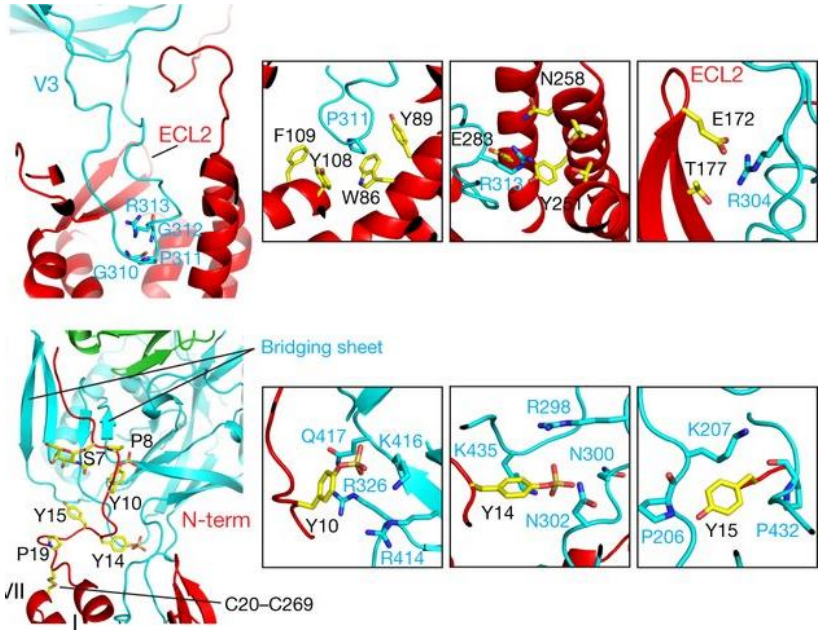


Fig. 16: Interactions of gp120 and coreceptor. Source: 10.1038/s41586-018-0804-9

447-52D as a potential lead:

As discussed above, numerous antibodies have been reported in infected populations, that neutralize the surface Env. However, the use of these antibodies in a therapeutic setting is limited by efficacy and delivery issues. To overcome these factors, compounds with potential anti-HIV-1 activity were developed by taking an antibody as a lead. With its high variability, HIV-1 has been evolving at a rate many times faster than its host and has been successful in establishing a fatal infection. HIV-1 uses its sequence variability as its shield even in evading the immune system and the surface exposed V3 loop of gp120 is known for its extensive sequence variation and its immunogenic property. But the crown region with residues GPGR/GPGQ is well conserved and acts as a hole in the shield. Thus, the antibodies elicited by this region are less type specific unlike other antibodies elicited by other regions of the V3 loop.

447-52D, identified and isolated by Zolla-Pazner and colleagues from an HIV-1 infected individual is an example of one such antibody. The antibody efficiently neutralized all possible laboratory strains and showed 50% neutralization of Clade B isolates. The neutralization was also observed to be independent of the tropism of the virus strain. Crystal Structure of 447-52D with V3 loop peptide was solved by Zolla-Pazner group, and it led to the understanding of the underlying interactions. The crystal structure reported three strands of mixed β- sheets, two anti-parallel sheets are formed by the heavy chain and CDR H3 of

the antibody and the third strand was formed by the V3 loop itself. The conformation of V3 loop in complex with 447-52D was similar to the conformations it takes with other antibodies, but the unique aspect of the complex was the mode of recognition by the loop through specific interactions. Antibodies usually recognize the epitopes via interaction with their sidechains and hence are sensitive to sequence mutations. However, the recognition of 447-52D is sequence dependent only in the crown region and is specific to GPxR, and other residues spanning the epitope are allowed to be substituted. Hence by binding to the conserved residues, the antibody attains specificity, and by binding to the main chain atoms of the neighbouring residues, it satisfies the requirement for surface area to achieve strong affinity (Kwong, 2004).

Four important interactions were defined in the complex i) π-cation stacking between the Arg315 of V3 loop and aromatic residues, TyrH100J and TrpH33 of the 447-52D heavy chain. ii) Salt bridge formed between anionic AspH95 of the antibody and cationic V3 loop Arg315 iii) van der Waals interaction in parallel and orthogonal orientation between Pro313 of V3 loop - TrpL91 and V3 loop -TrpL96 respectively iv) Interaction of the main chain atoms of N-terminal of the V3 crown with CDRH3 of the antibody. This involved, interactions mediated by six main chain atoms of the V3 crown N-terminus and C-terminus of CD4H3 (Killikelly et al., 2013).

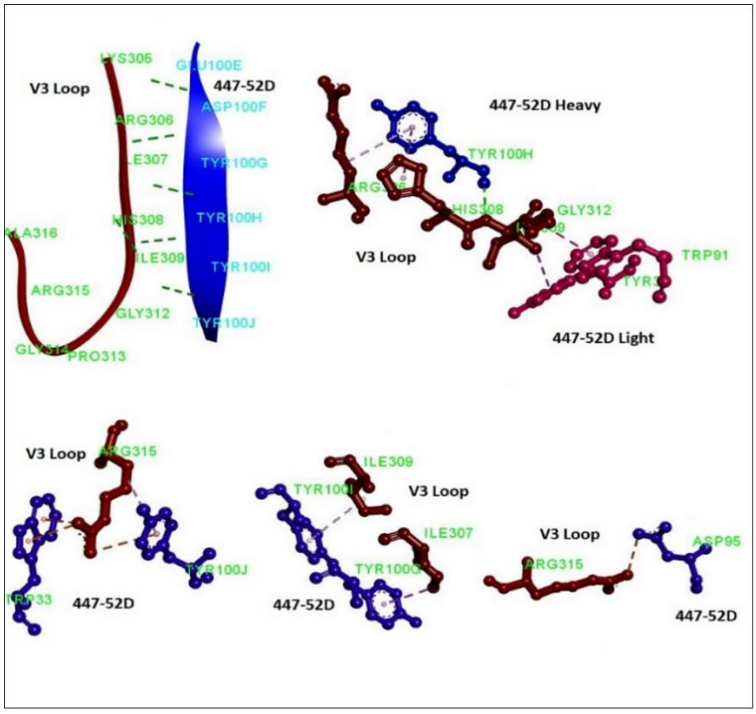


Fig. 17: The interaction interface between 447-52D and HIV-1 gp120 V3 loop.

Mutational analysis conducted to understand changes in binding affinity with respect to the four types of interactions observed, reported a) 56 and 150-fold decrease in affinity when TrpH33Ala and TyrH100J mutations that destroyed the stacking with Arg315, were introduced respectively b) 170-fold decrease in affinity when AspH95Asn was introduced breaking the salt-bridge c) 230 and 76-fold decrease in affinity when TrpL91Ala and TrpL96Ala were introduced. A correlation of 0.63 was observed between the decrease in binding affinity and reduced neutralization (Killikelly et al., 2013).

The complex crystal structure and the impact of interactions on affinity and neutralization were taken as a cue to develop novel compounds with retained properties and activities.

EPAP-1 as a potential Lead:

A glycoprotein of molecular weight of 90kDa was isolated from first-trimester pregnant woman who underwent medically terminated pregnancy (MTP). The status of the woman was HIV negative, and the protein was identified and isolated from the placental tissue of the woman. This protein was named Early pregnancy-associated protein-1 (EPAP-1). Sambucus nigra agarose lectin was used to isolate and purify the protein by affinity chromatography. Observations from previous studies on Epap-1 showed (K.Kondapi et al., 2002; Rani et al., 2006)

- The involvement of the envelope epitopes with Epap-1. gp160JR-FL was used for the study. It was observed that the gp160-Epap-1 complex was formed by strong interactions with V2 and V3 loop regions.

- The effect of CD4 on the binding of Epap-1 to gp160 was studied. It was observed that the binding of Epap-1 to CD4-gp160 complex blocks the exposure of V2, V4 and C1, C2, C5 regions. However, the interaction with the V3 loop remained unaltered, strong and specific.

- To examine the regions involved in interactions with Epap-1, the binding of the CD4-gp160-CXCR4 with Epap-1 was checked. The results showed that the interaction of Epap-1 at V3 region was unaltered even in the presence of CD4 and CXCR4.

- Epap-1 inhibited recognition of gp120-specific monoclonals 670-30D, V3-21 and F-425, C-5 (Rani et al., 2006). The studies show that the interaction is mainly in the V3 region, and the interaction of Epap-1 may result in conformational change of gp120, resulting in the exposure of other epitopes.

• The analysis of Epap-1 localization experiments show that CD4 binding is not completely abolished by Epap-1 binding. The results also suggest that the Epap-1 interaction with V3 loop epitopes is highly conserved. The interaction of Epap- 1 at C5 epitope in soluble gp120 is observed to be minimal but in CD4–gp120 binary complexes, the interaction was inhibited. (K.Kondapi et al., 2002; Rani et al., 2006).

Interestingly, Epap-1 inhibits HIV-1 entry by freezing, a complex comprising of Epap-1, gp120-41, CD4 and CXCR4. Thus, a transient conformation of the envelope in this posture will be a dynamic conformational element, which may probably be conserved in HIV-1 isolates. This idea is reinforced by our study that shows that the action of EPAP-1 is conserved among HIV-1M-N, HIV-191US056, HIV-1VB7, HIVVB-66 strains.

Our group used in vitro approaches to synthesize peptides derived from Epap-1 based on the interaction interface with gp120. Analysis of peptide regions identified two decamer peptides (P2: H2N-RSTCALTAATAKAYATRSLEHRVVYRILHDC-COOH) (P3: H2N-VYNLDCLFRSICLHPWWWWMGVYNLVRDFITLH-COOH) with significant binding to gp120 at V3 domain and surrounding regions involving (Glu267, Glu268, Thr290, Ser334, Arg335, Ala336, Asn340, Gly410). These interactions show the requirement of a surface volume of 900 to 1000 Å3 with 744 to 745 Å2 surface area with 17 donor atoms and 16 acceptor atoms (Bhaskar et al., 2013). These properties of the peptides provided specificity of interactions with the V3 loop regions.

Major issues associated with peptide therapeutics include delivery, stability and expensive production (Haggag et al., 2018). The development of small molecular peptidomimetics will help tackle these issues. The configuration of the peptide-gp120 complex provided indications on the molecular frameworks that can mimic the interaction provided by peptides. The objective envisages the development of anti-HIV-1 active small molecular organic frame works with specific interaction at the V3 loop domain of gp120 mimicking that of Epap-1 derived peptides.

This chapter is further divided into two parts

a) Design, Development and Evaluation of peptidomimetics derived from broadly neutralizing Antibody 447-52D
b) Design, Development and Evaluation of peptidomimetics derived from broadly active protein EPAP-1

Methods

1. Computational design of compounds from 447-52D

1.1 447-52D-V3 loop complex:

The structure with RCSB PDB id 1Q1J solved by Stanfield et al. with a resolution of 2.5 Å was the closest crystal structure record found, for the antibody-V3 loop complex. The complex consisted of 3 unique structures with a total of 924 residues. L and M are light chains of the antibody, 447-52D, H and I are heavy chains of the antibody, P and Q are the 16 residue, V3 loop. UniProt Id: P05877

The PDB file was processed to hold only the V3 peptide by removing the antibody and the water molecules.

1.2 Virtual Screening :

Pep:MMs: MIMIC server was used to screen the MMs database for molecules with desired physical properties of the 447-52D-V3 loop interaction interface. The V3 residues I309, I307, H308, R315 etc. were given as the interacting residues. The scoring was based on both pharmacophoric(60%) and shape(40%) parameters.

1.3 Screening of the reported molecules :

The molecules reported by Pep:MMs: MIMIC were further screened to obtain hits with good binding affinity to V3 loop. GOLD docking suite was used for screening of the reported molecules. The processed 1Q1J file was used as the receptor, and the docking was performed in batch. The binding site residue was specified as Arg315 and 5Å was set as the radius of the cavity. Genetic algorithm was selected as the default program for exploring the conformational space for ligand binding. The parameters for GA run were set to default as described; 100 was set as the population size, the selection pressure was set to 1.1, the number of operations was given as 10,000, and one island and niche size of 2 was set per run. Mutation and cross over rates were set at 100. The GOLD score was used as a metric to score the molecules. The output was set to 10 solutions per molecule.

2. Computational design of compounds from EPAP-1

2.1 Modelling peptides obtained from EPAP-1

Two peptides, (P2: H2NRSTCALTAATAKAYATRSLEHRVVYRILHDC-COOH) (P3: H2N-VYNLDCLFRSICLHPWWWWMGVYNLVRDFITLH-COOH) derived from EPAP-1 identified to be active by in-vitro studies, were used for the study. The peptides were modelled using PEPFOLD3.0 and the number of iterative simulations was set to 100. The generator ts5 was used as the peptide contained more than ten amino acids. sOPEP energy was used to sort the peptide clusters generated after completion of the run. Internal model quality assessment was performed using Apollo. The number of steps for the Monte Carlo simulation was set to 30,000 steps. The generated structures and clusters were analysed and validated using RAMPAGE. The structures with acceptable results as per RAMPAGE were selected, and the energy minimization protocol was followed with the Steepest Descent algorithm and 50,000 steps. The minimized structures were taken to perform docking and obtain interacting regions.

2.2 Receptor, gp120 structure

Huang. C et al. deposited the gp120 crystal structure, 2B4C derived from JR-FL and consisting of V3 loop. This structure was retrieved and processed to use as a receptor to understand the interactions of the obtained lead molecules. The resolution of the structure was 3.3 Ao and consisted of 4 unique units and a total of 975 residues. The structure consisted of env glycoprotein gp120 with V3 loop, CD4 receptor, Heavy and Light chains of X5 antibody. The complex also consisted of some naturally occurring ligands and some stabilizing agents. The water molecules, ligands, CD4 and antibody fragments were removed for processing of the gp120 structure. The final structure was used for further studies.

2.3 Docking of modelled peptides to gp120.

HEX8.0.0 software suite was used for the peptide-protein docking of EPAP-1 P2 and P3 peptides with processed gp120 structure. The steric scan was set at a default of 16 and final search at 25. 3D FFT mode was set to search the angles of ligand and receptor. To search the space, the Distance Range was set to 40, The Scan step was set to 0.75, and the substeps was set to 2. All other parameters were set to default. The Docking was performed, and results were analysed.

2.4 Compound design by Fragment Building

EPAP-1 peptide P-2 was fragmented to give smaller 3 amino residue peptides. These short fragments were docked to gp120 using Autodock Vina and ranking of the fragments was done based on Autodock Vina scoring function. Fragments with higher rank were taken for further process. The process was repeated to obtain best hits. Pharmacophore was generated for the best hits obtained. Base structures were chosen from known inhibitors by taking constraints of volume, interacting residues etc. The functional groups and charges were added based on the obtained Pharmacophore.

2.5 Screening of compounds obtained from Fragment building

Molecular Docking was used for ranking the molecules based on their binding to the V3 loop. It was carried out, using Autodock -Vina v1.2. MGL Tools and PyMol were used for visualization of the results. Further, the grid was set with co-ordinates X=99.363, Y=-135.363, Z=148.056 and volume 60*60*60 Å. Each docking run was iterated ten times with each ligand. The iteration runs by Autodock Vina was automated by Bash scripts.

3. gp120 structure

Huang. C et al. deposited the gp120 crystal structure, 2B4C derived from JR-FL consisting of V3 loop. This structure was retrieved and processed to use as a receptor to understand the interactions of the obtained lead molecules. The resolution of the structure was 3.3 Å and consisted of 4 unique units and a total of 975 residues. The structure consisted of ENV glycoprotein gp120 with V3 loop, CD4 receptor, Heavy and Light chains of X5 antibody. The complex also consisted of some naturally occurring ligands and some stabilizing agents. The water molecules, ligands, CD4 and antibody fragments were removed for processing of the gp120 structure. The final structure was used for further studies.

3.1 Molecular Dynamics of gp120 structure

The processed PDB structure of gp120 was subjected to molecular dynamics simulation using the GROMOS96 43a1 force field. The simulation was carried in water using the SPC/E water model and a cubic box was defined to build the system(protein in water). The protein was placed in the centre of the box, 1nm from the edge. Thirteen water molecules in the solvent were replaced by negatively charged Chloride ions to balance the charge and neutralize the

system. Before proceeding with the simulations, the structure was subjected to energy minimization using the Steepest Descent Algorithm with 50000 iteration steps and a step size of 0.01 KCal/mol. The maximum force on the system was set to 1000KJ/mol/nm, and the minimization would stop if the force falls below the level. The minimized system was equilibrated at constant pressure and constant volume for 200ps each. The equilibrated system was subjected to final dynamic simulation for a 20ns run. The simulated structure was evaluated by RMSD and radius of gyration. GROMACS 5.0 was used for setting up and running the MD simulation.

The final structure obtained was used for further docking studies.

4. Docking of obtained hits with gp120 structures

Autodock Vina v1.2 was used to dock the selected, redesigned, synthesized and tested molecules with X-Ray structure of gp120 obtained from PDB record (2B4C) and the minimized and processed structure of gp120. Configuration files to run the docking for each set were prepared separately by placing a grid box on the receptor and defining the boundaries for binding region scan.

4.1 447-52D

a) For crystal structure of gp120 from 2B4C, the grid box size was set as 60*34*84 and was centred at 85.185, -133.514, 178.973 coordinates for X, Y and Z respectively. Since a large search space was defined, the exhaustiveness was set to 10. The output of 10 conformations per molecule was set as default.

b) For structure of molecular dynamics processed gp120, the grid box size was set as 32*40*58 and was centred at 59.09, 95.94, 58.77 coordinates for X, Y and Z respectively. Since a large search space was defined, the exhaustiveness was set to 10. The output of 10 conformations per molecule was set as default.

4.2 EPAP-1

a) For crystal structure of gp120 from 2B4C, the grid box size was set as 40*40*40 and was centred at 90.625, -133.514, 136.598 coordinates for X, Y and Z respectively. Since a large search space was defined, the exhaustiveness was set to 10. The output of 10 conformations per molecule was set as default.

b) For the structure subjected to Molecular Dynamics, the grid box size was set as 40*40*40 and was centred at 52.31, 90.18, 54.4 coordinates for X, Y and Z respectively. Since a large search space was defined, the exhaustiveness was set to 10. The output of 10 conformations per molecule was set as default.

5. Visualization and Analysis of Results

Discovery studio and AutoDock Tools were used to visualize the docked conformations and the interactions between the receptor and ligand.

6. Cell Culture Assays

6.1 Propagation of Cell lines

SupT1 cells and U-937 cells were cultured and maintained in RPMI 1640 supplemented with 10% FBS, 100U/ml and 100µg/ml of penicillin and streptomycin, respectively. SK-N-SH cells were maintained and propagated in Minimum Essential Media (MEM) supplemented with 2mM L-glutamine and Non-Essential Amino Acids and 10% serum. HL2/3 were propagated using Dulbecco's Minimum Essential Media (DMEM) supplemented with 10% serum. All the cell lines were maintained in the presence of 100U/ml of penicillin, and 100µg/ml of streptomycin. All cells were cultured in an incubator humidified with 5% CO2 at 37°C. The adherent cells were scraped using 0.5% Trypsin EDTA for experiments.

6.2 Measurement of cytotoxicity by MTT assay

Reduction of 3-(4,5-dimethylthiozol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, Sigma) was chosen as a measurement of cytotoxicity. SUP-T1 cells were seeded in a 96-well plate at a density of 0.2x106 cells/well and incubated at 37°C in a 5% CO2 incubator for 4 hours. After incubation, the cells were treated with increasing concentrations of synthesized and characterized compounds dissolved in their respective solvents. The cells were incubated for 16 hours in the 5% CO2 incubator. The cells after incubation were pelleted at 1200 rpm for 7 min and re-suspended in complete medium. From the working concentration of 5 mg/ml pre-dissolved MTT, 20µl was added to each well and incubated for 4 hr. Following incubation, the cells were pelleted at 1200 rpm for 10 minutes, the media in the wells was discarded, and MTT-formazan crystals were dissolved in DMSO by adding 100uL into each well. Following incubation of the plate in the dark for 5 mins, the colour change was recorded in an ELISA reader at 595 nm. The triplicate experiments with standard deviation were plotted to represent

cell survival. Survival in cell lines U937 and HL2/3 was also tested using the above-mentioned protocol.

7. Measurement of anti-viral activity in-vitro:

7.1 Propagation of Virus:

HIV-1 primary isolate 93IN101 was propagated in T lymphocytic cell line SupT1, promonocytic cell line U937 and neuroblastoma cell line SK-NSH. Primary isolate, 93RW024 and laboratory-adapted strain NL4-3 were propagated in SupT1 Cell line. Laboratory strain, Ba-L was propagated in SupT1 and U937. 5 * 107 cells were incubated with stored, undiluted virus, in medium supplemented with 0.1% FBS for four hours in 5% CO2 incubator at 37◦C. After incubation, the cells were washed, pelleted and resuspended with complete medium. The washed cells were incubated for 96 hrs. The supernatant was collected and tested for the p24 antigen after 96hrs. The virus was propagated for multiple passages, and upon achieving the good yield, the virus was collected and used for further experiments.

7.2 p24 quantification:

HIV-1 infection was quantified by detecting the levels of viral envelope antigen p24. After incubating the cells for 96 hrs, the viral core p24 antigen in the cell supernatant was quantified by using p24 Antigen ELISA Kit. The protocol for the quantification was as per the manufacturer's guide. Briefly, supernatants were collected and diluted 10-fold with complete media. 100µl of the dilutions were added to each well of the precoated ELISA plate. To each well, 25µl disruption buffer was added, and the plate was sealed. Incubation for 1 hr at 37°c was followed by aspiration of the contents of each well. The wells were washed with wash buffer 3x times, and 100µl of the conjugate solution was added to each well. The sealed plate was incubated for 1 hour. After 1hr, the solution was aspirated, and the plate was washed for three times. The aspirated solution was replaced with 100µl of substrate solution in each well, and the pate was sealed. The reaction was stopped after incubation at room temperature for 20 min by adding 100µl of 1N HCl. The plate was quantified for optical density at 450nm by ELISA reader. A curve was established using standard p24 protein dilutions provided by the manufacturer. The p24 concentration of propagated virus was calculated according to the standard curve obtained.

7.3 HIV-1 neutralization assay

Co-Infection:

106 SupT1 cells suspended in RPMI1640 supplemented with 0.1% FBS were seeded per well. Increasing concentrations of synthesized compounds were added to the cells. The cells were then infected with 1ng/ml p24 equivalent of 93IN101 and incubated for 4hrs at 37°C in 5% CO2 incubator. After incubation, the cells were pelleted at 1200rpm for 5 mins, and the supernatant was discarded. The pellets were washed in RPMI 1640 supplemented with 10% FBS. The washed pellets were resuspended with complete fresh medium, and the cells were incubated for 96h.

The supernatants collected after incubation were analysed for the p24 antigen using the protocol described above. p24 levels in the absence of compounds were considered as 0% inhibition and were taken as the negative control. The infection levels in the presence of T20(Enfuvirtide) was considered as the positive control.

Pre-incubation with Virus:

The 1ng p24 equivalent virus was incubated with the synthesized compounds at 37°C in 5% CO2 incubator for 1hr. 1ng of incubated without compounds was considered negative control and virus incubated with T-20 was considered positive control. 106 SupT1 cells with 100% viability, suspended in RPMI 1640 supplemented with 0.1% FBS were seeded in each well. After 1hr, the incubated virus is added to the cells and further incubated for 4hrs at 37°C in 5% CO2 incubator. The cells were pelleted, washed and resuspended in complete media. The cells were incubated for 96hrs in 5% CO2 incubator. After the incubation period, p24 antigen levels were quantified in the supernatant after incubation using the method described previously.

7.4 Broad spectrum anti-HIV-1 activity

Activity against HIV-1 subtypes

The activity of the compounds against HIV-1 Subtypes A, B and C were tested by propagating different primary isolates and laboratory strains in the method described above. Cells were infected with the propagated virus, either by co-infection or by incubating cells with the pre-incubated virus and finally quantifying the p24 levels.

Activity in different cell types

T-lymphocytic, promonocytic and neuroblastoma cell lines were infected with the respective virus as described above. The cells were treated with virus and compounds in different concentration through co-infection. The virus was preincubated with compounds and after incubation was added to different cell lines. The p-24 levels were quantified by ELISA method defined previously.

8. Cell Fusion by Dye Redistribution Assay

a) HL2/3 labelling :

HL2/3 cells expressing HIV-1IIIB Env on the surface were incubated with 0.5 μM of Calcein AM for 1 h at 37 °C. After incubation, the cells were washed thoroughly and were incubated in fresh medium at 37 °C for 30 mins. To reproduce the preincubation experiment, the cells were incubated with the synthesized compounds for 1hr after loading the dye.

b) SupT1 labelling:

SupT1 cells were incubated for 1 hr with 20 μM of Calcein Blue at 37 °C. After incubation, the cells were washed by spinning at 1200 rpm. After washing, the pellet was resuspended in fresh medium and incubated 30 min at 37 °C.

c) Fusion Assay:

The Env expressing cell line(HL2/3) and CD4+, CCR5+, CXCR4+ (SupT1) previously labelled with fluorescent dyes were co-cultured at a ratio of 1:1 and incubated at 37 °C for 2hrs. The labelled SupT1 cells at a cell count approximately equal to HL2/3 were added to and incubated as specified above. After incubation, the % of fused cells was observed using a Leica Fluorescent microscope. The fusion was monitored at IC80 concentration of the compounds, and the fusion in the presence of T20 was considered positive control. Cell fusion in the absence of compounds was considered as control. The images were acquired from the microscope and to confirm uniformity, different regions were scanned.

Interacting residue: V3 loop	Interacting residue: 447-52D	Type of Interaction
Arg306:NE	Tyr100H:OH	Hydrogen Bonding
Ile307:N	Asp100F:O	
Ile307:O	Tyr100H:N	
His308:ND1	Tyr100J:OH	
Ile309:N	Tyr100H:O	
Ile309:O	Tyr100J:N	
His308	Tyr100H	Non-bonded
Ile307:CD1	Tyr100G	Hydrophobic
Ile309:CG2	Tyr32	Non-bonded
Ile309:C,O;Gly312N	Trp91	Non-bonded
Gly312:C,O:Pro313N	Trp91	Non-bonded
Arg315:NH1	Trp33	Cation-π
Arg315:NH2	Asp95:OD2	Salt Bridge
Arg306	Tyr100H	Non-bonded
Ile309	Tyr100I	Hydrophobic
Arg315	Tyr100J	Cation-π

Tab. 5: Interacting residues between 447-52D and V3 Loop giving as input to pep:MMs:MIMIC

MMsINC_ID	Score	MMsINC_ID	Score	MMsINC_ID	Score
MMs02188495_c2	0.4	MMs00886437_c2	0.35	MMs03493210_c1	0.35
MMs02514398_c1	0.39	MMs02514241_c3	0.35	MMs02460893_c2	0.35
MMs02188494_c4	0.37	MMs02564531_c5	0.35	MMs03112165_c1	0.35
MMs03114524_c3	0.37	MMs03113826_c4	0.35	MMs00886451_c3	0.35
MMs02895170_c3	0.36	MMs02895165_c2	0.35	MMs02460892_c2	0.35
MMs02895206_c2	0.36	MMs02898061_c1	0.35	MMs03928968_c1	0.35
MMs03115698_c2	0.36	MMs02898147_c3	0.35	MMs02489200_c4	0.35
MMs02895032_c1	0.36	MMs03867811_c5	0.35	MMs03867835_c5	0.35
MMs02820793_c5	0.36	MMs03867812_c5	0.35	MMs02001984_c4	0.35
MMs02188495_c2	0.4	MMs00030268_c5	0.35	MMs03867827_c4	0.35

Tab. 6: Hits obtained from virtual screening of MMs database by pep:MMs:MIMIC. 200 hits were reported. Top 30 are represented here sorted in descending order of the reported score

Results and Discussion:

1. **447-52D**

1.1 447-52D-V3 loop complex:

The PDB file, 1QIJ indicated the interactions, specified in **Tab. 5,** between IIIB V3 loop and 447-52D antibody. The V3 loop extracted from the PDB complex along with the residues 'Arg306:Side chain', 'Ile307:P Side chain', 'Ile309:P Side chain', Arg'315:P Side chain', 'Pro313:P Side chain', 'Gly312:P Backbone CO', 'His308:P Side chain' were given as an input to pep:MMs:MIMIC server.

1.2 Virtual Screening:

A pharmacophore developed by the server using the parameters specified was used to screen the MMsINC database of small molecules to return hits. Two hundred small molecules were reported and scored, based on the match to the pharmacophore. Top 30 molecules are reported along with scores in **Tab. 6**.

1.3 Screening of reported molecules

The 200 molecules reported by pep:MMs: MIMIC were docked using GOLD by the protocol described previously. The top 50 (**Tab. 7a**) from the hit library were docked to gp120 whole structure (**Fig 19**). The amino acid residues, Arg306 and Lys307, were seen to be involved in the interaction with most molecules. These residues were also crucial in the interaction interface of 447-52D and V3 loop. Apart from these residues, Cys297 and Cys331 were also important for the interaction. Based on the number, type, other parameters of interactions and the GOLD score, ten molecules **Tab. 7b** were selected for synthesis and validation. The selected molecules are given in **Fig 20**.

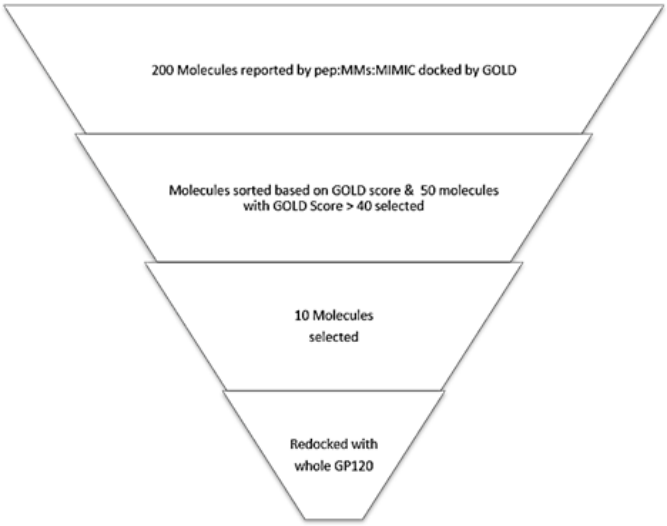


Fig. 19: The filtering of hits to obtain potential lead molecules for in-vitro validation

a)

Redocked molecules	GOLD score	Redocked molecules	GOLD score
MMs02001984	60.19	MMs03115698	50.04
MMs02489197	44.10	MMs00458616	51.85
MMs02898061	44.65	MMs03493215	53.45
MMs02489200	45.39	MMs03655090	54.89
MMs03113826	46.42	MMs03635663	54.21
MMs03606040	47.34	MMs00886401	56.99
MMs02186806	47.67	MMs00886437	59.12
MMs03895345	48.38	MMs00886371	59.43
MMs02242938	48.40	MMs02596171	57.19
MMs03112165	48.68	MMs00172364	49.09

b)

MMsINNC Id	GOLD score	No of Interactions	Overlapped interactions
MMs02001984	**60.19**	**09**	**Arg306**
MMs00886371	59.43	12	Arg 306,Lys 307
MMs00886437	**59.12**	**14**	**Arg306**
MMs00886428	**57.92**	**19**	**Arg306,Lys307**
MMs02596171	57.19	09	Arg306
MMs00886401	**56.99**	**11**	**Arg306**
MMs03655090	**54.89**	**15**	**Arg306**
MMs02273758	**54.36**	**21**	**Arg306,Lys307**
MMs03635663	54.21	11	Arg306,Lys 307
MMs03493210	54.18	09	Arg306,Lys307
MMs03493215	**53.45**	**11**	**Arg306**
MMs00458616	**51.85**	**08**	-
MMs03115698	**50.04**	**15**	**Arg306,Lys307**
MMs03732486	**44.01**	**15**	**Arg306,Lys307**

Tab. 7: a) Top 20 molecules obtained on redocking given along with GOLD score. b) Top 15 molecules with score, number of interactions and the residues observed in the binding region of V3 loop. The molecules selected are highlighted in bold

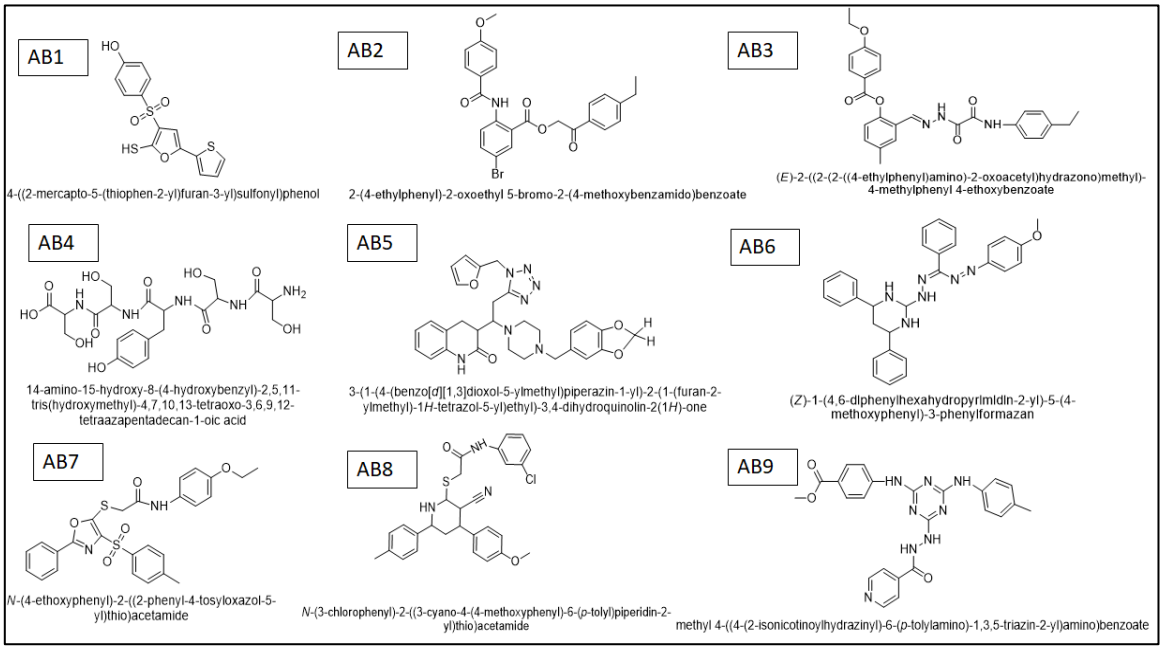


Fig. 20: Top nine molecules selected from the virtual screening protocol for 447-52 mimics

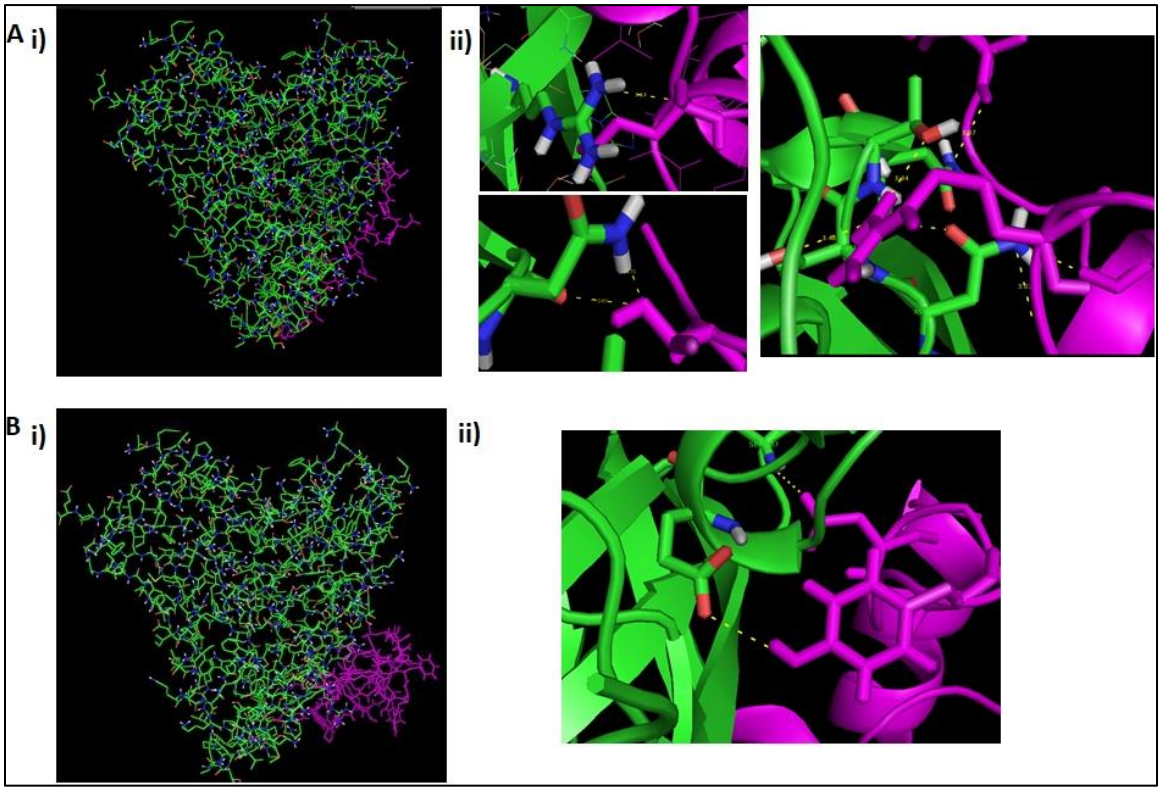


Fig. 21: A i) Docking of the peptide 2(magenta) with gp120(green) performed using HEX Docking suite. **ii)** Depicts the interactions between the peptide and the residues of V3 loop; The interacting residues of V3 loop are ASN386, SER387,THR388,ASN392 and ARG419.The interacting residues of Peptide 2 include CYS4, VAL23, ARG26. **B i)** Docking of the peptide 3(magenta) with gp120(green) performed using Hex Docking suite. **ii)** Depicts the interactions between the peptide and the residues of V3 loop; The interacting residues of V3 loop are GLU370,SER387.The interacting residues of Peptide 3 include TYR2,ASP5.

2.2 Compound design by Fragment Building

The peptide -RSTCALTAATAKAYATRSLEHRVVYRILHDC- was broken into shorter fragment(Sp) and the structure of the fragments was modelled and optimized. The optimized structures were docked to the using the protocol described. A total of 27 short three amino acid fragments were screened for favourable interactions with gp120. The fragments were ranked based on the docking score (**Tab. 8**). The fragments with score above 5.5 Kcal/mol were selected for further analysis.

The best performing fragments were Sp10, Sp13, Sp15, Sp16, Sp 18, Sp 21, Sp22, Sp 24, Sp25, the scores are given in Table 9. The residues commonly observed interacting with these peptide fragments included Asn280, Arg 335, Thr 408 and residues in the vicinity. The interacting interface of the fragments closely matched the interaction observed for peptide P2.

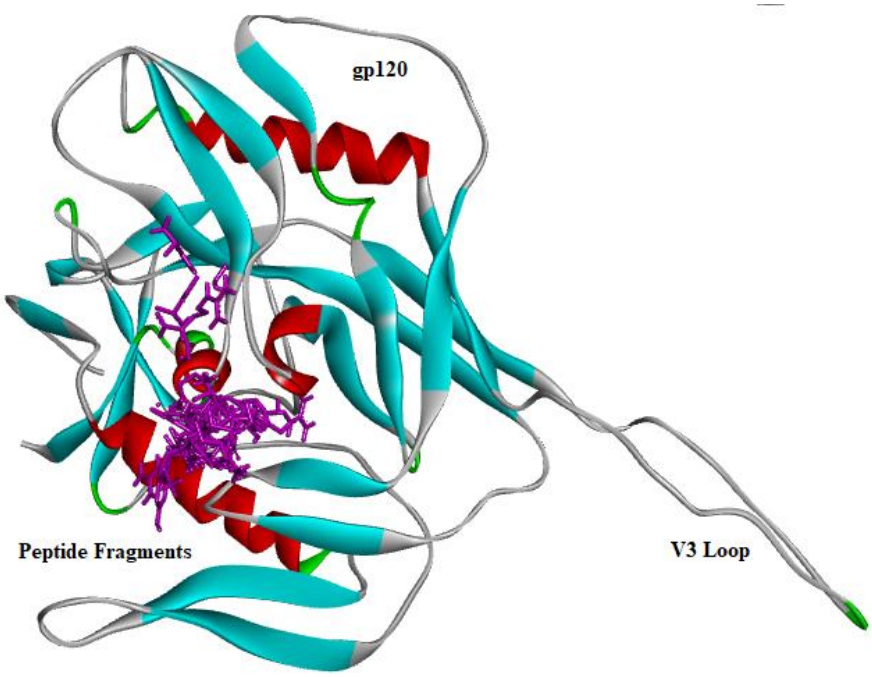


Fig 22: The docked complex of gp120 with the selected peptides. The peptides are represented in pink

S.No	Pep	Binding Energy (Kcal/mol)	S.No	Pep	Binding Energy (Kcal/mol)
Sp1	ILH	-5.2	**Sp15**	**YAT**	**-5.7**
Sp2	LHD	-5.2	**Sp16**	**ATR**	**-6.1**
Sp3	HDC	-5.0	**Sp17**	TRS	-4.8
Sp4	RST	-4.5	**Sp18**	**RSL**	**-6.0**
Sp5	STC	-4.8	**Sp19**	SLE	-5.2
Sp6	TCA	-4.5	**Sp20**	LEH	-4.6
Sp7	CAL	-4.8	**Sp21**	**EHR**	**-6.2**
Sp8	ALT	-4.9	**Sp22**	**HRV**	**-6.7**
Sp9	AAT	-5.0	**Sp23**	RVV	-5.0
Sp10	**ATA**	**-5.5**	**Sp24**	**VVY**	**-5.6**
Sp11	TAK	-4.7	**Sp25**	**YRI**	**-6.3**
Sp12	AKA	-4.2	**Sp26**	RIL	-5.2
Sp13	**KAY**	**-5.7**	**Sp27**	TAA	-5.4
sp14	AYA	-5.1			

Tab. 8: Three residue long peptide fragments obtained from peptide P-2. The AUTODOCK Vina score is mentioned next to the fragment. The fragments that showed binding with the binding energy scoring fuction greater than 5.5 KCal/mol are highlighted in bold.

2.3 Pharmacophore and Molecules:

To understand the contributions to binding, pharmacophore was developed for the selected fragments. Features like the volume (calculated by Peptide Property Calculator) and interactions were assessed. The pharmacophore was developed using PharmaGist server. The pharmacophore was generated by overlapping the fragments considering Sp 22 (high scoring) as an anchor. The common features of the aligned fragments were assessed and reported. One Hydrophobic contribution, one negatively charged group, five hydrogen bond acceptors and three donors were the reported features of the pharmacophore. Apart from these the length and volume of the fragments were considered important to establish binding with gp120. (**Fig. 23**). Considering the pharmacophoric features of the high binding fragments, molecules were designed to satisfy the properties. The base structures considered for fragment building included coumarin and curcumin to satisfy the volume and area considerations and also since they are an active part of many bioactive compounds (Stefanachi et al., 2018). The best nine compounds are represented in **Fig. 24**.

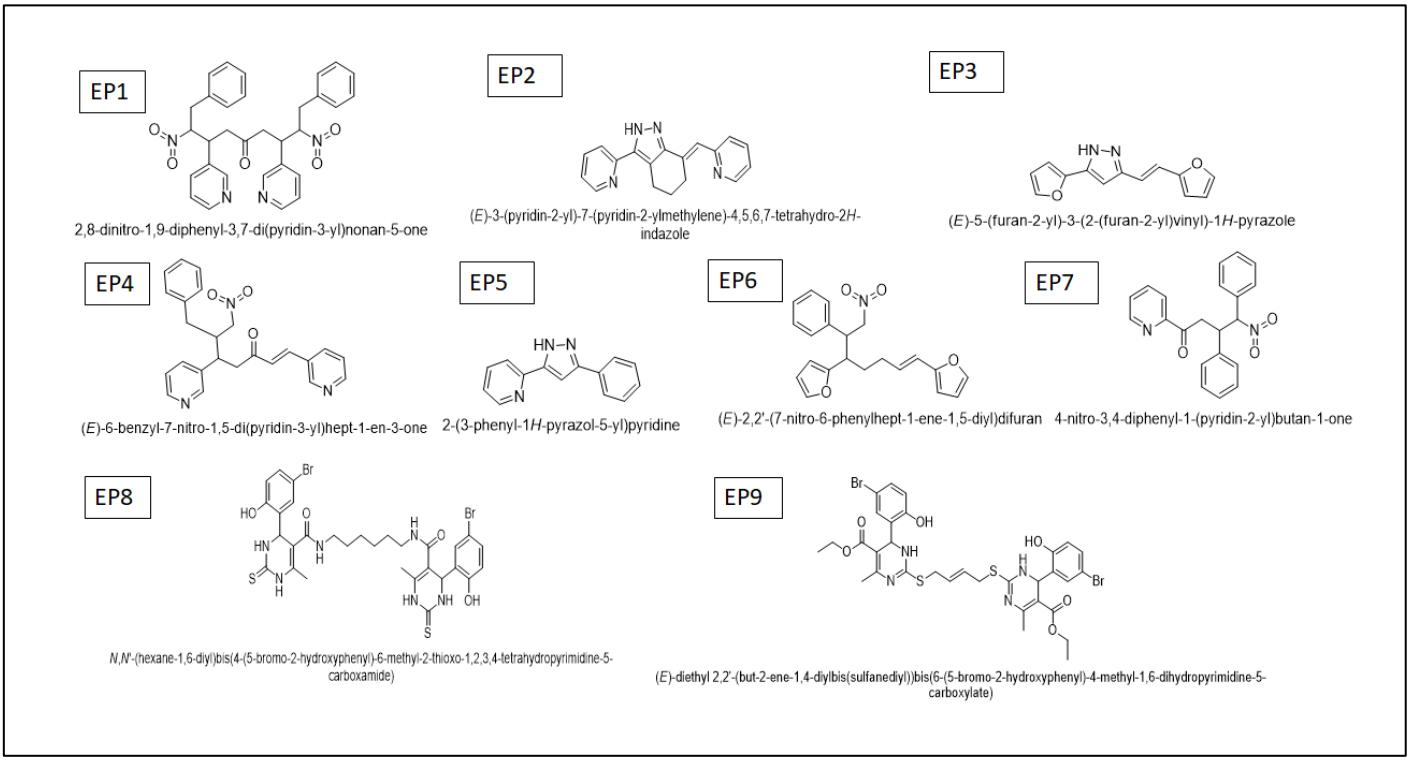
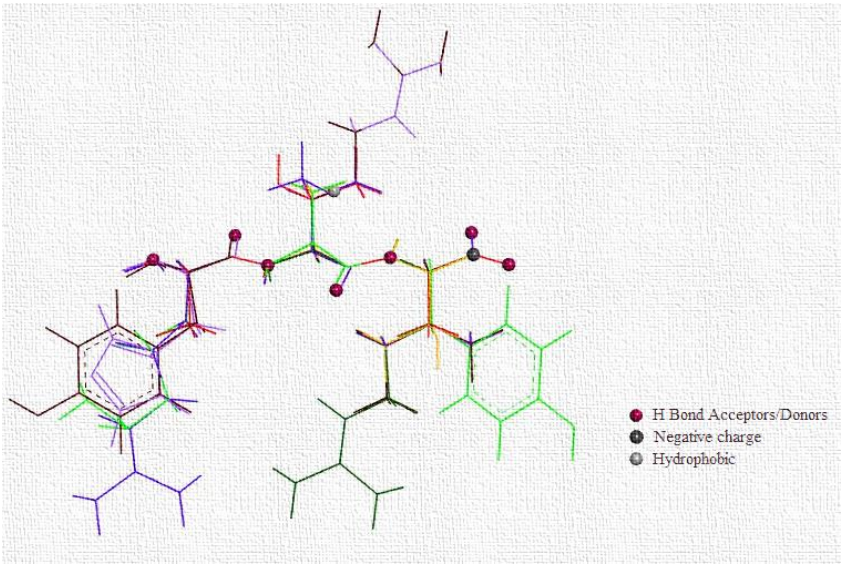


Fig. 24: Compounds designed based on the pharmacophore obtained from the high affinity fragments.

S.No	Peptide	Binding Energy (Kcal/mol)	Volume in Å
Sp10	ATA	-5.5	316
Sp13	KAY	-5.7	461
Sp15	YAT	-5.7	427
Sp16	ATR	-6.1	419
Sp18	RSL	-6.0	453
Sp21	EHR	-6.2	532
Sp22	HRV	-6.7	496
Sp24	VVY	-5.6	459
Sp25	YRI	-6.3	545

Tab. 9: The selected fragments with docking score and volume.



Score	Features	Spatial Features	Aromatic	Hydrophobic	Donors	Acceptors	Negatives	Positives
37.908	10	9	0	1	3	5	1	0

Fig.23: Pharmacophore generated by PharaGist based on the selected nine fragments. The fragments were aligned to assess various features common in the fragmenst. The table elucidates the pharmacophoric features observed.

The docking of final compounds with gp120 structure reported good affinity with residues from the region of fragment binding . The volume constraints have also been satisfied. The score along with residues and Volume observed are given in **Tab. 10**.

Molecule Id	Binding affinity (kcal/mol)	Interacting residues	Volume in Å
EP1	-6.7	ASN362, SER447	706
EP2	-6.4	VAL255, ASN377	402
EP3	-5.8	VAL255, ASN377	301
EP4	-6.4	GLY379, SER447	440
EP5	-6.3	GLY379	512
EP6	-6.2	-	320
EP7	-5.8	-	461
EP8	-6.2	ARG252, GLY379, PHE210, ASN377	820
EP9	-6.7	ARG252, PHE210, ASN377	645

Tab.10 : Table depicts the compounds developed from fragment building, Auntodock scoring function, the interacting reisdues. The volume of the compounds is also reported.

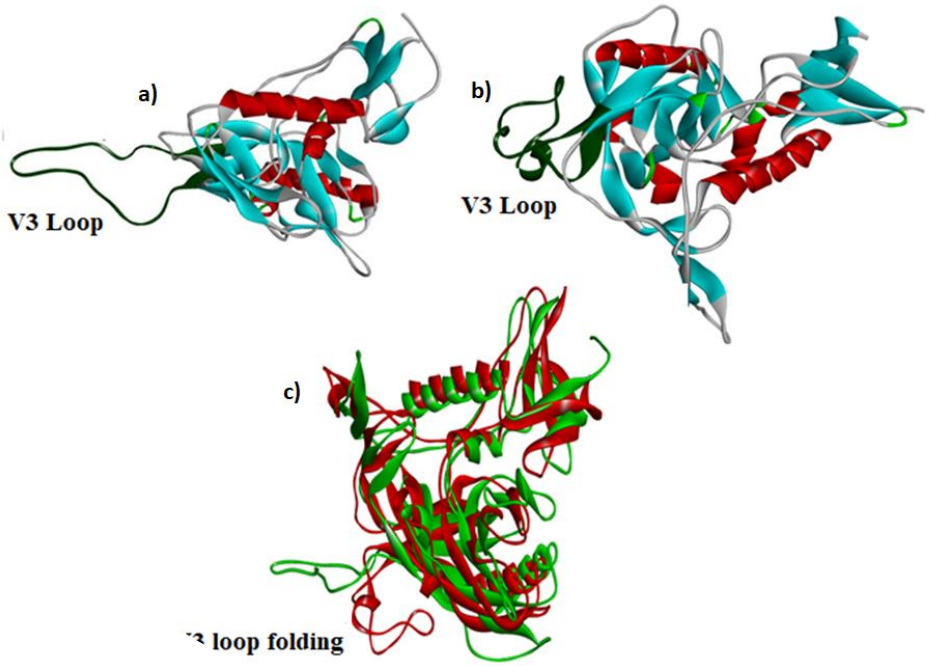


Fig. 25: a) X-Ray crystal structure of gp120 with intact V3 loop (2B4C) initial structure b) gp120 structure obtained after 20ns MD simulation run. Final structure c) Overlapped initial and final structures of gp120 molecular dynamics run. A clear change in conformation of V3 loop is observed in the overlapped structure.

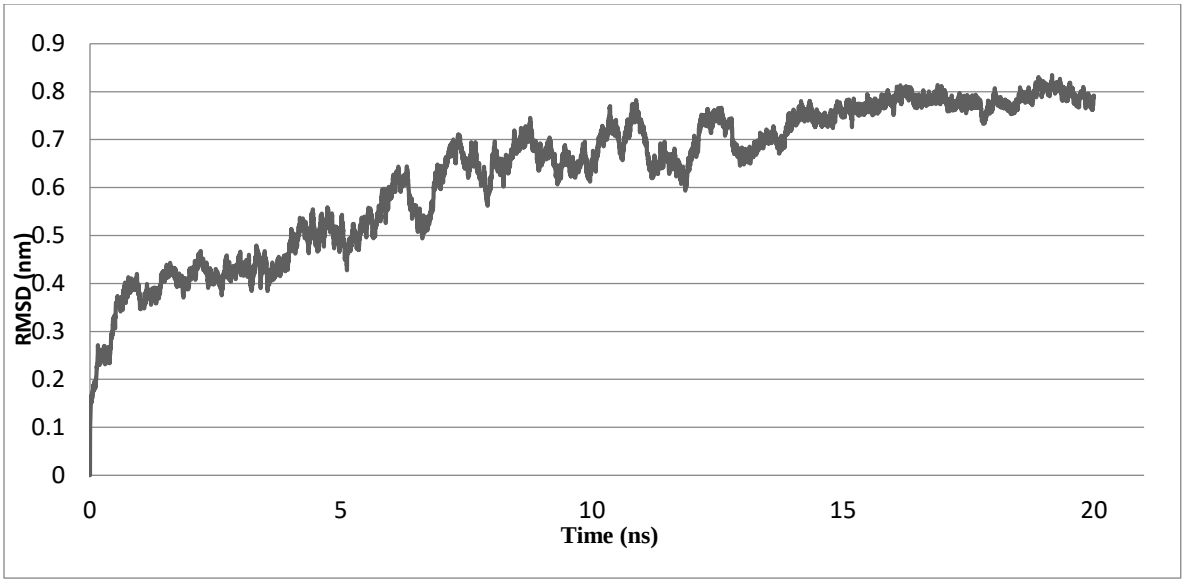


Fig. 26: Graph plotted between time of simulation in nanoseconds and RMSD in nm observed between previous and current structures in the trajectory. The graph is unstable up to 15ns after which it stabilizes between 15-20ns

3. gp120 structural dynamics

The structure of the gp120, complexed with CD4 and 447-52D antibody, retrieved from RCSB PDB is depicted in **Fig. 25a**. The simulation was carried out using the standard method with the parameters mentioned previously. RMSD graph plotted showed high variation until 15ns after which it stabilized with minor variations between 15ns to 20ns. The structure of gp120 after the 20ns simulation is depicted in **Fig. 25b**. The RMSD graph (**Fig. 26**) shows the stabilization of the structure after 15ns. The Carbon backbone of the structure before and after simulation was overlapped to monitor the structural changes. Major structural change was seen in the V3 loop of the gp120 structure, where an open structure seen when complexed with CD4 was changed to a closed structure when subjected to the molecular dynamics protocol. This change in conformation between CD4 bound and unbound gp120 has been reported in several studies.

One research paper submitted by Xiang et al. showed the difference in structures while investigating the conformational dynamics of V3 loop in the presence of antibodies bound to different regions of gp120. The structural change reported was similar to **Fig. 25c**. In another research article reported by Bowder et al., 2018 on the contribution of a V3 loop hydrophobic patch in stabilizing gp120 trimer, the conformational dynamics of V3 loop was discussed. They overlapped an unbound gp120 crystal structure of sgp140 SOSIP.664 and the gp120 structure from the bound crystal structure. The overlapped structures reflected a similar observation as the one in this study. This change in conformation was attributed to the engagement of gp120 and CD4, leading to the unfolding of the V3 loop to allow contacting the co-receptor. One study reported by Hollingsworth et al., 2018 conducted a similar molecular dynamics simulation of gp120 structures obtained from various laboratory-adapted viral strains and performed docking of 112 anionic ligands. This study showed a similar pattern of dynamics of the V3 loop with gp120 with some strains folding back to native state faster than others. The group conducted docking studies of the ligands in both open and closed structures to identify residues important for interactions despite the conformational dynamics. The study concluded that the recognition of gp120 binding regions by small molecules is dependent on the V3 structural dynamics as some molecules recognize open, and some recognize closed structures better. A transient structure also exists that can be recognized by both.

With the above reports and many more studies that emphasize the importance of V3 conformation in the stabilization of the gp120 structure, engagement with coreceptor and in

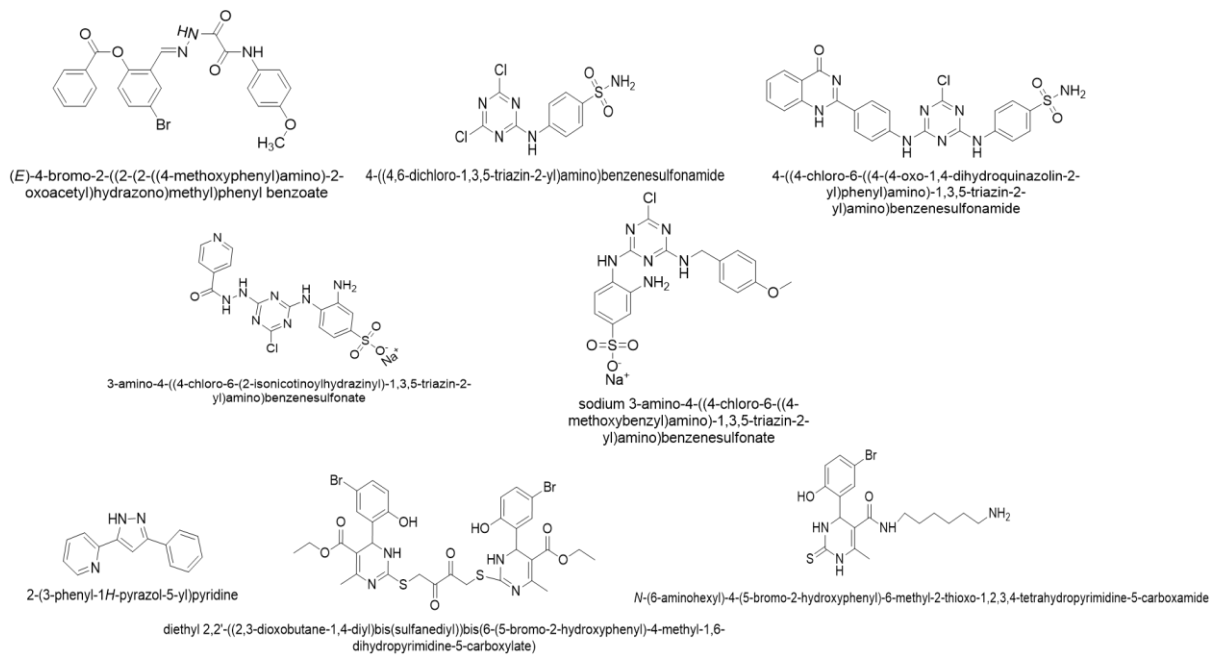


Fig. 27: The final molecules designed for validation of in-vitro activity. Compounds D1-D5 are derived from 447-52D and compounds D6-D8 are derived from EPAP-1 peptide P2.

Compounds	Closed V3 Loop conformation (Kcal/mol)	Open V3 Loop conformation (Kcal/mol)	Interacting Residues (Closed V3 Loop)	Interacting Residues (Open V3 Loop)
D1	-7.0	-6.1	Arg 304, Gly312,Asn 332	Gly 312, Arg 315
D2	-6.6	-5.7	Arg 304, Ser 387, Asn 295	Gly312,Gly314, Arg 315
D3	-8.5	-7.7	Arg 304,Tyr 318, Thr 394	Gly 312, Pro 313,Gly 314
D4	-7.8	-6.9	Gly 312, Asn 332, Arg 304	Gly 312, Pro313,Gly 314
D5	-7.2	-6.1	Arg304, Ile309, Asn 295	Gly 312, Pro314, Gly314
D6	-6.1	-6.3	Gln 344, Asn 406	Asp 368, Trp 427
D7	-7.8	-7.8	Asn262, Phe 376	Ser 365, Asn 280
D8	-6.8	-7.1	Asn 339, Thr 394	Glu 370, Asn 425

Tab. 11: Molecular Docking results of the compounds D1-D8. The Docking was performed with both native and CD4 bound V3 loop structures (close and open). The residues in the interacting regions are tabulated.

the neutralization of virus, both open and closed loop gp120 structures were used in the study to understand the binding and interactions of the synthesized compounds.

4. Synthesized and validated compounds

The compounds obtained from virtual screening using 447-52D as a base and from fragment building using EPAP-1 were considered for synthesis. Based on the ease and possibility of synthesis, 8 molecules in total were filtered. Changes were made to the side chains of the compounds to satisfy some properties (**Fig. 27**).

Since the 447-52D derivatives were obtained from MMs database, the molecules were known to possess biological activities. The 1,3,5-triazine-2,4,6-triamine scaffold of molecule AB9 was previously established to possess anti-HIV property (Modh et al., 2013). This was used as a base to design molecules D2, D3, D4 and D5 as the frame. The molecule AB3 was modified to include certain functional groups to accommodate electrostatic interaction with V3 loop of gp120 and it was labelled as D1.

To satisfy the pharmacophoric features, EPAP-1 peptide derivatives EP2, EP3, EP5 molecules were designed with pyrazole rings, which has been reported in many bioactive compounds. Hence, though not satisfying the length and volume constraints, EP5 was selected for further synthesis and labelled D1. Compounds EP 8 and EP 9 were taken further, and modifications were made. To simplify the synthesis, EP8 retained only one of the Tetra hydropyramidne thione ring branched with 5-Bromo-2-hydroxyphenyl ring. In EP9 the aliphatic chain used for increasing the length of the compound was modified to include carbonyl groups to facilitate participation in hydrogen bond . EP8 was designated as D8 and EP9 was designated D7.

The designed molecules D1-D8 commonly contained sodium sulfonate , sulphonamide, quinazolin ,pyridine, methoxy aniline structural groups that are commonly observed in bioactive compound.

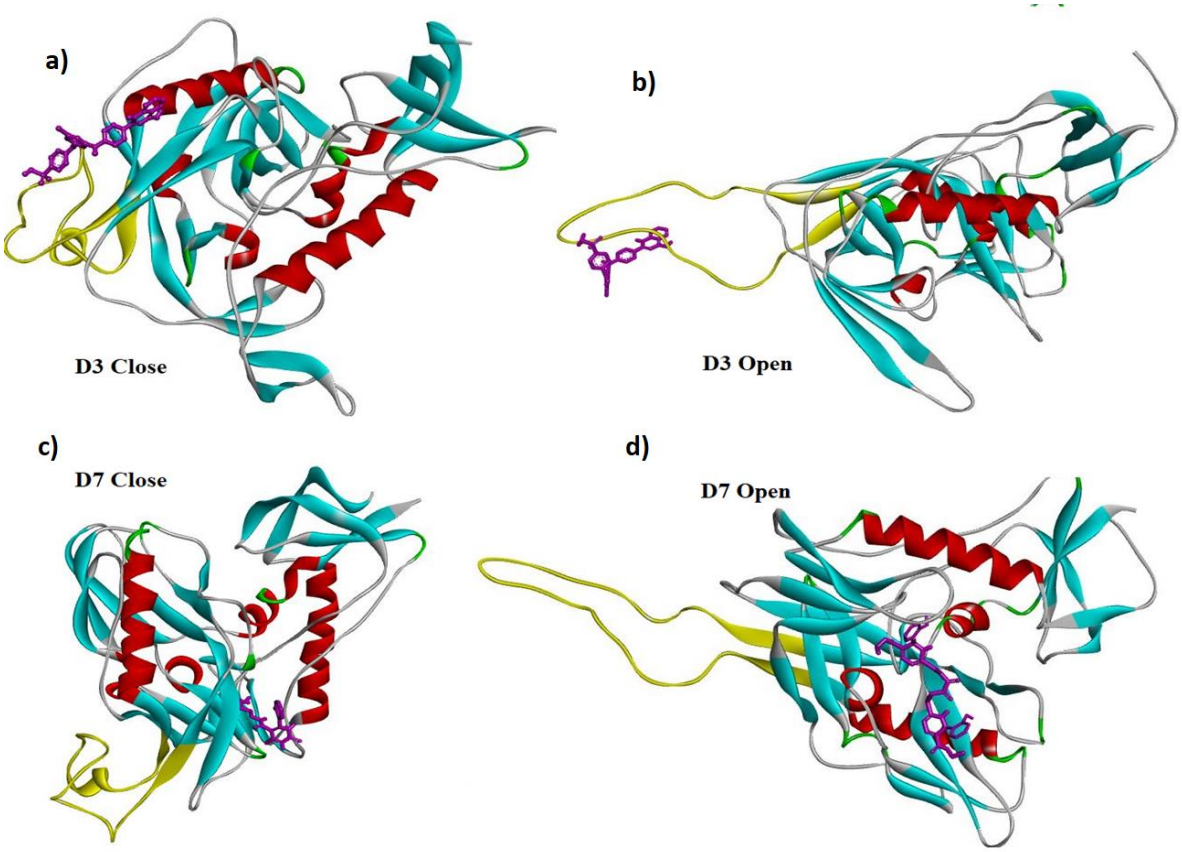


Fig. 28: Docked structure of compounds D3 and D7. a) D3 compound docked with closed V3 loop conformation of gp120 b) D3 compound docked with open loop conformation of gp120 c) D7 compound docked with closed V3 loop conformation of gp120 d) D7 compound docked with open loop conformation of gp120.

The selected molecules were docked with the open(2B4C), and closed structures(post-MD 2B4C) of gp120 and the binding energies were analysed to understand the mode of action of the compounds (**Tab. 11**) (**Fig 28**).

From the docking, it was seen that EPAP-1 mimetics showed a greater affinity of binding to the V3 loop base and surrounding residues. Hydrogen Bonding and cation-π interactions were commonly observed. The binding affinity to the regions were intact irrespective of the conformation of gp120. This was seen in the similar AUTODOCK Vina score and interacting region for closed and open conformations (**Fig 29a and b**). The residues in the binding region for both conformations were in the vicinity of one another. We assume that the folding of the V3 loop does not interfere with the interacting interface of the compounds despite slight structural changes in the conformation, thus resulting in similar affinities. It was observed that the volume and area formed by D7 had a major role in defining the affinity of the compound.

From the docking scores and the interacting residues, it was observed that the binding region for 447-52D mimetics was similar to the antibody interacting interface. Electrostatic interactions, Hydrogen bonding and cation-π were observed in the binding of the compounds, while electrostatic interactions were dominant. The compounds designed to possess an overall negative charge tend to interact with greater affinity due to the dominant positive charge of the V3 loop. The binding region also included amino acids from the base of the V3 loop in closed structures, most prominent being Asn332, Thr297 and Asn295. When scores for open and closed V3 loop conformation was compared, better affinity was observed for closed structures than open structures in case of 447-52D mimetics. This was also reflected in the interacting residues, where 447-52D mimetics preferred a region of the closed structure where the stem, crown and base of the V3 loop and other regions of gp120 were in proximity, due to the folded V3 conformation (Fig 29a). Among various residues outside V3 loop involved in the binding were Arg444, Thr413, Val292, Lys337 in the closed loop conformation.

The compound D3,D4 developed from 447-52D and compound D7 developed from EPAP-1 showed good affinity by in-silico method. To understand and validate the developed compounds further, in-vitro assays were performed.

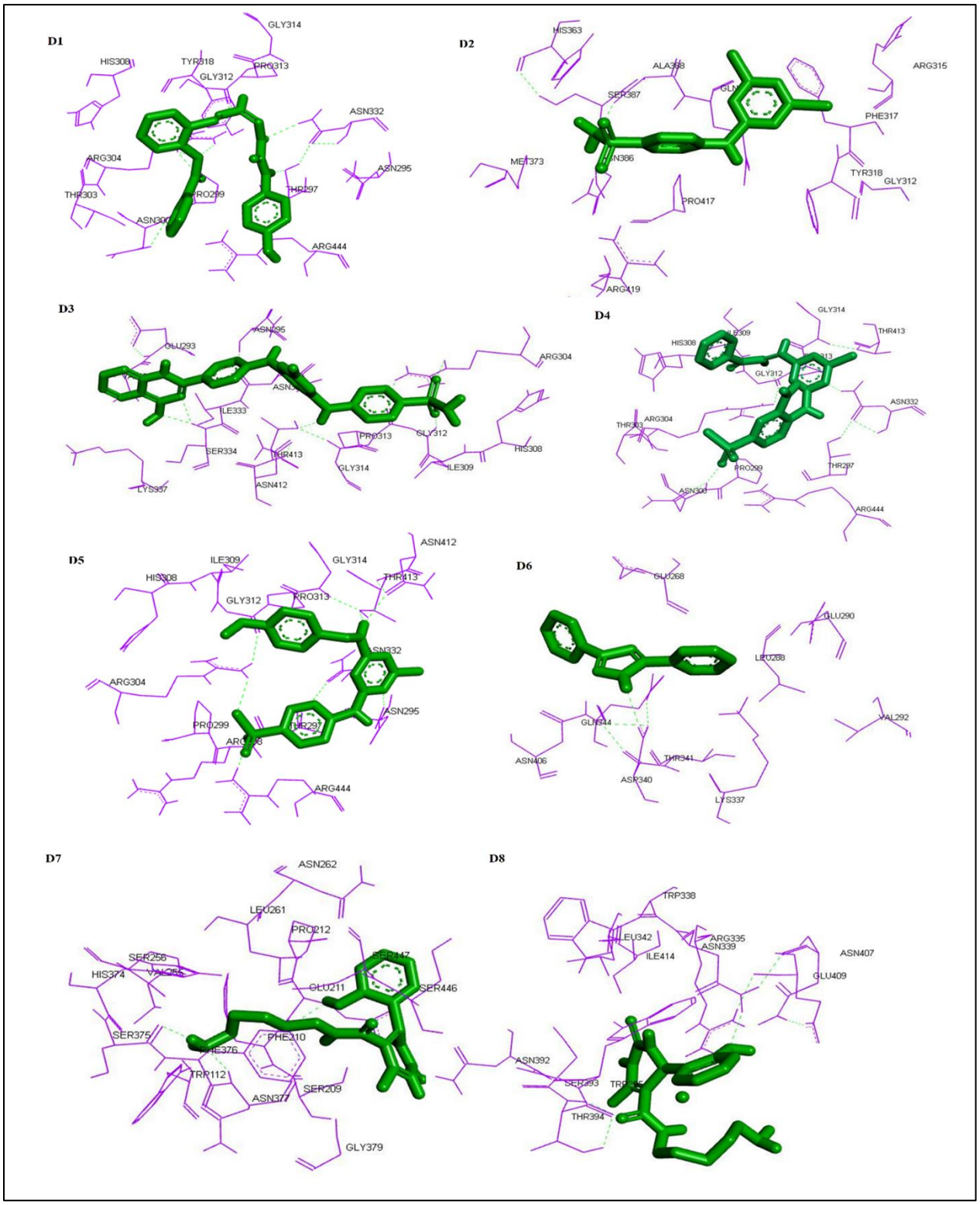


Fig. 29a) Interacting regions and residues of closed loop-gp120 docked with compounds D1- D8

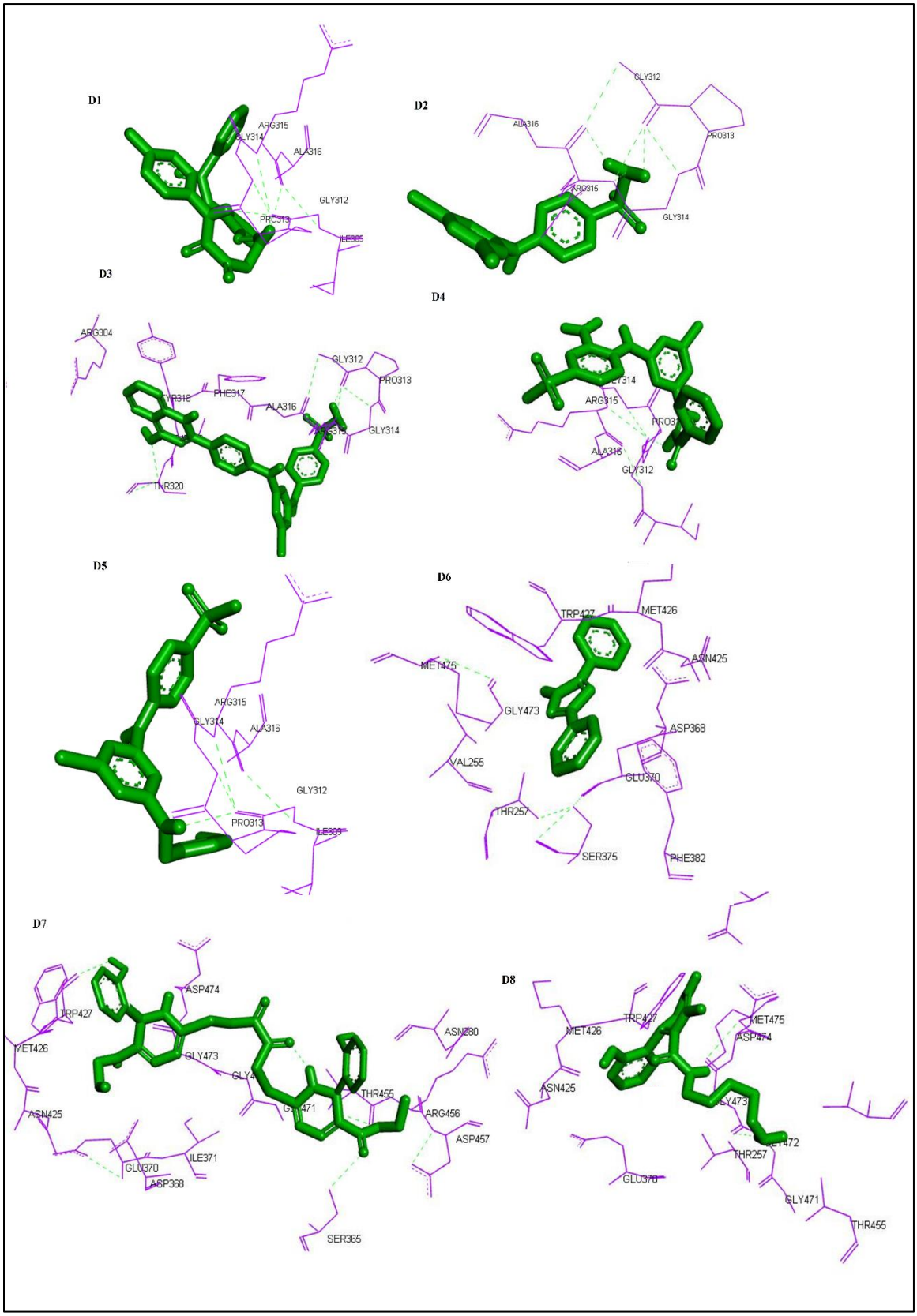
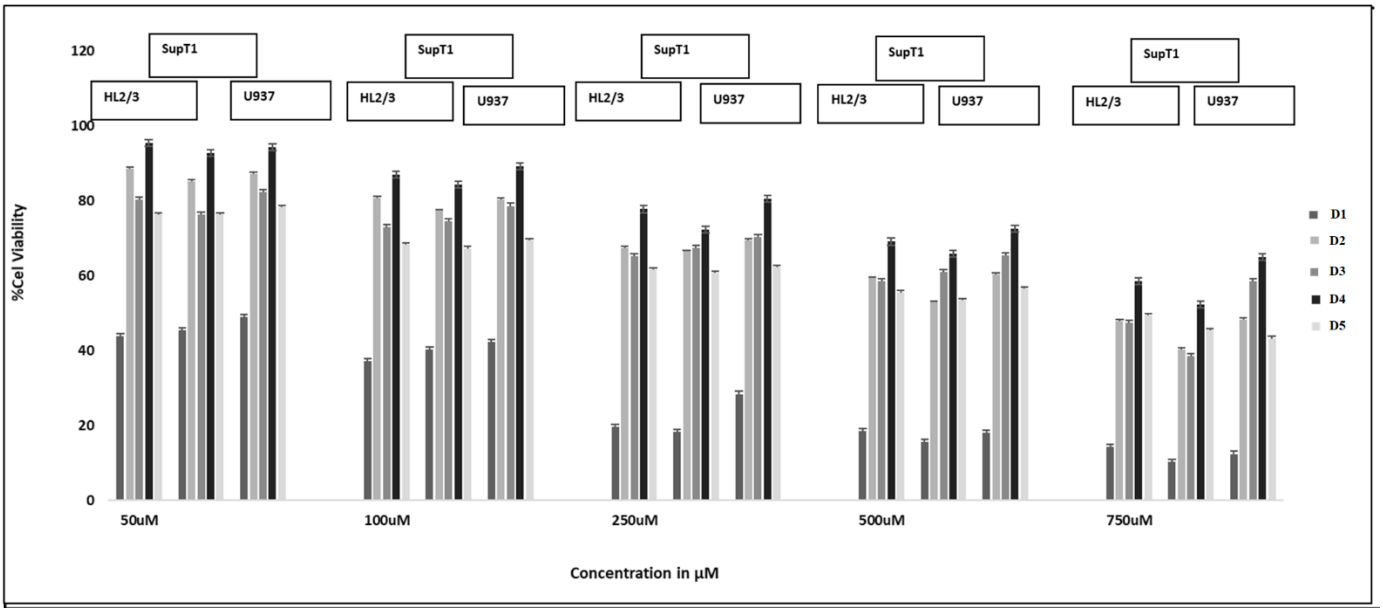


Fig. 29b) Interacting regions and residues of open loop-gp120 docked with compounds D1- D8

a)



b)

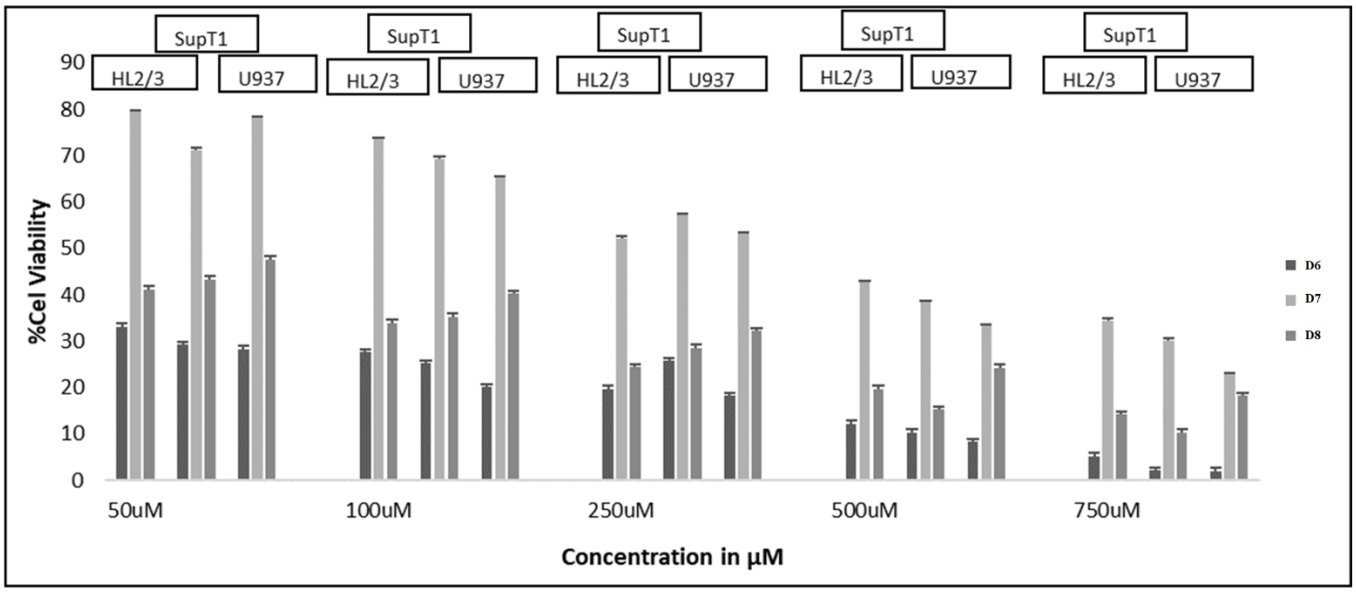


Fig. 30: Cytotoxicity of the compounds tested on cell lines SupT1, U937 and HL2/3 a) Cytotoxicity of compounds D1-D5 derived from 447-52D b) D6-D8 derived from EPAP-1. The toxicity was measured by MTT assay at increasing concentration of the compounds and the % cytotoxicity was plotted against the concentration. Mean±SD was plotted. P value was ≤ 0.05.

5. Cell Cytotoxicity Assay:

The toxicity of the designed compounds on different cell lines was evaluated by MTT assay as per the protocol described. The toxicity of the drugs was tested at 5 concentrations, 50, 100, 250, 500 and 750µM. In **Fig. 30a** depicting 447-52D mimetics, drugs D2-D5 showed lower cell death of <50% even at concentrations as high as 750µM. However, drug D1 showed high levels of toxicity ~60% at 50µM concentration. The least toxic drug D4 showed the highest cell survival of 65% at a concentration of 750µM. Followed by drug D4, drugs D2, D3, D5 showed cell survivals of 50%, 60% and 45% respectively at 750µM concentration. The three cell lines tested showed similar levels of sensitivity to all drugs. Drug D7 showed a ~100% survival at low concentration of 50µM .

In **Fig. 30b** depicting the EPAP-1 mimetics, drugs D6, D8 showed ~70% and ~60% cell death at low concentration of 50µM, respectively. Drug D7 showed low toxicity even at higher concentration of 500µM. The drug showed the lowest cell death of ~20% at 50µM concentration. Similar cell viability was observed in all cell lines. The drug D6 showed the highest toxicity, and the cell lines U937 and SupT1 showed similar sensitivity to the drug.

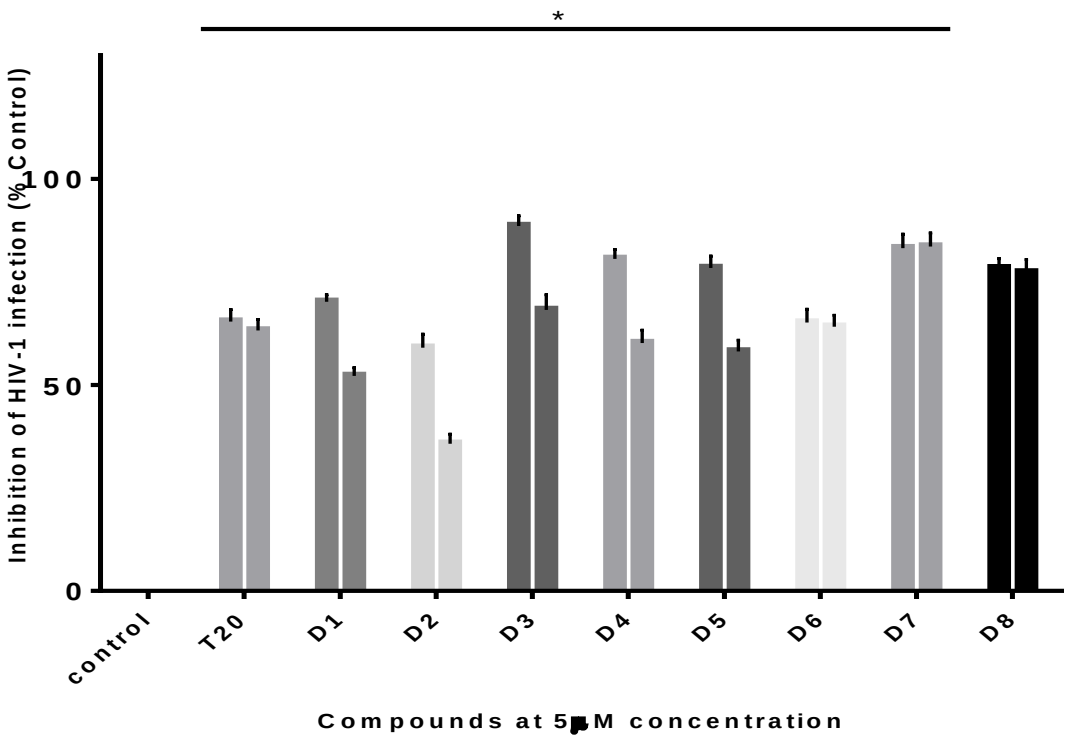


Fig. 31a): Anti HIV activity of compounds D1-D8 tested at 5µM concentration. The activity was tested with p24 assay and infection without inhibitors was considered as negative control and in presence of T20 was considered negative control. First bar of each compound is with pre-treated virus and second bar is con-infected virus. The % inhibition is plotted with SD and p value was ≤ 0.05

6. Anti-HIV activity

The compounds were tested for in-vitro HIV-1 inhibition in two modes. In the first mode(pre-incubation), the compounds were pre-incubated with virus before challenging the T-Cell line (SupT1) . In the second mode,(co-incubation) the compounds and the virus were added to the cells simultaneously. Pre-incubating the virus allows access to closed V3 loop conformation of gp120. Challenging the T-Cells with virus and compounds simultaneously allows for exposure of CD4 bound gp120 with open V3 loop conformation.(**Fig. 31a**)

The compounds developed from 447-52D, D1-D5 were tested at 5µM concentration in both pre-incubation and co-incubation experiments. In this set of compounds, the inhibition was higher when the virus was pre-incubated with the compounds and >15% difference in inhibition was observed in the two experiments as seen by p-24 protein quantification using ELISA assay. All the compounds showed >50% inhibition at 5µM concentration and compounds D3 and D4 showed high activity with ~80% reduction in p24 quantity. The activity difference observed in both experiments could be related to the better affinity of 447-52D mimetics to closed V3 loop conformation of gp120 that exists either in the cell-free virus or in the early stages of viral entry.

The compounds developed from EPAP-1, D6,D7 and D8 were tested at 5µM concentration in both pre-incubation and co-incubation experiments. The compounds showed similar levels of inhibition in both modes of infection as seen by p-24 protein quantification using ELISA assay. The compounds showed >60% inhibition at 5µM concentration, and compound D7 showed the highest activity with ~80% reduction in p24 quantity. Since similar activity was seen in both experiments, it could be assumed that HIV-1 inhibition by EPAP-1 mimetics was independent of the conformation of the V3 loop.

447-52D mimics	IC50 (NL4-3,Clade B)	IC50(93IN101,Clade C)
T-20	25nM	25nM
D1	2µM±0.27	3.2µM±0.32
D2	5µM±0.19	>5µM
D3	< 0.5µM	0.75µM
D4	1.6 µM±0.23	1.8 µM±0.3
D5	1.4 µM±0.2	2 µM±0.34
D6	>5 µM	3.2 µM±0.12
D7	1.2 µM±0.2	0.5 µM±0.1
D8	2.6 µM±0.15	1.5 µM±0.02

Tab. 12: HIV-1 Inhibition IC50 concentrations calculated for compounds D1-D8 .

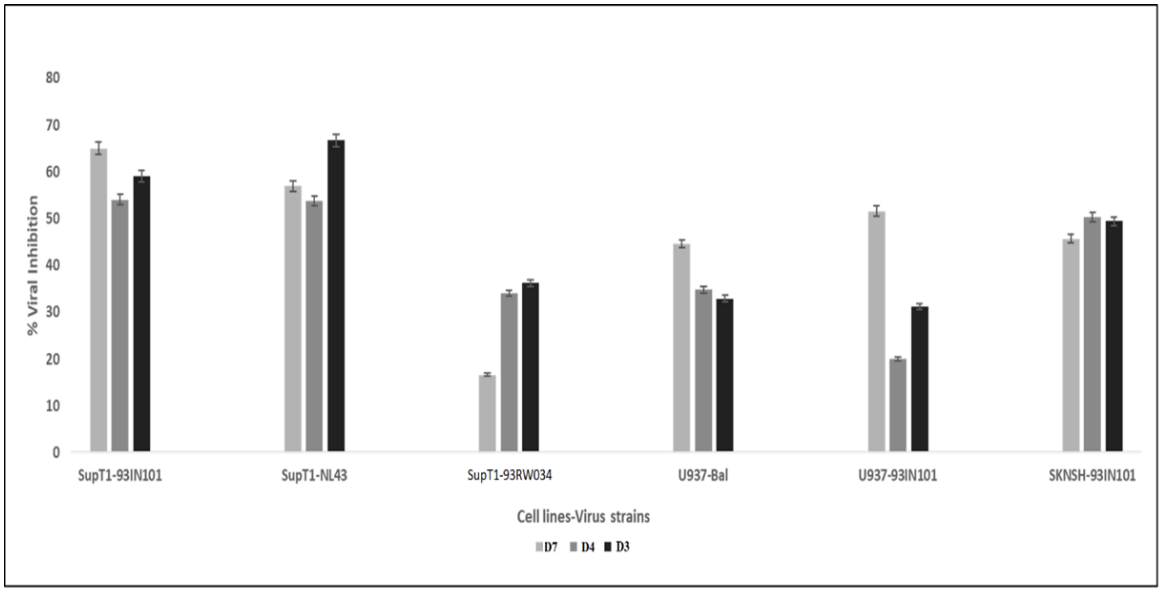


Fig. 31b): Broad spectrum activity of the compounds D3, D4 and D7 tested at 2µM concentration against Subtype A, B and C in cell lines SupT1, U937 and SKN-SH.

7. Dose dependent Viral inhibition:

The inhibitors were further tested in a dose-dependent manner in co-incubation for EPAP-1 based inhibitors, and 447-52D in preincubated experiments, respectively (**Tab. 12**).

The dose-dependent inhibition was tested by challenging T-Cells with Subtype B and Subtype C virus, NL4-3 and 93IN101, respectively in the presence of increasing concentrations of compounds. IC50 concentration of the compounds was calculated using standard formula.

Among compounds developed from 447-52D, the inhibition of the subtype B virus was higher than the inhibition of subtype C virus. The IC50 of the highest active compound, D3, against strain B virus was < 0.5uM and inhibition of C strain virus were seen at IC50 of 0.75uM. Compounds D4 and D5 showed inhibition of NL4-3 virus with an IC50 of 1.6uM and 1.4uM, respectively. The compounds showed an inhibition with IC50 of 2uM and 1.8uM respectively against clade C HIV-1 virus, 93IN101.

Compound D7 showed better activity in comparison to other EPAP-1 derived molecules. The inhibition of HIV-1 clade C virus was better with an IC50 of 0.5µM±0.1 than Clade B virus with an IC50 of 1.2uM±0.2. Similar activity was observed for compounds D6 and D8 with better activity in Clade C virus than in Clade B virus. Compound D8 showed inhibition of Subtype C strain with an IC50 of 1.5uM±0.02, while IC50 for inhibition of Subtype B strain was 2.6uM±0.15.

8. Broad spectrum activity of the compounds

Compounds, D3, D4 and D7 that showed better inhibition were selected for further in-vitro experiments. The broad range activity of the compounds was assessed, by testing the inhibition of compounds on HIV-1 subtypes A, B and C strains. The inhibition was also tested on R5 and R4 tropic viruses in T-Cell line and macrophages. SK-NSH neuronal cell line was challenged with HIV-1 in the presence of compounds to study the inhibition in CD4 absent cell line. The compounds were tested at 2µM concentration, in a pre-incubation experiment for compound D3, D4 and co-incubation experiment for compounds D7.

At 2µM concentration, in 447-52D derivatives, compounds D3 and D4 showed greater inhibition of X4 tropic laboratory-adapted virus NL4-3. ~50% and ~65% inhibition was seen in the presence of compounds D3 and D4 respectively. Inhibition of these compounds was

observed to be better than D7 in neuronal cells challenged with virus. Pro-monocytic cell line challenged with R5 tropic virus, Ba-L was considerably better inhibited than 93IN101 by these compounds. Subtype A virus 93RW024 was better inhibited by D3 and D4.

The compound D7, developed from EPAP-1 showed considerable inhibition in T-Cell as well as promonocytic cell lines challenged with isolates HIV-1 93IN101 and Ba-L virus. The compound was seen to specifically inhibit R5 tropic virus Ba-L and 93IN101 strains at ~45% and 65% respectively. Considerable activity was also noticed when neuronal cell line was challenged with Subtype C isolate (**Fig. 31b**).

9. Inhibition of Cell entry:

Inhibition of viral entry was tested by monitoring a cell-mediated fusion of gp160 expressing HL2/3 and SupT1 cells. The compounds D3, D4 and D7, were tested for inhibition by dye redistribution assay as described previously. The compounds were tested at respective IC80 concentrations with D7 in the co-incubation experiment and D3, D4 in preincubation experiment where the compounds were incubated with HL2/3 cells before co-culturing them with SupT1 at 1:1 ratio. Compound D3 and D7 showed better blocking of entry, as seen in **Fig. 32**, with fewer fused cells(cyan coloured cells) in comparison to control cells. Compound D4 showed considerable blocking of entry. Greater blocking of cell fusion of HL2/3 cells that express HIV-1 IIIB derived gp160, an X4 strain, by compound D3 is consistent with the observed anti-HIV-1 activity.

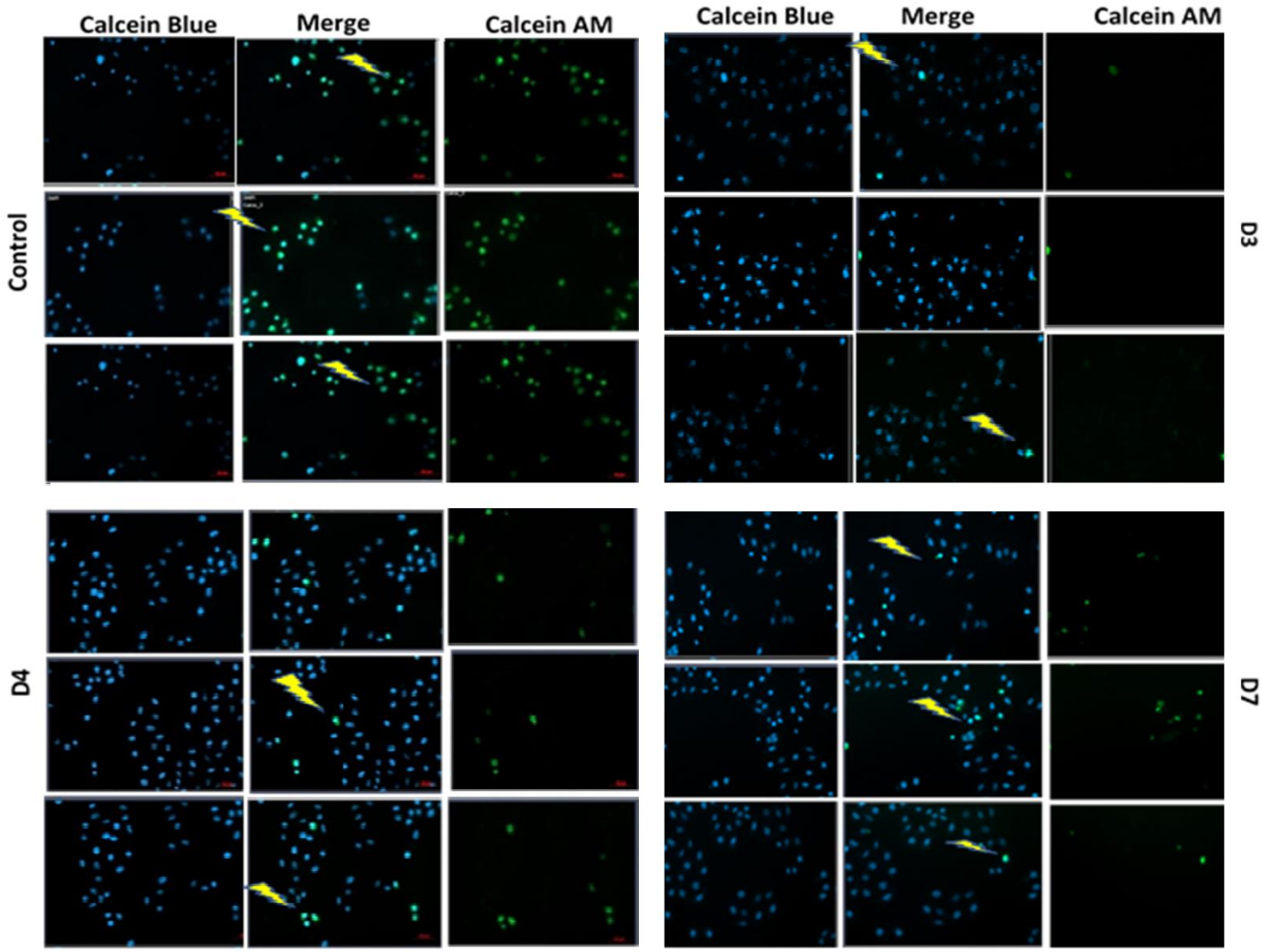


Fig. 32: Dye transfer assay to monitor inhibition of cell fusion in presence of compounds D3,D4 and D7. The images were captured by Leica Fluorescent Microscope as different regions of coverslip. Fusion in absence of compounds was considered as control. Fused cells exhibit cyan color and are indicated by yellow flash for visualization

Summary

Despite great efforts in the development of inhibitors targeting HIV-1 entry and fusion, an only limited number are in clinical use. A large number of promising molecules enter clinical trials periodically but fail to succeed. In an attempt to identify novel scaffolds and molecules that can inhibit this early stage of inhibition, two macromolecules with well-studied and established experimental proof of activity were taken as leads to guide the process of designing compounds that can inhibit HIV-1 infection.

V3 loop region of gp120 has been reported to play a role in HIV-1 Entry and also in deciding the tropism of the virus. Being highly exposed, V3 loop elicits a large number of neutralizing antibodies, but its high variability complicates it as a drug target. However, in between the high variability, the loop consists of a conserved region, the crown, which can be a considerable target. A broadly neutralizing antibody, 447-52D targeting V3 loop was identified to be active against many different laboratory strains of the virus. Similarly, a large number of natural molecules, including proteins, have been identified and reported, to be active against HIV-1. One such protein is EPAP-1, identified and characterized in our lab to possess high anti-HIV activity among different strains of the virus.

The challenges accompanied by the use of protein and peptides as therapeutics include metabolic instability, low bioavailability and reduced membrane permeability. Shorter half-lives of linear peptides obstruct the delivery of effective drug dosage to the target organ/tissue. Oral administration of peptide therapeutics is ineffective due to the break down by digestive enzymes and thus are more commonly administered through intravenous or intraperitoneal injections. These problems can be overcome by mimetic compounds developed from active peptides, designed to possess better properties. Peptidomimetics are compounds which mimic the pharmacophore of a peptide in the 3-dimensional space, retaining the interactions with its receptor and result in the same biological activity.

In this objective, compounds designed based on V3-Loop targeting antibody and protein are reported. A total of twenty compounds were rationally designed, and in-silico studies showed the interaction of compounds with the V3 loop and surrounding region. Eight compounds were chemically synthesized, validated and tested for activity in-vitro. Five compounds, of the eight were selected from the rationally designed compounds, with slight modifications. Three compounds, were modified by substituting groups that have been previously reported

to participate in interactions with V3 loop and groups that lower toxicity. The modified compounds showed better binding with the V3 loop in in-silico studies.

In In-vitro studies, the cytotoxicity of the compounds was evaluated in four different cell lines of varying origins. Compounds derived from 447-52D were observed to have a better cytotoxic profile in comparison to compounds from EPAP-1. Compounds D6, D8 of EPAP-1 origin and compound, D1 of 447-52D origin showed high toxicity with <50% cell survival at 50µM concentration. Compound D7 showed cell survival >50% at 250µM and compounds D2-D5 showed >60% survival even at high concentrations. To validate the ability of the compounds to inhibit viral replication, the p24 antigen was quantified using ELISA. The compounds were tested for viral inhibition in two methods, one by co-incubating them with receptor cells and virus. The other method involved pre-incubating the virus with the compounds for a period of 1 hrs before challenging the receptor cells. Compounds D6-D8 showed no influence of incubation conditions on activity, while compounds D1-D5 showed higher activity when pre-incubated with the virus. The viral inhibition was tested at 5µM concentration.

IC50 concentrations for viral inhibition in isolate 93IN101 and laboratory adapted strain NL4-3 showed highest inhibition for compounds D3 and D7 of 447-52D and EPAP-1 respectively followed by D2 and D5. Compounds D1 and D8 also showed considerable activity but displayed high cytotoxicity. Further studies were carried out to characterize the active compounds D3, D4 and D7. To understand the breadth of inhibition, different receptor cell lines and multiple viral strains were tested. Compound D7 showed better inhibition of CCR5 tropic strain Ba-L, while compounds D3 and D4 showed better inhibition of X4 tropic NL4-3. R5X4 tropic 93IN101 was consistently inhibited by all the three compounds. The results reflected the studies previously reported. The sensitivity of X4 tropic and R5X4 tropic viruses and resistance of R5 tropic virus to 447-52D neutralization was suggested in one study. Another study conducted to explore the conformation of V3 loop during antibody binding reported a change in conformation from closed to open loop upon binding with 447-52D suggesting preference of the antibody to bind to closed loop conformation. The blocking of viral entry was demonstrated by the inhibition of cell fusion via the dye redistribution assay. Minimal fusion of cells in presence of compounds D3,D4 and D7 was observed in the experiment with D3 showing highest inhibition of viral entry. The active compounds illustrated in this objective identify novel scaffolds that can be further developed to yield better molecules.

CHAPTER V

OBJECTIVE 3 : Insights into the binding pocket of HIV-1 Associated Topoisomerase II Beta Kinase using in-silico studies.

Introduction

Most of the signalling and cellular processes are controlled by phosphorylation. During Phosphorylation, protein kinases transfer ɣ-Phosphate of a nucleoside triphosphate to the substrate partner's amino acid residues (Cheng et al., 2011). A total of 518 protein kinases have been identified and well characterized in eukaryotes (Roskoski, 2015). These kinases are classified as typical and atypical based on their properties. Typical kinases are further classified as Tyrosine and Serine/Threonine kinases based on the phosphorylating residue (Miranda-saavedra et al., 2007). Atypical kinases are characterized by kinase specific activity but lack sequence similarity with existing eukaryotic kinases (Cheek et al., 2002). Structurally, the kinase family proteins are distinguished by the structures adopted in active and inactive states.

Viral kinases have been a source of interest since reports on kinase activity in virions have been published (Jacob et al., 2011). A major challenge associated with these kinases is the possibility of the origin being host cellular and hence require rigorous characterization before defining the source as viral. Sequence comparison and phylogenetic analysis of catalytic domains of viral and cellular kinases showed that the viral kinases branched as a separate family and did not project any cellular homologue (Leader, 1993).

The association of Topoisomerase II in HIV-1 infection has been well studied and documented. The binding and cleavage sites of Topoisomerase II in the genome (LTR-Region) of HIV-1 was studied and demonstrated by Howard and Griffith, thus suggesting a possible interaction between Topoisomerase II Beta and HIV-1 proviral DNA (Michael T.Howard et al., 1993). The presence of highly phosphorylated Topoisomerase isoforms during HIV-1 infection suggests their activation during infection (Kondapi et al., 2005). Later, it was demonstrated that HIV-1 replication is affected upon regulating Topoisomerase II, suggesting its role specifically in HIV-1 replication (Kondapi et al., 2006). Some reports suggest the association of isoforms with preintegration complex along with Vpr and reverse transcriptase. This association was also confirmed by the disruption of preintegration complex formation by Topoisomerase II Beta (TopoIIβ) poisons (Kondapi et al., 2006). The role of TopoIIβ in the transcription of viral genes was demonstrated to be through the recruitment of the enzyme to HIV-1 LTR for the Tat-dependent transactivation (Anil Chekuri et al., 2016).

Assays performed using purified virus lysates showed its ability to phosphorylate both Topo II α and β isoforms. Analysis of isopycnic glycerol gradients of the viral lysate showed the presence of two distinct kinases that can phosphorylate α and β independently. A lower density fraction showed the ability to phosphorylate TopoIIα and TopoIIβ phosphorylation activity was seen in a higher density fraction (Kondapi et al., 2005) .

Analysis of the phosphorylating activity of the fractions in presence of pan kinase inhibitors, staurosporine and PD98059 showed a decrease in activity of the α phosphorylating fraction, suggesting the kinase in the fraction has a common catalytic domain like that of Protein Kinase C and MAPK, which have been previously reported to be associated with the virions (Kondapi et al., 2005). Apart from these kinases, others like ERK-2, cellular protein kinase-A, NDR-1, NDR-2 have been reported to be virion-associated (Giroud et al., 2011), and hence like other retroviruses, the TopoII α kinase activity of the virion fraction could be of a cellular kinase packaged during budding of the virus. However, the β phosphorylating activity of the second fraction demonstrated very low sensitivity to the inhibitors tested, suggesting a different catalytic domain and thus a different kinase. The kinase from both fractions was purified and characterized. The kinase with the ability to phosphorylate TopoIIβ was labelled HIV-1 associated Topoisomerase II Beta Kinase(TopoIIβKHIV-1) (Kondapi et al., 2005).

Apart from the two pan-kinase inhibitors, experiments to understand the ability of TopoIIβKHIV-1 to phosphorylate the β isoform was conducted in the presence of 20 inhibitors targeting different Serine/Threonine kinases(STK) and Tyrosine Kinases(TK). The assay revealed high TopoIIβKHIV-1 activity (no inhibition) in the presence of TK inhibitors. However, in the presence of STK inhibitors showed varied inhibition of activity, suggesting the purified TopoIIβKHIV-1 could be an STK and also a novel catalytic site (Ponraj et al., 2013).

Phosphorylation experiments conducted in the presence of different classes of organic molecules showed significant inhibition in the presence of pyridine derivatives and demonstrated low cytotoxicity. The tested derivatives were used to construct a 3D-QSAR model, to assist in optimizing and designing compounds (Ponraj et al., 2013). Based on the contours obtained by the study, further compounds were designed and synthesized. In the process of designing and development the compounds based on the QSAR studies, Coumarin scaffold was identified as potent inhibitors of the. Coumarin compounds have been associated with broad biological

activities, and the coumarin ring system is found in pharmacologically active compounds, especially anti-tumour and anti-bacterial compounds (Kammari et al., 2017).

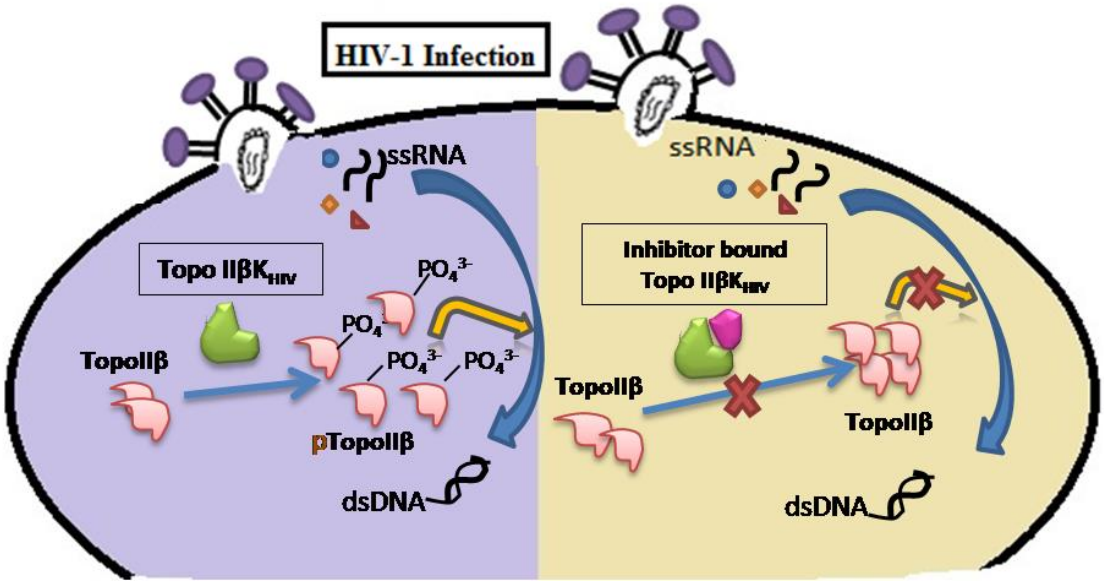
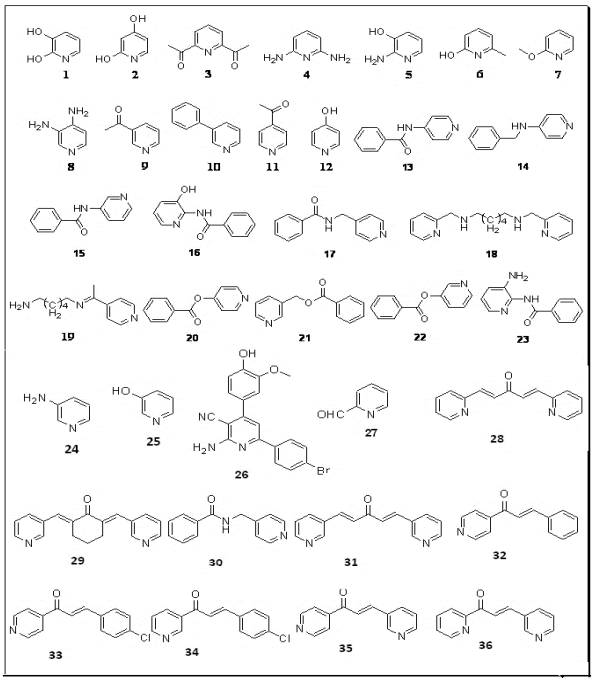


Fig.33: Possible mode of action of TopoIIBK$_{HIV}$ Inhibitors in controlling HIV-1 infection.

In this objective, the QSAR studies on the inhibitors is described and based on the studies and the pharmacophores, we tried to explore the properties of the binding pocket of the novel Kinase.

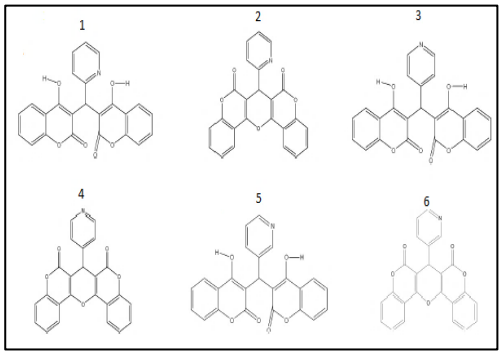
Data Set

Training Set

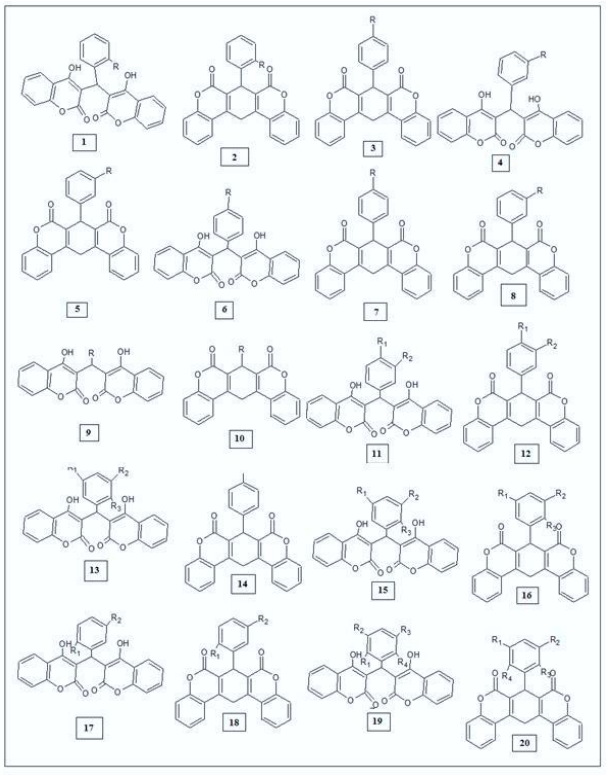


Pyridine Derivatives

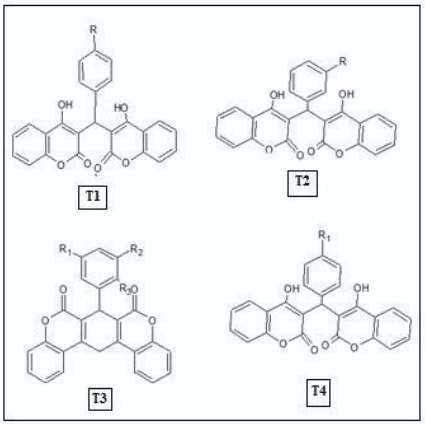
Test Set



Coumarin Derivatives



Training Set



Test Set

Compound. No	R1	R2	R3	R4
T1	NO2	-	-	-
T2	Cl	-	-	-
T3	Br	CH3	OH	
T4	OCH3	-	-	-

Methods:

1. Dataset:

As described previously, the dataset comprised of pyridine and coumarin derivatives which were used to develop two QSAR models respectively.

1.1 For Model 1, a QSAR equation previously developed from a small set of pyridine derivatives was further refined by taking a total of 36 molecules(23+13). These 36 pyridine derivatives were synthesized, characterised and tested for TopoIIBK$_{HIV}$ activity and used for training the model. Six compounds designed based on Electrostatic and Steric contours obtained from the initial 3D QSAR model were used as a test set to validate the developed model.

1.2 A higher activity observed in the coumarin substituted pyridine derivatives, further prompted the design and development of dicoumarol derivatives. A set of 20 dicoumarol compounds used as a training set for the Model 2 and four compounds were used as a test set to validate the model.

The same protocol was used to develop both Model1 and Model2.

2. Preparation of Molecules for QSAR Analysis

Sketching and preliminary cleaning of the molecular structures was performed using Chemsketch (ACD Labs). Energy optimization of the sketched structures was performed in Sybyl by initializing partial atomic charges with Gastieger Huckel and using Tripos forcefield with a convergence criterion of 0.005 kcal/(Å mol). All the molecules were optimized using the same method.

The activity of compounds depends on positioning and conformation of the compounds in the receptor active site. Before developing a model, alignment was performed by positioning the molecules over one another by taking the high active molecule as a template to depict the interacting conformation. The alignment is usually performed with reference to a common core which defines an index to monitor the changes in orientation during the alignment. For Model1 pyridine ring was considered a common core and coumarin ring was considered for the Model2. 10 from dataset one and compound 19 from dataset two were used as a template for aligning the compounds.

3. CoMFA

For the first time, Comparative Molecular Field Analysis (CoMFA), a method proposed by Cramer and Milne, associated the 3D structures of the molecules to their activities. The method follows an underlying assumption that similar pharmacophoric side chains of molecules result in similar orientation and hence, similar interaction with the receptor.

The alignment of the molecules obtained previously was enclosed in a 3-D grid box of 2.0 Å dimension, for calculation of interaction energy. The Van der Waals potential and Coulombic and interactions were calculated using Tripos forcefield at each intersection of the lattice by using a probe of +1.0 charge, usually an sp3 carbon. Electrostatic interactions were calculated using a 1.0r dielectric constant. Cut-offs were set for both fields to normalize the values, +30.0 kcal/mol was used as a cut-off to truncate the steric field at lattice points where the value exceeds it. Electrostatic fields were ignored at points where a high steric field was calculated.

4. Regression Analysis

Partial Least Square Regression (PLS) algorithm was employed to correlate the biological activity of the molecules with the interaction energies previously calculated. The activity of the molecules represented as the negative logarithm of IC50 values (pIC50) are used for developing the regression model. The Optimum Number of Components (ONC) required for the run was set at a default of eight for a Leave One Out (LOO) cross-validated PLS. Of the models generated, the one with higher q2 and lower standard error was selected, and the reported ONC was noted. The model generated after cross-validation was subjected to another run of PLS without validation to obtain r2 (conventional). Equal weightage was given to both steric and electrostatic fields during regression analysis, and a cut-off of (σ) 2.0KCal/mol was set to filter columns to reduce noise. The robustness of the model was evaluated by a 100-run bootstrapping analysis was performed. The bootstrap analysis randomly selects samples from the original dataset and generates new datasets to perform the statistical calculations for each of the new datasets. The bias in the original calculation is measured by obtaining the difference between calculations of the original, and the average of the bootstrapped datasets.The external validation of the model was performed by using a test set of 6 molecules for model 1, and a set of 4 molecules for model 2. The accuracy of prediction was plotted between the predicted biological activity obtained from the CoMFA equation and the experimental values.

5. Contour Map

Contour maps are a figurative depiction of favoured and disfavored substitutions on the common core of the derivatives (pyridine and dicoumarol rings). The levels for both steric and electrostatic fields are fixed at 80% and 20% for favoured and disfavored respectively. Specific colour coding for electrostatic and steric allows for better understanding of the maps. Green and Yellow depict favoured and disfavoured regions for substitution of steric groups. Similarly, Blue and Red depict the favoured electropositive and electronegative substitutions. These maps are a product of standard deviation and coefficients. (StDev $*$ Coeff)

6. Effect of charge on binding of compounds

6.1 Peptide Sequence and structure:

Peptide sequences of 25 amino acid length were generated with residues selected at random by RANDSEQ with user specified amino acid content. The peptide sequences were generated with increasing positive charge by varying the number of charged residues incorporated. The residues majorly altered included, R, K, H. Peptide sequences generated without specifying charge were selected as control. The total charge, the percentage of positive charge in the peptide and the stability of the peptide based on the instability index was calculated using Expasy ProtParam. The tool calculates physical and chemical parameters of a user-specified protein/peptide sequence. A total of 9 peptides were chosen based on the stability, charge and content of the peptide sequence.

The selected sequences were modelled using the server, PEPFOLD 3.0 to obtain secondary structures of the peptides. The algorithm employed, predicts the structures by a de-novo approach based on a Hidden Markov model sampling approach. The generated structures were validated using RAMPAGE. Validated structures were taken forward for docking with the compounds.

6.2 Molecular Docking with compounds

Autodock Vina v1.2 was used to dock the highest active compound in both pyridine (10) and coumarin analogues (19) with the modelled peptides. Configuration files to run the docking for each peptide were prepared separately by placing a grid box on the peptide and defining the boundaries for binding region search. The grid was set at 29.27,27.02,20.17 X, Y, Z coordinates and the dimensions of the box was fixed to 10*18*16. Each docking run was iterated 10 times with each ligand and the pose and Binding energy that is repeated the most times was considered for analysis. The iteration runs were automated by Bash scripts.

7. Binding with other kinases

7.1 Kinase selection

A representative kinase from every class of Ser/Thr Kinase family were selected to study the affinity of high active compound to the kinases. Casein Kinase1 (2CSN), WNK-1 (3FPQ), STK11 (2J7T), IRAK4 (2NRY), Protein Kinase B (1O6L), DAPK-1 (1IG1), Caesin Kinase-II (3FWQ) were considered for the study based on the availability of good quality X-Ray crystal structure.

7.2 Molecular Docking

Molecular docking to the ATP binding pocket and allosteric binding pockets of the kinases was performed with the high active compound of coumarin derivative. The kinase crystal structures were pre-processed to remove water molecules and existing ligands. The grid box for binding was set up based on the binding site of the receptors as per literature. Autodock Vina was used for docking the peptides to the compounds. Each run was Bootstrapped ten times. The peptides were pre-processed and were converted to the autodock readable format. The binding was tested with high active compound (Coumarin 19).

8. TopoIIBK$_{HIV}$ peptide from MALDI Analysis

The peptide sequences were modelled using the server, PEPFOLD 3.0 to obtain secondary structure of the peptides. The algorithm employed, predicts the structures by a de-novo approach based on a Hidden Markov model sampling approach. The generated structure was validated using RAMPAGE. The validated structure was docked with the high active compounds 10 and 19 using AUTODOCK Vina.

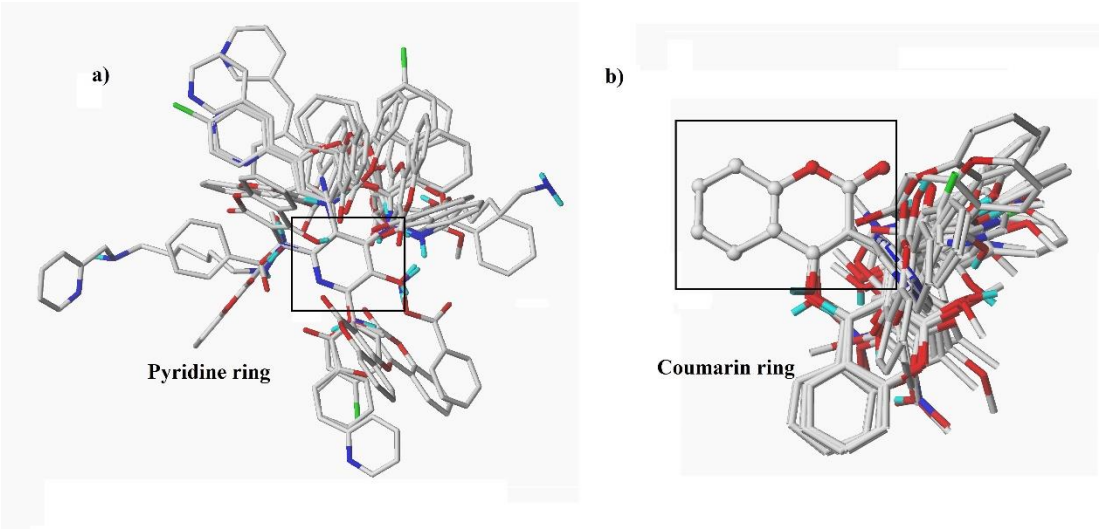
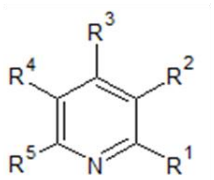
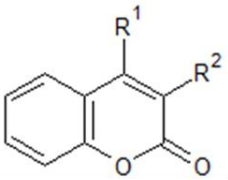


Fig. 34: Superimposition of Analogs on the common structural framework of a)Pyridine ring b)Coumarin ring



Pyridine Analogs



Coumarin Analogs

PLS Statistics	Pyridine analogs	Coumarin analogs
q^2 (LOO) c-v predictive power of model	0.633	0.743
r^2 (correlation coefficient squared)	0.997	0.978
Standard Error of Estimate (SEE)	0.160	0.091
F values	316.65	139
Prob.of r^2=0	0	0
N (optimum number of components)	7	4
Relative Contributions		
CoMFA (1782 vars) (Steric)	0.393	0.330
CoMFA(1782vars)(Electrostatic)	0.607	0.670

Tab. 13: PLS Statistics of the models developed from Pyridine and Coumarin Analogues. The cross validated r^2 (q^2) for both the models was reported to be > 0.6. The standard error estimates were <0.2 and r^2 value observed was > 0.9 for both models. These statistics validate the model as acceptable

Results and Discussion

1. 3D QSAR model

1.1 Model Quality and statistics:

Validation is a major requisite for developing a useful QSAR model. It is important for the predictions made by the QSAR models to be accurate and the predictivity be well validated. The model is built and analysed based on a dataset divided into a training and testing set. The model is trained on the dataset known as the training set and an internal assessment done on this set is known as internal validation. The data set used for external evaluation of the model is known as the test set, and the trained model is unaware of the data points.

Internal validation by Leave One Out cross-validation is done by building a QSAR model on a subset of the training set, by leaving out one data each time and iterating until every data point is left out at least once. This method is employed in situations where the number of samples in the data set is limited. Cross-validated r2 (q2), calculated in internal validation, is used to evaluate the predictive power of an equation. Following the internal validation, the external validation is performed on the test set molecules by comparing the values obtained by the model prediction and their experimental responses. Test data prediction enables assessing the model's ability to predict the activity of unknown compounds with accuracy. To statistically represent the accuracy of the model in external validation, the commonly used measures include, coefficient of determination (r2). The overall r2 is usually observed to be higher than the cross-validated r2(q2). The acceptable statistical metrics for an accurate QSAR model have been specified as:

$q2>0.5$

$r2>0.6$

Another parameter considered for assessing the QSAR model is the Root mean squared error (RMS) that quantifies the relative error in prediction.

a) Pyridine Derivatives

The alignment of the pyridine derivatives is given in **Fig.34a**. A cross-validated r2 (q2) value of 0.633 with seven principal components and a non-cross-validated r2 value of 0.997 was obtained for the model derived from pyridine derivatives. After refining the dataset to remove outliers, a set of 26 compounds constituted the training set. The r2 (q2) value obtained post-internal validation by Leave One Out(LOO) method is good to accept the model. The standard error estimate (SEE) of 0.160 and F values of 316.65 was obtained for the model. Further, for internal validation, a Bootstrap analysis of 100 runs was performed. The statistics are given in **Tab. 13**. The accuracy of the prediction of the model was evaluated with a test set of 6 molecules. A graph was plotted with the predicted-activity data against the experimental response values. The error of prediction of the model is seen to be less than 1 log value (<10 Units) for all molecules.

b) Coumarin Analogs

The alignment of the coumarin derivatives is given in **Fig.34b**. A cross-validated r2 (q2) value of 0.743 with seven principal components and a non-cross-validated r2 value of 0.978 was obtained for the model derived from Coumarin derivatives. After refining the dataset to remove outliers, a set of 20 compounds constituted the training set. Like the pyridine value, the r2 (q2) value obtained post-internal validation by Leave One Out(LOO) method is good to accept the model. The standard error estimate (SEE) of 0.091 and F values of 139 were obtained for the model. Further, for internal validation a Bootstrap analysis of 100 runs was performed. The statistics are given in the **Tab. 13**. The accuracy of the prediction of the model was evaluated with the test set of 4 molecules. A graph was plotted with the predicted-activity data against the experimental data for both the data set molecules. The error of prediction of the model is seen to be less than 1 log value (<10 Units) for all molecules.

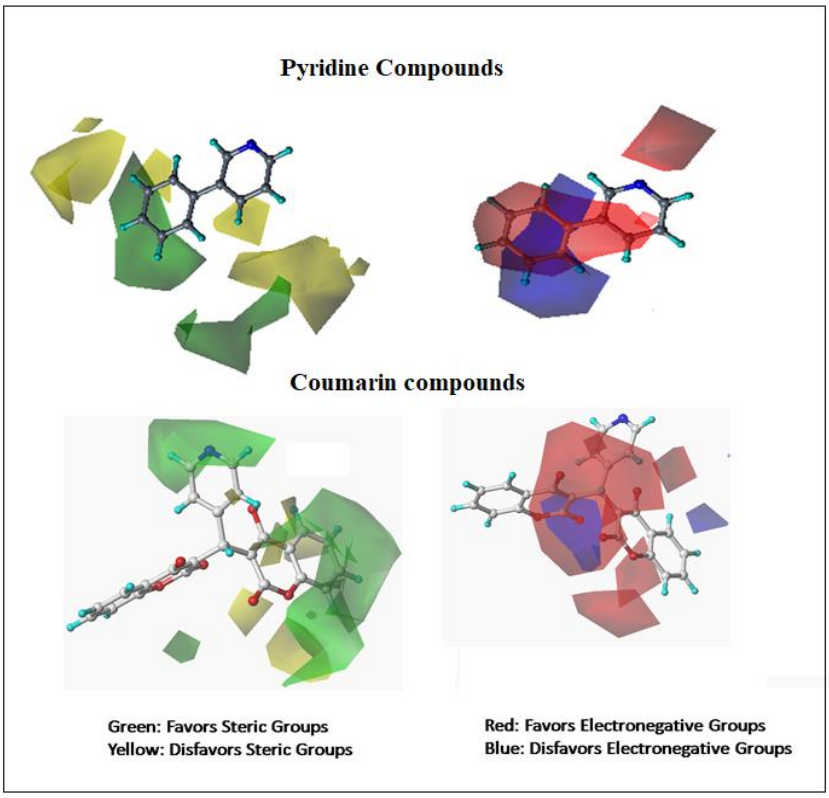


Fig. 35: CoMFA StDev * Coeff contour plots generated based on the developed 3D-QSAR equation green & yellow suggesting steric substitutions, blue and red suggesting electrostatic substitutions . The bulk dicoumarol ring of the compound in panel b) is seen overlapped with the green polyhedron indicating its role as a bulk group enhancing the activity of the compound. The overlap of red polyhedron and the negatively charged enolic Oxygen group points to the role of electronegative groups in the enhancement of TopoIIβK_{HIV} antagonism.

1.2 Contour Maps

Along with the prediction model described previously, the interaction fields at each grid point is calculated by CoMFA to generate an equation depicting the contributions of the field at each point. The contributions are usually depicted as contour maps, for easy visualization. These maps guide the understanding of functional groups and positions contributing to the activity of the molecules and depict specific regions in the molecular space important for structural modifications that significantly alter the activity. They also guide in exploring the binding pocket of the receptor(ccr5). At each region depicted, both positive and negative contours are shown suggesting favourable and less favourable substitution affecting the activity of the designed molecule. CoMFA calculates steric and electrostatic contributions, and the contours depict favourable and less favourable steric(green, yellow) and electrostatic substitutions (red, blue).

The steric and electrostatic contributions corresponding to the structural activity of the Pyridine and Coumarin compounds were studied based on the contours generated through CoMFA. The model developed using pyridine derivatives suggested an increased activity of the compounds upon addition of Steric groups in the third and fourth positions. Substitution of Steric groups aligning them between second and third positions was disfavored as it would result in reduced binding affinity. Similarly, the addition of electropositive group at the fourth position such that it aligns between third and fourth positions increases the activity. The model shows a distinct electronegative group preference at positions around the heterocyclic N and mostly at third and fourth positions. For, the model developed using coumarin analogues; the second position showed an increased preference for steric groups and positions second and third groups showed a preference for electronegative groups. The best active molecules have been overlayed with the contours and depicted in the **Fig. 35**.

On comparing the features of both the models, the role of steric groups and electronegative groups in defining the interactions of the compounds with the residues in the binding pocket of TopoIIβKHIV is very prominent. The region on one side of the scaffold of both models is available for substitutions of bulk groups. A distinct preference for smaller group substitutions on the other side was observed, which might suggest the presence of steric hindrances in the pocket resulting from the side chains of bulky amino acids. In the case of electrostatic substitution, an increased preference for electronegative group substitution was observed. This

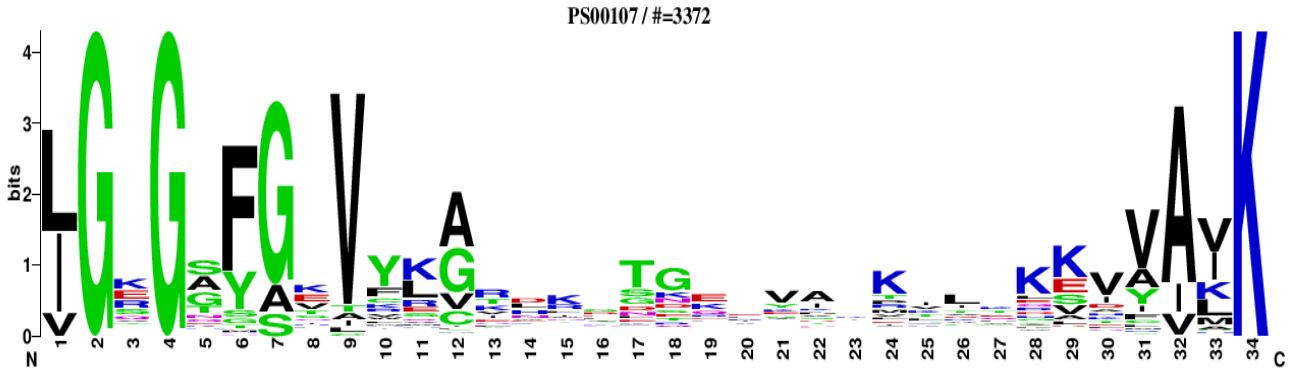


Fig. 36: Conserved signature of Kinase ATP binding domain obtained from PROSITE. The signature shows a well conserved Lysine residue and Glycine residues. Valine is also seen to be conserved in the signature. The constitution of the site does not imply a dominance of charge.

observation was supported by a trend in the dataset with high active molecules consisting of negatively charged and high electronegative groups. This points to a probability of positively charged amino acids, Arg, His, Lys in the binding site.

Given the above observations, the conserved signature of the Kinase binding domain retrieved from Prosite (Prosite ID: PS00107) (**Fig:36**) did not show the presence of dominant positive charge except for a conserved Lysine, K. This Prosite signature was derived from 3372 Kinases from Eukaryotic, Bacterial and viral origin. However, the profile showed dominant hydrophobic residues in the site. Interestingly, a study to trace the genomic origin of the TopoIIBKHIV using in-silico prediction to identify a Kinase domain in the translated HIV genome using PROSITE resulted in a negative response.

According to a study that summarizes the ligand interactions in 220 kinase complexes obtained from PDB crystal structures showed only 3% of the ligand set contained negatively charged groups, thus being very uncommon. While groups such as tertiary amines that are positively charged are common with approximately 37% of compounds being positively charged in the dataset analysed. Comparing the contour maps of both studies, reveals a significant difference in binding sites of TopoIIBKHIV and other kinases.

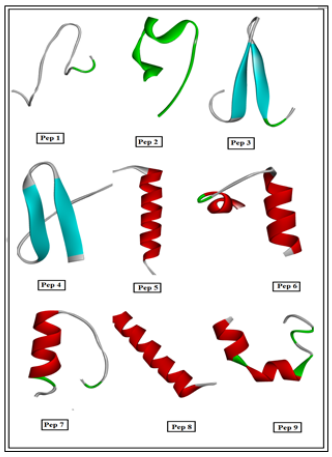
Assuming, the binding pocket to be composed of the predicted positive charge, molecular docking was employed to understand the effect of charge on interactions.

2. Effect of charge on binding of compounds

The hypothesis of the distinct charge based environment in the binding pocket of TopoIIBKHIV was further explored by the binding of the highly active compound to the positively charged peptides. Nine peptides of twenty-five amino acid length with varying charge were used to understand the affinity of the compound. Positively charged amino acids, Histidine, Lysine and Arginine were incorporated into the peptides at 0,4,12,20,24 and 28% of the total length of the peptide. Randomly generated peptides were used as control. The validated structures of the peptides were used for the study (**Fig. 37a and b**).

The method of binding a ligand to multiple protein/peptide targets has been widely used to identify off-targets and also for lead optimization. This method, reverse docking, was developed in the last few years and also has the application of predicting the binding site of protein targets

a)



b)

Peptide	Peptide sequence	Positively charged aminoacids(%)	Total charge of the protein	Binding Energy(Kcal /mol)
Pep 1	-FTEDQILLTIS GDCCYVAMGSNVWP-	0%	0.8	-4.6
Pep 2	-PDLSFLCWGQ TMMTGNVSEYACVIA-	0%	-2.3	-5.4
Pep 3	-PAMWNSTDE TFGDKVIYMLSVQIGC-	4%	-2.3	-5.3
Pep 4	-AHGMFCKPWE VTSIMILKYSGQNDR-	12%	0.8	-5.1
Pep 5	-RDSTCVRGCM FNVIEAHRYQKWPLK-	20%	2.8	-6.0
Pep 6	-KQTLKNSIRM WFGPCVCRYRHKAAL-	24%	5.8	-6.8
Pep 7	-EKSGRKRGFP VARMCKAWIQCYTLN-	24%	4.6	-6.4
Pep 8	-KYCIKQCFRW RKTGRELSKNAAGPM-	28%	5.6	-6.0
Pep 9	-MYRSKQKFKPL GETIGCRARKNAWC-	28%	5.6	-6.4

c)

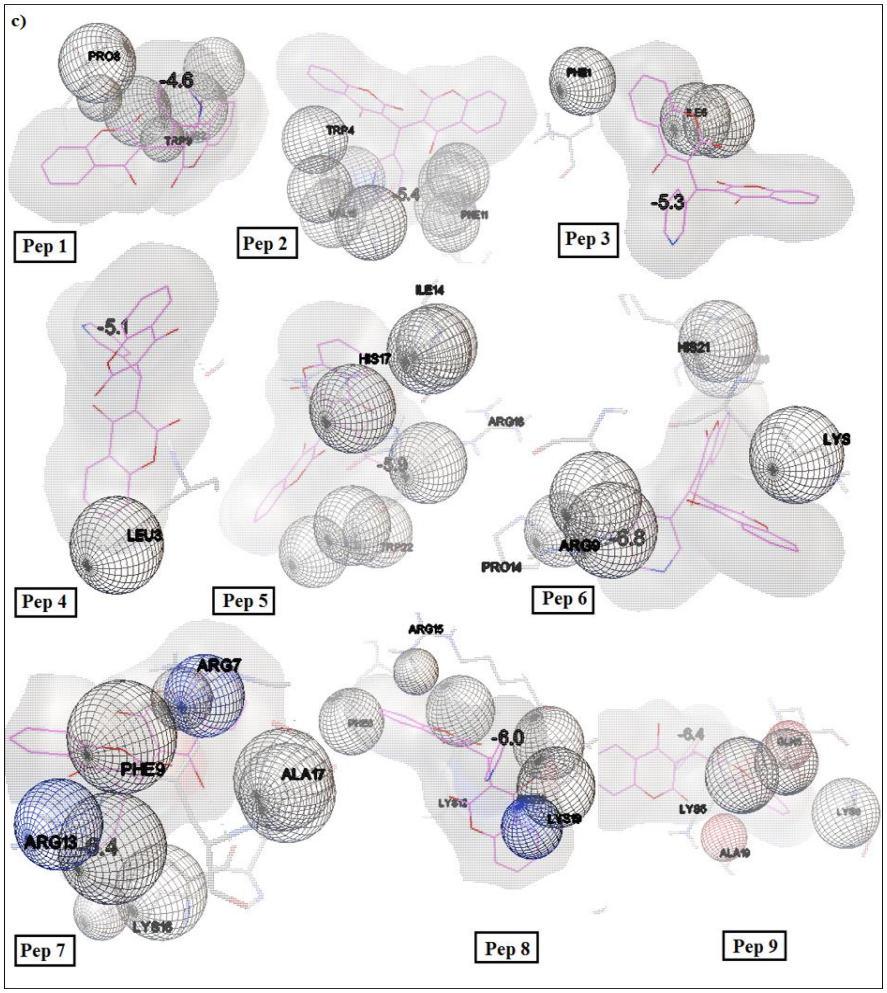


Fig. 37: a) Randomly designed peptides with varying charges. b) Table containing designed peptides with the % of positive amino acid residues, total charge and the binding energy obtained upon docking with the active molecule c) Interactions of the active compound with the designed peptides. As the positive charge residues in the peptides increase, the interacting region is concentrated by residues R, K and H. The electrostatic interactions and increased positive charge appear to be resulting in a better AUTODOCK energy score. Apart from charge residues, amino acids with bulk chains (F, A) are also observed in the interacting region.

(Lee et al., 2016). A similar approach was used to bind the active compound to the designed and modelled peptides. Autodock Vina has been used and reported by various research groups to be efficient in reverse docking (Xu et al., 2018). The sequence of the peptides, the positive charge content and overall charge of the peptides are given in the table. The structures of the peptides are presented in Fig A and Fig B presents the docked interface between the compound AKKM-2 and the peptides. The role of electrostatic forces of the binding pocket in interaction with charged ligands has been well established in complexes such as superoxide dismutase (SOD) with its positively charged residues in the active site attracts the ligand, superoxide which is negatively charged (Klapper et al., 1986). The enzyme triose-phosphate isomerase (TIM) moves its substrates towards the active site, especially the positively charged regions (Wade et al., 1998). The docking results showed prominent interactions of the compound with the positively charges amino acids in the peptides with better Binding Affinity scores reported by Autodock Vina. The affinity was observed to be low for peptides with an overall negative charge despite the presence of positive amino acids. The affinity in KCal/mol decreased as the number of positively charged amino acids in the peptides increased. Peptides with 24 and 28% positive charged amino acids showed better binding affinity than the others (**Fig 37c**). These results strengthen the basis of the hypothesis, suggesting a strong positively charged environment in the binding site of the novel TopoIIBKHIV.

3. Binding studies with other Kinases

Protein Kinases are majorly classified based on their substrate specificity and the specific amino acid sequence of the binding and catalytic domains. The former classification method is based on the amino acid of the substrate, phosphorylated by the kinase and is mainly divied as Tyrosine Kinases(TK) and Serine-Threonine Kinases(STK). As previously mentioned, the TopoIIBK$_{HIV}$ has been observed to selectively phosphorylate Serine-Threonine residues(Kanna) thus classifying it as a STK. STKs have been observed to possess a glycine rich catalytic core with a central conserved Aspartic acid residue necessary for its catalytic activity. Similarly, another method of classification based on the sequence of the catalytic domains divides the kinases into seven main families AGC family: The members of this family include protein kinases A, G and C; CAMK family: The protein Ca2+/CAM-dependent PK is a member of this family CAMK;

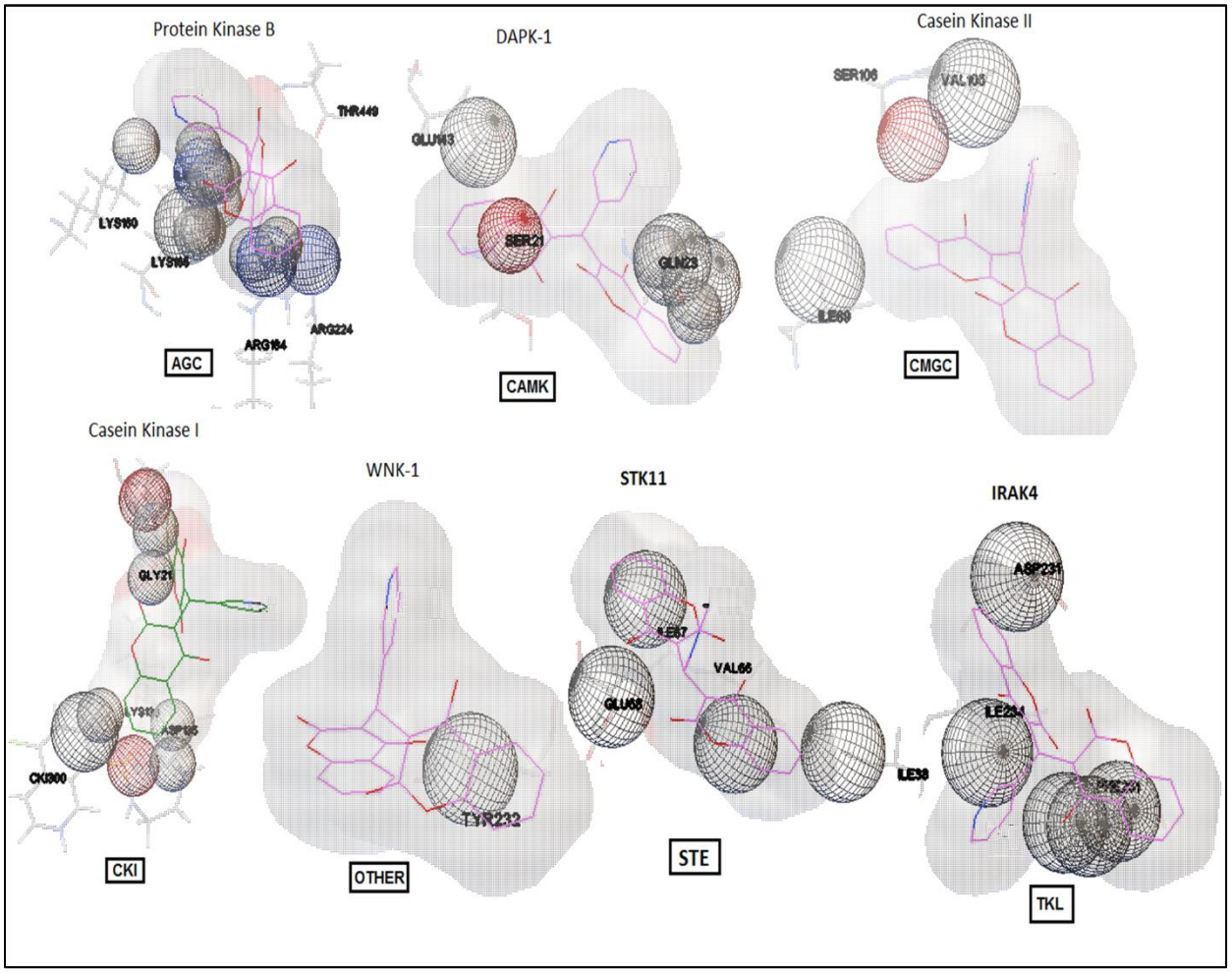


Fig. 38: Interactions of the active compounds with members of different Kinases. From the interaction of the compound with the Kinases, it could be noted that the pockets lack a dominant positive charge-based environment with the exception of AGC group representative.

CK1 family: consisting of the protein casein kinase 1; CMGC family: Includes CDK, MAPK, GSK3, CLKs; STE family :The commonly observed members are homologues of yeast sterile 7, 11, 20 kinases; TKs : the group of tyrosine kinases belong to this family, TKL: tyrosine kinase-like PKs fall under this family. Crystal structures of a large number of PKs have been solved in an attempt to obtain a better insight into the mechanisms of action of the enzymes. The ATP binding site of the kinases is conserved with similar over-all structure of the site. However, minor modifications confer specificity to the inhibitors that bind the site, thus affecting only one class and not the others. To understand the possible binding mode of the compounds with other kinases, a representative from the seven classes of Kinase family were selected. To study the affinity of high active compound to the kinases, molecular docking to the ATP binding pocket and allosteric binding pockets of the kinases was performed with the compounds. The results depicted in Fig, show bound conformation of the compound in the ATP binding pocket of the kinases. The pockets lack the presence of positively charged amino acids which is assumed to be important for better binding of the compounds. The affinity of the compound as seen by the AUTODOCK energy function was low. The lowest being STK11 of STE family and the highest being Casein Kinase 1 of CK1 family.(**Fig.38**) The presence of R and K in the pocket of AGC Family kinases, might suggest a phylogenetic link between TopoIIBK$_{HIV}$ and the family proteins. Despite the dominnat positive charge, the affinity of the compound was observed to be low with the member of this family, possibly explaining the lack of inhibition of the TopoIIBK$_{HIV}$ activity by PKA inhibitor H8 and PKC inhibitor Rottlerin.

4. MALDI derived TopoIIBK$_{HIV}$ peptide

According to the CoMFA model proposed, one side of the common scaffold of the compounds disfavours the substitution of steric bulk while the other side favours the substitution. The latter suggests the probability of the positions interacting with the Kinase. This was also seen in the model previously reported. The involvement of amide hydrogen and carbonyl oxygens of the Kinase region hinge peptide in interaction with the pyridine scaffold of the compounds in a manner similar to the adenine ring of ATP. This suggests that the inhibitors might occupy ATP binding pocket of the TopoIIBK$_{HIV}$.

Peptide	Peptide Sequence	Positively charged amino acids	Total charge of the protein	Binding Energy(Kcal/mol)	
1	-NITNNAKTIVHLNKSVEINCTRPFN NTRTSIRIGPGQVFYRTKWNETLKQ VARKLKGEFFYCNTTKLFN-	17%	8.7	-6.5	-7.7

Tab. 14: TopoIIBK$_{HIV}$ peptide derived from MALDI analysis. The total charge of the protein, number of positive residues and the Binding energies of compounds 10 and 19 with the peptide are given.

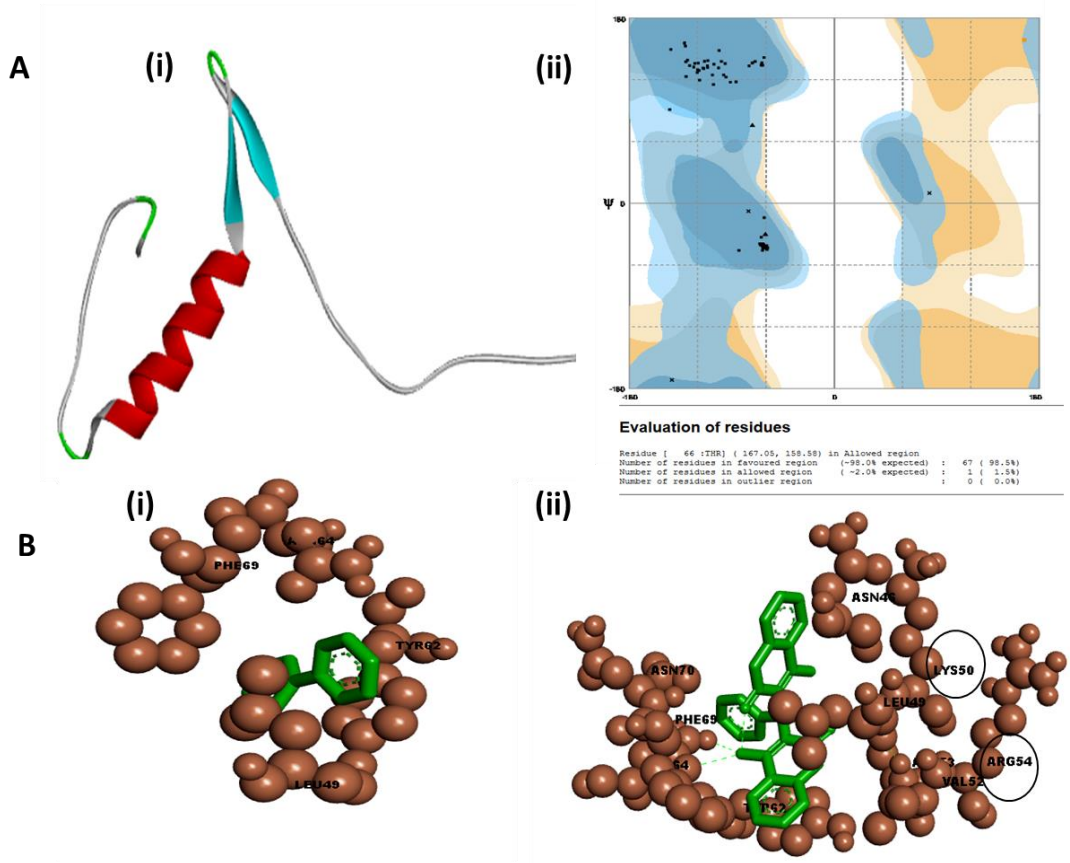


Fig 39: A) i). Peptide obtained upon MALDI TOFF analysis of the 72KDa TopoIIBKHIV ii).Fitness of the model as validated by RAMPAGE. B) i). Interaction of the pyridine active compounds with the modeled peptide. ii). Interaction of the coumarin active compound with the peptide.

Previous unpublished work on MALDI analysis of purified TopoIIβKHIV from viral lysate reported short peptide fragments. Screening the peptides for previously described properties presented a 70-residue peptide. The sequence along with two other peptides was modelled using PEPFOLD3 and validated by RAMPAGE. Many inhibitors designed to target kinases, specifically bind the ATP-binding pocket, often forming similar hydrogen bonds that are formed by the adenine of ATP. Thus, identifying the ATP binding pocket of the kinases becomes important. Docking of the modelled MALDI peptide with ATP was performed as described previously. The binding region residues interacting directly with the ATP structure consisted of a Lysine reside which resonated with the conserved Lysine observed in the ATP binding domain signature obtained from PROSITE.

The high active compounds from both the groups were docked to the modelled 70 residue peptide and the affinity was observed to be high as the charge and placement of the amino acid residues satisfied the conditions. The overall charge of the peptide was 8.6 with 17% of the residues being positively charged and the affinity reported by AUTODOCK Vina was -7.7Kcal/mol for coumarin derivative and -6.4Kcal/mol for pyridine derivative (Tab. 14). The difference in affinity between the two derivatives can be explained by the bulkier groups on coumarin moiety satisfying the steric requirement specified by the contour, compared to the pyridine compounds (**Fig 39**) .

Summary and Discussion

In this objective, 3D QSAR models for pyridine and coumarin derivatives of Topoisomerase II Beta Kinase inhibitors were developed. 36 Pyridine derivatives and 34 Coumarin derivatives were used to develop these models with good statistical parameters. q^2 value of 0.603 was obtained for pyridine compounds with a standard error estimate of 0.61. For Coumarin derivatives , a q^2 value of 0.743 was obtained and a standard error estimate for the model was 0.091. Prediction accuracy of the model were assessed, and it showed a good correlation between predicted and experimental values. The 3D contours reported by CoMFA were used to guide the design of molecules with better activity. Potent Molecules designed from the Contour models are given at the end of the section.

The contours and the pattern of activity of the compounds suggested a predominant positive charge-based environment in the binding site of the novel kinase. 3D-QSAR studies conducted on inhibitors of other Ser/Thr Kinases reported contours (**Fig 40**) that showed the increase in activity of the compounds depended upon the substitution of electropositive groups, unlike the contours of TopoIIBK$_{HIV}$ that were dominated by the electronegative substitutions. The 3D-QSAR studies reported in the figure represent the contours of Glycogen synthase kinase 3β, a Ser/Thr Kinase belonging to CMCG family (Nigus Dessalew et al., 2007), B-Raf Ser/Thr kinase (Xie et al., 2015), Caesin Kinase II (Liu et al., 2011), Aurora B Kinase (Zhang et al., 2010) and p38 MAP Kinase (Lan et al., 2010). The blue polyhedrons seen in the figure represent the favourable regions for substitution of electropositive groups which dominate the pharmacophore. The contours reported for TopoIIBK$_{HIV}$ are dominated by the red polyhedrons representing the electronegative substitutions. An increased selection of positively charged residues by the compound in the binding interface was observed in the docking studies conducted using peptides designed in-silico. The specificity of the compounds was studied by docking to conducted on kinases from different kinase families. A peptide derived from the MALDI analysis of the TopoIIBK$_{HIV}$ reflected properties observed in the CoMFA studies and also showed affinity to the active compounds. The peptide showed good binding with ATP with the binding region consisting of a Lysine Residue which was observed as conserved in Kinases.. These observations provide an insight into the binding site of the Kinase. The results obtained in the in-silico studies are required to be validated by in-vitro experiments.

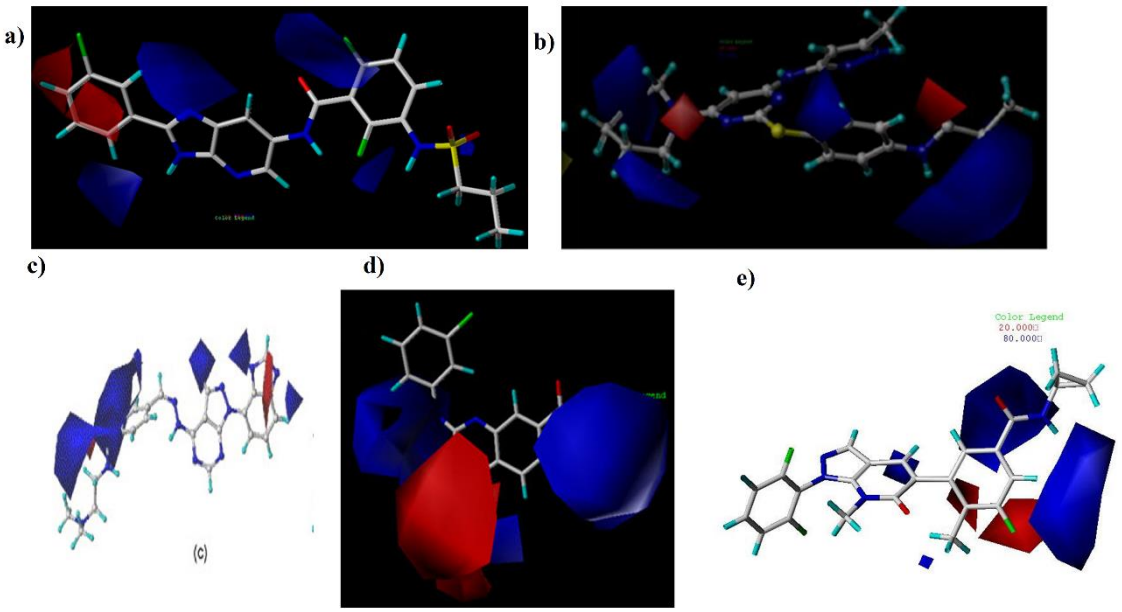
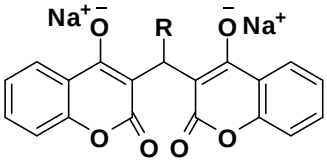
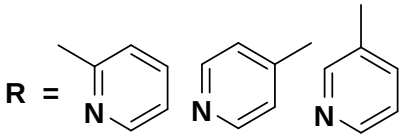


Fig. 40: Represents the contour plots of various Ser/Thr Kinase inhibitors obtained from Literature. a) B-raf Kinase b) Aurora Kinase c) glycogen synthase kinase 3β d) Caesin Kinase II e) p38 MAP Kinase. The contours of electrostatic contributions of the high active inhibitors for the kinase show a higher requirement of electropositive groups to enhance activity unlike the TopoIIBK$_{HIV}$, where a requirement for electronegative groups was observed.

High active Compounds designed using 3D-QSAR studies

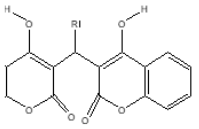
a) Pyridine Compounds



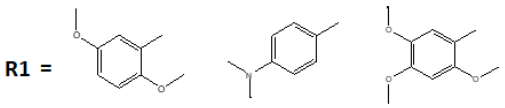


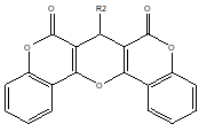
Compound	Kinase Inhibition IC50 (pM)
UHAKKM-7	0.9±0.03
UHAKKM-8	0.8±0.09
UHAKKM-9	0.8±0.05

b) Coumarin compounds

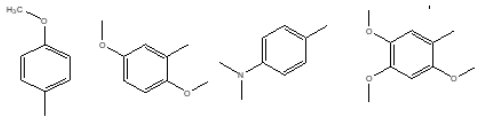


Pyridine dicoumarol





Pyridylidene-4,4[1]-epoxy Dicoumarin



S.No	R	Kinase InhibitionIC50 (pM)
UHAKKM-20	R2	8.5878±0.1
UHAKKM-21	R1	7.012±0.08
UHAKKM 22	R2	4.410±0.04
UHAKKM-23	R1	5.760±0.23
UHAKKM-24	R2	5.510±0.07
UHAKKM-25	R1	4.417±0.17
UHAKKM-26	R2	7.352±0.3

CHAPTER VI

OBJECTIVE 4 : Facilitating cellular localization of Topoisomerase II Beta Kinase inhibitors using Lactoferrin Nanoparticles.

Introduction

Pyridine epoxy dicoumarin derivatives (UHAKKM 1-3) targeting TopoIIβKinase developed from the 3D-QSAR, showed high kinase inhibition at low concentration but showed no anti HIV-1 activity even at high concentration (uM). Assuming the rigidity of the structures, Pyridine dicoumarol derivatives (UHAKKM 4-6) with an open ring and de-cyclized structure were synthesized to increase flexibility (Fig. 41). However, the compounds showed a similar pattern of activity. The formation of intramolecular hydrogen bonds in these structures was suspected to cause rigidity of the compounds. From the above observations, the rigidity in the compounds was assumed to be responsible for the decreased anti-HIV-1 activity due to the incapability to reach the intracellular target (Kammari et al., 2017).

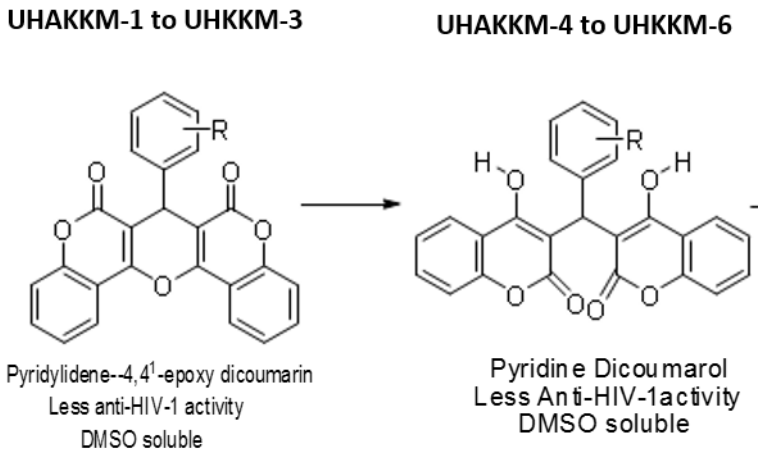


Fig. 41: Rigid structures of compounds UHAKKM1-6

Application of Nanoparticles in the field of medicine, especially for delivering an optimal dosage of the drug to the intended target; thus, affecting the efficacy, toxicity profile and patient outcomes, have been attracting attention (Khan et al., 2017). Many nano-vehicles like liposomes, micelles, quantum dots, dendrimers etc. have been developed for various applications (Jain et al., 2018). The use of protein-based carriers has been reported in the form of prodrugs, nanoparticles, conjugates, microcapsules, gels and scaffolds. The advantage that proteins confer as carriers are the characteristics like non-toxic, easy metabolism and greater biocompatibility. Apart from these, proteins are degradable inside the human body and also elicit a very weak immune

response (Jain et al., 2018). The enhanced solubility and increased bioavailability of the nanoparticles are due to their smaller size and large surface area, which also promotes crossing of the Blood-Brain Barrier and facilitates entry into the pulmonary system (Rizvi et al., 2018). Facilitating cellular uptake of drug molecules that are not membrane permeable is one of the important application of nanoparticles, i.e. making the payload available at the target, which could be a tissue/ cell (Senapati et al., 2018).

Many proteins like albumin, fibrin, gelatin, gliadin and others that have been derived from sources like insects, plants, animals and other protein expression systems (Jain et al., 2018) β-lactoglobulin and lactoferrin are known for their ability to form gels and their nutrient value. The use of Transferrin and lactoferrin nanoparticles as drug carriers has been widely demonstrated. Intravenous delivery of doxorubicin-loaded apo-transferrin and lactoferrin nanoparticles in rats with hepatocellular carcinoma showed enhanced efficacy and bioavailability of the chemotherapeutic agent along with reduced cardiotoxicity (Golla et al., 2013 a). Similar results were observed in the oral administration of the nanoparticle formulation (Golla et al., 2013 b). Carboplatin loaded apo-transferrin and lactoferrin nanoparticles demonstrated better intracellular uptake of chemo agent in retinoblastoma cell lines. High anti-proliferative activity (Apo-nano-carbo IC50=4.31 μg ml(-1), Lacto-nano-carbo IC50=4.16 μg ml(-1), Sol-carbo IC50=13.498 μg ml(-1)) was observed in these formulations (Ahmed et al., 2014). Lactoferrin nanoparticle loaded with 5-fluorouracil has been demonstrated to have higher efficiency in targeted delivery of the agent in melanoma cells, shown by in-vitro studies (Kumari et al., 2017 a). For use in glioma, Temozolomide(TMZ) loaded Lactoferrin nanoparticles showed the capability of crossing the blood-brain barrier, as seen by in-vivo studies in mice (Kumari et al., 2017 b). Further, for treatment of Glioblastoma Multiforme, Aurora B Kinase siRNA was delivered with and without TMZ via Lactoferrin Nanoparticles. This formulation demonstrated increased life expectancy in mice when siRNA loaded nanoparticles were used as treatment (Kumari et al., 2018).

The lactoferrin (Lf) from bovine milk is a transferrin family glycoprotein consisting of two globular lobules and is of interest owing to its properties such as antibacterial, high iron binding capacity, antiviral, immunomodulatory (Martins et al., 2016). The use of Lactoferrin nanoparticle as a carrier to deliver anti-HIV compounds is also well documented. Lactoferrin, as a drug carrier, used to deliver the anti-HIV drug, Zidovudine in rats showed increased efficacy, decreased

toxicity and enhanced bioavailability (Kumar et al., 2015). A triple drug combination of Zidovudine, Efavirenz and Lamivudine of the First line Antiretroviral therapy delivered using Lf nanoparticle showed increased bioavailability, better PK profile and reduced toxicity (Kumar et al., 2017). Lf nanoparticles were also shown to be effective in Pre-exposure Prophylaxis (PrEP) in a study conducted with the nanoparticles as triple combination microbicide with loaded Curcumin and Efavirenz. This combination showed increased bioavailability in the vaginal fluid, thus constituting an effective microbicide, they were also shown to be safe (Lakshmi et al., 2016).

Methods:

1. Preparation of Lactoferrin Nanoparticles

Previously established and patented sol-oil method was used for the preparation of lactoferrin nanoparticles. Lactoferrin dissolved in 1X Phosphate Buffered Saline (pH:7.4) was taken at a concentration of 20mg/ml and mixed with 5mg of drugs AKKKM 1-3 and curcumin dissolved in 50 DMSO for different preparations. Following incubation on ice for 60min, the mixture was added to 25ml olive oil. The mixture was allowed to set on ice for 1hr. Sonication was applied to the oil-water interface formed after incubation, with the probe of an ultrasonic homogenizer for 15mins at 20kHz and 60% efficiency. The sonicated mixture was further incubated on ice for 4 hours, followed by centrifugation for 30 minutes at 6000rpm. The excess oil was discarded, and the pellet was washed with cold Diethyl ether to remove the remaining oil drops. The washed pellet was dissolved in pH 7.4 Phosphate Buffered Saline (1x PBS) and stored at 4ºC for further use. The blank Lactoferrin nanoparticles were prepared by the same method but did not involve the incubation with drugs. The NPs loaded with drugs were labelled KM1NPs, KM2NPs, KM3NPs, Curcumin loaded Nanoparticle was labelled CurNPs and blank Nanoparticles were labelled blank NPs.

2. Characterization of Nanoparticles

2.1 Morphological and Size of nanoparticles:

Field emission scanning microscopy (FE-SEM) was used to calculate the diameter of the nanoparticles. The sample was prepared by spotting the freshly prepared nanoparticles suspended in 1ml milliQ, on a sterile glass slide and drying overnight. The dried slide was coated with gold and analysed under FE-SEM and images were captured as per instructions. For Transmission Electron Microscopy (TEM), 10 µl of suspended nanoparticles were spotted and evenly distributed onto a copper grid of size (Carbon type B 200 mesh). The grid was dried, stained with 2% Uranyl acetate and analysed using TEM.

2.2 Hydrodynamic Radius and Charge of the nanoparticles:

To confirm the stability of the nanoparticles, the zeta potential was measured using ZetaSizer at 25ºC. The sample was prepared by diluting the nanoparticles in MilliQ water and injecting into the instrument. The image was captured at the lowest background noise.

2.3 Encapsulation Efficiency

Quantifying the drug encapsulated in nanoparticles is essential to understand the concentration of the drug used for in-vitro treatments. The drug-loaded nanoparticles were incubated in 1X PBS(pH 5.0) at room temperature for 24hrs to release the drug from the particles into the aqueous phase. Further, to precipitate the Lactoferrin protein, 200 µl of 30% $AgNO_3$ was added to the previously incubated nanoparticles. To the precipitated mixture, 1ml of HPLC grade methanol was added to facilitate the movement of the released drug into the organic phase. Centrifugation of the mixture was carried out at 12000rpm for 20mins followed by passing the supernatant through 0.4-micron filter. The quantification of the drug in the filtrate was performed by semi-preparative HPLC (Waters 2487). The encapsulation efficiency is calculated as follows:

$$\boldsymbol{EE}\% = (\boldsymbol{Amount\ of\ drug\ present\ in\ NP} \div \boldsymbol{Initial\ amount\ of\ drug\ used}) \times \mathbf{100}$$

3. Nanoparticle localization Assay:

The efficiency of nanoparticles to facilitate cellular uptake of the loaded drug was monitored through fluorescent studies. For this purpose, intrinsically fluorescent Curcumin with λEx 458 nm and λEm 530 nm was loaded into Lactoferrin Nanoparticles and used for the studies. For the treatment, 15,000 SupT1 cells were allowed to adhere for 2hrs onto Poly-D-Lysine coated coverslips. Similarly, 15,000 HL2/3 (adherent cell line) were allowed to adhere on sterilized coverslips. The adhered cells were treated with increasing concentrations, ranging from 1 to 10 µM of Cur-Lf NPs. The treatment was carried out at different time points. The excess of nanoparticles after incubation was washed thrice with 1X PBS (pH 7.4) and the cells were stained with DAPI. The coverslips were mounted on glass slides and visualized under Fluorescent Microscope.

4. Cell Culture Assays

a) Propagation of Cell lines

SupT1 cells were maintained in RPMI 1640 complete medium supplemented with 10% FBS, 100U/ml of penicillin, and 100µg/ml of streptomycin. The cells were cultured in a 5% CO_2 incubator at 37°C. HL2//3 cells were maintained in DMEM complete medium with 10% FBS, 100U/ml of penicillin, and 100µg/ml of streptomycin. The cells were cultured in a 5% CO_2 incubator at 37°C.

b) Measurement of cytotoxicity by MTT assay

$0.2x10^6$ SupT1 cells were seeded per well in a 96-well plate and incubated for 4 hours in a 5% CO_2 incubator. Following incubation, the cells were treated with increasing concentrations of nanoparticles AKKM-1 NP, AKKM-2 NP and AKKM-3 NP dissolved in PBS and soluble compounds UHAKKM-1, UHAKKM-2, UHAKKM-3 dissolved in DMSO. The cells were incubated for 16 hours at 37°C in 5% CO_2 incubator. The treatments were performed in triplicate. After incubation, the cells were pelleted by spinning the plate at 1200 rpm for 7 min. After discarding the media, the pellet was re-suspended in complete medium. 20µl from 5mg/ml working solution of MTT dissolved in DMSO was added to each well. The cells were incubated for 4hrs. MTT-formazan crystals formed upon incubation were pelleted by spinning at 1200 rpm for 7 min. The formazan crystals were dissolved by adding 100uL of DMSO into each well and incubated in the dark for 5 min. The colour change was quantified in an ELISA reader at 595 nm. The average of the triplicates with standard deviation was plotted to represent cell survival.

5. Measurement of anti-viral activity in-vitro:

a) Propagation of Virus:

SupT1 cells were infected with HIV-1 primary isolate 93IN101, and the virus was propagated. 5*107 SupT1 cells were incubated with the undiluted virus, in RPMI 1650 medium supplemented with 0.1% FBS. Following four-hour incubation in 5% CO2 incubator at 37◦C, the cells were pelleted at 1200 rpm for 7 min. The pelleted cells were incubated in complete medium further for 96 hrs. The supernatant was collected and tested for the p24 antigen after incubation. The

virus was propagated for multiple passages and upon achieving a good yield, the virus was collected and used for further experiments.

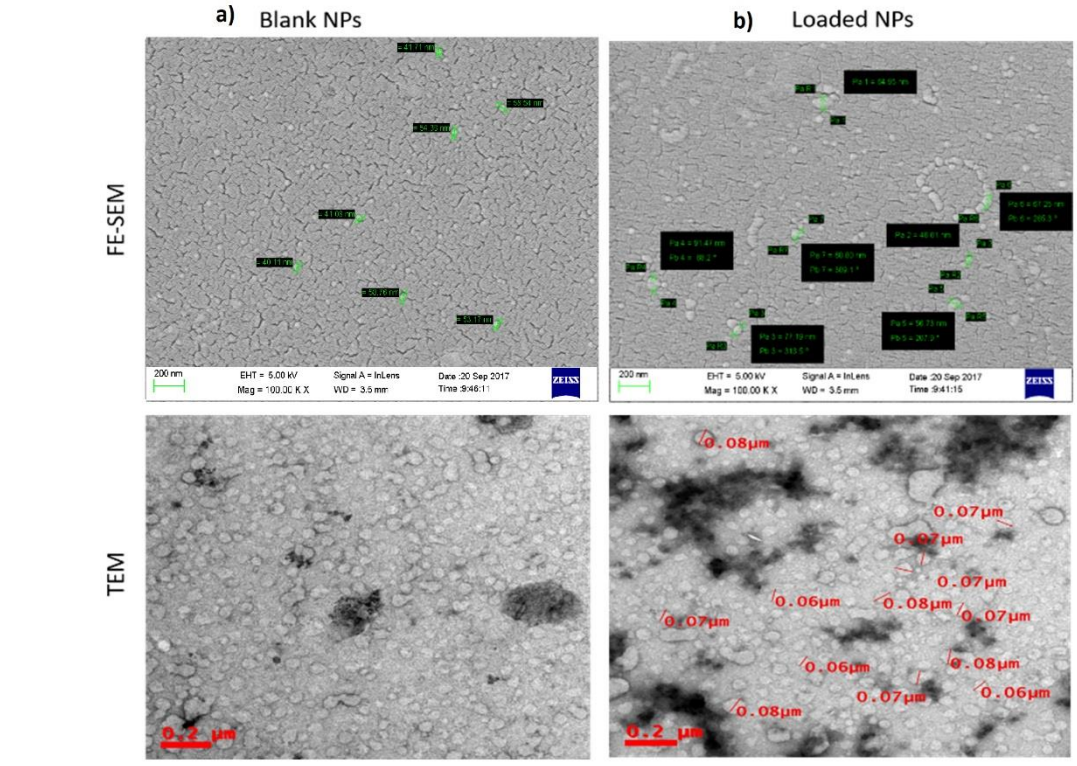
b) p24 quantification:

The levels of viral envelope antigen p24 were detected, to quantify HIV-1 infection, and the levels were detected in the cell supernatant after 96hr incubation period using p24 Antigen ELISA Kit. The protocol specified by the manufacturer was used for the experiment.

A 10-fold dilution of the collected cell culture supernatants was made with complete media. 100µl each of this dilution was added to the wells of the precoated ELISA plate. 25µl disruption buffer was added to each well, and the plate was sealed. Incubation for 1hr at 37°c was followed by aspirating the contents and washing the wells 3x times with 1x wash buffer. 100µl of the conjugate solution was added to each well; the plate was sealed and incubated at 37°C for 1 hour. Following the previous washing steps, the well was aspirated and washed 3x times with 1x wash buffer. To the washed wells, 100µl of substrate solution was added, and the plate was sealed. 20 min incubation at room temperature led to the change in colour of the solution. The reaction was stopped after 20 minutes by adding 100µl of 1N HCl. At 450nm, the optical density of the wells was recorded with an ELISA Reader. The p24 concentration of propagated virus was calculated according to the standard curve. The standard curve was established using standard p24 protein dilutions provided by the manufacturer.

c) HIV-1 neutralization assay:

SupT1 cells evenly suspended in RPMI1640 were seeded at a concentration of 106cells/well supplemented with 0.1% FBS. The cells were challenged with the 1ng/ml p24 equivalent of the 93IN101 virus in the presence of increasing concentrations of compounds. The cells were incubated for 4hrs at 37°C in 5% CO2 incubator. The cells were pelleted at 1200rpm for 7 mins and washed in RPMI 1640 supplemented with 10% FBS. The washed pellets were resuspended in complete fresh medium and incubated for 96h. After 96hr incubation, the cells were pelleted, and the supernatants were collected. The supernatants were analysed for the p24 antigen using the protocol described above. p24 levels in the absence of compounds were considered as 0% inhibition and were taken as the negative control. The infection levels in the presence of AZT(Zidovudine) was considered as positive control.



c)

Delsa™ Nano
Common

BECKMAN COULTER

Distribution Graph S/N : 139411

User	: Common	Group	: sunil	Repetition	: 1/1
Date	: 10/17/2011	File Name	: lacto NP		
Time	: 03:54:21	Sample Information	: lacto NP_1		
SOP Name	: polomi			Security	: No Security

Version 3.73 / 2.30

Mobility Distribution

Measurement Results

Zeta Potential	: -3.82	(mV)	Doppler shift	: 4.33	(Hz)
Mobility	: -5.946e-005	(cm²/Vs)	Base Frequency	: 126.1	(Hz)
Conductivity	: 7.2857	(mS/cm)	Conversion Equation	: Smoluchowski	

Zeta Potential of Cell			Diluent Properties		
Upper Surface	: 7.84	(mV)	Diluent Name	: WATER	
Lower Surface	: -29.63	(mV)	Temperature	: 25.0	(°C)
Cell Condition			Refractive Index	: 1.3328	
Cell Type	: Flow Cell		Viscosity	: 0.8878	(cP)
Avg. Electric Field	: -15.09	(V/cm)	Dielectric Constant	: 78.3	
Avg. Current	: -5.49	(mA)			

d)

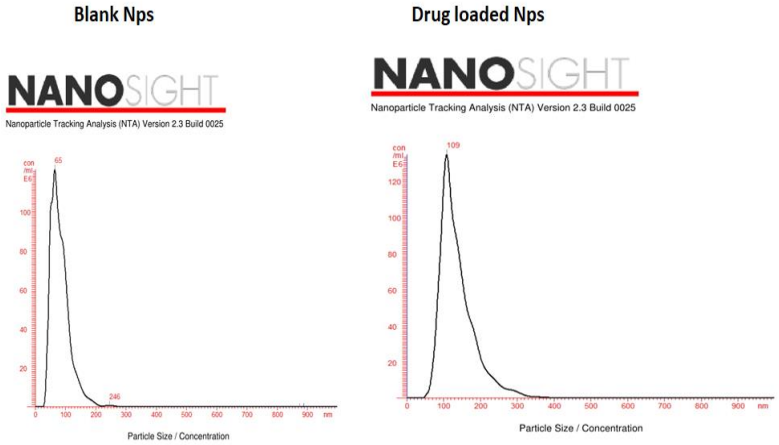


Fig. 42: Characterization of size and shape of a)Blank Lf-NPs b) Compound loaded Lf-NPs by TEM and FE-SEM. Compound loaded Lf NPs showed a size range of 65nm to 80nm and Blank Lf-MES NPs were in the size range of 40nm to 60nm.c) Zeta potential measured by Zeta-sizer reported a value of -3.82 mV.d) The size observed by DLS was in the range of 60-110 nm in PBS pH 7.4

Results:

1. Nanoparticle preparation and Characterization

1.1 Size and Morphology:

The rate of absorption and uptake of nanoparticles is dependent mainly on the shape and size of the particles. Moreover, it has been recently observed that the details of the NP surface topology have considerable effect on cellular uptake. The use of advanced microscopy techniques makes possible to visualise the surface topology of the most effective nanoscale vector present in nature, the virus and suggests that the presence of a pattern or domain on the NP surface facilitates cellular endocytosis by matching specific targets of the cellular surface. Blank, Drug and Curcumin (BlankNPs, UHAKKM-4-6, CurNPs) loaded Lactoferrin nanoparticles were prepared as described previously. Transmission Electron Microscopy (TEM) analysis showed nearly uniform particles with even size and shape. The particles were viewed to be spherical in shape. The results also showed an increase in the size of the drug loaded particles as seen by the diameter of loaded particles . FE-SEM analysed the size of the particles as 40-60 nm, and an increased size of 65-80nm after drug/curcumin loading (**Fig. 42a and b**).

1.2 Hydrodynamic radius and Charge:

Pharmacokinetic properties and half-life of the particles in biological fluids is decided by the size distribution in a solvent, known as soluble size distribution. The distribution in solvent was calculated in milli Q water and was measured at high resolution using Nanosight with particle by particle validation of data by NTA software. The size was measured to be 60-102nm for blank Lf Nps and 100-110nm for loaded Lf-MES-Nps. The Zeta potential was measured using Zetasizer and was recorded as -3.82mV measured for loaded Lf-NPs **(Fig. 42c and d)**.

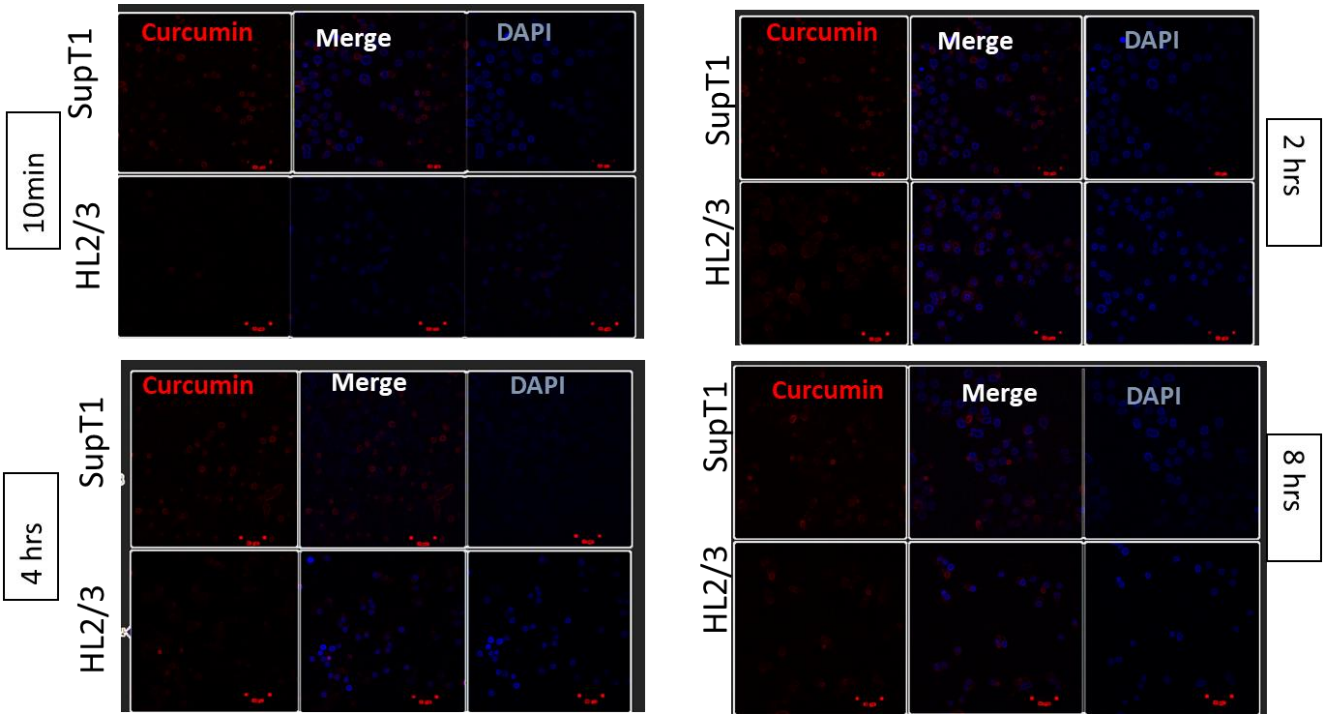


Fig. 43: The localization of Curcumin loaded Lactoferrin Nanoparticles in HL2/3 and SupT1 cells was monitored . Curcumin is labelled in Red, DAPI in blue. Localization was observed at 10 min time point and was at the highest at 4hr time point.

1.3 Encapsulation Efficiency and Drug Loading Capacity:

Therapeutic effectiveness of a nanoparticle formulation is determined by its loading capacity and the encapsulation efficiency. Encapsulation efficiency (EE) depends on the method applied to prepare the nanoparticles and the amount of drug initially used for the preparation . The drug loading mainly depends on the structure , physical and chemical properties of the nanomaterial used as a carrier and low loading capacity is a result of poor drug and nanoparticle interaction. The particles with low loading require administration in high dosages to maintain drug availability, this leads to high toxicity and also problems with excretion and metabolism. The loading capacity and encapsulation efficiency was observed to be 6.5% and 55% respectively.

2. Cellular uptake study:

The cellular uptake study was conducted to understand the release of compounds into the cells when delivered as a nanoparticle formulation, in-vitro. Curcumin was used to facilitate visualization due to its intrinsic fluorescent property. The uptake was studied in SupT1 cells and HL2/3 Cells. The localization was similar in both cell lines. From the **Fig 43**, the early uptake of curcumin in CurNPs treated cells, was observed. .The uptake in CurNPs started at 10min timepoint and was seen to be highest at 4hr time point and curcumin continued to be present in the cells even at 8hr timepoint. This suggests a better uptake of compounds by cells when delivered by nanoparticles.

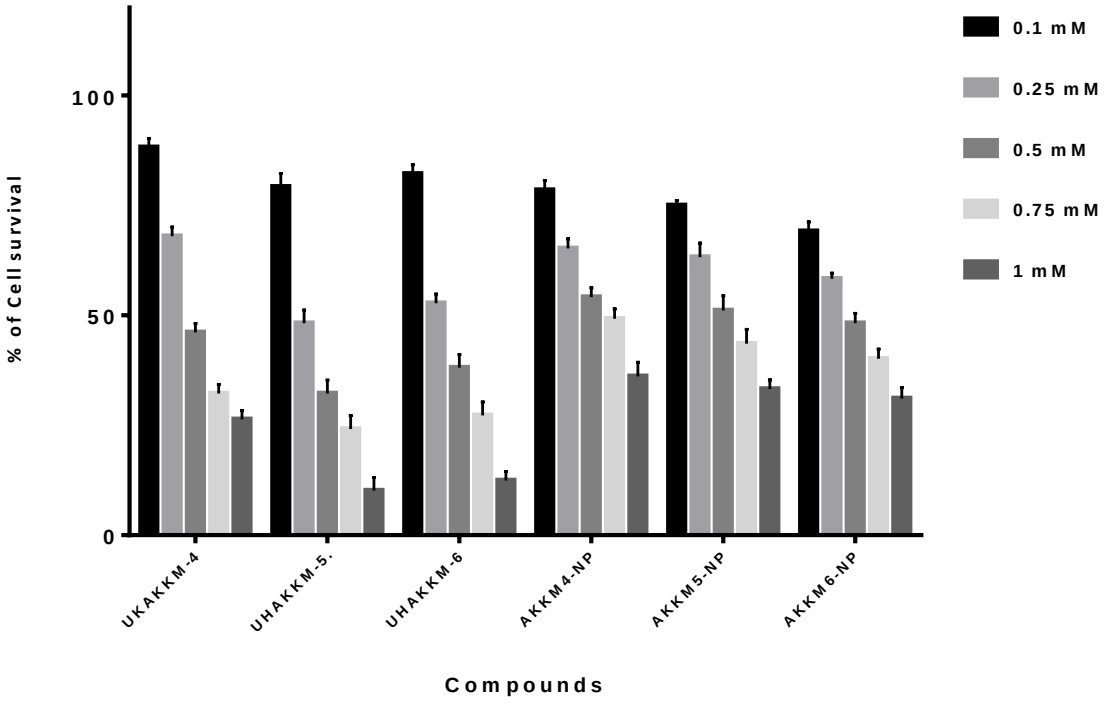


Fig 44: The graph depicts the SupT1 cell survival in presence of nanoparticle formulations KKM4-NP to KKM6-NP. The MTT assay was performed in increasing concentrations (0.1mM,0,25mM,0.5mM,0.75mM, 1mM). The nanoparticles showed low toxicity at concentrations up to 0.5mM. Overall, nanoparticles showed good cell survival when compared to free drugs UHAKKM4-6. The experiments were conducted in triplicate and std deviation was calculated and plotted. p value was ≤ 0.01.

3. MTT Assay:

Cytotoxicity of the nanoparticles was assessed by MTT Assay in SupT1 cells. The cells were incubated at different concentrations of the compounds ranging from 0.1mM to 1mM and the cell survival was estimated by the method described previously. The graph plotted between %cell survival and concentration of the particles showed 40% survival of cells even at particle concentration of 0.75mM suggesting lower toxicity of the nano-formulations (**Fig. 44**). This could be a result of slow and sustained release of the drug in the host cells.

4. Antiviral Assay:

The inhibition of viral propagation by the compound loaded nanoparticles was tested using the method described previously. The compounds despite being highly active in inhibiting TopoIIβK_{HIV} activity in-vitro, showed very poor anti-viral property. Assuming, cell permeability issue, the compounds were loaded into lactoferrin nanoparticles and their anti-viral activity was examined. At 0.5µM concentration, the nanoparticles showed greater than 60% viral inhibition while the soluble compounds showed less than 30% inhibition (**Fig. 45**). Previous studies conducted using nanoparticles for delivering ART drugs suggested an increase in the activity when compared to soluble drugs, but the increase was not significant. In this study, all the nanoparticle formulations showed high potency with significant difference from their soluble counterparts. This might suggest better availability of the active compound to its target resulting from the intracellular delivery mediated by the nanoparticles, thus overcoming the problem of cell permeability.

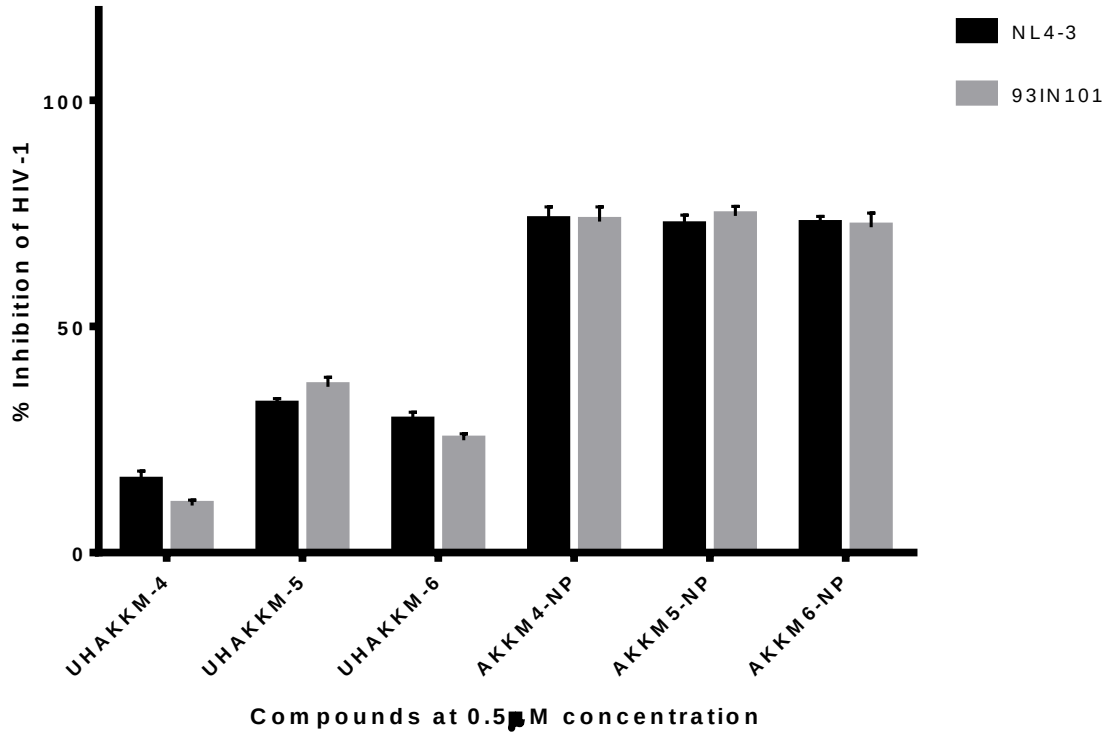


Fig 45: The graph depicts the inhibition of HIV-1 virus propagation in SupT1 cells as quantified by p-24 ELISA Assay. The inhibition was checked at 0.5µM concentration of AKKM-4 to 6 compounds and KKM4-6NP nanoparticles, in Subtype C (93IN101) and Subtype B (NL4-3) strains of the virus. A defined increase in the inhibition can be seen in nanoparticle formulations of the compounds in comparison to soluble compounds. The experiments were conducted in triplicate and std deviation was calculated and plotted. P-value was ≤ 0.0001.

Discussion and Summary:

Biological activity of a compound is largely dependent on its cellular uptake which is based on the interaction of the molecule with the plasma membrane. Molecule can cross the membrane through a simple process called the passive diffusion. In this process, the molecule diffuses across the phospholipid layer by dissolving in it. Upon crossing the membrane the molecules dissolve in the aqueous solution of the cytoplasm (Cooper, 2000). large size of nanoparticles necessitates a different method for its cellular uptake. This is usually through endocytic pathways which leads to an efficient internalization of the particles which is not possible in case of molecules that cannot internalize on their own. This efficient uptake warrants the use of NPs to facilitate cellular entry of small molecules (Mosquera et al., 2018).

Compounds UHAKKM-4 to 6 loaded nanoparticles were prepared by oil in water method. The formed nanoparticles were characterized by TEM,SEM and Nanosight. Drug loading was estimated, and the encapsulation efficiency was observed to be approximately 55%. The nanoparticle induced cytotoxicity was seen to be optimal with greater than 40% cell survival even at high concentrations (0.5mM). The inhibition of HIV-1 propagation was analyzed by p-24 ELISA method. The inhibition was observed to be enhanced when the drugs were delivered through nanoparticles in comparison to soluble compounds . At a concentration of 0.5µM, the compounds delivered through nanoparticles showed an inhibition greater than 60% as opposed to approximately 30% when soluble compounds were tested. This partly confirms that the Rigid structure formed between hydroxyl groups of compounds UHAKKM-4 to 6 which prevents flexibility and thus cell entry. This was overcome by delivery using a drug carrier nanoparticle.

CHAPTER VII

SUMMARY AND CONCLUSION

Broadly, outcome of the thesis can be divided into four parts, each exploring an area pertaining to HIV-1. One of the very first drugs introduced in the clinics was targeted to reverse transcriptase and both nucleoside and non-nucleoside drugs form a major part of the treatment regimens. To understand the Reverse Transcriptase Drug resistance pattern in the Indian population, an analysis was carried out to determine its prevalence under Objective I. The increasing resistance, toxicities and costs associated with the current treatments warrants the discovery of novel drugs and delivery systems. In Objectives II and III, we explored compounds derived from new leads and diverse structural entities developed against novel/known protein targets. Computational methods served as a torch guiding through the seemingly complicated and strenuous process of identifying and developing these novel molecules. In the last objective, Objective IV we explored the use of nanoparticles for delivery of molecules to facilitate cellular entry and potentiate anti-HIV-1 activity.

Objective I:

- A study to explore the Drug Resistance to HIV-1 RT drugs, in the context of Indian Subcontinent was performed.
- The sequences retrieved from Genbank, are from studies conducted in various research and ART centres across the country.
- Analysis to understand the distribution of sequences showed large number of sequence submissions from West region of India (63%).
- Most prevalent Subtype in the population was “Subtype C” which is consistent with reports from various research groups in the country.
- Region wise study reported an increased circulation of resistant mutations in East. The DRM prevalence in South was next to East followed by West. NRTI based mutations were commonly observed in East and North. NNRTI based mutations were commonly observed in South and West.
- Most commonly observed mutations were TAMs and M184V in NRTI. K103N and Y181C/I/V were commonly observed in NNRTI.
- Some major polymorphisms were observed in the sequences, in well conserved regions across Subtype B and C. Some commonly observed are at positions I135, I178, I5, W229, K32, H221. The highest being I135, I178 and I5. W229, K32 and H221 though present in small number are reported first time in India.

Objective II:

- The protein EPAP-1 and broadly neutralizing antibody 447-52D known to possess neutralizing property, were used as base structures to guide the designing of novel inhibitors targeting HIV-1 entry.
- The compounds obtained from virtual screening using 447-52D and from fragment building using EPAP-1 were assessed for chemical synthesis and based on the ease and possibility of synthesis, 8 molecules in total were selected.
- On docking, it was observed that Epap-1 mimetics interacted with a region of gp120 that was in the vicinity of interacting interface of peptide P2. The binding affinity to the regions was intact irrespective of the open or closed loop conformation of gp120. They showed a greater affinity of binding to the base and surrounding residues of V3 loop.
- For 447-52D mimetics, the binding region was similar to the antibody interacting interface. The compounds designed to possess an overall negative charge tend to interact mostly with positively charged amino acids (Arg304). When scores for open and closed V3 loop conformation were compared, better affinity was observed for closed structures than open structures in case of 447-52D mimetics.
- In-vitro studies, to evaluate cytotoxicity of the compounds were conducted on four different cell lines of varying origins. Compounds derived from 447-52D were observed to have better cytotoxic profile in comparison to compounds from Epap-1.
- In p24-ELISA Assay conducted to measure anti-HIV activity, compounds D1-D5 (447-52D) showed higher activity when pre-incubated with virus, while compounds D6-D8 showed no influence of incubation conditions with virus on activity.
- IC50 concentrations for viral inhibition in isolate 93IN101 and laboratory adapted strain NL4-3 showed highest inhibition for compounds D3 and D7 of 447-52D and EPAP-1 respectively.
- The blocking of viral entry was demonstrated by the inhibition of cell fusion via dye redistribution assay. Minimal fusion of cells in presence of compounds D3,D4 and D7 was observed in the experiment with D3 showing highest inhibition of cell fusion.
- The active compounds illustrated in this objective identify novel scaffolds that can be further be developed to yield better molecules.

Objective III:

- Two 3D-QSAR models for pyridine derivative and coumarin derivatives were developed with training molecules.
- On comparing the features of CoMFA models generated for pyridine and coumarin derivatives, the role of electronegative groups was very prominent. Presence of electronegative and negatively charged groups on high active molecules in the dataset, points at a probability of presence of positively charged amino acids, Arg, His, Lys in the binding site of TopoIIBK$_{HIV}$.
- In-silico binding of high active compound with peptide sequences generated with varying positive charge showed increase in affinity to the peptides with increasing positive charge.
- Molecular docking to the ATP binding pocket and allosteric binding pockets of the various Ser/Thr kinases reported lowest affinity to STK11 of STE family and the highest affinity to Casein Kinase 1 of CK1 family.
- A 70-residue peptide derived from MALDI analysis of TopoIIBK$_{HIV}$ showed highest affinity and had the charge and placement of the amino acid residues that satisfied the conditions from above study.
- Residues of the peptide interacting with ATP as per the docked conformation consisted of a Lysine reside which resonated with the conserved Lysine observed in the ATP binding domain signature obtained from PROSITE.
- A prominent charged environment in the binding pocket with a possible Lysine residue is hypothesized in this objective. A peptide fragment identified by MALDI Analysis showed good affinity with the highest active compound.

Objective IV:

- Compound loaded Nanoparticles were prepared by a sol-oil method described in the Prashanth et al.
- Nanoparticle Size of prepared Lactoferrin Nanoparticles was observed to be in the range of 45-60 nm for blank and 65-80 nm for loaded NPs and the zeta potential was -3.82 mV. Nanoparticles in the obtained size and potential have been reported to be optimum for use as delivery systems.
- The cellular uptake study showed an uptake in CurNPs at 10min time point and was seen to be highest at 4hr time point and curcumin continued to be present in the cells even at

8hr timepoint. This suggests a better uptake of compounds by cells when delivered by nanoparticles.

- Cytotoxicity of the nanoparticles was assessed by MTT Assay in SupT1 cells. The graph plotted between %cell survival and concentration of the particles showed 40% survival of cells even at particle concentration of 0.75mM suggesting lower toxicity of the nano-formulations. This could be a result of slow and sustained release of the drug in the host cells.
- At 0.5µM concentration, the nanoparticles showed greater than 60% viral inhibition. While the soluble compounds showed less than 30% inhibition.
- In this study, we unequivocally showed that anti-HIV-1 compounds devoid of activity in cellular infection could be potentiated by enhanced localization using lactoferrin nanoparticles. Thus, suggesting higher availability of the active compound to its target resulting from the intracellular delivery mediated by the nanoparticles, thus overcoming the problem of cell permeability.

Computational Design, Development, Delivery and Validation of HIV-1 Entry and Early Replication Inhibitors

by Akhila Bommakanti

Submission date: 26-Jul-2019 06:29PM (UTC+0530)
Submission ID: 1155154224
File name: (5.91M)
Word count: 36186
Character count: 191470

Computational Design, Development, Delivery and Validation of HIV-1 Entry and Early Replication Inhibitors

ORIGINALITY REPORT

5%	2%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	C. Bhaskar, Palakolanu S. Reddy, K. Sarath Chandra, Sudeep Sabde, Debashis Mitra, Anand K. Kondapi. "Identification of the potential regions of Epap-1 that interacts with V3 loop of HIV-1 gp120", Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2013 Publication	<1%
2	Submitted to University of Hyderabad, Hyderabad Student Paper	<1%
3	Farhan Ahmed, Mohammad Javed Ali, Anand K. Kondapi. "Carboplatin loaded protein nanoparticles exhibit improve anti-proliferative activity in retinoblastoma cells", International Journal of Biological Macromolecules, 2014 Publication	<1%
4	Submitted to University of Hong Kong Student Paper	<1%
5	www.tandfonline.com Internet Source	

REFERENCES

- Ahmed, F.; Ali, M.; Kondapi, A., 2014: Carboplatin loaded protein nanoparticles exhibit improve anti-proliferative activity in retinoblastoma cells. *Internation Journal of Biological Macromolecules*., **70**, 572–582.
- Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J., 1997: Gapped BLAST and PSI-BLAST : a new generation of protein database search programs, **25**, 3389–3402.
- Anil Chekuri; C.Bhaskar; Satish Bollimpelli, V.; K.Kondapi, A., 2016: TopoisomeraseIIβ in HIV-1 transactivation. *Archives of Biochemistry and Biophysics*., **593**, 90–97.
- Anstett, K.; Brenner, B.; Mesplede, T.; Wainberg, M. A., 2017: HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology*., **14**, 1–16.
- Archin, N. M.; Sung, J. M.; Garrido, C.; Soriano-Sarabia, N.; Margolis, D. M., 2014: Eradicating HIV-1 infection: Seeking to clear a persistent pathogen. *Nature Reviews Microbiology*., **12**, 750–764.
- Arts, E. J.; Hazuda, D. J.; Bushman, E. F. D.; Nabel, G. J.; Swanstrom, R., 2012: HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harbor Perspectives in MedicinePerspectives in Medicine*., **2**, 1–23.
- Barbara S. Taylor, M.D., Magdalena E. Sobieszczyk, M.D., M.P.H., Francine E. McCutchan, Ph.D., and Scott M. Hammer, M. D., 2009: The Challenge of HIV-1 Subtype Diversity. *New England Journal of Medicine*., **358**, 1590–1602.
- Barré-Sinoussi, F., 1996: HIV as the cause of AIDS. *Lancet*., **348**, 31–35.
- Baxter, J. D.; Chasanov, W. M., 2016: An Update on HIV-1 Protease Inhibitor Resistance. *Journal of AIDS & Clinical Research*., **7**, 1–7.
- Bennett, D. E.; Bertagnolio, S.; Sutherland, D.; Gilks, C. F., 2006: The World Health Organization ' s global strategy for prevention and assessment of HIV drug resistance. *Antiretroviral Therapy*., **13**, 1–13.
- Bhaskar, C.; Reddy, P. S.; Chandra, K. S.; Sabde, S.; Mitra, D.; Kondapi, A. K., 2013: Identification of the potential regions of Epap-1 that interacts with V3 loop of HIV-1 gp120. *BBA - Proteins and Proteomics*., **1834**, 780–790.
- Blanco, J. L.; Whitlock, G.; Milinkovic, A.; Moyle, G., 2015: HIV integrase inhibitors: a new era in the treatment of HIV. *Expert Opinion on Pharmacotherapy*., **16**, 1313–1324.
- Bonsignori, M.; Liao, H. xin; Gao, F.; Williams, W. B.; Alam, S. M.; Montefiori, D. C.; Haynes, B. F., 2017: Antibody virus co- - evolution in HIV infection : paths for HIV vaccine development. *Wiley Immunological Reviews*., **275**, 145–160.
- Bowder, D.; Hollingsead, H.; Durst, K.; Hu, D.; Wei, W., 2018: Contribution of the gp120 V3 loop to envelope glycoprotein trimer stability in primate immunode fi ciency viruses. *Virology*., **521**, 158–168.

- Brady, M. T.; Oleske, J. M.; Williams, P. L.; Elgie, C.; Mofenson, L. M.; Dankner, W. M.; Van Dyke, R. B., 2010: Declines in mortality rates and changes in causes of death in HIV-1-infected children during the haart era. *Journal of Acquired Immune Deficiency Syndromes.*, **53**, 86–94.
- Brenner, B.; Turner, D.; Oliveira, M.; Moisi, D.; Detorio, M.; Carobene, M.; Marlink, R. G.; Schapiro, J.; Roger, M.; Wainberg, M. A., 2003: A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. *AIDS.*, **17**.
- Camproux, A. C.; Gautier, R.; Tuffe, P., 2004: A Hidden Markov Model Derived Structural Alphabet for Proteins. *Journal of Molecular Biology.*, **339**, 591–605.
- Carlos, Luiz Alcantara, J.; Cassol, S.; Libin, P.; Deforche, K.; Pybus, O. G.; Ranst, M. Van; Galva, B., 2009: A standardized framework for accurate , high-throughput genotyping of recombinant and non-recombinant viral sequences, **37**, 634–642.
- Chan, K. W.; Pan, R.; Costa, M.; Gorny, M. K.; Wang, S.; Lu, S.; Kong, X. P., 2018: Structural Comparison of Human Anti-HIV-1 gp120 V3 MAbs of the Same Gene Usage Induced by Vaccination and Chronic Infection. *Journal of Virology.*, **92**, 1–23.
- Chaplin, B.; Eisen, G.; Idoko, J.; Onwujekwe, D.; Idigbe, E.; Adewole, I.; Gashau, W., 2011: Impact of HIV Type 1 Subtype on Drug Resistance Mutations in Nigerian Patients Failing First-Line Therapy. *AIDS Research and Human Retroviruses.*, **27**, 71–80.
- Cheek, S.; Zhang, H.; Grishin, N. V, 2002: Sequence and Structure Classification of Kinases. *Journal of Molecular Biology.*, **320**, 855–881.
- Cheng, H. chin; Qi, R. Z.; Paudel, H.; Zhu, H. jian, 2011: Regulation and Function of Protein Kinases and Phosphatases. *SAGE Enzyme Resarch.*, **2011**, 7–9.
- Chris Beyrer; Pozniak, A., 2017: HIV Drug Resistance — An Emerging Threat to Epidemic Control. *New England Journal of Medicine.*, **377**, 1605–1607.
- Chu, Y.; Wang, Y., 2015: HIV protease inhibitors : a review of molecular selectivity and toxicity. *HIV/AIDS-Research and Palliative Care.*, **7**, 95–104.
- Cooper, G., 2000: *Transport of Small Molecules*. *The Cell: A Molecular Approach. 2nd edition.* 2nd edn.
- Deeks, S. G., 2017: HIV infection. *Viral Infections in Children.*, **1**, 69–100.
- Deshpande, A.; Karki, S.; Recordon-Pinson, P.; Fleury, H. J., 2011: Drug Resistance Mutations in HIV Type 1 Isolates from Naive Patients Eligible for First Line Antiretroviral Therapy in JJ Hospital, Mumbai, India. *AIDS Research and Human Retroviruses.*, **27**, 4–7.
- Dogo-isonagie, C.; Lam, S.; Acharya, P.; Shahzad-ul-hussan, S.; Marius, G.; Kwong, P. D.; Bewley, C. A., 2012: Peptides from Second Extracellular Loop of C-C Chemokine Receptor Type 5 (CCR5) Inhibit Diverse Strains of HIV-1. *The Journal of Biological Chemistry.*, **287**, 15076–15086.

- Duff, A., 2014: Bidirectional interactions between hiv/aids and indian culture aoife duff. *HIV/AIDS and Indian culture BIDIRECTIONAL*., 188–212.
- Dutta, N.; Nandi, S.; Guha, S. K.; Saha, M. K., 2017: Emergence of HIV drug-resistant mutations in East Indian population after failure of first-line antiretroviral therapy. *HIV & AIDS Review*., **16**, 258–264.
- Engh, R. A.; Huber, R., 1991: Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement, 392–400.
- Farzan, M.; Mirzabekov, T.; Kolchinsky, P.; Wyatt, R.; Cayabyab, M.; Gerard, N. P.; Gerard, C.; Sodroski, J.; Choe, H., 1999: Tyrosine Sulfation of the Amino Terminus of CCR5 Facilitates HIV-1 Entry Beth Israel Hospital and. *Cell*., **96**, 667–676.
- Finzi, D.; Hermankova, M.; Pierson, T.; Carruth, L. M.; Buck, C.; Chaisson, R. E.; Quinn, T. C.; Chadwick, K.; Margolick, J.; Brookmeyer, R.; Gallant, J.; Markowitz, M.; Ho, D. D.; Richman, D. D.; Siliciano, R. F., 1997: Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. *Science*., **278**, 1295–1300.
- Fiser, A.; Sali, A., 2003: ModLoop : automated modeling of loops in protein structures, **19**, 2500–2501.
- Floris, M.; Masciocchi, J.; Fanton, M.; Moro, S., 2011: Swimming into peptidomimetic chemical space using pepMMsMIMIC, **39**, 261–269.
- Ford, N.; Migone, C.; Calmy, A.; Kerschberger, B.; Kanters, S.; Nsanzimana, S.; Mills, E. J.; Meintjes, G.; Vitoria, M.; Doherty, M.; Shubber, Z., 2018: Benefits and risks of rapid initiation of antiretroviral therapy. *Aids*., **32**, 17–23.
- Ghoorah, A. W.; Devignes, M. dominique; Sma, M.; Lorraine, D., 2013: Protein Docking Using Case-Based Reasoning. *Proteins: Structure, Function and Bioinformatics*., **81**, 2150–2158.
- Giroud, C.; Chazal, N.; Briant, L., 2011: Cellular kinases incorporated into HIV-1 particles : passive or active passengers ? *Retrovirology*., **8**, 71.
- Golla, K.; Bhaskar, C.; Ahmed, F.; Kondapi, A., 2013a: A target-specific oral formulation of Doxorubicin-protein nanoparticles: efficacy and safety in hepatocellular cancer. *Journal of Cancer*., **4**, 644–652.
- Golla, K.; Reddy, P.; Bhaskar, C.; Kondapi, A., 2013b: Biocompatibility, absorption and safety of protein nanoparticle-based delivery of doxorubicin through oral administration in rats. *Drug Delivery*., **20**, 156–167.
- Gregson, J.; Kaleebu, P.; Marconi, V. C.; Vuuren, C. Van; Ndembi, N.; Hamers, R. L.; Kanki, P.; Hoff, C. J., 2017: Occult HIV-1 drug resistance to thymidine analogues following failure of fi rst-line tenofovir combined with a cytosine analogue and nevirapine or efavirenz in sub Saharan Africa : a retrospective multi-centre cohort study. *The Lancet Infectious Diseases*., **17**, 296–304.

- Gu, W. G.; Zhang, X.; Yuan, J. F., 2014: Anti-HIV Drug Development Through Computational Methods. *The AAPS Journal*., **16**, 674–680.
- Guha, S. K.; Haldar, S. N., 2012: Drug resistance in HIV. *Medicine Update*., **22**, 84–88.
- Gupta, R. K.; Gregson, J.; Parkin, N.; Haile-selassie, H.; Tanuri, A.; Forero, L. A.; Kaleebu, P.; Watera, C.; Kugathasan, R.; Cutino, T.; Hunt, G.; Rios, S. A.; Doherty, M.; Jordan, M. R.; Bertagnolio, S., 2017: HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries : a systematic review and meta-regression analysis. *The Lancet Infectious Diseases*., **18**, 346–355.
- Haggag, Y. A.; Donia, A. A.; Osman, M. A.; El-gizawy, S. A., 2018: Peptides as Drug Candidates : Limitations and Recent Development Perspectives. *BiomedicalJournal of Scientific and Technical Review*., **5**, 6659–6662.
- Hall, T. A., 1999: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*., **41**, 95–98.
- Hassan Baig, M.; Ahmad, K.; Roy, S.; Mohammad Ashraf, J.; Adil, M.; Haris Siddiqui, M.; Khan, S.; Amjad Kamal, M.; Provazník, I.; Choi, I., 2016: Computer Aided Drug Design: Success and Limitations. *Current Pharmaceutical Design*., **22**, 572–581.
- Hemelaar, J., 2012: The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine*., **18**, 182–192.
- Herman, B. D.; Sluis-Cremer, N., 2012: *Molecular Pharmacology of Nucleoside and Nucleotide HIV-1 Reverse Transcriptase Inhibitors*. *Pharmacology*.
- Hillis, D. M., 2000: Origins of HIV. *Science Perspectives: AIDS*., **288**, 1757–1759.
- Hollingsworth, L. R.; Brown, A. M.; Gandour, R. D.; Bevan, R., 2018: Computational study of HIV gp120 as a target for polyanionic entry inhibitors : Exploiting the V3 loop region. *PLoS ONE*., **13**, 1–22.
- Huang, C. chin; Tang, M.; Zhang, M. yun; Majeed, S.; Montabana, E.; Stanfield, R. L.; Dimitrov, D. S.; Korber, B.; Sodroski, J.; Wilson, I. A.; Wyatt, R.; Kwong, P. D., 2005: Structure of a V3-Containing HIV-1 gp120 Core. *Science*., **310**, 1025–1028.
- Iqbal, S.; Maqsood, S.; Zafar, A.; Zakar, R.; Zakar, M. Z.; Fischer, F., 2019: Determinants of overall knowledge of and attitudes towards HIV/AIDS transmission among ever-married women in Pakistan: evidence from the Demographic and Health Survey 2012–13. *BMC Public Health*., **19**, 1–14.
- Jacob, T.; Broeke, V. Den; Favoreel, H. W., 2011: Viral Serine / Threonine Protein Kinases. *Journal of Virology*., **85**, 1158–1173.
- Jain, A.; Singh, S. K.; Arya, S. K.; Kundu, S. C.; Kapoor, S., 2018: Protein Nanoparticles : Promising Platforms for Drug Delivery Applications. *ACS Biomaterials Science & Engineering*., **4**, 3939–3961.

- James, M.; Murtola, T.; Schulz, R.; Smith, J. C.; Hess, B.; Lindahl, E., 2015: GROMACS : High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*., **2**, 19–25.
- Jiang, X.; Burke, V.; Totrov, M.; Williams, C.; Cardozo, T.; Gorny, M. K.; Zolla-pazner, S.; Kong, X. peng, 2010: Conserved structural elements in the V3 crown of. *Nature Structural and Molecular Biology*., **17**, 955–962.
- JOHNSON, V. A., 2009: Combination Therapy: More Effective Control of HIV Type 1? *AIDS Research and Human Retroviruses*., **10**, 907–912.
- Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., 1997:Development and Validation of a Genetic Algorithm for Flexible Docking . *Journal of Molecular Biology*., **267**, 727-748.
- K.Kondapi, A.; M.A.Hafiz; T.Sivaram, 2002: Anti-HIV activity of a glycoprotein from first trimester placental tissue. *Antiviral Research*., **54**, 47–57.
- Kammari, K.; Devaraya, K.; Bommakanti, A.; Kondapi, A. K., 2017: Development of pyridine dicoumarols as potent anti HIV-1 leads, targeting HIV-1 associated topoisomeraseIIβ kinase. *Future Medicinal Chemistry*., **9**, 1597–1609.
- Kandathil, A.; Kannangai, R.; Abraham, O.; Pulimood, S.; Jensen, M.; Sridharan, G., 2009: A comparison of interpretation by three different HIV type 1 genotypic drug resistance algorithms using sequences from non-clade B HIV type 1 strains. *AIDS Res Hum Retroviruses*., **25**, 315–318.
- Kannangai, R.; David, S.; Sundaresan, V. C.; Sachithanandham, J.; Mani, M.; Cherian, O., 2015: Frequency of Transmitted Drug Resistance Mutations Among ¨ ve HIV-1-Infected Individuals at a Tertiary Care Centre in South India. *Molecular Diagnosis & Therapy*., **19**, 273–275.
- Kans, J., 2013: *Entrez Direct : E-utilities on the UNIX Command Line*.
- Karade, S. K.; Kulkarni, S. S.; Ghate, M. V; Patil, A. A.; Londhe, R.; Salvi, S. P.; Kadam, D. B.; Joshi, R. K.; Rewari, B.; Gangakhedkar, R. R., 2017: Antiretroviral resistance following immunological monitoring in a resource- limited setting of western India : A cross- sectional study. *PLoS ONE*., **12**, 1–11.
- Katz, I. T.; Ehrenkranz, P.; El-Sadr, W., 2018: The global HIV epidemic: What will it take to get to the finish line? *JAMA - Journal of the American Medical Association*., **319**, 1094–1095.
- Khan, I.; Saeed, K.; Khan, I., 2017: Nanoparticles : Properties , applications and toxicities. *Arabian Journal of Chemistry*.
- Kijewski, S. D. G.; Gummuluru, S., 2015: A mechanistic overview of dendritic cell-mediated HIV-1 trans infection : the story so far. *Future Virology*., **10**, 257–269.
- Killikelly, A.; Zhang, H. tang; Spurrier, B.; Williams, C.; Gorny, M. K.; Zolla-pazner, S.; Kong, X. peng, 2013: Thermodynamic Signatures of the Antigen Binding Site of mAb 447 − 52D Targeting the Third Variable Region of HIV - 1 gp120. *Biochemistry*., **52**, 6349–6257.

- Kim, K. H., 1995: 12 Comparative molecular field analysis (CoMFA). *Molecular Similarity in Drug Design*., pp. 291–331.
- Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H., 2016: PubChem Substance and Compound databases, **44**, 1202–1213.
- Klapper, I.; Hagstrom, R.; Fine, R.; KSharp; Honig, B., 1986: Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins*., **1**, 47–59.
- Kondapi, A. K.; Padmaja, G.; Satyanarayana, N.; Mukhopadyaya, R.; Reitz, M. S., 2005: A biochemical analysis of topoisomerase II and kinase activity found in HIV-1 infected cells and virus ☪. *Archives of Biochemistry and Biophysics*., **441**, 41–55.
- Kondapi, A. K.; Satyanarayana, N.; Saikrishna, A. D., 2006: A study of the Topoisomerase II activity in HIV-1 replication using the ferrocene derivatives as probes ☪. *Archives of Biochemistry and Biophysics*., **450**, 123–132.
- Kumar, P.; Lakshmi, Y.; Bhaskar, C.; Golla, K.; Kondapi, A., 2015: Improved Safety, Bioavailability and Pharmacokinetics of Zidovudine through Lactoferrin Nanoparticles during Oral Administration in Rats. *PLoS ONE*., **10**, 1–17.
- Kumar, P.; Lakshmi, Y.; Kondapi, A., 2017: Triple Drug Combination of Zidovudine, Efavirenz and Lamivudine Loaded Lactoferrin Nanoparticles: an Effective Nano First-Line Regimen for HIV Therapy. *Pharmaceutical Research*., **34**, 257–268.
- Kumar, R.; Id, J.; Mehendale, S. M., 2019: Determinants of consistently high HIV prevalence in Indian Districts : A multi-level analysis. *PLoS ONE*., **14**, 1–14.
- Kumari, S.; Kondapi, A., 2017a: Lactoferrin nanoparticle mediated targeted delivery of 5-fluorouracil for enhanced therapeutic efficacy. *Internation Journal of Biological Macromolecules*., **95**, 232–237.
- Kumari, S.; Ahsan, S.; Kumar, J.; Kondapi, A.; Rao, N., 2017b: Overcoming blood brain barrier with a dual purpose Temozolomide loaded Lactoferrin nanoparticles for combating glioma (SERP-17-12433). *Scientific Reports*., **7**, 6602.
- Kumari, S.; Bhattacharya, D.; Rangaraj, N.; Chakarvarty, S.; Kondapi, A.; Rao, N., 2018: Aurora kinase B siRNA-loaded lactoferrin nanoparticles potentiate the efficacy of temozolomide in treating glioblastoma. *Nanomedicine (Lond)*., **13**, 2579–2596.
- Kurle, S.; Gangakhedkar, R.; Sen, S.; Hayatnagarkar, S.; Tripathy, S.; Paranjape, R., 2007: Emergence of NNRTI drug resistance mutations after single-dose nevirapine exposure in HIV type 1 subtype C-infected infants in India. *AIDS Res Hum Retroviruses*., **23**, 682–685.

- Kwong, P. D., 2004: The 447-52D Antibody : Hitting HIV-1 Where Its Armor Is Thickest. *Elsevier Structure*., **12**, 173–174.
- Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Hendrickson, W. A.; Biophysics, M.; Diseases, I.; Orleans, N.; Pharmaceuticals, S. B., 2017: Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*., **393**, 648–659.
- Lakshmi, Y. S.; Kumar, P.; Golla, K.; Bhaskar, C.; Kondapi, A., 2016: Triple combination MPT vaginal microbicide using curcumin and efavirenz loaded lactoferrin nanoparticles. *Scientific Reports*., **6**, 1–13.
- Lall, M.; Gupta, R.; Sen, S.; Kapila, K.; Tripathy, S.; Paranjape, R., 2008: Profile of primary resistance in HIV-1-infected treatment-naive individuals from Western India. *AIDS Res Hum Retroviruses*., **24**, 987–990.
- Lamiable, A.; Thevenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuff, P., 2016: PEP-FOLD3 : faster denovo structure prediction for linear peptides in solution and in complex, **44**, 449–454.
- Lan, P.; Huang, Z. J.; Sun, J. R.; Chen, W. M., 2010: 3D-QSAR and Molecular Docking Studies on Fused Pyrazoles as p38α Mitogen-Activated Protein Kinase Inhibitors. *International Journal of Molecular Sciences*., **11**, 3357–3374.
- Lange, J.; Levy, J. A.; Cohen, M. S.; Hellmann, N.; DeCock, K., 2008: The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. *Journal of Clinical Investigation*., **118**, 1244–1254.
- Leader, D. P., 1993: VIRAL PROTEIN KINASES AND PROTEIN PHOSPHATASES. *Pharmacology and Therapeutics*., **59**, 343–389.
- Lee, A.; Lee, K.; Kim, D., 2016: Using reverse docking for target identification and its applications for drug discovery. *Expert Opinion on Drug Discovery*., **11**, 707–715.
- Leelananda, S. P.; Lindert, S., 2016: Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry*., **12**, 2694–2718.
- Liu, H.; Wang, X.; Wang, J.; Li, G., 2011: Structural Determinants of CX-4945 Derivatives as Protein Kinase CK2 Inhibitors: A Computational Study. *International Journal of Molecular Sciences*., **12**, 7004–7021.
- Liu, T. F.; Shafer, R. W., 2006: Web Resources for HIV Type 1 Genotypic-Resistance Test Interpretation. *Clinical Infectious Diseases*., **42**, 1608–1618.
- Lobritz, M. A.; Ratcliff, A. N.; Arts, E. J., 2010: HIV-1 entry, inhibitors, and resistance. *Viruses*., **2**, 1069–1105.
- Lovell, S. C.; Davis, I. W.; Iii, W. B. A.; Bakker, P. I. W. De; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., 2003: Structure Validation by C‿ Geometry: ?,? and C$^{\eta}$ Deviation, **450**, 437–450.

- Lu, D. Y.; Wu, H. Y.; Yarla, N. S.; Xu, B.; Ding, J.; Lu, T. R., 2017: HAART in HIV/AIDS Treatments: Future Trends. *Infectious Disorders - Drug Targets*., **18**, 15–22.
- Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M. dominique; Ritchie, D. W., 2010: HexServer : an FFT-based protein docking server powered by graphics processors, **38**, 445–449.
- Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. et al., 1998: All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †, **5647**, 3586–3616.
- Mali, S. N.; Sapkal, P. M., 2015: HIV Drug Resistance: An Overview. *Human Journals Review Article October*., **1**, 72–82.
- Martins, J. T.; Santos, S. F.; Bourbon, A. I.; Pinheiro, A. C.; González-fernández, Á.; Pastrana, L. M.; Cerqueira, M. A.; Vicente, A. A., 2016: Lactoferrin-based nanoparticles as a vehicle for iron in food applications – Development and release pro fi le. *Food Research International*., **90**, 16–24.
- Maupetit, J.; Tuffery, P.; Derreumaux, P., 2007: and structure prediction. *Proteins: Structure, Function and Bioinformatics*., **69**, 394–408.
- Mayr, L. M.; Zolla-pazner, S., 2015: Antibodies Targeting the Envelope of HIV-1. *Microbiology Spectrum*., **3**, 1–11.
- Michael T.Howard; D.Griffith, J., 1993: A Cluster of Strong Topoisomerase II Cleavage Sites is Located Near an Integrated Human Immunodeficiency Virus. *Journal of Molecular Biology*., **232**, 1060–1068.
- Mignani, S.; Huber, S.; Tomás, H.; Rodrigues, J.; Majoral, J. P., 2016: Why and how have drug discovery strategies in pharma changed? What are the new mindsets? *Drug Discovery Today*., **21**, 239–249.
- Miranda-saavedra, D.; Barton, G. J., 2007: Classification and functional annotation of eukaryotic protein kinases. *Wiley InterScience*., **68**, 893–914.
- Mohs, R. C.; Greig, N. H., 2017: Drug discovery and development: Role of basic biological research. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*., **3**, 651–657.
- Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J.; Al, M. E. T., 1998: Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function, **19**, 1639–1662.
- Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., 2010: AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*., **30**, 2785–2791.
- Mosquera, J.; García, I.; Liz-Marzán, L. ., 2018: Cellular Uptake of Nanoparticles versus Small Molecules: A Matter of Size. *Accounts of Chemical Research*., **51**, 2305−2313.

- NACO, (n.d.) Members of the Network of Indian Institutions / Organisations for HIV / AIDS Research (NIIHAR).
- Neogi, U.; Bontell, I.; Shet, A.; Costa, A. De; Gupta, S.; Diwan, V., 2012: Molecular Epidemiology of HIV-1 Subtypes in India : Origin and Evolutionary History of the Predominant Subtype C. *PLoS ONE*., **7**, 1–8.
- Nigus Dessalew; S.PateL, D.; V.Bharatam, Lp., 2007: 3D-QSAR and molecular docking studies on pyrazolopyrimidine derivatives as glycogen synthase kinase-3β inhibitors. *Journal of Molecular Graphics and Modelling*., **25**, 885–895.
- O'Connell, K. A.; Bailey, J. R.; Blankson, J. N., 2009: Elucidating the elite: mechanisms of control in HIV-1 infection. *Trends in Pharmacological Sciences*., **30**, 631–637.
- Pachamuthu, B.; Shanmugam, S.; Nagalingeswaran, K.; Ss, S.; Solomon, S., 2006: HIV-1 drug resistance among untreated patients in India: Current status. *J Postgrad Med*., **52**, 183–186.
- Pantophlet, R.; Burton, D. R., 2006: GP120: Target for Neutralizing HIV-1 Antibodies. *Annual Review of Immunology*., **24**, 739–769.
- Paranjape, R. S.; Challacombe, S. J., 2016: HIV/AIDS in India: An overview of the Indian epidemic. *Oral Diseases*., **22**, 10–14.
- Persaud, D.; Pierson, T.; Ruff, C.; Finzi, D.; Chadwick, K. R.; Margolick, J. B.; Ruff, A.; Hutton, N.; Ray, S.; Siliciano, R. F., 2000: A stable latent reservoir for HIV-1 in resting CD4+ T lymphocytes in infected children. *Journal of Clinical Investigation*., **105**, 995–1003.
- Phillips, M. A.; Stewart, M. A.; Woodling, D. L.; Xie, Z. R., 2018: Has Molecular Docking Ever Brought us a Medicine? *Molecular Docking*., pp. 141–178.
- Pitman, M. C.; Lewin, S. R., 2015: Towards a cure for human immunodeficiency virus, **40**, 1291–1296.
- Ponraj, K.; Prabhakar, M.; Rathore, R. S.; Bommakanti, A.; Kondapi, A. K., 2013: HIV-1 Associated Topoisomerase II Kinase : A Potential Pharmacological Target for Viral Replication. *Current Pharmaceutical Design*., **19**, 4776–4786.
- Poongavanam, V.; Namasivayam, V.; Vanangamudi, M.; Al Shamaileh, H.; Veedu, R. N.; Kihlberg, J.; Murugan, N. A., 2018: Integrative approaches in HIV-1 non-nucleoside reverse transcriptase inhibitor design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*., **8**, 1–26.
- Pugazhendhi, D., 2013: International Journal of Advanced Research in Insilico Methods in Drug Discovery - A Review, **3**, 680–683.
- Pyne, M. T.; Hackett, J.; Holzmayer, V.; Hillyard, R., 2013: Large-Scale Analysis of the Prevalence and Geographic Distribution of HIV-1 Non-B Variants in the United States, **51**, 2662–2669.
- Rai, M. A.; Pannek, S.; Fichtenbaum, C. J., 2018: Emerging reverse transcriptase inhibitors for HIV-1 infection. *Expert Opinion on Emerging Drugs*., **23**, 149–157.

- Rajesh, L.; Ramesh Karunaianantham, Paranji R. Narayanan, S. S., 2009: Antiretroviral Drug-Resistant Mutations at Baseline and at Time of Failure of Antiretroviral Therapy in HIV Type 1-Coinfected TB Patients. *AIDS Research and Human Retroviruses*., **25**, 1179–1185.
- Rani, K. P. R.; Pelluru, D.; K.Kondapi, A., 2006: A conserved molecular action of native and recombinant Epap-1 in inhibition of HIV-1 gp120 mediated viral entry. *Archives of Biochemistry and Biophysics*., **456**, 79–92.
- Resistance, I. U. D., 2007: Update of the drug resistance mutations in HIV-1: 2007. *Topics in HIV Medicine*., **15**, 119–125.
- Rhee, S. yon; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W., 2003: Human immunodeficiency virus reverse transcriptase and protease sequence database, **31**, 298–303.
- Rizvi, S. A. A.; Saleh, A. M., 2018: Applications of nanoparticle systems in drug delivery technology. *Saudi Pharmaceutical Journal*., **26**, 64–70.
- Rodriguez, M. A.; Ding, M.; Ratner, D.; Chen, Y.; Tripathy, S. P.; Kulkarni, S. S.; Chatterjee, R.; Tarwater, P. M.; Gupta, P., 2009: High replication fi tness and transmission ef fi ciency of HIV-1 subtype C from India : Implications for subtype C predominance. *Virology*., **385**, 416–424.
- Roskoski, R., 2015: A historical overview of protein kinases and their targeted small molecule inhibitors. *Pharmacological Research*., **100**, 1–23.
- Sanner, M. F., 2009: Python: a programming language for software integration and development. *J. Mol. Graphics Mod*., **Vol 17**, 57–61.
- Santos, L. H.; Ferreira, R. S.; Caffarena, E. R., 2015: Computational drug design strategies applied to the modelling of human immunodeficiency virus-1 reverse transcriptase inhibitors. *Memorias do Instituto Oswaldo Cruz*., **110**, 847–864.
- Saravanan, S.; Kausalya, B.; Gomathi, S.; Sivamalar, S.; Balakrishnan Pachamuthu, P. S.; Pradeep, A.; Sunil, S.; Mothi, S. N.; Smith, D. M.; Kantor, R., 2017: Etravirine and Rilpivirine Drug Resistance. *AIDS Research and Human Retroviruses*., **33**, 567–575.
- Sathasivam Sivamalar, Thongadi Ramesh Dinesha , Jayaseelan Boobalan , Selvamurthi Gomathi , Ambrose Pradeep , Selvamuthu Poongulali , Sunil Solomon , Suniti Solomon , Pachamuthu Balakrishnan, S. S., 2016: Pattern of HIV-1 drug resistance mutations among patients failing thymidine analogue and non-thymidine analogue based first-line therapy in South India. *International Journal of Infectious Diseases*., Vol. 45p. 31.
- Senapati, S.; Shukla, R.; Tripathi, Y. B.; Mahanta, A. K.; Rana, D.; Maiti, P., 2018: Engineered Cellular Uptake and Controlled Drug Delivery Using Two Dimensional Nanoparticle and Polymer for Cancer Treatment. *Molecular Pharmaceutics*., **15**, 679–694.
- Shafer, R. W., 2006: Rationale and Uses of a Public HIV Drug-Resistance Database, **94305**, 51–58.

- Shaik, M.; Peng, H.; Lu, J.; Rits-volloch, S.; Xu, C.; Liao, M.; Chen, B., 2019: Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature*., **565**, 318–323.
- Shao, Y.; Williamson, C., 2012: The HIV-1 epidemic: Low- to middle-income countries. *Cold Spring Harbor Perspectives in Medicine*., **2**, 1–17.
- Sharma, A. L.; Singh, T. R.; Singh, L. S.; Devi, K. R., 2016: Prevalence of drug resistance associated mutations among the anti retroviral therapy exposed HIV-1 infected individuals in Manipur, Northeast India. *Current HIV Research*., **14**, 360–370.
- Sharp, P. M.; Hahn, B. H., 2001: Origins of HIV and the AIDS epidemic. *Cold Spring Harbor Perspectives in Medicine*., **1**, 1–22.
- Sharp, P. M.; Hahn, B. H., 2010: The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society B: Biological Sciences*., **365**, 2487–2494.
- Siepel, A. C.; Halpern, A. L.; Macken, C.; Korber, B. T. M., 1995: Computer Program Designed to Screen Rapidly for HIV Type 1 Intersubtype Recombinant Sequences, **11**, 1413–1416.
- Sinha, S.; Ahmad, H.; Shekhar, R. C.; Kumar, N.; Dar, L.; Samantaray, J. C.; Sharma, S. K.; Bhargava, A.; Pandey, R. M.; Mitsuyasu, R. L.; Fahey, J. L., 2012: Prevalence of HIV Drug Resistance Mutations in HIV Type 1 Isolates in Antiretroviral Therapy Naive Population from Northern India. *AIDS Research and Treatment*., **2012**, 1–6.
- Sinha, S.; Gupta, K.; Khan, N. H.; Mandal, D.; Kohli, M.; Das, B. K.; Pandey, R. M., 2018: Higher Frequency of HIV-1 Drug Resistance and Increased Nucleoside Reverse Transcriptase Inhibitor Mutations among the HIV-1 Positive Antiretroviral Therapy – Naïve patients Coinfected With Mycobacterium tuberculosis Compared With Only HIV Infection in In. *Infectious Diseases: Research and Treatment*., **11**, 1–7.
- Solomon, S.; Solomon, S. S.; Ganesh, A. K., 2006: AIDS in India. *Postgraduate Medical Journal*., **82**, 545–547.
- Sun, M.; Li, Y.; Zheng, H.; Shao, Y., 2016: Recent Progress toward engineering HIV-1-Specific Neutralizing Monoclonal Antibodies. *Frontiers in Immunology*., **7**, 1–8.
- Suzek, B. E.; Wang, Y.; Huang, H.; Mcgarvey, P. B.; Wu, C. H.; Consortium, U., 2015: UniRef clusters : a comprehensive and scalable alternative for improving sequence similarity searches, **31**, 926–932.
- Thevenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tuffery, P., 2012: PEP-FOLD : an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides, **40**, 288–293.
- Thirunavukarasu, D.; Udhaya, V.; Iqbal, H. S.; Umaarasu, T., 2016: Patterns of HIV-1 Drug-Resistance Mutations among Patients Failing First-Line Antiretroviral Treatment in South India. *Antiretroviral Therapy*., **15**, 261–268.

- Thorat, S. R.; N., D.; Chaturbhuj; Hingankar, N. K.; Velura Chandrasekhar, Rajasekhar Koppada, S. R. D.; Srikantiah, P.; Garg, R.; Kabra, S.; Haldar, P.; Dandu C.S. Reddy, Damodar Bachani, S. P. T.; Paranjape, R. S., 2011: Surveillance of Transmitted HIV Type 1 Drug Resistance Among HIV Type 1-Positive Women Attending an Antenatal Clinic in Kakinada, India. *AIDS Research and Human Retroviruses*., **27**, 1291–1297.
- Tilton, J. C.; Doms, R. W., 2010: Entry inhibitors in the treatment of HIV-1 infection. *Antiviral Research*., **85**, 91–100.
- Trott, O.; Olson, A. J., 2011: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem.*, **31**, 455–461.
- Turner, B. G.; Summers, M. F., 1999: Structural Biology of HIV HIV : Introduction. *Journal of Molecular Biology*., **285**, 1–32.
- Vella, S.; Schwartländer, B.; Sow, S. P.; Eholie, S. P.; Murphy, R. L., 2012: The history of antiretroviral therapy and of its implementation in resource-limited areas of the world. *Aids*., **26**, 1231–1241.
- Wade, R.; Gabdoulline, R.; Ludemann, S.; Lounnas, V., 1998: Electrostatic Steering and Ionic Tethering in Enzyme-Ligand Binding - Insights from Simulations. *Proc. Natl. Acad. Sci.*, **95**, 5942–5949.
- Wainberg, M. A.; Brenner, B. G., 2010: Role of HIV Subtype Diversity in the Development of Resistance to Antiviral Drugs. *Viruses*., **2**, 2493–2508.
- Ward, A. B.; Wilson, I. A., 2015: Insights into the trimeric HIV-1 envelope glycoprotein structure. *Trends in Biochemical Sciences*., **40**, 1–7.
- Watkins, D. A.; Jamison, D. T.; Mills, A.; Atun, R.; Danforth, K.; Glassman, A.; Horton, S.; Jha, P.; Kruk, M. E.; Norheim, O. F.; Qi, J.; Verguet, S.; Wilson, D.; Alwan, A.; Soucat, A., 2018: *Universal Health Coverage and Essential Packages of Care*. *Disease Control Priorities, Third Edition (Volume 9): Improving Health and Reducing Poverty*.
- Wilen, C. B.; Tilton, J. C.; Doms, R. W., 2012: HIV : Cell Binding and Entry. *Cold Spring Harbor Perspectives in Medicine*., **2**, 1–14.
- Xie, H.; Chen, L.; Zhang, J.; Xie, X.; Qiu, K.; Fu, J., 2015: A Combined Pharmacophore Modeling, 3D QSAR and Virtual Screening Studies on Imidazopyridines as B-Raf Inhibitors. *International Journal of Molecular Sciences*., **16**, 12307–12323.
- Xu, D.; Zhang, Y., 2011: Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. *Biophysj*., **101**, 2525–2534.
- Xu, X.; Huang, M.; Xiaoqin, Z., 2018: Docking-based inverse virtual screening: methods, applications, and challenges. *Biophysical reports*., **4**, 1–16.

- Yadav, G.; Kumar, P.; Kumar, Y.; Singh, P. K., 2018: Dolutegravir, Second Generation Integrase Inhibitor: A New Hope for HIV Patient. *European Journal of Molecular and Clinical Medicine*., **5**, 20–29.
- Yokoyama, M.; Naganawa, S.; Yoshimura, K.; Matsushita, S.; Sato, H., 2012: Structural Dynamics of HIV-1 Envelope Gp120 Outer Domain with V3 Loop. *PLoS ONE*., **7**, 4–7.
- Zhan, P.; Pannecouque, C.; De Clercq, E.; Liu, X., 2016: Anti-HIV Drug Discovery and Development: Current Innovations and Future Trends. *Journal of Medicinal Chemistry*., **59**, 2849–2878.
- Zhang, B.; Li, Y.; Zhang, H.; Ai, C., 2010: 3D-QSAR and molecular docking studies on derivatives of MK-0457, GSK1070916 and SNS-314 as inhibitors against Aurora B Kinase. *International Journal of Molecular Sciences*., **11**, 4326–4347.
- Zhou, H.; Zhou, Y., 2002: Distance-scaled , finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *protein science*., **11**, 2714–2726.

Software:

- Dassault Systèmes BIOVIA, Discovery Studio Visualizer2015, 17.2.0, San Diego: Dassault Systèmes, 2019
- The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
- Sybyl version 8.0, Tripos International, St. Louis, USA.