Design of inhibitors of the HIV-1 integrase core domain using virtual screening

Preetom Regon[1], Dhrubajyoti Gogoi[2]*, Ashok Kumar Rai[1], Manabjyoti Bordoloi[3] & Rajib Lochan Bezbaruah[2]

[1]Centre for Bioinformatics Studies, Dibrugarh University, Dibrugarh, Assam; [2]DBT-Bioinformatics Infrastructure Facility, Biotechnology Division, CSIR-North East Institute of Science and Technology (CSIR), Jorhat, Assam; [3]Natural Product Chemistry Division, CSIR-North East Institute of Science and Technology (CSIR), Jorhat, Assam; Dhrubajyoti Gogoi - Email: dhruba.bio.du@gmail.com; *Corresponding author

Abstract:
Acquired immunodeficiency syndrome (AIDS) is a disease of the human immune system caused by the human immunodeficiency virus (HIV). The integrase (IN) enzyme of HIV interacts with several cellular and viral proteins during the integration process. Thus, it represents an appropriate target for antiretroviral drugs (ARVs). We performed virtual screening of database compounds and designed analogues using Elvitegravir (EVG) as a standard compound. The 378 screened compounds were retrieved from ZINC, ChemSpider, PubChem, and ChemBank Chemical Databases based on chemical similarity and literature searches related to the structure of EVG. The Physiochemical properties, Bioactivity, Toxicity and Absorption, Distribution, Metabolism and Excretion of Molecules (ADME) of these compounds were predicted and docking Experiments were conducted using Molegro Virtual Docker software. The docking and ADME suggested very significant results in regard to EVG. The MolDock and Rerank scores were used to analyze the results. The compounds ZINC26507991 (-84.22), Analogue 9 (-68.49), ZINC20731658 (-66.79), ZINC00210363 (-43.44) showed better binding orientation with IN receptor model with respect to EVG (182.52). The ZINC26507991 has showed significant ADME result.

Background:
HIV Integrase (IN) is an enzyme that enables its genetic material to be integrated into the DNA of the infected cell. HIV integrase (IN) is an essential enzyme in HIV replication and an important target for drug design **[1, 2, 3]**. HIV-1 IN is comprised of three domains: the catalytic core, the C-terminal domain, and the N-terminal domain. These designations are based on the susceptibility of the linker regions to proteolysis, functional studies, and the structures of the domains, which have been individually determined by x-ray crystallography and NMR **[4]**. The N-terminal domain of HIV-1 IN contains a conserved pair of His and Cys residues, a motif similar to the zinc-coordinating residues of zinc fingers **[4, 5]**. The C-terminal domain binds DNA non-specifically. Because the sites of

integration into target DNA are relatively nonspecific, it has been suggested that this domain may interact with target DNA. **[6, 7]**. The catalytic core domain contains the invariant triad of acidic residues, the D, D-35-E motif, residues Asp64, Asp116, and Glu152 **[8]**.

The Elvitegravir (EVG) is an ARV (antiretroviral) drug, which inhibit the viral replication by preventing integration of viral DNA into the host cell **[9]**. The drug was approved by U.S. Food and Drug Administration on August 27, 2012 for people starting HIV treatment for the first time as part of the fixed dose combination known as Stribild **[10]**. In this current work, an endeavour was made for identification of EVG-like compound in chemical libraries using virtual screening

approach. The docking results suggested good binding interaction of five candidate compounds against IN receptor.
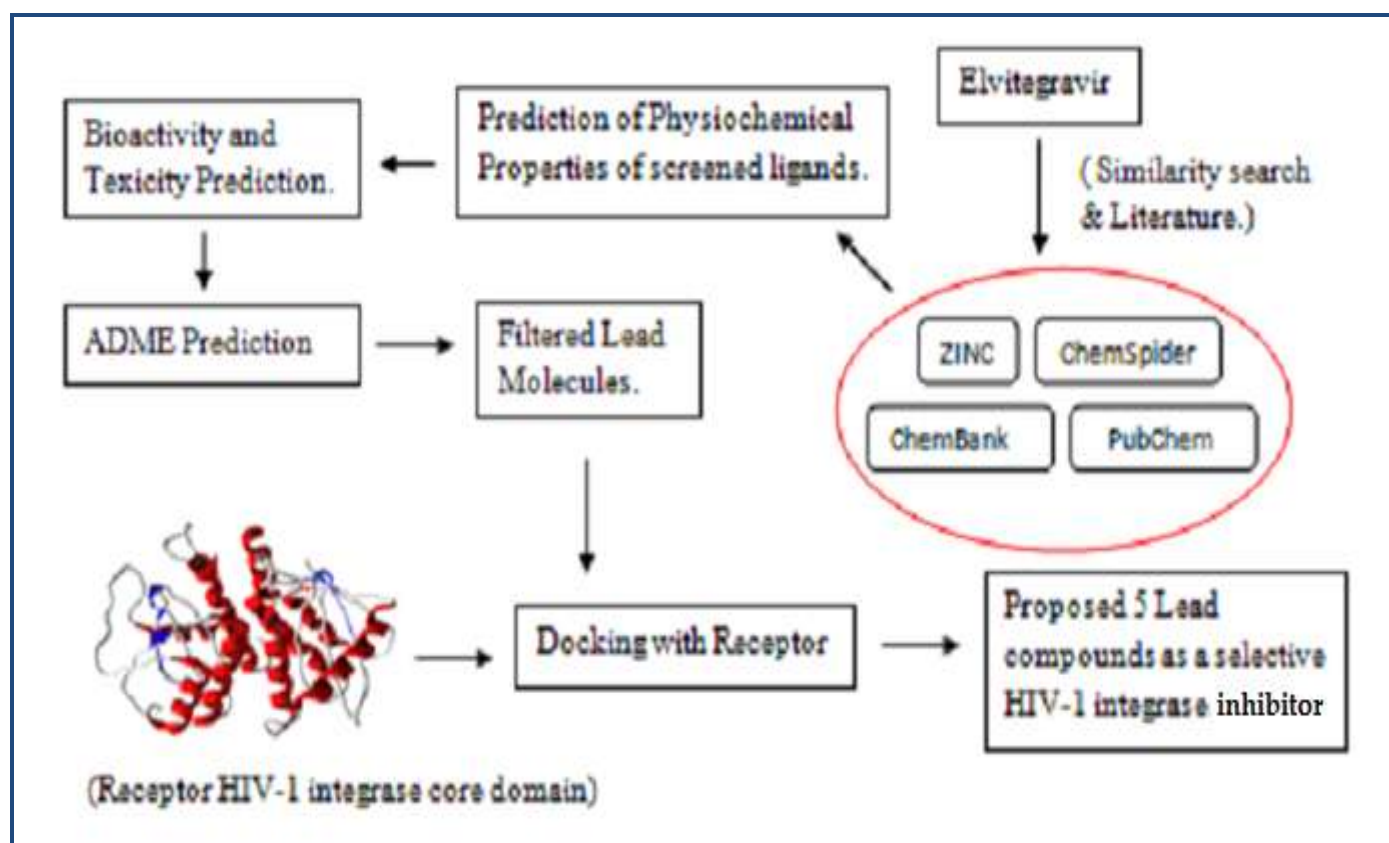

Figure 1: The flow chart showing different step of the methodology in Investigation

Methodology:
The methodologies applied in this investigation are as follows and shown in **(Figure 1).**

Target selection
The three dimensional Crystal structure of target HIV-1 IN (PDB ID: 3L3U) was retrieved from Protein Data Bank (http://www.pdb.org/) **[11].**

Analogue Design for Elvitegravir
An analogue is a drug whose structure is related to that of another drug, but whose chemical and biological properties may differ. In drug development, large series of structural analogs of an initial lead compound are created and tested as part of a structure-activity relationship study. Using EVG as a reference, analogues were designed. The structures were design by changing the functional groups keeping the backbone of the EVG fixed. The functional groups used to design analogues are F, Cl, NH2, OH etc. The ChemBio Office was used to drawn the different analogue of EVG.

Compound Retrieval and Similarity Search
The screening was performed against different chemical databases. These databases are ZINC (zinc.docking.org/), ChemSpider (www.chemspider.com/), PubChem (www.ncbi.nlm.nih.gov/pccompound) and ChemBank (chembank.broadinstitute.org/). The EVG -like compounds were retrieved from these chemical databases based on fingerprint, substructure, Physical properties, etc. Most of the compounds were found in the ZINC database. ZINC is a free

database of chemical compounds that contains 13 million compounds **[12, 13].** MolSoft ICM-Browser was used to perform Chemical Similarity Search.

Physiochemical Property Calculations
Chemical behaviour of molecules through chemical structures is an integral part of drug discovery process. In vivo pharmacokinetics parameters, such as absorption, distribution, metabolism and excretion are strongly influenced by the physiochemical properties of drug. Based on Lipinski's rules **[14]**, physiochemical properties were calculated for each of the screened molecule and Analogues. We selected only those molecules whose properties do not violate the Lipinski rule of five for subsequent steps. The remainer of the molecules that violated the Lipinski's rules was discarded from screening.

Bioactivity and Toxicity Prediction
Bioactivity of molecules describes compounds that have a specific effect on system- inhibiting, activating, or otherwise modulating a cellular process. In drug discovery processes, one cannot designate a molecule as a drug only by predicting its activity. It is important to predict or examine the toxicity of the molecules as well. If a molecule is biochemically active, but contains some toxic effects, then one cannot accept the molecule as a drug. One can designate a molecule as a drug if it is biochemically active and as well as non-toxic. Therefore, during the drug discovery process, prediction of activity and toxicity of molecules is very important. Using the PASS software, we predicted all those molecules that passed Lipinski Filtration. PASS (Prediction of Activity Spectra of Substances) predicts

BIOINFORMATION

drug-likeness as well as toxicity of molecules **[15].** This study selected only those molecules that have the drug-likeness property of 50% or above. During our search process, out of 211 molecules, 48 molecules showed toxicity. The molecules that showed toxicity (teratogen, carcinogenic and embryo-toxic) were rejected from the screening process.

Energy minimization

The Energy minimization of molecules was carried out using software ChemBio 3D Ultra. The potential energy calculated by summing the energies of various interactions is a numerical value for a single conformation. This number can be used to evaluate a particular conformation, but it may not be a useful measure of a conformation because it can be dominated by a few lower grade interactions. For instance, a large molecule with an optimum conformation for nearly all atoms can have a large overall energy because of a single bad interaction. For example, two atoms too near each other in space that has a huge Van der Waals repulsion energy. It is often preferable to carry out energy minimization on a conformation to find the best nearby conformation. Energy minimization is usually performed by gradient optimization: atoms are moved to reduce the net forces among them. The energies of selected ligands were predicted in ChemBio Office Ultra using MM2 module before performing docking.

Active Site Prediction and Cavity Detection

The region of a protein that interacts with a ligand is generally referred to as the "active site." The active site generally lies on the surface of the protein. In some cases, the active site is buried within the protein. Residues with reactive groups (Asp, Glu, Ser, Cys, His, Lys, Arg) tend to be abundant in protein active sites. The Ser-His-Asp (sometimes Ser-His-Glu) "catalytic triad" is a motif commonly found in enzyme active sites. We used Molegro Virtual Docker for protein active site prediction and cavity detection. Cavity detection is often the first step for functional analysis, since binding sites in proteins usually lie in cavities. The predicted cavities are shown in **Table 1 (see supplementary material).**

Docking Studies and Interpretation

During docking, at first, the molecules were prepared and bonds, bond orders, explicit hydrogen, flexible torsion were assigned if they were missing by the Molegro Virtual Docker program to both the protein and ligands. The analysis of side chain flexibility affords insights valuable to improve docking algorithms and can provide an index of amino-acid side-chain flexibility, potentially useful in molecular docking **[16].** The search algorithm is taken as Moldock SE and numbers of runs are taken and maximum interactions were 1000. The population size was set at 50 with energy threshold of 100. The potential binding site were predicted and docking performed with cavity-1 having the having a volume of 79.36 A3 cube and a surface area of 259.84 A2 square. The screen compounds shown in **(Table 1)** along with EVG were imported to Molegro Virtual Docker workspace. The bonds flexibility of the ligands was set and the side chain flexibility of the amino acids in the binding cavity was set with a tolerance of 1.10 and strength of 0.90 for docking simulations. RMSD threshold for multiple cluster poses was set at <2.00 Å. The docking algorithm was set at a maximum iteration of 1500 with a simplex evolution size of 50 and a minimum of 20 runs **[17-19].**

ADME Prediction

The ADME properties were predicted using PreADMET (preadmet.bmdrc.org) server to know whether the compounds have the potential of adverse effect in human.

Results & Discussion:

The virtual screening of chemical libraries for HIV integrase inhibitor resulted in four compounds: ZINC26507991, ZINC01585513, ZINC01459700, and ZINC20731658 **Table 2 (see supplementary material).** These compounds were reported from ZINC compounds and showed significant resemblance with the reference drug EVG with better ranking scores. The Docking of these compounds including one analogue of EVG have shown good rerank and Moldock scores. The ZINC26507991 shows highest rerank score of -84.22 whereas EVG has shown -22.93. The other compounds as in **Table 2** shows moderately good ranking with respect to EVG. In second and third cavity, the EVG not showing any significant interaction, on the other hand ZINC26507991, ZINC01585513 predicted better score and suitable interaction with receptor. The ADME prediction results of these five candidate compounds possess all the minimum characteristics of a drug. The predicted scores were shown in Table 3. The results of Docking and ADME reflected the novelty of these candidate compounds as IN inhibitors. The in vitro and in vivo experiments may be recommended with these molecules as future candidate drug of AIDS.

Conclusion:

The virtual screening and analogue design of EVG revealed five compounds with more potent activity than EVG: ZINC26507991, ZINC01585513, ZINC01459700, ZINC20731658 and Analogue-9. The ADME prediction also showed very encouraging results. Hence, these compounds may be tested in vitro as future IN inhibitors.

References:

[1] http://www.ncbi.nlm.nih.gov/books/NBK19376/
[2] Hindmarsh P & Leis J, *Microbiol Mol Biol Rev*. 1999 **63**: 836 [PMID: 10585967]
[3] Engelman A & Craigie R, *J Virol*. 1992 **66**: 6361 [PMID: 1404595]
[4] Yang W & Steitz TA, *Structure* 1995 **3**: 131 [PMID:7735828]
[5] Rice P *et al. Curr Opin Struct Biol*. 1996 **6:** 76 [PMID: 8696976]
[6] Lodi PJ *et al. Biochemistry* 1995 **34:** 9826 [PMID: 7632683]
[7] Eijkelenboom AP *et al. Nat Struct Biol*. 1995 **2**: 807 [PMID: 7552753]
[8] Jenkins TM *et al. Proc Natl Acad Sci U S A*. 1995 **92**: 6057 [PMID: 7597080]
[9] Klibanov OM, *Curr Opin Investig Drugs.* 2009 **10**: 190 [PMID: 19197797]
[10] Sax PE *et al. The Lancet* 2012 **379:** 2439 [PMID: 22748591]
[11] Wielens J *et al. FEBS Lett*. 2010 **584**: 1455 [PMID: 20227411]

BIOINFORMATION

[12] http://as.wiley.com/WileyCDA/WileyTitle/productCd-0471672009.html

[13] Irwin JJ *et al. J Chem Inf Model.* 2005 **45**: 177 [PMID: 15667143]

[14] Lipinski CA *et al. Adv Drug Deliv Rev.* 2001 **46**: 3 [PMID: 11259830]

[15] Filimonov DA *et al. Eksp Klin Farmakol.* 1995 **58**: 56 [PMID: 7773095]

[16] Shoichet Brian K, *Nature* 2004 **432:** 862 [PMID: 15602552]

[17] Gogoi D *et al. Bioinformation* 2012 **8**: 812 [PMID: 23139590]

[18] Najmanovich R *et al. Proteins* 2000 **39**: 261 [PMID: 10737948]

[19] Thomsen R *et al. J Med Chem.* 2006 **49:** 3315 [PMID: 16722650]

BIOINFORMATION

Supplementary material:
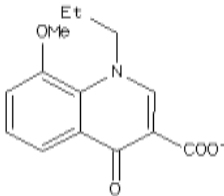
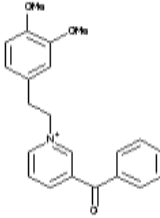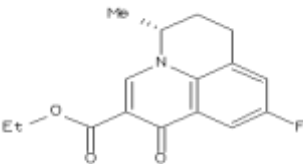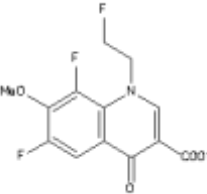**Table 1:** Best ligands from ZINC database

ZINC26507991

ZINC01585513

ZINC01459700

ZINC20731658

Table 2: Docking result of candidate compounds (analyzed based on Rerank Score, MolDock Score and H Bond Energy Score)

ZINC ID	Molecular formula	MolDock score	Rerank score	H bond
ZINC26507991	$C_{14} H_{15} N O_4$	-97.94	-84.22	-6.52
Analogue 9	$C_{23} H_{22} Br Cl F N O_4$	-103.07	-68.49	-13.31
ZINC20731658	$C_{13} H_{10} F_3 N O_4$	-79.94	-66.79	-7.15
ZINC00210363	$C_{12} H_9 N O_3$	-66.24	-43.44	-2.00
Elvitegravir	$C_{23} H_{23} Cl F N O_5$	-22.93	182.52	-0.26
ZINC01682662	$C_{15} H_{19} N O_3$	-13.10	294.03	-5.95
Cavity Name	**Volume ($Å^3$)**	**Surface ($Å^3$)**		
1	79.36	259.84		
2	68.096	244.48		
3	43.008	148.48		
4	42.496	154.88		
5	35.328	140.80		

Table 3: ADME result of ZINC compounds

Compounds	HIA (%)	C2C (nm/sec)	MDCK (nm/sec)	Logkp (cm/hour)	PPB (%)	BBBP (c.brain/c.blood)
ZINC26507991	96.70	0.38	46.40	-3.63	71	1.22
ZINC20731658	96.21	22.00	79.30	-4.15	62	1.16
ZINC00210363	96.87	20.90	25.10	-4.07	93	0.86
ZINC01682662	97.91	17.34	57.30	-3.59	63	1.57

NB: HIA= Human Intestinal Absorption, C2C=CaCo2 cell, MDCK=Madin darby canine kidney cell, PPB=plasma protein binding, BBBP=blood brain barrier permeability.