Virtual Screening of Novel HIV-RT Inhibitors using Zinc Database

P.V.R.D. Prasada Rao
Associate Professor , Dept of CSE, K.L.University
Vijayawada, Andhra Pradesh, India

K.Lakshmi Sunitha
Asst Professor , KLUBS, K.L.University
Vijayawada, Andhra Pradesh, India

K.Srinivas
Research Scholar, JNTUK

A. Govardhan, PhD.
Director of Evaluation , JNTU Hyderabad, India

ABSTRACT
HIV–1 (human immunodeficiency virus type–1) is the pathogenic retrovirus and causative agent of AIDS, When viral RNA is translated into a polypeptide sequence, it is assembled in a long polypeptide chain, which includes several individual proteins namely, reverse transcriptase, protease, integrase, etc. Before these enzymes become functional they have to be cut from polypeptide chain, The dipyrido diazepinone Nevirapine is a potent and highly specific inhibitor of the reverse transcriptase (RT) from human immunodeficiency virus type 1 (HIV-1). In this paper, we implemented better than existing system by virtual screening analysis of HIV-RT from PDB database versus chemical compounds from ZINC database using eHiTS software. Using molecular constraint search, 884 ligands were extracted and docking analysis resulted in 59 best hits.

Keywords: HIV, reverse transcriptase, protease , virtualscreening, RMSD, Zincdatabase, ehits , docking, RCPlot, Clustering

1. INTRODUCTION
Human immunodeficiency virus (HIV) is a retrovirus that causes acquired immunodeficiency syndrome (AIDS), a condition in humans in which the immune system begins to fail, leading to life-threatening opportunistic infections [1]. Acquired immune deficiency syndrome (AIDS) is a formidable pandemic that is still wreaking havoc worldwide. The causative moiety of the disease is human immunodeficiency virus (HIV), which is a retrovirus of the lentivirus family . The following enzymes reverse transcriptase, protease and integrase encoded by the gag and gag–pol genes of HIV play an important role in the virus replication cycle. Among them, viral reverse transcriptase (RT) catalyzes the formation of proviral DNA from viral RNA, the key stage in viral replication. Its central role in viral replication makes RT a prime target for anti–HIV–therapy, two main categories of HIV RT inhibitors have been discovered to date. The first category of inhibitors is nucleoside analogues (e.g., AZT, 3TC, ddI, ddC) and the second category of inhibitors is nonnucleoside analogues. Nevirapine, delaviridine and efavirenz are the only nonnucleoside reverse transcriptase inhibitors (NNRTI) that have received regulatory approval with several NNRTIs (MKC442, Troviridine, S–1153/ AG1549. PNU142721, ACT and HBY1293/GW420867X) are currently undergoing clinical trials. Efavirenz 9 was the first potent anti–HIV drug to be approved by FDA and studies have shown that efavirenz penetrates into a common viral sanctuary in cerebrospinal fluid,. The therapeutic efficacy of the drug is mainly restricted due to the development of viral resistance associated with mutations

HIV is different in structure from other retroviruses [3]. It is about 120 nm in diameter and roughly spherical. It is composed of two copies of positive single- stranded RNA that codes for the virus's nine genes enclosed by a conical capsid composed of 2,000 copies of the viral protein p24. The single-stranded RNA is tightly bound to nucleocapsid proteins, p7 and enzymes needed for the development of the virion such as reverse transcriptase, proteases, ribonuclease and integrase [4]. A matrix composed of the viral protein p17 surrounds the capsid ensuring the integrity of the virion particle. This is, in turn, surrounded by the viral envelope which is composed of two layers of fatty molecules called phospholipids taken from the membrane of a human cell when a newly formed virus particle buds from the cell [5] ,dipyridodiazepinone Nevirapine is a potent and highly specific inhibitor of the reverse transcriptase (RT) from human immunodeficiency virus type 1 (HIV-1). It is a member of an important class of non nucleoside drugs [8] that appear to share part or all of the same binding site on the enzyme but are susceptible to a variety of spontaneous drug-resistance mutations [9].In this study, a virtual screening routine was reported by utilizing 1VRT from protein data bank and screening based on docking ZINC database ligands for effective HIV-RT inhibitor.

2. MATERIALS AND METHODS
Traditional drug discovery is time consuming and expensive in modern drug discovery computational methods are generally involved in identifying and modifying lead compounds, most commonly used is molecular docking in docking involves two processes 1a) Geometric sampling of potential ligand / protein binding modes b) Scoring , Usually using an equation and specific parameters to estimate a ligands binding affinity.

Computational molecular docking comprises of sampling and scoring of Protien –Ligand complexes to predict the binding orientation of a given ligand ,Virtual screening docks a large number of different ligands to the target protein, seeking to predict the relative affinity and activity of the ligands , docking is performed by a computer program to generate computer representations of ligand binding modes in the binding site of the proteins and we will use RMSD is metric used to measure the performance of a docking algorithm

2.1 Steps in PDB data processing:

Data processing consists of data deposition, annotation and validation. These steps are shown in Figure1. Data (atomic coordinates, structure factors and NMR restraints) may be submitted via email or via the AutoDep Input Tool (ADIT) developed by the RCSB.

Step-1: *After a structure has been deposited using ADIT, a PDB identifier is sent to the author automatically. The entry is then annotated and validated followed by checking errors and inconsistencies in files.*
Step-2: *The annotated file, validation information is sent back to the depositor.*
Step-3: *Steps-2 and 3 may be repeated after revised file submitted by author.*
Step-4: *Entry released in PDB for distribution.*

Entries before release are categorized as
'in processing' (PROC),
'in depositor review' (WAIT),
'to be held until publication' (HPUB) or
'on hold until a depositor-specified date' (HOLD).

Validation

Two types of validation are performed for PDB files.

a. Structure Validation: for assessing the quality of deposited atomic models.
b. Experimental Validation: for assessing how well these models fit the experimental data.

The following checks are performed under data validation.

1. *Covalent bond distances and angles*
2. *Stereochemical validation*
3. *Atom nomenclature*
4. *Close contacts*
5. *Ligand and atom nomenclature*
6. *Sequence comparison*
7. *Distant waters*

For docking we need protein and ligand here proteins are taken from protein database(PDB), Ligands are taken from ZINC Database in protein databank there are number of 3D structures and 3million compounds are available in zinc database , so question is which should be chosen for docking, but in HIV there are 12 structures, out of these 12 structure , which is best structure some criteria should be there to decide , with the help of Ramachandran plot we can judge which is best and which is effective for docking ,

RC Plot specifies the quality of protein structure which is best than other structures here number of disallowed regions should not be there ,80% should be there in most allowed regions , *1VRT* is found to be best , for this we need one bound ligand (drug Nivarapine) like nivarapine there are many drugs but their structures are not good like nivarapine

Complete structure was searched in zinc database the structure of nevirapine was drawn by using a tool in zinc database. There were no similar structures available in zinc database, Hence, property based search was used to screen ZINC database for similar compounds.

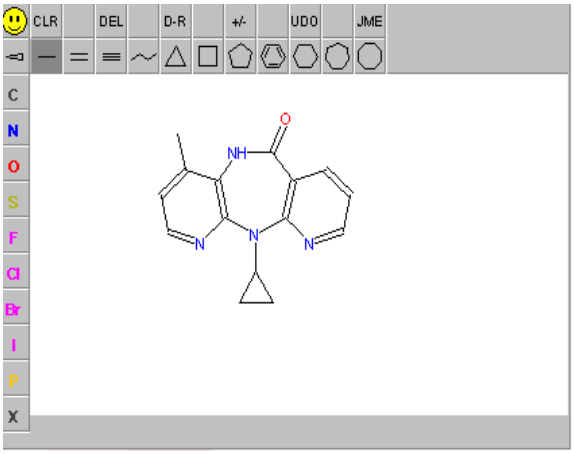

Figure-1:Nivarapine Structure and complete structure search in ZINC database.

Table1. For different Protein ID's, parameters for RC plot are shown below

PDB ID	No of Residues in most favoured regions	No of Residues in additional allowed regions	No of Residues in generously allowed regions	No of Residues in disallowed regions
1TKT	696	86	2	2
1TKZ	696	86	2	2
1SLT	189	39	0	0
1SLU	189	39	0	0
1SLW	189	39	0	0
1JLB	708	114	5	2
1JLC	652	152	6	0
1JLF	605	147	7	1
1LW2	681	108	6	1
1JKH	694	118	9	3
1S6P	599	235	13	2
1S6Q	659	177	19	2

1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	696	88.1%*
Additional allowed regions	[a,b,l,p]	86	10.9%
Generously allowed regions	[~a,~b,~l,~p]	6	0.8%
Disallowed regions	[XX]	2	0.3%*
Non-glycine and non-proline residues		790	100.0%
End-residues (excl. Gly and Pro)		9	
Glycine residues		49	
Proline residues		67	
Total number of residues		915	

After docking we got 59 best , Once again by Python programming we reduced to 39 best , from these we have chosen top 3 ligands were chosen with affinity around 8.5 , which is best compound better than Nivarapine (6). 1VRT is best new identified compound till now we got computational structure then this has to go through experimental structure

1VRT was downloaded from Protein Data Bank and used as receptor structure for virtual screening program. Chemical library, ZINC database and the docking program eHiTS (electronic High Throughput Screening) was employed in the study, ZINC database has over 4.6 million compounds in ready-to-dock formats. The database was screened for compounds with either similar geometrical features or Lipinski compliant [10]. The physico-chemical properties such as log P value, H-bond donors, H-bond acceptors, molecular weight and rotational bonds, of nevirapine ligand were calculated using Tsar software. , eHiTS has a novel flexible ligand docking method and generates poses that avoid severe steric clashes between receptor and ligand. The algorithm is exhaustive on the conformations and the conformers are compatible with the steric and chemistry constraints.

3. RESULTS AND DISCUSSION

Before screening ZINC database, the ehits docking protocol was validated. 1VRT protein bound ligand nevirapine was docked into the binding pocket and the RMSD (Root Mean Square Deviation) of the docked pose was 0.55 A° (Figure 2) with co-crystallized ligand, indicating that the parameters for docking simulation are good in reproducing the X-ray crystal structure.

A structure based search using structural features that are similar to nevirapine resulted in no hits. Hence, a molecular constraint search was employed using physico-chemical properties of nevirapine which resulted in 884 ligands. All these ligands are found to be Lipinski compliant. All compounds are docked and the binding compatibility of each pose with the receptor was evaluated based on docked energies. The technique used in the study identified diverse geometrical ligands but specific in displaying binding compatibilities with the receptor active site region. From the screening analysis of 884 ligands, a total of 59 molecules resulted in high dock scores (>-6.58 to -8.179 kcal.mol) than the original nevirapine molecule (-6.5818 kcal/mol). The ZINC id's along with binding energy scores for top three molecules are given in Table 2.

Table 2: The best three ZINC hits

S. No.	ZINC ID	e-hits score(kcal/mol)
1	1VRT bound Nevirapine	-6.582
2	ZINC04923148	-8.179
3	ZINC05442451	-7.886
4	ZINC04923002	-7.424

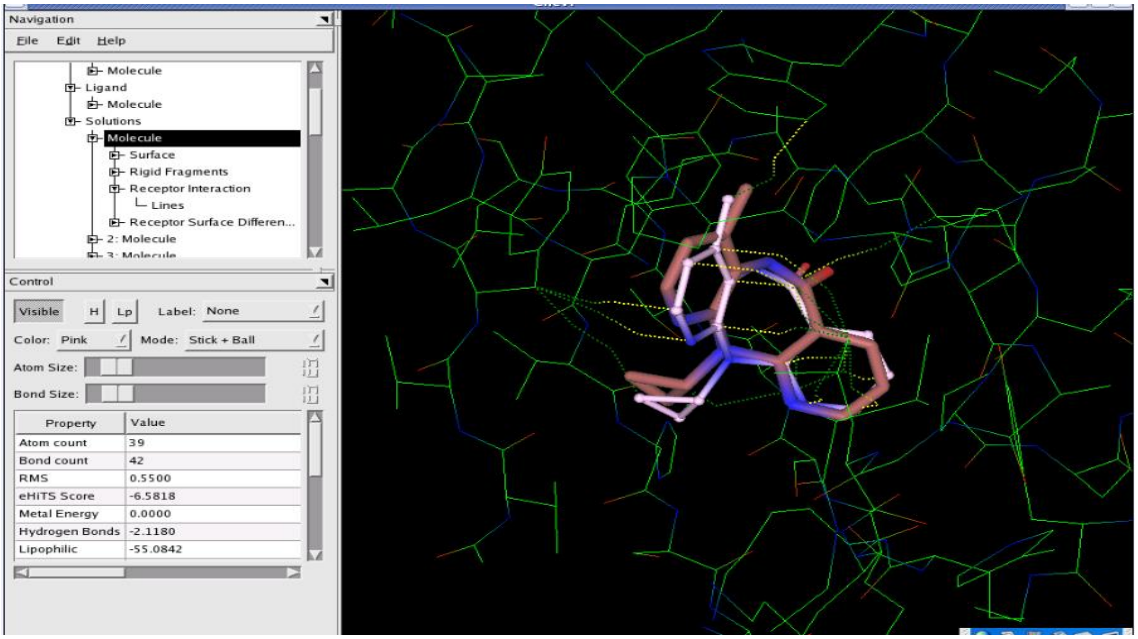

Figure 2: 1VRT bound nevirapine (-6.5818 kcal/mol) with active site residue interactions displaying protein structure in background and RMSD 0.55 A°.

Figure 2 shows the image of original ligand bound within active site region of 1VRT protein with a conformer representing the RMSD value of 0.55 A^o with e-hits score of -6.5818 kcal/mol. Therefore, when a screening analysis is performed against 1VRT protein, any such molecule which binds to 1VRT with a score better than -6.5818 kcal/mol is of prime interest in this screening schedule. Therefore, from the analysis, as given in Table 2, it became evident that there existed about three best ligand conformers with netter binding compatibilities than 1VRT bound ligand. The number of interacting residues for nevirapine, ZINC04923148, ZINC05442451, and ZINC04923002 molecules are given below.

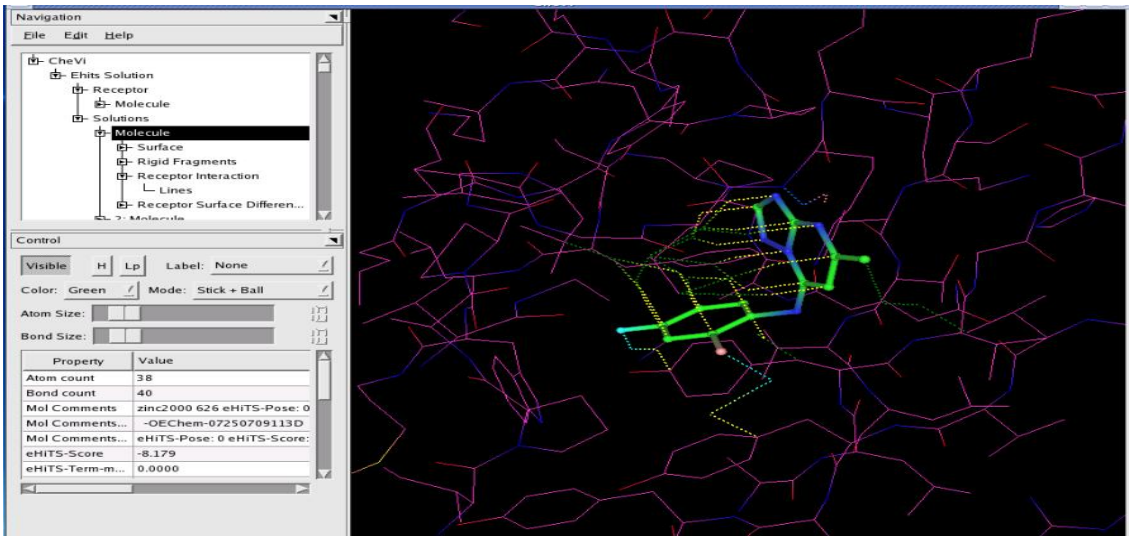

Figure 3: 1VRT with docked pose of ZINC04923148 showing e-hits score of -8.179 kcal/mol.

ZINC04923148 represented better orientation (Figure 3, -8.179 kcal/mol) with possible H-bond interactions being 19 and the Lipinski data are: H-bond donors 1, H-bond acceptors 5, molecular weight 277.69, logp 3.17 and number of rotatable bonds 2, respectively. From the table it is also evident that the molecule exhibited more number of interactions than the original ligand. In order to study the probable reason behind difference in number of interactions, the residue wise atomic

interactions for each molecule was evaluated. From the interaction list, individual interactions between atomic coordinates of 1VRT active site residues and ZINC ligand displayed the high score for 18th interaction showing TYR-188 residue CZ atom interaction with ligand. This was mainly due to the Lone electron pair of a halogen atom of ligand and Pi electron of an aromatic ring of 1VRT. The best interaction from 19 interacting atoms between receptor and ligand of ZINC04923148 was:

Interaction 18
==========

Receptor SPT	[16] Pi electron of an aromatic ring
Ligand SPT	[21] Lone electron pair of a halogen atom (F,Cl,I)
Receptor angle	11.49
Ligand angle	11.85
Dihedral angle	162.20
Distance	3.9245
Score	-2.5779
Receptor atom	Index:206 Residue: CZ TYR-188 Type:C
Ligand atom	Index:18 Type:F

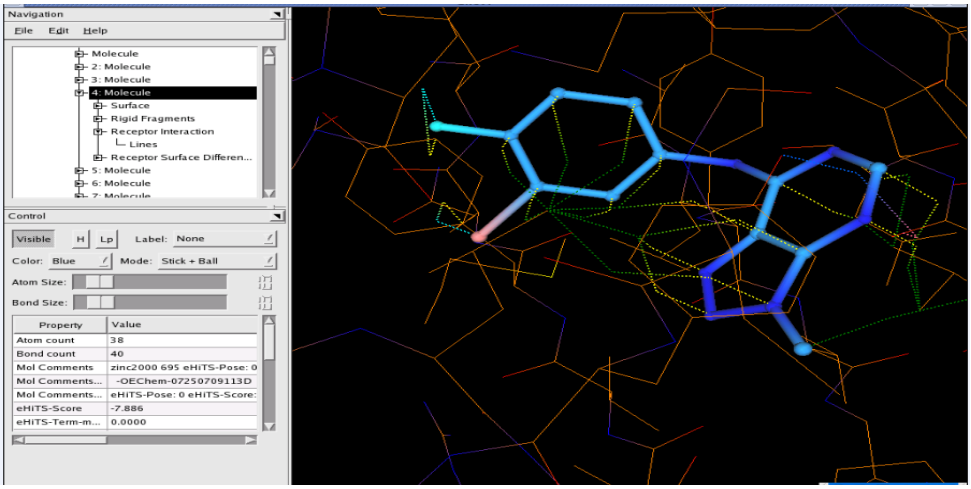

Figure 4: 1VRT vs ZINC05442451 showing e-hits score of -7.886 kcal/mol.

ZINC05442451 ligand (Figure 4, -7.886 kcal/mol) with about 19 interactions and the Lipinski data are: H-bond donors 1, H-bond acceptor 6 and molecular weight 278.678, logp 2.63 and number of rotatable bonds 2, respectively. Individual interactions between atomic coordinates of 1VRT active site residues and ZINC ligand displayed the high score for 15th interaction showing TYR-181 residue CG atom interaction with ligand. This was mainly due to the Lone electron pair of a halogen atom of ligand and Pi electron of an aromatic ring of 1VRT. The best interaction from 19 interacting atoms between receptor and ligand of ZINC05442451 was:

Interaction 15
==========

Receptor SPT	[16] Pi electron of an aromatic ring
Ligand SPT	[21] Lone electron pair of a halogen atom (F,Cl,I)
Receptor angle	6.76 , Ligand angle 25.22
Dihedral angle	166.51 , Distance 3.2061
Score	-1.7645
Receptor atom	Index:154 Residue: CG TYR-181 Type:C
Ligand atom	Index:15 Type:Cl

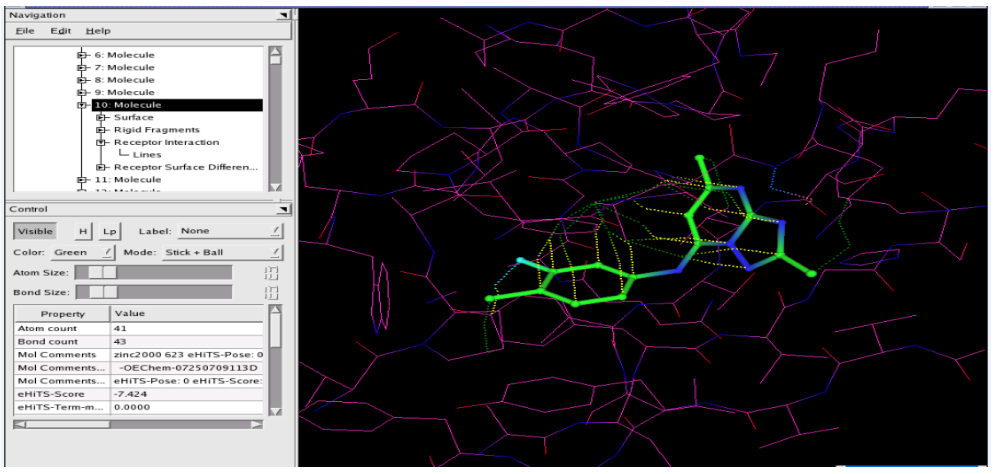

Figure 5: 1VRT protein with ZINC04923002 ligand with e-hits score of -7.424 kcal/mol

ZINC04923002 ligand (Figure 5, -7.424 kcal/mol) with about 20 interactions and the Lipinski data are: H-bond donors 1, H-bond acceptor 5 and molecular weight 271.299, logp 3.02 and number of rotatable bonds 2, respectively. From the interaction list, 20th interaction showing LYS-101 residue N atom interaction with ligand was found to be the best, mainly due to the strong (primary) hydrogen bond acceptor lone pair atom of ligand and Strong (primary) hydrogen bond donor H (polar-atom-H) of 1VRT. The best interaction from 20 interacting atoms between receptor and ligand of ZINC04923002 was:

Interaction 20
==========
Receptor SPT [3] Strong (primary) hydrogen bond donor H (polar-atom-H)
Ligand SPT [7] Strong (primary) hydrogen bond acceptor lone pair
Receptor angle 29.80
Ligand angle 41.16
Dihedral angle 72.43
Distance 2.6059
Score -2.4245
Receptor atom Index:48 Residue: N LYS-101 Type:N
Ligand atom Index:17 Type:N

In Nevirapine, the major active site residues participated in interactions are: Leu100, Val106, Tyr181, Trp229 and Leu234. Whereas, in ZINC04923148, the majority residue interactions are formed by: Leu100, Lys101, Val179, Tyr181, Tyr188 and Leu234. From the above data, it is evident that the high dock score obtained for ZINC04923148 was due to new residue interactions formed by Lys101 and Tyr188 respectively.

3.1 SUPERPOSED ORIGINAL LIGAND

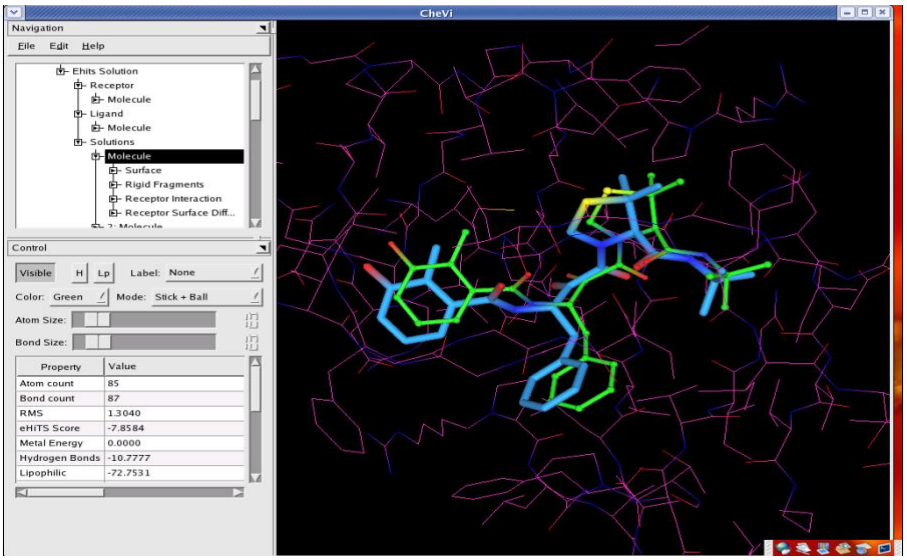

Figure 6: 1MRX original ligand superposed image showing h-bond score of -10.777 kcal/mol

In Fig 6:1MRX ligand shows the image of original ligand bound within active site region of 1MRX protein with a conformer representing the RMSD value of 1.07A^o. The RMSD value obtained for the original 1MRX ligand is within 2.0 A^o limit. Figure 1 shows the original ligand conformer displaying a h-bond score of –10.777 kcal/mol. Therefore, when a screening analysis is performed against 1MRX protein, any such molecule which shows Hydrogen bond interactions greater than 1MRX ligand is regarded as the best ligand than the original one. From this screening analysis, as given in table 6 it became evident that there existed about 6 best ZINC/PubChem/NCI ligands which represent higher score than bound ligand in 1MRX protein. That means these 6 ligands would act as inhibitors against1MRX protein and such screening runs form the first step when lakhs of ligand libraries are available such as in ZINC/PubChem/NCI database or others.

3.2 Filtering Data by Clustering analysis

Bioinformatics and data mining provide exciting and challenging research and in many application areas for computational science. Bioinformatics is the science of managing, mining, and interpreting information from biological sequences and structures. Advances such as genome-sequencing initiatives, microarrays, proteomics, and functional and structural genomics have pushed the frontiers of human knowledge.

Clustering is defined as the process of dividing patterns into several groups without prior knowledge of which pattern falls into which group and each group is called a cluster. In other words, clustering is dividing patterns, such as $X = \{x_1, x_2, ..., x_n\}$, into K groups, such a, $C = \{C_1, C_2, ..., C_k\}$ in such a way that the following conditions are satisfied:
$C_1 \cup C_2 \cup ... \cup C_k = X$, $C_i \neq \emptyset$, $Ci \cap Cj = \emptyset$ for $i \neq j$.

Generally, clustering algorithms are of two categories 1. Hierarchical and 2. Partitioning algorithms . In the hierarchical method, advantages of the hierarchical algorithm is that neither initialization nor determination of the number of cluster is required here we will be using tree structure called dendrograms . Hierarchical method have three shortcomings; first, since only the local neighbors are considered at each

stage, here in our work we used agglomerative approach for clustering analysis, this was carried out by clustering compounds based on similar or nearer properties such as 'MW Vs HBA', 'MW Vs HBD', 'MW Vs LogP', 'MW Vs RB', 'HBA Vs HBD', 'HBA Vs LogP', 'HBA Vs RB', 'HBD Vs LogP', 'HBD Vs RB', 'LogP Vs RB' respectively. In order to perform the task, those molecules which have any two similar properties are segregated to form clusters , For example, molecular weight versus hydrogen bond acceptors, and others. A python program was written to cluster sets of compounds that share better scores than original co-crystallized nevirapine (-6.58 kcal/mol). Initially, clusters are generated using python program. From the screening analysis of 884 ligands, a total of 59 molecules resulted in high dock scores (>-6.58 to -8.179 kcal.mol) than the original nevirapine molecule (-6.5818 kcal/mol), this was found by computational analysis then by experimental analysis effective drug can be formulated for HIV through reverse transcriptase.

4. CONCLUSION

In this work, we recognized that the best molecule by docking ligand and PDB protein , 1VRT resulted in about 59 such molecules from 4.6 million molecule ZINC database. 1VRT bound co-crystallized ligand displayed an e-hits score of -6.5818 kcal/mol. Screening procedures carried out using selected criteria resulted in top three best molecules, represented by ZINC04923148, ZINC05442451 and ZINC04923002 with e-hits scores of -8.179, -7.886 and -7.424 kcal/mol respectively.

5. REFERENCES

[1] HIV reverse transcriptase: structural interpretation of drug resistant genetic variants from India. Kandathil AJ, Joseph AP, Kannangai R, Srinivasan N, Abraham OC, Pulimood SA, Sridharan G.Bioinformation. 2009 Aug 20;4(1):36-45.

[2] Alexaki, A., Y. J. Liu, and B. Wigdahl. 2008. Cellular reservoirs of HIV-1 and their role in viral persistence. Curr. HIV Res. 6:388-400.

[3] Blankson, J. N., D. Persaud, and R. F. Siliciano. 2002. The challenge of viral reservoirs in HIV-1 infection. Annu. Rev. Med. 53**:**557-593.

[4] Finzi, D., and R. F. Siliciano. 1998. Viral dynamics in HIV-1 infection. Cell 93:665-671.

[5] KOHLSTAEDT LA, WANG J, FRIEDMAN JM, RICE PA, STEITZ TA. 1992. Crystal structure at 3.5A° resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256: 1783 – 1790.

[6] SPENCE RA, KATI WM, ANDERSON KS, JOHNSON KS. 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. Science 267: 988 – 993.

[7] RICHMAN DD. 1993. Resistance of clinical isolates of human immunodeficiency virus to antiretroviral agents. Antimicrob. Agents. Chemother 37: 1207-1213.

[8] Synthesis and biological evaluation of novel 2-arylalkylthio-4-amino-6-benzyl pyrimidines as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. Qin H, Liu C, Zhang J, Guo Y, Zhang S, Zhang Z, Wang X, Zhang L, Liu J. Bioorg Med Chem Lett. 2010 May 1;20(9):3003-5.

[9] ROMERO DL, BUSSA, M, TAN CK, REUSSER F, PALMER JR, POPPE SM, ARISTOFF PA, DOWNERY KM, SO AG, RESNICK L, TARPLEY WG. 1991. Nonnucleoside reverse transcriptase inhibitors that potently and specifically block human immunodeficiency virus type 1 replication. Proc Natl Acad Sci U.S.A. 88: 8806-8810.

[10] A. Lipinski, J Pharmacol Toxicol Methods., 44: 235 (2000).

AUTHOR'S PROFILE

P.V.R.D.Prasada Rao received B.Tech degree in Computer Science & Engineering from K. L. College of Engineering in 1995, Masters in Computer Science and Technology from Andhra University in 1998

Presently working as a Associate professor in department of CSE in K.L.University

Dr.A.Govardhan received Ph.D in Computer Science & Engineering from JNTU in 2003. M.Tech from JNU in 1994. and B.E from Osmania University in 1992.

He is Presently working as a Director of Evaluation in JNTU, Hyderabad. His research interests includes Data Mining, Bioinformatics , Network security, Information Retrieval & Search Engines. He has Published more than 120 papers in various national and international Journals, resource person and keynote speaker for many international conferences.

.