In silico identification of Novel HIV-Protease inhibitors (PIs) using ZINC drug Database

K.K. Srivastava[1], Shubha Srivastava[2], Tanweer Alam[1], Rituraj[1*]

[1]Department of Chemistry, Vinoba Bhave University, Hazaribagh, India
[2]Department of Chemistry, KBW College, Hazaribagh, India

Abstract

The Human immunodeficiency virus type-1 protease is one of the most important target of highly active anti retrovirus therapy (HAART) for the treatment of all acquired immune deficiency syndrome (AIDS). Protease inhibitor Darunavir is most recent included as a PI in the list of HARRT, more effective against mutant type and wild type of Protease with increased no. of H-bonding then precursors approved by FDA, So herein we taken Darunavir as a base structure for virtually identification of more/similar efficient drug like leads then Darunavir using PDB structure (3BGR) of Protease from PDB database 'RCSB' versus chemical compounds database 'ZINC' using Schrodinger and Discovery Studio software. Using molecular constraint search with similarity coefficient 'Tanimoto', 1,65,000 ligands were extracted and docking analysis resulted in some efficient in docking and in other computational medicinal parameters, we are reporting such leads, and, they may further undergo through high end extensive virtual investigation and beyond.

Keywords: HIV, 3BGR, Protease inhibitors (PIs), Darunavir, docking, Tanimoto, Schrodinger

Introduction

Human immunodeficiency virus (HIV) is retrovirus that causes acquired immunodeficiency syndrome (AIDS), a condition of immunity to fail in human body to begin life threatening Infections lifelong. Acquired immunodeficiency syndrome (AIDS) is one of the leading cause of death in the world[1]. After rigorous multidisciplinary research worldwide successful development of vaccine is still elusive. Human immunodeficiency virus type-1 protease (HIV-1 PR) is a catalytic protein, in a role to cleaves the Gag and Gag-Pol viral poly-proteins, allowing the virus to efficiently infect new host cells. The HIV-1 PR, encoded in the 5' end of the pol gene, is expressed as part of the gag-pol poly-protein. This gene encodes a 99 amino-acids protein. Homodimeric of this protein, i.e. protease is a C2 symmetric enzyme consisting 99 amino acid monomer. Each monomer contributes an aspartic acid residue that is essential for catalysis [2,3,4]. The two chains of this homodimeric form a tunnel with a 'flap" from each protein chain helping to secure the poly-protein in place [5]. The Darunavir and many others inhibitory drugs interact with amino acids in between these dimeric protein flaps. In HIV-1 Protease PIs target to disrupt an essential function in the life cycle of HIV by breaking up the viral polypeptide into components that can be used to form mature virus particles [6]. Darunavir and other PIs act as non-covalent inhibitor of HIV protease and compete with the natural substrate to occupy the active site. When a protease inhibitor binds, the HIV life cycle is halted as the protein components for new viral particles are not able to produce.[4]On this background we extended to search to find out as similar and effective potent PIs as like Darunavir.

Darunavir

Darunavir is an anti-retroviral drug under the umbrella of Protease inhibitors, which is used for hindering the activity of the virus protease [6]. In the work herein Darunavir is taken as reference molecule and find out 1% of similar molecules of each retrieved files of zinc drug bank(sd file) using similarity coefficient "Tanimoto" in D.S. 2.5 in a single job around 1350 molecules was found as similar to Darunavir, we performed as like. 118 jobs and a total 118x1390 molecules, we found and perform docking in Schrodinger.

PDB	LIGAND	RESOLUTION	R VALUE	R FREE	MUTATION
3BVB	Darunavir	1.30	0.170(Obs.)	0.210	D25N

Protease

The HIV-1 protease is one of the most important targets of antiretroviral therapy used in the treatment of AIDS, this HIV protein has an important key role in viral replication as a catalytic protein. The chemical activity of the HIV-1 protease depends on the two residues in the active site, Asp 25, Asp25', one from each copy of the homodimer. Darunavir interacts with these catalytic aspartates and the backbone of the active site through hydrogen bonds, specifically binding to residues Asp 25, Asp25', Asp29, Asp 30, Asp30' and Gly27. This interaction prohibits viral

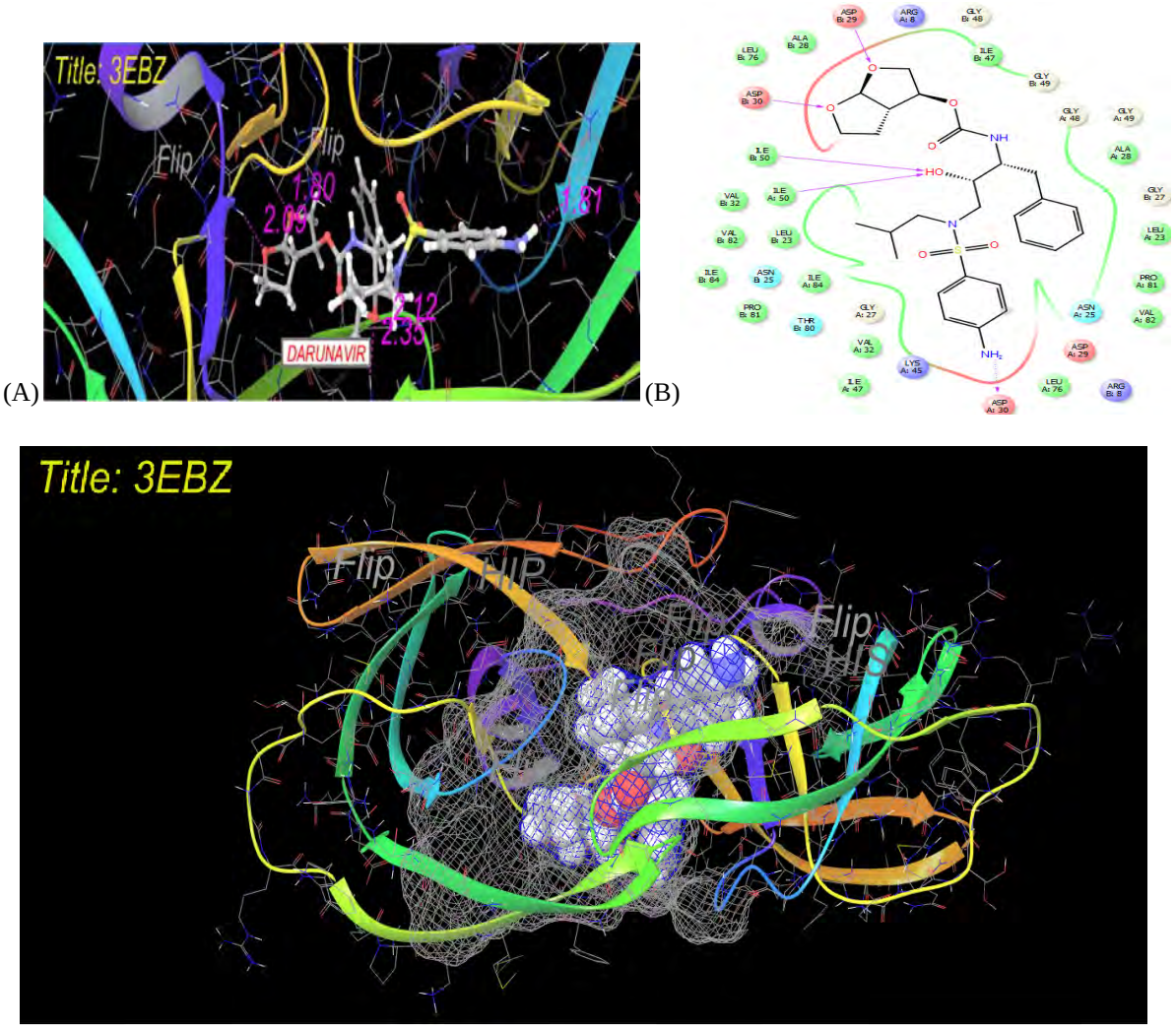

FIG. 1.1(A): DARUNAVIR DOCKED IN PDB 3BGR, (B) ITS INTERACTION DIAGRAM (C)SURFACE MAP OF DARUNAVIR AND PR CAVITY(CORRESPONDING BLUE AND BROWN NET)

replication as it competitively inhibits the viral polypeptides from gaining access to the active site and strongly binds to the enzymatic portions of this protein. Protease inhibitors (PI) were developed to inhibit cleavage function of HIV1 protease by mimicking the reaction intermediates that arises during the hydrolysis of the substrate, disabling the enzyme to cleave the Gag and Gag-Pol viral polyproteins, the virus to efficiently infect new host cells. X-ray crystal structures of ligand-protein co-complexes have been important tools for medicinal chemists in the discovery, design, and optimization of drug candidates [7,8,9]. These structural data, along with the computational analysis tools that have been developed to implement structure-based drug design(SBDD), in in-vitro analysis high cost and the extensive time frame requirement to find out drug like leads, make it still impractical to use these conventional methods to evaluate the effect of each mutation in view of the genetic background of HIV-1 protease. In this context computational methods are growing up as a very important tool to medicinal chemist and now popularity of such methodology seeing on floor. Fast up-gradation in computation algorithm and easy to use, improve the screening analysis much easier, revealing the role of individual mutation and its impact on the protein function, proved to be very successful in medicinal chemistry. As a greater number of X-ray crystal structures become available to medicinal chemists, with the advent of structural genomics [10], computational methods that take advantage of protein-ligand structural data are becoming more critical to the drug design process, in this regard we retrieved 3BVB (see Table-1) Pbd file from rcsb.org for Protease as target

having co-complexed with inhibitor Darunavir (most recent HIV protease inhibitor to reach the market in 2009). HIV-1 protease is one of the most important drugable target in new drug design/discovery. Protease inhibitors restrain the viral maturation by inhibiting the functional and structural proteins formation in virus so HIV produced immature, non-infectious. A single Mutation in gene of HIV-PR causes double mutation in enzyme since HIV-PR is homodimeric protein, containing 99 amino acids in each chain with an active site located at the dimeric interface. The crystal structure analysis shows HIV-PR docked with Darunavir in the surrounding pocket of amino acids-Asp-25(A), 30(A), 29(A), 129(B), 130(B), Ala-28(A), 128(B), Gly-48(A), 49(A),127(B), 148(B), 27(A), 149(B),Leu-123(B), 23(A), Ile-150(B), 182(B), 184(B), 132(B), 50(A), 82(A), 84(A), 32(A), Ser-31(A), Met-76(A), Thr-180(B), Val-47(A), 147(B), Ash-125(B), Arg-8(A), Pro-181(B), 81(A), from literature these amino acids constitutes three regions; catalytic-core (Asp-25(A), Gly-27(A), Ala-28(A), Asp-29(A) and Asp-30(A), flap(Ile-47(A), 54(A), Met-46(A) Phe-53(A), Gly-48(A) ,49(A) 51(A), 52(A) and Ile-50(A)) and the C-terminal region (Pro-81(A), Ile-84(A)). According to literature, Asp-25(A), Gly-27(A), Ala-28(A), Asp-29(A) and Gly-49(A) are known to be highly conserved residue in which a potential protease inhibitor bind effectively. Mutation of HIV-PR at Val-32(A) Ile-50(A), 84(A) (hydrophobic residues, close to binding pocket) are responsible to lesser binding efficacy.

We retrieved 3BVB (see table 1.1) Pdb file from rscb.org for protease as target having in vivo docked with inhibitor Darunavir (PR, a non-peptidic analogus of Amprenavir, at the critical change at the terminal tetrahydrofuran(THF) group. Instead of a single THF group, Darunavir contains two groups fused in the compound, form a bis-THF moiety which makes it more effective than Amprenavir. Crystal structure analysis of HIV Protease enzyme show that the HIV protease, encoded in the 5' end of the pol gene, is expressed as part of the gag-pole polyprotein, a C2 symmetric homodimeric enzyme consisting of two 99 amino acids monomers contributes an aspartic acid residue that is essential for catalysis.

Experiments: Retrived pdb file (see Table 1.1)from rcsb.org were prepared in protein preparation wizard of maestro with following steps- preprocess(default settings), deleting all unnecessary water molecules and other structures except Darunavir, added hydrogen, generated it states, optimization, and minimization(with OPLS2005 forcefield) with default constraint of the 0.3Å of RMSD and corresponding Grids are generated in these prepared pdb with the center defined by the co-crystalized ligand Darunavir with default settings included partial charge and saved all in pre-created directory folder. Ligands extracted as previously mentioned procedure as similar to Darunavir with DS V2.5 in job " find similar molecules" with settings 1% similar molecules to 'Darunavir' with similarity coefficient 'tanimoto' which is very well known similarity measures, remaining are almost default. Similar ligands are prepared for docking jobs in 'ligprep' with forcefield OPLS2005 using epik with deselected options 'desalt' and selected 'generate tautomer' and finally with the ligands whose docking score more than Darunavir selectively prepared in 'ligprep' with forcefield OPLS 2005 using 'ioniser' with included setting as previous, generate default no. of low ring conformations and all combination as default and docked in corresponding grids of pdb. All docking calculations are performed using the "Extra Precision"(XP) mode of Glide Program with settings including sampling ligands 'flexible', optionally available various protocols for ligands constraints as rewards measure, partial charge of ligands and similarity measures to 'Darunavir' were included, finally extracted all best ligands in DS prepared in 'ionizer' with settings as default and checked for its ionized state influence on docking scores sometimes they appear lift up the scores up to the 2 units or more; due to interactions elevated. All jobs were done on Intel i7-3770K (unlocked) quad core machine with bios setting 3.9-4.4GHz with GSkill 8GB RAM & Corsair H-70 liquid cooling system. Medicinal parameters were calculated using qikprop (table-2-6)

Results and Discussion

In our virtual investigation we find following 'ZINC' molecules close similar in docking score and other essential computed medicinal parameters in comparison to Darunavir (Table 1.2).

Table-1.2 3BVB DOCKING SCORE AND OTHER CALCULATED PROPERTIES DETAILS

Title	D.S.	LIP	#rtvFG	CNS	dipole	SASA	QPlogHERG	QPlogBB	QPPMDCK	QPlogKp	#metab	QPlogKhsa	PHOAbs
Darunavir	-13.7	-7.5	1	-2	9.277	787.344	-5.716	-1.643	190.071	-1.945	3	-0.21	76.87
ZINC78487241	-12.60	-5.3	1	-2	2.926	650.261	-2.939	-1.672	83.222	-3.575	4	-0.724	60.758
ZINC09060710	-12.50	-6.5	0	-2	8.907	727.848	-4.054	-1.54	106.32	-2.492	6	-0.19	76.47
ZINC72320180	-12.37	-5.1	0	-2	3.745	599.221	-2.474	-1.835	16.188	-3.136	5	-0.605	62.248
ZINC78487244	-12.22	-5.3	1	-2	2.816	647.065	-2.895	-1.638	87.579	-3.536	4	-0.729	61.146
ZINC78487242	-12.15	-5.0	1	-2	7.816	640.264	-2.839	-1.764	63.111	-3.793	4	-0.755	57.933
ZINC20756562	-12.01	-6.3	0	-2	8.731	692.828	-0.872	-1.343	38.186	-2.729	5	-0.779	68.228
ZINC21532338	-11.83	-5.7	0	-2	5.557	724.991	-4.359	-2.026	31.472	-4.368	4	-0.457	62.337
ZINC04166489	-11.66	-4.5	1	-2	8.352	625.05	-5.492	-2.187	18.173	-4.822	4	-0.562	57.809
ZINC65562423	-11.39	-5.86	0	-2	3.726	608.11	-4.697	-1.357	131.763	-2.592	5	-0.282	78.322
ZINC09492777	-11.57	-6.2	2	-2	13.207	748.356	-6.01	-1.988	48.114	-3.467	3	-0.909	72.361
ZINC13728984	-11.44	-5.7	2	-2	5.422	654.216	-6.861	-1.536	142.047	-2.306	2	0.233	90.384
ZINC02244300	-11.38	-5.6	0	-2	1.225	577.75	-5.204	-1.048	275.202	-2.282	3	-0.057	90.713
ZINC14888955	-11.3	-6.1	0	-2	4.495	727.053	-4.09	-1.208	236.8	-2.599	4	0.279	91.111
ZINC65562426	-11.22	-5.55	0	-2	3.398	616.868	-4.766	-1.457	107.04	-2.746	5	-0.246	76.512
ZINC66142186	-11.21	-5.4	0	-2	6.856	623.076	-4.761	-1.76	104.184	-2.533	5	-0.668	71.783
ZINC39667724	-11.29	-6	0	-2	7.8	619.517	-3.745	-1.555	60.697	-3.615	2	-0.174	73.44
ZINC65562428	-11.18	-5.31	0	-2	5.123	612.868	-4.752	-1.373	130.272	-2.596	5	-0.268	78.513
ZINC14541646	-11.08	-5.45	0	-2	5.101	732.213	-3.833	-1.233	225.736	-2.424	4	-0.314	77.055
ZINC14745685	-11.28	-5.63	0	-2	4.018	712.53	-2.778	-1.291	273.337	-2.737	4	-0.735	66.466
ZINC14888955	-11.09	-6.2	0	-2	5.577	694.038	-3.348	-1.031	295.019	-2.63	4	0.238	92.365
ZINC04834564	-11.03	-7.2	0	-2	9.183	679.069	-6.674	-1.523	136.534	-2.113	4	0.463	94.24
ZINC13552738	-11.45	-7.0	0	-2	8.474	738.46	-3.767	-1.017	279.402	-1.881	8	0.299	90.875
ZINC16363188	-11.37	-4.6	1	-2	8.306	627.594	-5.649	-1.778	45.243	-4.177	3	-0.308	69.862
ZINC32202801	-11.21	-5.9	1	-2	3.292	678.993	-5.947	-1.135	240.977	-2.619	3	0.251	95.049
ZINC76938860	-11.45	-4.0	1	-2	5.694	532.76	-4.389	-1.882	30.383	-4.302	4	-0.859	56.221
ZINC54481755	-11.15	-5.6	0	-2	7.408	655.603	-2.106	-1.304	17.04	-5.791	6	-1.444	34.035

(D.S. (Docking Score, kcal/mol), Lip(Lipophilicity), rtvFG(no. of reactive functional groups, 0 – 2), CNS(Predicted

central nervous system activity on a –2 (inactive) to +2 (active) scale),Dipole(computed dipole moment, 1.0 – 12.5), SASA(Total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius, RANGE- 300.0 – 1000.0), QPlogHERG (Predicted IC50 value for blockage of HERG K+ channels,

concern below –5), QPlogBB(Predicted brain/blood partition coefficient, –3.0 – 1.2), QPPMDCK(Predicted apparent MDCK cell permeability in nm/sec, <25 poor, >500 great), QPlogKp (Predicted skin permeability, log *K*p, –8.0 – –1.0), metab (Number of likely metabolic reactions, 1 – 8), QPlogKhsa(Prediction of binding to human serum albumin, –1.5 – 1.5), PHOAbs(Predicted human oral absorption on 0 to 100% scale, >80% is high, <25%ispoor)

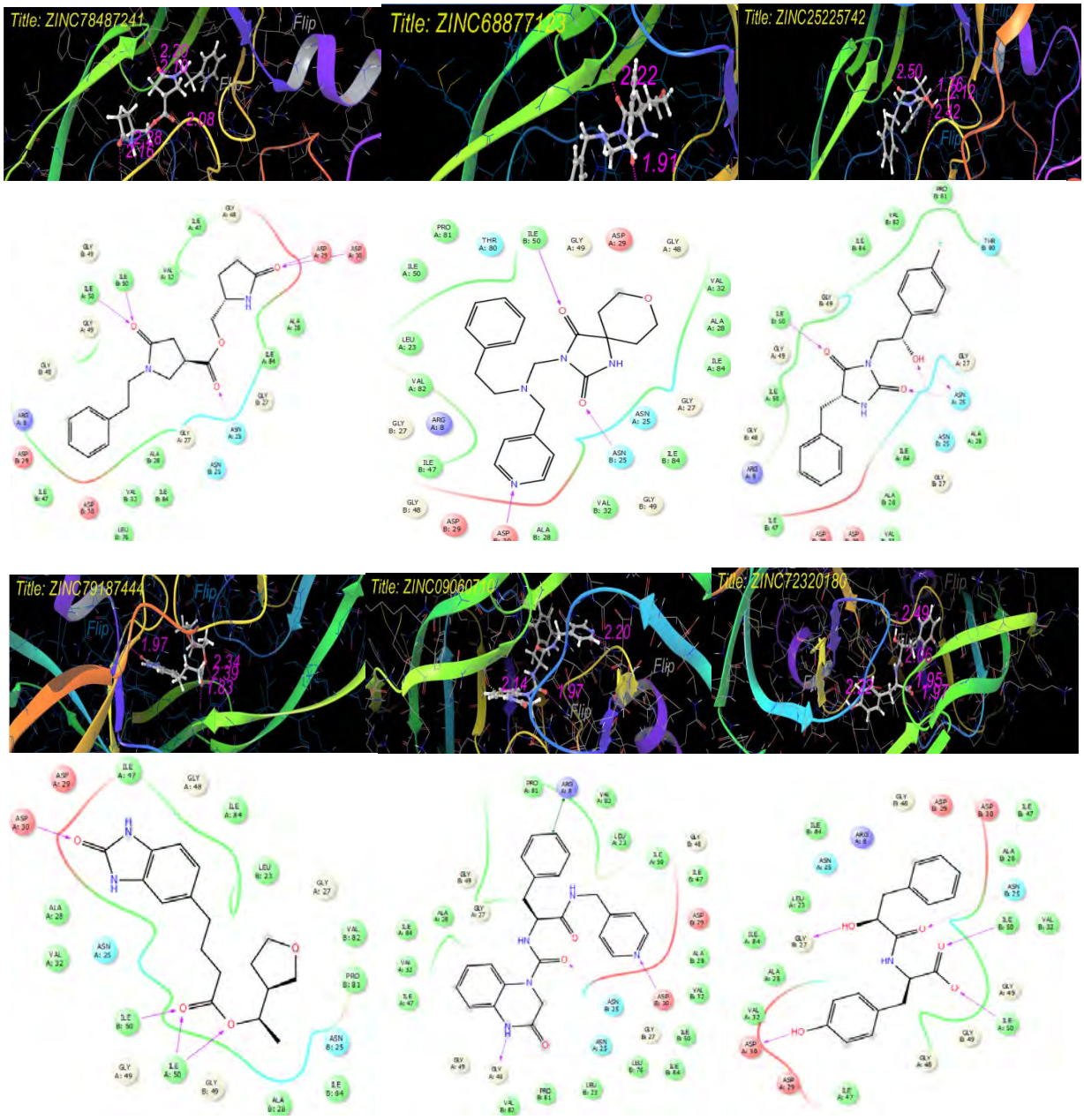

FIG.1.2 3BGR(PR) WITH DOCKED LIGAND ZINC78487241, 68877123, 25225742, 79187444, 09060710, 72320180 AND CORRESPONDING INTERACTION DIAGRAMS(BELOW)

CNS(Predicted central nervous activity, which is -2 in inactive state), here, all above at -2(table-1.2), so very importantly, for identification as a drug like lead for any medicinal chemist, on the another hand QPlogBB (Predicted brain/blood partition coefficient all in case below -1 and up to very low level in depth, showing very good computed data, evenly below -2 in some cases, so good result in this context, in during days, medicinally identification of molecular leads the brain/blood partition coefficient value is very important raw data to think so to consider a molecule as like a drug leads; this is very primarily important to drug discovery process. Herein we seeing above all are in likely to be in very comfortable zone in this regard. All above shortlisted molecules have average to good lipophilic nature as computed data showing, which is very essential parameter in others including QPPMDCK(Predicted apparent MDCK cell permeability computed data (in nm/sec) are in required limit(MDCK cells are considered to be a good mimic for the blood-barrier, data are for non-active transport),

#metab showing no. of likely metabolic reactions, QPlogKp predicting skin permeability ability, QPlogKhsa predicting the binding ability of above molecules to human serum albumin and PHOAbs are in desired limit.

The docking analysis and interaction diagram showing that the interaction of the ZINC78487241 with a new amino acid ASN-25(A) with the acyclic esteric part of that but losing the interaction of ASP-29(B)(which is present with bis-oxolane part in the darunavir) so some new interaction are in new outcome frame, the influence of such new kind of interaction and on efficacy of drug potency may make a floor for big discussion but the docking calculation are in some diminished comparatively to Darunavir herein about one unit which are some drawback but this backbone moiety considerably attentional to discovery for better leads.

Conclusion

In this work, we have tried to recognized some more/similar potent drug like leads instead 'Darunavir' may be more effective, we used five different RT crystallographic structures for better identification/verification for our results, ZINC78487241, ZINC09060710, ZINC72320180, ZINC78487244 & ZINC78487242 are showing very fine computed properties therefore, this study verify the importance of small drug like molecules libraries as like 'ZINC.docking.org'and their use certainly help scientific groups to enhance their capabilities in drug discovery with reducing time, including drug discovery process prior synthesis. Meanwhile all herein identified molecules may further investigate instead "in silico".

ACKNOWLEDGEMENTS:

The authors wish to thank with proud and privilege to Will Richard, Raghu Rangaswamy and Vinod Dewarjee of Schodinger ltd. for providing the Schrodinger Suite Software.

References

[1] Rituraj and Alam. IJPSR(2014), Vol.-5(7): 2947-2954. "VIRTUAL SCREENING OF NOVEL HIV-RT NNRT INHIBITORS USING ZINC DATABASE"
[2] De Clercq, E. Strategies in the design of antiviral drugs. Nat. Rev. Drug Discovery 2002, 1, 13-25
[3] Meadows, D.C.; Gervay-Hague, J. Current developments in HIV chemotherapy. ChemBioChem 2006, 1, 16-29
[4] Dennis M. Junqueira2,3, Sabrina E. M. Almeida2, Gustavo F. Vieira1,3* New Insights into the In Silico Prediction of HIV Protease Resistance to Nelfinavir, www.plosone.org January 2014 | Volume 9 | Issue 1 | e87520
[5] Goodsell, D. "HIV-1 Protease" Protein Data Bank. June 2000 Molecule of the Month
[6] Mittal, S., Bandaranayake, R. M. et al "Structural and Thermodynamic Basis of Amprenavir/Darunavir and Atazanavir Resistance in HIV-1 Protease with Mutations at Residue 50" J. Virology. 2013, 87, 4176-4184
[7] Babine, R. E.; Bender, S. L. Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. Chem. Rev. 1997, 97, 1359-1472.
[8] Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. J. Med. Chem. 1994,37, 1035-1054.
[9] Kuntz, I. D. Structure-based strategies for drug design and discovery. Science 1992, 257, 1078-1082
[10] Stevens, R. C.; Yokoyama, S.; Wilson, I. A. Global efforts in structural genomics. Science 2001, 294, 89-92