Research Article

In silico Screening of Various Natural Compounds to Predict the Potential Inhibitors that Target the HIV-1 Protease Sub Type – A

Jitender Singh, Ashvinder Raina*
Dept of Experimental Medicine and Biotechnology, Post Graduate Institute of Medical Education and Research, Chandigarh, India.
***Corresponding author's E-mail:** drashvinder@gmail.com

Accepted on: 02-08-2016; **Finalized on:** 30-09-2016.

ABSTRACT

HIV virus is the member of the retrovirus which causes the AIDS disease. HIV-1 protease (PR) sub type-A is an aspartic protease (human immunodeficiency virus type 1 protease) that plays a crucial role in the in the virus assembly and budding process and thus involved in virus replication cycle. During last decades emphasis was to develop effective molecules that inhibit the HIV-1 PR which has resulted in the discovery of various drugs. Many natural products have been shown to possess promising activities that could help in the prevention and betterment of the disease. In this study we have explored the activity and predicted the binding pocket of this HIV-1 protease sub type–A which was then considered as a drug target site for the structure based drug design. Molecular docking studies were employed to find the best inhibitors against uncomplexed HIV-1 protease sub type A. After screening 20 compounds, five compounds from natural plants showing good docking score and glide energy as well as interaction with specific amino acid residues were shortlisted as a drug lead compound. After analyzing the *in silico* study results, we concluded that all these natural compound inhibitors can be potential inhibitors for HIV-1 protease sub type-A uncomplexed protein and can work by stopping the virus assembly and budding processes. Thus this data could be used in the research and development of new therapeutic agents for AIDS treatment.

Keywords: HIV-1 protease sub type-A, inhibitors, pharmacokinetics, molecular docking, Binding pocket.

INTRODUCTION

The human immunodeficiency virus (HIV) is a lentivirus (a sub group of retrovirus) that cause HIV infection and acquired immunodeficiency syndrome (AIDS)[1,2] and is one of the most important public health problems, affecting many people every day. It is still one of the most challenging diseases of the early 21st century. AIDS is a condition in which there is complete failure of immune system which allows various life-threatening opportunistic infections and cancers to prosper. HIV infects the important cells in the immune system just like helper T cells, dendritic cells and macrophages. The virus interacts with target cells using envelope glycoprotein that are recognized by CD4 receptors and CCR5 in macrophage or CXCR4in T cell co-receptors leading to membrane fusion followed by virus entry and subsequent integration to the host genome[3]. HIV infection leads to decrease in the number of T-helper cells ($CD4^+$ T) cells through various mechanisms like apoptosis of uninfected bystander cells[4], pyroptosis of abortively infected T cells,[5] direct or viral killing of infected cells, and killing of infected $CD4^+$ T cell by cytotoxic T lymphocytes ($CD8^+$ T) that recognize infected cells[6]. HIV-1 protease (human immunodeficiency virus type 1 protease) is a member of the aspartyl-protease family enzyme essential for the life–cycle of HIV and plays a critical role in the virus replication cycle[7,8]. It acts by cleaving the *gag* and *pol* viral polyproteins at the appropriate sites of active site to process and helps in viral maturation[9-11], thus resulting in the production of mature protein components of an infectious HIV virion[12]. This ultimately results in the formation of an infectious virus. As maturation is an essential step in HIV life cycle, and HIV-1 protease plays a key role in virus biology like proteolytic cleavage of polyproteins for the maturation and infectivity of the virus[13], it is considered to be an important target for development of anti-HIV-1 drugs[14]. For AIDS therapy HIV-1 protease inhibitors have been successfully used. However, the development of drug resistance by the virus represents an important problem on AIDS-chemotherapy. Additionally, there is not yet a definitive cure for this disease[15]. At the molecular level, resistance takes place in the form of mutations within the protein, which maintains the viable catalytic site but preferentially lower the affinity of protease inhibitors with respect to protease substrates[16]. Mutations associated with drug resistance occur within the active site as well as non-active distal sites[17]. HIV-1 protease is a dimer of C2 symmetry with each monomer consisting of 99 amino acid residues.

Each dimer is of 11 kDa, which contains a conserved catalytic site of Asp-Thr-Gly (Asp25, Thr26 and Gly27), and the amino acid site sequences that can be catalyzed by HIV protease are Phe-Pro, Pro-Tyr and Leu-Phe in polyprotein[18]. Each monomer in its secondary structure contains one α helix and two antiparallel β sheets. It is gated by two extended β hairpin loops (residues 46−56) known as flaps[19]. HIV protease has two molecular "flaps" which move a distance of up to 7A when the enzyme becomes associate with a substrate[20]. The crystal structure of unbound HIV-1 sub type –A protease (pr) has been determined to 1.7 A resolution and refined as a homodimer in the hexagonal space group P6(1) to an R(cryst) of 20.5%. In this study our interest was to predict

the potential inhibitors from the selected natural plant compounds and further analyze important role of the target conformation by the molecular docking results.

MATERIALS AND METHODS

Protein Preparation

The 99 amino acids residue long two chains (A,B) crystal structure of HIV-1 protease sub type-A protein of HIV virus was retrieved in FASTA format from National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) database with accession number: 3IXO. The 3D structure available of this accession number that was taken from PDB database (http://www.rcsb.org/)(Figure1A). In addition to also target FASTA sequences were constructed using different servers CPH Model (http://www.cbs.dtu.dk/services/CPHmodels/) and phyre2 (http://www.sbg.bio.ic.ac.uk/phye2/html/page.cgi?Id=index). The best model was selected on the basis of Ramachandarn plot and protein stability analysis by using SAVES online software tool (https://services.mbi.ucla.edu).

Binding Site Prediction

Active sites of the receptor were analyzed by metapocket software (http://metapocket.eml.org). Binding sites are the sites where active site is surrounded by residues. In the centre of the active site-1 we selected a grid map value for the preparation of the grids for docking score.

The active site lies between the identical subunits and has the characteristic Asp-Thr-Gly (Asp25, Thr26 and Gly27) sequence common to aspartic proteases. The two Asp25 residues (one from each chain) act as the catalytic residues. These subunits occurs in the active site-1 only.[21]

Ligand Preparation

The structure of natural compounds from selected natural plants was downloaded from pubchem data base (http://pubchem.ncbi.nlm.nih.gov/). (Table 1) The drug files was downloaded in sdf file format and were converted into pdb files using open Pymol software. The ligands preparation included 2D-3D conversion, verifying and optimizing the structures. We have shortlisted 20 ligands for further analysis.

Docking

Autodock was used to dock protein and ligand molecules. The molecular docking of each natural compound at the binding site of proteins a grid is generated by autodock[22]. Preparation of required input files for autodock using autodock 4.2 version. Preparation of files through autodock 4.2 involved addition of polar hydrogen atoms and charges. We studied the ligand–protein binding energy and selected top five ligands as good inhibitors.

Pharmacokinetic Properties

FAF Drugs3[23] server was used to analyze the absorption, distribution, metabolism, and excretion properties of top 5 best hits. ADMET properties of top 5 successive hits were checked in optimal descriptors (hydrogen bonds, charge) in Ph=7.4. And also analyze the oral toxicity properties by using PROTOX web server[24].

RESULTS AND DISCUSSION

To know the active binding sites of target protein taken from PDB with ID: 3IXO (Figure 1A), first of all we validated the 3D structure of the target protein using SAVES online software tool https://services.mbi.ucla.edu (took help of PROCHECK and Ramachandran plot-Figure 1B) and we found 87.88% of the residues had an averaged 3D-1D score >= 0.2, and at least 80% of the amino acids have scored >= 0. 2 in the 3D/1D profile, thus passing the criteria of verifying 3D predicted structure. After validation of this Predicted 3D structure of target protein we used this protein to know the ligand binding sites using Metapocket 2.0 online tool. Top three active sites predicted are shown in Figure 1C in the form of spheres. Site prediction was followed by analyzing the number of important amino acid residues in catalytic region.

We found that there are important amino acid residues present in top three metapocket sites which are shown in Figure 2. Out of these three metapocket sites we selected metapocket 1 for further analysis as metapocket 1 contained conserved catalytic site of Asp-Thr-Gly (Asp25, Thr26 and Gly27). While finalizing metapocket site, we screened 20 different compounds from natural plants. These 20 compounds were then subjected to Lipinski rule of five to evaluate druglikness (www.scfbio-iitd.res.in)[25]. Out of 20 compounds 14 followed the Lipinski rule of five suggesting their role as a drug compound. (Table 1). These 14 drug compounds were then checked for their binding with target protein and also check the docking score by using autodock Docking software. Grid map was set in X, Y AND Z centre with 50, 50, 50A° spacing respectively. For each compound, many docking poses were obtained with different binding energy score. Out of many docking poses, only those were selected which had the best binding affinity (on the basis of docking score) and showed good hydrogen bond interaction.

This was repeated for all the 14 drug compounds. Among all these we selected only top five based on their binding energy (Docking score).

Their various chemical characteristics have been shown in Table 2. So from these results we concluded that these five drug compounds can be selected as potential inhibitors against the target proteins having good docking score (as shown in the Table 3 and Figure 3).

Among these top selected drug compounds the Curcumin shows highest affinity to bind with target protein with energy value - 8.55 kcal/mol compared to other drug compounds like demethoxycurcumin, Flavonoid, Formononetion and Oleocanthal. (Figure 4A, B, C, D and E).

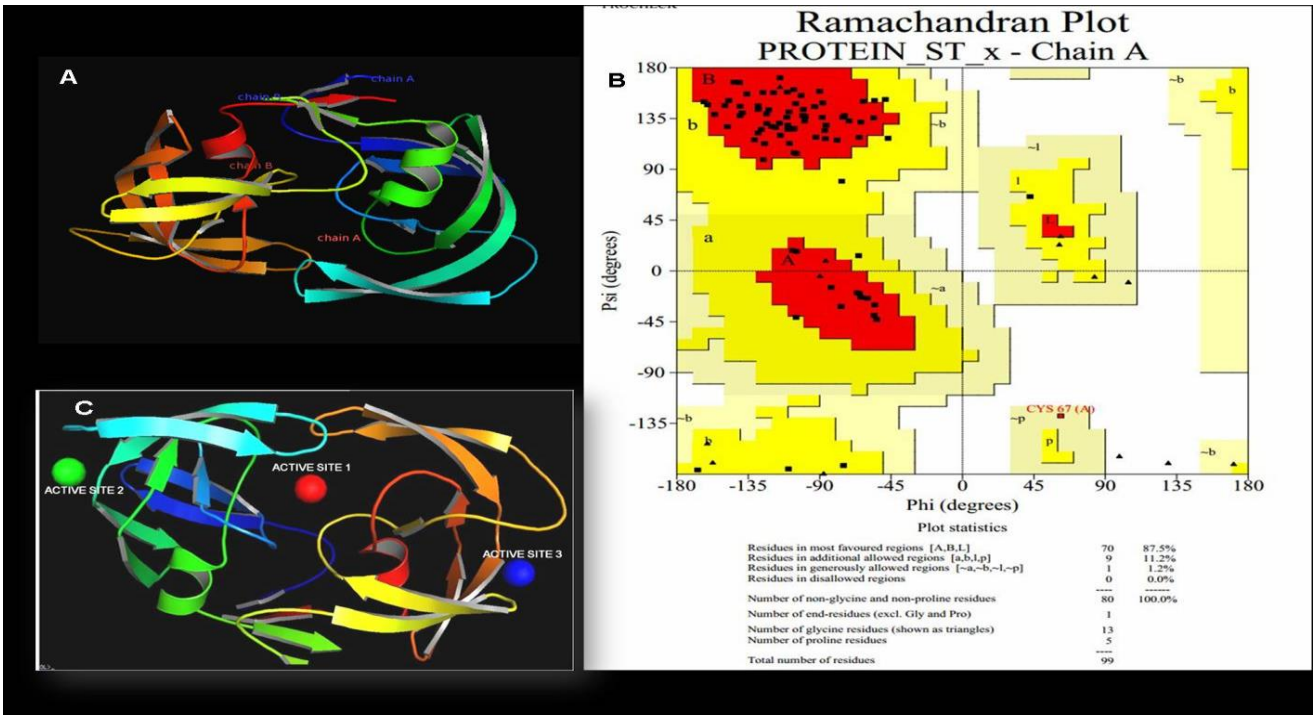
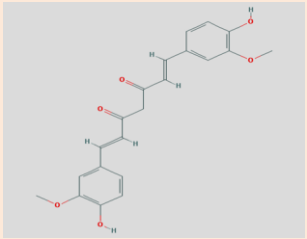


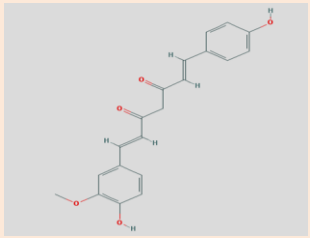
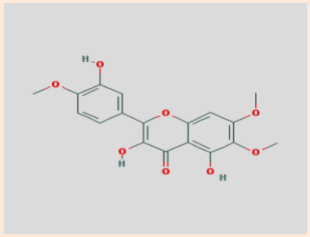
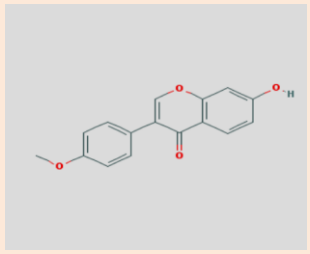
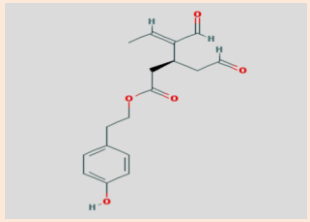
Figure 1A: 3D structure of HIV-1 protease subtype-A, **Figure 1B:** Ramachandran plot of HIV-1 protease subtype-A, **Figure 1C:** Protein structure showing with their active sites

Table 1: Compound's follows Lipinski Five Rules

S. No.	Compound Name	Molecular Mass	LOgP	Hydrogen Bond Donors	Hydrogen Bond Acceptors	Molar Refractivity
1	Cineole	154	2.7	0	1	45.526985
2	Citronellol	156	2.751	1	1	49.531
3	Curcumin	368	3.36	2	6	102.0165
4	Demethylcurcumin	338	3.361	2	5	95.4645
5	Elenolic Acid	240	-1.31	0	6	52.361
6	Eggenol	164	2.129	1	2	48.5597
7	Farnesol	222	4.3978	1	1	72.498779
8	Flavonoid	360	2.6254	3	8	90.376
9	Formononetion	268	3.0166	1	4	74.0362
10	Geraniol	154	2.6714	1	1	49.5077
11	Hydroxytyrosol	154	0.632	3	3	40.561
12	Isoflevene	208	3.6195	0	1	66.2499
13	Oleocanthal	304	2.218	1	5	81.37877
14	Teroinenol	154	2.5	1	1	47.39

Table 2: Showing Top Five Compounds and their Chemical Properties

Name of Compound	Pubchem ID	Chemical Formula	Structure
Curcumin	969516	$C_{21}H_{20}O_6$	

Demethoxy curcumin	5469424	$C_{20}H_{18}O_5$	
Flavonoid	5317287	$C_{18}H_{16}O_8$	
Formononetion	5280378	$C_{16}H_{12}O_4$	
Oleocanthal	11652416	$C_{17}H_{20}O_5$	

Binding site ID: 1					
RESI	ARG_B^8^	SER_B^81^	VAL_B^82^	ASP_A^29^	LEU_B^23^
RESI	GLY_A^27^	ARG_A^87^	THR_A^26^	ALA_A^28^	ILE_B^84^
RESI	ASP_B^25^	THR_B^80^	PHE_A^53^	GLY_A^49^	ILE_A^50^
RESI	GLY_A^48^	VAL_B^32^	ASP_A^30^	ASP_A^25^	PRO_B^79^
RESI	ALA_B^28^	ILE_B^54^	ILE_A^47^	LYS_A^45^	ASP_B^30^
RESI	THR_B^31^	GLY_A^51^	MET_A^46^	ILE_B^47^	LEU_B^76^
RESI	ASP_B^29^	GLY_B^27^	GLY_A^52^	ILE_B^50^	VAL_A^32^
RESI	ILE_A^84^	GLY_B^48^	GLY_B^49^	LEU_A^23^	VAL_A^82^
RESI	ARG_A^8^	SER_A^81^	MET_B^46^	ARG_B^87^	GLY_B^52^
RESI	LEU_A^10^	GLN_B^7^	PHE_B^53^		
Binding site ID: 2					
RESI	LEU_A^38^	PRO_A^39^	ASP_A^60^	GLN_A^61^	ILE_A^62^
RESI	LYS_A^43^	GLN_A^58^	TYR_A^59^	TRP_A^42^	LYS_A^41^
RESI	GLY_A^40^				
Binding site ID: 3					
RESI	LYS_B^14^	LEU_B^63^	GLU_B^65^	LYS_B^70^	ILE_B^15^
RESI	ILE_B^64^	LYS_B^69^	ALA_B^71^	GLY_B^16^	GLY_B^17^
RESI	LEU_B^38^	PRO_B^39^	TYR_B^59^	ILE_B^62^	GLN_B^61^
RESI	ASP_B^60^	LYS_A^41^	TRP_A^42^	LYS_A^43^	GLN_A^58^
RESI	TYR_A^59^	ASP_A^60^	GLY_B^40^	LYS_B^41^	TRP_B^42^
RESI	GLN_B^58^	LYS_B^43^			

Figure 2: Representing Top Three Active Sites with Amino Acid Residues

Table 3: Docking Score and Overall Properties of top 5 Successive Ligands

Chemical ID	Binding Affinity	LD50	Molecular Weight	Rotatable Bonds	Flexibility	H acceptor	H donor	Ring	Carbon Atom	Hetero Atom	Oral Bioavailability
969516	-8.55	2000 mg/kg	368.38 g/mol	8	0.33	6	2	2	21	6	good
5469424	-8.24	2000 mg/kg	338.35 g/mol	7	0.3	5	2	2	20	5	good
5317287	-7.51	5000 mg/kg	360.32 g/mol	4	0.18	8	3	2	18	8	good
5280378	-6.34	2500 mg/kg	268.26 g/mol	2	0.1	4	1	2	16	4	good
11652416	-6.17	5000 mg/kg	304.34 g/mol	10	0.5	5	1	1	17	5	good

Table 4: ADMET Properties of Top 5 compounds

Properties	Ligand-1 Curcumin	Ligand-2 Demethoxy-Curcumin	Ligand-3 Flavonoid	Ligand-4 Formononetion	Ligand-5 Oleocanthal
Human Ether-a-go-go-Related Gene Inhibition	Weak inhibitor	Weak inhibitor	Weak inhibitor	Weak inhibitor	Weak inhibitor
	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor
AMES Toxicity	Non AMES toxic	Non AMES toxic	Non AMES toxic	Non AMES toxic	Non AMES toxic
Carcinogens	Non-carcinogens	Non-carcinogens	Non-carcinogens	Non-carcinogens	Non-carcinogens
Fish Toxicity	High FHMT	High FHMT	High FHMT	High FHMT	High FHMT
Tetrahymena Pyriformis Toxicity	High TPT	High TPT	High TPT	High TPT	High TPT
Honey Bee Toxicity	High HBT	High HBT	High HBT	High HBT	High HBT
Biodegradation	Not ready biodegradable	Not ready biodegradable	Not ready biodegradable	Not ready biodegradable	Ready biodegradable
Acute Oral Toxicity	III	III	III	III	III

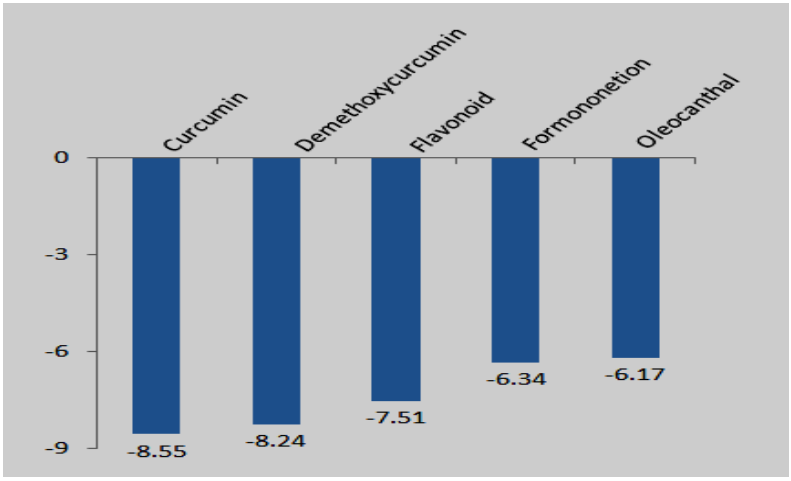


Figure 3: Binding Affinity of top-5 Ligands

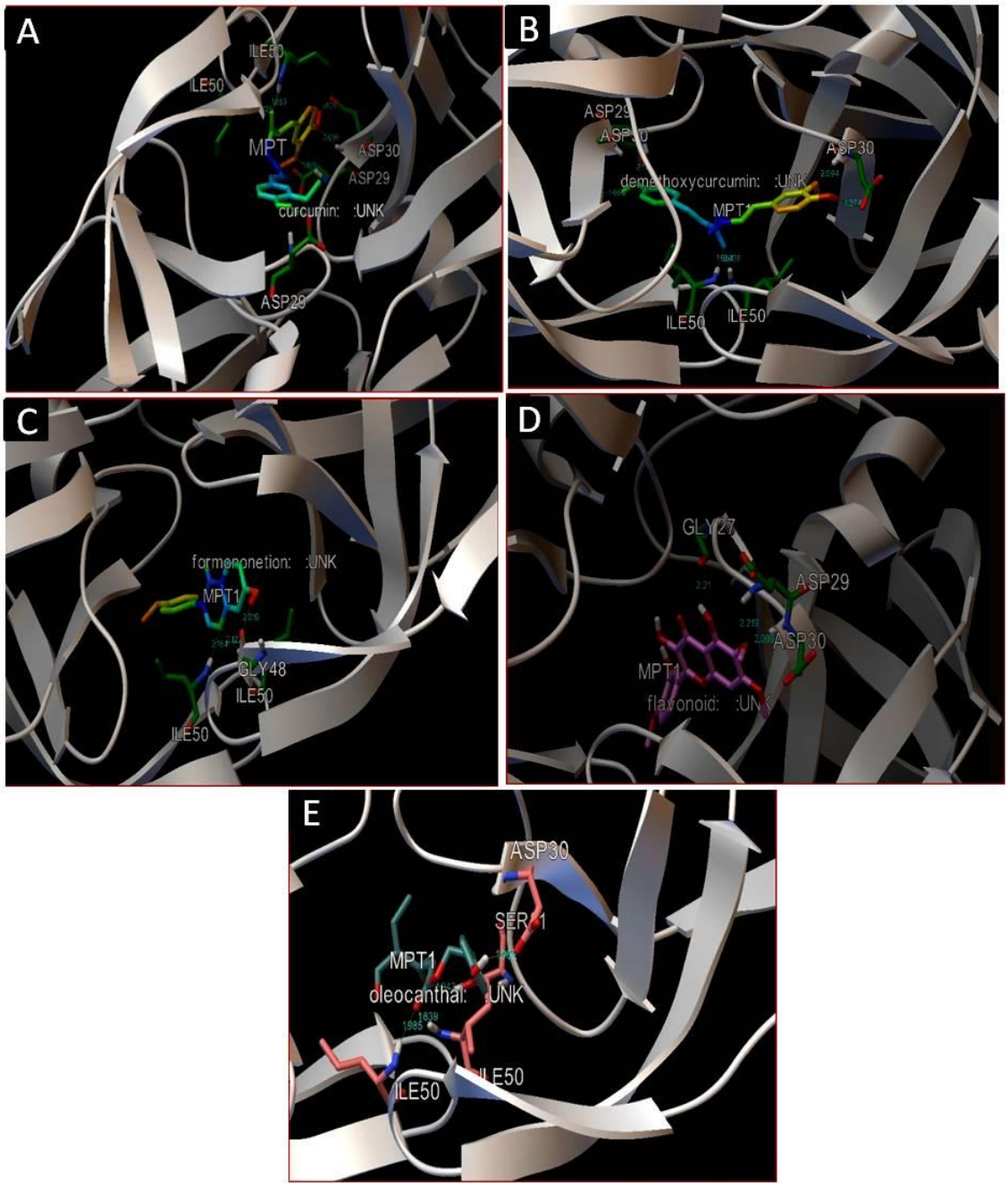


Figure 4: Structure showing Interaction of Top Five Protein-Ligand Complexes

Curcumin interacts with target protein amino acid residues, in chain A with ILE50, ASP 29 and in chain B with ILE 50, ASP29, and ASP30, resulting in the formation of a total of 6 hydrogen bonds (Figure 4A).

The hydrogen bonding is of considerable importance in the interaction of molecules. We also checked the pharmacokinetic properties of the top five drug compounds.

FAF Drugs3[23] server was used to analyze the absorption, distribution, metabolism, and excretion properties of top 5 best hits. ADMET properties of top 5 successive hits were checked in optimal descriptors (hydrogen bonds, charge) in Ph=7.4, and also analyzed the oral toxicity properties by using PROTOX web server[24] (Table 4).

On observing the results we found top hits which have high potential to be used as the base structure for the structure based drug design process.

On the basis of molecular docking energy score to five as suggested may act as natural inhibitor that may be used for the treatment of HIV-1 virus.

Hence this information may be useful for the designing of the novel drug molecules against HIV-1 Protease sub type-A protein of HIV.

CONCLUSION

This study introduced 5 drug like compounds i.e. Curcumin, demethoxycurcumin, Flavonoid, Formononetion and Oleocanthal, which could specifically bind to HIV-1 Protease sub type-A.

The study also showed considerable binding affinity and the results of pharmacokinetic studies also confirmed the potential of these structures to be considered as a base structure for anti-HIV treatment.

REFERENCES

1. Weiss RA "How does HIV cause AIDS?" Science, 260, 1993, 1273–9.
2. Douek DC, Roederer M, Koup RA. "Emerging Concepts in the Immunopathogenesis of AIDS". Annu. Rev. Med. 60, 2009, 471–84.
3. Naider F, Anglister J. Peptides in the treatment of AIDS. Curr Opin Struct Biol; 19, 2009, 1-10.
4. Garg H, Mohl J, Joshi A. "HIV-1 induced bystander apoptosis". Viruses, 4, 2012, 3020–43.
5. Doitsh Gilad, Galloway, Nicole L. K., Geng Xin, Yang Zhiyuan, Monroe Kathryn M., Zepeda Orlando, Hunt Peter W., Hatano Hiroyu, Sowinski Stefanie, Muñoz-Arias Isa, Greene Warner C. (2014). "Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection." Nature, 505, 2014, 509–514.
6. Kumar Vinay. Robbins Basic Pathology (9thed.). 2012, p. 147. ISBN 9781455737871.
7. Davies DR. "The structure and function of the aspartic proteinases". Annu Rev Biophys Biophys Chem, 19(1), 1990, 189–215.
8. Brik A, Wong CH. "HIV-1 protease: mechanism and drug discovery". Org. Biomol. Chem. 1, 2003, 5–14.
9. Katz RA: The retroviral enzymes. Annu Rev Biochem. 63, 1994, 133-173.
10. Wlodawer A: Rational approach to AIDS drug design through structural biology. Ann Red Med, 53, 2002, 595-614.
11. Kurup A, Mekapati SB, Garg R, Hansch C. HIV-1 protease inhibitors: a comparative QSAR analysis. Curr Med Chem, 10, 2003, 1679-1688.
12. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS: Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA, 85, 1988, 4686-4690.
13. Fun A. HIV-1 protease inhibitor mutations affect the development of HIV-1 resistance to the maturation inhibitor bevirimat. Retrovirology, 8, 2011, 70.
14. Kohl N.E., Emini E.A., Schleif W.A. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. USA., 85, 1988, 4686-4690.
15. Cunha EFF, Sippl W, Ramalho TC, Antunes OAC, Alencastro RB, Albuquerque MG. 3D-QSAR CoMFA/CoMSIA models based on theoretical active conformers of HOE/BAY-793 analogs derived from HIV-1 protease inhibitor complexes. Eur J Med Chem. 44, 2009, 4344-52.
16. Freire E. Overcoming HIV-1 resistance to protease inhibitors. Drug Discov Today, 3, 2006, 281-286.
17. Muzammil S, Ross P, Freire E: A major role for a set of non-active site mutations in the development of HIV-1 protease drug resistance.Biochemistry, 42, 2003, 631-638.
18. Oroszlan S. Biosynthesis and proteolytic processing of retroviral proteins: an overview. Current communications in molecular biology viral protenases as targets for chemotherapy. 1989, 87-100. Cold spring harbor laboratory press, New York.
19. Mager PP. The active site of HIV-1 protease. Med Res Rev, 21, 2001, 348-353.
20. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB, Wlodawer A. "Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution". Science, 246, 1989, 1149–52.
21. Bingding Huang. Meta Pocket: a Meta approach to improve protein ligand binding site prediction, Omics, 13, 2009, 325-330.
22. Trott O. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem., 31, 2010, 455-461.
23. David Lagorce, Olivier Sperandio, Hervé Galons, Maria A Miteva and Bruno O Villoutreix. FAF-Drugs2: Free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC Bioinformatics, 9, 2008, 396.
24. Drwal MN, Banerjee P, Dunkel M, Wettig MR, Preissner R.ProTox: a web server for the *in silico* prediction of rodent oral toxicity. Nucleic Acids Res. 42, 2014, W53-8.
25. Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol. 1, 2004, 337-41.

Source of Support: Nil, **Conflict of Interest:** None.