

# A Combinatorial approach: To design inhibitory molecules on Hemagglutinin protein of H1N1 virus (Swine Flu)

Chekkara Venkata Satya Siva Prasad\*, Kamal Kumar Chaudhary & Parul Dinkar

Division of Applied Sciences & IRCB, Indian Institute of Information Technology, Deoghat, Jhalwa, Allahabad 211012, India; Chekkara Venkata Satya Siva Prasad - Email: shiva@iiita.ac.in; Tel: +91532 2922201; Fax: +91532 2430006; \*Corresponding author

Received June 06, 2013; Accepted June 07, 2013; Published June 29, 2013

## Abstract:

The Hemagglutinin (HA) is a protein of influenza A virus. It is present on the surface of influenza A virus and it is a glycoprotein. The HA is identified as potential drug target. H1N1 thiazolides, proved to be a potent drug in the inhibition of H1N1 replication. It is also known as inhibitor of other strains of influenza A virus. Thiazolide drug represses viral HA's maturation at a level which exists just before the resistance from digestion of endoglycosidase-H and thereby it hampers, HA insertion in host membrane. Blocking the appropriate active site of hemagglutinin protein helps in the disease control. In the present work, we have generated diverse combinatorial library based ligands on known inhibitor thiazolides and they were used for virtual screening by Molegro virtual docker program. K-means clustering approach was used for finding new inhibitory molecules with more appropriate features. These resulted molecules are may be helpful in the treatment of swine flu and many other related diseases.

**Keywords:** Hemagglutinin protein, combinatorial library, virtual screening, Hydrogen bond interaction, K-means clustering.

## Background:

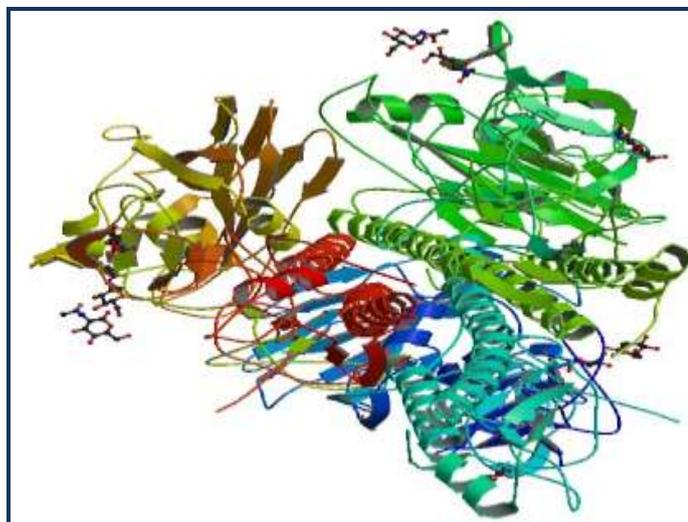
The pandemic flu of influenza A viruses had caused deaths of millions of people Worldwide. It is proved in April 2009 that swine flu is contagious disease and it is spread from human to human [1]. Hemagglutinin (HA) protein of influenza A virus is a surface glycoprotein of virus [2]. Swine flu subtype H1N1 belongs to the Orthomyxoviridae influenza family of viruses [3]. There are 16 different HA antigens subtype present in swine flu [4]. These subtypes are labeled H1 through H16 (H1, H2, H3...H16) [5]. The structure of HA differs from strain to strain. HA occurs as mushroom shaped structure found on the virus surfaces and it is spread 135Å from the membrane [6]. The molecular size of HA is approximately 75kDa [7]. It is an antigenic glycoprotein and having capability for binding the cell of the host. HA glycoprotein binds to sialic acid receptors in the host's membrane cells [8]. After HA binds to these receptors, the viral membranes and cellular membranes are fuse, and the virus is taken entry into the cell with a small envelope of cellular membrane. According to the updates of

WHO (2009 June 10) 27,737 cases of H1N1 had been reported among 74 different countries. The fatality rate of H1N1 is 0.4% (range 0.3%-1.5%) [9]. There are four drugs commonly used in the treatment of the above said disease, Amantadine and Rimantadine drugs worked by blocking the proton channel M2 and two drugs Zanamivir and Oseltamivir inhibit neuraminidase protein function [10]. These antiviral drugs are given to treat those people who become severely ill.

Now a days it has been reported that thiazolide is a potent drug that inhibits the replication of swine flu virus by selectively blocking the maturation of HA, thereby it hampers HA by preventing its insertion in the host's membrane [11]. This study is basically aimed to develop new type of drug molecule for which, we used combinatorial library approach and K-means clustering approaches for finding drug like molecules. The newly designed drug molecules may help in the treatment of swine flu and many other related diseases.

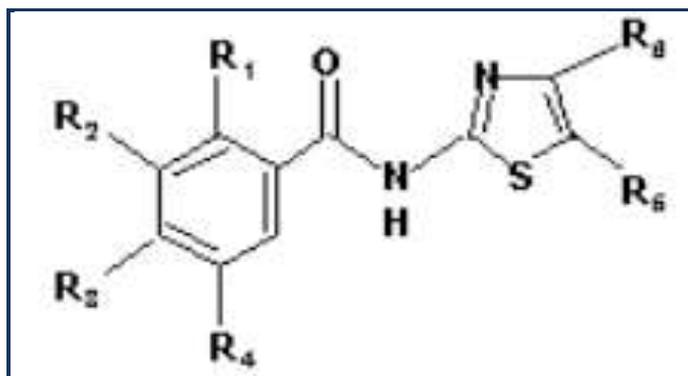
## Methodology:

The structure of HA protein of swine flu H1N1 along with a ligands and water molecules were downloaded from RCSB Protein Data Bank (PDB ID: 1RUY in website <http://www.rcsb.org/pdb>) (Figure 1). 1RUY protein 3D molecules and anti drug molecule thiazolide were used for designing new inhibitor molecules. We took thiazolide as seed molecule and chemical structure was drawn using software ACD ChemSketch into as a mol file (Figure 2).



**Figure 1:** 3D structure of Hemagglutinin protein (Pdb Code: 1RUY).

Then mol file was converted into .sdf file by using Open Babel software. Combinatorial library designed by using known inhibitor thiazolide with the help of random library generation process [12]. For the high quality molecular diversity library generation, ILib diverse software was used [13]. This program was also designed for obtaining compound libraries with an optimal molecular diversity [14] in which molecules were filtered by using the Lipinski's rule of 5 which is the most approved filter for discernment between drug-like and non drug-like molecules [15]. For the conversion of 2D SDFfile format into 3D SDFfile format Corina software was used. It automatically generates three-dimensional atomic coordinates from the structure of a molecule and expresses it as a connection table or linear code, for large databases. The generated 3D library was used for virtual screening for finding new inhibitor for HA [16].



**Figure 2:** 2D structure of Thiazolide.

## Molegro Virtual Docker

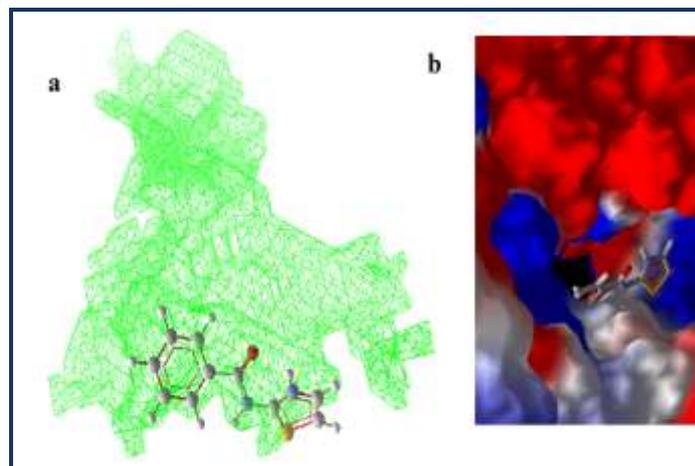
Molegro Virtual Docker (MVD) is software for studying and predicting ligands interaction towards macromolecules. The docking performed by using flexible docking method of MVD [17]. Docking was performed for designed combinatorial library molecules on binding site of 1RUY protein.

## K-mean clustering

On the Top 200 molecules were taken from a combinatorial library in which these molecules having all information about molecular descriptor for K-means clustering purpose [18-19]. These molecules having many descriptors values such as log p value, hydrogen bond acceptor, hydrogen bond donor, topological polar surface area, rotatable bond count, rings numbers and molecular weight etc were calculated. Cluster these molecules by k-mean clustering algorithm. Two descriptors were selected in which molecular weight present on the X-axis and log p value present on the Y-axis. K-means clustering was done by Matlab.

## Results & Discussion:

The structure of HA was taken from protein data bank with the resolution of 2.70Å, R-factor of 0.248 and R-free: 0.265. Ramachandran plot analysis showed about 83.2% residues falling in the most favorable region, 15.2% residues falling in the allowed region and 1% generously allowed regions while 0.7% in the disallowed region.



**Figure 3:** a) Hydrogen bond interaction between Thiazolide and 1RUY; b) Possible binding site for Thiazolide drug in 1RUY

## Docking result of Thiazolide with 1RUY protein

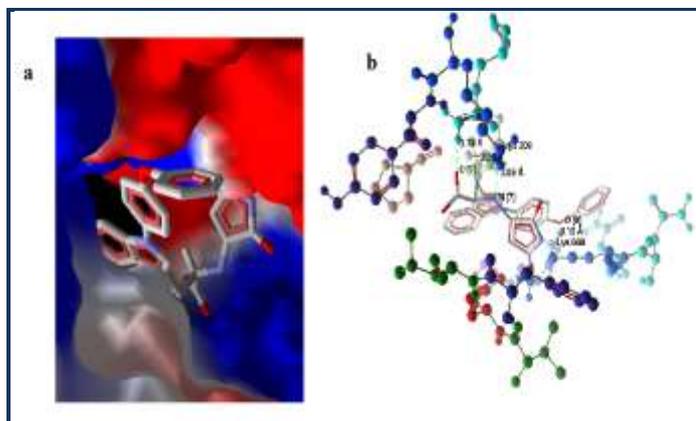
Figure 3b shows thiazolide drug molecule was docked with 1RUY protein, and the energy score comes out -77.5712KJ/mol. (Figure 3a) Illustrate the interaction between thiazolide and the active site of 1RUY. Thiazolide shows one hydrogen bond interaction within the active site of 1RUY. This hydrogen bond is donated by ligand molecule. Hydrogen bond energy of the protein-ligand complex is -1.13817. The residues that are distantly interacted with the ligand are Asp 585 was obtained by Molegro virtual docker.

## Residue involved in active site

Possible binding site of 1RUY protein for design drug like top 10 ligands:

## Molecule\_1899

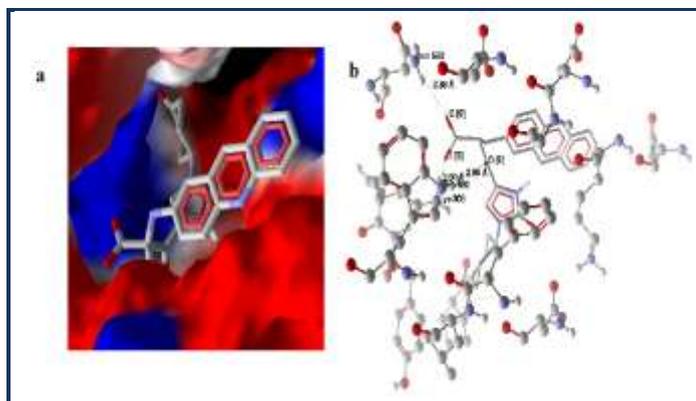
In (Figure 4a), Molecule\_1899 was docked with 1RUY protein, and giving the energy score of -169.672KJ/mol. In (Figure 4b), shows the hydrogen bond interaction between Molecule 1899 and the active site of 1RUY. Molecule\_1899 shows four hydrogen bond interaction and one strong electrostatic interaction within the active site of 1RUY. All the four hydrogen bonds are donated by target molecule (1RUY). This protein-ligand complex is having hydrogen bond energy of -5.01256.



**Figure 4:** a) Possible binding site for Molecule\_1899 in 1RUY protein; b) Hydrogen bond interactions between Molecule\_1899 and 1RUY

## Molecule\_1730

In (Figure 5a), Molecule\_1730 was docked with 1RUY protein, and giving the energy score -162.303KJ/mol. In (Figure 5b), illustrate the interaction between Molecule\_1730 and the active site of 1RUY. Molecule\_1730 shows three hydrogen bond and two strong electrostatic interactions within the active site of 1RUY. All the three hydrogen bonds are donated by target molecule (1RUY) and is having hydrogen bond energy of -5.20299.

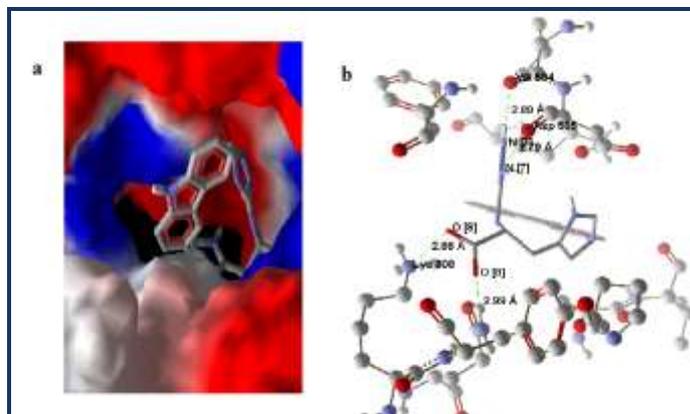


**Figure 5:** a) Possible binding site for Molecule\_1730 in 1RUY protein; b) Hydrogen bond interactions between Molecule\_1730 and 1RUY

## Molecule\_4488

In (Figure 6a), Molecule\_4488 was docked with 1RUY protein, and giving the energy score of -161.455 KJ/mol. In (Figure 6b), it illustrates the interaction between Molecule\_4488 and the active site of 1RUY. Molecule\_4488 shows four hydrogen bond and two strong electrostatic interactions within the active site of

1RUY. Out of four hydrogen bonds, two hydrogen bonds are donated by target molecule and the remaining two hydrogen bonds are donated by ligand molecule. Hydrogen bond energy of the protein-ligand complex is -6.41833. These ten molecules having the high negative energy score compare to thiazolide molecule were docked with 1RUY protein. These ten molecules had energy scores ranging from -151.601 to -169.672KJ/mol



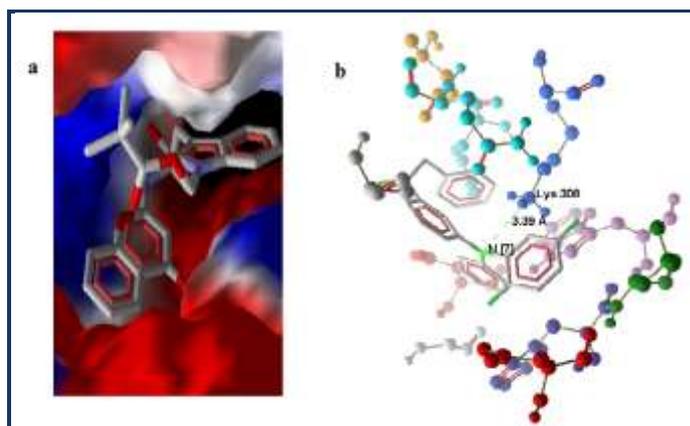
**Figure 6:** a) Possible binding site for Molecule\_4488 in 1RUY protein; b) Hydrogen bond interactions between Molecule\_4488 and 1RUY

## Clustering result

K-means clustering approach used in a top 200 molecules of combinatorial library were found assign five clusters [12]. For each cluster two molecules were represented in all clusters centroid. Which are nearest to the cluster centroid. For k-means clustering were selected descriptors molecular weight and log p value. Select top five ligands which were found from virtual screening by using Molegro program had energy scores ranging from -90.893 to -120.842KJ/mol.

## Molecule\_45

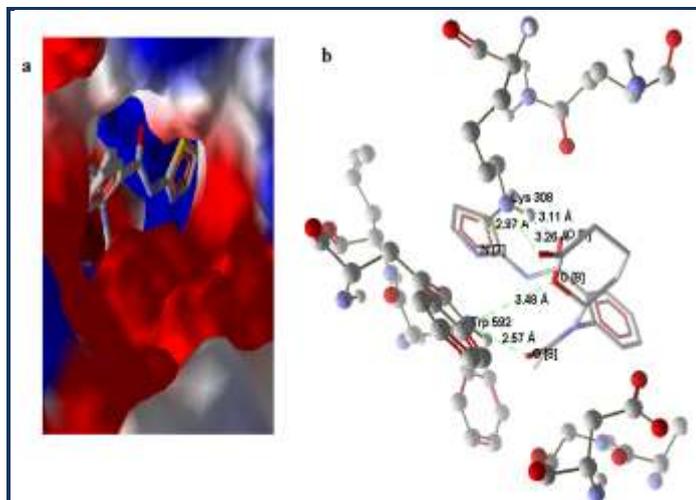
In (Figure 7a) Molecule\_45 was docked with 1RUY protein, and giving the energy score -120.842KJ/mol. In (Figure 7b), illustrate the interaction between Molecule\_45 and the active site of 1RUY. Molecule\_45 shows one hydrogen bond interactions within the active site of 1RUY. This hydrogen bond is donated by target molecule (1RUY) and energy of this hydrogen bond is -0.328907.



**Figure 7:** a) Possible binding site for Molecule\_45 in 1RUY protein; b) Hydrogen bond interaction between Molecule\_45.

## Molecule\_77

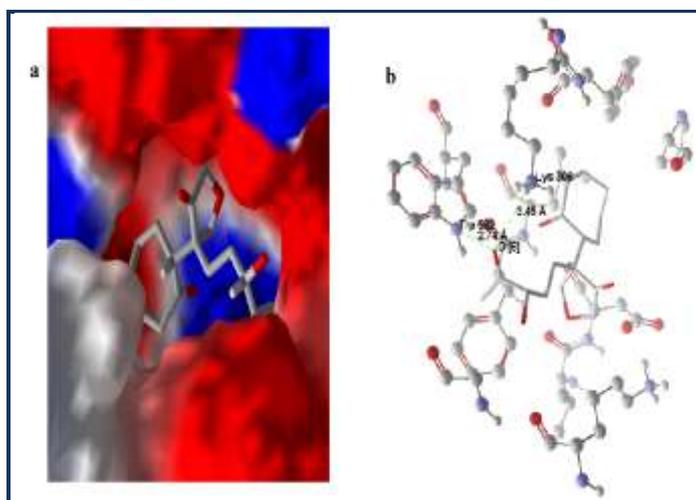
In (Figure 8a), Molecule\_77 was docked with 1RUY protein, and giving the energy score -106.101 KJ/mol. In (Figure 8b), it illustrates the interaction between Molecule\_77 and the active site of 1RUY. Molecule\_77 shows five hydrogen bond interactions within the active site of 1RUY. All five hydrogen are donated by target molecule (1RUY). Hydrogen bond energy of this protein-ligand complex is -5.17897.



**Figure 8:** a) Possible binding site for Molecule\_1779 in 1RUY protein, b) Hydrogen bond interactions between Molecule\_77 and 1RUY.

## Molecule\_35

Figure 9a Molecule\_35 was docked with 1RUY protein, and giving the energy score -92.996KJ/mol. In (Figure 9b), it illustrate the interaction between Molecule\_35 and the active site of 1RUY. Molecule\_35 shows 2 hydrogen bond interactions within the active site of 1RUY. All the hydrogen bonds are donated by target molecule (1RUY). The hydrogen bond interaction energy of this protein-ligand complex is -0.517533. After docking, the above said top five molecules were found with better energy scores comparison the energy score of thiazolide drug molecule.



**Figure 9:** a) Possible binding site for Molecule\_35 in 1RUY protein; b) Hydrogen bond interactions between Molecule\_35 and 1RUY.

The structure of hemagglutinin (HA) protein along with ligands and water molecules are taken from PDB (PDB Code: 1RUY) [20]. To prevent H1N1 replication thiazolide, proved to be a potent drug. Thiazolide drug represses viral HA's maturation at a level which exists just before the resistance from digestion of endoglycosidase-H and thereby it hampers HA preventing its insertion in host membrane. Hence we have used this protein structure for virtual screening to find out thiazolide specific new inhibitors [21]. We first built a diverse combinatorial library of 5000 molecules by incorporating the features of known inhibitor [22]. This library was pruned by using Lipinski rule of 5 [23]. Each compound was automatically docked into the binding region of protein and given a score. These scores were given by the quality of fitting compound to the target site. So after careful docking generated more ligands in which top 10 ranking ligands displaying the lowest predicted energy were selected. These molecules having more negative energy score (KJ/mol).

After docking of thiazolide antidrug molecules with 1RUY, the energy score was found to be -77.5712 KJ/mol. Top ten ligands which were found from docking had energy scores ranging from -151.601 to -169.672 KJ/mol. These 10 molecules had better energy compare to thiazolide energy score in Table 2 (see supplementary material). Then we checked the hydrogen bond interaction between 1RUY protein and top10 molecules were obtained by Molegro virtual docker. K-means clustering approach was used for finding more new inhibitor by virtual screening. After this process top 5 ligands were selected. These 5 molecules had energy score ranging from -90.893 to -120.842 KJ/mol. These molecules had also high negative energy compare to thiazolide, in Table 2 (see supplementary material). Then we had checked the hydrogen bond interaction between 1RUY protein and top 5 molecules. More negative the energy score (KJ/mol) more is the binding affinity. These new ligands molecules will help for the prevention of swine flu and many other related diseases.

## Conclusion:

It was found that drug like molecules for the inhibition of hemagglutinin, thiazolides antidrug molecule play an important role for designing a new novel drug like molecules. The structure of HA (PDB Code: 1RUY) was considered as target receptor. HA protein was used for virtual screening for generating new ligands. These ligands having better energy scores than known inhibitor thiazolides by generated diverse combinatorial library ILib diverse and docking based virtual screening of this library by using the Molegro Virtual Docker program. In which top 10 ligands having better energy scores compared to thiazolide. K-mean-clustering approach was also used and found top five molecules having energy ranging from -90.893 to -120.842 KJ/mol. Energies of these molecules were also high in compare to thiazolide. These ligands may function as a potent and specific inhibitor for HA and will help the prevention of swine flu and other related diseases.

## References:

- [1] Majhdi AI *et al.* *Science Directs.* 2009 **7**: 113
- [2] Rossignol JF *et al.* *JBC Papers.* 2009 **284**: 297 [PMID: 19638339]
- [3] Abhilash K & Nandhini K, *International Journal of Pharma Sciences and Research.* 2010 **1**: 40

- [4] Russella J Rupert *et al.* *PANS*. 2008 **105**: 46 [PMID: 19004788]
- [5] Bedell CR *et al.* *J Pharmacol*. 1976 **57**: 201 [PMID: 938794]
- [6] Jie Yang *et al.* *Viruses*. 2013 **5**: 352 [PMID: 23340380].
- [7] Kashyap AK *et al.* *PLoS Pathog*. **6**: e1000990 [PMID: 20628565]
- [8] Igarashi M *et al.* *PLoS ONE*. 2010 **5**: e8553 [PMID: 20049332]
- [9] Rungrotmongkol T *et al.* *Curr Pharma Des*. 2011 **17**: 1720 [PMID: 21619529]
- [10] De Waterbeemed *et al.* *Pure and Applied Chemistry*. 2004 **69**: 1137
- [11] Dr Richard M, Casey B – *eye*. 2006
- [12] Rajashree Dash *et al.* *International Journal of Engineering Science and Technology*. 2010 **2**: 59
- [13] Gillet J Valerie, *Chemical Data Analysis Large*. 2000
- [14] Graham Richards W, *Pure Appl Chem*. 1994 **66**: 1589
- [15] Clark DE & Pickett SD, *Drug Discovery Today*. 2000 **5**: 49 [PMID: 10652455]
- [16] CLMJ & Hol WG, *Science directs*. 2004 577
- [17] Lowe Gordon *et al.* *Pure Appl Chem*. 1998 **70**: 11.
- [18] Kouji Kitajima & Shander Ahmed, *Genome Informatics*. 2002 **2**: 59
- [19] Thierry Langer & Gerhard Wolber, *Pure and Applied Chemistry*. 2004 **76**: 991
- [20] <http://www.rcsb.org/pdb>
- [21] Wenlock JMC *et al.* *Med Chem*. 2003 **146**: 1250 [PMID: 12646035]
- [22] Lipinski CA *et al.* *Adv Drug*. 2001 **46**: 3 [PMID: 11259830]
- [23] Tao P & Lai L, *J Comput aided Mol Des*. 2001 **15**: 429 [PMID: 11394737]

Edited by P Kanguane

Citation: Prasad *et al.* *Bioinformation* 9(11): 565-571 (2013)

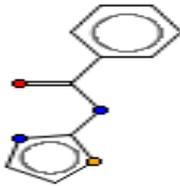
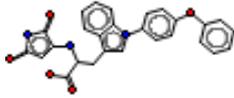
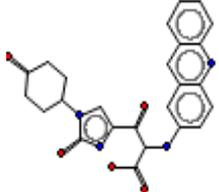
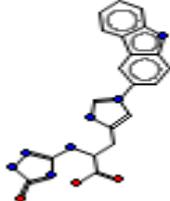
**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

## Supplementary material:

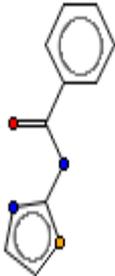
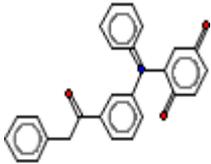
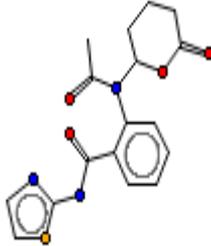
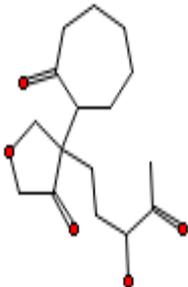
**Table 1:** Residue involved in active site

S.No.	Residues ( amino acid)	Bond Length
1899	Gln562,Lys 308,Lys 568	2.79 Å,2.36 Å-3.59,3.10 Å
1730	Asn560,Trp 592, Lys 308	3.38 Å,2.90 Å ,2.60 Å,
4488	Val584,Asp 585,Lys 308	2.89 Å,2.79 Å,2.88 Å ,2.99 Å
2830	Lys 308	3.11 Å
2977	Asp 590,Phe 563	3.02 Å ,2.94 Å ,
21	Asn 560,Lys 308,Gly 587,	307 Å ,260 Å,3.10 Å
1194	Gln 562, Lys 308, Trp 592, Leu 589, Phe 588, Lys 592.	3.27 Å ,3.22 Å , 2.55 Å ,3.33 Å ,3.50 Å ,2.60 Å.
1702	Lys 308	2.77Å, 3.03Å
1779	Lys 308,Trp 592	3.09 Å ,.6 Å
45	Lys 308	3.39Å
77	Trp 592, Lys 308	( 3.48Å ,2.57Å ) ,(2.97Å,3.11 Å , 3.26 Å )
35	Trp 592,Lys 308	2.74Å, 3.45 Å
187	Lys 583	3.14 Å
69	Lys 308	2.86 Å

**Table 2:** Comparison of energy score of 1RUY with drug like top ligands after Docking

Molecules	2D structure of molecules	E-Inter	E-Intar	Energy score KJ/mol
Thiazolide (Reference Molecule)		-88.9887	11.4175	-77.5712
Molecule_1899 (2-[(2,5-dioxopyrrol-3-yl)amino]-3-[1-(4-phenoxyphenyl)-1H-indol-3-yl]propanoic acid)		-186.703	17.0314	-169.672
Molecule_1730 (2-(acridin-2 ylamino)-3-oxo-3-[2 oxo-1-(4oxocyclohexyl)imidazol-4-yl] propanoic acid)		-171.222	8.91901	-162.303
Molecule_4488 (1-(carbazol-3-yl)-4-[2-carboxy-2-[(3-oxo-2H- ( 2, 4-triazol-5-yl) amino] ethyl]-3H, (3-imidazol-1-ylum)		153.1	8.35533	-161.455

**Table 3:** Comparison of energy score of 1RUY with nearest two molecules for each cluster centroid after docking was found top molecules

Name	2D structure of molecules	E-Inter	E-Intar	Energy score KJ/mol
Thiazolide (Reference Molecule)		-88.9887	11.4175	-77.5712
<b>Molecule_45</b> (2-[[3-(2-phenylacetyl) phenyl] (phenylidene) amino] cyclohexa-2, 5-diene-1, 4-dione)		-141.589	20.7473	-120.842
<b>Molecule_77</b> (2-[N-(6-oxooxan-2-yl) acetamido]-N-(1, 3-thiazol-2-yl) benzamide)		-115.915	27.0106	-88.9046
<b>Molecule_35</b> (4-(3-hydroxy-4-oxopentyl)-4-(2-oxocycloheptyl) oxolan-3-one)		-109.183	20.985	-88.1978