Available online at www.derpharmachemica.com

Scholars Research Library

Der Pharma Chemica, 2014, 6(4):80-89
(http://derpharmachemica.com/archive.html)

ISSN 0975-413X
CODEN (USA): PCHHAX

3D-QSAR analysis on 6-(1-benzyl-1*H*-pyrrol-2-yl)-2, 4-dioxo-5-hexenoic acid derivatives as recombinant HIV-1 integrase inhibitors

Bawneet K. Narang[1], Virender Singh[2], Manish K. Gupta[1] and Ravindra K. Rawal[1*]

[1]*Department of Pharmaceutical Chemistry, Indo-Soviet Friendship College of Pharmacy, Ferozepur Road, Ghal Kalan, Moga, Punjab, India*
[2]*Department of Chemistry, Dr. B. R. Ambedkar National Institute of Technology (NIT), Jalandhar, Punjab, India*

ABSTRACT

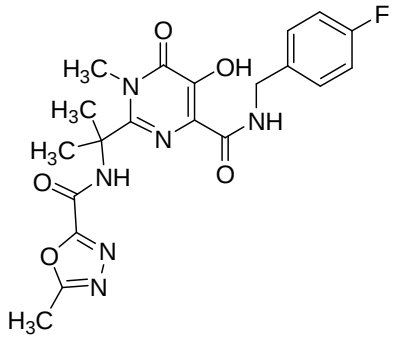
HIV-1 integrase is a fascinating target for designing of novel HIV-1 Integrase inhibitors. Due to the development of resistance by the use of already developed inhibitors the novel inhibitors are designed that can target Integrase with higher selectivity and less toxicity profiles. The present work describes the 3D QSAR studies on series of 6-(1-Benzyl-1H-pyrrol-2-yl)-2, 4-dioxo-5-hexenoic acids for establishing quantitative relationship between biological activity and their physicochemical properties. This study was performed with 47 compounds (data set) using manual selection method and simulated annealing algorithm for the division of the data set into training and test set. In this analysis, three statistical significant models were obtained using PLS as the statistical method. The most significant model is having (q^2 = 0.608) and (pred_r^2 =0.699). Model showed that steric (S_881), (S_184) and electrostatic (E_496) interactions play important role in modulating the HIV-1 integrase inhibitory activity.

INTRODUCTION

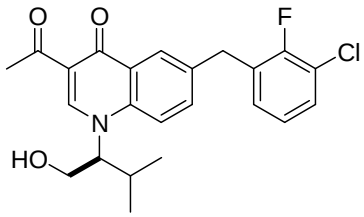
Acquired immuno deficiency syndrome (AIDS) is a major global public health issue of the 21st century that is characterized by a cascade of infections resulting from the damage to the immune system caused by Lentivirus i.e. human immunodeficiency virus (HIV) belonging to the family Retroviridae [1,2]. AIDS is the most serious stage of HIV due to which the body becomes susceptible to numerous infections and diseases. The number of T cells that plays a major role in the immune system is declined to a severe level with the alteration in the immune system [3]. In the HIV life cycle the enzymes which play a vital role are reverse transcriptase (RT), protease (PR) and integrase (IN). With the help of the reverse transcriptase (RT) there is the transcription of the viral RNA genome into viral complementary DNA (cDNA). The IN helps in the insertion of viral cDNA into the host cell genome and the cleavage of newly synthesized polypeptides into single viral proteins occurs by viral PR [4]. Since last decades, the RT and PR are seen as the main targets for the inhibition of HIV replication.

The regimen known as highly active antiretroviral therapy (HAART) for the treatment of HIV includes the use of nucleoside RT inhibitors (NRTIs), non-nucleoside RT inhibitors (NNRTIs) and protease inhibitors (PRIs) resulting in the resistance due to the long term use. So a remedial intervention is needed to stop the epidemic completely from the roots [5]. Hence, as a result the enzyme IN has emerged as an attractive target for discovery of safe, efficient and effective HIV-IN inhibitors [6]. There is no known human counterpart to integrase hence integrase as a target is an attractive approach which will prove to be more advantageous over existing therapies due to high selectivity and less toxicity profiles [7]. Moreover, there is an assumption that the addition of the HIV-IN inhibitor into the regimen can

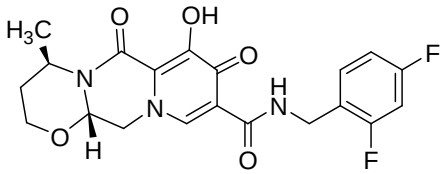
increase the life expectancy with less expensive therapy in comparison to the HAART regimen. Among the drugs currently used in HAART regimens, till date Raltegravir, Elvitegravir and Dolutegravir have got approval for their use in antiretroviral therapy (ART) by the US Food and Drug administration (FDA) and European Medicines Agency (Fig.1) [8].



Raltegravir (1)



Elvitegravir (2)



Dolutegravir (3)

Fig.1- FDA Approved HIV-1 Integrase inhibitors

Amongst these drugs raltegravir was first potent HIV-1 integrase inhibitor approved at the end of 2007 and was clinically effective against viruses resistant to other classes of antiretroviral agents [8] but had number of reported side effects such as rashes, swelling of lips, swelling of face, trouble sleeping etc. [9]. Its long use resulted in resistance due to the number of mutations like E92Q, G140S, Q148H, N155H, and E157Q. The drawbacks of raltegravir led to the development of another first-generation IN inhibitor i.e. elvitegravir by Gilead Sciences. Side effects from its use may include kidney failure, bone problems, an amplified quantity of fat in the head and neck, and changes in the immune system resulting in a condition called immune reconstitution syndrome making the condition of the patient worse [10]. But the need to arrest the viral replication in an efficient manner; the evolution of next generation IN inhibitor with improved features of dose and resistance led to the development of second generation IN inhibitor in August 2013 by the US FDA i.e. Dolutegravir.

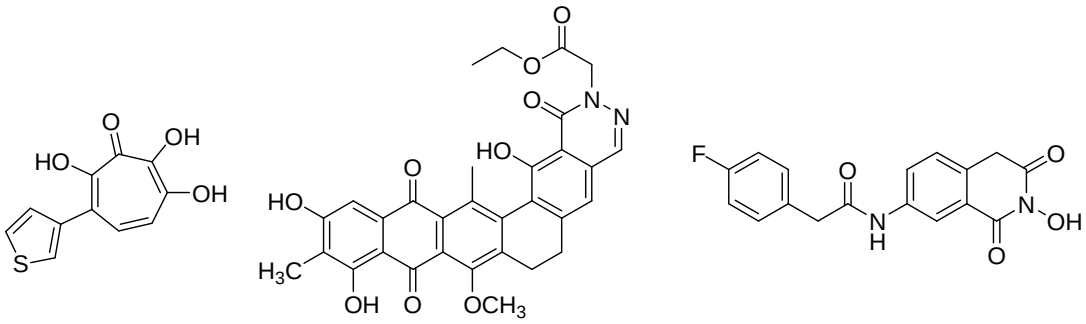
IN belongs to the polynucleotidyl transferase family sharing the structural similarities of its active site with the active site of HIV-1 reverse transcriptase (RT)-associated ribonuclease H (RNase H) [11]. As depicted through number of studies the resistance and mutations has occurred due to the long term use of the already approved drugs having integrase inhibitory action. These trends have accentuated the need for new, safe and effective agents. The keenness for breakthrough in this field led to the monitoring of drugs having HIV Integrase inhibitory action.

The QSAR studies generate a model that seeks to develop mathematical relationships between chemical structure and biological activity of the compound. This approach does not take into account any experimental properties but the molecular descriptors, which needs the availability of the molecular structure single-handedly. On the establishment of a reliable model inhibitory activity of compounds is predicted and the structural factors that influence the activity are envisaged. Here we have undertaken MFA based 3D QSAR study and the results are presented here.

MATERIALS AND METHODS

Dataset and Data Selection

6-(1-Benzyl-1*H*-pyrrol-2-yl)-2, 4-dioxo-5-hexenoic acid derivatives and their HIV -1 IN inhibitory activities against ST were taken from the literature reported by Roberta Costi et al. [12]. The compounds included diketo acid (DKA) derivatives like 4-[5- (benzoylamino)thien-2-yl]-2,4-dioxobutanoic acid and DNA aptamers. The different kind of structurally different compounds as inhibitors were reported including tropolones (**4**), madurahydroxylactone derivatives (**5**) and 2-hydroxyisoquinolin- 1, 3(2*H*, 4*H*)-diones (**6**) as shown in Fig.2.



4 **5** **6**

Fig. 2- Tropolones (4), Madurahydroxylactone (5), 2-hydroxyisoquinolin- 1, 3(2*H*,4*H*)-diones (6)

A data set of 47 molecules (n=47) are considered for the analysis along with their reported strand transfer inhibitory activity (IC_{50}, inhibitory concentration, in µM) expressed in terms of logarithm of inverse of their inhibitory concentration (-log IC_{50}) as shown in **Table 1**. The activity data (IC_{50}) was given in µM value. All values have been converted to the logarithmic scale [pIC50] and the data saved as .txt file. After the conversion, structures are saved as .mol2 file in QSAR Plus 3D window. The analogues as reported in **Table 1** have undergone 3D QSAR studies for drug design along with their general core structure using V Life Molecular Design Suite (MDS) Software. All the molecules in series were aligned based on atom overlapping alignment. The derived 3D-QSAR PLS model for the chemical series reported as anti-HIV integrase agents give insight into the influence of various interactive fields on the activity and thus help in designing the HIV integrase inhibitory activity of novel compounds.

The core 3D structures have been created and generated using Molecular Operating Environment (MOE) which were subjected to energy minimization using force field: Merck Molecular Force Field (MMFF94x) and Austin Model 1 (AM1) with 0.01 and 0.001 kcal/mol as the energy gradients. The force field was applied in order to obtain and ensure the minimum energy conformer. All the other structures mentioned in **Table 1** were generated in MOE using the force fields as mentioned above with the same gradients by modifying the core structure. All these energy minimized structures of compounds reported in **Table 1** have been harbored to the 3D-QSAR of QSARPLUS module for computational studies, hence computing the parameters.

Descriptors Evaluation

In these steric, electrostatic and hydrophobic parameters have been calculated using V Life MDS Software.

Model Development

In 3D-QSAR analysis for model development three methods i.e. Random selection method, Manual data selection method, Sphere Exclusion method were used for creation of training and test set. In this paper we have selected Manual data selection method and Random Selection method by conflicting dispartate values. After the creation of training and test set, the statistical method like PLS (Partial Least Square) have been used for model building. PLS method helps in the establishment of relationship between the dependent variables (biological activity) and the independent parameters (electrostatic, steric and hydrophobic fields).

Variable Selection

In PLS method Stepwise variable selection is done. Simulated annealing method is used for 3D-QSAR analysis.

RESULTS AND DISCUSSION

The various regression methods were performed on all the 47 molecules of the selected series for QSAR analysis. The following considerable and significant 3D-QSAR models with equations were obtained using the alignment of energy minimized confirmations of diketo acid derivatives and PLS method as mentioned above.

Model 1

$$-\log IC_{50} = (8.368)\ S\_716 + (0.719)\ E\_496 + (19.810)\ S\_861 - (3.926)\ H\_285 + 11.01$$
$$n = 38,\ r^2 = 0.712,\ q^2 = 0.624,\ r^2\ T = 0.625,\ \text{F-Test} = 43.294$$

Here n is the number of compounds in training set and r^2 T is the r^2 of test set compounds. The equation of model 1 shows the effects of different parameters at grid points 716, 496, 861 and 285. The steric and electrostatic parameters give the positive contribution at points 716, 861 and 496 whereas the contribution by the hydrophobic parameter is negative at point 285. Figure 3a shows that S_861 correspond to the substitution at X of –COOX and S_716 signifies the position 2nd of the diketo acid chain. The E_496 indicates the electrostatic effect at the position linking N of the pyrrol and the phenyl ring whereas the negative contribution by hydrophobic parameter is unfavorable at R_4 substitution of the phenyl ring. The compounds in the test set were 9. The linear regression is obtained by fitting a best fit straight line to the data. The input parameters included cross correlation limit 0.8; iteration at given temperature 10; seed 1; decrease in temperature 1; perturbation limit 2; terms in model 4; variance cut off 0; auto scaling and four number of components.

Model 2

$$-\log IC50 = (92.520)\ S\_881 + (35.297) S\_184 + (0.065) E\_496 - (9.640) H\_71$$
$$n = 38,\ r^2 = 0.686,\ q^2 = 0.6990,\ r^2\ T = 0.608,\ \text{F-Test} = 38.352$$

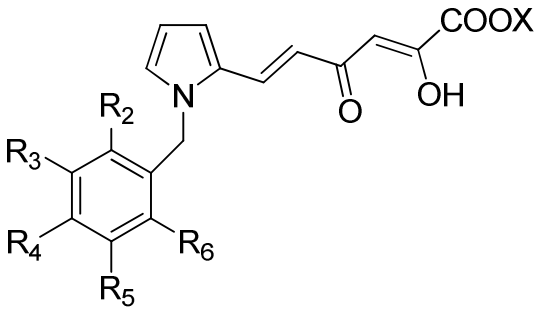
The equation of the model shows the effects of different parameters at grid points 71, 881, 184 and 496. In this equation also the steric and electrostatic parameters the positive contribution is provided at points 881,184 and 496 whereas the contribution by the hydrophobic parameter is negative at point 71. Figure 3b shows that S_881 correspond to the substitution at X of –COOX and S_184 signifies the substation at R_4 of the phenyl ring. The E_496 indicates the electrostatic effect at the position linking N of the pyrrol and the phenyl ring whereas the negative contribution by hydrophobic parameter is unfavorable at R_5 substitution of the phenyl ring. The compounds in the test set were 9. The linear regression is obtained by fitting a best fit straight line to the data. The input parameters included cross correlation limit 0.5; iteration at given temperature 10; seed 2; decrease in temperature 2; perturbation limit 2; terms in model 4; variance cut off 0; auto scaling and four number of components.

Model 3

$$-\log IC_{50} = (69.42)\ S\_94 + (290.61)\ S\_891 + (0.065) E\_496 - (5.420) H\_192$$
$$n = 38,\ r^2 = 0.687,\ q^2 = 0.598,\ r^2\ T = 0.831,\ \text{F-Test} = 37.473$$

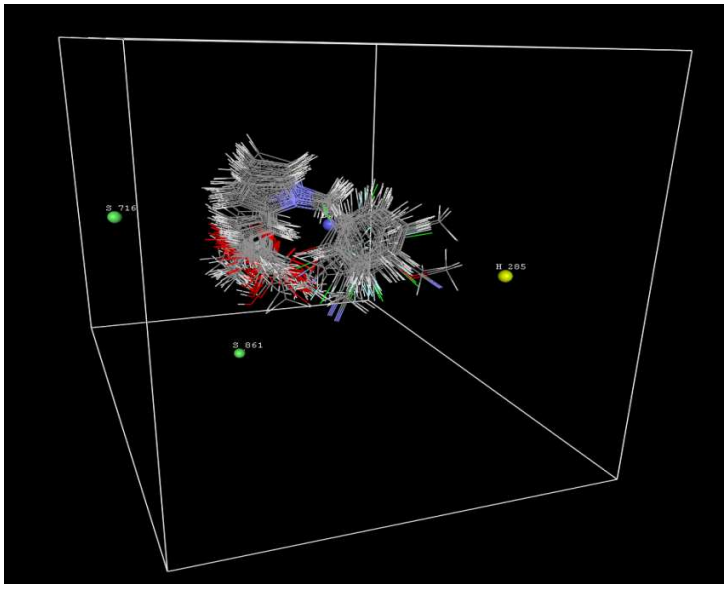
The equation of model 3 shows the effects of different parameters at grid points 94,891,496 and 192. This equation signifies the positive contribution by steric and electrostatic parameters at points 94,891 and 496 whereas the contribution by the hydrophobic parameter is negative at point 192. Figure 3c shows that S_891 correspond to the substitution at X of –COOX and S_94 signifies the substation at R_4 of the phenyl ring. The E_496 indicates the electrostatic effect at the position linking N of the pyrrol and the phenyl ring whereas the negative contribution by hydrophobic parameter is unfavorable at R_5 substitution of the phenyl ring. The linear regression is obtained by fitting a best fit straight line to the data. The compounds in the test set were 10. The input parameters included cross correlation limit 0.7; iteration at given temperature 10; seed 2; decrease in temperature 1; perturbation limit 2; terms in model 4; variance cut off 0; auto scaling and four number of components. The test set prediction by equations 1, 2 and 3 is in agreement with their experimental values as shown in Table 1.

Table 1- Observed and Predicted activity of diketo acid derivatives

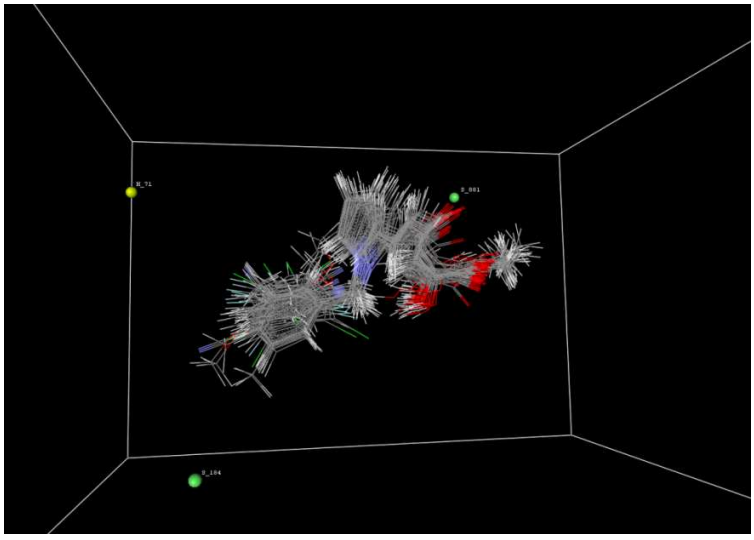


	Substitutions							-log IC_{50}			
Compound	R_2	R_3	R_4	R_5	R_6	X	IC_{50}	Observed	EMC (Model1)	EMC (Model2)	EMC (Model3)
01	H	H	H	H	H	Et	15	4.823	4.778	5.0028	5.085
02	Me	H	H	H	H	Et	32	4.494	4.315	3.961	4.153
03	H	Me	H	H	H	Et	8	5.096	4.742	4.742	4.627
04	F	H	H	H	H	Et	1	6.000	5.086	5.444	5.670
05 [a,b,c]	H	F	H	H	H	Et	11	4.958	5.152	5.440	5.355
06	H	H	F	H	H	Et	98	4.008	5.013	5.127	5.105
07	H	Cl	H	H	H	Et	6	5.221	4.916	5.149	4.139
08	H	H	Cl	H	H	Et	42	4.376	4.535	4.840	4.813
09	CN	H	H	H	H	Et	9	5.045	4.930	5.379	5.245
10	H	CN	F	H	H	Et	13	4.886	5.032	5.079	4.871
11	OMe	H	H	H	H	Et	23	4.638	4.938	4.856	4.676
12 [a,b,c]	H	H	OMe	H	H	Et	110	3.958	4.733	4.011	3.855
13	OEt	H	H	H	H	Et	12	4.920	4.998	4.807	4.798
14	H	Me	H	H	H	Et	19	4.721	4.571	4.424	4.563
15	F	F	H	H	H	Et	0.53	6.275	6.168	5.974	5.938
16	F	H	F	Me	H	Et	1	6.000	6.012	5.764	5.817
17 [a,b,c]	F	H	H	H	H	Et	0.45	6.346	5.546	5.577	5.924
18	F	H	H	H	H	Et	4.00	5.397	6.315	6.460	6.436
19	H	F	F	F	H	Et	0.60	6.221	6.011	5.928	5.934
20	H	F	H	H	F	Et	0.49	6.309	5.891	6.179	6.002
21	Cl	H	H	H	H	Et	1.70	5.769	5.729	5.588	5.490
22 [a,c]	H	Cl	H	F	H	Et	8.00	5.096	5.167	5.412	5.352
23	H	H	H	H	H	H	0.09	7.045	6.408	6.602	6.078
24	Me	H	H	H	Cl	H	0.17	6.769	6.258	6.330	6.043
25 [c]	H	Me	H	Cl	H	H	1.3	5.886	6.242	6.326	6.146
26	H	H	Me	H	H	H	1.2	5.920	6.003	5.782	5.777
27 [a,b,c]	F	H	H	H	H	H	0.98	6.008	6.699	6.710	6.807
28	H	F	H	H	H	H	0.92	6.036	6.349	6.566	6.407
29	H	H	F	H	H	H	0.026	7.585	6.269	6.345	6.432
30 [a,b,c]	H	Cl	H	H	H	H	0.31	6.508	5.928	5.989	6.009
31	H	H	Cl	H	H	H	4.1	5.387	5.663	5.794	5.827
32	CN	H	H	H	H	H	6	5.221	6.020	6.358	6.318
33	H	CN	H	H	H	H	0.75	6.124	5.974	6.146	6.083
34 [a,b,c]	H	H	CN	H	H	H	1.7	5.769	6.136	5.802	5.685
35	OMe	H	H	H	H	H	0.53	6.275	6.066	5.930	5.974
36	H	H	OMe	H	H	H	4.1	5.387	5.971	5.263	5.283
37	OEt	H	H	H	H	H	0.31	6.508	6.544	6.473	6.694
38	H	Me	H	Me	H	H	1.6	5.795	5.715	5.573	5.770
39 [a,b,c]	F	F	H	H	H	H	0.059	7.229	6.534	6.487	6.632
40	F	H	F	H	H	H	0.042	7.376	6.922	6.859	6.957
41 [c]	F	H	H	F	H	H	0.052	7.283	7.310	7.207	7.339
42	F	H	H	H	F	H	0.15	6.823	6.569	6.577	6.570
43	H	F	F	H	H	H	1.60	5.795	6.428	6.320	6.367
44 [a,b,c]	H	F	H	F	H	H	1.2	5.920	6.223	5.884	5.979
45	Cl	H	Cl	H	H	H	4.9	5.309	5.965	5.704	5.785
46	Cl	H	H	H	Cl	H	0.17	6.769	6.810	6.324	6.476
47 [b]	H	Cl	H	Cl	H	H	0.97	6.013	6.037	6.010	6.279

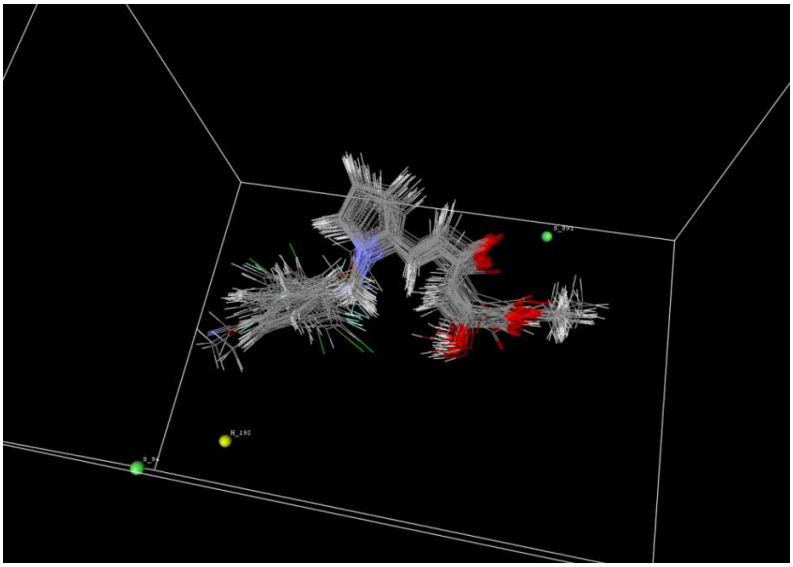
[a,] test set compound in model 1; [b,] test set compound in model 2; [c,] test set compound in model 3.



Model 1(Fig. 3a)



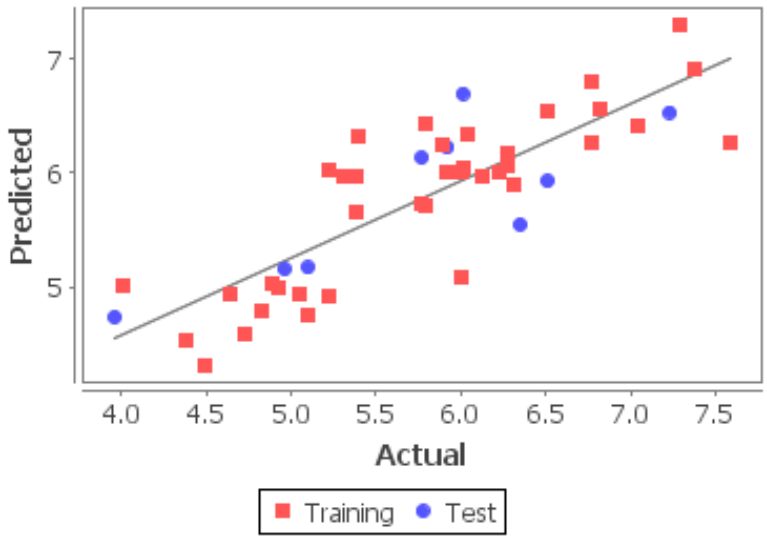
Model 2 (Fig. 3b)



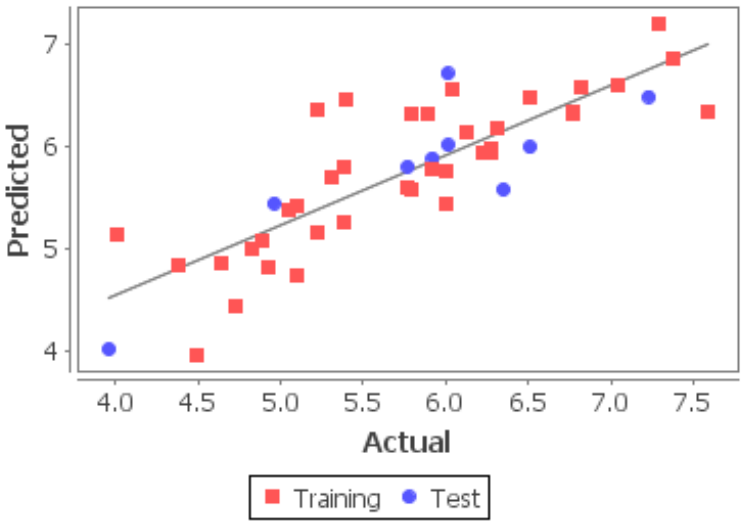
Model 3 (Fig. 3c)

Fig.3- Molecular rectangular field grid displaying the effect of data points (s: steric effect, E: electrostatic effect, H:hydrophobic effect)

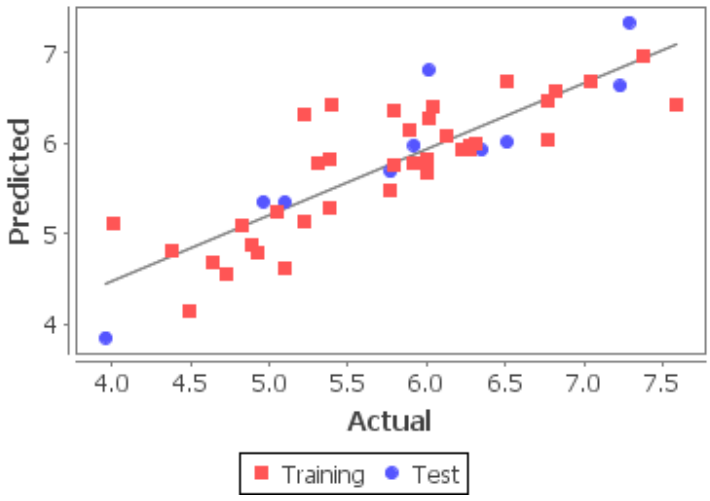
The fitness plot and the plot between the predicted and the actual biological activity for the test set and the training set for the equation 1, 2 and 3are shown in figure 4 and 5 respectively.



Model 1

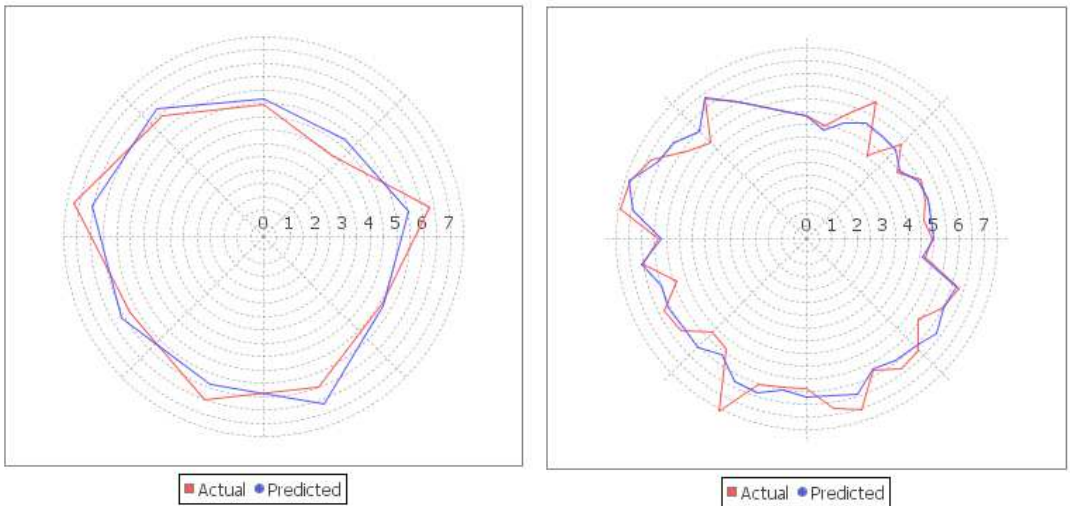


Model 2

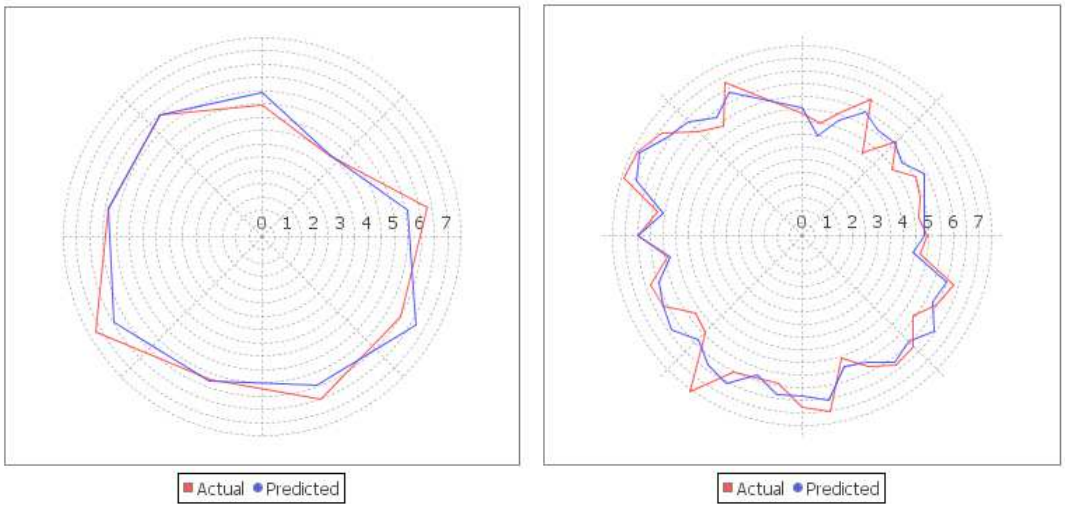


Model 3

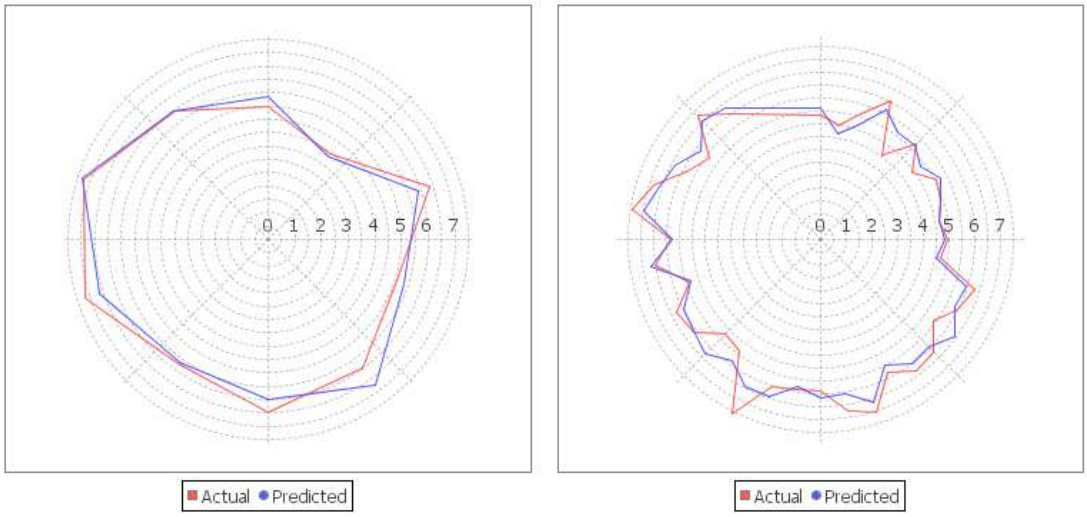
Fig.4- Fitness graphs between the actual and the predicted biological activities for equation 01, 02 and 03 respectively



Model 1



Model 2



Model 3

Fig.5- Graph between actual and predicted biological activity for test set and training set

Acknowledgements
We are heartly thankful to GOD for giving such a great opportunity. After that we are grateful to Shri Parveen Garg (Chairman), I.S.F. College of Pharmacy, Ferozepur Road, Moga 142001, INDIA for their necessary support and motivation.

REFERENCES

[1] R. J. Hu, J. P. Doucet, M. Delamar, R. S. Zhang, *Eur. J. Med. Chem*., **2009,** 44, 2158-2171.
[2] K. Ryan, C. Ray, J. Sherris, *J. Med. Microbiol*., **2004**, 4th ed., 481-491.
[3] C. Gallo Robert, L. Montagnier, *Engl. J. Med*., **2003**, 349, 2283-2285.
[4] S. Broder, AS. Fauci, *Public Health Rep*., **1988**, 103(3), 224–229.
[5] C. Katlama, S. G. Deeks, B. Autran, J. Martinez-Picado, J. van Lunzen, C. Rouzioux, M. Miller, S. Vella, J. E. Schmitz, J. Ahlers, *The Lancet*, **2013**, 381, 9883, 2109 – 2117.
[6] D. J. Hazuda, P. Felock, M. Witmer, A. Wolfe, K. Stillmock, J. A. Grobler, A. Espeseth, L. Gabryelski, W. Schleif, C. Blau, *Science*, **2000,** 287 (5453), 646-650.
[7] G. Zhao, C. Wang, C. Liu, H. Lou, *Mini- Rev. in Med. Chem.*, **2007,** *7* (7), 707-725.
[8] E. A. Semenova, C. Marchand, Y. Pommier, *Adv.Pharmacol.*, **2008**, 56, 199-228.
[9] J. J. Eron Jr, J. K. Rockstroh, J. Reynes, J. Andrade-Villanueva, J. V. Ramalho-Madruga, L.G. Bekker, B. Young, C. Katlama, J. M. Gatell-Artigas, J. R. Arribas, *Lancet Infect. Dis.*, **2011,** 11 (12), 907-915.
[10] A. Tappuni, *Adv. Dent. Res*., **2011,** 23 (1), 90-96.
[11] F. Esposito, A. Corona, E. Tramontano, *Mol. Biol. Int*., **2012**, 586401.
[12] R. Costi, M. Métifiot, F. Esposito, G. C. Crucitti, L. Pescatori, A. Messore, L. Scipione, S. Tortorella, L. Zinzula, E. Novellino, Y. Pommier, E. Tramontano, C.Marchand, R. Santo, *J. Med. Chem.*, **2013**, 56, 8588−8598.