Journal of Computational Methods in Molecular Design, 2016, 6 (3):8-19

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Monte Carlo method based QSAR and docking studies of Pyrazoline and Benzoxazole derivatives as antitubercular agents

***Kalpana Devi, Monica Kachroo and Lokesh Pathak**

Department of Pharmaceutical Chemistry, Al-Ameen College of Pharmacy, Bangalore, India

---

ABSTRACT

World Health Organization has reported that 14 million people worldwide are infected with active tuberculosis and over 1.7 million deaths occur every year. There are many drugs available in the market for treating tuberculosis, but the emergence of tuberculosis is due to the appearance of Multi Drug Resistance (MDR) against one or more of the 1st line antimycobacterial drug. Therefore, there is a need to explore and develop newer structural moiety as antitubercular drug. In the present study CORAL software was used for constructing large-scale QSAR models for predicting the antitubercular activity of 24 pyrazoline and benzoxazole based chalcones on the Monte Carlo approach. Further these 24 target molecules were subjected to docking for finding out the interactions of the molecules with various targets of mycobacterium species. Computational results indicated that this approach can satisfactorily predict the desired activity with very good statistical significance. For best built model statistical parameters were R^2=0.8813 and Q^2=0.8031 for test set and R^2=0.6124 and Q^2=0.4914 for training set. Additionally, molecular docking study was performed for finding out the interactions of the molecules with various targets of mycobacterium species. Monte Carlo method proved to be an efficient approach to build up a robust model for estimating. Based on QSAR and molecular docking studies, some important physicochemical parameters of pyrazoline moiety could be assessed for antitubercular drugs.

Key words*:* Pyrazoline, Benzoxazole, , Multi drug resistance, Anti tubercular drugs.

INTRODUCTION

Tuberculosis is currently the leading killer of youth, women and patients suffering from AIDS. The duration of the present antitubercular therapy leads to patient non-compliance and this in turn has contributed to the development of drug resistance. In recent years, the pandemic of AIDS poses a major impact on the world wide spread of TB. However, since 1980s, the disease has seen resurgence due to variety of changes in social, medical and economic factors. Concomitant with the resurgence of TB, is the appearance of multidrug-resistant TB which exposes the frailties of the current drug armamentarium.

The importance of quantitative structure-reactivity relationship (QSAR) studies in modern drug design is well established since QSAR can make the early prediction of activity-related characteristics of drug candidates and can eliminate molecules with undesired properties [1]. The main goal of QSAR approach is to correlate the biological activity of a series of compounds with the calculated molecular properties in terms of descriptors [2]. Thousands of molecular descriptors are used in QSAR studies for the purpose of encoding molecules chemical and structural features [3, 4] with great importance of topological descriptors calculated on the basis of molecular graphs [5]. The simplified molecular input line entry system (SMILES) is an alternative to molecular graphs and it can be used for representation of molecular structures [6].

A common problem in the development of QSAR/QSPR models can arise from: (i) selection of an appropriate subset of molecular descriptors from the masses of available descriptors, (ii) the vagueness of interpreting certain

descriptors obtained from QSAR/QSPR modeling and (iii) the need to geometrically optimize structures if three dimensional descriptors are to be used. Therefore CORAL software (available at http://www.insilico.eu/coral) is a tool that allows QSAR/ QSPR analysis as a function of conformation-independent and SMILES based descriptors while complying with the OECD principles [7-10].

Molecular docking has been frequently used to predict the prominent and acknowledged geometry of a protein-ligand complex and to understand the interaction studies of the target with specific ligands. Docking is often used with scoring functions to predict binding affinities of ligands in virtual screening experiments [11]. It is also important in studying the structure activity relationship of the newly synthesized compounds [12-13]. The function of docking is to define the energetics of the system and the efficiency of the ligand molecule to bind to its target, as it forms the basis of the docking algorithms attempt

AutoDock Vina is a new open source program for drug discovery, molecular docking and virtual screening, offering multi-core capability, high performance and enhanced accuracy and ease of use. AutoDock Vina significantly improves the average accuracy of the binding mode predictions.

Thus, the present study employed CORAL software for constructing large-scale QSAR models for predicting the antitubercular activity of 24 pyrazoline and benzoxazole based chalcones on the Monte Carlo approach. Such models afford a simple and versatile approach for discerning the origins of investigated activities directly from the SMILES notation that had been used for encoding molecular structures. Further these 24 target molecules were subjected to docking for finding out the interactions of the molecules with various targets of *mycobacterium* species.

MATERIALS AND METHODS

Data

A dataset of 24 pyrazoline derivatives with determined antitubercular activity against *M.tuberculosis* $H_{37}Rv$ was selected for QSAR study [14]. Figure 1 presents general structures of used pyrazoline compounds for QSAR modeling. The negative logarithmic IC_{50} values of antitubercular activity (pIC50) were selected as the endpoint for QSAR analysis which was converted by microsoft excel sheet using the formula fx = -log10.

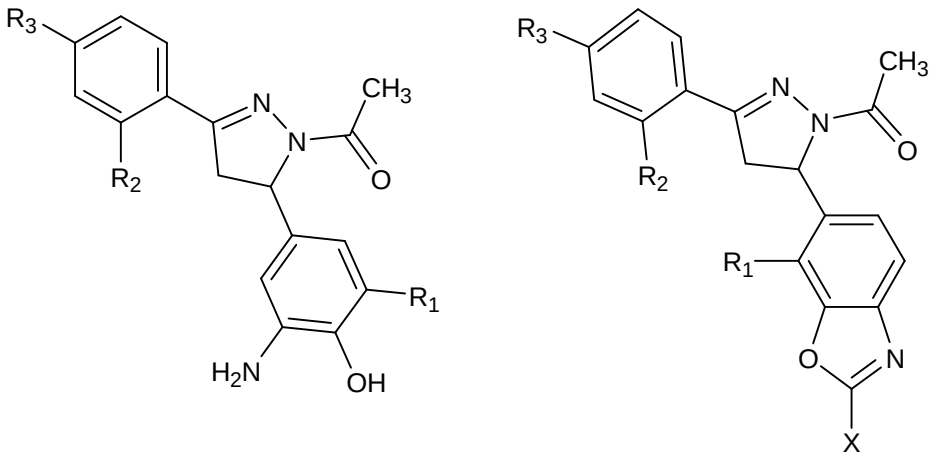


Figure 1: General molecular structures of used molecules

Canonical SMILES for all compounds were generated with the ACD/ChemSketch program (ACD/Chem Sketch v.11.0) in order to preserve consistency because different software may generate different SMILES notations. The role of the training set is in developing of the model. The role of test set is selection of preferable values for the number of epoch of the Monte Carlo optimization and the threshold value.

Optimal descriptor

Optimal descriptors for constructing QSAR models are based on SMILES notation as described according to the following equation [15]:

Whereas DCW represents descriptor correlation weight for SMILES/molecular graph descriptors. 'T' represents threshold which in turn describes rare SMILES or molecular graph attribute which is used in Monte Carlo algorithm e.g. If 'T' is defined as 1 then any SMILES/Graph attribute occurring less than instance '1' is considered as rare

attribute whose correlation weight (CW) is fixed as 0. N_{epoch} is number of epoch of Monte Carlo algorithms which defines the cycle of modifications in correlation weight calculations for building up model. Hence, each Monte Carlo based QSAR model is a function of T and N_{epoch} value

In the present study, each SMILES based DCWs were calculated according to following equation-

$DCW(SMILES) = \alpha\sum Sk + \beta\sum SSk + \gamma\sum SSSk$ --------Equation 2

Whereas α, β, γ represents correlation weights for the SMILES attributes, Sk represent single SMILE component in SMILES representation of the molecules e.g. atoms like C, H, N,O etc, SSk represents two consecutive letter in the SMILES attribute. With respect to molecular structures this could be equivalent to two elements in the structures joined by any bond. SSSk represents three consecutive SMILES attributes. This could be considered as a representation of molecular fragments in the structure. In the present study correlation weights of the SSSk components were used in the DCW calculation of the SMILES optimal descriptors.

The DCW values obtained were then correlated with pIC_{50} values by using least square methodology using following Equation.

$pIC_{50} = C0 + C1 \times DCW(T, N_{epoch})$ -------Equation 3

The quality and robustness of the QSAR model developed was assessed by subjecting it to various statistical techniques as follows:

1) Internal validation was performed by Leave One Out (LOO) cross validation technique on the training set
2) External validation was performed on Test set of compounds which also assess predictive power of the model on the compounds not included in the training set
3) The Randomization test or Y-scrambling test to rule out any chance correlation of descriptors involved in the model to anticancer activity of the compounds.

For calculation of cross validated squared correlation coefficient (q^2), one molecule was randomly deleted from the training set and test set and model was rebuilt and regression coefficient was calculated by following formula-

$Q^2(\text{training/test set}) = 1 - \sum [Aexp(train/test)-Apred(train/test)]^2 /$

$\sum [Aexp(train/test)- \bar{A}(train/test)]^2$.....Equation 4

Where Aexp represents the experimental pIC50 values of the compound in training/test set, Apred indicates predicted activities of the compounds after deletion of random molecules from the training or test set, Ā represents mean experimental pIC_{50} values of the training or test set molecules.

In the present model novel statistical parameter for assessing the predictive power of the QSAR model known as (Rm^2) was used [16] . Rm^2 presents the stricter test of validation where built model is penalized for large differences in the experimental and predicted activities of the compounds [17].

Y-randomization or scrambling test were performed where activity fields were randomly assigned to compounds in training and test set and model was rebuilt. The squared correlation coefficients of the randomized model (Rr^2) were compared to squared correlation coefficients of non randomized model (R^2). A new statistical parameter (CR^2p) which represents squared correlation coefficient of the model after penalizing model for small differences between Rr^2 and R^2 was used to assess robustness quality of the model [18].

$^{SMILES}(DCW) = {}^{SMILES}DCW(T, N_{epoch})$ -------Equation 1

Docking Studies

Enzyme structure

The X-ray crystal structures of *mycobacterium* enoyl reductase (InhA) (PDB ID: 2H7I which was obsolete and now changed to 4UOJ[27]), Cytochrome P-450-14-alpha sterol demethylase from *Mycobacterium tuberculosis* (PDB ID: 1H5Z[28]), human and tubercular DHFR, complexed with folate and trimethoprim (TMP) respectively with (PDB ID 1DRF[29] and 1DG5[30]), Glucosamine-1-Phosphate-N-Acetyl Transferase (GLmU) (PDB ID: 3D8V[31]), Shikimate Kinase (SK) (PDB ID: 1L4Y[32]) were obtained from the protein data bank. . (http://www.rcsb.org/pdb).

RESULTS

QSAR

Table 1: Molecules in SMILES format with their descriptor correlation weight (DCW) values and experimental and calculated pIC_{50} values and residuals

Code		SMILES	DCW	Exp	Calc	Residual
KK1	+	:O=C(C)N2N=C(CC2c1cc(N)c(O)c(OC)c1)c3ccccc3	44.32750	4.113	4.213	-0.100
KK3	+	:O=C(C)N2N=C(CC2c1cc(N)c(O)c(OC)c1)c3ccc(F)cc3	45.19088:	4.438	4.416	0.022
KK5	+	:O=C(C)N2N=C(CC2c1ccc(O)c(N)c1)c3ccccc3	48.51100	5.373	5.197	0.176
KK7	+	:O=C(C)N2N=C(CC2c1cc(N)c(O)c(Cl)c1)c3ccccc3	48.38800	5.005	5.168	-0.163
KK9	+	:CC(=O)N3N=C(CC3c1ccc2nc(N)oc2c1OC)c4ccccc4	48.86056	5.748	5.280	0.468
KK10	+	:Clc1ccc(c(Cl)c1)C=4CC(c2ccc3nc(N)oc3c2OC)N(N=4)C(C)=O~	45.65425	4.524	4.525	-0.001
KK11	+	:Fc1ccc(cc1)C=4CC(c2ccc3nc(N)oc3c2OC)N(N=4)C(C)=O	46.22394	4.769	4.659	0.110
KK13	+	:CC(=O)N3N=C(CC3c1ccc2nc(N)oc2c1)c4ccccc4	43.89481	4.107	4.111	-0.004
KK15	+	:CC(=O)N3N=C(CC3c1ccc2nc(N)oc2c1Cl)c4ccccc4	46.89281	4.149	4.816	-0.667
KK17	+	:CC(=O)N3N=C(CC3c1ccc2nc(S)oc2c1OC)c4ccccc4	48.83031	4.769	5.273	-0.504
KK19	+	:Fc1ccc(cc1)C=4CC(c2ccc3nc(S)oc3c2OC)N(N=4)C(C)=O	46.19369	4.487	4.652	-0.165
KK20	+	:Brc1ccccc1C=4CC(c2ccc3nc(S)oc3c2OC)N(N=4)C(C)=O	45.43931	4.551	4.474	0.077
KK21	+	:CC(=O)N3N=C(CC3c1ccc2nc(S)oc2c1)c4ccccc4	43.86456	4.129	4.104	0.025
KK22	+	:Brc1ccccc1C=4CC(c2ccc3nc(S)oc3c2)N(N=4)C(C)=O	45.49981	4.520	4.489	0.031
KK23	+	:CC(=O)N3N=C(CC3c1ccc2nc(S)oc2c1Cl)c4ccccc4	46.86256	5.471	4.809	0.662
KK24	+	:Brc1ccccc1C=4CC(c2ccc3nc(S)oc3c2Cl)N(N=4)C(C)=O	46.98456	4.856	4.838	0.018
KK2	#	:O=C(C)N2N=C(CC2c1cc(N)c(O)c(OC)c1)c3ccc(Cl)cc3Cl	44.90500	4.497	4.349	0.148
KK4	#	:O=C(C)N2N=C(CC2c1cc(N)c(O)c(OC)c1)c3ccccc3Br	45.32850	4.508	4.448	0.060
KK6	#	:O=C(C)N2N=C(CC2c1ccc(O)c(N)c1)c3ccccc3Br	49.51200	5.776	5.433	0.343
KK8	#	:O=C(C)N2N=C(CC2c1cc(N)c(O)c(Cl)c1)c3ccccc3Br	49.38900	5.513	5.404	0.109
KK12	#	:Brc1ccccc1C=4CC(c2ccc3nc(N)oc3c2OC)N(N=4)C(C)=O	45.46956	4.534	4.481	0.053
KK14	#	:Brc1ccccc1C=4CC(c2ccc3nc(N)oc3c2)N(N=4)C(C)=O	45.53006	4.202	4.496	-0.294
KK16	#	:Brc1ccccc1C=4CC(c2ccc3nc(N)oc3c2Cl)N(N=4)C(C)=O	47.01481	5.122	4.845	0.277
KK18	#	:Clc1ccc(c(Cl)c1)C=4CC(c2ccc3nc(S)oc3c2OC)N(N=4)C(C)=O~	45.62400	4.843	4.518	0.325

Table 2: The data showing statistical parameters of the constructed model for the molecules of test and training set

Set	N	r^2	q^2	MAE	s	F
Train.	16	0.6124	0.4914	0.201	0.313	22
Test.	8	0.8813	0.8031	0.200	0.247	45

N= Number of molecules; r^2= squared correlation coefficient, q^2= leave one out cross validated squared correlation coefficient, s= standard deviation in the activity and F= Fischer coefficient

Table 3: The data representing the details of ten randomization tests

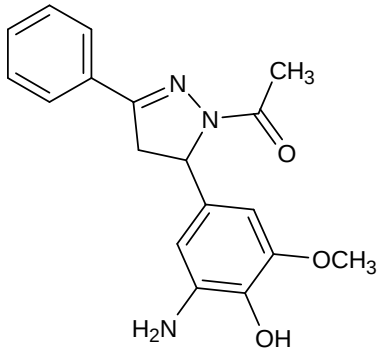
No. of Randomization runs	Training test (r^2)	Test set (r^2)
1	0.0033	0.8813
2	0.2650	0.1077
3	0.2034	0.0035
4	0.0481	0.0715
5	0.4880	0.0509
6	0.3612	0.3967
7	0.2302	0.1883
8	0.1543	0.2499
9	0.0418	0.1504
10	0.1168	0.0965
AverageR2 of randomized	0.1912	0.1361
R2 of Non Randomized Model	0.5079	0.8105:
CRp2=R*sqrt(R2-Rr2)	0.6124	0.8813

Table 4: List of SMILES attributes contributing in the antitubercular activity of the pyrazoline compounds

Promoter of antitubercular activity	Demoter of antitubercular activity	Insignificant attributes
(...C...(...: 1.26181,	(...c...(...: -0.30950	. (...Cl.(...: 0.93950
(...F...(...: 1.48238	=...N...(...: -0.06750	(...N...(...: 0.93750
(...O...(...: 2.00200	C...O...(...: -0.74700	(...S...(...: 0.46675
1...Cl..(...: 1.18750	O...C...(...: -0.75300	1...c...(...: 0.24700
2...Cl..(...: 1.00200	c...(...O...: -0.00300	2...c...(...: 0.38281
2...N...(...: 1.44050	c...o...(...: -0.37600	2...c...1...: 0.53525
3...c...(...: 1.40925		3...N...(...: 0.11338
=...4...(...: 1.59375		3...c...1...: 0.48038
=...C...(...: 1.87200		3...c...2...: 0.87200
=...C...1...: 1.09375		4...c...(...: 0.06550
=...N...2...: 2.00300		C...(...4...: 0.51181
=...N...3...: 1.87600		C...(...=...: 0.12600
=...O...(...: 1.52925		C...(...C...: 0.52044

C...(...1...: 2.03625		C...4...=...: 0.06250
C...C...(...: 2.53225		C...=...4...: 0.12881
C...C...2...: 3.15125		C...C...3...: 0.35056
C...O...1...: 2.94050		C...C...4...: 0.27844
N...(...Cl..: 2.28225		C...O...2...: 0.87200
N...2...N...: 1.97075		F...c...1...: 0.73937
N...3...N...: 1.53625		Br..c...1...: 0.71375
N...=...4...: 1.03425		Cl..c...1...: 0.62200
N...=...C...: 1.51963		N...(...C...: 0.62500
O...(...N...: 1.51963		N...(...N...: 0.62781
O...=...C...: 1.27825		O...=...(...: 0.54188
c...(...F...: 1.37200		c...(...1...: 0.03125
c...(...c...: 1.15525		c...(...C...: 0.37800
c...1...Cl..: 1.46975		c...(...Cl..: 0.99800
c...1...O...: 3.05750		c...(...N...: 0.34775
c...1...c...: 1.62900		c...(...S...: 0.12700
c...2...(...: 2.87200		c...1...(...: 0.62600
c...2...C...: 2.37600		c...1...C...: 0.50400
c...2...Cl..: 1.00500		c...2...c...: 0.35938
c...2...O...: 2.00000		c...3...c...: 0.10538
c...3...C...: 1.85838		c...c...2...: 0.61037
c...3...Br..: 1.00100		c...c...4...: 0.44031
c...3...Cl..: 1.00500		n...2...c...: 0.78625
c...4...c...: 1.93850		n...c...(...: 0.19150
c...c...(...: 1.06750		o...c...2...: 0.14263
c...c...1...: 2.25781		o...c...3...: 1.96675
c...c...3...: 1.67188		
c...c...c...: 2.74900		
c...n...2...: 1.96775		
c...n...3...: 1.12981		
n...3...c...: 2.28525		
o...(...N...: 1.05950		
o...(...S...: 1.72075		

Table : DCW and CWs values for a pyrazoline (KK-1)



SMILES notation: O=C(C)N2N=C(CC2c1cc(N)c(O)c(OC)c1)c3ccccc3

(SA)	CW(SA)
O...=...C...	1.2782
=...C...(...	1.8720
C...(...C...	0.5204
(...C...(...	1.2618
N...(...C...	0.6250
2...N...(..	1.4405
N...2...N...	1.9707
=...N...2...	2.0030
N...=...C...	1.5196
=...C...(...	1.8720
C...(...C...	0.5204
C...C...(...	2.5322
C...C...2...	3.1513
c...2...C...	2.3760
2...c...1...	0.5353
c...1...c...	1.6290
c...c...1...	2.2578
c...c...(...	1.0675
c...(...N...	0.3478
(...N...(...	0.9375

c...(...N...	0.3478
(...c...(...	-0.3095
c...(...O...	-0.0030
(...O...(...	2.0020
c...(...O...	-0.0030
(...c...(...	-0.3095
c...(...O...	-0.0030
C...O...(...	-0.7470
O...C...(...	-0.7530
c...(...C...	0.3780
1...c...(...	0.2470
c...1...(...	0.6260
c...(...1...	0.0313
3...c...(...	1.4092
c...3...c...	0.1054
c...c...3...	1.6719
c...c...c...	2.7490
c...c...c...	2.7490
c...c...c...	2.7490
c...c...3...	1.6719

Descriptor correlation weight (DCW) = 44.32750
SA= Structural attribute (SMILES); CWs= Correlation weights

DOCKING

The binding energy scores of the targeted benzoxazole based pyrazoline derivatives with the various proteins of *mycobacterium* species have been depicted in the tables from 5-9.

Table 5: Benzoxazole based pyrazoline derivatives with their binding energy scores (Kcal/mol) and H-bonds interactions against mycobacterium enoyl reductase (InhA) (PDB ID: 2H7I which is obsolete and changed to 4UOJ)

Ligand Code	R_1	R_2	R_3	X	Binding energy score (Kcal/mol)	pMIC50 uM ($H_{37}Rv$)	H-bond Interacting Residues
KK1	O-CH_3	H	H	-	-9.2	4.113	-
KK2	O-CH_3	Cl	Cl	-	-7.6	4.497	Gly 96
KK3	O-CH_3	H	F	-	-7.8	4.438	Meth 98
KK4	O-CH_3	Br	H	-	-8.4	4.508	Gly 96
KK5	H	H	H	-	-8.3	5.373	Gly 96
KK6	H	Br	H	-	-8.5	5.776	-
KK7	Cl	H	H	-	-8.5	5.005	Gly 96
KK8	Cl	Br	H	-	-8.7	5.513	Gly 96
KK9	O-CH_3	H	H	NH_2	-8.5	5.748	Gly 96
KK10	O-CH_3	Cl	Cl	NH_2	-7.4	4.524	-
KK11	O-CH_3	H	F	NH_2	-8.4	4.769	-
KK12	O-CH_3	Br	H	NH_2	-8.6	4.534	-
KK13	H	H	H	NH_2	-8.7	4.107	-
KK14	H	Br	H	NH_2	-8.8	4.202	Gly 96
KK15	Cl	H	H	NH_2	-8.8	4.149	Gly 96
KK16	Cl	Br	H	NH_2	-8.9	5.122	Gly 96
KK17	O-CH_3	H	H	SH	-8.5	4.769	-
KK18	O-CH_3	Cl	Cl	SH	-7.8	4.843	-
KK19	O-CH_3	H	F	SH	-6.3	4.487	Threo 96
KK20	O-CH_3	Br	H	SH	-8.4	4.551	-
KK21	H	H	H	SH	-7.1	4.129	-
KK22	H	Br	H	SH	-8.6	4.520	-
KK23	H	H	H	SH	-8.8	5.471	-
KK24	H	Br	H	SH	-8.7	4.856	Threo 196

Table 6: Benzoxazole based pyrazoline derivatives with their binding energy scores (in Kcal/mol) and H-bonds interactions against Cytochrome P-450-14-Alpha Sterol Demethylase demethylase from *Mycobacterium tuberculosis* (PDB ID: 1H5Z)

Ligand Code	R_1	R_2	R_3	X	Binding energy score (Kcal/mol)	pMIC50 uM ($H_{37}Rv$)	No. of H-Bonds / H-bond Interacting Residues
KK1	O-CH_3	H	H	-	-9.0	4.1134	Arg 326, Tyr 76
KK2	O-CH_3	Cl	Cl	-	-8.9	4.497	Arg 326
KK3	O-CH_3	H	F	-	-8.7	4.438	Arg 326
KK4	O-CH_3	Br	H	-	-9.0	4.508	Arg 326
KK5	H	H	H	-	-9.4	5.373	Arg 326, Tyr 76
KK6	H	Br	H	-	-9.1	5.776	Arg 326
KK7	Cl	H	H	-	-9.1	5.005	Arg 326
KK8	Cl	Br	H	-	-9.8	5.513	Leu 324, Arg 326, Prol 386
KK9	O-CH_3	H	H	NH_2	-8.6	5.748	His 259
KK10	O-CH_3	Cl	Cl	NH_2	-9.0	4.524	Ala 256
KK11	O-CH_3	H	F	NH_2	-9.1	4.769	-
KK12	O-CH_3	Br	H	NH_2	-9.0	4.534	-
KK13	H	H	H	NH_2	-9.1	4.107	-
KK14	H	Br	H	NH_2	-8.9	4.202	-
KK15	Cl	H	H	NH_2	-8.8	4.149	-
KK16	Cl	Br	H	NH_2	-8.7	5.122	-
KK17	O-CH_3	H	H	SH	-9.0	4.769	-
KK18	O-CH_3	Cl	Cl	SH	-8.7	4.843	-
KK19	O-CH_3	H	F	SH	-9.3	4.487	-
KK20	O-CH_3	Br	H	SH	-8.9	4.551	-
KK21	H	H	H	SH	-8.3	4.129	-
KK22	H	Br	H	SH	-9.1	4.520	Arg 326
KK23	H	H	H	SH	-9.4	5.471	-
KK24	H	Br	H	SH	-9.4	4.856	Arg 326

Table 7: Benzoxazole based pyrazoline derivatives with their binding energy scores (in Kcal/mol) and H-bonds interactions against human and tubercular DHFR, complexed with folate and trimethoprim (TMP) respectively, was obtained from the protein data bank (PDB ID 1DRF[32] and 1DG5[33] respectively

a) PDB ID 1DRF

Ligand Code	R_1	R_2	R_3	X	Binding energy score (Kcal/mol)	pMIC50 uM ($H_{37}Rv$)	No. of H-Bonds / H-bond Interacting Residues
KK1	O-CH_3	H	H	-	-9.0	4.113	-
KK2	O-CH_3	Cl	Cl	-	-9.2	4.497	-
KK3	O-CH_3	H	F	-	-9.2	4.438	-
KK4	O-CH_3	Br	H	-	-8.8	4.508	-
KK5	H	H	H	-	-9.4	5.373	-
KK6	H	Br	H	-	-9.3	5.776	-
KK7	Cl	H	H	-	-9.4	5.005	-
KK8	Cl	Br	H	-	-9.3	5.513	-
KK9	O-CH_3	H	H	NH_2	-10.0	5.748	-
KK10	O-CH_3	Cl	Cl	NH_2	-9.9	4.524	-
KK11	O-CH_3	H	F	NH_2	-10.1	4.769	-
KK12	O-CH_3	Br	H	NH_2	-9.3	4.534	-
KK13	H	H	H	NH_2	-10.4	4.107	Glu 30
KK14	H	Br	H	NH_2	-10.1	4.202	Glu 30
KK15	Cl	H	H	NH_2	-10.2	4.149	-
KK16	Cl	Br	H	NH_2	-9.8	5.122	-
KK17	O-CH_3	H	H	SH	-10.2	4.769	-
KK18	O-CH_3	Cl	Cl	SH	-9.2	4.843	-
KK19	O-CH_3	H	F	SH	-10.0	4.487	-
KK20	O-CH_3	Br	H	SH	-9.5	4.551	-
KK21	H	H	H	SH	-10.0	4.129	Peptide bond between Gly 117 & gly 116
KK22	H	Br	H	SH	-10.0	4.520	-
KK23	H	H	H	SH	-10.2	5.471	-
KK24	H	Br	H	SH	-9.4	4.856	-

a) PDB ID 1DG5

Ligand Code	R_1	R_2	R_3	X	Binding energy score (Kcal/mol)	pMIC50 uM ($H_{37}Rv$)	H-bond Interacting Residues
KK1	$O\text{-}CH_3$	H	H	-	-8.4	4.113	-
KK2	$O\text{-}CH_3$	Cl	Cl	-	-8.1	4.497	-
KK3	$O\text{-}CH_3$	H	F	-	-8.3	4.438	-
KK4	$O\text{-}CH_3$	Br	H	-	-8.3	4.508	-
KK5	H	H	H	-	-7.8	5.373	Gly 96
KK6	H	Br	H	-	-7.7	5.776	Threo 46
KK7	Cl	H	H	-	-8.5	5.005	-
KK8	Cl	Br	H	-	-8.0	5.513	Threo 46
KK9	$O\text{-}CH_3$	H	H	NH_2	-8.5	5.748	Gly 96
KK10	$O\text{-}CH_3$	Cl	Cl	NH_2	-8.6	4.524	-
KK11	$O\text{-}CH_3$	H	F	NH_2	-8.7	4.769	Gly 96
KK12	$O\text{-}CH_3$	Br	H	NH_2	-8.6	4.534	Threo 46
KK13	H	H	H	NH_2	-8.9	4.107	-
KK14	H	Br	H	NH_2	-8.9	4.202	-
KK15	Cl	H	H	NH_2	-8.4	4.149	Gly 97
KK16	Cl	Br	H	NH_2	-8.8	5.122	Threo 46
KK17	$O\text{-}CH_3$	H	H	SH	-8.5	4.769	Gly 97
KK18	$O\text{-}CH_3$	Cl	Cl	SH	-8.6	4.843	Gly 96
KK19	$O\text{-}CH_3$	H	F	SH	-8.6	4.487	Gly 97
KK20	$O\text{-}CH_3$	Br	H	SH	-8.4	4.551	-
KK21	H	H	H	SH	-8.3	4.129	Gly 96
KK22	H	Br	H	SH	-8.7	4.520	Gly 97
KK23	H	H	H	SH	-8.4	5.471	Gly 96
KK24	H	Br	H	SH	-8.4	4.856	Threo 46

Table 8: Benzoxazole based pyrazoline derivatives with their binding energy scores (in Kcal/mol) and H-bonds interactions against Glucosamine-1-Phosphate-N-Acetyl Transferase (GLmU) (PDB ID: 3D8V)

Ligand Code	R_1	R_2	R_3	X	Binding energy score (Kcal/mol)	pMIC50 uM ($H_{37}Rv$)	No. of H-Bonds / H-bond Interacting Residues
KK1	$O\text{-}CH_3$	H	H	-	-8.1	4.113	-
KK2	$O\text{-}CH_3$	Cl	Cl	-	-8.7	4.497	-
KK3	$O\text{-}CH_3$	H	F	-	-8.7	4.438	-
KK4	$O\text{-}CH_3$	Br	H	-	-8.6	4.508	-
KK5	H	H	H	-	-8.5	5.373	Lys 26
KK6	H	Br	H	-	-8.6	5.776	Lys 26
KK7	Cl	H	H	-	-8.3	5.005	-
KK8	Cl	Br	H	-	-8.9	5.513	Gly 88
KK9	$O\text{-}CH_3$	H	H	NH_2	-8.8	5.748	-
KK10	$O\text{-}CH_3$	Cl	Cl	NH_2	-9.2	4.524	Lys 26 and Asp 114
KK11	$O\text{-}CH_3$	H	F	NH_2	-8.9	4.769	Lys 26 and Asp 114
KK12	$O\text{-}CH_3$	Br	H	NH_2	-9.1	4.534	Gly 15
KK13	H	H	H	NH_2	-8.8	4.107	Lys 26
KK14	H	Br	H	NH_2	-8.7	4.202	Lys 26
KK15	Cl	H	H	NH_2	-8.8	4.149	Lys 26 and Asp 114
KK16	Cl	Br	H	NH_2	-8.7	5.122	Ser 112
KK17	$O\text{-}CH_3$	H	H	SH	-8.2	4.769	Peptide bond between Gly 15 & Ala 14
KK18	$O\text{-}CH_3$	Cl	Cl	SH	-8.6	4.843	Asp 114
KK19	$O\text{-}CH_3$	H	F	SH	-8.2	4.487	Asp 114
KK20	$O\text{-}CH_3$	Br	H	SH	-8.4	4.551	Peptide bond between Gly 15 & Ala 14 and Ala 13 (side amino acid)
KK21	H	H	H	SH	-8.4	4.129	Ala 182
KK22	H	Br	H	SH	-8.4	4.520	Peptide bond between Ala 182 & Asn 181
KK23	H	H	H	SH	-8.4	5.471	Peptide bond between Ala 182 & Asn 181
KK24	H	Br	H	SH	-8.4	4.856	Peptide bond between Threo 89 & Asp 114

Table 9: Benzoxazole based pyrazoline derivatives with their binding energy scores (in Kcal/mol) and H-bonds interactions against Shikimate Kinase (SK) (PDB ID: 1L4Y)

Ligand Code	R_1	R_2	R_3	X	Binding energy score (Kcal/mol)	pMIC50 uM ($H_{37}Rv$)	No. of H-Bonds / H-bond Interacting Residues
KK1	O-CH_3	H	H	-	-7.9	4.113	-
KK2	O-CH_3	Cl	Cl	-	-8.4	4.497	Ser 16 and Lys 15
KK3	O-CH_3	H	F	-	-8.0	4.438	Arg 117
KK4	O-CH_3	Br	H	-	-8.4	4.508	Ser 16
KK5	H	H	H	-	-7.9	5.373	-
KK6	H	Br	H	-	-8.3	5.776	-
KK7	Cl	H	H	-	-8.3	5.005	-
KK8	Cl	Br	H	-	-8.4	5.513	-
KK9	O-CH_3	H	H	NH_2	-8.8	5.748	-
KK10	O-CH_3	Cl	Cl	NH_2	-8.8	4.524	-
KK11	O-CH_3	H	F	NH_2	-8.9	4.769	Gly 81
KK12	O-CH_3	Br	H	NH_2	-9.2	4.534	Gly 81
KK13	H	H	H	NH_2	-8.7	4.107	-
KK14	H	Br	H	NH_2	-8.9	4.202	-
KK15	Cl	H	H	NH_2	-9.0	4.149	-
KK16	Cl	Br	H	NH_2	-9.1	5.122	-
KK17	O-CH_3	H	H	SH	-8.2	4.769	-
KK18	O-CH_3	Cl	Cl	SH	-8.0	4.843	-
KK19	O-CH_3	H	F	SH	-8.3	4.487	-
KK20	O-CH_3	Br	H	SH	-8.3	4.551	-
KK21	H	H	H	SH	-8.6	4.129	-
KK22	H	Br	H	SH	-8.5	4.520	-
KK23	H	H	H	SH	-8.4	5.471	-
KK24	H	Br	H	SH	-7.8	4.856	

The interaction the compounds KK-5, KK-8 (as these two derivatives have shown hydrogen bonding with all the targeted proteins) are shown in Figures 1-4.

KK-5 **KK-8**

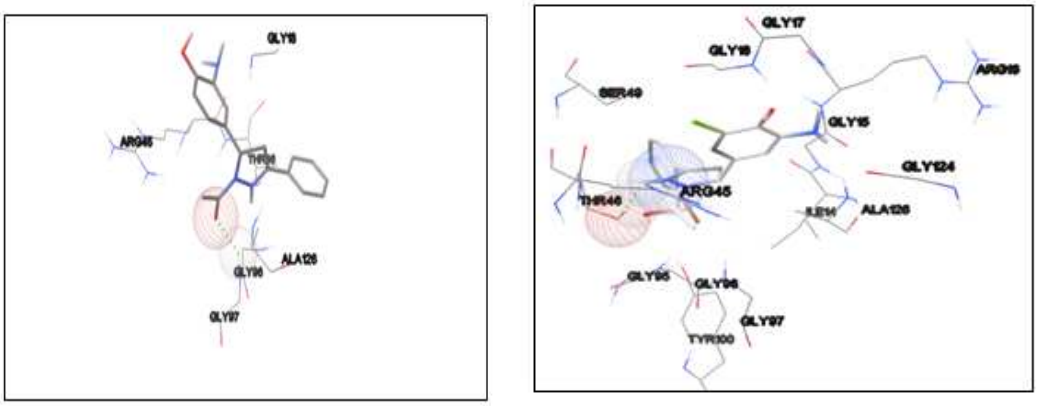


Fig 01: Interaction of KK-5 and KK-8 against 1DG5

KK-5 **KK-8**

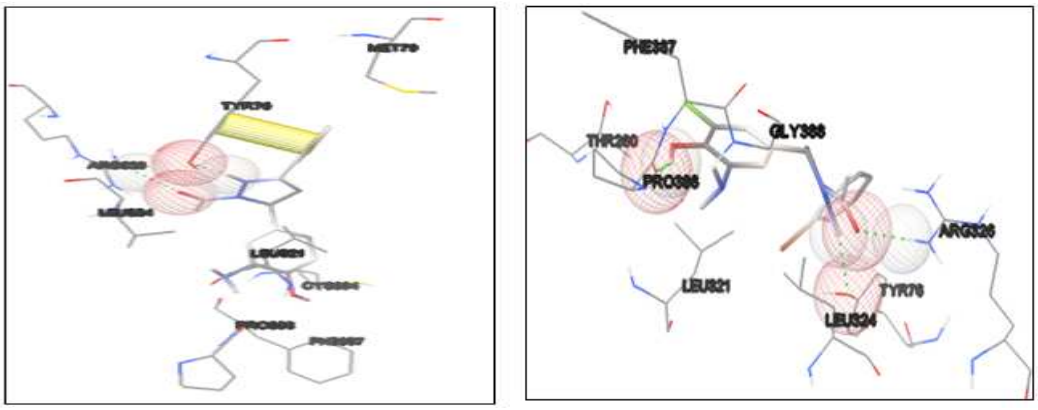


Fig 02: Interaction of KK-5 and KK-8 against 1H5Z

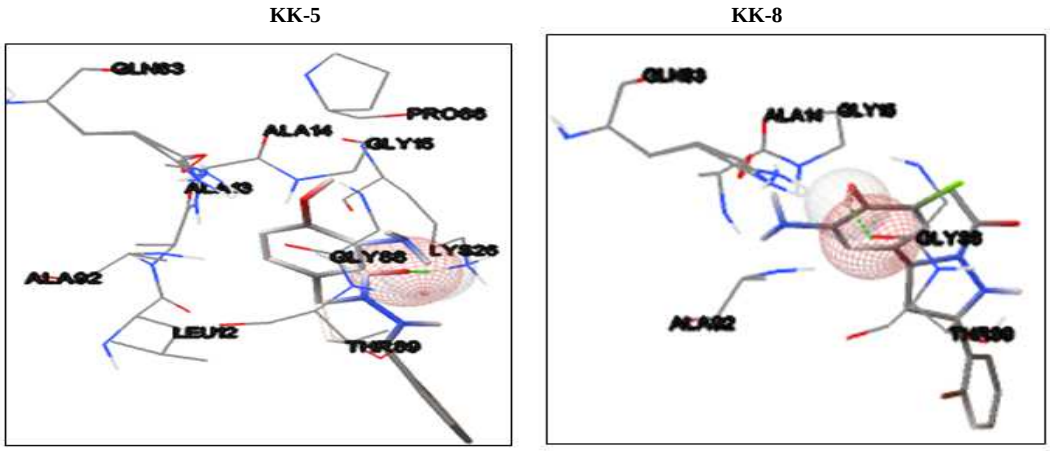


Fig 03: Interaction of KK-5 and KK-8 against 3D8V

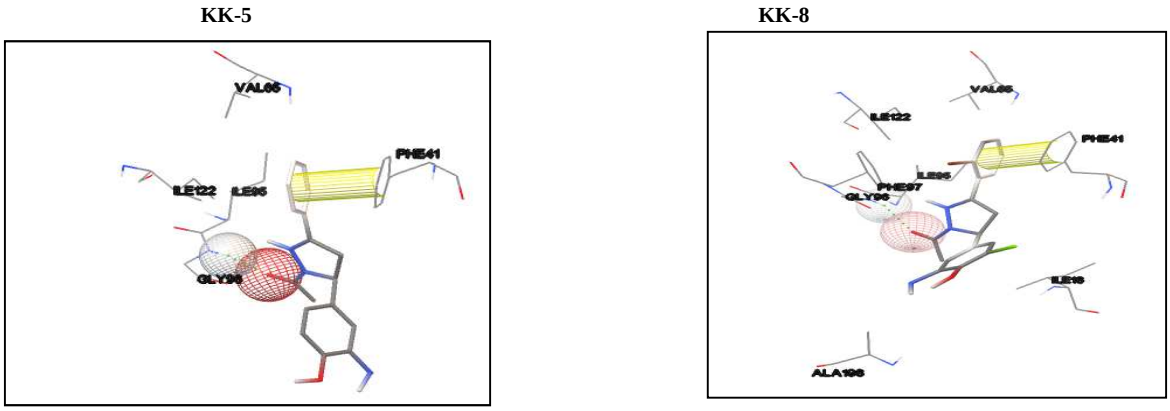


Fig 04: Interaction of KK-5 and KK-8 against 4UOJ

DISCUSSION

QSAR

Each and every SMILES based optimal descriptors were evaluated for obtaining best statistical parameters for the QSAR model.After several trials, SSSk were found to be suitable for these molecules. The threshold and epoch values for these optimal descriptors obtained were 0 and 4 respectively. The experimental and predicted values for the training and test set molecules have been shown in Table 1 along with their DCW values. For better understanding of DCW calculation, component CWs of one of the compound used in model building has been depicted in Table 8. The statistical parameters obtained for the model are represented in Table 2.The squared correlation coefficients obtained for training and test set were 0.6124 and 0.8813 which represents good correlation of the compound activity with the predicted acivity which has been calculated by SMILES. The cross validated squared correlation coefficient values were also in satisfactory range. The model is considered highly predictive when Q^2 values presented by both training and test set is greater than 0.5 . The stringent statistical parameters like Rm^2 which penalizes the model for greater difference between experimental and predicted values have also fulfilled the minimum criteria for its predicitive ability. The model is considered as predictive when Rm^2 values is greater than 0.5.The average Rm^2 for this model was found to be 0.7743. During model building there is possibility that descriptors under consideration are correlated with the activity values by mere chance and not because they are certainly useful for the activity. Hence randomization or Y-scrambling tests were performed. After randomly shuffling the activity fields, descriptors which were found to be useful for model building of non-randomized dataset did not correlate well with the randomized activity fields. The squared correlation coefficient values were substantially less in such cases as compared to R^2values of non-randomized model. Rr^2values for randomized datasets are represented in Table 3.Data have shown that any possibility of chance correlation could be eliminated as Rr^2values are substantially less for the built model. Also stringent statistical parameter CR^2P which penalizes model for small differences between randomized and non-randomized correlation coefficients were calculated. The CR^2P values for training and test set of the model were 0.51 and 0.81. For considering the model to be of high quality and robust this value should be greater than 0.5.

Docking

Interaction of Ligands with the active site of mycobacterium

The result showed that the binding interaction was good for the protein 3D8V, 1DG5, 2H7I, 1H5Z in most of the compounds as compared to the protein 1DRF, 1L4Y. The amino acids which are involved in the bond formation between the ligands and the protein were glycine 15, 81, 96, 97,98, 116, 117, methionine 98, threonine 46, 89 ,196, arginine 117, 326, tyrosine 76, leucine 324, proline 386, histidine 259, alanine 13, 14, 182, 256, glutamine 30, lysine 15, 26, aspartate 114, serine 16 ,112.

In case of 1DG5, most of the hydrogen bonding interactions were due to the presence of electron donating groups like methoxy, pyrazoline moiety and methoxy groups of benzoxazole moiety. Therefore hydrogen bonding acceptor amino acids residues are predominantly in the outer site of proteins of mycobacterium tuberculosis because majority of the hydrogen bond interactions to the ligands were mediated through ligand donating groups. Hence for the designing of novel analogues incorporation of majority of hydrogen bond acceptors group in the ligands should be considered

Docking studies of these 24 ligands against DHFR of mycobacteria and human proteins have revealed important information regarding selectivityof these molecules on intended microorganism. However results were discouraging. It was observed that optimum distance between two aromatic moieties of benzoxazole substituted pyrazolines have led to higher scores for these molecules whereas in case of mycobacterium this distance was of secondary importance as most of interaction were polar groups dominated. For shift in the selectivity of the binding affinity of these scaffolds increase in distance between two aromatic groups were suggested to impart lesser toxicity of these molecules to human host

In case of 1H5Z, planar group of pyrazoline and benzoxazole moiety played an important role in pi-pi kind of non bonding interaction. In addition it also facilitated hydrogen bonding interactions of the substitutent group with the polar amino acids . For example in KK-1, KK-2, KK-5, KK-22 hydrogen bond interaction have been observed mostly with arginine 326 residues. This emphasizes substitutent of electron rich groups on the aromatic moiety for enhanced interaction with cationing arginine groups of active site at physiological pH.

The interactions of ligands with protein 4UOJ have shown that non bonding interactions were uniformly distributed along the length of the molecule in the form of pi-cation type of interaction (lysine 196) and hydrogen bonding interactions with aliphatic amino acid residues like threonine 196 and aromatic hydroxyl group like tyrosine 168. However thiol group could not offer any interaction as aliphatic hydrophobic amino acid were surrounding the sulphur substituted ring as found in KK-19 making thiol group frivolous

In case of 1L4Y, predominant polar interactions with amino acids residue like arginine 117 were observed. Analysis of active site has shown predominance of aliphatic neutral amino acids justifying why aromatic ligands fail to interact optimally with shikimate acid protein. This was corroborated by lower energy scores in the range of -7.8 to -9.2. Hence it was clear that shikimate acid is not a favoured target for these molecules for its antitubercular activity.
Polar mediated interactions with predominantly cationic active site were observed with the target 3D8V. Pyrazoline moiety is of secondary importance with regard to 3D8V. Binding affinity scores range from -8.1 to -9.2

It was observed from these data that N of pyrazole moiety contributed significantly in the antitubercular activity as all the SMILES attribute representing pyrazole nitrogens have shown higher positive correlation weight in the range of 2.00-1.440.Attributes showing methoxy groups were represented as demoter of activity which is also confirmed by the docking studies . It was also observed that cyclization of amino and hydroxyl groups was unfavourable for the targeted activity as SMILE attribute[(c...o...(...: -0.37600)] has shown negative CWs.Atomic radius of halogen substitution plays an important role as increase in atomic radius might have caused congestion in the active site as evidenced by decreased CWs of halogens in the order of F>Cl>Br. Electronegativty also follows the same order for activity as that of atomic radius. In case of benzoxazole derivatives thiol substitutent is more potent than NH_2 Based on these observations and DCW calculations, QSAR equation was obtained for calculating anticancer activity of the compounds and is represented as follows-

pMIC (H37Rv) = -6.2221000 (± 0.6479070) + 0.2354000 (± 0.0143569) * DCW(0,4) Where DCW is descriptor correlation weights calculated for individual compounds.

CONCLUSION

After analyzing docking and QSAR studies and based upon availability of chemicals proposed molecule ought to posses the following properties

- Presence of fluorine and chlorine is preferred as compared to bromine and iodine
- Presence of thiol is preferred as compared to amino group
- OH or CH_3 instead of OCH_3

The compounds which have been identified by QSAR will be considered for design and development of new molecules by Lead Grow module and the same will be analyzed for its binding affinity towards the selected protein receptor by docking studies and further carried out for synthesis.

REFERENCES

[1] C Hansch; D Hoekman; H Gao. *Chem Rev*; **1996**, 96, 3, 1045-76.
[2] A Tropsha; P Gramatica; VK Gombar. *QSAR Comb Sci*; **2003**, 22, 1, 69-77.
[3] M Karelson. Molecular Descriptors in QSAR/QSPR. Wiley- Interscience, New York, **2000**.
[4] Todeschini R, Consonni V. Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, Germany, **2000**.
[5] PR Duchowicz; A Talevi; LE Bruno-Blanch; EA Castro. *Bioorg Med Chem*; **2008**, 16, 17, 7944-55.
[6] Daylight Chemical Information Systems, Inc., **2008**.
[7] AA Toropov; AP Toropova; E Benfenati; G Gini; D Leszczynska; J Leszczynski. *Biochem Biophys Res Commun*; **2013**, 432, 214–25.
[8] AA Toropov, AP Toropova, E Benfenati, G Gini, D Leszczynska, J Leszczynski. *Chemosphere*; **2013**, 90, 877–80.
[9] AP Toropova; AA Toropov. *Eur J Pharm Sci*; **2014**, 52, 21–25.
[10] A Worachartcheewan; C Nantasenamat; NAC Isarankura; V Prachayasittikul. *Lett Drug Des Discov*; **2014**, 11, 420–27.
[11] I Muegge, M Rarey. Reviews in computational chemistry. John Wiley & Sons, Inc., **2001**, pp 1-60.
[12] KM Holloway et al. *J Med Chem*; **1995**, 38, 305-17.
[13] M Vieth, DJ Cummins. *J Med Chem*; **2000**, 43,16, 3020-32.
[14] NR Dharmarajsinh, TC Mahesh, KS Nisha, SB Pathik. *Med Chem Res*; **2014**, 23, 2218-28.
[15] AP Toropova, AA Toropov, E Benfenati, G Gini, D Leszczynska, J Leszczynski. *J Comput Chem*; **2011**, 32, 2727–33.
[16] PP Roy, K Roy. *Chem biol & drug design*; **2009**, 73, 442-55.
[17] PK Ojha, I Mitra, RN Das, K Roy. *Chemometrics & Intelligent Lab Syst*; **2011**, 107, 194-205.
[18] PK Ojha, K Roy. *Chemometrics & Intelligent Lab Syst*; **2011**, 109, 146-61.
[19] X He , A Alian , R Stroud , MPR Ortiz de . *J Med Chem;* **2006**, 49, 21, 6308-23.
[20] LM Podust , LV Yermalitskaya , GI Lepesheva , VN Podust , EA Dalmasso , etal. *Structure*; **2004**, 12, 11, 1937- 45.
[21] C Oefner, DA Arcy, FK Winkler. *Eur J Biochem*; **1988**, 174, 377-85.
[22] R Li, R Sirawaraporn, P Chitnumsub, et al. *J Mol Biol*; **2000**, 295, 307-23.
[23] Z Zhang, EM Bulloch, RD Bunker, EN Baker, CJ Squire. *Acta Crystallogr D Biol Crystallogr*, **2009**, 65, 275-83.
[24] Y Gu, L Reshetnikova, Y Li, Y Wu, H Yan, etal. *J Mol Biol*; **2002**, 319, 779-89.