DESIGN OF POTENTIAL ANTITUBERCULAR AGENTS CONTAINING UREA ANALOGUES, USING PHARMACOPHORE OPTIMIZATION BY MOLECULAR MODELING STUDIES

Vikrant V. Nawale[1*], Shashikant V. Bhandari[2], Kunal G. Raut[2], Pavan P. Wankhade[1]
[1]K. T. Patil College of Pharmacy, Barshi Road, Siddharth Nagar, Osmanabad-413501 (M.S.)
[2]AISSM's College of Pharmacy, Near R.T.O, Kennedy Road, Pune-01, Maharashtra, India

Abstract:
Tuberculosis disease has been the leading cause of morbidity and mortality among the infectious diseases. To address these issues, research and developmental activities to develop novel and potent new chemical entities are necessary. Molecular modelling studies are an approach which is used to narrow down a library containing an extraordinarily high number of random molecules into a smaller list of the potentially effective inhibitors. Two dimensional 2D and three dimensional 3D Quantitative Structure activity relationship (QSAR) studies were performed for correlating chemical composition of 1, 3-disubstituted urea analogues and antitubercular activity using Multiple Linear Regression (MLR) Analysis. The developed QSAR models were found to be statistically significant with respect to training ($r^2 > 0.7$), cross-validation ($q^2 > 0.5$), and external validation ($pred\_r^2 > 0.5$). The best model shows $r^2 = 0.9743$, $q^2 = 0.9134$ and $Pred\_r^2 = 0.5650$. Binding affinities of designed NCEs were studied on Epoxide Hydrolase at specific binding site using docking studies and their ADME properties were also predicted. All the designed compounds show better affinity to the receptor site than the reference compound i.e. thioacetazon. This study will help for the synthesis of potential antitubercular agents.
Keywords: *Epoxide hydrolase, 2D QSAR, 3D QSAR, Docking, ADME Prediction.*

*Corresponding author:
Mr. Nawale Vikrant Vijaykumar, (M.Pharm)
K. T. Patil College of Pharmacy,
Barshi Road, Siddharth Nagar,
Osmanabad-413501
Ph. No.: +917620610555
Email: vikrant.nawale71@gmail.com

QR code

1. INTRODUCTION:

Tuberculosis (TB), is a major chronic infectious diseases caused by Mycobacterium tuberculosis (M. tuberculosis) and to a lesser degree by Mycobacterium bovis and Mycobacterium Africanum affects nearly 32% of the World's population with about 9.4 million Worldwide and 1.6-2.4 million cases alone in India [1,2]. The disease has been the leading cause of morbidity and mortality among the infectious diseases. Multidrug-resistant tuberculosis (MDR-TB) is a form of TB caused by bacteria that do not respond to, at least, isoniazid and rifampicin, the two most powerful, first-line anti-TB drugs. In some cases more severe drug resistance can develop. Extensively drug-resistant TB, XDR-TB, is a form of multi-drug resistant tuberculosis that responds to even fewer available medicines, including the most effective second-line anti-TB drugs [3, 4,5].To address these issues, research and developmental activities to develop novel and potent new chemical entities are necessary.

Recently, it was found that M. tuberculosis epoxide hydrolase (EH) enzyme B (EphB) is a promising target for the anti-tubercular drug [6,7]. The tuberculosis genome contained at least six putative EH enzymes [8]. These large number of EHs compared to other bacteria, suggests that these enzymes play an important roles in the physiology of M. tuberculosis; notably, lipid metabolism and detoxification of reactive oxygen species derived from the host's immune system. Epoxide hydrolase enzyme catalyzes the hydrolysis of epoxide to diol [6,9]. Currently, there are no reports on the whole cell anti-tubercular activity of EH inhibitors, though molecules with similar structures have recently been described with good inhibition of M. tuberculosis EphB6 or antitubercular MICs [10,11].

Thioacetazone is the only drug which is urea derivative acts as antimycobacterial epoxide hydrolaseinhibitor. But it has weak activity against Mycobacterium tuberculosis and not used alone to treat tuberculosis. Thioacetazone is ineffective when given intermittently and causes skin reaction in HIV positive patients. Urea molecules show antibacterial [12], anticancer [13], antifungal [14], antidiabetic [15], antihypertensive [16] as well as antidepressant [17] activities. Recently disubstituted urea derivatives were designed and synthesized by many researchers and their studies shows disubstituted urea derivatives cause inhibition of epoxide hydrolaseenzyme of mycobacterium tuberculosis [18-21].

Molecular Modelling studies is an approach which is used to narrow down a library containing an extraordinarily high number of random molecules into a smaller list of the potentially effective inhibitors [22,23]. Thus in the present study we have focused on development of Two Dimensional (2D) and Three Dimensional (3D) QSAR studies using Multiple Linear Regression (MLR) Analysis and Simulated Annealing k Nearest Neighbour Molecular Field Analysis (SA-kNN MFA), respectively for a series of disubstituted urea derivatives as novel epoxide hydrolase inhibitors. Due to this one of the main objective of our study is the prediction of new potent and selective drug like biologically active compounds on the basis of previously synthesized ones in an attempt to save the cost of blind synthetic process, manual labour as well as time.
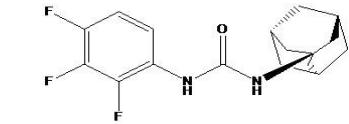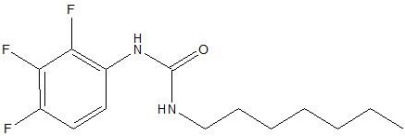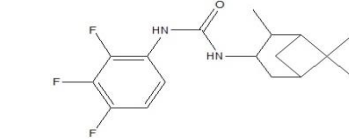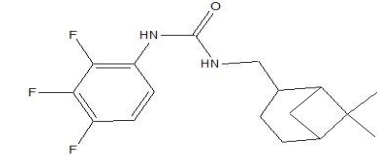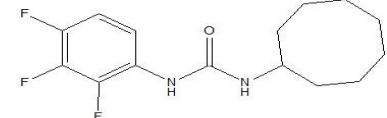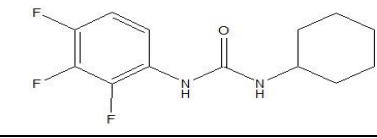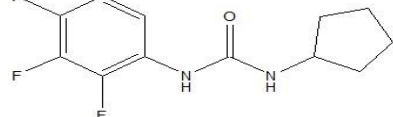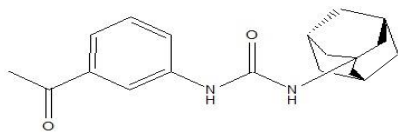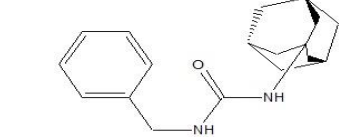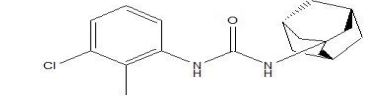
2. MATERIALS AND METHODS:
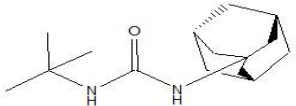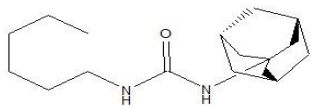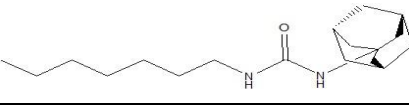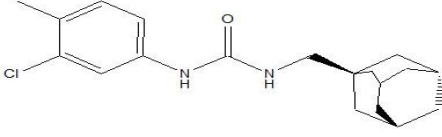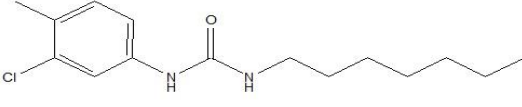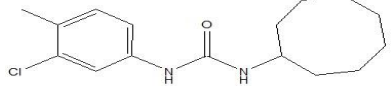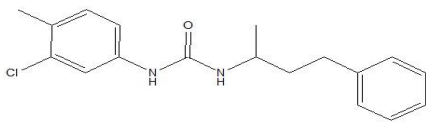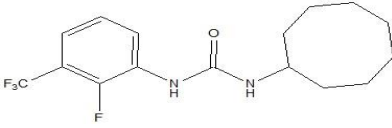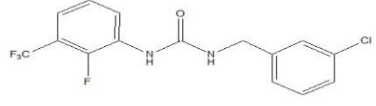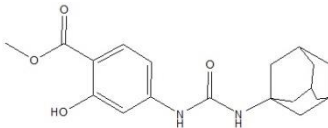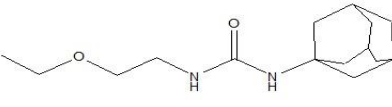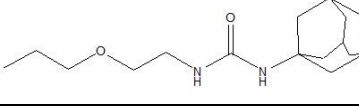
QSAR
Biological dataset
For this study, total 22 molecules of disubstituted urea series reported for antitubercular activity (**Table 1**) has been chosen to develop QSAR models [18].

Table 1: Selected series of compounds containing urea pharmacophore

Sr. No.	Compound Name	Structure	MIC (µg/ml)	Pmic
1	v01		0.4	5.9089
2	v02		6	4.6817
3	v03		0.02	7.2132
4	v04		0.2	6.2126
5	v05		0.8	5.5744
6	v06		12.5	4.3380
7	v07		50	3.7130
8	v08		0.1	6.4948
9	v09		3.1	4.9625
10	v10		1.6	5.2994

11	v11		12.5	4.3016
12	v12		0.01	7.4449
13	v13		0.2	6.1650
14	v14		0.02	7.2212
15	v15		3.1	4.9601
16	v16		0.01	7.4698
17	v17		6	4.7226
18	v18		0.8	5.6185
19	v19		1.6	5.3358
20	v20		0.4	5.9350
21	v21		50	3.7265
22	v22		25	4.0498

Molecular modelling tools
All QSAR studies were performed using VLife Molecular Design Suite 3.5 [24]. Molecules were optimized by Merck Molecular Force Field (MMFF) energy minimization method [25].
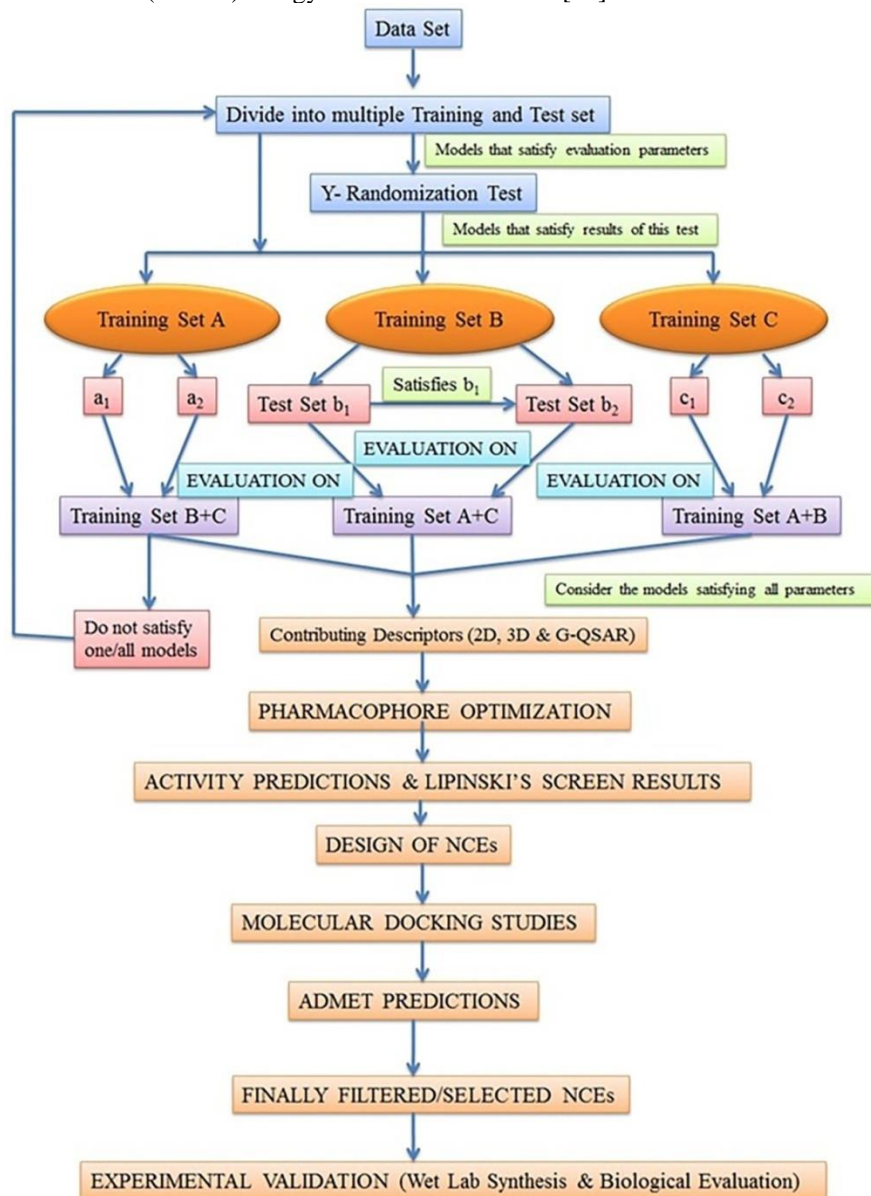
Figure 1. Experimental Design

Two-dimensional QSAR (2D QSAR) studies
Experimental design for 2D QSAR
Dataset of 22 molecules was divided into multiple training and test sets by using manual data selection method and no. of sets were generated using different combinations of molecules in training and test sets such that it covers each molecule in different set every time in an attempt to ensure robustness of QSAR model and increase its predictive ability. From these sets, training and test sets which followed all model evaluation parameters were subjected to Y-randomization test (Fig. 1.).

Evaluation parameters of Y-randomization test:
n = Number of molecules
df = Degree of freedom (n-k-1) (higher is better)
k = Number of descriptors in a model ($\leq$ n/5)
r^2 = Coefficient of determination (>0.7)
q^2 = Cross validated r^2 (>0.5)
SEE = Standard error of estimate (smaller is better)
Pred_r^2 = r^2 of external test set (>0.5)
F-test = Statistical significance of the model (higher is better for same descriptors and compounds)
Best_ran_r^2 = Highest r^2 value in the Y-randomization test (as low as compared to r^2)

Best_ran_q^2 = Highest q^2 value in the Y-randomization test (as low as compared to q^2)

Z score = It is calculated by the Y-randomization test (higher is better)

Alpha = Statistical significance parameter by randomization test (<0.01)

Three models were selected which satisfied the results of Y-randomization test and were named as Training set-A, Training set-a, Training set-B and Training set-b. These models were subjected to two times for external validation by splitting them into two test sets viz. test set a_1, a_2 for training set-a and test set-b_1, b_2 for training set-b in order to avoid the chance correlated results. Only those models which satisfied both the test sets were selected for design of NCEs. We have ensured that selected training and test sets also satisfied the following criteria:

- Representative points of the test set must be close to those of the training set;
- Representative points of the training set must be close to representative points of the test set;
- Training set must have wide chemical and biological diversity.

Uni-column statistics

The comparative statistical parameters of training and test sets created by manual data selection method are reported in **Table 2**. The min and max values in both training and test set should be compared in a way that-

i.   The max of the test should be less than max of training set.

ii.   The min of the test should be greater than min of training set.

It shows that the test set is interpolative i.e. derived within the min-max range of the training set. The mean and standard deviation of the training and test set provides insight to the relative difference of mean and point density distribution (along mean) of the two sets. Standard deviation of Training set A, a, B and b with test set a_1, a_2 and b_1, b_2 respectively were found to be nearly close to each other. This showed that even though the selected molecules in training or test sets are different, but the distribution pattern with respect to the biological activity of molecules in both the selection methods is quite similar.

Table 2. Uni-Column statistics for training sets and test sets

Para-meters	Training Set B+C	Test set a_1	Test set a_2	Training Set A+C	Test set b_1	Test set b_2	Training Set A+B	Test set c_1	Test set c_2
Avg.	5.2783	5.5923	5.2889	5.5290	5.4970	5.0072	5.3630	5.1106	5.3590
Max	7.2220	7.7210	7.9830	7.4698	7.4449	7.4720	7.2220	7.0130	7.2220
Min	3.5500	3.7030	3.6130	3.7130	3.7265	3.5510	3.5890	3.5900	3.7070
S.D.	1.7059	1.7004	1.6957	1.1276	1.3113	1.3200	1.6321	1.6199	1.6119
Sum	69.352	52.000	51.012	71.876	49.472	50.125	70.584	50.768	49.818

Table 3. Correlation matrix of descriptors (2D QSAR)

Descriptor	T_C_C_4	T_C_F_4	SaaCHcount	Quadrupole3	ZcompDipole	Quadrupole1
T_C_C_4	1	-0.3722	0.1323	-0.3333	-0.2230	0.4033
T_C_F_4	-0.3722	1	0.3378	0.6169	0.1033	-0.8515
SaaCHcount	0.1323	0.3378	1	0.0547	-0.1165	-0.3531
Quadrupole3	-0.3333	0.6169	0.0547	1	0.0064	-0.8420
ZcompDipole	-0.2230	0.1033	-0.1165	0.0064	1	-0.0590
Quadrupole1	0.4033	-0.8515	-0.3531	-0.8420	-0.0590	1

Descriptor selection

Various 2D descriptors (a total of 612) were calculated and pre-processing of them was carried out by removing invariable columns. It has been reported that there is high probability of chance correlation between the observed and predictive activity; especially when No. of descriptors are comparable or more than the No. of compounds in dataset for any QSAR analysis [24]. Thus, reduction in no. of descriptors is a very important step which is required to avoid the occurrences of chance correlation and inclusion of irrelevant descriptors in final QSAR model. We applied combinations of different descriptor selection methods viz. forward, forward-backward, genetic algorithm, simulated annealing etc. as well as different QSAR methods on same

molecules sets and finally considered the results of forward variable selection method with Multiple Linear Regression (MLR) after comparing all results to improve performance as well as predictability of QSAR model.

Correlation matrix

It is very popular and crucial technique used for QSAR studies. We have considered the correlation between descriptor with activity as well as their inter-correlation i.e. descriptor-descriptor correlation. We have shown only those descriptors contributed for the selected series of compounds in 2D QSAR studies; which show either direct or inverse correlation with activity.

Fitness plot

Correlation coefficient cannot give information about data spread between the descriptor and activity. There may be some descriptors showing chance correlation with activity because each variable selection method is based upon correlation between descriptor and activity and not on the type of data spread. To avoid above said pitfall, the proper observation of fitness plot between descriptor and activity is needed.
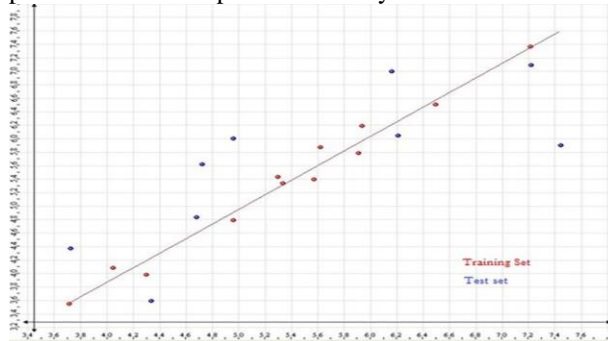

Fig. 2: Fitness plot for descriptor

The following are few important points that we have taken into consideration while selecting proper descriptors for QSAR model generation:

We have ensured that percentage distribution of data points on both sides of best fit line should be nearly 50-50%. (we preferred slope value more than 0.15)

In case of topological descriptor, No. of occurrences of particular data point was observed in fitness plot which gave information about the frequency of occurrence of each particular substituent in series. Thus although the particular descriptor shows good correlation with activity as well as comes in the QSAR final model result, but we cannot take it into final consideration unless and until it shows well spread fitness plot. In conclusion, we can say that careful observation and right analysis of a fitness plot helped us to reduce no. of descriptors.

Variance

Another significant way to find out unimportant descriptors is by using information of variance of descriptors [26]. There were some descriptors which showed consistently high variance even if there was small change in physicochemical properties and vice versa. After close analysis of the output of our study, we conclude that we should focus more on correlation between descriptors and activity instead of considering descriptors of highest variance, as final results rely more on correlation than on variance.

The algorithm we followed for variable reduction is as follows

1.   Define appropriate correlation cutoff value between descriptor & activity, which is mentioned as $A_{max.}$ Remove all descriptors which have value less than A_{max}

2.   Define appropriate cross correlation cutoff value between descriptor-descriptor, which is defined as $C_{max.}$ Remove all descriptors which have values larger than $C_{max.}$

3.   Define variance cutoff value for descriptor which is mentioned as V_{max}. The descriptors having variance value less than V_{max} were removed.

We observed that this algorithm reduced No. of descriptors nearly up to 50%. After this, we applied manual variable selection method and multiple regression analysis (MLR) as it is ensured that each remaining descriptor is significantly contributing for QSAR model. The only thing we have to find out is No. of descriptors in final equation should be as low as possible which must be contributing highly and should be seen in the structural features of the reported compounds of the series as well.

Three-dimensional QSAR (3D QSAR) studies

Alignment of molecules

Proper alignment of molecules is the prerequisite for studying 3D QSAR as well as in almost all the fields of drug discovery for getting reliable results. So after optimization, we carried out alignment of all molecules using MolSign which also serves as the basic tool to identify the common pharmacophore features as well as the individual molecular feature.
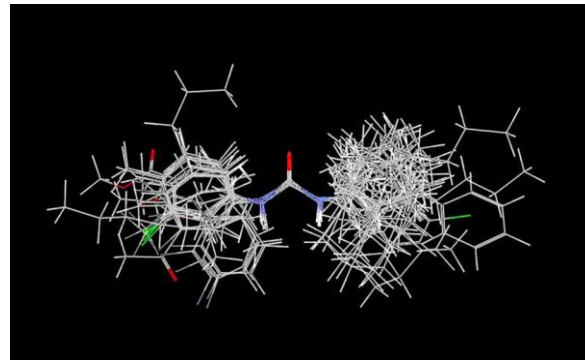

Fig.3: Identified pharmacophore features and alignment of molecules by MolSign

The colour scheme for identification of various chemical features is as follows:

Hydrogen bond donor: Magenta colour
Hydrogen bond acceptor: Buff colour
Hydrophobic: Orange colour
Aliphatic: Orange colour
Negative ionizable: Green colour
Positive ionizable: Violet colour

The larger tessellated spheres are indicative of the common pharmacophore features identified in the molecules and the smaller solid features are of the individual molecules.

3D QSAR by SA-kNN-MFA

3D QSAR studies were performed by generation of numerous models by taking same molecules in the respective training and test sets as in 2D QSAR by using k-Nearest Neighbour-Molecular Field Analysis (kNN–MFA) methodology with Simulated Annealing (SA) variable selection method as it has been reported as the more relevant and suitable method to perform 3D QSAR [27,28,29]. kNN–MFA requires suitable alignment of given set of molecules after optimization which had already been carried out by MolSign, but it was again carried out to generate a folder of aligned molecules to proceed for 3D QSAR by atom based alignment which gives alignment based on each and every individual atom of the pharmacophore. Molecular alignment was used to visualize the structural diversity in the given set of molecules. It was followed by generation of common rectangular grid around the molecules. Steric and electrostatic interaction energies were computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values at the grid points are considered for relationship generation using kNN method and utilized as descriptors for obtaining distances within this method. Resulting set of aligned molecules was then used to build 3D QSAR model.

Design of New Chemical Entities (NCEs) containing urea pharmacophore

The information obtained from 2D and 3D-QSAR studies was utilized in optimizing urea pharmacophore and to design potent anti-tubercular NCEs. Substitution pattern around pharmacophore, shown in Fig. 6 was used to design NCEs using CombiLib tool of VLife MDS software. Designed compounds were subjected to Lipinski's screen [30] to ensure their drug like pharmacokinetic profile in order to improve their bioavailability. The following parameters were used as Lipinski's filters (values in parenthesis indicate ideal requirements):

1. Number of Hydrogen Bond Acceptor (A) (<10)
2. Number of Hydrogen Bond donor (D) (<5)
3. Number of Rotatable Bond (R) (<10)
4. XlogP (X) (<5)
5. Molecular weight (W) (<500 g/mol)
6. Polar surface area (S) is (<140 Å)

Molecular Docking studies

All the designed compounds that showed good predicted activity by all QSAR studies and followed Lipinski's rule as well as reported series molecules for comparison purpose were subjected to molecular docking for studying the binding mode of designed compounds and were further screened to sort out the best compounds having good binding affinity compared with binding mode of standard diaryl urea. The main molecular docking tool used was GLIDE (Maestro; Schrödinger Inc., USA) for protein-ligand docking studies in to the receptor epoxide hydrolase enzyme binding pocket (Glide, Schrödinger9.0, LLC, New York, USA). The crystal structures of EH were obtained from Protein Data Bank (PDB Code: 2ZJF) [31]. All structures were prepared for docking using 'Protein preparation wizard' and 'Ligand preparation wizard' in Maestro wizard of Schrödinger 9.0. In the refinement component, a restrained impact minimization of the co-crystallized complex was carried out. It uses the OPLS-AA force field for this purpose. The co-crystallized ligand was removed from active site and the grids were defined by centering them on the ligand in the crystal structure. Then our structures (ligands) were imported in the project table, built using maestro structure builder panel and prepared by Ligprep module which produces the low energy conformers of ligands using Merck Molecular force field. The lower energy conformations of the ligands were selected and docked into the grid generated from protein structure using extra precision (XP) docking mode. In this docking method, the ligands are flexible and receptor is rigid, except the protein active site which has slight flexibility. The final evaluation is done with glide score (docking score) and single best pose is generated as the output for particular ligand.

G-score = a*vdw + b*coul+ Lipo + H-bond + Metal + BuryP + Rot B + Site

where, vdW, Vander Waal energy; Coul, Coulombic energy; Lipo, Lipophilic contact term; HBond, Hydrogen-bonding term; Metal, Metal-binding term; BuryP, Penalty for buried polar groups; RotB, Penalty for freezing rotatable bonds; Site, Polar

interactions at the active site. The coefficients of vdW and Coul are: a = 0.065, b = 0.130 respectively.

The accurate prediction of protein-ligand interaction geometries is essential for the success of virtual screening approach in structure-based drug design. The docking results were evaluated based on Glide Score (G-Score), Hydrogen bonds (H-bond) and Vander Waals (vdW) interactions between ligand and receptor.

Prediction of ADME properties

Sometimes compounds that show very high activity in vitro however are proved later to have no in vivo activity, or to be highly toxic in in-vivo models. Lack of in vivo activity may be attributed to undesirable pharmacokinetic properties and the toxicity may result from the formation of reactive metabolites. The failure of NCEs at latter stages of drug discovery process due to lack of drug like pharmacokinetic profile has forced us to set filters of ADMET properties. Thus we have ensured that only drug-like NCEs would be selected for experimental validation. All designed compounds which showed good binding affinity were filtered by predicting their Absorption, Distribution, Metabolism and Excretion (ADME) properties by means of QikProp tool of Schrodinger 9.0 [32].Prediction of ADME properties was used as the last screen to sort out those compounds which already followed Lipinski's rule, showed good predicted activity as well as good binding affinity with EH enzyme.

ADME prediction by QikProp, Schrödinger 9.0

It predicts both physicochemical significant descriptors and pharmacokinetically relevant properties. It also evaluates the acceptability of analogues based on Lipinski's rule of 5, which is essential to ensure drug like pharmacokinetic profile while using rational drug design. All the analogues were neutralized before being used by QikProp. This program is designed using the BOSS program and the OPLS-AA force field. It uses Monte Carlo statistical mechanics simulations on organic solutes in periodic boxes of explicit water molecules to perform all

predictions. This process resulted in configurationally averages for a No. of descriptors, including H-bond counts and solvent-accessible surface area (SASA). Correlations of these descriptors to determine properties experimentally were obtained and then algorithms that mimic the full Monte Carlo simulations and produce comparable results were developed by the QikProp tool.

3. RESULTS AND DISCUSSION:

2D QSAR models

Using MLR, one set of 6 meaningful descriptors, out of which Quadrupole1showed up to 35% contribution for antitubercular activity.

pMIC= 0.6229 T_C_C_4 – 0.2090 T_C_F_4 – 0.080 SaaCHcount - 0.04262 Quadrupole3 + 0.4757 ZcompDipole – 0.0814 Quadrupole1+ 0.0235

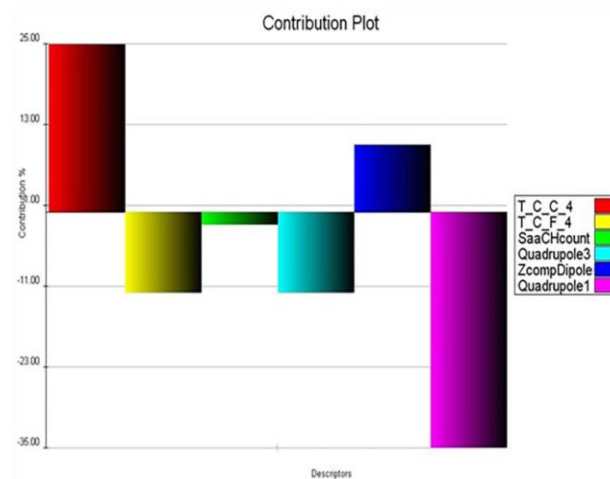$r^2 = 0.9743$, $q^2 = 0.9134$, F-test = 37.93, Pred_r^2 = 0.5650.

Fig.4: Contribution plot of selected descriptors

From the results, it is known that Quadrupole1alone satisfies all evaluation parameters. It shows highest correlation with activity (as shown in correlation matrix) and also shows proper distribution of data points. To increase the predictive power, different combinations of descriptors were made.

Table 4: Statistical parameters of developed QSAR models for descriptor by forward variable selection method and MLR

Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a_1	Test set a_2	Test Set b_1	Test Set b_2	Test Set c_1	Test Set c_2
N	22	22	22	22	22	22
r^2	0.9723	0.8631	0.9743	0.9032	0.9000	0.8579
r^2 se	0.3319	0.5056	0.2556	0.3010	0.3909	0.5314
q^2	0.7816	0.8252	0.9134	0.6996	0.7884	0.7267
q^2 se	0.9317	0.8235	0.4693	0.6869	0.5686	0.7368
F-Test	15.93	18.52	37.93	13.46	38.99	10.06
Pred_r^2	0.5173	0.5010	0.5650	0.4865	0.5024	0.5822
Pred_r^2 se	1.0545	1.2375	0.4652	1.5619	1.1266	0.9779
	(+)VELY CONTRIBUTING			(-)VELY CONTRIBUTING		
Descriptors (Test set b_1)	T_C_C_4 (25%) ZcompDipole (11%)			T_C_F_4 (12%) SaaCHcount (2%) Quadrupole3 (12%) Quadrupole1 (35%)		

Accuracy of model

The value of residuals is a key factor in validating the accuracy of the model. As the value of residual is near to zero, the model is considered as more accurate since it shows minimum (≈0) difference in actual and predicted activity.

Table 5: Test set b_1 and b_2 and Training set A+C along with biological activity,predicted activity and residuals data

Sr. no.	Compd.	Biological Activity (pMIC)	Predicted Activity	Residuals
		Training Set A+C		
1	v24	7.4698	7.5968	-0.1270
2	v14	5.2994	5.4303	-0.1309
3	v13	4.9625	4.7889	0.1727
4	v16	4.3016	3.9820	0.3196
5	v40	4.0498	4.0850	-0.0352
6	v12	6.4948	6.5071	-0.0123
7	v02	5.9089	5.7853	0.1236
8	v05	7.2132	7.3612	-0.1480
9	v07	5.5744	5.3965	0.1779
10	v09	3.713	3.5457	0.1673
11	v28	5.6185	5.8710	-0.2525
12	v30	5.3358	5.3375	-0.0017
13	v36	5.9350	6.1888	-0.2538
		Test set b1		
1	v23	4.9601	5.9978	-1.3077
2	v26	4.7226	5.6169	-0.8943
3	v22	7.2212	7.0910	0.1302
4	v20	7.4449	5.9001	1.5448
5	v21	6.1650	6.9926	-0.8276
6	v37	3.7265	4.3677	-0.6412
7	v06	6.2126	6.0430	0.1696
8	v04	4.6817	4.8293	-0.1476
9	v08	4.3380	3.5887	0.7493

Residuals = Actual Biological Activity (MIC) – Predicted Activity

Similar calculations were carried out for Test sets a_1 & a_2 and c_1 & c_2 of Set-I as well as for Set-II. In both the sets, models b_1 & b_2 (evaluated on training set A+C) were found to be the best models.

Interpretation of 2D QSAR

The present QSAR model reveals that Quadrupole1 as well as T_C_C_4 descriptors have major contribution in explaining variation in activity. Descriptors T_X_Y_Z can be defined as total count of fragments formed with atom types X and Y separated by topological distance of Z bonds (Baumann, 2002; Hall and Kier, 1995). Interpretation of descriptors which contributed significantly for QSAR models are given below, the value given in parenthesis are percentile contribution of descriptor for the activity:

1)   T_C_C_4 (25%)**:** No. of carbon atoms separated from any carbon atom by 4-bond distance
2)   T_C_F_4 (-12%)**:** No. of carbon atoms separated from any fluorine atom by 4-bond distance
3)   SaaCHcount (-2%)**:** Total no. of carbon atoms connected with a hydrogen along with two aromatic bonds
4)   Quadrupole3 (-12%)**:** Magnitude of third tensor of quadrupole moments

5)   ZcompDipole (11%)**:** Z componant of dipole moment
6)   Quadrupole1 (-35%)**:** Magnitude of first tensor of quadrupole moments.

Careful observation of descriptors in models suggests that Quadrupole1 is an indicator variable which negatively contributes for QSAR equation (35%) and this signifies antitubercular activity. Other descriptor like T_C_F_4 which is inversely proportional to activity shows the presence of -F on ring is detrimental for biological activity. Whereas the other descriptors are contributing positively in more or less percentage revealing the importance of respective atoms/groups at different respective position on the ring for potential antitubercular molecule design.

3D QSAR models

Using SA-kNN-MFA, 3 descriptors were finalized which were satisfying all statistical parameters in the generated models. After calculating residuals for each model, models b_1 and b_2 were found to be the best models.

Table 6: Comparison of the various statistical results of 3D QSAR generated by SA-kNN-MFA method

| Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a1	Test Set a2	Test Set b1	Test Set b2	Test Set c1	Test Set c2
N	22	22	22	22	22	22
k-NN	2	2	2	2	2	2
q^2	0.8480	0.9358	0.8168	0.8096	0.7523	0.7895
q^2 se	0.2626	0.2961	0.3010	0.4021	0.5236	0.4756
Pred_r^2	0.7620	0.6871	0.9285	0.8231	0.7486	0.8421
Pred_r^2 se	0.6452	0.8067	0.3855	0.5421	0.6742	0.5123
	(+)VELY CONTRIBUTING			(-)VELY CONTRIBUTING		
DESCRIPTORS (Test Set b_1)	H_865 (0.789483, 0.798317) H_832 (0.37942, 0.421762)			S_790 (-0.018441, -0.015148)		
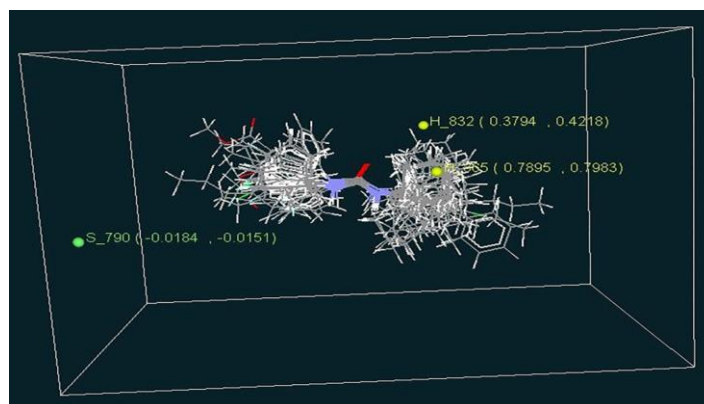
Fig. 5: Grid points generated by SA-kNN–MFA method in 3D rectangular grid showing contributions of electrostatic, steric and hydrophobic functional groups for significant antitubercular activity (test set b₁)

Interpretation of 3D QSAR

3D QSAR was used to optimize the electrostatic, steric and hydrophobic requirements around Urea pharmacophore. The property values for the generated data points helped us for the design of potent NCEs. The ranges of data point values were based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbour set. Points generated in SA-kNN-MFA 3D QSAR model are H_865 (0.789483, 0.798317), H_832 (0.37942, 0.421762) and S_790 (-0.018441, -0.015148) i.e. two hydrophobic and a steric data points at lattice points of 865, 832 and 790 respectively.

Positive value in hydrophobic data points indicated the requirement of more hydrophobic substituent for enhancing biological activity.

Low range of positive steric value indicated that highly bulky groups (i.e.-C_6H_5, -adamantly) are required to increase activity.

Design of New Chemical Entities (NCEs) containing N-phenyl-2, 2-dichloroacetamide pharmacophore

A total of 392 NCEs were designed by applying interpreted 2D and 3D QSAR results. Out of 392, only 15 NCEs exhibited good predicted activity as compared to compounds of the original series by applying all two QSAR predictions along with Lipinski's screen score 5/6 and hence were selected for further in-silico studies.
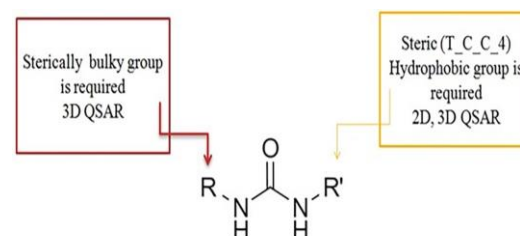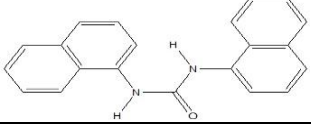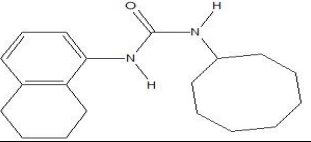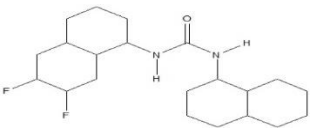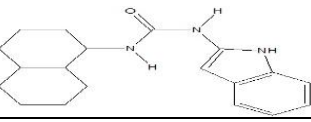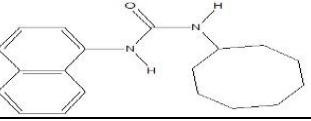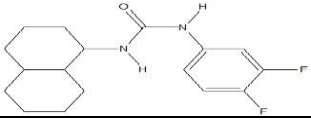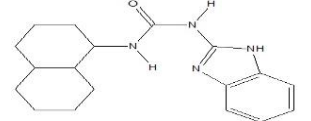

Fig. 6: Pharmacophoric requirements around Urea nucleus from 2D and 3D QSAR

Table 7: Structures of designed NCEs along with predicted activity obtained by 2D (both descriptor sets) and 3D QSAR equations (arranged in descending order)

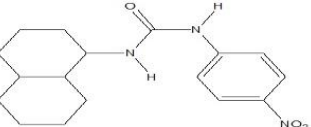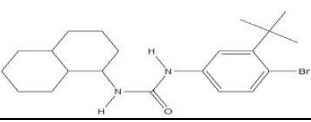| Sr. No.	Molecule Name	Structure	H accept bond	xLogP	M.W.	Polar surface Area	Screen Result	Screen Score	Predicted Activity
1	VN1		3	5.331	332.529	41.13	ADRWS	5	10.1128
2	VN2		3	5.2150	322.450	41.13	ADRWS	5	9.3555
3	VN3		3	4.9810	326.482	41.13	ADRXWS	6	9.2736
4	VN4		3	4.8650	316.402	41.13	ADRXWS	6	8.5335
5	VN5		3	4.6310	320.434	41.13	ADRXWS	6	8.2895
6	VN6		3	5.1370	306.491	41.13	ADRWS	5	8.2865

7	VN7		3	5.0990	312.370	41.13	ADRWS	5	7.9130
8	VN8		3	4.7870	300.444	41.13	ADRXWS	6	7.4408
9	VN9		5	4.7650	368.510	41.13	ADRXWS	6	7.0709
10	VN10		3	4.5580	311.427	56.92	ADRXWS	6	7.0279
11	VN11		3	5.0210	296.412	41.13	ADRWS	5	6.8322
12	VN12		6	3.8010	317.388	86.95	ADRXWS	6	6.6863
13	VN13		3	5.0060	365.313	41.13	ADRWS	5	6.6174
14	VN14		5	4.1330	308.371	41.13	ADRXWS	6	6.5272
15	VN15		3	3.7400	312.414	69.81	ADRXWS	6	6.5176

Results of molecular docking studies

In molecular docking results (Maestro, Schrödinger 9.0), it was found that urea analogues mimic Bsu 360 (A) and bind to the Bsu 360 (A) binding region of epoxide hydrolase active site.

Evaluation of molecular docking results with enzyme EH (PDB code: 2ZJF)

G-score

The scoring function of GLIDE docking program is presented in G-score form. G-score indicates the binding affinity of the designed compound to the receptor or enzyme. G-score of standard compound Thioacetazone was found to be -7.8346; whereas out of 15 designed NCEs, only 6 showed better G-score than the standard. The G-score of the designed NCEs VN14, VN13, VN11, VN06, VN02 and VN07 was found to be -10.4141, -9.1733, -8.8281, -8.5703, -8.3076 and -7.83468 respectively. More negative is the value of G-score, higher the binding affinity of that compound. The close analysis of these results

suggests that the designed NCEs have more binding affinity with enzyme than standard.

H-Bond interactions

H-bond is one of the most widely used parameter for the evaluation of the docking results, as it is an influential parameter in the activity of drug compound. The numbers of H-bond interactions as well as their length in standard were compared with that of the designed NCEs. Thioacetazone itself involves single important H-bonding interactions with the key binding amino acids Phe-36 of the protein backbone as reported in the standard ligand plot. Here also, out of 15 NCEs, only 6 compounds

showed better results than standard as well the most potent compound of the series (v24). The H-bond length is also an important parameter in molecular docking studies as if it is found lesser than the H-bond length of standard, then it means that our compounds bind with higher affinity with the respective amino acid. Here we found all the compounds have shorter length of both the H-bonds compared to standard. Compound VN14 showed the highest binding affinity with the key binding amino acids Asp-104 and Tyr-164 in the binding pocket of EH.

Table 8: Results of molecular docking studies performed using extra precision mode of Glide (Maestro, Schrödinger) (arranged in descending order) with EH

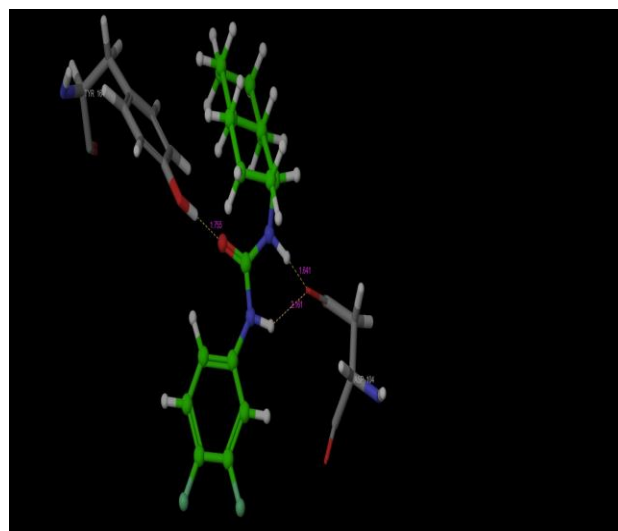| Sr. No	Title	G-score	E-Model	H-Bond
1	VN14	-10.4141	41.4112	0
2	VN13	-9.1733	13.9871	0
3	VN11	-8.8281	9.9924	-1.2704
4	VN06	-8.5703	23.4568	-1.575
5	VN02	-8.3076	54.0869	-0.6946
6	VN07	-7.8346	20.9319	0
7	STD (Thioacetazon)	-1.9539	-41.4952	0

Fig.7: Binding pose of compound VN14 in receptor binding pocket of EH (PDB Code: 2ZJF)
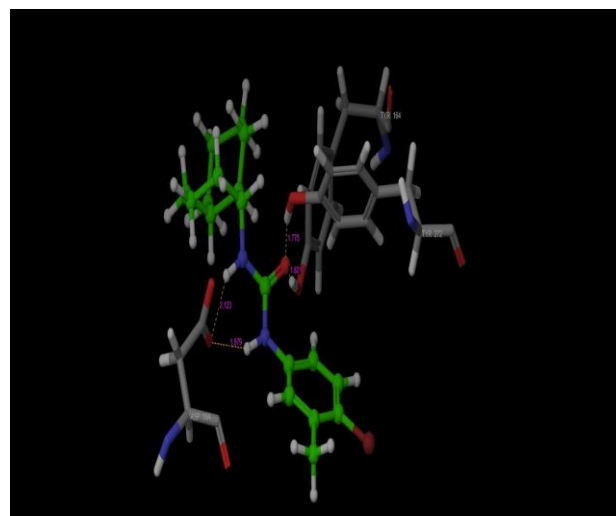

Fig.8: Binding pose of compound VN13 in receptor binding pocket of EH (PDB Code: 2ZJF)
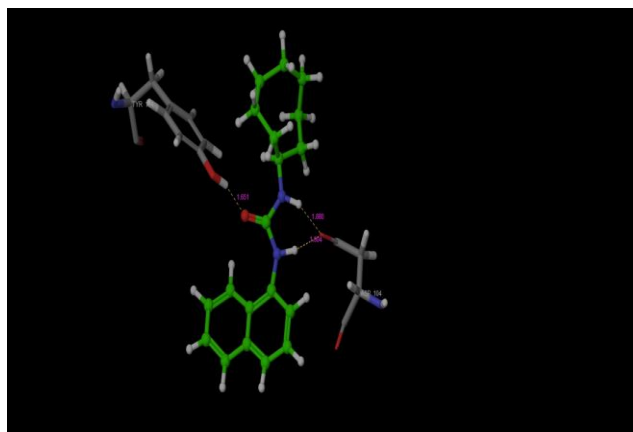
Fig.9: Binding pose of compound VN11 in receptor binding pocket of EH (PDB Code: 2ZJF)
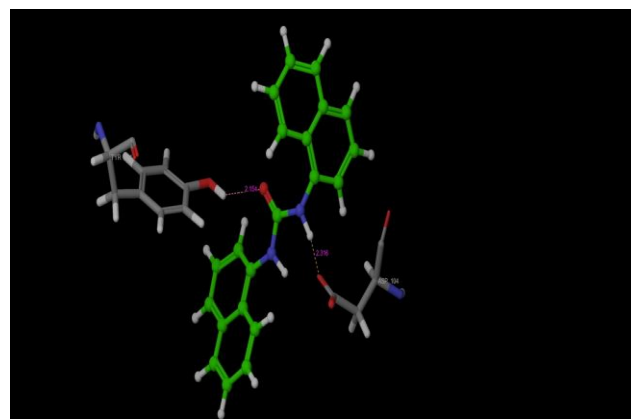

Fig.12:Binding pose of compound VN07 in receptor binding pocket of EH (PDB Code: 2ZJF)
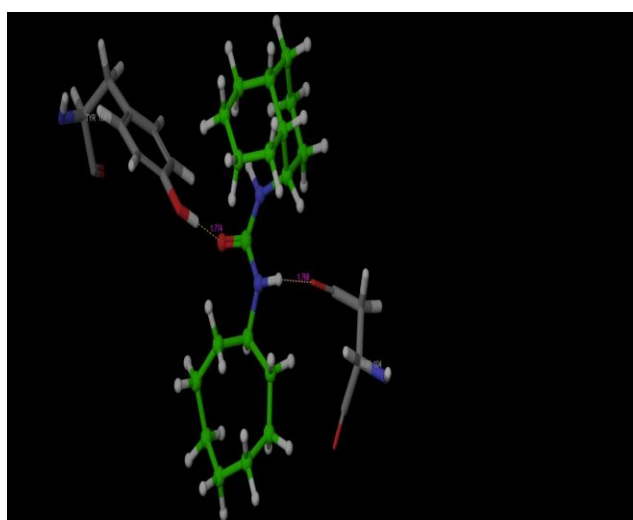

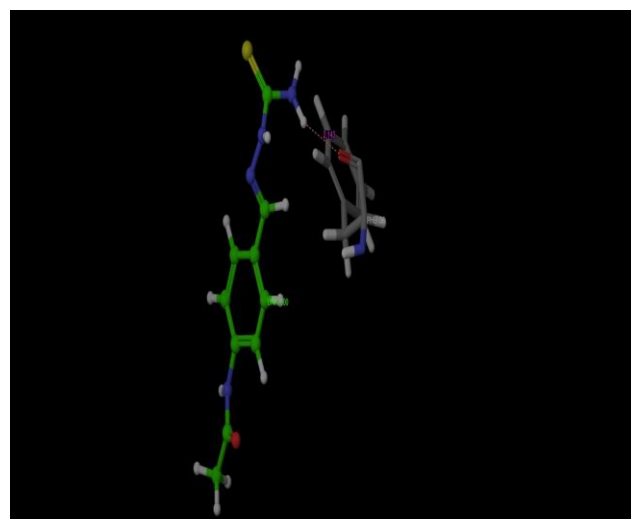**Fig.10: Binding pose of compound VN06 in receptor binding pocket of EH (PDB Code: 2ZJF)**

Fig.13: Binding pose of standard Thioacetazon in receptor binding pocket of EH (PDB Code: 2ZJF)

ADME predictions

The NCEs generated using CombiLib were analyzed by Lipinski's rule to ensure their drug-like pharmacokinetic profile while designing. In addition to that, their other ADME properties were also predicted and compared with their ideal ranges for a chemical entity to act as a drug. Out of designed entities 8 NCEs showed satisfactory results within the ideal ranges.

ADME prediction by QikProp, Schrödinger

Numbers of properties of designed analogues were predicted by QikProp tool, Schrödinger 9.0 which was used as last screening tool to select the final NCEs. Here we have reported only descriptors which contributed significantly for predicting drug like properties of the molecule. These properties are as follow: (figures in parenthesis indicate ideal values in order the test compounds to have drug like pharmacokinetic properties)

Fig.11: Binding pose of compound VN02 in receptor binding pocket of EH (PDB Code: 2ZJF)

1)   Lipinski's Rule: Compounds that satisfy this rule were expected to have drug like pharmacokinetic profile.
2)   Brain/blood partition coefficient (CNS) (-2 to 2)
3)   Per cent Human Oral absorption (>80% is high, <25% is poor)
4)   Number of possible metabolites (should range from 1-8)
Ideal range of Lipinski's rule properties have been satisfied by the finally considered 6 NCEs. Hence, it concludes that these NCEs can act as drug and have drug like bioavailability.
CNS parameter is related with absorption of entity through Blood brain barrier; standard limit for CNS is -2 to +2, where -2 show inactive CNS penetration and +2 shows active CNS penetration. All the designed compounds must have CNS parameter value as -2 (inactive BBB penetration), which has been satisfied

by all 6 finally considered NCEs. Therefore all are considered as nontoxic compounds.
Per cent human oral absorption parameter is related with extent of oral absorption of drug, indicating suitable route of administration and exhibit the extent of oral bioavailability profile. All NCEs showed 100% human oral absorption, hence we can say that all finally considered designed NCEs can be orally absorbed and will exhibit bioavailability profile.
All compounds showed No. of metabolites in ideal range as well as relatively more No. of metabolites as compared to standard so that we can say that they can be metabolized and excreted easily (mostly via urine) and will not cause any side effect or will not produce any toxic metabolite.
So after detailed analysis of ADME predictions, it will worth saying that all the 6 selected compounds can exhibit drug like pharmacokinetic profile.

Table 9. Results of ADMET properties

Sr. No.	Comp	Mol. Wt.	Donor HB	Acceptor HB	QPlog Po/w	Percent Human Oral Absorption %	CNS	No.Of possible metabolites
1	VN14	308.37	2	2	4.02	100	1	0
2	VN13	365.31	2	2	4.25	100	1	1
3	VN11	296.41	2	2	3.97	100	1	0
4	VN06	306.49	2	2	4.16	100	1	0
5	VN02	322.44	2	2	4.31	100	1	0
6	VN07	312.37	2	2	4.22	100	1	0

From the results of molecular modelling studies, six compounds (VN14, VN13, VN11, VN06, VN02, VN07) have been selected for further experimental validation.

4. CONCLUSION:
The results of 2D and 3D QSAR studies using different interpretation approaches have yielded detailed insights of proper working and ways of thinking in this research area. In the present study objective of optimization of the selected Urea pharmacophore using molecular modeling studies was found to be achieved; as predicted activity of NCEs was found significantly greater (VN02=9.3555) than most potent compound of original series (v24=7.4698). The resulting QSAR models were found to be have generated significantly good statistical results i.e. ($r^2 > 0.7$), cross-validation ($q^2 > 0.6$), and external validation (pred_$r^2 > 0.6$), it indicates high predictive ability of all models. In all it is worth concluding that rational used for optimization of Urea pharmacophore using 2D, 3D

QSAR, ADME, Molecular Docking studies was found to be significantly accurate.
The results of activity prediction show significant improvement of activity. Ultimately QSAR studies are useful in understanding the structural requirement to design novel potent molecules. The binding affinity for design NCEs were checked by subjecting them to docking studies in active binding site of protein receptor. A screening approach has thus facilitated the identification of suitable compounds from designed library for antitubercular activity.

necessary infrastructure facilities to carry out this research work.

REFERENCES:

1.Pablos-Mendez A, Raviglione MC, Laszlo A, Binkin N, Reider HL, Bustreo F, Cohn DL, Lambregts-van Weezenbeek CS, Kim SJ, Chaulet P, Nunn P. Global surveillance for antituberculosis-drug resistant. N Engl J. Med., 1998; 338:1641-1649.

2.WHO Report, Global Tuberculosis Control Surveillance, Financing and Planning, 2010.

3.Multidrug-resistant tuberculosis (MDR-TB) October 2013 Update, WHO.

4.Louw GE, Warren RM, Gey Van Pittius NC, McEvoy CRE, Van Helden PD, Victor TC. A Balancing Act: Efflux/Influx in Mycobacterial Drug Resistance. Antimicrobial Agents and Chemotherapy, 2009; 53:3181–3189.

5.Narang SK. Extensively Drug Resistant Tuberculosis(XDR-TB). Jour of Med Edu and Res., 2009; 11:102-103.

5.Biswal BK, Morisseau C, Garen G, Cherney MM, Garen C, Niu C, Hammock BD, James MN. The molecular structure of epoxide hydrolase B from Mycobacterium tuberculosis and its complex with a urea-based inhibitor**.** J Mol Biol., 2008; 381: 897-912.

6.Morisseau C, Hammock BD. Epoxide hydrolases: mechanisms, inhibitor designs, and biological roles. Annu Rev Pharmacol Toxicol., 2005; 45: 311-333.

7.Biswal BK, Garen G, Cherney MM, Garen C, James MN. Cloning, expression, purification, crystallization and preliminary X-ray studies of epoxide hydrolases A and B from Mycobacterium tuberculosis. Acta Crystallogr Sect F Struct Biol Crystallogr Commun., 2006; 62: 136-138.

8.Silver LL. Multi-targeting by monotherapeutic antibacterials. Nat Rev Drug Disc., 2007; 6: 41-55.

9.Ananthan S, Faaleolea ER, Goldman RC, Hobrath JV, Kwong CD, Laughon BE, Maddry JA, Mehta A, Rasmussen L, Reynolds RC, Secrist 3rd JA, Shindo N, Showe DN, Sosa MI, Suling WJ, White EL. High-throughput screening for inhibitors of Mycobacterium tuberculosis H37Rv. Tuberculosis (Edinb)., 2009; 89: 334-353.

10.Usha V, Gurcha SS, Lovering AL, Lloyd AJ, Papaemmanouil A, Reynolds RC, Besra GS. Identification of novel diphenyl urea inhibitors of Mt-GuaB2 active against Mycobacterium tuberculosis. Microbiology., 2011; 157: 290-299.

11.Seth PP, Ranken R, Robinson DE, Osgood SA, Risen LM, Rodgers EL, Migawa MT, Jefferson EA, Swayze EE. Aryl urea analogs with broad-spectrum antibacterial activity. Bioorganic and Medicinal Chemistry Letters, 2004; 14: 5569–5572.

12.Li HQ, Lv PC, Yan T, Zhu HL. Urea derivatives as anticancer agents. Anticancer Agents Med Chem., 2009; 9: 471-480.

13.Kocyigit-Kaymakcioglu B, Celen AO, Tabanca N, Ali A, Khan SI, Khan IA, Wedge DE. Synthesis and biological activity of substituted urea and thiourea derivatives containing 1,2,4-triazole moieties. Molecules, 2013; 18: 3562-3576.

14.Skillman TG, Feldman JM. The pharmacology of sulfonylureas. The American Journal of Medicine, 1981; 70: 361–372.

15.Vajragupta O, Pathomsakul A, Matayatsuk C, Ruangreangyingyod L, Wongkrajang Y, Foye WO. Synthesis and antihypertensive activity of N-(alkyl/alkenyl/aryl)-N-heterocyclic ureas and thioureas. J Pharm Sci., 1996; 85: 258-61.

16.Perveen S, Mustafa S, Khan MA, Dar A, Khan KM, Voelter W. Substituted Urea Derivatives: A Potent Class of Antidepressant Agents. Medicinal Chemistry, 2015; 11: 330-336.

17.Brown JR, North EJ, Hurdle JG, Morisseau C, Scarborough JS, Sun D, Korduláková J, Scherman MS, Jones V, Grzegorzewicz A, Crew RM, Jackson M, McNeil MR, Lee RE. The structure-activity relationship of urea derivatives as anti-tuberculosis agents. Bioorg Med Chem., 2011; 19: 5585-5595.

18.In-Hae Kim, Nishi K, Kasagami T, Morisseau C , Liu JY, Tsai HJ, Hammock BD. Biologically active ester derivatives as potent inhibitors of the soluble epoxide hydrolase. Bioorganic and Medicinal Chemistry Letters,2012; 22: 5889–5892.

19.Scherman MS, North EJ, Jones V, Hess TN, Grzegorzewicz AE, Kasagami T, Kim H, Merzlikin O, Lenaerts AJ, Lee RE, Jackson M, Morisseau C, McNeil MR. Screening a library of 1600 adamantyl ureas for anti-Mycobacteriumtuberculosis activity in vitro and for better physical chemical properties for bioavailability. Bioorganic and Medicinal Chemistry, 2012; 20: 3255–3262.

20.North EJ, Scherman MS, Bruhn DF, Scarborough JS, Maddox MM, Jones V, Grzegorzewicz A, Yang L, Hess T, Morisseau C, Jackson M, McNeil MR, Lee RE. Design, synthesis and anti-tuberculosis activity of 1-adamantyl-3-heteroaryl ureas with improved in vitro pharmacokinetic properties. Bioorganic and Medicinal Chemistry, 2013; 21: 2587–2599.

21.Dudek AZ, Arodz T, Galvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): A review. Comb Chem High T Scr., 2006; 9: 213-228.

22.Tropsha A. Best practices for QSAR model development, validation and exploitation. Mol. Inf., 2010; 29: 478-488.

23.VLife MDS 3.5, 2004. Molecular Design Suite. VLife Sciences Technologies Pvt. Ltd., Pune, India. (www.vlifesciences.com)

24.Halgren TA. Merck molecular force field. II. MMFF94 Vander Waals and electrostatic parameters for intermolecular interactions. J Comp Chem., 1996; 17: 520-552.

25.Topliss TG, Edwards RP. Chance factors in studies of quantitative structure- activity relationships. J Med Chem., 1979; 22: 1238-1244.

26.Zheng W, Tropsha A. Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbour principle. J Chem Inf Comput Sci., 2000; 4: 185-194.

27.Shen M, Xiao Y, Golbraikh A, Gombar VK, Tropsha A. Development and validation of k-nearest-neighbor, QSPR models of metabolic stability of drug candidates. J Med Chem., 2003; 46: 3013-3020.

28.Ajmani S, Jadhav K, Kulkarni SA. Three-dimensional QSAR using the k-nearest neighbour method and its interpretation. J Chem Inf Model., 2006; 46: 24–31.

29.Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development setting. Adv Drug Deliv Rev., 2001; 46: 3-26.

30.www.rscb.org (Accessed on 13-09-2014)

31.QikProp, Schrödinger Inc., version 9.0, LLC, New York, USA