Available online at http://jddtonline.info

RESEARCH ARTICLE

RATIONAL DESIGN OF ANTIBACTERIAL THIENOPYRIMIDINES BY 2D-QSAR STUDY

Mulla Jameel Ahmed S [1], Palkar Mahesh B[1], Maddi Veeresh S [1], *Khazi Imtiyaz Ahmed M[2]

[1]K.L.E. University's College of Pharmacy, Vidyanagar, Hubli-580031, INDIA
[2]Department of Chemistry, Karnatak University, Dharwad-580003, INDIA
*Corresponding author's E-mail: drimkorgchem@gmail.com; Phone: +91 9900768968; Fax: +91-836-2771275.

Received 29 Jan 2012; Revised 03 March 2012; Accepted 04 March 2012, Available online 15 March 2012

ABSTRACT

QSAR studies were performed on a set of 43 analogs of thienopyrimidine using V-Life Molecular Design Suite (MDS 3.5) QSAR plus module by using Multiple Linear Regression (MLR) and Partial Least Squares (PLS) Regression methods against a gram positive (*S.aureus*) and a gram negative (*E.coli*) bacteria. MLR method has shown a very promising prediction results in both *S.aureus* and *E.coli*. QSAR model was generated by a training set of 34 molecules with correlation coefficient (r^2) of 0.9849, 0.8719, significant cross validated correlation coefficient (q^2) of 0.8881, 0.7811 and F test of 40.4301, 40.4768 respectively. In the selected descriptors, alignment independent descriptors such as T_C_C_7, T_N_O_3, T_2_N_1, T_N_O_1, T_O_O_7 and T_N_Cl_4 were the most important descriptors in predicting antibacterial activity.

Keywords: Thienopyrimidine, Antibacterial, Multiple Linear Regression (MLR), Partial Least Squares (PLS) Regression

INTRODUCTION

Inappropriate and irrational use of antimicrobial medicines provides favourable conditions for resistant microorganisms to emerge, spread and persist. Infections caused by resistant microorganisms often fail to respond to conventional treatment, resulting in prolonged illness and greater risk of death. Antimicrobial resistance (AMR) is resistance of a microorganism to an antimicrobial medicine to which it was previously sensitive. Resistant organisms (they include bacteria, viruses and some parasites) are able to withstand attack by antimicrobial medicines, such as antibiotics, antivirals, and antimalarials, so that standard treatments become ineffective and infections persist and may spread to others. AMR is a consequence of the use, particularly the misuse, of antimicrobial medicines and develops when a microorganism mutates or acquires a resistant gene. AMR reduces the effectiveness of treatment because patients remain infectious for longer, thus potentially spreading resistant microorganisms to others. The achievements of modern medicine are put at risk by AMR. Without effective antimicrobials for care and prevention of infections, the success of treatments such as organ transplantation, cancer chemotherapy and major surgery would be compromised[1].

Quantitative structure activity relationship (QSAR) searches information relating chemical structure to biological and other activities by developing a QSAR model. Several molecular descriptors are used to quantify the structural feature of lead molecule. The purpose of using QSAR-Descriptors is to calculate the properties of molecules that serve as numerical descriptions or characterizations of molecules in other calculations such as diversity analysis or combinatorial library design. Using such an approach one could predict the activities of newly designed compounds before a decision is being made whether these compounds should be really synthesized and tested. Recently, Cao H *et al*., have reported the 3D QSAR study on a series of thienopyrimidines as highly selective inhibitors of three receptor tyrosine kinases (RTKs)[2]. Singh M *et al*., have reported a novel QSAR model of a series of thienopyrimidine derivatives for evaluating and predicting the inhibition activity of H_1-receptor antagonists[3].

Thienopyrimidines and other fused pyrimidines continue to attract considerable attention of researchers in different countries because of their great practical usefulness, primarily, due to a wide spectrum of their biological activities. Thienopyrimidines occupy a special position among these compounds. Along with some other pyrimidine systems containing an annelated five membered heteroaromatic ring, thienopyrimidines are structural analogs of biogenic purines and can be considered as potential nucleic acid antimetabolites[4]. Consequently, thienopyrimidines[5,6] have become a well sought-privileged class of compounds in drug discovery programs due to their wide variety of interesting biological activities observed for these compounds, such as antimicrobial[7-11], anticancer[12], antiviral[13], antitumor[14] and anti-inflammatory activity[15].

In view of the above facts and scope a 2D-QSAR study is performed on structurally-related thienopyrimidine derivatives against a gram positive (*S.aureus*) and a gram

negative (*E.coli*) bacteria; in order to get a better understanding of their structural features and antibacterial activity.

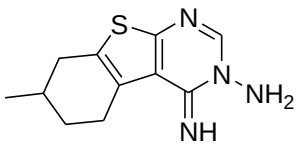
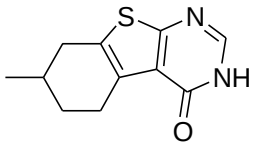
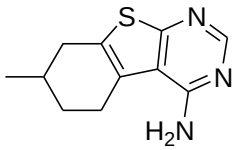
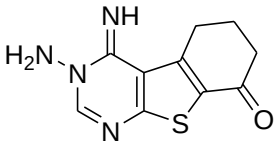
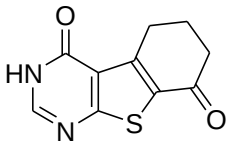
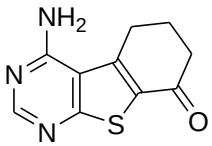
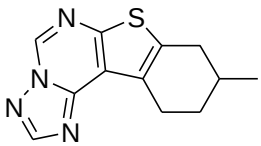
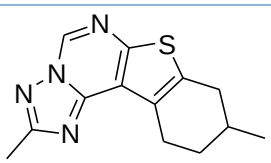
MATERIALS AND METHODS

2D-QSAR methodology

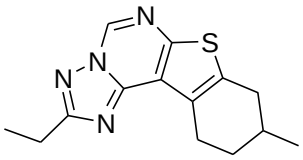
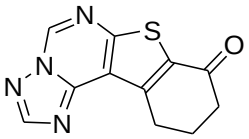
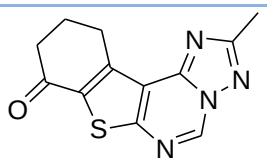
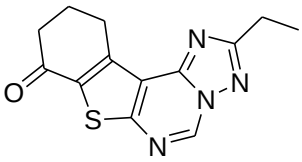
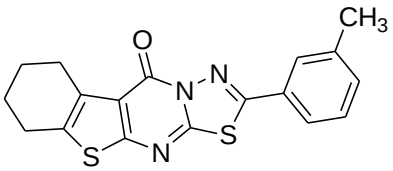
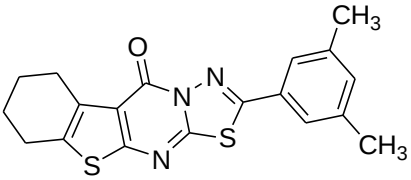
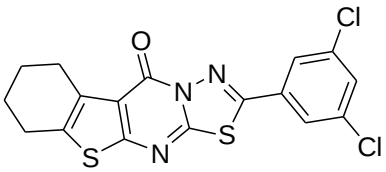
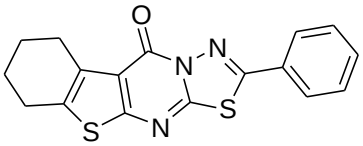
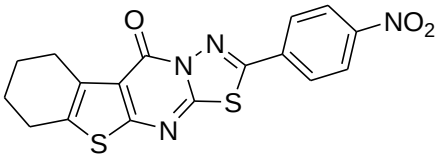
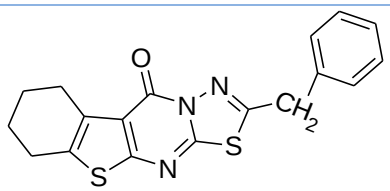
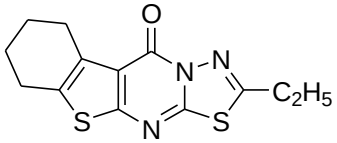
Data set

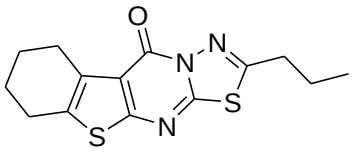
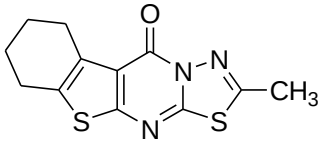
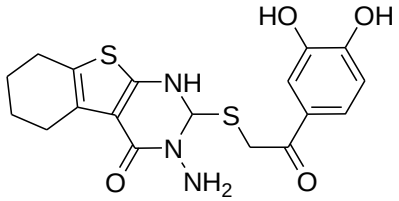
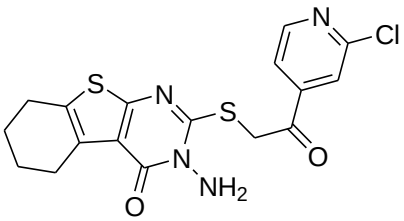
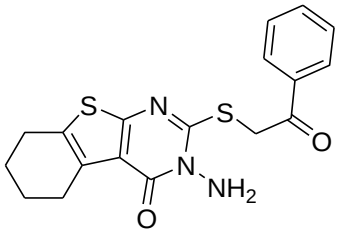
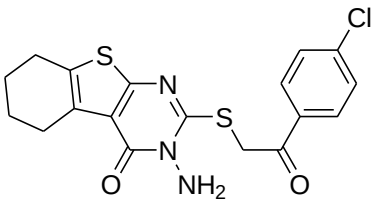
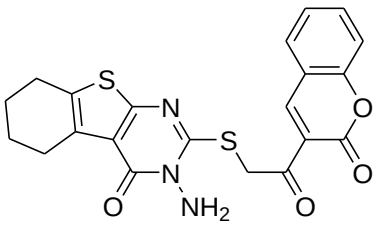
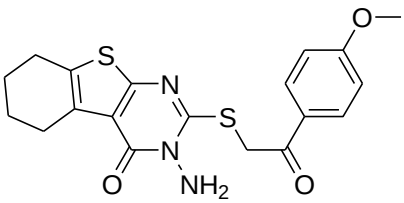
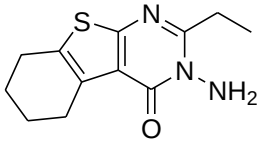
Forty-three molecules belonging to thienopyrimidine derivatives as antibacterial were taken from the literature[8,9,16,17] and used for QSAR analysis. The above reported series of thienopyrimidine derivatives showed wide variations in their structures and potency profiles. The 2D-QSAR models were generated for this series using multiple linear regression (MLR) and partial least squares (PLS) regression methods against *S.aureus* and *E.coli* and those models which come out with promising results are discussed here. QSAR models were generated by a training set of 34 molecules for each model. Predictive power of the resulting models was evaluated by a test set of 9 molecules with uniformly distributed biological activities. The structures of all the compounds along with their actual and predicted biological activities are presented in Table 1.

Table 1: Structure, Experimental and Predicted Activity of Thienopyrimidines

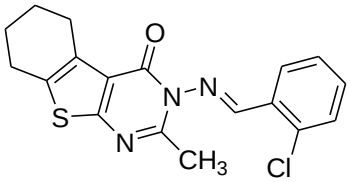
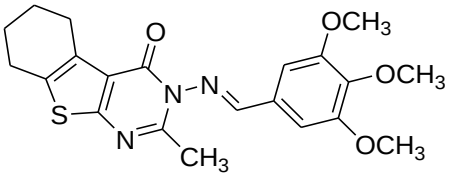
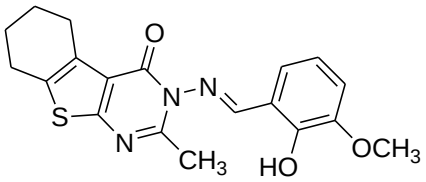
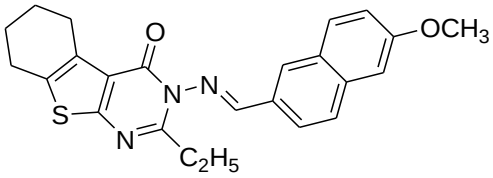
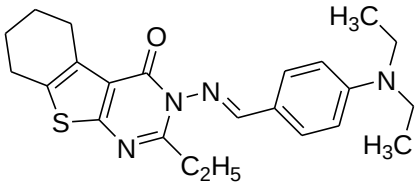
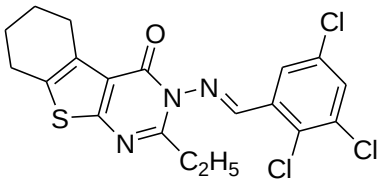
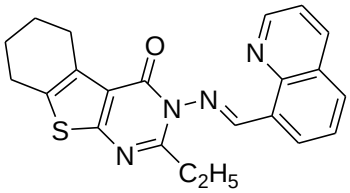
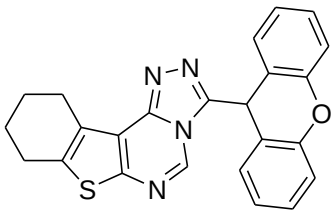
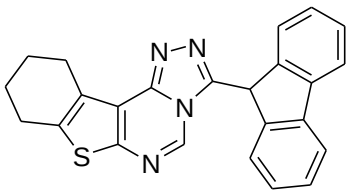
Sl.No.	Structure	pMIC[a] *(S.aureus)*			pMIC[a] *(E.coli)*		
		Exp.	Pred.	Residual	Exp.	Pred.	Residual
1		3.29	3.47	-0.18	3.59	3.66	-0.07
2		3.29	3.37	-0.08	3.59	3.66	-0.07
3		3.59	3.47	0.12	3.89	3.66	0.23
4		3.29	3.45	-0.16	3.59	3.62	-0.03
5		3.29	3.34	-0.05	3.59	3.62	-0.03
6		3.59	3.45	0.14	3.59	3.62	-0.03
7		3.59	3.80	-0.21	3.59	3.70	-0.11
8		3.89	3.85	0.04	3.59[T]	3.78	-0.19

 ISSN: 2250-1177

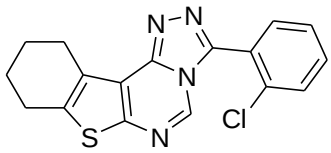
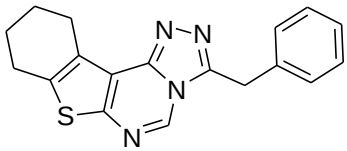
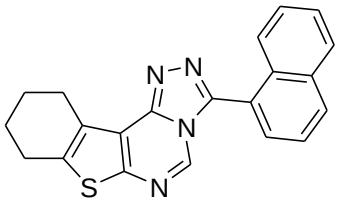
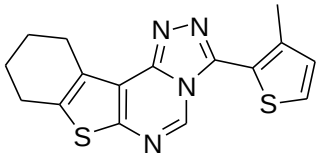
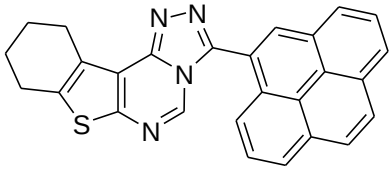
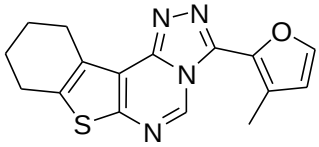
9		3.89	3.90	-0.01	3.59	3.86	-0.27
10		3.89	3.78	0.11	3.89	3.66	0.23
11		3.59[T]	3.83	-0.24	3.89[T]	3.74	0.15
12		3.89	3.88	0.01	3.59	3.82	-0.23
13		5.0	4.51	0.49	4.0	4.55	-0.55
14		4.0[T]	4.54	-0.54	4.22[T]	4.59	-0.37
15		4.0	4.49	-0.49	4.0	4.51	-0.51
16		4.22[T]	4.49	-0.27	4.0	4.51	-0.51
17		4.0	3.99	0.01	4.0[T]	4.51	-0.51
18		4.0	4.51	-0.51	4.09	4.55	-0.46
19		4.0[T]	4.34	-0.34	4.09[T]	4.27	-0.18

20		5.0	4.39	0.61	5.0	4.35	0.65
21		5.0[T]	4.31	0.69	5.0[T]	4.23	0.77
22		4.22	4.05	0.17	4.09	4.12	-0.03
23		4.09	4.44	-0.35	4.0	4.08	-0.08
24		4.90	4.26	0.64	4.0	4.12	-0.12
25		4.0	4.26	-0.26	4.0	4.12	-0.12
26		5.20[T]	4.31	0.89	4.90[T]	4.20	0.7
27		4.0	4.28	-0.28	4.0	4.16	-0.16
28		4.90[T]	4.11	0.79	4.90	4.23	0.67

 ISSN: 2250-1177

29		5.0	4.88	0.12	5.20	4.71	0.49
30		5.0	4.61	0.39	4.90	4.71	0.19
31		4.30	4.61	-0.31	4.90	4.36	0.54
32		5.0	4.79	0.21	5.0	4.98	0.02
33		5.0	4.87	0.13	5.0	4.94	0.06
34		4.69	4.98	-0.29	5.0	4.86	0.14
35		5.0	5.02	-0.02	4.90	5.02	-0.12
36		4.19	4.48	-0.29	4.49	4.61	-0.12
37		4.49[T]	4.43	0.06	4.49[T]	4.53	-0.04

 ISSN: 2250-1177

38		4.79	4.69	0.1	4.49	4.09	0.4
39		4.49	4.18	0.31	4.49	4.13	0.36
40		4.49	4.31	0.18	4.49	4.33	0.16
41		4.19[T]	4.13	0.06	4.49	4.05	0.44
42		4.49	4.51	-0.02	4.49	4.64	-0.15
43		5.39	5.46	-0.07	4.79[T]	4.97	-0.18

Expt. = Experimental activity; Pred. = Predicted activity; [a] = -Log (MICx10^{-6}); [T] = Test Set

Biological activities

The biological activities (MIC) were converted into the corresponding pMIC values (Eq. 1), where MIC value represents the lowest concentration of drug in microgram that inhibited the visible growth of a microorganism after overnight incubation. The MIC values of reference compounds were checked to ensure that no difference occurred between different groups. The pMIC values of the molecules under study spanned a range from 3.29-5.39 and 3.59-5.20 for *S.aureus* and *E.coli*, respectively.

$$pMIC = -\log(MIC \times 10^{-6}) \quad \text{..... Eq. (1)}$$

Computational data

The antibacterial activity data used for the QSAR analysis contains 43 molecules belonging to thienopyrimidine derivatives. All the structures of the compounds were drawn in 2D-APPL mode of software. The modeling analyses, calculations, and visualizations for 2D-QSAR were performed using the V-Life Molecular Design Suite 3.5 (Vlife MDS 3.5). The compounds were then subjected to conformational analysis and energy minimization using Montocarlo conformational search with RMS gradient of 0.001 kcal/mol using a MMFF force field. Montocarlo conformational search method is similar to the RIPS method that generates a new molecular conformation by randomly perturbing the position of each coordinate of each atom in molecule. Most stable structure for each compound was generated after energy minimization and used for calculating various physicochemical descriptors.

Molecular descriptors

The various descriptors selected for 2D QSAR were Individual, Retention Index (chi), Atomic valence connectivity index (chiv), Path Count, Cluster, Kappa.

Selection of training and test set

In order to obtain a validated QSAR model for the purpose of meaningful prediction, an available dataset should be divided into the training and test sets. For the prediction statistics to be reliable, the test set must include at least five compounds[18]. Ideally, the division into the training and test set must satisfy the following three conditions: (i) All representative compound-points of the test set in the multidimensional descriptor space must be close to those of the training set. (ii) All representative points of the training set must be close to those of the test set. (iii) The representative points of the training set must be distributed within the whole area occupied by the entire dataset[18].

 ISSN: 2250-1177

The dataset of 43 molecules was divided into training and test sets by sphere exclusion (SE) method for MLR and PLS model using pMIC activity field as dependent variable and various 2D descriptors as independent variables. In classical sphere exclusion algorithm the molecules are selected whose similarities with each of the other selected molecules are not higher than a defined threshold[19]. Each selected molecule generates a hyper-sphere around it, so that any molecule inside the sphere is excluded from the selection in the train set and driven towards the test set. The number of compounds selected and the diversity among them can be determined by adjusting the radius of the sphere (R). The different statistical models were developed using multiple linear regression (MLR) and partial least squares (PLS) regression methods. The equations were found to derive 2D-QSAR equation from different model building method (MLR and PLS) coupled with stepwise forward–backward variable selection method for assuming the biological activity with the help of physico-chemical descriptor values. The equations are discussed in Section 3 which come out with promising results from MLR and PLS.

Statistical parameters

Statistical measures used for the evaluation of models were the number of compounds in regression n, the regression coefficient r^2, the F-test (Fischer's value) for statistical significance F, the cross-validated correlation coefficient q^2 and the standard error of estimation r^2 and q^2. The regression coefficient r^2 is a relative measure of fit by the regression equation. It represents the part of the variation in the observed data that is explained by the regression. The correlation coefficient values closer to 1.0 represent the better fit of the regression. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F-test indicate that the model is statistically significant. Validation parameter, predictive r^2(r^2_pred) was calculated for evaluating the predictive capacity of the model. A value of r^2_pred greater than 0.5 indicates the good predictive capacity of the QSAR model.

Model validation

This is done to test the internal stability and predictive ability of the QSAR models. Developed QSAR models were validated by the following procedure.

Internal validation

Internal validation was carried out using leave-one-out (q^2, LOO) method. For calculating q^2, each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The q^2 was calculated using the equation (Eq. 2) which describes the internal stability of a model.

$$q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{mean})^2} \quad \text{..... Eq. (2)}$$

where y_i and $\hat{y}_i$ are the actual and predicted activity of the *i*th molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

External validation

The predictive ability of the selected model was also confirmed by external validation of test set compounds which is also denoted with pred_r^2. The pred_r^2 value is calculated as follows (Eq. 3):

$$Pred\_r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{mean})^2} \quad \text{..... Eq. (3)}$$

where y_i and $\hat{y}_i$ are the actual and predicted activity of the *i*th molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

Randomization test

To evaluate the statistical significance of the QSAR model for an actual dataset, one-tail hypothesis testing was used[20].The robustness of the models for training sets was examined by comparing these models to those derived for random datasets. Random sets were generated by rearranging the activities of the molecules in the training set. The statistical model was derived using various randomly rearranged activities (random sets) with the selected descriptors and the corresponding q^2 were calculated. The significance of the models hence obtained was derived based on a calculated Z score[21]. A Z score value is calculated by the following formula (Eq. 4):

$$Z\,score = \frac{(h - \mu)}{\sigma} \quad \text{..... Eq. (4)}$$

where h is the q^2 value calculated for the actual dataset, μ the average q^2, and σ is its standard deviation calculated for various iterations using models built by different random datasets.

QSAR by Multiple Linear Regression (MLR) Analysis

Multiple regression is the standard method for multivariate data analysis. It is also called as ordinary least squares regression (OLS). This method of regression estimates the values of the regression coefficients by applying least squares curve fitting method. For getting reliable results, dataset having typically 5 times as many data points (molecules) as independent variables (descriptors) is required.

The regression equation takes the form

$$Y = b1*x1 + b2*x2 + b3*x3 + c \quad \text{..... Eq. (5)}$$

Where, Y is the dependent variable, the 'b's are regression coefficients for corresponding 'x's (independent variable), 'c' is a regression constant or intercept. In the present study QSAR model was developed using multiple regression by forward-backward variable selection method with pMIC activity field as dependent variable and physicochemical descriptors as independent variable having cross correlation limit of 20 and number of variables in final equation were 5 and 3 for *S.aureus* and *E.coli* respectively. Selection of test and training set was done by sphere exclusion method.

QSAR by Partial Least Squares (PLS) Regression Method

PLS is an effective technique for finding the relationship between the properties of a molecule and its structure. In

 ISSN: 2250-1177

mathematical terms, PLS relates a matrix Y of dependent variables to a matrix X of molecular structure descriptors, i.e., a latent variable approach to modeling the covariance structures in these two spaces. PLS have two objectives: to approximate the X and Y data matrices, and to maximize the correlation between them. Whereas the extraction of PLS components is performed stepwise and the importance of a single component is assessed independently, a regression equation relating each Y variable with the X matrix is created. PLS decomposes the matrix X into several latent variables that correlate best with the activity of the molecules. PLS can be done using NIPALS or SIMPLS iterative algorithm, with consecutive estimates obtained using the residuals from previous iterations as the new dependent variable

The regression equation takes the form

$$Y = b0 + b1X1 + b2X2 + ... + bpXp \quad \text{..... Eq. (6)}$$

In this equation b0 is the regression coefficient for the intercept and the bi values are the regression coefficients (for variables 1 through p) computed from the data.

The QSAR model was developed using partial least squares by forward-backward variable selection method with pMIC activity field as dependent variable and physicochemical descriptors as independent variable having cross correlation limit of 20 and number of variables in final equation were 3 each for *S.aureus* and *E.coli*. Selection of test and training set was done by sphere exclusion method.

RESULT AND DISCUSSION

With regard to QSAR modeling, our first goal was to establish a predictive model with a reasonable number of input features to ensure good generalization performance. While correlating various descriptors with biological activity is the most important means to study structure–activity relationships, the interest lies in deciding when to stop adding a new descriptor to the model. Thus, the optimal model should use the minimum number of descriptors to obtain the best fit.

Training set of 34 and 09 of test set of thienopyrimidines having different substitution, were employed. The 2D-QSAR models were developed separately for *S.aureus* and *E.coli*. Following statistical measure was used to correlate biological activity and molecular descriptors; n ,number of molecules; k ,number of descriptors in a model; df ,degree of freedom; r^2 ,coefficient of determination; q^2 , cross validated r^2; pred_r^2 , r^2 for external test set; pred_r^2se , coefficient of correlation of predicted data set; Z score, Z score calculated by the randomization test; best_ran_r^2;best_ran_q^2,highest q^2 value in the randomization test; α, statistical significance parameter obtained by the randomization test.

Generation of QSAR Models

Model– 1 (Multiple Linear Regression (MLR) Analysis)

After 2D QSAR study by Multiple Linear Regression method using forward-backward stepwise variable selection method, the final QSAR equation was developed having 5 variables as follows.

pMIC (*S.aureus*) = + 0.0251(± 0.0010) T_C_C_7 + 0.6640(± 0.0982) T_N_O_3 + 0.1022(± 0.0167) T_2_N_1 - 0.3491(± 0.1583) T_N_O_1 + 0.2680(± 0.1309) T_N_Cl_4 + 2.9396

Model – 1 has a correlation coefficient (r^2) of 0.9849, significant cross validated correlation coefficient (q^2) of 0.8881, F test of 40.4301 and degree of freedom 28. The model is validated by α_ran_r^2 = 0.0000, α_ran_q^2 = 0.0000, best_ran_r^2 = 0.29216, best_ran_q^2 = 0.03433, Z score_ran_r^2 = 7.94639 and Z score_ran_q^2 = 8.71551. The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. The observed and predicted pMIC along with residual values are shown in Table 1**.** Statistical data is shown in Table 2. The plot of observed vs. predicted activity is shown in Figure 1. According to this Model 1, the alignment independent descriptors like T_C_C_7, T_N_O_3, T_2_N_1 and T_N_Cl_4 are directly; whereas T_N_O_1 is indirectly proportional to the antibacterial activity against *S.aureus* and explains the importance of aromaticity in antibacterial activity. The descriptors which contribute for the pharmacological action are shown in Table 3 and Figure 2.

After 2D QSAR study by Multiple Linear Regression method using forward-backward stepwise variable selection method, the final QSAR equation was developed having 3 variables as follows.

pMIC (E.coli) = + 0.0394(± 0.0011) T_C_C_7 + 0.4560(± 0.1252) T_N_O_3 - 0.3499(± 0.1635) T_O_O_7 + 3.6257

Model – 1 has a correlation coefficient (r^2) of 0.8719, significant cross validated correlation coefficient (q^2) of 0.7811, F test of 40.4768 and degree of freedom 30. The model is validated by α_ran_r^2 = 0.0000, α_ran_q^2 = 0.0000, best_ran_r^2 = 0.23848, best_ran_q^2 = 0.03342, Z score_ran_r^2 = 7.65114 and Z score_ran_q^2 = 8.54198. The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. The observed and predicted pMIC along with residual values are shown in Table 1**.** Statistical data is shown in Table 2.The plot of observed vs. predicted activity is shown in Figure 3. Along with the alignment independent descriptors like T_C_C_7 and T_N_O_3 are directly; whereas T_O_O_7 is directly proportional to the antibacterial activity against *E.coli*. The percentage contribution of these descriptors is given in Table 3 and Figure 4.

 ISSN: 2250-1177

Table 2: Statistical parameters of MLR and PLS

Parameters	*S.aureus*		*E.coli*	
	MLR	**PLS**	**MLR**	**PLS**
N	34	34	34	34
Df	28	32	30	31
r^2	0.9849	0.7051	0.8719	0.6710
q^2	0.8881	0.6574	0.7811	0.5883
F test	40.4301	33.5104	40.4768	31.6084
r^2 se	0.7058	0.3350	0.6269	0.3220
q^2 se	0.8682	0.3610	0.7693	0.3602
pred_r^2	0.6342	0.5287	0.6195	0.5158
best_ran_r^2	0.29216	0.21576	0.23848	0.18480
best_ran_q^2	0.03433	0.02140	0.03342	0.01996
Z score_ran_r^2	7.94639	6.45317	7.65114	6.07437
Z score_ran_q^2	8.71551	8.68220	8.54198	8.19551
α_ran_r^2	0.0000	0.0000	0.0000	0.00000
α_ran_q^2	0.0000	0.0000	0.0000	0.00000
Z score_pred_r^2	0.55371	0.48822	0.43250	0.39181
Best_ran_Pred_r^2	0.42210	0.32077	0.23959	0.24636
α_ran_pred_r^2	0.0000	0.0000	0.00000	0.00000

MLR = multiple linear regression, PLS = partial least squares, N = number of molecules of training set, Df= degree of freedom, r^2 = coefficient of determination, q^2 =cross-validated r^2, pred_r^2= r^2 for external test set, Z score = the Z score calculated by q^2 in the randomization test, best_ran_q^2 = the highest q^2 value in the randomization test and a_ran_q^2 =the statistical significance parameter obtained by the randomization test.

.

Table 3: Molecular descriptors contributing in the present study

Descriptor	Description
Alignment independent descriptors	
T_C_C_7	This is the count of number of carbon atoms separated from carbon atoms by seven bond distance
T_N_O_3	This is the count of number of nitrogen atoms separated from oxygen atoms by three bond distance
T_2_N_1	This is the count of number of double bonded atoms separated from nitrogen atoms by a single bond distance
T_N_O_1	This is the count of number of nitrogen atoms separated from oxygen atom by single bond distance
T_O_O_7	This is the count of number of oxygen atoms separated from oxygen atoms by seven bond distance
T_N_Cl_4	This is the count of number of nitrogen atoms separated from chlorine atoms by four bond distance

 ISSN: 2250-1177

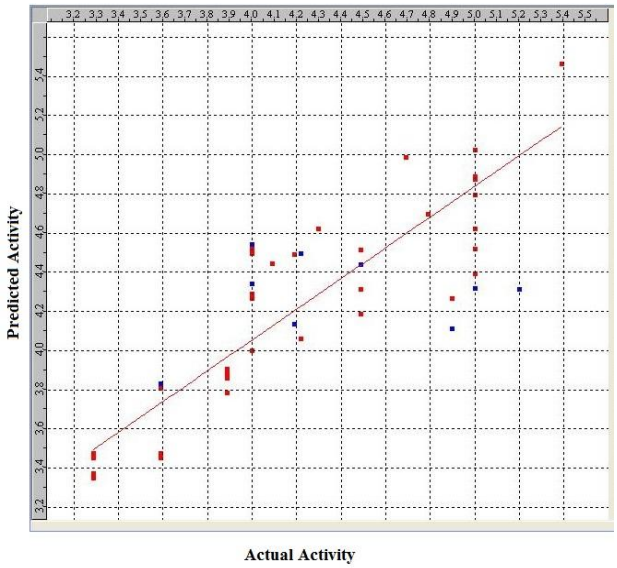


Figure 1: Graph of Actual vs. Predicted activities for training and test set molecules from the Multiple Linear Regression (MLR) model. (A) Training set (Red dots) (B) Test Set (Blue dots) (S.aureus)

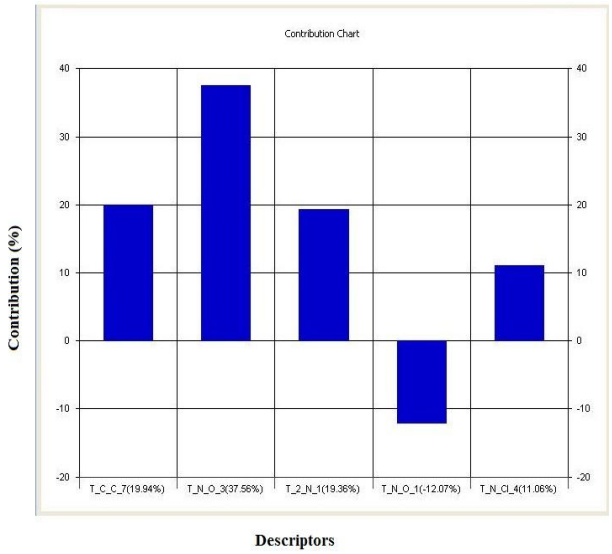


Figure 2: Plot of percentage contribution of each descriptor in developed MLR model explaining variation in the activity (*S.aureus*).

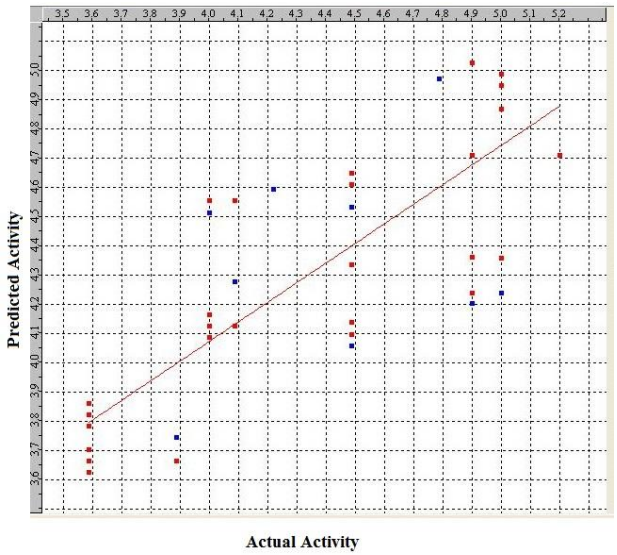


Figure 3: Graph of Actual vs. Predicted activities for training and test set molecules from the Multiple Linear Regression (MLR) model. (A) Training set (Red dots) (B) Test Set (Blue dots) (*E.coli)*

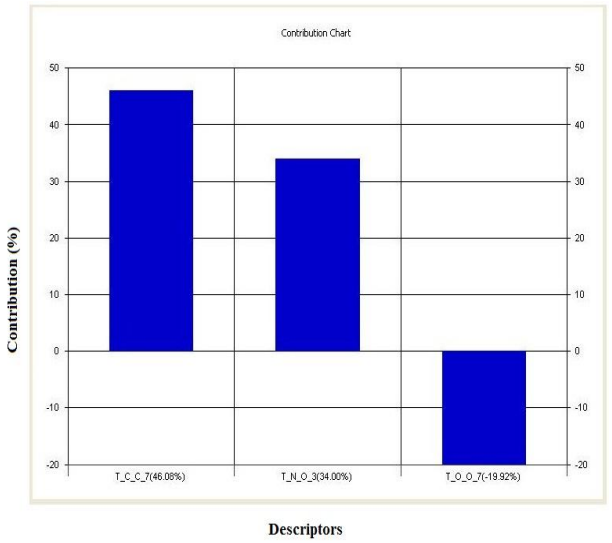


Figure 4: Plot of percentage contribution of each descriptor in developed MLR model explaining variation in the activity (*E.coli)*

Model – 2 (Partial Least Squares (PLS) Analysis)

Model - 2 is having following QSAR equation with 3 variables.

pMIC (*S.aureus*) = + 0.0383 T_C_C_7 + 0.4863 T_N_O_3 + 0.0689 T_2_N_1 + 3.1226

The Model -2 gave correlation coefficient (r^2) of 0.7051, significant cross validated correlation coefficient (q^2) of 0.6574, F test of 33.5104 and degree of freedom 32. The model is validated by $\alpha\_ran\_r^2$ = 0.0000, $\alpha\_ran\_q^2$ = 0.0000, $best\_ran\_r^2$ = 0.21576, $best\_ran\_q^2$ = 0.02140, Z score_ran_r^2 = 6.45317 and Z score_ran_q^2 = 8.68220. The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. The observed and predicted pMIC along with residual values are shown in Table 1. Statistical data is shown in Table 2. The plot of observed vs. predicted activity is shown in Figure 5**.** According to Model 2, the antibacterial activity of thienopyrimidine derivatives against *S.aureus* is mainly attributed to the alignment independent descriptors like T_C_C_7, T_N_O_3 and T_2_N_1; this clearly signifies the importance of aromaticity in antibacterial activity. The descriptors which contribute for the pharmacological action are shown in Table 3 and Figure 6.

Model - 2 is having following QSAR equation with 3 variables.

pMIC (*E.coli*) = + 0.0385 T_C_C_7 + 0.3497 T_N_O_3 + 0.3186 T_N_Cl_4 + 3.5918

The model -2 gave correlation coefficient (r^2) of 0.6710, significant cross validated correlation coefficient (q^2) of 0.5883, F test of 31.6084 and degree of freedom 31. The model is validated by $\alpha\_ran\_r^2$ = 0.0000, $\alpha\_ran\_q^2$ = 0.0000, best_ran_r^2 = 0.18480, best_ran_q^2 = 0.01996, Z score_ran_r^2 = 6.07437 and Z score_ran_q^2 = 8.19551. The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. The observed and predicted pMIC along with residual values are shown in Table 1. Statistical data is shown in Table 2. The plot of observed vs. predicted activity is shown in Figure 7**.** The alignment independent descriptor like T_C_C_7, T_N_O_3 and T_N_Cl_4 are directly proportional to the antibacterial activity. The descriptors which contribute for the pharmacological action are shown in Table 3 and Figure 8.

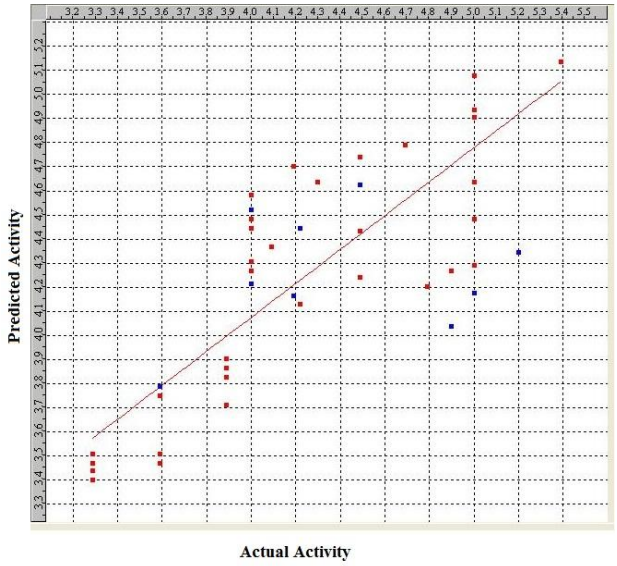


Figure 5: Graph of Actual vs. Predicted activities for training and test set molecules from the Partial Least Squares (PLS) model. (A) Training set (Red dots) (B) Test Set (Blue dots) (*S.aureus*)

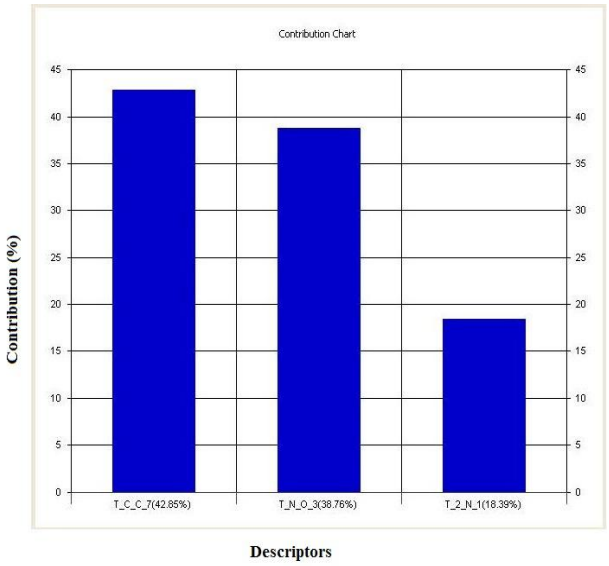


Figure 6: Plot of percentage contribution of each descriptor in developed PLS model explaining variation in the activity (*S.aureus*)

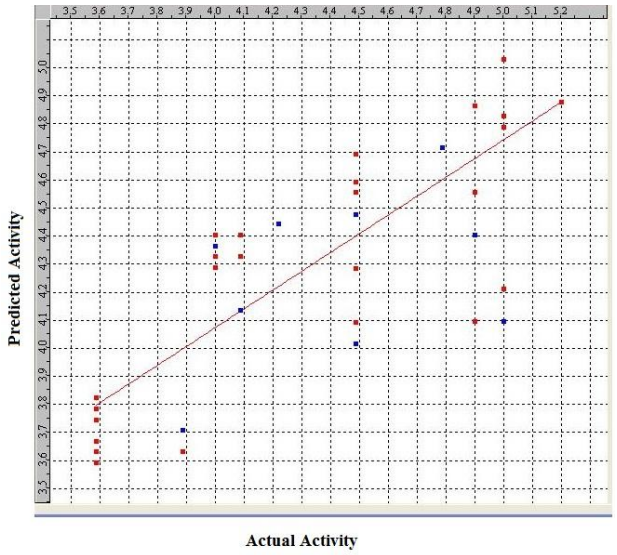


Figure 7: Graph of Actual vs. Predicted activities for training and test set molecules from the Partial Least Squares (PLS) model. (A) Training set (Red dots) (B) Test Set (Blue dots) (*E.coli*)

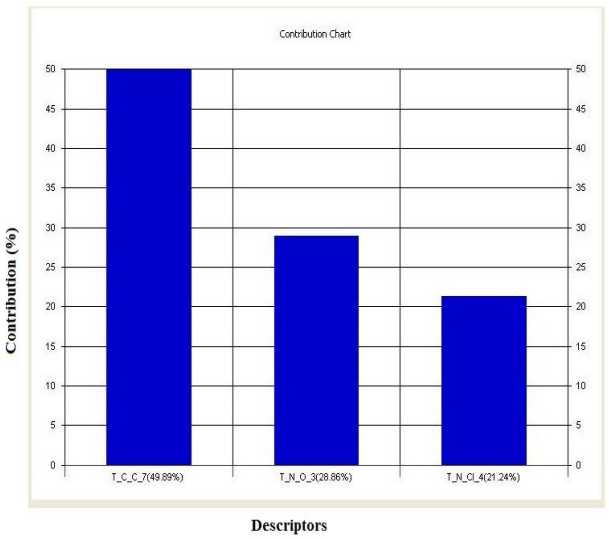


Figure 8: Plot of percentage contribution of each descriptor in developed PLS model explaining variation in the activity (*E.coli*)

 ISSN: 2250-1177

The above study leads to the development of statistically significant QSAR model, which allows understanding of the molecular properties/features that play an important role in governing the variation in the activities. In addition, this QSAR study allowed investigating influence of very simple and easy-to-compute descriptors in determining biological activities, which could shed light on the key factors that may aid in design of novel potent molecules.

CONCLUSION

In conclusion, the 2D-QSAR models developed to predict the structural features of thienopyrimidines to inhibit the growth of *S.aureus* (gram positive) and *E.coli* (gram negative), reveals useful information about the structural features requirement for the molecule. In all the optimized models, MLR method has given very significant results both in *S.aureus* and *E.coli*. This model could be considered as best one in terms of excellent internal and external predictive abilities. According to model I and II, the alignment independent descriptors help in understanding the effect of substituent at different position of thienopyrimidine derivatives. The results obtained from this 2D-QSAR study are in agreement with the observed SAR of thienopyrimidine studied. Hence, the model proposed in this work is useful and can be employed to design new derivatives of thienopyrimidines as prospective antibacterial agents.

ACKNOWLEDGEMENTS

The authors are thankful to the Principal, K.L.E. University's College of Pharmacy, Hubli, for providing necessary facilities to carry out this research work. The authors sincerely acknowledge AICTE, New-Delhi (India), for financial support under RPS Scheme (File No. 8023/BOR/RID/RPS-170/2008-09).

REFERENCES

1. http://www.who.int/mediacentre/factsheets/fs194/en/ dtd: 13.2.2012
2. Cao H, Zhang H, Zheng X, Gao D, 3D QSAR studies on a series of potent and high selective inhibitors for three kinases of RTK family, J Mol Grap Mod, 2007, 26, 236-245.
3. Singh M, Singh SK, Chhabria MT, A novel QSAR model for evaluating and predicting the inhibition activity of H_1-receptor antagonists: a series of thienopyrimidine derivatives, J Drug Del Ther, 2012, 2, 1-10
4. Litvinov VP, Thienopyrimidines: synthesis, properties, and biological activity, Russ Chem Bull, 2004, 53, 487-516
5. Vorobev EV, Kletskii ME, Krasnikov VV, Mezheritskii VV, Steglenko DV, Studies on mechanisms of the rearrangement of thieno[3,2-*e*][1,2,4]triazolo[4,3-c]pyrimidines into thieno[3,2-*e*][1,2,4]triazolo[1,5-c]pyrimidines, Russ Chem Bull, 2006, 55, 2247-2255.
6. Shishoo CJ, Devani MB, Ullas GV, Ananthan S, Bhadthi VS, Studies in the synthesis and interconversion of isomeric triazolothienopyrimidines, J Het Chem, 1981, 18, 43-46.
7. Chambhare RV, Khase BG, Bobde AS, Bahekar RH, Synthesis and preliminary evaluation of some N-[5-(2-furanyl)-2-methyl-4-oxo-4H-thieno[2,3-*d*]pyrimidin-3-yl]-carboxamide and 3-substituted-5-(2-furanyl)-2-methyl-3H-thieno[2,3-*d*]pyrimidin-4-ones as antimicrobial agents, Eur J Med Chem, 2003, 38, 89-100.
8. Shetty NS, Lamani RS, Khazi IM, Synthesis and antimicrobial activity of some novel thienopyrimidines and triazolothienopyrimidines, J Chem Sci, 2009, 121, 301-307.
9. Ashalatha BV, Narayana B, Vijaya Raj KK, Suchetha Kumari N, Synthesis of some new bioactive 3-amino-2-mercapto-5,6,7,8-tetrahydro[1]benzothieno[2,3-d]pyrimidin-4(3H)-one derivatives, Eur J Med Chem, 2007, 42, 719-728.
10. Alagarsamy V, Meena S, Ramseshu KV, Solomon VR, Thirumurugan K, Dhanabal K, Murugan M, Synthesis, analgesic, anti-inflammatory, ulcerogenic index and antibacterial activities of novel 2-methylthio-3-substituted-5,6,7,8-tetrahydrobenzo (b) thieno[2,3-d]pyrimidin-4(3H)-ones, Eur J Med Chem, 2006, 41, 1293-1300.
11. Hozien ZA, Abdel-Wahab AA, Hassan KM, Atta FM, Ahmed SA, Synthesis of some biologically active agents derived from thieno[2,3-d]pyrimidine derivatives, Die Pharmazie, 1997, 52, 753-758.
12. Petrie CR, Cottam HB, Mckernan PA, Robins RK, Revankar GR, Synthesis and biological activity of 6-azacadeguomycin and certain 3,4,6-trisubstituted pyrazolo[3,4-d]pyrimidine ribonucleosides, J Med Chem, 1985, 28, 1010-1016.
13. Nasr MN, Gineinah MM, Pyrido[2, 3-d]pyrimidines and pyrimido[5',4':5, 6]pyrido[2, 3-d]pyrimidines as new antiviral agents: synthesis and biological activity, Arch Pharm, 2002, 335, 289-295.
14. Baraldi PG, Pavani MG, Nunez MC, Brigidi P, Vitali B, Gambari R, Romanoli R, Antimicrobial and antitumor activity of N-heteroimmine-1,2,3-dithiazoles and their transformation in triazolo-, imidazo-, and pyrazolopirimidines, Bioorg Med Chem, 2002, 10, 449-456.
15. Sondhi SM, Johar M, Rajvanshi S, Dastidar SG, Shukla R, Raghubir R, Lown JW, Anticancer, Anti-inflammatory and Analgesic Activity Evaluation of Heterocyclic Compounds Synthesized by the Reaction of 4-Isothiocyanato-4-methylpentan-2-one with Substituted o-Phenylenediamines, o-Diaminopyridine and (Un)Substituted o, Australian J Chem, 2001, 54, 69-74 .
16. Panchamukhi SI, Mulla JS, Shetty NS, Khazi MA, Khan AY, Kalashetti MB, Khazi IM,Benzothieno [3,2-e][1,2,4] triazolo[4,3-c]pyrimidines: Synthesis, Characterization, Antimicrobial Activity, and Incorporation into Solid Lipid Nanoparticles, Arch Pharm Chem Life Sci, 2011,11, 358-365.
17. Narayana B, Ashalatha BV, Vijaya Raj KK, Suchetha Kumari N, Synthesis of 3-amino-2-methyl/ethyl-5,6,7,8-tetrahydro[1]benzothieno[2,3-d]pyrimidin-4[3*H*]-one and its Schiff bases as possible antimicrobial and non-steroidal anti-inflammatory agents, Indian J Chem, 2006, 45B, 2696-2703.
18. Golbraikh A, Tropsha A, Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, J Comput Aided Mol Des, 2002, 16, 357-369.
19. Shen M, LeTiran A, Xiao Y, Golbraikh A, Kohn H, Tropsha A, Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods, J Med Chem, 2002, 45, 2811-2823.
20. Golbraikh A, Tropsha A, QSAR modeling using chirality descriptors derived from molecular topology, J Chem Inf Comput Sci, 2003, 43, 144-154.
21. Cramer RD, Patterson DE, Bunce JD, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J Am Chem Soc, 1988, 110, 5959-5967.

 ISSN: 2250-1177