AMERICAN JOURNAL OF PHARMTECH RESEARCH
Journal home page: http://www.ajptr.com/

2D and 3D Quantity Structure Activity Relationship Studies on 1,4-Benzothiazine Derivatives for Designing Potent Antifungal

Nitin P. Jain[1*], Chandrashekhar D. Upasani[2], Girish A. Kashid[1], Usha N. Jain[1]
1. Department of Pharmaceutical Chemistry(PG), S N D College of Pharmacy, Babhulgaon, Yeola, Nashik, Maharashtra, India-423401
2. Department of Pharmacology, SNJB's S S D J College of Pharmacy, Chandwad, Nashik. Maharashtra, India-423101.

ABSTRACT

1,4-Benzothiazine derivatives having variable antifungal activity against four species of fungus such as *Candida albicans, Trichophyton rubrum, Epidermophyton floccosum* and *Malassazia furfur* were selected to develop 2D and 3DQSAR models. The best 2DQSAR model was selected, having correlation coefficient r^2 (0.8747) and cross validated squared correlation coefficient q^2 (0.7145) with external predictive ability of pred_r^2 (0.9658). 2DQSAR parameters are Quadrople1, SssCH2count, XA Average hydrophobicity, X Amost hydrophilic area and Polar surface area excluding P and S contributed in the model. The best 3DQSAR model having the correlation coefficient r^2= 0.8635 was selected for further study. The model was further validated by means of crossed squared correlation coefficient q^2 =0.7614 and pred_r^2 =0.6654, which show that the model has good predictive ability and was developed by Forward SW-MLR. From QSAR model it concluded that the bulky substitution is require at para and meta position of Phenyl ring and less bulky substitution at ortho position of phenyl ring for antifungal activity.

Keyword: 2DQSAR, 3DQSAR, Antifungal, 1,4-Benzothiazine, Vlife MDS and MLR

*Corresponding Author Email: nitinpjain06@yahoo.co.in
Received 01 August 2014, Accepted 07 August 2014

Please cite this article as: Jain N *et al.*, 2D and 3D Quantity Structure Activity Relationship Studies on 1,4-Benzothiazine Derivatives for Designing Potent Antifungal. American Journal of PharmTech Research 2014.

INTRODUCTION

During the past two decades, the frequency of invasive and systemic fungal infections has increased dramatically in the population with altered immunity,[1,2] fungal infection has become an important complication and a major cause of morbidity and mortality in immunocompromised individuals such as patients undergoing anticancer chemotherapy or organ transplants and patients with AIDS.[3] Current available therapy in treating fungal infections can suffer from drug related toxicity, hazardous drug– drug interactions, non-optimal pharmacokinetics and development of drug resistance.[4] Fungal infections remain a one major cause of morbidity and mortality, specially in immunocompromised host where the incidence of life threatening fungal infections has risen dramatically.[5] Clinically, candidosis, aspergillosis and cryptococcosis are the three major fungal infections in immunocompromised individuals.[6] The 1,4-Benzothiazine compounds reported having good antifungal activity against the *Candida albicans, Trichophyton rubrum, Epidermophyton floccosum* and *Malassazia furfur.*[7] In our continual interest in rational antifungal drug design, we have constructed two & three-dimensional (3D and 2D) models of QSAR from antifungal activity against *Candida albicans, Trichophyton rubrum, Epidermophyton floccosum* and *Malassazia furfur.*

This study is aimed to elucidate the structural features of benzothiazine derivatives required for antifungal activity and to obtain predictive 2D and 3DQSAR models to guide the rational synthesis of novel antifungal drug.

MATERIALS AND METHOD

All molecular modeling studies (2D and 3DQSAR) were performed using the Molecular Design Suite on a DELL PC with a Pentium IV processor and a Windows XP operating system. Structures were sketched using the 2D draw application and converted to 3D structures[8].

Biological activity dataset for QSAR analysis

The antifungal activity [MIC(µmol)] data against four fungal species (*Candida albicans, Trichophyton rubrum, Epidermophyton floccosum* and *Malassazia furfur*) of substituted 1,4-Benzothiazine derivatives were taken from the reported work,[9] the results presented in table 1. The total 37 set of compounds was divided into a training set (29 compounds) for generating 2D and 3DQSAR models and a test set (8 compounds) for validating the quality of the models. Selection of the training set and test set molecules was done on the basis of structural diversity and a wide range of activity such that the test-set molecules represent a range of biological activity similar to that of the training set; thus, the test set is truly representative of the training set. The structure of

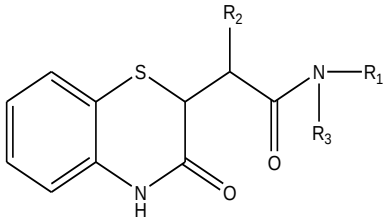
37 molecules given in table 2.The biological activity values [MIC] reported in micromolar units(µM) were converted to their microgram units(µgm) and then further to negative logarithmic scale and subsequently used as the dependent variable for the QSAR analysis.

Table 1. Antifungal activity data (MIC) of Benzothiazine derivatives against four fungal species reported in Ph D thesis.

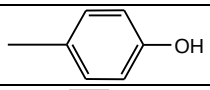
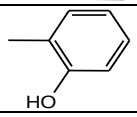
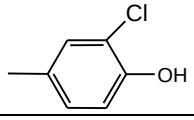
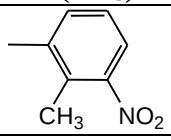
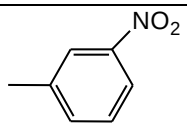
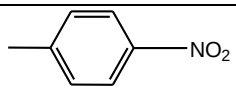
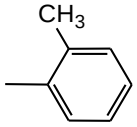
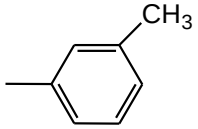
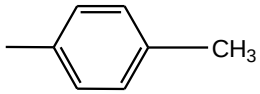
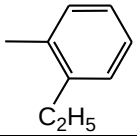
Sr.No.	Compound Code	*C. albicans*	*E. floccosum*	*T. rubrum*	*M. furfur*
1	BTA-4	0.250	0.125	0.250	0.125
2	BTA-8	0.125	0.125	0.125	0.250
3	BTA-5	0.125	0.125	0.125	0.125
4	BTA-47	0.0625	0.0625	0.0625	0.125
5	BTA-20	0.0625	0.0625	0.125	0.125
6	BTA-43	0.0312	0.0625	0.125	0.125
7	BTA-64	0.250	0.250	0.250	0.250
8	BTA-35	0.125	0.125	0.125	0.0625
9	BTA-25	0.125	0.250	0.0625	0.125
10	BTA-66	0.250	0.250	0.125	0.250
11	BTA-69	0.125	0.125	0.125	0.250
12	BTA-63	0.250	0.125	0.250	0.250
13	BTA-57	0.250	0.250	0.125	0.250
14	BTA-58	0.250	0.250	0.125	0.125
15	BTA-56	0.0625	0.0625	0.0312	0.0312
16	BTA-65	0.250	0.250	0.250	0.250
17	BTA-59	0.125	0.125	0.250	0.250
18	BTA-60	0.250	0.250	0.250	0.125
19	BTA-61	0.250	0.0625	0.125	0.250
20	BTA-67	0.250	0.250	0.125	0.125
21	BTA-68	0.125	0.250	0.125	0.125
22	BTA-24	0.0625	0.0625	0.0625	0.0625
23	BTA-70	0.125	0.125	0.250	0.250
24	BTMA-11	0.250	0.250	0.250	0.250
25	BTMA-12	0.125	0.125	0.250	0.125
26	BTMA-24	0.125	0.0625	0.0312	0.0312
27	BTMA-33	0.250	0.250	0.250	0.250
28	BTMA-36	0.0625	0.0625	0.250	0.125
29	BTMA-48	0.250	0.250	0.250	0.250
30	BTMA-50	0.0625	0.0625	0.125	0.125
31	BTMA-59	0.250	0.250	0.250	0.250
32	BTMA-60	0.250	0.250	0.250	0.250
33	BTMA-74	0.250	0.250	0.250	0.250
34	BTMA-77	0.125	0.250	0.250	0.250
35	BTMA-80	0.250	0.250	0.250	0.125
36	BTMA-83	0.250	0.250	0.250	0.250
37	BTMA-88	0.250	0.250	0.0625	0.0625
38	Ketaconazole	0.0312	0.0312	0.0312	0.0312

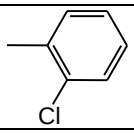
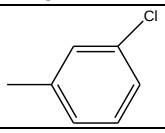
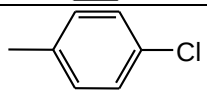
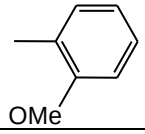
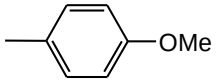
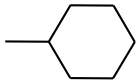
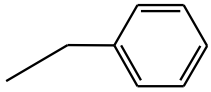
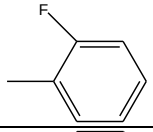
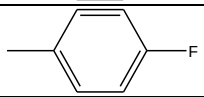
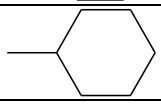
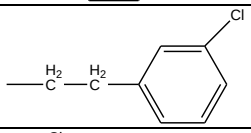
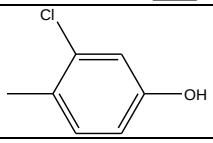
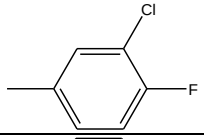
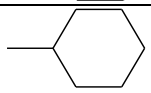
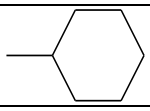
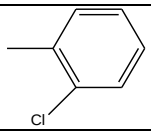
#All MIC value in µmol/ml

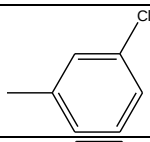
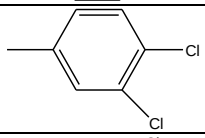
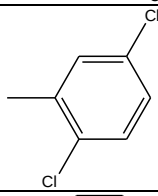
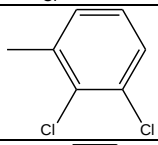
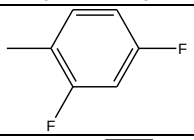
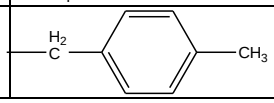
Table 2. Structures of dataset used for MLR QSAR analysis.



General structure of Benzothiazine compounds

compounds	R_1	R_2	R_3
BTA-4	$-CH_3$	H	H
BTA-8	$-C(CH_3)_3$	H	H
BTA-5	$-CH_2CH_3$	H	H
BTA-47	$-CH_2CH_2CH_2CH_3$	H	H
BTA-20	$-CH_2CH_2OH$	H	H
BTA-43		H	H
BTA-64		H	H
BTA-35		H	H
BTA-25	$-CH(CH_3)CH_2OH$	H	H
BTA-66		H	H
BTA-69		H	H
BTA-63		H	H
BTA-57		H	H
BTA-58		H	H
BTA-56		H	H
BTA-65		H	H

BTA-59		H	H
BTA-60		H	H
BTA-61		H	H
BTA-67		H	H
BTA-68		H	H
BTA-24		H	H
BTA-70		H	H
BTMA-11		CH_3	H
BTMA-12		CH_3	H
BTMA-24		CH_3	H
BTMA-33		CH_3	H
BTMA-36		CH_3	H
BTMA-48		CH_3	H
BTMA-50		CH_3	
BTMA-59		CH_3	H

BTMA-60		CH_3	H
BTMA-74		CH_3	H
BTMA-77		CH_3	H
BTMA-80		CH_3	H
BTMA-83		CH_3	H
BTMA-88		CH_3	H

Computational details

The structures of all 37 compounds were drawn in 2DDrawApp (MDS 3.5 2010). The 2D structures were converted to 3D structures by sending them to MDS. Each compound was energy minimized and batch optimized by using Merck Molecular Force Field (MMFF) and charges.[10]

MOLECULAR MODELING FOR 2D-QSAR

Calculation of 2D Molecular descriptors

2D descriptors (physicochemical and alignment independent) were calculated for the optimized compounds on QSAR plus work sheet given in table 3. The invariable descriptors (the descriptors that are constant for all the molecules) were removed, as they do not contribute to QSAR. Most stable structure for each compound was generated after energy minimization and used for calculating various physico-chemical descriptors like thermodynamic, steric and electronic. The energy-minimized geometry was used for the calculation of the various 2D descriptors (Individual, Chi, ChiV, Path count, Chi Chain, ChiV Chain, Chain path count, Cluster, Path cluster, Kappa, Element Count, Estate number, Estate contribution, Semi-empirical, Hydrophillic–hydrophobic, Polar surface area and Alignment independent) and was considered as independent variables in the present study.

Table 3. List of descriptors to be used in the most significant 2DQSAR models of Benzothiazine derivatives.

compounds	Average -ve potential	T_2_N_4	T_N_O_3	Polar surface area excluding P and S
BTA-4	-0.026674	3	1	78.43
BTA-8	-0.023032	3	0	58.2
BTA-5	-0.039886	3	1	78.43
BTA-47	-0.025048	4	0	90.46
BTA-20	-0.030644	4	0	90.46
BTA-43	-0.020051	3	0	70.23
BTA-64	-0.03927	3	0	58.2
BTA-35	-0.034731	4	0	58.2
BTA-25	-0.031684	4	0	58.2
BTA-66	-0.030379	4	0	58.2
BTA-69	-0.034371	4	0	58.2
BTA-63	-0.035662	3	0	58.2
BTA-57	-0.027043	4	0	58.2
BTA-58	-0.027197	4	0	58.2
BTA-56	-0.031298	5	0	104.02
BTA-65	-0.042909	4	1	90.46
BTA-59	-0.027412	4	0	58.2
BTA-60	-0.035032	4	1	78.43
BTA-61	-0.04089	4	1	78.43
BTA-67	-0.035853	4	0	67.43
BTA-68	-0.036566	6	0	104.02
BTA-24	-0.032069	5	0	58.2
BTA-70	-0.025151	3	0	58.2
BTMA-11	-0.045988	4	0	58.2
BTMA-12	-0.030422	4	0	58.2
BTMA-24	-0.027504	3	0	58.2
BTMA-33	-0.028509	5	0	75.27
BTMA-36	-0.036055	4	0	78.43
BTMA-48	-0.035548	4	0	58.2
BTMA-50	-0.022405	3	0	49.41
BTMA-59	-0.032697	4	0	58.2
BTMA-60	-0.036003	4	0	58.2
BTMA-74	-0.027679	4	0	58.2
BTMA-77	-0.035112	4	0	58.2
BTMA-80	-0.027635	4	0	58.2
BTMA-83	-0.036541	4	0	58.2
BTMA-88	-0.026871	5	0	75.27

Table 3. (Continued)

compounds	Quadrople1	SssCH2 count	XA Average hydrophobicity	XA most hydrophilic area
BTA-4	-8.729997	3	0.101808	-0.38321
BTA-8	10.980475	6	0.103928	-0.339453

BTA-5	20.182372	2	0.085463	-0.336955
BTA-47	7.238034	1	0.097466	-0.334198
BTA-20	13.940225	1	0.086118	-0.333989
BTA-43	13.59071	4	0.100732	-0.349021
BTA-64	20.818424	1	0.088521	-0.364035
BTA-35	14.167039	1	0.10151	-0.332373
BTA-25	13.536532	1	0.103813	-0.338032
BTA-66	20.820471	1	0.106272	-0.338528
BTA-69	18.786904	1	0.106009	-0.336452
BTA-63	-0.993766	2	0.090317	-0.365866
BTA-57	14.913566	1	0.109389	-0.334626
BTA-58	7.187541	1	0.113167	-0.338511
BTA-56	-11.651467	1	0.090158	-0.334959
BTA-65	34.220524	1	0.095852	-0.353091
BTA-59	15.093787	2	0.106281	-0.339453
BTA-60	28.069249	1	0.090936	-0.332505
BTA-61	-0.871857	1	0.090583	-0.335021
BTA-67	25.268817	1	0.090515	-0.342981
BTA-68	28.937069	1	0.091519	-0.346353
BTA-24	23.94469	2	0.104131	-0.487664
BTA-70	13.696868	1	0.091005	-0.328196
BTMA-11	28.209626	0	0.106578	-0.329532
BTMA-12	-12.20455	0	0.10309	-0.335529
BTMA-24	15.388179	5	0.113231	-0.336017
BTMA-33	20.463742	2	0.121812	-0.33471
BTMA-36	-27.010808	0	0.098876	-0.326067
BTMA-48	-11.822584	0	0.113317	-0.334272
BTMA-50	15.489776	10	0.128473	-0.329571
BTMA-59	17.212399	0	0.129612	-0.396205
BTMA-60	14.14887	0	0.119364	-0.339349
BTMA-74	10.367645	0	0.126172	-0.338166
BTMA-77	26.791118	0	0.133635	-0.485718
BTMA-80	-23.810701	0	0.132143	-0.345419
BTMA-83	17.224041	0	0.120847	-0.416222
BTMA-88	18.121406	1	0.114855	-0.341004

Training and test set selection

Random selection method was adopted for division of training and test set. This is a rational selection method which takes into consideration both biological and chemical space for division of dataset. A commonly used ratio of training to validation objects (test set), which was also adopted in this work, is 80:20%. This approach resulted in selection of 8 compounds as test set for validating the quality of the models and the remaining 29 compounds as the training set for generating QSAR models (table 6).

Table 6. Data set for 2DQSAR and 3DQSAR model.

compounds	2D QSAR set	3DQSAR set
BTA-4	Training	Training
BTA-8	Training	Test
BTA-5	Training	Training
BTA-47	Test	Test
BTA-20	Training	Training
BTA-43	Training	Training
BTA-64	Training	Training
BTA-35	Test	Training
BTA-25	Training	Training
BTA-66	Training	Training
BTA-69	Training	Training
BTA-63	Training	Training
BTA-57	Training	Test
BTA-58	Training	Training
BTA-56	Training	Training
BTA-65	Test	Training
BTA-59	Training	Training
BTA-60	Training	Training
BTA-61	Training	Test
BTA-67	Training	Test
BTA-68	Training	Training
BTA-24	Training	Training
BTA-70	Test	Training
BTMA-11	Training	Training
BTMA-12	Training	Test
BTMA-24	Training	Training
BTMA-33	Test	Training
BTMA-36	Training	Test
BTMA-48	Test	Training
BTMA-50	Training	Training
BTMA-59	Training	Training
BTMA-60	Test	Training
BTMA-74	Training	Training
BTMA-77	Training	Training
BTMA-80	Training	Training
BTMA-83	Training	Training
BTMA-88	Test	Test

Variable selection and model development

Variable selection is a key step in QSAR analysis. The reduced set of descriptors was then treated by forward stepwise variable selection for further reduction of non-significant descriptors and finally the optimum models with four significant descriptors were considered in our 2D-QSAR analysis. The variable/feature selection method can be used together with MLR regression analysis for constructing a 2D-QSAR model. MLR is the traditional and standard approach for multivariate

data analysis. Multivariate analysis is the analysis of multidimensional data metrics by using statistical methods. It relates the dependent variable (biological activity) to a number of independent (predictor) variables (molecular descriptors) by using linear equations. This method of regression estimates the values of the regression coefficients by applying least square curve fitting method.

Statistical parameters for 2D-QSAR models

The program computes the best model on the basis of coefficient of determination/squared correlation coefficient r^2; cross-validated correlation coefficient/cross-validated explained variance q^2, which is a relative measure of quality of fit; F-test (Fischer's value) which represents F-ratio between the variances of calculated and observed activity; pred_r^2, r^2 for external test set; DF, degree of freedom. Zscore, calculated by q^2 in the randomization test; best_rand_q^2, the highest q^2 value in the randomization test and alpha_rand_q^2 obtained by randomization test. The coefficient of determination/ squared correlation coefficient r^2 is a relative measure of fit by the regression equation. The coefficient of determination is the percent of the variation that can be explained by the regression equation. It represents the explained variance of the model and is used as a measure of the goodness-of-fit of the model. 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. Cross-validated correlation coefficient/cross-validated explained variance q^2 is used as a measure of the internal performance and sometime used to estimate predictivity. Validation parameter, pred_r^2 was calculated for evaluating the predictive capacity of the model. A value of pred_r^2 is greater than 0.5 indicates that good predictive capacity of the QSAR model. However, a QSAR model is considered to be predictive, if the following conditions are satisfied: r^2 >0.6, q^2 >0.6 and pred_r^2 > 0.5[11]. The low standard error of pred_r^2se, q^2se and r^2se shows absolute quality of fitness of the model. The generated QSAR model was validated for predictive ability inside the model by using cross validation[12] (LOO) for q^2 and external validation which is a more robust alternative method by dividing the data into training set & test set and calculating pred_r^2. The high pred_r^2 and low pred_r^2se show high predictive ability of the model.

MOLECULAR MODELING FOR 3D-QSAR

Molecular alignment

Proper alignment of molecules is the most critical input and a critical step in the ligand based 3D-QSAR modeling method to obtain good results.[13,14] Energy-minimized and geometry optimized

structures of molecules were aligned by the template-based method[15], where a template structure is defined and used as a basis for alignment of a set of molecules. The template structure, i.e. 1,4-Benzothiazine ring, was used for the alignment by considering the common elements of the series as shown in Figure 1. The reference molecule is chosen in such a way that it is the most active among the series of molecules considered. The reference molecule is the molecule on which the other molecules of the align dataset get aligned based on the chosen template[9]. Compound BTA-56 possessed very high inhibitory activities against all fungus species so it is a valid lead molecule and therefore, was chosen as a reference molecule. After optimizing, the template structure and the reference molecule were used to superimpose all molecules from the series using the template alignment method in VLife MDS 3.5 software[9] to obtain optimal alignment between the molecular structures necessary for ligand–receptor interactions. To adjusts the geometry of the molecules such that their steric and electrostatic fields match the fields of the template molecule[16]. The superimposition of all molecules based on minimizing RMS deviation is shown in Figure 2.

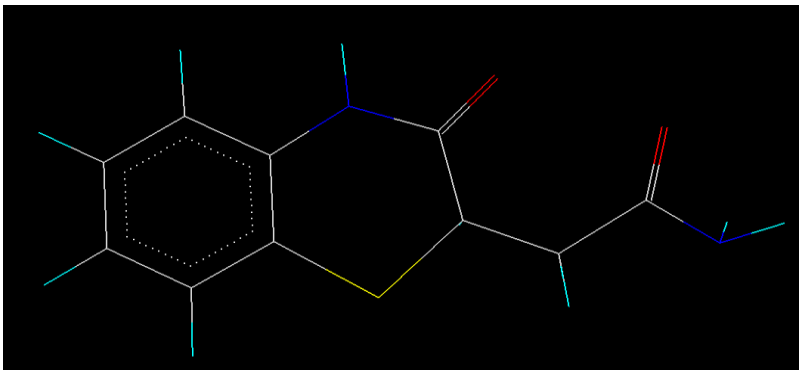


Figure 1: 1, 4-Benzothiazine ring as template used for alignment of Benzothiazine derivatives.

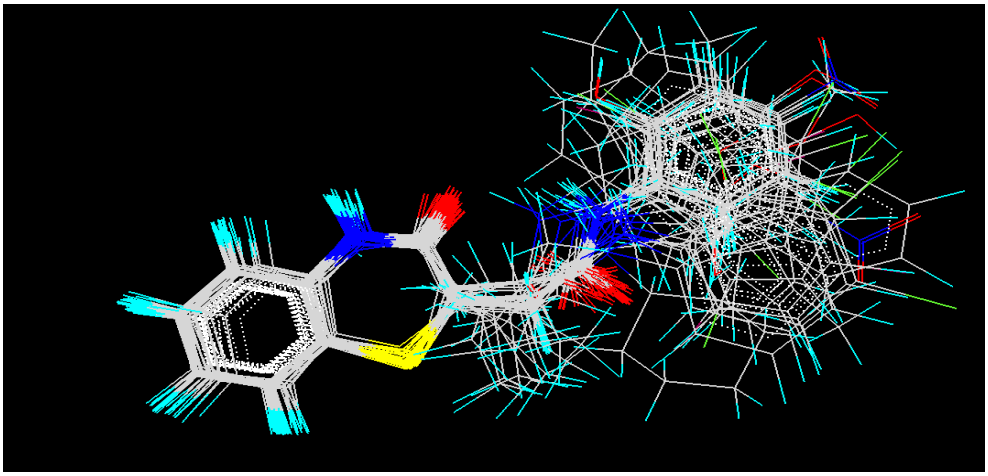


Figure 2: 3D view of template based alignment of benzothiazine derivatives on the base template.

Molecular descriptors

MFA is a method for quantifying the interaction energy between a probe molecule and a set of aligned target molecules in a rectangular grid box and can be useful in establishing QSAR[17]. The aligned biologically active conformations of benzothiazine were used for the calculation of molecular fields. Molecular fields are the electrostatic, steric and hydrophobic interaction energies which are used to formulate a relationship among electrostatic, steric and hydrophobic properties together with the biological activities of compounds. Descriptors were calculated using a sp3 carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 with default cut-off energy 30 kcal/mol to generate steric field, electrostatic and hydrophobic fields. Molecular descriptors such as steric, electrostatic and hydrophobic fields have been calculated utilizing VLife MDS 3.5 software which allows the user to choose probe, grid size, and grid interval for the generation of descriptors.[18] This is done by generating 3D rectangular grids around the molecule and calculating the interaction energy between the molecule and probe group placed at each grid point. Using Tripos force field steric, electrostatic and hydrophobic fields were computed at each grid point considering Gasteiger-Marsili charges.[19] A value of 1.0 was assigned to the distance-dependent dielectric constant. A total of 3568 three dimensional descriptors were calculated, these included electrostatic, steric and hydrophobic field descriptors for all the compounds in separate columns. The value electrostatic, steric and hydrophobic field descriptors those contribute for the development of QSAR model is given in table 4 for 37 compounds.

Table 4. List of descriptors to be used in the most significant 3DQSAR models of Benzothiazine derivatives.

compounds	S_1183	E_1260	E-862	E_303	S_1282	S-645	S-599	S-1181
BTA-4	0.580001	0.580001	-0.005852	1	-0.07138	-0.42075	0.564542	-0.0255
BTA-8	0.557059	0.557059	-0.009034	1	-0.06991	-0.46951	1	0.02843
BTA-5	0.64944	0.64944	-0.063532	1	-0.02385	-0.44642	1	0.01368
BTA-47	0.873942	0.873942	1	1	1	-0.39684	1	0.13773
BTA-20	0.762806	0.762806	0.202939	1	0.943209	-0.33717	1	0.11647
BTA-43	0.989148	0.989148	0.347948	1	-0.27101	-0.37526	1	1
BTA-64	0.595234	0.595234	-1	1	-0.01382	-0.32547	1	0.01482
BTA-35	0.628927	0.628927	1	1	-0.05521	-0.37334	1	0.05731
BTA-25	0.611516	0.611516	0.470856	0.466765	-0.01839	-0.30956	1	0.10647
BTA-66	1.139961	1.139961	1	1	-0.33138	-0.36001	1	0.16363
BTA-69	0.583955	0.583955	1	0.094056	-0.0305	-0.48585	1	0.26296
BTA-63	0.51078	0.51078	-1	1	-0.02404	-0.29915	0.198069	0.01959
BTA-57	0.611383	0.611383	0.648202	0.554134	-0.17779	-0.43877	1	0.28776
BTA-58	0.484497	0.484497	0.445752	1	-0.22432	-0.26391	1	0.04512
BTA-56	0.531015	0.531015	1	0.590331	0.24956	-0.39274	1	0.33708
BTA-65	0.665322	0.665322	-1	1	-0.06313	-0.30427	1	0.05997

BTA-59	0.555864	0.555864	1	0.411531	-0.12605	-0.46276	1	-0.2878
BTA-60	0.573486	0.573486	-0.972175	0.85332	-0.18001	-0.3792	1	0.14568
BTA-61	0.592811	0.592811	1	0.002429	-0.01368	-0.52002	1	0.09447
BTA-67	0.542406	0.542406	1	0.163258	-0.23031	-0.30972	1	1
BTA-68	0.60239	0.60239	-0.271593	1	-0.18664	-0.35569	1	-0.0721
BTA-24	0.677583	0.677583	-1	1	-0.14337	-0.40317	-0.6571	0.02362
BTA-70	0.609883	0.609883	-0.251746	1	-0.02559	-0.39073	1	0.01331
BTMA-11	0.519141	0.519141	-1	0.044701	-0.1366	-0.40996	1	0.09289
BTMA-12	0.491647	0.491647	1	0.709869	-0.22705	-0.3914	1	-0.1406
BTMA-24	0.643958	0.643958	0.94504	1	-0.07527	-0.52935	1	0.06119
BTMA-33	0.587606	0.587606	-0.390461	1	-0.25716	-0.39735	1	0.03848
BTMA-36	0.433419	0.433419	-1	1	-0.07038	-0.37585	1	-0.0315
BTMA-48	0.446881	0.446881	1	0.338075	-0.12117	-0.16861	1	0.13232
BTMA-50	0.645352	0.645352	0.076746	0.960193	-0.01808	-0.48374	1	0.01172
BTMA-59	0.542049	0.542049	-1	1	-0.02038	-0.23009	1	0.01315
BTMA-60	0.527248	0.527248	-1	1	-0.01487	-0.4392	1	0.01424
BTMA-74	0.524367	0.524367	-0.516255	1	-0.05588	-0.20538	1	0.02058
BTMA-77	0.553561	0.553561	-0.604009	1	-0.12532	-0.35548	1	0.03214
BTMA-80	0.519407	0.519407	-1	1	-0.01379	-0.13492	1	-0.0098
BTMA-83	0.549184	0.549184	1	1	-0.11246	-0.47036	1	0.04878
BTMA-88	0.606991	0.606991	-0.802098	1	-0.01537	-0.31258	1	0.00875

Table 4. (*continued*).

S_1183	**S_535**	**E_624**	**E_963**	**S_864**	**S_535**	**E_636**	**E_439**	**S_424**
0.580001	-0.39961	-1	-0.00585	0.053404	0.787744	-0.82388	0.012042	1
0.557059	-0.38582	-1	-0.00903	-0.54568	0.84625	-1	-0.08062	1
0.64944	-0.45055	-1	-0.06353	-0.48413	0.379355	-0.63526	0.015137	1
0.873942	0.388055	0.426943	1	0.019873	1	0.213263	0.315823	1
0.762806	-0.18448	0.466599	0.202939	-0.34558	1	0.25406	0.285551	1
0.989148	-0.39507	0.211897	0.347948	-0.49983	0.5314	-0.01225	0.116229	1
0.595234	-0.36553	-1	-1	-1	0.676218	-1	-0.10531	1
0.628927	0.287436	-0.48741	1	-0.21496	1	0.109672	0.319897	1
0.611516	0.790722	-1	0.470856	-0.75872	1	0.494288	0.427669	0.616868
1.139961	-0.384	-0.13946	1	-0.86398	1	-0.33623	0.110788	1
0.583955	-0.44145	-0.21931	1	-1	0.535303	1	0.38762	-0.05237
0.51078	-0.28799	1	-1	-1	0.609459	0.855825	0.221703	1
0.611383	-0.20524	1	0.648202	-0.31441	1	1	0.49099	0.558084
0.484497	-0.31407	-0.13433	0.445752	-0.25873	1	-0.20267	0.142683	1
0.531015	-0.16293	-0.07617	1	0.162814	1	0.964687	0.527287	0.712211
0.665322	-0.38002	-1	-1	-1	0.521308	-1	-0.05749	1
0.555864	-0.31954	1	1	-0.9082	1	1	0.466438	0.368356
0.573486	-0.1019	1	-0.97218	-0.95839	1	0.503466	0.335157	0.673514
0.592811	-0.44123	-1	1	1	0.035708	1	0.351953	-0.1929
0.542406	-0.22979	0.349882	1	-0.25622	1	1	0.473092	0.350852
0.60239	0.079675	1	-0.27159	-0.57881	1	0.728942	0.402637	1
0.677583	0.73522	-1	-1	-1	1	-0.47409	0.175536	0.806824
0.609883	-0.37005	-1	-0.25175	-0.57062	1	-0.42762	0.104952	1
0.519141	0.199039	0.302293	-1	-1	1	0.461275	0.275462	0.138835

0.491647	0.183815	1	1	-0.63578	1	0.906081	0.485586	0.944606
0.643958	0.508387	-1	0.94504	-0.60539	1	-1	0.130387	1
0.587606	0.042721	1	-0.39046	-0.27776	1	0.982479	0.491693	1
0.433419	0.306491	0.666396	-1	-1	1	0.40861	0.314135	0.870606
0.446881	-0.47745	1	1	-0.91925	1	1	0.454112	0.335696
0.645352	-0.31716	-1	0.076746	-0.04013	-1	-1	-0.28742	0.826947
0.542049	-0.22262	1	-1	-1	1	0.635574	0.29341	1
0.527248	-0.24686	1	-1	-0.6824	1	0.624389	0.331996	1
0.524367	-0.20897	0.963725	-0.51626	-1	-0.376	0.207937	0.062129	0.898303
0.553561	-0.33059	1	-0.60401	-1	1	0.388657	0.247939	1
0.519407	-0.17068	1	-1	-0.9939	1	0.472022	0.264454	1
0.549184	-0.49262	0.552367	1	-1	0.256272	0.02105	0.062226	1
0.606991	-0.3246	-1	-0.8021	-0.86535	0.861901	-1	-0.07406	1

Division of a dataset into training and test sets

The random selection method was adopted for division of training and test data sets in order to assess the similarity of the distribution pattern of the compounds in the generated sets, statistical parameters (with respect to the biological activity) i.e. mean, maximum, minimum and standard deviation were calculated for the training and test sets. For selection of training and test sets, we were ensured that the compounds have uniform spread (training and test) in terms of both activity and chemical space. Random selection method resulted in the selection of 8 compounds as the test set for validating the quality of the models and the remaining 29 compounds as the training set for generating 3D-QSAR models (table 6). The test was used to ascertain the predictive power of the model.

Forward stepwise as feature (variable) selection method

Chance correlations and multi-colinearity are two major problems often encountered when attempting to find generalized QSAR models for use in drug design. Feature selection is a key step in QSAR analysis. An integral aspect of any model building exercise is the selection of an appropriate set of features with low complexity and good predictive accuracy. This process forms the basis of a technique known as feature selection or variable selection. Among several search algorithms, stepwise (SW), genetic algorithm (GA) and simulated annealing (SA) based feature selection procedures are most popular for building QSAR models and can explain the situation more effectively.[20-22] In SW forward variable selection algorithm, the search procedure begins with developing a trial model step by step with a single independent variable and to each step, independent variables are added one at a time, examining the fit of the model by using the MLR procedure. Thus the model is repeatedly altered from the previous one by adding or removing a predictor variable in accordance with the 'stepping criteria' (in this case, F= 4 for inclusion for the forward selection method). The method continues until there is no more significant variable

remaining outside the model. In the selected equations, the cross-correlation limit was set at 0.5, the number of variables at 4 and the term selection criteria are r^2. An F value was specified to evaluate the significance of a variable. The variance cut-off was set at 0.0 and scaling as none.

MLR METHODOLOGY FOR BUILDING QSAR MODELS

Model validation and evaluation

This is done to test the internal stability and predictive ability of the QSAR models.

Internal and external validations

Internal validation was carried out using leave-one-out (q^2, LOO) method[11]. For calculating q^2, each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted using the model developed by the remaining molecules. The cross-validated coefficient q^2 was calculated using Eq. (1).

$$q^2 = 1- \Sigma(y_i-\hat{y}_i)^2 / \Sigma(y_i-y_{mean})^2 \quad (1)$$

Where y_i and $\hat{y}_i$ are the actual and the 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. However, a high q^2 value does not necessarily give a suitable representation of the real predictive power of the model. So an external validation is also carried out in this study. The external predictive power of the model is assessed by predicting pMIC value of eight test set molecules, which are not included in the QSAR model development. The predictive ability of the selected model is also confirmed by pred_r^2. For external validation, the activity of each molecule in the test set was predicted using the model developed by the training set. The pred_r^2 value is calculated as follows Eq. (2)

$$Pred\_r^2 = 1 - \Sigma(y_i-\hat{y}_i)^2/ \Sigma(y_i-y_{mean})^2 \quad (2)$$

Where y_i and $\hat{y}_i$ are the actual and the predicted activity of the i^{th} molecule in the test set respectively and y_{mean} is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. Thus the pred_r^2 value is indicative of the predictive power of the current MLR model based on the external test set.

Randomization test

To evaluate the statistical significance of a QSAR model for a real data set, one tail hypothesis testing was used[23]. The robustness of the models for training sets was examined by comparing these models to those derived from random data sets. 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 using the calculated Zscore; Eq. (3). A Zscore value is calculated by the equation 3.

Zscore = h - μ / Б (3)

where h is the q^2 value calculated for the actual data set, π is the average q^2, and r is its standard deviation. The probability (α) of significance of the randomization test is derived by using calculated Zscore value as given in the literature[24,25].

Evaluation of the quantitative model.

The developed 3D-QSAR model was evaluated using the following statistical measures: N, number of observations (molecules) in the training set; q^2, cross-validated r^2 (by leave one out method) which is a relative measure of quality of fit; pred_r^2, r^2 for external test set; q^2se, standard error of cross-validation and pred_r^2se, standard error of external test set prediction. However, a QSAR model is considered to be predictive, if the following conditions are satisfied: $q^2 > 0.6$ and pred_r^2 >0.5[32]. The low standard error of pred_r^2se and q^2se shows absolute quality of fitness of the model. The high pred_r^2 and low pred_r^2se show high predictive ability of the model. The q^2 and pred_r^2 values were used as deciding factors in selecting the optimal models.

RESULTS AND DISCUSSION

For the development of two and three dimensional QSAR models of benzothiazine compounds, MLR methodology was used based on forward Step Wise selection method. A total number of 37 Benzothiazine derivatives have been considered for the QSAR study using software Vlife MDS 3.5. For the 37 molecules, which are the dataset is divided into training and test sets for an effective QSAR modeling (table 6). For selection of training and test sets, we were ensured that the molecules have uniform spread (training and test) in terms of both activity and chemical space. pMIC was selected as dependent variable and remaining all the variables were selected as independent variables. Selection of compounds in the training set and test set is a key and important feature of any QSAR model. Therefore care was taken in such a way that biological activities of all compounds in test lie within the maximum and minimum value range of biological activities of the training set of compounds. A unicolumn statistics for training set and test set were generated to check correctness of selection criteria for training and test set compounds. The training and test-set molecules for this group of compounds are selected by the random selection method and the models are validated by both internal and external validation procedures. Some statistically significant QSAR models were chosen for discussion.

MODEL I 2DQSAR (*E. Floccusom)*

pMIC =0.0463(±0.0006) SssCH2count-0.0039(±0.0000) Quadrupole1-3.4940(±0.9602) XA Average Hydrophobicity -0.8135(±0.2384) XAMostHydrophilic+0.6802

Statistics:

n = 29, Degree of freedom = 24, r^2 = 0.7672, q^2 = 0.6779, F test = 12.0265, r^2 se = 0.0667, q^2 se = 0.0835, pred_r^2 = 0.6622, pred_r^2se = 0.1372, Zscore_r^2 =6.209, Zscore_q^2 =4.39348, best_rand_r^2 = 0.2143, alpha_rand_q^2 =0.00

The derived model from multiple linear regression(MLR) with forward stepwise shows good correlation between biological activity and parameters SssCH2count, Quadrupole1, XA Average Hydrophobicity and XA Most Hydrophilic as the coefficient of determination, (r^2)= 0.7672 capable of explaining 76% of variance in the observed activity values. All the descriptors contributed well for the generation of model which is show in figure 4. The low standard error of r^2_se = 0.0667 demonstrates accuracy of the model. The leave-one-out procedure was used for internal validation of the model. The model showed an internal predictive power cross validated r^2 (q^2 =0.6779) of 68% and predictivity for external test set (pred_r^2 = 0.6622) about 66% and low q^2_se = 0.0835 values reflect good internal predictive power of the model. In addition, the randomization test shows confidence of ~ 99.9% that the generated model is not random and hence it is chosen as the QSAR model. The F-test= 12.0265 shows the overall statistical significance level of 99.99% of the model which means the probability of failure of the model is 1 in 10,000. The graph of actual versus predicted activity is shown in figure 3. Since positively contributed descriptor in the above model is SssCH2count, this descriptor denotes count of CH_2 groups attached with 2 single bonds to any atoms in the molecule. The positive correlation suggests that antifungal activity of 1,4-benzothiazine derivatives may be improved by increasing the number of such CH_2 groups present in the molecules. SssCH_2count defines hydrophobic state indices for number of CH_2 group connected with two single bonds. The descriptors show positive correlation among the parameters selected for the derived QSAR model. The positive coefficients suggest that inclusion of such carbon atoms in the molecules lead to increased antifungal activity. Quadrupole1 descriptor signifies magnitude of first tensor of quadrupole moments. Its negative contribution in the QSAR model implies that will lead to decrease potency. Its negative value suggests that decreasing the number of such atom that increase the dipole moment will lead to better antifungal potency. The XA Average Hydrophobicity and XA Most Hydrophilic descriptor are type of hydrophobic hydrophilic scale and its negative contribution for the antifungal activities indicate that optimum hydrophilic and hydrophobic groups provide good antifungal activity. The

inter-correlation among the selected descriptors was very less due to auto scaling and cross correlation limit permitted was 0.6.

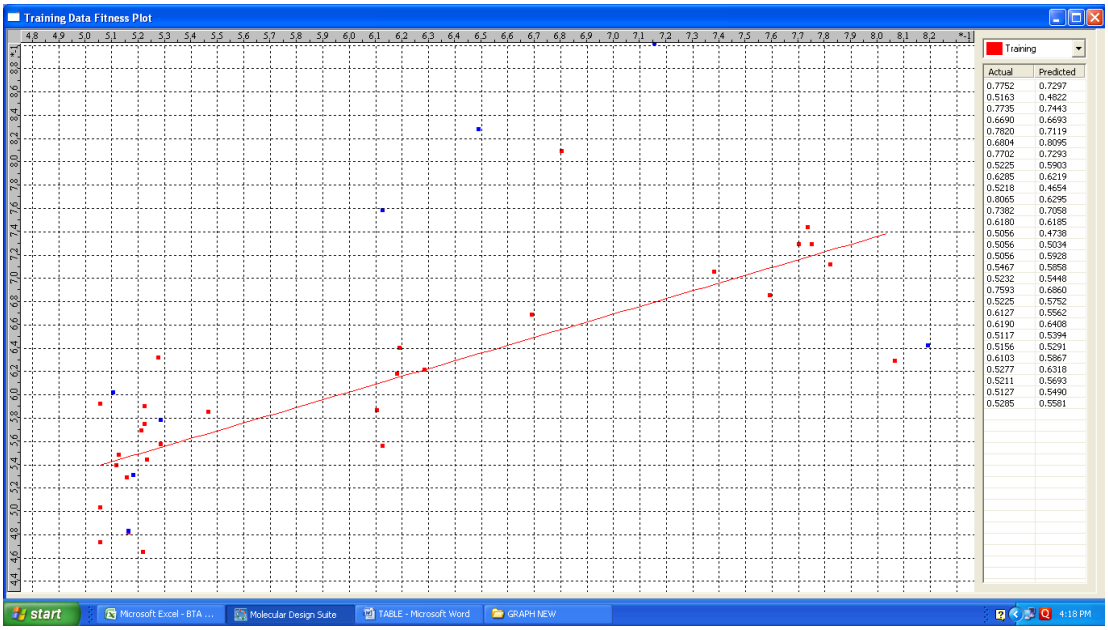


Figure 3: Fitness plot between observed activity Vs predicated activity for model I(2DQSAR).

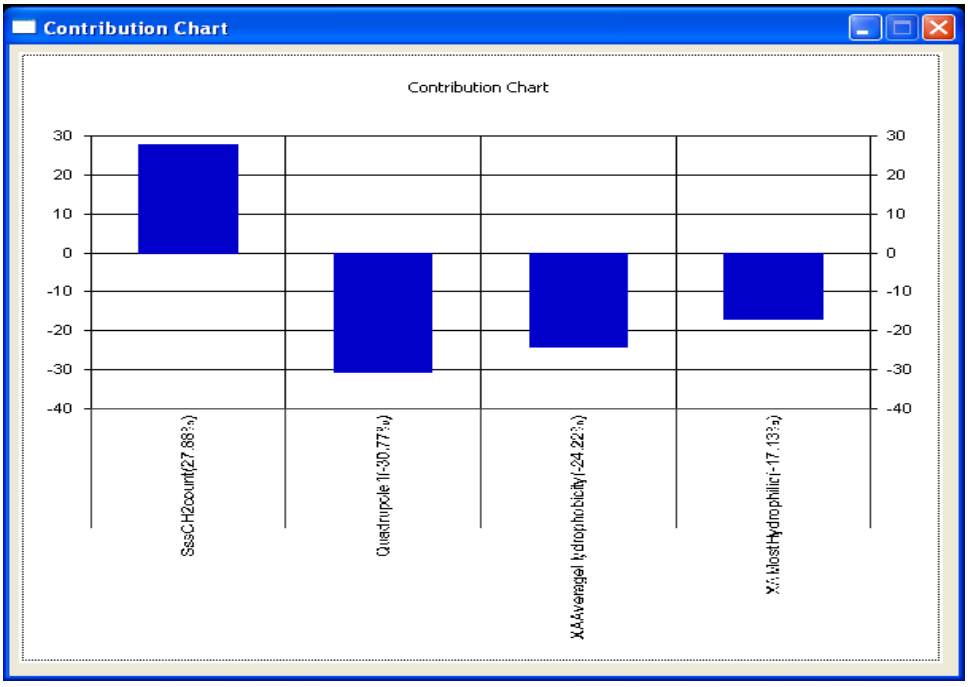


Figure 4: Contribution charts of the descriptors for model I (2DQSAR).

Model II 2DQSAR (*C. albicans*)

pMIC =6.8093(±2.1626) Average-vePotential+0.0073(±0.0000) Polar Surface Area Excluding PandS-0.1103(±0.0006) T_2_N_4-0.1272(±0.0016) T_N_O_3+0.7751

Statistics:

n = 29, Degree of freedom = 24, r^2 = 0.8747, q^2 = 0.7145, F test = 12.4439, r^2 se = 0.0739, q^2 se = 0.0902 , pred_r^2 = 0.9658, pred_r^2se = 0.1557, Zscore_r^2 =5.531, Zscore_q^2 =4.766, best_rand_r^2 = 0.3325, alpha_rand_q^2 =0.00

To improve the external predictivity of the model, MLR analysis with stepwise forward method was performed, Model II is a fourth parametric model generated with coefficient of determination, r^2 = 0.8747 which is capable of explaining 87.47% of variance in the observed activity values. The model showed an internal predictive power (q^2 = 0.7145) of 71.45% and predictivity for external test set (pred_r^2 =0.9658) about 96%. The F-test = 12.4439 shows the overall statistical significance level of 99.99 % of the model which means the probability of failure of the model is 1 in 10, 000. The randomization test suggest that the proposed QSAR model has a probability of less than 0.01of being generated by chance.

In this QSAR model, the positive coefficient of Average-ve Potential and Polar Surface Area Excluding P and S showed that increase in the values of these descriptors is beneficial for the antifungal activity. Model II reveals that the descriptor Polar Surface Area Excluding P and S plays the most important role (95.58%) in determining the antifungal activity (figure 5).

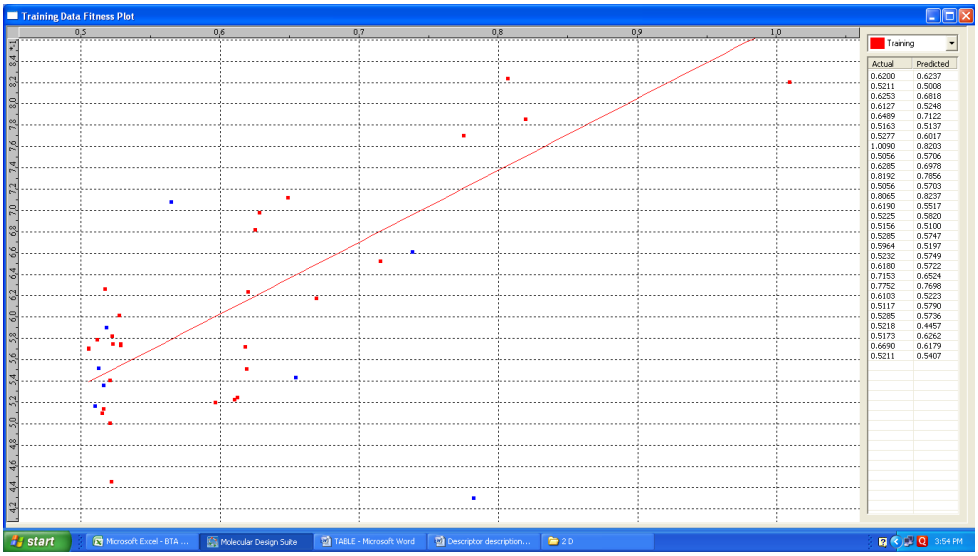


Figure 5: Fitness plot between observed activity Vs predicated activity for model II(2DQSAR).

Descriptors Polar Surface Area Excluding P and S signifies total polar surface area Excluding phosphorus and sulphur, which carries a positive sign in this model meaning that the increase in polarity of the molecule increases the activity. The next two important factors governing variation of the activity are T_2_N_4 (2.37%), T_N_O_3 (2%) are directly proportional to the activity. The

positive coefficient of Average-ve Potential but not such important for activity due to less contribute (0.11%). The positive coefficient of Average-ve Potential shows that increase in the Average-ve Potential is detrimental for the activity. The contribution of descriptors show in figure 6

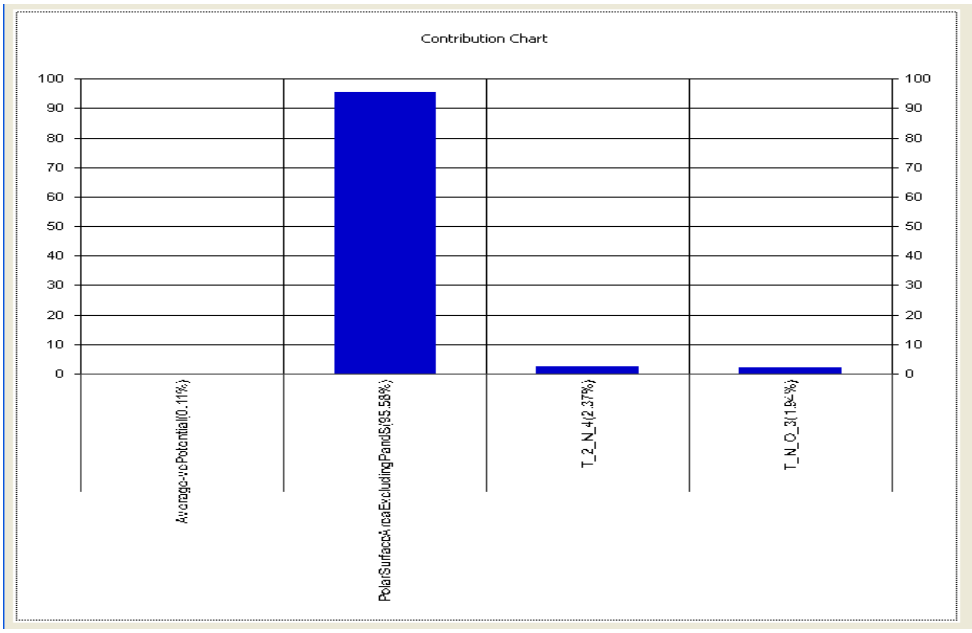


Figure 6: Contribution charts for model II(2DQSAR).

Model III 3D QSAR (*M. furfur*)

pMIC = 0.0340(±0.0070)S_535 - 0.3543(±0.0035)E_636 + 1.3214(±0.0707)E_439 - 0.3310(±0.0569)S_424 + 0.3970

Statistics:

n = 29, Degree of freedom = 24, r^2 = 0.8635, q^2 = 0.7614, F test = 37.9421, r^2 se = 0.0437, q^2 se = 0.0578, pred_r^2 = 0.6654, pred_r^2se = 0.2059, Zscore_r^2 =5.104, Zscore_q^2 =3.213, best_rand_r^2 = 0.3108, alpha_rand_q^2 =0.01

The descriptors that were selected in a given model were the field points of steric nature at particular locations in a common grid around reported set of molecules. The model selection criterion is the value of r^2, correlation coefficient and q^2, the internal predictive ability of the model, and that of pred_r^2, the ability of the model to predict the activity of external test set. In addition, the randomization test shows confidence of ~ 99.9% that the generated model is not random and hence it is chosen as the QSAR model. As the cross-validated correlation coefficient (q^2) is used as a measure of reliability of prediction, the correlation coefficient suggests that our model is reliable and accurate. A data set of compounds containing eight molecules was selected as the test set from the original data of 37 compounds for the validation experiments (table 6). The

q^2 value obtained (0.7614) is the indicative power of the model. According to this model pMIC is a function of independent variables and dependent variables are steric fields. Values of q^2 = 0.7614, F test = 37.9421, r^2 se = 0.0437, q^2 se = 0.0578, pred_r^2 = -0.6654, pred_r^2se = 0.2059 prove that QSAR equation so obtained is statistically significant and shows the predictive power of the model as 76.13% (internal). The observed and predicative activity of model III is given in table 5 along with minimum residual indicates good correlation. The descriptors that were selected in a given model are the field points of steric and electrostatic nature at particular locations in a common grid around reported set of molecules. The descriptors S_535, E_636, E_439 and S_424 were the steric and electrostatic field energy of interactions between probe (CH_3) and compounds at their corresponding spatial grid points of 535, 636, 439 and 424. The contribution of electrostatic fields indicates that electric field is more predominant and contribution chart of selected descriptors are represented in Figure 8. It is evident that the predicted activities of all the compounds in the test set are in good agreement with their corresponding experimental activities and optimal fit is obtained (figure 7)

Table 5. Observed and predicted activity by 2QSAR and 3DQSAR equations along with the residuals.

	For model I (2DQSAR) *E. Floccusom*			**For model II (2DQSAR) *C. albicans***		
Compound	**Observed**	**predicted**	**residuals**	**Observed**	**predicted**	**residuals**
BTA20	0.5218	0.7058	-0.184	0.5056	0.5703	-0.0647
BTA24	0.7735	0.7443	0.0292	0.6253	0.6818	-0.0565
BTA25	0.7593	0.5562	0.2031	0.7153	0.6524	0.0629
BTA35	0.6489	0.8283	-0.1794	0.5647	0.7079	-0.1432
BTA43	0.5467	0.686	-0.1393	0.5232	0.5749	-0.0517
BTA47	0.6543	0.6427	0.0116	0.6543	0.5431	0.1112
BTA4	0.669	0.6693	-0.0003	0.6127	0.5248	0.0879
BTA56	0.6285	0.6295	-0.001	0.8192	0.7856	0.0336
BTA57	0.5056	0.5928	-0.0872	0.5156	0.51	0.0056
BTA58	0.5127	0.549	-0.0363	0.669	0.6179	0.0511
BTA59	0.619	0.5291	0.0899	0.5117	0.579	-0.0673
BTA5	0.7702	0.7293	0.0409	0.5277	0.6017	-0.074
BTA60	0.6804	0.8095	-0.1291	0.5163	0.5137	0.0026
BTA61	0.7752	0.7297	0.0455	0.62	0.6237	-0.0037
BTA63	0.5227	0.6408	-0.1181	0.7752	0.7698	0.0054
BTA64	0.5277	0.5693	-0.0416	0.5173	0.6262	-0.1089
BTA65	0.5127	0.1589	0.3538	0.5127	0.552	-0.0393
BTA66	0.8065	0.6185	0.188	0.8065	0.8237	-0.0172
BTA67	0.5163	0.4822	0.0341	0.5211	0.5008	0.0203
BTA68	0.6804	0.5903	0.0901	1.009	0.8203	0.1887
BTA69	0.5225	0.4654	0.0571	0.6285	0.6978	-0.0693
BTA70	0.782	0.5784	0.2036	0.782	0.4302	0.3518

BTA8	0.5117	0.5867	-0.075	0.5285	0.5736	-0.0451
BTMA11	0.5285	0.5581	-0.0296	0.5211	0.5407	-0.0196
BTMA12	0.782	0.7119	0.0701	0.6489	0.7122	-0.0633
BTMA24	0.5056	0.5448	-0.0392	0.5964	0.5197	0.0767
BTMA33	0.7382	0.6025	0.1357	0.7382	0.6609	0.0773
BTMA36	0.6103	0.6318	-0.0215	0.5218	0.4457	0.0761
BTMA48	0.5105	0.9022	-0.3917	0.5105	0.5168	-0.0063
BTMA50	0.7702	0.6219	0.1483	0.5056	0.5706	-0.065
BTMA59	0.5232	0.5752	-0.052	0.618	0.5722	0.0458
BTMA60	0.5163	0.4838	0.0325	0.5163	0.5362	-0.0199
BTMA74	0.6127	0.5394	0.0733	0.6103	0.5223	0.088
BTMA77	0.5056	0.5858	-0.0802	0.5285	0.5747	-0.0462
BTMA80	0.618	0.5034	0.1146	0.5225	0.582	-0.0595
BTMA83	0.7382	0.4738	0.2644	0.619	0.5517	0.0673
BTMA88	0.5183	0.5316	-0.0133	0.5183	0.5902	-0.0719

Table 5. (*continued*).

For model III (3DQSAR) *M. furfur*			**For model IV (3DQSAR) *C. albicans***		
0.6285	0.5796	0.0489	0.5285	0.4835	0.045
0.5285	0.5995	-0.071	0.6285	0.6099	0.0186
0.6489	0.7439	-0.095	0.8065	0.8739	-0.0674
0.5056	0.469	0.0366	0.5964	0.5536	0.0428
0.6571	0.6179	0.0392	0.8192	0.7628	0.0564
0.5429	0.6106	-0.0677	0.689	0.5571	0.1319
0.669	0.657	0.012	0.669	0.6494	0.0196
0.62	0.6168	0.0032	0.62	0.6024	0.0176
0.5127	0.321	0.1917	0.5127	1.14	-0.6273
0.6804	0.6887	-0.0083	0.5647	0.58	-0.0153
1.0034	0.8466	0.1568	0.6253	0.644	-0.0187
0.5886	0.6419	-0.0533	0.7153	0.6454	0.0699
0.5277	0.5829	-0.0552	0.5277	0.5952	-0.0675
0.5173	0.4448	0.0725	0.5173	0.5108	0.0065
0.5105	0.5365	-0.026	0.5105	0.4469	0.0636
0.7477	0.7785	-0.0308	0.6103	0.6289	-0.0186
0.5218	0.5495	-0.0277	0.5218	0.5191	0.0027
0.5163	0.5275	-0.0112	0.5163	0.5272	-0.0109
1.0118	0.5739	0.4379	0.7752	0.531	0.2442
0.5156	0.5469	-0.0313	0.5156	0.5492	-0.0336
0.5056	0.5154	-0.0098	0.5056	0.5244	-0.0188
0.6543	0.7465	-0.0922	0.6543	0.6115	0.0428
0.5232	0.7311	-0.2079	0.5232	0.6653	-0.1421
0.5173	0.4836	0.0337	0.6127	0.584	0.0287
0.619	0.6304	-0.0114	0.619	0.4916	0.1274
0.6274	0.6245	0.0029	1.009	0.9881	0.0209
0.5211	0.523	-0.0019	0.618	0.5559	0.0621
0.604	0.6371	-0.0331	0.7382	0.4334	0.3048
0.782	0.7503	0.0317	0.782	0.6776	0.1044
0.5964	0.558	0.0384	0.5056	0.5194	-0.0138

0.5117	0.494	0.0177	0.5117	0.5876	-0.0759
0.5211	0.5021	0.019	0.5211	0.5928	-0.0717
0.5285	0.5392	-0.0107	0.5285	0.6114	-0.0829
0.618	0.555	0.063	0.5211	0.5735	-0.0524
0.62	0.5828	0.0372	0.5225	0.5424	-0.0199
0.5163	0.5324	-0.0161	0.5163	0.542	-0.0257
0.7534	0.6352	0.1182	0.5183	0.607	-0.0887

Table 5. (*continued*).

For model V (3DQSAR) *E. floccosum*			**For model VI (3DQSAR) *T. rubrum***		
0.5232	0.5235	-0.0003	0.6571	0.6096	0.0475
0.5285	0.4551	0.0734	0.5056	0.4563	0.0493
0.5056	0.5535	-0.0479	0.8148	0.8222	-0.0074
0.5225	0.6282	-0.1057	0.5218	0.6464	-0.1246
0.5467	0.5317	0.015	0.6489	0.7061	-0.0572
0.7752	0.7195	0.0557	1.0118	0.6454	0.3664
0.7735	0.6213	0.1522	0.62	0.5987	0.0213
0.5127	0.491	0.0217	0.5218	0.5597	-0.0379
0.6285	0.6075	0.021	0.5285	0.6001	-0.0716
0.5285	0.5744	-0.0459	0.5277	0.5307	-0.003
0.669	0.6526	0.0164	0.5886	0.6105	-0.0219
0.7153	0.6841	0.0312	0.782	0.7728	0.0092
0.7702	0.7009	0.0693	0.5056	0.4649	0.0407
0.7382	0.5827	0.1555	0.5111	0.5756	-0.0645
0.6489	0.6584	-0.0095	0.6103	0.6615	-0.0512
0.5225	0.5341	-0.0116	0.6285	0.6226	0.0059
0.6127	0.6255	-0.0128	0.8065	0.7342	0.0723
0.5218	0.5866	-0.0648	0.618	0.6954	-0.0774
0.5105	0.3912	0.1193	0.5105	0.5025	0.008
0.5211	0.5448	-0.0237	0.5173	0.4744	0.0429
0.7593	0.7034	0.0559	0.5163	0.5375	-0.0212
0.6103	0.5808	0.0295	0.62	0.6266	-0.0066
0.5056	0.4507	0.0549	0.5163	0.4341	0.0822
0.782	0.7787	0.0033	0.5211	0.5958	-0.0747
0.5056	0.4053	0.1003	0.5056	0.5563	-0.0507
0.5117	0.5555	-0.0438	0.6062	0.5929	0.0133
0.5277	0.5577	-0.03	0.6274	0.5704	0.057
0.5156	0.6456	-0.13	0.5156	0.5339	-0.0183
0.5163	0.4798	0.0365	1.0034	0.892	0.1114
0.8065	0.8353	-0.0288	0.6127	0.5643	0.0484
0.619	0.5443	0.0747	0.6285	0.567	0.0615
0.5163	0.6428	-0.1265	0.5117	0.609	-0.0973
0.6804	0.6693	0.0111	0.669	0.6286	0.0404
0.6127	0.6499	-0.0372	0.5211	0.5159	0.0052
0.8192	0.777	0.0422	0.5647	0.6041	-0.0394
0.618	0.6058	0.0122	0.5232	0.5301	-0.0069
0.5183	0.5477	-0.0294	0.7534	0.5669	0.1865

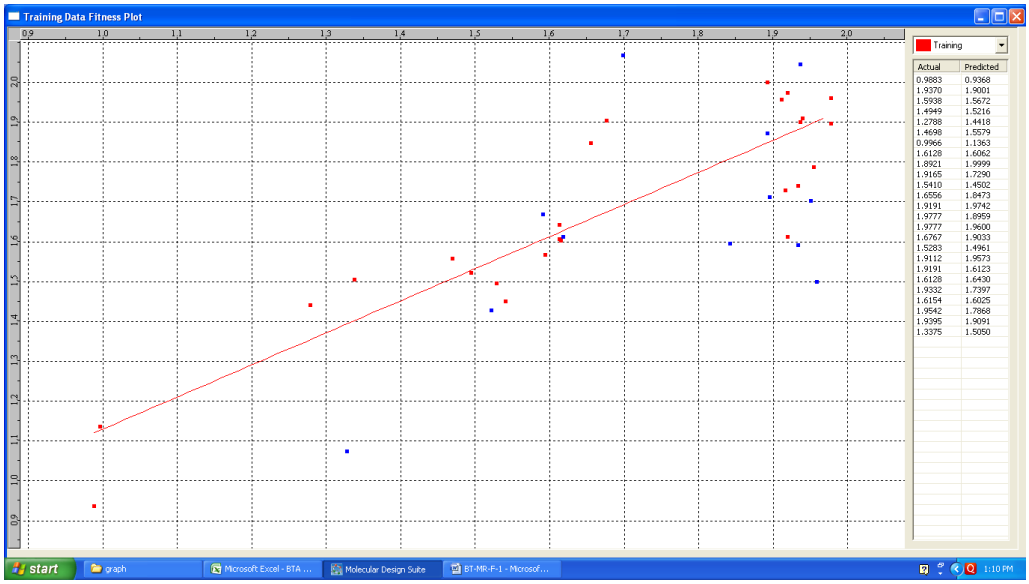


Figure 7: Fitness plot between observed activity Vs predicated activity of model III.

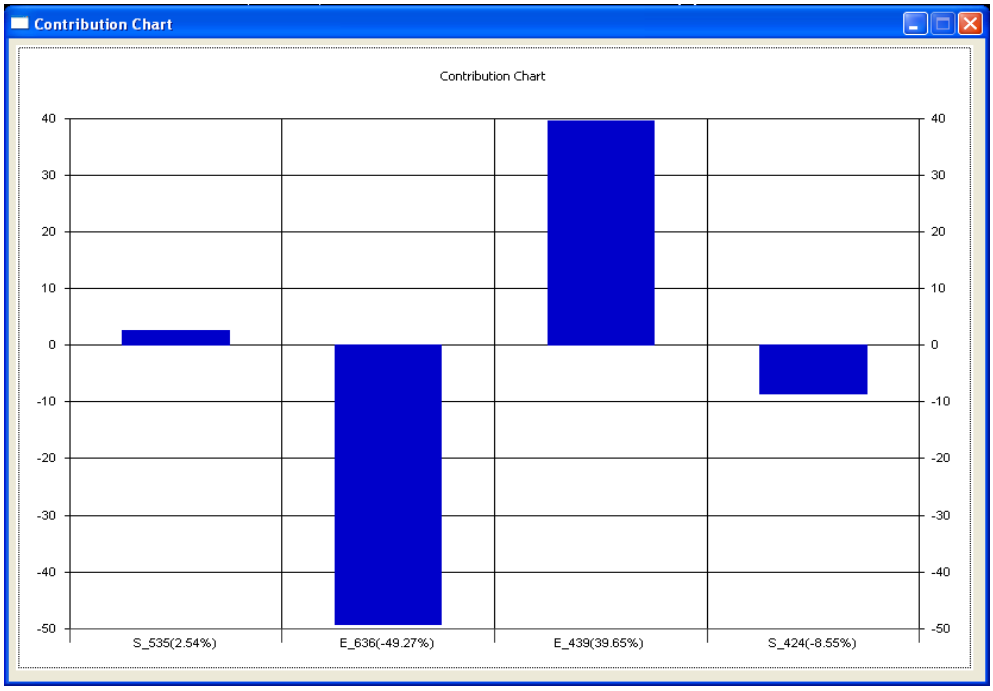


Figure 8: Contribution charts of the descriptors for model III.

Model IV-3D QSAR (*C. albicans*)

pMIC = 0.3294(±0.0445)S_1183 + 0.1210(±0.0185)E_1260 + 0.0658(±0.0124)E_962 + 0.1064(±0.0315)E_303 + 0.5420

Statistics:

n = 29, Degree of freedom = 24, r^2 = 0.8335, q^2 = 0.7352, F test = 30.04, r^2 se = 0.0533, q^2 se = 0.0672, pred_r^2 = 0.6556, pred_r^2se = 0.2925, Zscore_r^2 =2.895, Zscore_q^2 =2.395, best_rand_r^2 = 0.3039, alpha_rand_q^2 =0.01

The descriptors S_1183 are the steric field energy of interactions between probe (CH_3) and compounds at their corresponding spatial grid points 1183. The only one of steric fields is more predominant and contribution chart of selected descriptors are represented in Figure 11. It is evident that the predicted activities of all the compounds in the test set are in good agreement with their corresponding experimental activities and optimal fit is obtained. The contribution plot of steric field interactions indicates relative regions of the local fields (steric) around the aligned molecules. The green-colored balls (figure12) specify the positions of the steric descriptors and the descriptors with positive or negative coefficients show a region where bulky substituent is favored or disfavored, respectively. From 3D-QSAR model it is observed that steric descriptors like S_1183 with positive coefficient signifying positive range of steric descriptors indicate that positive steric potential is favorable for activity and more bulky substituent is preferred in that region. The E_1260, E_962 and E_303 descriptor are electrostatic field in nature and contribute around 67%, positive coefficient means at that particular point electron withdrawing group will favor the activity. The correlation plot of the actual versus predicted activity is shown in Figure 10. Robustness of the QSAR model for experimental training sets was examined by comparing this model to those derived for sphere exclusion dataset. In addition, the randomization test suggests that QSAR model are generate not by chance. The QSAR model was evaluated using the following statistical measures; numbers of observations, i.e., molecules in data set (n =29); cross validated r^2 (q^2 = 0.7352); predicted r^2 for the external test set (pred_r^2 = 0.6556); standard error of cross validated r^2 (pred_r^2se = 0.2925).

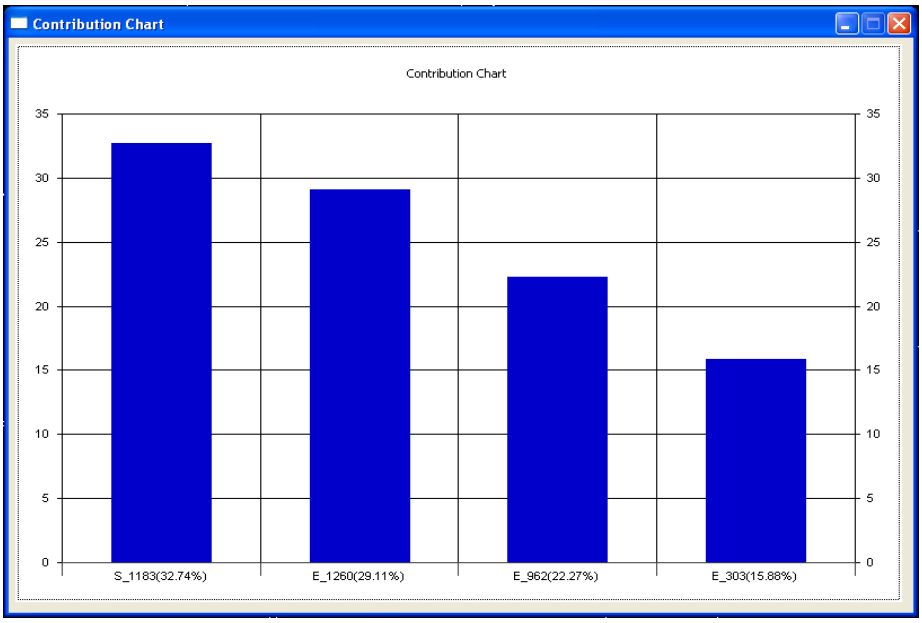


Figure 11: Contribution charts of the descriptors for model IV.

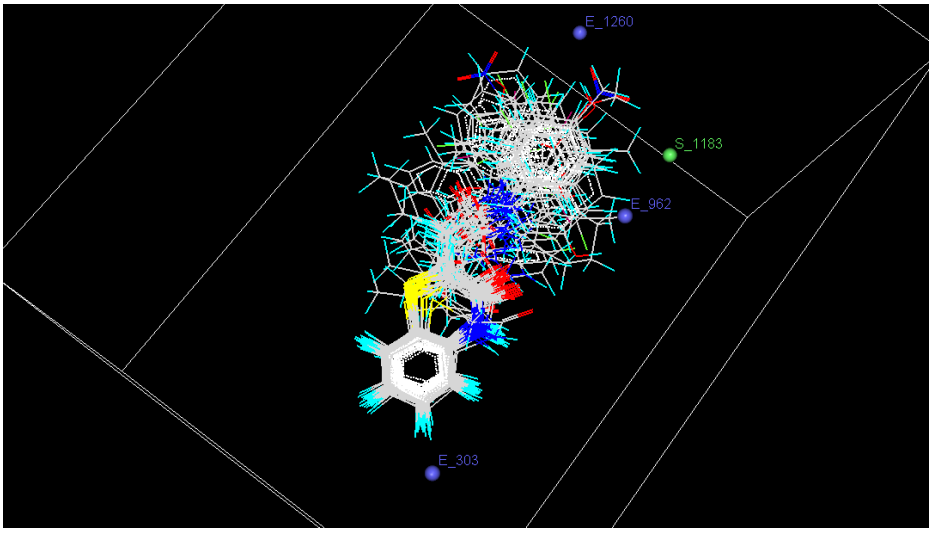


Figure 12: 3D View of aligned molecule and contribution of descriptors of model IV.

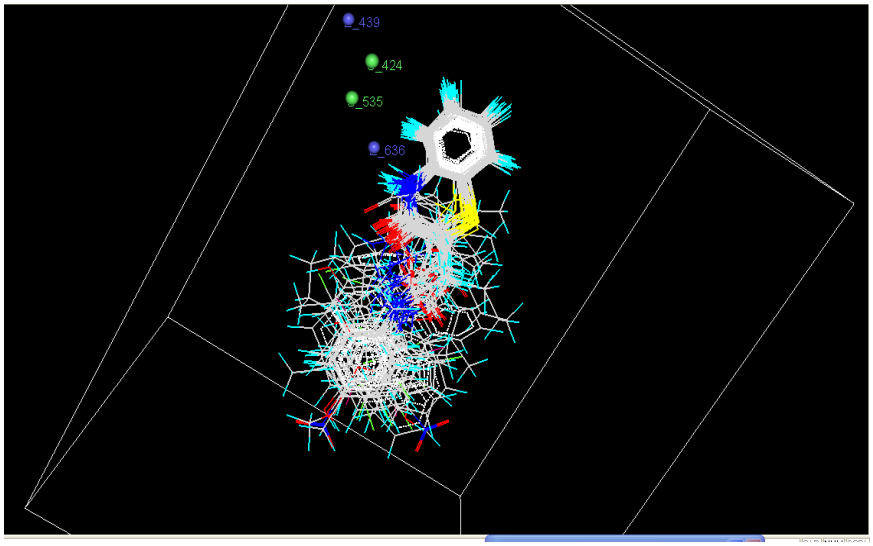


Figure 9: 3D view of aligned molecule with contribution of descriptors (model III)

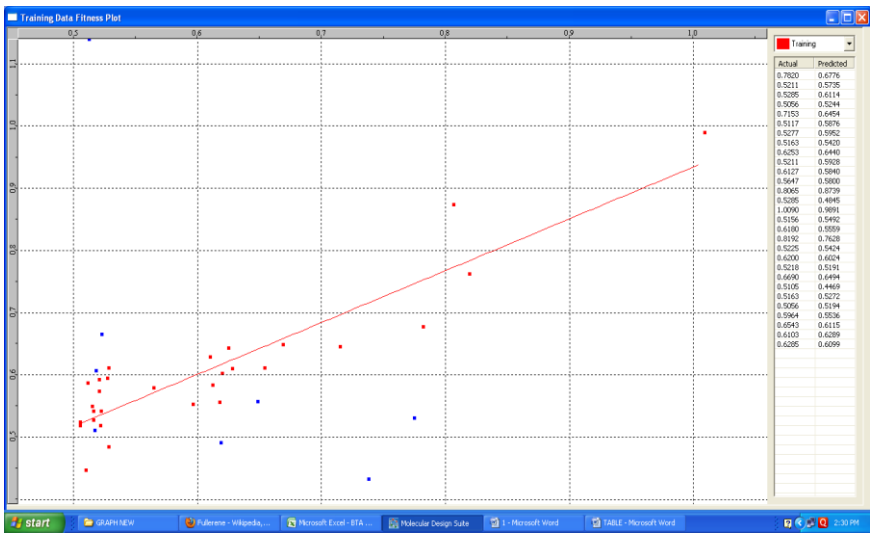


Figure 10: Fitness plot between observed activity Vs predicated activity of model IV.

Model V-3D QSAR (*E. floccosum*)

pMIC = 0.2333 (±0.0368) S_1282 - 0.8029 (±0.1470) S_645- 0.1147 (±0.0311) S_599 + 0.1317(±0.0315)S_1181+0.4162

Statistics:

n = 29, Degree of freedom = 24, r^2 = 0.7518, q^2 = 0.7137, F test = 18.1776, r^2 se = 0.0573, q^2 se = 0.0802, pred_r^2 = 0.6281, pred_r^2se = 0.1021, Zscore_r^2 = 4.652, Zscore_q^2 =3.11, best_rand_r^2 = 0.406, alpha_rand_q^2 =0.01

Model V are resulted from the activity against *E. floccosum*. Model V with coefficient of determination (r^2) = 0.7518 which is capable of explaining 75.18% of variance in the observed activity values. The model selection criterion is the value of q^2, the internal predictive ability of the model, and that of pred_r^2, the ability of the model to predict the activity of external test set. As the cross-validated correlation coefficient (q^2) is used as a measure of reliability of prediction, the correlation coefficient suggests that our model is reliable and accurate. The randomization tests suggest that the proposed QSAR model has a probability of less than 0.01 of being generated by chance. S_1282, S_645, S_599 and S_1182 are steric descriptors contributing to model V (figure 13).

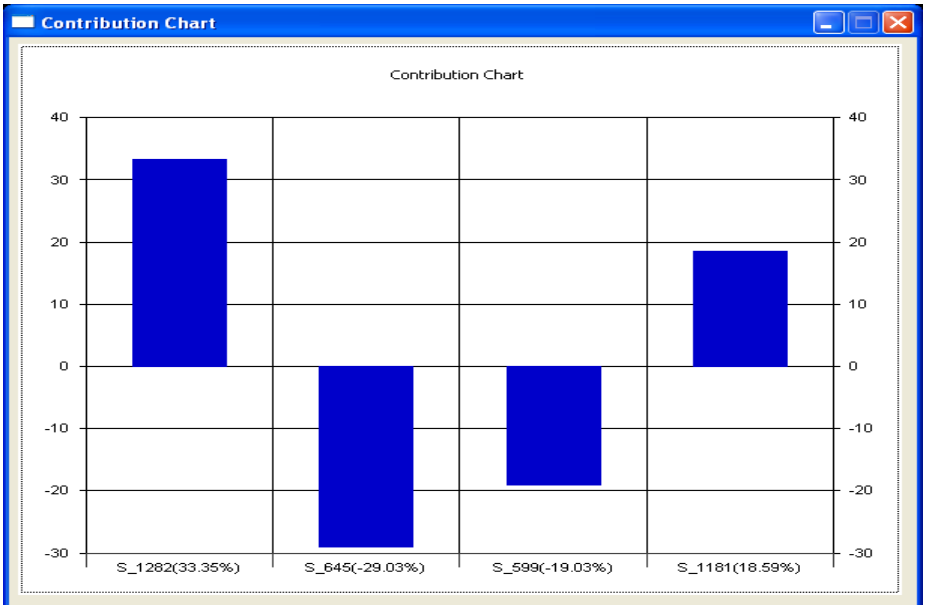


Figure 13: Contribution charts of the descriptors for model V.

The q^2 value obtained (0.7137) is the indicative power of the current MLR model. Values of r^2 = 0.7518, q^2 = 0.7137, F test = 18.1776, r^2 se = 0.0573, q^2 se = 0.0802, pred_r^2 = 0.6281, pred_r^2se = 0.1021 prove that QSAR equation are obtained is statistically significant and shows that the predictive power of the model is 71.37% (internal validation) and 62.81 % (external validation).

Forward method of variable selection indicating its crucial role in predicting antifungal activity and signifying negative range of steric descriptors indicate that negative steric potential is favorable for activity and less bulky substituent is preferred in that region. The steric effect, as shown in figure 15 with green color ball around the phenyl ring at ortho or meta position, implies about the preferred substitution (less or more bulky group) to produce higher antifungal activity. Electrostatic descriptor with positive coefficient (E_918) around position of the phenyl ring corroborates that electropositive (electron-withdrawing) group is preferred at 4-position of phenyl ring.

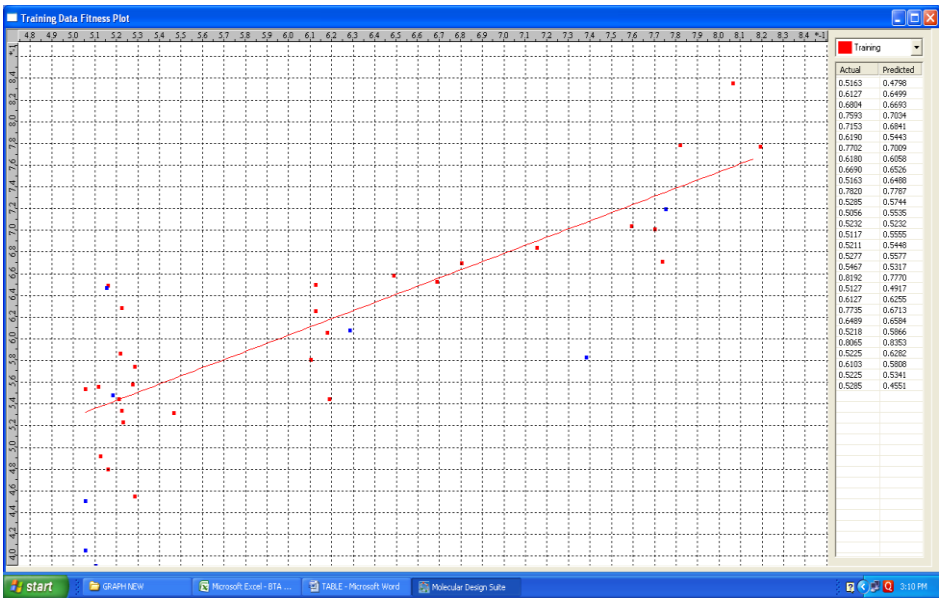


Figure 14: Fitness plot between observed activity Vs predicated activity of model V.

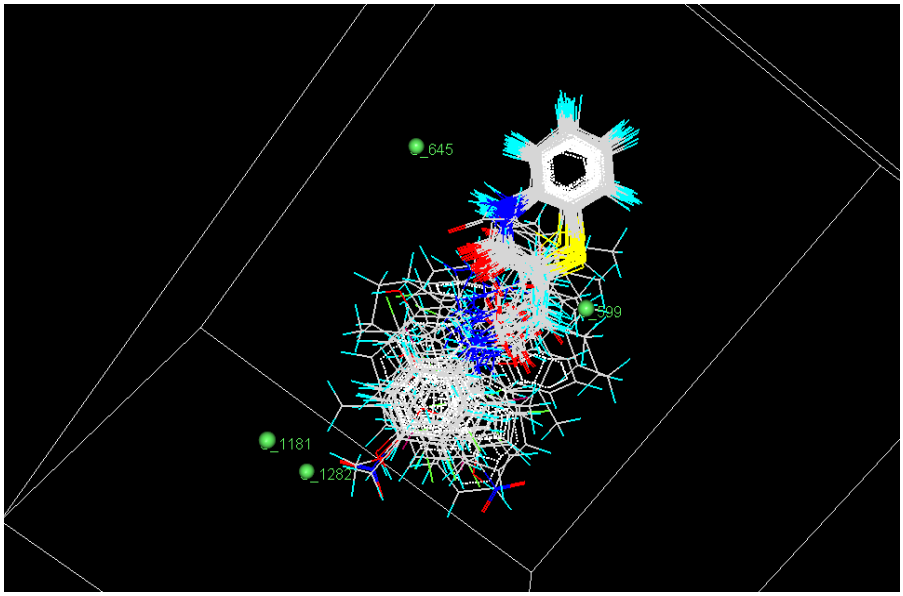


Figure 15: 3D View of aligned molecule and contribution of descriptors of model V.

Model VI (*T. rubrum*)

pMIC = 0.2168(±0.0321)S_535-0.0614(±0.0131)E_624+0.0670(±0.0134)E_963 + 0.0839 (±0.0291)S_864+0.6296

STASTICS

n = 29, Degree of freedom = 24, r^2 = 0.7727, q^2 = 0.7516, F test = 20.3915, r^2 se = 0.0592, q^2 _se = 0.0733, pred_r^2 = 0.6075, pred_r^2se = 0.1639, Zscore_r^2 =4.197, Zscore_q^2 =3.44, best_rand_r^2 = 0.441, alpha_rand_q^2 =0.01

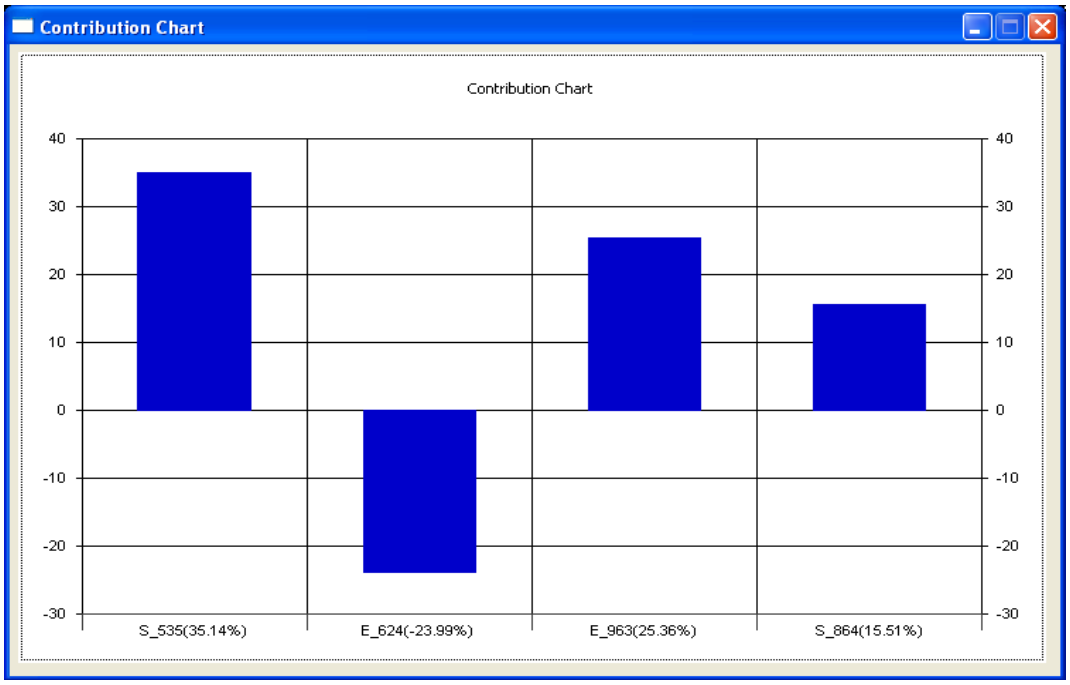


Figure 16: Contribution charts of the descriptors for model VI.

The descriptors that get selected in a given model are the field points either of steric and electrostatic nature at particular locations in a common grid around reported set of molecules. For 3D QSAR a MLR with Forward method resulted in several statistically significant models, of which the corresponding best model VI is reported herein. In this model high cross validated q^2 = 0.7516 and low q^2_se= 0.0733 value, reflects the very good internal predictive power of the model. The model selection criterion is the value of r^2 fitness of model, q^2 the internal predictive ability of the model and pred_r^2, the ability of the model to predict the activity of external test set. As the cross-validated correlation coefficient (q^2) is used as a measure of reliability of prediction, the correlation coefficient suggests that our model is reliable and accurate. According to this model pMIC is a function of independent variables and dependent variables are steric and electrostatic fields. Values of r^2 = 0.7727, q^2 = 0.7516, F test = 20.3915, r^2 se = 0.0592, q^2 _se = 0.0733, pred_r^2 = 0.6075, pred_r^2se = 0.1639, prove that QSAR equation so obtained is statistically significant and

shows the predictive power of the model is 75.16% (internal) and 60.75% (external). In addition, the randomization test shows confidence of 99 % (alpha_rand_q^2 =0.01) that the generated model is not random and hence may be chosen as the QSAR model. The descriptors S_535, E_624, E_963 and S_864 are the steric and electrostatic field energy of interactions between probe (CH_3) and compounds at their corresponding spatial grid points of 535, 624, 963 and 864 show in 3D view (figure 18). The contributions of steric and electrostatic fields indicate that both fields are more important and the contribution chart of selected descriptors is represented in Figure 16. It is evident that the predicted activities of all the compounds in the test set are in good agreement with their corresponding experimental activities and optimal fit is obtained show in Figure17.

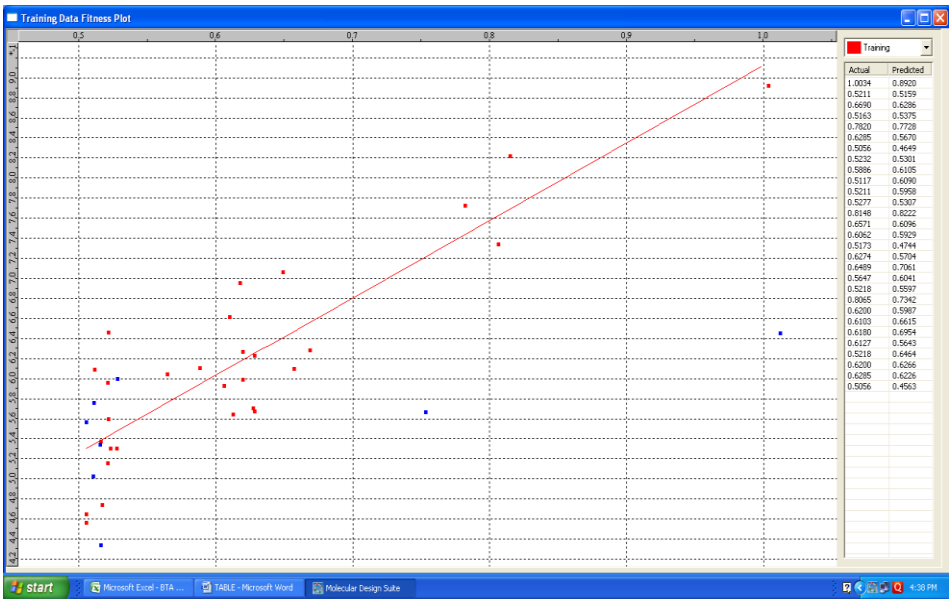


Figure 17: Fitness plot between observed activity Vs predicated activity of model VI.

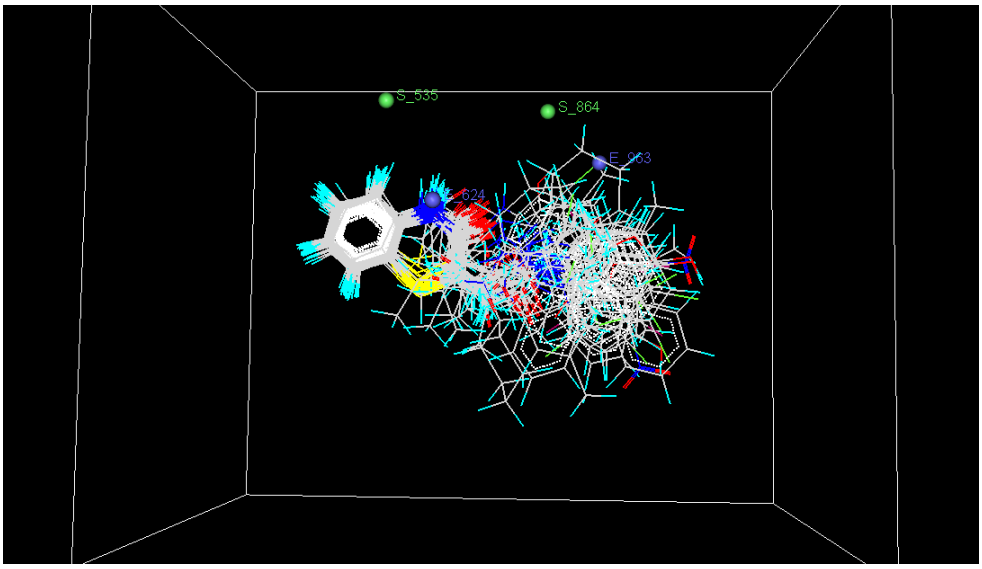


Figure 18: 3D View of aligned molecule and contribution of descriptors of model VI.

CONCLUSION

QSAR models were statistically significant, thus, from above QSAR investigations it could be concluded that 2D and 3D descriptors properties of 1,4-Benzothiazine derivatives are mainly involved in antifungal activity. The good correlation between experimental and predicted biological activity in the test set further highlights the reliability of the constructed QSAR model. The requirements for the antifungal activity are explored with 2D and 3DQSAR studies. The 2DQSAR analysis indicates the importance of Quadrople1, SssCH2count, XA Average hydrophobicity, XA most hydrophilic area and Polar Surface Area Excluding P and S of the compounds on the activity. From the QSAR model concluded that the bulky substitution is important at para and meta position of Phenyl ring and less bulky substitution at ortho position of phenyl ring. Electron-withdrawing group is preferred at para position of phenyl ring. NH-CO group of lactam of 1,4- Benzothiazine is important for the antifungal property as they showed interactions shown in 3DQSAR model.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the ‘‘Board of College and University Development, University of Pune, Pune for providing funding to institute for VLife MDS software. Authors like to acknowledge Principal of the institute for providing facilities to carry out the work. The authors also thank the anonymous reviewers whose valuable comments and suggestions greatly helped in improving the manuscript.

REFERENCES

1. Imwidthaya P, Poungvarin . Cryptococcosis in AIDS. Postgrad Med J 2000; 76: 85-88.
2. Jain NP, Upasani CD, Jain UN. 3DQSAR Studies of *N*-[Alkyl or Un/substituted phenyl]-2-(3-oxo-3,4-dihydro-2*H*-benzo[b] [1,4] thiazin-2-yl)acetamide. Int J Pharmaceutical and Phytopharmacological Res 2013; 3(2): 124-133.
3. Fridkin SK, Jarvis WR. Epidemiology of nosocomial fungal infections. Clin Microbiol Rev 1996; 9: 499-511.
4. Maertens JA, Boogaerts MA. Fungal cell wall inhibitors: emphasis on clinical aspects. Curr Pharm Des 2000; 6: 225-239.
5. Watkins WJ, Renau TE. Burger’s Medicinal Chemistry and Drug Discovery. Edn 6. Vol 5, John Wiley & Sons Inc, New York. 2003.
6. Latge JP. *Aspergillus fumigatus* and Aspergillosis. Clin Microbiol Rev 1999; 12: 310-350.

7. Jain NP, Upasani CD, Jain UN. Synthesis and antifungal Activity of 1,4-Benzothiazine Derivatives. Am J PhamTech Res 2013; 3(5): 548-558.
8. VLife MDS 3.5, Molecular Design Suite. Vlife Sciences Technologies Pvt. Ltd. Pune, 2010.
9. N.P. Jain, Dissertation, Vinayaka Mission University, 2013;74-75.
10. Halgren T A. Merck molecular force field: III. Molecular geometries and vibrational frequencies for MMFF94. J Comput Chem 1996; 17: 553–586.
11. Golbraikh A, Tropsha A. Beware of q2!. J Mol Graph Model 2002; 20: 269–276.
12. Cramer III 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.
13. Tanneeru K, Reddy BM, Guruprasad L. Three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis and molecular docking of ATP-competitive triazine analogs of human mTOR inhibitors. Med Chem Res 2011. Availabel at: http://dx.doi.org/10.1007/s00044-011-9629-x. Accessed on 15 December 2012.
14. Sharma MC, Kohli DV. A comprehensive structure–activity analysis 2,3,5-trisubstituted 4,5-dihydro-4-oxo-3H-imidazo [4,5-c] pyridine derivatives as angiotensin II receptor antagonists: using 2D- and 3D-QSAR approach. Med Chem Res. 2012. Availabel at: http://dx.doi.org/10.1007/s00044-012-0040-z. Accessed on 17 December 2012
15. Ajmani S, Jadhav K, Kulkarni SA. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. J Chem Inf Model 2006; 46: 24–31.
16. Radhika V, Kanth SS, Vijjulatha M. CoMFA and CoMSIA studies on inhibitors of HIV-1 integrase-bicyclic pyrimidinones. J Chem 2010; 7(S1): S75–S84.
17. Nakka S, Guruprasad L. The imidazolidone analogs as phospholipase D1 inhibitors: analysis of the three-dimensional quantitative structure–activity relationship. Med Chem Res 2012; 21:2517–2525.
18. Sahu NK, Bari SB, Kohli DV. Molecular modeling studies of some substituted chalcone derivatives as cysteine protease inhibitors. Med Chem Res. (2011). Availabel at: http://dx.doi.org/10.1007/s00044-011-9900-1. Accessed on 17 December 2012
19. Ghosh P, Bagchi MC. QSAR modeling for quinoxaline derivatives using genetic algorithm and simulated annealing based feature selection. Curr Med Chem 2009; 16: 4032–4048.

20. Clark M, Cramer III RD, Van ON. Validation of the general purpose Tripose 5.2 force field. J Comput Chem 1989; 10: 982–1012
21. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity a rapid access to atomic charges. Tetrahedron 1980; 36: 3219–3228.
22. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003; 3: 1157–1182.
23. Regression and Linear Models, RB. Darlington, New York, USA: McGraw-Hill; 1990.
24. Zheng W, Tropsha A. Novel variable selection quantitative structure property relationship approach based on the k-nearest neighbor principle. J Chem Inf Comput Sci 2000; 40: 185–194.
25. Shen M, Xiao Y, Golbraikh A, Gombar VK and Tropsha A, Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates J Med Chem 2003; 46:3013.3DQSAR Studies of *N*-[Alkyl or Un/substituted phenyl]-2-(3-oxo-3,4-dihydro-2*H*-benzo[b] [1,4] thiazin-2-yl)acetamide .