Journal of Computational Methods in Molecular Design, 2012, 2 (1):24-38

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

2D QSAR Analysis of 3', 4', 5'- Trimethoxychalcone analogues as inhibitors of nitric oxide production and tumor cell proliferation

Steffi I. Gonsalves and A. R. Shaikh*

Bhujbal Knowledge City, METs Institute of Pharmacy, Adgaon, Nashik, India

---

ABSTRACT

Two dimensional quantitative structure activity relationship (2D QSAR) studies by means of multiple linear regression (MLR) method was performed on a series of 3', 4', 5'-trimethoxychalcone analogues as anti inflammatory and antitumor agents using software QSARpro (VLifeScience). This study was performed with 23 compounds (data set) using random and manual data selection methods for the division of the data into training and test set. MLR methodology with stepwise (SW) forward-backward variable selection method was used for building the QSAR models. Statistically significant QSAR models were developed. Among them most significant models has squared correlation coefficient (r^2), cross validated correlation coefficient (q^2) and predictive correlation coefficient (pred_r^2) for Anti inflammatory activity (0.9751, 0.7518, 0.4496), antitumor Hep G2 activity (0.7737, 0.5955, 0.5270), and antitumor Colon 205 activity (0.9067, 0.7759, 0.2928) respectively. The first QSAR model indicates that the descriptors Quadrupole 2 [signifies magnitude of second tensor of quadrupole moments]; SK Most Hydrophobic Hydrophilic Distance [signifies distance between most hydrophobic and hydrophilic point on the vdW surface (By Kellog Method using Slogp)], XK Most Hydrophobic Hydrophilic Distance [signifies distance between most hydrophobic and hydrophilic point on the vdW surface (By Kellog Method using Xlogp)] , T_O_O_5 [count of number of oxygen atoms (single double or triple bonded) separated from any other oxygen atom (single double or triple bonded) by 5 bonds in a molecule], T_N_O_5 [count of number of nitrogen atoms (single double or triple bonded)] contributing to anti inflammatory activity. Similar inferences were drawn for the other activities also.

Keywords: 2D-QSAR, MLR, Antitumor agents, 3', 4', 5'-trimethoxychalcone.

INTRODUCTION

Chalcones constitute an important group of natural products and serve as precursors for the synthesis of different classes of flavonoids, which are common substances in plants. Chalcones are open-chain flavonoids in which two aromatic rings are joined by a three carbon a,b-unsaturated carbonyl system (1,3-diphenyl-2-propen-1-ones)[1]. Chalcone derivatives have received a great deal of attention due to their relatively simple structures, and wide variety of

pharmacological activities reported for these compounds include anti-inflammatory[2], anti-bacterial[1], anti-fungal[3-5], and anti-tumor activities[6-9]. These activities are largely attributed due to the a,b-unsaturated ketone moiety. Introduction of various substituents into the two aryl rings is also a subject of interest because it leads to useful structure–activity relationship (SAR) conclusions and thus helps to synthesize pharmacologically active chalcones[6].

In recent years, noteworthy advancement has been made by computational chemistry which led new challenges to drug discovery. Quantitative structure activity relationship (QSAR) which has become an reputable tool for establishing quantitative relationship between biological activity and physicochemical properties of the compounds in a series using various statistical methods (linear regression and non- linear regression analysis) and it helps to calculate the biological activities of newly designed analogues contributing to the drug discovery process.

The core idea of the present study is the search for novel 3', 4', 5'-trimethoxychalcone analogues that would show a promise to become useful as inhibitors of nitric oxide production and tumor cell proliferation**.**

A series of 3', 4', 5'-trimethoxychalcone analogues which were reported[16] are chosen for QSAR study in order to establish quantitative relationship between physiochemical properties and biological activities of the compounds using QSARpro software (VlifeScience)[17].

MATERIALS AND METHODS

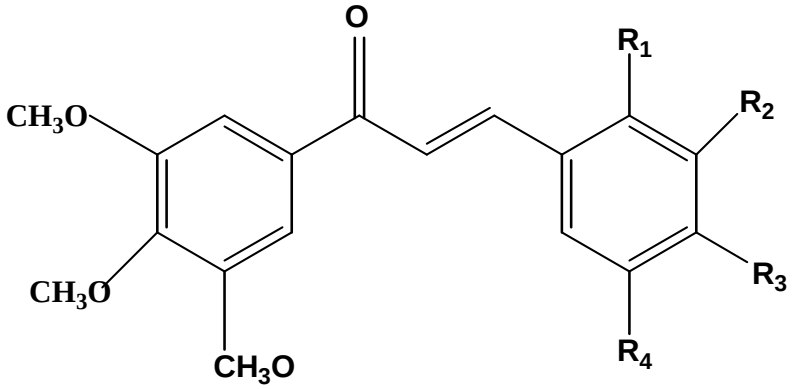
Data set:

In the present study a data set of 3',4' ,5'-trimethoxychalcone analogues as inhibitors of nitric acid production and tumor cell proliferative agents (23 molecules) has been taken from the literature for QSAR studies (**Table 1**). The synthesis and determination of the activity of these compounds have already been reported in literature [16]. The biological data have been converted to logarithmic scale (pIC50) in mathematical operation mode of software to reduce skewness of data set and then used for subsequent QSAR analysis as dependent variables.

Molecular modeling:

Molecular modeling and Multiple linear regression (MLS) were studies were performed on HP computer having genuine Intel Pentium Processor with 32-bit Windows 7 as operating system using the software QSARpro (VlifeScience). Structures were drawn using the 2D draw application and converted to 3D structures. Structures were optimized by energy minimization and geometry optimization was done using Merk molecular force field method and Modified Qeq Charge with 10000 as maximum number of cycles, 0.01 as convergence criteria (root mean square gradient) and 1.0 as constant (medium's dielectric constant which is 1 for in vacuo) in dielectric properties. The default values of 30.0 and 10.0 Kcal/mol were used for electrostatic and steric energy cutoff. Complete geometry optimization was performed taking the most extended conformations as starting geometries. The basis of energy minimization is that the drug binds to effectors/receptor in the most stable form i.e. minimum energy state form.

Table 1: 3',4',5'-Trimethoxychalcone analogues with activity



COMPOUND	R1	R2	R3	R4	ACTIVITY		
					NO*	HEP G2**	COLON 205***
1	OMe	OMe	H	H	2.4±0.3	11.5±1.4	13.2± 2.0
2	H	OMe	OMe	H	4.5±0.5	20.3±1.9	>100
3	H	OMe	OMe	OMe	2.8±0.3	19.6±2.1	19.2±1.7
4	H	H	$N(CH3)_2$	H	27±4.0	30.0±3.8	75±5.5
5	OMe	H	OMe	OMe	4.6±1.1	16.1±2.9	18.6±3.5
6	H	H	OH	H	5.0±0.7	16.0±1.8	29.7±4.3
7	H	OMe	OH	H	0.3±0.1	>100	>100
8	OMe	H	OMe	H	>50	>100	82.5±5.0
9	OH	H	H	OH	>50	>100	>100
10	OH	OMe	H	H	3.0±0.2	13.50±0.8	13.0±1.2
11	H	OH	OH	H	1.5±0.4	14.9±2.1	19.8±2.7
12	H	OH	H	H	13.5±1.2	94.0±4.6	49±3.2
13	H	H	OMe	H	7.6±1.9	49.5±2.6	46.4±5.2
14	H	OH	OMe	H	1.3±0.3	10.6±1.3	11.2±2.4
15	H	OMe	H	H	0.7±0.1	1.8±0.3	2.2±0.7
16	H	H	F	H	17.6±3.6	22.0±2.5	25.7±2.8
17	H	H	Br	H	20.0±4.1	80.0±4.9	22.5±3.2
18	H	H	NO2	H	30.0±4.5	>100	43±5.0
19	H	H	Me	H	4.5±1.2	52.1±4.1	41.5±3.2
20	OH	H	H	NO2	4.4±0.7	>100	76±4.7
21	CHO	H	H	H	27.0±2.5	>100	88±5.1
22	H	CHO	H	H	>50	>100	22.5±2.3
23	H	H	H	CHO	>50	>100	68.8±5.4

Number of descriptors was calculated after optimization or minimization of the energy of the data set molecules. Various types of physicochemical descriptors were calculated: Individual (Molecular weight, H-Acceptor count, H-Donor count, XlogP, slogP, SMR, polarisablity, etc.), retention index (Chi), atomic valence connectivity index (ChiV), Path count, Chi chain, ChiV chain, Chain PathCount, Cluster, Pathcluster, Kappa, Element count (H, N, C, S count etc.), and Polar surface area. More than 200 alignment independent descriptors were also calculated using the following attributes. A few examples are T_2_O_7, T_N_N_5, T_2_2_6, T_C_O_1, T_O_Cl_5 etc.

Structural descriptors	Selected Attributes
*Topological	2
Range	T (Any)
Range	C
Min-0	N
Max- 7	O
	F
	Cl

Generation of training and test set of compounds:
In order to evaluate the QSAR model, data set was divided into training and test set using Random data selection and Manual data selection method. Training set is used to develop the QSAR model for which biological activity data are known. Test set is used to challenge the QSAR model developed based on the training set to assess the predictive power of the model which is not included in model generation.

Random data selection method: the data was selected randomly entering the percentage of training set molecules to be selected. The percentage value was adjusted subsequently in order to get the different sets of training and test molecule. This is based on trial and error method to get the desired test set molecules.

Manual data selection method: Data set is divided manually into training and test sets on the basis of the result obtained in random data selection method.

Multiple linear regressions
MLR method was used for model generation. The multiple linear regression models and its estimation using ordinary least squares (OLS) is doubtless the most widely used tool. The multiple linear regression models assumes a linear (in parameters) relationship between a dependent variable yi and a set of explanatory variables xi = (xi0; xi1;:::; xiK). Xik is also called an independent variable, a covariate or a regressor. The first regressor xi0 = 1 is a constant unless otherwise specified.

Table 2a: Results of MLR analysis using RANDOM data selection method for NO inhibition activity

TRIAL	TEST SET MOLECULES	r^2	q^2	Pred_r^2	r^2 se	$^{q2\ se}$	Pred_r^2 se	F TEST
1	**10,2,17,23,3,7,9**	**0.9751**	**0.7518**	**0.4496**	**0.1221**	**0.3854**	**0.5962**	**78.34**
2	10,2,23,7,9	0.9331	0.6639	0.4678	0.1881	0.4205	0.6835	33.4856
3	10,2,3,7,9	0.9481	0.7516	0.4471	0.1714	0.3748	0.6682	43.8134
4	10,2,23,3,7	0.9711	0.8808	0.3448	0.1278	0.2597	0.7223	80.7281
5	2,23,3,7,9	0.9524	0.7578	0.4535	0.1584	0.3573	0.6951	48.005
6	10,17,2, 23,3,7	0.9775	0.8890	0.3514	0.1165	0.2590	0.6543	95.7140

Table 2b: Results of MLR analysis using MANUAL data selection method for NO inhibition activity

TRIAL	%	TEST SET MOLECULES	r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2 se	F TEST
1	66	1,13,17,2,23,6,7,9	0.9218	0.6061	0.5406	0.2291	0.5144	0.5002	21.2304
2	66	12,14,18,2,20,6,7,9	0.8882	0.6332	0.5037	0.2607	0.4728	0.5037	19.8669
3	75	18,20,22,23,4,9	0.7724	O.5747	0.4044	0.3424	0.4681	0.7664	10.1828
4	75	10,2,23,3,7,9	0.9683	0.7529	0.4343	0.1333	0.3721	0.6442	67.1937
5	75	15,19,22,23,7,9	0.9135	0.7667	0.2107	0.1746	0.2668	0.9726	23.2265
6	80	13,2,3,7,9	0.9394	0.7559	0.4154	0.1878	0.3769	0.6577	37.2170
7	85	17,21,22,9	0.8136	0.6994	0.4884	0.3156	0.4042	0.6971	15.2765

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable *x* is associated with a value of the dependent variable *y*. Here all the calculated descriptors were considered as independent variable and biological activity as dependent variable.

Table 3a : Results of MLR analysis using RANDOM data selection method for antitumor hepatic G2 activity.

TRIAL	%	TEST SET MOLECULES	r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F TEST
1	85	18,2,22,9	0.4759	0.2099	0.4151	0.3698	0.4550	0.3933	7.2637
2	80	10,14, 15, 3,9	0.8749	0.7197	0.0674	0.1527	0.2286	0.8307	16.7785
3	75	13,16, 18,22, 6, 9	0.9176	0.8064	-2.9950	0.1744	0.2674	0.8424	24.4929
4	70	1,12,17,19,23,3, 9	0.7253	0.5185	-0.0922	0.2825	0.3766	0.5460	10.5596

Table 3b: Results of MLR analysis using MANUAL data selection method for antitumor hepatic G2 activity

TRIAL	TEST SET MOLECULES	r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2 se	F TEST
1	11,15,16,19,3,9	0.8813	0.7402	0.0773	0.1668	0.2467	0.6800	16.3281
2	1,11,13,14,15,22,7,9	0.9158	0.7685	0.2048	0.2048	0.2162	0.6056	19.5727
3	**13,18, 23, 9**	**0.7737**	**0.5955**	**0.5270**	**0.2692**	**0.3599**	**0.3698**	**8.8874**

Table 4a : Results of MLR analysis using RANDOM data selection method for antitumor COLON 205 activity

TRIAL	%	TEST SET MOLECULES	r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F TEST
1	70	1,12,13,14,18,6, 15	0.9108	0.7712	0.2789	0.1161	0.1906	0.5264	20.4196
2	70	1,12 14,15,18,19,6	0.9102	0.7730	0.2754	0.1190	0.1899	0.5294	20.2821
3	75	1,14, 15, 21,22,6	0.9041	0.7314	0.1972	0.1111	0.1860	0.6329	20.7381
4	75	14, 15, 2,21,6,9	0.7123	0.4558	0.1578	0.1739	0.2391	0.6189	10.7281
5	85	13,16,2,9	0.7999	0.4806	-1.7280	0.2195	0.3519	0.7320	10.2666

Table 4b: Results of MLR analysis using MANUAL data selection method for antitumor COLON 205 activity

TRIAL	TEST SET MOLECULES	r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2 se	F TEST
1	1, 12, 14, 15, 21,22,6	0.9707	0.7212	0.1922	0.1161	0.1985	0.5780	18.9758
2	1,14,15,18.21,22,6	0.8896	0.6650	0.2356	0.1251	0.2178	0.5640	16.1095
3	1,14, 15, 18, 22, 6	0.8686	0.7325	0.2486	0.1340	0.1913	0.6139	14.5456
4	**1,12, 14, 15, 18, 6**	**0.9067**	**0.7759**	**0.2928**	**0.1161**	**0.1799**	**0.5721**	**21.3786**
5	1, 12, 14, 15, 18, 20,6	0.9031	0.7497	0.2872	0.1216	0.1954	0.5235	18.6304
6	1, 12, 14, 15, 18, 6, 8	0.9178	0.7887	0.2925	0.1112	0.1783	0.5227	22.3390

RESULTS AND DISCUSSION

Selected data set 3', 4', 5'-trimethoxychalcone analogues were subjected to multiple linear regression analysis method for model building. Result of MLR analysis using random data selection and manual data selection methods is shown in Table 2a, 2b, 3a, 3b, 4a and 4b respectively for all three activities. The statistically significant model obtained is shown in Table 5a, 5b and 5c.

Multiple linear regression analysis (MLR) in conjunction with stepwise (SW) forward-backward was applied for building QSAR models. The resulting models were validated by leave-one-out cross-validation procedures to check their predictivity and robustness.

Data fitness plot for model 1(best model) is shown in Fig. 2a. Result of the observed and predicted biological activity for the training and test compounds for the Models is shown in Table 6a, 6b and 6c respectively. The plot of observed vs predicted activity of training and test sets for models is shown in Fig.3**.** From the plot it can be seen that model is able to predict the activity of training set quite well (all points are close to regression line) as well as external.

In the above QSAR equations, n is the number of molecules (Training set) used to derive the QSAR model, r^2 is the squared correlation coefficient, q^2 is the cross-validated correlation coefficient, pred_r^2 is the predicted correlation coefficient for the external test set, *F* is the Fisher ratio [20] which 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. r^2 se, q^2 se and pred_r^2se are the standard errors terms for r^2, q^2 and pred_r^2 (smaller is better).

All compound from the data for MLR resulted in the generated model with improved statistical significance and predictive ability, this generated model can be developed for the series. All these models were screened on the basis of $q^2 > 0.7$ and the intercept to best fit line. Hence the best statistical results are reported in Table 2 and actual activity and predicted activity of best model for all the activities are shown in Table 3. The plots of cross-validated calculated activity and the corresponding residuals against the experimental values are represented in Fig. 3a and 3b, respectively. The residual plot shows the relatively uniform distribution of data around the zero line.

From Table 5a, for NO inhibition activity model 1 explains 97.51 % (r^2= 0.9751) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 75.18% and 40.96% respectively. The F-test = 78.34 shows the statistical significance of 99.98% of the model. In addition randomization test shows confidence of 99.9% that the generated model is not random and hence it can be selected as the QSAR model (**model 1**). Model 2 explains 96.83% (r^2= 0.9683) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 75.29% and 43.43% respectively. Model 3 explains 95.24% (r^2= 0.9524) of the total variance in the training set as well as it has internal (q^2) and external (pred_r2) predictive ability of 75.78% and 45.35% respectively. Model 4 explains 94.81% (r^2= 0.9481) of the total variance in the training set as well as it has internal (q^2) and external (pred_ r^2) predictive ability of 75.16% and 46.78% respectively. Model 5 explains 93.94% (r^2= 0.9394) of the total variance in the training set as well as it has internal (q^2) and external (pred_ r^2) predictive ability of 75.59% and 41.54% respectively.

From Table 5b, for Antitumor Hep G2 activity model 1 explains 77.37 % (r^2= 0.7737) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 59.55% and 52.70% respectively. The F-test = 8.88 shows the statistical significance of the model. Model 2 explains 91.58% (r^2= 0.9158) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 76.35% and 20.48% respectively. From both of these models first model is selected as QSAR model. **(Model 2)**

From Table 5c, for Antitumor Colon 205 activity model 1 explains 91.78 % (r^2= 0.9178) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 78.87% and 29.25% respectively. The F-test = 22.3390 shows the statistical significance of the model. Model 2 explains 90.67% (r^2= 0.9067) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 77.59% and 29.28% respectively. Here the second model is selected as QSAR model. (**Model 3**)

Table 6a, 6b and 6c represents the actual and predicted biological activity for the models.

Contribution chart for models is represented in Figure-11, 1b and 1c. Data fitness plot for models is shown in Figure-2a, 2b and 2c. The plot of observed vs. predicted activity provides an idea about how well the model was trained and how well it predicts the activity of the external test set. From the plot (Figure-3a-3f) it can be seen that the model is able to predict the activity of the training set quiet well as well as external test set, providing confidence of the model.

Table 5a: Statistical significant models generated for NO inhibition activity

MODEL	TRIAL TYPE	TEST SET MOLECULE	EQUATION
1	Manual	10,2,17,23,3,7,9	pIC_{50} = - 0.0193(± 0.0001)Quadrupole2 - 0.2807(± 0.0062) SK Most Hydrophobic Hydrophilic Distance+ 0.2621(± 0.0282) XK Most Hydrophobic Hydrophilic Distance + 0.8545(± 0.1306) T_O_O_5 0.6429(± 0.1268) T_N_O_5 - 0.1825 n = 16, Degree of freedom = 10, r^2 = 0.9751 ,q^2 = 0.7518,F test = 78.3483, r^2 se = 0.1221, q^2 se = 0.3854, pred_r^2 = 0.4496, pred_r^2 se = 0.5962, Alpha Rand R^2 = 0.0003; Alpha Rand Q^2 = 0.01; Alpha Rand Pred R^2 = 0.0000
2	Random	10,2,23,3,7,9	pIC_{50} = - 0.0198(± 0.0001) Quadrupole2 - 0.2788(± 0.0068) SK Most Hydrop Hydrophilic Distance + 0.2610(± 0.0309) XK Most Hydrophobic Hydrophilic Dista 0.8353(± 0.1422) T_O_O_5 - 0.6651(± 0.1382) T_N_O_5 -0.1438 n = 17 , Degree of freedom = 11 , r^2 = 0.9683 , q^2 = 0.7529 F test = 67.1937 , r^2 se = 0.1333 , q^2 se = 0.3721 ,pred_r = 0.4343 pred_r^2 se = 0.6492 , Alpha Rand R^2 = 0.01; Alpha Rand Q^2 = 0.001; Alpha Rand Pred R^2 = 0.00000
3	Manual	2,23,3,7,9	pIC_{50} = - 0.0198(± 0.0001) Quadrupole2 - 0.2567(± 0.0078) SK Most Hydrophobic Hydrophilic Distance + 0.2490(± 0.0376) XK Most Hydrophobic Hydrophilic Distance + 0.7410(± 0.1673) T_O_O_5 - 0.6752(± 0.1639) T_N_O_5 - 0.0475 n = 18, Degree of freedom = 12, r^2 = 0.9524 , q^2 = 0.7578, F test = 48.0005, r^2 se = 0.1584 ,q^2 se = 0.3573, pred_r^2 = 0.4535 pred_r^2 se = 0.6951, Alpha Rand R^2 = 0.00; Alpha Rand Q^2 = 0.01; Alpha Rand Pred R^2 = 0.00000
4	Manual	10,2,3,7,9	pIC_{50} = - 0.0188(± 0.0001)Quadrupole2 - 0.2566(± 0.0083)SK Most Hydrophobic Hydrophilic Distance + 0.2329(± 0.0375) XK Most Hydrophobic Hydrophilic Distance + 0.7706(± 0.1821) T_O_O_5 - 0.6239(± 0.1775) T_N_O_5 -0.0548 n = 18, Degree of freedom = 12 ,r^2 = 0.9481, q^2 = 0.7516, F test = 43.8134 ,r^2 se = 0.1714 ,q^2 se = 0.3748, pred_r^2 = 0.4471 pred_r^2 se = 0.6682, Alpha Rand R^2 = 0.00003; Alpha Rand Q^2 = 0.00008; Alpha Rand Pred R^2 = 0.00000
5	Random	13,2 3, 7 ,9	pIC_{50} = - 0.0186(± 0.0001) Quadrupole2 - 0.2441(± 0.0097) SK Most Hydrophobic Hydrophilic Distance + 0.2206(± 0.0420) XK Most Hydrophobic Hydrophilic Distance + 0.6984(± 0.1981) T_O_O_5 - 0.6443(± 0.1946) T_N_O_5 + 0.0837 n = 18, Degree of freedom = 12, r2 = 0.9394, q2 = 0.7559, F test = 37.2170, r2 se = 0.1878, q2 se = 0.3769, pred_r2 = 0.4154 pred_r2se = 0.6577, Alpha Rand R^2 = 0.000; Alpha Rand Q^2 = 0.00025; Alpha Rand Pred R^2 = 0.01

Table 5b: Statistical significant models generated for ANTITUMOR HEP G2 ACTIVITY

MODEL	TRIAL TYPE	TEST SET MOLECULE	EQUATION
1	Manual	13, 18 , 23, 9	pIC_{50} = - 0.0164(± 0.0000) Quadrupole2 - 0.1270(± 0.0035) SK Most Hydrophobic Hydrophilic Distance + 0.2809(± 0.0021) T_2_C_1+ 0.2153(± 0.0016) Y comp Dipole - 18.7436(± 6.9159) SA Most Hydrophobic -1.3272 n = 19, Degree of freedom = 13, r^2 = 0.7737, q^2 = 0.5955, F test = 8.8874 , r^2 se = 0.2692 , q^2 se = 0.3599, pred_r^2 = 0.5270,pred_r^2 se = 0.3698, Alpha Rand R^2 = 0.01; Alpha Rand Q^2 = 0.05; Alpha Rand, Pred R^2 = 0.000
2	Manual	1,11,13,14,15,22,7,9	pIC_{50} = - 0.0740(± 0.0001) SK Most Hydrophobic Hydrophilic Distance + 0.083 (± 0.0002) T_2_2_2 - 0.0431(± 0.0000) SA Most Hydrophobic Hydrophilic Distance + 25.4922(± 7.8831) Average -ve Potential- 86.7603(± 33.9704) XA Most Hydrophilic -25.6308 n = 15 , Degree of freedom = 9 , r^2 = 0.9158 , q^2 = 0.7685 , F test = 19.5727, r^2 se = 0.1304 , q^2 se = 0.2162 , pred_r^2 = 0.2048 , pred_r^2 se = 0.6056, Alpha Rand R^2 = 0.01; Alpha Rand Q^2 = 0.01; Alpha Rand Pred R^2 = 0.05

Table 5c: Statistical significant models generated for ANTITUMOR COLON 205 ACTIVITY

MODEL	TRAIL TYPE	TEST SET MOLECULE	EQUATION
1	Manual	1,12,14,14, 15, 18, 6, 8	pIC_{50} = - 40.5200(± 3.4004) SA Most Hydrophobic + 0.1195(± 0.0001) T_2_C_2 - 15.4446(± 2.8415) XA Average - 0.0710(± 0.0003) SK Most Hydrophobic Hydrophilic Distance + 0.3013(± 0.0052) T_O_O_5 + 11.2367 n = 16, Degree of freedom = 10 , r^2 = 0.9178,q^2 = 0.7887, F test = 22.3390, r^2 se = 0.1112 ,q^2 se = 0.1783,pred_r^2 = 0.2925, pred_r^2 se = 0.5227, , Alpha Rand R^2 = 0.01; Alpha Rand Q^2 = 0.001; Alpha Rand Pred R^2 = 0.000
2	Manual	1,12,14,15,18, 6	pIC_{50} = - 40.4423(± 3.5490) SAMostHydrophobic + 0.1177(± 0.0001) T_2_C_2 - 16.3424(± 2.9393) XAAverage - 0.0779(± 0.0004) SK Most Hydrophobic Hydrophilic Distance + 0.2865(± 0.0055) T_O_O_5 + 11.4760 n = 17, Degree of freedom = 11, r^2 = 0.9067, q^2 = 0.7759 F test = 21.3786, r^2 se = 0.1161 , q^2 se = 0.1799, pred_r^2 = 0.2928 , pred_r^2 se = 0.5721, Alpha Rand R^2 = 0.00005; Alpha Rand Q^2 = 0.001; Alpha Rand Pred R^2 = 0.0000

Table 6a - Actual and Predicted data of NO inhibition activity model 1

TRAINING SET		TEST SET	
ACTUAL ACTIVITY	PREDICTED ACTIVITY	ACTUAL ACTIVITY	PREDICTED ACTIVITY
0.3802	0.3622		
0.1761	0.1549		
1.1303	1.1138		
0.8808	1.0212		
0.1139	-0.1351	0.4771	0.0890
-0.5229	-0.3158	1.3010	1.0742
1.2455	1.2396	0.6532	-0.1972
1.4771	1.4855	1.6990	2.1389
0.6532	0.6407	0.4472	1.1038
0.6435	0.6435	-0.5229	0.2403
1.4314	1.4938	1.6990	1.6724
1.6990	1.6725		
1.4314	1.3109		
0.6628	0.6628		
0.6990	0.7912		
1.6990	1.6876		

Table 6b - Actual and Predicted data of antitumor hep g2 activity model 2

TRAINING SET		TEST SET	
ACTUAL ACTIVITY	**PREDICTED ACTIVITY**	**ACTUAL ACTIVITY**	**PREDICTED ACTIVITY**
1.2068	1.0684		
1.9031	1.5589		
1.0253	0.9608		
1.3424	1.5734		
1.2041	1.4424		
2.000	1.9987		
2.000	1.6115		
1.607	1.3187		
1.4771	1.7530	1.6946	1.5339
2.000	1.8463	2.0000	1.8957
2.000	2.0238	2.0000	2.3564
0.2553	0.3277	2.0000	1.5036
2.0000	1.9494		
1.1303	1.4467		
1.3075	1.4234		
1.1732	0.9464		
1.9731	1.5644		
1.7160	1.8200		
1.2923	1.4515		

Table 6c - Actual and Predicted data of antitumor COLON 205 activity model 3

TRAINING SET		TEST SET	
ACTUAL ACTIVITY	**PREDICTED ACTIVITY**	**ACTUAL ACTIVITY**	**PREDICTED ACTIVITY**
1.1139	1.1884		
1.2967	1.1602		
1.6665	1.7143		
1.4099	1.4518		
1.3522	1.3055		
1.6181	1.6872	1.1206	2.0757
2.000	1.8113	1.6902	1.6253
1.8808	1.8771	1.0492	1.3053
1.9445	1.7872	0.3424	0.9591
1.3522	1.4407	1.6335	2.0452
1.8376	1.8601	1.4728	1.7514
1.2833	1.3766	1.9165	1.7436
1.8751	1.8650		
1.2695	1.2687		
2.0000	2.1051		
2.0000	2.0008		

Interpretation of the models:

Among the five significant models generated for NO inhibition activity (Table-5a), model 1 is the most significant one. The equation explains 97.51% (r^2 = 0.9751) of the total variance in the training set and has an internal (q^2) and external (pred_r^2) predictive ability of ~75% and ~45% respectively. The F test shows the statistical significance of 99.99 % of the model which means that probability of failure of the model is 1 in 10000. In addition, the randomization test shows confidence of 99.9999 (Alpha Rand Pred R^2 = 0.00000) that the generated model is not random and hence chosen as the QSAR model (**model 1**).

In the QSAR model 1, the negative coefficient value of Quadrupole2 [this descriptor signifies magnitude of second tensor of quadrupole moments] on the biological activity indicated that lower value leads to better anti inflammatory activity whereas higher value leads to decrease anti inflammatory activity . Negative coefficient value of SK Most Hydrophobic Hydrophilic

Distance [This descriptor signifies distance between most hydrophobic and hydrophilic point on the vdW surface. (By Kellog Method using Slogp)] on biological activity indicated that shorter the distance leads to good anti inflammatory activity while longer the distance leads to reduced anti inflammatory activity. The positive coefficient value of XK Most Hydrophobic Hydrophilic Distance [This descriptor signifies distance between most hydrophobic and hydrophilic point on the vdW surface. (By Kellog Method using Xlogp)] on biological value indicates that longer the distance leads to better activity and shorter the distance leads to lower activity. The positive coefficient of T_O_O_5 [This is the count of number of oxygen atoms (single double or triple bonded) separated from any other oxygen atom (single double or triple bonded) by 5 bonds in a molecule] on biological activity indicates that higher the value leads to better activity and lower the value leads to poor activity. The negative coefficient of T_N_O_5 [This is the count of number of nitrogen atoms (single double or triple bonded) separated from oxygen atom by 5 bond distance in a molecul] on biological activity indicates that lower value leads to better anti inflammatory activity and vice versa.

Contribution chart for **model 1** reveals that the descriptors XK Most hydrophobic hydrophilic Distance, T_O_O_5 contributing 16.94%, 14.22% respectively. Three more descriptors Quardrupole 2, SK Most Hydrophobic Hydrophilic Distance and T_N_O_5 and are contributing inversely 26.23%, 31.92% and 10.70% respectively to biological activity.

The observed vs. predicted activity provides an idea about how well the model was trained and how well it predicts the activity of the external test set. From the plot it can be seen that model is able to predict the activity of training set quite well (all points are close to the regression line) as well as external test set providing confidence in the predictive ability of the model.

Among the two significant models generated for Antitumor Hep G2 activity (Table-5b), model 1 is the most significant one. The equation explains 77.37% (r^2 = 0.9751) of the total variance in the training set and has an internal (q^2) and external (pred_r^2) predictive ability of ~60% and ~53% respectively. The F test shows the statistical significance of 99.99 % of the model which means that probability of failure of the model is 1 in 10000. In addition, the randomization test shows confidence of 99.9999 (Alpha Rand Pred R^2 = 0.00000) that the generated model is not random and hence chosen as the QSAR model **(mode**l **2**).

In the QSAR **model 2**, the negative coefficient value of Quadrupole2[this descriptor signifies magnitude of second tensor of quadrupole moments] on the biological activity indicated that lower value leads to better anti tumor Hep G2 activity whereas higher value leads to decrease anti tumor activity . Negative coefficient value of SK Most Hydrophobic Hydrophilic Distance [This descriptor signifies distance between most hydrophobic and hydrophilic point on the vdW surface. (By Kellog Method using Slogp)] on the biological activity indicated that shorter the distance leads to good anti tumor activity while longer the distance leads to reduced anti tumor activity. The positive coefficient value of T_2_C_1 [This is the count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from carbon atom by 1 bond in a molecule] on the biological value indicates that higher value leads to better antitumor activity and lower value leads to lower activity. The positive coefficient of YcompDipole [This descriptor signifies the y component of the dipole moment (external coordinates)] on biological activity indicates that higher the value leads to better activity and lower the value leads to poor

activity. The negative coefficient of SAMostHydrophobic [Most hydrophobic value on the vdW surface (By Audry Method using Slogp)] on biological activity indicates that lower value leads to better anti inflammatory activity and vice versa.

Contribution chart for **model 2** reveals that the descriptors T_2_C_1, YcompDipole contributing 22.23%, 21.21% respectively. Three more descriptors Quardrupole 2, SK Most Hydrophobic Hydrophilic Distance, SA Most Hydrophobic, are contributing inversely 23.39%, 318.99% and 14.18% respectively to biological activity.

Among the two significant models generated for Antitumor colon 205 activity (Table-5c), model 2 is the most significant one. The equation explains 90.67% (r^2 = 0.9067) of the total variance in the training set and has an internal (q^2) and external (pred_r^2) predictive ability of ~77% and ~30% respectively. The F test shows the statistical significance of 99.99 % of the model which means that probability of failure of the model is 1 in 10000. In addition, the randomization test shows confidence of 99.9999 (Alpha Rand Pred R^2 = 0.00000) that the generated model is not random and hence chosen as the QSAR model (**model 3)**.

In the QSAR **model 3**, the negative coefficient value of SAMostHydrophobic [Most hydrophobic value on the vdW surface. (By Audry Method using Slogp)] on the biological activity indicated that lower value leads to better anti tumor colon 205 activity whereas higher value leads to decrease anti tumor activity . The positive coefficient of T_2_C_2 [This is the count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from carbon atom by 2 bonds in a molecule] on the biological value indicates that higher value leads to better antitumor activity and lower value leads to lower activity. The negative coefficient of XAAverage [Average hydophobicity function value (By Audry Method using Xlogp)] on the biological activity indicated that lower value leads to better anti tumor colon 205 activity whereas higher value leads to decrease anti tumor activity. Negative coefficient value of SK Most Hydrophobic Hydrophilic Distance [This descriptor signifies distance between most hydrophobic and hydrophilic point on the vdW surface. (By Kellog Method using Slogp)] on the biological activity indicated that shorter the distance leads to good anti tumor activity while longer the distance leads to reduced anti tumor activity. The positive coefficient of T_O_O_5 [This is the count of number of oxygen atoms (single double or triple bonded) separated from any other oxygen atom (single double or triple bonded) by 5 bonds in a molecule] on biological activity indicates that higher the value leads to better activity and lower the value leads to poor activity.

Contribution chart for **model 3** reveals that the descriptors T_2_C_2, T_O_O_5 contributing 22.58%, 10.47% respectively. Three more descriptors SAMostHydrophobic , SK Most Hydrophobic Hydrophilic Distance, XAAverage, are contributing inversely 36.55%, 13.78% and 16.62% respectively to biological activity.

Figure 1a- Contribution chart of descriptor for model 1

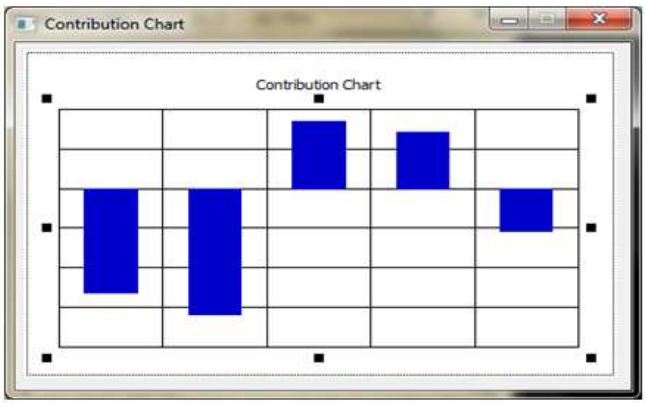


Figure 1b- Contribution chart of descriptor for model 2

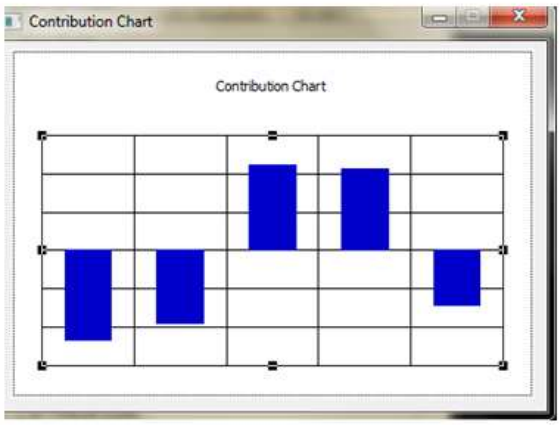


Figure 1c - Contribution chart of descriptor for model 3

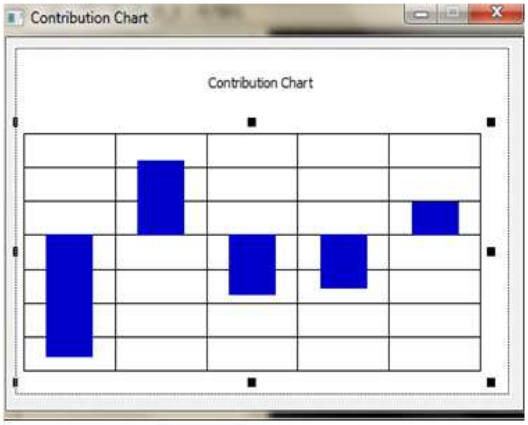


Figure2a: Data fitness plot for model 1

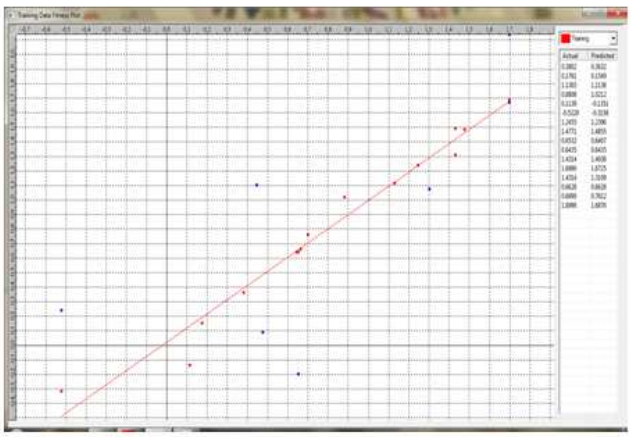


Figure 2b: Data fitness plot for model 2

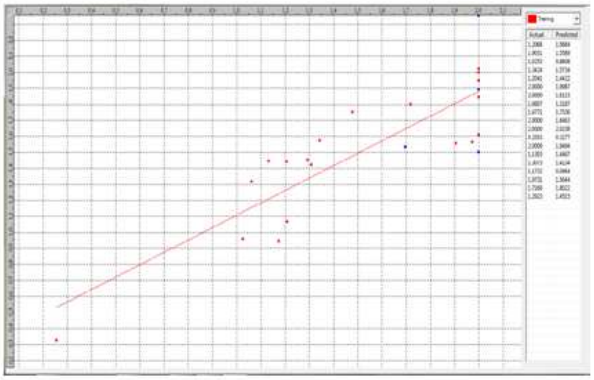


Figure 2c: Data fitness plot for model 3

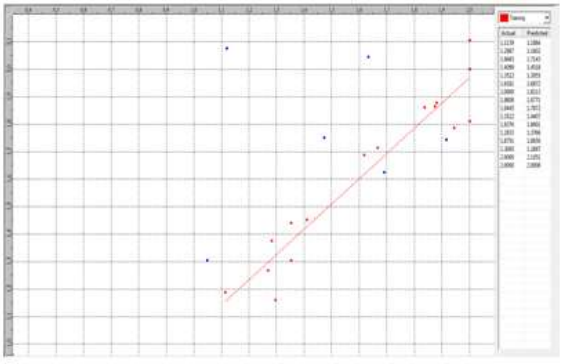


Figure 3a: Graph between actual and predicted biological activity of TRAINING SET for Model-1

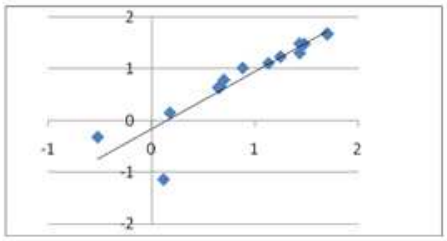


Figure 3b: Graph between actual and predicted biological activity of TEST SET for Model-1

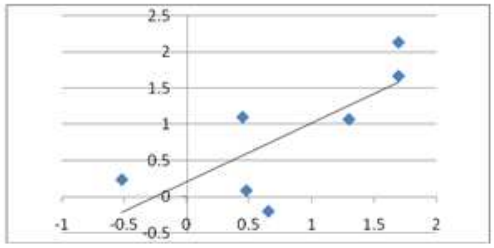


Figure 3c: Graph between actual and predicted biological activity of TRAINING SET for Model-2

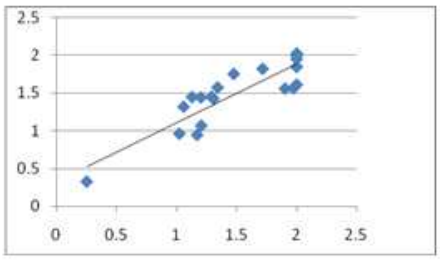


Figure 3d: Graph between actual and predicted biological activity of TEST SET for Model-2

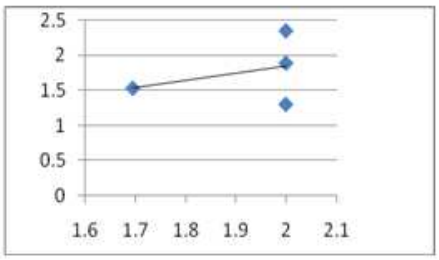


Figure 3e: Graph between actual and predicted biological activity of TRAINING SET for Model-3

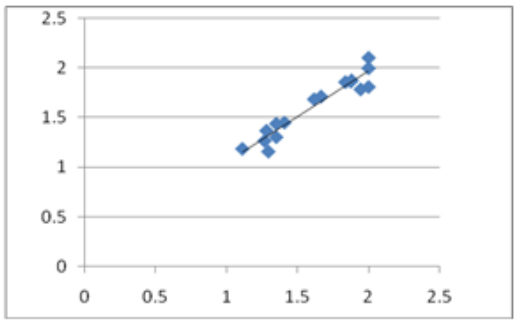
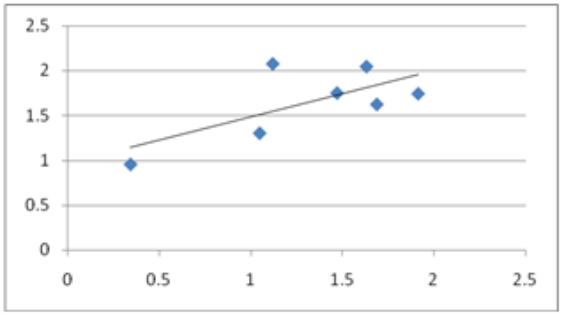


Figure 3f: Graph between actual and predicted biological activity of TEST SET for Model-3



CONCLUSION

Two dimensional quantitative structure activity relationship (2D QSAR) studies by means of multiple linear regression (MLR) method was performed on a series of 3', 4', 5'-trimethoxychalcone analogues as anti inflammatory and antitumor agents using software QSARpro (VLifeScience). This study was performed with 23 compounds (data set) using random data selection and manual selection methods for the division of the data set into training and test set. MLR methodology with stepwise (SW) forward-backward variable selection method

was used for building the QSAR models. Statistically significant QSAR models were generated. Among them most significant model is the one which is developed for NO inhibitory activity i.e. anti inflammatory activity having squared correlation coefficient (r^2), cross validated correlation coefficient (q^2) and predictive correlation coefficient (pred_r^2) 0.9752, 0.7518 and 0.4496 respectively. From all the models generated for all three activities, models for anti inflammatory activity have more predictive correlation coefficient (pred_r^2). It means the models for No inhibition activity has more predicting ability hence they are proved as the best models.

From this it was concluded that the compounds have promising anti inflammatory activity as compared to the antitumor activity.

Structural information obtained can be used for predicting the activity of the newer compounds with more potent activity.

Acknowledgement
Authors are thankful to the principal Dr. P. S. Gide and the trustee Bhujbal Knowledge City for providing the necessary facilities to carry out the research work.

REFERENCES

[1] H. P Avila, E. F Smania, F. D. Monache, A.Smania, *Jr. Bioorg. Med. Chem.***2008**, 16, 9790.
[2] J. H. Cheng; C. F. Hung; S. C.Yang; J. P.Wang; S. J.Won; C. N.Lin*; Bioorg. Med.Chem.* **2008**, 16, 7270.
[3] 3. M. Sortino; P. Delgado; S. Juarez; J. Quiroga; R. Abonia; B. Insuasty; M. Nogueras; L. Rodero; F. M. Garibotto;, R. D. Enriz; S. A. Zacchino; *Bioorg. Med. Chem*. **2007**, 15, 484.
[4] M. L. Y. Vargas; M. V. Castelli; V. V. Kouznetsov; G. J. M. Urbina; S. N. Lopez; M. Sortino; R. D. Enriz; J. C. Ribas; S. Zacchino; *Bioorg. Med. Chem*, **2003**, 11, 1531
[5] S. N. Lopez; M. V. Castelli; S. A. Zacchino; J. N. Dominguez; G. Lobo; J. Charris-Charris; J. C. Cortes; J. C. Ribas; C. Devia; A. M. Rodriguez; R. D. Enriz; *Bioorg. Med. Chem*. **2001**, 8, 1999.
[6] A. M. Katsori; D. Hadjipavlou-Litina; *Curr. Med. Chem*, **2009**, 16, 1062.
[7] G. Achanta; A. Modzelewska; L. Feng; S. R. Khan; P. Huang; *Mol. Pharmacol.* **2006,** 70,426.
[8] A. Modzelewska; C. Pettit, G. Achanta; N. E. Davidson; P. Huang; S. R. Khan; *Bioorg. Med. Chem.* **2006**, 14 3491.
[9] S. K. . Kumar; E. Hager; C. Pettit; H. Gurulingappa; N. E. Davidson; S. R. Khan; *J. Med. Chem*, **2003**, 46,2813.
[10] Y. K. Rao; S. H. Fang; Y. M. Tzeng; *Bioorg. Med. Chem*, **2004**, 12, 2679.
[11] Y.K. Rao;S. H. Fang; Y. M. Tzeng; *Bioorg. Med. Chem*, **2005**, 13, 6850.
[12] D. Alloatti; G. Giannini W. Cabri; I. Lustrati; M. Marzi; A. Ciacci; G. Gallo; M. O. Tinti; M. Marcellini; T. Riccioni; M. B. Guglielmi; P. Carminati; C. Pisano, *J. Med. Chem*, **2008**, 51, 2708.
[13] G. C.Tron; F. Pagliai;, E. Del Grosso; A. A. Genazzani; G. J. Sorba,; *Med. Chem.* **2005**, 48, 3260.

[14] S. J. Won; C. T. Liu; L. T. Tsao; J. R. Weng; H. H. Ko; J. P. Wang; C. N. Lin; *Eur. J. Med. Chem*, **2005**, 40, 103.
[15] D. Batovska; S. Parushev; B. Stamboliyska; I. Tsvetkova; M. Ninova;H. Najdenski; *Eur. J. Med. Chem*, **2009**, 44, 2211.
[16] Y. K. Rao; S. H. Fang; Y. M. Tzeng, *Bioorg. Med. Chem*, 2009,17, 7909
[17] QSARpro, VLife Sciences Technologies, Pvt. Ltd. Pune, India, www.vlifesciences.com **2011**
[18] R. K. Prasad, R. Sharma *J. Comput. Method. Mol. Design*, **2011**, 1 (1), 7
[19] S K. JAIN; A. K. YADAV; P. NAYAK; *International Journal of Current Pharmaceutical Research*, **2011**, 3, 4,
[20] S.Boltan, C. Bon. 'Pharmaceutical statistics- Practical & Clinical applications', 5th ed, informa healthcare, New York, **2009,** pp. 65,119,120