Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2012, 4(6):3215-3223

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

QSAR Analysis on 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides as epidermal growth factor receptor (EGFR) kinase inhibitors

Sanmati K. Jain*, Rahul Jain, Lokesh Sahu and Arvind K. Yadav

SLT Institute of Pharmaceutical Sciences, Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh-495009, India

---

ABSTRACT

Quantitative structure activity relationship (QSAR) study was performed on a series of 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides possessing epidermal growth factor receptor (EGFR) kinase inhibitory activity for establishing quantitative relationship between biological activity and their physicochemical / structural properties. Several statistical regression equations were obtained using partial least squares regression (PLSR) analysis. Most statistical significant model generated, explains 77% (r2 = 0.7705) of the total variance in the training set as well as it has internal (q2) and external (pred_r2) predicative ability of ~51% (q2 = 0.5065) and 61% (pred_r2 = 0.6112) respectively. In this model positive coefficient value of T_C_C_4 [This is the count of number of carbon atoms separated from any carbon atom (single, double or triple bonded) by 4 bond distance in a molecule] and SssOcount [ ] on the biological activity indicated that higher value leads to better epidermal growth factor receptor (EGFR) kinase inhibitory activity whereas lower value leads to decrease activity. Negative coefficient value of SKMostHydrophobicHydrophilicDistance [most hydrophobic value on vdW surface] on the biological activity indicated that lower values leads to good epidermal growth factor receptor (EGFR) kinase inhibitory activity while higher value leads to reduced activity. Contribution chart reveals that the descriptors T_C_C_4, SKMostHydrophobicHydrophilicDistance and SssOcount contributing 46.75 %, 32.74% and 20.51% respectively.

Key words: 2D-QSAR, epidermal growth factor receptor (EGFR) kinase inhibitors, 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides.

INTRODUCTION

Cancer chemotherapy has entered a new era of molecularly targeted therapeutics, which is highly selective and not associated with the serious toxicities of conventional cytotoxic drugs [1]. Receptor protein tyrosine kinase plays a key role in signal transduction pathways that regulate cell division and differentiation. Among the growth factor receptor kinases that have been identified as being important in cancer is epidermal growth factor receptor (EGFR) kinase. Activation of EGFR may be because of overexpression, mutations resulting in constitutive activation, or autocrine expression of ligand [2,3]. The role of EGFR has been most thoroughly studied in breast cancer, where it is over expressed in 25–30% of cases and is correlated with a poor prognosis. EGFR over expression is also seen in ovarian cancer [4], lung cancer (especially lung adenocarcinomas) [5-7] and in hormone-refractory prostate cancer [8]. Compounds that inhibit the kinase activity of EGFR after binding of ligand are of potential interest as new therapeutic antitumor agents [9,10].

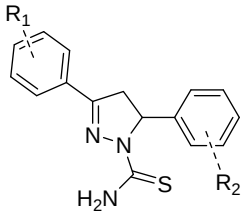
The thiourea and urea derivatives play significant role in anticancer agents because of their inhibitory activity against receptor tyrosine kinases (RTKs), protein tyrosine kinases (PTKs), and NADH oxidase, which play important roles in many aspects of tumorigenesis [11-13].

Various pyrazole derivatives are known to possess a wide range of bioactivities. The pyrazole motif makes up the core structure of abundant biologically active compounds. Thus, various representatives of this heterocycle exhibit anti-viral/anti-tumor [15-17], antibacterial [18-21], antiinflamatory [22], analgesic [23], fungistatic [24] and anti-hyperglycemic activity [25,26]. Pyrazofurin a natural pyrazole C-glycoside, which demonstrated a broad spectrum of antimicrobial activity [27], a great deal of attention, was paid to pyrazole as a potential antimicrobial agent. A few reports regarding EGFR inhibitory activity of pyrazole derivatives containing thiourea skeleton are there in the literature [28,29]. In the present work, quantitative structure activity relationship (QSAR) study was performed on 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides possessing epidermal growth factor receptor (EGFR) kinase inhibitory activity for establishing quantitative relationship between biological activity and their physicochemical / structural properties.

EXPERIMENTAL SECTION

***Data set*:** A dataset of 30 molecules has been taken from the literature [30]. Selected data set, their biological activity is shown in Table-1. Biological data's represented as epidermal growth factor receptor (EGFR) kinase inhibitory activity, IC_{50} values (nM) were converted into log ($1/IC_{50}$) [pIC_{50}] for computational work.

Table 1: General structure of 3,5-disubstituted-4,5-dihydropyrazole-1-carbothioamides and their biological activities (data set of 30 molecules)



S.NO	Compound	R_1	R_2	EGFR ($log1/IC_{50}$)
1	C01	3,4-2 CH_3	4-F	6.08
2	C02	3,4-2 CH_3	4-Cl	5.87
3	C03	3,4-2 CH_3	4-Br	5.66
4	C04	3,4-2 CH_3	4-CH_3	6.47
5	C05	3,4-2 CH_3	4-OCH_3	7.15
6	C06	3,4-2 CH_3	4-OH	6.89
7	C07	3,4-2 CH_3	4-NO_2	5.51
8	C08	3,4-2 CH_3	2-F	5.28
9	C09	3,4-2 CH_3	2-Cl	5.41
10	C10	3,4-2 CH_3	2-Br	5.37
11	C11	3,4-2 Cl	4-F	5.20
12	C12	3,4-2 Cl	4-Cl	5.13
13	C13	3,4-2 Cl	4-Br	5.16
14	C14	3,4-2 Cl	4-CH_3	5.13
15	C15	3,4-2 Cl	4-OCH_3	5.24
16	C16	3,4-2 Cl	4-OH	5.28
17	C17	3,4-2 Cl	4-NO_2	5.11
18	C18	3,4-2 Cl	2-F	5.18
19	C19	3,4-2 Cl	2-Cl	5.14
20	C20	3,4-2 Cl	2-Br	5.19
21	C21	3,4-2 Br	4-F	5.05
22	C22	3,4-2 Br	4-Cl	5.09
23	C23	3,4-2 Br	4-Br	4.99
24	C24	3,4-2 Br	4-CH_3	5.01
25	C25	3,4-2 Br	4-OCH_3	5.09
26	C26	3,4-2 Br	4-OH	4.94
27	C27	3,4-2 Br	4-NO_2	4.97
28	C28	3,4-2 Br	2-F	4.87
29	C29	3,4-2 Br	2-Cl	4.91
30	C30	3,4-2 Br	2-Br	4.96

***QSAR Analysis*:** Structure of the compounds of selected series were drawn using 2D Draw application option of QSAR Plus [31] and converted to 3D structure by exporting to QSAR Plus window. Energy minimizations of the compounds were done by using Merck Molecular Force Field (MMFF) method [Charge-Modified Qeq charge; Maximum number of cycles = 10,000; Convergence criteria (root mean square gradient) = 0.01; Gradient type=analytical and 1.0 as constant (medium's dielectric constant which is 1 for in vacuo) in dielectric properties. The default values of 20.0 and 10.0 Kcal/mol were used for electrostatic and steric energy cutoff] followed by batch optimization. After optimization, number of physicochemical (Individual (H-Acceptor count, H-Donor count , X logP, SMR, polarisablity, etc.), retention index (Chi), atomic valence connectivity index (ChiV), Path count, Chi chain, Chiv chain, Chain Path Count, Cluster, Path cluster, Kapa, Element count (H, N, C, S, O, Cl, Br, I), Estate numbers (SsCH3 Count, SdCH2 Count, SssCH2 Count, StCH count etc.), Estate contribution (SsCH3-index., SdCH2- index, SssCH2 – index , StCH index) and Polar surface area), alignment (for example, T_2_O_7, T_2_N_5, T_2_2_6, T_C_O_1, T_O_Cl_5 etc.) and atom type (based on MMFF atom types and their count in each molecule. In MMFF, there are 99 atom types and hence 99 descriptors indicating number of times that atom has occurred in a given molecule are generated) independent descriptors were calculated for the data set. Calculated descriptors and biological activity were taken as independent and dependent variables respectively. Random, manual and sphere exclusion methods were used for creation of training and test data set. Partial least squares regression (PLSR) statistical method was used to generate QSAR models. Following statistical parameters were considered to select the statistical significance QSAR models: squared correlation coefficient (r^2), F-test (F-test for statistical significance of the model), and cross-validated squared correlation coefficient (q^2).

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 Sphere Exclusion, random and manual data selection methods. 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.

Sphere Exclusion method: In this method dissimilarity value provides an idea to handle training and test set size. It needs to be adjusted by trial and error until a desired division of training and test set is achieved. Increase in dissimilarity value results in increase in number of molecules in the test set.

Random selection: In order to construct and validate the QSAR models, both internally and externally, the data sets were divided into training (90%, 85%, 80% and 75% of total data set) set and test sets (10%, 15%, 20% and 25% of total data set) in a random manner. Ten trials were run in each case.

Manual data selection: Whole range of activities was sorted on the basis of results obtained in sphere exclusion and random methods.

After the creation of training and test set, Min and Max value of the test and training set is checked, using the QSAR tool, if the values are not following the Min – Max, then the training / test set is again set and procedure is repeated. If the Min – Max is following, then Partial Least Squares Regression (PLSR) used for model building (Cross correlation Limit – < 0.5; No. of variables – 1/5th of total training set; Term selection – r2; F test: In – 4.00, Out – 3.99; Model building criteria – Cross validation).

Partial least square regression (PLSR): PLSR was used for model generation. PLSR is an expansion of the multiple linear regression (MLR). In its simplest form, a linear model specifies the (linear) relationship between a dependent variable and a set of predictor variables. In PLSR, prediction functions are represented by factors extracted from the Y'XX'Y matrix. The number of such prediction functions that can be extracted typically will exceed the maximum of the number of Y and X variables. PLSR is probably the least restrictive of the various multivariate extensions of the multiple linear regression models. PLSR can be used as an exploratory analysis tool to select suitable predictor variables and to identify outliers before classical linear regression. All the calculated descriptors were considered as independent variable and biological activity as dependent variable.

RESULTS AND DISCUSSION

All the 30 molecules of the selected series were subjected to partial least squares regression (PLSR) analysis, results of random selection method (Table-2-5), sphere exclusion method (Table-6) and manual data selection method

(Table-7) are shown in the Table 2 to 7. Statistically significant QSAR models with equations obtained for epidermal growth factor receptor (EGFR) kinase inhibitors is shown in Table-8.

Table 2: List of predictive QSAR models with equation generated from PLSR by Random data selection method (90%)

Training set%	Trial no	Test Set	Stepwise-forward backward (SW-FB)						
			r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F test
90%	1	c20,c29,c10	0.6830	0.6046	-0.1624	0.3410	0.3808	0.4094	53.8690
90%	2	c03,c08,c10	0.7389	0.6690	-9.5279	0.3125	0.3518	0.6973	70.7625
90%	3	c21,c07,c10	0.7009	0.6241	-2.8854	0.3338	0.3742	0.5030	58.5884
90%	4	c12,c14,c10	0.6956	0.6193	-2.0083	0.3367	0.3765	0.4594	57.1313
90%	5	c19,c03,c10	0.6918	0.6117	-1.8906	0.3386	0.3801	0.4488	56.1138
90%	6	c27,c02,c10	0.6730	0.5861	0.2823	0.3438	0.3868	0.3800	51.4465
90%	7	c24,c06,c10	0.8042	0.6443	0.4675	0.2387	0.3217	0.8174	49.2773
90%	8	c28,c06,c10	0.7854	0.6127	0.5192	0.2480	0.3331	0.7914	43.9250
90%	9	c16,c23,c10	0.6836	0.6055	-0.7439	0.3426	0.3826	0.3903	54.0030
90%	10	c16,c21,c10	0.6855	0.6079	-1.3255	0.3423	0.3822	0.3946	54.4876

Table 3: List of predictive QSAR models with equation generated from PLSR by Random data selection method (85%)

Training set%	Trial no	Test Set	Stepwise-forward backward (SW-FB)						
			r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F test
85%	1	c16,c22,c25,c03,c10	0.7069	0.6293	-2.0213	0.3420	0.3846	0.4532	55.4633
85%	2	c13,c17,c19,c21,c10	0.6867	0.6077	-0.0765	0.3523	0.3942	0.3222	50.4069
85%	3	c20,c21,c27,c06,c10	0.7616	0.6504	0.5407	0.2664	0.3226	0.5474	73.4769
85%	4	c11,c15,c04,c09,c10	0.7488	0.6347	0.3174	0.2989	0.3604	0.4699	68.5581
85%	5	c20,c24,c30,c03,c10	0.7059	0.6250	-0.3953	0.3391	0.3829	0.4066	55.2033
85%	6	c15,c21,c27,c03,c10	0.7574	0.6520	-0.9827	0.3090	0.3701	0.4478	71.7869
85%	7	c26,c01,c03,c08,c10	0.7242	0.6387	-0.2836	0.3223	0.3689	0.5019	60.3979
85%	8	c14,c29,c30,c07,c10	0.7009	0.6208	-0.1700	0.3407	0.3836	0.4062	53.8942
85%	9	c15,c18,c30,c03,c10	0.8578	0.7675	-5.3932	0.2485	0.3177	0.7310	42.2192
85%	10	c15,c29,c30,c04,c10	0.8547	0.7160	-0.9525	0.2329	0.3256	0.8839	41.1741

Table 4: List of predictive QSAR models with equation generated from PLSR by Random data selection method (80%)

Training set%	Trial no	Test Set	Stepwise-forward backward (SW-FB)						
			r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F test
80%	1	c25,c29,c01,c03,c07,c10	0.7164	0.6326	-0.4459	0.3319	0.3777	0.5113	55.5776
80%	2	c20,c28,c29,c01,c04,c10	0.6066	0.4565	0.8319	0.3597	0.4228	0.2734	33.9193
80%	3	c12,c15,c16,c25,c09,c10	0.8499	0.7536	-19.7937	0.2571	0.3293	0.9646	59.4540
80%	4	c17,c18,c20,c22,c29,c10	0.6846	0.6038	0.2623	0.3573	0.4005	0.3070	47.7557
80%	5	c22,c26,c03,c08,c09,c10	0.8307	0.5970	-4.6725	0.2642	0.4076	0.6609	107.9299
80%	6	c21,c23,c29,c03,c07,c10	0.6951	0.6100	-0.0418	0.3501	0.3960	0.3616	50.1517
80%	7	c12,c14,c21,c04,c06,c10	0.7705	0.5064	0.6112	0.2479	0.3635	0.5563	35.2447
80%	8	c15,c17,c28,c29,c03,c10	0.7895	0.6674	-0.4469	0.2960	0.3721	0.4694	39.3799
80%	9	c19,c21,c27,c29,c03,c10	0.6736	0.5861	0.4560	0.3604	0.4058	0.2875	45.3943
80%	10	c18,c26,c28,c01,c02,c10	0.7810	0.5686	-0.1167	0.2930	0.4112	0.5196	37.4413

Table 5: List of predictive QSAR models with equation generated from PLSR by Random data selection method (75%)

Training set%	Trial no	Test Set	Stepwise-forward backward (SW-FB)						
			r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F test
75%	1	c12,c13,c15,c21,c27,c02,c09,c10	0.7703	0.6610	-0.4263	0.3174	0.3856	0.3800	67.0888
75%	2	c11,c14,c24,c27,c29,c01,c06,c10	0.7975	0.6618	0.4309	0.2540	0.3282	0.5264	37.4045
75%	3	c15,c16,c23,c24,c27,c30,c08,c10	0.7974	0.7020	-0.2329	0.2954	0.3582	0.4319	78.6938
75%	4	c14,c27,c28,c29,c04,c06,c09,c10	**0.7891**	**0.6090**	**0.6053**	0.2411	0.3282	0.4905	35.5498
75%	5	c15,c16,c21,c23,c02,c05,c07,c10	0.6352	0.5199	0.6963	0.3212	0.3685	0.4120	34.8198
75%	6	c15,c16,c21,c27,c06,c07,c08,c10	0.8896	0.8009	0.1529	0.1992	0.2674	0.5648	76.5357
75%	7	c16,c17,c20,c21,c22,c25,c03,c10	0.7050	0.6254	-0.4465	0.3623	0.4083	0.3598	47.8072
75%	8	c13,c17,c19,c21,c24,c27,c06,c10	**0.7027**	**0.5949**	**0.5538**	0.3126	0.3649	0.4279	47.2791
75%	9	c19,c20,c21,c27,c03,c06,c08,c10	0.8729	0.7636	0.2594	0.2117	0.2887	0.5402	65.2164
75%	10	c11,c15,c21,c22,c29,c04,c09,c10	0.7432	0.6068	0.4256	0.3162	0.3913	0.3689	57.8683

Table 6: List of predictive QSAR models with equation generated from PLSR by by sphere exclusion method

Trial	Dissimilarity value	Test Set	Stepwise-forward backward (SW-FB)						
			r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F test
1	2.6	c23,c12,c14	0.6601	0.5785	0.6519	0.3526	0.3927	0.2385	48.5553
2	3.0	c22,c12,c13	0.6604	0.5797	0.5574	0.3541	0.3939	0.2278	48.6093
3	3.1	c02,c03,c12,c13,c23,c24	0.6643	0.5682	0.5724	0.3649	0.4138	0.2492	43.5316
4	3.2	c02,c03,c10,c12,c13,c29,c23,c24	0.6910	0.5961	0.3086	0.3619	0.4137	0.3169	44.7280
5	3.5	c02,c03,c10,c12,c13,c19,c29,c23,c24	0.6913	0.5955	0.3558	0.3693	0.4227	0.3030	42.5427

Table 7: List of predictive QSAR models with equation generated from PLSR by manual data selection method

Trial	Test Set	Stepwise-forward backward (SW-FB)						
		r^2	q^2	Pred_r^2	r^2 se	q^2 se	Pred_r^2se	F test
1	c12,c13,c23,c02,c03	0.6499	0.5535	0.7725	0.3675	0.4150	0.1800	42.6906
2	c12,c13,c24,c02,c03	0.6700	0.5763	0.4500	0.3569	0.4044	0.2780	46.7013
3	c12,c23,c24,c02,c03	0.6622	0.5665	0.5886	0.3590	0.4067	0.2618	45.0952
4	c12,c13,c23,c24,c02	0.6608	0.5703	0.6385	0.3601	0.4053	0.2468	44.8035
5	c12,c13,c23,c24,c03	0.6721	0.5861	0.4396	0.3569	0.4010	0.2768	47.1336
6	c12,c13,c23 ,c02	0.6477	0.5570	0.8392	0.3624	0.4064	0.1636	44.1327
7	c12,c13, c19,c23, c02	0.6492	0.5579	0.7700	0.3680	0.4131	0.1824	42.5632
8	c12,c13,c23 ,c02,c10	0.6772	0.5898	0.2663	0.3543	0.3994	0.3029	48.2537
9	c12, c23,c29, c02	0.6360	0.5422	0.9547	0.3646	0.4089	0.1022	41.9392
10	c19,c23, c24,c29,c02	0.7663	0.6189	0.2277	0.3023	0.3860	0.4129	36.0753

Table 8: List of significant QSAR models with equation generated from PLSR

Model. no	Method	Test set	Equation
1	Random selection method 75%/trial 4/PLS	C14,C27,C28,C29,C 04,C06,C09,C10	log1/IC_{50}= 0.2794 T_C_C_4 – 0.1336 SKMostHydrophobicHydrophilicDistance + 0. 1223 T_N_Cl_6 + 2.1328 Optimum Components = 2 n = 22 Degree of freedom = 24 F test = 35.5498 r2 = 0.7891 q2 = 0.6090 pred_r2 = 0.6053 r2 se = 0.2411 q2 se = 0.3282 pred_r2se = 0.4905 Alpha rand R^2 = 0.00000 ; Alpha rand Q^2 = 0.00000; Alpha rand Pred R^2 = 0.01000
2	Random selection method 80%/trial 7/PLS	C12,C14,C21,C04,C 06,C10	log1/IC_{50} = 0.1809 T_C_C_4 – 0.1428 SKMostHydrophobicHydrophilicDistance + 0.3643 SssOcount + 3.7834 Optimum Components = 2 n = 24 Degree of freedom = 21 F test = 35.2447 r2 = 0.7705 q2 = 0.5065 pred_r2 = 0.6112 r2 se = 0.2479 q2 se = 0.3635 pred_r2se = 0.5563 Alpha rand R^2 = 0.00000 ; Alpha rand Q^2 = 0.00000; Alpha rand Pred R^2 = 0.00100

In the above QSAR models, n is the number of molecules (Training set) used to derive the QSAR model, r^2 is the squared correlation coefficient, q2 is the cross-validated correlation coefficient, pred_r^2 is the predicted correlation coefficient for the external test set, *F* is the Fisher ratio, 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, q2 se and pred_r^2se are the standard errors terms for r^2, q^2 and pred_r^2 (smaller is better). R^2 is the correlation coefficient for observed vs. predicted biological activity.

Table 9: Actual and predicted biological activity for Training set and test set

S. No.	Compound	Actual (pIC_{50})	Predicted Biological Activity (pIC_{50})	
			Model 1	Model 2
1	C01	6.08	6.00	5.92
2	C02	5.87	5.83	5.60
3	C03	5.66	5.75	5.64
4	C04	6.47	6.29*	6.11*
5	C05	7.15	6.53	6.61
6	C06	6.89	5.84*	5.74*
7	C07	5.51	5.84	5.74
8	C08	5.28	5.81	5.71
9	C09	5.41	5.72*	5.62
10	C10	5.37	5.35*	5.22*
11	C11	5.20	5.26	5.22
12	C12	5.13	5.24	5.07*
13	C13	5.16	5.20	5.15
14	C14	5.13	5.51*	5.36*
15	C15	5.24	5.38	5.47
16	C16	5.28	5.19	5.14
17	C17	5.11	5.25	5.20
18	C18	5.18	4.91	4.84
19	C19	5.14	5.13	5.08
20	C20	5.19	5.13	5.07
21	C21	5.05	5.02	5.22*
22	C22	5.09	5.00	5.07
23	C23	4.99	4.99	5.19
24	C24	5.01	5.24	5.34
25	C25	5.09	5.08	5.42
26	C26	4.94	4.95	5.14
27	C27	4.97	4.54*	4.70
28	C28	4.87	4.65*	4.82
29	C29	4.91	4.64*	4.82
30	C30	4.96	4.63	4.81

**indicates compounds are in the test set for the corresponding model and rest are in the training set.*

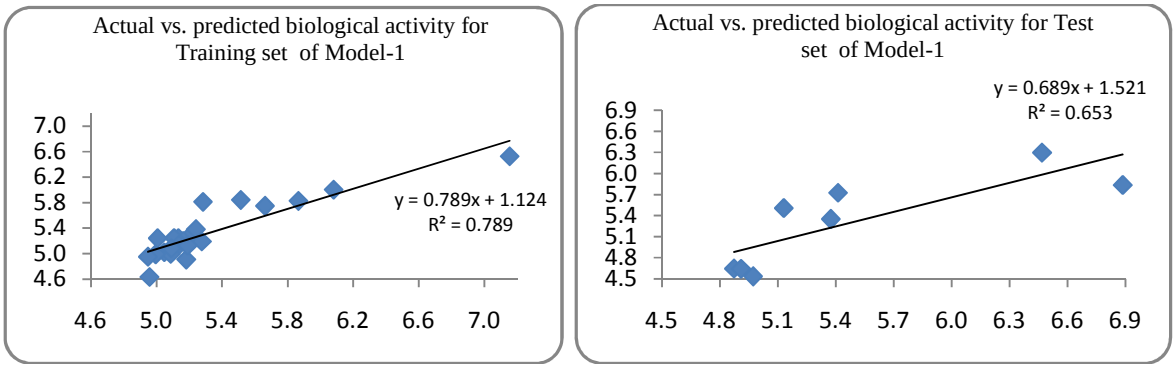


Figure-01: Graph between actual and predicted biological activity of training and test set for Model-1.

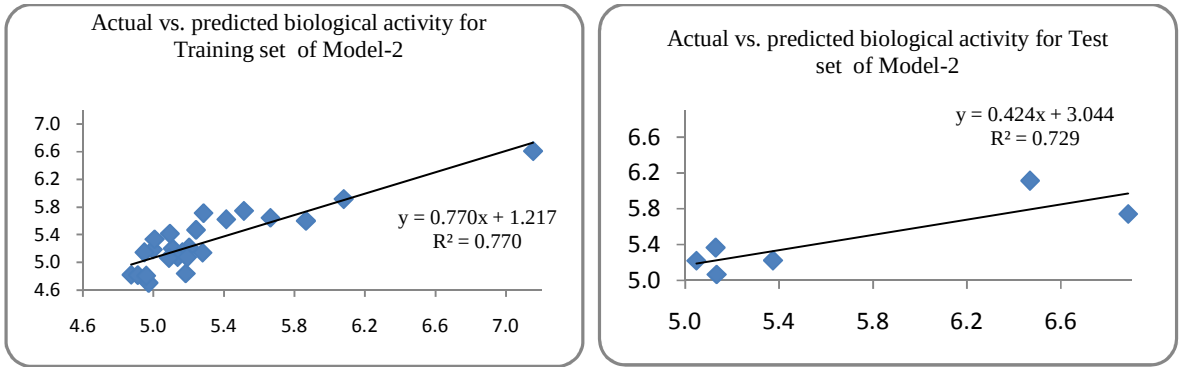
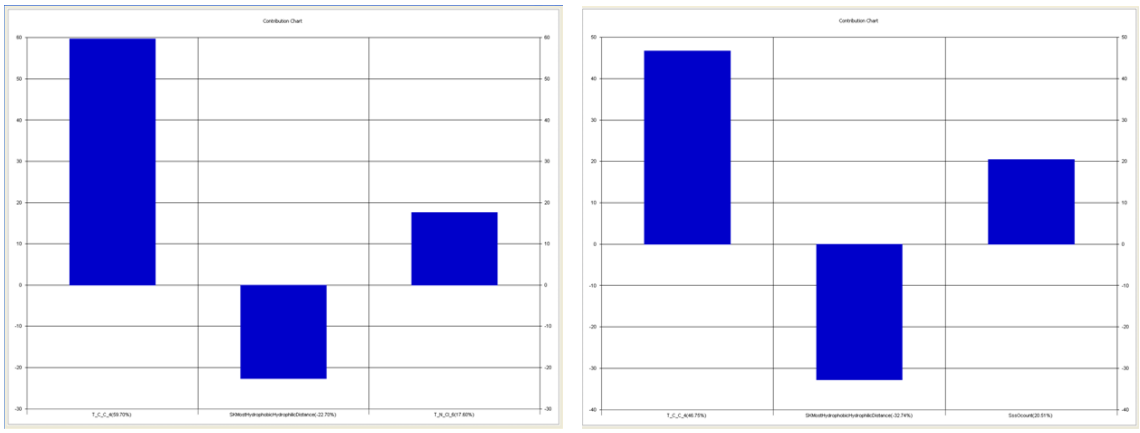
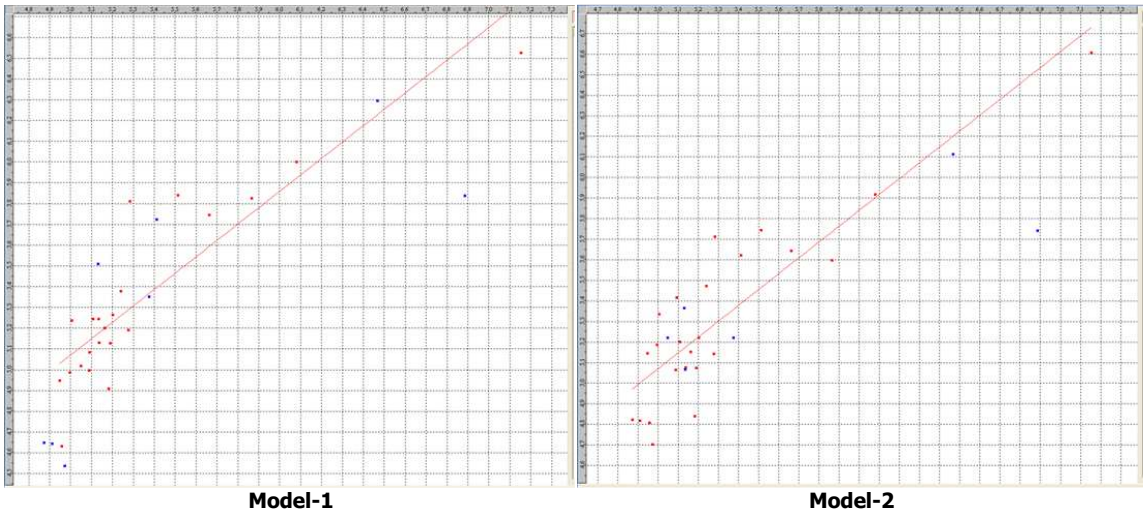


Figure-02: Graph between actual and predicted biological activity of training and test set for Model-2.



Model-1 **Model-2**

Figure-03: Contribution plot for Model 1-2.



Model-1 **Model-2**

Figure-04: Data fitness plot for Model 1-2.

From the table-8, the equation of Model-01 explains 79% (r2=0. 0.7891) of the total variance in the training set as well as it has internal (q2) and external (pred_r2) predicative ability of 61% (q2 =0.6090) and 61% (pred_r2 = 0.6053) respectively. Model- 02 explains 77% (r2 = 0.7705) of the total variance in the training set as well as it has internal (q2) and external (pred_r2) predicative ability of 51% (q2 = 0.5065) and 61% (pred_r2 = 0.6112) respectively.

Table-09 represents the predicted biological activity by the model for training and test set. 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-02 and 3) 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.

Interpretation of the Model 2 (Most significant)

Among the significant models generated (Table-02), model 2 is the most significant one as it is having the higher correlation coefficient value for the test set (R^2 = 0.729) as compared to model 1 (R^2 = 0.653) displayed in Figure 2 and 1 respectively.

The equation 2 explains 77% (r2 = 0.7705) of the total variance in the training set and has an internal (q2) and external (pred_r2) predictive ability of ~51% (q2 = 0.5065) and ~61% (pred_r2 = 0.6112) 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 99.9 (Alpha Rand Pred R^2 = 0.001) that the generated model is not random and hence may be chosen as the QSAR model. In the QSAR model 2, the positive coefficient value of T_C_C_4 [This is the count of number of carbon atoms separated from any carbon atom (single, double or triple bonded) by 4 bond distance in a molecule] and SssOcount [this descriptor defines the total number of oxygen atom connected with two single bonds] on the biological activity indicated that higher value leads to better epidermal growth factor receptor (EGFR) kinase inhibitory activity (compound c05,c01,c02,c07 etc.) whereas lower value leads to decrease activity (compound c28,c13,c26,c29,c30 etc.). Negative coefficient value of SKMostHydrophobicHydrophilicDistance [most hydrophobic value on vdW surface] on the biological activity indicated that lower values leads to good epidermal growth factor receptor (EGFR) kinase inhibitory activity (compound c05,c01,c07, etc.) while higher value leads to reduced activity (compound c28,c15,c29,c30,c25,c18 etc.). Figure-03 represents the contribution chart showing contribution of the various descriptors playing important role in determining the epidermal growth factor receptor (EGFR) kinase inhibitory activity for model 01-02 and Figure-04 represents the data fitness plot for model 01-02. Contribution chart for model 2 reveals that the descriptors T_C_C_4, SKMostHydrophobicHydrophilicDistance and SssOcount contributing 46.75 %, 32.74% and 20.51% respectively.

Interpretation of the Model 01

The equation 1 explains 79% (r2=0. 0.7891) of the total variance in the training set as well as it has internal (q2) and external (pred_r2) predicative ability of 61% (q2 =0.6090) and 61% (pred_r2 = 0.6053) 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 (Alpha Rand Pred R^2 = 0.01) that the generated model is not random and hence may be chosen as the QSAR model.

In the QSAR model 1, the positive coefficient value of T_C_C_4 [This is the count of number of carbon atoms separated from any carbon atom (single, double or triple bonded) by 4 bond distance in a molecule] and T_N_Cl_6 [this is the count of number of Nitrogen atoms (single, double or triple bonded) separated from any Chlorine atom by 6 bonds in a molecule] on the biological activity indicated that higher value leads to better epidermal growth factor receptor (EGFR) kinase inhibitory activity (compound c05,c01,c02,c07 etc.) whereas lower value leads to decrease activity (compound c28,c13,c26,c29,c30 etc.). Negative coefficient value of SKMostHydrophobicHydrophilicDistance on the biological activity indicated that lower values leads to good epidermal growth factor receptor (EGFR) kinase inhibitory activity (compound c05,c01,c07, etc.) while higher value leads to reduced activity (compound c28,c15,c29,c30,c25,c18 etc.). Contribution chart for model 1 reveals that the descriptors T_C_C_4, SKMostHydrophobicHydrophilicDistance and T_N_Cl_6 contributing 59.70%, 22.70%, and 17.60 % respectively.

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.

From Figure 1, and 2, it is seen that the plots of observed vs. predicated activity for different models provide an idea about how well the models were trained and how well they predict the activity of the external test set.

Acknowledgement
The authors are indebted to the Head, SLT Institute of Pharmaceutical Sciences, Guru Ghasidas Vishwavidyalaya, Bilaspur (CG) for providing necessary facilities. RJ and LS is thankful to AICTE for GPAT scholarship.

REFERENCES

[1]L Seymore. *Cancer Treat. Rev.*, **1999**, 25, 301.
[2]DJ Slamon, GM Clark, SG Wong, WJ Levin, A Ullrich, WL McGuire. *Science*, **1987**, 235, 177.
[3]DJ Slamon, W Godolphin, LA Jones. *Science*, **1989**, 244, 707.
[4]D Scheurle, M Jahanzeb, RS Aronsohn, L Watzek, RH Narayanan. *Anticancer. Res.*, **2000**, 20, 2091.
[5]G Cox, M Vyberg, B Melgaard, J Askaa, A Oster, KJ O'Byrne. *Int. J. Cancer*, **2001**, 92, 480.
[6]WJ Gullick. *Br. Med. Bull.*, **1991**, 47, 87.
[7]DK Moscatello, M Holgado-Mudruga, AK Godwin, G Ramirez, G Gunn, PW Zoltick, JA Biegel, RL Hayes, AJ Wong. *Cancer Res.*, **1995**, 55, 5536.
[8]CJ Wikstrand, RE McLendon, A Friedman, DD Bigner. *Cancer Res.*, **1997**, 57, 4130.
[9]AJ Bridges. *Curr. Med. Chem.*, **1999**, 6, 825.
[10]DH Boschelli. *Drugs Future*, **1999**, 24, 515.
[11]A Ullrich, Schlessinger. *Cell*, **1990**, 61, 203.
[12]SR Hubbard, JH Till. *Rev. Biochem.*, **2000**, 69, 373.
[13]Y Dai, Y Guo, RR Frey, Z Ji, ML Curtin, AA Ahmed, DH Albert, L Arnold, SS Arries, T Barlozzari, JL Bauch, JJ Bouska, PF Bousquet, GA Cunha, KB Glaser, J Guo, J Li, PA Marcotte, KC Marsh, MD Moskey, LJ Pease, KD Stewart, VS Stoll, P Tapang, N Wishart, SK Davidsen, MR Michaelides. *J. Med. Chem.*, **2005**, 48, 6066.
[14]Y Dai, K Hartandi, Z Ji, AA Ahmed, DH Albert, L Arnold, SS Arries, T Barlozzari, JL Bauch, JJ Bouska, PF Bousquet, GA Cunha, KB Glaser, J Guo, J Li, PA Marcotte, KC Marsh, MD Moskey, LJ Pease, KD Stewart, VS Stoll, P Tapang, N Wishart, SK Davidsen, MR Michaelides. *J. Med. Chem.*, **2007**, 50, 1584.
[15]S Manfredini, R Bazzanini, PG Baraldi, M Guarneri, D Simoni, ME Marongiu, A Pani, E Tramontano, P La Colla. *J. Med. Chem.*, **1992**, 35, 917.
[16]S Manfredini, R Bazzanini, PG Baraldi, M Bonora, M Marangoni, D Simoni, A Pani, F Scintu, E Pinna. *Anti-Cancer Drug Des.*, **1996**, 11, 193.
[17]HA Park, K Lee, SJ Park, B Ahn, JC Lee, HY Cho, KI Lee. *Bioorg. Med. Chem. Lett.*, **2005**, 15, 3307.
[18]A Tanitame, Y Oyamada, K Ofuji, M Fujimoto, N Iwai, Y Hiyama, K Suzuki, H Ito, M Wachi, J Yamagishi, *J. Med. Chem.*, **2004**, 47, 3693.
[19]SG Küçükgüzel, S Rollas, H Erdeniz, M Kiraz, AC Ekinci, A Vidin. *Eur. J. Med. Chem.*, **2000**, 35, 761.
[20]MJ Genin, DA Allwine, DJ Anderson, MR Barbachyn, DE Emmert, SA Garmon, DR Graber, KC Grega, JB Hester, DK Hutchinson, J Morris, RJ Reischer, CW Ford, GE Zurenko, JC Hamel, RD Schaadt, D StapertBH Yagi. *J. Med. Chem.*, **2000**, 43, 953.
[21]AA Bekhit, HTY Fahmy, SAF Rostom, AM Baraka. *Eur. J. Med. Chem.*, **2003**, 38, 27.
[22]TD Penning, JJ Talley, SR Bertenshaw, JS Carter, PW Collins, S Docter, MJ Graneto, LF Lee, JW Malecha, JM Miyashiro, RS Rogers, DJ Rogier, SS Yu, GD Anderson, EG Burton, JN Cogburn, SA Gregory, CM Koboldt, WE Perkins, K Seibert, AW Veenhuizen, YY Zhang, PC Isakson. *J. Med. Chem.*, **1997**, 40, 1347.
[23]G Menozzi, L Mosti, P Fossa, F Mattioli, M Ghia. *J. Heterocycl. Chem.*, **1997**, 34, 963.
[24]R Sridhar, PT Perumal, S Etti, G Shanmugam, MN Ponnuswamy, VR Prabavathy, N Mathivanan. *Bioorg. Med. Chem. Lett.*, **2004**, 14, 6035.
[25]KL Kees, JJ Fitzgerald, KE Steiner, JF Mattes, B Mihan, T Tosi, D Mondoro, ML McCaleb. *J. Med. Chem.*, **1996**, 39, 3920.
[26]GR Bebernitz, G Argentieri, B Battle, C Brennan, B Balkan, BF Burkey, M Eckhardt, J Gao, P Kapa, RJ Strohschein, HF Schuster, M Wilson, DD Xu. *J. Med. Chem.*, **2001**, 44, 2601.
[27]RN Comber, RJ Gray, JA Secrist. *Carbohydr. Res.*, **1992**, 216, 441.
[28]PC Lv, CF Zhou, J Chen, PG Liu, KR Wang, WJ Mao, HQ Li, Y Yang, J Xiong, HL Zhu, *Bioorg. Med. Chem.*, **2010**, 18, 314.
[29]PC Lv, KR Wang, QS Li, J Chen, J Sun, HL Zhu. *Bioorg. Med. Chem.*, **2010**, 18, 1117.
[30]PC Lv, HQ Li, J Sun, Y Zhou, HL Zhu. *Bioorg. Med. Chem.*, **2010**, 18, 4606.
[31]VLifeSciences Technology Pvt. Ltd., Pune-411045, Web: vlifesciences.com.