



Scholars Research Library

Der Pharmacia Lettre, 2012, 4 (4):1060-1070  
(<http://scholarsresearchlibrary.com/archive.html>)



## QSAR Analysis on 3,4-disubstituted-1H-pyrazol-5(4H)-ones as KDR/VEGFR-2 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 3,4-disubstituted-1H-pyrazol-5(4H)-ones possessing KDR/VEGFR-2 kinase inhibitory activity for establishing quantitative relationship between biological activity and their physicochemical/ structural properties. Several statistical regression expressions were obtained using partial least squares regression (PLSR) analysis. Most statistical significant model generated, explains 80% ( $r^2 = 0.8044$ ) of the total variance in the training set and has an internal ( $q^2$ ) and external ( $pred_r^2$ ) predictive ability of ~71% ( $q^2 = 0.7077$ ) and ~90% ( $pred_r^2 = 0.8959$ ). In this model the positive coefficient value of  $T\_T\_N\_7$  [This is the count of number of any atoms separated from any Nitrogen atom (single, double or triple bonded) by 7 bond distance in a molecule] and  $T\_N\_Br\_4$  [this is the count of number of Nitrogen atoms (single, double or triple bonded) separated from any Bromine atom by 4 bonds in a molecule] on the biological activity indicated that higher value leads to better KDR/VEGFR-2 kinase inhibitory activity whereas lower value leads to decrease activity. Negative coefficient value of  $T\_C\_N\_3$  [This is the count of number of Carbon atoms (single, double or triple bonded) separated from any Nitrogen atom (single, double or triple bonded) by 3 bond distance in a molecule] and AveragePotential on the biological activity indicated that lower values leads to good KDR/VEGFR-2 kinase inhibitory activity while higher value leads to reduced activity. Contribution chart reveals that the descriptors  $T\_T\_N\_7$ ,  $T\_C\_N\_3$ , AveragePotential and  $T\_N\_Br\_4$  contributing 22.56%, 34.34%, 24.87% and 18.22 % respectively.

**Keywords:** 2D-QSAR, KDR/VEGFR-2 kinase inhibitors, 3,4-disubstituted-1H-pyrazol-5(4H)-ones

### INTRODUCTION

Recognition of the novel receptor tyrosine kinase (RTK) inhibitors continue to be an passionate area of exploration in anti-tumor research [1-3], due to the fact that over expression of a few of their analogous receptor proteins contributes to constitutive RTK signaling, resulting in deregulated cell growth and cancer [4,5]. Such uncontrolled RTK signaling leading to tumor growth has provided the impetus for the design of RTK inhibitors as potential anti-tumor agents. The approval of Sorafenib [6], Sunitinib [7] and Iressa [8] has validated such mechanistic approaches. One of the most comprehensively studied pathways in this area is vascular endothelial growth factor (VEGF) [9] and its cell surface receptor in human KDR (kinase domain containing receptor or VEGFR-2) [10,11] due to their important roles in angiogenesis [12] which is crucial for survival and proliferation of tumor cells. KDR receptors, shown to be expressed primarily in endothelial cells [13], upon binding to VEGF get activated and their intracellular kinase domains undergo autophosphorylation, which in turn triggers signaling pathways leading to sprouting of

blood vessels toward the tumor cells. Therefore, inhibition of KDR kinase and subsequent blockage of angiogenesis could provide an alternate approach to cancer therapy [14].

Heterocyclic substituted pyrazolones (Figure-1) as the core structure were reported in the literature as a novel class of KDR kinase inhibitors [15-18]. Also 1,2,3-Thiadiazole substituted pyrazolones were reported as potent KDR/VEGFR-2 kinase inhibitors [19].

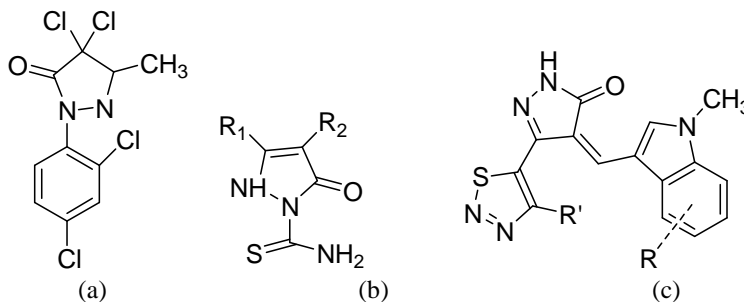
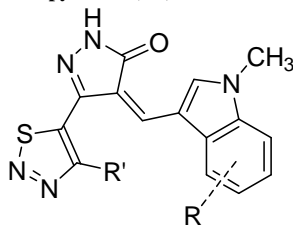


Figure-1: Heterocyclic substituted pyrazolones

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



S. No.	Compound	R	R'	log 1/IC <sub>50</sub>
1	8a	H	CH <sub>3</sub>	7.0223
2	8b	5-F	CH <sub>3</sub>	7.0044
3	8c	5-Cl	CH <sub>3</sub>	6.8182
4	8e	5-OCH <sub>3</sub>	CH <sub>3</sub>	7.0458
5	8g	4-F	CH <sub>3</sub>	6.8633
6	8h	4-Cl	CH <sub>3</sub>	7.3468
7	8i	4-Br	CH <sub>3</sub>	7.4685
8	8j	4-OCH <sub>3</sub>	CH <sub>3</sub>	7.7212
9	8k	4-OC <sub>2</sub> H <sub>5</sub>	CH <sub>3</sub>	7.4202
10	8l	4-CH <sub>3</sub>	CH <sub>3</sub>	7.2366
11	8m	4-Br, 5-OCH <sub>3</sub>	CH <sub>3</sub>	7.6383
12	8n	7-OCH <sub>3</sub>	CH <sub>3</sub>	6.9626
13	8p	4,7 bis-OCH <sub>3</sub>	CH <sub>3</sub>	7.6990
14	14a	5-F	H	7.3188
15	14b	5-Cl	H	7.4685
16	14c	5-CN	H	6.7595
17	14d	5-COOCH <sub>3</sub>	H	7.1367
18	14e	5-OCH <sub>3</sub>	H	7.5528
19	14f	5-Br	H	6.9586
20	14g	5-CH <sub>3</sub>	H	7.4559
21	14h	4-F	H	7.7959
22	14i	4-Cl	H	7.7959
23	14j	4-Br	H	8.0969
24	14k	4-OCH <sub>3</sub>	H	7.8861
25	14l	4-OC <sub>2</sub> H <sub>5</sub>	H	7.6021
26	14m	4-COOCH <sub>3</sub>	H	8.2218
27	14n	4-Br,5-OCH <sub>3</sub>	H	7.6576
28	14o	6-Cl	H	7.1367
29	14p	4-Br, 6-CH <sub>3</sub>	H	7.9208
30	14q	4,7 bis-OCH <sub>3</sub>	H	8.0000

## MATERIALS AND METHODS

**Data set:** A dataset of 30 molecules has been taken from the literature [19]. Selected data set, their biological activity is shown in Table-1. Biological data's represented as KDR/VEGFR-2 kinase inhibitory activity, IC<sub>50</sub> values (nM) were converted into log (1/IC<sub>50</sub>) [pIC<sub>50</sub>] for computational work.

**QSAR Analysis:** Structure of the compounds of selected series were drawn using 2D Draw application option of QSAR Plus [20] 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 (HAcceptorcount, 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% & 75% of total data set) set and test sets (10%, 15%, 20% & 25%) 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 –  $r^2$ ; 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). 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. 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 KDR/VEGFR-2 kinase inhibitory activity 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 <sup>2</sup>	q <sup>2</sup>	Pred_r <sup>2</sup>	r <sup>2</sup> se	q <sup>2</sup> se	Pred_r <sup>2</sup> se	F test
90%	1	8l,14f,14q	0.8686	0.8169	-0.3095	0.1537	0.1815	0.6199	50.6857
90%	2	8b,8e, 14q	0.8261	0.7679	0.8922	0.1753	0.2026	0.1887	36.430
90%	3	8g,14j,14q	0.8267	0.7057	0.5331	0.1652	0.2153	0.5114	36.5815
90%	4	8k,14j,14q	0.8134	0.7186	0.8738	0.1781	0.2187	0.2353	33.4229
90%	5	8j,14i,14q	0.8587	0.8067	0.7300	0.1594	0.1864	0.2964	46.5847
90%	6	14d,14i,14q	0.8387	0.7597	-1.0125	0.1710	0.2087	0.7583	39.8618
90%	7	14a,14m,14q	0.8452	0.7982	0.7312	0.1706	0.1948	0.2375	41.8660
90%	8	14e,14m,14q	0.8019	0.6580	-3.1501	0.1926	0.2531	0.9660	31.0415
90%	9	8h,8p,14q	0.8443	0.7867	0.8360	0.1707	0.1998	0.1905	41.5850
90%	10	8h,14j,14q	0.8453	0.7610	0.8627	0.1621	0.2015	0.2445	41.9031

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 <sup>2</sup>	q <sup>2</sup>	Pred_r <sup>2</sup>	r <sup>2</sup> se	q <sup>2</sup> se	Pred_r <sup>2</sup> se	F test
85%	1	8c,8i,8k,8m,14q	0.8025	0.6797	0.5207	0.1935	0.2464	0.2989	28.4432
85%	2	8l,8m, 14d, 14m,14q	0.8511	0.7278	-1.6686	0.1721	0.2326	0.6121	40.0056
85%	3	8a,8g, 14k, 14p,14q	0.8776	0.8196	0.4221	0.1419	0.1723	0.4304	50.1762
85%	4	8l,14a,14g,14j,14q	0.8044	0.7077	0.8959	0.1900	0.2323	0.1523	28.7873
85%	5	8g,8m,14d,14j,14q	0.8760	0.7813	0.3276	0.1437	0.1908	0.4574	49.4542
85%	6	8j,8l,8n,14f,14q	0.7570	0.4346	-0.0279	0.2105	0.3211	0.4805	21.8010
85%	7	8n,14j,14o,14p,14q	0.7952	0.6654	0.6000	0.1815	0.2321	0.3726	27.1867
85%	8	8m,14f,14j,14n,14q	0.7687	0.6451	0.7946	0.1804	0.2234	0.3138	23.2579
85%	9	8b,8e,8m,14m,14q	0.7858	0.5998	0.3014	0.2014	0.2752	0.3619	25.6852
85%	10	8g,8i,14f,14j,14q	0.8352	0.7114	0.6519	0.1636	0.2164	0.3387	35.4705

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 <sup>2</sup>	q <sup>2</sup>	Pred_r <sup>2</sup>	r <sup>2</sup> se	q <sup>2</sup> se	Pred_r <sup>2</sup> se	F test
80%	1	8k,8m,14d, 14f, 14j, 14q	0.7636	0.6446	-0.0738	0.2060	0.2526	0.4984	21.5299
80%	2	8j,14c, 14e, 14h, 14i, 14q	0.8024	0.6722	0.5496	0.1872	0.2410	0.3334	27.0713
80%	3	8e, 14a, 14d, 14h, 14m, 14q	0.7690	0.6466	-2.1828	0.2089	0.2584	0.6979	34.9480
80%	4	8g,8p, 14d, 14g, 14o, 14q	0.8437	0.7717	-1.1493	0.1729	0.2090	0.6228	35.9920
80%	5	8e, 14d, 14e, 14f, 14m, 14q	0.7992	0.6013	0.4943	0.1972	0.2779	0.2941	26.5395
80%	6	8g,8h,8m,8p,14l,14q	0.8386	0.7373	-0.8147	0.1778	0.2268	0.5464	34.6446
80%	7	8h,8m,14d,14m,14n,14q	0.8081	0.6529	0.7693	0.1836	0.2470	0.2460	28.0750
80%	8	8h,8m,8n,14b,14j,14q	0.8075	0.7138	0.8213	0.1876	0.2287	0.1993	27.9690
80%	9	8c,8i,8k,8m,14d,14q	0.8400	0.7490	-2.3664	0.1765	0.2210	0.7486	35.0063
80%	10	8l,8m,14d,14m,14o,14q	0.8531	0.7679	-2.2568	0.1735	0.2180	0.6413	38.7003

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 <sup>2</sup>	q <sup>2</sup>	Pred_r <sup>2</sup>	r <sup>2</sup> se	q <sup>2</sup> se	Pred_r <sup>2</sup> se	F test
75%	1	8a,8g,8m,8n,14f, 14k, 14p, 14q	0.7541	0.5823	0.2172	0.2016	0.2627	0.4428	18.3987
75%	2	8l,8n,14a, 14f, 14g, 14h, 14j, 14q	0.8350	0.7357	0.1435	0.1746	0.2210	0.4149	30.3646
75%	3	8g,8l,8m,8p,14d, 14j, 14n, 14q	0.8241	0.7082	-0.0307	0.1620	0.2087	0.5644	28.1167
75%	4	8j,14a, 14b, 14c, 14h, 14j, 14o, 14q	0.7906	0.6569	0.6086	0.1918	0.2454	0.2997	22.6545
75%	5	8e,8g,8p,14f, 14g, 14n, 14o, 14q	0.8186	0.7202	0.2727	0.1716	0.2131	0.4328	27.0742
75%	6	8g,8h,8i,8j,14d, 14g, 14j, 14q	0.8583	0.7484	0.0215	0.1646	0.2193	0.4275	36.3314
75%	7	8g,14b, 14e, 14f, 14g, 14k, 14l, 14q	0.7461	0.5687	-1.0632	0.2253	0.2936	0.5777	17.6314
75%	8	14b, 14c, 14f,14h, 14j, 14k,14n, 14q	0.7116	0.5216	0.7194	0.1874	0.2413	0.3300	14.8042
75%	9	8c, 8l, 14e, 14f,14h, 14j, 14k, 14q	0.7635	0.6525	0.2936	0.1963	0.2380	0.4307	19.3689
75%	10	8b,8g,8h,8i,14b,14c,14p,14q	0.8716	0.7725	0.2854	0.1504	0.2002	0.4013	40.7214

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

Trial	Dissimilarity value	Test Set	Stepwise-forward backward (SW-FB)						
			r <sup>2</sup>	q <sup>2</sup>	Pred_r <sup>2</sup>	r <sup>2</sup> se	q <sup>2</sup> se	Pred_r <sup>2</sup> se	F test
1	2.8	8n,14b	0.8405	0.7887	0.8411	0.1728	0.1989	0.1943	42.1649
2	3.0	8j,8n,14b	0.8618	0.7280	0.4876	0.1629	0.2285	0.2809	47.8246
3	3.1	14f,14l,8j,8n,14b,14i,8h,8g	0.9399	0.8968	-0.0031	0.1107	0.1451	0.3945	93.8463
4	3.4	14f,14l,8j,8n,14b,14i,8h,14h,8g	0.9457	0.8963	0.0004	0.1066	0.1474	0.3830	98.7702
5	3.5	14j,14e,8j,8n,14l,14b,14i,8h,14h,8g	0.9809	0.8995	0.3167	0.0628	0.1440	0.3351	273.5704

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 <sup>2</sup>	q <sup>2</sup>	Pred_r <sup>2</sup>	r <sup>2</sup> se	q <sup>2</sup> se	Pred_r <sup>2</sup> se	F test
1	8b, 8k, 14j, 14m, 14q	0.7987	0.6154	0.3923	0.1880	0.2599	0.4061	27.767
2	8g, 8h, 14f, 14j, 14q	0.8352	0.7131	0.6487	0.1634	0.2156	0.3406	35.487
3	8h, 8m, 8n, 14j, 14q	0.7329	0.5907	0.8530	0.2158	0.2671	0.2011	19.206
4	8m, 8n,14b, 14d, 14j	0.7919	0.6939	-0.0263	0.1974	0.2393	0.4516	26.634
5	8m, 8n,8p,14q	0.8482	0.7733	-0.4736	0.1673	0.2045	0.5787	40.971
6	8l,14g,14j,14q	0.8127	0.7094	0.8740	0.1817	0.2264	0.1937	31.8296
7	8l,14a,14j,14q	0.8090	0.6983	0.8792	0.1835	0.2306	0.1886	31.0539
8	14a,14g,14j,14q	0.8136	0.7036	0.8702	0.1818	0.2292	0.1953	32.0156
9	8l,14a,14g,14j	0.7436	0.5916	0.1157	0.2219	0.2801	0.3856	21.2652
10	8m, 8n,14b, 14j, 14q	0.7411	0.5938	0.8189	0.2124	0.2643	0.2246	20.0321
11	8h, 8n,14b, 14j, 14q	0.7446	0.6069	0.8049	0.2125	0.2636	0.2248	20.4098
12	8h, 8m,14b, 14j, 14q	0.8174	0.7297	0.7951	0.1827	0.2222	0.2240	31.3266
13	8h, 8m,8n,14j,14q	0.7329	0.5907	0.8530	0.2158	0.2671	0.2011	19.2058
14	8h, 8m,8n,14b,14q	0.5807	0.4553	-0.1237	0.2743	0.3127	0.4102	31.8527
15	8h, 8m,8n,14b,14j	0.7589	0.6289	0.7463	0.2145	0.2661	0.2142	22.0302
16	8m, 14j,14m,14q	0.7932	0.6048	0.3896	0.1882	0.2601	0.4689	28.1244
17	8m, 8n,14b, 14j, 14q	0.7411	0.5989	0.8189	0.2124	0.2643	0.2246	20.0328
18	8h, 8m,8n,14b	0.8200	0.7386	0.7139	0.1908	0.2299	0.1638	33.4037
19	8h, 8m,8n,14j	0.7518	0.6257	0.7949	0.2127	0.2612	0.2218	22.2091
20	8m, 8n,14j,14q	0.7330	0.5959	0.8537	0.2108	0.2594	0.2318	20.1364

In the above QSAR models, n is the number of molecules (Training set) used to derive the QSAR model, r<sup>2</sup> is the squared correlation coefficient, q<sup>2</sup> is the cross-validated correlation coefficient, pred\_r<sup>2</sup> 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<sup>2</sup> se, q<sup>2</sup> se and pred\_r<sup>2</sup>se are the standard errors terms for r<sup>2</sup>, q<sup>2</sup> and pred\_r<sup>2</sup> (smaller is better). R<sup>2</sup> is the correlation coefficient for observed vs. predicted biological activity.

Table-8: List of significant QSAR models with equation generated from PLSR (Best Five)

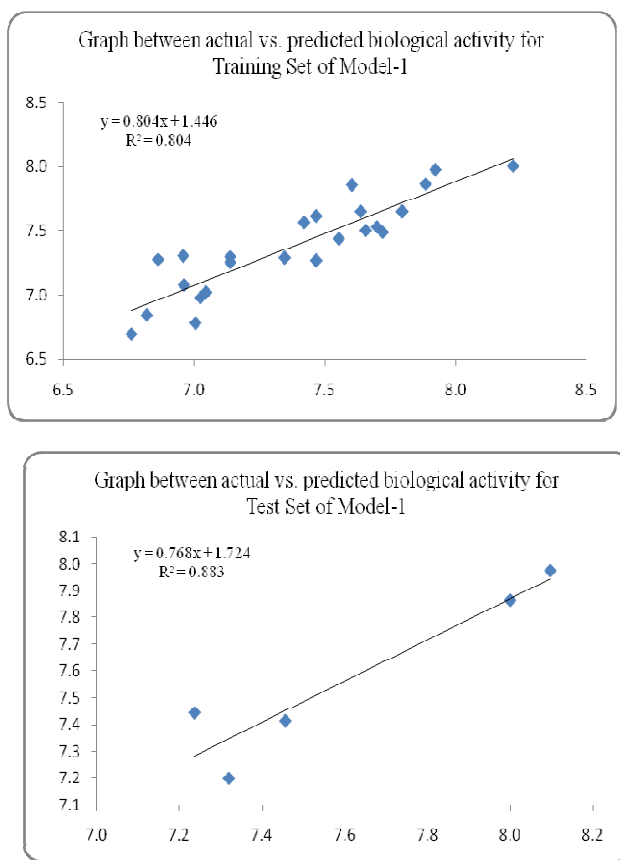
Model no	Method	Test set	Equation
1	Random selection method 85%/ trial 4/ PLS	08l,14a,14g, 14j, 14q	$pIC_{50} = 0.1439 T\_T\_N\_7 - 0.3681 T\_C\_N\_3 - 36.6941$ Average Potential + 0.3044 $T\_N\_Br\_4 + 9.0773$ Optimum Components = 3 n = 25 Degree of freedom = 21 F test = 28.7873 r2 = 0.8044 q2 = 0.7077 pred_r2 = 0.8959 r2 se = 0.1900 q2 se = 0.2323 pred_r2se = 0.1523
2	Random selection method 90%/ trial 2/ PLS	08b,08e,14q	$pIC_{50} = 0.4312 T\_N\_Br\_4 - 0.3947 T\_C\_N\_3 +$ 12.6985 BalabanIndexJ - 0.9659 XKMostHydrophobic - 6.7634 Optimum Components = 3 n = 27 Degree of freedom = 23 F test = 36.4300 r2 = 0.8261 q2 = 0.7679 pred_r2 = 0.8922 r2se = 0.1753 q2se = 0.2026 pred_r2se = 0.1887
3	Manual selection method/ trial 7/ PLS	08l,14a, 14j,14q	$pIC_{50} = -107.72$ chiV6chain - 0.1889 Radius Of Gyration + 16.3623 chi5chain + 0.0062 -ve Potential SurfaceArea + 6.5062 Optimum Components = 3 n = 26 Degree of freedom = 22 F test = 31.0539 r2 = 0.8090 q2 = 0.6983 pred_r2 = 0.8792 r2se = 0.1835 q2se = 0.2306 pred_r2se = 0.1886
4	Manual selection method/ trial 6/ PLS	08l,14g, 14j,14q	$pIC_{50} = -105.198$ chiV6chain - 0.1915 Radius Of Gyration + 16.7349 chi5chain + 0.0062 -ve Potential SurfaceArea + 6.3632 Optimum Components = 3 n = 26 Degree of freedom = 22 F test = 31.8296 r2 = 0.8127 q2 = 0.7094 pred_r2 = 0.8740 r2 se = 0.1817 q2 se = 0.2264 pred_r2se = 0.1937
5	Random selection method 90%/ trial 4/ PLS	08k,14j,14q	$pIC_{50} = -0.3751 T\_C\_N\_3 + 10.9998$ BalabanIndexJ + 0.3447 $T\_N\_Br\_4 - 8.0463$ XAAverageHydrophobicity - 4.2831 Optimum Components = 3 n = 27 Degree of freedom = 23 F test = 33.4229 r2 = 0.8134 q2 = 0.7186 pred_r2 = 0.8738 r2 se = 0.1781 q2 se = 0.2187 pred_r2se = 0.2353

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

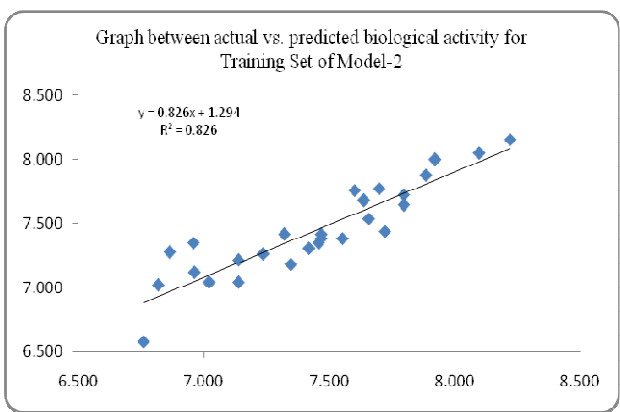
S. No.	Compound	Actual ( $pIC_{50}$ )	Predicted Biological Activity ( $pIC_{50}$ )	
			Model 1	Model 2
1	08a	7.022	6.979	7.039
2	08b	7.004	6.782	7.016*
3	08c	6.818	6.843	7.019
4	08e	7.046	7.022	7.011*
5	08g	6.863	7.279	7.276
6	08h	7.347	7.29	7.177
7	08i	7.468	7.614	7.417
8	08j	7.721	7.492	7.438
9	08k	7.42	7.566	7.308
10	08l	7.236	7.445*	7.264
11	08m	7.638	7.652	7.68
12	08n	6.962	7.076	7.119
13	08p	7.699	7.534	7.774
14	14a	7.319	7.2*	7.416
15	14b	7.468	7.269	7.379
16	14c	6.759	6.699	6.576
17	14d	7.137	7.299	7.039
18	14e	7.553	7.439	7.384
19	14f	6.958	7.301	7.348
20	14g	7.456	7.415*	7.347
21	14h	7.796	7.65	7.724
22	14i	7.796	7.652	7.643
23	14j	8.097	7.974*	8.05
24	14k	7.886	7.866	7.877
25	14l	7.602	7.856	7.758
26	14m	7.657	7.506	7.537
27	14n	8.221	8.005	8.156
28	14o	7.137	7.255	7.213
29	14p	7.921	7.975	7.999
30	14q	8	7.864*	8.264*

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

Graph plotted between actual and predicted biological activity for Model 1 and 2 are shown in Figure 2 and 3 respectively.



**Figure-02: Graph between actual and predicted biological activity of training and test set for Model-1.**



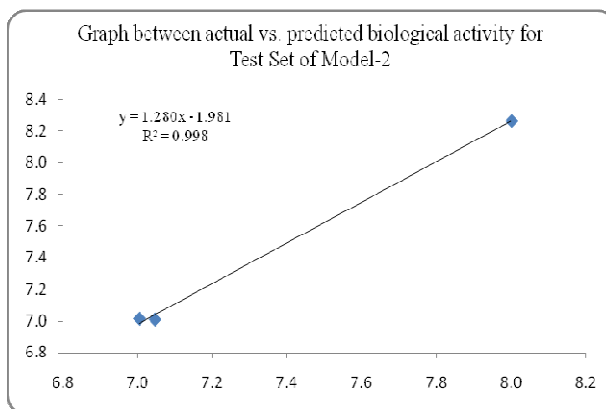


Figure-03: Graph between actual and predicted biological activity of training and test set for Model-2.

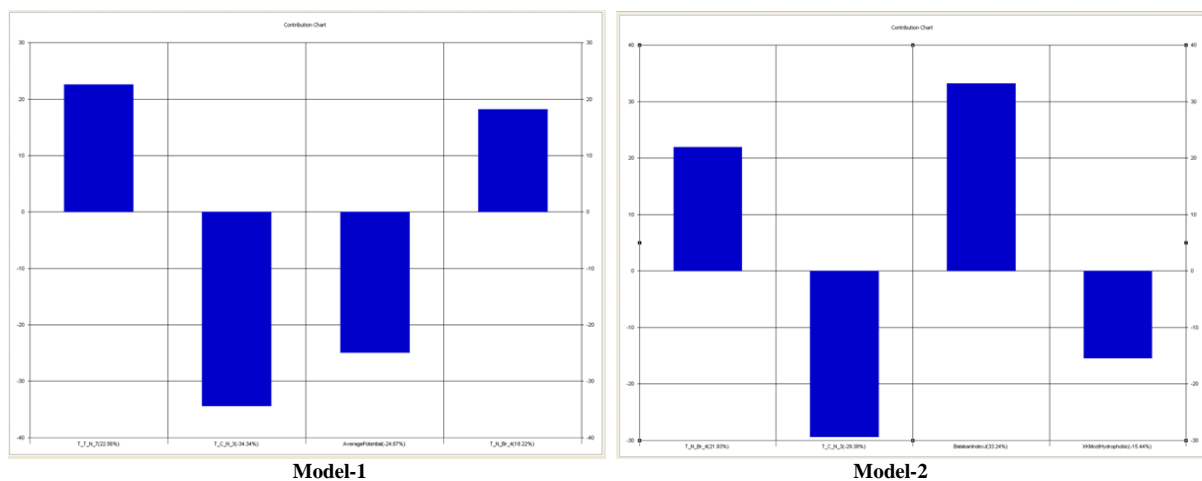


Figure-04: Contribution plot for Model 1-2.

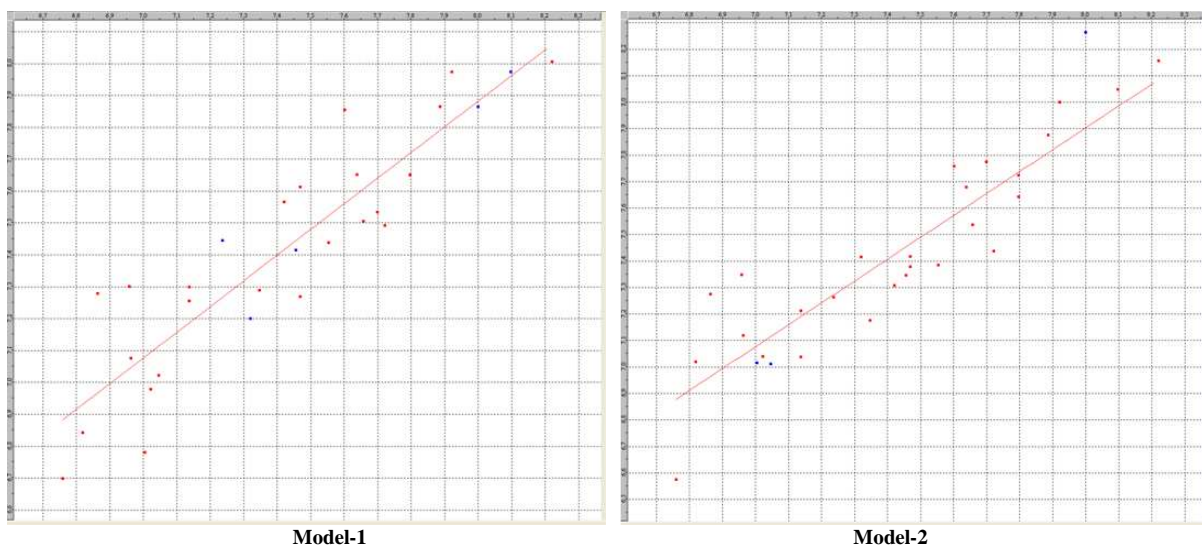


Figure-05: Data fitness plot for Model 1-2.

From table-08, the equation of Model-01 explains 82% ( $r^2=0.8163$ ) of the total variance in the training set as well as it has internal ( $q^2$ ) and external ( $pred\_r^2$ ) predicative ability of 61 % and 99% respectively. Model- 02 explains 81% ( $r^2= 0.81$ ) of the total variance in the training set as well as it has internal ( $q^2$ ) and external ( $pred\_r^2$ ) predicative ability of 62% and 94% respectively. Model-03 explains 82% ( $r^2= 0.8164$ ) of the total variance in the training set as well as it has internal ( $q^2$ ) and external ( $pred\_r^2$ ) predicative ability of 64 % and 94 % 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-2, 3) it can be seen that the model is able to predict the activity of the training set quite well as well as external test set, providing confidence of the model.

#### **Interpretation of the Model 01 (Most significant)**

Among the significant models generated (Table-08), model 1 is the most significant one as it is having the highest predicted correlation coefficient value. The equation 1 explains 80% ( $r^2 = 0.8044$ ) of the total variance in the training set and has an internal ( $q^2$ ) and external ( $pred\_r^2$ ) predictive ability of ~71% ( $q^2 = 0.7077$ ) and ~90% ( $pred\_r^2 = 0.8959$ ) 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\_T\_N\_7 [This is the count of number of any atoms separated from any Nitrogen atom (single, double or triple bonded) by 7 bond distance in a molecule] on the biological activity indicated that higher value leads to better KDR/VEGFR-2 kinase inhibitory activity (compound 14n, 14p, 14h, 8j, 8p, 8m, etc.) whereas lower value leads to decrease activity (compound 8c, 14f, 8a, 8b, 8n, 14o, 14b, etc.). Also the positive coefficient value of T\_N\_Br\_4 [this is the count of number of Nitrogen atoms (single, double or triple bonded) separated from any Bromine atom by 4 bonds in a molecule] on the biological activity indicated that higher value leads to better KDR/VEGFR-2 kinase inhibitory activity (compound 14n, 14p, 14k, 8j, 8p, 8m, etc.) whereas lower value leads to decrease activity (compound 8c, 14f, 14c, 8g, 14o, 14d, etc.). Negative coefficient value of T\_C\_N\_3 [This is the count of number of Carbon atoms (single, double or triple bonded) separated from any Nitrogen atom (single, double or triple bonded) by 3 bond distance in a molecule] on the biological activity indicated that lower values leads to good KDR/VEGFR-2 kinase inhibitory activity (compound 14n, 14p, 14k, 8j, 8p, 8m, etc.) while higher value leads to reduced activity (compound 14c, 8n, 8g, 8c, etc.). Also negative coefficient value of AveragePotential on the biological activity indicated that lower values leads to good KDR/VEGFR-2 kinase inhibitory activity (compound 14k, 8k, 8p, 8j, 14l, 14e, etc.) while higher value leads to reduced activity (compound 14c, 8b, 8c, 14f, 8g, etc.).

Figure-04 represents the contribution chart showing contribution of the various descriptors playing important role in determining the KDR/VEGFR-2 kinase inhibitory activity and Figure-05 represents the data fitness plot. Contribution chart for model 1 reveals that the descriptors T\_T\_N\_7, T\_C\_N\_3, AveragePotential and T\_N\_Br\_4 contributing 22.56%, 34.34%, 24.87% and 18.22% respectively.

#### **Interpretation of the Model 2**

Among the significant models generated (Table-08), model 2 is the significant one as it is having the higher predicted correlation coefficient value. The equation 2 explains 83% ( $r^2 = 0.8261$ ) of the total variance in the training set and has an internal ( $q^2$ ) and external ( $pred\_r^2$ ) predictive ability of ~77% ( $q^2 = 0.7679$ ) and ~89% ( $pred\_r^2 = 0.8922$ ) 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 90 (Alpha Rand Pred  $R^2 = 0.1$ ) 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\_N\_Br\_4 [this is the count of number of Nitrogen atoms (single, double or triple bonded) separated from any Bromine atom by 4 bonds in a molecule] on the biological activity indicated that higher value leads to better KDR/VEGFR-2 kinase inhibitory activity (compound 14n, 14j, 14p, 8m, 8i, etc.) whereas lower value leads to decrease activity (compound 14f, 8g, 8c, 14c, 8a, etc.). Also positive coefficient value of BalabanIndexJ [distance based topological descriptor] on the biological activity indicated that higher value leads to better KDR/VEGFR-2 kinase inhibitory activity (compound 8p, 8m, 8j, 8k, 14h, etc.) whereas lower value leads to decrease activity (compound 14o, 14d, 14c, 14e, 14f, 8c, 8a, etc.). Negative coefficient value of T\_C\_N\_3 [This is the count of number of Carbon atoms (single, double or triple bonded) separated from any Nitrogen atom (single, double or triple bonded) by 3 bond distance in a molecule] on the biological activity

indicated that lower values leads to good KDR/VEGFR-2 kinase inhibitory activity (compound 14n, 14m, 14k, 14l, 8p, 8m, etc.) while higher value leads to reduced activity (compound 14c, 8c, 8a, 8n, etc.). Also negative coefficient value of XKMostHydrophobic [most hydrophobic value on vdW surface] on the biological activity indicated that lower values leads to good KDR/VEGFR-2 kinase inhibitory activity (compound 14k, 14h, 14e, 14l, etc.) while higher value leads to reduced activity (compound 8k, 8h, 14m, 8m, 8i, etc.).

Figure-04 represents the contribution chart showing contribution of the various descriptors playing important role in determining the KDR/VEGFR-2 kinase inhibitory activity and Figure-05 represents the data fitness plot. Contribution chart for model 2 reveals that the descriptors T\_N\_Br\_4, T\_C\_N\_3, BalabanIndexJ and XKMostHydrophobic contributing 21.97%, 29.30%, 33.24% and 15.44 % 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 2 and 3, 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.

### CONCLUSION

Quantitative structure activity relationship (QSAR) study was performed on a series of 3,4-disubstituted-1H-pyrazol-5(4H)-ones possessing KDR/VEGFR-2 kinase inhibitory activity for establishing quantitative relationship between biological activity and their physicochemical/ structural properties. Several statistical regression expressions were obtained using partial least squares regression (PLSR) analysis. Most statistical significant model generated, explains 80% ( $r^2 = 0.8044$ ) of the total variance in the training set and has an internal ( $q^2$ ) and external ( $pred\_r^2$ ) predictive ability of ~71% ( $q^2 = 0.7077$ ) and ~90% ( $pred\_r^2 = 0.8959$ ). In this model the positive coefficient value of T\_T\_N\_7 [This is the count of number of any atoms separated from any Nitrogen atom (single, double or triple bonded) by 7 bond distance in a molecule] on the biological activity indicated that higher value leads to better KDR/VEGFR-2 kinase inhibitory activity (compounds 14n, 14p, 14h, 8j, 8p, 8m, etc.) whereas lower value leads to decrease activity (compounds 8c, 14f, 8a, 8b, 8n, 14o, 14b, etc.). Also the positive coefficient value of T\_N\_Br\_4 [this is the count of number of Nitrogen atoms (single, double or triple bonded) separated from any Bromine atom by 4 bonds in a molecule] on the biological activity indicated that higher value leads to better KDR/VEGFR-2 kinase inhibitory activity (compounds 14n, 14p, 14k, 8j, 8p, 8m, etc.) whereas lower value leads to decrease activity (compounds 8c, 14f, 14c, 8g, 14o, 14d, etc.). Negative coefficient value of T\_C\_N\_3 [This is the count of number of Carbon atoms (single, double or triple bonded) separated from any Nitrogen atom (single, double or triple bonded) by 3 bond distance in a molecule] on the biological activity indicated that lower values leads to good KDR/VEGFR-2 kinase inhibitory activity (compounds 14n, 14p, 14k, 8j, 8p, 8m, etc.) while higher value leads to reduced activity (compounds 14c, 8n, 8g, 8c, etc.). Also negative coefficient value of AveragePotential on the biological activity indicated that lower values leads to good KDR/VEGFR-2 kinase inhibitory activity (compounds 14k, 8k, 8p, 8j, 14l, 14e, etc.) while higher value leads to reduced activity (compound 14c, 8b, 8c, 14f, 8g, etc.). Contribution chart reveals that the descriptors T\_T\_N\_7, T\_C\_N\_3, AveragePotential and T\_N\_Br\_4 contributing 22.56%, 34.34%, 24.87% and 18.22 % respectively.

In the present study an attempt has been made to identify the necessary structural and substituent requirements. From the present QSAR analysis, three best models were generated among which any one can be used for predicting the activity of the newly designed compounds in finding some more potent molecules. Finally, it is concluded that the work presented here will play an important role in understanding the relationship of physiochemical parameters with structure and biological activity. By studying the QSAR model one can select the suitable substituent for active compounds with maximum potency.

### Acknowledgement

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

## REFERENCES

- [1] D. Bour's, Y. Kusumanto, C. Meijer, N.H. Mulder, G.A.P. Hosper, *Pharmacol. Res.*, **2006**, 53, 89.
- [2] D. N. Kim, B. Lu, D. E. Hallahan, *Curr. Opin. Investig. Drugs*, **2004**, 5, 597.
- [3] P. Cohen, *Curr. Opin. Chem. Biol.*, **1999**, 3, 459.
- [4] A. Bennasroune, A. Gardin, D. Aunis, G. Cre´mel, P. Hubert, *Critical Rev. Oncol./Hematol.*, **2004**, 50, 23.
- [5] J. G. Christensen, R. Schreck, J. Burrows, P. Kuruganti, E. Chan, P. Le, J. Chen, X. Wang, L. Ruslim, R. Blake, K.E. Lipson, J. Ramphal, S. Do, J.J. Cui, J. M. Cherrington, D.B. Mendel, *Cancer Res.*, **2003**, 63, 7345.
- [6] S. Wilhelm, C. Carter, M. Lynch, T. Lowinger, J. Dumas, R.A. Smith, B. Schwartz, R. Simantov, S. Kelley, *Nat. Rev. Drug Discov.*, **2006**, 5, 835.
- [7] E. Cabebe, H. Wakelee, *Drugs Today (Barc)*, **2006**, 42, 387.
- [8] D. McKillop, E.A. Partridge, J.V. Kemp, M.P. Spence, J. Kendrew, S. Barnett, P.G. Wood, P.B. Giles, A.B. Patterson, F. Bichat, N. Guilbaud, T.C. Stephens, *Mol. Cancer Ther.*, **2005**, 4, 641.
- [9] L.B. Jakeman, M. Armanini, H.S. Phillips, N. Ferrara, *Endocrinology*, **1993**, 133, 848.
- [10] C. De Vries, J.A. Escobedo, H. Ueno, K. Houck, N. Ferrara, L.T. Williams, *Science*, **1992**, 255, 989.
- [11] B.I. Terman, M. Dougher-Vermazen, M.E. Carrion, D. Dimitrov, D.C. Armellino, D. Gospodarowicz, P. Bohlen, *Biochem. Biophys. Res. Commun.*, **1992**, 187, 1579.
- [12] M. Klagsbrun, M.A. Moses, *Chem. Biol.*, **1999**, 6, R217.
- [13] L.M. Strawn, G. McMahon, H. App, R. Schreck, W.R. Kuchler, M.P. Longhi, T.H. Hui, C. Tang, A. Levitzki, A. Gazit, I. Chen, G. Keri, L. Orfi, W. Risau, I. Flamme, A. Ullrich, K.P. Hirth, L.K. Shawver, *Cancer Res.*, **1996**, 56, 3540.
- [14] K. Paz, Z. Zhu, *Front. Biosci.*, **2005**, 10, 1415.
- [15] R. Tripathy, A. Reiboldt, P.A. Messina, M. Iqbal, J. Singh, E.R. Bacon, T.S. Angeles, S.X. Yang, M.S. Albom, C. Robinson, H. Chang, B.A. Ruggeri, J.P. Mallamo, *Bioorgan. Med. Chem. Lett.*, **2006**, 16, 2158.
- [16] Y. Kakiuchi, N. Sasaki, S.M. Masuoka, H.M. Murofushi, K.M. Murofushi, *Biochem Biophys Res Com.*, **2004**, 320, 1351–1358.
- [17] J.S. Casas, E.E. Castellano, J. Ellena, M.J. Perez-Paralle, A. Sanchez, A.S. Gonzalez, J. Sordo, A. Touceda, *J. Inorg. Biochem.*, **2008**, 102, 33–45.
- [18] M.F. Bran, A. Gradillas, A.G. Ovalles, B. Lo´pez, A. Nuria, L. Francisco, D.M. Mingarroa, *Bioorg Med Chem.*, **2006**, 14, 9–16.
- [19] R. Tripathy, A. Ghose, J. Singh, E.R. Bacon, T.S. Angeles, X.S. Yang, M.S. Albom, L.D. Aimone, J.L. Herman, J.P. Mallamo, *Bioorg Med Chem Lett.*, **2007**, 17, 1793-1798.
- [20] VLifeSciences Technology Pvt. Ltd., Pune-411045, Web: vlifesciences.com.