

QSAR study on N- (Aryl)-4-(Azolyethyl) Thiazole-5-Carboxamides: Novel Potent Inhibitors of VEGF Receptors I and II

Asha Patel^a, C.Karthikeyan^b, N.S. Hari Narayana Moorthy^b and Piyush Trivedi^{b,*}

^aDepartment of Pharmacy, SGSITS, Indore-452003, Madhya Pradesh, India; ^bSchool of Pharmaceutical Sciences, Rajiv Gandhi Technological University, Airport Bypass Road, Gandhi Nagar, Bhopal – 462036, Madhya Pradesh, India

Abstract: QSAR study on N- (Aryl)-4-(azolyethyl) thiazole-5-carboxamides analogues, which are novel potent inhibitor of VEGF receptor II and I, were performed using topological, electronic and physicochemical descriptors. The results obtained demonstrate in detail, which specify that topological descriptors of the compounds play a significant role in developing QSAR models. The significance of presence and absence of substituents on particular position is successfully explored with the help of indicator variables. The results are critically discussed on the basis of multiple linear regression parameters.

Key Words: QSAR, VEGF receptor, Topological descriptors.

INTRODUCTION

Angiogenesis, the growth and expansion of blood vessels from pre-existing vasculature, is involved in a variety of disease states that include diabetic retinopathy, rheumatoid arthritis, and tumor growth [1, 2]. Although angiogenesis is a highly complex process, there is a large body of evidence that suggests that an endothelial cell-specific mitogen, vascular endothelial growth factor (VEGF), is a major regulator of these events [3, 4]. The biological effects of VEGF are mediated by two-receptor tyrosine kinases known as VEGFR-1 (also known as Flt-1) and VEGFR-2. It has been demonstrated that the disruption of VEGF signaling can retard angiogenesis and inhibit tumor growth. Numerous compounds, such as PTK 787, ZD 6474, SU6668, SU11248, CHR200131, CP547632, AG13736, CEP7055/CEP5214, KR633 have been shown to be effective in this manner and other molecules progressing to the clinic for further evaluation [5, 6]. Preclinical and clinical data strongly support the involvement of specific VEGFR in the formation and progression of a subset of solid tumors.

Quantitative structure–activity relationships (QSAR) have been broadly used for some years mainly in medical research [7-11]. It is basically concerned with the correlation of structure with activity. Several physicochemical, topological and electronic descriptors are used in QSAR studies. The present study involves QSAR analysis of N- (Aryl)-4-(azolyethyl) thiazole-5-carboxamides, which are novel potent inhibitors of VEGF receptors 1 and 2 [12].

EXPERIMENTAL

Biological Activity

The series consist of total 31 compounds but only 23 compounds have well defined activity against VEGFR-2. All

the biological activity data (IC_{50} in μM) have been converted to negative logarithmic molar dose (pIC_{50}) in order to linearly relate free energy change of the interaction of compounds with receptor and to reduce the skewness of the data set. The general structure of these analogues along with their biological activity data are shown in Table 1.

Molecular Descriptors

The Compounds in the series were sketched using Chem Draw module of ChemOffice 200 [13] and the sketched structures were subsequently used for the calculation of molecular descriptors available in QSAR software Modeslab 1.5 [14].

Topological [15] and physicochemical descriptors were calculated for all molecules using QSAR software Modeslab whereas electronic descriptors were calculated on the Chem3D software using the “Compute Properties Module”.

The structures were energy minimized to calculate electronic descriptors.

(i) MM₂ server was used for energy minimization; minimum RMS gradient used was 0.100.

(ii) In the subsequent step MOPAC server was used for energy minimization, minimum RMS gradient used was 0.001.

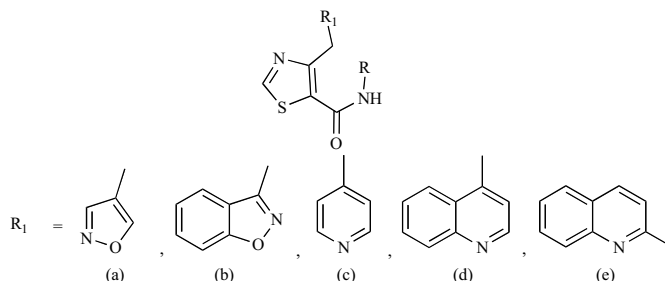
Indicator Variables

Indicator variables are used to account for certain features which can't be described by the continuous variables. In QSAR equation they normally used to describe a certain structural element, be it a substituent or another molecular fragment. The present study employs two indicator variables I_1 and I_2 which represents the 4-pyridine nucleus and isoxazole nucleus respectively.

REGRESSION ANALYSIS

Biological activity was taken as dependent variable and the calculated molecular descriptors were taken as independent variables. A correlation analysis was performed between

*Address correspondence to this author at the School of Pharmaceutical Sciences, Rajiv Gandhi Technological University, Airport Bypass Road, Gandhi Nagar, Bhopal – 462036, Madhya Pradesh, India;
Tel: +91-755-2678883; Fax: +91-755-2742006;
E-mail: prof_piyushtrivedi@yahoo.com

Table 1. Structure of N- (Aryl)-4-(azolyethyl) thiazole-5-carboxamides Analogues Along with their Biological Activity Data

Comp. No.	R	R1	Activity		
			VEGFR-2	VEGFR-1	VEGFR-2 Cell Based ELISA
1	4-Cl (C6H4)	a	5.7329	5.3372	5.6383
2	4-Cl (C6H4)	c	6.8861	6.5086	7.0223
3	4-Cl (C6H4)	d	5.4711	5.1643	NA
4	4-t-Bu (C6H4)	a	5.8386	NA	5.7878
5	46-t-Bu (C6H4)	c	6.9586	6.1427	7.1367
6	4-t-Bu (C6H4)	d	5.6003	5.2628	NA
7	4-i-Pr (C6H4)	a	5.7932	5.1931	NA
8	4-i-Pr (C6H4)	c	6.8539	6.0757	6.9208
9	4-ClF ₂ CO (C6H4)	a	5.7520	NA	5.6478
10	4-ClF ₂ CO (C6H4)	c	6.9586	6.1192	7.0706
11	3-Me (C6H4)	c	6.2518	5.3316	NA
12	4-N-Morpholino- (C6H4)	a	5.4341	NA	NA
13	4-N-Morpholino- (C6H4)	c	6.1487	5.6326	6.2366
14	3,4-Cl (C6H4)	a	5.9031	5.2660	NA
15	3,4-Cl (C6H4)	c	6.7447	6.2291	7.0315
16	4-Cl-3-CF ₃ (C6H3)	a	5.6737	5.2741	NA
17	4-Cl-3-CF ₃ (C6H3)	c	6.5686	6.1427	6.7212
18	2-F-4-Me (C6H3)	a	5.7799	5.5199	5.5607
19	2-F-4-Me (C6H3)	c	6.5086	6.2007	6.6383
20	3,4-methylenedioxy	c	6.4318	6.1249	6.4815
21	4-Br (C6H4)	c	6.0655	5.4145	5.9914
22	4-Ph (C6H4)	c	5.9666	NA	NA
23	4-PhO (C6H4)	c	5.7167	NA	NA

dependent and independent variables. QSAR models were generated through forward stepwise multiple linear regression analysis using the method of least squares adopted by statistical program SYSTAT [16]. In this method, descriptors were selected on the basis of F-statistics and tolerance.

Descriptor having F value greater than or equal to 4 were selected and less than or equal to 3.9 were removed. The statistical measures used for the evaluating the generated models are correlation coefficient (r), squared correlation coefficient (r²), Fischer ratio value and standard deviation. Furthermore,

Table 2. Correlation Matrix for Models 1-3

Models	Descriptors	I ₂	X (Ch)(6)	Xv (PC)(5)
Model 1	I ₂	1.0000		
	X (Ch)(6)	0.6026	1.0000	
	Xv (PC)(5)	0.2573	0.1424	1.0000
Model 2	Descriptors	I ₁	Xv (P)(5)	
	I ₁	1.0000		
	Xv (P)(5)	0.6345	1.0000	
Model 3	Descriptors	I ₁	DL	
	I ₁	1.0000		
	DL	0.0377	1.0000	

the obtained models were also checked for multicollinearity problem by calculating correlation matrix for the descriptors in the QSAR models (Table 2). The selected models were validated by Leave One Out (LOO) [17] method employing validation software VALSTAT [18] and the validation parameters (Q₂, r²_{pred}, S_{PRESS} and S_{DEP}) were calculated for generated models. The Z score method was adopted for the detection of outliers. Z Score can be defined as absolute difference between the value of the model and the activity field, divided by the square root of the mean square error of the data set. Any compound which shows a value of Z score higher than 2.5, during generation of a particular QSAR model, is considered as outlier (Table 3).

RESULTS AND DISCUSSION

Correlations between different VEGFR inhibitory activities and calculated variables were established through forward stepwise multiple linear regressions using the method of least squares. Various models were generated and best models were selected on the basis of their statistical significance. It was observed from QSAR study that VEGFR inhibitory activity of N-(Aryl)-4-(azolyethyl) thiazole-5-carboxamides can be successfully explained in terms of topology of the molecule.

Statistically significant QSAR models generated for different inhibitory activity data are as follow:

Against VEGFR-2

Model-1

$$BA = [5.35111 (\pm 0.639971)] + I_2 [1.21889 (\pm 0.28205)] + X (Ch)(6) [-4.40829 (\pm 1.95195)] + Xv (PC)(5) [0.399077 (\pm 0.307898)]$$

n=23, r=0.905808, r²=0.820487, std=0.229413, F=28.9474 (F_(3,19) = 4.94) Q²=0.753, S_{PRESS}=0.268718, S_{DEP}=0.244236

Against VEGFR-1

Model-2

$$BA = [9.20894 (\pm 0.694867)] + I_1 [-1.21417 (\pm 0.190454)] + Xv (P)(5) [-1.49191 (\pm 0.317222)]$$

n=16, r=0.968464, r²=0.937922, std=0.125256, F=98.2067 (F_(2,13) = 6.36), Q²= 0.897918, S_{PRESS}= 0.160621, S_{DEP}= 0.144782

Table 3. Z-Score Values of Compounds in QSAR Models 2 and 3

Compound No.	Z-Score Value	
	Model 2	Model 3
1	-0.5728	-1.4250
2	1.7513	0.2127
3	-2.0527	0.2788
4	0.2142	1.3129
5	0.3934	-0.1671
6	-0.7658	0.8123
7	-0.9185	1.6984
8	-0.1370	-1.2502
9	-2.8640*	0.2240
10	1.4548	-0.6878
11	0.0883	0.3340
12	0.6265	-1.5231
13	0.7705	0.1803
14	0.4977	-2.8243*
15	0.4799	
16	-1.4044	
17	-0.4254	
18	-3.1201*	

*outlier.

Against VEGFR-2 (Cell Based ELISA)**Model-3**

BA = [7.16667 (\pm 0.251731)] + I₁ [-1.18213 (\pm 0.242998)] + DL [-0.102785 (\pm 0.0607741)]

n=13, r=0.963524, r²=0.928378, std=0.178149, F=64.8112 (F_(2,10) = 7.56), Q²= 0.876577, S_{PRESS}= 0.233862, S_{DEP}= 0.20511

Where, n is the number of data points, r is correlation coefficient, r² is squared correlation coefficient, std is standard deviation, F represents Fischer ratio between the variances of calculated and observed activities, values in parentheses that follows calculated F-value, are tabulated F-values at 99% confidence interval, figures with \pm sign which are

given in the parentheses within the equation, are 95% confidence limits associated with the regression coefficients, chance is the ratio of the equivalent regression equations to the total number of randomized sets, Q² is cross validated squared correlation coefficient, S_{PRESS} and S_{DEP} correspond to standard deviation based on predicted residual sum of squares and standard deviation of error of prediction respectively.

All the QSAR models are significant at 99% level, which is shown by their greater calculated F value in comparison to the tabulated F value. Each model can explain a good percentage of inhibitory activity, lowest being 82% for inhibition of VEGFR-2 and highest being 93% for inhibition of VEGFR-1 as shown by their r² values. Accuracy in the analysis is shown by low values of std. High Q² and low

Table 4. Descriptors Involved in QSAR Models

Sr. No.	Model 1(VEGFR-2)			Model 2(VEGFR-1)		Model 3(VEGFR-2, Cell Based ELISA)	
	I ₂	X (Ch)(6)	Xv (PC)(5)	I ₁	Xv (P)(5)	I ₁	DL
1	0	0.0833	1.624979	1	1.736517	1	1.11412
2	1	0.1854	1.685507	0	1.94684	0	1.74132
3	0	0.2347	2.305011	0	2.550603	-	-
4	0	0.0833	2.596807	-	-	1	2.35507
5	1	0.1854	2.657335	0	2.072006	0	2.36905
6	0	0.2347	3.276839	0	2.675768	-	-
7	0	0.0833	1.945869	1	1.818022	-	-
8	1	0.1854	2.006398	0	2.028345	0	2.12754
9	0	0.0833	2.026945	-	-	1	4.56119
10	1	0.1854	2.087473	0	2.06029	0	3.62213
11	1	0.1854	1.735986	0	1.862689	-	-
12	0	0.1854	2.085257	-	-	-	-
13	1	0.2874	2.145786	0	2.510828	0	7.07095
14	0	0.0680	2.0627	1	1.835941	1	2.26624
15	1	0.1701	2.1232	0	2.046263	0	1.66934
16	0	0.0680	2.2122	1	1.883844	1	3.87878
17	1	0.1701	2.2728	0	2.094166	0	3.24524
18	0	0.0680	1.7229	1	1.696306	1	4.65229
19	1	0.1701	1.7834	0	1.906628	0	2.73091
20	1	0.1701	1.8540	0	2.033897	0	6.95145
21	1	0.1854	1.8077	0	2.091344	0	1.70648
22	1	0.2875	1.9966	-	-	-	-
23	1	0.2874	1.7912	-	-	-	-

Compounds have no well-defined activity.

Table 5. Observed and Predicted Activities for VEGFR Inhibition Calculated by Using the QSAR Models

Sr. No.	Model 1(VEGFR-2)		Model 2(VEGFR-1)		Model 3(VEGFR-2, Cell Based ELISA)	
	Obs. Activity	Pred. Activity	Obs. Activity	Pred. Activity	Obs. Activity	Pred. Activity
1	5.732828	5.60561	5.337242	5.42121	5.638272	5.99274
2	6.886057	6.36819	6.508638	6.26516	7.022276	6.98001
3	5.471083	5.13712	5.164309	5.49741	NA	NA
4	5.838632	6.07109	NA	NA	5.787812	5.72612
5	6.958607	6.75492	6.142668	6.11475	7.136677	6.88831
6	5.600326	5.65363	5.262807	5.18238	NA	NA
7	5.793174	5.7555	5.193142	5.30488	NA	NA
8	6.853872	6.52617	6.075721	6.19743	6.920819	6.95294
9	5.752027	5.7982	NA	NA	5.647817	5.46071
10	6.958607	6.54909	6.119186	6.13712	7.070581	6.75973
11	6.251812	6.46717	5.331614	NA	NA	NA
12	5.434152	5.35011	NA	NA	NA	NA
13	6.148742	6.16134	5.632644	5.40849	6.236572	6.57579
14	5.90309	5.86979	5.266001	5.2531	7.031517	6.98668
15	6.744727	6.65778	6.229148	6.14676	6.721246	6.84731
16	5.673664	5.97851	5.274088	5.16023	5.560667	5.48309
17	6.568636	6.75428	6.142668	6.0782	6.638272	6.92126
18	5.779892	5.72951	5.519993	5.44889	6.481486	6.43408
19	6.508638	6.53445	6.200659	6.40239	5.9914	NA
20	6.431798	6.5734	6.124939	6.18117	NA	NA
21	6.065502	6.51459	5.414539	NA	NA	NA
22	5.966576	6.1309	NA	NA	NA	NA
23	5.716699	6.11351	NA	NA	NA	NA

S_{PRESS} and S_{DEP} values in each model reflect their good predictive potential.

The best correlation for modeling VEGFR-2 inhibition for N- (Aryl)-4-(azolyethyl) thiazole-5-carboxamides comprises of I_2 , Xv (PC)(5) and X (Ch)(6). I_2 is the indicator variable given for the presence of pyridine nucleus. Indicator variables are used to account for certain features which can't be described by the continuous variables. In QSAR equation they normally used to describe a certain structural element, be it a substituent or another molecular fragment (Table 4). I_2 is positively correlated with the activity. It may be observed from Table 1, compounds 2, 5, 8, 10, 11, 13, 17, 19, 20, 21 having pyridine nucleus show more activity in comparison to other compounds. It shows importance of pyridine nucleus for the activity. X (Ch)(6) is a topological descriptor. It is the sixth order chain type molecular connectivity index. The chain-type molecular connectivity indices describe the type

of rings that are present in a molecule as well as the substitution patterns on those rings. X (Ch)(6) corresponds to the 4-pyridine nucleus. It is negatively correlated with the activity in the model. Thus, 4-pyridinyl moiety is important for the VEGFR-2 inhibitory activity. Xv (PC)(5) is a topological descriptor. It denotes fifth order valence path cluster type molecular connectivity index. The path/cluster indices describe mainly local structural properties, such as the extent or degree of branching in a molecule. They are highly sensitive to changes in branching, and their value rapidly increases with the degree of branching. It is positively correlated with the activity indicates that inhibitory activity will decrease with molecular branching.

Model for VEGFR-1 inhibitory activity highlights the linear relationship between VEGFR-1 inhibitory activity of the molecules and the molecular descriptors I_1 and Xv (P)(5). I_1 is the indicator variable given for the presence of isoxazole

nucleus in the structure. I_1 is negatively correlated with the VEGFR-1 inhibitory activity. It shows isoxazole nucleus is inimical to the VEGFR-1 inhibitory activity. X_v (P)(5) is a topological descriptor. It is the fifth order path type valence connectivity indices. It is negatively correlated with the activity in model 2. X_v (P)(5) is similar to X (P)(5) with additional heteroatom and valence state information. It's negative correlation with activity indicates that VEGFR-1 inhibitory activity will increase with presence of heteroatom.

The best correlation for modeling VEGFR-2 (Cell based ELISA) inhibition N- (Aryl)-4-(azolylethyl) thiazole-5-carboxamides comprises of I_1 and DL. I_1 is the indicator variable given for the presence of isoxazole nucleus in the structure. I_1 is negatively correlated with the VEGFR-2 (Cell based ELISA) inhibitory activity. It shows isoxazole nucleus is not conducive for the VEGFR-2 inhibitory activity. DL is a electronic descriptor. It is determined by the dipole moment. The dipole moment is the first derivative of the energy with respect to an applied electric field. It measures the asymmetry in the molecular charge distribution and is reported as a vector in three dimensions. The dipole length is the summarized vector of three dimensions. If the branches have narrow dimension, calculated dipole length will be high, if wide enough to be a plane, the value is 0. It is negatively correlated with the VEGFR-2 (Cell based ELISA) activity, which implies that if dipole length will be high than the molecule will be less active.

Although, generation of QSAR models with good statistical significance is of paramount importance, the models should also exhibit good predictive ability. The predictive ability of the models was gauged by a cross-validation procedure following a Leave-One-Out scheme. Furthermore, values and predicted activity values calculated by using the obtained models in Table 5 and graphical representation of the same are depicted in Graph 1. This comparison together with the graphical plot provides ample evidence for the good predictive potential of the models generated for modeling VEGFR inhibitory activity of N- (Aryl)-4-(azolylethyl) thiazole-5-carboxamides.

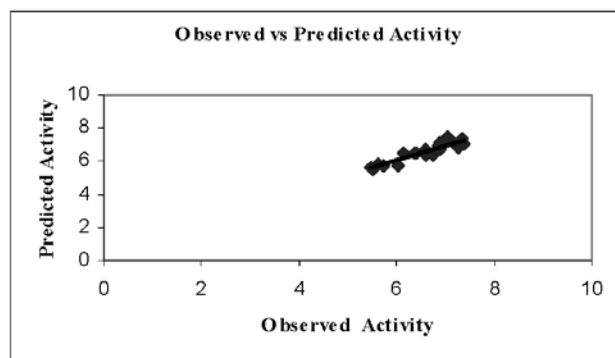
CONCLUSION

The present study gives rise to quantitative models to be capable of good prediction of VEGFR inhibition by N- (Aryl)-4-(azolylethyl) thiazole-5-carboxamides. The results of the QSAR study suggest that I_2 and X_v (PC)(5) are positively correlated with the VEGFR-2 inhibitory activity. It indicates pyridine nucleus is important for VEGFR-2 inhibitory activity and activity will increase with decrease in molecular branching. Further, negative correlation of the X (Ch)(6) with VEGFR-2 inhibitory activity indicates pyridine nucleus is important for the activity.

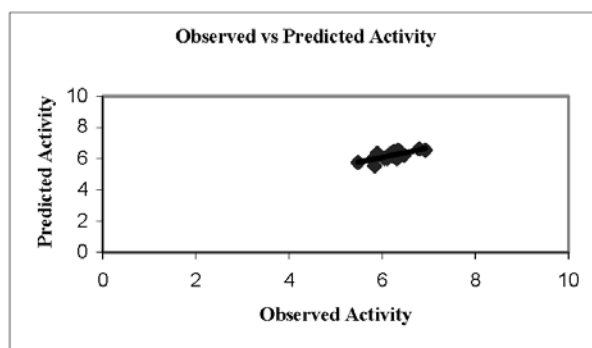
I_1 is negatively correlated with the VEGFR-1 inhibitory activity, which indicates that isoxazole nucleus is not required for VEGFR-1 inhibitory activity. X_v (P)(5) is negatively correlated with VEGFR-1 inhibitory activity, which implies that inhibitory activity will increase with presence of heteroatom.

I_1 is negatively correlated with the VEGFR-2 (cell based ELISA) inhibitory activity indicates isoxazole nucleus is not

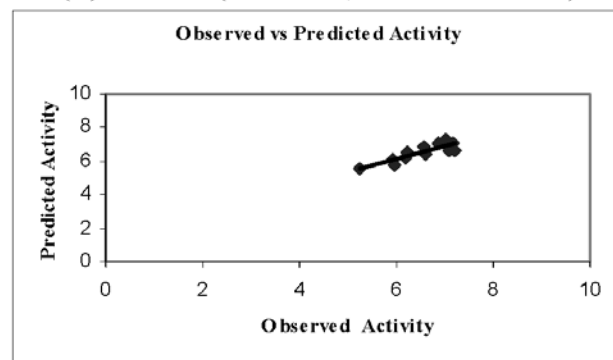
(A) Model-1 (VEGFR-2)



(B) Model-2 (VEGFR-1)



(C) Model-3 (VEGFR-2, cell based ELISA)



Graph 1. Scatter plot between predicted activity and observed activity for models (A) Model-1 (VEGFR-2).

required for VEGFR-2 inhibitory activity. Dipole length is negatively correlated with the VEGFR-2 (cell based ELISA) inhibitory activity, which indicates substituents which are capable of forming dipole should not present in the molecule for VEGFR-2 inhibitory activity.

ACKNOWLEDGEMENTS

The author likes to thank All India Council for Technical Education, New Delhi for providing fellowship. The authors wish to thank the Director, Shri G.S. Institute of Technology & Science, Indore for providing the necessary facilities for

undertaking this research work. One of the author Mr. C. Karthikeyan wish to acknowledge CSIR, India for providing Senior research fellowship.

REFERENCES

- [1] Folkman, J. Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nat. Med.*, **1995**, *1*(1), 27-31.
- [2] Carmeliet, P.; Jain, R.K. Angiogenesis in cancer and other diseases. *Nature*, **2000**, *407*(6801), 249-257.
- [3] Robinson, D.R.; Wu, Y.M.; Lin, S.F. The protein tyrosine kinase family of the human genome. *Oncogene*, **2000**, *19*(49), 5548-5557.
- [4] Jakeman, L.B.; Winer, J.; Bennett, G.L.; Altar, C.A.; Ferrara, N. Binding sites for vascular endothelial growth factor are localized on endothelial cells in adult rat tissues. *J. Clin. Invest.*, **1992**, *89*(1), 244-253.
- [5] <http://www.clinicaltrials.gov>.
- [6] Hoekman K.; Pinedo, H.M. Angiogenesis: a potential target for therapy of soft tissue sarcomas. *Cancer Treat. Res.*, **2004**, *120*, 169-180.
- [7] Gonzalez-Diaz, H.; Bastida, I.; Castan˜edo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H.S.; de Armas, R.R. Simple stochastic fingerprints towards mathematical modelling in biology and medicine. 1. The treatment of coccidiosis. *Bull. Math. Biol.*, **2004**, *66*(5), 1285-1311.
- [10] Gonzalez-Diaz, H.; Gia, O.; Uriarte, E.; Hernadez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J.A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L.A.; Cabrera, M.A. Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design I: discovery of anticancer compounds. *J. Mol. Model.*, **2003**, *9*(6), 395-407.
- [11] Gonzalez-Diaz, H.; Caballero, J.; Tundidor-Camba, A.; Helguera, A.M.; Fernandez, M. Modeling of farnesyltransferase inhibition by some thiol and non-thiol peptidomimetic inhibitors using genetic neural networks and RDF approaches. *Bioorg. Med. Chem.*, **2006**, *14*(1), 200-213.
- [12] Kiselyov, A.S.; Semenova, M.; Semenov, V.V. Heteroaryl imidazoles: a novel dual inhibitors of VEGF receptors I and II. *Bioorg. Med. Chem. Lett.*, **2006**, *16*(5), 1440-1444.
- [13] CS Chem office 2001 software package, supplied by Cambridge Software Company, 100, Cambridge Park Drive, Cambridge MA 02140-2317, USA.
- [14] www.modeslab.com.
- [15] Kier, L.B.; Hall, L.H. *Molecular Connectivity in Chemical and Drug Research*, Academic press: New York, **1976**, 320.
- [16] SYSTAT is statistical software, copyright SYSTAT Inc.
- [17] Kubinyi, H. An evolutionary algorithm. *Quant. Struct. Act. Relat.*, **1994**, *13*, 285-294.
- [18] VALSTAT is a Statistical Program Developed by Arun Kumar Gupta, SGSITS, Indore, **2004**.