informa
healthcare

Modeling of vascular endothelial growth factor receptor 2 (VEGFR2) kinase inhibitory activity of 2-anilino-5-aryloxazoles using chemometric tools

B. K. SHARMA[1], S. K. SHARMA[1], P. SINGH[1], SUSHEELA SHARMA[2], & Y. S. PRABHAKAR[3]

[1]*Department of Chemistry, S.K. Government College, Sikar 332 001, India,* [2]*Department of Engineering Chemistry, Sobhasaria Engineering College, Sikar 332 021, India, and* [3]*Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow 226 001, India*

(*Received 25 August 2007; accepted 16 November 2007*)

Abstract

The structure-activity models of the VEGFR2 kinase inhibitory activity of the derivatives of 2-anilino-5-aryloxazole have been investigated using Combinatorial Protocol in Multiple Linear Regression (CP-MLR) with nearly 500 topological descriptors which were calculated from DRAGON software. Among the descriptor classes considered collectively in the study, the inhibitory activity was, however, correlated with simple functional (FUN), topological (TOPO), atom centered fragments (ACF), molecular walk counts (MWC) and 2D-autocorrelation (2D-AUTO) descriptors. The developed models and participating descriptors in them have suggested that the substitutional modifications in the 2-anilino-5-aryloxazole moiety may have sufficient scope in optimization of prevailing inhibitory activity of these analogues.

Keywords: *Quantitative structure-activity relationship (QSAR), vascular endothelial growth factor receptor 2 kinase, 2-anilino-5-aryloxazoles, combinatorial protocol in multiple linear regression (CP-MLR), VEGFR2 inhibition*

Introduction

Mitogenic endogeneous proteins have important function in the formation and growth of solid tumors (angiogenesis) [1]. Among these proteins vascular endothelial growth factor (VEGF) controls a key step in the angiogenesis. It is up regulated by tumor cells and induces mitogenic response on binding to the tyrosine kinase receptor VEGFR2 (KDR/Flk-1) of nearby endothelial cells [2]. Hence, this pathway has attracted widespread interest in anticancer therapy [3]. A variety of compounds such as anilinophthalazine [4], anilinoquinazoline [5], indolinone [6] and isothiazole [7] are known to inhibit the VEGFR2. Recently, Harris et al. [8] have explored a new chemical class, 2-anilino-5-aryloxazoles (Figure 1), as VEGFR2 inhibitors. The variation in the chemical space of these analogues is focused around the substituents of 2-anilino and 5-phenyl moieties of the structure. In order to investigate the scope of chemical space of 2-anilino-5-aryloxazoles as VEGFR2 inhibitors, a high dimensional quantitative structure-activity relationship (QSAR) study has been undertaken on these analogues to rationalize their activity profile. For this, it is necessary to characterize the molecules or their varying structural fragments from different perspectives. Among different methods, graph theoretical approaches provide large number of structural indices characteristic to the molecules and their functional units [9–13]. Moreover, when dealing with a large number of descriptors, for the optimum utilization of information content of the generated data sets, it is necessary to adopt typical protocol(s) to identify the best models as well as information rich descriptors corresponding to the

Correspondence: P. Singh, Department of Chemistry, S.K. Government College, Sikar 332 001, India. E-mail: psingh_sikar@rediffmail.com

ISSN 1475-6366 print/ISSN 1475-6374 online © 2009 Informa UK Ltd.
DOI: 10.1080/14756360801915351

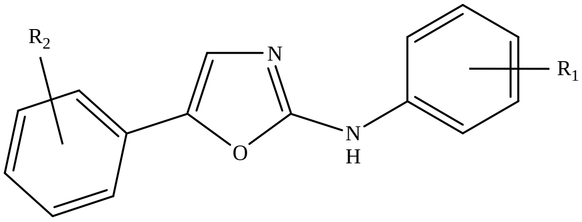

Figure 1. Structure of 2-anilino-5-aryloxazole derivatives.

phenomenon under investigation. The Combinatorial Protocol in Multiple Linear Regression (CP-MLR) [14–19] is an approach among many others to address the model evolution in high-dimensional QSAR studies. The aim of present communication is therefore, to establish QSAR between the reported VEGFR2 inhibitory activity of 2-anilino-5-aryloxazoles and the molecular descriptors calculated from DRAGON software [12] using the CP-MLR analysis.

Materials and method

Data set

In this study 2-anilino-5-aryloxazole analogues (Table I) have been considered from the literature report [8] along with their tyrosine kinase receptor, vascular endothelial growth factor receptor 2 (VEGFR2) (KDR/Flk-1) inhibitory activity in the form of logarithm of the inverse of inhibitory concentration (pIC_{50} where IC_{50} is in moles per liter against VEGFR2). The structures, for the varying R_1 and R_2 substituents of respective compounds, have been generated in ChemDraw [20] using the standard procedure. These structures have been ported to

Table I. The observed, calculated and predicted VEGFR2 inhibition activity values of 2-anilino-5-aryloxazole derivatives (Figure 1 for structures).

			pIC_{50} (M)			
				Calcd		
S. No.	R1	R2	Obsd[*]	Eq(19)	Eq(20)	Prctd LOO
1	H	H	5.92	6.00	5.98	5.99
2	2-CN	H	6.06	6.20	6.18	6.20
3	2-OMe	H	6.42	6.34	6.33	6.32
4	3-CN	H	6.03	6.01	5.99	5.98
5	3-OPh	H	5.85	6.02	5.99	6.03
6	3-CH_2OH	H	6.30	6.08	6.06	6.03
7	3-SO_2Et	H	6.14	6.26	6.24	6.25
8	3-CH_3	H	6.06	5.96	5.94	5.92
9	4-SO_2NH_2	H	6.30	6.51	6.50	6.52
10	4-OPh	H	–[†]	6.05	6.02	–
11	3,4-$(OMe)_2$	H	6.85	6.71	6.71	6.67
12	3,5-$(OMe)_2$	H	6.77	6.71	6.71	6.70
13	3,4,5-$(OMe)_3$	H	7.00	7.07	7.07	7.09
14	2-OMe, 5-SO_2NEt_2	H	7.64	7.57	7.54	7.53
15	2-OMe, 5-SO_2Et	H	7.30	7.01	6.98	6.97
16	2-OMe, 5-SO_2Et	2-Cl	6.44	7.07	7.05	7.07
17[‡]	2-OMe, 5-SO_2Et	3-Cl	7.66	6.96	6.93	7.07
18	2-OMe, 5-SO_2Et	3-CN	6.85	7.16	7.13	7.15
19	2-OMe, 5-SO_2Et	3-OMe	7.44	7.37	7.35	7.34
20	2-OMe, 5-SO_2Et	3-F	7.03	7.01	6.99	6.99
21	2-OMe, 5-SO_2Et	4-OMe	7.06	7.35	7.33	7.35
22	2-OMe, 5-SO_2Et	4-Cl	7.09	7.12	7.10	7.10
23	2-OMe, 5-SO_2Et	4-F	7.10	7.05	7.03	7.03
24	2-OMe, 5-SO_2Et	4-CN	7.29	7.24	7.22	7.21
25	2-OMe, 5-SO_2Et	4-$CONH_2$	7.74	7.35	7.33	7.30
26	2-OMe, 5-SO_2Et	3-Et	7.54	7.28	7.25	7.23
27	2-OMe, 5-SO_2Et	3-COMe	7.04	7.32	7.30	7.31
28	2-OMe, 5-SO_2Et	3-(2-thiophenyl)	7.80	7.74	7.74	7.71
29	2-OMe, 5-SO_2Et	3-(3-thiophenyl)	7.82	7.70	7.70	7.64
30	2-OMe, 5-SO_2Et	3-(2-pyridyl)	7.66	7.24	7.20	7.15
31	2-OMe, 5-SO_2Et	3-(3-pyridyl)	7.12	7.27	7.24	7.25
32	2-OMe, 5-SO_2Et	3-(4-pyridyl)	7.11	7.31	7.27	7.29
33	2-OMe, 5-SO_2Et	3-{5-(1-methylimidazole)}	7.60	7.78	7.78	7.86
34	2-OMe, 5-SO_2Et	3-phenyl	7.10	7.27	7.23	7.25
35	2-OMe, 5-SO_2Et	3-(2-F-phenyl)	7.05	7.28	7.24	7.26
36	2-OMe, 5-SO_2Et	3-(2-Cl-phenyl)	7.41	7.25	7.21	7.19

[*] The inhibition concentration on molar scale, taken from Ref. [11]; [†] Uncertain activity; [‡] 'Outllier' compound of present study.

DRAGON software for computing the parameters corresponding to 0D-, 1D-, and 2D-descriptor classes. A total number of 491 descriptors corresponding to these classes were generated. The descriptor classes along with their definitions and scope in addressing the structural features are given in Table II. As the total number of descriptors involved in this study is very large, only the names of descriptor classes and the actual descriptor involved in the models have been addressed in the discussion. The QSAR model generation and validation have been done using the combinatorial protocol in multiple linear regression (CP-MLR) analysis [14].

CP-MLR

The CP-MLR is a 'filter'-based variable selection procedure for the development of statistical models in high dimensional QSAR studies [14–19]. It involves a combinatorial strategy with approximately placed 'filters' interfaced with MLR and extracts diverse models having unique combination of descriptors from the dataset. The filters set the thresholds for the descriptors in terms of inter-parameter correlation cutoff limits in subset regressions (filter-1), *t*-values of the regression coefficients (filter-2), internal explanatory power (filter-3; square root of adjusted multiple correlation coefficient of regression equation, *r*-bar), and the external consistency (filter-4; Q^2 i.e. cross-validated R^2 from the leave-one-out procedure). Throughout this study, the thresholds for the filters-1, 2 and 4 were assigned as 0.3, 2.0, and $0.3 \leq Q^2 \leq 1.0$, respectively. In the initial attempt, the base line models were generated by selecting a value of 0.71 for filter-3. In order to collect the descriptors with higher information content, the initial threshold of filter-3 was successively incremented with increasing the number of descriptors (per model) by considering the r-bar value of the preceding optimum model as the new threshold for the next generation.

Descriptor classification protocol

The three-stage descriptor classification protocol [18] is implemented with two-descriptor combinations (baseline models), as they are the simplest to understand and to explain the activity. In the first stage of the classification protocol, the correlations of the activity with two descriptor combinations from the individual descriptor classes (DCs) of the dataset were used to sort them into four categories. They are primary contributors (category I: a DC forms a model with its constituent descriptors), collective contributors (category II: a DC unable to form a model with its constituent descriptors, but forms model(s) in combination with a descriptor from another such DC), secondary contributors (category III: a DC forms the model(s) only in combination with category I) and noncontributors (category IV: a DC unable to form a model(s) in any manner like that of category I, II, and III). The sorted DCs were collated in the second stage to identify all the 3-descriptor models across the

Table II. Descriptor classes used for the analysis of VEGFR2 inhibitory activity of derivatives of 2-anilino-5-aryloxazole and identified categories in modeling the activity.

Descriptor class (acronyms)*	Definition and scope	Descriptors' category†
Constitutional (CONS)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations	I
Topological and (TOPO)	2D-descriptor from molecular graphs independent conformations	I
Molecular walk walks counts (MWC)	2D-descriptors representing self-returning counts of different lengths	I
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms	I
Galvez topological charge indices (GVZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix	I
2D-autocorrelations (2DAUTO)	Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the *lag*)	I
Functional groups (FUN)	Molecular descriptors based on the counting of the chemical functional groups	I
Atom centered fragments (ACF)	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen	I
Empirical (EMP)	1D-descriptors represent the counts of non-single bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule	I
Properties (PROP)	1D-descriptors representing molecular properties of a molecule	I

* Reference [12]; † Descriptor categories identified at the end of second stage; in this the filter values are: filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.71, and filter-4 as $0.3 \leq Q^2 \leq 1.0$, the number of compounds in each dataset was 35.

categories. In the last stage, the individual descriptors emerged in all three-descriptor models were pooled to discover the higher models for quantification of the activity.

All the identified models have been put to the randomization test [16,21] by repeated randomization of the activity to discover the chance correlations, if any, associated with them. For this every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted to express the percent chance correlation of the model under scrutiny. The model development procedure has been finally validated by creating divergent training and test sets from complete data set.

Results and discussion

Initially, the VEGFR2 inhibitory activity of 35 analogues of 2-anilino-5-aryloxazole was investigated with a variety of 0D-, 1D- and 2D-descriptors obtained from DRAGON software. Several of these descriptors recognized from ten different classes, have shown significant correlations and are identified as the primary contributors (category I) in modeling the inhibitory activity of these compounds. Table III lists various models (Equations 1–17) derived in such descriptors from each class along with their statistical parameters.

Among the one and two descriptor models, the inhibition activity of the compounds have been best explained by CONS, TOPO, BCUT and 2DAUTO descriptor classes. The CONS descriptors appeared in model 2, favors the flexibility in molecular structure (RBN, number of rotatable bonds) in addition to the preference of five membered rings (nR05, number of five membered rings) in a structure for enhanced activity. The TOPO class descriptors, TIC2 (total information content index of neighborhood symmetry of order-2) and CIC5 (complementary information content of neighborhood symmetry of order-5) in model 4, favor the enhancement of these information contents for improving the activity. The BCUT descriptor BELe6 (lowest eigenvalue 6 of Burden matrix/weighted by atomic Sanderson electronegativities) alone have explained 74 per cent of variance in the activity (model 7). The 2DAUTO descriptors in model 11, ATS7m (Broto-Moreau autocorrelation of *lag* 7/weighted by atomic masses) and GATS2p (Geary autocorrelation of *lag* 2/weighted by atomic polarizabilities) suggest the importance of *lags* 7 and 2 weighted by respective properties.

The significance of various emerged models, listed in Table III, may be ascertained through the statistical parameters, the correlation coefficient *r*, the standard error of the estimate *s*, and the Fisher's

Table III. The QSAR models emerged in primary descriptors* from ten different classes.

Des. class	Constant	Des-1	Des-2	*r*	Q^2	*s*	*F*	*AIC*	*FIT*	Eqn.
CONS	5.343	0.249 RBN		0.829	0.659	0.337	72.51	0.128	2.014	(1)
	4.734	0.237 RBN	0.633 nR05	0.881	0.741	0.289	55.87	0.099	2.865	(2)
TOPO	−39.869	19.482 LP1		0.849	0.691	0.318	85.24	0.114	2.368	(3)
	4.382	0.012 TIC2	0.826 CIC5	0.875	0.735	0.297	52.11	0.105	2.673	(4)
MWC	4.877	3.359 MWC08		0.809	0.616	0.354	62.55	0.141	1.738	(5)
	4.877	9.252 MWC09	0.051 SRW05	0.837	0.656	0.334	37.60	0.133	1.928	(6)
BCUT	2.760	4.107 BELe6		0.862	0.716	0.306	95.04	0.105	2.640	(7)
	51.406	−25.536 BELm2	1.333 BEHv3	0.794	0.536	0.372	27.33	0.164	1.402	(8)
GVZ	5.350	1.325 GGI4		0.848	0.691	0.319	84.54	0.114	2.348	(9)
2DAUTO	4.975	0.020 ATS5e		0.845	0.686	0.322	82.55	0.116	2.293	(10)
	7.960	0.032 ATS7m	−2.290 GATS2p	0.870	0.725	0.300	50.25	0.108	2.577	(11)
FUN	6.108	0.480 nCp		0.786	0.582	0.372	53.61	0.155	1.489	(12)
	4.887	0.871 nRSR	0.342 nHAcc	0.848	0.694	0.324	41.16	0.125	2.111	(13)
ACF	6.032	0.398 C-005	0.728 C-006	0.832	0.645	0.340	35.89	0.137	1.841	(14)
EMP	5.345	1.172 Ui	−5.464 ARR	0.830	0.640	0.342	35.33	0.139	1.812	(15)
PROP	5.560	0.019 PSA		0.774	0.557	0.382	49.36	0.163	1.371	(16)
	5.095	0.029 MR	−0.353 MLOGP	0.809	0.592	0.360	30.38	0.154	1.558	(17)

*RBN, number of rotatable bonds; nR05, number of five membered rings; LP1, eigenvalue based Lovasz-Pelikan index; TIC2, total information content index of neighborhood symmetry of order-2; CIC5, complementary information content of neighborhood symmetry of order-5; MWC08 and MWC09, molecular walk count of order-08 and -09 respectively; SRW05, self returning walk count of order-5; BELe6, lowest eigenvalue no. 6 of Burden matrix/weighted by atomic Sanderson electronegativities; BELm2, lowest eigenvalue no. 2 of Burden matrix/weighted by atomic masses; BEHv3, highest eigenvalue no. 3 of Burden matrix/weighted by atomic van der Waals volumes; GGI4, topological charge index of order-4; ATS5e, Broto-Moreau autocorrelation of *lag* 5/weighted by atomic Sanderson electronegativities; ATS7m, Broto-Moreau autocorrelation of *lag* 7/weighted by atomic masses; GATS2p, Geary autocorrelation of *lag* 2/weighted by atomic polarizabilities; nCp, number of total primary sp^3 carbon atoms; nRSR, number of sulphurs; nHAcc, number of hydrogen bond acceptor N, O, F atoms; C-005, $-CH_3X$; C-006, $-CH_2X-$; Ui, the unsaturation index; ARR, ratio of the number of aromatic bonds over the total number of non-H bonds; PSA, fragment based polar surface area; MR, Ghose-Crippon molar refractivity; MLOGP, Moriguchi octanol-water partition coefficient.

ratio *F*. Additionally, the cross-validated index Q^2, obtained from leave-one-out (LOO) procedure, may assist in identifying the robustness of these models. In a comparative study, where a large number of QSAR models are generated from the descriptors belonging to different categories, the other important statistics such as the Kubinyi function, *FIT* [22,23] and the Akaike's information criterion, *AIC* [24,25] are very important in explaining the best predictive model equation. Even in stepwise development of a QSAR equation, these statistical parameters may play crucial role in ascertaining the overall significance of final model. The *FIT* function is closely related to the *F*-statistic but proved to be a useful parameter for the assessment of the quality of the models. The disadvantage of the *F* value is its sensitivity to changes in the number of independent variables, *k* in the equation that describes the model. The *F* value is more sensitive if *k* is small, whereas it is less sensitive if *k* is large. The *FIT* function, on the other hand, is less sensitive to a lower number *k* but is more sensitive to a larger number *k*. The best model would yield the highest value for this function. The *AIC* takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The model that produces the lower *AIC* value should be considered potentially the most useful. The physical interpretation of resultant descriptors is given briefly in the footnotes under Table III.

In search of statistical significant models, the equations in three descriptors were further derived and identified relevant descriptors were categorized under two different pools. Firstly, 25 descriptors (Table III) have emerged from the category analysis. These primary contributors were then subjected to CP-MLR. The resulting best model, in three descriptors, is shown in Equation (18)

$$\begin{aligned}\mathrm{pIC}_{50} &= 0.451(0.168)\mathrm{nR05}\\ &\quad + 0.190(0.002)\mathrm{ATS5e}\\ &\quad - 0.095(0.025)\mathrm{nCaH} + 5.526\\ n &= 35,\ r = 0.907,\ Q^2 = 0.777,\\ s &= 0.277,\ F = 47.735,\\ FIT &= 3.255, AIC = 0.087 \end{aligned} \tag{18}$$

In above equation, nR05, a CONS class descriptor, is accounting for the number of 5-membered rings in the structure under consideration. The positive regression coefficient of this descriptor recommends 5-membered rings in a structure. Likewise, ATS5e, which is Broto-Moreau autocorrelation of topological structure with path length (*lag*) 5 in the graph weighted by atomic Sanderson electronegativities, belonging to 2DAUTO class, indicated that the higher path lengths rich in electronic content would be favorable for the improvement of inhibition activity. The FUN class descriptor, nCaH, is indicative of the number of unsubstituted aromatic C (sp2) in a molecule. The associated negative regression coefficient of this parameter, therefore, demands for higher number of substituted aromatic carbons.

To explore models superior to the model in Equation (18), a second pool of descriptors was formulated from all the 10 classes considered collectively. This pool now comprising of 481 descriptors, was able to generate 16 models through the CP-MLR analysis. The 18 descriptors participated in these models along with their average regression coefficients, and total incidences are listed in Table IV. In these 16 models, Equation (18) was obtained once again in addition to a slightly superior Equation (19)

$$\begin{aligned}\mathrm{pIC}_{50} &= 0.574(0.167)\mathrm{nR05}\\ &\quad + 3.727(0.412)\mathrm{BEHe7}\\ &\quad - 3.125(0.903)\mathrm{MATS4v} - 4.908\\ n &= 35,\ r = 0.909,\ Q^2 = 0.798,\\ s &= 0.259,\ F = 49.424,\\ FIT &= 3.370, AIC = 0.084 \end{aligned} \tag{19}$$

The statistical parameter *r* is now able to account for 83% of variance in the observed activity and Q^2 pointed to the robustness of this model. The newly appeared descriptors in above Equation, BEHe7 and MATS4v, belong to BCUT and 2DAUTO classes respectively. The former descriptor representing the highest eigenvalue n.7 of Burden matrix/weighted by atomic Sanderson electronegativities contributes positively to improve the activity while the later one being the Moran autocorrelation of topological structure with *lag* 4 in the graph weighted by atomic van der Waals volumes causes detrimental effect to the activity. Equation (19) was further subjected to randomization process, where 100 simulations were carried out but none of the identified models in these simulations has shown any chance correlation. The above equation was, therefore, used to calculate pIC_{50} values which were in close agreement with observed ones for all the compounds except the congener **17**. The residual [pIC_{50}(observed) − pIC_{50}(calculated)] obtained for congener **17** was more than two times the standard deviation. This data point was, therefore, eliminated from the data set and the derived new

Table IV. Descriptors identified for modeling the VEGFR2 inhibitory activity along with the average regression coefficient, standard deviation and the total incidence.

Descriptor*	Avg reg coeff(sd) total incidence†		Descriptor*	Avg reg coeff(sd) total incidence†
nR05	0.558(0.082)6	CONS	nBnz	− 0.407(0.066)2
X4v	0.534(0.000)1	TOPO	TIC1	0.018(0.000)1
SRW07	0.006(0.001)5	MWC	SRW09	0.001(0.000)2
BEHe7	3.875(0.158)7	BCUT	–	–
GGI6	2.004(0.169)3	GVZ	–	–
ATS7m	0.035(0.000)1	2DAUTO	ATS5e	0.019(0.000)1
ATS8e	0.020(0.000)2		MATS4v	− 3.279(0.365)4
MATS4p	− 3.247(0.824)2		GATS4v	2.974(0.501)4
GATS2e	0.893(0.022)2		GATS2p	− 1.822(0.000)1
nCaH	− 0.103(0.007)3	FUN	–	–
C-024	− 0.100(0.000)1	ACF	–	–

* The descriptors are identified from the three parameter models emerged from CP-MLR protocol with filter-1 as 0.3; filter-2 as 2.0; filter-3 as 0.89; filter-4 as $0.3 \leq Q^2 \leq 1.0$; number of compounds in the study are 35; CONST: nR05 is number of 5-membered rings; TOPO: X4v is valence connectivity index chi-4; TIC1 was total information content index (neighborhood symmetry of 1-order); MWC: SRW07 and SRW09 are self-returning walk count of order 07 and 09 respectively; BCUT: BEHe7 highest eigenvalue n.7 of Burden matrix/weighted by atomic Sanderson electronegativities; GALV: GGI6 topological charge index of order 6; 2DAUTO: ATS7m is Broto-Moreau autocorrelation of a topological structure—*lag* 7/weighted by atomic masses; ATS5e and ATS8e are Broto-Moreau autocorrelation of a topological structure—*lag* 5 and 8 respectively/weighted by atomic Sanderson electronegativities, MATS4v and MATS4p are Moran autocorrelation—*lag* 4/weighted by atomic van der Waals volumes and polarizabilities respectively, GATS4v Geary autocorrelation—*lag* 4/weighted by atomic van der Waals volumes, GATS2e and GATS2p are Geary autocorrelation—*lag* 2/weighted by Sandersons electronegativities and polarizabilities, respectively; FUN: nCaH number of unsubstituted aromatic C(sp2); ACF: C-024 R-CH-R; also see ref. [23]; † The average regression coefficient of the descriptor corresponding to all models and the total number of its incidences; the arithmetic sign represents the sign of the regression coefficient in the models.

model is as in Equation (20)

$$
\begin{aligned}
pIC_{50} &= 0.610(0.148)\mathrm{nR05} \\
&\quad + 3.654(0.365)\mathrm{BEHe7} \\
&\quad - 3.284(0.799)\mathrm{MATS4v} - 4.776 \\
n &= 34,\ \ r = 0.929,\ \ Q^2 = 0.837, \\
s &= 0.228,\ \ F = 63.480, \\
FIT &= 4.429, AIC = 0.066 \qquad (20)
\end{aligned}
$$

All the statistical parameters of Equation (20) have improved over to that of Equation (19). This, in turn, reflected upon the superiority of newly derived model. The outlier behavior of compound **17** is not immediately apparent. The calculated pIC_{50} values, using Equation (20), and predicted from LOO analysis are in close agreement with the observed ones (Table I). The plot of observed versus calculated and predicted pIC_{50} values is given in Figure 2 to demonstrate the goodness of fit and to show systematic variations between them in the present congeneric series. From Equation (20), it has appeared that the more number of 5-membered rings and the highest eigenvalue n.7 of Burden matrix/weighted by atomic Sanderson electronegativities contributes positively to improve the activity while the Moran autocorrelation of topological structure with *lag* 4 in the graph weighted by atomic van der Waals volumes results in detrimental effect to it.

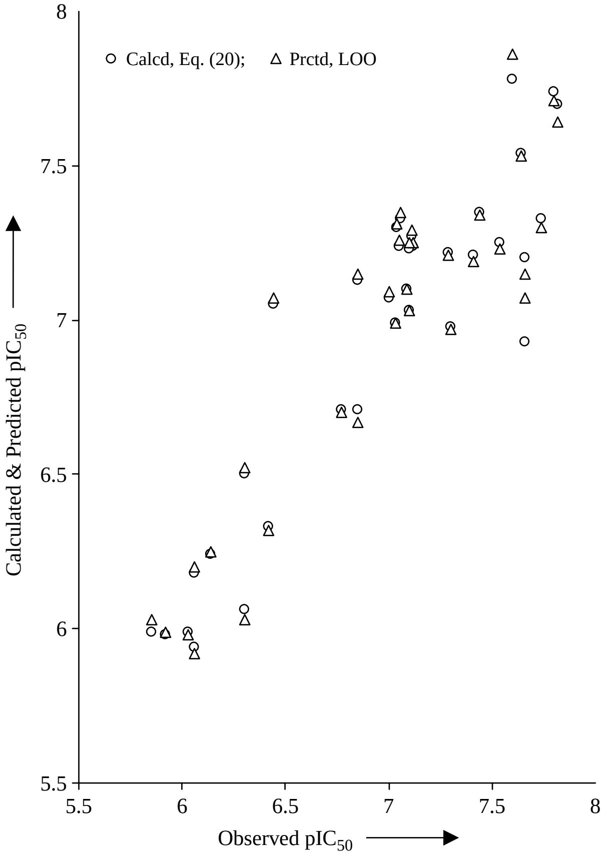

Figure 2. Plots of observed versus calculated and predicted pIC_{50} values.

Table V. Predicted residual activity of different test sets (9 compounds each) of the compounds listed in Table I.

	Residual*		
Compd. No.	Des Clus†	Act Clus‡	Random Clus¶
1	−0.01		−0.11
3		0.13	
5	−0.15	−0.19	−0.16
8		0.10	
9	−0.23		−0.25
10	6.03	6.07	6.05
13			−0.10
14		0.05	
15	0.34	0.30	
17	6.90	6.96	6.93
18		−0.32	−0.25
19	0.14		
22	0.01		−0.01
26			0.31
28	0.09	0.09	
30	0.52		0.50
31	−0.06		
34		−0.19	−0.10
35		−0.25	
Training set§			
r	0.918	0.917	0.929
s	0.231	0.244	0.226
F	37.828	37.112	44.283
Q^2	0.795	0.788	0.828
Test set‖			
r^2	0.883	0.899	0.856

*The difference of observed and predicted pIC_{50} values; the training model have 25 compounds each; †Test set from the cluster analysis of full descriptor data set of the compounds; ‡Test set from the cluster analysis of the activity of the compounds; ¶Test set from the random selection of the compounds; §The statistics for the training sets were derived with the descriptors of Equation (20); ‖In computing the r^2 values, the descriptors were considered as in Equation (20) for nine compounds in each of three test sets.

Equation (20) was further validated through three test sets, each containing 9 compounds out of the 34 active ones listed in Table I. Of the three test sets, two were generated in the SYSTAT [26] using the single linkage hierarchical cluster procedure involving the Euclidean distances with regard to the descriptors and to the activity values. In either case, the selection of the test set from the cluster tree was done in such a way as to keep the test compounds at a maximum possible distance from each other. The third test set of the compounds corresponds to the random selection procedure. The three test sets, selected in this manner, represent different cross-sections of all the compounds in present series. The remaining 25 data points in each of the three training sets were then used to derive new models. These models were next used to predict activities of compounds in test sets and of compound **10** having uncertain activity value and compound **17**, the outlier congener. The residuals of the predictions and the corresponding predictive r^2, s, Q^2 and F-values have been given in Table V. The predictions for compounds corresponding to three test sets are within the reasonable limits of their actual values.

In conclusion, the present study has provided structure-activity relationship of the VEGFR2 kinase inhibitory activity of 2-anilino-5-aryloxazole analogues in terms of structural requirements. The inhibitory activity has, therefore become the function of the cumulative effect of different structural features which were identified in terms of individual descriptors. In order to improve the inhibitory activity of a compound, the descriptors, nR05 and BEHe7 have advocated, respectively, the presence of 5-membered rings in the structural frame work and the electronic content associated to eigenvalue n.7 of the Burden matrix while the descriptor MATS4v emphasized the requirement of the path length 4, weighted by atomic van der Waals volumes. The derived models and participating descriptors in them have suggested that the substituent of 2-anilino-5-aryloxazole moiety have sufficient scope for further modification. Thus, our study may provide a ground for modeling aspects of 2-anilino-5-aryloxazoles as the inhibitors of vascular endothelial growth factor receptor 2 (VEGFR2) kinase.

Acknowledgements

We express our sincere gratitude to our Institutions for providing the necessary facilities to complete this work.

References

[1] Folkman J. Fundamental concepts of the angiogenic process. Curr Mol Med 2003;3:643–651.

[2] Veikkola T, Karkkainen M, Claesson-Welsh L, Alitalo K. Regulation of angiogenesis via vascular endothelial growth factor receptors. Cancer Res 2000;60:203–212.

[3] Glade-Bender J, Kandel JJ, Yamashiro DJ. VEGF blocking therapy in the treatment of cancer. Expert Opin Biol Ther 2003;3:263–276.

[4] Bold G, Altman K-H, Frei J, Lang M, Manley PW, Traxler P, Weitfeld B, Bruggen J, Buschdunger E, Cozens R, Ferrari S, Furet P, Hoffman F, Martiny-Baron G, Mestan J, Rosel J, Sills M, Stover D, Acemoglu F, Boss E, Emmenegger R, Lasser L, Masso E, Roth R, Schlachter C, Vetterli W, Wyxx D, Wood JM. New anilinophthalazines as potent and orally active well absorbed inhibitors of the VEGF receptor tyrosine kinase useful as antagonists of tumor-driven angiogenesis. J Med Chem 2000;43:2310–2323.

[5] Hennequin LF, Stokes ES, Thomas AP, Johnstone C, Ple PA, Ogilvie DJ, Dukes M, Wedge SR, Kendrew J, Curwen JO. Novel 4-anilinoquinazolines with C-7 basic side chains: Design and structure-activity relationship of a series of potent orally active, VEGF receptor tyrosine kinase inhibitors. J Med Chem 2002;45:1300–1312.

[6] Mendel DB, Laird AD, Xin X, Louie SG, Christensen JG, Li G, Schreck RE, Abrams TJ, Ngai TJ, Lee LB, Murray LJ, Carver J, Chan E, Moss KG, Haznedar JO, Sukbuntherng J, Blake RA, Sun L, Tang C, Miller T, Shirazian S, McMahon G, Cherrington JM. *In vivo* antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet derived growth factor receptors: Determination of a pharmacokinetic/pharmacodynamic relationship. Clin Cancer Res 2003;9:327–337.

[7] Beebe JS, Jani JP, Knauth E, Goodwin P, Higdon CE, Rossi AM, Emerson E, Finkelstein M, Floyd E, Harriman S, Atherton J, Hillerman S, Soderstrom C, Kou K, Grant T, Noe MC, Foster B, Rastinejad F, Marx MA, Schaeffer T, Whalen PM, Roberts WG. Pharmacological characterization of CP-547,632, a novel vascular endothelial growth factor receptor-2 tyrosine kinase inhibitor for cancer therapy. Cancer Res 2003;63:7301–7309.

[8] Harris PA, Cheung M, Hunter III RN, Brown ML, Veal JM, Nolte RT, Wang L, Liu W, Crosby RM, Johnson JH, Epperly AH, Kumar R, Luttrell DK, Stafford JA. Discovery and evaluation of 2-anilino-5-aryloxazoles as a novel class of VEGFR2 kinase inhibitors. J Med Chem 2005;48: 1610–1619.

[9] Basak SC, Harriss DK, Magnuson VR. POLLY. Duluth, MN: University of Minnesota; 1988.

[10] Molconn Z, ver. 2.07, eduSoft Lc, a Virginia Corporation, Ashland, VA 23005 USA. www.edusoft-lc.com.

[11] (a) Katritzky AR, Lobnov V, Karelson M. CODESSA (Comprehensive descriptors for structural and statistical analysis). Gainesville, FL: University of Florida; 1994. (b) Katritzky AR, Perumal S, Petrukhin R, Kleinpeter E. CODESSA-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities. J Chem Inf Comput Sci 2001;41:569–574.

[12] DRAGON software version 3.0-2003. By Todeschini R, Consonni V, Mauri A, Pavan M. Milano, Italy. http//disat.unimib.it/chm/Dragon.htm.

[13] Gonzalez MP, Helguera AM. TOPS-MODE verces Dragon descriptors to predict permeability coefficients through low-density polymer. J Comput-Aided Mol Des 2003;17: 665–672.

[14] Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: Analysis of Selwood et al. data set–A case study. QSAR Comb Sci 2003;22: 583–595.

[15] Gupta MK, Prabhakar YS. Topological descriptors in modeling the antimalarial activity of 4-(3′,5′-disubstituted aniline)quinolines. J Chem Inf Model 2006;46:93–102.

[16] Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR Comb Sci 2004;23:234–244.

[17] Prabhakar YS. A combinatorial protocol in multiple linear regression to model gas chromatographic response factor of organophosphonate esters. Internet Electron J Mol Des 2004; 3: 150–162, http://www.biochempress.com.

[18] Gupta MK, Sagar R, Shaw AK, Prabhakar YS. CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols–Topological descriptors in modeling the activity. Bioorg Med Chem 2005;13:343–351.

[19] Saquib M, Gupta MK, Sagar R, Prabhakar YS, Shaw AK, Kumar R, Maulik PR, Gaikwad AN, Sinha S, Srivastava AK, Chaturvedi V, Srivastava R, Srivastava BS.C-3 Alkyl/arylalkyl-2,3-dideoxy hex-2-enopyranosides as antitubercular agents: Synthesis, biological evaluation and QSAR study. J Med Chem 2007;50:2942–2950.

[20] Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA.

[21] So SS, Karplus M. Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks. 1. Methods and validations. J Med Chem 1997;40:4347–4359.

[22] Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quant Struct–Act Relat 1994;13: 285–294.

[23] Kubinyi H. Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. Quant Struct–Act Relat 1994;13:393–401.

[24] Akaike H. Information theory and an extension of the minimum likelihood principle. In: Petrov BN, Csaki F, editors. Second international symposium on information theory. Akademiai Kiado: Budapest; 1973. p 267–281.

[25] Akaike H. A new look at the statistical identification model. IEEE Trans Autom Control 1974;AC-19:716–723.

[26] SYSTAT, Version 7.0; SPSS Inc., 444 North Michigan Avenue, Chicago, IL, 60611.