QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP STUDY OF SUBSTITUTED-[1,2,4] OXADIAZOLES AS S1P₁ AGONISTS

P. SINGH[*]

Department of Chemistry, S. K. Government College, SIKAR – 332001 (Raj.) INDIA

(*Received : 30.01.2013; Revised : 07.02.2013; Accepted : 08.02.2013*)

ABSTRACT

A quantitative structure-activity relationship (QSAR) study has been performed on agonist actions of [1,2,4] oxadiazoles for S1P₁ receptor. The mechanism of action of 3-substuted-[1,2,4] oxadiazoles appeared to be different from that of 5-(substituted phenyl)-[1,2,4] oxadiazoles, therefore, these compounds, in that order, have been divided in two categories, CTGA and CTGB to develop statistical models that may explain the agonist action of the series. From highest significant models for compounds in CTGA, it appeared that the higher values of the descriptors, MATS1v, MATS4m and MATS5p are conducive in improving the agonist action while the functionality such as Al-O-Ar or Ar-O-Ar or R-O-R/R-O-C=X, imparting detrimental effect to it, is undesirable. Thus, the Moran autocorrelation (MATS) of lags 1, 4 and 5 of a molecule weighted, respectively, by atomic van der Waals volumes, atomic masses and atomic polarizabilities remain the leading reasons during interaction with receptor. For compounds in CTGB, the higher values of the descriptors BELe2 (the lowest eigenvalue no. 2 of Burden matrix/weighted by atomic Sanderson electronegativities) and MATS7p or the descriptors MATS7p and C-025 (accounting for the fragment R--CR--R) are beneficial in improving the agonist action of a compound. Therefore, the electronic and polarization effects or polarization effect in addition to the structural fragment, R--CR--R appeared to be governing features during interaction. PLS analysis has further confirmed the dominance of the CP-MLR identified descriptors. The guidelines, based on the statistically validated models, may facilitate in exploring more potential analogues of the series. Applicability domain analysis revealed that the suggested models have acceptable predictability. Except one obvious outlier compound from CTGB, all other compounds were within the applicability domain of the proposed models and were evaluated correctly.

Key words: [1,2,4]-Oxadiazole derivatives, S1P₁ agonists, Combinatorial protocol in multiple linear regression (CP-MLR) analysis, Molecular descriptors, QSAR study.

INTRODUCTION

The lysophospholipid sphingosine-1-phosphate (S1P) binds five specific G-protein coupled receptors $(S1P_1-S1P_5)$[1-4] and exerts a variety of biological activities such as vascular maturation, cell survival, proliferation, differentitation, migration and chemotaxis[5-7]. The S1P₁ modulates egress of T-lymphocytes from thymus and peripheral lymphoid organs[8], and it was demonstrated[9,10] that targeting S1P₁ is sufficient to cause lymphocyte sequestration to thymus and lymphoid organs, without affecting the innate immune system and even cellular reactivity of lymphocytes to antigen challenge. Selective S1P₁ receptor agonists are promisingly developed as a novel immunomodulator. S1P₁ receptor has been considered a potential target for a variety of immune-mediated diseases, including rheumatoid arthritis, psoriasis and multiple sclerosis disease. Recently, a novel class of S1P₁ receptor agonists based on the 1,2,4-oxadiazole

---

scaffold with variations at 3- and 5-positions has been reported[11]. The category A (CTGA) comprise of compounds in which position-5 is fixed to 3,5-bis(trifluoromethyl) phenyl moiety and position-3 is varied. In category B (CTGB), position-3 is fixed to 1-(4-hydroxy-4-oxo-butanoyl) indolin-5-yl moiety and a phenyl ring, attached to position-5, is varied with different *meta-* and *para-*substituents. Due to differing mode of actions of two categories, the compounds [Fig. 1 (CTGA and CTGB)] of these structures have been documented separately in Table 1 and Table 2. The reported study delineating structure-activity relationships (SARs) was targeted at the alterations of substituents at different positions and provided no rationale to reduce the trial-and-error factors. Hence, in the present communication a 2D-quantitative SAR (2D-QSAR) for these analogues has been conducted to provide the rationale for drug-design and to explore the possible mechanism of their action. In the congeneric series, where a relative study is being carried out, the 2D-descriptors may play important role in deriving the significant correlations with activity profiles of the compounds. The novelty and importance of a 2D-QSAR study is due to its simplicity for the calculations of different descriptors and their interpretation (in physical sense) to explain the biological activity of compounds at molecular level.

EXPERIMENTAL

Material and methods

Data set

In the present work, the compounds [Table 1-2; Fig. 1 (CTGA and CTGB)] along with their *in vitro* agonist activity values have been taken from the literature[11]. The human $S1P_1$ EC_{50} values, determined in the HTRF cAMP assay, represents the concentration of the drug needed to suppress lymphocyte levels by 50% and the same, expressed as $-\log EC_{50}$ on a molar basis, are listed in Table 1 and 2. For modeling purpose, the data-set was divided into training- and test-sets to insure external validation of derived models. The selection of test-set compounds was made through SYSTAT[12] using the single linkage hierarchical cluster procedure involving the Euclidean distances of the agonist action, $-\log EC_{50}$ values. Nearly 25% of the total compounds were selected for this purpose from the generated cluster tree in such a way to keep them at a maximum possible distance from each other. In SYSTAT, by default, the normalized Euclidean distances are computed to join the objects of cluster. The normalized distances are root mean-squared distances. The single linkage uses distance between two closest members in clustering. It generates long clusters and provides scope to choose objects at intervals. Due to this reason, a single linkage clustering procedure was applied.

Table 1: Observed and modeled $S1P_1$ agonist actions of 3-substituted-5-(3,5-bis(trifluoromethyl) phenyl)-1,2,4-oxadiazoles [Fig. 1 (CTGA) for structures]

Compd.	X	$-\log EC_{50}$ (M)			
		Obsd.[a]	Calcd.		
			Eq. (1)	Eq. (2)	PLS
1		8.14	8.24	7.95	8.12
2		8.01	7.59	7.69	7.91
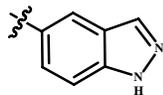
Compd.	X	–logEC$_{50}$ (M)			
		Obsd.[a]	Calcd.		
			Eq. (1)	Eq. (2)	PLS
3		8.31	7.95	8.28	8.24
4		8.38	8.35	8.89	8.56
5		4.00	4.24	4.04	3.91
6[b]		6.53	5.95	6.20	5.85
7[b]		6.39	6.67	7.13	6.98
8		6.00	6.16	6.64	6.28
9		7.62	7.37	7.09	7.33
10		7.12	8.08	7.84	7.97
11		8.37	7.49	7.99	7.73
12[b]		7.43	7.98	7.76	7.86
13		7.27	7.69	7.23	7.55
14		6.43	6.61	6.33	6.52
15		7.39	7.51	7.61	7.67
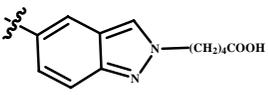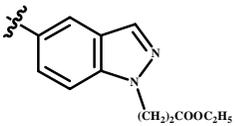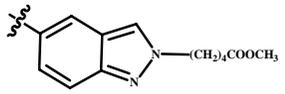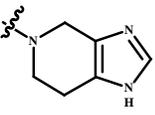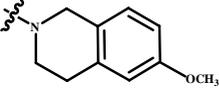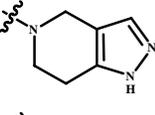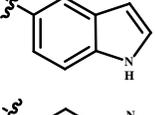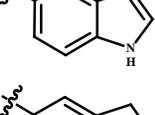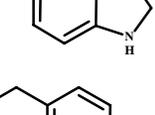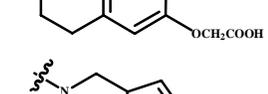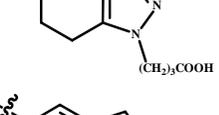
Compd.	X	–logEC$_{50}$ (M)			
		Obsd.[a]	Calcd.		
			Eq. (1)	Eq. (2)	PLS
16[b]		8.92	8.87	8.36	8.85
17		8.34	7.40	7.57	7.52
18		8.21	8.01	8.32	8.04
19[b]		8.18	7.62	7.90	7.70
20		6.99	7.51	7.60	7.42
21		8.79	9.16	8.30	8.59
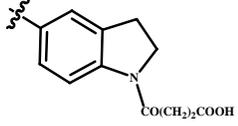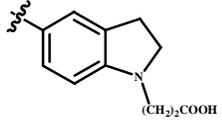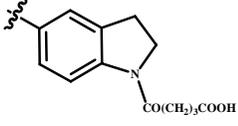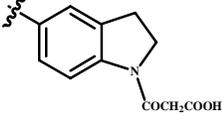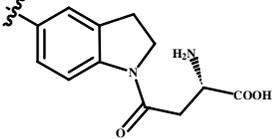
[a]Agonist activity; taken from ref. (11). [b]Test-set compound

Table 2: Observed and modeled S1P$_1$ agonist activity of 4-[5-(5-substituted-phenyl-[1,2,4] oxadiazol-3-yl)-2,3-dihydroindol-1-yl]-4-oxobutyric acid derivatives [Fig. 1 (CTGB) for structures]

Compd.	R$_1$	R$_2$	–logEC$_{50}$ (M)			
			Obsd.[a]	Calcd.		
				Eq.(3)	Eq.(4)	PLS
22[b]	Cl-		9.09	8.73	8.59	8.74
23	CF$_3$-		8.89	9.23	9.19	9.24
24	CN-		7.74	7.92	7.91	7.69
25	Cl-		7.35	7.65	7.61	7.54
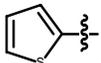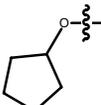
Compd.	R_1	R_2	$-\log EC_{50}$ (M)			
			Obsd.[a]	Calcd.		
				Eq.(3)	Eq.(4)	PLS
26[b]	CH_3O-	CHF_2O-	8.82	8.67	8.63	8.72
27	$CN-$	$(CH_3)_2CHO-$	8.96	8.83	8.84	8.64
28	CF_3O-	CH_3O-	9.00	8.80	8.79	8.71
29	$Br-$	$(CH_3)_2CHO-$	8.40	8.48	8.49	8.70
30	CF_3-	C_2H_5-	8.57	8.36	8.50	8.43
31[b]	CF_3O-	$Cl-$	8.28	8.55	8.63	8.56
32	$CN-$	CH_3O-	8.20	7.91	7.92	7.98
33	CF_3-	CH_3O-	7.96	7.67	7.72	7.91
34	CH_3-	CH_3O-	7.74	7.80	7.76	7.83
35[b]	$Cl-$	CH_3O-	7.96	7.55	7.58	7.88
36[c]	CH_3SO_2-	C_2H_5-	7.42	9.83	9.34	9.66
37	H	CF_3O-	7.17	7.32	7.25	7.31

[a]Agonist activity; taken from ref. (11), [b]Test-set compound. [c]Outlier compound

(CTGA)   **(CTGB)**
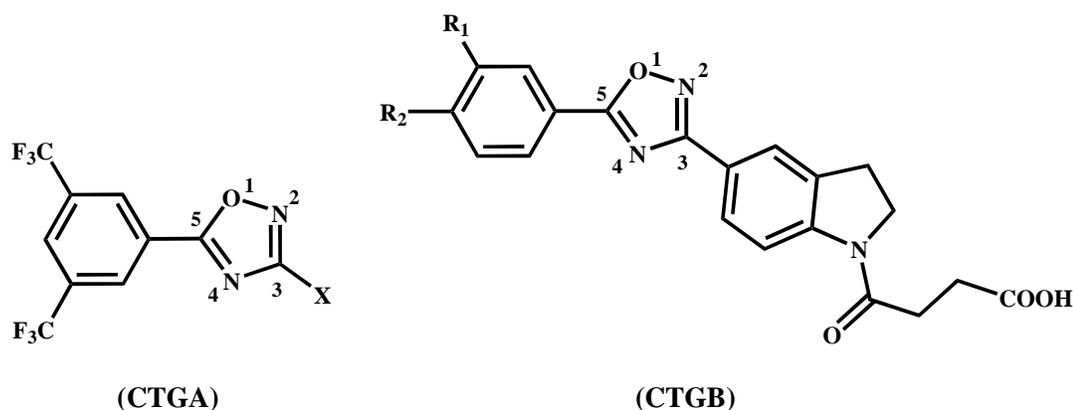
Fig. 1: [1,2,4]-Oxadiazoles having substituent variations at position-3 (CTGA) and position-5 (CTGB)

Molecular descriptors

The structures of the compounds (Table 1 and 2) under study have been drawn in 2D ChemDraw[13] using the standard procedure. These structures were converted into 3D objects using the default conversion procedure implemented in the CS Chem3D Ultra. The generated 3D structures of the compounds were subjected to energy minimization in the MOPAC module, using the AM1 procedure for closed shell systems, implemented in the CS Chem3D Ultra. This will ensure a well defined conformer relationship across the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software[14] for computing the descriptors corresponding to 0D-, 1D-, and 2D-classes. Table 3 provides the definition and scope of these descriptor classes in addressing the structural features which were employed in present QSAR work. The combinatorial protocol in multiple linear regression (CP-MLR) computational procedure[15] has been used for present work in developing QSAR models.

Table 3: Descriptor classes used for the analysis of agonist activity of substituted-[1,2,4] oxadiazoles

Descriptor class (acronyms)	Definition and scope
Constitutional (CONST)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations.
Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations.
Molecular walk counts (MWC)	2D-descriptors representing self-returning walk counts of different lengths.
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms.
Galvez topological charge indices (GLVZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix.
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).
Functional groups (FUNC)	Molecular descriptors based on the counting of the chemical functional groups.
Atom-centred fragments (ACF)	Molecular descriptors based on the counting of 120 atom-centred fragments, as defined by Ghose-Crippen.
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.
Properties (PROP)	1D-descriptors representing molecular properties of a molecule.

Model development

The CP-MLR is a 'filter'-based variable selection procedure for model development in QSAR studies[15]. The procedural aspects and implementation of this procedure are discussed in our recent publications[16-21]. The developed computer program, based on CP-MLR procedure, is interfaced with four filters which make the variable selection process efficient and lead to a unique solution. Filter-1 seeds the variables by way of limiting inter-parameter correlations to predefined level (upper limit ≤ 0.79); filter-2 controls the variables entry to a regression equation through t-values of coefficients (threshold value ≥ 2.0); filter-3 provides comparability of equations with different number of variables in terms of square root of adjusted multiple correlation coefficient of regression equation, r-bar; filter-4 estimates the consistency of the equation in terms of cross-validated Q^2 with leave-one-out (LOO) cross-validation as default option (threshold value $0.3 \leq Q^2 \leq 1.0$). In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation. Furthermore, in order to discover any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been put to a randomization test[22,23] by repeated randomization of the activity to ascertain 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.

Applicability domain

The utility of a QSAR model is based on its accurate prediction ability for new compounds. A model is valid only within its training domain and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain is evaluated by the leverage values for each compound[24,25]. The Williams plot (the plot of standardized residuals versus leverage values, h) can then be used for an immediate and simple graphical detection of both the response outliers (Y-outliers) and structurally influential chemicals (X-outliers) in the model. In this plot, the applicability domain is established inside a squared area within $\pm \beta \times$ s.d. and a leverage threshold h^*. The threshold h^* is generally fixed at $3(k + 1)/n$ (n is the number of training-set compounds and k is the number of model parameters) whereas $\beta = 2$ to 3. Prediction must be considered unreliable for compounds with a high leverage value (h > h^*). On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training-set compounds.

RESULTS AND DISCUSSION

In the initial attempt, all compounds (Table 1 and 2) have been considered together to quantify their observed agonist activities as a function of molecular descriptors. However, no significant relationship was obtained, suggesting that all compounds collectively may not follow similar trend and require to review their functionality variations at 3- and 5-positions of 1,2,4-oxadiazole scaffold. When position-3 was varied, position-5 was fixed to 3,5-bis(trifluoromethyl)phenyl moiety (CTGA) and when position-5 was varied, position-3 was fixed to 1-(4-hydroxy-4-oxo-butanoyl)indolin-5-yl moiety (CTGB). In either CTGA or CTGB the occurrence of benzimidazole or indoline rings with or without an acid side chain, at 3-position of 1,2,4-oxadiazole scaffold, are invariably accommodated in the $S1P_1$ receptor pocket[11]. Therefore the mode of action of 3-substituents in these categories may not be differentiated. On the other hand, the influence of substituents in 5-phenyl ring of 1,2,4-oxadiazole scaffold may be discriminated easily in the two categories. In CTGA the 5-phenyl ring has fixed *meta*-substitutents (3-CF_3 and 5-CF_3) while in CTGB the same is varied with different *meta*- and *para*-substituents. Obviously, the behavior of *para*-substituents differs significantly from that of *meta*-substituents which may possibly be reflected during interaction with receptor site(s).

A total number of 452 and 448 descriptors, belonging to 0D-, 1D- and 2D- classes, were computed, respectively, for the compounds of CTGA and CTGB utilizing DRAGON software. In either category, the data-set was further divided into training-set for model development and test-set for external validation of developed models. The selection of compounds for test-set was made through SYSTAT using the single linkage hierarchical cluster procedure involving the Euclidean distances of the $-logEC_{50}$ values. Five compounds (S. Nos. **6, 7**, **12**, **16** and **19**; Table 1) from CTGA and four compounds (S. Nos. **22**, **26**, **31** and **35**; Table 2) from CTGB were selected from the generated cluster trees in such a way to keep them at a maximum possible distance from each other. The computed descriptors were subjected to CP-MLR to develop models which may explain their agonist actions. The internal consistency, for each generated models, was achieved through LOO and L5O procedures while external validation was ascertained through r^2_{Test}. Three tri-variant models for CTGA and twenty bi-variant models for CTGB remained statistically significant. However, two highest significant models, in increasing level of significance, for each of these categories have been given through Eqs. (1)-(4). Thus for compounds in CTGA:

$$-logEC_{50} = 3.679 + 2.241 \, (0.612) \, MATS \, 4 \, m + 3.258 \, (0.525) \, MATS1v + 2.471 \, (1.098) \, MATS \, 5 \, p$$

$$n = 16, \, r = 0.911, \, s = 0.556, \, F \, (3, 12) = 19.533, \, Q^2_{LOO} = 0.687, \, Q^2_{L5O} = 0.778, \, r^2_{Test} = 0.776 \qquad ...(1)$$

$$-\log EC_{50} = 5.306 + 2.911 \, (0.550) \, MATS \, 4 \, m + 3.427 \, (1.023) \, MATS \, 5 \, p - 1.996 \, (0.283) \, O\text{-}060$$

$$n = 16, \, r = 0.928, \, s = 0.503, \, F \, (3, 12) = 24.758, \, Q^2_{LOO} = 0.763, \, Q^2_{L5O} = 0.737, \, r^2_{Test} = 0.750 \quad ...(2)$$

and for compounds in CTGB:

$$-\log EC_{50} = 7.318 + 0.934 \, (0.247) \, BELe \, 2 + 1.456 \, (0.225) \, MATS \, 7 \, p$$

$$n = 11, \, r = 0.933, \, s = 0.259, \, F(2,8) = 26.748, \, Q^2_{LOO} = 0.676, \, Q^2_{L5O} = 0.755, \, r^2_{Test} = 0.699 \quad ...(3)$$

$$-\log EC_{50} = 7.253 + 1.542 \, (0.193) \, MATS \, 7 \, p + 0.897 \, (0.188) \, C\text{-}025$$

$$n = 11, \, r = 0.952, \, s = 0.220, \, F(2, 8) = 38.493, \, Q^2_{LOO} = 0.795, \, Q^2_{L5O} = 0.789, \, r^2_{Test} = 0.571 \quad ...(4)$$

In above equations, n, r, s and F represent, respectively, the number of compounds, correlation coefficient, standard deviation and F-ratio between the variances of observed to calculated activity values. In fact three models, obtained for CTGA, have shared 6 descriptors and twenty models, obtained for CTGB have shared 14 descriptors. The brief description, average regression coefficients and the total incidence of these participated descriptors are given in Table 4. In all emerged models, the F-values remained significant at 99% level [$F_{3,12}(0.01) = 5.953$ and $F_{2,8}(0.01) = 8.649$] and the standard errors of regression coefficients (data within the parentheses) were significant at more than 95% level. For derivation of models from CTGB, one compound **36** (Table 2) was eliminated from the data-set due to it indifferent behavior. In fact, this was the lone compound which contains a *meta*-methanesulfonyl substituent in phenyl ring attached to 5-position of [1,2,4] oxadiazole scaffold. Possibly, the bigger size of this substiuent may not allow proper molecular binding with receptor site. Eqs. (1)-(4) are able to explain more than 85% of the variance in observed agonist actions of the compounds. The indices Q^2_{LOO} and Q^2_{L5O} (> 0.5) have accounted for internal robustness of the developed models while the index r^2_{Test} greater than 0.5 specified that the selected test-sets are accountable for external validation of these models. The signs of the regression coefficients have indicated the direction of influence of explanatory variables in a given model; the positive regression coefficient associated to a descriptor will augment the activity profile of a compound while the negative coefficient will cause detrimental effect to it.

Table 4: Identified descriptors[a] along with their physical meaning, average regression coefficient and incidence[b], in modeling the S1P$_1$ agonist activity of 1,2,4-oxadiazoles

S. No.	Descriptor	Class	Physical meaning	Average regression coefficient (incidence)	
				CTGA	**CTGB**
1	PW3	TOPO	Path/walk 3-Randic shape index.	-1.045(1)	
2	PW4	TOPO	Path/walk 4-Randic shape index.		46.620 (1)
3	VEA1	TOPO	Eigenvector coefficient sum from adjacency matrix.		0.921 (1)
4	VRA1	TOPO	Randic-type eigenvector-based index from adjacency matrix.		-0.003 (2)

Cont…

S. No.	Descriptor	Class	Physical meaning	Average regression coefficient (incidence)	
				CTGA	CTGB
5	VRA2	TOPO	Average Randic-type eigenvector-based index from adjacency matrix.		-0.128 (2)
6	BELm1	BCUT	Lowest eigenvalue no. 1 of Burden matrix/ weighted by atomic masses.		25.892 (1)
7	BELm2	BCUT	Lowest eigenvalue no. 2 of Burden matrix/ weighted by atomic masses.		18.646 (4)
8	BELe2	BCUT	Lowest eigenvalue no. 2 of Burden matrix/ weighted by atomic Sanderson electronegativities.		21.367 (3)
9	BELp4	BCUT	Lowest eigenvalue no. 4 of Burden matrix/ weighted by atomic polarizabilities.		-3.719 (1)
10	MATS4m	2DAUTO	Moran autocorrelation – lag 4/ weighted by atomic masses.	2.576 (2)	
11	MATS1v	2DAUTO	Moran autocorrelation – lag 1/ weighted by atomic van der Waals volumes.	3.509 (2)	
12	MATS7v	2DAUTO	Moran autocorrelation – lag 7/ weighted by atomic van der Waals volumes.	-2.563 (1)	7.556 (4)
13	MATS5p	2DAUTO	Moran autocorrelation – lag 5/ weighted by atomic polarizabilities.	2.949 (2)	
14	MATS7p	2DAUTO	Moran autocorrelation – lag 7/ weighted by atomic polarizabilities.		7.291 (6)
15	GATS7v	2DAUTO	Geary autocorrelation – lag 7/ weighted by atomic van der Waals volumes.		-7.082 (7)
16	GATS7p	2DAUTO	Geary autocorrelation – lag 7/ weighted by atomic polarizabilities.		-5.427 (3)
17	C-025	ACF	Corresponds to: R--CR--R.		0.399 (4)
18	C-026	ACF	Corresponds to: R--CX--R.		-0.327 (1)
19	O-060	ACF	Corresponds to: Al-O-Ar/ Ar-O-Ar/ R..O..R / R-O-C=X.	-1.996 (1)	

[a]The descriptors have been identified from the models, emerged from CP-MLR protocol with a training-set of 16 and 11 compounds for $S1P_1$ agonist activity the compounds in CTGA and CTGB respectively.
[b]The average regression coefficient of the descriptor corresponding to all models and the total number of its incidence. The arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models

In above equations, the descriptors MATS4m, MATS1v, MATS5p and MATS7p represent the Moran autocorrelations in which weighting components m, p and v represent, respectively, atomic masses, atomic polarizabilities and atomic van der Waals volumes and penultimate number represents the lag (path). These spatial autocorrelation descriptors are calculated on a H-depleted molecular graph. The descriptor, BELe2 represents the lowest eigenvalue no. 2 of Burden matrix/ weighted by atomic Sanderson electronegativities. It is calculated for a H-included molecular graph from the Burden matrix. Lastly, the descriptors, C-025 and O-060 correspond to the functionalities R--CR--R and Al-O-Ar/Ar-O-Ar/R-O-R/R-O-C=X respectively.

From Eqs. (1) and (2), it appeared that the higher values of the descriptors, MATS1v, MATS4m and MATS5p are conducive in improving the agonist action of a compound from CTGA while the functionality such as Al-O-Ar or Ar-O-Ar or R-O-R/R-O-C=X, imparting detrimental effect to it, is undesirable. Thus, the lags 1, 4 and 5 of a molecule weighted, respectively, by atomic van der Waals volumes, atomic masses and atomic polarizabilities are the dominating factors during interaction with receptor. Both these equations have been used to calculate the $-logEC_{50}$ values of the compounds. These calculated values were found in close agreement with the observed ones (Table 1). A graphical display, showing the plot of observed versus calculated $-logEC_{50}$s, is included in Fig. 2. The systematic variations, between the two, have been observed for the model Eqs. (1) and (2).

The descriptor MATS7p, participated in Eqs. (3) and (4), has revealed positive influence on agonist action of compounds from CTGB. Similarly, the descriptors, BELe2 and C-025 have also shown incremental effect on the activity. The higher values of the descriptors BELe2 and MATS7p (Eq. 3) or the descriptors MATS7p and C-025 (Eq. 4) are helpful in augmenting the activity of a compound. Thus the electronic and polarization effects or polarization effect in addition to structural fragment; R--CR--R appeared as dominating factors during interaction with receptor sites. The calculated $-logEC_{50}$ values, using Eqs. (3) and (4), listed in Table 2 remained in parity with the observed ones. The plot, revealing the goodness of fit and the systematic behavior, between observed and calculated $-logEC_{50}$s from both these models are given in Fig. 2.

Further, the PLS analysis[26-28] have also been performed using 6 and 14 identified descriptors for compounds of CTGA and CTGB respectively and the results are summarized in Table 5. In the study, the descriptors were autoscaled (zero mean and unit standard deviation) to provide each one of them equal weightage. In the PLS cross-validation, two-components remained optimum for these descriptors and they have explained 88.4% and 87.0% of variances in the observed agonist activities of the compounds from CTGA and CTGB respectively. The PLS equations of optimum two-components and MLR-like PLS coefficients of identified descriptors are given in Table 5. The calculated activity values of training- and test-set compounds remained in close agreement to that of the observed ones and are listed in Table 1 and 2. For comparison, the plot between observed and calculated activities (through PLS analyses) for the training- and test-set compounds is given in Fig. 2. Fig. 3 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity (Table 5). Different orders, indicating the level of significance of 6 and 14 participated descriptors for compounds of CTGA and CTGB respectively, are included in Table 5. For a given descriptor, lower is the order higher would be its significance in addressing the biological activity. The descriptors having positive contribution will augment the activity and their higher values are desirable to further improve it. On the other hand, the descriptors having negative contribution will diminish the activity. The lower or more negative values of such descriptors may, therefore, enhance the activity of a compound.

Table 5: PLS and MLR-like PLS models from the descriptors of three and two parameter CP-MLR models for S1P$_1$ agonist actions of compounds from CTGA and CTGB

A: PLS equation		
PLS components	**PLS coefficient (s.e.)[a]**	
	CTGA	**CTGB**
Component-1	0.790 (0.082)	0.266 (0.037)
Component-2	-0.309 (0.102)	-0.056 (0.034)
Constant	7.461	8.180

B: MLR-like PLS equation						
S. No.	**CTGA**			**S. No.**	**CTGB**	
	Descriptor	**MLR-like coefficient (f.c.)[b]**	**Order**		**Descriptor**	**MLR-like coefficient (f.c.)[b]**	**Order**
1	PW3	-0.485 (-0.064)	6	1	PW4	0.172 (0.056)	8
2	MATS4m	1.643 (0.182)	3	2	VEA1	0.196 (0.064)	6
3	MATS1v	1.975 (0.254)	1	3	VRA1	-0.158 (-0.061)	7
4	MATS7v	-1.338 (-0.138)	5	4	VRA2	-0.195 (-0.080)	5
5	MATS5p	2.270 (0.141)	4	5	BELm1	0.085 (0.032)	11
6	O-060	-1.003 (-0.220)	2	6	BELm2	0.039 (0.014)	13
	Constant	6.013		7	BELe2	0.136 (0.050)	9
				8	BELp4	-0.036 (-0.015)	12
				9	MATS7v	0.311 (0.134)	4
				10	MATS7p	0.388 (0.157)	1
				11	GATS7v	-0.366 (-0.149)	2
				12	GATS7p	-0.334 (-0.136)	3
				13	C-025	0.110 (0.046)	10
				14	C-026	-0.017 (-0.007)	14
					Constant	8.208	

C: PLS regression statistics		
Symbol	**Value**	
	CTGA	**CTGB**
n	16	11
r	0.942	0.933
s	0.434	0.259
F	49.347	26.895
Q^2_{LOO}	0.838	0.643
Q^2_{L5O}	0.843	0.598
r^2_{Test}	0.736	0.833

[a]Regression coefficient of PLS factor and its standard error. [b]Coefficients of MLR-like PLS equation in terms of descriptors for their original values; f.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data.
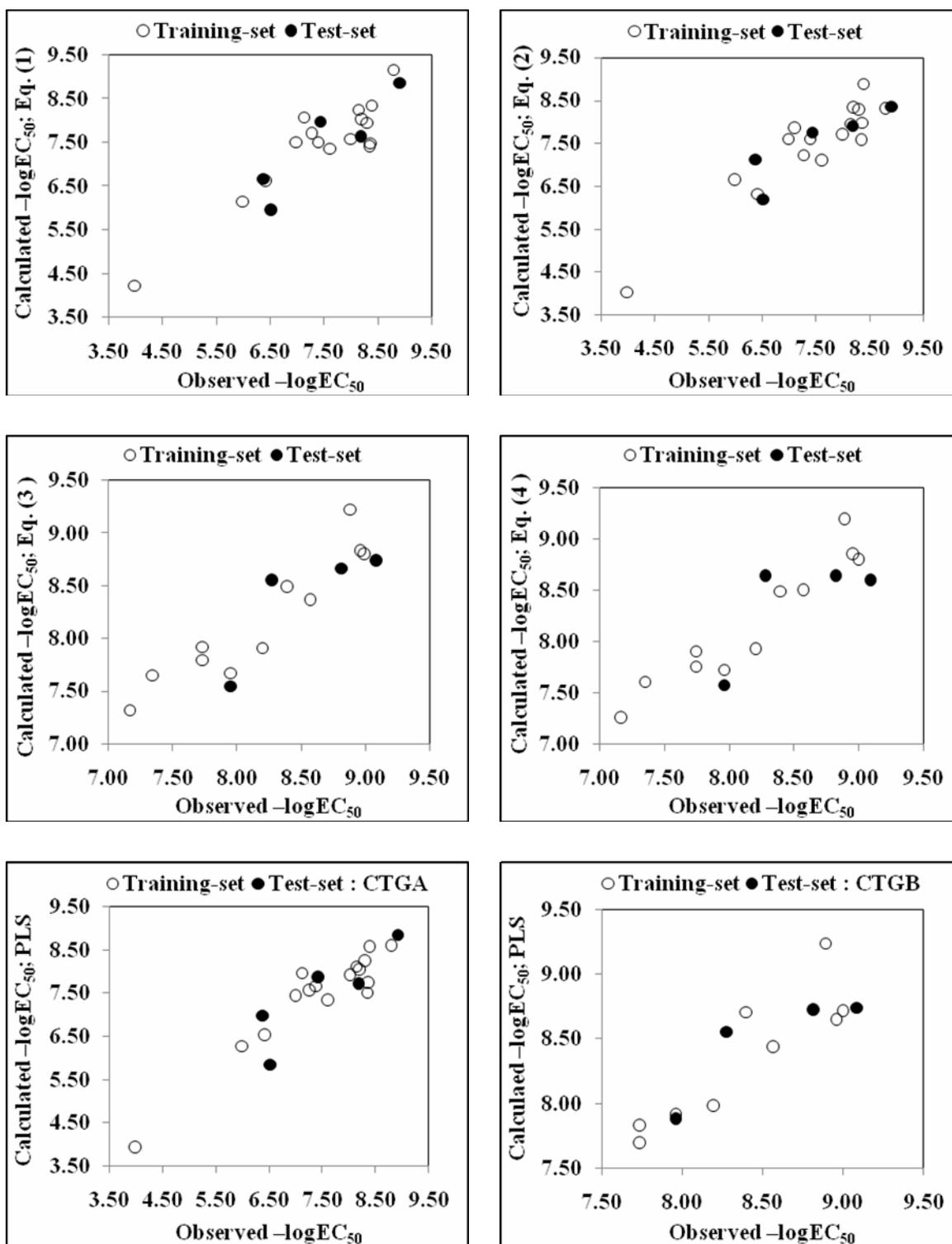
Fig. 2: Plot of observed *versus* caculated agonist activity, –logEC₅₀, values of training-set and test-set compounds from CTGA and CTGB
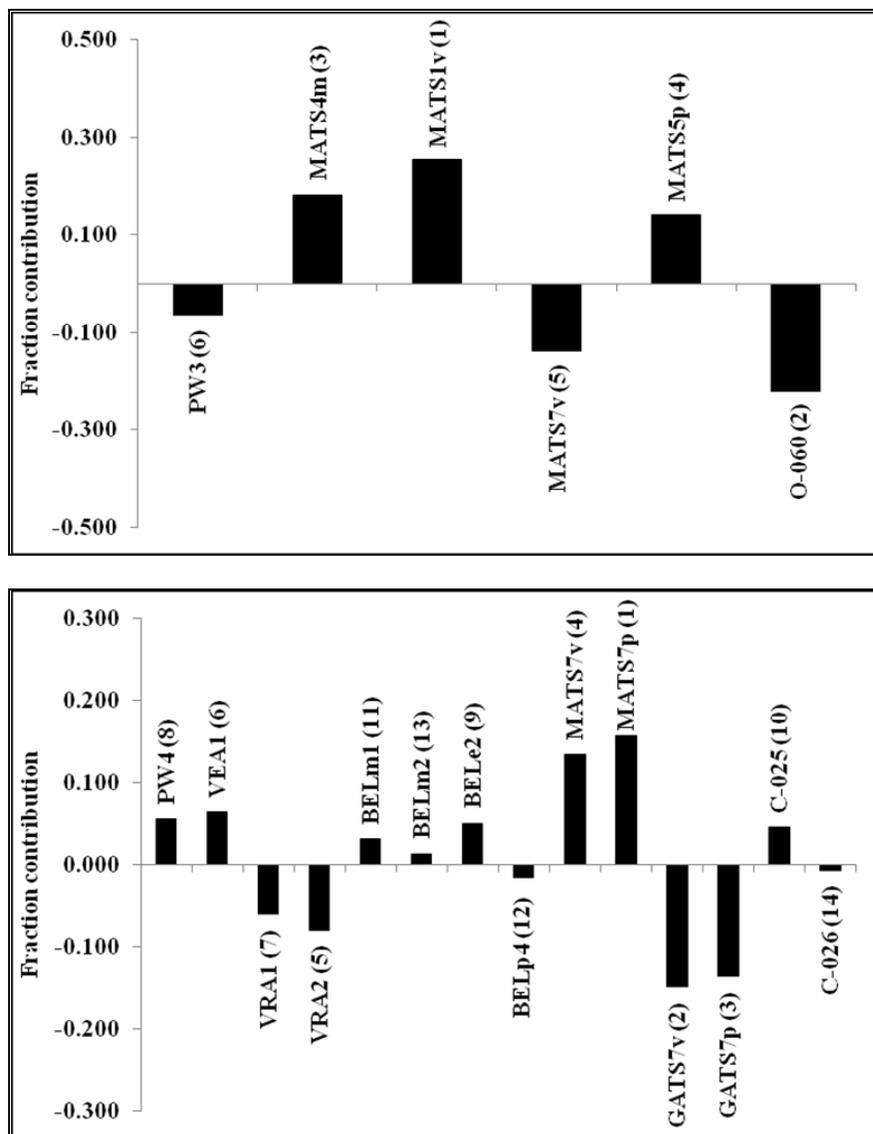
Fig. 3: Plot between fraction contribution of MLR-like PLS coefficients (normalized) and 6 and 14 identified descriptors (Table 5) associated with agonist actions of the compounds from CTGA and CTGB

The applicability domain (AD) has been analyzed for the models based on whole data-set relating to the activity profiles of the compounds. It is characterized by the Williams plot of standardized residuals *versus* leverage (h_i) values. For this purpose the participated descriptors in the most significant Eqs. (1)-(4) have been considered to derive corresponding models based on whole data sets which are given in Eqs. (5)-(8). The standardized residuals and leverage values, calculated in conjunction with them, are further used to ascertain their ADs. For compounds in CTGA:

$$-\log EC_{50} = 3.703 + 2.009\ (0.534)\ MATS4m + 3.243\ (0.405)\ MATS1v + 2.957\ (0.937)\ MATS5p$$

$$n = 21,\ r = 0.910,\ s = 0.519,\ F(3,17) = 27.168,\ Q^2_{LOO} = 0.704,\ Q^2_{L5O} = 0.550 \qquad\qquad ...(5)$$

$$-\log EC_{50} = 5.415 + 2.699\ (0.508)\ MATS4m + 3.649\ (0.912)\ MATS5p - 2.112\ (0.246)\ O\text{-}060$$

n = 21, r = 0.920, s = 0.490, F(3,17) = 31.082, Q^2_{LOO} = 0.767, Q^2_{L5O} = 0.661 ...(6)

and for compounds in CTGB:

$-logEC_{50}$ = 7.309 + 1.013 (0.236) BELe2 + 1.405 (0.201) MATS7p

n = 15, r = 0.912, s = 0.273, F(2,12) = 29.473 ...(7)

$-log\ EC_{50}$ = 7.326 + 1.489 (0.207) MATS7p + 0.890 (0.210) C-025

n = 15, r = 0.910, s = 0.275, F(2,12) = 29.026 …(8)

The limits of normal values for the standardized residuals (response or Y-outliers) were set as ±2.5 × s.d. while leverage threshold as h*. The graphical representations for these models are given in Fig. 4. From this figure, it is appeared that compound **36**, as expected, remains the obvious outlier from CTGB. The standardized residuals of this compound, calculated using the descriptors of Eqs. (3) and (4), have been found much deviated from the normal limits (±2.5 × s.d.). Except this congener, all remaining compounds (training-set and test-set), from CTGA and CTGB, remained within the squares, indicated that ADs are fully justified and identified models in Eqs. (1)-(4) have been evaluated correctly. Also, the derived models match the high quality parameters with good fitting power and capability of assessing external data.
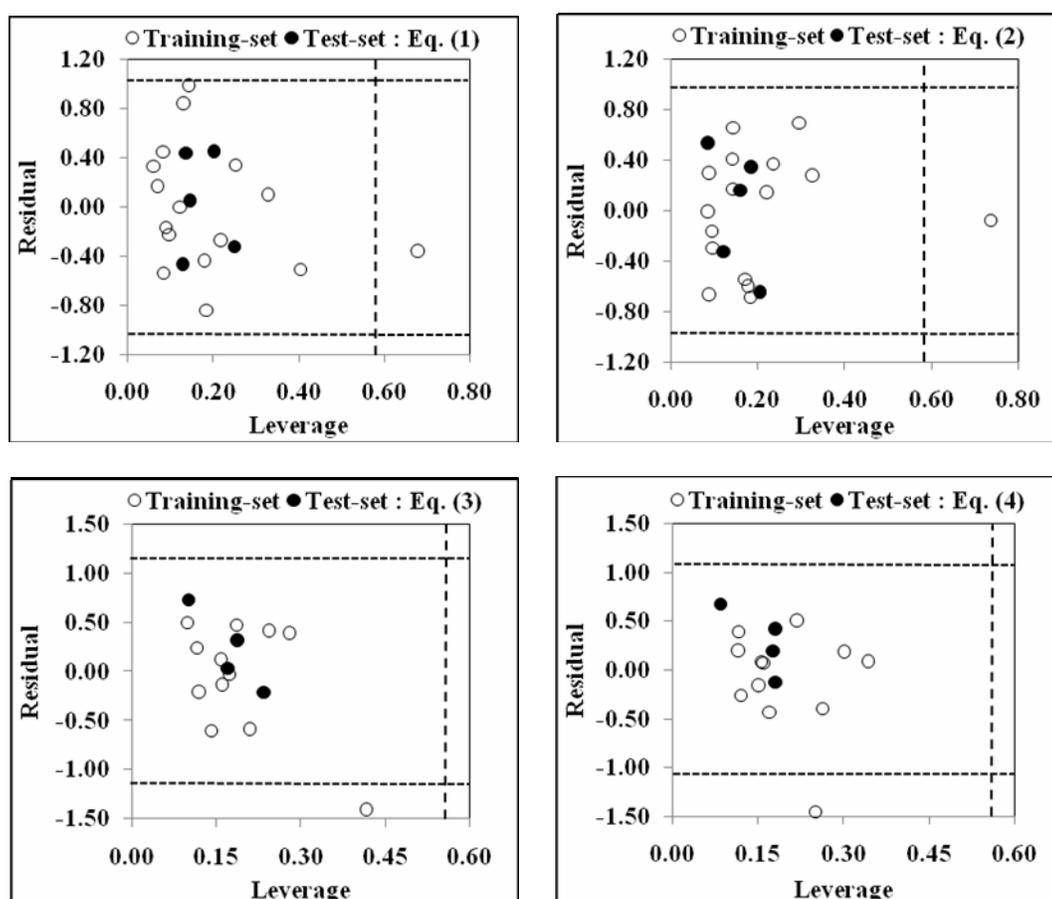
Fig. 4: Williams plot for whole data-set for agonist actions of compounds from CTGA and CTGB, listed in Table 1 and 2 respectively (h* values, in that order, are 0.571 and 0.563 and residual limits are ± 2.0 × s.d.)

CONCLUSIONS

The agonist activity of substituted-[1,2,4]oxadiazoles has been quantitatively analyzed in terms of molecular descriptors. The mode of interaction of 3-substuted-[1,2,4]oxadiazoles (CTGA) appeared significantly different from that of 5-(substituted phenyl)-[1,2,4]oxadiazoles (CTGB), therefore, compounds in these two categories have been considered separately to derive statistical models which were able to explain their agonist actions. For compounds in CTGA, tri-variant models have been identified to address the biological actions of the compounds. From highest significant models, it appeared that the higher values of the descriptors, MATS1v, MATS4m and MATS5p are conducive in improving the agonist action of a compound while the functionality such as Al-O-Ar or Ar-O-Ar or R-O-R/R-O-C=X, imparting detrimental effect to it, is undesirable. Thus, the Moran autocorrelations (MATS) of lags 1, 4 and 5 of a molecule weighted, respectively, by atomic van der Waals volumes (v), atomic masses (m) and atomic polarizabilities (p) remain the dominating factors during interaction with receptor. For compounds in CTGB, bi-variant models remained significant to describe their activity profiles. The higher values of the descriptors BELe2 (the lowest eigenvalue no. 2 of Burden matrix/ weighted by atomic Sanderson electronegativities) and MATS7p (Moran autocorrelation – lag 7/ weighted by atomic polarizabilities) or the descriptors MATS7p and C-025 (the fragment R--CR--R) are helpful in augmenting the activity of a compound. Thus the electronic and polarization effects or polarization effect in addition to structural fragment, R--CR--R appeared to be governing features during interaction. PLS analysis has further confirmed the dominance of the CP-MLR identified descriptors. The guidelines, based on the statistically validated models, may facilitate in exploring more potential analogues of the series. Applicability domain analysis revealed that the suggested models have acceptable predictability. Except one obvious outlier compound from CTGB, all remaining compounds were within the applicability domain of the proposed models and were evaluated correctly.

ACKNOWLEDGEMENT

REFERENCES

1.   S. E. Gardell, A. E. Dubi and J. Chun, Trends Mol. Med., **12**, 65 (2006).

2.   I. Ishii, N. Fukushima, X. Ye and J. Chun, Annu. Rev. Biochem., **73**, 321 (2004).

3.   J. Chun, E. J. Goetzl, T. Hla, Y. Igarashi, K.R. Lynch, W. Moolenaar, S. Pyne and G. Tigyi, Pharmacol. Rev., **54**, 265 (2002).

4.   H. Rosen, P. J. Gonzalez-Cabrera, M. G. Sanna and S. Brown, Annu. Rev. Biochem., **78**, 743 (2009).

5.   S. Spiegel and S. Milstien, Nature Rev. Mol. Cell Biol., **4**, 397 (2003).

6.   A. Kihara, S. Mitsutake, Y. Mizutani and Y. Igarashi, Prog. Lipid Res., **46**, 126 (2007).

7.   N. C. Hait, C. A. Oskeritzian, S. W. Paugh, S. Milstien and S. Spiegel, Biochim. Biophys. Acta., **758**, 2016 (2006).

8.   M. Matloubian, C. G. Lo, G. Cinamon, M. J. Lesneski, Y. Xu, V. Brinkmann, M. L. Allende, R. L. Prola and J. G. Cyster, Nature, **427**, 355 (2004).

9.   J. J. Hale, W. Neway, S. G. Mills, R. Hajdu, C. A. Keohane, M. Rosenbach, J. Milligan, G. J. Shei, G. Chrebet, J. Bergstrom *et al.*, Bioorg. Med. Chem. Lett., **14**, 3351 (2004).

10. M. L. Allende, J. L. Dreier, S. Mandala and R. L. Proia, J. Biol. Chem., **279**, 15396 (2004).

11. D. Buzard, S. Han, L. Thoresen, J. Moody, L. Lopez, A. Kawasaki, T. Schrader, C. Sage, Y. Gao, J. Edwards, J. Barden, J. Thatte, L. F. M. Solomon, L. Liu, H. Al-Shamma, J. Gatlin, M. Le, C. Xing, S. Espinola and R. M. Jones Bioorg. Med. Chem. Lett., **21**, 6013 (2011).

12. SYSTAT, Version 7.0; SPSS Inc., 444 North Michigan Avenue, Chicago, IL 60611.

13. ChemDraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA.

14. DRAGON Software, Version 3.0-2003; by R. Todeschini, V. Consonni, A. Mauri and M. Pavan, Milano, Italy. Available from: http//www. disat.unimib.it/chm/Dragon.htm.

15. Y. S. Prabhakar, QSAR Comb. Sci., **22**, 583 (2003).

16. S. Sharma, Y. S. Prabhakar, P. Singh and B. K. Sharma, Eur. J. Med. Chem., **43**, 2354 (2008).

17. S. Sharma, B. K. Sharma, S. K. Sharma, P. Singh and Y. S. Prabhakar, Eur. J. Med. Chem., **44**, 1377 (2009).

18. B. K. Sharma, P. Pilania, P. Singh and Y. S. Prabhakar, SAR QSAR Environ. Res., **21**, 169 (2010).

19. B. K. Sharma, P. Singh, K. Sarbhai and Y. S. Prabhakar, SAR QSAR Environ. Res., **21**, 369 (2010).

20. B. K. Sharma, P. Pilania, K. Sarbhai, P. Singh and Y.S. Prabhakar, Mol. Divers., **14**, 371 (2010).

21. B. K. Sharma, P. Singh, M. Shekhawat, K. Sarbhai and Y.S. Prabhakar, SAR QSAR Environ. Res., **22**, 365-383 (2011).

22. S.-S. So and M. Karplus, J. Med. Chem., **40**, 4347 (1997).

23. Y. S. Prabhakar, V. R. Solomon, R. K. Rawal, M. K. Gupta and S. B. Katti, QSAR Comb. Sci., **23**, 234 (2004).

24. P. Gramatica, QSAR Comb. Sci., **26**, 694 (2007).

25. L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell and P. Gramatica, Environ. Health Perspec., **111**, 1361 (2003).

26. S. Wold, Technomet., 20, 397 (1978).

27. N. Kettaneh, A. Berglund and S. Wold, Comput. Stat. Data Anal., **48**, 69 (2005).

28. L. Stahle and S. Wold, In: G. P. Ellis, G. B. West, (Eds.), Progress in Medicinal Chemistry, **Vol. 5**, Elsevier Science Publishers, BV (1988) p. 291.