GSC Advanced
Research and Reviews

Global Scholarly Communication
Online Press

(RESEARCH ARTICLE)

CrossMark
click for updates

CP-MLR derived QSAR study on the PPARγ agonists: The benzylpyrazole acylsulfonamide derivatives

Raghuraj Parihar and Brij Kishore Sharma *

Department of Chemistry, Government College, Bundi-323 001 (Rajasthan), India.

Abstract

The PPARγ transactivation activity of benzylpyrazole acylsulfonamide derivatives have been quantitatively analyzed in terms of 0D- to 2D-Dragon descriptors. This study has provided a rational approach for the development of titled derivatives as PPARγ agonists. The descriptors identified in CP-MLR analysis for the PPARγ transactivation activity have highlighted the role of atomic properties (mass, electronegativity, van der Waals volumes and polarizability) in terms of weighted 2D autocorrelations and BCUT descriptors and electronic content in terms of Galvez charge indices and maximal electrotopological positive variation (MAXDP). Additionally, Balaban's U and centric indices (Uindex and BAC, respectively), Lopping centric index (Lop), topological distance between N and O atom and hydrophobicity accounting parameter MLOGP have also shown prevalence to optimize the PPARγ transactivation of titled compounds. PLS analysis has further confirmed the dominance of the CP-MLR identified descriptors and applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Keywords: QSAR; PPARγ transactivation; Combinatorial protocol in multiple linear regression (CP-MLR) analysis; Dragon descriptors; Benzylpyrazole acylsulfonamides.

1. Introduction

Elevated plasma glucose in the presence of high endogenous insulin levels is characteristic of type 2 diabetes (T2D). T2D is a complex metabolic disorder because insulin resistance and impaired insulin secretion lead to abnormal metabolism of glucose, lipids and amino acids. The quality of life of diabetic patients slowly decreases due to developed long-term micro- and macro- vascular complications like neuropathy, retinopathy, nephropathy, myocardial infarction, stroke, and lower limb amputation as the progression of the disease progresses [1,2]. The prevalence of T2DM in developed and developing countries is rising speedily and it is expected that number of diabetics to reach 380 million by 2025 [3.4]. Thus in this scenario development of new and safer antidiabetic agents which may lower hemoglobin A_{1c} (HbA_{1c}) levels and improve the lipid profile of patients simultaneously is ardently needed [5-8].

Peroxisome proliferator-activated receptors (PPARs), belonging to the nuclear receptor superfamily, are ligand activated transcription factors. At present three PPAR subtypes, PPARα, PPARγ, and PPARδ (also known as PPARβ), have been cloned and characterized. PPARs exert their effects through transcription of a constellation of genes encoding proteins. This transcription is ligand dependent and regulates nutrient metabolism, energy homeostasis, and cell differentiation. Among the PPAR subtypes the most extensively investigated subtype is PPARγ. It is expressed predominantly in adipose tissue, in a lesser extent in the intestine, mammary gland, endothelium, liver, skeletal muscle and in other tissues throughout the body. PPARγ plays a pivotal role in many physiological processes such as adipogenesis, glucose and lipid homeostasis, insulin sensitivity, inhibition of inflammatory responses, cell proliferation

* Corresponding author: Sharma Brij Kishore

and promotion of terminal differentiation [9-11]. Introduction of troglitazone, pioglitazone hydrochloride and rosiglitazone maleate (the representatives of thiazolidinediones (TZDs)) as insulin sensitizers and the fact that TZDs are high-affinity PPARγ ligands [12] has opened new avenues for extensive research in the area of antidiabetic drug discovery and development [13-15].

Efforts were made to indentify novel classes of PPAR ligands, based on several approaches such as PPARα/γ dual agonists, PPARγ/δ dual agonists and PPARα/γ/δ pan agonists, as second-generation insulin sensitizers [13]. Numerous reported non-TZD PPARγ ligands belonging to different chemical classes are mostly carboxylic acids. A novel class of benzylpyrazole acylsulfonamides as non-thiazolidinedione (TZD), non-carboxylic-acid-based selective PPARγ agonists has been reported by Rikimaru et al. [16]. The aim of present communication is to establish the quantitative relationships between the reported activities and molecular descriptors unfolding the substitutional changes in titled compounds.

2. Material and methods

2.1. Biological actions and theoretical molecular descriptors

The reported twentyeight benzylpyrazole acylsulfonamides are considered as the data set for this study [16]. The structural variations of these analogues are mentioned in Table 1. These derivatives were evaluated for their transactivation activity against human PPARγ stably expressed in Chinese hamster ovary (CHO) cells. Transactivation activities were assessed by a luciferase reporter gene assay using (R)-5-(3-{4-[(2-Furan-2-yl-5-methyl-1,3-oxazol-4-yl)methoxy]-3-methoxyphenyl}propyl)-1,3-oxazolidine-2,4-dione [17] as the reference PPARγ agonist and were reported as EC50 and the same are also presented in Table 1 as pEC50 on molar basis. For modeling purpose the data set has been sub-divided into training set (for model development) and test set (for external prediction or validation). The selection of test set compounds was made using an in-house written randomization program. The test and training set compounds are also mentioned in Table 1.
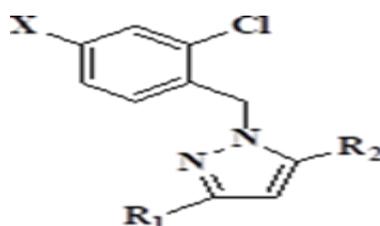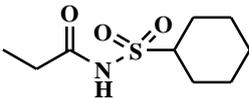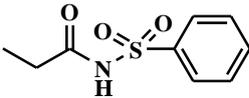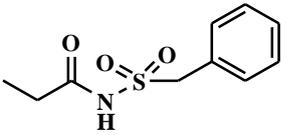

Table 1 Structural variations and reported PPARγ transactivation activities of benzylpyrazole acylsulfonamides.

Cpd.	X	R_1	R_2	$EC_{50}(M)^a$
1	Cl	Isopropoxy		6.35
2[b]	Cl	Isopropoxy		6.80
3	Cl	Isopropoxy		7.80
4	Cl	Isopropoxy		7.40
5[b]	Cl	Isopropoxy		7.68
6	Cl	Isopropoxy		7.77

No.				Value
7	Cl	Isopropoxy		6.82
8	Cl	Isopropoxy		6.96
9	Cl	Isopropoxy		6.59
10	Cl	Butoxy	$(CH_2)_4Me$	8.12
11[b]	Cl	$MeO(CH_2)_2O$	$(CH_2)_4Me$	7.51
12[b]	Cl	Benzyloxy	$(CH_2)_4Me$	7.85
13	Cl	2-Pyridylmethoxy	$(CH_2)_4Me$	6.92
14	Cl	Isopropyl	$(CH_2)_4Me$	7.89
15[b]	Cl	phenyl	$(CH_2)_4Me$	7.96
16	Cl	Butoxy	$(CH_2)_4Me$	-[c]
17	Cl	Butoxy	$(CH_2)_4Me$	-[d]
18[b]	Cl	Butoxy	$(CH_2)_4Me$	7.17
19	Cl	Butoxy	$(CH_2)_4Me$	7.54
20	CF₃	Isopropoxy	$(CH_2)_4Me$	8.00
21[b]	CF₃	Isopropyl	$(CH_2)_4Me$	8.08
22	CF₃	Cyclopropyl	$(CH_2)_4Me$	7.89
23	Cl	Isopropoxy	$(CH_2)_4Me$	7.70
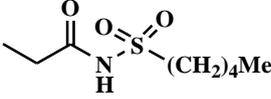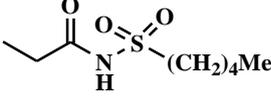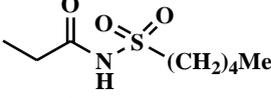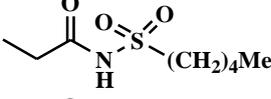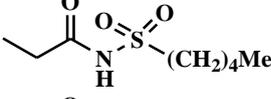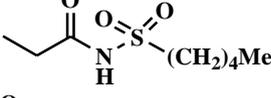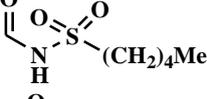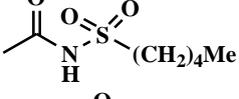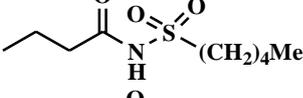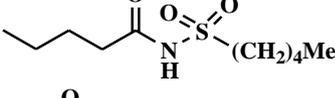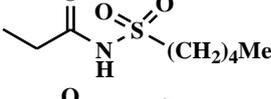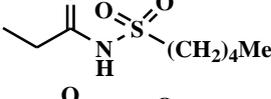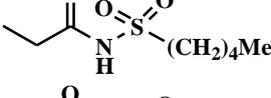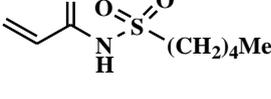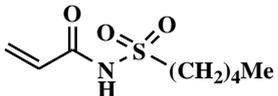
24	CF_3	Isopropoxy		8.03
25	CF_3	Isopropoxy		8.08
26	CF_3	$MeO(CH_2)_2O$		8.05
27	Cl	Isopropyl		8.09
28	CF_3	Cyclopropyl		7.92
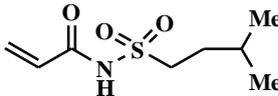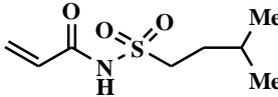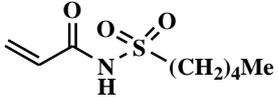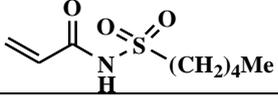
[a]EC50 (the effective concentration for 50% response of a given compound's intrinsic maximum response) on molar basis, taken from reference [16]; [b]Compound included in test set; [c]Inactive compound, not part of data set; [d]Compound with uncertain activity, not part of data set.

Table 2 Descriptor classes used for the modeling of PPARγ transactivation activity of benzylpyrazole acylsulfonamides.

S. No.	Descriptor Class (Acronyms)[a]	Definition and Scope
1	Constitutional (**CONST**)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
2	Topological (**TOPO**)	2D-descriptor from molecular graphs and independent conformations
3	Molecular walk counts (**MWC**)	2D-descriptors representing self-returning walk counts of different lengths
4	Modified Burden eigenvalues (**BCUT**)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights of the diagonal elements and atoms
5	Galvez topological charge indices (**GALVEZ**)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix
6	2D-autocorrelatons (**2D-AUTO**)	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)
7	Functional groups (**FUN**)	Molecular descriptors based on the counting of the chemical functional groups
8	Atom centered fragments (**ACF**)	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen
9	Empirical (**EMP**)	1D-descriptors represent the counts of nonsingle bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule
10	Properties (**PROP**)	1D-descriptors representing molecular properties of a molecule

[a]Reference [19].

The structures of the all the compounds (listed in Table 1) were drawn in 2D ChemDraw [18] and subjected to energy minimization in the MOPAC using the AM1 procedure for closed shell system after converting these into 3D modules. The energy minimization was carried out to attain a well defined conformer relationship among the congeners under study. The 0D- to 2D-molecular descriptors of titled compounds was computed using DRAGON software [19]. This software offers a large number of descriptors corresponding to ten different classes of 0D- to 2D-descriptor modules.

The different descriptor classes include the constitutional, topological, molecular walk counts, BCUT descriptors, Galvez topological charge indices, 2D-autocorrelations, functional groups, atom-centered fragments, empirical descriptors and the properties describing descriptors. These descriptors offer characteristic structural information specific to the descriptor class. The definition and scope of these descriptor's classes is given in Table 2.

A total number of 484 descriptors, belonging to 0D- to 2D- modules, have been computed to obtain most appropriate models describing the biological activity. Prior to model development procedure, all those descriptors that are inter-correlated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This procedure has reduced the total descriptors from 484 to 107 as relevant ones to explain the biological actions of titled compounds.

2.2. Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR) [20-24] and partial least squares (PLS) [25-27] procedures were used in the present work for developing QSAR models. The CP-MLR is a "filter"-based variable selection procedure, which employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure–activity models, each having unique combination of descriptors from the generated dataset of the compounds under study. The embedded filters make the variable selection process efficient and lead to unique solution. Fear of "chance correlations" exists where large descriptor pools are used in multilinear QSAR/QSPR studies [28,29]. In view of this, to find out any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been subjected to randomization test [30,31] by repeated randomization (100 simulation runs) of the biological responses. The datasets with randomized response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test its prediction and generalization within the study domain. For each model, derived by involving n data points, a number of statistical parameters such as r (the multiple correlation coefficient), s (the standard deviation), F (the F ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to access its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5. Additional statistical parameters such as, the Akaike's information criterion, AIC [32,33], the Kubinyi function, FIT [34,35] and the Friedman's lack of fit, LOF [36], have also been calculated to further validate the derived models. 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 FIT, closely related to the F-value, proved to be a useful parameter for assessing the quality of the models. A model which is derived in k independent descriptors, its F-value will be more sensitive if k is small while it becomes less sensitive if k is large. The FIT, on the other hand, will be less sensitive if k is small whereas it becomes more sensitive if k is large. The model that produces the lowest AIC value and highest FIT value is considered potentially the most useful and the best. The LOF factor takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large number of parameters.

2.3. Applicability domain

The usefulness of a model is based on its accurate prediction ability for new congeners. 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 (AD) is evaluated by the leverage values for each compound [37]. A Williams plot (the plot of standardized residuals versus leverage values (h)) is constructed, which can be used for a simple graphical detection of both the response outliers (*Y* outliers) and structurally influential chemicals (*X* outliers) in the model. In this plot, the AD is established inside a squared area within ±*x* standard deviations and a leverage threshold h^*, which 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 $x = 2$ or 3. If the compounds have a high leverage value ($h > h^*$), then the prediction is not trustworthy. 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.

3. Results and discussion

3.1. QSAR results

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modeling study, 7 compounds have been included in the test set for the validation of the models derived from 19 training set compounds. A total number of 107 significant descriptors from 0D- to 2D- classes have been subjected to CP-MLR analysis with default "filters" set in it. Statistical models in two descriptors have been derived to achieve the best relationship correlating PPARγ transactivation activity. A total number of seven models in two descriptors, having $r^2_{Test} > 0.5$, were obtained through CP-MLR. The selected models in two descriptors are given below.

$pEC_{50} = 6.337 + 1.221(0.229)BELm5 + 1.017(0.225)JGI4$

$n = 19, r = 0.871, s = 0.291, F = 25.371, Q^2_{LOO} = 0.658, Q^2_{L50} = 0.663$

$r^2_{Test} = 0.549, FIT = 2.206, LOF = 0.114, AIC = 0.116$ (1)

$pEC_{50} = 6.319 + 1.342(0.254)BELm5 + 0.810(0.218)JGI2$

$n = 19, r = 0.841, s = 0.321, F = 19.337, Q^2_{LOO} = 0.594, Q^2_{L50} = 0.594$

$r^2_{Test} = 0.503, FIT = 1.681, LOF = 0.139, AIC = 0.142$ (2)

$pEC_{50} = 6.387 + 1.080(0.265)BELm5 + 1.055(0.300)GGI4$

$n = 19, r = 0.831, s = 0.330, F = 17.991, Q^2_{LOO} = 0.597, Q^2_{L50} = 0.583$

$r^2_{Test} = 0.680, FIT = 1.564, LOF = 0.147, AIC = 0.149$ (3)

$pEC_{50} = 6.405 + 1.189(0.276)BELm5 + 0.836(0.274)GGI2$

$n = 19, r = 0.809, s = 0.349, F = 15.186, Q^2_{LOO} = 0.538, Q^2_{L50} = 0.501$

$r^2_{Test} = 0.597, FIT = 1.320, LOF = 0.165, AIC = 0.167$ (4)

where n, r, s and F represent respectively the number of data points, the multiple correlation coefficient, the standard deviation and the F-ratio between the variances of calculated and observed activities. In above and all follow-up regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models. 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. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation.

Most of the descriptors GGI2, GGI4, JGI2 and JGI4 participated in above models are from the GALVEZ class and the remained one BELm5 is the modified Burden eigenvalue (BCUT class descriptor). All the descriptors have shown positive influence on the activity as evident from the signs of regression coefficients. Thus a higher value of Galvez descriptors GGI2 (2nd order topological charge index), GGI4 (4th order topological charge index), JGI2 (2nd order mean topological charge index) and JGI4 (4th order mean topological charge index) in addition to a higher value of the lowest eigenvalue n.5 of Burden matrix/weighted by atomic masses (descriptor BELm5) would be beneficiary to the activity.

The two descriptor models could estimate nearly 76% in observed activity of the compounds. Considering the number of observation in the dataset, models with up to three descriptors were explored. It has resulted in 21 three-parameter models with test set $r^2 > 0.50$. These models (with 107 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this, the optimum *r*-bar value of the preceding level model (=0.854) has been used as the new threshold of filter-3 for the next generation. These models have shared 26 descriptors among them. All these 26 descriptors along with their brief meaning, average regression coefficients, and total incidence are listed in Table 3, which will serve as a measure of their estimate across these models.

Table 3 Identified descriptors[a] along with their class, average regression coefficient and incidence[b], in modeling the PPARγ transactivation activities of benzylpyrazole acylsulfonamides.

Descriptor class	Average regression coefficient and (incidence)
Topological descriptors (TOPO):	HNar, -0.692(1); MAXDP, 0.941(2); BAC, 1.180(2); Lop, 1.714(10); Uindex, 1.527(1); BIC3, 0.678(1); T(N..O),-0.778(3)
Modified Burden Eigen values (BCUT):	BELm5, 1.069(6); BEHv2, 0.986(6); BELv8, 1.368(1); BEHm3, 0.659(1)
Galvez Topological charge indices (GALVEZ):	GGI2, 0.621(1); GGI4, 0.885(4); GGI7, 0.659(1); JGI2, 0.517(1); JGI4, 0.825(4); JGT, 0.583(1)
2D autocorrelations (2D-AUTO):	MATS8m, -0.765(2); MATS4v, 0.890(6); MATS3e, 1.538(2); MATS3p, -0.690(1); MATS5p, -1.079(1); GATS5p, 0.603(2)
Functional groups (FUNC):	nCrH2, -0.856(1)
Empirical descriptors (EMP):	Hy, -7.354(1)
Properties (PROP):	MLOGP, 0.523(1)

[a]The descriptors are identified from the three parameter models for PPARγ binding activity transactivation activity emerged from CP-MLR protocol with filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.854 and filter-4 as $0.3 \leq q2 \leq 1.0$ with a training set of 19 compounds. [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. TOPO: HNar, Narumi harmonic index; MAXDP, maximal electrotopological positive variation; BAC, Balaban centric index; Lop, Lopping centric index; Uindex, Balaban U index; BIC3, bond information content of 3rd order neighborhood symmetry; T(N..O), sum of topological distances between N..O; BCUT: BEHm3, highest eigenvalue n.3 of Burden matrix/weighted by atomic masses; BELm5, lowest eigenvalue n.5 of Burden matrix/weighted by atomic masses; BEHv2, highest eigenvalue n.2 of Burden matrix/weighted by van der Waals volumes; BELv8, lowest eigenvalue n.8 of Burden matrix/weighted by van der Waals volumes; GALVEZ: GGI2, topological charge index of order 2; GGI4, topological charge index of order 4; GGI7, topological charge index of order 7; JGI2, mean topological charge index of order 2; JGI4, mean topological charge index of order 4; JGT, global topological charge index; 2D-AUTO: MATS8m, Moran autocorrelation of lag-8/ weighted by atomic masses; MATS4v, Moran autocorrelation of lag-4/ weighted by atomic van der Waals volumes; MATS3e, Moran autocorrelation of lag-3/ weighted by atomic Sanderson electronegativities; MATS3p, Moran autocorrelation of lag-3/weighted by atomic polarizabilities; MATS5p, Moran autocorrelation of lag-5/ weighted by atomic polarizabilities; GATS5p, Geary autocorrelation of lag-5/ weighted by atomic polarizabilities; FUNC: nCrH2, number of ring secondary C(sp3); EMP: Hy, hydrophilic factor; PROP: MLOGP, Moriguchi octanol-water partition coefficient (logP).

Following are the selected three-descriptor models for the PPARγ transactivation activities of benzylpyrazole acylsulfonamides emerged through CP-MLR.

pEC_{50} = 5.456 + 1.120(0.297)MAXDP + 1.679(0.208)Lop + 0.807(0.169)JGI4

n = 19, r = 0.936, s = 0.215, F = 35.559, Q^2_{LOO} = 0.718, Q^2_{L5O} = 0.606

r^2_{Test} = 0.523, FIT = 3.809, LOF = 0.078, AIC = 0.071   (5)

pEC_{50} = 6.380 + 1.267(0.185)Lop – 0.583(0.222)T(N..O) + 1.022(0.208)GGI4

n = 19, r = 0.927, s = 0.229, F = 30.866, Q^2_{LOO} = 0.760, Q^2_{L5O} = 0.733

r^2_{Test} = 0.512, FIT = 3.307, LOF = 0.088, AIC = 0.080   (6)

pEC_{50} = 5.986 + 1.243(0.193)Lop + 0.894(0.228)GGI4 + 0.523(0.235)MLOGP

n = 19, r = 0.920, s = 0.240, F = 27.678, Q^2_{LOO} = 0.704, Q^2_{L5O} = 0.631

r^2_{Test} = 0.532, FIT = 2.965, LOF = 0.097, AIC = 0.088   (7)

pEC_{50} = 5.683 + 0.761(0.349)MAXDP + 1.560(0.241)Lop + 0.892(0.230)GGI4

n = 19, r = 0.919, s = 0.241, F = 27.382, Q^2_{LOO} = 0.685, Q^2_{L5O} = 0.773

r^2_{Test} = 0.556, FIT = 2.933, LOF = 0.098, AIC = 0.089   (8)

These models have accounted for nearly 88% variance in the observed activities. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The values greater than 0.5 of Q^2 index is in accordance to a reasonable robust QSAR model. The pEC_{50} values of training set compounds calculated using Eqs. (5) to (8) and predicted from LOO procedure have been included in Table 4.

Table 4 Observed and modeled PPARγ transactivation activity of benzylpyrazole acylsulfonamides.

S. No.	pEC$_{50}$(M)[a]										
		Eq. (5)		Eq. (6)		Eq. (7)		Eq. (8)		PLS	
	Obsd[b]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]
1	6.35	6.48	7.08	6.89	7.09	6.61	6.86	6.46	7.05	6.57	6.64
2[d]	6.8	6.91	-d	7.09	-d	6.87	-d	6.84	-d	6.85	-d
3	7.8	7.70	7.69	7.57	7.53	7.54	7.50	7.65	7.62	7.58	7.55
4	7.4	7.57	7.58	7.48	7.49	7.38	7.37	7.46	7.47	7.24	7.22
5[d]	7.68	7.81	-d	7.65	-d	7.69	-d	7.81	-d	7.84	-d
6	7.77	7.93	7.95	7.75	7.75	7.69	7.68	7.78	7.78	7.62	7.61
7	6.82	6.86	6.89	6.67	6.60	6.72	6.67	6.81	6.80	6.87	6.88
8	6.96	6.63	6.49	6.67	6.54	6.73	6.63	6.65	6.52	6.69	6.60
9	6.59	6.64	6.67	6.73	6.80	6.77	6.85	6.72	6.78	6.66	6.68
10	8.12	7.79	7.68	7.66	7.58	7.70	7.61	7.80	7.69	7.73	7.62
11[d]	7.51	7.76	-d	7.40	-d	7.38	-d	7.78	-d	7.46	-d
12[d]	7.85	7.43	-d	7.46	-d	7.61	-d	7.59	-d	7.57	-d
13	6.92	7.42	7.53	7.00	7.11	7.29	7.35	7.58	7.68	7.33	7.38
14	7.89	8.12	8.17	8.00	8.01	7.92	7.92	7.89	7.89	7.95	7.96
15[d]	7.96	7.62	-d	7.64	-d	7.71	-d	7.65	-d	7.68	-d
16[e]	-e	-e	-e	-e	-e	-e	-e	-e	-e	-e	-e
17[e]	-e	-e	-e	-e	-e	-e	-e	-e	-e	-e	-e
18[d]	7.17	7.54	-d	7.56	-d	7.76	-d	7.69	-d	7.74	-d
19	7.54	7.57	7.59	7.64	7.68	7.99	8.25	7.80	7.86	7.83	7.97
20	8.00	7.95	7.94	7.98	7.98	8.01	8.01	8.01	8.01	8.03	8.04
21[d]	8.08	8.37	-d	8.42	-d	8.39	-d	8.25	-d	8.38	-d
22	7.89	7.93	7.94	8.04	8.07	8.05	8.09	7.91	7.91	7.87	7.87
23	7.70	7.57	7.55	7.57	7.55	7.50	7.47	7.56	7.53	7.63	7.63
24	8.03	7.81	7.80	7.98	7.98	7.97	7.96	7.92	7.90	8.09	8.09
25	8.08	8.00	7.99	8.17	8.20	8.12	8.13	8.05	8.04	8.17	8.21
26	8.05	8.08	8.08	8.01	7.99	7.97	7.87	8.20	8.24	8.05	8.05
27	8.02	7.99	7.98	8.00	8.00	7.88	7.87	7.80	7.78	8.02	8.02
28	7.92	7.80	7.78	8.04	8.06	8.02	8.03	7.81	7.80	7.92	7.92

[a]On molar basis; [b]Taken from ref. [16]; [c]Leave-one-out (LOO) procedure; [d]Compound included in test set; [e]Compound with uncertain activity or inactive, not part of data set.

The models (5) to (8) are validated with an external test set of 7 compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the same is reported in Table 4. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.
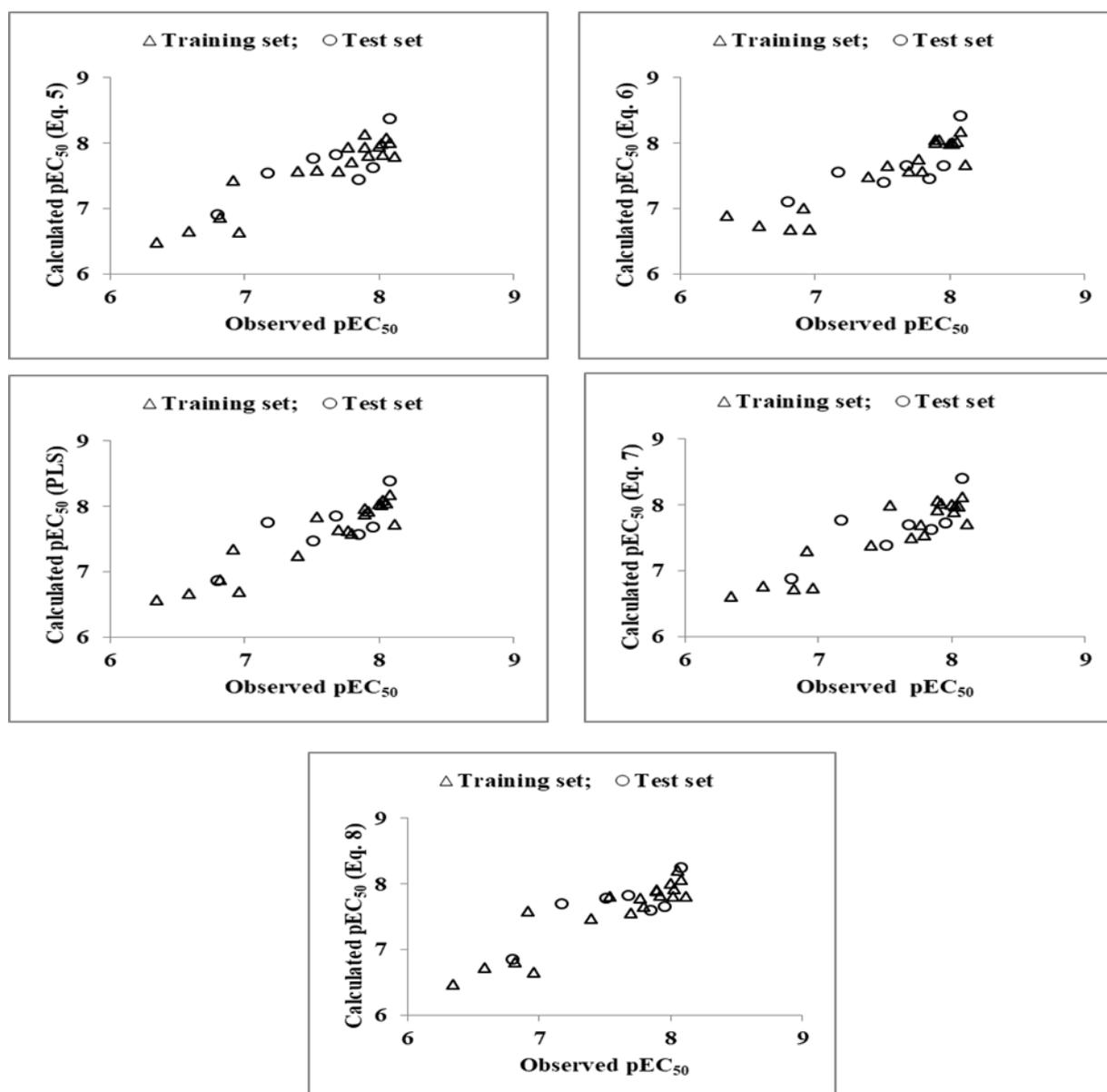
Figure 1 Plot of observed and calculated pEC_{50} values of training- and test-set compounds for PPARγ transactivation.

The newly appeared descriptors in above models, MAXDP, Lop and T(N..O) are topological class descriptors whereas MLOGP belongs to properties class. Descriptors MAXDP, Lop and MLOGP have shown positive and descriptor T(N..O) negative correlation to the activity. The signs of regression coefficients advocated that higher values of maximal electrotopological positive variation (descriptor MAXDP), Lopping centric index (descriptor Lop) and Moriguchi octanol-water partition coefficient i.e. logP (descriptor MLOGP) would be incremental to the activity. On the other hand a higher value of sum of topological distances between N..O would be deleterious to the activity.

A partial least square (PLS) analysis has been carried out on these 26 CP-MLR identified descriptors (Table 3) to facilitate the development of a "single window" structure–activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, two components are found to be the optimum for these 10 descriptors and they explained 88.36% variance in the activity ($r^2 = 0.940$, $Q^2_{LOO} = 0.819$, $s = 0.202$, $F = 60.955$, $r^2_{Test} = 0.517$). The MLR-like PLS coefficients of these 26 descriptors are given in Table 5.

Table 5 PLS and MLR-like PLS models from the descriptors of three parameter CP-MLR models for PPARγ transactivation activities.

A: PLS equation								
PLS components					PLS coefficient (s.e.)[a]			
Component-1					-0.171(0.016)			
Component-2					-0.078(0.021)			
Constant					7.571			
B: MLR-like PLS equation								
S. No.	Descriptor	MLR-like coefficient[b]	(f.c.)[c]	Order	S. No.	Descriptor	MLR-like coefficient[b]	(f.c.)[c]	Order
1	HNar	-0.149	-0.043	10	14	GGI7	0.116	0.037	12
2	MAXDP	-0.026	-0.006	24	15	JGI2	0.053	0.020	20
3	BAC	0.156	0.048	9	16	JGI4	0.100	0.033	14
4	Lop	0.287	0.092	1	17	JGT	-0.018	-0.005	25
5	Uindex	0.194	0.064	5	18	MATS8m	-0.204	-0.063	6
6	BIC3	0.111	0.026	18	19	MATS4v	0.216	0.065	4
7	T(N..O)	-0.151	-0.040	11	20	MATS3e	0.310	0.067	3
8	BEHm3	0.061	0.019	21	21	MATS3p	-0.022	-0.006	23
9	BELm5	0.272	0.088	2	22	MATS5p	-0.006	-0.002	26
10	BEHv2	-0.111	-0.033	13	23	GATS5p	0.094	0.029	17
11	BELv8	0.210	0.061	7	24	nCrH2	-0.117	-0.032	15
12	GGI2	0.068	0.022	19	25	Hy	-0.442	-0.016	22
13	GGI4	0.110	0.031	16	26	MLOGP	0.194	0.053	8
Constant					6.558			
C: PLS regression statistics					Values			
n					19			
r					0.940			
s					0.202			
F					60.955			
FIT					5.300			
LOF					0.055			
AIC					0.056			
Q^2_{LOO}					0.819			
Q^2_{L5O}					0.797			
r^2_{Test}					0.517			

[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; [c]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.

For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is also given in Figure 1. Figure 2 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity.
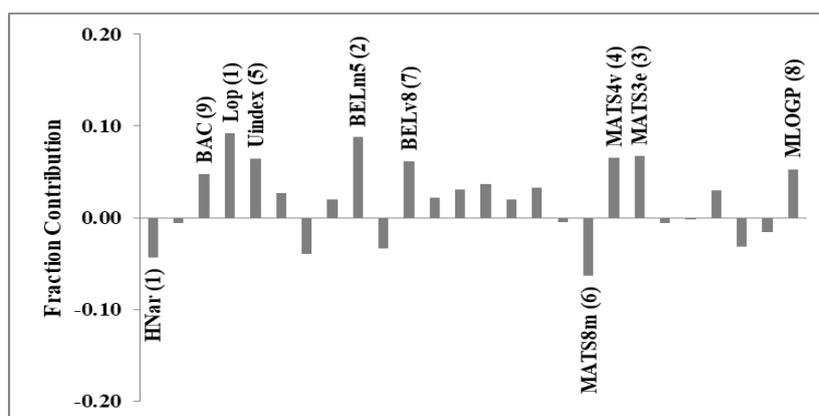
Figure 2 Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 26 CP-MLR identified descriptors (Table 4) associated with PPARγ transactivation activity of benzylpyrazole acylsulfonamide derivatives

The PLS analysis has suggested Lop as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 4 in Table 5; Figure 2). The other nine significant descriptors in decreasing order of significance are BELm5, MATS3e, MATS4v, Uindex, MATS8m, BELv8, MLOGP, BAC and HNar. Descriptors Lop, BELm5 and MLOGP are part of Eqs. (1) to (8) and convey same inference in the PLS model as well.

It is inferred from the PLS analysis that a higher values of 2D autocorrelation descriptors MATS3e (Moran autocorrelation of lag-3/ weighted by atomic Sanderson electronegativities) and MATS4v (Moran autocorrelation of lag-4/ weighted by atomic van der Waals volumes), topological descriptors Uindex (Balaban U index) and BAC (Balaban centric index); and modified Burden eigenvalue class descriptor BELv8 (lowest eigenvalue n.8 of Burden matrix/weighted by van der Waals volumes) would be advantageous to the activity. Based on the similar grounds a lower value of Moran autocorrelation of lag-8/ weighted by atomic masses (descriptor MATS8m) and Narumi harmonic index (descriptor HNar) will be supportive to the activity. It is also observed that PLS model from the dataset devoid of CP-MLR identified 26 descriptors (Table 3) is inferior in explaining the activity of the analogues.

3.2. Applicability domain (AD)

On analyzing the model AD in the Williams plot (Figure 3) of the model based on the whole dataset (Table 6), it has appeared that none of the compounds were identified as an obvious outlier for the PPARγ transactivation activities if the limit of normal values for the *Y* outliers (response outliers) was set as 3 (standard deviation) units. One compound listed in Table 1 at S. No. 1 found to have leverage (h) values greater than the threshold leverage (h*) suggesting them as chemically influential compounds. For both the training-set and test-set, the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data. Furthermore, all of the compounds were within the applicability domain of the proposed model and were evaluated correctly.

Table 6 Models derived for the whole data set (n = 26) for the PPARγ transactivation activity in descriptors identified through CP-MLR.

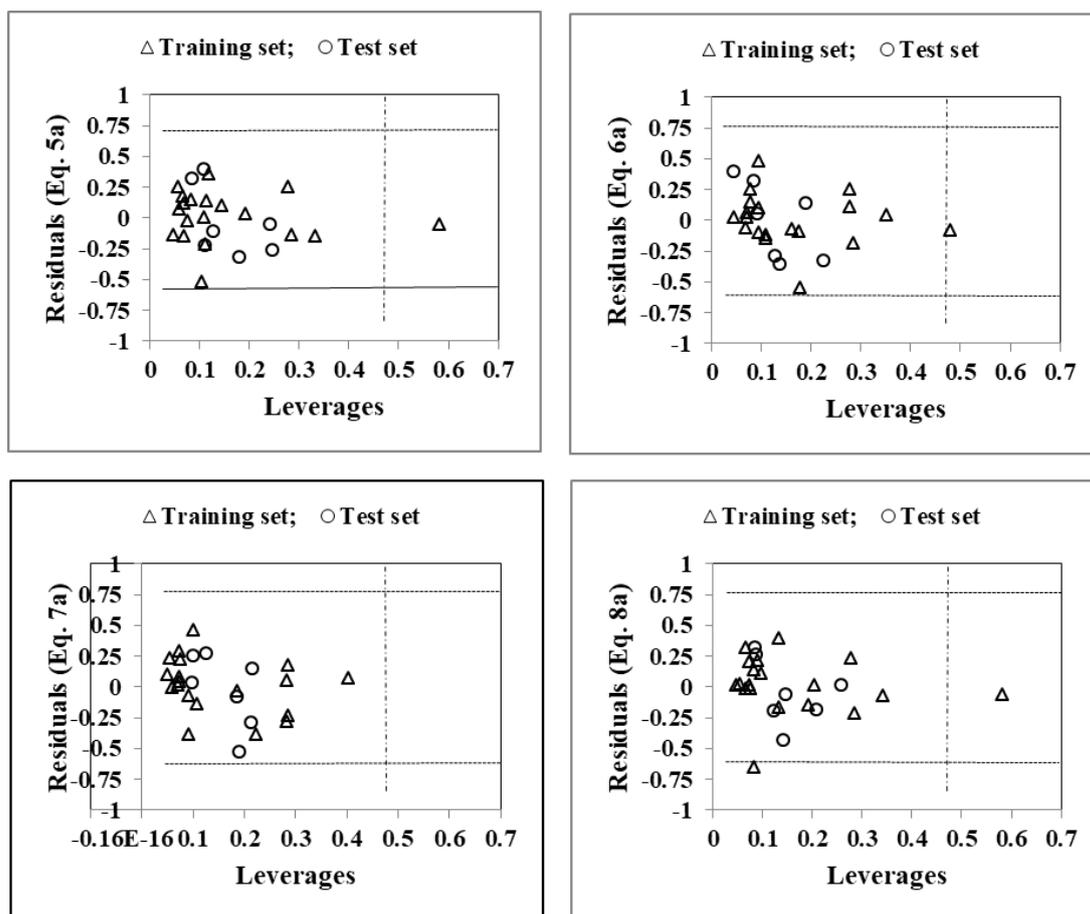| Model	r	s	F	Q^2_{LOO}	Eq.
pEC50 = 5.430 + 1.259(0.267)MAXDP + 1.583(0.197)Lop + 0.795(0.140)JGI4	0.905	0.237	33.591	0.756	(5a)
pEC50 = 6.419 + 1.191(0.189)Lop – 0.572(0.219)T(N..O) + 1.029(0.185)GGI4	0.893	0.251	29.140	0.721	(6a)
pEC50 = 6.063 + 1.135(0.190)Lop + 0.922(0.199)GGI4 + 0.471(0.189)MLOGP	0.891	0.254	28.302	0.702	(7a)
pEC50 = 5.699 + 0.796(0.290)MAXDP + 1.409(0.210)Lop + 0.977(0.186)GGI4	0.896	0.248	30.001	0.741	(8a)

Figure 3 Williams plot for the training-set and test- set compounds for PPARγ transactivation activity. The horizontal dotted line refers to the residual limit (±3×standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.46).

4. Conclusion

The PPARγ transactivation activity of benzylpyrazole acylsulfonamide derivatives have been quantitatively analyzed in terms of 0D- to 2D-Dragon descriptors. This study has provided a rational approach for the development of titled derivatives as PPARγ agonists. The descriptors identified in CP-MLR analysis for the PPARγ transactivation activity have highlighted the role of atomic properties (mass, electronegativity, van der Waals volumes and polarizability) in terms of weighted 2D autocorrelations and BCUT descriptors and electronic content in terms of Galvez charge indices and maximal electrotopological positive variation (MAXDP). Additionally, Balaban's Uand centric indices (Uindex and BAC, respectively), Lopping centric index (Lop), topological distance between N..O and hydrophobicity accounting parameter MLOGP have also shown prevalence to optimize the PPARγ transactivation of titled compounds. PLS analysis has further confirmed the dominance of the CP-MLR identified descriptors and applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare no conflict of interest.

References

[1] American Diabetes Association. (2004). Diagnosis and classification of diabetes mellitus. Diabetes Care, 27, S5.

[2] Porte D. Jr. and Schwartz MW. (1996). Diabetes complications--Why is glucose potentially toxic? Science, 272, 699.

[3] Kopelman PG and Hitman GA. (1998). Diabetes. Exploding type II. Lancet, 352, Suppl. 4, SIV5.

[4] King H, Aubert RE and Herman WH. (1998). Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. Diabetes Care, 21, 1414-1431.

[5] Ohkubo Y. Kishikawa H, Araki E, Miyata T, Lsami S, Motoyoshi S, Kojima Y, Furuyoshi N and Shichiri M. (1995). Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. Diabetes Research and Clinical Practice, 28, 103-117.

[6] Haffner SM, Alexander CM, Cook TJ, Boccuzzi SJ, Musliner TA, Pedersen TR, Kjekshus J and Pyörälä K. (1999). Reduced coronary events in simvastatin-treatedpatients with coronary heart disease and diabetes or impaired fasting glucose levels: Subgroup analyses in the Scandinavian simvastatin survival study. Archives of Internal Medicines, 159, 2661-2667;

[7] Pyörälä K, Pedersen TR, KJekshus J, Faergeman O, Olsson AG and Thorgeirsson G. (1997). Cholesterol lowering with simvastatin improves prognosis of diabetic patients with coronary heart disease. A subgroup analysis of the Scandinavian Simvastatin Survival Study (4S). Diabetes Care, 20, 614-620.

[8] Sacks FM, Pfeffer MA, Moye LA, Rouleau JL, Rutherford ID, Cole TG, Brown L, Warnica JW, Arnold JMO, Wun CC, Davis BR and Braunwald E. (1996). The effect of pravastatin on coronary events after myocardial infarction in patients with averagecholesterol levels. Cholesterol and recurrent events trial investigators. New England Journal of Medicine, 335, 1001-1009.

[9] Vamecq J and Latruffe N. (2000). Peroxisome proliferator-activated receptors (PPARs) and their implications in diseases. Current Opinion in Endocrinology and Diabetes, 7, 8-18.

[10] Evans RM, Barish GD and Wang Y-X. (2004). PPARs and the complex journey to obesity. Nature Medicine, 10, 355-361.

[11] Berger J and Moller DE. (2002). The mechanisms of action of PPARs. Annual Review of Medicine, 53, 409-435.

[12] Lehmann JM, Moore LB, S-Oliver TA, Wilkison WO, Willson TM and Kliewer SA.(1995). An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). The Journal of Biological Chemistry, 270, 12953-12956.

[13] Cho N and Momose Y. (2008). Peroxisome proliferator-activated receptor gamma agonists as insulin sensitizers: from the discovery to recent progress. Current Topics in Medicinal Chemistry, 8, 1483-1507.

[14] Willson TM, Brown PJ, Sternbach DD and Henke BR. (2000). The PPARs: from orphan receptors to drug discovery. Journal of Medicinal Chemistry, 43, 527-550.

[15] Berger JP, Akiyama TE and Meinke PT. (2005). PPARs: therapeutic targets for metabolic disease. Trends in Pharmacological Sciences, 26, 244-251.

[16] Rikimaru K, Wakabayashi T, Abe H, Imoto H, Maekawa T, Ujikawa O, Murase K, Matsuo T, Matsumoto M, Nomura C, Tsuge H, Arimura N, Kawakami K, Sakamoto J, Funami M, Mol CD, Snell GP, Bragstad KA, Sang B-C, Dougan DR, Tanaka T, Katayama N, Horiguchi Y and Momose Y. (2012). A new class of non-thiazolidinedione, non-carboxylic-acid-based highly selective peroxisome proliferator-activated receptor (PPAR)γagonists: Design and synthesis of benzylpyrazole acylsulfonamides. Bioorganic and Medicinal Chemistry, 20, 714-733.

[17] Momose Y, Maekawa T, Yamano T, Kawada M, Odaka H, Ikeda H and Sohda T. (2002).Novel 5-substituted 2,4-thiazolidinedione and 2,4-oxazolidinedione derivatives as insulins ensitizers with antidiabetic activities. Journal of Medicinal Chemistry, 45, 1518-1534.

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

[19] Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy. http://www.talete.mi.it/dragon.htm

[20]   Prabhakar YS. (2003). A combinatorial approach to the variable selection in multiple   linear   regression: analysis of Selwood et al. Data Set-a case study. QSAR and Combinatorial Science, 22, 583-595.

[21]   Sharma S, Prabhakar YS, Singh P and Sharma BK. (2008). QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR   approach. European Journal of Medicinal Chemistry, 43, 2354-2360.

[22]   Sharma S, Sharma BK, Sharma SK, Singh P and Prabhakar YS. (2009). Topological descriptors in modeling the agonistic activity of human A3 adenosine receptor ligands: The derivatives of 2-Chloro-N6-substituted-4'-thioadenosine-5'-uronamide. European Journal of Medicinal Chemistry, 44, 1377-1382.

[23]   Sharma BK, Pilania P, Singh P and Prabhakar YS. (2010). Combinatorial protocol inmultiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. SAR and QSAR inEnvironmental Research, 21, 169-185.

[24]   Sharma BK, Singh P, Sarbhai K and Prabhakar YS. (2010). A quantitative structure-activity relationship study on serotonin 5-HT6 receptor ligands: Indolyl and piperidinyl sulphonamides. SAR and QSAR in Environmental Research, 21, 369-388.

[25]   Wold S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. Technometrics, 20, 397–405.

[26]   Kettaneh N, Berglund A and Wold S. (2005). PCA and PLS with very large data sets. Computational Statistics and Data Analysis, 48, 69-85.

[27]   Stahle L and Wold S. (1988). Multivariate data analysis and experimental design. In: Ellis GP and West WB. (Eds.), Biomedical research. Progress in medicinal chemistry. Elsevier Science Publishers, BV, Amsterdam.25, 291–338.

[28]   Topliss JG and Edwards RP. (1979). Chance factors in studies of quantitative structure–activity relationships. Journal of Medicinal Chemistry, 22, 1238–1244.

[29]   Katritzky AR, Dobchev DA, Slavov S and Karelson M. (2008). Legitimate utilization of large descriptor pools for QSAR/QSPR models. Journal of Chemical Information and Modeling, 48, 2207–2213.

[30]   So S-S and Karplus M. (1997). Three-dimensional quantitative structure–activity relationship from molecular similarity matrices and genetic neural networks. 1. Method   and validation. Journal of Medicinal Chemistry, 40, 4347–4359.

[31]   Prabhakar YS, Solomon VR, Rawal RK, Gupta MK and Katti SB. (2004). CP-MLR/PLS   directed   structure–activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR and Combinatorial Science, 23, 234-244.

[32]   Akaike H. (1973). Information theory and an extension of the minimum likelihood principle. In: Petrov BN and Csaki F. (Eds.). Second international symposium oninformation theory. Budapest: Akademiai Kiado, 267–281.

[33]   Akaike H. (1974). A new look at the statistical identification model. IEEE Transactions on   Automatic   Control, AC-19, 716–723.

[34]   Kubinyi H. (1994). Variable selection in QSAR studies. I. An evolutionary algorithm. Quantitative Structure-Activity Relationship, 13, 285–294.

[35]   Kubinyi H. (1994). Variable selection in QSAR studies. II. A highly efficientcombination ofsystematic search and evolution. Quantitative Structure-Activity Relationship, 13, 393–401.

[36]   Friedman J. In Technical Report No. 102. Laboratory for Computational Statistics, Stanford University: Stanford; 1990.

[37]   Gramatica P. (2007). Principles of QSAR models validation: internal and external. QSAR and Combinatorial Science, 26, 694-701.

How to cite this article