

Available online at [GSC Online Press Directory](https://www.gsconlinepress.com/)

GSC Biological and Pharmaceutical Sciences

e-ISSN: 2581-3250, CODEN (USA): GBPSC2

Journal homepage: <https://www.gsconlinepress.com/journals/gscbps>

(RESEARCH ARTICLE)



CP-MLR derived QSAR rationales for the PPAR γ agonistic activity of the pyridyloxybenzene-acylsulfonamide derivatives

Parihar Raghuraj, Jahan Afsar and Sharma Brij Kishore *

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

Publication history: Received on 20 July 2020; revised on 29 July 2020; accepted on 30 July 2020

Article DOI: <https://doi.org/10.30574/gscbps.2020.12.1.0231>

Abstract

QSAR rationales have been obtained for the PPAR γ transactivation activity of pyridyloxybenzene-acylsulfonamides in terms of 0D- to 2D-Dragon descriptors. The descriptors identified in CP-MLR analysis have highlighted the role of atomic mass, van der Waals volumes and polarizability through weighted 2D autocorrelations (GATS1v and GATS1p), modified Burden eigenvalue (BEHm4) and molecular weight (MW). Sum of topological distances between O and S atoms (descriptor T(O..S)), and N and Cl atoms (descriptor T(N..Cl)), average connectivity index chi-1(X1A) and Quadratic index (Qindex) have also shown dominance to optimize the PPAR γ transactivation. Descriptors RBN and RBF suggested presence of rotatable bonds in a molecular structure for better PPAR γ activity. 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; PPAR γ transactivation; combinatorial protocol in multiple linear regression (CP-MLR) analysis; Dragon descriptors; Pyridyloxybenzene-acylsulfonamides.

1. Introduction

Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder. It is characterized by impaired insulin secretion, insulin resistance and hyperglycemia. More than 90% of diabetic patients are T2DM cases. These metabolic disorders develop, in a long term, many other disorders like dyslipidemia, hypertension and coronary heart disease. Sedentary lifestyle and obesity in association with these risk factors increases morbidity and mortality [1]. It is expected that by the year 2025 there would be 380 million T2DM cases [2]. To cure T2DM a class of new therapeutic agents thiazolidinediones (TZDs), the representatives of which are pioglitazone and rosiglitazone, emerged as insulin sensitizers. The identification and optimization of TZDs was devoid of the knowledge of the target protein. In the field of antidiabetic drug discovery and development, the findings that TZDs are high affinity ligands for peroxisome proliferator-activated receptor γ (PPAR γ) [3], opened channels for the extensive research [4-9]. The binding of TZD activates PPAR γ which functions as an essential transcriptional regulator of glucose and lipid homeostasis. PPAR γ is the most broadly studied subtype among the three PPAR subtypes (namely designated as PPAR α , PPAR γ , and PPAR δ). PPAR γ , expressed predominantly in adipose tissue, regulates the expression of a constellation of genes which is closely related to adipocyte differentiation, glucose and lipid metabolism, insulin sensitivity, inflammatory responses and cell proliferation [3,10]. The majority of reported PPAR γ ligands like TZD, oxazolidinone and tetrazole possess a carboxylic acid or its heterocyclic bioisostere [11-16]. There is also an example of non-TZD and non-carboxylic acid PPAR γ agonists [17].

* Corresponding author: Brij Kishore

A novel class of pyridyloxybenzene-acylsulfonamides as non-thiazolidinedione (TZD), non-carboxylic-acid-based selective PPAR γ agonists has been reported by Rikimaru et al. [18]. 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 thirty four pyridyloxybenzene-acylsulfonamides are considered as the data set for this study [18]. These derivatives were evaluated for their transactivation activity against human PPAR γ stably expressed in Chinese hamster ovary (CHO) -K1 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 [19] as the reference PPAR γ agonist and were reported as EC₅₀. The general structure of these compounds is given in Figure 1.

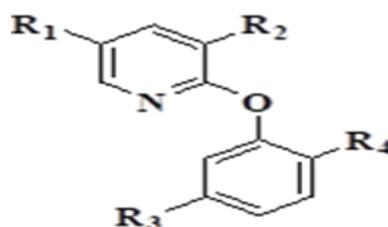
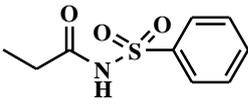
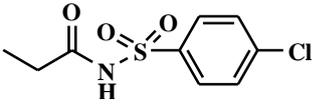
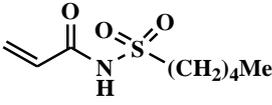
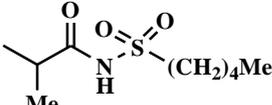
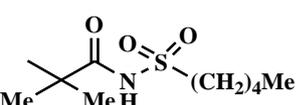
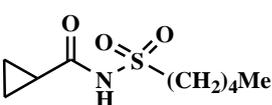
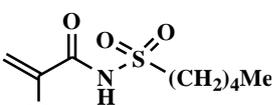
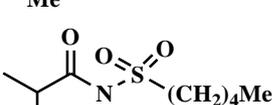
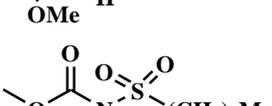
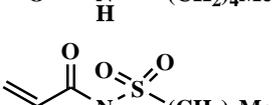
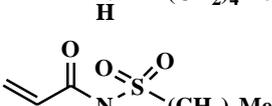
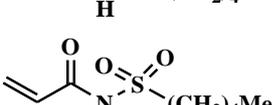
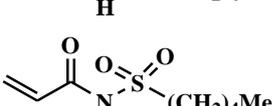
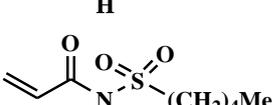
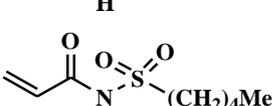


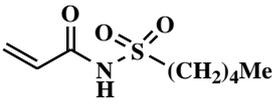
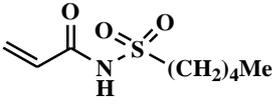
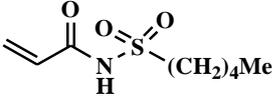
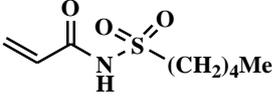
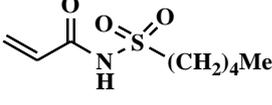
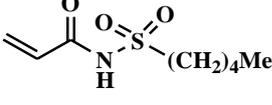
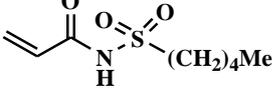
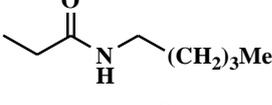
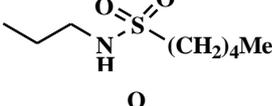
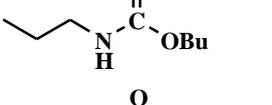
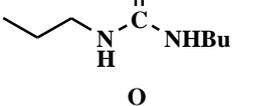
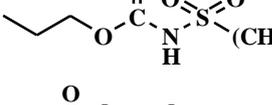
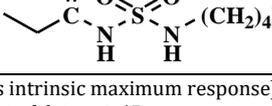
Figure 1 General structure of titled compounds

The structural variations of these analogues along with their reported pEC₅₀, on molar basis, are mentioned in Table 1. The data set has been sub-divided into training and test set. The models developed from training set have externally validated through test set. The test set compounds was selected 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 PPAR γ transactivation activities of pyridyloxybenzene-acylsulfonamides

Cpd.	R ₁	R ₂	R ₃	R ₄	pEC ₅₀ (M) ^a
1	CF ₃	Cl	H		7.21
2	CF ₃	Cl	<i>i</i> -PrO		8.59
3 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		8.82
4	CF ₃	Cl	Et ₂ NC(=O)CH ₂ O		8.16
5	CF ₃	Cl			8.47
6	CF ₃	Cl	MeO(CH ₂) ₂ O		7.96

7 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		7.49
8 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		8.07
9 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		7.82
10	CF ₃	Cl	MeO(CH ₂) ₂ O		8.64
11 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		8.51
12	CF ₃	Cl	MeO(CH ₂) ₂ O		8.26
13 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		8.28
14 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		8.17
15	CF ₃	Cl	MeO(CH ₂) ₂ O		8.00
16 [*]	CF ₃	Cl	MeO(CH ₂) ₂ O		7.03
17 ^c	CF ₃	H	MeO(CH ₂) ₂ O		-c
18 ^c	H	Cl	MeO(CH ₂) ₂ O		-c
19	Cl	Cl	MeO(CH ₂) ₂ O		6.72
20 ^c	NH ₂	Me	MeO(CH ₂) ₂ O		-c
21 ^c	MeCONH	Me	MeO(CH ₂) ₂ O		-c

22	CF ₃	Cl	<i>i</i> -PrO-		7.59
23	CF ₃	Cl	<i>c</i> -PrO(CH ₂) ₂ O		8.21
24	CF ₃	Cl	MeO(CH ₂) ₃ O		8.13
25 ^b	CF ₃	Cl	HO(CH ₂) ₂ O		7.10
26	CF ₃	Cl	MeC(=O)(CH ₂) ₃ O		7.23
27	CF ₃	Cl	MeSO ₂ (CH ₂) ₃ O		6.49
28	CF ₃	Cl	NC(CH ₂) ₃ O		7.82
29 ^b	CF ₃	Cl	<i>i</i> -PrO		6.54
30	CF ₃	Cl	MeO(CH ₂) ₂ O		7.49
31	CF ₃	Cl	MeO(CH ₂) ₂ O		7.42
32 ^b	CF ₃	Cl	MeO(CH ₂) ₂ O		7.32
33	CF ₃	Cl	MeO(CH ₂) ₂ O		8.35
34	CF ₃	Cl	MeO(CH ₂) ₂ O		8.96

^aEC50 (the effective concentration for 50% response of a given compound's intrinsic maximum response) on molar basis, taken from reference [18];

^bCompound included in test set; ^cCompound with uncertain activity, not part of data set; *Benzene ring instead of pyridine ring.

The structures of the all the data set compounds of Table 1 were drawn in 2D ChemDraw [20] 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 [21]. 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, modified Burden eigenvalues, 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.

Table 2 Descriptor classes used for the modeling of PPAR γ transactivation activity of pyridyloxybenzene-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

^aReference [21].

Dragon software computed a total number of 496 descriptors, belonging to 0D- to 2D-modules. These descriptors have been utilized to obtain most appropriate models describing the biological activity. Prior to model development procedure, all those descriptors that are inter-correlated beyond 0.90 (descriptor versus descriptor, $r > 0.9$) and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. In doing so, 120 descriptors appeared as significant ones to explain the biological activity of titled compounds.

2.2. Development and validation of model

In the present study, QSAR models have been developed using the combinatorial protocol in multiple linear regression (CP-MLR) [22-26] procedure. It is a "filter"-based variable selection procedure, which employs a combinatorial strategy with MLR to result in selected subset regressions for the pulling out of diverse structure-activity models. Each derived model has 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 [27,28]. The fear of any chance correlations associated with the models recognized in CP-MLR, overcome by randomization test [29,30] in which each cross-validated model has been subjected to 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 unscramble 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 assess 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 [31,32], the Kubinyi function, FIT [33,34] and the Friedman's lack of fit, LOF [35], 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 [36]. 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 $\pm x$ 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, a model equation(s) along the descriptor class provides a prospect to unravel the phenomenon under study i.e. the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modeling study, one third of total active compounds (10) have been included in the test set for the validation of the models derived from remaining 20 training set compounds. A total number of 120 relevant descriptors from 0D- to 2D- classes, which were obtained after the reduction of descriptor data set, have been subjected to CP-MLR analysis with default "filters" set in it. Statistical models in two, three and four descriptors have been explored to achieve the best relationship correlating PPAR γ transactivation activity. The obtained two and three descriptor models are given below.

$$pEC_{50} = 7.153 + 2.563(0.541)Q_{index} - 2.460(0.554)BEHm4$$

$$n = 20, r = 0.769, s = 0.452, F = 12.352, Q^2_{LOO} = 0.490, Q^2_{L50} = 0.225$$

$$r^2_{Test} = 0.297, FIT = 1.027, LOF = 0.271, AIC = 0.276 \quad (1)$$

$$pEC_{50} = 7.420 + 2.279(0.575)Wap - 1.875(0.536)BEHm4$$

$$n = 20, r = 0.712, s = 0.496, F = 8.768, Q^2_{LOO} = 0.303, Q^2_{L50} = 0.185$$

$$r^2_{Test} = 0.141, FIT = 0.730, LOF = 0.327, AIC = 0.333 \quad (2)$$

$$pEC_{50} = 6.582 + 1.811(0.554)MW - 3.195(0.680)T(O..S) + 1.699(0.527)GATS1p$$

$$n = 20, r = 0.823, s = 0.413, F = 11.268, Q^2_{LOO} = 0.540, Q^2_{L50} = 0.573$$

$$r^2_{\text{Test}} = 0.505, \text{FIT} = 1.165, \text{LOF} = 0.279, \text{AIC} = 0.256 \quad (3)$$

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.

The descriptors Qindex, Wap and T(O..S) participated in above models are topological descriptors. Descriptors BEHm4, MW and GATS1p are from the constitutional (CONST), modified Burden eigenvalue (BCUT) and 2D-autocorrelation (2D-AUTO) classes, respectively. Except BEHm4 and T(O..S), all the descriptors have shown positive influence on the activity as evident from the signs of regression coefficients. Thus a higher value of descriptors Qindex (Quadratic index), Wap (all-path Wiener index), MW (molecular weight) and GATS1p (Geary autocorrelation of lag-1/weighted by atomic polarizabilities) in addition to a lower value of the highest eigenvalue n.4 of Burden matrix/weighted by atomic masses (descriptor BEHm4) and sum of topological distances between O and S atoms (descriptor T(O..S)) would be beneficiary to the activity.

Table 3 Identified descriptors^a along with their class, average regression coefficient and incidence^b, in modeling the PPAR γ transactivation activities of pyridyloxybenzene-acylsulfonamides.

Descriptor class, average regression coefficient and (incidence)	
Constitutional descriptors (CONST):	MW, 2.286 (17); nBM, -2.033 (1); nCIC, 0.826 (2); ARR, -1.692 (2); RBN, 1.203 (2); RBF, 0.966 (3); nDB, -1.756 (2); nN, 1.083 (1)
Topological descriptors (TOPO):	AAC, 1.986 (3); Qindex, 2.617 (6); GNar, 1.813 (1); JhetZ, -1.360 (5); MAXDP, 1.319 (1); X1A, -1.363 (7); X2A, -1.112 (1); X1Av, -1.388 (1); S2K, -1.479 (2); Lop, -1.586 (1); IDDE, 1.574 (6); SIC2, -1.944 (1); VEA1, 1.165 (1); T(N..Cl), 0.990 (6); T(O..S), -2.922 (16)
Modified Burden Eigen values (BCUT):	BEHm4, 1.720 (1) and -2.330 (15); BEHm7, 1.462 (4); BELm7, -1.572 (2); BELm8, -1.897 (1); BEHv1, -0.976 (3); BELv4, -2.323 (4); BELv8, 2.764 (1); BELp3, 1.379 (2)
Galvez Topological charge indices (GALVEZ):	GGI4, 1.946 (2); JGI3, 1.927 (2); JGI4, 1.330 (1); JGI5, 0.796 (1); JGT, 1.588 (2)
2D autocorrelations (2D-AUTO):	MATS1v, -0.963 (1); MATS2e, 1.600 (1); MATS3e, 1.048 (1); GATS1v, 1.554 (2); GATS1p, 2.198 (11)
Atom centered fragments (ACF):	H-046, 2.005 (2); H-047, -1.255 (1)

^aThe descriptors are identified from the three parameter models for PPAR γ binding activity transactivation activity emerged from CP-MLR protocol with filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.786 and filter-4 as $0.3 \leq q_2 \leq 1.0$ with a training set of 20 compounds. ^bThe 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. **CONST**: MW, molecular weight; nBM, number of multiple bonds; nCIC, number of rings; ARR, aromatic ratio; RBN, number of rotatable bonds; RBF, rotatable bond fraction; nDB, number of double bonds; nN, number of Nitrogen atoms; **TOPO**: AAC, mean information index on atomic composition; Qindex, Quadratic index; GNar, Narumi geometric topological index; JhetZ, Balaban-type index from Z weighted distance matrix (Barysz matrix); MAXDP, maximal electrotopological positive variation; X1A, average connectivity index chi-1; X2A, average connectivity index chi-2; X1Av, average valence connectivity index chi-1; S2K, 2-path Kier alpha-modified shape index; Lop, Lopping centric index; IDDE, mean information content on the distance degree equality; SIC2, structural information content (neighborhood symmetry of 2-order); VEA1, eigenvector coefficient sum from adjacency matrix; T(N..Cl), sum of topological distances between N..Cl; T(O..S), - sum of topological distances between O..S; **BCUT**: BEHm4, highest eigenvalue n.4 of Burden matrix/weighted by atomic masses; BEHm7, highest eigenvalue n.7 of Burden matrix/weighted by atomic masses; BELm7, lowest eigenvalue n.7 of Burden matrix/weighted by atomic masses; BELm8, lowest eigenvalue n.8 of Burden matrix/weighted by atomic masses; BEHv1, highest eigenvalue n.1 of Burden matrix/weighted by van der Waals volumes; BELv4, lowest eigenvalue n.4 of Burden matrix/weighted by van der Waals volumes; BELv8, lowest eigenvalue n.8 of Burden matrix/weighted by van der Waals volumes; BELp3, lowest eigenvalue n.3 of Burden matrix/weighted by atomic polarizabilities; **GALVEZ**: GGI4, topological charge index of order 4; JGI3, mean topological charge index of order 3; JGI4, mean topological charge index of order 4; JGI5, mean topological charge index of order 5; JGT, global topological charge index; **2D-AUTO**: MATS1v, Moran autocorrelation of lag-1/ weighted by atomic van der Waals volumes; MATS2e, Moran autocorrelation of lag-2/ weighted by atomic Sanderson electronegativities; MATS3e, Moran autocorrelation of lag-3/ weighted by atomic Sanderson electronegativities; GATS1v, Geary autocorrelation of lag-1/ weighted by atomic van der Waals volumes; GATS1p, Geary autocorrelation of lag-1/weighted by atomic polarizabilities; **ACF**: H-046, H attached to C0(sp3) no X attached to next C atom; H-047, H attached to C1(sp3) / C0(sp2).

The three descriptor model could estimate nearly 68% in observed activity of the compounds. Considering the number of observation in the dataset, models with up to four descriptors were explored. It has resulted in 37 models with test set $r^2 > 0.50$. These models (with 120 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 ($=0.786$) has been used as the new threshold of filter-3 for the next generation. These models have shared 43 descriptors among them. All these shared 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.

Following are the selected four-descriptor models for the PPAR γ transactivation activities of pyridyloxybenzene-acylsulfonamides emerged through CP-MLR.

$$\begin{aligned} pEC_{50} &= 6.025 + 2.640(0.407)Q_{index} + 1.131(0.345)T(N..Cl) - 2.345(0.437)BEHm4 \\ &\quad + 1.131(0.392)GATS1v \\ n &= 20, r = 0.899, s = 0.329, F = 15.918, Q^2_{L00} = 0.716, Q^2_{L50} = 0.661 \\ r^2_{Test} &= 0.602, FIT = 1.768, LOF = 0.225, AIC = 0.180 \end{aligned} \quad (4)$$

$$\begin{aligned} pEC_{50} &= 6.554 + 0.828(0.308)RBN + 2.519(0.427)Q_{index} + 0.840(0.357)T(N..Cl) \\ &\quad - 2.369(0.447)BEHm4 \\ n &= 20, r = 0.894, s = 0.337, F = 14.988, Q^2_{L00} = 0.691, Q^2_{L50} = 0.636 \\ r^2_{Test} &= 0.548, FIT = 1.665, LOF = 0.236, AIC = 0.189 \end{aligned} \quad (5)$$

$$\begin{aligned} pEC_{50} &= 6.422 + 0.645(0.257)RBF + 2.773(0.422)Q_{index} + 1.013(0.358)T(N..Cl) \\ &\quad - 2.246(0.477)BEHm4 \\ n &= 20, r = 0.889, s = 0.344, F = 14.205, Q^2_{L00} = 0.665, Q^2_{L50} = 0.655 \\ r^2_{Test} &= 0.512, FIT = 1.578, LOF = 0.247, AIC = 0.197 \end{aligned} \quad (6)$$

$$\begin{aligned} pEC_{50} &= 7.084 + 1.332(0.501)MW - 0.924(0.339)X1A - 3.321(0.577)T(O..S) \\ &\quad + 2.193(0.481)GATS1p \\ n &= 20, r = 0.885, s = 0.349, F = 13.683, Q^2_{L00} = 0.648, Q^2_{L50} = 0.628 \\ r^2_{Test} &= 0.573, FIT = 1.520, LOF = 0.254, AIC = 0.203 \end{aligned} \quad (7)$$

These models have accounted for nearly 81% 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. (4) to (7) and predicted from LOO procedure have been included in Table 4.

The models (4) to (7) are validated with an external test set of 10 compounds mentioned 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 2.

The newly appeared descriptors in above models are, T(N..Cl) and X1A (topological descriptors), RBN and RBF (constitutional descriptors), and GATS1v (a 2D-AUTO class descriptor). Descriptors T(N..Cl), RBN, RBF and GATS1v have correlated positively to the PPAR γ transactivation whereas descriptor X1A influenced it negatively. Thus from the signs of regression coefficients of these descriptors it is evident that higher values of the sum of topological distances between N and Cl atoms (descriptor T(N..Cl)), presence of more number of rotatable bonds (descriptor RBN), higher value of rotatable bond fraction (descriptor RBF) in a molecular structure and a higher value of Geary autocorrelation of lag-1/weighted by atomic polarizabilities (GATS1p) would be beneficial to the activity, whereas a lower value of descriptor X1A (average connectivity index chi-1) would be advantageous to the activity.

Table 4 Observed and modeled PPAR γ transactivation activity of pyridyloxybenzene-acylsulfonamides.

S. No.	pEC50(M) ^a								
	Obsd ^b	Eq. (4)		Eq. (5)		Eq. (6)		Eq. (7)	
		Calc	Pred ^c						
1	7.21	7.23	7.24	7.29	7.33	7.43	7.48	7.29	7.35
2	8.59	8.24	8.18	8.32	8.28	8.39	8.36	8.14	8.06
3 ^d	8.82	8.08	- ^d	8.12	- ^d	8.14	- ^d	7.96	- ^d
4	8.16	8.43	8.52	8.56	8.79	8.45	8.57	8.55	8.72
5	8.47	8.62	8.70	8.52	8.55	8.49	8.50	8.53	8.56
6	7.96	8.21	8.25	8.12	8.14	8.24	8.29	8.16	8.20
7 ^d	7.49	7.58	- ^d	7.85	- ^d	8.01	- ^d	7.40	- ^d
8 ^d	8.07	7.65	- ^d	7.84	- ^d	8.24	- ^d	7.19	- ^d
9 ^d	7.82	7.61	- ^d	7.51	- ^d	7.49	- ^d	7.67	- ^d
10	8.64	8.20	8.16	8.28	8.23	8.24	8.19	8.19	8.15
11 ^d	8.51	8.52	- ^d	8.62	- ^d	8.56	- ^d	8.58	- ^d
12	8.26	8.50	8.60	8.46	8.55	8.53	8.64	8.26	8.26
13 ^d	8.28	7.67	- ^d	7.61	- ^d	7.53	- ^d	7.91	- ^d
14 ^d	8.17	8.33	- ^d	8.35	- ^d	8.30	- ^d	8.42	- ^d
15	8.00	8.21	8.24	8.12	8.14	8.24	8.28	8.19	8.24
16	7.03	7.36	7.43	7.40	7.47	7.26	7.33	7.39	7.42
17 ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e
18 ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e
19	6.72	6.81	6.91	6.76	6.82	6.77	6.85	6.39	6.14
20 ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e
21 ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e	- ^e
22	7.59	7.73	7.75	7.68	7.70	7.70	7.73	7.85	7.89
23	8.21	7.69	7.55	7.67	7.52	7.65	7.48	8.06	8.02
24	8.13	7.63	7.59	7.58	7.53	7.45	7.37	7.77	7.73
25 ^d	7.10	7.31	- ^d	7.42	- ^d	7.41	- ^d	7.06	- ^d
26	7.23	7.54	7.58	7.68	7.71	7.60	7.63	7.60	7.67
27	6.49	6.49	6.49	6.50	6.51	6.57	6.88	6.55	6.84
28	7.82	7.77	7.74	7.85	7.86	7.92	7.96	7.53	7.42
29 ^d	6.54	7.20	- ^d	7.28	- ^d	7.26	- ^d	6.89	- ^d
30	7.49	7.86	7.92	7.81	7.85	7.70	7.73	7.94	8.03
31	7.42	7.22	7.13	7.18	7.07	7.20	7.09	7.65	7.75
32 ^d	7.32	7.63	- ^d	7.58	- ^d	7.59	- ^d	7.37	- ^d
33	8.35	8.37	8.38	8.39	8.39	8.31	8.30	8.20	8.16
34	8.96	8.62	8.49	8.56	8.43	8.57	8.43	8.49	8.37

^aOn molar basis; ^bTaken from ref. [18]; ^cLeave-one-out (LOO) procedure; ^dCompound included in test set; ^eCompound with uncertain activity, not part of data set.

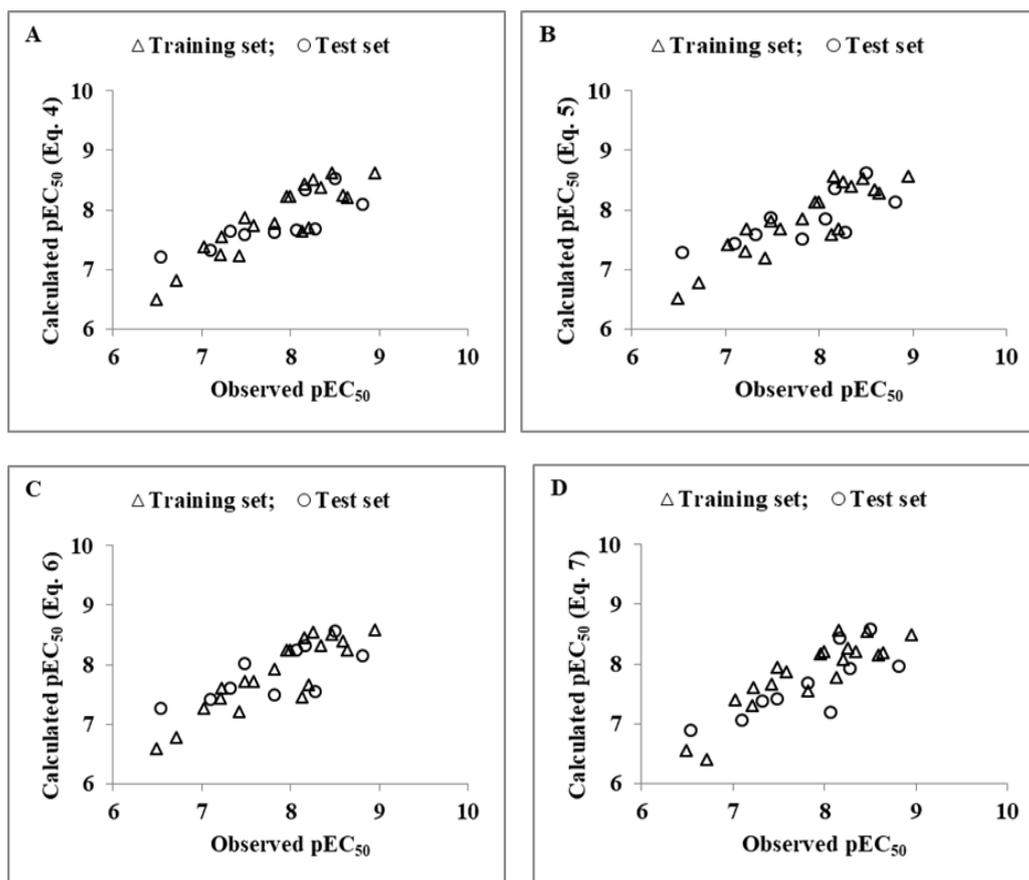


Figure 2 Plot of observed and calculated pEC₅₀ values (A, B, C and D by equations (4), (5), (6) and (7), respectively) of training- and test-set compounds for PPAR γ transactivation.

3.2. Applicability domain (AD)

On analyzing the model AD in the Williams plot, shown in Figure 3, of the model based on the whole dataset (Table 5), it has appeared that none of the compounds were identified as an obvious outlier for the PPAR γ transactivation activities if the limit of normal values for the Y outliers (response outliers) was set as 3 (standard deviation) units. Two compounds listed in Table 1 at S. No. **8** and **27** 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 5 Models derived for the whole data set (n = 30) for the PPAR γ transactivation activity in descriptors identified through CP-MLR.

Model	r	s	F	Q2L00	Eq.
pEC ₅₀ = 5.889 + 2.887(0.392)Qindex + 1.221(0.303)T(N..Cl) - 2.381(0.461)BEHm4 + 1.164(0.317)GATS1v	0.865	0.359	18.664	0.668	(4a)
pEC ₅₀ = 6.468 + 0.850(0.281)RBN + 2.669(0.429)Qindex + 0.890(0.319)T(N..Cl) - 2.416(0.494)BEHm4	0.846	0.382	15.847	0.637	(5a)
pEC ₅₀ = 6.399 + 0.663(0.248)RBF + 2.872(0.430)Qindex + 0.919(0.328)T(N..Cl) - 2.353(0.526)BEHm4	0.836	0.393	14.569	0.621	(6a)
pEC ₅₀ = 7.116 + 1.730(0.450)MW - 0.903(0.321)X1A	0.860	0.366	17.816	0.582	(7a)

$$- 3.351(0.577)T(O.S) + 1.853(0.359)GATS1p$$

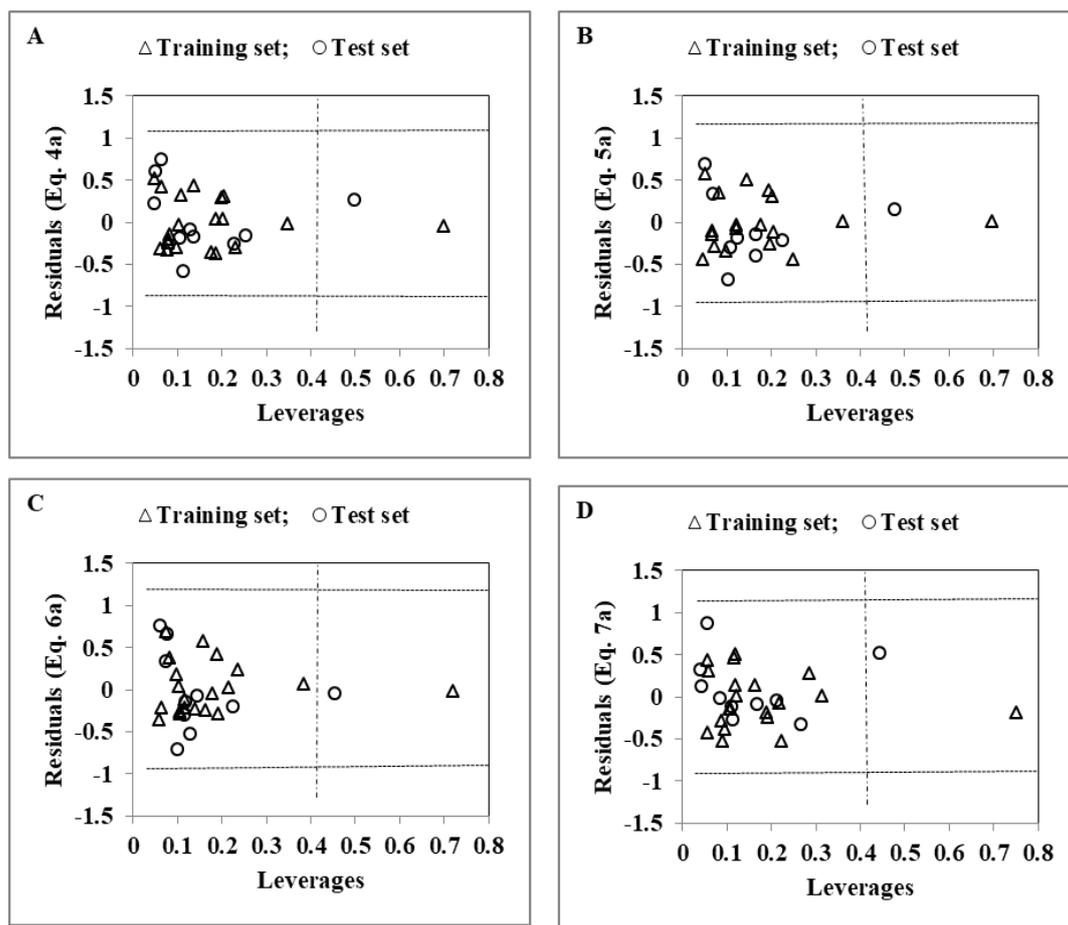


Figure 3 Williams plot for the training-set and test- set compounds for PPAR γ transactivation activity (A, B, C and D based on the whole data set equations (4a), (5a), (6a) and (7a), respectively). The horizontal dotted line refers to the residual limit ($\pm 3 \times$ standard deviation) and the vertical dotted line represents threshold leverage h^* ($= 0.40$).

4. Conclusion

QSAR rationales have been obtained for the PPAR γ transactivation activity of pyridyloxybenzene-acylsulfonamides in terms of 0D- to 2D-Dragon descriptors. The descriptors identified in CP-MLR analysis have highlighted the role of atomic mass, van der Waals volumes and polarizability through weighted 2D autocorrelations (GATS1v and GATS1p), modified Burden eigenvalue (BEHm4) and molecular weight (MW). Sum of topological distances between O and S (descriptor T(O..S)), and N and Cl (descriptor T(N..Cl)), average connectivity index chi-1(X1A) and Quadratic index (Qindex) have also shown dominance to optimize the PPAR γ transactivation. Descriptors RBN and RBF suggested presence of rotatable bonds in a molecular structure for better PPAR γ activity. 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

Acknowledgments

Authors are thankful to their institution for providing necessary facilities to complete this study.

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 (Suppl.1), S5–S10.
- [2] Wild S, Roglic G, Green A, Sicree R and King H. (2004). Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care*, 27, 1047-1053.
- [3] 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.
- [4] Evans RM, Barish GD and Wang Y-X. (2004). PPARs and the complex journey to obesity. *Nature Medicine*, 10, 355-361.
- [5] 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.
- [6] Berger J and Moller DE. (2002). The mechanisms of action of PPARs. *Annual Review of Medicine*, 53, 409-435.
- [7] Henke BR. (2004). Peroxisome proliferator-activated receptor gamma (PPARgamma) ligands and their therapeutic utility. *Progress in Medicinal Chemistry*, 42, 1-53.
- [8] 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.
- [9] Berger JP, Akiyama TE and Meinke PT. (2005). PPARs: therapeutic targets for metabolic disease. *Trends in Pharmacological Sciences*, 26, 244-251.
- [10] Loviscach M, Rehman N, Carter L, Mudaliar S, Mohadeen P, Ciaraldi TP, Veerkamp JH and Henry RR. (2000). Distribution of peroxisome proliferator-activated receptors (PPARs) in human skeletal muscle and adipose tissue: relation to insulin action. *Diabetologia*, 43, 304-311.
- [11] Minoura H, Takeshita S, Ita M, Hirosumi J, Mabuchi M, Kawamura I, Nakajima S, Nakayama O, Kayakiri H, Oku T, Okubo-Suzuki A, Fukagawa M, Kojo H, Hanioka K, Yamasaki N, Imoto T, Kobayashi Y and Mutoh S. (2004). Pharmacological characteristics of a novel non-thiazolidinedione insulin sensitizer, FK614. *European Journal of Pharmacology*, 494, 273-281.
- [12] Minoura H, Takeshita S, Yamamoto T, Mabuchi M, Hirosumi J, Takakura S, Kawamura I, Seki J, Manda T, Ita, M and Mutoh S. (2005). Ameliorating effect of FK614, a novel non-thiazolidinedione peroxisome proliferator-activated receptor gamma agonist, on insulin resistance in Zucker fatty rat. *European Journal of Pharmacology*, 519, 182-190.
- [13] Kyowa Hakko Kirin Company Limited. Yanagisawa A, Otsubo W, Ueno K, Suzuki M, Matsubara M, Saeki S, Yao K, Hamaguchi A and Aratake S. (2008). Biphenyl derivative.
- [14] Kowa Company Limited. Miura T, Onogi K, Gomi N, Araki T, Tagashira J, Sekimoto R, Ishida R, Aoki H and Ohgiya T. (2011). Novel sulfonamide derivative and pharmaceutical product containing same. *PCT Int. Appl. WO2011/024468*.
- [15] Hopkins CR, O'Neil SV, Laufersweiler MC, Wang Y, Pokross M, Mekel M, Evdokimov A, Walter R, Kontoyianni M, Petrey ME, Sabatakos G, Roesgen JT, Richardson E and Demuth TP. (2006). Design and synthesis of novel N-sulfonyl-2-indole carboxamides as potent PPAR-gamma binding agents with potential application to the treatment of osteoporosis. *Bioorganic and Medicinal Chemistry Letters*, 16, 5659-5663.
- [16] Etgen GJ and Mantlo N. (2003). PPAR ligands for metabolic disorders. *Current Topics in Medicinal Chemistry*, 3, 1649.
- [17] 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) γ agonists: Design and synthesis of benzylpyrazole-acylsulfonamides. *Bioorganic and Medicinal Chemistry*, 20, 714-733.
- [18] Rikimaru K, Wakabayashi T, Abe H, Tawaraishi T, Imoto H, Yonemori J, Hirose H, 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). Structure-activity relationships and key structural feature of pyridyloxybenzene-acylsulfonamides as new, potent, and selective peroxisome proliferator-activated receptor (PPAR) γ agonists. *Bioorganic and Medicinal Chemistry*, 20, 3332-3358.

- [19] 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 insulin sensitizers with antidiabetic activities. *Journal of Medicinal Chemistry*, 45, 1518-1534.
- [20] Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. <http://www.cambridgesoft.com>
- [21] Dragon software (version 1.11-2001) by Todeschini R; Consonni V. Milano, Italy. <http://www.taletе.mi.it/dragon.htm>
- [22] 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.
- [23] 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.
- [24] 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.
- [25] Sharma BK, Paliania P, Singh P and Prabhakar YS. (2010). Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. *SAR and QSAR in Environmental Research*, 21, 169-185.
- [26] 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.
- [27] Topliss JG and Edwards RP. (1979). Chance factors in studies of quantitative structure-activity relationships. *Journal of Medicinal Chemistry*, 22, 1238-1244.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] Akaike H. (1973). Information theory and an extension of the minimum likelihood principle. In: Petrov BN and Csaki F. (Eds.). *Second international symposium on information theory*. Budapest: Akademiai Kiado, 267-281.
- [32] Akaike H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- [33] Kubinyi H. (1994). Variable selection in QSAR studies. I. An evolutionary algorithm. *Quantitative Structure-Activity Relationship*, 13, 285-294.
- [34] Kubinyi H. (1994). Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. *Quantitative Structure-Activity Relationship*, 13, 393-401.
- [35] Friedman J. (1990). In Technical Report No. 102. Laboratory for Computational Statistics, Stanford University: Stanford.
- [36] Gramatica P. (2007). Principles of QSAR models validation: internal and external. *QSAR and Combinatorial Science*, 26, 694-701.

How to cite this article

Parihar R, Jahan A and Sharma BK. (2020). CP-MLR derived QSAR rationales for the PPAR_γ agonistic activity of the pyridyloxybenzene-acylsulfonamide derivatives. *GSC Biological and Pharmaceutical Sciences*, 12(1), 273-285.
