Research Article
AMERICAN JOURNAL OF PHARMACY AND HEALTH RESEARCH
www.ajphr.com
2020, Volume 8, Issue 08
ISSN: 2321–3647(online)

QSAR study on the hGPR119 Agonistic Activity of Triazolopyridines

Brij Kishore Sharma*, Raghuraj Parihar
Department of Chemistry, Government College, Bundi-323 001 (Rajasthan), India

ABSTRACT

The hGPR119 agonistic activity of triazolopyridines has been analysed with topological and molecular features with DRAGON software. Analysis of the structural features in conjunction with the biological endpoints in combinatorial protocol in multiple linear regression (CP-MLR) led to the identification of 10 descriptors for modelling the activity. The study clearly suggested the role of path/walk 5-Randic shape index (PW5), mean information vertex degree equality (IVDE), Lovasz-Pelikan index (LP1), atomic properties (mass, van der Waals volume and Sanderson electronegativities) in terms of weighted 2D-autocorrelations (MATS4m, MATS2e, MATS4e and MATS5e) and modified Burden eigenvalues (BELm7 and BEHv8) and total primary sp3 hybridized carbon atoms (nCp) in a molecular structure to optimize the hGPR119 agonistic activities of titled compounds. 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; hGPR119 agonistic activity; Combinatorial protocol in multiple linear regression (CP-MLR) analysis; Dragon descriptors; Triazolopyridine derivatives.

*Corresponding Author Email: bksharma_sikar@rediffmail.com
Received 10 July 2020, Accepted 04 August 2020

Please cite this article as: Sharma BK *et al*., QSAR study on the hGPR119 Agonistic Activity of Triazolopyridines. American Journal of Pharmacy & Health Research 2020.

INTRODUCTION

More than 400 million people worldwide are adversely affected with diabetes and it is supposed that this total will reach to 642 million by 2040 [1]. Type 2 diabetes mellitus (T2DM) is the most common form of diabetes which is a metabolic disorder and is characterized by hyperglycemia. Impaired insulin secretion and insulin resistance causes hyperglycemia which in long-term increases risk of micro- and macro-vascular complications that may cause blindness, renal failure, diabetic foot disorders, heart attacks and strokes. Multiple oral antidiabetic agents like sulfonylureas, meglitinides, biguanides, thiazolidinediones, β-glucosidase inhibitors and dipeptidyl-peptidase-4 (DPP-4) inhibitors have been used to cure T2DM but many patients failed to achieve glycemic control at desired level [2,3]. The glucose-lowering effect of sodium-dependent glucose co-transporter 2 (SGLT2) inhibitor is devoid of hypoglycemia or weight gain. The need to develop new antidiabetic agents with greater safety and efficacy still exist.

GPR119, a G-protein coupled receptor (GPCR), is expressed predominantly in the pancreatic β-cells and gastrointestinal L-cells. The identified endogenous agonists for the GPR119 receptor are oleoyl-lysophosphatidylcholine and oleoylethanolamide (OEA) [4,5]. Glucose-dependent insulin secretion from pancreatic β-cells increases due to increased cellular cAMP levels on activation of the GPR119 receptor [6]. Release of incretins like glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP), from enteroendocrine cells are the results of the activation of the GPR119 receptor in the gut [7]. The stimulation of insulin secretion from □-cells in a glucose-dependent manner by GLP-1 and GIP protects β-cells against apoptosis [8,9]. The glucose-dependent dual mechanism of action of GPR119 agonists may improve glycemic control without inducing hypoglycemia.

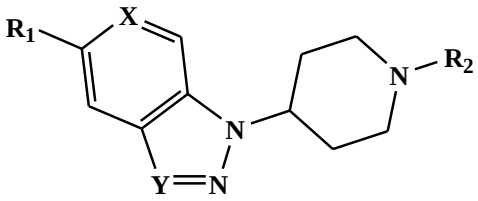
The investigations of several research groups [10,11] on multiple small-molecule GPR119 agonists led to the development of clinical compounds which include APD668 [12], GSK1292263 [13] and MBX-2982 [14]. Poor aqueous solubility of agonists causes low bioavailability, produces erratic assay results in *in vitro* studies and carries a high risk of not advancing due to potential toxicity which may not be recognized during preclinical studies [15,16]. As an attempt to improve aqueous solubility of GPR119 agonist a novel series of triazolopyridine derivatives have been reported by Matsuda et al. [17]. These derivatives are based on 3*H*-[1,2,3]triazolo[4,5-*c*]pyridine scaffold and having variations at central spacer, left-hand aryl group and right-hand piperidine N-capping group. The present communication is aimed at to establish the quantitative relationships between the reported activities and descriptors unfolding the substitutional changes in titled molecules.

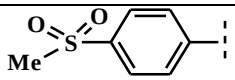
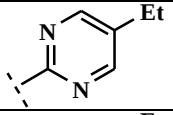
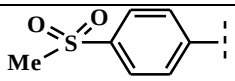
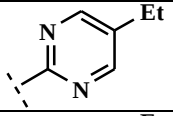
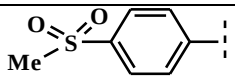
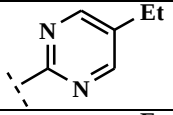
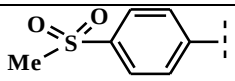
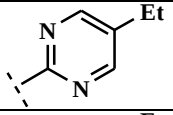
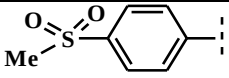
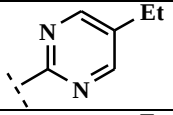
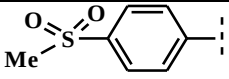
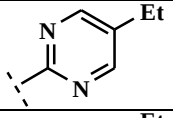
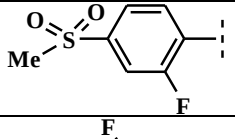
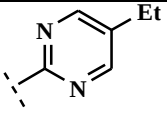
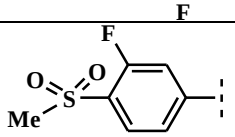
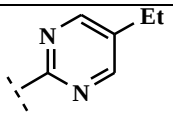
MATERIALS AND METHOD

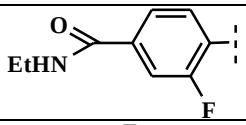
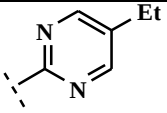
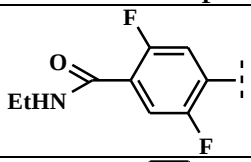
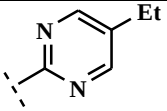
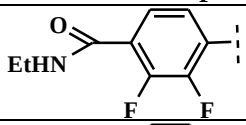
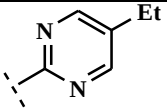
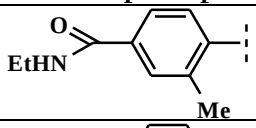
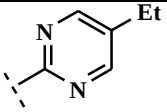
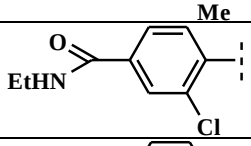
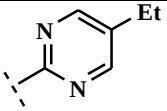
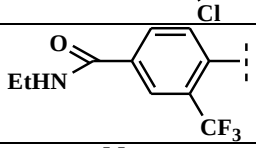
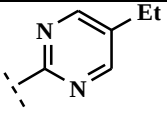
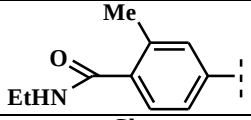
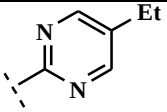
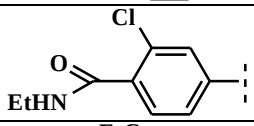
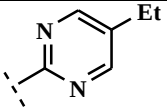
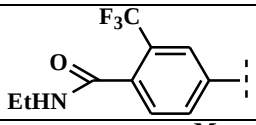
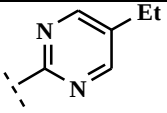
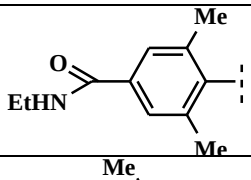
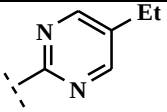
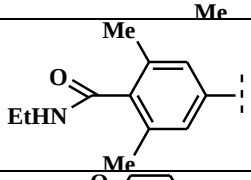
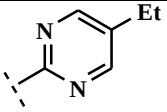
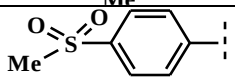
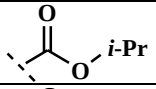
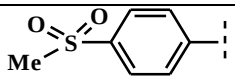
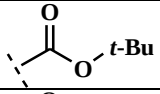
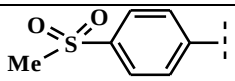
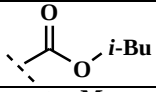
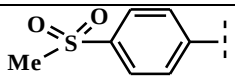
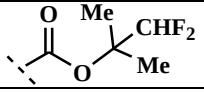
Biological actions and theoretical molecular descriptors

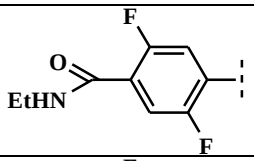
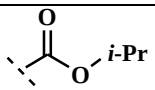
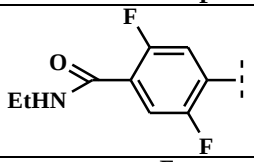
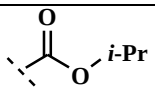
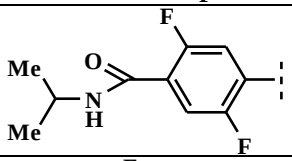
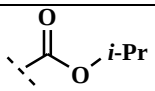
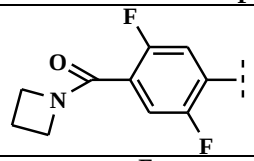
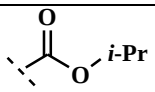
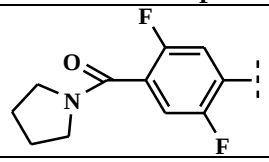
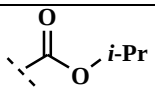
The reported twenty eight derivatives of triazolopyridine the data set for present study [17]. These derivatives were evaluated for their agonistic activity against human GPR119 overexpressed in Flp-In-T-Rex-HEK293 cells by measuring changes in the cellular cAMP levels and were reported as EC_{50}. The reported activity on molar basis (as pEC_{50}) along with the structural variations of these analogues is shown in Table 1. The data set was sub-divided into training set to develop models and test set to validate the models externally. The test set compounds which were selected using an in-house written randomization program, are also mentioned in Table 1.

Table 1: Structural variations and reported hGPR119 agonistic activities of triazolopyridine derivatives.



Cpd.	R_1	X	Y	R_2	$pEC_{50}(M)^a$
1		N	CH		7.89
2		CH	CH		7.85
3[b]		CMe	CH		7.15
4*		CH	CH		5.47
5[b]		CH	N		7.51
6		N	N		7.68
7[b]		N	N		8.15
8		N	N		8.05

9[b]		N	N		7.82
10		N	N		8.70
11		N	N		8.00
12		N	N		7.72
13		N	N		7.89
14		N	N		7.22
15		N	N		7.74
16		N	N		7.74
17		N	N		7.41
18		N	N		7.08
19		N	N		7.57
20[b]		N	CH		7.38
21[b]		N	CH		7.70
22[b]		N	CH		7.96
23		N	CH		7.64

24		N	CH		8.00
25		N	N		7.43
26		N	N		7.48
27		N	N		7.21
28		N	N		7.19

[a]EC_{50} (the concentration of the test compound required to achieve 50% of the maximal response) on molar basis, taken from reference [17]; [b]Compound included in test set; *4-Methyl substituted indazole.

The structures of the all the data set compounds of Table 1, drawn in 2D ChemDraw [18], were 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. Descriptors, belonging to 0D-, 1D- and 2D-classes, of titled compounds were computed using DRAGON software [19]. This software offers a large number of descriptors corresponding to ten different classes of 0D- to 2D-descriptor modules which 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. Characteristic structural information specific to the descriptor class is offered by these descriptors. The definition and scope of these descriptor's classes is given in Table 2.

Table 2: Descriptor classes used for modeling the hGPR119 agonistic activity of triazolopyridines.

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].

A total number of 492 descriptors, belonging to 0D- to 2D- modules, computed by Dragon software have been utilized to obtain most appropriate models describing the biological activity. The descriptors pool has been reduced by eliminating those descriptors which 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, $r < 0.1$), prior to model development procedure. In this way, 99 descriptors appeared as significant ones to explain the biological actions of titled compounds.

Development and validation of model

QSAR models have been developed, in the present study, using a "filter"-based variable selection procedure namely the combinatorial protocol in multiple linear regression (CP-MLR) [20-24]. This procedure employs a combinatorial strategy with MLR to result in selected subset regressions to pull out the diverse structure–activity models and 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. The fear of existence of "chance correlations" in using large descriptor pools for multilinear QSAR/QSPR studies [25,26] overcome by randomization test [27,28] in which each cross-validated CP-MLR recognized 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 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. 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, for each model derived in n data points. 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 [29,30], the Kubinyi function, FIT [31,32] and the Friedman's lack of fit, LOF [33], 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.

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 [34]. 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.

RESULTS AND DISCUSSION

QSAR results

A derived model equation(s), using a pool of descriptors of different descriptor classes, provides an opportunity 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, 07 (one fourth of total active) compounds have been included in the test set for the validation of the models derived from remaining 21 training set compounds. A total number of 99 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 three descriptors have been explored to achieve the best relationship correlating hGPR119 agonistic activity. All the models obtained in three descriptors were having the r^2_{Test} value less than 0.5. Considering the number of observation in the dataset, models with up to four descriptors were explored. It has resulted in 04 models with test set $r^2 > 0.50$. These models (with 99 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 (=0.814, r-bar value of the three parameter model having highest r^2_{Test}) has been used as the new threshold of filter-3 for the next generation. These models have shared 10 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.

Table 3. Identified descriptors[a] along with their class, average regression coefficient and incidence[b], in modeling the hGPR119 agonistic activities of triazolopyridines.

Descriptor class, average regression coefficient and (incidence)	
Topological descriptors (**TOPO**):	PW5, 2.374(3); IVDE, 1.128(2); LP1, -2.367(4)
Modified Burden Eigen values (**BCUT**):	BELm7, 0.980(1); BEHv8, -0.539(1)
2D autocorrelations (**2D-AUTO**):	MATS4m, 1.430(1); MATS2e, -0.722(1); MATS4e, 0.983 (1); MATS5e, 1.223(1)
Functional group counts (**FUNC**):	nCp, -0.412(1)

[a]The descriptors are identified from the four parameter models for PPARγ binding activity transactivation activity emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.814 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 20 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:** PW5, path/walk 5-Randic shape index; IVDE, mean information vertex degree equality; LP1; Lovasz-Pelikan index (leading eigenvalue); **BCUT:** BELm7, lowest eigenvalue n.7 of Burden matrix/weighted by atomic masses; BELm8, lowest eigenvalue n.8 of Burden matrix/weighted by atomic masses; BEHv8, highest eigenvalue n.8 of Burden matrix/weighted by van der Waals volumes; **2D-AUTO:** MATS4m, Moran autocorrelation of lag-4/ weighted by atomic masses; MATS2e, Moran autocorrelation of lag-2/ weighted by atomic Sanderson electronegativities; MATS4e, Moran autocorrelation of lag-4/ weighted by atomic Sanderson electronegativities; MATS5e, Moran autocorrelation of lag-5/ weighted by atomic Sanderson electronegativities; **FUNC:** nCp, number of total primary C(sp3).

The models in four descriptors emerged through CP-MLR are mentioned below.

$pEC_{50} = 6.436 + 2.274(0.444)PW5 - 1.967(0.305)LP1 + 1.430(0.464)MATS4m - 0.721(0.266)MATS2e$

$n = 21, r = 0.899, s = 0.297, F = 16.833, Q^2_{LOO} = 0.513, Q^2_{L5O} = 0.591$

$r^2_{Test} = 0.532, FIT = 1.819, LOF = 0.175, AIC = 0.143$ (1)

$pEC_{50} = 6.587 + 2.349(0.536)PW5 + 0.916(0.454)IVDE - 2.424(0.424)LP1 - 0.539(0.238)BEHv8$

$n = 21, r = 0.885, s = 0.315, F = 14.465, Q^2_{LOO} = 0.545, Q^2_{L5O} = 0.557$

$r^2_{Test} = 0.661, FIT = 1.563, LOF = 0.198, AIC = 0.161$ (2)

$pEC_{50} = 6.865 - 2.308(0.323)LP1 + 0.980(0.365)BELm7 + 0.983(0.248)MATS4e$

+ 1.223(0.390) MATS5e

n = 21, r = 0.883, s = 0.317, F = 14.207, Q^2_{LOO} = 0.534, Q^2_{L5O} = 0.528

r^2_{Test} = 0.579, FIT = 1.535, LOF = 0.200, AIC = 0.164 (3)

pEC_{50} = 6.474 + 2.500(0.556)PW5 + 1.340(0.474)IVDE – 2.770(0.409)LP1

– 0.412(0.203)nCp

n = 21, r = 0.879, s = 0.323, F = 13.596, Q^2_{LOO} = 0.562, Q^2_{L5O} = 0.598

r^2_{Test} = 0.522, FIT = 1.469, LOF = 0.207, AIC = 0.169 (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.

The participated descriptors, PW5, IVDE and LP1, in above models belong to topological class. It is apparent from the above mentioned equations that a higher value of path/walk 5-Randic shape index (PW5), and mean information vertex degree equality (IVDE) and a lower value of Lovasz-Pelikan index (LP1) would be helpful to elevate the agonistic activity. Modified Burden eigenvalue (BCUT) class descriptors BELm7 (lowest eigenvalue n.7 of Burden matrix/weighted by atomic masses) and BEHv8 (highest eigenvalue n.8 of Burden matrix/weighted by van der Waals volumes) have shown positive and negative contribution, respectively, to the activity suggesting a higher value of BELm7 and a lower value of BEHv8 beneficiary to the activity. Except MATS2e, all the participated 2D-autocoorelation descriptors namely MATS4m, MATS4e and MATS5e contributed positively to the activity. Thus it may be inferred that a lower value of MATS2e (Moran autocorrelation of lag-2/weighted by atomic Sanderson electronegativities) and higher values of MATS4m (Moran autocorrelation of lag-4/weighted by atomic masses), MATS4e (Moran autocorrelation of lag-4/weighted by atomic Sanderson electronegativities) and MATS5e (Moran autocorrelation of lag-5/weighted by atomic Sanderson electronegativities) would be helpful for better activity. Additionally, presence of higher number of total sp3 hybridized carbon atoms in a molecular structure (nCp, functional group class descriptor) would be detrimental to the activity.

Nearly 81% variance in the observed activity has been accounted by these models. None of the CP-MLR identified model has shown any chance correlation in the randomization study (100 simulations per model). The values of Q^2 index, greater than a specified cutoff (0.5), hint that derived models are reasonable robust QSAR models. The pEC_{50} values of training set compounds calculated using Eqs. (1) to (4) and predicted from LOO procedure have been included in Table 4.

Table 4: Observed and modeled hGPR119 activity of triazolopyridines.

S. No.	pEC50(M)[a]								
	Obsd[b]	Eq. (1)		Eq. (2)		Eq. (3)		Eq. (4)	
		Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]	Calc	Pred[c]
1	7.89	7.95	7.96	7.86	7.86	7.89	7.89	7.90	7.90
2	7.85	7.82	7.82	7.86	7.87	7.92	7.93	7.90	7.91
3[d]	7.15	6.72	-[d]	7.02	-[d]	6.79	-[d]	6.81	-[d]
4	5.47	5.67	6.18	5.71	6.22	5.91	6.44	5.70	6.19
5[d]	7.51	7.82	-[d]	7.86	-[d]	7.73	-[d]	7.90	-[d]
6	7.68	7.96	8.00	7.86	7.90	7.80	7.82	7.90	7.94
7[d]	8.15	8.16	-[d]	8.02	-[d]	8.10	-[d]	8.14	-[d]
8	8.05	7.80	7.77	7.80	7.76	8.30	8.38	7.92	7.90
9[d]	7.82	8.04	-[d]	7.88	-[d]	7.91	-[d]	7.81	-[d]
10	8.7	8.23	8.03	8.09	7.91	8.33	8.04	8.07	7.89
11	8.00	7.68	7.61	7.49	7.40	7.80	7.76	7.42	7.35
12	7.72	7.64	7.62	7.63	7.61	7.51	7.48	7.60	7.57
13	7.89	7.97	7.98	7.88	7.88	7.55	7.51	7.81	7.79
14	7.22	6.90	6.59	7.00	6.95	6.59	6.39	7.19	7.19
15	7.74	7.83	7.85	7.87	7.90	7.52	7.46	7.83	7.85
16	7.74	8.11	8.16	8.07	8.15	7.95	7.98	8.03	8.09
17	7.41	7.56	7.72	7.85	7.97	7.73	7.90	8.09	8.35
18	7.08	7.09	7.10	7.02	6.99	7.18	7.22	7.07	7.07
19	7.57	7.37	7.31	7.40	7.28	7.53	7.51	7.53	7.51
20[d]	7.38	7.19	-[d]	7.70	-[d]	7.39	-[d]	7.69	-[d]
21[d]	7.70	7.68	-[d]	7.73	-[d]	7.55	-[d]	7.62	-[d]
22[d]	7.96	7.93	-[d]	7.97	-[d]	7.62	-[d]	7.97	-[d]
23	7.64	7.26	6.69	7.43	6.98	7.85	7.96	7.52	7.30
24	8.00	7.69	7.66	7.97	7.97	7.70	7.65	7.85	7.82
25	7.43	7.71	7.74	7.97	8.12	7.45	7.46	7.85	7.96
26	7.48	7.80	7.88	7.68	7.72	7.57	7.61	7.36	7.31
27	7.21	7.36	7.37	7.26	7.27	7.19	7.19	7.22	7.22
28	7.19	7.57	7.65	7.24	7.25	7.70	7.77	7.20	7.20

[a]On molar basis; [b]Taken from ref. [17]; [c]Leave-one-out (LOO) procedure; [d]Compound included in test set.

The models (1) to (4) are validated with an external test set of 7 compounds mentioned in Table 1. The test set r^2 (r^2_{Test}) values greater than 0.5 of these models reflect that these models have satisfactory external validation capability. The predicted activity values of test set compounds are in tune to the observed ones and the same is mentioned 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.

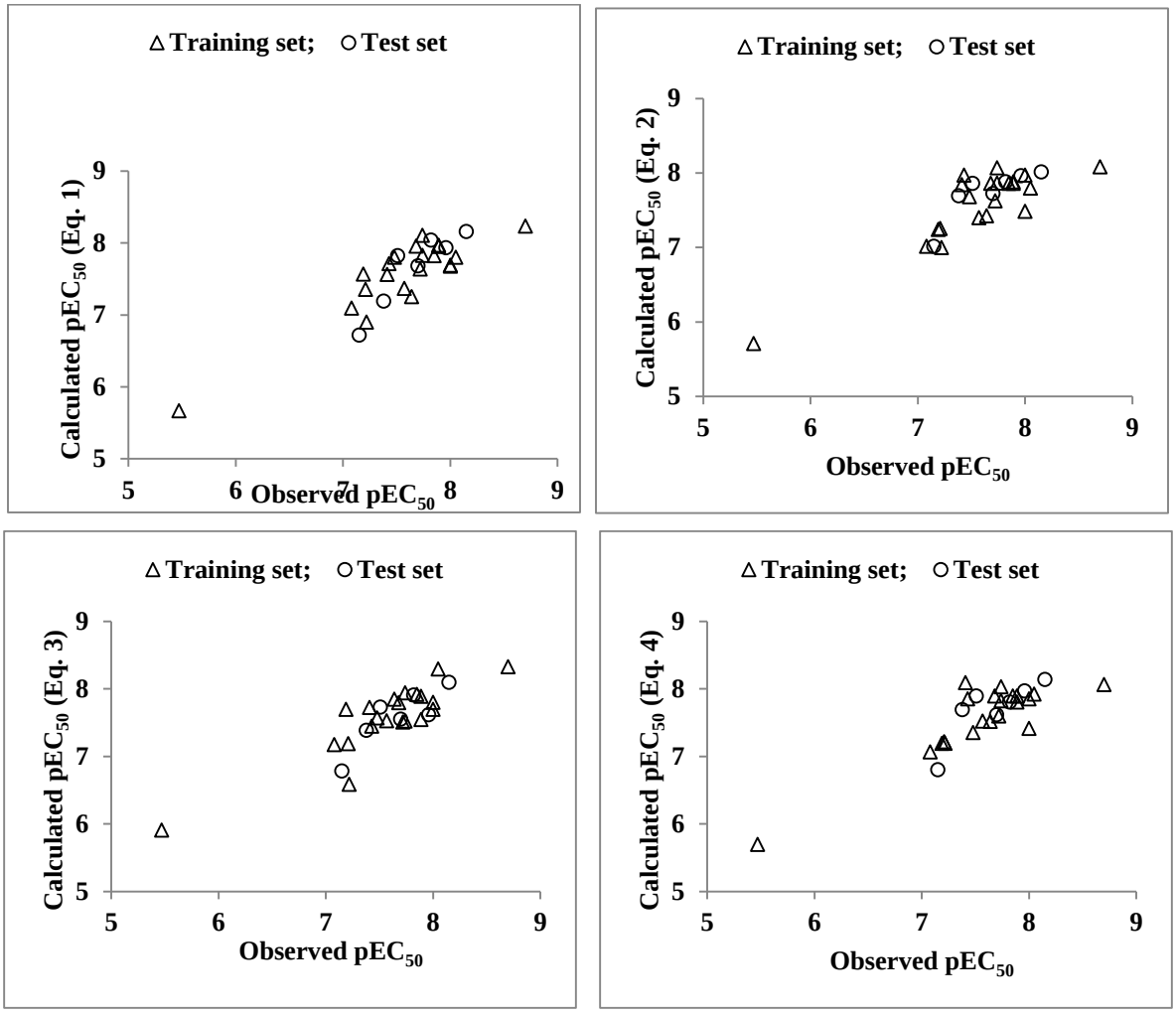


Figure 1: plot of observed and calculated pEC_{50} values of training- and test-set compounds for hGPR119 agonistic activity of triazolopyridines.

Applicability domain (AD)

On analyzing the model AD in the Williams plot, shown in Figure 2, 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 hGPR119 activity of triazolopyridines 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. **4** found to have leverage (h) values greater than the threshold leverage (h*) suggesting this training set compound as chemically influential compound. 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 = 28) for the hGPR119 agonistic activity in descriptors identified through CP-MLR.**

Model	r	s	F	Q^2_{LOO}	Eq.
pEC_{50} = 6.525 + 2.074(0.353)PW5 – 1.783(0.254)LP1 + 1.362(0.349)MATS4m – 0.685(0.223)MATS2e	0.887	0.273	21.332	0.610	(1a)
pEC_{50} =6.391 + 2.536(0.444)PW5 + 1.032(0.337)IVDE – 2.370(0.329)LP1 – 0.559(0.199)BEHv8	0.882	0.279	20.182	0.634	(2a)
pEC_{50}=6.956– 2.210(0.270)LP1 + 0.895(0.262)BELm7 + 0.924(0.204)MATS4e + 1.140(0.263)MATS5e	0.874	0.288	18.632	0.584	(3a)
pEC_{50}= 6.326 + 2.595(0.468)PW5 + 1.418(0.371)IVDE – 2.650(0.332)LP1 – 0.380(0.169)nCp	0.869	0.293	17.803	0.590	(4a)

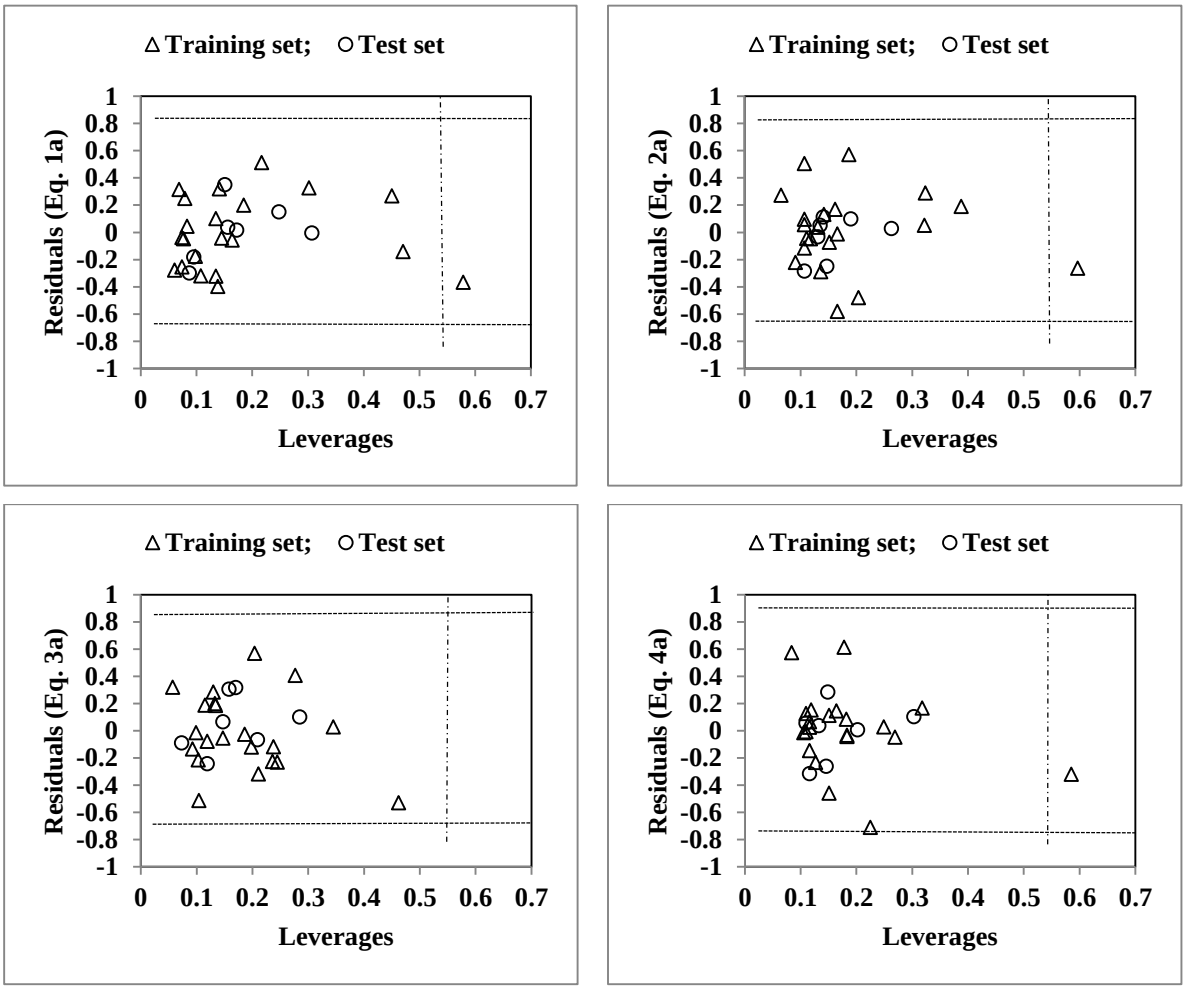


Figure 2. **Williams plot for the training-set and test- set compounds for hGPR119 agonistic activity. The horizontal dotted line refers to the residual limit (±3×standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.540).**

CONCLUSION

QSAR study has been carried out on the hGPR119 agonistic activity of triazolopyridines in 0D- to 2D-Dragon descriptors. The descriptors identified in CP-MLR analysis have highlighted the role

of molecular topology accounting features path/walk 5-Randic shape index (PW5), mean information vertex degree equality (IVDE), Lovasz-Pelikan index (LP1) in addition to atomic properties such as mass, van der Waals volume, and Sanderson electronegativity through weighted 2D autocorrelations (MATS4m, MATS2e, MATS4e and MATS5e) and modified Burden eigenvalues (BELm7 and BEHv8). Counts of total primary sp3 hybridized carbon atoms in a molecular structure (descriptor nCp) have also shown significance to optimize the hGPR119 agonistic 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

Acknowledgements

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. International Diabetes Federation, IDF diabetes atlas seventh edition 2015. http://www.diabetesatlas.org/.
2. Koro CE, Bowlin SJ, Bourgeois N and Fedder DO. Glycemic control from 1988 to 200 among U.S. adults diagnosed with type 2 diabetes: a preliminary report. Diabetes Care. 2004;27(1):17-20.
3. Stein SA, Lamos EM and Davis SN. A review of the efficacy and safety of oralantidiabetic drugs. Expert Opin Drug Saf. 2013;12(2):153-175.
4. Overton HA, Babbs AJ, Doel SM, Fyfe MCT, Gardner LS, Griffin G, Jackson HC,Procter MJ, Rasamison CM, Tang-Christensen M, Widdowson PS, Williams GM and Reynet C. Deorphanization of a G protein-coupled receptor for oleoylethanolamide and its use in the discovery of small-molecule hypophagic agents Cell Metab. 2006;3:167- 175.
5. Soga T, Ohishi T, Matsui T, Saito T, Matsumoto M, Takasaki J, Matsumoto S-I,Kamohara M, Hiyama H, Yoshida S, Momose K, Ueda Y, Matsushime H, Kobori M and Furuichi K. Lysophosphatidylcholine enhances glucose-dependent insulin secretion via an orphan G-protein-coupled receptor. Biochem Biophys Res Commun. 2005;326:744- 751.
6. Chu Z-L, Jones RM, He H, Carroll C, Gutierrez V, Lucman A, Moloney M, Gao H, Mondala H, Bagnol D, Unett D, Liang Y, Demarest K, Semple G, Behan DP and Leonard

J. A role for beta-cell-expressed G protein-coupled receptor 119 in glycemic control by enhancing glucose-dependent insulin release. Endocrinology. 2007;148:2601- 2609.

7. Chu Z-L, Carroll C, Alfonso J, Gutierrez V, He H, Lucman A, Pedraza M, Mondal H, Gao H, Bagnol D, Chen R, Jones RM, Behan DP and Leonard J. A role for intestinal endocrine cell-expressed g protein-coupled receptor 119 in glycemic control by enhancing glucagon-like Peptide-1 and glucose-dependent insulinotropic Peptide release. Endocrinology. 2008;149:2038-2047.
8. Farilla L, Bulotta A, Hirshberg B Calzi SL, Khoury N, Noushmehr H, Bertolotto C, Mario UD, Harlan DM, and Perfetti R. Glucagon-like peptide 1 inhibits cell apoptosis and improves glucose responsiveness of freshly isolated human islets. Endocrinology. 2003;144:5149-5158.
9. Trümper A, Trümper K, Trusheim H, Arnold R, Göke B, Hörsch D. Glucose-dependent insulinotropic polypeptide is a growth factor for beta (INS-1) cells by pleiotropic signaling. Mol Endocrinol. 2001;15(9):1559-1570.
10. Jones RM, Leonard JN, Buzard DJ, Lehmann J. GPR119 agonists for the treatment of type 2 diabetes. Expert Opin Ther Pat. 2009;19(10):1339-1359.
11. Ritter K, Buning C, Halland N, Pöverlein C, Schwink L. G Protein-Coupled Receptor 119 (GPR119) Agonists for the Treatment of Diabetes: Recent Progress and Prevailing Challenges. J Med Chem. 2016;59(8):3579-3592.
12. Semple G, Ren A, Fioravanti B, Pereira G, Calderon I, Choi K, Xiong Y, Shin Y-J, Gharbaoui T, Sage CR, Morgan M, Xing C, Chu Z-L, Leonard JN, Grottick AJ, Al-Shamma H, Liang Y, Demarest KT and Jones RM. Discovery of fused bicyclic agonistsof the orphan G-protein coupled receptor GPR119 with in vivo activity in rodent models of glucose control. Bioorg Med Chem Lett. 2011;21:3134-41.
13. Polli JW, Hussey E, Bush M, Generaux G, Smith G, Collins D, McMullen S, Turner and Nunez DJ. Evaluation of drug interactions of GSK1292263 (a GPR119 agonist) with statins: from in vitro data to clinical study design. Xenobiotica. 2013;43:498-508.
14. http://clinicaltrials.gov/ct2/show/NCT01035879.
15. Matsuda D, Kobashi Y, Mikami A, Kawamura M, Shiozawa F, Kawabe K, Hamada M, Oda K, Nishimoto S, Kimura K, Miyoshi M, Takayama N, Kakinuma H and Ohtake N. Design and synthesis of 1*H*-pyrazolo[3,4-*c*]pyridine derivatives as a novel structural class of potent GPR119 agonists. Bioorg Med Chem Lett. 2016;26:3441-3446.

16. Alelyunas YW, Empfield JR, McCarthy D, Spreen RC, Bui K, Pelosi-Kilby L and Shen C. Experimental solubility profiling of marketed CNS drugs, exploring solubility limit of CNS discovery candidate. Bioorg Med Chem Lett. 2010;20:7312-7316.
17. Matsuda D, Kobashi Y, Mikami A, Kawamura M, Shiozawa F, Kawabe K, Hamada M, Nishimoto S, Kimura K, Miyoshi M, Takayama N, Kakinuma H and Ohtake N. Novel 3*H*-[1,2,3]triazolo[4,5-*c*]pyridine derivatives as GPR119 agonists: Synthesis and structure-activity/solubility relationships. Bioorg Med Chem. 2017;25:4339–4354.
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. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: analysis of Selwood et al. Data Set-a case study. QSAR and Comb Sci. 2003; 22:583-595.
20. Sharma S, Prabhakar YS, Singh P and Sharma BK. QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach, Eur J Med Chem 2008;43:2354-2360.
21. Sharma S, Sharma BK, Sharma SK, Singh P and Prabhakar YS. Topological descriptors in modeling the agonistic activity of human A_3 adenosine receptor ligands: The derivatives of 2-Chloro-N^6-substituted-4'-thioadenosine-5'-uronamide, Eur J Med Chem. 2009; 44:1377-1382.
22. Sharma BK, Pilania P, Singh P and Prabhakar YS. 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 QSAR Environ Res. 2010; 21: 169-185.
23. Sharma BK, Singh P, Sarbhai K and Prabhakar YS. A quantitative structure-activity relationship study on serotonin 5-HT6 receptor ligands: Indolyl and piperidinyl sulphonamides, SAR QSAR Environ Res. 2010; 21:369-388.
24. Topliss JG and Edwards RP. Chance factors in studies of quantitative structure–activity relationships. J Med Chem 1979;22:1238–1244.
25. Katritzky AR, Dobchev DA, Slavov S and Karelson M. Legitimate utilization of large descriptor pools for QSAR/QSPR models. J Chem Inf Model 2008;48:2207–2213.
26. So S-S and Karplus M. Three-dimensional quantitative structure–activity relationship From molecular similarity matrices and genetic neural networks. 1. Method and validation. J Med Chem 1997;40:4347–4359.

27. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK and Katti SB. CP-MLR/PLS directed structure–activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3 thiazolidin-4-ones. QSAR Comb Sci 2004;23:234–244.
28. Akaike H. Information theory and an extension of the minimum likelihood principle. In Petrov BN and Csaki F, editors. Second international symposium on information theory. Budapest: Akademiai Kiado; 1973. p 267–281.
29. Akaike H. A new look at the statistical identification model. IEEE Trans Autom Control 1974;AC-19:716–723.
30. Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quant Struct-Act Relat 1994;13:285–294.
31. Kubinyi H. Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. Quant Struct-Act Relat 1994;13:393–401.
32. Friedman J. In Technical Report No. 102. Laboratory for Computational Statistics Stanford University: Stanford; 1990.
33. Gramatica, P. Principles of QSAR models validation: internal and external. QSAR Comb. Sci. 2007;26:694-701.