EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 9
Issue: 10
218-234
Year: 2022

CP-MLR/PLS DIRECTED QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP STUDY ON THE SMALL-MOLECULE MODULATORS OF THE MEDIATOR COMPLEX-ASSOCIATED KINASES CDK8 AND CDK19: THE PYRIDINE DERIVATIVES

Dinesh Kumar Meena[a], Mukesh Meena[a], Brij Kishore Sharma[a,*] and Raghuraj Parihar[b]

[a]Department of Chemistry, Government College, Bundi-323 001, India.
[b]Department of Chemistry, Government College, Kota-324 001, India.

***Corresponding Author: Brij Kishore Sharma**
Department of Chemistry, Government College, Bundi-323 001, India.

Article Received on 24/07/2022 **Article Revised on 14/08/2022** **Article Accepted on 04/09/2022**

ABSTRACT
The CDK8 and CDK19 inhibition activity of pyridine derivatives has been quantitatively analyzed in terms of Dragon descriptors. The statistically validated quantitative structure-activity relationship (QSAR) models provided rationales to explain the inhibition activities of these congeners. The descriptors identified through combinatorial protocol in multiple linear regression (CP-MLR) analysis for the CDK8 inhibitory activity have highlighted the role of 1st and 2nd order information content indices (IC1 and IC2, respectively), 2D Petitjean shape index (PJI2), Galvez topological charge index of 10th order (GGI10) and 1st order mean topological charge index (IGI1) to rationalize the activity. Atomic properties such as atomic van der Waals volumes and polarizabilities in terms of Mv, and Moran and Geary autocorrelations (MATS3v, GATS1v and GATS3p) in addition to the number of oxygen atoms (nO) have shown prevalence to model the CDK8 inhibitory activity. PLS analysis has also corroborated the dominance of CP-MLR identified descriptors. 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. The derived QSAR models for the CDK19 inhibitory activity have revealed that the structural information content of 2nd order neighborhood symmetry (SIC2), atomic polarizabilities weighted Geary autocorrelations of lag 5 and 7 (GATS5p and GATS7p) and atomic masses weighted lowest eigenvalue n.2 of Burden matrix (BELm2) played a pivotal role in rationalization of CDK19 inhibition activity of titled compounds. Additionally, presence of H attached to C0(sp3) with 1X attached to next carbon (H-052) and CH3R/CH4 (C-001) type structural fragment and lesser number or absence of tertiary aliphatic amines functionality (nNR2) are also predominant to explain CDK19 inhibition actions of pyridine derivatives.

KEYWORDS: QSAR, CDK8 and CDK19 inhibitors, Combinatorial protocol in multiple linear regression (CP-MLR) analysis, PLS, Dragon descriptors, Pyridine derivatives.

INTRODUCTION
The regulation of gene transcription in many contexts such as stem cell function, the immune response, inflammation, cell adhesion, the epithelial to mesenchymal transition and development is associated to the multiprotein assembly, the Mediator complex.[1-5] The kinase module of the Mediator complex is formed by the cyclin C-dependent kinases CDK8 and CDK19 along with MED12 and MED13.[6,7] CDK8 regulates basal transcription by phosphorylation of RNA polymerase II and phosphorylates E2F1 that leads to activation of WNT signaling.[8,9] In colon and gastric cancers CDK8 gene expression is associated with activation of □-catenin which is a core transcriptional regulator of canonical WNT signaling.[10,11] The increased mortality in colorectal, breast and ovarian camcers is also associated CDK8 gene expression.[12] In melanoma, CDK8 has overexpression and necessary for cell proliferation.[13] The location of CDK8 is in the region of chromosome 13 which is known to undergo copy number gain in about 60 percent of colorectal cancers. The growth of HT29 and Colo205 was reduced, by inducible shRNA-mediated knock-down of CDK8 protein, in human tumor xenograft animal models harboring CDK8 gene amplification.[14] The implications of the kinase function of CDK8 in oncogenesis are due to the fact that expression of CDK8 transforms NIH3T3 cells into a malignant phenotype.[15] A kinase-dead mutant is devoid of this transformation. CDK19 form Mediator complexes and that are independent of CDK8

but the context-dependent roles are still the subject of further studies.[7]

The small molecule ligands for CDK8 and its paralog CDK19 are the steroidal natural product cortistatin A[16] and its analogs[17,18], sorafenib[19] and its derivatives[20], linifanib and ponatinib[21], senexin B.[22] In patent literature other small molecule inhibitors of CDK8 are also reported.[23-25] Mallinger *et al.*, in a cell-based pathway screen, reported CCT251545 as a small molecule inhibitor of WNT signaling.[26] A novel series of trisubstituted pyridine derivatives considering primary targets of protein kinases CDK8 and CDK19 has been reported by Mallinger *et al.*[27] 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.

MATERIAL AND METHODS

Data-set

For present work the reported fifty five pyridine derivatives have been considered as the data set.[27] The general structure of these compounds is represented in Figure 1 and structural variations are mentioned in Table 1.

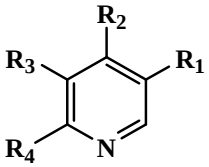
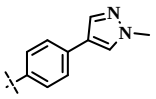
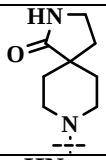
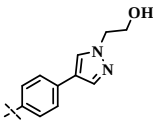
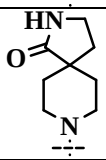
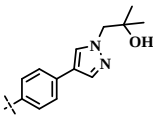
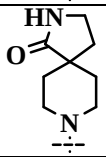
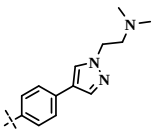
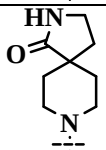
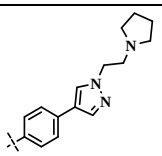
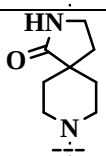
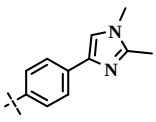
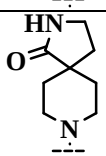
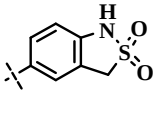
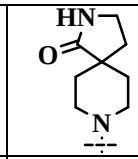
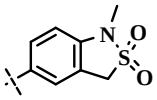
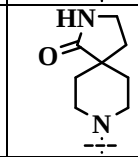
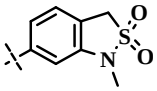
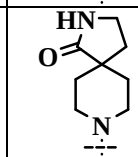
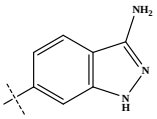
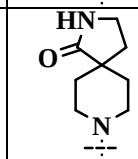
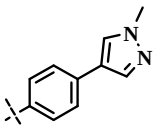
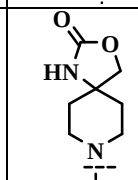
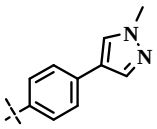
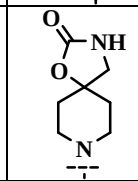
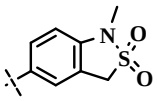
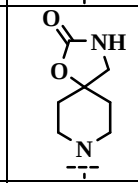
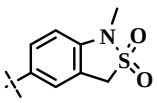
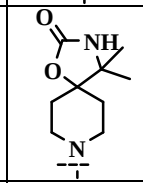
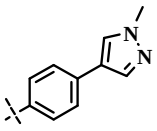
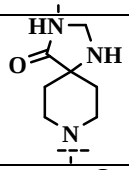
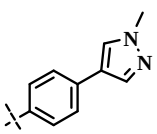
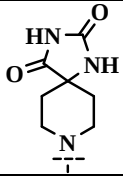
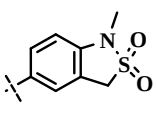
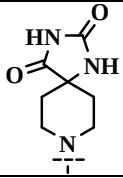


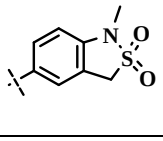
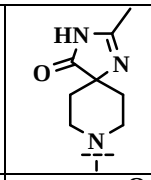
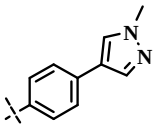
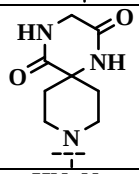
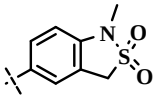
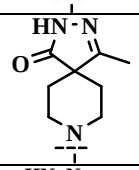
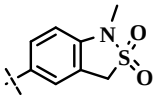
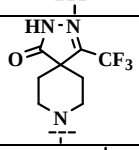
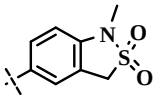
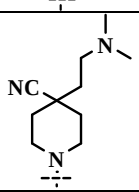
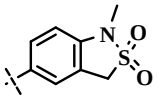
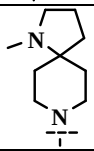
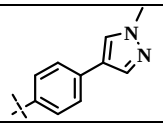
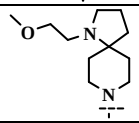
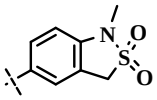
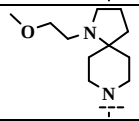
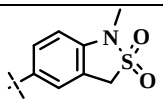
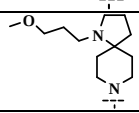
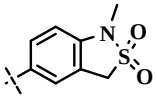
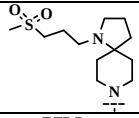
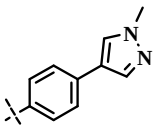
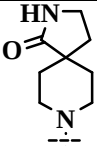
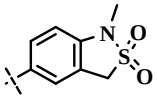
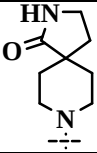
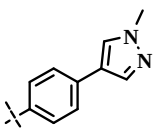
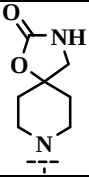
Fig. 1: General structure of pyridine derivatives.

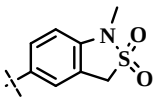
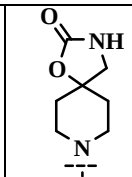
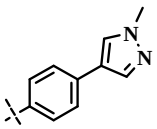
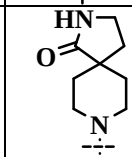
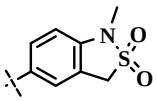
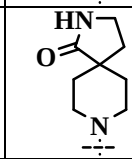
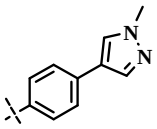
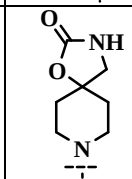
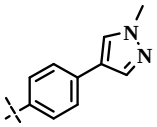
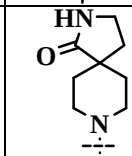
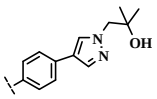
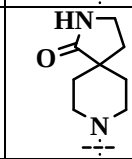
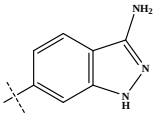
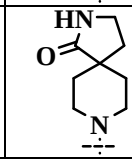
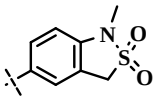
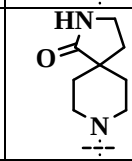
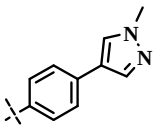
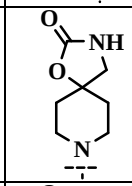
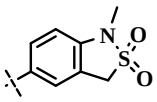
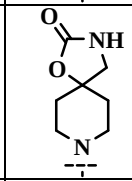
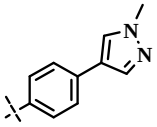
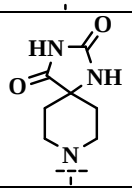
These derivatives were evaluated for their inhibition of CDK8 and CDK19. Both the inhibition activities have also been reported in Table1.[27] The inhibition activity, IC_{50}, represents the concentration of a compound to achieve 50% inhibition of CDK8 and CDK19. The same is expressed as pIC_{50} on a molar basis and considered as the dependent variable for the present quantitative analysis. In the dataset, the initial assessment of activity with all descriptors has suggested the compound **11** as potential outlier. An outlier to a QSAR can indicate the limits of applicability of QSAR models. This outlier is not part of the data set. 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.

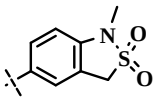
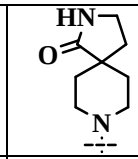
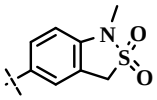
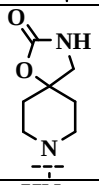
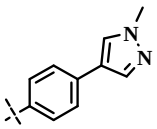
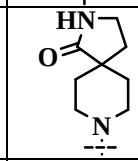
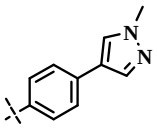
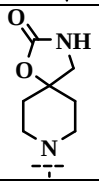
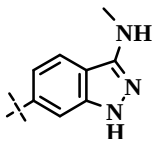
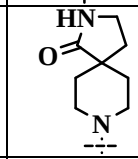
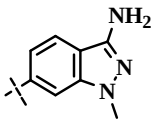
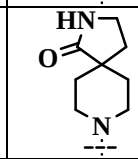
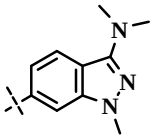
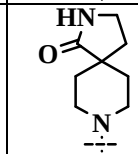
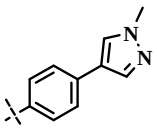
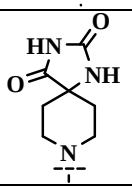
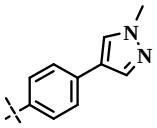
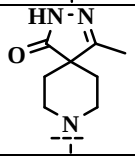
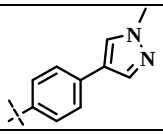
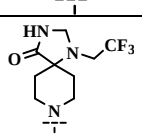
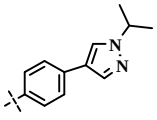
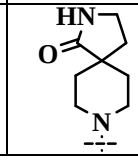
Table 1: Structural variations and observed CDK8 and CDK19 inhibition activities of pyridine derivatives.

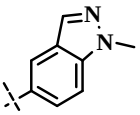
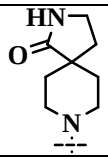
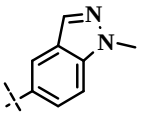
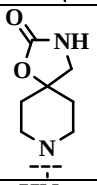
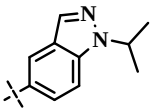
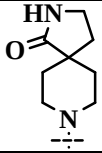
Cpd.	R_1	R_2	R_3	R_4	pIC_{50}[a]	
					CDK8	CDK19
1[b]			Cl	H	8.14	8.22
2[b]			Cl	H	8.26	8.60
3			Cl	H	8.80	8.85
4			Cl	H	7.55	-[d]
5[b]			Cl	H	7.57	7.69
6			Cl	H	7.69	-[d]

7			Cl	H	8.62	8.60
8[b]			Cl	H	8.62	8.74
9			Cl	H	8.43	-[d]
10			Cl	H	8.77	8.48
11[c]			Cl	H	_c	-[d]
12			Cl	H	7.97	8.12
13			Cl	H	8.64	8.74
14[b]			Cl	H	7.85	-[d]
15[b]			Cl	H	8.41	8.08
16			Cl	H	8.15	8.25
17			Cl	H	8.62	-[d]

18			Cl	H	8.35	8.32
19[b]			Cl	H	8.38	-[d]
20			Cl	H	8.57	9.00
21			Cl	H	7.92	-[d]
22			Cl	H	7.98	-[d]
23			Cl	H	7.43	7.64
24			Cl	H	6.15	-[d]
25			Cl	H	7.31	7.74
26			Cl	H	6.85	7.04
27			Cl	H	6.84	-[d]
28[b]			F	H	7.10	-[d]
29			F	H	8.42	8.52
30			F	H	7.69	-[d]

31			F	H	8.37	-[d]
32			CF_3	H	7.83	-[d]
33			CF_3	H	8.48	8.72
34[b]			CF_3	H	7.48	8.11
35[b]			Cl	NH_2	8.30	8.46
36			Cl	NH_2	7.79	-[d]
37			Cl	NH_2	8.68	-[d]
38[b]			Cl	NH_2	8.64	8.51
39			Cl	NH_2	8.02	8.24
40			Cl	NH_2	8.72	8.68
41			Cl	NH_2	8.55	8.48

42[b]			F	NH_2	8.48	8.57
43			F	NH_2	8.26	-[d]
44[b]			CF_3	NH_2	8.03	-[d]
45			CF_3	NH_2	7.92	8.02
46			Cl	H	8.55	8.49
47			Cl	H	8.49	-[d]
48[b]			Cl	H	7.33	-[d]
49			Cl	H	7.76	-[d]
50			Cl	H	8.24	-[d]
51[d]			Cl	H	-[d]	-[d]
52			Cl	NH_2	7.53	-[d]

53			Cl	NH_2	8.64	8.59
54[b]			Cl	NH_2	7.94	-[d]
55[b]			Cl	NH_2	7.27	-[d]

[a]Reference[27], pIC_{50} on molar basis, IC_{50} represents the concentration of a compound to bring out 50% inhibition of CDK8 and CDK19; [b]Compound included in test set; [c]"Outlier" compound not included in data set; [d]Not reported.

Molecular descriptors

The structures of the compounds (Table 1), under study, have been drawn in 2D ChemDraw[28] and were converted into 3D objects using the default conversion procedure implemented in the CS Chem3D Ultra. The generated 3D-structures of the compounds were subjected to energy minimization in the MOPAC module, using the AM1 procedure for closed shell systems, implemented in the CS Chem3D Ultra. This will ensure a well defined conformer relationship across the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software[29] for computing the descriptors corresponding to 0D-, 1D-, and 2D-classes.

Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR)[30-34] and partial least squares (PLS)[35-37] procedures were used in the present work for developing QSAR models. The CP-MLR is a "filter"-based variable selection procedure, which employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure–activity models, each having unique combination of descriptors from the generated dataset of the compounds under study. The embedded filters make the variable selection process efficient and lead to unique solution. Fear of "chance correlations" exists where large descriptor pools are used in multilinear QSAR/QSPR studies. Furthermore, in order to discover any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been put to a randomization test[38,39] by repeated randomization of the activity to ascertain the chance correlations, if any, associated with them. For this, every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted, to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test its prediction and generalization within the study domain. For each model, derived by involving n data points, a number of statistical parameters such as r (the multiple correlation coefficient), s (the standard deviation), F (the F ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to access its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5.

Applicability Domain

The utility of a QSAR model is based on its accurate prediction ability for new compounds. A model is valid only within its training domain and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain is assessed by the leverage values for each compound.[40] The Williams plot (the plot of standardized residuals versus leverage values, h) can then be used for an immediate and simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the applicability domain is established inside a squared area within ± x (s.d.) and a leverage threshold h^*. The threshold h^* is generally fixed at 3(k + 1)/n (n is the number of training-set compounds and k is the number of model parameters) whereas x = 2 or 3. Prediction must be considered unreliable for compounds with a high leverage value ($h > h^*$). On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training-set compounds.

RESULTS AND DISCUSSION

QSAR results

For the compounds in Table 1, a total number of 514 descriptors belonging to 0D- to 2D- classes of DRAGON have been computed. Prior to model development procedure, all those descriptors that are inter-correlated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This procedure has reduced the total descriptors from 514 to 131 as relevant ones to explain the biological actions of titled compounds and these were subjected to CP-MLR analysis with default "filters" set in it. The descriptors have been scaled between the intervals 0 to 1 to ensure that a descriptor will not dominate simply because it has larger or smaller pre-scaled value compared to the other descriptors. In this way, the scaled descriptors would have equal potential to influence the QSAR models.

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds.

The 53 compounds were divided into training-set and test-set. Sixteen compounds (nearly 30% of total population) have been selected for test-set. The identified test-set was then used for external validation of models derived from remaining thirty seven compounds in the training-set. The squared correlation coefficient between the observed and predicted values of compounds from test-set, r^2_{Test}, was calculated to explain the fraction of explained variance in the test-set which is not part of regression/model derivation. It is a measure of goodness of the derived model equation. A high r^2_{Test} value is always good. But considering the stringency of test-set procedures, often r^2_{Test} values in the range of 0.5 to 0.6 are regarded as logical models. Following the strategy to explore only predictive models, CP-MLR resulted into 03 models in two descriptors and 45 models in three descriptors for the CDK8 inhibitory activity. The selected models are mentioned in Table 2.

Table 2: **Highest significant models in two and three parameters derived for training set through CP-MLR for CDK8 inhibitory activity.**

Model	r	s	F	Q^2_{LOO}	r^2_{Test}	Eq.
pIC_{50} = 8.887 +0.786(0.203)nO –2.187(0.314)GATS1v	0.770	0.400	24.878	0.505	0.574	(1)
pIC_{50} = 8.042 +1.056(0.167)nO –2.054(0.245)GATS1v +1.127(0.231)GATS3p	0.873	0.310	35.547	0.690	0.579	(2)
pIC_{50} = 8.829 –0.798(0.192)GGI10 +1.096(0.204)JGI1 – 1.920(0.238)GATS1v	0.869	0.315	33.955	0.676	0.560	(3)
pIC_{50} = 9.218 +0.833(0.165)nO –1.063(0.246)MATS3v -1.880(0.264)GATS1v	0.860	0.325	31.393	0.667	0.618	(4)
pIC_{50} = 7.508 –0.288(0.127)PJI2 +1.729(0.210)IC1 - 0.732(0.212)GGI10	0.857	0.328	30.626	0.650	0.528	(5)

The signs of the regression coefficients have indicated the direction of influence of explanatory variables in above 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 above model Eqs., (1-5), the descriptors participated are from constitutional class (descriptor nO), topological class (descriptors PJI2 and IC1), 2D-autocorrelations (GATS1v, GATS3p and MATS3v) and Galvez topological charge indices (GGI10 and JGI1).

The positive sign of regression coefficients of descriptors nO (number of oxygen atoms), GATS3p (atomic polarizabilities weighted Geary autocorrelation of lag 3), IC1 (information content index of 1st order neighborhood symmetry) and JGI1 (1st order mean topological charge index) suggested that a higher value of these descriptors would be beneficial to augment the CDK8 inhibitory activity. On the other hand, a lower value of descriptors GATS1v (atomic van der Waals volume weighted Geary autocorrelation of lag 1), MATS3v (atomic van der Waals volume weighted Moran autocorrelation of lag 3), PJI2 (2D Petitjean shape index) and GGI10 (10th order topological charge index) would be supportive to the CDK8 inhibition.

Considering the number of observation in the dataset, models with up to four descriptors were explored. It has resulted in 45 four-parameter models with test set $r^2 > 0.50$. These models (with 131 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 (=0.861) has been used as the new threshold of filter-3 for the next generation. These models have shared 35 descriptors among them. All these 35 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 physical meaning, average regression coefficient and incidence[b], in modeling the CDK8 inhibitory activities.

Descriptor class	Descriptor (physical meaning), avg reg coeff (incidence)
Constitutional	AMW (average molecular weight), 1.550(4); Mv (mean atomic van der Waals volume scaled on Carbon atom), 0.812(8); Me (mean atomic Sanderson electronegativity scaled on Carbon atom), 0.966(1); Mp (mean atomic polarizability scaled on Carbon atom), 0.903(1); nAB (number of aromatic bonds), -0.321(1); nN (number of Nitrogen atoms), -0.690(2); nO (number of Oxygen atoms), 0.907(9)
Topological	SPI (superpendentic index), -0.968(7); PJI2 (2D Petitjean shape index), -0.303(2); IC1 (information content index of 1st order neighborhood symmetry), 1.316(9); TIC1 (total information content index of 1st order neighborhood symmetry), 0.870(1); IC2 (information content index of 2nd order neighborhood symmetry), 0.814(5); piPC10 (molecular multiple path count of order 10), -0455(1); T(N..O) (sum of topological distances between N..O), 0.738(1)
BCUT	BEHm3 (atomic masses weighted highest eigenvalue n.3 of Burden matrix), -1.443(7)
Galvez topological charge indices	GGI5 (5th order topological charge index), 0.771(1); GGI6 (6th order topological charge index), -0.757(2); GGI10 (10th order topological charge index), -0.872(25); JGI1 (1st order mean topological charge index), 1.284(3)
2D autocorrelations	MATS2m (atomic masses weighted Moran autocorrelation of lag 2), 1.449(1); MATS4m (atomic masses weighted Moran autocorrelation of lag 4), 0.663(5); MATS1v (atomic van der Waals volume weighted Moran autocorrelation of lag 1), 0.751(2); MATS3v (atomic van der Waals volume weighted Moran autocorrelation of lag 3), -1.028(19); MATS4e (atomic Sanderson electronegativities weighted Moran autocorrelation of lag 4), -0.848(1); GATS1v (atomic van der Waals volume weighted Geary autocorrelation of lag 1), -1.693(27); GATS8v (atomic van der Waals volume weighted Geary autocorrelation of lag 8), 0.515(1); GATS3p (atomic polarizabilities weighted Geary autocorrelation of lag 3), 1.043(14)
Functional groups	nCp (number of total sp3 hybridized primary carbon atoms), -0.849(3); nCaR ((number of substituted sp2 hybridized aromatic carbon atoms), -0.444(1); nCONHR (number of secondary aliphatic amides), 0.290(1); nCONN (number of urea derivatives), -0.438(1); nHDon (number of donor atoms for H-bonds with N and O), 0.785(2)
Atom centred fragments	C-004 (CR4), 0.257(1); H-049 (H attached to C3(sp3)/C2(sp2)/C3(sp2)/ C3(sp)), 0.535(1); O-058 (O=), 1.363(9)

[a]The descriptors are identified from the four parameter models for activity emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.861 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 37 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.

Following are the selected four-descriptor models for the CDK8 inhibitory activities emerged through CP-MLR.

pIC_{50} = 8.072 +0.392(0.173)Mv +1.078(0.207)IC1 –0.934(0.146)GGI10 –1.006(0.180)MATS3v
n = 37, r = 0.927, s = 0.242, F = 48.998, Q^2_{LOO} = 0.780, Q^2_{L5O} = 0.789, r^2_{Test} = 574 (6)

pIC_{50} = 8.210 +0.755(0.152)Mv +0.707(0.167)IC2 –0.927(0.159)GGI10 –1.173(0.194)MATS3v
n = 37, r = 0.912, s = 0.264, F = 40.050, Q^2_{LOO} = 0.747, Q^2_{L5O} = 0.685, r^2_{Test} = 0.547 (7)

pIC_{50} = 8.250 +0.943(0.146)nO –0.596(0.166)GGI10 –1.886(0.215)GATS1v +1.117(0.198)GATS3p
n = 37, r = 0.911, s = 0.265, F = 39.523, Q^2_{LOO} = 0.754, Q^2_{L5O} = 0.759, r^2_{Test} = 0.613 (8)

pIC_{50} = 9.448 +0.713(0.143)nO –0.650(0.171)GGI10 –1.098(0.208)MATS3v –1.686(0.229)GATS1v
n = 37, r = 0.906, s = 0.274, F = 36.697, Q^2_{LOO} = 0.741, Q^2_{L5O} = 0.743, r^2_{Test} = 0.638 (9)

These models have accounted for nearly 86% 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 pIC_{50} values of training set compounds calculated using Eqs. (6) to (9) have been included in Table 4. The models (6) to (9) are validated with an external test set of 16 compounds listed in Table 4. 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.

Table 4: Observed and calculated CDK8 and CDK19 inhibition activities of pyridine derivatives.

Cpd.	CDK8 Inhibition $pIC_{50}(M)$[a]							CDK19 Inhibition $pIC_{50}(M)$[a]				
	Obs.	Calculated						Obs.	Calculated			
		Eq.(6)	Eq.(7)	Eq.(8)	Eq.(9)	PLS			Eq.(10)	Eq.(11)	Eq.(12)	Eq.(13)
1[b]	8.14	8.14	8.12	8.15	8.21	8.12		8.22	8.07	8.24	8.46	8.31
2[b]	8.26	7.76	7.70	8.08	7.93	7.84		8.60	8.53	8.45	8.37	8.31
3	8.8	8.18	8.19	8.27	8.40	8.26		8.85	8.71	8.70	8.79	8.87
4	7.55	7.76	7.75	7.68	7.74	7.60		-[d]	-[d]	-[d]	-[d]	-[d]
5[b]	7.57	7.35	7.17	7.40	7.43	7.26		7.69	7.72	7.45	7.46	7.74
6	7.69	7.74	7.77	7.57	7.64	7.79		-[d]	-[d]	-[d]	-[d]	-[d]
7	8.62	8.69	8.72	8.24	8.92	8.54		8.6	8.77	8.74	8.74	8.62
8[b]	8.62	8.53	8.64	8.59	8.62	8.59		8.74	8.46	8.57	8.73	8.62
9	8.43	8.40	8.51	8.50	8.53	8.49		-[d]	-[d]	-[d]	-[d]	-[d]
10	8.77	8.94	8.81	9.08	9.02	8.95		8.48	8.79	8.70	8.42	8.56
11[c]	-[c]	-[c]	-[c]	-[c]	-[c]	-[c]		-[d]	-[d]	-[d]	-[d]	-[d]
12	7.97	7.91	7.92	7.93	7.91	7.83		8.12	8.05	8.09	8.29	8.19
13	8.64	8.41	8.48	8.55	8.46	8.37		8.74	8.44	8.44	8.59	8.55
14[b]	7.85	8.10	8.23	8.16	8.17	8.15		-[d]	-[d]	-[d]	-[d]	-[d]
15[b]	8.41	8.19	8.23	7.99	8.09	8.18		8.08	8.25	8.23	8.29	8.25
16	8.15	8.23	8.23	8.28	8.36	8.17		8.25	8.26	8.23	8.24	8.27
17	8.62	8.72	8.86	8.83	8.84	8.72		-[d]	-[d]	-[d]	-[d]	-[d]
18	8.35	8.30	8.45	8.06	8.16	8.27		8.32	8.35	8.57	8.53	8.70
19[b]	8.38	8.03	8.03	7.99	8.09	7.94		-[d]	-[d]	-[d]	-[d]	-[d]
20	8.57	8.56	8.82	8.58	8.58	8.73		9	9.15	8.99	8.92	8.89
21	7.92	8.40	8.48	8.15	8.22	8.42		-[d]	-[d]	-[d]	-[d]	-[d]
22	7.98	7.71	7.78	7.71	7.81	7.79		-[d]	-[d]	-[d]	-[d]	-[d]
23	7.43	7.61	7.66	7.55	7.68	7.65		7.64	7.69	7.50	7.64	7.52
24	6.15	6.63	6.62	6.62	6.68	6.52		-[d]	-[d]	-[d]	-[d]	-[d]
25	7.31	6.97	7.09	7.17	7.09	6.96		7.74	7.69	7.54	7.48	7.47
26	6.85	6.90	6.99	7.13	7.02	6.99		7.04	7.23	7.38	7.31	7.43
27	6.84	6.79	6.95	7.07	7.21	6.99		-[d]	-[d]	-[d]	-[d]	-[d]
28[b]	7.1	7.95	7.89	7.92	7.94	7.92		-[d]	-[d]	-[d]	-[d]	-[d]
29	8.42	8.33	8.41	8.32	8.32	8.37		8.52	8.60	8.57	8.83	8.60
30	7.69	7.68	7.50	7.66	7.61	7.55		-[d]	-[d]	-[d]	-[d]	-[d]
31	8.37	8.27	8.24	8.26	8.14	8.17		-[d]	-[d]	-[d]	-[d]	-[d]
32	7.83	7.98	7.89	7.86	7.80	8.02		-[d]	-[d]	-[d]	-[d]	-[d]
33	8.48	8.51	8.47	8.41	8.37	8.55		8.72	8.38	8.57	8.54	8.60
34[b]	7.48	7.66	7.52	7.64	7.51	7.65		8.11	7.93	8.10	8.05	8.13
35[b]	8.3	8.18	8.07	8.17	8.14	8.21		8.46	8.34	8.39	8.50	8.31
36	7.79	8.19	8.13	8.28	8.32	8.32		-[d]	-[d]	-[d]	-[d]	-[d]
37	8.68	8.93	8.72	9.18	9.03	8.99		-[d]	-[d]	-[d]	-[d]	-[d]
38[b]	8.64	8.59	8.52	8.71	8.64	8.67		8.51	8.70	8.71	8.76	8.62
39	8.02	7.84	7.67	7.95	7.85	7.81		8.24	8.28	8.23	8.34	8.19
40	8.72	8.56	8.53	8.68	8.49	8.54		8.68	8.63	8.55	8.62	8.55
41	8.55	8.21	8.15	8.29	8.29	8.22		8.48	8.49	8.37	8.30	8.27
42[b]	8.48	8.50	8.41	8.53	8.45	8.55		8.57	8.83	8.71	8.86	8.60
43	8.26	8.43	8.23	8.49	8.29	8.35		-[d]	-[d]	-[d]	-[d]	-[d]
44[b]	8.03	7.89	7.63	7.91	7.71	8.01		-[d]	-[d]	-[d]	-[d]	-[d]
45	7.92	7.66	7.44	7.69	7.43	7.72		8.02	8.17	8.23	8.21	8.13
46	8.55	8.48	8.42	8.54	8.38	8.35		8.49	8.40	8.53	8.32	8.56
47	8.49	8.39	8.28	8.32	8.28	8.40		-[d]	-[d]	-[d]	-[d]	-[d]
48[b]	7.33	7.70	7.53	7.72	7.66	7.57		-[d]	-[d]	-[d]	-[d]	-[d]
49	7.76	7.98	8.05	7.96	7.98	7.91		-[d]	-[d]	-[d]	-[d]	-[d]
50	8.24	8.11	8.25	7.79	7.89	8.09		-[d]	-[d]	-[d]	-[d]	-[d]
51[d]	-[d]	-[d]	-[d]	-[d]	-[d]	-[d]		-[d]	-[d]	-[d]	-[d]	-[d]
52	7.53	7.68	7.69	7.57	7.52	7.66		-[d]	-[d]	-[d]	-[d]	-[d]
53	8.64	8.51	8.45	8.38	8.34	8.51		8.59	8.45	8.59	8.44	8.56
54[b]	7.94	8.35	8.23	8.25	8.14	8.23		-[d]	-[d]	-[d]	-[d]	-[d]
55[b]	7.27	7.41	7.36	7.44	7.35	7.44		-[d]	-[d]	-[d]	-[d]	-[d]

[a]IC_{50} on molar basis, taken from reference[27]; [b]Compound included in test set; [c]"Outlier" compound not included in data set; [d]Not reported.

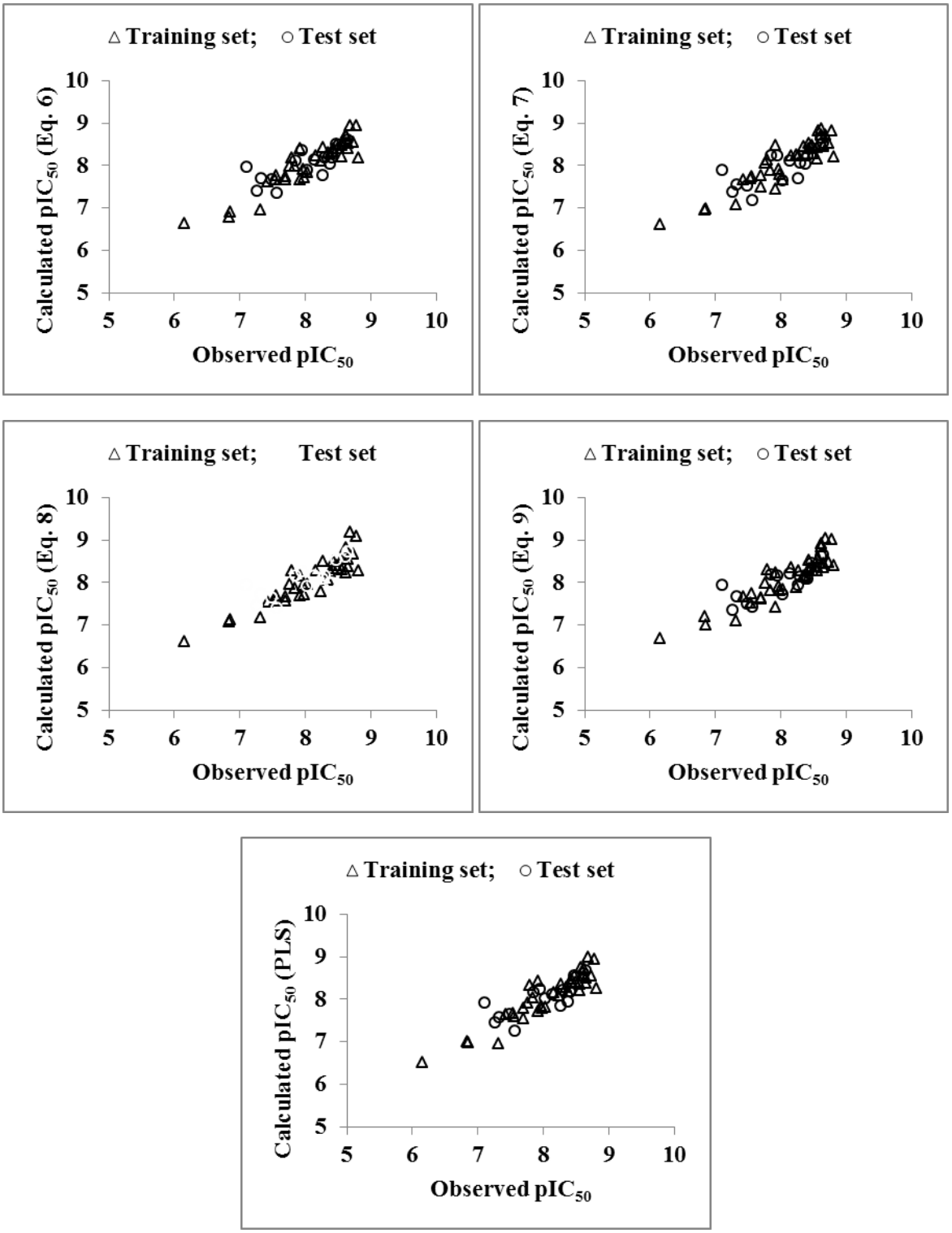


Fig. 2: Plot of observed versus caculated pIC_{50} values for training- and test-set compounds for CDK8 inhibition.

The newly appeared descriptors in above models are Mv (constitutional class) and IC2 (topological class). Both of these descriptors have shown positive correlation to the activity. The signs of regression coefficients advocated that higher values of mean atomic van der Waals volume scaled on Carbon atom (descriptor Mv) and information content index of 2nd order neighborhood symmetry (descriptor IC2) would be incremental to the activity.

A partial least square (PLS) analysis has also been carried out on the descriptors emerged in CP-MLR identified models {Eqs. (1) to (9)}, and discussed above to facilitate the development of a "single window" structure–activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, two components are found to be the optimum for these 10 descriptors and

they explained 88.26% variance in the activity. The MLR-like PLS coefficients of these 10 descriptors are given in Table 5.

Table 5: PLS and MLR-like PLS models from the 10 descriptors of four parameter CP-MLR models for CDK8 inhibitory activities.

A: PLS equation										
PLS components						PLS coefficient (s.e.)[a]				
Component-1						0.315(0.023)				
Component-2						0.111(0.032)				
Constant						8.069				
B: MLR-like PLS equation										
S. No.	**Descriptor**	**MLR-like coefficient[b]**	**(f.c.)[c]**	**Order**		**S. No.**	**Descriptor**	**MLR-like coefficient[b]**	**(f.c.)[c]**	**Order**
1	Mv	0.20	0.11	5		6	GGI10	-0.32	-0.19	1
2	nO	0.08	0.05	10		7	JGI1	0.09	0.05	9
3	PJI2	-0.12	-0.07	8		8	MATS3v	-0.22	-0.12	2
4	IC1	0.21	0.12	3		9	GATS1v	-0.21	-0.12	4
5	IC2	0.13	0.08	7		10	GATS3p	0.17	0.10	6
							Constant = -1.475			
C: PLS regression statistics						**Values**				
n						37				
r						0.924				
s						0.240				
F						99.690				
Q^2_{LOO}						0.817				
Q^2_{L5O}						0.815				
r^2_{Test}						0.615				

[a]Regression coefficient of PLS factor and its standard error. [b]Coefficients of MLR-like PLS equation in terms of descriptors for their original values; [c]f.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data.

For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is also given in Figure 2. Figure 3 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity. The PLS analysis has suggested GGI10 as the most determining descriptor for modeling the activity of the compounds (Figure 3). The other descriptors in decreasing order of significance are MATS3v, IC1, GATS1v, Mv, GATS3p, IC2, PJI2, JGI1 and nO. All these descriptors convey same inference in the PLS model as well. It is also observed that PLS model from the dataset devoid of CP-MLR identified 10 descriptors is inferior in explaining the activity of the analogues.

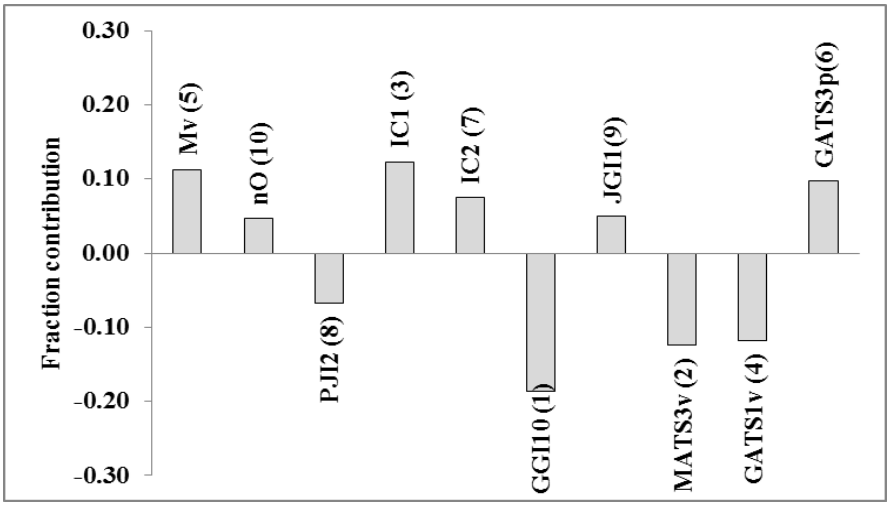


Fig. 3: Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 10 CP-MLR identified descriptors associated with CDK8 inhibitory activity of pyridine derivatives.

CP-MLR analysis has also been carried out for another reported inhibition activity CDK19 using the same test set. Following are the selected (from the 66 models, sharing 51 descriptors) highly significant three-descriptor models for the CDK19 inhibitory activities emerged through CP-MLR.

pIC_{50} = 6.850 +1.500(0.154)SIC2 +0.533(0.155)GATS7p +1.008(0.157)H-052
n = 19, r = 0.939, s = 0.178, F = 37.338, Q^2_{LOO} = 0.804, Q^2_{L5O} = 0.800, r^2_{Test} = 0.672 (10)

pIC_{50} = 7.710 +0.857(0.253)SIC2 –0.572(0.185)nNR2 +0.724(0.161)H-052
n = 19, r = 0.933, s = 0.186, F = 33.757, Q^2_{LOO} = 0.763, Q^2_{L5O} = 0.755, r^2_{Test} = 0.778 (11)

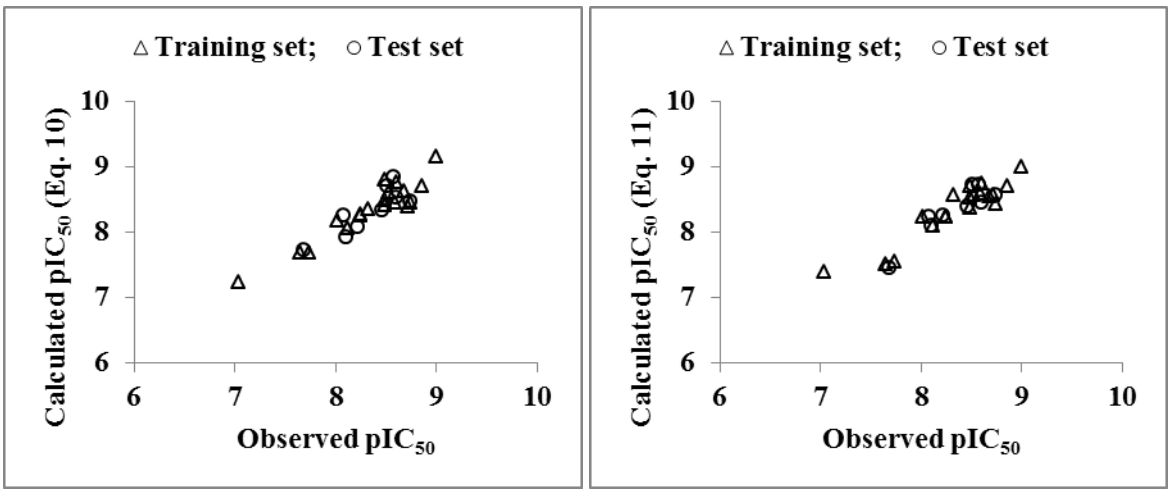
pIC_{50} = 8.833–0.628(0.150)GATS5p –1.109(0.124)nNR2 +0.394(0.168)C-001
n = 19, r = 0.931, s = 0.188, F = 32.867, Q^2_{LOO} = 0.758, Q^2_{L5O} = 0.667, r^2_{Test} = 0.593 (12)

pIC_{50} = 8.696 –0.951(0.294)BELm2 –0.941(0.125)nNR2 +1.120(0.238)H-052
n = 19, r = 0.930, s = 0.190, F = 32.247, Q^2_{LOO} = 0.698, Q^2_{L5O} = 0.689, r^2_{Test} = 0.798 (13)

The descriptors participated in above models are SIC2 (topological class descriptor; structural information content of 2nd order neighborhood symmetry), GATS5p and GATS7p (2D-autocorrelation descriptors; atomic polarizabilities weighted Geary autocorrelations of lag 5 and 7, respectively), BELm2 (BCUT descriptor; atomic masses weighted lowest eigenvalue n.2 of Burden matrix), H-052 and C-001 (atom-centered fragments; representing H attached to C0(sp3) with 1X attached to next carbon and CH3R/CH4, respectively) and nNR2 (functional group descriptor; number of tertiary aliphatic amines).

Descriptors SIC2, GATS7p, H-052 and C-001 have shown positive correlation to the CDK19 activity suggesting that higher values of structural information content of 2nd order neighborhood symmetry, atomic polarizabilities weighed Geary autocorrelation of lag 7 in addition to presence of certain atom-centered fragments such as H attached to C0(sp3) with 1X attached to next carbon and CH3R/CH4 in a molecular structure will be supportive to enhance the inhibition activity. The negative contribution of descriptors GATS5p, BELm2 and nNR2 advocated a lower value of atomic polarizabilities weighted Geary autocorrelations of lag 5, atomic masses weighted lowest eigenvalue n.2 of Burden matrix, and absence of tertiary aliphatic amine functionality in a molecular structure will augment the inhibition activity.

These models have accounted for nearly 88% variance in the observed activities. The values greater than 0.5 of Q^2 index is in accordance to a reasonable robust QSAR model. The pIC_{50} values of training set compounds calculated using Eqs. (10) to (13) have been included in Table 4. The models (10) to (13) are validated with an external test set of 9 compounds listed in Table 4. 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 4.



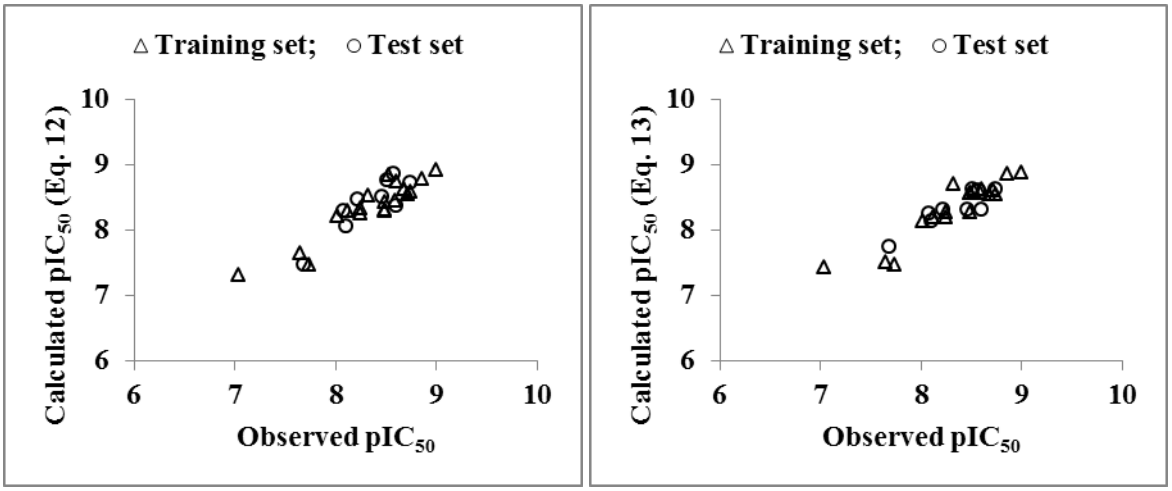


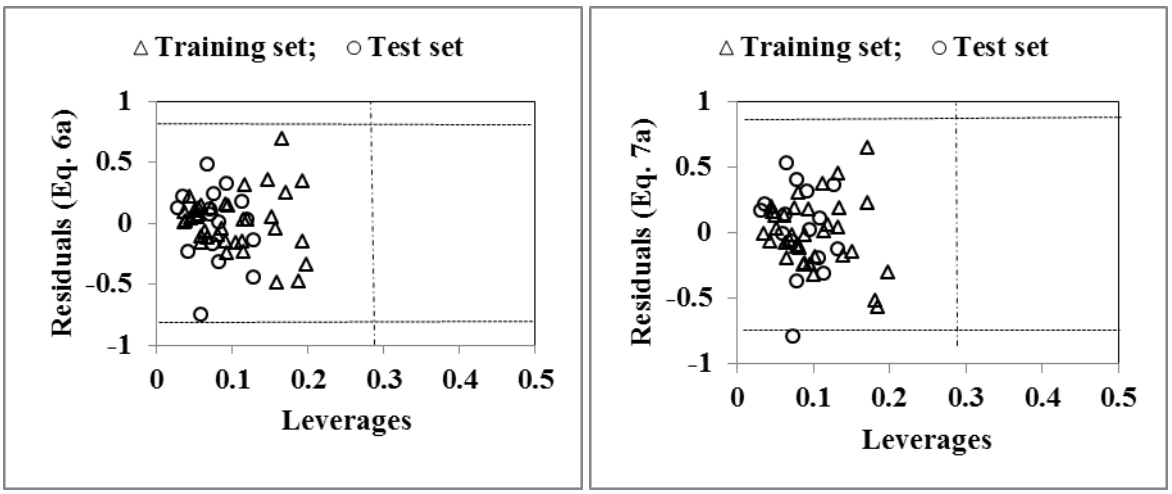
Fig. 4: Plot of observed versus caculated pIC_{50} values for training- and test-set compounds for CDK19 inhibition.

Applicability domain

On analyzing the applicability domain (AD) for the CDK8 inhibitory actions in the Williams plot (Figure 5) of the model based on the whole data set (Table 6), No any compound has been identified as an obvious 'outlier' for the CDK8 inhibitory activity if the limit of normal values for the Y outliers (response outliers) was set as 3×(standard deviation) units. None of the compound was found to have leverage (h) values greater than the threshold leverage (h*). For both the training-set and test-set, the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data. Furthermore, all of the compounds were within the applicability domain of the proposed model and were evaluated correctly.

Table 6: Models derived for the whole data set (n = 53) in descriptors identified through CP-MLR for CDK8 inhibitory actions.

Model	r	s	F	Q^2_{LOO}	Eq.
pIC_{50} = 7.972 +0.457(0.158)Mv +1.086(0.190)IC1 –0.920(0.141)GGI10 –0.913(0.174)MATS3v	0.893	0.270	47.247	0.738	(6a)
pIC_{50} = 8.154 +0.781(0.143)Mv +0.685(0.142)IC2 –0.894(0.151)GGI10 –1.075(0.184)MATS3v	0.877	0.287	40.366	0.704	(7a)
pIC_{50} = 8.360 +0.952(0.128)nO –0.555(0.152)GGI10 –2.032(0.209)GATS1v +0.997(0.177)GATS3p	0.887	0.277	44.367	0.732	(8a)
pIC_{50} = 9.436 +0.741(0.123)nO –0.590(0.154)GGI10 –0.988(0.180)MATS3v –1.845(0.215)GATS1v	0.885	0.279	43.388	0.726	(9a)



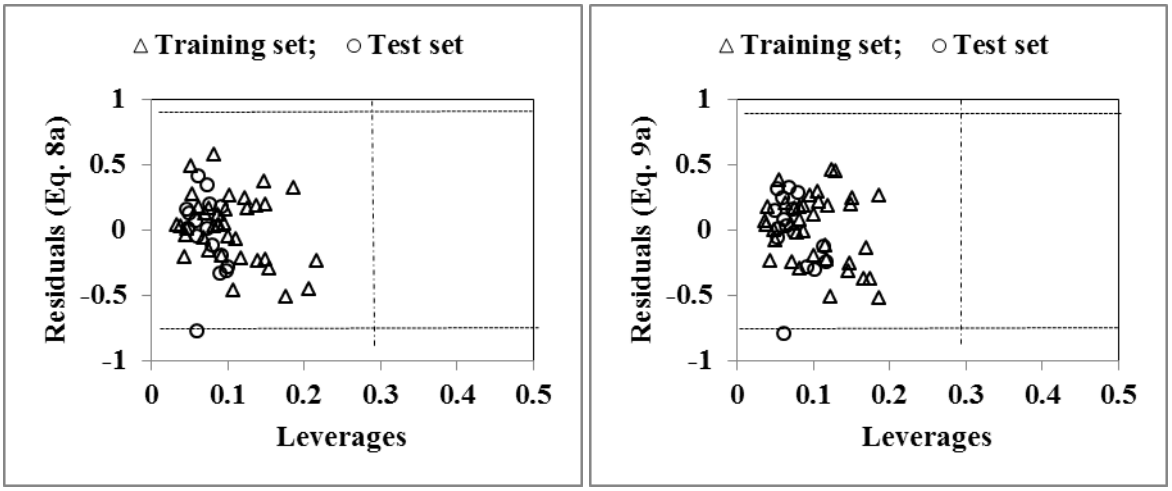


Fig. 5: Williams plot for the training-set and test-set for CDK8 inhibition activity of compounds in Table 1. The horizontal dotted line refers to the residual limit (±3×standard deviation) and the vertical dotted line represents threshold leverage h* (=0.283).

CONCLUSION

The CDK8 and CDK19 inhibition activity of pyridine derivatives has been quantitatively analyzed in terms of Dragon descriptors. The statistically validated quantitative structure-activity relationship (QSAR) models provided rationales to explain the inhibition activities of these congeners. The descriptors identified through combinatorial protocol in multiple linear regression (CP-MLR) analysis for the CDK8 inhibitory activity have highlighted the role of 1st and 2nd order information content indices (IC1 and IC2, respectively), 2D Petitjean shape index (PJI2), Galvez topological charge index of 10th order (GGI10) and 1st order mean topological charge index (IGI1) to rationalize the activity. Atomic properties such as atomic van der Waals volumes and polarizabilities in terms of Mv, and Moran and Geary autocorrelations (MATS3v, GATS1v and GATS3p) in addition to the number of oxygen atoms (nO) have shown prevalence to model the CDK8 inhibitory activity. PLS analysis has also corroborated the dominance of CP-MLR identified descriptors. 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.

The derived QSAR models for the CDK19 inhibitory activity have revealed that the structural information content of 2nd order neighborhood symmetry (descriptor SIC2), atomic polarizabilities weighted Geary autocorrelations of lag 5 and 7 (descriptors GATS5p and GATS7p) and atomic masses weighted lowest eigenvalue n.2 of Burden matrix (descriptor BELm2) played a pivotal role in rationalization of CDK19 inhibition activity of titled compounds. Additionally, presence of H attached to C0(sp3) with 1X attached to next carbon (descriptor H-052) and CH3R/CH4 (descriptor C-001) type structural fragment and lesser number or absence of tertiary aliphatic amines functionality (descriptor nNR2) are also predominant to explain CDK19 inhibition actions of pyridine derivatives.

ACKNOWLEDGEMENT

Authors are thankful to their Institution for providing necessary facilities to complete this work.

FUNDING

Nil.

AUTHORS CONTRIBUTIONS

All the authors have contributed equally.

CONFLICT OF INTERESTS

Declared none.

REFERENCES

1. Allen BL, Taatjes DJ. The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol, 2015; 16: 155-166.
2. Carlsten JO, Zhu X, Gustafsson CM. The multitalented Mediator complex. Trends Biochem Sci., 2013; 38: 531-537.
3. Kim S, Xu X, Hecht A, Boyer TG. Mediator is a transducer of Wnt/beta-catenin signaling. J Biol Chem., 2006; 281: 14066-14075.
4. Schiano C, Casamassimi A, Vietri MT, Rienzo M, Napoli C. The roles of mediator complex in cardiovascular diseases. Biochim Biophys Acta Gene Regul Mech., 2014; 1839: 444-451.
5. Schiano C, Casamassimi A, Rienzo M, de Nigris F, Sommese L and Napoli C. Involvement of Mediator complex in malignancy. Biochim Biophys Acta Rev Cancer, 2014; 1845: 66-83.
6. Galbraith MD, Donner AJ, Espinosa JM. CDK8: a positive regulator of transcription. Transcription, 2010; 1: 4-12.

7. Tsutsui T, Fukasawa R, Tanaka A, Hirose Y, Ohkuma Y. Identification of target genes for the CDK subunits of the Mediator complex. Genes Cells, 2011; 16: 1208-1218.
8. Rickert P, Seghezzi W, Shanahan F, Cho H, Lees E. Cyclin C/CDK8 is a novel CTD kinase associated with RNA polymerase II. Oncogene, 1996; 12: 2631-2640.
9. Morris EJ, Ji JY, Yang F, Di Stefano L, Herr A, Moon NS, Kwon EJ, Haigis KM, Naar AM, Dyson NJ. E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8. Nature, 2008; 455: 552-556.
10. Firestein R, Shima K, Nosho K, Irahara N, Baba Y, Bojarski E, Giovannucci EL, Hahn WC, Fuchs CS, Ogino S. CDK8 expression in 470 colorectal cancers in relation to beta-catenin activation, other molecular alterations and patient survival. Int J Cancer, 2010; 126: 2863-2873.
11. Kim MY, Han SI, Lim SC. Roles of cyclin-dependent kinase 8 and beta-catenin in the oncogenesis and progression of gastric adenocarcinoma. Int J Oncol, 2011; 38: 1375-83.
12. Porter DC, Liang J, Kaza V, Chumanevich AA, Altillia S, Farmaki E, Chen M, Schools GP, Chatzistamou I, Pena MM, Friedhoff LT, Wentland MP, Broude E, Kiaris H, Roninson IB. Targeting the seed and the soil of cancers with selective small molecule inhibitors of CDK8/19: Chemopotentiating, chemopreventive, anti-invasive and anti- metastatic activities. Presented at the AACR Annual Meeting 2014, San Diego, CA, 2014.
13. Kapoor A, Goldberg MS, Cumberland LK, Ratnakumar K, Segura MF, Emanuel PO, Menendez S, Vardabasso C, Leroy G, Vidal CI, Polsky D, Osman I, Garcia BA, Hernando E, Bernstein E. The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. Nature, 2010; 468: 1105-1109.
14. Adler AS, McCleland ML, Truong T, Lau S, Modrusan Z, Soukup TM, Roose-Girma M, Blackwood EM, Firestein R. CDK8 maintains tumor dedifferentiation and embryonic stem cell pluripotency. Cancer Res., 2012; 72: 2129-2139.
15. Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, Freed E, Ligon AH, Vena N, Ogino S, Chheda MG, Tamayo P, Finn S, Shrestha Y, Boehm JS, Jain S, Bojarski E, Mermel C, Barretina J, Chan JA, Baselga J, Tabernero J, Root DE, Fuchs CS, Loda M, Shivdasani RA, Meyerson M, Hahn WC. CDK8 is a colorectal cancer oncogene that regulates betacatenin activity. Nature, 2008; 455: 547-551.
16. Cee VJ, Chen DY, Lee MR, Nicolaou KC. Cortistatin A is a high-affinity ligand of protein kinases ROCK, CDK8, and CDK11. Angew Chem Int Ed., 2009; 48: 8952-8957.
17. Shair MD. Cortistatin analogues and syntheses and uses thereof. WO2015100420 A1, 2015.
18. Pelish HE, Liau BB, Nitulescu II, Tangpeerachaikul A, Poss ZC, Da Silva DH, Caruso BT, Arefolov A, Fadeyi O, Christie AL, Du K, Banka D, Schneider EV, Jestel A, Zou G, Si C, Ebmeier CC, Bronson RT, Krivtsov AV, Myers AG, Kohl NE, Kung AL, Armstrong SA, Lemieux ME, Taatjes DJ, Shair MD. Mediator kinase inhibition further activates super enhancer-associated genes in AML. Nature, 2015; 526: 273-276.
19. Schneider EV, Bottcher J, Blaesse M, Neumann L, Huber R, Maskos K. The structure of CDK8/CycC implicates specificity in the CDK/cyclin family and reveals interaction with a deep pocket binder. J Mol Biol., 2011; 412: 251-266.
20. Schneider EV, Bottcher J, Huber R, Maskos K, Neumann L. Structure-kinetic relationship study of CDK8/CycC specific compounds. Proc Natl Acad Sci USA, 2013; 110: 8081-8086.
21. Dale T, Clarke PA, Esdar C, Waalboer D, Adeniji-Popoola O, Ortiz-Ruiz M, Mallinger A, Samant RS, Czodrowski P, Musil D, Schwarz D, Schneider K, Stubbs M, Ewan K, Fraser E, TePoele R, Court W, Box G, Valenti M, de Haven Brandon A, Gowan S, Rohdich F, Raynaud F, Schneider R, Poeschke O, Blaukat A, Workman P, Schiemann K, Eccles SA, Wienke D, Blagg J. A selective chemical probe for exploring the role of CDK8 and CDK19 in human disease. Nat Chem Biol, 2015; 11: 973-980.
22. Roninson IB, Porter DC, Wentland MP. CDK8/CDK19 selective inhibitors and their use in anti-metastatic and chemopreventative methods for cancer. WO2013116786 A1, 2013.
23. Hu H, Jiang M, Jin T, Niu R, Wang J, Yang S, Yuan T, Zhou C, Wang M, Zhou Z. Novel phenyl-pyridine/pyrazine amides for the treatment of cancer. WO2014029726 A1, 2014.
24. Rzymski T, Zarebski A, Dreas A, Osowska K, Kucwaj K, Fogt J, Cholody M, Galezowski M, Czardybon W, Horvath R, Wiklik K, Milik M, Brzózka K. Substituted tricyclic benzimidazoles as kinase inhibitors. WO2014072435 A1, 2014.
25. Romero DL, Chaudhary D, Robinson S, Masse CE, Morin MJ. CDK8 inhibitors and uses thereof. WO2014194245 A3, 2014.
26. Mallinger A, Crumpler S, Pichowicz M, Waalboer D, Stubbs M, Adeniji-Popoola O, Wood B, Smith E, Thai C, Henley AT, Georgi K, Court W, Hobbs S, Box G, Ortiz-Ruiz MJ, Valenti M, de Haven Brandon A, TePoele R, Leuthner B, Workman P, Aherne W, Poeschke O, Dale T, Wienke D, Esdar C, Rohdich F, Raynaud F, Clarke PA, Eccles SA, Stieber F, Schiemann K, Blagg J. Discovery of potent, orally bioavailable, small-molecule inhibitors of WNT signaling from a cell-based pathway screen. J Med Chem., 2015; 58: 1717-1735.A
27. Mallinger A, Schiemann K, Rink C, Stieber F, Calderini M, Crumpler S, Stubbs M, Adeniji-Popoola O, Poeschke O, Busch M, Czodrowski P, Musil D, Schwarz D, Ortiz- Ruiz M-J, Schneider R, Thai C, Valenti M, de Haven Brandon A, Burke R, Workman P, Dale T, Wienke D, Clarke PA, Esdar C, Raynaud FI, Eccles SA, Rohdich F, Blagg J.

Discovery of potent, selective, and orally bioavailable small-molecule modulators of the Mediator complex-associated kinases CDK8 and CDK19. J Med Chem., 2016; 59: 1078- 1101.
28. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. http://www.cambridgesoft.com
29. Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy. http//www.talete.mi.it/dragon.htm
30. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: analysis of Selwood et al. Data Set-a case study. QSAR Comb Sci., 2003; 22: 583-595.
31. Sharma S, Prabhakar YS, Singh P, 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.
32. Sharma S, Sharma BK, Sharma SK, Singh P, Prabhakar YS. Topological descriptors in modeling the agonistic activity of human A3 adenosine receptor ligands: The derivatives of 2-Chloro-N^6-substituted-4'-thioadenosine-5'-uronamide. Eur J Med Chem., 2009; 44: 1377-1382.
33. Sharma BK, Pilania P, Singh P, 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.
34. Sharma BK, Singh P, Sarbhai K, 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.
35. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics, 1978; 20: 397-405.
36. Kettaneh N, Berglund A, Wold S. PCA and PLS with very large data sets. Comput Stat Data Anal, 2005; 48: 69-85.
37. Stahle L, Wold S. Multivariate data analysis and experimental design. In: Ellis GP, West WB. Eds., Biomedical research. Progress in medicinal chemistry. Elsevier Science Publishers, BV, Amsterdam, 1988; 25: 291-338.
38. So S-S, 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.
39. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure–activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3- thiazolidin-4-ones. QSAR Comb Sci., 2004; 23: 234-244.
40. Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci., 2007; 26: 694-701.