GSC Biological and Pharmaceutical Sciences
eISSN: 2581-3250 CODEN (USA): GBPSC2
Cross Ref DOI: 10.30574/gscbps
Journal homepage: https://gsconlinepress.com/journals/gscbps/

(RESEARCH ARTICLE)

Check for updates

Quantitative structure-activity relationship study on the CDK2 inhibitory activity of 6-substituted 2-arylaminopurines

Dinesh Kumar Meena [1], Brij Kishore Sharma [1,*] and Raghuraj Parihar [2]

[1] *Department of Chemistry, Government College, Bundi-323 001, India.*
[2] *Department of Chemistry, Government College, Kota-324 001, India.*

GSC Biological and Pharmaceutical Sciences, 2022, 20(03), 107–119

Publication history: Received on 30 July 2022; revised on 03 September 2022; accepted on 05 September 2022

Article DOI: https://doi.org/10.30574/gscbps.2022.20.3.0344

Abstract

QSAR study has been carried out on the CDK2 inhibitory activity of 6-substituted 2-arylaminopurines in 0D- to 2D-Dragon descriptors. The derived QSAR models have revealed that the reciprocal hyper-detour index (descriptor Rww) and path/walk 5 Randic shape index (descriptor PW5) played a pivotal role in rationalization of CDK2 inhibition activity of titled compounds. Molecular weight (MW), mean atomic volume scaled on Carbon atom (Mv) and atomic properties such as mass and atomic Sanderson electronegativity in terms of atomic properties weighted descriptors MATS1m, MATS3e, MATS4e, GATS3e and GATS8e, certain atom centred fragments such as H attached to C0(sp3) no X attached to next C (descriptor H-046),R--CH--X (descriptor C-027) and R--CX--X (descriptor C-029) are also predominant to explain CDK2 inhibition actions of 6-substituted 2-arylaminopurines.

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.

Keywords: QSAR; CDK2 inhibitors; Combinatorial protocol in multiple linear regression (CP-MLR) analysis; PLS; Dragon descriptors; 6-Substituted 2-arylaminopurines.

1. Introduction

The cyclin-dependent kinase (CDK) not only plays a vital role in the eukaryotic cell cycle regulation but also have imperative role in apoptosis, transcription, differentiation and neuronal function [1,2]. The main feature of human cancer is the deregulation of the cell cycle. This deregulation is often has association with aberrant CDK activity via the mechanisms which include mutation or overexpression of CDKs, and modulation of the CDK activity through mutations in genes encoding proteins [3-5]. The common feature of most of the human tumors are the overexpression of cyclin E and/or suppression of p27Kip1 [6,7]. The partner cyclins A and E when associated with CDK2 causes activation of CDK2 on the other hand, the endogenous proteins p21Cip1 and p27Kip1 inhibits CDK2. Considering, the inhibition of CDKs as a potential therapeutic target a large number of ATP-competitive inhibitors have been reported [5-9]. The poor kinase selectivity and uncertainty as to which CDK constitutes the most appropriate therapeutic target hindered the development of CDK inhibitors, clinically [3].

The knockdown experiments where the loss of CDK2 failed to induce cell cycle arrest in tumour cell lines [10] and mouse knockout experiments with viable animals [11,12] originally questioned the validity of CDK2 as a cancer therapeutic target. It is however also anticipated the utility of CDK inhibitors where the cancer has addiction to enhanced CDK

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

activity or in cases with identified synthetic lethalities. Studies, with maintained CDK2 expression and inhibited kinase activity, based on chemical genetic approach are supportive to CDK2 as a valid cancer target. A marked growth inhibition was observed by the selective CDK2 inhibition in human cancer cells having transformation with various oncogenes [13]. CDK2-selective inhibitors in combination with phosphatidylinositol-3-kinase induced apoptosis in malignant glioma xenografts through a synthetic-lethal interaction [14]. The evidence of CDK2 implications in BRCA-deficient cancers [15], euroblastoma [16] and ovarian cancer [17] with supportive clinical data [18] put CDK2 inhibitors as cancer therapeutic agents [19,20]. The cell toxicity caused by inhibition of CDK1, which have close structural homology to CDK2, must be avoided.

A variety of selective ATP-competitive CDK2 inhibitors based on purine scaffold, utilizing a structure-lead approach, has been reported [21-23]. The phosphorylation of RNA polymerase II by CDKs leads to transcriptional regulation [2,24].

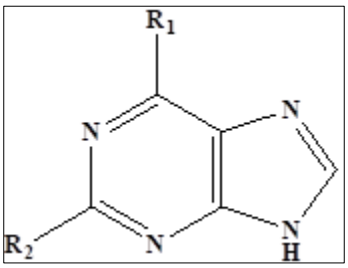
The pharmacological effects of selective CDK2 inhibitors such as seliciclib, dinaciclib and SNS-032 were concealed by the modulation of protein synthesis by off-target inhibition of CDK7/CDK9 [3, 25, 26]. The inhibitor interactions within the ATP-binding site has been revealed by the crystal structures of purines complexed with T160-phosphorylated CDK2-cyclin A [21-23, 27, 28]. A novel series of compounds based on purine scaffold, as selective CDK2 inhibitors, has been reported by Coxon *et al.* [29]. The aim of present communication is to establish the quantitative relationships between the reported activities and molecular descriptors unfolding the substitutional changes in titled compounds.

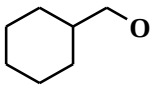
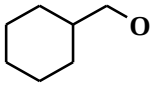
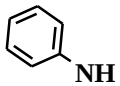
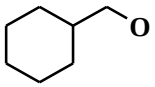
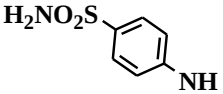
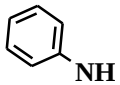
2. Material and methods

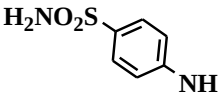
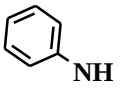
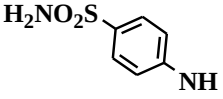
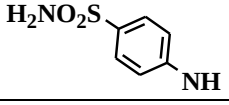
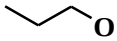
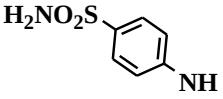
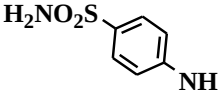
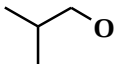
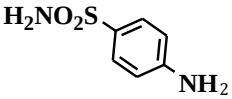
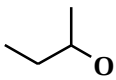
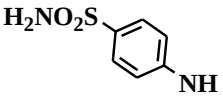
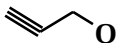
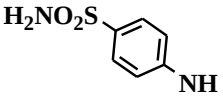
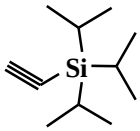
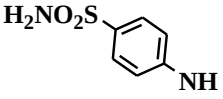
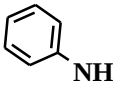
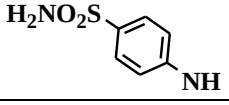
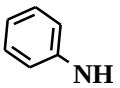
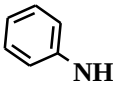
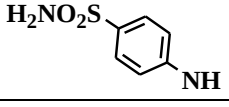
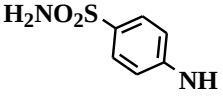
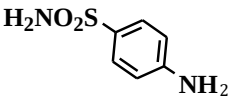
2.1. Biological actions and theoretical molecular descriptors

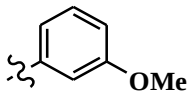
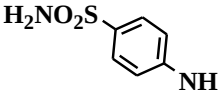
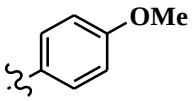
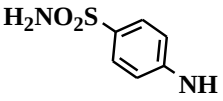
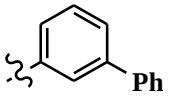
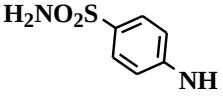
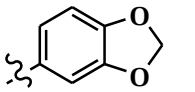
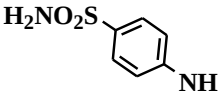
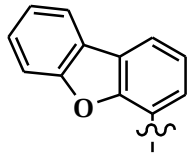
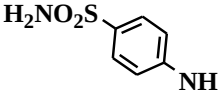
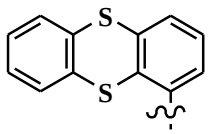
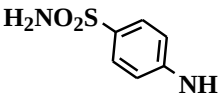
The reported twentyeight purine derivatives are considered as the data set for present study [29]. These derivatives were evaluated for their CDK2 inhibitory activities and were reported as IC_{50}. The reported CDK2 activity on molar basis (as pIC_{50}) along with the structures of these analogues is 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 Structures, observed and calculated CDK2 inhibitory activities of purine derivatives



Cpd.	R_1	R_2	pIC50(M)[a]			
			Obsd.	Eq. (2)	Eq. (3)	PLS
1		NH_2	4.77	5.71	5.13	5.71
2			6.01	5.76	6.08	5.75
3[b]			8.30	7.81	7.32	7.35
4	H	NH_2	5.40	5.43	5.02	5.34
5	H		4.21	4.28	4.75	4.35

6[b]	H		5.82	6.08	6.39	6.36
7	OH(C=O)		4.47	3.89	3.78	3.80
8[b]	OH(C=O)		4.28	5.82	6.11	6.13
9[b]			7.59	8.02	7.29	8.00
10			8.10	7.94	7.48	7.90
11			8.00	8.41	8.32	8.61
12			8.52	7.96	7.85	7.98
13			9.00	8.36	8.54	8.60
14			7.72	7.99	7.09	7.48
15			7.77	8.54	8.86	7.90
16			4.68	4.79	5.28	4.80
17			6.08	6.51	6.88	6.90
18			5.19	5.36	6.00	5.14
19[b]			5.05	4.17	5.40	4.61
20			6.66	6.63	6.86	7.16
21[b]			7.72	6.65	6.89	7.20
22			7.62	6.66	6.30	6.60

23			7.41	6.55	7.06	7.08
24			7.28	7.05	7.07	7.19
25			7.36	6.91	6.55	6.56
26[b]			6.24	7.61	7.16	7.43
27			5.85	6.35	6.53	6.34
28			5.17	6.21	5.85	6.07

[a]IC_{50} on molar basis, Taken from reference [29]; [b]Compound included in test set.

Table 2 Descriptor classes used for the modeling of CDK2 inhibitory activity of purine derivatives

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 [31].

The structures of the all the compounds (listed in Table 1) were drawn in 2D ChemDraw [30] and subjected to energy minimization in the MOPAC using the AM1 procedure for closed shell system after converting these into 3D modules. The energy minimization was carried out to attain a well defined conformer relationship among the congeners under study. The 0D- to 2D-molecular descriptors of titled compounds was computed using DRAGON software [31]. This software offers a large number of descriptors corresponding to ten different classes of 0D- to 2D-descriptor modules. The different descriptor classes include the constitutional, topological, molecular walk counts, BCUT descriptors, Galvez topological charge indices, 2D-autocorrelations, functional groups, atom-centered fragments, empirical descriptors and the properties describing descriptors. These descriptors offer characteristic structural information specific to the descriptor class. The definition and scope of these descriptor's classes is given in Table 2.

A total number of 494 descriptors, belonging to 0D- to 2D- modules, have been computed to obtain most appropriate models describing the biological activity. Prior to model development procedure, all those descriptors that are inter-correlated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This procedure has reduced the total descriptors from 494 to 77 as relevant ones to explain the biological actions of titled compounds.

2.2. Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR) [32-36] and partial least squares (PLS) [37-39] 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 [40,41]. In view of this, to find out any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been subjected to randomization test [42,43] by repeated randomization (100 simulation runs) of the biological responses. The datasets with randomized response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test its prediction and generalization within the study domain. For each model, derived by involving n data points, a number of statistical parameters such as r (the multiple correlation coefficient), s (the standard deviation), F (the F ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to access its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5. Additional statistical parameters such as, the Akaike's information criterion, AIC [44,45], the Kubinyi function, FIT [46,47] and the Friedman's lack of fit, LOF [48], have also been calculated to further validate the derived models. The AIC takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The FIT, closely related to the F-value, proved to be a useful parameter for assessing the quality of the models. A model which is derived in k independent descriptors, its F-value will be more sensitive if k is small while it becomes less sensitive if k is large. The FIT, on the other hand, will be less sensitive if k is small whereas it becomes more sensitive if k is large. The model that produces the lowest AIC value and highest FIT value is considered potentially the most useful and the best. The LOF factor takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large number of parameters.

2.3. Applicability domain

The usefulness of a model is based on its accurate prediction ability for new congeners. A model is valid only within its training domain and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain (AD) is evaluated by the leverage values for each compound [49]. A Williams plot (the plot of standardized residuals versus leverage values (h)) is constructed, which can be used for a simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the AD is established inside a squared area within $\pm x$ standard deviations and a leverage threshold h^*, which is generally fixed at $3(k + 1)/n$ (n is the number of training set compounds and k is the number of model parameters), whereas x = 2 or 3. If the compounds have a high leverage value ($h > h^*$), then the prediction is not trustworthy. On the other hand,

when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training set compounds.

3. Results and discussion

3.1. QSAR results

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modeling study, 3 compounds have been included in the test set for the validation of the models derived from 14 training set compounds. A total number of 77 significant descriptors from 0D- to 2D- classes have been subjected to CP-MLR analysis with default "filters" set in it. Statistical models in one, two and three descriptors have been derived to achieve the best relationship correlating CDK2 inhibitory activity. One model in two descriptors having r^2_{Test}> 0.5, were obtained through CP-MLR. The analysis resulted in to 06 three parameter models which have shared 14 descriptors among them. All these 14 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, physical meaning, average regression coefficient and incidence[b]

Descriptor class, average regression coefficientand (incidence)	
Descriptor class	**Descriptor (physical meaning), avg reg coeff (incidence)**
Constitutional descriptors (CONST):	MW (molecular weight), 3.272(1);AMW (average molecular weight) -2.835(1); Mv (mean atomic volume scaled on Carbon atom), -2.415(1)
Topological descriptors (TOPO):	Rww (reciprocal hyper-detour index), 5.067(2); PW5 (path/walk 5 Randic shape index), -6.348(1)
2D autocorrelations (2D-AUTO):	MATS1m (Moran autocorrelation - lag 1 / weighted by atomic masses), -3.955(1); MATS6v, (Moran autocorrelation - lag 6 / weighted by atomic van der Waals volumes), -4.391(1); MATS3e (Moran autocorrelation lag-3/ weighted by atomic Sanderson electronegativities), -3.922(1); MATS3e (Moran autocorrelation lag-4/ weighted by atomic Sanderson electronegativities), -5.515(1), GATS3e (Geary autocorrelation of lag-2/ weighted by atomic Sanderson electronegativities), 3.165(2);GATS8e (Geary autocorrelation of lag-8/ weighted by atomic Sanderson electronegativities), -2.719(1)
Atom centred fragments (ACF):	C-027 (R--CH--X), -2-210(2); C-029 (R--CX--X), 1.764(2); H-046 (H attached to C0(sp3) no X attached to next C), 1.428(1)

[a]The descriptors are identified from the three parameter models for activity emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.773 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 21 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.

The representative model in two descriptors and highly significant models in three descriptors are given below.

$$pIC_{50} = 3.898 + 3.431(0.735)MATS6e + 1.659(0.434)C\text{-}029$$

$$n = 21, r = 0.799, s = 0.927, F = 15.917, Q^2_{LOO} = 0.509, Q^2_{L5O} = 0.651$$

$$r^2_{Test} = 0.517, FIT = 1.273, LOF = 1.125, AIC = 1.147 \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$pIC_{50} = 5.427 + 4.879(0.664)Rww\text{-}2.719(0.619)GATS8e + 1.461(0.296)C\text{-}029$$

$$n = 21, r = 0.922, s = 0.612, F = 32.429, Q^2_{LOO} = 0.728, Q^2_{L5O} = 0.742$$

$$r^2_{Test} = 0.507, FIT = 3.242, LOF = 0.595, AIC = 0.551 \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

pIC_{50} = 2.562 + 5.255(0.790)Rww +3.198(0.733)GATS3e + 1.428(0.566)H-046

n = 21, r = 0.889, s = 0.725, F = 21.511, Q^2_{LOO} = 0.611, Q^2_{L50} = 0.575

r^2_{Test} = 0.529, FIT = 2.151, LOF = 0.834, AIC = 0.773……..…(3)

where; n, r, s and F represent respectively the number of data points,the multiple correlation coefficient, the standard deviation and the F-ratio between the variances of calculated and observed activities. In above 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.

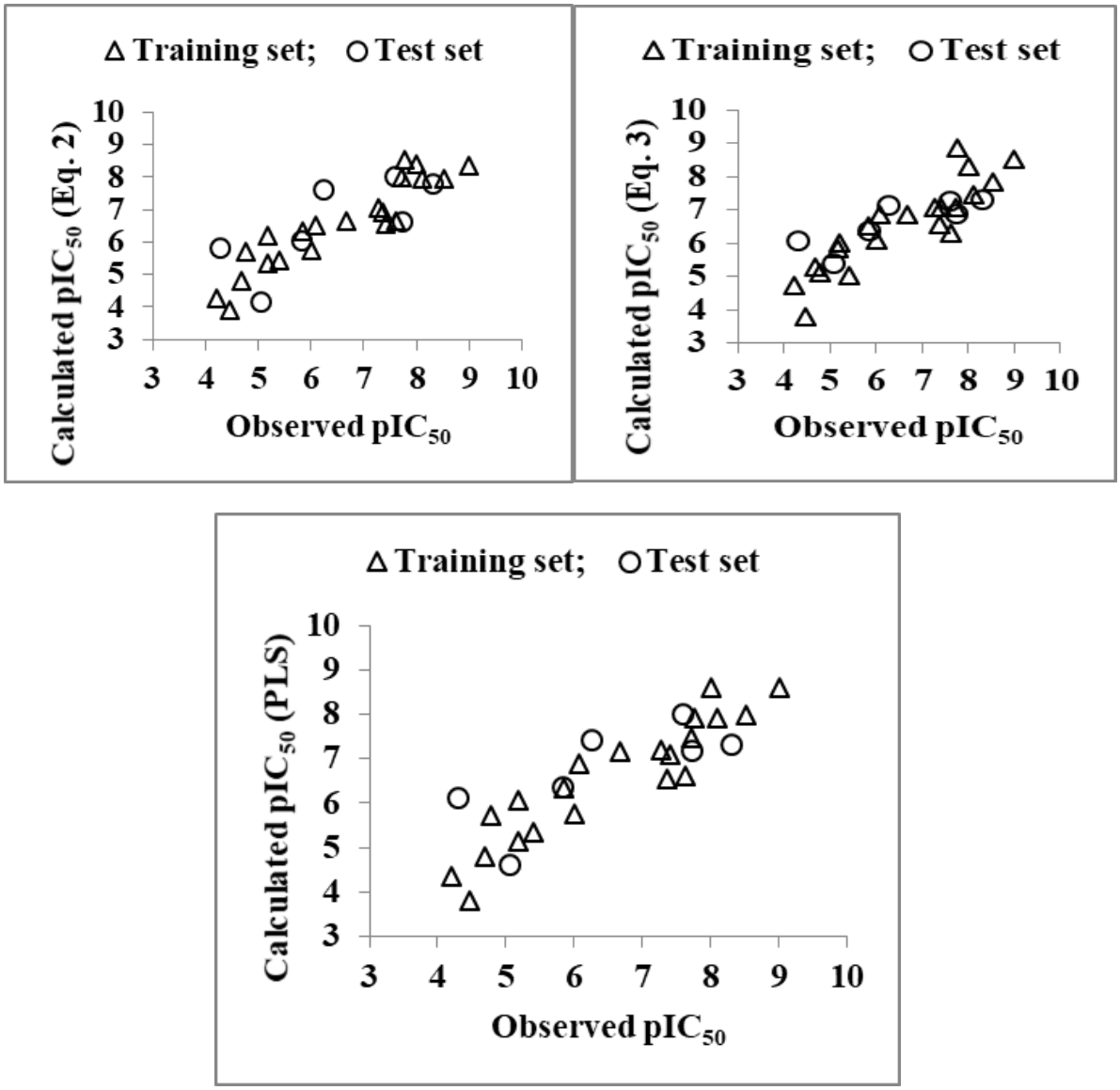


Figure 1 Plot of observed and calculated pIC50 values of training- and test-set compounds for CDK2 inhibition

The descriptor Rww participated in above models is the topological descriptor representing reciprocal hyper-detour index. The positive influence of descriptors Rww on the activity suggested that higher value of it would be beneficiary to the activity. The descriptor MATS6e, GATS3e and GATS8e are 2D-autocorrelations. Descriptors MATS6e and GATS3e showed positive contribution and descriptor GATS8e negative contribution to the activity. Thus, lower value of atomic Sanderson electronegativities weighted Geary autocorrelations of lag-8 and higher values of atomic Sanderson

electronegativities weighted Moran autocorrelations of lag-6 and atomic Sanderson electronegativities weighted Geary autocorrelations of lag-3 would be favorable to the activity. Additionally, positive sign of atom centered fragment class descriptors C-029 and H-046 representing R--CX--X and H attached to C0(sp3) no X attached to next C, respectively, advocated that more number of such structural fragments in a molecule would be supportive to elevated activity.

These models have accounted for nearly 85% 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. (2) and (3) have been included in Table 1. The models (2) and (3) are validated with an external test set of 7 compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the same is reported in Table 1. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

A partial least square (PLS) analysis has been carried out on these 14 CP-MLR identified descriptors (Table 3) to facilitate the development of a "single window" structure–activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, two components are found to be the optimum for these 14 descriptors and they explained 85.74% variance in the activity. The MLR-like PLS coefficients of these 14 descriptors are given in Table 4.

Table 4 PLS and MLR-like PLS models from the 14 descriptors of three parameter CP-MLR models for CDK2 inhibitory activities

A: PLS equation				
PLS components		**PLS coefficient (s.e.)[a]**		
Component-1		-0.777(0.076)		
Component-2		0.180(0.088)		
Constant		6.536		
B: MLR-like PLS equation				
S. No.	**Descriptor**	**MLR-like coefficient[b]**	**(fraction contribution)[c]**	**Order**
1	MW	0.931	0.082	6
2	AMW	0.089	0.008	13
3	Mv	-0.809	-0.086	4
4	Rww	1.830	0.135	1
5	PW5	-1.073	-0.083	5
6	MATS1m	-0.573	-0.054	10
7	MATS6v	-0.715	-0.051	11
8	MATS3e	-0.760	-0.082	7
9	MATS4e	-0.713	-0.070	9
10	GATS3e	1.132	0.098	3
11	GATS8e	-0.590	-0.048	12
12	C-027	-0.734	-0.078	8
13	C-029	0.707	0.121	2
14	H-046	0.038	0.004	14
		Constant	7.164	
C: PLS regression statistics			**Values**	
n			21	

r	0.926
s	0.584
F	53.841
FIT	4.307
LOF	0.446
AIC	0.454
Q^2_{LOO}	0.814
Q^2_{L5O}	0.825
r^2_{Test}	0.507

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

For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is also given in Figure 1. Figure 2 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity.

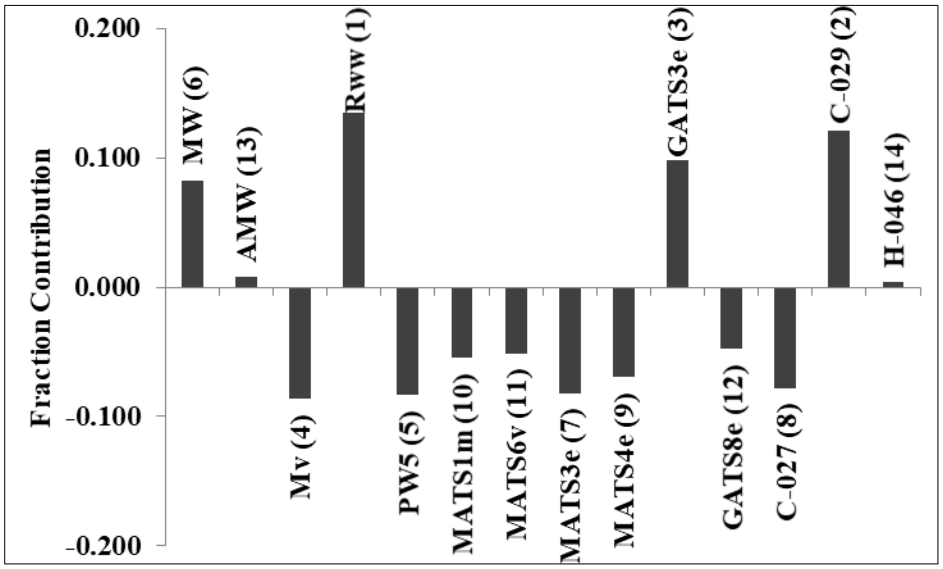


Figure 2 Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 14 CP-MLR identified descriptors (Table 3) associated with CDK2 inhibitory activity of purine derivatives

The PLS analysis has suggested Rww as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 4 in Table 4; Figure 2). The other nine descriptors in decreasing order of significance are C-029, GATS3e, Mv, PW5, MW, MATS3e, C-027, MATS4e and MATS1m. Descriptor Rww, C-029 and GATS3e are part of Eqs. (2) and (3) and convey same inference in the PLS model as well.

It is inferred from the PLS analysis that a higher value of descriptors MW (molecular weight) and lower values of descriptors Mv (mean atomic volume scaled on Carbon atom), PW5 (path/walk 5 Randic shape index), MATS3e and MATS4e (atomic Sanderson electronegativities weighted Moran autocorrelations of lag 3 and 4, respectively) and MATs1m ((Moran autocorrelation of lag-1/ weighted by atomic masses) in addition to the absence or lower number of R--CH--X type atom centered fragment (descriptor C-027) in a molecular structure would be advantageous to the activity. It is also observed that PLS model from the dataset devoid of CP-MLR identified 14 descriptors (Table 3) is inferior in explaining the activity of the analogues.

3.2. Applicability domain (AD)

On analyzing the model AD in the Williams plot, shown in Figure 3, of the model based on the whole dataset (Table 5), it has appeared that none of the compound was identified as an obvious outlier for the CDK2 inhibitory activity if the limit of normal values for the *Y* outliers (response outliers) was set as 2.5 times of standard deviation units. None of the compounds listed in Table 1 were found to have leverage (h) values greater than the threshold leverage (h*=0.571). 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) in descriptors identified through CP-MLR

Model	r	s	F	Q2LOO	Eq.
pIC50 = 5.416 + 4.478(0.761)Rww -2.479(0.606)GATS8e + 1.534(0.302)C-029	0.879	0.727	27.459	0.697	(2a)
pIC50 = 2.136 + 5.275(0.795)Rww +3.658(0.724)GATS3e + 2.123(0.505)H-046	0.865	0.766	23.983	0.629	(3a)

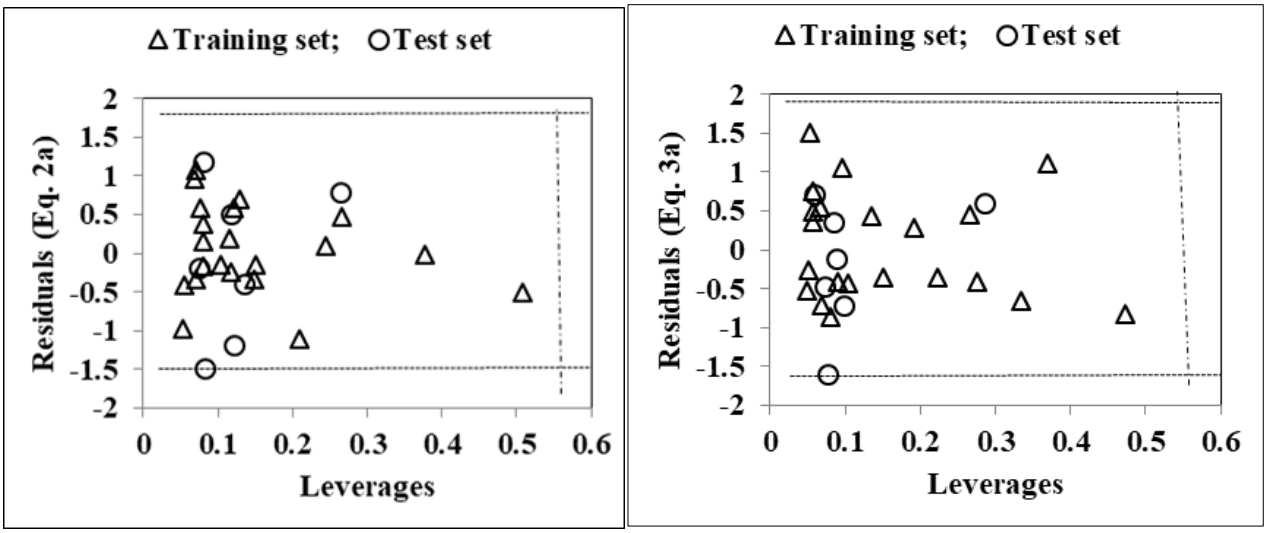


Figure 3 Williams plot for the training-set and test- set compounds for MMP-13 inhibitory activity. The horizontal dotted line refers to the residual limit (±2×standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.571)

4. Conclusion

QSAR study has been carried out on the CDK2 inhibitory activity of 6-substituted 2-arylaminopurines in 0D- to 2D-Dragon descriptors. The derived QSAR models have revealed that the reciprocal hyper-detour index (descriptor Rww) and path/walk 5 Randic shape index (descriptor PW5) played a pivotal role in rationalization of CDK2 inhibition activity of titled compounds. Molecular weight (MW), mean atomic volume scaled on Carbon atom (Mv) and atomic properties such as mass and atomic Sanderson electronegativity in terms of atomic properties weighted descriptors MATS1m, MATS3e,MATS4e, GATS3e and GATS8e, certain atom centred fragments such as H attached to C0(sp3) no X attached to next C (descriptor H-046),R--CH--X (descriptor C-027) and R--CX--X (descriptor C-029) are also predominant to explain CDK2 inhibition actions of 6-substituted 2-arylaminopurines.

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.

Compliance with ethical standards

Acknowledgments

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

Disclosure of conflict of interest

The authors declare no conflict of interest.

References

[1] Lim S and Kaldis P. Cdks, Cyclins and CKIs: Roles Beyond Cell Cycle Regulation. Development 2013, 140, 3079-3093.

[2] Malumbres M. Cyclin-Dependent Kinases. Genome Biol. 2014, 15, 122-132.

[3] Malumbres M and Barbacid M. Cell Cycle, CDKs and Cancer: A Changing Paradigm. Nat. Rev. Cancer 2009, 9, 153-166.

[4] Casimiro MC, Velasco-Velazquez M, Aguirre-Alvarado C and Pestell RG. Overview of Cyclins D1 Function in Cancer and the CDK Inhibitor Landscape: Past and Present. Expert Opin. Investig. Drugs 2014, 23, 295-304.

[5] Sherr CJ, Beach D and Shapiro GI. Targeting CDK4 and CDK6: From Discovery to Therapy. Cancer Discovery 2016, 6, 353-367.

[6] Bruyere C and Meijer L. Targeting Cyclin-Dependent Kinases in Anti-Neoplastic Therapy. Curr. Opin. Cell Biol. 2013, 25, 772-779.

[7] Asghar U, Witkiewicz AK, Turner NC and Knudsen ES. The History and Future of Targeting Cyclin-Dependent Kinases in Cancer Therapy. Nat. Rev. Drug Discov. 2015, 14, 130-146.

[8] Bose P, Simmons GL and Grant S. Cyclin-Dependent Kinase Inhibitor Therapy for Hematologic Malignancies. Expert Opin. Investig. Drugs 2013, 22, 723-738.

[9] Aleem E and Arceci RJ. Targeting Cell Cycle Regulators in Hematologic Malignancies. Front. Cell Dev. Biol. 2015, 3, 16-37.

[10] Tetsu O and McCormick F. Proliferation of Cancer Cells Despite CDK2 Inhibition. Cancer Cell 2003, 3, 233-245.

[11] Ortega S, Prieto I, Odajima J, Martin A, Dubus P, Sotillo R, Barbero JL, Malumbres M and Barbacid M. Cyclin-Dependent Kinase 2 is Essential for Meiosis but not for Mitotic Cell Division in Mice. Nat. Genet. 2003, 35, 25-31.

[12] Berthet C, Aleem E, Coppola V, Tessarollo L and Kaldis P. Cdk2 Knockout Mice are Viable. Curr. Biol. 2003, 13, 1775-1785.

[13] Horiuchi D, Huskey NE, Kusdra L, Wohlbold L, Merrick KA, Zhang C, Creasman KJ, Shokat KM, Fisher RP and Goga A. Chemical-Genetic Analysis of Cyclin Dependent Kinase 2 Function Reveals an Important Role in Cellular Transformation by Multiple Oncogenic Pathways. Proc. Natl. Acad. Sci. U. S. A. 2012, 109, E1019-1027.

[14] Cheng CK, Gustafson WC, Charron E, Houseman BT, Zunder E, Goga A, Gray NS, Pollok B, Oakes SA, James CD, Shokat KM, Weiss WA and Fan QW. Dual Blockade of Lipid and Cyclin-Dependent Kinases Induces Synthetic Lethality in Malignant Glioma. Proc. Natl. Acad. Sci. U. S. A. 2012, 109, 12722-12727.

[15] Deans AJ, Khanna KK, McNees CJ, Mercurio C, Heierhorst J and McArthur GA. Cyclin-Dependent Kinase 2 Functions in Normal DNA Repair and is a Therapeutic Target in BRCA1-Deficient Cancers. Cancer Res. 2006, 66, 8219-8226.

[16] Molenaar JJ, Ebus ME, Geerts D, Koster J, Lamers F, Valentijn LJ, Westerhout EM, Versteeg R and Caron HN. Inactivation of CDK2 is Synthetically Lethal to MYCN Over-Expressing Cancer Cells. Proc. Natl. Acad. Sci. U. S. A. 2009, 106, 12968-12973.

[17] Yang L, Fang D, Chen H, Lu Y, Dong Z, Ding HF, Jing Q, Su SB and Huang S. Cyclin-Dependent Kinase 2 is an Ideal Target for Ovary Tumors with Elevated Cyclin E1 Expression. Oncotarget 2015, 6, 20801-20812.

[18] Guha M. Cyclin-Dependent Kinase Inhibitors Move into Phase III. Nat. Rev. Drug Discov. 2012, 11, 892-894.

[19] Galons H, Oumata N, Gloulou O and Meijer L. Cyclin-Dependent Kinase Inhibitors Closer to Market Launch? Expert Opin. Ther. Pat. 2013, 23, 945-963.

[20] Santo L, Siu KT and Raje N. Targeting Cyclin-Dependent Kinases and Cell Cycle Progression in Human Cancers. Semin. Oncol. 2015, 42, 788-800.

[21] Arris CE, Boyle FT, Calvert AH, Curtin NJ, Endicott JA, Garman EF, Gibson AE, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Lawrie AM, Newell DR, Noble ME, Sausville EA, Schultz R and Yu W. Identification of Novel Purine and Pyrimidine Cyclin-Dependent Kinase Inhibitors with Distinct Molecular Interactions and Tumor Cell Growth Inhibition Profiles. J. Med. Chem. 2000, 43, 2797-2804.

[22] Davies T-G, Bentley J, Arris C-E, Boyle FT, Curtin N-J, Endicott J-A, Gibson A-E, Golding B-T, Griffin R-J, Hardcastle I-R, Jewsbury P, Johnson L-N, Mesguiche V, Newell D-R, Noble M-EM, Tucker J-A, Wang L and Whitfield H-J. Structure-Based Design of a Potent Purine-Based Cyclin-Dependent Kinase Inhibitor. Nat. Struct. Biol. 2002, 9, 745-749.

[23] Hardcastle IR, Arris CE, Bentley J, Boyle FT, Chen Y, Curtin NJ, Endicott JA, Gibson AE, Golding BT, Griffin RJ, Jewsbury P, Menyerol J, Mesguiche V, Newell DR, Noble ME, Pratt DJ, Wang LZ and Whitfield HJ. N2-Substituted O6-Cyclohexylmethylguanine Derivatives: Potent Inhibitors of Cyclin-Dependent Kinases 1 and 2. J. Med. Chem. 2004, 47, 3710-3722.

[24] Jonkers I and Lis JT. Getting up to Speed with Transcription Elongation by RNA Polymerase II. Nat. Rev. Mol. Cell Biol. 2015, 16, 167-177.

[25] Fischer PM and Gianella-Borradori A. Recent Progress in the Discovery and Development of Cyclin-Dependent Kinase Inhibitors. Expert Opin. Investig. Drugs 2005, 14, 457-477.

[26] Parry D, Guzi T, Shanahan F, Davis N, Prabhavalkar D, Wiswell D, Seghezzi W, Paruch K, Dwyer MP, Doll R, Nomeir A, Windsor W, Fischmann T, Wang Y, Oft M, Chen T,Kirschmeier P and Lees EM. Dinaciclib (SCH 727965), A Novel and Potent Cyclin-Dependent Kinase Inhibitor. Mol. Cancer Ther. 2010, 9, 2344-2353.

[27] Pratt DJ, Bentley J, Jewsbury P, Boyle FT, Endicott JA and Noble ME. Dissecting the Determinants of Cyclin-Dependent Kinase 2 and Cyclin-Dependent Kinase 4 inhibitor Selectivity. J. Med. Chem. 2006, 49, 5470-5477.

[28] Gibson AE, Arris CE, Bentley J, Boyle FT, Curtin NJ, Davies TG, Endicott JA, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Mesguiche V, Newell DR, Noble MEM, Tucker JA and Whitfield HJ. Probing the ATP Ribose-Binding Domain of Cyclin-Dependent Kinases 1 and 2 with O(6)-Substituted Guanine Derivatives. J. Med. Chem. 2002, 45, 3381-3393.

[29] Coxon CR, Anscombe E, Harnor SJ, Martin MP, Carbain B, Golding BT, Hardcastle IR, Harlow LK, Korolchuk S, Matheson CJ, Newell DR, Noble MEM, Sivaprakasam M, Tudhope SJ, Turner DM, Wang LZ, Wedge SR, Wong C, Griffin RJ, Endicott JA and Cano C. Cyclin-Dependent Kinase (CDK) Inhibitors; Structure-Activity Relationships and Insights into the CDK-2 Selectivity of 6-Substituted 2-Arylaminopurines. J. Med. Chem. 2017, 60, 1746-1767.

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

[31] Dragon software (version 1.11-2001) by Todeschini R.; Consonni V. Milano, Italy.

[32] 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.

[33] 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.

[34] 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.

[35] Sharma, B. K.; Pilania, P.; Singh, P.; Prabhakar, Y. S. 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.

[36] 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.

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

[38] Kettaneh N, Berglund A and Wold S. PCA and PLS with very large data sets. Comput Stat Data Anal 2005, 48, 69–85.

[39] Stahle L and 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.

[40] Topliss JG and Edwards RP. Chance factors in studies of quantitative structure–activity relationships. J. Med. Chem. 1979, 22, 1238-1244.

[41] 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.

[42] So S-S and Karplus M. Three-dimensional quantitative structure–activity relationship frommolecular similarity matrices and genetic neural networks. 1. Method and validation. J. Med. Chem. 1997, 40, 4347-4359.

[43] 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.

[44] Akaike H. Information theory and an extension of the minimum likelihood principle. In:Petrov BN, Csaki F, editors. Second international symposium on information theory. Budapest: Akademiai Kiado; 1973, p 267-281.

[45] Akaike H. A new look at the statistical identification model. IEEE Trans. Autom. Control. 1974, AC-19, 716-723.

[46] Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quant. Struct.-Act. Relat. 1994, 13, 285–294.

[47] 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.

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

[49] Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb. Sci. 2007, 26, 694-701.