

**QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR)
STUDY ON THE CDK2 INHIBITORY ACTIVITY OF 4-
BENZOYLAMINO-1H-PYRAZOLE-3-CARBOXAMIDE DERIVATIVES**

Meena Dinesh Kumar and Sharma Brij Kishore*

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

Article Received on
21 July 2022,

Revised on 11 August 2022,
Accepted on 01 Sept. 2022

DOI: 10.20959/wjpr202212-25468

***Corresponding Author**

Sharma Brij Kishore

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

ABSTRACT

The CDK 2 and 1 inhibition activity of 4-benzoylamino-1H-pyrazole-3-carboxamide 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 CDK2 inhibitory activity have highlighted the role of E-state topological parameter (TIE), path/walk 4-Randic shape index (PW4), atomic masses weighted Moran autocorrelation of lag 6 (MATS6m) and fragment based polar surface area (PSA). Atom

centered fragments, R--CH--X (C-027) and RCO-N</>N-X=X (N-072) have also shown prevalence to model the CDK2 inhibitory activity. The statistics emerged from the test sets have validated the identified significant models. 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. For the CDK1 inhibitory activity, a higher value of descriptors MATS7m (atomic masses weighted Moran autocorrelations of lag 7), SPI (superpendentic index), MSD (Balaban mean square distance index) and MLOGP (Moriguchi octanol-water partition coefficient, logP) will be helpful to augment the activity. On the other hand lower values of descriptors nO (number of oxygen atoms), MATS8v (atomic van der Waals volumes weighted Moran autocorrelations of lag 8), TIE (E-state topological parameter) and JGI4 (mean topological charge index of order 4) advocates that a lower value of these will be beneficial for the activity.

KEYWORDS: QSAR, CDK2 inhibitors, Combinatorial protocol in multiple linear regression (CP-MLR) analysis, Dragon descriptors, 4-benzoylamino-1H-pyrazole-3-carboxamide derivatives.

1. INTRODUCTION

Noteworthy roles in cell proliferation, apoptosis and transcription are played by cyclin dependent kinases (CDKs).^[1] CDKs are serine/threonine protein kinases. Based on the biological function the 20 known CDKs in human cells may be divided into two subfamilies such as cell cycle-related family (e.g. CDK1-6) and transcription-related family (e.g. CDK7-13).^[2] The regulation of kinase activity of CDKs is due to its cyclin regulatory partners and inhibitory proteins (CKIs).

In regulation of cell cycle progression CDK2 plays a critical role. The initiation of S phase of cell cycle is because of the involvement of CDK2-cyclin E in late G1 to make the full phosphorylation of retinoblastoma (Rb), whereas the S/G2 transition is facilitated by CDK2-cyclin A. Moreover, significant role in adaptive immune response, apoptosis, cell differentiation and normal DNA repair is also played by CDK2.^[3-6] The loss of proliferation control leading to cancer is accountable to CDK2 and its regulatory proteins.^[7-9] In metastasis of prostate cancer CDK2 is involved critically.^[10] The inhibition of CDK2 might curb proliferation of ovarian cancer cells with amplified CCNE1 expression,^[11] induce apoptosis of MYCN-amplified neuroblastoma cells^[12] and cause the death of BRCA1-deficient cancers.^[6] The CDK2 inhibitors in combination with phosphatidylinositol-3-kinase (PI3K) inhibitors induced cell death in colorectal cancer whereas the dual inhibition of CDK2 and bromodomain-containing protein 4 (BRD4) caused apoptosis in MYC-amplified medulloblastoma cells.^[13,14] The acquired resistance of CDK4/6 inhibitors may be reversed by inhibition of CDK2.^[15] Thus, in cancer therapy, CDK2 is a promising drug target.

In the recent past, a variety of scaffolds have been utilized to discover CDK inhibitors that are ATP-competitive inhibitors occupying the catalytic ATP binding site. The first pan-CDK inhibitor, Flavopiridol, inhibited CDKs and other protein kinases is in clinical trials to treat leukemia, breast cancer, colorectal cancer and gastric cancer.^[16,17] Other more selective CDK inhibitors identified are Seliciclib,^[18] Dinaciclib,^[19,20] AT7519,^[21,22] BAY-1000394,^[23,24] PHA-848125^[25,26] and et cetera.^[27,28] The poor selectivity of most of these lacking of balance between efficacy and safety might hindered further development. The importance of selectivity is evinced by the fact that, so far, Palbociclib,^[29] Abemaciclib^[30] and Ribociclib^[31]

are the three selective CDK4/6 inhibitors which have been approved for the treatment of breast cancer by U.S. Food and Drug Administration (FDA).

To improve the selectivity with anti-proliferative activity of CDK inhibitors towards CDK2 is a challenging issue as the reported most selective CDK2 inhibitor, C-73, had almost no anti-proliferative activity against cancer cells.^[32,33] AT7519, a pan-CDK inhibitor, inhibited effectively the CDK1, CDK2 and CDK9 kinase activity, showed high anti-proliferation effect against various tumor cell lines, though potentially affecting the viability of normal cell line MRC-5.^[34] AT7519 is in phase I clinical trials for the treatment of lymphocytic leukemia.

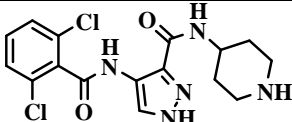
With the intention to optimize the selectivity of AT7519 towards CDK2 a novel series of 4-benzoylamino-1H-pyrazole-3-carboxamide derivatives as CDK2 inhibitors has been reported by Lin *et al.*^[35] The aim of present communication is to establish the quantitative relationships between the reported activities and molecular descriptors unfolding the substitutional changes in titled compounds.

2. MATERIAL AND METHODS

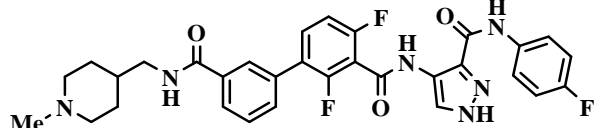
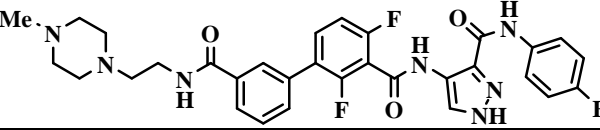
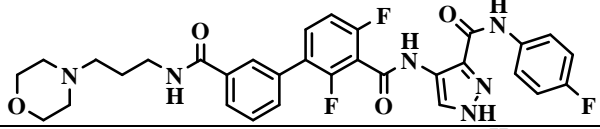
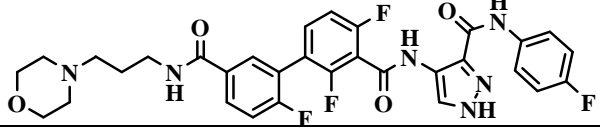
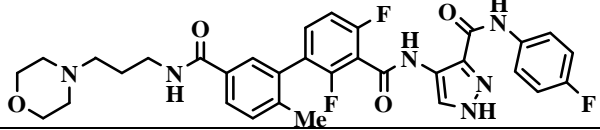
2.1 Data-set

For present work the reported nineteen pyrazole-3-carboxamide derivatives have been considered as the data set.^[35] The structures of these compounds are represented in Table 1. These derivatives were evaluated for their inhibition of CDK1 and CDK2. Both the inhibition activities have also been reported in Table1.^[35] The inhibition activity, IC_{50} , represents the concentration of a compound to achieve 50% inhibition of CDK1 and CDK2. The same is expressed as pIC_{50} on a molar basis and considered as the dependent variable for the present quantitative analysis. 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 and Observed CDK1 and CDK2 inhibition activities of 4-benzoylamino-1H-pyrazole-3-carboxamide derivatives.

Cpd.	Structure	IC_{50} (μ M) ^a	
		CDK2	CDK1
1		0.032	0.089

2		0.250	0.594
3		0.006	0.014
4		0.400	0.306
5		0.086	0.432
6		0.051	0.278
7		0.164	1.850
8		0.144	0.712
9		0.120	1.262
10 ^b		0.058	1.019
11 ^b		0.026	0.236
12		0.076	0.537
13		0.210	0.496
14		0.008	0.069

15 ^b		0.082	0.743
16		0.067	0.350
17		0.073	0.723
18 ^b		0.169	1.299
19		0.485	3.689

^aRef.,^[35] IC₅₀ represents the concentration of a compound to bring out 50% inhibition of CDK1 and CDK2; ^bCompound included in test set.

2.2 Molecular descriptors

The structures of the compounds (Table 1), under study, have been drawn in 2D ChemDraw^[36] 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^[37] for computing the descriptors corresponding to 0D-, 1D-, and 2D-classes.

2.3 Development and Validation of model

The combinatorial protocol in multiple linear regression (CP-MLR)^[38] and partial least squares (PLS)^[39-41] procedure has been 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^[42,43] 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 assess its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5.

2.4 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.^[44] 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 $\pm x$ (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.

3. RESULTS AND DISCUSSION

3.1 QSAR results

For the compounds in Table 1, a total number of 475 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 475 to 60 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^[45] 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 19 compounds were divided into training-set and test-set. Four compounds (nearly 20% of total population) have been selected for test-set. The identified test-set was then used for external validation of models derived from remaining fifteen 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 01 model in two descriptor and 03 models in three descriptors all having $r^2_{\text{Test}} > 0.5$, for the CDK2 inhibitory activity. All these models are mentioned below.

$$\text{pIC}_{50} = 7.966 - 1.354(0.404)\text{TIE} - 0.910(0.341)\text{C-027} \\ n = 15, r = 0.745, s = 0.393, F = 7.499, Q^2_{\text{LOO}} = 0.381, Q^2_{\text{L3O}} = 0.389, r^2_{\text{Test}} = 0.595 \quad (1)$$

$$\text{pIC}_{50} = 7.497 - 0.883(0.147)\text{PW4} + 0.885(0.309)\text{MATS6m} - 1.493(0.247)\text{PSA} \\ n = 15, r = 0.903, s = 0.263, F = 16.350, Q^2_{\text{LOO}} = 0.615, Q^2_{\text{L3O}} = 0.653, r^2_{\text{Test}} = 0.570 \quad (2)$$

$$\text{pIC}_{50} = 8.221 - 0.802(0.285)\text{TIE} - 0.814(0.150)\text{PW4} - 0.946(0.240)\text{PSA} \\ n = 15, r = 0.901, s = 0.266, F = 15.996, Q^2_{\text{LOO}} = 0.642, Q^2_{\text{L3O}} = 0.567, r^2_{\text{Test}} = 0.527 \quad (3)$$

$$\text{pIC}_{50} = 8.041 - 1.598(0.355) \text{ TIE} + 1.115(0.254) \text{ N-072} - 0.961(0.285) \text{ PSA}$$
$$n = 15, r = 0.866, s = 0.307, F = 11.042, Q^2_{\text{LOO}} = 0.565, Q^2_{\text{L3O}} = 0.521, r^2_{\text{Test}} = 0.544 \quad (4)$$

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,^[1-4] the descriptors TIE and PW4 are topological class descriptors. The topological class descriptors are based on a graphical representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labelling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching, and cyclicity and can also encode chemical information concerning atom type and bond multiplicity.

The descriptor MATS6m is a 2D-autocorrelation descriptor, C-027 and N-072 are atom centered fragments whereas PSA is a property class descriptor. The 2D-autocorrelations are molecular descriptors, which describe how a considered property is distributed along a topological molecular structure. The 2D-auto descriptors have their origin in autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS). The computation of these descriptors involves the summations of different autocorrelation functions corresponding to the different fragment lengths and leads to different autocorrelation vectors corresponding to the lengths of the structural fragments. Also a weighing component in terms of a physicochemical property has been embedded in this descriptor. As a result, these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag n (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighing component— m for mass or v for volume or e for Sanderson electronegativity or p for polarizability—for its computation. Atom centered fragment descriptors are simple molecular descriptors defined as the number of specific atom types in a molecule and their calculation is based on the knowledge of the molecular composition and atom connectivities.

The positive sign of regression coefficients of descriptors MATS6m and N-072 suggested that a higher value of atomic masses weighted Moran autocorrelation of lag 6 and presence of RCO-N</>N-X=X type structural fragments in a molecule would be beneficial to augment the CDK2 inhibitory activity. On the other hand, a lower value of descriptors TIE (E-state topological parameter), PW4 (path/walk 4-Randic shape index), PSA (fragment based polar surface area) and absence of R--CH--X type structural fragment would be supportive to the CDK2 inhibition.

These models have accounted for nearly 82% 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. (2) to (4) have been included in Table 2. The models (2) to (4) are validated with an external test set of 4 compounds listed in Table 2. 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 2. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

Table 2: Observed and Calculated CDK2 and CDK1 inhibition activities of 4-benzoylamino-1H-pyrazole-3-carboxamide derivatives.

Cpd.	pIC50(M) ^a					pIC50(M) ^a			
	CDK2 inhibition					CDK1 inhibition			
	Obsd.	Calculated				Obsd.	Calculated		
		Eq.(2)	Eq.(3)	Eq.(4)	PLS		Eq.(5)	Eq.(6)	Eq.(7)
1	7.49	6.98	7.03	7.10	7.15	7.05	6.84	6.97	6.84
2	6.60	6.72	6.75	6.54	6.74	6.23	6.60	6.16	6.38
3	8.22	8.01	8.02	8.04	8.16	7.85	7.71	7.87	7.94
4	6.40	6.60	6.62	6.87	6.31	6.51	6.58	6.30	6.19
5	7.07	7.20	7.03	7.29	7.24	6.36	6.37	6.30	6.65
6	7.29	7.30	7.36	7.32	7.26	6.56	6.64	6.60	6.52
7	6.79	6.89	6.93	6.47	6.62	5.73	5.61	5.35	6.04
8	6.84	6.80	6.81	6.72	6.77	6.15	6.09	6.25	6.30
9	6.92	7.04	7.23	7.07	7.07	5.90	5.95	6.22	6.15
10 ^b	7.24	6.89	6.95	7.55	6.92	5.99	6.19	6.29	5.88
11 ^b	7.59	7.43	7.47	7.76	7.71	6.63	6.32	6.63	6.42
12	7.12	7.32	7.23	7.49	7.45	6.27	6.66	6.34	6.27
13	6.68	6.59	6.64	6.51	6.93	6.30	6.45	6.42	6.09
14	8.10	8.27	8.23	7.87	7.98	7.16	6.96	6.82	6.98
15 ^b	7.09	6.96	7.15	7.25	7.10	6.13	5.79	5.88	5.99
16	7.17	6.99	7.07	7.18	7.06	6.46	6.10	6.27	6.01

17	7.14	6.76	6.65	6.94	6.78	6.14	5.87	6.39	6.14
18 ^b	6.77	6.76	6.49	6.61	6.54	5.89	5.89	5.68	5.48
19	6.31	6.65	6.55	6.73	6.62	5.43	5.67	5.86	5.60

^aIC₅₀ on molar basis, taken from reference,^[35] ^bCompound included in test set.

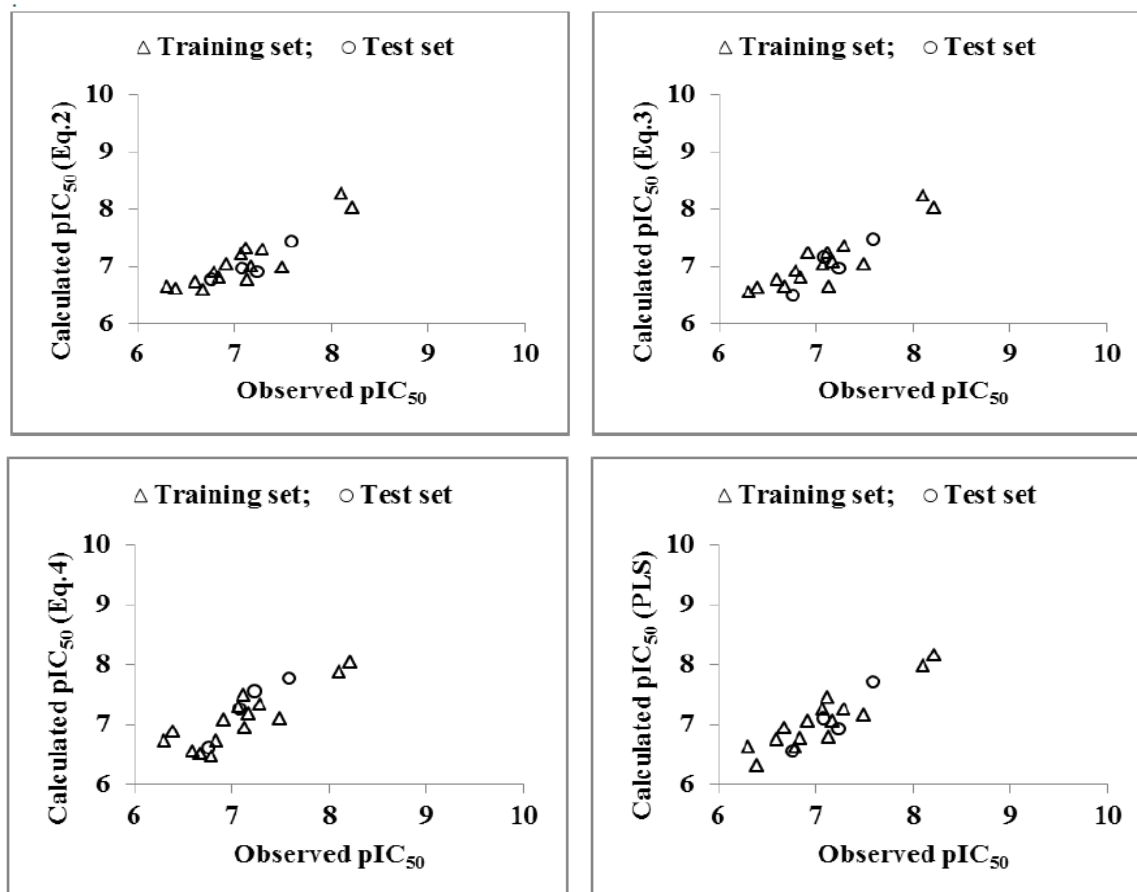


Figure 1: Plot of observed versus calculated pIC₅₀ values for Training and Test-set compounds for CDK2 inhibition.

A partial least square (PLS) analysis has been carried out on these 6 CP-MLR identified descriptors discussed above (Eqs. (1)-(4)) 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 6 descriptors and they explained 84.12% variance in the activity. All these descriptors convey same inference in the PLS model as well. It is also observed that PLS model from the dataset devoid of these 6 descriptors is inferior in explaining the CDK2 activity of the analogues. The MLR-like PLS coefficients of these 6 descriptors are given in Table 3. 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 (Table 3).

Table 3: PLS and MLR-like PLS models from the 6 descriptors of Two and Three parameter CP-MLR models for CDK2 inhibitory activities.

parameter of MLR models for CDH2 inhibitory activities.									
A: PLS equation									
PLS components					PLS coefficient (s.e.) ^a				
Component-1					-0.463(0.060)				
Component-2					0.090(0.045)				
Constant					7.076				
B: MLR-like PLS equation									
S. no.	Descriptor	MLR-like coefficient ^b	(f.c.) ^c	Order	S. No.	Descri ptor	MLR-like coefficient ^b	(f.c.) ^c	Order
1	TIE	-0.575	- 0.369	1	4	C-027	-0.369	-0.183	4
2	PW4	-0.488	0.098	2	5	N-072	0.098	0.049	5
3	MATS6m	0.011	- 0.472	6	6	PSA	-0.472	-0.235	3
						Constant = 19.504			
C: PLS regression statistics					Values				
n					15				
r					0.917				
s					0.235				
F					31.792				
Q ² _{LOO}					0.749				
Q ² _{L50}					0.713				
r ² _{Test}					0.547				

^aRegression coefficient of PLS factor and its standard error. ^bCoefficients of MLR-like PLS equation in terms of descriptors for their original values; ^cf.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data.

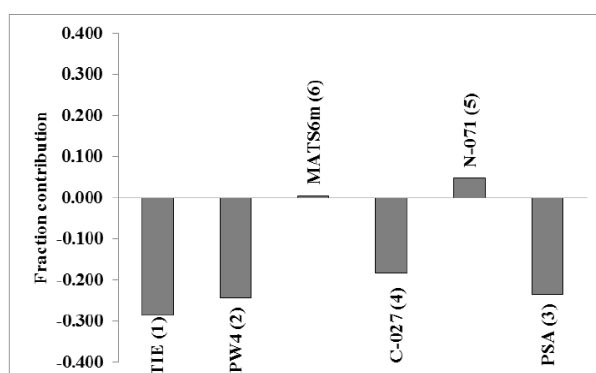


Figure 2: Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 6 CP-MLR identified descriptors associated with CDK2 inhibitory activity of 4-benzoylamino-1H-pyrazole-3-carboxamide derivatives.

CP-MLR analysis has also been carried out for another reported inhibition activity CDK1 using same test set. Following are the selected three-descriptor models (out of 16 models) for the CDK1 inhibitory activities emerged through CP-MLR.

$$\text{pIC}_{50} = 6.840 - 0.948(0.180) \text{ nO} + 1.764(0.262) \text{ MATS7m} - 1.857(0.276) \text{ MATS8v}$$
$$n = 15, r = 0.925, s = 0.253, F = 22.018, Q^2_{\text{LOO}} = 0.739, Q^2_{\text{L3O}} = 0.761, r^2_{\text{Test}} = 0.555 \quad (5)$$

$$\text{pIC}_{50} = 8.342 + 0.613(0.195) \text{ SPI} - 2.322(0.310) \text{ TIE} - 1.335(0.222) \text{ JGI4}$$
$$n = 15, r = 0.923, s = 0.258, F = 21.179, Q^2_{\text{LOO}} = 0.524, Q^2_{\text{L3O}} = 0.635, r^2_{\text{Test}} = 0.655 \quad (6)$$

$$\text{pIC}_{50} = 5.814 + 1.582(0.271) \text{ MSD} - 0.991(0.288) \text{ TIE} + 1.054(0.359) \text{ MLOGP}$$
$$n = 15, r = 0.920, s = 0.262, F = 20.358, Q^2_{\text{LOO}} = 0.729, Q^2_{\text{L3O}} = 0.750, r^2_{\text{Test}} = 0.570 \quad (7)$$

It is evinced from the models mentioned above that descriptors MATS7m (atomic masses weighted Moran autocorrelations of lag 7), SPI (superpendentic index), MSD (Balaban mean square distance index) and MLOGP (Moriguchi octanol-water partition coefficient, logP) have shown positive contribution to activity suggesting higher values of these descriptors for the elevated activity. The negative correlation to the activity of descriptors nO (number of oxygen atoms), MATS8v (atomic van der waals volumes weighted Moran autocorrelations of lag 8), TIE (E-state topological parameter) and JGI4 (mean topological charge index of order 4) advocates that a lower value of these will be beneficial for the activity.

These models have accounted for nearly 86% 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. (5) to (7) have been included in Table 2. The models (5) to (7) are validated with an external test set of 4 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 2. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 3.

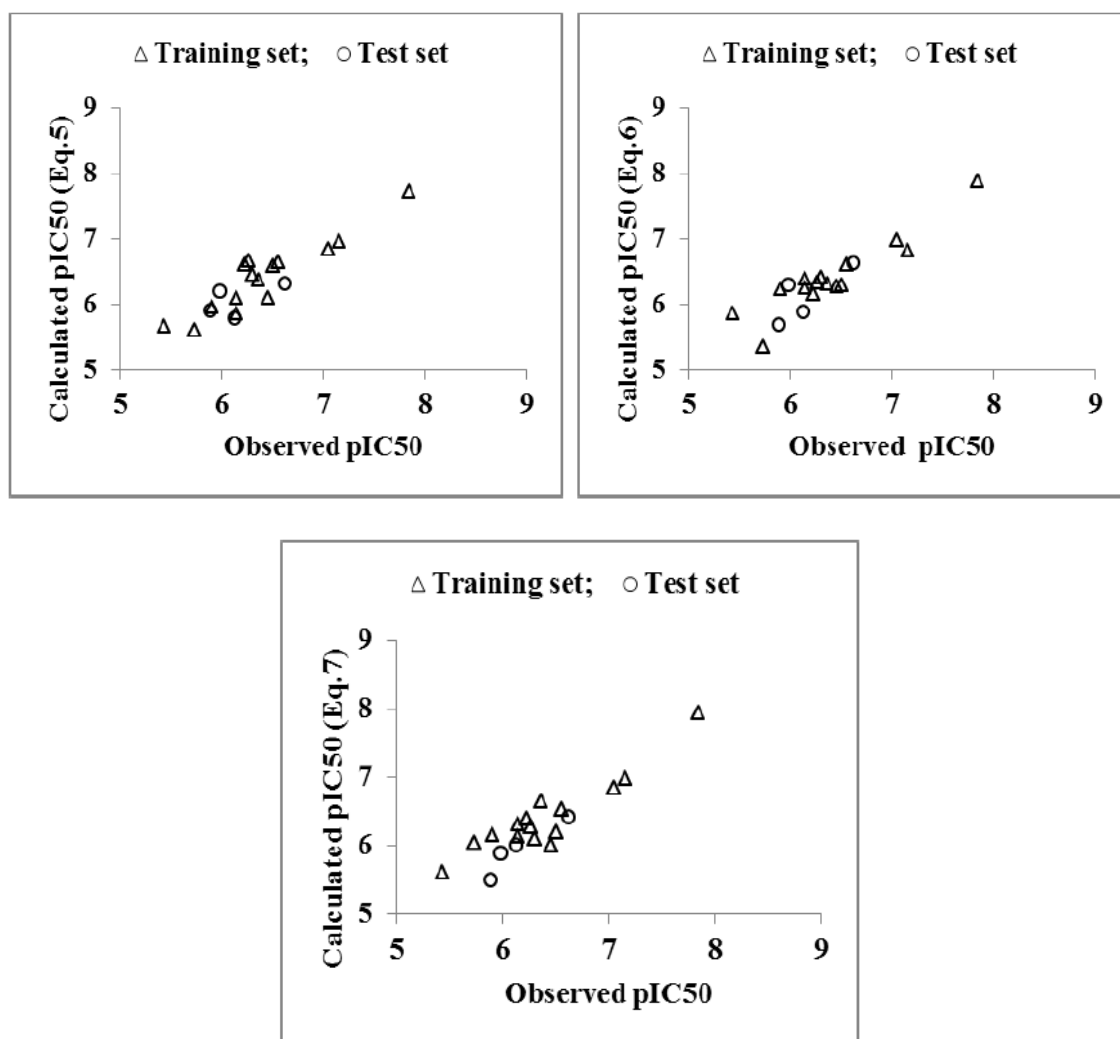


Figure 3: Plot of observed versus calculated pIC₅₀ values for Training and Test-set compounds for CDK1 inhibition.

3.2 Applicability domain

On analyzing the applicability domain (AD) for the CDK2 inhibitory actions in the Williams plot (Figure 4) of the model based on the whole data set (Table 4), No any compound has been 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 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 4: Models derived for the whole data set (n = 19) in descriptors identified through CP-MLR for CDK2 inhibitory actions.

Model	r	s	F	Q ² _{LOO}	Eq.
$pIC_{50} = 7.494 - 0.891(0.134)PW4 + 0.927(0.251)MATS6m - 1.467(0.205)PSA$	0.895	0.245	20.263	0.635	(2a)
$pIC_{50} = 8.210 - 0.834(0.231)TIE - 0.789(0.134)PW4 - 0.838(0.201)PSA$	0.893	0.247	19.749	0.675	(3a)
$pIC_{50} = 7.944 - 1.421(0.280)TIE + 1.014(0.208)N-072 - 0.934(0.243)PSA$	0.862	0.279	14.501	0.603	(4a)

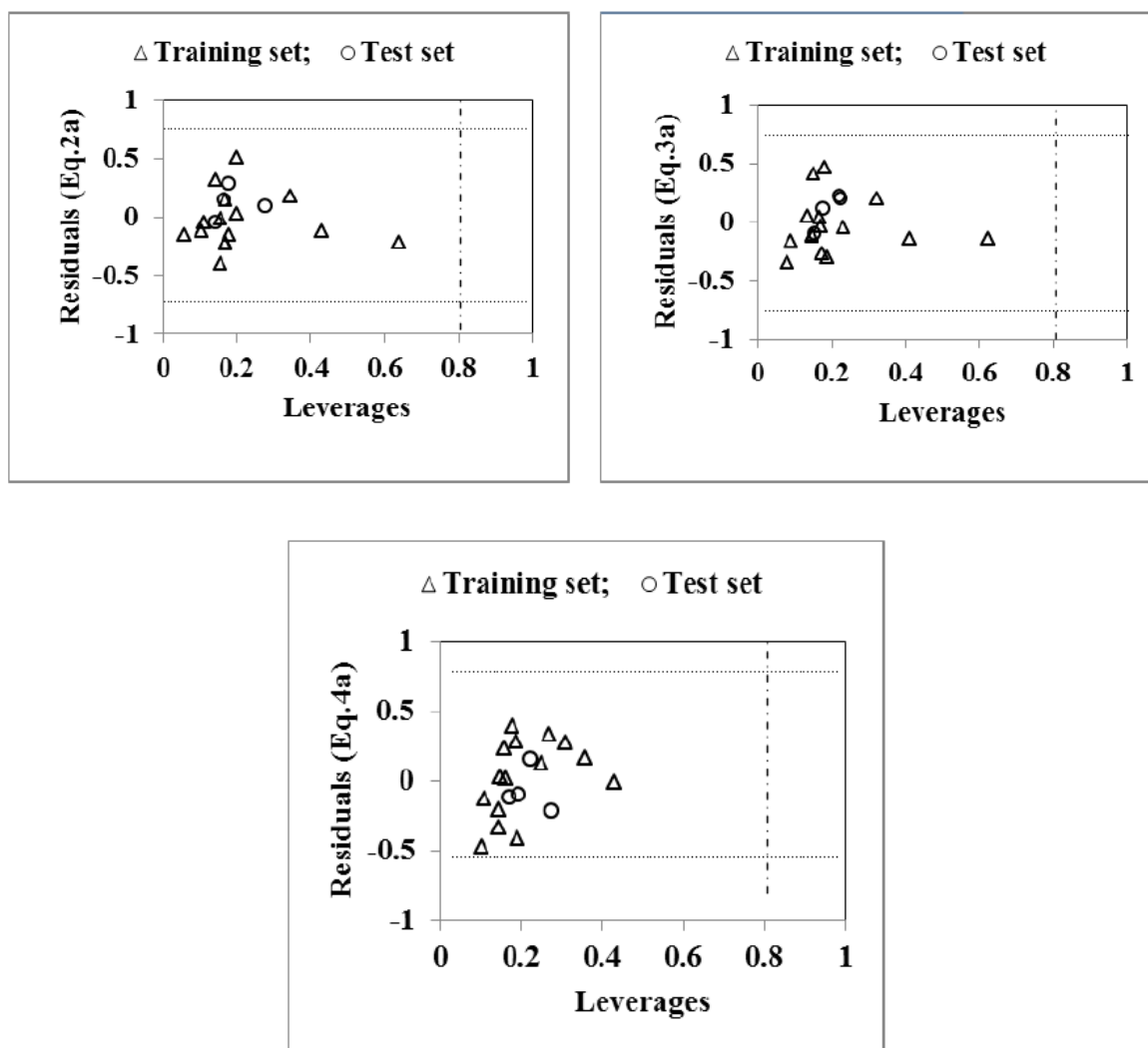


Figure 4: Williams plot for the training-set and test-set for CDK2 inhibition activity of compounds in Table 1. The horizontal dotted line refers to the residual limit ($\pm 3 \times$ standard deviation) and the vertical dotted line represents threshold leverage h^* ($=0.8$).

4. CONCLUSION

The CDK 2 and 1 inhibition activity of 4-benzoylamino-1H-pyrazole-3-carboxamide 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 CDK2 inhibitory activity have highlighted the role of E-state topological parameter (TIE), path/walk 4-Randic shape index (PW4), atomic masses weighted Moran autocorrelation of lag 6 (MATS6m) and fragment based polar surface area (PSA). Atom centered fragments, R--CH--X (C-027) and RCO-N</>N-X=X (N-072) have shown prevalence to model the CDK2 inhibitory activity.

The positive sign of regression coefficients of descriptors MATS6m and N-072 suggested that a higher value of descriptor MATS6m and presence of RCO-N</>N-X=X type structural fragments in a molecule would be beneficial to augment the CDK2 inhibitory activity. On the other hand, a lower value of descriptors TIE, PW4 and PSA, and absence of R--CH--X type structural fragment would be supportive to the CDK2 inhibition. The statistics emerged from the test sets have validated the identified significant models. 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.

For the CDK1 inhibitory activity, a higher value of descriptors MATS7m (atomic masses weighted Moran autocorrelations of lag 7), SPI (superpendentic index), MSD (Balaban mean square distance index) and MLOGP (Moriguchi octanol-water partition coefficient, logP) will be helpful to augment the activity. On the other hand lower values of descriptors nO (number of oxygen atoms), MATS8v (atomic van der Waals volumes weighted Moran autocorrelations of lag 8), TIE (E-state topological parameter) and JGI4 (mean topological charge index of order 4) advocates that a lower value of these will be beneficial for the activity.

Compliance with ethical standards**ACKNOWLEDGEMENTS**

Authors are thankful to their institutions 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 Biology*, 2014; 15: 122.
3. Chunder N, Wang L, Chen C, Hancock WW and Wells AD. Cyclin-dependent kinase 2 controls peripheral immune tolerance. *Journal of Immunology*, 2012; 189: 5659-5666.
4. Saurus P, Kuusela S, Dumont V, Lehtonen E, Fogarty CL, Lassenius MI, Forsblom C, Lehto M, Saleem MA, Groop PH and Lehtonen S. Cyclin-dependent kinase 2 protects podocytes from apoptosis. *Scientific Reports*, 2016; 6: 21664.
5. Granes F, Roig MB, Brady HJ and Gil-Gomez G. Cdk2 activation acts upstream of the mitochondrion during glucocorticoid induced thymocyte apoptosis, *European Journal of Immunology*, 2004; 34: 2781-2790.
6. 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 Research*, 2006; 66: 8219-8226.
7. Malumbres M and Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm, *Nature Reviews Cancer*, 2009; 9: 153-166.
8. Wang J, Yang T, Xu G, Liu H, Ren C, Xie W and Wang M. Cyclin-dependent kinase 2 promotes tumor proliferation and induces radio resistance in glioblastoma. *Translation Oncology*, 2016; 9: 548-556.
9. Ying M, Shao X, Jing H, Liu Y, Qi X, Cao J, Chen Y, Xiang S, Song H, Hu R, Wei G, Yang B and He Q. Ubiquitin-dependent degradation of CDK2 drives the therapeutic differentiation of AML by targeting PRDX2. *Blood*, 2018; 131: 2698-2711.
10. Yin XF, Yu J, Zhou Y, Wang CY, Jiao ZM, Qian ZN, Sun H and Chen BH. Identification of CDK2 as a novel target in treatment of prostate cancer, *Future Oncology*, 2018; 14: 709-718.

11. Au-Yeung G, Lang F, Azar WJ, Mitchell C, Jarman KE, Lackovic K, Aziz D, Cullinane C, Pearson RB, Mileskin L, Rischin D, Karst AM, Drapkin R, Etemadmoghadam D and Bowtell DD. Selective targeting of cyclin E1-amplified high-grade serous ovarian cancer by cyclin-dependent kinase 2 and AKT inhibition, *Clinical Cancer Research*, 2017; 23: 1862-1874.
12. Molenaar JJ, Ebus ME, Geerts D, Koster J, Lamers F, Valentijn LJ, Westerhout EM, Versteega R and Caron HN. Inactivation of CDK2 is synthetically lethal to MYCN over-expressing cancer cells, *Proceedings of the National Academy of Sciences USA*, 2009; 109: 12968-12973.
13. Beale G, Haagenen EJ, Thomas HD, Wang LZ, Revill CH, Payne SL, Golding BT, Hardcastle IR, Newell DR, Griffin RJ and Cano C. Combined PI3K and CDK2 inhibition induces cell death and enhances in vivo antitumour activity in colorectal cancer. *British Journal of Cancer*, 2016; 115: 682-690.
14. Bolin S, Borgenvik A, Persson CU, Sundstrom A, Qi J, Bradner JE, Weiss WA, Cho YJ, Weishaupt H and Swartling FJ. Combined BET bromodomain and CDK2 inhibition in MYC-driven medulloblastoma. *Oncogene*, 2018; 37: 2850-2862.
15. Herrera-Abreu MT, Palafox M, Asghar U, Rivas MA, Cutts RJ, Garcia-Murillas I, Pearson A, Guzman M, Rodriguez O, Grueso J, Bellet M, Cortes J, Elliott R, Pancholi S, Baselga J, Dowsett M, Martin LA, Turner NC and Serra V. Early adaptation and acquired resistance to CDK4/6 inhibition in estrogen receptor-positive breast cancer. *Cancer Research*, 2016; 76: 2301-2313.
16. Senderowicz AM. Flavopiridol: the first cyclin-dependent kinase inhibitor in human clinical trials. *Investigational New Drugs*, 1999; 17: 313-320.
17. Deep A, Marwaha RK, Marwaha MG, Jyoti, Nandal R and Sharma AK. Flavopiridol as cyclin dependent kinase (CDK) inhibitor: a review. *New Journal of Chemistry*, 2018; 42: 18500-18507.
18. Alvi AJ, Austen B, Weston VJ, Fegan C, MacCallum D, Gianella-Borradori A, Lane DP, Hubank M, Powell JE, Wei W, Taylor AM, Moss PA and Stankovic T. A novel CDK inhibitor, CYC202 (R-roscovitine), overcomes the defect in p53-dependent apoptosis in B-CLL by down-regulation of genes involved in transcription regulation and survival. *Blood*, 2005; 105: 4484-4491.
19. Kumar SK, LaPlant B, Chng WJ, Zonder J, Callander N, Fonseca R, Fruth B, Roy V, Erlichman C, Stewart AK, Mayo C and Phase. Dinaciclib, a novel CDK inhibitor,

- demonstrates encouraging single-agent activity in patients with relapsed multiple myeloma. *Blood*, 2015; 125: 443-448.
20. Parry D, Guzi T, Shanahan F, Davis N, Prabhavalkar D, Wiswell D, Seghezzi W, Paruch Dwyer MP, Doll R, Nomeir A, Windsor W, Fischmann T, Wang Y, Oft M, Chen T, Kirschmeier P and Lees EM. Dinaciclib, A novel and potent cyclin-dependent kinase inhibitor. *Molecular Cancer Therapeutics*, 2010; 9: 2344-2353. SCH 727965.
21. Seftel MD, Kuruvilla J, Kouroukis T, Banerji V, Fraser G, Crump M, Kumar R, Chalchal HI, Salim M, Laister RC, Crocker S, Gibson SB, Toguchi M, Lyons JF, Xu H, Powers J, Sederias J, Seymour L and Hay AE. The CDK inhibitor AT7519M in patients with relapsed or refractory chronic lymphocytic leukemia (CLL) and mantle cell lymphoma. A Phase II study of the Canadian Cancer Trials Group. *Leukemia and Lymphoma*, 2017; 58: 1358-1365.
22. Wyatt PG, Woodhead AJ, Berdini V, Boulstridge JA, Carr MG, Cross DM, Davis DJ, Devine LA, Early TR, Feltell RE, Lewis EJ, McMenamin RL, Navarro EF, O'Brien MA, O'Reilly M, Reule M, Saxty G, Seavers LC, Smith DM, Squires MS, Trewartha G, Walker MT and Woolford AJ. Identification of N-(4-piperidiny)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *Journal of Medicinal Chemistry*, 2008; 51: 4986-4999.
23. Cho BC, Dy GK, Govindan R, Kim DW, Pennell NA, Zalcmann G, Besse B, Kim JH, Koca G, Rajagopalan P, Langer S, Ocker M, Nogai H and Barlesi F. Phase Ib/II study of the pan-cyclin-dependent kinase inhibitor roniciclib in combination with chemotherapy in patients with extensive-disease small-cell lung cancer. *Lung Cancer*, 2018; 123: 14-21.
24. Lücking U, Jautelat R, Krüger M, Brumby T, Lienau P, Schöfer M, Briem H, Schulze J, Hillisch A, Reichel A, Wengner AM and Siemeister G. The lab oddity prevails: discovery of pan-CDK inhibitor (R)-S-cyclopropyl-S-(4-{[4-{[(1R,2R)-2-hydroxy-1-methylpropyl]oxy}-5-(trifluoromethyl)pyrimidin-2-yl]amino}phenyl)sulfoximide (BAY :1000394) for the treatment of cancer. *ChemMedChem*, 2013; 8: 1067-1085.
25. Degrassi A, Russo M, Nanni C, Patton V, Alzani R, Giusti AM, Fanti S, Ciomei M, Pesenti E and Texido G. Efficacy of PHA-848125, a cyclin-dependent kinase inhibitor, on the K-Ras(G12D)LA2 lung adenocarcinoma transgenic mouse model: evaluation by multimodality imaging. *Molecular Cancer Therapeutics*, 2010; 9: 673-681.
26. Brasca MG, Amboldi N, Ballinari D, Cameron A, Casale E, Cervi G, Colombo M, Colotta F, Croci V, D'Alessio R, Fiorentini F, Isacchi A, Mercurio C, Moretti W, Panzeri

- A, Pastori W, Pevarello P, Quartieri F, Roletto F, Traquandi G, Vianello P, Vulpetti A, and Ciomei M. (2009). Identification of N,1,4,4-tetramethyl-8-{[4-(4-methylpiperazin-1-yl)phenyl]amino}-4,5-dihydro-1H-pyrazolo[4,3-h]quinazoline-3-carboxamide (PHA-48125), a potent, orally available cyclin dependent kinase inhibitor. *Journal of Medicinal Chemistry*, 2010; 52: 5152-5163.
27. Reddy MV, Akula B, Cosenza SC, Athuluridivakar S, Mallireddigari MR, Pallela VR, Billa VK, Subbaiah DR, Bharathi EV, Vasquez-Del Carpio R, Padgaonkar A, Baker SJ and Reddy EP. Discovery of 8-cyclopentyl-2-[4-(4-methyl-piperazin-1-yl)-phenylamino]-7-oxo-7,8-dihydro-pyrid o[2,3-d]pyrimidine-6-carbonitrile (7x) as a potent inhibitor of cyclin-dependent kinase 4 (CDK4) and AMPK-related kinase 5 (ARK5). *Journal of Medicinal Chemistry*, 2014; 57: 578-599.
28. Jorda R, Hendrychova D, Voller J, Reznickova E, Gucky T and Krystof V. How selective are pharmacological inhibitors of cell-cycle-regulating cyclin dependent kinases? *Journal of Medicinal Chemistry*, 2018; 61: 9105-9120.
29. Clark AS, Karasic TB, DeMichele A, Vaughn DJ, O'Hara M, Perini R, Zhang P, Lal P, Feldman M, Gallagher M and O'Dwyer PJ. Palbociclib (PD0332991)-a selective and potent cyclin-dependent kinase inhibitor: a review of pharmacodynamics and clinical development. *JAMA Oncology*, 2016; 2: 253-260.
30. Goldman JW, Shi P, Reck M, Paz-Ares L, Koustenis A and Hurt KC. Treatment rationale and study design for the juniper study: a randomized phase III study of Abemaciclib with best supportive care versus erlotinib with best supportive care in patients with stage IV non-small-cell lung cancer with a detectable KRAS mutation whose disease has progressed after platinum-based chemotherapy. *Clinical Lung Cancer*, 2016; 17: 80-84.
31. Syed YY. Ribociclib: first global approval. *Drugs*, 2017; 77: 799-807.
32. Coxon CR, Anscombe E, Harnor SJ, Martin MP, Carbain B, Golding BT, Hardcastle IR, Harlow LK, Korolchuk S, Matheson CJ, Newell DR, Noble ME, Sivaprakasam M, Tudhope SJ, Turner DR, 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-arylamino purines. *Journal of Medicinal Chemistry*, 2017; 60: 1746-1767.
33. Cheng W, Yang Z, Wang S, Li Y, Wei H, Tian X and Kan Q. Recent development of CDK inhibitors: an overview of CDK/inhibitor co-crystal structures, *European Journal of Medicinal Chemistry*, 2019; 164: 615-639.

34. Squires MS, Feltell RE, Wallis NG, Lewis EJ, Smith DM, Cross DM, Lyons JF and Thompson NT. Biological characterization of AT7519, a small molecule inhibitor of cyclin-dependent kinases, in human tumor cell lines, *Molecular Cancer Therapeutics*, 2009; 8: 324-332.
35. Lin T, Li J, Liu L, Li Y, Jiang H, Chen K, Xu P, Luo C and Zhou B. Design, synthesis, and biological evaluation of 4-benzoylamino-1H-pyrazole-3-carboxamide derivatives as potent CDK2 inhibitors. *European Journal of Medicinal Chemistry*, 2021; 215: 113281-92.
36. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. <http://www.cambridgesoft.com>
37. Dragon software (version 1.11-2001) by Todeschini R and Consonni V. Milano, Italy. <http://www.taletе.mi.it/dragon.htm>
38. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: analysis of Selwood et al. Data Set-a case study. *QSAR and Combinatorial Science*, 2003; 22: 583-595.
39. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, 1978; 20: 397-405.
40. Kettaneh N, Berglund A and Wold S. PCA and PLS with very large data sets. *Computational Statistics and Data Analysis*, 2005; 48: 69-85.
41. Stahle L and Wold S. Multivariate data analysis and experimental design. In: Ellis GP and West WB. (Eds.), *Biomedical research. Progress in medicinal chemistry*. Elsevier Science Publishers, BV, Amsterdam, 1988; 25: 291-338.
42. So S-S and Karplus M. Three-dimensional quantitative structure–activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation. *Journal of Medicinal Chemistry*, 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 and Combinatorial Science*, 2004; 23: 234-244.
44. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR and Combinatorial Science*, 2007; 26: 694-701.
45. Golbraikh A and Tropsha A. Beware of q^2 ! *Journal of Molecular Graphics and Modeling*, 2002; 20: 269-276.