

**BJMHR**

British Journal of Medical and Health Research

Journal home page: www.bjmhr.com

Quantitative Structure-Activity Relationship Study on the CDK4/6 Inhibitory Activity: The 4-Thiazol-N-(pyridin-2-yl)pyrimidin-2-amines

Dinesh Kumar Meena¹, Brij Kishore Sharma^{1,*}, Raghuraj Parihar²

1.Department of Chemistry, Government College, Bundi-323 001, India

2.Department of Chemistry, Government College, Kota-324 001, India

ABSTRACT

Cyclin D dependent kinases namely CDK4 and CDK6 regulate entry into S phase of the cell cycle. These are emerging validated targets for anti-cancer drug discovery. A QSAR study has been carried out on the 4-thiazol-N-(pyridin-2-yl)pyrimidin-2-amine derivatives, which were reported as highly potent and selective inhibitors of CDK4 and CDK6, in terms of Dragon descriptors with the aim to establish the quantitative relationships between the reported activities and molecular descriptors unfolding the substitutional changes. In deriving QSAR models, combinatorial protocol in multiple linear regression (CP-MLR) approach was used. These models have accounted for nearly 79 and 83% variance in the CDK4 and CDK6 observed activities, respectively. The statistically validated QSAR models and the descriptors emerged in these models provided rationales to explain the CDK4 and CDK6 inhibitory activities of these congeners. 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, CDK4 and CDK6 inhibitors, Combinatorial protocol in multiple linear regression (CP-MLR) analysis, PLS, Dragon descriptors, 4-Thiazol-N-(pyridin-2-yl)pyrimidin-2-amines.

*Corresponding Author Email: bksharma_sikar@rediffmail.com

Received 09 August 2022, Accepted 28 September 2022

Please cite this article as: Sharma BK *et al.*, Quantitative Structure-Activity Relationship Study on the CDK4/6 Inhibitory Activity: The 4-Thiazol-N-(pyridin-2-yl)pyrimidin-2-amines. British Journal of Medical and Health Research 2022.

INTRODUCTION

The ordered sequence of events in cell cycle not only leads to the transition from quiescence or cytokinesis to cell proliferation but also ensures genome stability¹. The four sequential phases are (i) S phase (DNA synthesis occurs), (ii) M phase (cell divides into daughter cells), (iii) G1 phase after mitosis and before S phase (cellular biosynthetic events take place at high rate and cells grow in size) (iv) G2 phase occurring between S and M phases (during which cells prepare for mitosis). Cells make a decision to enter S phase or remain quiescent utilizing signaling pathways which link extracellular cues (e.g. growth factors) to G1 phase of the cell cycle². The retinoblastoma tumor suppressor proteins (Rbs) regulate the progression through G1 phase, repressing the activity of E2 promoter binding factor (E2F) transcription factors that are functionally requisite for transition from G1 to S phase. The transcription of genes encoding necessary proteins for DNA replication (such as cyclin A and cyclin E) is promoted by release of E2F transcription factors on phosphorylation of Rb by G1-phase CDKs.

In mammalian cells, CDK4 and CDK6 are the primary kinases which phosphorylate Rb proteins in G1 phase³. The disruption of CDK4/6-Rb-E2F pathway in 90% of cancers was shown in earlier studies⁴⁻⁷. Moreover, it is established in genetic studies that CDK4/6 are not essential for the mitotic cell cycle. The evidences in favor of this are the facts that mice lacking CDK4/6 are viable, and the proliferation of specific cell types was only affected by the inactivation of the CDK4/6 genes^{8,9} rendering CDK4/6 as rational targets to develop small-molecules to intervene therapeutically in cancers¹⁰. The approval of first CDK4/6 inhibitor, palbociclib, to treat metastatic breast cancer by FDA revived the area of selective CDK inhibitors¹¹. Now a days, researches on CDK4/6 inhibitors got recognition in answering complex biological questions^{12,13}.

The selectivity of kinases is still unconquered due to the binding of most of the CDK inhibitors to the highly conserved ATP binding site^{14,15}. Not only the drug safety be impaired by off-targets but differential cellular potencies and dissimilar clinical responses also. Therefore, new pharmacophores might be with affordable desired kinase selectivity profiles to offer minimized undesirable side effects. The evolutionary conservation of the ATP-binding site is a challenging task in discovering highly selective small-molecule ATP-antagonistic CDK inhibitors^{10,16-18}. The 2-anilino-4-(thiazol-5-yl)pyrimidine pharmacophore was identified as both potent and selective ATP competitive CDK2 or CDK9 inhibitors¹⁹⁻²¹. Analogues based on this pharmacophore were synthesized and screened against a wide panel of CDKs, and several of these were found with modest CDK4 inhibitory potency¹⁹⁻²¹. A novel series of compounds based on the structural modifications to the 2-anilino-4-(thiazol-5-yl)pyrimidine pharmacophore as CDK4/6 inhibitors was synthesized and evaluated by

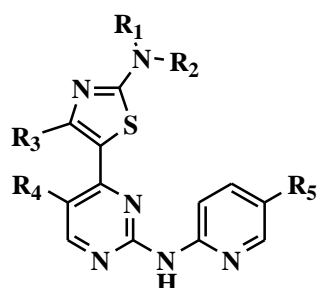
Tadesse *et al.*²². The aim of present communication is to establish the quantitative relationships between the reported activities and molecular descriptors unfolding the substitutional changes in titled compounds.

MATERIALS AND METHOD

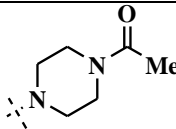
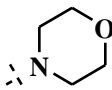
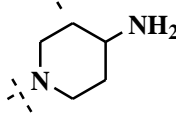
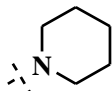
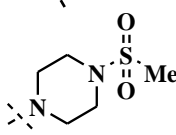
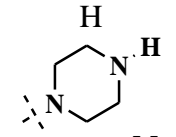
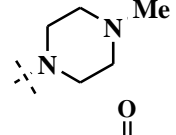
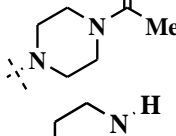
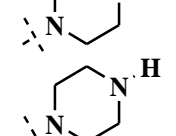
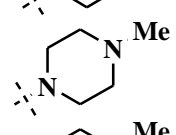
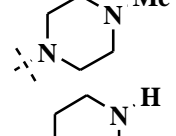
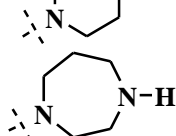
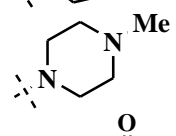
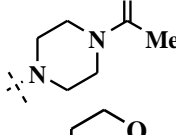
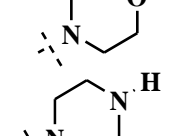
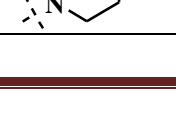


Data-set

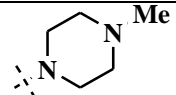
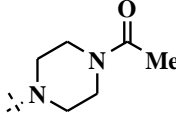
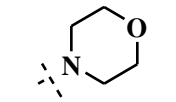
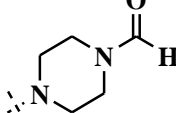
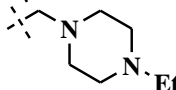
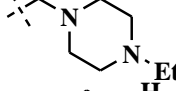
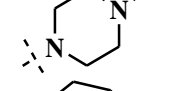
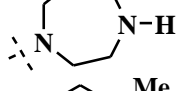
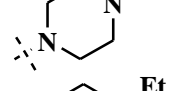
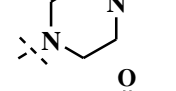
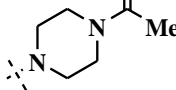
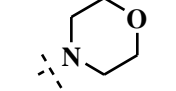
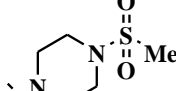
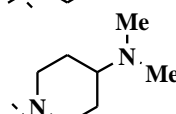
For present work the reported thirty seven 4-Thiazol-N-(pyridin-2-yl)pyrimidin-2-amines have been considered as the data set²². The structural variations are mentioned in Table 1. These derivatives were evaluated for their inhibition of CDK4 and CDK6. Apparent inhibition constants (K_i) were calculated using the half maximal inhibition (IC_{50}) and the appropriate $K_m(ATP)$ of each kinase²². Both the inhibition activities have also been reported in Table 1. The same is expressed as pK_i 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 compounds **1** and **15** as potential outliers. An outlier to a QSAR can indicate the limits of applicability of QSAR models. These outliers are 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 CDK4 and CDK6 inhibition activities of 4-thiazol-N-(pyridin-2-yl) pyrimidin-2-amines.



Cpd.	R ₁	R ₂	R ₃	R ₄	R ₅	CDK Inhibition ($\square M$) ^a	
						4D1 ^b	6D3 ^b
1	c-Pent	H	Me	H		0.001	0.034
2	c-Pent	H	Me	H		0.001	0.008
3 ^c	c-Pent	H	Me	H		0.002	0.009
4	c-Pent	H	Me	H		0.002	0.011

5	c-Pent	H	Me	H		0.006	0.093
6 ^c	c-Pent	H	Me	H		0.004	0.030
7	c-Pent	H	Me	H		0.003	0.133
8	c-Pent	H	Me	H		0.070	0.257
9	c-Pent	H	Me	H		0.007	0.055
10	c-Pent	H	Me	H	H	0.570	-
11	c-Pent	Me	Me	H		0.002	0.010
12 ^c	c-Pent	Me	Me	H		0.002	0.010
13	c-Pent	Me	Me	H		0.010	0.031
14 ^c	c-Pent	c-Pent	Me	H		0.071	0.539
15	Ph	H	Me	H		0.005	0.066
16	Ph	H	Me	H		0.019	0.485
17	Ph	Me	Me	H		0.026	0.100
18	<i>i</i> -Pr	H	Me	H		0.005	0.011
19 ^c	<i>i</i> -Pr	H	Me	H		0.016	0.028
20	<i>i</i> -Pr	H	Me	H		0.003	0.015
21	<i>i</i> -Pr	H	Me	H		0.021	0.105
22	<i>i</i> -Pr	H	Me	H		0.041	0.082
23	c-Pent	H	CF ₃	H		0.008	0.002

24	c-Pent	H	CF ₃	H		0.001	0.003
25	c-Pent	H	CF ₃	H		0.004	0.006
26	c-Pent	H	CF ₃	H		0.008	0.011
27	c-Pent	H	CF ₃	H		0.011	0.007
28	c-Pent	H	Me	H		0.006	0.009
29 ^c	c-Pent	H	Me	F		0.003	0.014
30	c-Pent	H	Me	F		0.003	0.007
31 ^c	c-Pent	H	Me	F		0.014	0.010
32 ^c	c-Pent	H	Me	F		0.001	0.003
33	c-Pent	H	Me	F		0.002	0.006
34 ^c	c-Pent	H	Me	F		0.034	0.023
35 ^c	c-Pent	H	Me	F		0.006	0.020
36	c-Pent	H	Me	F		0.039	0.101
37	c-Pent	H	Me	F		0.001	0.031

^aReference [22]; ^b4D1 and 6D3 represent CDK4-cyclin D1 and CDK6-cyclin D3, respectively; ^c compound included in test set.

Molecular descriptors

The structures of the compounds (Table 1), under study, have been drawn in 2D ChemDraw²³ 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 software²⁴ 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)²⁵⁻²⁹ and partial least squares (PLS)³⁰⁻³² 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^{33,34} 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.

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 compound³⁵. 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 $\pm x(\text{s.d.})$ 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 491 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 491 to 84 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 35 compounds were divided into training-set and test-set. Ten 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 twenty five 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 three descriptors shown below

$$pK_i = 8.171 - 1.784(0.540)\text{CIC3} + 2.143(0.407)\text{nNR2} - 1.801(0.345)\text{ARR}$$

$n = 25$, $r = 0.828$, $s = 0.388$, $F = 15.314$, $Q^2_{LOO} = 0.576$, $Q^2_{L50} = 0.537$

$r^2_{Test} = 0.440$, $FIT = 1.351$, $LOF = 0.220$, $AIC = 0.208$

(1)

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), the identified model has not shown any chance correlation.

The descriptors CIC3, nNR2 and ARR participated in above model are topological, functional and empirical class descriptors, respectively. The negative influence of descriptors CIC3 and ARR on the activity suggested that lower values of complimentary information content of 3rd order neighborhood symmetry (descriptor CIC3) and aromatic ratio (descriptor ARR) would be beneficiary to the activity. The positive contribution of other participated descriptor nNR2 suggested that higher number of tertiary aliphatic amines in a molecular structure would be favorable to the activity. This model in three descriptors could account for nearly 69% variance in the observed activities. Considering the number of observations models up to four descriptors have been explored. A total number of 12 models in four descriptors (sharing 20 descriptors) for the CDK4 inhibitory activity were obtained. All these 20 descriptors along with their brief meaning, average regression coefficients, and total incidence are listed in Table 2, which will serve as a measure of their estimate across these models.

Table 2: Identified descriptors^a along with their class, physical meaning, average regression coefficient and incidence^b

Descriptor class	Descriptor (physical meaning), avregcoeff (incidence)
Constitutional descriptors (CONST):	Ss(sum of Kier-Hall electrotopological states), -1.029(1); Mv (mean atomic volume scaled on Carbon atom), -1.907(1); Mp (mean atomic polarizability scaled on Carbon atom), -1.805(1); nBM (number of multiple bonds), -1.853(1); nCIC (number of rings), 2.015(1); RBN (number of rotatable bonds), -1.033(1);
Topological descriptors (TOPO):	SPI (superpendentic index), -0.647(2); IDDE (mean information content on the distance degree equality), -0.992(5); IC2 (information content index of 2 nd order neighborhood symmetry), 0.748(1); SIC2 (structural information content of 2 nd order neighborhood symmetry), 0.903(1); CIC3 (complimentary information content of 3 rd order neighborhood symmetry), -1.595(2);

BCUT	BEHv1 (highest eigenvalue n.1 of Burden matrix/ weighted by atomic van der Waals volumes), -14-223(1); BEHv3 (highest eigenvalue n.3 of Burden matrix/ weighted by atomic van der Waals volumes), 0.858(1); BELv4 (lowest eigenvalue n.4 of Burden matrix/ weighted by atomic van der Waals volumes), 1.176(1)
2D autocorrelations (2D-AUTO):	MATS7m (Moran autocorrelation - lag 7 / weighted by atomic masses), -0.995(7); MATS2p, (Moran autocorrelation - lag 2 / weighted by atomic polarizabilities), 2.877(1)
Functional groups	nNR2 (number of tertiary aliphatic amines), 1.805(4); nHDon (number of donor atoms for H-bonds with N and O), 1.579(2);
Atom centered fragments (ACF)	H-047 (H attached to C1(sp ³) / C0(sp ²), 1.566(6)
Empirical Descriptors (EMP)	ARR (aromatic ratio), -1.891(2)

^aThe 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 0.800 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 25 compounds. ^bThe average regression coefficient of the descriptor corresponding to all models and the total number of its incidence. The arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

The selected models in four descriptors are mentioned below

$$pK_i = 9.783 - 2.051(0.302)Mv - 1.447(0.371)IDDE - 1.358(0.235)MATS7m + 1.743(0.454)H-047$$

$$n = 25, r = 0.887, s = 0.327, F = 18.618, Q^2_{LOO} = 0.677, Q^2_{L50} = 0.638$$

$$r^2_{Test} = 0.655, FIT = 1.816, LOF = 0.185, AIC = 0.160 \quad (2)$$

$$pK_i = 6.627 - 1.853(0.385)nBM - 0.750(0.221)MATS7m + 1.580(0.389)nHDon + 2.993(0.469)H-047$$

$$n = 25, r = 0.873, s = 0.347, F = 16.024, Q^2_{LOO} = 0.612, Q^2_{L50} = 0.527$$

$$r^2_{Test} = 0.651, FIT = 1.563, LOF = 0.208, AIC = 0.180 \quad (3)$$

$$pK_i = 8.321 - 0.739(0.360)IDDE - 1.591(0.511)CIC3 + 2.495(0.416)nNR2 - 1.907(0.325)ARR$$

$$n = 25, r = 0.860, s = 0.362, F = 14.284, Q^2_{LOO} = 0.582, Q^2_{L50} = 0.579$$

$$r^2_{Test} = 0.609, FIT = 1.393, LOF = 0.227, AIC = 0.196 \quad (4)$$

$$pK_i = 8.269 - 0.556(0.274)SPI - 1.598(0.512)CIC3 + 2.058(0.382)nNR2 - 1.874(0.324)ARR$$

$$n = 25, r = 0.859, s = 0.363, F = 14.199, Q^2_{LOO} = 0.580, Q^2_{L50} = 0.688$$

$$r^2_{Test} = 0.513, FIT = 1.385, LOF = 0.228, AIC = 0.197 \quad (5)$$

These models have accounted for nearly 79% 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 pK_i values of training set compounds calculated using Eqs. (2) to (5) have been included in Table

3. The models (2) to (5) are validated with an external test set of 10 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 3. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

Table 3: Observed and calculated activities for the CDK4 inhibition.

Cpd.	CDK4 Inhibition $pK_i(M)^a$						CDK6 Inhibition $pK_i(M)^a$				
	Obs.	Calculated					Obs.	Calculated			
		Eq.(2)	Eq.(3)	Eq.(4)	Eq.(5)	PLS		Eq.(6)	Eq.(7)	Eq.(8)	Eq.(9)
1 ^b	9.00	_{-b}	_{-b}	_{-b}	_{-b}	_{-b}	7.47	_{-b}	_{-b}	_{-b}	_{-b}
2	9.00	8.38	8.37	8.82	8.82	8.67	8.10	7.89	8.20	7.66	7.98
3 ^c	8.70	9.14	8.73	8.95	8.80	8.76	8.05	7.89	8.60	7.66	7.98
4	8.70	9.08	8.71	8.97	8.95	8.70	7.96	7.81	7.79	7.82	7.90
5	8.22	7.79	7.83	7.76	8.02	8.06	7.03	7.51	7.71	7.15	7.53
6 ^c	8.40	7.95	7.94	8.16	8.21	8.09	7.52	7.06	8.18	7.24	7.65
7	8.52	8.31	8.57	8.28	8.36	8.81	6.88	7.08	6.95	6.60	7.16
8	7.15	7.51	7.19	7.70	7.75	7.54	6.59	6.86	6.84	7.07	6.76
9	8.15	8.18	8.08	8.19	8.26	7.99	7.26	7.28	7.10	7.14	7.44
10	6.24	6.43	6.44	6.62	6.60	6.00	_{-d}	_{-d}	_{-d}	_{-d}	_{-d}
11	8.70	8.75	8.28	8.28	8.14	8.50	8.00	7.89	8.27	7.82	7.69
12 ^c	8.70	8.60	8.54	8.48	8.54	8.51	8.00	7.81	8.12	7.95	7.62
13	8.00	8.02	7.60	7.70	7.96	8.05	7.51	7.42	7.55	7.38	7.24
14 ^c	7.15	7.27	7.38	7.22	7.23	7.81	6.27	6.90	6.55	6.45	6.75
15 ^b	8.30	_{-b}	_{-b}	_{-b}	_{-b}	_{-b}	7.18	_{-b}	_{-b}	_{-b}	_{-b}
16	7.72	7.46	7.56	7.46	7.58	7.62	6.31	5.91	6.37	6.41	6.13
17	7.59	7.75	7.98	7.93	7.77	7.79	7.00	7.19	6.91	7.07	7.13
18	8.30	7.64	8.34	7.65	7.70	8.20	7.96	8.03	7.61	7.59	7.58
19 ^c	7.80	7.81	8.24	8.30	8.43	8.39	7.55	7.93	7.36	7.69	7.58
20	8.52	8.56	8.60	8.35	8.28	8.47	7.82	7.93	7.89	7.70	7.58
21	7.68	7.85	7.70	7.54	7.36	7.65	6.98	7.55	6.94	7.19	7.12
22	7.39	7.64	7.81	7.71	7.76	7.82	7.09	7.10	7.50	7.29	7.25
23	8.10	8.29	8.83	8.26	8.42	8.54	8.70	8.60	8.51	8.47	8.37
24	9.00	8.82	9.03	9.08	8.99	8.63	8.52	8.48	8.43	8.58	8.29
25	8.40	8.30	8.16	7.99	7.91	7.97	8.22	8.07	7.86	8.28	7.91
26	8.10	7.84	8.33	8.32	8.48	8.24	7.96	7.68	8.04	8.38	8.37
27	7.96	8.35	7.99	8.26	8.03	8.05	8.15	8.19	8.06	8.25	8.29
28	8.22	8.46	8.53	8.30	8.45	8.56	8.05	7.74	7.76	8.13	8.60
29 ^c	8.52	7.84	8.31	8.27	8.36	8.31	7.85	7.91	7.72	8.32	8.60
30	8.52	8.43	8.28	8.34	8.08	8.42	8.15	8.19	8.32	7.86	7.98
31 ^c	7.85	8.02	8.18	8.64	8.76	8.56	8.00	8.09	7.88	7.87	7.91
32 ^c	9.00	8.72	8.51	8.77	8.72	8.64	8.52	8.09	8.27	7.99	7.91
33	8.70	8.56	8.49	8.73	8.86	8.55	8.22	8.00	7.76	8.15	7.90
34 ^c	7.47	7.74	7.64	7.72	7.86	7.91	7.64	7.70	7.68	7.64	7.53
35 ^c	8.22	8.43	7.76	8.40	8.14	8.07	7.70	7.26	7.85	7.71	7.58
36	7.41	7.87	7.86	8.05	7.84	7.61	7.00	7.47	6.79	7.68	7.37
37	9.00	9.00	8.74	9.02	8.94	8.84	7.51	7.10	7.80	7.29	7.42

^aOn molar basis, reference [22]; ^bOutlier compound in present study; ^cCompound included in test set and ^dActivity not reported.

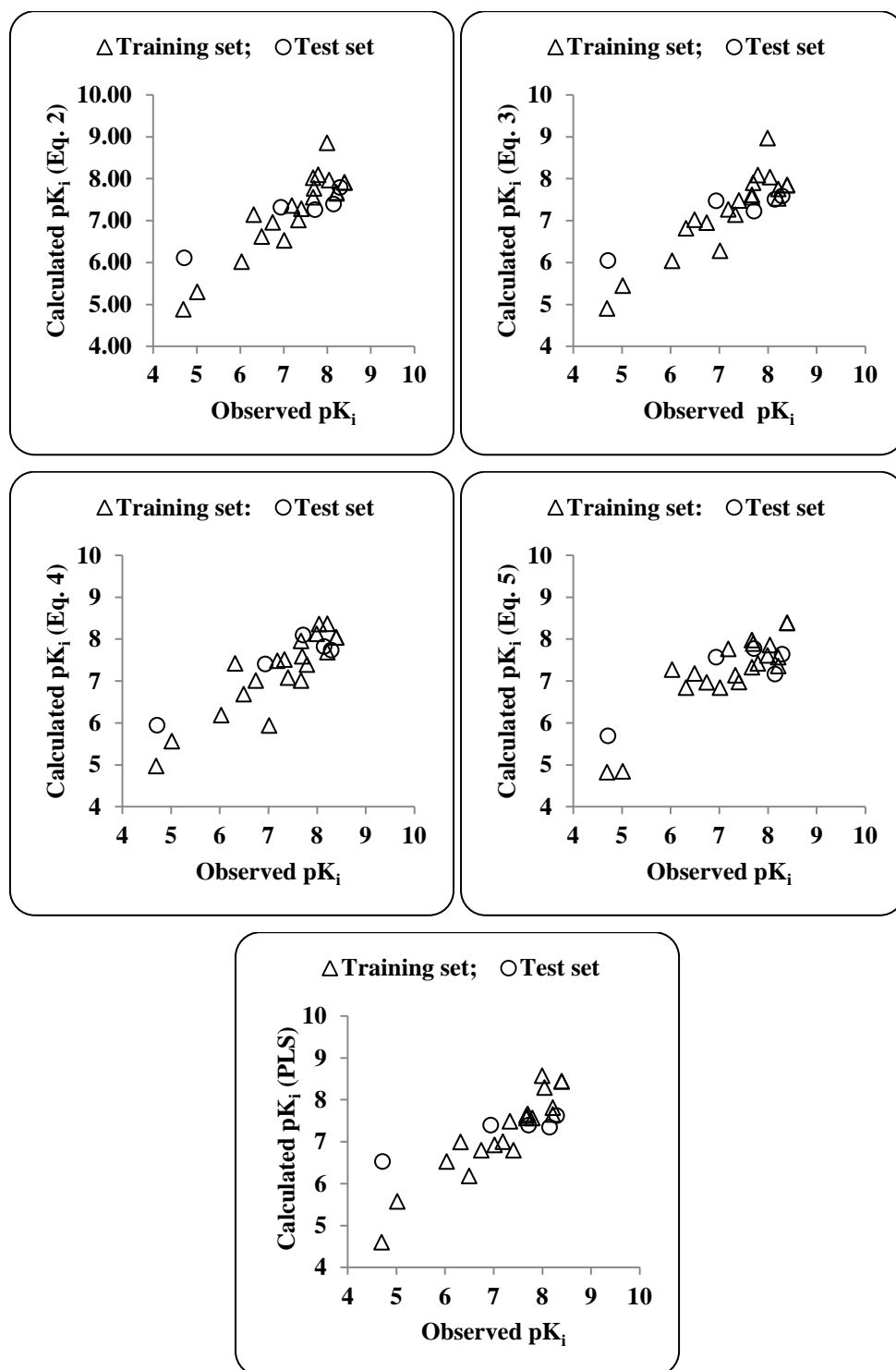


Figure 1: Plot of observed and calculated pK_i values of training- and test-set compounds for CDK4 inhibition.

It is evident from the signs of the regression coefficients that the newly appeared topological class descriptors IDDE (mean information content on the distance degree equality) and SPI (super pendent index); 2D-autocorrelation descriptor MATS7m (Moran autocorrelation - lag 7 / weighted by atomic masses); and constitutional descriptors Mv (mean atomic volume scaled on Carbon atom) and nBM (number of multiple bonds) contributed negatively to the activity. Thus lower values of descriptors IDDE, SPI, MATS7m, Mv and nBM would be beneficiary to the activity. The other participated descriptors in above models are H-047 from

atom centered class and nHDon from functional group. The positive correlation to the activity of both of these descriptors advocated that presence of structural fragment H attached to C1(sp³)/C0(sp²) (descriptor H-047) and number of donor atoms for H-bonds with N and O (descriptor nHDon) would be favorable to the activity.

A partial least square (PLS) analysis has been carried out on these 20 CP-MLR identified descriptors (Table 2) to facilitate the development of a “single window” structure–activity model. For the purpose of PLS, the descriptors have been auto scaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, three components are found to be the optimum for these 20 descriptors and they explained 85% variance in the activity. The MLR-like PLS coefficients of these 20 descriptors are given in Table 4.

Table 4: PLS and MLR-like PLS models from the descriptors of four parameter CP-MLR models

A: PLS equation				
PLS components		PLS coefficient (s.e.) ^a		
Component-1		-0.233(0.024)		
Component-2		0.127(0.028)		
Component-3		0.127(0.037)		
Constant		8.131		
B: MLR-like PLS equation				
S. No.	Descriptor	MLR-like coefficient ^b	(fraction contribution) ^c	Order
1	Ss	-0.040	-0.006	18
2	Mv	-0.587	-0.084	4
3	Mp	-0.739	-0.107	1
4	nBM	-0.285	-0.040	12
5	nCIC	0.149	0.040	13
6	RBN	-0.163	-0.025	15
7	SPI	-0.444	-0.076	5
8	IDDE	-0.171	-0.028	14
9	IC2	0.297	0.049	11
10	SIC2	0.033	0.005	19
11	CIC3	-0.376	-0.049	10
12	BEHv1	0.798	0.015	17
13	BEHv3	-0.011	-0.002	20
14	BELv4	0.495	0.057	8
15	MATS7m	-0.324	-0.063	7
16	MATS2p	0.366	0.051	9
17	nNR2	0.621	0.103	3
18	nHDon	0.809	0.103	2
19	H-047	0.603	0.074	6
20	ARR	-0.157	-0.024	16
		Constant	7.530	
C: PLS regression statistics		Values		
n		25		
r		0.922		
s		0.267		
F		40.072		

FIT	3.535
LOF	0.104
AIC	0.099
Q^2_{LOO}	0.738
Q^2_{L5O}	0.823
r^2_{Test}	0.507

^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.

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.

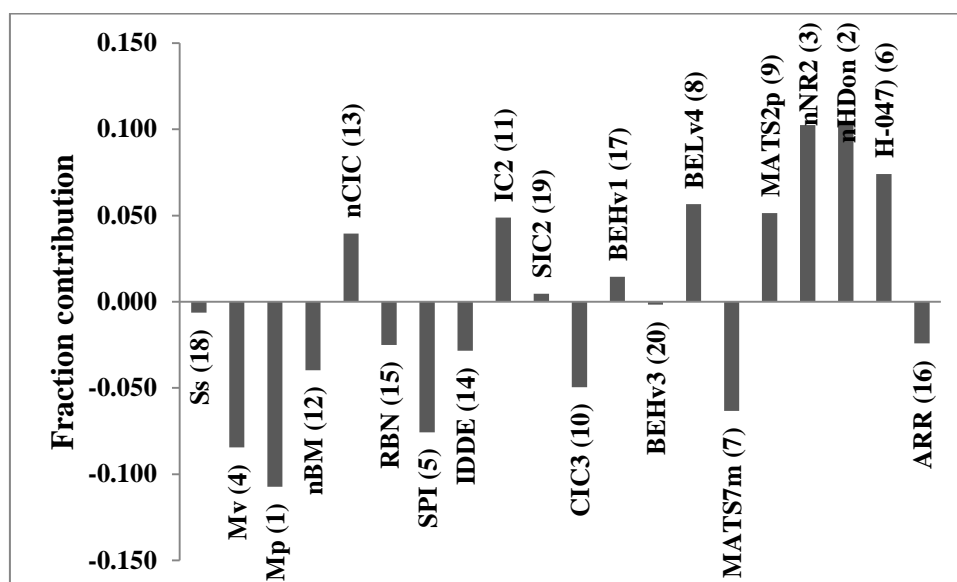


Figure 2: Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 20 CP-MLR identified descriptors (Table 2) associated with CDK4 inhibitory activity of 4-Thiazol-*N*-(pyridin-2-yl)pyrimidin-2-amines.

The PLS analysis has suggested Mp (mean atomic polarizability scaled on Carbon atom) as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 3 in Table 4; Figure 2). The other nine descriptors in decreasing order of significance are nHDon, nNR2, Mv, SPI, H-047, MATS7m, BELv4 (lowest eigenvalue n.4 of Burden matrix/ weighted by atomic van der Waals volumes), MATS2p (Moran autocorrelation - lag 2 / weighted by atomic polarizabilities) and CIC3. It is also observed that PLS model from the data set devoid of CP-MLR identified 20 descriptors (Table 2) is inferior in explaining the activity of the analogues.

CP-MLR analysis has also been carried out for another reported inhibition activity CDK6 using the same test set. Following are the selected (from the 66 models, sharing 51 descriptors) highly significant four-descriptor models for the CDK6 inhibitory activities emerged through CP-MLR.

$$pK_i = 6.503 + 0.976(0.202)AMW - 3.135(0.375)nBM + 2.051(0.380)nN - 1.631(0.403)C-008$$

$$n = 24, r = 0.910, s = 0.292, F = 23.063, Q^2_{LOO} = 0.575, Q^2_{L50} = 0.602$$

$$r^2_{Test} = 0.617, FIT = 2.306, LOF = 0.151, AIC = 0.130 \quad (6)$$

$$pK_i = 5.272 + 3.245(0.398)nCIC - 32.175(4.165)BEHv1 + 3.609(0.540)MATS2p + 0.800(0.155)nNHR$$

$$n = 24, r = 0.910, s = 0.292, F = 22.926, Q^2_{LOO} = 0.793, Q^2_{L50} = 0.775$$

$$r^2_{Test} = 0.688, FIT = 2.292, LOF = 0.152, AIC = 0.130 \quad (7)$$

$$pK_i = 5.312 + 1.518(0.229)GATS7e + 2.546(0.509)C-005 + 6.261(0.874)C-006 - 5.799(1.011)H-047$$

$$n = 24, r = 0.905, s = 0.300, F = 21.551, Q^2_{LOO} = 0.702, Q^2_{L50} = 0.694$$

$$r^2_{Test} = 0.747, FIT = 2.155, LOF = 0.160, AIC = 0.137 \quad (8)$$

$$pK_i = 7.705 - 1.600(0.356)Mp - 8.256(2.917)BEHv1 - 1.537(0.308)C-001 + 2.244(0.467)C-006$$

$$n = 24, r = 0.894, s = 0.315, F = 19.054, Q^2_{LOO} = 0.635, Q^2_{L50} = 0.658$$

$$r^2_{Test} = 0.568, FIT = 1.905, LOF = 0.177, AIC = 0.152 \quad (9)$$

The descriptors participated in above models are from constitutional class (AMW, Mp, nN, nBM and nCIC) and atom centered class (H-047, C-001, C-005, C-006 and C-008). Among the constitutional class descriptors, AMW (average molecular weight), nN (number of nitrogen atoms) and nCIC (number of rings) have shown positive and descriptors Mp (mean atomic polarizability scaled on Carbon atom) and nBM (number of multiple bonds) negative correlation to the CDK6 activity. Thus a higher value of descriptors AMW, nN and nCIC, and a lower value of descriptors Mp and nBM would be beneficial to the activity. Atom centered descriptors C-005 and C-006 contributed positively and descriptors H-047, C-001 and C-008 negatively to the activity advocating that presence of CH3X (descriptor C-005) and CH2RX (descriptor C-006), and absence of H attached to C1(sp3)/C0(sp2) (descriptor H-047), CH3R/CH4 (descriptor C-001) and CHR2X (descriptor C-008) type atom centered fragments in a molecular structure would be favorable to the activity. On the similar ground, a lower value of BCUT class descriptor BEHv1 (highest eigenvalue n.1 of Burden matrix/ weighted by atomic van der Waals volumes), higher values of 2D-autocorrelation descriptors MATS2p (Moran autocorrelation - lag 2 / weighted by atomic polarizabilities) and GATS7e

(Geary autocorrelation of $-\text{lag}7$ weighted by atomic Sandersons electro negativities), and presence of secondary aliphatic amine functionality (descriptor nNHR, a functional group class descriptor) in a molecular structure will augment the inhibition activity.

These models have accounted for nearly 83% 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 pK_i values of training set compounds calculated using Eqs. (6) to (9) have been included in Table 3. The models (6) to (9) are validated with an external test set of 10 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 3. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 3.

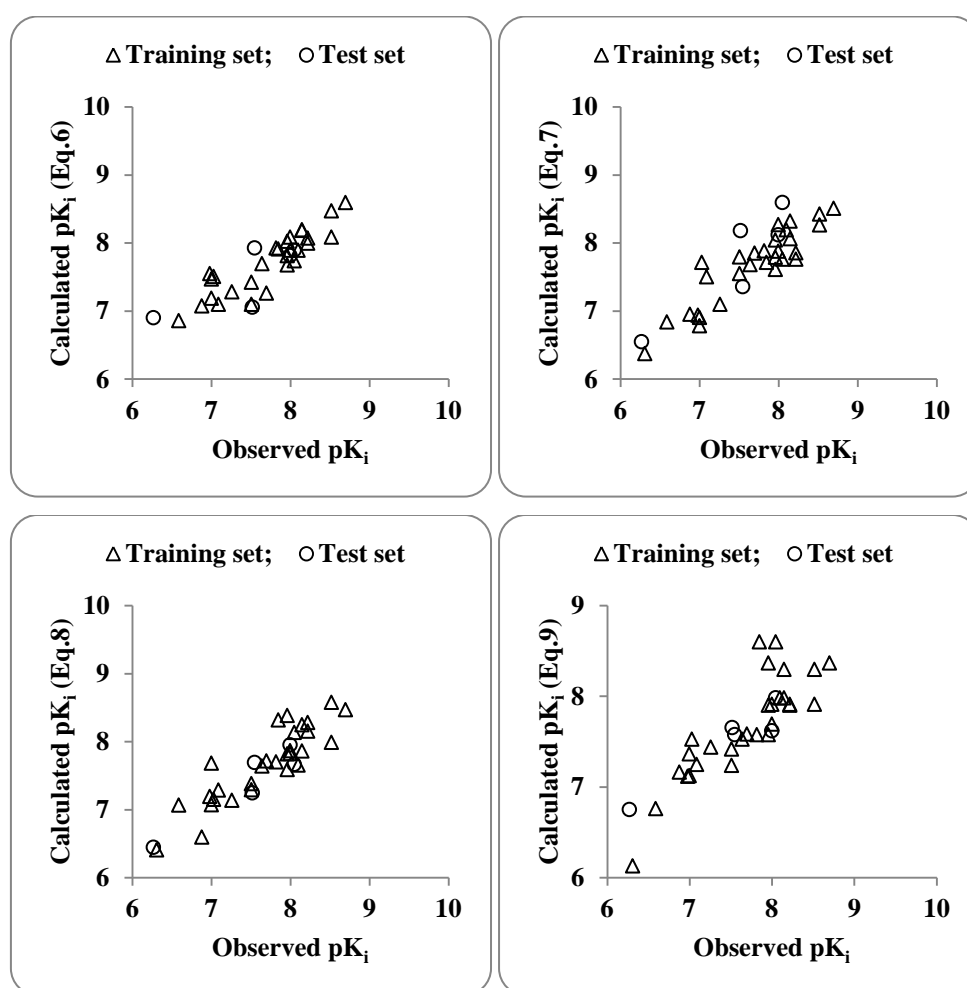


Figure 3: Plot of observed and calculated pK_i values of training- and test-set compounds for CDK6 inhibition.

Applicability domain

On analyzing the applicability domain (AD) for the CDK4 inhibitory actions in the Williams plot (Figure 4) of the model based on the whole data set (Table 5), no any compound has been identified as an obvious ‘outlier’ for the CDK4 inhibitory activity if the limit of normal values for the Y outliers (response outliers) was set as $3 \times (\text{standard deviation})$ 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 5: Models derived for the whole data set (n = 35) in descriptors identified through CP-MLR.

Model	r	s	F	Q ² _{LOO}	Eq.
$pK_i = 9.621 - 2.002(0.281)Mv - 1.256(0.300)IDDE - 1.266(0.176)MATS7m + 1.707(0.392)H-047$	0.872	0.325	23.920	0.687	(2a)
$pK_i = 6.820 - 2.012(0.334)nBM - 0.844(0.163)MATS7m + 1.293(0.336)nHDon + 3.002(0.404)H-047$	0.864	0.335	22.136	0.629	(3a)
$pK_i = 8.100 - 0.756(0.284)IDDE - 1.278(0.281)CIC3 + 2.491(0.317)nNR2 - 1.774(0.301)ARR$	0.853	0.347	20.047	0.625	(4a)
$pK_i = 8.269 - 0.556(0.274)SPI - 1.598(0.512)CIC3 + 2.058(0.382)nNR2 - 1.874(0.324)ARR$	0.840	0.361	18.056	0.592	(5a)

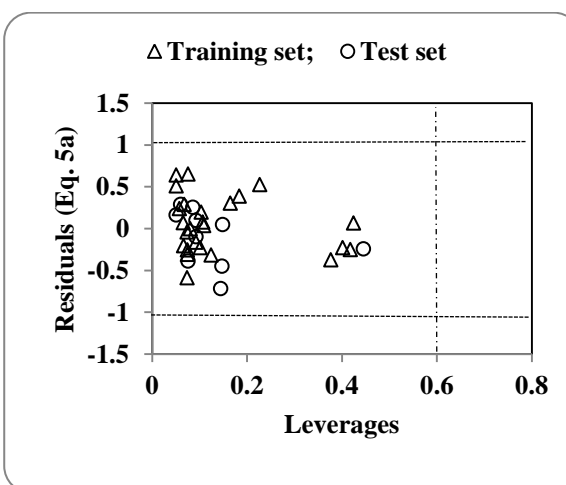
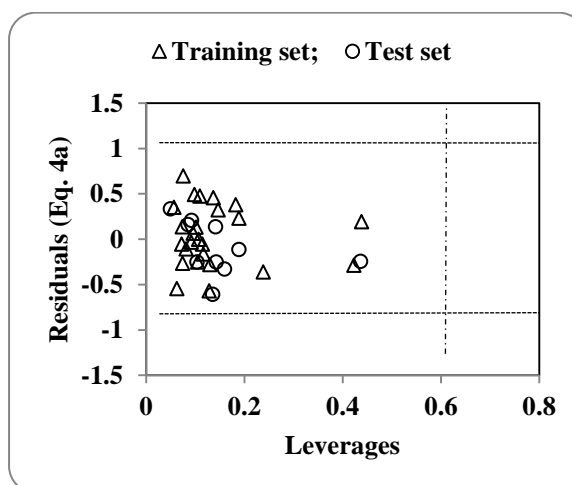
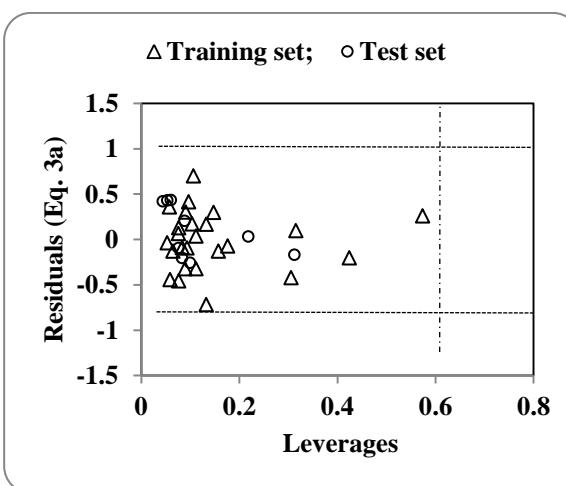
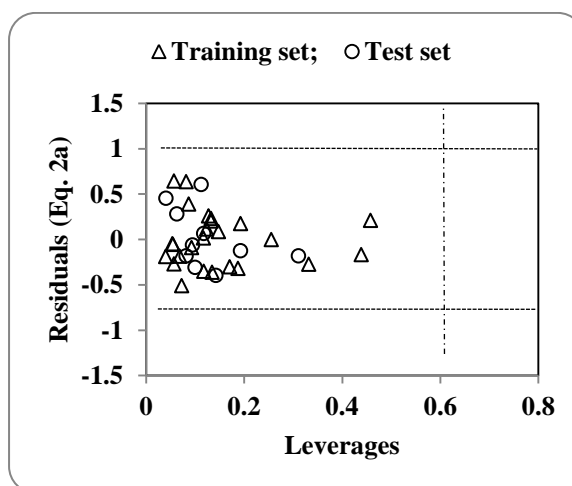


Figure 4: Williams plot for the training- and test- set compounds. The horizontal dotted line refers to the residual limit ($\pm 3 \times$ standard deviation) and the vertical dotted line represents threshold leverage h^* (= 0.6).

CONCLUSION

The CDK4 and CDK6 inhibition activity of 4-Thiazol-*N*-(pyridin-2-yl)pyrimidin-2-amines have 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 CDK4 inhibitory activity have highlighted the role of complimentary information content of 3rd order neighborhood symmetry (CIC3), mean information content on the distance degree equality (IDDE) and super pendent index (SPI) to rationalize the activity. Atomic properties such as atomic van der Waals volumes and polarizabilities in terms of *M_v*, and Moran autocorrelation (MATS7m) in addition to the aromatic ration (ARR) number of multiple bonds (nBM), tertiary aliphatic amines (nNR2) and number of donor atoms for H-bonds with N and O (nHDon) functionality, and structural fragment H attached to C1(sp³)/C0(sp²) (H-047) have also shown prevalence to model the CDK4 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 CDK6 inhibitory activity have revealed that the average molecular weight (descriptor AMW), mean atomic polarizability scaled on Carbon atom (descriptor Mp), number of multiple bonds (nBM), number of nitrogen atoms (nN), number of rings (nCIC), number of secondary aliphatic amines (nNHR), highest eigenvalue n.1 of Burden matrix/ weighted by atomic van der Waals volumes (BEHv1), Moran autocorrelation - lag 2 / weighted by atomic polarizabilities (MATS2p) and Geary autocorrelation of - lag 7/ weighted by atomic Sandersons electronegativities (GATS7e) played a pivotal role in rationalization of CDK6 inhibition activity of titled compounds. Additionally, absence of attached to C1(sp³)/C0(sp²) (H-047), CH3R/CH4 (C-001) and CHR2X (C-008), and presence of CH3X (C-005) and CH2RX (C-006) type structural fragment in a molecular structure are also predominant to explain CDK6 inhibition actions of 4-Thiazol-*N*-(pyridin-2-yl)pyrimidin-2-amines.

ACKNOWLEDGEMENT

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

REFERENCES

1. Pardee ABA. Restriction point for control of normal animal cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* 1974, 71, 1286-1290.
2. Siskin JE, Morasca L. Intra population kinetics of the mitotic cycle. *J. Cell Biol.* 1965, 25, 179-189.
3. Bartek J, Bartkova J, Lukas J. The retinoblastoma protein pathway in cell cycle control and cancer. *Exp. Cell Res.* 1997, 237, 1-6.
4. Malumbres M, Barbacid M. Milestones in cell division: to cycle or not to cycle: a critical decision in cancer. *Nat. Rev. Cancer* 2001, 1, 222-231.
5. Ortega S, Malumbres M, Barbacid M. Cyclin D-dependent kinases, INK4 inhibitors and cancer. *Biochim. Biophys. Acta. Rev. Cancer* 2002, 1602, 73-87.
6. Malumbres M, Ortega S, Barbacid M. Genetic analysis of mammalian cyclin dependent kinases and their inhibitors. *J. Biol. Chem.* 2000, 381, 827-838.
7. Fry DW, Harvey PJ, Keller PR, Elliott WL, Meade M, Trachet E, Albassam M, Zheng X, Leopold WR, Pryer NK, Toogood PL. Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts. *Mol. Cancer Ther.* 2004, 3, 1427-1438.
8. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer* 2009, 9, 153-166.
9. Nevins JR. The Rb/E2F pathway and cancer. *Hum. Mol. Genet.* 2001, 10, 699-703.
10. Tadesse S, Yu M, Kumarasiri M, Le BT, Wang S. Targeting CDK6 in cancer: state of the art and new insights. *Cell Cycle* 2015, 14, 3220-3230.
11. DiPippo AJ, Patel NK, Barnett CM. Cyclin-dependent kinase inhibitors for the treatment of breast cancer: past, present, and future. *Pharmacotherapy* 2016, 36, 652-667.
12. Uras IZ, Walter GJ, Scheicher R, Bellutti F, Prchal-Murphy M, Tigan AS, Valent P, Heidel FH, Kubicek S, Scholl C, Frohling S, Sexl V. Palbociclib treatment of FLT3-ITD+ AML cells uncovers a kinase-dependent transcriptional regulation of FLT3 and PIM1 by CDK6. *Blood* 2016, 127, 2890-2902.
13. Goel S, Wang Q, Watt AC, Tolaney SM, Dillon DA, Li W, Ramm S, Palmer AC, Yuzugullu H, Varadan V, Tuck D, Harris LN, Wong KK, Liu XS, Sicinski P,

- Winer EP, Krop IE, Zhao JJ. Overcoming therapeutic resistance in HER2-positive breast cancers with CDK4/6 inhibitors. *Cancer Cell* 2016, 29, 255-269.
14. Sumi NJ, Kuenzi BM, Knezevic CE, Rensing Rix LL, Rix U. Chemoproteomics reveals novel protein and lipid kinase targets of clinical CDK4/6inhibitors in lung cancer. *ACS Chem. Biol.* 2015, 10, 2680-2686.
 15. Chen P, Lee N, Hu W, Xu M, Ferre RA, Lam H, Bergqvist S, Solowiej J, Diehl W, He YA, Yu X, Nagata A, VanArsdale T, Murray BW. Spectrum and degree of CDK drug interactions predicts clinical performance. *Mol. Cancer Ther.* 2016, 15, 2273-2281.
 16. Muller S, Chaikuad A, Gray NS, Knapp S. The ins and outs of selective kinase inhibitor development. *Nat. Chem. Biol.* 2015, 11, 818-821.
 17. Wu P, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. *Trends.Pharmacol. Sci.* 2015, 36, 422-439.
 18. 18. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 2011, 29, 1046-1051.
 19. Wang S, Meades C, Wood G, Osnowski A, Anderson S, Yuill R, Thomas M, Mezna M, Jackson W, Midgley C, Griffiths G, Fleming I, Green S, McNae I, Wu SY, McInnes C, Zheleva D, Walkinshaw MD, Fischer PM. 2-Anilino-4-(thiazol-5-yl)pyrimidine CDK inhibitors: synthesis, SAR analysis, X-ray crystallography, and biological activity. *J. Med. Chem.* 2004, 47, 1662-1675.
 20. Shao H, Shi S, Huang S, Hole AJ, Abbas AY, Baumli S, Liu X, Lam F, Foley DW, Fischer PM, Noble M, Endicott JA, Pepper C, Wang S. Substituted 4-(thiazol-5-yl)-2- (phenylamino)pyrimidines are highly active CDK9 inhibitors: synthesis, X-ray crystal structures, structure-activity relationship, and anticancer activities. *J. Med. Chem.* 2013, 56, 640-659.
 21. Wang S, Griffiths G, Midgley CA, Barnett AL, Cooper M, Grabarek J, Ingram L, Jackson W, Kontopidis G, McClue SJ, McInnes C, McLachlan J, Meades C, Mezna M, Stuart I, Thomas MP, Zheleva DI, Lane DP, Jackson RC, Glover DM, Blake DG, Fischer PM. Discovery and characterization of 2-anilino-4- (thiazol-5-yl) pyrimidine transcriptional CDK inhibitors as anticancer agents. *Chem. Biol.* 2010, 17, 1111-1121.
 22. Tadesse S, Yu M, Mekonnen LB, Lam F, Islam S, Tomusange K, Rahaman MH,Noll B, Basnet SKC, Teo T, Albrecht H, Milne R, Wang S. Highly Potent, Selective and Orally Bioavailable 4-Thiazol-*N*-(pyridin-2-yl)pyrimidin-2-amine

- Cyclin-Dependent Kinase 4 and 6 Inhibitors as Anticancer Drug Candidates: Design, Synthesis and Evaluation. *J. Med. Chem.* 2017, 60, 1892-1915.
23. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. <http://www.cambridgesoft.com>
 24. Dragon software (version 1.11-2001) by Todeschini R.;Consonni V. Milano, Italy. <http://www.taletе.mi.it/dragon.htm>
 25. 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.
 26. 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.
 27. 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-N⁶-substituted-4'-thioadenosine-5'-uronamide. *Eur J Med Chem.* 2009, 44, 1377-1382.
 28. Sharma BK, Paliania 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.
 29. 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.
 30. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 1978, 20, 397-405.
 31. Kettaneh N, Berglund A, Wold S. PCA and PLS with very large datasets. *Comput Stat Data Anal.* 2005, 48, 69-85.
 32. 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.
 33. 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.
 34. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure–activity modeling of the HIV-1RT inhibitory activity of 2,3-diaryl-1,3- thiazolidin-4-ones. *QSAR Comb Sci.* 2004, 23, 234-244.

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

BJMHR is

- **Peer reviewed**
- **Monthly**
- **Rapid publication**
- **Submit your next manuscript at**

editor@bjmhr.com

