Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 41 (2006) 1190–1195
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://france.elsevier.com/direct/ejmech

Original article

Exploring QSAR on 3-aminopyrazoles as antitumor agents for their inhibitory activity of CDK2/cyclin A

Soma Samanta[a], Bikash Debnath[a], Anindya Basu[a], Shovanlal Gayen[a]
Kolluru Srikanth[b], Tarun Jha[a,*]

[a] *Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, P.O. Box 17020, Jadavpur University, Kolkata 700032, India*
[b] *Department of Medicinal Chemistry, College of Pharmacy, 8-125 Weaver Densford Hall, 308, Harvard St SE, University of Minnesota, Minneapolis MN 55455, USA*

Received in revised form 3 May 2006; accepted 11 May 2006
Available online 30 June 2006

Abstract

Chemical inhibitors of cyclin-dependent kinases have great therapeutic potential against various proliferative and neurodegenerative disorders. The pharmacophoric requirement of 3-aminopyrazole, inhibitors of CDK2/cyclin A as antitumor agents was explored. QSAR study was performed using ETSA index, RTSA index, indicator parameters and atomic charges to consider quantitatively the effect of the structural variation on the antitumor activity of 3-aminopyrazole. Result showed that atom number 5 is important for the activity. It plays some electronic roles in the interaction of these compounds with enzymes as well as assumed to be involved through the dispersive/van der Waals interactions with enzyme. Presence of meta substitutions on the phenyl ring indicate the detrimental effects towards the activity. The presence of substituted biphenyl/2-thenyl phenyl at R1 are favorable towards the activity. QSAR study also indicates that with increasing the electronegativity of oxygen at position 8, the activity increases.

Keywords: CDK2/cyclin A inhibitor; 3-Aminopyrazole; Quantitative structure–activity relationship; Electrotopological state atom index; Refractotopological state atom index

1. Introduction

Eukaryotic cell cycle progression is governed by cyclindependent kinases (CDKs), a family of enzymes that are regulated by phosphorylation and activated by their association with cyclins [1,2]. CDKs play a key role in cell-cycle control (e.g. CDKs 1–4 as well as 6–7), in thymocyte apoptosis (CDK2), in neuronal functions (CDK5) and in transcription (CDKs 7–9). Even, CDKs and their regulators are very frequently deregulated in human tumors. On the other hand, CDK5 is important in neurodegenerative disorders such as Alzheimer's disease and Parkinson's disease. The search for the chemical inhibitors of CDKs has intensified over the past few years. All CDK inhibitors, identified so far, function by competing with ATP for binding to the catalytic site. Various CDKs, which are serine/threonine protein kinase enzymes, act by complexing with corresponding cyclin subunits, and thus regulate the eukaryotic cell cycle. The frequent deregulation of cell cycle progression in cancer has prompted search for kinase inhibitors with high affinity and specificity for Cdks [3]. CDKs are involved in controlling normal cell proliferation through retinoblastoma pathway. Disruption of the normal cell cycle mechanism in abnormal proliferation of cancer cells can be occur by making complexes between CDKs and cyclins by inhibiting their kinase activity. Chemical inhibitors of CDKs have great therapeutic potential against various proliferative and neurodegenerative disorders [4]. Pevarello et al. [5] reported a new class of CDK2/cyclin A inhibitors and based on SAR find a lead as antitumor agents with good in vivo bioavailability and efficacy in an animal tumor model. Here, we tried to find pharmacophoric requirement of 3-aminopyrazole which are inhibitors of CDK2/cyclin A as antitumor agents. As a part of our composite program of drug design, discovery and development [6–8], QSAR study was performed using ETSA index, RTSA index, indicator parameters and atomic charges to consider quantita-

* Corresponding author.
E-mail address: tjjupharm@yahoo.com (T. Jha).

0223-5234/$ - see front matter © 2006 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.05.007

tively the effect of the structural variation on the antitumor activity of 3-aminopyrazole, inhibitors of CDK2/cyclin A. The structural details and activity data were collected from published work by Pevarello et al. [5].

2. Materials and methods

2.1. Parameters and data set used to develop QSAR model

For the development of QSAR models on 3-aminopyrazole derivatives, the antitumor activity by inhibiting CDK2/cyclin A was utilized. Electrotopological state atom index (ETSA), refractotopological state atom index (RTSA), indicator parameters and atomic charges were used for QSAR study. Arbitrary numbering (shown in Fig. 1) was used to calculate ETSA and RTSA indices. Negative logarithm of CDK2/cyclin A inhibitory activity of 3-aminopyrazole (pIC_{50}) were used for developing QSAR model to get linear relationship with independent variables. pIC_{50} values of all the compounds are given in Table 1.

From the total data set eight compounds (**cpd no 6, 10, 14, 19, 24, 27, 36, 42**) were chosen as the test set and the remaining 34 compounds were treated as the training set. k-means

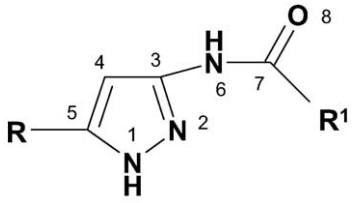

Fig. 1. General structure of 3-aminopyrazole.

Table 1
CDK2/cyclin A inhibitory activity, standardized value of ETSA indices, RTSA index, atomic charge at atom 8 of 3-aminopyrazoles

Compound[a]	R	R_1	S_5	S_6	R_5	q8	I_1	I_2	pIC_{50}
1	CH_3-	C_6H_5-	0.184	0.536	0.090	0.360	0.000	0.000	–3.176
2	CH_3-	C_3H_7-	0.292	0.495	0.113	0.281	0.000	0.000	–4.000
3	C_3H_5-	C_3H_7-	1.000	0.707	0.766	0.281	0.000	0.000	–2.350
4	C_3H_5-	C_6H_5-	0.892	0.749	0.740	0.360	0.000	0.000	–2.462
5	C_3H_5-	4-Br-C_6H_4-	0.868	0.759	0.652	0.180	0.000	0.000	–1.531
6[b]	C_3H_5-	4-Cl-C_6H_4-	0.819	0.687	0.726	0.202	0.000	0.000	–1.929
7	C_3H_5-	4-OMe-C_6H_4-	0.833	0.738	0.785	0.281	0.000	0.000	–2.848
8	C_3H_5-	4-COOH-C_6H_4-	0.628	0.492	0.832	0.169	0.000	0.000	–2.217
9	C_3H_5-	4-$CONH_2$-C_6H_4-	0.684	0.573	0.787	0.225	0.000	0.000	–1.833
10[b]	C_3H_5-	3-Br-C_6H_4-	0.865	0.764	0.634	0.315	0.000	0.000	–2.996
11	C_3H_5-	3-Cl-C_6H_4-	0.806	0.666	0.723	0.247	0.000	0.000	–3.000
12	C_3H_5-	3-OMe-C_6H_4-	0.819	0.736	0.797	0.258	0.000	0.000	–2.778
13	C_3H_5-	3-CF_3-C_6H_4-	0.267	0.000	1.000	0.157	0.000	0.000	–2.699
14[b]	C_3H_5-	2-Br-C_6H_4-	0.858	0.777	0.605	0.101	1.000	0.000	–3.176
15	C_3H_5-	2-Cl-C_6H_4-	0.781	0.635	0.719	0.146	1.000	0.000	–3.000
16	C_3H_5-	2-OMe-C_6H_4-	0.802	0.733	0.816	0.022	1.000	0.000	–4.000
17	C_3H_5-	2.6-diCl-C_6H_4-	0.674	0.521	0.697	0.000	1.000	0.000	–4.000
18	C_3H_5-	3,4-diCl-C_6H_4-	0.733	0.606	0.707	0.292	0.000	0.000	–1.919
19[b]	C_4H_7-	4-Br-C_6H_4-	0.903	0.811	0.532	0.292	0.000	0.000	–1.954
20	C_5H_9-	4-Br-C_6H_4-	0.941	0.865	0.411	0.202	0.000	0.000	–1.699
21	C_6H_{11}-	4-Br-C_6H_4-	0.965	0.904	0.333	0.337	0.000	0.000	–3.000
22	CH_3-	4-Br-C_6H_4-	0.160	0.547	000.0	0.326	0.000	0.000	–3.778
23	C_2H_5-	4-Br-C_6H_4-	0.465	0.635	0.087	0.326	0.000	0.000	–2.799
24[b]	n-C_3H_7-	4-Br-C_6H_4-	0.587	0.697	0.071	0.315	0.000	0.000	–2.813
25	i-C_3H_7-	4-Br-C_6H_4-	0.483	0.687	0.173	0.225	0.000	0.000	–2.301
26	Sec-C_4H_9-	4-Br-C_6H_4-	0.601	0.749	0.156	0.348	0.000	0.000	–4.000
27[b]	tert-C_4H_9-	4-Br-C_6H_4-	0.427	0.731	0.258	0.326	0.000	0.000	–4.000
28	C_6H_5-	4-Br-C_6H_4-	0.000	0.725	0.189	0.314	0.000	0.000	–4.000
29	$CH_2C_6H_5$-	4-Br-C_6H_4-	0.326	0.767	0.038	0.202	0.000	0.000	–4.000
30	C_3H_5-	Phenylacetyl	0.927	0.832	0.721	0.888	0.000	0.000	–1.681
31	C_3H_5-	4-$CONH_2$-phenylacetyl	0.757	0.692	0.766	0.775	0.000	0.000	–1.869
32	C_3H_5-	3-OMe-phenylacetyl	0.872	0.824	0.771	0.921	0.000	0.000	–1.462
33	C_3H_5-	4-(2-Pyrrolidin-1-yl)-ethoxy	0.920	0.930	0.870	0.933	0.000	0.000	–1.978
34	C_3H_5-	4-OCF_3-phenylacetyl	0.507	0.422	0.924	0.865	0.000	0.000	–1.826
35	C_3H_5-	4-Biphenylacetyl	0.906	0.951	0.681	0.899	0.000	0.000	–1.748
36[b]	C_3H_5-	4-(3-Flurobiphenyl)acetyl	0.799	0.842	0.728	0.888	0.000	1.000	–0.602
37	C_3H_5-	4-(3-Methylbiphenyl)acetyl	0.906	0.972	0.674	0.899	0.000	1.000	–0.954
38	C_3H_5-	4-(4-Carboxybiphenyl)acetyl	0.674	0.850	0.723	0.876	0.000	1.000	–1.041
39	C_3H_5-	4-(4-Carboxamidobiphenyl)acetyl	0.792	0.868	0.702	0.876	0.000	1.000	–0.602
40	C_3H_5-	4-(2-Thenyl)-phenylacetyl	0.924	0.969	0.631	0.899	0.000	1.000	–0.477
41	C_3H_5-	4-(2-Naphthyl)acetyl	0.917	0.956	0.702	0.888	0.000	0.000	–1.568
42[b]	C_3H_5-	4-(1-Naphthyl)acetyl	0.913	1.000	0.700	1.000	0.000	0.000	–1.580

[a] compound number.
[b] compound chosen for test set using k-means cluster analysis.

cluster analysis (k-MCA) was used in designing training set and test set. QSAR study was performed upon these 34 compounds, which were considered as the training set. Useful models were obtained by using ETSA indices, RTSA indices and atomic charges as well as indicator variables.

2.1.1. Electrotopological state atom index

The electrotopological state atom index [9–13] is a structural descriptor encoding both electronic feature and topological environment of each skeletal atom on a molecule. It has two basic components:

- Intrinsic electrotopological state of an atom;
- Perturbation factor (effect of environment).

This index has been projected to determine pharmacophore moieties of biologically active congeneric compounds. E-state indices atom wise were calculated using the computer program "mouse" [14]. In the program, molecular connection table in specified format and the intrinsic values of different atoms are given as input and output result is the ETSA indices of atoms. Before the calculation the atoms of the molecules were numbered arbitrarily keeping the serial number of atoms same in all molecules.

2.1.2. Refractotopological state atom index

Refractotopological state atom indices [7,15] are based on the atomic refractivities and the topological environment of the atom and sum of atomic refractivities, that is, molar refractivity is directly proportional to the polarizability of a substance, which determines London force/dispersive force between nonpolar molecules. Thus, R-state indices are important for modeling the dispersive/van der Waals interactions with the receptor. The same initial consideration of E-state formalism was taken for R-state index. Atom-wise R state indices were calculated using the computer program "mouse" [14].

2.1.3. Indicator variables

Indicator parameter was used along with the ETSA and RTSA indices in order to find out the role of specific substituent/substituent pattern at specific position towards the activity.

2.1.4. Atomic charge

Atomic charge is the difference between charge on the core and the electron density on the atom. The charge on each atom was calculated using AM1 method on Hyperchem Release 7.0 Pro.Package [16]. For the calculation of charge structure of the compounds were drawn and model build by Hyperchem software. Structures were optimized by AM1 method using Polok Ribiere algorithm with RMS gradient of 0.1 kcal/(A^0 mol).

2.2. Statistical analysis

All the statistical analysis was done by the software "Multi Regress" [17] developed in our laboratory. Correlation analysis [18] of the independent parameters and biological activity was carried out. Intercorrelated parameters were eliminated stepwise depending on their individual correlation with biological activity. Multiple regression analysis [19,20] was carried out using CDK2/cyclinA inhibitory activity of 3-aminopyrazole as the dependent variables and ETSA indices, RTSA indices as well as indicator parameters as independent variables. The statistical quality of the regression equations were justified by parameters like correlation coefficient (R), percentage of explained variance (%EV), adjusted R^2 (R^2_A), variance ratio (F) at specified degrees of freedom (df), probability factor related to F-ratio (P), standard error of estimate (s). Significant level of regression coefficient, intercepts of all equations was determined by t-statistics and P-values of the corresponding parameters. Use of more than one variable in the multivariate equation was justified by correlation analysis. A compound was considered as an outlier if the standard residual is more than thrice of standard error of estimate.

2.3. Validation of the QSAR model

For the internal validation, we have used Leave-One-Out (LOO-) cross validation method [21]. By using this, the predictive powers of the equations were validated. Predicted residual sum of square (PRESS), total sum of squares (SSY), cross-validated R^2 (R^2_{CV}), standard deviation based on PRESS (S_{PRESS}) and standard deviation error of prediction (S_{DEP}) were considered for the validation of these models.

2.4. k-Means cluster analysis

With the aim of designing a test set and the training set k-means cluster analysis (k-MCA) was performed. Before going to the cluster analysis, the standardization of variables was done using the Eq. (1).

$$X_{ki} = (x_i - x_{i(minimum)})/(x_{i(Maximum)} - x_{i(Minimum)}) \quad (1)$$

Standardized values of the variables are shown in Table 1. A k-MCA splits 3-aminopyrazole derivatives in four clusters with 18, 1, 13 and 10 members. Selection of the training and test set was carried out in a random way using compounds belonging to each cluster. From each cluster, 20% members are considered for the designing test set. To ensure statistically acceptable data partition into several clusters the number of members in each cluster and the lower standard deviation of the variables in the cluster (as low as possible) was considered. All the variables showed P levels < 0.05 for Fisher test.

2.5. Prediction of the test set

On the basis of the models build using training set, predicted values of the test set compounds were calculated.

3. Result and discussion

For the QSAR study, total data set was divided in test set and training set as stated in materials and methods section. Correlation analysis was performed among the dependent variable and useful independent variables and the result shown in Table 2.

Table 2
Correlation matrix of biological activity and QSAR parameters

	S_5	S_6	R_5	I_1	I_2	q8	pIC_{50}
S_5	1.00	0.61	0.59	0.10	0.17	–0.30	0.56
S_6		1.00	–0.01	–0.10	0.36	–0.53	0.41
R_5			1.00	0.16	0.16	–0.27	0.51
I_1				1.00	–0.12	0.39	–0.36
I_2					1.00	–0.52	0.61
q8						1.00	–0.69
pIC_{50}							1.00

Intercorrelated parameter was eliminated depending on their individual correlation with the biological activity. All the possible combination of the parameters were considered and subjected to the multiple regression analysis. Statistical parameters used to explain the QSAR equation are given in Table 3.

Various combinations were tried to find the regression equation. The best equation was obtained with using topological parameters, indicator parameters and atomic charges as shown bellow

$$pIC_{50} = -4.129(\pm0.263) + 1.085(\pm0.429)S_5 + 1.027(\pm0.391)R_5 - 1.160(\pm0.361)I_1 + 1.114(\pm0.323)I_2 + 0.734(\pm0.389)q8 \quad (2)$$

$n = 34$; $R = 0.889$; %EV = 78.97; $F(5,28) = 21.04$; $P < 0.00001$; $s = 0.516$; SSY = 35.525; PRESS = 10.312; $R^2_{CV} = 0.709$; $S_{DEP} = 0.551$; $S_{PRESS} = 0.607$.

For the development of the above equation ETSA index (S_5), RTSA index (R_5) and charge at atom number 8 (q8) as well as indicator parameters (I_1 and I_2) were considered. Model explains the variance of the activity 78.97%. Equation (2) clearly showed that atom number 5 is important for the activity. Positive coefficient of the S_5 and R_5 indicate that increasing the value of S_5 and R_5 the CDK2/cyclin A inhibitory activity of the compounds increases. Since E-state indices are the measure of the availability of the π and/or lone pair electrons on the atoms, atom number 5 (C-5) might be playing some electronic roles in the interaction of the compounds with the enzymes. On the other hand, CDK2/cyclin A inhibitory activity of these compounds will increase with increasing the R-state index of atom 5. It can be assumed to be involved through the dispersive/van der Waals interactions with the enzyme. In this equation, two indicator parameters (I_1 and I_2) were used. I_1 stand for the presence of meta substitution of the phenyl ring. The negative coefficient implies the detrimental effects towards the activity. I_2 stands for presence of substituted bi-phenyl/2-thenyl phenyl. The positive coefficient of the I_2 indicates that the presence of substituted bi-phenyl/2-thenyl phenyl is favorable towards the activity. Positive coefficient of charge of atom number 8 indicates that with increase of the value of charge at atom 8 biological activity increases, i.e. with increasing the electronegativity of oxygen at position 8 the activity increases.

Table 3
Statistical parameter used to explained QSAR equations

Statistical parameters	Explanation
n	Number of data point
R	Correlation coefficient
%EV	Percentage of explained variance
R^2_A	Adjusted R^2
F	Ratio between the variances of observed and calculated activities
P	Probability factor related to F-ratio
S	Standard error of estimate
PRESS	Predicted residual sum of squares
SSY	Total sum of squares
R^2_{CV}	Squared cross validated correlation coefficient
S_{PRESS}	Standard error of PRESS
S_{DEP}	Standard deviation of error of prediction
R^2_{pred}	Squared predicted correlation coefficient

Stepwise deletion of the compound 25 and 5 improve the quality of the relation statistically as follows

$$pIC_{50} = -4.271(\pm0.242) + 0.913(\pm0.309)S_5 + 1.163(\pm0.356)R_5 - 1.002(\pm0.328)I_1 + 1.092(\pm0.289)I_2 + 0.974(\pm0.360)q8 \quad (3)$$

$n = 32$; $R = 0.917$; %EV = 84.03; $F(5,26) = 27.368$; $P < 0.00001$; $s = 0.416$; SSY = 34.669; PRESS = 8.107; $R^2_{CV} = 0.766$; $S_{DEP} = 0.503$; $S_{PRESS} = 0.558$.

DC = 25, 5

Here **DC** is the deleted compound treated as an outlier since they were exhibiting aberrant behaviors, which may be due to different mode of actions. The model explained 84.03% variance of the activity.

Another model was developed using ETSA index (S_6), RTSA index (R_5) and indicators parameters (I_1 and I_2) as shown bellow

$$pIC_{50} = -4.411(\pm0.446) + 1.229(\pm0.546)S_6 + 1.876(\pm0.343)R_5 - 1.425(\pm0.349)I_1 + 1.238(\pm0.330)I_2 \quad (4)$$

$n = 34$; $R = 0.862$; %EV = 74.31; $F(4,29) = 20.972$; $P < 0.00001$; $s = 0.561$; SSY = 35.525; PRESS = 12.096; $R^2_{CV} = 0.660$; $S_{DEP} = 0.596$; $S_{PRESS} = 0.646$.

Equation 4 explained 74.31% variance of the activity. Atom number 5 and 6 are important for the activity. Positive coefficient of S_6 implies that the CDK2/cyclin A inhibitory activity of the compounds will increase with increasing the ETSA index of the nitrogen (S_6). It can be assumed to be involved in the electronic interaction with the enzyme. The positive coefficient of R_5 also indicate that with increasing the value of RTSA index at atom number 5 anticancer activity also increases. Negative coefficient of the I_1 indicates that absence of the meta substitution of the phenyl ring favor the activity. Positive coefficient of the I_2 Implies that substituted bi-phenyl/2-thenyl phenyl is beneficial for the activity.

$$PIC_{50} = -4.499(\pm0.418) + 1.086(\pm0.510)S_6 + 2.073(\pm0.326)R_5 - 1.393(\pm0.322)I_1 + 1.322(\pm0.306)I_2 \quad (5)$$

$n = 32$; $R = 0.891$; %EV = 79.37; $F(4,27) = 25.970$; $P < 0.00001$; $s = 0.517$; SSY = 34.952; PRESS = 10.121; $R^2_{CV} = 0.710$; $S_{DEP} = 0.562$; $S_{PRESS} = 0.612$.

DC = 25, 20

Stepwise deletion of the compounds 25, 20 statistically improve the quality of the equation by increasing explained variance up to 79.37%.

In deriving the above equations, some compounds were deleted, viz. compounds 25 and 5 for Eq. (3), compounds 25 and 20 for Eq. (5). All these compounds exhibited aberrant behaviors. However, further study is necessary to find out convincing reasons behind such aberration of the excluded compounds.

Both the equations obtained are considered statistically significant because the associated probability values are <0.05 and therefore F is statistically significant. The student *t* values and associated probability values *P* of both the equations are given in Table 4. The Eq. (3) is considered to be the best model where the multiple correlation coefficient is 0.917 and this explains 84.03% of the variation of the biological activity data.

Table 4
t-Statistics and *P*-values for all the equations

Eq. number	Intercept/parameter	*t*-value	*P*-value	Eq. number	Intercept/Parameter	*t*-value	*P*-value
2	Intercept	–15.685	0.000	4	Intercept	–9.882	0.000
	S_5	2.528	0.017		S_6	2.252	0.032
	R_5	2.624	0.014		R_5	5.474	0.000
	q8	1.885	0.070*		I_1	–4.085	0.000
	I_1	–3.214	0.003		I_2	3.753	0.001
	I_2	3.448	0.002				
3	Intercept	–17.635	0.000	5	Intercept	–10.773	0.000
	S_5	2.339	0.027		S_6	2.128	0.042
	R_5	3.269	0.003		R_5	6.360	0.000
	q8	2.709	0.012		I_1	–4.333	0.000
	I_1	–3.054	0.005		I_2	4.313	0.000
	I_2	3.782	0.001				

*Confidence interval is less than 95%.

Table 5
Observed (Obs.), calculated (Calc.), residual (Res.), LOO-predicted (LOO-pred) and predicted residual (Pres.) activities of Eqs. (3) and (5)

Compound[a]	Obs	Eq. (3)				Eq. (5)			
		Calc	Res	LOO-pred	Pres.	Calc	Res	LOO-pred	Pres.
1	–3.176	–3.649	0.473	–3.744	0.568	–3.730	0.554	–3.834	0.658
2	–4.000	–3.599	–0.401	–3.540	–0.460	–3.726	–0.274	–3.673	–0.327
3	–2.350	–2.194	–0.156	–2.168	–0.182	–2.143	–0.207	–2.131	–0.219
4	–2.462	–2.246	–0.216	–2.226	–0.236	–2.152	–0.310	–2.133	–0.329
5	–1.531	–	–	–	–	–2.322	0.791	–2.363	0.832
7	–2.848	–2.324	–0.524	–2.267	–0.581	–2.070	–0.778	–2.018	–0.830
8	–2.217	–2.566	0.349	–2.626	0.409	–2.239	0.022	–2.242	0.025
9	–1.833	–2.512	0.680	–2.595	0.762	–2.245	0.412	–2.275	0.442
11	–3.000	–2.454	–0.546	–2.397	–0.603	–2.276	–0.724	–2.239	–0.761
12	–2.778	–2.345	–0.433	–2.294	–0.484	–2.048	–0.730	–1.998	–0.780
13	–2.699	–2.711	0.012	–2.719	0.020	–2.426	–0.273	–2.107	–0.592
15	–3.000	–3.582	0.582	–3.881	0.881	–3.713	0.713	–4.070	1.070
16	–4.000	–3.571	–0.429	–3.352	–0.648	–3.405	–0.595	–3.090	–0.910
17	–4.000	–3.847	–0.153	–3.770	–0.230	–3.881	–0.119	–3.818	–0.182
18	–1.919	–2.496	0.577	–2.540	0.621	–2.375	0.456	–2.400	0.481
20	–1.699	–2.737	1.038	–2.979	1.280	–	–	–	–
21	–3.000	–2.674	–0.326	–2.594	–0.406	–2.826	–0.174	–2.805	–0.195
22	–3.778	–3.808	0.030	–3.815	0.037	–3.905	0.127	–3.936	0.158
23	–2.799	–3.427	0.628	–3.521	0.722	–3.628	0.829	–3.759	0.960
25	–2.301	–	–	–	–	–	–	–	–
26	–4.000	–3.202	–0.798	–3.093	–0.907	–3.362	–0.637	–3.283	–0.717
28	–4.000	–3.745	–0.255	–3.663	–0.337	–3.319	–0.681	–3.245	–0.755
29	–4.000	–3.732	–0.268	–3.684	–0.316	–3.588	–0.412	–3.512	–0.488
30	–1.681	–1.72	0.041	–1.729	0.048	–2.101	0.420	–2.135	0.454
31	–1.869	–1.934	0.065	–1.941	0.072	–2.160	0.291	–2.177	0.308
32	–1.462	–1.682	0.220	–1.723	0.261	–2.007	0.545	–2.053	0.591
33	–1.978	–1.511	–0.467	–1.416	–0.562	–1.686	–0.292	–1.636	–0.342
34	–1.826	–1.890	0.064	–1.914	0.088	–2.124	0.298	–2.176	0.350
35	–1.748	–1.776	0.028	–1.782	0.034	–2.055	0.307	–2.097	0.349
37	–0.954	–0.693	–0.261	–0.603	–0.351	–0.725	–0.229	–0.647	–0.307
38	–1.041	–0.869	–0.172	–0.805	–0.236	–0.754	–0.287	–0.656	–0.384
39	–0.602	–0.786	0.184	–0.848	0.246	–0.779	0.177	–0.838	0.236
40	–0.477	–0.726	0.249	–0.815	0.338	–0.816	0.339	–0.931	0.454
41	–1.568	–1.753	0.185	–1.785	0.217	–2.005	0.437	–2.068	0.500

[a] compound number.

Table 6
Observed and predicted values of test set compounds

Compound[a]	Observed	Predicted value	
		Eq. (3)	Eq. (5)
6	–1.929	–2.483	–2.248
10	–2.996	–2.439	–2.355
14	–3.176	–3.688	–3.794
19	–1.954	–2.544	–2.515
24	–2.813	–3.346	–3.595
27	–4.000	–3.264	–3.17
36	–0.602	–0.738	–0.753
42	–1.580	–1.649	–1.962

$R^2_{pred} = 0.741$ for Eq. (3). $R^2_{pred} = 0.666$ for Eq. (5).
[a] compound number.

Useful informations are obtained at molecular level to the structural requirements of the 3-aminopyrazole analogues for their antitumor activity from the above two equations (Eqs. (3) and (5)). The predictive power of the equations was confirmed by Leave-One-Out (LOO-) cross validation method, where one compound is deleted at once and prediction of the activity of the deleted compound is made based on the QSAR model. The process is repeated after elimination of another compound until all of the compounds have been deleted at once. The observed, calculated, residual, LOO-predicted (pred) and predicted residual (pres) values of the equations are shown in Table 5.

On the basis of these two models predicted values of the test set compounds were calculated. The observed and predicted values of the test set compounds are given in Table 6. It shows significant R^2_{Pred} for the test set as follows

$R^2_{pred} = 0.741$ for Eq. (3).
$R^2_{pred} = 0.666$ for Eq. (5).

4. Conclusion

The two final QSAR models reveal pharmacophoric requirements of 3-aminopyrazoles for their CDK2/cyclin A inhibitory activity as antitumor agents. It is found that atom number 5 and 6 are important for the activity. Models show that with increasing the value of S_5 and S_6 (E-state index of atom number 5 and 6, respectively) increases the CDK2/cyclin A inhibitory activity of the compounds. Thus, atom number 5 and 6 might have some roles in the electronic interaction of the compounds with the enzymes. The CDK2/cyclin A inhibitory activity of the compounds increases with increasing the R-state index of atom 5. It implies that atom number 5 might be involved through the dispersive/van der Waals interactions of the compound with the enzyme. At R_1 position for the presence of substitution on the phenyl ring at 2 position or the meta position indicate the detrimental effects towards the activity. The presence of substituted biphenyl or 2-thenyl phenyl at R_1 position are favorable towards the activity. Charge at atom number 8 is important for the biological activity. Model indicates that with increase of the value of charge at atom 8 biological activity increases, i.e. with increasing the electronegativity of oxygen at position 8 the activity increases.

Acknowledgements

Authors are thankful to the University Grants Commission (UGC), New Delhi and All India Council for Technical Education (AICTE), New Delhi for providing financial assistance. Three of the authors S.S., B.D. and S.G. are grateful to University Grants Commission (UGC), New Delhi, the Council of Scientific and Industrial Research (CSIR), New Delhi and Indian Council of Medical Research for providing Junior Research Fellowship and Senior Research Fellowships, respectively. Authors are also grateful to the authority of Jadavpur University for their help and encouragement.

References

[1] M. Mihara, S. Shintani, A. Kiyota, T. Matsumura, D.T.W. Wong, Inter. J. Oncol. 21 (2002) 95–101.
[2] M. Meyerson, G.H. Enders, C.L. Wu, L.K. Su, C. Gorka, C. Nelson, E. Harlow, L.H. Tsai, EMBO J. 11 (1992) 2909–2917.
[3] N. Giocanti, R. Sadri, M. Legraverend, O. Ludwig, E. Bisagni, S. Leclerc, L. Meijer, V. Favaudona, Ann. N. Y. Acad. Sci. 886 (1999) 180–182.
[4] M. Knockaert, N. Gray, E. Damiens, Y.-T. Chang, P. Grellier, K. Grant, D. Fergusson, J. Mottram, M. Soete, J.-F. Dubremetz, K. Le Roch7, C. Doerig, P.G. Schultz, L. Meijer, Chem. Biol. 7 (2000) 411–422.
[5] P. Pevarello, M.G. Brasca, R. Amici, P. Orsini, G. Traquandi, L. Corti, C. Piutti, P. Sansonna, M. Villa, B.S. Pierce, M. Pulici, P. Giordano, K. Martina, E.L. Fritzen, R.A. Nugent, E. Casale, A. Cameron, M. Ciomei, F. Roletto, A. Isacchi, G. Fogliatto, E. Pesenti, W. Pastori, A. Marsiglio, K.L. Leach, P.M. Clare, F. Fiorentini, M. Varasi, A. Vulpetti, M.A. Warpehoski, J. Med. Chem. 47 (2004) 3367–3380.
[6] S. Samanta, K. Srikanth, S. Banerjee, B. Debnath, S. Gayen, T. Jha, Bioorg. Med. Chem. 12 (2004) 1413–1423.
[7] B. Debnath, S. Gayen, A. Basu, K. Srikanth, T. Jha, J. Mol. Mod. (2004) 328–334.
[8] S. Samanta, B. Debnath, S. Gayen, B. Ghosh, A. Basu, T. Jha, IL Farmaco 60 (2005) 818–825.
[9] C. de Gregorio, L.B. Kier, L.H. Hall, J. Comput. Aided Mol. Des. 12 (1998) 557–561.
[10] K. Rose, L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 42 (2002) 651–666.
[11] K. Roy, A.U. De, C. Sengupta, Drug Des. Discov. 18 (2002) 33–43.
[12] K. Roy, D.K. Pal, C. Sengupta, Drug Des. Discov. 17 (2001) 207–218.
[13] K. Roy, D.K. Pal, C. Sengupta, Drug Des. Discov. 17 (2000) 183–190.
[14] Mouse, a computer program written in C++ language by S.K. Naskar of Computer science and Engineering department, Jadavpur University and B. Debnath, one of the authors of this article.
[15] R. Carrasco-Vellar, J.A. Padron, J. Galvez, J. Pharm. Pharm. Sci. 7 (2004) 19–26.
[16] HyperChem Professional 7.0 of Hypercube, Inc.;http:// www.hyper.com.
[17] Multi Regress, a computer program written in C^{++} language by S.K. Naskar, Dept. of Computer Science and Engineering and B. Debnath, one of the authors of this article.
[18] G.W. Snedecor, W.G. Cochran, Multiple regression in Statistical Methods, Oxford & IBH, New Delhi, India, 1967.
[19] D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons, New York, USA, 1997.
[20] E.J. Lien, in: E.J. Ariens (Ed.), Drug Design, vol V, Academic Press, New York, USA, 1975, pp. 84–94.
[21] I.V. Tetko, V.Y. Tanchuk, A.E. Villa, J. Chem. Inf. Comput. Sci. 41 (2001) 1407–1421.