Available online at www.derpharmachemica.com

ISSN 0975-413X
CODEN (USA): PCHHAX

Der Pharma Chemica, 2017, 9(14):127-135
(http://www.derpharmachemica.com/archive.html)

Computational Multivariate Regression and Validation Analysis on a Set of AKT kinase Inhibitors

Siva Prasad N[1], Manideep Reddy M[1], Krishna Murthy P[*], Ajay Babu P[2]

[1]*Department of Chemistry, Bapatla Engineering College (Autonomous), Acharaya Nagarjuna University Bapatla-522102, A.P., India*
[2]*Department of Chemistry, St. Peter's University, Chennai-600054, Tamil Nadu, India*

ABSTRACT

AKT as a result of inactivation of tumor suppressor PTEN has been found in a variety of human tumors. AKT has long been considered an attractive target for the treatment of cancers. A computational multivariate regression was carried out on a set of 61 pyridine based analogs to study the influence of physico-chemical properties on AKT inhibition. A regression model was generated by dividing the complete set as a 51 molecule training set and a 6 molecule validation set based on selection criteria after rejecting outliers from the data set. The generated equation when applied on test set molecules suggested predictive ability. The model can be utilized to study the efficacy of AKT inhibition based on the properties evaluated.

Keywords: Multivariate regression, AKT Kinase, Correlation, Validation.

INTRODUCTION

Activation of oncogenes, such as Ras, ErbB-2, and Src or loss of tumor suppressor genes, such as PTEN, can lead to aberrant signaling in the PI3K/AKT signal transduction pathway [1]. Three isoforms of AKT kinases are known, such as AKT1 (PKBa), AKT2 (PKBb) and AKT3 (PKBc). All three are up-regulated in different types of cancers including NSCLC (non-small cell lung carcinoma), breast and prostate cancers, making them potential oncology targets. Several small molecule AKT inhibitors have recently been reported [2,3]. Kinase selectivity is important because long term inhibition of off-target kinases may cause undesired side effects and toxicity.

Protein kinases are a large family of diverse but related enzymes that regulate nearly all aspects of cell growth, differentiation, and division. Dys-regulation of one or more protein kinases has been associated with a wide spectrum of human cancers. Clinical success of Gleevec (inhibitor of BCR-ABL, PDGFR, and c-Kit), as well as Iressa (EGFR inhibitor) resulted in a search for small molecule inhibitors of protein kinases as anti-cancer chemotherapeutics [4,5]. Among the superfamily of protein kinases, protein kinase B, also called AKT, is a pivotal component of the phosphatidylinositol 30-kinase/Akt signal transduction pathway that regulates many processes crucial to carcinogenesis [6].

AKT as a result of, for example, inactivation of tumor suppressor PTEN has been found in a variety of human tumors [7,8]. Therefore, AKT has long been considered an attractive target for the treatment of cancers. There have been a number of small molecule inhibitors that partially target AKT. While the majority of these inhibitors are ATP-competitive, Lindsley et al. reported a series of selective allosteric and non-ATP-competitive diphenylquinoxaline- and diphenylpyridine-based inhibitors of AKT that target the Pleckstrin Homology (PH) domain of the protein kinase [9].

In view of the above, computational multivariate regressions were carried out on a set of 61 pyridine based analogs to study the influence of physico-chemical properties of side chain groups of these congeneric series of compounds.

MATERIALS AND METHODS

Data set

A dataset consisting of selective AKT inhibitors with experimental biological activity were considered from literature having oxindole-pyridine and 2,3,5-trisubstituted pyridine moieties [10,11]. The inhibitory activities of these derivatives reported in terms of IC_{50} in micromolar were transformed into their corresponding logarithmic values in order to overcome overlapping data.

Therefore, to obtain linear distribution of data, the inhibition converted to negative logarithmic values was used for subsequent analysis. The structures were sketched using ISIS Draw 2.3 software and the descriptors were calculated.

Multivariate regression analysis

Regression models were built on complete and training sets, respectively.

The relationship between dependent variable (log $1/IC_{50}$) and independent variables was established by linear regression analysis. Significant descriptors were chosen based on the statistical data of analysis. Statistical quality of the generated regression equation was judged based on the parameters like correlation coefficient (r), F-value, cross-validation r^2 etc.

Descriptors for regression

Nearly 30 descriptors which represent physico-chemical properties of chemical compounds obeying drug likeness parameters based on Lipinski rule of 5, such as molecular weight, hydrogen bond acceptors, hydrogen bond donors, logP, number of rotatable bonds and other parameters such as Dipole, lipole, 5, 6-membered aromatic rings, molecular surface area, and molecular volume, indices such as Kier, K Alpha, Randic, Balaban, Weiner are considered.

Validity of regression equation

Predictive validity of the regression model was estimated externally by predicting the activities of validation set. Apart from this, another criterion was proposed [12] which are based on the regression of observed activities against predicted activities and vice versa for validation set, if the following conditions are satisfied [13].

$$(R^2–R_0^2)/R^2<0.1 \text{ or } (R^2–R_0^2)/R^2<0.1$$
$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15$$

Calculations relating to the above equations and the slopes, *k* and *k'* are based on regression of observed values against predicted values and vice versa.

Outliers standardized residuals

A standardized residual is a ratio: The difference between the observed values and predicted values and the standard deviation of the predicted values. The standardized residual is a measure of the strength of the difference between observed and expected values.

$$Standardized\ Residual\ i = \frac{Residual\ i}{Standard\ Deviation\ of\ Residual\ i}$$

Standardized residuals greater than 2 and less than -2 are usually considered large [13]. Outliers should be removed in order to obtain the best statistical result [14].

RESULTS AND DISCUSSION

Multivariate regression analysis using regress It add-in Excel program resulted in few influential parameters displayed significant positive and negative contribution towards biological activity of AKT inhibitors. Equation 1 given below represents the regression model from a complete set of 61 AKT inhibitors. The complete data set along with independent variables is presented in Tables 1 and 2. A plot of actual values versus predicted values from Equation 1 was given in Figure 1. In order to produce better predictive values, outliers should be analyzed in data and should be removed from analysis. Therefore, standardized residuals were calculated and data was presented.

Complete set

Log ($1/IC_{50}$)=-0.105x DIPOLE Z
+0.063x LIPOLE X
+0.564x KAPPA1
-2.366x KAPPA3
+0.352x LogP
-0.691x Kier-chi-V1-bond
+8.050

r=0.71484; r^2=0.511; Adj. r^2; n=61 (1)

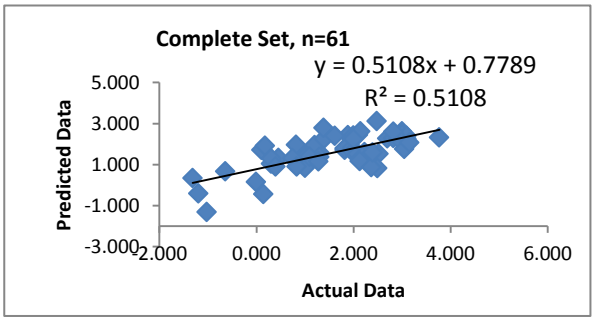

Figure 1: Actual vs. predicted activities of complete dataset (n=61)

Table 1: Complete set data with predicted and residual values: Outliers calculated by Standardized residuals, highlighted in color

Mol No.	Activity	Predictions	Residuals	standardized residuals	Molecular Surface Area	Moelcular Volume	Total Dipole	Total Lipole	Molecular Refractivity	Kappa 1	Kappa 2	Kappa 3	Randic	Balaban	Weiner	5-membered RINGS	6-membered RINGS
2_17A.mol	3.000	2.378	0.622	0.768	438.117	345.970	3.550	6.765	135.983	26.234	11.623	5.814	17.547	0.938	4390	3	3
2_17B.mol	3.097	2.250	0.847	1.047	438.575	348.257	2.511	5.997	135.983	26.234	11.623	5.814	17.547	0.940	4416	3	3
2_17C.mol	3.000	2.609	0.391	0.483	435.390	348.129	1.841	5.684	135.983	26.234	11.623	5.673	17.564	0.942	4387	3	3
2_18A.mol	2.097	2.439	-0.342	-0.422	422.443	336.541	1.545	9.538	133.562	25.288	11.396	5.579	17.153	0.947	4050	3	3
2_18B.mol	1.886	2.425	-0.539	-0.666	424.353	336.228	6.151	9.802	133.578	25.288	11.396	5.579	17.153	0.947	4050	3	3
2_18C.mol	2.000	2.412	-0.412	-0.509	424.462	336.230	6.047	10.233	133.238	25.288	11.396	5.579	17.153	0.947	4050	3	3
2_2A.mol	0.903	1.664	-0.761	-0.941	340.092	274.346	1.401	3.502	107.371	20.280	9.667	5.331	13.187	1.082	2136	1	3
2_9a.mol	1.301	1.589	-0.288	-0.356	414.874	332.772	2.343	5.415	132.167	24.684	11.823	6.228	16.170	1.077	3402	1	4
2_9b.mol	1.824	1.715	0.109	0.134	428.438	340.991	1.507	2.761	133.861	25.641	12.030	6.297	16.581	1.100	3637	1	4
2_9c.mol	1.102	1.128	-0.025	-0.031	425.855	340.750	2.312	2.249	133.861	25.641	12.030	6.478	16.564	1.084	3664	1	4
2_9d.mol	0.301	1.033	-0.732	-0.905	424.292	340.097	2.827	1.548	133.861	25.641	12.030	6.478	16.564	1.086	3691	1	4
2_9e.mol	1.000	1.342	-0.342	-0.423	467.348	354.744	2.931	3.118	138.630	26.601	12.698	6.533	17.119	1.117	3906	1	4
2_9f.mol	2.000	1.581	0.419	0.519	435.195	348.840	2.137	3.046	134.077	26.601	12.240	6.533	16.974	1.113	3904	1	4
2_9g.mol	1.201	1.947	-0.747	-0.923	449.722	356.680	3.009	3.781	134.294	27.563	12.453	6.605	17.385	1.124	4173	1	4
2_9h.mol	2.523	1.526	0.997	1.233	464.439	357.989	2.230	4.772	138.665	26.601	12.240	6.533	16.974	1.113	3904	1	4
2_9i.mol	0.102	1.712	-1.609	-1.990	395.146	318.548	2.363	1.651	124.558	23.728	11.160	5.814	15.670	1.071	3143	2	3
2_9j.mol	2.000	1.712	0.288	0.356	394.926	318.529	2.362	1.652	124.558	23.728	11.160	5.814	15.670	1.071	3143	2	3
2_9k.mol	1.102	1.130	-0.027	-0.034	404.993	335.686	2.798	2.455	131.001	23.728	11.160	5.814	15.670	1.071	3143	2	3
2_9l.mol	1.301	1.372	-0.071	-0.088	409.261	330.370	2.077	2.019	131.123	23.728	11.160	5.814	15.670	1.071	3143	2	3
2_9m.mol	1.824	1.785	0.039	0.048	402.551	320.447	1.823	3.149	126.803	23.728	11.160	5.814	15.670	1.071	3143	2	3
2_9N.mol	-1.021	-1.325	0.303	0.375	428.031	336.542	2.265	9.120	131.465	26.074	11.823	7.492	15.782	1.219	3484	1	3
2_9O.mol	0.143	-0.446	0.589	0.728	423.974	325.757	2.122	9.978	127.164	25.104	12.630	7.014	15.635	1.191	3256	1	3
2_9P.mol	3.000	2.456	0.544	0.672	429.710	338.451	2.406	6.628	135.766	25.288	11.396	5.579	17.153	0.947	4050	3	3
2_9Q.mol	3.097	2.397	0.700	0.865	469.101	353.795	1.125	4.696	140.916	26.234	11.623	5.673	17.564	0.964	4323	3	3
1.mol	2.699	2.276	0.423	0.523	370.024	278.792	3.122	9.073	117.804	21.825	10.292	5.333	14.759	0.958	2868	1	4
11A.mol	2.481	3.103	-0.622	-0.769	411.375	315.735	3.296	6.244	127.136	23.402	10.166	4.857	16.153	0.848	3662	1	3
11B.mol	2.229	1.606	0.623	0.771	429.824	342.165	3.454	8.246	139.883	26.234	12.027	5.961	17.653	0.841	4628	1	4
11c.mol	0.453	1.317	-0.864	-1.068	466.391	376.089	3.709	12.971	149.493	28.135	12.475	6.278	18.458	0.850	5330	1	4
11D.mol	1.983	1.809	0.174	0.215	428.850	338.764	4.064	4.469	137.695	26.234	12.027	5.961	17.653	0.841	4628	1	4

11E.mol	1.987	1.618	0.369	0.457	424.643	339.515	2.258	4.254	137.354	26.234	12.027	5.961	17.653	0.841	4628	1	4
11F.mol	2.824	2.283	0.540	0.668	417.011	329.897	2.121	5.875	136.236	25.288	11.396	5.579	17.153	0.845	4282	1	3
11G.mol	0.180	1.915	-1.735	-2.145	430.446	335.971	2.954	12.281	135.205	25.288	11.396	5.579	17.153	0.845	4282	1	3
11H.mol	2.086	1.511	0.575	0.711	431.665	355.340	28.720	8.313	142.131	26.234	11.623	5.673	17.564	0.850	4597	1	3
11I.mol	2.921	2.271	0.650	0.804	418.852	323.353	4.174	11.842	135.019	25.288	11.396	5.579	17.153	0.845	4282	1	3
11J..mol	0.815	1.955	-1.140	-1.409	475.614	367.863	2.896	5.105	150.970	27.810	12.313	5.819	19.136	0.782	5682	2	4
11K.mol	2.143	2.605	-0.462	-0.572	418.025	326.729	3.630	6.648	132.284	25.288	11.396	5.579	17.153	0.845	4282	2	3
11L.mol	1.398	2.250	-0.852	-1.053	490.775	380.784	4.713	4.646	154.352	29.089	12.701	6.080	18.923	0.858	5686	2	3
11M.mol	3.046	1.759	1.286	1.591	429.615	340.295	2.161	4.742	138.717	25.288	11.396	5.579	17.153	0.845	4282	2	3
11N.mol	3.770	2.312	1.458	1.803	421.113	328.084	3.308	4.798	132.274	25.288	11.396	5.579	17.153	0.845	4282	2	3
11O.mol	3.155	2.069	1.086	1.342	439.231	340.439	3.189	6.115	137.423	26.234	11.623	5.814	17.547	0.843	4627	2	3
11P.mol	2.000	2.230	-0.230	-0.285	458.707	353.564	1.924	6.255	142.464	27.184	11.852	5.898	17.958	0.846	4975	2	3
9a.mol	2.268	1.390	0.877	1.085	354.289	286.193	8.383	3.231	115.308	20.878	10.080	5.258	14.349	1.100	2352	2	3
9b.mol	0.312	1.289	-0.977	-1.208	355.823	286.702	6.700	3.321	115.308	20.878	10.080	5.258	14.349	1.100	2352	2	3
9c.mol	1.387	2.777	-1.390	-1.718	351.209	282.629	6.466	3.974	109.035	20.878	10.080	5.258	14.349	1.124	2352	2	3
9d.mol	-0.643	0.658	-1.302	-1.610	379.771	302.078	5.525	10.605	117.362	21.825	10.292	5.507	14.742	1.125	2587	1	2
9e.mol	2.495	0.824	1.671	2.066	345.844	295.522	4.759	6.705	117.362	21.825	10.292	5.507	14.742	1.125	2587	1	2
9F.mol	1.004	0.838	0.166	0.205	372.365	297.760	6.009	11.329	115.788	21.825	10.292	5.507	14.742	1.125	2587	1	2
9g.mol	0.747	1.295	-0.548	-0.678	351.218	298.803	6.939	8.547	118.150	22.776	10.508	5.399	15.170	1.140	2803	1	2
9H.mol	2.187	1.378	0.809	1.000	386.531	313.847	5.573	11.250	121.685	23.728	11.160	5.644	15.708	1.148	3050	1	2
9I.mol	0.833	0.894	-0.062	-0.076	370.418	322.405	6.827	7.105	126.205	24.684	11.823	6.049	16.208	1.150	3329	1	2
9J.mol	0.393	0.894	-0.502	-0.620	370.186	322.411	6.827	7.105	126.205	24.684	11.823	6.049	16.208	1.150	3329	1	2
9K.mol	-1.316	0.333	-1.649	-2.039	376.613	322.683	5.427	7.266	126.205	24.684	11.823	6.049	16.208	1.120	3389	1	2
9L.mol	2.131	1.168	0.963	1.190	366.081	297.316	6.617	9.190	113.838	21.825	10.292	5.507	14.742	1.125	2587	1	2
9M.mol	2.367	0.872	1.495	1.848	346.608	301.302	5.981	3.496	120.281	21.825	10.292	5.507	14.742	1.125	2587	1	2
9N.mol	-1.193	-0.420	-0.773	-0.955	360.559	306.422	5.698	8.933	119.487	22.776	10.092	5.742	15.056	1.130	2824	1	2
9O.mol	1.620	2.388	-0.768	-0.950	361.171	321.348	4.694	4.873	126.438	23.728	10.318	5.327	15.503	1.158	3021	1	2
9P.mol	2.398	1.582	0.816	1.009	350.555	308.042	4.756	5.894	121.937	22.776	10.508	5.399	15.170	1.140	2803	1	2
9Q.mol	1.097	1.372	-0.275	-0.340	364.263	321.649	4.733	5.359	126.538	23.728	11.160	5.644	15.708	1.148	3050	1	2
9r.mol	-0.009	0.142	-0.151	-0.186	496.635	375.890	5.655	11.195	146.633	27.184	13.109	6.840	18.226	1.036	4591	1	3
9s.mol	1.278	1.141	0.137	0.170	427.841	333.654	5.592	11.010	131.696	24.684	11.373	5.878	16.081	1.159	3299	1	2
9T.mol	2.824	2.595	0.229	0.283	383.802	318.366	6.728	8.939	117.330	23.728	10.318	5.327	15.503	1.158	3021	1	2
			Standard Deviation: 0.809														

Table 2: Complete set data with predicted and residual values

Mol No.	Molecular Weight	HB Acceptors	HB Donors	logP	Rotatable Bonds	Dipole moment X	Dipole moment Y	Dipole moment Z	Lipole X	Lipole Y	Lipole Z	KierChiV0 atoms	KierChiV1 bond	KierChiV2 path	KAlpha1	KAlpha2	KAlpha3
2_17A.mol	481.57	5	3	4.398	7	2.948	1.532	-1.251	-5.388	-3.819	1.463	19.586	11.666	8.928	23.411	9.812	4.742
2_17B.mol	481.57	5	3	4.398	7	1.843	-1.701	0.124	-5.139	-2.916	1.024	19.586	11.666	8.928	23.411	9.812	4.742
2_17C.mol	481.57	5	3	4.398	7	0.545	-1.758	0.051	-4.782	-2.89	1.043	19.586	11.672	8.9	23.411	9.812	4.623
2_18A.mol	464.57	6	3	3.705	7	0.414	-0.651	-1.339	-3.695	-8.16	3.275	19.155	11.426	8.637	22.848	9.812	4.653
2_18B.mol	464.57	6	3	3.412	7	5.777	1.663	-1.3	-2.313	-8.225	4.802	19.155	11.416	8.659	22.848	9.812	4.653
2_18C.mol	464.57	6	3	3.346	7	4.648	3.622	-1.36	-2.133	-8.666	5.008	19.155	11.426	8.637	22.848	9.812	4.653
2_2A.mol	358.48	4	2	3.683	6	-0.713	0.964	0.724	-3.337	0.181	-1.045	15.222	9.013	6.74	18.31	8.308	4.441
2_9a.mol	434.58	4	2	5.301	7	-1.544	-0.753	1.593	5.065	-1.593	1.062	18.532	11.101	8.263	21.977	9.97	5.066
2_9b.mol	450.58	5	3	5.017	7	1.056	-0.791	0.728	2.751	0.234	-0.004	18.902	11.241	8.415	22.89	10.168	5.139
2_9c.mol	450.58	5	3	5.017	7	-1.113	0.578	1.942	2.179	0.476	-0.285	18.902	11.235	8.448	22.89	10.168	5.293
2_9d.mol	450.58	5	3	5.017	7	-1.472	0.461	2.369	1.375	0.207	0.681	18.902	11.235	8.444	22.89	10.168	5.293
2_9e.mol	464.61	5	2	5.049	8	1.047	-2.644	0.708	1.142	2.581	-1.326	19.863	11.63	8.598	23.843	10.804	5.367
2_9f.mol	468.57	5	3	5.157	7	0.157	-1.1	1.826	3.028	0.244	-0.225	19.202	11.341	8.559	23.776	10.349	5.34
2_9g.mol	486.56	5	3	5.296	7	-1.923	0.997	2.088	3.737	-0.395	-0.423	19.503	11.446	8.681	24.665	10.534	5.402
2_9h.mol	485.02	5	3	5.535	7	0.073	-1.124	1.925	4.699	-0.235	-0.8	20.02	11.749	9.031	24.12	10.575	5.48
2_9i.mol	424.54	5	2	4.25	7	-0.835	-1.404	1.708	0.4	1.135	1.13	17.785	10.585	7.875	21.238	9.477	4.771
2_9j.mol	424.54	5	2	4.25	7	-0.834	-1.402	1.708	0.401	1.136	1.13	17.785	10.585	7.875	21.238	9.477	4.771
2_9k.mol	440.6	4	2	4.593	7	-0.381	-0.064	2.771	1.451	0.584	1.892	18.668	11.536	8.899	21.607	9.723	4.921
2_9l.mol	440.6	4	2	4.829	7	-0.848	-1.69	0.86	1.409	0.44	1.377	18.668	11.591	8.753	21.607	9.723	4.921
2_9m.mol	423.56	4	3	4.072	7	0.461	-1.077	-1.397	-1.53	1.441	2.344	17.877	10.684	7.981	21.238	9.477	4.771
2_9N.mol	440.59	5	3	3.876	9	1.618	1.393	0.754	-1.7	8.957	-0.254	19.092	10.914	8.706	23.613	10.161	6.283
2_9O.mol	426.56	5	3	3.636	10	1.62	1.07	-0.858	-1.684	9.833	0.15	18.007	10.61	7.673	22.647	10.857	5.838
2_9P.mol	463.58	5	3	4.259	7	1.451	1.381	-1.332	-4.972	-4.065	1.64	19.285	11.566	8.783	22.538	9.615	4.541
2_9Q.	477.61	5	3	4.27	7	1.035	-	-	-	-	1.42	20.20	11.9	9.16	23.	9.853	4.646

mol				3			0.388	0.207	4.324	1.158	1	8	96	2	477		
1.mol	394.51	4	2	3.868	6	-2.564	1.146	1.363	3.366	-8.307	-1.413	16.455	9.987	7.482	19.388	8.652	4.323
11A.mol	438.57	4	3	2.971	7	2.805	1.166	-1.279	1.628	5.557	2.335	18.252	11.491	9.15	20.819	8.545	3.937
11B.mol	474.6	4	3	3.965	7	2.973	1.104	-1.37	-5.826	-3.044	4.978	19.648	12.041	9.179	22.895	9.842	4.675
11c.mol	543.48	4	3	5.001	7	3.313	-1.077	-1.273	-9.945	-5.88	5.896	21.883	13.063	10.346	25.325	10.661	5.197
11D.mol	475.59	5	3	3.118	7	1.995	1.212	-3.326	-2.734	2.101	2.843	19.517	11.891	9.055	23.204	10.039	4.789
11E.mol	475.59	5	3	3.052	7	0.878	1.381	-1.556	-2.354	2.427	2.582	19.517	11.901	9.023	23.204	10.039	4.789
11F.mol	463.58	5	3	2.675	7	0.892	-0.554	-1.843	1.712	5.193	2.147	19.017	11.691	8.987	22.51	9.598	4.531
11G.mol	467.62	5	4	2.203	7	2.062	1.884	-0.965	5.857	10.782	0.51	19.46	12.225	9.516	22.651	9.687	4.582
11H.mol	478.62	4	3	2.845	7	-27.052	-0.224	9.642	7.623	3.269	-0.556	20.017	12.138	9.503	23.223	9.694	4.556
11I.mol	463.58	5	3	2.127	7	3.829	-0.978	-1.342	5.43	10.38	1.734	19.017	11.691	9.03	22.51	9.598	4.531
11J..mol	513.64	4	4	3.738	7	2.696	-0.964	0.437	-4.542	1.41	1.855	21.148	13.035	10.058	24.218	10.042	4.546
11K.mol	464.57	5	4	2.647	7	0.472	-0.43	-3.573	1.825	5.546	3.179	18.863	11.484	8.743	22.473	9.574	4.517
11L.mol	519.7	4	4	4.081	8	4.331	1.636	-0.879	-1.129	3.661	2.628	22.468	13.441	10.389	25.931	10.684	4.93
11M.mol	480.62	4	3	3.257	7	0.602	-0.714	-1.949	-1.447	2.46	3.787	19.784	12.476	9.842	22.529	9.609	4.537
11N.mol	464.56	5	3	2.914	7	2.705	1.139	-1.526	-0.484	3.608	3.125	18.901	11.525	8.784	22.163	9.379	4.406
11O.mol	478.59	5	3	2.928	7	2.715	1.14	-1.226	1.12	5.458	2.52	19.824	11.949	9.217	23.101	9.618	4.63
11P.mol	492.62	5	3	3.395	7	1.318	-0.79	-1.157	0.384	5.544	2.872	20.746	12.366	9.656	24.042	9.858	4.726
9a.mol	400.53	3	2	3.508	7	0.534	8.285	-1.162	-2.612	-1.848	0.45	16.591	10.412	8.1	19.021	8.808	4.464
9b.mol	400.53	3	2	3.508	7	4.742	4.583	1.181	-0.311	-3.275	-0.453	16.591	10.412	8.1	19.021	8.808	4.464
9c.mol	385.46	5	2	3.744	7	-5.152	-1.948	-3.387	3.515	0.131	1.849	15.578	9.417	6.872	18.965	8.77	4.44
9d.mol	400.52	3	3	1.489	7	-3.063	-4.425	1.249	0.531	-10.54	-1.067	16.545	10.282	7.847	19.463	8.702	4.5
9e.mol	400.52	3	3	1.489	7	-3.88	-2.605	-0.901	-0.421	-6.633	-0.887	16.545	10.282	7.847	19.463	8.702	4.5
9F.mol	401.51	3	4	1.371	7	-3.111	-5.005	1.172	1.479	-11.13	-1.511	16.338	10.059	7.555	19.425	8.677	4.485
9g.mol	414.5	4	3	0.934	7	-6.666	-1.808	-0.665	-2.205	-8.231	-0.662	16.746	10.263	7.775	20.094	8.729	4.314
9H.mol	429.52	5	4	1.591	8	-3.429	-4.046	1.712	2.49	-10.84	-1.71	17.233	10.482	7.896	21.38	9.571	4.682
9I.mol	440.55	4	3	1.326	8	-5.5	-1.119	-3.886	-1.87	-6.84	-0.437	17.94	10.895	8.138	21.778	9.836	4.834

9J.mol	440.55	4	3	1.326	8	-5.5	-1.119	-3.886	-1.872	-6.84	-0.438	17.94	10.895	8.138	21.778	9.836	4.834
9K.mol	440.55	4	3	1.326	8	-2.819	-4.265	1.82	-1.28	-7.129	-0.587	17.94	10.895	8.138	21.778	9.836	4.834
9L.mol	402.49	4	3	2.021	7	-4.382	-4.56	1.945	3.291	-8.469	-1.382	16.246	9.96	7.426	19.425	8.677	4.485
9M.mol	418.55	3	3	2.364	7	-5.1	-2.863	-1.251	1.773	-2.972	-0.497	17.129	10.911	8.67	19.792	8.92	4.636
9N.mol	437.56	5	4	0.934	7	-4.528	-3.084	-1.566	-2.307	-8.609	-0.59	17.537	11.7	9.275	20.632	8.703	4.822
9O.mol	428.58	3	3	2.716	7	-3.784	-2.633	-0.889	2.297	-4.217	-0.833	18.338	11.059	9.132	21.351	8.796	4.398
9P.mol	414.55	3	3	2.051	7	-3.856	-2.664	-0.809	0.884	-5.82	-0.287	17.416	10.72	8.286	20.405	8.931	4.435
9Q.mol	428.58	3	3	2.447	8	-3.849	-2.632	-0.813	1.912	-5.004	0.138	18.123	11.258	8.512	21.351	9.553	4.671
9r.mol	490.65	3	3	3.665	9	-2.776	-4.727	1.391	10.3 38	-3.827	1.951	20.509	12.815	9.794	24.079	10.995	5.531
9s.mol	440.59	3	3	2.395	8	-2.196	-4.547	2.404	4.655	-9.977	-0.09	18.838	11.309	8.824	22.053	9.618	4.799
9T.mol	436.5	3	3	2.386	7	-4.987	-4.288	1.415	4.447	-7.563	-1.714	17.094	10.437	8.034	21.219	8.713	4.349

New regression model–training and validation sets

A new regression model was attempted by dividing the complete set as training and validation sets based on selection criteria after rejecting outliers from the data set. The selection of molecules in the training set was made according to the activity data, so that representatives of a wide range of structures with different substituents, atoms and activity were included. The distribution of activity values for the validation set follows the similar distribution of the activity values for the training set [15].

Training set

A 61 molecule complete set was divided into 51 molecule training set and a 6 molecule validation set after rejecting 4 compounds as outliers. Several runs were performed on training set by varying the number of independent variables. After each analysis, the obtained equation was applied on validation set and graphs were plotted. Once the equation predicts activity of validation set then the predictive validity of the model was estimated (Figure 2).

$$\begin{aligned} \text{Log}(1/IC_{50}) = &-0.084x\ \text{DIPOLE Z} \\ &-0.070x\ \text{LIPOLE X} \\ &+0.595x\ \text{Kappa1} \\ &-2.357x\ \text{Kappa3} \\ &+0.680x\ \text{Kier-chi-v1-bond} \\ &+0.309x\ \text{LogP} \\ &+7.303 \end{aligned} \quad (2)$$

$r=0.7523$; $r^2=0.566$; $F=13.47$;
p-value=0.282, n=51

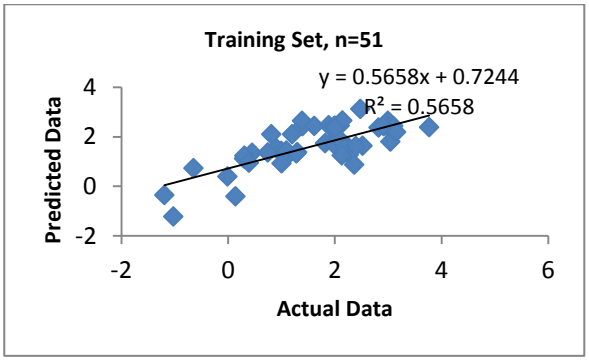

Figure 2: Actual versus predicted values of training set comprising of 51 compounds

Table 3: Validation dataset of 6 molecules

Mol	Actual	Predicted
2-9a.mol	1.301	1.625
2-9b.mol	1.823	1.760
2-9k.mol	1.102	1.166
1.mol	2.698	2.249
9-l.mol	0.832	0.933
9-t.mol	2.823	2.704

Validation of regression equation

After obtaining the regression model, it is important to determine the reliability and significance. These can be used to check if the size of the model is appropriate for the quantity of data available, as well as to provide some estimate of how well the model can predict the activity for new molecules.

One such validation procedure is dividing the set as training and test set and then applying training set equation on test set data. This will ensure applicability of the equation to ascertain values on external dataset. The above equation (2) was applied on test set molecules and the data is shown in Table 3 and the graphs are given in Figures 3 and 4, respectively.

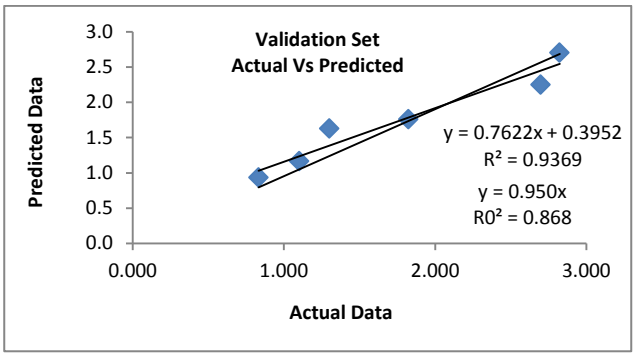

Figure 3: Actual versus predicted values of validation set compounds showing r2 and r_0^2

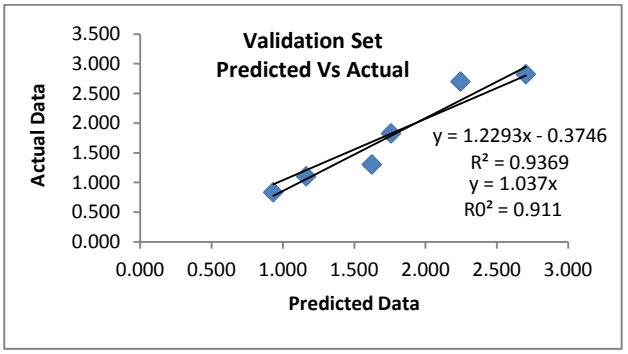

Figure 4: Predicted versus Actual values of validation set compounds showing r^2 and r_0^2

Figure 3 represents the predicted values of test set data when Equation (2) was applied and the regression coefficient (r^2) was obtained. However, apart from r^2, when the regression line passes through the origin, another regression coefficient (r_o^2) was plotted and it was observed that this value is also within the limits.

Alternatively, Regression plot between actual vs. predicted values of compounds from validation set justifies the predictive ability of regression model. However, a reverse graph, viz., Regression plot between predicted vs. actual values of compounds needs to be plotted in order to assess the predictive ability. Figure 4 represents predicted vs. actual values of test data set where r^2 and r'^2_o suggests the predictive validity of Equation (2).

Predictive validity

When analysis is run with varied independent variables, the validation set reported valid results by passing all the conditions set. The calculations are given below.

Calculation of k and k': All values of k are within the defined limits.

Formula: $k = \frac{\sum actual \; x \sum predicted}{predicted^2}$ $k' = \frac{\sum predicted \; x \sum actual}{actual^2}$

	Validation set	
Molecules	**Actual values**	**Predicted values**
2-9a.mol	1.301	1.625
2-9b.mol	1.824	1.760
2-9k.mol	1.102	1.166
1.mol	2.699	2.249
9-l.mol	0.833	0.933
9-t.mol	2.824	2.704
Summation	10.582	10.437

$k=\frac{10.582\ x\ 10.437}{(10.437^2)}=1.101$

	Validation set	
Molecules	**Predicted values**	**Actual values**
2-9a.mol	1.625	1.301
2-9b.mol	1.760	1.824
2-9k.mol	1.166	1.102
1.mol	2.249	2.699
9-l.mol	0.933	0.833
9-t.mol	2.704	2.824
Summation	10.437	10.582

$k'=\frac{10.437\ x10.582}{(10.582^2)}=0.986$

Interpretation of variables in regression equation

From equation 2, it can be observed that the logP, Kappa1 index and Kier ChiV1 properties on these inhibitors have positive contribution towards AKT inhibition. On the other hand, negative contribution of dipole X component and Kappa3 renders better AKT inhibition. The Kappa index [16] is a molecule shape index based on the assumption that the shape of a molecule is a function of the number of atoms and their bonding relationship. Kappa 1 shows the degree of complexity of a bonding pattern. Kappa 2 indicates the degree of linearity of bonding patterns. Kappa 3 indicates the degree of branching at the center of a molecule, larger for predominantly linear molecules with branching at the ends. Equation 2 suggests that a high value of kappa1 and a low value of kappa3 indices are favorable for activity.

CONCLUSION

AKT inhibitors have been widely studied in cancer progression and disease. Hence, an attempt made to evaluate the influential descriptors for AKT inhibition has been studied using multivariate regression analysis. Predictive validity of the model was estimated. The analysis resulted in few parameters displayed significant positive and negative contribution towards activity of AKT inhibitors. A regression model attempted by dividing the complete set (n=61) as a 51 molecule training set and a 6 molecule validation set resulted in a regression model.

The generated equations when applied on test set molecules resulted in better predictive values. Hence, designing or screening compound libraries for new compounds or analogs with possible increase in logP values, Kappa1 and Kier-ChiV bond parameters with decrease in diploe X component and Kappa3 index would enhance inhibitory activity against AKT.

REFERENCES

[1] C.W. Lindsley, S.F. Barnett, M.E. Layton, M.T. Bilodeau, *Curr. Cancer. Drug. Targets*., **2008**, 8, 7.
[2] X. Lin, J.M. Murray, A.C. Rico, M.X. Wang, D.T. Chu, Y. Zhou, M. Del Rosario, S. Kaufman, S. Ma, E. Fang, K. Crawford, A.B. Jefferson, *Bioorg. Med. Chem. Lett.*, **2006**, 16(16), 4163.
[3] I. Collins, J. Caldwell, T. Fonseca, A. Donald, V. Bavetsias, L.J.K. Hunter, M.D. Garrett, M.G. Rowlands, G.W. Aherne, T.G. Davies, V. Berdini, S.J. Woodhead, D. Davies, L.C.A. Seavers, P.G. Wyatt, P. Workman, E. McDonald, *Bioorg. Med. Chem*., **2006**, 14(4), 1255.
[4] S. Barnett, M. Bilodeau, C. Lindsley, *Curr. Topics Med. Chem*., **2005**, 5, 109.
[5] Q. Li, G.D. Zhu, *Curr. Topics Med. Chem*., **2002**, 2, 939.
[6] K.M. Nicholson, N.G. Anderson, *Cell. Signal*., **2002**, 14, 381.
[7] I. Vivanco, C. Sawyers, *Nat. Rev*., **2002**, 2, 489.
[8] J. Gills, P. Dennis, *Expert Opin. Investig. Drugs*, **2004**, 13, 787.
[9] C. Lindsley, Z. Zhao, H. William, R. Robinson, S. Barnett, D. Defeo Jones, R. Jones, G. Hartman, J. Huff, H. Huber, M. Duggan, *Bioorg. Med. Chem. Lett.*, **2005**, 15, 761.
[10] G.D. Zhu, V.B. Gandhi, J. Gong, Y. Luo, X. Liu, *Bioorg. Med. Chem. Lett.*, **2006**, 16, 3424-3429.
[11] H. Lin, D.S. Yamashita, J. Zeng, R. Xie, W. Wang, *Bioorg. Med. Chem. Lett.*, **2010**, 20, 673-678.
[12] A. Golbraikh, A. Tropsha. J. *Mol. Graph. Model*., **2002**, 20, 269-276.
[13] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, *Bioorg. Med. Chem*., 14, **2006**, 6686-6694.
[14] D. Kim, S. Hong, D. Lee, Int. J. Mol. Sci., **2006**, 7, 556-570.
[15] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, *J. Comput. Aided. Mol. Des.*, **2006**, 20, 83-95.
[16] L.B. Kier, L.H. Hall, *J. Pharmacol. Sci*., **1983**, 72, 1170-1173.