International Scholarly Research Network
ISRN Physical Chemistry
Volume 2012, Article ID 260171, 11 pages
doi:10.5402/2012/260171

Research Article

A QSPR Study for the Prediction of the p*Ka* of N-Base Ligands and Formation Constant *Kc* of Bis(2,2′-bipyridine)Platinum(II)-N-Base Adducts Using Quantum Mechanically Derived Descriptors

Selami Palaz,[1] Baki Türkkan,[2] and Erol Eroğlu[3]

[1] *Physics Department, Faculty of Sciences and Arts, Harran University, Osmanbey, 63300 Sanliurfa, Turkey*
[2] *Chemistry Department, Faculty of Sciences and Arts, Harran University, Osmanbey, 63300 Sanliurfa, Turkey*
[3] *Primary Science Education Department, Faculty of Education, Akdeniz University, Dumlupinar Bulvari, 07058 Antalya, Turkey*

Correspondence should be addressed to Erol Eroğlu, eeroglu@akdeniz.edu.tr

Received 9 July 2012; Accepted 29 July 2012

Academic Editors: J. G. Han and W. P. Hu

Quantitative structure-property relationship (QSPR) study on the acid dissociation constant, p*Ka* of various 22 N-base ligands including pyridines, pyrimidines, purines, and quinolines has been carried out using Codessa Pro methodology and software. In addition, the formation constant, *Kc* of these ligands with $Pt(II)(bpy)_2^{2+}$ (bpy = 2,2′-bipyridine) ion has also been modelled with the same methodology. Linear regression QSPR models of p*Ka* and *Kc* were established with descriptors derived from AM1 calculations. Among the obtained QSPR models of p*Ka* presented in the study, statistically the most significant one is a four parameters linear equation with the squared correlation coefficient, R^2 values of ca. 0.95 and the squared cross-validated correlation coefficient, R^2_{cv} values of ca. 0.89, and external the squared correlation coefficient, $R^2_{ext.}$ values of ca. 0.97. Statistically the most significant QSPR model of *Kc* is also a four parameters linear equation with the squared correlation coefficient, R^2 values of ca. 0.75 and the squared cross-validated correlation coefficient, R^2_{cv} values of ca. 0.55, and external the squared correlation coefficient, $R^2_{ext.}$ values of ca. 0.81. An analysis of descriptors that involved in the p*Ka* models indicate that reactivity index and charge distribution related descriptors play major roles to model acid dissociation constant of ligands of N bases.

1. Introduction

The acid dissociation constant, p*Ka,* which describes the extent to which a compound dissociates in an aqueous solution, is a fundamental physical property of a chemical. The p*Ka* value plays a fundamental role in many analytical procedures such as acid-base titrations, solvent extraction, complex formation, and ion transport, and is especially relevant in medicinal chemistry because it affects ADME and activity. The degree of ionization determines permeability and solubility, two properties widely used in pharmaceutical research to predict the pharmacokinetic profile of a compound. Differences in adsorption, toxicology, solubility, bioconcentration, and reactivity are common when comparing the properties of the ionized molecule to its neutral form [1]. All of these reasons make the p*Ka* a common physicochemical property as a descriptor for QSAR and QSPR (quantitative structure and property relationship) studies. The experimental p*Ka* values of organic compounds have been largely determined through well-established methods such as potentiometry, UV-visible absorption spectrometry, conductimetry, and competitive reactions. A p*Ka* value of a compound may have to be determined, as it may not be available in published literature. Therefore, it is of interest to develop modeling methods for estimating the p*Ka* of compounds, and use these methods to predict the properties of a chemical in a solvent environment. Appropriate modeling methods may minimize cost and time to predict accurately p*Ka* values of unknown compounds.

Numerous computational studies have attempted to predict p*Ka* of organic compounds by applying different theoretical models [2–13]. Three main approaches have been employed in these theoretical models. First approach relieds upon using linear free energy relationships (LFER). The primary disadvantage of the purely LFER-based approaches is the need to derive a vast number of fragment constants and correction factors which are used in the estimation methods. Some researchers have also criticized the use of LFER approaches on more philosophical grounds, arguing that the prediction of molecular properties by fragment constant methods lacks solid scientific support [14, 15]. A second approach is based on accurate energy calculation of the molecules in their neutral and deprotonated forms in the gas phase and solvent, to account for the solute-solvent interaction. In order to obtain adequately accurate energies to calculate solution phase dissociation constants, one must account for electron correlation at the *ab initio* or density functional theory (DFT) methods, and consider the effects of solvation on the molecule. When moderate- or relatively large-sized molecules are considered, due to the great computational cost of *ab initio* or DFT calculations this approach is not feasible, especially for virtual screening applications. Third is the QSPR approach that is a mathematical equation relating chemical structure to a wide variety of physical, chemical, biological, and technological property.

In the study herein, one of the p*Ka* modeling approaches described above, the QSPR method has been used to construct quantitative models to predict p*Ka* values of various 22 N-bases ligands including pyridines, pyrimidines, purines, and quinolines. In addition, we have developed QSPR models to predict the formation constant, *Kc* of reactions of these ligands with $Pt(II)(bpy)_2^{2+}$ (bpy = 2,2′-bipyridine) ion. After the discovery of the cell division-inhibiting effect of cisplatin by Rosenberg et al. [16], there is considerable interest in platinum chemistry. To date, six platinum complexes (including cisplatin) have been used in drug anticancer chemotherapy [17]. To obtain p*Ka* and *Kc* QSPR models, program Codessa Pro (comprehensive descriptors for structural and Statistical Analysis), Version 2.7.2 [18] was employed to build a multilinear regression (MLR) method using observed p*Ka* and *Kc* values taken from Kawanishi et al. [19] in combination with the descriptors which were calculated by the AMPAC [20] semiempirical quantum chemistry code.

2. Results and Discussion

The structures of 22 N-base ligands are shown in Figure 1. Table 1 shows the following information QSPR modeling of p*Ka* of compounds: (i) AM1-based calculated molecular descriptor values involved in the obtained models; (ii) experimental p*Ka* values taken from the original references; (iii) the predicted p*K* values using Model I, II, and III obtained in this study. Table 2 shows the obtained QSPR equations together with their statistical parameters for p*Ka* modeling. Table 3 shows the intercorrelation of descriptors involved in the p*Ka* models. The predicted values of p*Ka* using Model I is plotted versus the experimental values in Figure 2. Reliability of Model I was tested by the Y-randomization test. 250 random shuffles of the Y (p*Ka*) were chosen, and the modeling process was performed for all cases of Model I. Results are shown in Figure 3. It should be noted that all of the results of the 250 random shuffles of the Y were not included, only the highest 60 data points were taken in the Figure 3. The lower values of R^2 and R^2_{cv} in comparison with the real model's results support the hypothesis that the good statistical results obtained by the QSPR model are not due to a chance correlation, or structural dependency of the training set.

Table 4 shows the following information for *Kc* QSPR modeling: (i) AM1-based calculated molecular descriptor values involved in the models; (ii) experimental *Kc* values taken from the original references; (iii) the predicted *Kc* values using Model IV, V, and VI obtained in this study. Figure 4 shows plot of experimental versus calculated *Kc* using Model IV. Table 5 shows the obtained QSPR equations together with their statistical parameters for *Kc* modeling. Table 6 shows the intercorrelation of descriptors involved in *Kc* models. At this point, it should be observed that Tables 3 and 6 demonstrate that although some descriptors exhibit relatively high intercorrelation, they are not involved in the same models as seen in Tables 2 and 5. Reliability of the best *Kc* model (Model IV) was also tested by the Y-randomization test. 250 random shuffles of the Y (*Kc*) were chosen, and the modeling process was performed for all cases of Model IV from which the results are shown in Figure 5. It should be noted that all of the results of the 250 random shuffles of the Y were not included, only the highest 60 data points were taken in the Figure 5. The lower values of R^2 and R^2_{cv} in comparison with the real model's results support the hypothesis that the good statistical results obtained by the QSPR model are not due to a chance correlation, or structural dependency of the training set.

2.1. Analysis of pKa Models. A perusal of Table 2 shows that nine types of descriptors are involved in the three p*Ka* models. Among the three models, the best, statistically speaking, is Model I, as evident from the fact that it has a very good statistical fit ($R^2 = 0.95$, $F = 61.79$, $S^2 = 0.311$) and predictive ability ($R^2_{cv} = 0.89$, $R^2_{ext.} = 0.97$, RMSE = 0.51) parameters. By analyzing Model I, it is clear that the most representative descriptor is *HNMVF* with the positive coefficient and the highest *t*-test value. The magnitude of *HNMVF* of a molecule comes from O-H, N-H, or C-H stretching vibrations. Hybridization state (sp, sp^2, and sp^3) also affects the magnitude of *HNMVF* of a molecule. An increase in the magnitude of *HNMVF* favors the exhibitions of the p*Ka* of a N-base ligands. In this model, FNSA3 and FPSA which are charged partial surface area (CPSA) descriptors, also have positive coefficients. They describe the polar interactions between the molecules. *MNRI-C* has a negative coefficient, highlighting that the lower the magnitude of MNRI-C, the lower the p*Ka*. This descriptor estimates the relative reactivity of the atom (carbon) in the molecule, and is related to the activation energy of the corresponding chemical reaction.

Table 1: Molecular descriptors values involved in the QSPR models for N-bases and predicted p*Ka* values.

Comp. no.	*HNMVF*	*FNSA3*	*FPSA1*	*MNRI-C*	*MERI-C*	HDSA1/ *TMSA*	*WPSA2*	Q_{max}	*ESP-HDCA1/ TMSA*	Exp. p*Ka*	Model I p*Ka*	Model II p*Ka*	Model III p*Ka*
1	3.2116*E* + 03	−0.0121	0.6971	0.0236	0.0255	0.0000	67.2803	0.0306	0.0000	5.1700	5.3301	5.2767	4.7608
2[a]	3.2085*E* + 03	−0.0126	0.7220	0.0268	0.0298	0.0927	94.9803	0.0310	0.0175	5.6800	5.2118	5.8955	5.9027
3	3.2070*E* + 03	−0.0106	0.7566	0.0238	0.0254	0.0996	146.4419	0.0308	0.0176	6.0000	6.4286	6.6858	6.5446
4	3.2163*E* + 03	−0.0177	0.8605	0.0248	0.0283	0.0962	148.5683	0.0330	0.0197	6.4700	6.2635	5.5030	4.4868
5[a]	3.2047*E* + 03	−0.0228	0.5736	0.0246	0.0313	0.0000	45.5035	0.0339	0.0000	1.4500	0.7797	3.5730	1.4487
6	3.2032*E* + 03	−0.0208	0.6012	0.0227	0.0329	0.0000	84.2644	0.0325	0.0000	1.9000	1.7858	2.4557	2.0669
7	3.2167*E* + 03	−0.0284	0.7932	0.0270	0.0298	0.0825	99.7973	0.0866	0.0228	3.2300	2.5117	2.8526	1.3664
8[a]	3.2087*E* + 03	−0.0109	0.6039	0.0236	0.0222	0.0000	117.2082	0.0311	0.0000	4.8000	4.2731	5.5081	5.1457
9	3.2038*E* + 03	−0.0111	0.6304	0.0181	0.0201	0.0000	133.0226	0.0319	0.0000	5.4000	4.9922	5.9022	5.0948
10	3.2124*E* + 03	−0.0186	0.6039	0.0283	0.0359	0.0000	27.6513	0.0356	0.0000	2.3300	2.0476	2.2486	2.7407
11	3.1506*E* + 03	−0.0184	0.7204	0.0367	0.0353	0.0000	64.1077	0.0426	0.0000	1.3000	1.6466	1.5161	2.8224
12	3.2027*E* + 03	−0.0146	0.7057	0.0145	0.0180	0.0000	244.6293	0.0311	0.0000	4.8200	5.4933	4.8015	4.0070
13	3.5371*E* + 03	−0.0258	0.6914	0.0197	0.0246	0.2577	143.8780	0.0564	0.1023	9.7500	9.2987	9.9944	8.0521
14	3.4841*E* + 03	−0.0386	0.6667	0.0257	0.0359	0.3736	419.7091	0.0840	0.1063	3.3000	4.1432	3.5748	4.3398
15[a]	3.4711*E* + 03	−0.0442	0.5711	0.0383	0.0474	0.1807	59.1654	0.0808	0.0538	0.6000	0.0191	0.6357	−1.2911
16	3.5160*E* + 03	−0.0330	0.6513	0.0439	0.0373	0.2588	139.8377	0.0842	0.0976	4.5800	4.3433	4.5739	5.4486
17	3.2125*E* + 03	−0.0100	0.7168	0.0254	0.0314	0.0972	129.9713	0.0307	0.0199	5.9600	5.9760	4.9171	6.8860
18	3.2150*E* + 03	−0.0256	0.6869	0.0282	0.0338	0.0788	83.5174	0.0903	0.0209	1.2500	1.5557	1.5127	2.0909
19[a]	3.2055*E* + 03	−0.0272	0.6545	0.0226	0.0252	0.0905	158.3118	0.0890	0.0164	2.2100	2.3704	1.7022	2.9931
20	3.1934*E* + 03	−0.0255	0.6964	0.0226	0.0226	0.0874	163.0047	0.0833	0.0168	3.2600	2.4503	2.7545	2.7138
21	3.2156*E* + 03	−0.0195	0.7263	0.0274	0.0334	0.0962	127.2344	0.0461	0.0191	3.0600	3.6848	3.5426	3.8711
22	3.2076*E* + 03	−0.0081	0.7258	0.0245	0.0246	0.0895	186.8004	0.0233	0.0162	6.7100	6.5386	6.3780	7.1975

[a] Compounds in the test set.

Table 2: QSPR models of p*Ka* for N-base ligands.

Models	Descriptors involved	C[a]	t-test	Statistical parameters					
				Training set (N = 17)			Test set (N = 5)		
				R^2	R^2_{cv}	F	S^2	$R^2_{ext.}$	RMSE
Model I p*Ka*	*Intercept*	−68.69	−12.73	0.95	0.89	61.79	0.311	0.97	0.51
	HNMVF	0.0216	13.60						
	FNSA3	244.02	10.22						
	FPSA1	13.96	6.33						
	MNRI-C	−89.88	−3.92						
Model II p*Ka*	*Intercept*	16.23	16.43	0.94	0.89	56.62	0.33	0.85	1.11
	MERI-C	−328.98	−10.45						
	HDSA1/TMSA	26.94	11.58						
	WPSA2	−0.0163	−6.51						
	Q_{max}	−48.37	−6.38						
Model III p*Ka*	*Intercept*	8.55	10.30	0.76	0.61	23.37	1.32	0.91	0.93
	HDCA1/TMSA	73.89	6.43						
	FNSA3	312.25	6.29						

[a] C: indicates the coefficients of the descriptors involved in the models, R^2: the square of the regression correlation coefficient, R^2_{cv}: the cross-validated square of the regression correlation coefficient, F: the F-value for the regression, S^2: the standard deviation of the regression, $R^2_{pred.}$: the predicted square of the regression correlation coefficient, RMSE: the root mean square error.

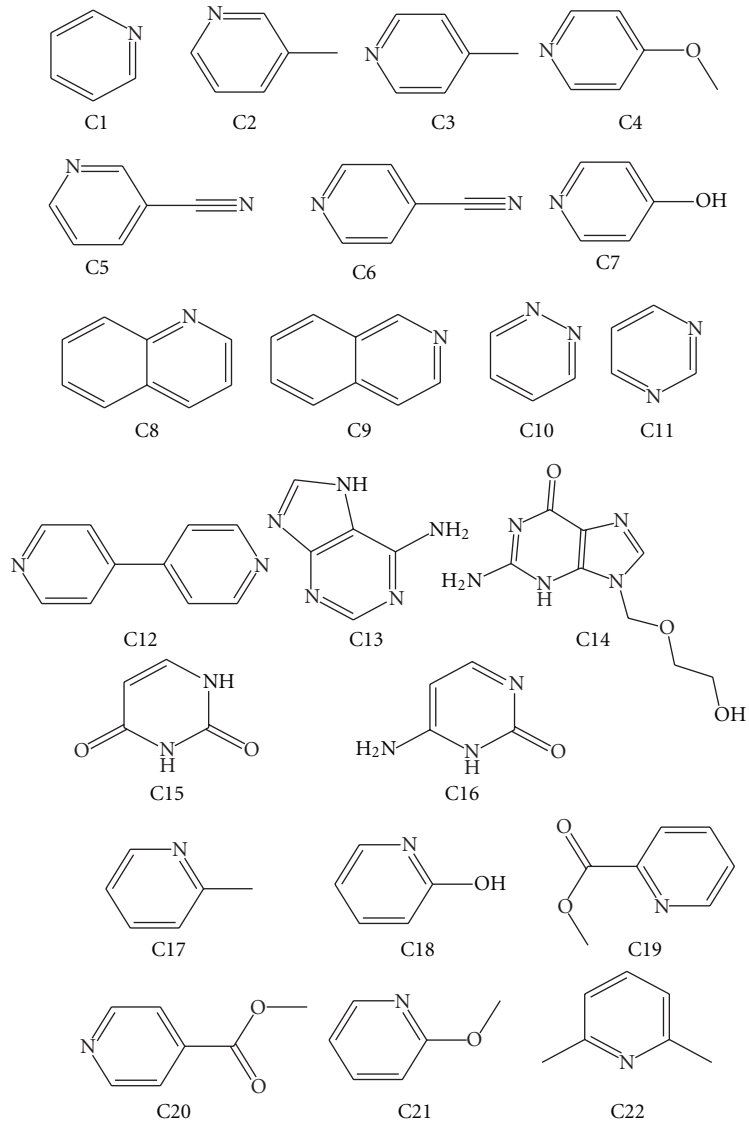

Figure 1: Structure of the 22 N-base ligands studied.

Table 3: Correlation matrix for the intercorrelation of various molecular descriptors involved in the Model I, II, and III.

	HNMVF	*FNSA3*	*FPSA1*	*MNRI-C*	*MERI-C*	*HDSA1/TMSA*	*WPSA2*	Q_{max}	*ESP-HA/HDCA*
HNMVF	1								
FNSA3	−0.665	1							
FPSA1	−0.198	0.152	1						
MNRI-C	0.252	−0.396	−0.034	1					
MERI-C	0.246	−0.508	−0.191	0.784	1				
ESP-HA/HDSA	0.888	−0.694	0.033	0.255	0.304	1			
ESP-WPSA2	0.451	−0.368	0.065	−0.245	−0.137	0.661	1		
Q_{max}	0.464	−0.87	−0.033	0.389	0.361	0.558	0.228	1	
ESP-HA/HDCA	0.971	−0.719	−0.072	0.289	0.293	0.965	0.542	0.559	1
Exp. p*Ka.*	0.390	0.304	0.302	−0.377	−0.546	0.306	0.201	−0.338	0.343

Table 4: Molecular descriptors values involved in the QSPR models and predicted *Kc* values for Bis(2,2′-bipyridine)platinum(II)-N-base adducts reactions.

Comp. no.	*DPSA1-D* in *CPSA*	*Int. entropy*	*ESP-FNSA3*	*Min e-eREP-N*	*BETA* Pol.	*Min Re.E-CN*	*Avg 1-e RI-C*	*Max n-n Re.-CN*	Exp. *Kc*	Model IV Pre. *Kc*	Model V Pre. *Kc*	Model VI Pre. *Kc*
1[a]	101.94	2.7079	−0.1000	131.2770	−28.47	11.0890	$0.0000E+00$	157.2510	2.700	3.1828	2.2619	4.3376
2	130.35	2.4482	−0.0810	131.1470	6.91	10.5260	$-7.2350E-04$	157.3980	8.600	7.8093	7.1075	7.2795
3	149.83	2.8800	−0.1108	131.4820	33.70	10.4940	$5.1936E-07$	157.2830	6.100	4.5208	4.2888	3.9780
4	216.70	2.8167	−0.1159	132.5050	84.58	10.6180	$4.1645E-05$	157.0280	7.000	7.2678	5.1309	5.3312
5	45.10	3.1756	−0.0879	125.3510	13.48	11.0730	$1.5784E-03$	157.3760	0.130	1.2985	−0.6526	−0.9633
6	61.70	3.1473	−0.1144	124.9290	44.92	11.0780	$-1.5936E-07$	157.3900	0.420	−0.1410	0.0666	0.6984
7[a]	159.34	2.9009	−0.1551	132.4970	59.29	11.0750	$1.6798E-04$	156.9030	0.720	1.5647	2.2465	3.5789
8	66.98	2.4303	−0.0970	130.8520	−21.34	11.0650	$1.4414E-03$	156.8270	3.800	4.6372	4.6507	4.2829
9[a]	82.19	2.4327	−0.1034	131.2590	−14.11	11.0710	$4.3553E-04$	157.1140	9.000	4.5363	4.7271	5.4907
10	50.59	2.9265	−0.0670	126.4770	−55.00	11.1310	$-1.6359E-05$	156.5620	0.510	1.7989	−0.0543	1.4260
11	97.85	2.9292	−0.1397	132.6640	−18.38	11.0550	$2.7492E-03$	156.6200	0.140	−0.5596	0.8269	0.2038
12[a]	152.49	2.6136	−0.1088	130.9570	−0.021	11.0830	$-3.1001E-07$	157.0540	5.400	3.7888	3.4930	5.7044
13[b]	—	—	—	—	—	—	—	—	—	—	—	—
14	154.02	3.1831	−0.1271	127.1010	68.32	10.5540	$-8.2424E-05$	159.2140	2.100	0.7716	2.2433	3.4761
15	40.09	3.3032	−0.1669	140.9640	−10.75	11.0700	$1.7900E-03$	157.9870	0.130	0.3506	−2.0398	−1.6271
16	84.92	3.1387	−0.2321	135.7030	206.26	10.9990	$-1.3541E-03$	155.2190	0.590	1.4546	3.0979	0.4967
17	126.62	2.8740	−0.0974	131.5120	−55.20	10.5360	$3.5327E-04$	157.2230	0.540	1.6894	2.7082	3.3170
18[a]	101.25	2.9361	−0.1388	131.5760	−100.44	11.0590	$-1.7656E-03$	156.0350	0.130	−5.0899	−0.5915	3.1158
19	105.84	3.2014	−0.1071	128.8170	4.00	10.6630	$6.5902E-04$	155.8870	0.100	0.7196	0.6116	−0.5665
20	137.24	3.1625	−0.1072	130.0300	−38.43	10.6630	$4.1303E-06$	157.0020	0.210	−0.4779	0.2191	1.5778
21	135.54	2.8629	−0.1047	131.6770	−79.47	10.6040	$1.6487E-03$	156.2280	0.190	−0.1194	2.1270	1.8236
22	143.20	3.0683	−0.0898	131.7610	−100.11	10.5380	$7.1287E-04$	154.6840	0.210	−0.2504	0.4383	0.0360

[a] Compounds in the test set.
[b] Compound 13 is an outlier in all the *Kc* model.

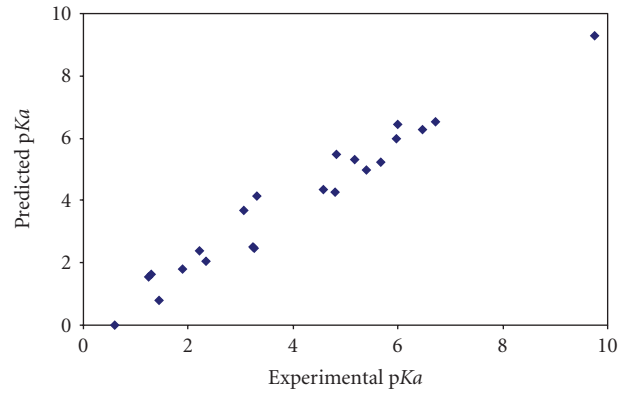

Figure 2: Plot of experimental versus calculated p*Ka* using Model I.

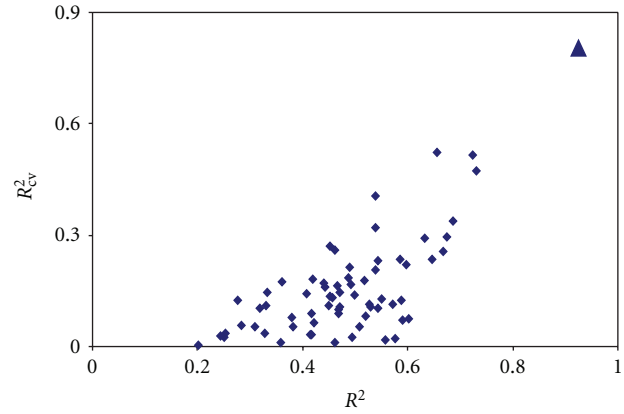

Figure 3: Y-randomization test associated to p*Ka* Model I. Squares represent the randomly ordered activities, and the triangle corresponds to the real model.

Model II is also a four-parametric model, similar to Model I. This model also has very good statistical fit (R^2 = 0.94, F = 56.62, and S^2 = 0.33) and good predictive ability (R^2_{cv} = 0.89, $R^2_{ext.}$ = 0.85, RMSE = 1.11) parameters. The descriptors involved in this model are *MERI-C*, *HDSA1/TMSA*, *WPSA2,* and Q_{max}. In this model, only *HDSA1/TMSA* has a positive coefficient. *HDSA1/TMSA* is statistically the most significant descriptor, as evident from the fact that it has the highest *t*-test value. *HDSA1* is the hydrogen bonding donor ability of a molecule and *TMSA* is total molecular surface area of a molecule. An increase in the magnitude of *HDSA1/TMSA* favours the exhibitions of the p*Ka* of N-base ligands. *MERI-C* is the the second statistically

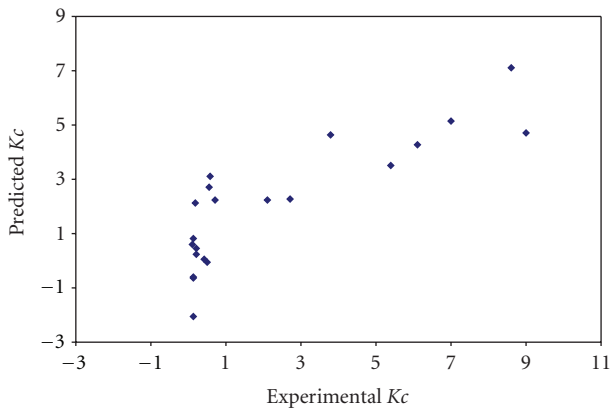

Figure 4: Plot of experimental versus calculated *Kc* using Model IV.

significant descriptor with a negative coefficient. This is a relative reactivity index, and corresponds to activation energy of a corresponding chemical reaction. *WPSA2,* and Q_{max} are related to charge distribution of the molecules. In this model, negative coefficients of *MERI-C*, WPSA2 and Q_{max} indicate that decreases in the magnitudes of *MERI-C*, *WPSA2* and Q_{max} are not favorable for an increase of p*Ka* of the molecules.

Model III is a two-parametric model. *HDCA1/TMSA* and *FNSA* descriptors are involved in this model with positive coefficients. These are CPSA descriptors, and describe the polar interactions between the molecules. Although, the goodness of fit of this model is only satisfactory, evident from the fact that it has $R^2 = 0.76$, $F = 23.37$, and $S^2 = 1.32$, its external predictive power is very good ($R^2_{ext.} = 0.91$). It is worthy here to mention that model validation is historically a hot topic for discussions within the QSAR community. Some researchers use only internal R^2_{cv} for model validation, and some others strongly believe that external validation is vital for a reliable QSAR model. If one surveys the QSAR literature, it can be shown that many QSAR models have been published only with internal validation until a study published by A. Golbraikh and A. Tropsha, entitled "Beware of *q*2!" [21]. In that study, the authors reexamined several published QSAR data sets, and demonstrated that there is lack of any relationship between R^2_{cv} and $R^2_{ext.}$. They concluded that the high value of R^2_{cv} appears to be the necessary, but not sufficient condition, for the model to have a high predictive power and the external validation is the only way to establish a reliable QSAR model. Our results regarding the p*Ka* models support these conclusive remarks. Model II has a higher internal validation value, $R^2_{cv} = 0.89$, but its external validation value, $R^2_{ext.}$ is only 0.85, whereas Model III has a lower internal validation value, $R^2_{cv} = 0.61$, and its external validation value, $R^2_{ext.}$ is very good 0.91.

2.2. Analysis of Kc Models. A perusal of Table 5 shows that nine types of descriptors are involved in the three *Kc* models. Among the three *Kc* models, Model IV has good fit ($R^2 = 0.89$, $F = 23.33$, and$S^2 = 1.16$) parameters, whereas its predictive ability only is acceptable ($R^2_{cv} = 0.74$, $R^2_{ext.} = 0.53$, RMSE = 3.18). *BETA* (β)*Polarizability, Internal Entropy, ESP-FNS3,* and *Min e-e REP-N* are involved in this model. *BETA* (β)*Polarizability* which has a positive coefficientisthe most representative descriptor,as evident from the fact that it has the highest *t*-test value. *Internal Entropy* divided by number of atoms in the molecules which has a negative coefficient in the model, is the least representative descriptor, as evident from the fact that it has the lowest *t*-test value. Also, the positive sign of coefficient of *ESP-FNS3* and *Min e-e REP-N* indicate that increases in the magnitude of *ESP-FNS3,* and *Min e-e REP-N* are favorable for an increase of *Kc* of Bis(2,2′-bipyridine)platinum(II)-N-base adduct reactions. Mechanistic interpretation of the model is quite complex due to the diverse nature of the involved descriptors. *BETA* (β)*Polarizability* reflects information about possible inductive interactions in the molecule. It also characterizes the properties of a molecule as an electron acceptor. *ESP-FNS3* is a CPSA descriptor, and describes the polar interactions between the molecules. *Min e-e REP-N* describes the electron repulsion driven processes in the molecule, and may be related to the atomic reactivity in the molecule [22].

In Model V, *BETA* (β)*Polarizability* and *Internal Entropy* have the same sing of coefficients as Model IV. *Internal Entropy* is the most representative descriptor in this model. The third descriptor is the *Min Re. E-CN* which is an energy-related descriptor with a positive sign of coefficient. Statistical parameters of this model are similar to Model III in term of validation parameters. Although this model has the lowest internal validation ($R^2_{cv} = 0.44$) value, its external predictive ability, $R^2_{ext.} = 0.77$ is relatively high. When these six *Kc* and p*Ka* models are considered together, one could draw a conclusion that there is not any relationship between R^2_{cv} and $R^2_{ext.}$.

The final model, Model VI has four descriptors. *Internal Entropy* is also the most representative descriptor in this model. *Internal Entropy* and avg. *1-e RI-C* have a positive sign of coefficient whereas *max n-n Re.-CN* and *DPSA1D* in *CPSA* have negative sign of coefficient. As observed for Model IV and V, an increase of in the magnitude of *Internal Entropy* divided by number of atoms in the molecules favours the exhibitions of the *Kc* for Bis(2,2′-bipyridine)platinum(II)-N-base adduct reactions. It may be concluded that the entropy changes seem to have an important effect on the Bis(2,2′-bipyridine)platinum(II)-N-base adduct reactions. Although, it has not been included as a Model in Table 5, but *Internal Entropy* divided by number of atoms in the molecules, itself as a monoparametric model has an acceptable statistical parameter, $R^2 = 0.48$, $F = 12.91$, $S^2 = 4.5$, $R^2_{cv} = 0.3$, $R^2_{ext.} = 0.97$ and RMSE = 1.89. It is worthy to mention that all of the *Kc* models have one nitrogen-related descriptor (*Min e-e REP-N*, *Min Re. E-CN* and *Max n-n Re.-CN*). When the proposed Bis(2,2′-bipyridine)platinum(II)-N-base adduct by formation of N-Pt(II) bond [19] is considered, the appearance of N-related descriptors in all of the models is not a surprise.

Finally, as mentioned in the introduction section, p*Ka* can be seen as a descriptor for QSAR and QSPR studies in literature. In this study, p*Ka* has been tested as a descriptor

Table 5: QSPR models of *Kc* for Bis(2,2′-bipyridine)platinum(II)-N-base adducts reactions.

Models	Descriptors involved	C^a	t-test	Statistical parameters					
				Training set ($N = 16$)			Test set ($N = 5$)		
				R^2	R^2_{cv}	F	S^2	$R^2_{ext.}$	RMSE
Model IV *Kc*	*Intercept*	−51.133	−2.9682	0.89	0.74	23.33	1.16	0.53	3.18
	(1/2)×BETA (β) polarizability	0.0965	6.5208						
	Internal entropy at 300 K/no. of atoms in mo.	−3.7361	−2.6391						
	ESP-FNS3	106.17	5.3226						
	Min e-e REP-N	0.5821	4.5888						
Model V *Kc*	*Intercept*	67.834	3.8027	0.74	0.44	11.61	2.58	0.77	2.23
	BETA (β) polarizability	0.0164	2.8619						
	Internal entropy at 300 K/no. of atoms in mo.	−7.8412	−4.6956						
	Min Re. E-CN	−3.9563	−2.3851						
Model VI *Kc*	*Intercept*	−98.836	−1.5907	0.75	0.55	8.7	2.64	0.81	2.53
	Internal entropy at 300 K/no. of atoms in mo.	−6.9666	−4.1294						
	Avg. 1-e RI-C	−741.89	−1.7774						
	Max n-n Re.-CN	0.7650	1.9331						
	DPSA1D in CPSA	0.0170	1.8440						

[a] C: the coefficients of the descriptors involved in the models, R^2: the square of the regression correlation coefficient, R^2_{cv}: the cross-validated square of the regression correlation coefficient, F: the F-value for the regression, S^2: the standard deviation of the regression.

Table 6: Correlation matrix for the intercorrelation of various molecular descriptors involved in the obtained QSPR *Kc* models.

	DPSA1-D in *CPSA*	*BETA* pol.	*Int. entropy*	*ESP-FNSA3*	*Min e-eREP-N*	*Min re.E-NH*	*Avg 1-eRI-C*	*Max n-n re.-CN*
DPSA-1 D in *CPSA*	1							
BETA pol.	0.0589	1						
Int. entropy	−0.2198	0.1726	1					
ESP-FNSA3	0.1137	−0.7118	−0.3991	1				
Min e-e REP-N	0.0489	0.1131	0.0110	−0.6106	1			
Min re. E-NH	−0.8457	0.1517	0.1730	−0.2553	−0.0182	1		
Avg 1-e RI-C	−0.2928	−0.5384	0.0386	0.1479	0.1486	0.3245	1	
Max n-n Re.-CN	0.0004	0.1123	0.0288	0.1108	−0.1606	−0.0185	0.0339	1
Exp. *Kc*	0.5104	0.2522	−0.6925	0.2009	0.0628	−0.4138	−0.3750	0.2586

for the modeling of *Kc*, but it has failed during the preselection of the descriptors by the software (Codessa Pro) due to its very low squared correlation coefficient of the one-parameter equation is less than R^2_{min} 0.01 by default.

3. Experimental Section

3.1. Quantum Chemical Calculations. All structures were optimized without geometry constraints using the standard AM1 Hamiltonian [23] within the AMPAC quantum chemistry code [20]. Structure fundamental vibrations were also calculated using the same method to check if there were true minima. All computations were carried out for the ground states of these molecules as single states. Output files of the molecules from AMPAC code were used as the input file for Codessa Pro [18] code for descriptor generation.

3.2. Descriptors Generation and Their Definitions. In the present work, more than five hundred descriptors were exploited by using Codessa Pro code, and they were divided into groups such as constitutional, topological, geometrical, electrostatic, quantum chemical, thermodynamic, and contracted. Constitutional descriptors are related to the number of atoms and bonds in each molecule. Topological descriptors include valence and nonvalence molecular connectivity indices calculated from the hydrogen-suppressed formula of the molecule, encoding information about the size, composition, and the degree of branching of a molecule. Geometrical descriptors are calculated from 3D atomic coordinates of the molecule and comprise moments of inertia, shadow indices, molecular volumes, molecular surface areas, and gravitation indices. Electrostatic descriptors reflect characteristics of the charge distribution of the molecule. Quantum chemical

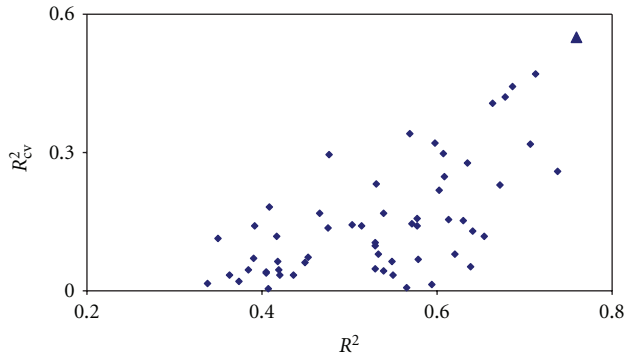

Figure 5: Y-randomization test associated to *Kc* Model IV. Squares represent the randomly ordered activities, and the triangle corresponds to the real model.

descriptors encode the polar interactions between molecules or their chemical reactivity and the activation energy of the corresponding chemical reaction. Thermodynamic descriptors are quantum mechanically calculated on the basis of the total partition function of the molecule, Q and its electronic, translational, rotational, and vibrational components. Codessa Pro also allows one to construct new descriptors by using the existing descriptors. In this way, the author constructed some common quantum chemical indices, namely, chemical hardness, electronegativity, and electrophilicity from HOMO and LUMO orbital energies. More than five hundred descriptors, andseventeen types of descriptors were involved in the selected p*Ka* and *Kc* models as shown in Table 1 and Table 4. Their definitions are adopted from the CODESSA Pro Manual [24] as described below.

Highest normal mode vibrational frequency (*HNMVF*) is an extreme maximum value of the normal mode vibrational frequencies in the molecule. Definition of the normal mode of vibration arises from the quantum mechanical Harmonic Oscillator model of a diatomic molecule.

Fractional PNSA (PNSA3/TMSA) or *FNSA3* is one of the electrostatics charged partial surface area (CPSA) descriptors, and was invented by Jurs et al. [25, 26]. Definition of *FNSA3* is atomic charge weighted partial negative surface area (PNSA) divided by total molecular surface area (TMSA) which has related features responsible for polar interactions between molecules.

Fractional CPSA (PPSA1/TMSA), FPSA1 is another CPSA descriptor and its definition is given as partial positive surface area (PPSA) divided by TMSA.

Maximum nucleophilic reaction index for a C atom, MNRI-C is one of the quantum-chemically calculated charge distribution-related reactivity indices. These descriptors represent or depend directly on the quantum-chemically calculated charge distribution in the molecules, and describes the polar interactions between molecules or their chemical reactivity. Definition of *MNRI-C* in a molecule is given as follows:

$$\text{MNRI-C} = \sum_{i=C} \frac{C_i\text{HOMO}^2}{(1 - \varepsilon_{\text{HOMO}})}, \tag{1}$$

where the summations are performed over all of the atomic orbitals, i of a Carbon atom in a molecule, C_i HOMO denotes the ith Atomic Orbital (AO) coefficient on the highest occupied molecular orbital (HOMO) and $\varepsilon_{\text{HOMO}}$ is the energy of HOMO orbital. The reactivity indices estimate the relative reactivity of the atoms in the molecule for a given series of compounds, and refer to the activation energy of the corresponding chemical reaction.

Maximum electrophilic reaction index for a C atom, MERI-C is another quantum-chemically calculated charge distribution-related reactivity index and its definition is given as follows:

$$\text{MERI-C} = \sum_{i=C} \frac{C_i\text{LUMO}^2}{(\varepsilon_{\text{LUMO}} + 10)}, \tag{2}$$

where the summations are performed over all atomic orbitals, i of a carbon atom in a molecule, C_i LUMO denotes the ith AO coefficient on the lowest unoccupied molecular orbital (LUMO), and $\varepsilon_{\text{LUMO}}$ is the energy of LUMO.

ESP-HA dependent HDSA1/TMSA, HDSA1/TMSA is one of the quantum-chemically calculated descriptors.

HDSA1 is the hydrogen bonding donor ability of a molecule and its definition is given as follows:

$$\text{HDSA1} = \sum_{D} s_D \quad D \in H_{\text{H-donor}}, \tag{3}$$

where S_D is the solvent-accessible surface area of H-bonding donor H atoms.

ESP-WPSA2 Weighted PPSA (*PPSA2* × *TMSA*/1000), *WPSA2* is the surface-weighted charged partial-positive charged surface area. This is one of the quantum-chemically calculated CPSA descriptors. Definition of *WPSA2* is given as follows:

$$\text{WPSA2} = \frac{(\text{PPSA2}) \times \text{TMSA}}{1000}, \tag{4}$$

where *PPSA2* is the total charge weighted by partial positive surface area and *TMSA* is the total molecular surface area.

Maximum partial charge, $Q_{\max}$ is one of the electrostatics descriptors and reflects characteristics of the charge distribution of the molecule. The empirical partial charges in the molecule are calculated using the approach proposed by Zefirov et al. [27, 28]. This method is based on the Sanderson electronegativity scale, and uses the concept which represents the molecular electronegativity as a geometric mean of atomic electronegativities. Definition of $Q_{\max}$ is given as follows:

$$Q_{\max} = \text{maximum } (Q^+), \tag{5}$$

where Q^+ is the positive atomic partial charges in the molecule.

ESP-HA/HDCA Dependent, HDCA1/TMSA is the hydrogen bonding donor ability of the molecule divided by *TMSA*. Definition of *HDCA1* is given as:

$$\text{HDCA1} = \sum_{D} s_D \quad D \in H_{\text{H-donor}}, \tag{6}$$

where S_D is the solvent-accessible surface area of the H-bonding donor H atoms, selected by threshold charge.

DPSA1 Difference in CPSAs (PPSA1-PNSA1) (Zefirov's PC), DPSA1D in CPSA is the difference between partial positively- and negatively-charged surface areas:

$$\text{DPSA1-D} \quad \text{in CPSA-} = \text{PPSA1-PNSA1}. \tag{7}$$

ALPHA (α) *and BETA* (β) *Polarizability* is the polarization of a molecule by an external electric field, and is given in terms of the nth order susceptibility tensors of the molecular bulk. The first-order term and the second-order term that contain information about possible inductive interactions in a molecule are referred to as α and β polarizabilitity, respectively, [29, 30]:

$$\mu' = \mu + \alpha E + \frac{1}{2}\beta E^2 + \ldots, \tag{8}$$

where μ is the permanent dipole moment of the molecule, μ' is the induced dipole moment of the molecule and E is the external electric field.

Internal Entropy at 300 K/(Number of atoms in the molecules) where the definition of internal entropy is given as

$$S_{\text{internal}} = S_{\text{vibrational}} + S_{\text{rotational}}, \tag{9a}$$

$$S_{\text{vibrational}} = \sum_{j=1}^{\alpha} \left\{ \frac{h\nu_j \exp\left(-\left(h\nu_j/2kT\right)\right)}{kT\left[1-\exp\left(-h\nu_j/2kT\right)\right]} - \ln\left[1-\exp\left(-\frac{h\nu_j}{2kT}\right)\right] \right\}, \tag{9b}$$

$$S_{\text{rotational}} = Nk\ln\left[\frac{\pi^{1/2}}{\sigma}\prod_{j=1}^{3}\left(\frac{8\pi^2 I_j kT}{h^2}\right)^{1/2}\right], \tag{9c}$$

where ν_j are the frequencies of normal vibrations in the molecule, I_j are the principal moments of inertia of the molecule, σ is the symmetry number of the molecule, h is the Planck's constant, $k$is the Boltzmann's constant and T is theabsolute temperature (K).

ESP-FNSA3 Fractional PNSA (PNSA3/TMSA) [Quantum-Chemical PC], ESP-FNS3 is the fractional atomic-charge weighted partial negative surface area and its definition is given as

$$\text{FNSA3} = \frac{\text{PNSA3}}{\text{TMSA}}, \tag{10}$$

where *PNSA3* is the total charge-weighted partial negatively charged molecular surface area.

Minimum e-e repulsion for an N atom, Min e-e REP-N where the extreme (maximum or minimum) values of the electron-electron repulsion energy are for a given atomic species (N) in the molecule, calculated as follows:

$$\text{Eee}(A) = \sum_{B\neq A}\sum_{\mu,\nu\in A}\sum_{\lambda,\sigma\in B} P_{\mu\nu}P_{\lambda\sigma}\langle\mu\nu \mid \lambda\sigma\rangle, \tag{11}$$

where $P_{\mu\nu}$ and $P_{\lambda\sigma}$ are the density matrix elements over atomic basis $\{\mu\nu\lambda\sigma\}$ and $\langle\mu\nu \mid \lambda\sigma\rangle$ are the electron repulsion integrals on the atomic basis $\{\mu\nu\lambda\sigma\}$. The electron-electron repulsion energy describes the electron repulsion driven processes in the molecule, and may be related to the conformational (rotational, inversional) changes or atomic reactivity in the molecule [31].

Minimum resonance energy for a C-H bond, Min Re. E-CN is a quantum mechanical energy-related descriptor and its definition is given as

$$E_R(AB) = \sum_{\mu\in A}\sum_{\nu\in \text{B}} P_{\mu\nu}\beta_{\mu\nu}, \tag{12}$$

where A is a given atomic species, B is the another atomic species, $P_{\mu\nu}$ is the density matrix elements over atomic basis $\{\mu\nu\}$, and $\beta_{\mu\nu}$ is the resonance integrals on atomic basis $\{\mu\nu\}$.

Average 1-electron reaction index for a C atom, Avg. 1-e RI-C is a charge distribution-related reactivity index and it estimates the relative reactivity of the atoms in the molecule for a given series of compounds, and this corresponds to the activation energy of the corresponding chemical reaction. Definition is given as

$$R_A = \frac{\sum_{i\in A}\sum_{j\in A} c_{i\text{HOMO}}c_{j\text{LUMO}}}{\varepsilon_{\text{LUMO}} - \varepsilon_{\text{HOMO}}}, \tag{13}$$

where $c_{i\text{HOMO}}$ is the highest occupied molecular orbital MO coefficients, and $c_{j\text{LUMO}}$ is the lowest unoccupied molecular orbital MO coefficients.

Maximum n-n repulsion for a C-N bond, Max n-n Re.-CN is a quantum mechanical energy-related descriptor. This energy describes the nuclear repulsion-driven processes in the molecule and may be related to the conformational (rotational, inversional) changes or atomic reactivity in the molecule [27, 28]. Maximum nuclear repulsion energy between two given atomic species (atoms A and B) in the molecule is calculated as follows:

$$E_{nn}(AB) = \frac{Z_A Z_B}{R_{AB}}, \tag{14}$$

where Z_A and Z_B are the nuclear (core) charges of atoms A and B, respectively, and R_{AB} is the distance between them.

3.3. Statistical Analysis. Codessa Pro was also used for statistical analysis. This code uses diverse statistical structure property/activity correlation techniques for the analysis of experimental data in combination with the calculated molecular descriptors. Heuristic and Best Multi-Linear Regression methods implemented in Codessa Pro were employed for selecting the "best" regression models [24].

In the study herein, the statistical quality of obtained QSPR models were assessed by the statistical parameters R^2, R^2_{cv}, F, S^2, $R^2_{\text{ext.}}$, RMSE, and the Y-randomization test. R^2, the squared correlation coefficient, is a measure of the fit of the regression equation. R^2_{cv}, the "leave one out" (LOO) cross-validated squared correlation coefficient, is an internal validation parameter for a model. The LOO approach involves developing a number of models with one sample

omitted at a time. F, the Fisher test value, reflects the ratio of the variance explained by the model and the variance due to the error in the model. Higher values of F-test indicate the significance of the equation. S^2 is the standard deviation of the regression. $R^2_{\text{ext.}}$ is the predicted square of the correlation coefficient for external validation, and is calculated from the test set by applying the equation developed on the training set. Prediction accuracy of the models is also given with root mean square error (RMSE) values. Reliability of the models is indicated by the Y-randomization test.

3.4. Data Set. The experimental acidic dissociation constants p*Ka* data of some N-base ligands and their formation constant, *Kc* with Bis(2,2′-bipyridine)platinum(II)$^{2+}$ ion were taken from [19]. The data set consists of 22 N-base ligands. For p*Ka* QSPR modeling, 5 of the ligands were selected as a test set. For *Kc* QSPR modeling, 4 compounds were selected as a test set. In the *Kc* QSPR modeling, compound 11 was selected as an outlier due to the fact that its reactivity is several magnitudes higher than the mean value of the set. While selection of training and test set of the compounds, attention was paid to all of the sets for spanning structural and activity diversity for compounds.

4. Conclusions

In the present study herein, the quantum chemical structural descriptors of 22 N-base ligands including pyridines, pyrimidines, purines, and quinolines have been correlated with their experimental p*Ka* using Codessa Pro methodology. In addition, the formation constant (*Kc*) of reactions of these ligands with Pt(II)$(\text{bpy})_2^{2+}$ (bpy = 2,2′-bipyridine) ion has also been modeled using the methodology. We have introduced three models each for p*Ka* and *Kc*. p*Ka* models demonstrated relatively better statistics than the *Kc* models.

The best obtained p*Ka* Model I is the four-parametric regression equation displaying very good statistical fit and predictive power as evident from its $R^2 = 0.95$, $F = 61.79$, $s2 = 0.311$, and $R^2_{\text{cv}} = 0.89$, $R^2_{\text{ext.}} = 0.97$, RMSE = 0.51 values. An analysis of descriptors that are involved in the p*Ka* models indicate that reactivity index and charge distribution-related descriptors play major roles in the model acid dissociation constant of N-base ligands.

Obtained *Kc* models exhibit interesting results in terms of validations parameters. If one analyses these models, and conclusion could be drawn that there is not any relationship between R^2_{cv} and $R^2_{\text{ext.}}$. An analysis of descriptors that are involved in the *Kc* models indicate that polarity and internal entropy considered with the Nitrogen-related index of N base ligands have dominant effects on the formation constant of reactions of these ligands with the Pt(II)$(\text{bpy})_2^{2+}$ (bpy = 2,2′-bipyridine) ion.

Acknowledgment

E. Eroğlu thankfully acknowledges the BAP Unit of Akdeniz University for the partial financial support.

References

[1] W. J. Lyman, W. F. Reehl, and D. H. Rosenblatt, *Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Compounds*, American Chemical Society, Washington, DC, USA, 1992.

[2] E. Demchuk and R. C. Wade, "Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins," *Journal of Physical Chemistry*, vol. 100, no. 43, pp. 17373–17387, 1996.

[3] P. J. Martel, C. M. Soares, A. M. Baptista et al., "Comparative redox and p*Ka* calculations on cytochrome c3 from several Desulfovibrio species using continuum electrostatic methods," *Journal of Biological Inorganic Chemistry*, vol. 4, no. 1, pp. 73–86, 1999.

[4] M. Schaefer, M. Sommer, and M. Karplus, "pH-dependence of protein stability: absolute electrostatic free energy differences between conformations," *Journal of Physical Chemistry B*, vol. 101, no. 9, pp. 1663–1683, 1997.

[5] G. Schüürmann, M. Cossi, V. Barone, and J. Tomasi, "Prediction of the pK_a of carboxylic acids using the *ab initio* continuum-solvation model PCM-UAHF," *Journal of Physical Chemistry A*, vol. 102, no. 33, pp. 6706–6712, 1998.

[6] C. O. Silva, E. C. Da Silva, and M. A. C. Nascimento, "Ab initio calculations of absolute p*Ka* values in aqueous solution II. Aliphatic alcohol, thiols, and halogenated carboxylic acids," *Journal of Physical Chemistry A*, vol. 104, no. 11, pp. 2402–2409, 2000.

[7] H. Li, A. W. Hains, J. E. Everts, A. D. Robertson, and J. H. Jensen, "The prediction of protein pK_a's using QM/MM: the pK_a of lysine 55 in turkey ovomucoid third domain," *Journal of Physical Chemistry B*, vol. 106, no. 13, pp. 3486–3494, 2002.

[8] J. E. Davies, N. L. Doltsinis, A. J. Kirby, C. D. Roussev, and M. Sprik, "Estimating p*Ka* values for pentaoxyphosphoranes," *Journal of the American Chemical Society*, vol. 124, no. 23, pp. 6594–6599, 2002.

[9] E. Soriano, S. Cerdán, and P. Ballesteros, "Computational determination of p*Ka* values. A comparison of different theoretical approaches and a novel procedure," *Journal of Molecular Structure*, vol. 684, no. 1–3, pp. 121–128, 2004.

[10] P. R. Duchowicz and E. A. Castro, "QSPR study of the acidity of carbon acids in aqueous solutions," *Mendeleev Communications*, vol. 12, no. 5, pp. 187–189, 2002.

[11] J. Ghasemi, S. Saaidpour, and S. D. Brown, "QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis," *Journal of Molecular Structure*, vol. 805, no. 1–3, pp. 27–32, 2007.

[12] M. Hennemann and T. Clark, "A QSPR-approach to the estimation of the pKHB of six-membered nitrogen-heterocycles using quantum mechanically derived descriptors," *Journal of Molecular Modeling*, vol. 8, no. 4, pp. 95–101, 2002.

[13] M. Goodarzi, M. P. Freitas, C. H. Wu, and P. R. Duchowicz, "pK_a modeling and prediction of a series of pH indicators through genetic algorithm-least square support vector regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 101, no. 2, pp. 102–109, 2010.

[14] M. J. Citra, "Estimating the pK_a of phenols, carboxylic acids and alcohols from semi-empirical quantum chemical methods," *Chemosphere*, vol. 38, no. 1, pp. 191–206, 1999.

[15] N. Bodor and M. J. Huang, "An extended version of a novel method for the estimation of partition coefficients," *Journal of Pharmaceutical Sciences*, vol. 81, no. 3, pp. 272–281, 1992.

[16] B. Rosenberg, L. Van Camp, and T. Krigas, "Inhibition of cell division in *Escherichia coli* by electrolysis products from a platinum electrode," *Nature*, vol. 205, no. 4972, pp. 698–699, 1965.

[17] G. Momekov and D. Momekova, "Recent developments in antitumor platinum coordination compounds," *Expert Opinion on Therapeutic Patents*, vol. 16, no. 10, pp. 1383–1403, 2006.

[18] CODESSA PRO (Comprehensive Descriptors for Structural and Statistical Analysis), Semichem, 7204, Mullen, Shawnee, KS, 66216 USA, Copyright, Semichem and the University of Florida, 1995 –2004.

[19] Y. Kawanishi, T. Funaki, T. Yatabe et al., "Spectral evidence and DFT calculations on the formation of bis(2,2′-bipyridine)platinum(II)-N-base adducts," *Inorganic Chemistry*, vol. 47, no. 9, pp. 3477–3479, 2008.

[20] AMPAC 9, 1992 –2008 Semichem, Inc. 12456 W 62nd Terrace—Suite D, Shawnee, KS, 66216 USA.

[21] A. Golbraikh and A. Tropsha, "Beware of q^2!," *Journal of Molecular Graphics and Modelling*, vol. 20, no. 4, pp. 269–276, 2002.

[22] E. Clementi, *Computational Aspects of Large Chemical Systems*, Springer, New York, NY, USA, 1980.

[23] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, "AM1: a new general purpose quantum mechanical molecular model," *Journal of the American Chemical Society*, vol. 107, no. 13, pp. 3902–3909, 1985.

[24] http://www.codessa-pro.com/manuals/Manual-codessa_pro.doc#_Toc92790360.

[25] D. T. Stanton and P. C. Jurs, "Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies," *Analytical Chemistry*, vol. 62, no. 21, pp. 2323–2329, 1990.

[26] D. T. Stanton, L. M. Egolf, P. C. Jurs, and M. G. Hicks, "Computer-assisted prediction of normal boiling points of pyrans and pyrroles," *Journal of Chemical Information and Computer Science*, vol. 32, pp. 306–316, 1992.

[27] N. S. Zefirov, M. A. Kirpichenok, F. F. Izmailov, and M. I. Trofimov, "Scheme for the calculation of the electronegativities of atoms in a molecule in the framework of Sanderson's frinciple," *Doklady Akademii Nauk SSSR*, vol. 296, pp. 883–887, 1987.

[28] M. A. Kirpichenok and N. S. Zefirov, "Electronegativity and molecular geometry I. General basis of the developed approach and determination of the effect of closer electrostatic interactions on bond lengths in organic molecules," *Zhurnal Organicheskoi Khimii*, vol. 23, pp. 673–703, 1987.

[29] A. Cartier and J. L. Rivail, "Electronic descriptors in quantitative structure-activity relationships," *Chemometrics and Intelligent Laboratory Systems*, vol. 1, no. 4, pp. 335–347, 1987.

[30] H. Sklenar and J. Jager, "Molecular structure-biological activity relationships on the basis of quantum-chemical calculations," *International Journal of Quantum Chemistry*, vol. 16, pp. 467–484, 1979.

[31] A. C. Gaudio, A. Korolkovas, and Y. Takahata, "Quantitative structure-activity relationships for 1,4-dihydropyridine calcium channel antagonists (nifedipine analogues): a quantum chemical/classical approach," *Journal of Pharmaceutical Sciences*, vol. 83, no. 8, pp. 1110–1115, 1994.