

Prediction of Acidity Constants of Thiazolidine-4-Carboxylic Acid Derivatives by Qsar Method

Shailendra Kumar Lagarkha*, Shailja Sachan#, Anand Kumar Lakhera*

*Department of Chemistry, A.P.S. University, Rewa (M.P.) INDIA – 486003 #Department of Chemistry, M. S. Golvalkar College, Rewa (M.P.) INDIA – 486001

Submitted: 15-03-2022

Accepted: 28-03-2022

ABSTRACT: Modelling of the acidity constant of thiazolidine-4-carboxylic acid derivatives as a function of molecular structures was established by means of the partial least squares algorithm. The subset of descriptors, which resulted in a low prediction error, was selected by genetic algorithm. This model was applied for the prediction of the acidity constant of some thiazolidine-4-carboxylic acid derivatives, which were not in the modelling procedure. Results of statistical analysis found with value of Variance as 0.8033, Cross validated regression coefficient and Fisher- value as 0.7552 and 13.885 respectively which may be useful for (medicinal) chemists in selecting the most suitable substituent for the development of more potent, effective and selective Thiazolidine-,4- carboxylic acid based acidity constant in future.

Key words: QSAR, Thiazolidine-4-carboxylic acid, PLS.

I. INTRODUCTION

Acidity constants can be a crucial parameter for understanding and quantifying chemical marvels, similar as response rates, natural exertion, natural uptake, natural transport, and environmental fate.⁽¹⁾ It has been shown that acid-base parcels affect the toxicity⁽²⁻³⁾, chromatographic retention gets, and pharmaceutical properties⁽⁴⁾ of organic acids and bases. Important of the theoretical foundation of ultramodern organic chemistry is grounded on the observation of the goods on acid- base equilibrium of changing molecular structure.⁽⁵⁾ Ali NIAZI, Turk J Chem. 30 (2006), 619 – 628.

A successful strategy for the vatic nation of the acidity constant is the construction of quantitative structure- exertion connections (QSARs).⁽⁶⁾ QSARs are fine equations relating chemical structure to a wide variety of physical, chemical, natural and technological parcels. QSAR models can be used to prognosticate parcels of composites as yet unmeasured or indeed unknown. Therefore, the QSAR approach saves coffers and expedites the process of development of new motes. $^{(7)}$

A major step in constructing QSAR models is chancing one or further molecular descriptors that represent variation in the structural property of the motes by a number. A wide variety of descriptors have been reported to be used in analysis.^(8–10) Recent progress OSAR in computational tackle and the development of effective algorithms have supported the routine development of molecular amount chemical computations. Quantum chemical computations are therefore an seductive source of new molecular descriptors, which can, in principle, express all of the electronic and geometric parcels of motes and their relations.⁽¹¹⁾ Infinitesimal charges, loftiest engaged molecular orbital (HOMO) and smallest unoccupied molecular orbital (LUMO) powers, molecular polarizability, dipole moments, and powers of patch are exemplifications of quantum chemical descriptors used in QSAR studies.

Multiple direct retrogression (MLR) is generally used in QSAR modelling.⁽¹²⁾ The colinearity problem of the MLR system has been overcome through the development of the partial least places (PLS) system, which plays an important part in QSAR analysis PLS is a factor analysis- grounded system that was firstly suggested and chemically applied by Wold et al.⁽¹⁴⁾ We've lately reported the operation of PLS modelling in spectrophotometric multivariate estimation.⁽¹⁵⁻²²⁾ PLS is used in confluence with optimization ways for point selection.⁽²³⁾ It has formerly been shown that inheritable algorithms (GAs) ⁽²⁴⁻³⁰⁾ can be successfully used as a point selection fashion.⁽³¹⁻³⁵⁾

II. PRESENTATION OF DATA

In present study table-1 represents the structure of thiazolidine-4-carboxylic acid derivatives, while table-2 shows the calculated Constitutional, topological & connectivity descriptors with acidity constants of thiazolidine-4carboxylic acid derivatives; table-3 represents the



correlation matrix between different constitutional, topological and connectivity descriptors.

Descriptor and acidity constants are given in table-2, table-3 and table-4 represents the residual report from best model of, topological and connectivity descriptors. Table-5 represent the Cross validation of best models. Ridge regression (fig-3) is representing the multicollinearity is not present in this study.

III. RESULT AND DISCUSSION

Multiple linear regression analysis and other statistical analysis were carried out on all the 23 molecules. The outlier molecules were then removed to improve the equation's predictive power. Descriptors were selected for the final equation based on their correlation coefficients and those descriptors having intercorrelation coefficient below 0.7 were considered, to select the best equation. Cross validation by leave one out method was carried out on these final set of 23 molecules to further enhance and validate the predictive power of the equation. Acceptability of the regression equation was judged by examining the statistical parameters.

The activity data pKa representing the concentration of compounds that inhibited the visible growth in various bacterial species is used as dependent variable to get a linear relationship in the QSAR models. These were correlated with molecular descriptors different like the descriptor; MW, topological constitutional descriptors; χt , Ram, Pol, MSD, ZM¹, ZM¹V and connectivity descriptor; χ^0 , χ^1 , χ^2 . The values of the selected descriptors used in the regression analysis are presented in Table-2 were calculated for the lowest energy conformers of the compounds in the series of software E-Dragon developed by VCC lab.

The essential feature of multiple regression analysis is cross-validation which asses the productivity of the computed model. Crossvalidation provides the values of PRESS, SSY, S_{PRESS} and R^2cv from which we can investigate the predictive power of the proposed model.

In order to determine the correlation between the observed biological activity, in terms pKa of the reported compounds and their structural parameters, QSAR investigation has been carried out model proposed by Hansch et. al. By using the data of table-3, correlation matrixes as well as the collinearity among the descriptors were calculated. A high inter relationship is observed between **Ram**, **MSD**, χ^2 , **ZM¹V and MW** (R² = 0.8033), while low inter relationship is observed $\mathbf{ZM}^{1}\mathbf{V}$ ($\mathbf{R}^{2} = 0.6152$). The results table-3 shows that some of the descriptors are mutually correlated. Thus, if a combination of them is present in the regression expression, them the model may suffer from a defect due to collinearity.

pKa = 7.3619, -0.0057(±0.0010) **ZM¹V**1

N=23, MSE= 7.1396, R^2 = 0.6152, AR^2 = 0.5969, Q-VALUE= 0.1098

 $pKa = 6.4486, -0.2616(\pm 0.1724) \quad \chi 0, -0.0513(\pm 0.0208) \text{ ZM}^1 \dots 2$

N=23, MSE= 7.0384, R^2 = 0.6387, AR^2 = 0.6026, Q-VALUE= 0.1135

pKa = 6.8545, 0.4603(\pm 0.2273)Ram, 0.7615(\pm 0.3022)MSD, -1.0500(\pm 0.3433) χ^2 3

N=23, MSE= 6.2339, R^2 = 0.696, AR^2 = 0.648, Q-VALUE= 0.1338

N=23, MSE = 5.1131, $R^2 = 0.7638$, $AR^2 = 0.7113$, Q-VALUE = 0.1709

The developed QSAR model eq. 4, demonstrated the importance of topological and connectivity descriptors which are used in the modeling especially topological. **Ram** coefficient is negative indicates that as their values increases the biological activity decreases, positive coefficient of χ^2 is directly proportional to activity. The correlation coefficient between the descriptors and activity is $R^2 = 0.7638$, which is quite good with the variance of 76.38% with the smallest standard error of estimation.

N=23, MSE = 4.5082, $R^2 = 0.8033$, $AR^2 = 0.7454$, Q-VALUE = 0.1988

The QSAR model described by eq. 5, demonstrated the importance of constitutional, connectivity and topological indices in describing the biological activity. The positive correlation is shown by MSD, χ^2 , ZM¹V biological activity reveals that increase in value of topological descriptors decrease in biological activity. While negative coefficient is shown by Ram and MW with biological activity reveals that increase in value of constitutional descriptor MW will lead in activity.

The correlation coefficient between the descriptors and activity is $R^2 = 0.8033$, which is quite good with the variance of 80.33% with the



smallest standard error of estimation. The PRESS/SSY value is 0.2448 indicates that the developed model is reasonable model to explain the biological activity.

The ratio PRESS/SSY can be used to calculate approximate confidence intervals of prediction of new compounds. To be a reasonable QSAR model PRESS/SSY should be smaller 0.4 and the value of this ratio smaller than 0.1 indicates an excellent model. Also if PRESS VALUE is transformed in a dimension less term by relating it to the initial sum of squares, be obtain $R^2_{CV}(Q^2)$ i.e. the complement to the trace on of unexplained variance over the total variance. Thus, PRESS and R^2_{CV} have good properties. However, for practical purposes of end users the use of square-root of PRESS/N, which is called PSE (predictive square error), is directly related to the uncertainty of the prediction. This parameter, namely PSE is much

more useful when S_{PRESS} (uncertainty of prediction) comes out to be the same as MSE (mean square error). All cross-validated parameters given in Table-5 are in accordance with the aforementioned findings.

The validity of the models has been tested using cross-validation method and the tetra parametric model discussed above has been found to be the best. The R^2_{CV} comes out to be 0.7552 also the lowest value of $S_{press.}$ A comparison between observed and estimated activity has also been demonstrated in figure-1.

According the result of biological screening summary of biological activity analogues a graph is plotted between observed and predicted pKa (Fig-1). Further a bar graph is also obtained to show the reliability of selected model between observed biological activities (Fig-2).



Table-1: The structures of compounds studied and their pKa activity



Com.no.	рКа	χt	Ram	Pol	MSD	χ^0	χ^1	χ^2	ZM^1	ZM ¹ V	MW
1	6.19	0.449	2	6	2.435	5.983	3.805	3.289	36	197	126.12
2	6.17	0.431	3	8	2.682	6.853	4.198	3.922	42	213	138.13
3	5.86	0.42	4	10	2.813	7.776	4.512	4.911	50	229	150.14
4	5.73	0.397	4	13	3.066	8.483	5.072	4.911	54	245	162.15
5	6.12	0.385	3	11	3.414	8.268	5.236	4.472	50	245	162.15
6	5.86	0.393	4	12	3.211	8.431	5.109	4.821	52	285	170.13
7	6.08	0.366	3	12	3.834	8.975	5.736	4.825	54	261	174.16
8	6.1	0.374	4	12	3.641	9.138	5.592	5.313	56	261	174.16
9	5.94	0.336	3	14	4.71	10.389	6.736	5.533	62	293	198.18
10	5.31	0.317	4	17	3.898	9.966	6.771	5.975	70	293	198.18
11	5.5	0.311	5	19	4.194	10.836	7.165	6.597	76	309	210.19
12	5.67	0.311	5	20	3.965	10.836	7.182	6.503	76	329	214.18
13	5.51	0.311	5	19	4.194	10.836	7.165	6.597	76	329	214.18
14	5.35	0.297	4	18	4.856	11.38	7.754	6.765	78	325	222.2
15	5.8	0.301	5	21	4.537	11.544	7.703	6.766	80	345	226.19
16	4.95	0.311	5	20	3.965	10.836	7.182	6.503	76	342	233.63
17	5.24	0.311	5	19	4.194	10.836	7.165	6.597	76	342	233.63
18	5.83	0.296	6	23	4.772	12.414	8.075	7.496	86	350	236.21
19	5.01	0.296	6	23	4.772	12.414	8.075	7.496	86	381	242.19
20	4.7	0.296	6	23	4.494	12.414	8.075	7.508	86	390	244.19
21	5.39	0.296	6	25	4.317	12.414	8.13	7.199	86	381	242.19
22	5.53	0.305	6	22	4.117	11.707	7.575	7.137	82	378	294.08
23	5.17	0.287	6	26	4.598	13.121	8.651	7.497	90	375	250.22

Table-2: Calculated Constitutional, topological & connectivity descriptors and pKa activity of Compound

Detailed Name of Descriptors:-

S. No.	Name of Descriptors	Detailed Name of Descriptors					
1	Xt	Total structure connectivity index					
2	Ram	Ramification index					
3	Pol	Polarity number					
4	MSD	Mean square distance index					
5	χ^0	Connectivity index of order 0					
6	χ^1	Connectivity index of order 1					
7	χ^2	Connectivity index of order 2					
8	ZM^1	First Zagreb index					
9	$ZM^{1}V$	First Zagreb index by valence vertex					
		degrees					
10	MW	Molecular weight					

Table-3: Correlation Matrix of different descriptors

	pK a	Xt	Ram	P 0 1	MSD	χ ⁰	χ ¹	χ^2	$\mathbf{Z}\mathbf{M}^{1}$	$\mathbf{Z}\mathbf{M}^{1}\mathbf{V}$	MW
рКа	1	0.730 3	- 0.7256	- 0 7 6 0 8	- 0.5843	- 0.7271	- 0.7328	- 0.7703	- 0.7727	- 0.7843	- 0.7227
Xt		1	-	-	-	-	-	-	-	-	-

| Impact Factor value 7.429 | ISO 9001: 2008 Certified Journal Page 634



1	I	i	0 00 15		0.0000	0.0672		0.0565	0.0706	0.0221	0.0107
			0.8045	0	0.9333	0.9673	0.9858	0.9565	0.9706	0.9331	0.9187
				•							
				9							
				4							
				2							
				6							
				0							
				•							
Ram			1	9	0.6597	0.8806	0.8411	0.9266	0.9106	0.9095	0.8732
				3							
				1							
				1	0.0001		0.0.00	0.0777	0.0004	0.0.00	0.0004
Pol				I	0.8321	0.9757	0.969	0.9755	0.9884	0.9692	0.9334
MSD					1	0.9248	0.9366	0.8779	0.8803	0.8456	0.8218
χ°						1	0.9918	0.9825	0.9856	0.9646	0.9327
χ^1							1	0.973	0.9857	0.9556	0.9283
χ^2								1	0.9937	0.9626	0.9371
ZM^1									1	0.967	0.9404
ZM^1										1	0.0616
V										1	0.9616
MW											1

C. No.	Obs. pKa	Est. pKa	Residual
1	6.033	6.19	0.157
2	6.196	6.17	-0.026
3	5.846	5.86	0.014
4	6.006	5.73	-0.276
5	6.147	6.12	-0.027
6	5.829	5.86	0.031
7	6.092	6.08	-0.012
8	6.06	6.1	0.04
9	6.019	5.94	-0.079
10	5.39	5.31	-0.08
11	5.616	5.5	-0.116
12	5.267	5.67	0.403
13	5.403	5.51	0.107
14	5.312	5.35	0.038
15	5.471	5.8	0.329
16	5.255	4.95	-0.305
17	5.391	5.24	-0.151
18	5.634	5.83	0.196
19	5.301	5.01	-0.291
20	4.902	4.7	-0.202
21	5.152	5.39	0.238
22	5.439	5.53	0.091
23	5.251	5.17	-0.081

Table-4: Residual Report









Figure-2: Graph plotted between the residual and observed activity



TABLE – 5 – Result of Cross Validation

Model No	Ν	PRESS	SSY	PRESS/SSY	R ²	R^2_{CV}	PSE	Sncess
1	23	1.4993	2.3968	0.6255	0.6152	0.3745	0.0532	0.2670
2	23	1.4076	2.4884	0.5656	0.6387	0.4344	0.0515	0.2651
3	23	1.1844	2.7117	0.4367	0.6960	0.5633	0.0473	0.2495
4	23	0.9203	2.9758	0.3092	0.7638	0.6908	0.0417	0.2260
5	23	0.7663	3.1297	0.2448	0.8033	0.7552	0.0380	0.2121

 $\begin{aligned} \text{PRESS} &= \sum (Y_{obs} - Y_{calc})^2 \\ \text{SSY} &= \sum (Y_{obs} - Y_{mean})^2 \\ \text{S}_{press} &= \left[\text{ press } / \left(\text{n-k-1} \right) \right]^{1/2} \end{aligned}$

$$PSE = \sqrt{press}/n$$
$$R^{2}_{cv} = 1 - \frac{PRESS}{SSY}$$

| Impact Factor value 7.429 | ISO 9001: 2008 Certified Journal Page 636



Volume 7, Issue 2 Mar-Apr 2022, pp: 631-637 www.ijprajournal.com ISSN: 2456-4494

Where, Y_{obs} , Y_{calc} and Y_{mean} are observed, calculated and mean values; n is number of the compounds, k is number of parameters.

IV. CONCLUSION

The following conclusions are obtained from this analysis:

- (1) Topological, Constitutional indices & Connectivity parameters may be used for modeling of these compounds.
- (2) Connectivity indices parameters are more effective in this QSAR study.

(3) χt , **Ram**, **Pol**, **MSD**, χ^0 , χ^1 , χ^2 , **ZM**¹, **ZM**¹V and **MW** parameters is useful for this study.

(4) The highest value $R^2 = 0.8033$ are obtained in QSAR models.

Acknowledgements

We are thankful to Dr. V. K. Agrawal professor, Dept. of Chemistry, A. P. S. University Rewa for giving their valuable suggestions to complete this study.

REFERENCE

- [1]. D. Kara and M. Alkan, Spectrochim. Acta Part A, 56, 2753 (2000).
- [2]. Y.H. Zhao, L.H. Yuan and L.S. Wang, Bull. Environ. Contam. Toxicol. 57, 242 (1996).
- [3]. P. Alines, J. Planar Chromatogr. Mod. TLC, 9, 52 (1996).
- [4]. G.H. Rochester, Acidity Functions, Academic Press, New York, 1971.
- [5]. L.P. Hammet, **Physical Organic Chemistry**, McGraw-Hill, New York, 1940.
- [6]. V. Tantishaiyakul, J. Pharm. Biomed. Anal. 37, 411 (2005).
- [7]. A.R. Katrizky, R. Petrukhin and D. Tatham, J. Chem. Inf. Comput. Sci. 41, 679 (2001).
- [8]. R. Todeschini and V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, (Germany), 2000.
- [9]. V. Consonni, R. Todeschinni, M. Pavan and P. Gramatica, J. Chem. Inf. Comput. Sci. 42, 693 (2002).
- [10]. G. Krenke, E.A. Castro and A.A. Toropov, J. Mol. Struct. (Theochem) 542, 107 (2001).
- [11]. R. Carbo-Dorca, L. Amat, E. Besalu, X. Girone's and D. Robert, J. Mol. Struct. (Theochem) 504, 181(2001).
- [12]. D.C. Montgomery and E.A. Peck, JohnW iley, New York, 1982.
- [13]. H.J.M. Verhaar, E.V. Ramos and J.L.M. Hermens, J. Chemometr. 10, 149 (1996).

- [14]. K.G. Joreskog and H. Wold, North Holland, Amsterdam, 1982.
- [15]. J. Ghasemi and A. Niazi, Microchem. J. 68, 1 (2001).
- [16]. J. Ghasemi, A. Niazi and A. Safavi, Anal. Lett. 34, 1389 (2001).
- [17]. J. Ghasemi, R. Amini and A. Niazi, Anal. Lett. 35, 533 (2002).
- [18]. J. Ghasemi and A. Niazi, Talanta 65, 1168 (2005).
- [19]. J. Ghasemi and A. Niazi, Anal. Chim. Acta 533, 169 (2005).
- [20]. A. Niazi, J. Ghasemi and A. Yazdanipour, Anal. Lett. 38, 2377 (2005).
- [21]. J. Ghasemi, A. Nikrahi and A. Niazi, Turk. J. Chem. 29, 669 (2005).
- [22]. A. Niazi and M. Sadeghi, Chem. Pharm. Bull. 54, (2006).
- [23]. F. Ros, M. Pintore and J.R. Chretien, Chemometr. Intell. Lab. Syst. 63, 15 (2002).
- [24]. M.J. Arcos, M.C. Ortiz and L.A. Sarbia, Anal. Chim. Acta 339, 63 (1997).
- [25]. U. Depczynski, V.J. Frost and K. Molt, Anal. Chim. Acta 420, 217 (2000). 627
- [26]. C.B. Lucasius and G. Kateman, Chemometr. Intell. Lab. Syst. 19, 1 (1993).
- [27]. D.B. Hibbert, Chemometr. Intell. Lab. Syst. 19, 227 (1993).
- [28]. R. Leardi, R. Boggia and M. Terrile, J. Chemometr. 6, 267 (1992).
- [29]. R. Leardi, J. Chemometr. 8, 65 (1994).
- [30]. R. Leardi, J. Devillers (ed.), Academic Press, London, 1996.
- [31]. R. Leardi and A.L. Gonzalez, Chemomet. Intell. Lab. Syst. 41, 195 (1998).
- [32]. R. Leardi, J. Chemometr. 14, 643 (2000).
- [33]. R. Leardi, J. Chemometr. 15, 559 (2001).
- [34]. J. Ghasemi, A. Niazi and R. Leardi, Talanta 59, 311 (2003).
- [35]. J. Ghasemi, D.M. Ebrahimi, R. Leardi and A. Niazi, J. Anal. Chem. 61, 92 (2006).