ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES Knowledge to Innovation

Vol 6, Suppl 5, 2013

ISSN - 0974-2441

Research Article

A QSAR STUDY ON THE SCHIFF BASES OF 2, 4, 6-TRICHLOROPHENYLHYDRAZINE USING FREELY AVAILABLE ONLINE 2D DESCRIPTORS

SUPRATIM RAY

Department of Pharmaceutical Sciences Assam University: Silchar 788011, Assam, India. Email: supratimray75@gmail.com

Received: 3 September 2013, Revised and Accepted: 25 September 2013

ABSTRACT

Objective: This study gives a quantitative structure activity relationship (QSAR) correlation of the thirty schiff bases of 2, 4, 6-Trichlorophenylhydrazine reported by Khan et al as DPPH radical scavengers.

Method: Only 2D descriptors available on freely available PaDEL were considered for the present study. Stepwise regression was used as chemometric tool. The developed model was rigorously validated using several validation tools.

Results and Conclusion: The model indicates the importance of count of E-States for (strong) hydrogen bond donors, sum of E-State descriptors of strength for potential hydrogen bonds of path length 4 and count of E-State descriptors of strength for potential hydrogen bonds of path length 9 necessary for DPPH radical scavenging activity.

Keywords: QSAR, schiff base, stepwise regression, validation,

INTRODUCTION

Schiff base-derived antioxidants have gained much attention for their capacity in scavenging free radicals. 2-oxo-quinoline-3-carbaldehyde Schiff-base derivatives showed better activity when compared with commercial antioxidants ascorbic acid, BHT and BHA, employing several assay techniques such as DPPH assay, ABTS assay etc [1]. The schiff bases of thymol and carvacrol exhibited much better antioxidant activity than thymol and carvacrol in DPPH assay [2]. The schiff bases of 4-amino-1, 5-dimethyl-2-phenylpyrazole-3-one derivatives also showed good activity for inhibition of nitric oxide [3].

The search for Schiff base-derived antioxidants has received much attention and effort in order to identify the compounds having high capacity in scavenging free radicals related to various disorders and diseases associated with oxidative damage, caused by reactive oxygen species (ROS). Presently, synthetic antioxidants are widely used because they are effective and cheaper than natural antioxidants. Currently a number of schiff-base metal complexes have been investigated as effective scavengers of ROS, acting as antioxidants [4].

Hydrazones represent a special group of compounds in the Schiff base family. Most of the hydrazones show biological activities, and therefore, these compounds are potentially used in the treatment of diseases like tuberculosis, mental disorder, antitumor and leprosy [5]. Based on the findings a QSAR study has been performed on the Schiff bases of 2, 4, 6-Trichlorophenylhydrazine reported by Khan et al [6] as DPPH Radical scavengers to find out the structural requirement of the compounds for activity.

MATERIALS & METHODS

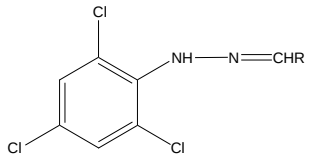
The *in vitro* DPPH radical scavenging activity (IC_{50}) of thirty schiff base derivatives were in µM range which was converted to mM range and then to logarithmic scale [log (10^3 / IC_{50})] were used as response variable (pC) for subsequent QSAR analyses (Table 1).

Table 1: Molecular scaffolds of the compounds along with their activity

Sl No	R	DPPH radical scavenging activity (µM)[C]	pC=log (1000/C)
1	3,4,5-Trimethoxy benzene	255.4	0.592779
2	3-Methoxy-2-hydroxy benzene	7.21	2.142065
3	Benzene	185.25	0.732242
4	2-Fluoro benzene	231.58	0.635299
5	4-Methylsulfanyl benzene	95.2	1.021363
6	4-Dimethylamino benzene	115.54	0.937268
7	2-Hydroxy-5-methyl benzene	24.42	1.612254
8	2,5-Dihydroxy benzene	5.85	2.232844
9	2,4-Dichloro benzene	136.26	0.865632
10	4-Chloro benzene	353.9	0.451119
11	2-Chloro benzene	295.85	0.528928
12	2-Hydroxy 3,5-dichloro benzene	6.32	2.199283
13	3,4-Dichloro benzene	240.39	0.619084
14	3,4-Dihydroxy benzene	4.49	2.347754
15	4-Hydroxy benzene	6.3	2.200659
16	3,4-Dimethoxy benzene	364.85	0.437886
17	2,3,4-Trihydroxy benzene	4.05	2.392545
18	2,3-Dihydroxy benzene	4.41	2.355561
19	3-Thiophene	291.43	0.535466
20	4-Pyridine	125.27	0.902153
21	2-Bromo benzene	329.94	0.481565
22	2-Hydroxy benzene	30.25	1.519275
23	3-Chloro benzene	278.73	0.554816
24	1-Phenanthrene	324.65	0.488585
25	3-Pyridine	251.51	0.599445
26	2-Naphthalene	330.66	0.480618
27	2-Methyl benzene	253.8	0.595508
28	2,4,6-Trihydroxy benzene	4.23	2.37366
29	2,4-Dihydroxy benzene	20.09	1.69702
30	4-Methyl benzene	369.3	0.432621

Descriptors

The structures of thirty compounds were sketched using Chem Draw Ultra version 6.0 [7] and saved in mol. format which is one of the suitable input formats for PaDEL. The energies of structural configuration were minimized by AM-1 method using Chem 3D Ultra version 6.0 and used as input structure for descriptor calculations. Only 2D descriptors available on freely available PaDEL were considered for the present study [8]. Initially 256 descriptors were calculated using PaDEL software version 2.12. Then descriptors having value of zero and constant value were deleted. Finally pruned 114 descriptors were chosen for QSAR analysis of selected data set.



Model development

For the development of model, the whole data set (n=30) was divided into training (n=23, 75% of the total number of compounds) and test (n=7, 25% of the total number of compounds) sets by *k*-means clustering technique applied on standardized descriptor matrix. The QSAR model was developed using the training set compounds (optimized by Q^2), and then the developed models were validated (externally) using the test set compounds. Stepwise regression was used as chemometric tool [9]. The stepping criterion was based on F value (F = 4.0 for inclusion; F = 3.9 for exclusion). MINITAB version 14 software [10] was used for stepwise regression method. *K*-means clustering, standardization of the variables was performed in SPSS version 9.0 software [11]. STAISTICA version 7 software [12] was used for the determination of the LOO (leave-one-out) values of the training set compounds.

Model validation

The statistical qualities of developed equation were judged by calculating several metrics namely determination coefficient (R^2) as a measure of the total variance of the response explained by the regression models (fitting), explained variance (R_a^2) and variance ratio (F) at specified degrees of freedom (df) [13].

Both internal and external validations are performed to assess to reliability and the predictive potential of the developed model. To determine the predictive quality of the models, models are required to be further validated using different validation techniques: (a) internal validation or cross-validation using the training set compounds, (b) external validation using the test set compounds

Internal validation

The internal validation of generated model was performed by the leave-one out procedure (Q^2_{int}) [14]. It can be expressed as follows:

The internal validation of generated model was performed by the leave-one out procedure (Q^2_{int}) [14]. It can be expressed as follows:

$$Q^2_{\text{int}} = 1 - \frac{\sum \left(Y_{obs} - Y_{cal}\right)^2}{\sum \left(Y_{obs} - \bar{Y}_{training}\right)^2} \qquad \textbf{(i)}$$

Where Y_{obs} and Y_{cal} indicate observed and calculated activity of training set compounds. $\bar{Y}_{training}$ indicates mean of activity of training set respectively

External validation

The developed models were judged by different external validation parameters like $Q^2_{ext(F1)}$, $Q^2_{ext(F2)}$ [15, 16], $Q^2_{ext(F3)}$ **[17]. They are defined as follows:**

$$Q^2_{ext(F1)} = 1 - \frac{\sum \left(Y_{obs(test)} - Y_{cal(test)}\right)^2}{\sum \left(Y_{obs(test)} - \bar{Y}_{training}\right)^2} \qquad \textbf{(ii)}$$

$$Q^2_{ext(F2)} = 1 - \frac{\sum \left(Y_{obs(test)} - Y_{cal(test)}\right)^2}{\sum \left(Y_{obs(test)} - \bar{Y}_{test}\right)^2} \qquad \textbf{(iii)}$$

$$Q^2_{ext(F3)} = 1 - \frac{\left[\sum \left(Y_{obs(test)} - Y_{cal(test)}\right)^2\right] / n_{test}}{\left[\sum \left(Y_{obs(test)} - \bar{Y}_{training}\right)^2\right] / n_{training}}$$

(iv)

Where $Y_{obs(test)}$ **and** $Y_{cal(test)}$ **indicate observed and** calculated activity of test set compounds. $\bar{Y}_{training}$ and $\bar{Y}_{test}$ indicate mean of activity of training and test set respectively. $n_{training}$ and n_{test} are the number of compounds in training and test set respectively.

Further test on external validation

As external validation is the optimum tool for establishing the predictive QSPR models, so beside the above parameters two more external validation parameters were also employed to check the predictive ability of the developed models.

The r_m^2 matrices ($\overline{r_m^2}$ and Δr_m^2) are employed to indicate better both the internal and external predictive capacities of a model and to ascertain the proximity in the values of the predicted and observed response data [18, 19]. They are calculated as follows:

$$\overline{r_m^2} = (r_m^2 + r'^2_m)/2 \qquad \textbf{(v)}$$

$$\Delta r_m^2 = \left|(r_m^2 - r'^2_m)\right| \qquad \textbf{(vi)}$$

Where $r_m^2 = r^2 * (1 - \sqrt{r^2 - r_0^2})$ **and** $r'^2_m = r^2 * (1 - \sqrt{r^2 - r'^2_0})$

Squared correlation coefficient values between the observed and predicted values of the test set compounds (leave-one out predicted values for training set compounds) with intercept (r^2) and without intercept (r_0^2) were calculated for determination of r_m^2 Change of the axes gives the value of $r'_0{}^2$ and the r'^2_m metric is calculated based on the value of $r'_0{}^2$. The $\overline{r_m^2}$ and Δr_m^2 matrices are applied for internal validation of training set compounds ($\overline{r^2_{m(LOO)}}$ as well as

$\Delta r^2_{m(LOO)}$, external validation of test set compounds ($\overline{r^2_{m(test)}}$ as well as $\Delta r^2_{m(test)}$) and overall validation for all compounds ($\overline{r^2_{m(overall)}}$ $\Delta r^2_{m(overall)}$). Those with $\overline{r^2_m}$ values above the threshold of 0.5 and with a Δr^2_m value less than 0.2 are considered to be predictive and reliable ones.

RESULTS AND DISCUSSION

Membership of compounds in different clusters generated using *k*-means clustering technique is shown in Table 2. The test set size was set to approximately 25% to the total data set size [20] and the test set members along with their observed and calculated activity are given in Table 3. The result obtained from developed method is described below and the interpretations of the equations are also depicted. Using stepping criteria based on F value (F = 4.0 for inclusion; F = 3.9 for exclusion), the best equation is derived as follows:

Table 2: *k*-Means clustering of compounds using standardized descriptors

Cluster No.	No. of compounds in different clusters	Compounds (Sl nos.) in each clusters
1	**1**	**24**
2	**12**	**1 2 5 6 9 12 13 16 17 21 26 28**
3	**17**	**3 4 7 8 10 11 14 15 18 19 20 22 23 25 27 29 30**

Table 3: Observed and calculated DPPH radical scavenging activity from developed modelTable 3 near here

Sl. No.	Observed DPPH radical scavenging activity (pC)[a]	Calculated activity [b]
	Training Set	
1	0.592779	0.613115
2	2.142065	1.784956
3	0.732242	0.602496
4	0.635299	0.609878
6	0.937268	0.586885
8	2.232844	2.134527
9	0.865632	0.592339
10	0.451119	0.623902
11	0.528928	0.617977
12	2.199283	1.770215
15	2.200659	1.769861
16	0.437886	0.62491
17	2.392545	2.797975
18	2.355561	2.069919
19	0.535466	0.617479
21	0.481565	0.621584
22	1.519275	1.945408
23	0.554816	0.616006
24	0.488585	0.621049
25	0.599445	0.612608
28	2.37366	0.263655
29	1.69702	2.565203
30	0.432621	0.62531
	Test Set	
5	1.021363	0.61124
7	1.612254	1.85724
13	0.619084	0.61124
14	2.347754	2.16324
20	0.902153	0.61124
26	0.480618	0.61124
27	0.595508	0.61124

aObserved activity (ref. 6); b Calculated from eq. (1);

$$pC = -0.63476+1.246nHBd-0.069SHBint4-0.47nHBint9$$

$$R^2 = 0.903, R^2_a = 0.887, PRESS = 6.51, F = 58.93(df = 3,19),$$

$$Q^2_{\text{int}} = 0.533, n_{training}=23,\ n_{\text{test}} = 7, Q^2_{ext(F1)} = 0.87,\ Q^2_{ext(F2)}=0.86,\ Q^2_{ext(F3)}=0.57$$

(1)

Eq. (1) could explain 88.7% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 53.3%. The descriptor nHBd has positive coefficient of activity. The parameter indicates the importance of count of E-States for (strong) hydrogen bond donors. It is observed that molecule 17 and 28 contain four hydrogen bond donor atoms (three OH groups and one NH group) showing highest DPPH radical scavenging activity. But compounds contain three hydrogen bond donor atoms (like compound 29), two hydrogen bond donor atoms (like compound 7) as well as one hydrogen bond donor atom (like compounds 1 and 30) showing comparatively less activity than compounds contain four hydrogen bond donor atoms.

The parameter SHBint4 signifying sum of E-State descriptors of strength for potential hydrogen bonds of path length 4 has negative coefficient of activity. Compound like 21 showing higher value of SHBint4 possess comparatively lower activity.

The descriptor nHBint9 signifying count of E-State descriptors of strength for potential hydrogen bonds of path length 9 has negative coefficient of activity. Compound like 28 showing absence of E-State descriptors of strength for potential hydrogen bonds of path length 9 possesses comparative better antioxidant activity. From Table 4 it is seen that the r^2_m matrices of the developed model are above the threshold value except for the value of $\overline{r^2_{m(LOO)}}$, which is near to threshold value.

Table 4: Further test on external validation

$\overline{r^2_m}$			Δr^2_m		
Test	**Training**	**Overall**	**Test**	**Training**	**Overall**
0.82	0.47	0.51	0.1	0.02	0.01

OVERVIEW AND CONCLUSIONS

The whole dataset (n=30) was divided into a training set (23 compounds) and a test set (7 compounds) based on *k*-means clustering of the standardized descriptor matrix and model was developed from the training set. The predictive ability of the models was judged from the prediction of the activity of the test set compounds. Three different external validations tool like $Q^2_{ext(F1)}$ $Q^2_{ext(F2)}$, $Q^2_{ext(F3)}$ were used to check the predictive ability of the model. Finally the r^2_m matrices ($\overline{r^2_m}$ and Δr^2_m) are employed to indicate better both the internal and external predictive capacities of a model and to ascertain the proximity in the values of the predicted and observed response data. The model indicates the importance of count of E-States for (strong) hydrogen bond donors, sum of E-State descriptors of strength for potential hydrogen bonds of path length 4 and count of E-State descriptors of strength for potential hydrogen bonds of path length 9.

REFERENCES

1. Zhang Y, Fang Y, Liang H, Wang H, Hu K, Liu X, Yi X, Peng Y. Synthesis and antioxidant activities of 2-oxo-quinoline-3-carbaldehyde Schiff-base derivatives. Bioorg Med Chem Lett 2013; 23:107-111.
2. Benna Kumar D, Rawat DS. Synthesis and antioxidant activity of thymol and carvacrol based Schiff bases. Bioorg Med Chem Lett 2013; 23: 641-645.
3. Alam MS, Choi JH, Lee DU. Synthesis of novel Schiff base analogues of 4-amino-1, 5-dimethyl-2-phenylpyrazol-3-one and their evaluation for antioxidant and anti-inflammatory activity. Bioorg Med Chem 2012; 20: 4103-4108.
4. Kostova I, Saso L. Advances in Research of Schiff-base Metal Complexes as Potent Antioxidants. Curr Med Chem. 2013 Jun 25 (E Pub ahead of schedule).
5. Avaji PG, Kumar CHV, Patil SA, Shivananda KN, Nagaraju C. Synthesis, spectral characterization, *in vitro* microbiological evaluation and cytotoxic activities of novel macrocyclic *bis* hydrazone. Eur J Med Chem 2009; 44: 3552-3559.
6. Khan KM, Shaha Z, Ahmad VU, Khan M, Tahaa M, Rahima F, Alia S, Ambreena N, Perveend S, Choudharya MI, Voelter W. 2, 4, 6-Trichlorophenylhydrazine Schiff Bases as DPPH Radical and Superoxide Anion Scavengers. Med Chem 2012; 8: 452-461.
7. CS ChemOffice is software of Cambridge Soft Corporation, USA, www.cambridgesoft.com.
8. Yap CW. PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. J Comput Chem 2011; 32: 1466-1474.
9. Darlington RB. Regression and linear models. New York: McGraw Hill; 1990.
10. MINITAB version 14 is statistical software of Minitab Inc, USA, http://www.minitab.com.
11. SPSS version 9.0 software is statistical software of IBM Corporation, http://www-01.ibm.com/software/analytics/spss.
12. STATISTICA version 7 is statistical software of Stat Soft Inc, www.statsoft.com.
13. Snedecor GW, Cochran WG. Statistical Methods. New Delhi: Oxford & IBH Publishing Co Pvt Ltd; 1967.
14. Wold S, Eriksson L. Validation tools. In: Van de Waterbeemd H, editor. Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry). New York: Weinheim-VCH; 1995. p 312–317.
15. Hawkins DM. 2004. The problem of overfitting. J Chem Inf Comp Sci 2004; 44: 1-12.
16. Schuurmann G, Ebert RU Chen J, Wang B, Kuhne R. External validation and prediction employing the predictive squared correlation coefficient-test set activity mean vs training set activity mean. J Chem Inf Model 2008; 48: 2140-2145.
17. Consonni V, Ballabio D, Todeschini R. 2009. Comments on the definition of the Q^2 parameter for QSAR validation. J Chem Inf Model 2009; 49: 1669-1678.
18. Ojha PK, Mitra I, Das RN, Roy K. Further exploring r_m^2 metrics for validation of QSPR models dataset. Chemon Intell Lab Sys 2011; 107: 194–205.
19. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci 2008; 27: 302-313.
20. Roy PP, Leonard JK, Roy K. Exploring the impact of the size of training sets for the development of predictive QSAR models. Chemom Intell Lab Sys 2008; 90: 31-42.