Academic Sciences Asian Journal of Pharmaceutical and Clinical Research
Vol 5, Suppl 4, 2012 ISSN - 0974-2441

Research Article

A QSAR STUDY AT AM1 SEMI EMPIRICAL LEVEL OF 1, 3- DIARYL PYRAZOLE DERIVATIVES AS ANTITUMOR AGENTS AGAINST HUMAN DU145 PROSTRATE CANCER CELL LINE

SUPRATIM RAY*, CHARUSMRITI

Division of Pharmaceutical Chemistry Dr. B C Roy College of Pharmacy & Allied Health Sciences, Bidhannagar, Durgapur, 713 206, India;
Email: supratimray_in@yahoo.co.in

Received: 7 September 2012, Revised and Accepted: 16 October 2012

ABSTRACT

A QSAR study was performed by quantum chemical calculation only at the AM1 semi empirical levels to calculate the mulliken's charges and dipole moment of common atoms for twenty six 1, 3-diaryl pyrazole derivatives as antitumor agents against human DU 145 prostate cancer cells. Stepwise regression analysis is used as chemometric tool. The model indicates the importance of hydroxyl group at various position of the moiety. Presence of double bond attached pyrazole nucleus and the ester or acid group is beneficial for antitumor activity. The statistical quality of the model showed acceptable internal validation (Q^2_{int} =0.678 and $r^2_{m(LOO)}$ =0.666), external validation ($Q^2_{ext(F1)}$ =0.711, $Q^2_{ext(F2)}$ =0.693, $r^2_{m(test)}$ =0.637), and overall validation ($r^2_{\text{m(overall)}}$ =0.676). The study shows that mulliken's charges can be projected as a useful tool in the context of QSAR studies.

Keywords: QSAR, Mulliken's charge, Stepwise regression, Pyrazole, Cancer.

INTRODUCTION

Cancer is one of the deadliest diseases in human history. Despite significant advances, the cure rates for various malignancies are limited. So much more attention was given for the discovery and development of more potent anticancer agents. It was observed that incorporation of heterocyclic residues into perspective pharmaceutical lead candidates constitutes an important strategy which provides activity and safety features [1, 2]. Among several heterocyclic moiety pyrazole derivatives exhibit a wide range of biological properties including promising antitumor activity. It was reported that 4-arylmethyl-1-phenylpyrazole and 4-aryloxy-1-phenylpyrazole derivatives act as novel androgen receptor antagonists [3]. A study shows that 3-(1H-indol-3-yl)-1H-pyrazole-5-carbohydrazide derivatives have activity against human HepG-2, BGC823 and BT474 cell lines [4]. PC et al reported that pyrazole derivatives containing thiourea skeleton as anticancer agents [5]. Vujasinovic et al reported anticancer activity and QSAR analysis of some novel pyrazole derivatives [6]. The activity of 3, 5-diaryl-4, 5-dihydropyrazole derivatives against EAC cell line and CoMFA analysis was also reported [7]. Some novel 4, 5-dihydro-1H-pyrazole niacinamide showed promising activity against BRAF (V600E) and WM266.4 human melanoma cell line [8]. Considering the above findings, we have performed a QSAR study at AM1 semi empirical levels for twenty six 1, 3-diaryl pyrazole derivatives as antitumor agents against human DU 145 prostate cancer cell lines reported by Christodoulou et al [9]

MATERIALS & METHODS

The Data-set and descriptors

The antitumor activities of substituted pyrazole derivatives against human DU 145 prostate cancer cells was reported by Christodoulou et al [9] were used as the model data-set for the present QSAR analysis (**Table 1 & 2**). The reported cell viability (C) of the compounds was converted to logarithmic scale as log (1/C). The whole data set contain twenty six compounds and all the compounds contain 19 common atoms (excluding hydrogen). The atoms of the molecules were numbered keeping serial numbers of the common atoms same in all the compounds (as shown in **Figure 1**). The mulliken's charges and dipole moment of common atom for twenty six pyrazole derivatives were calculated by CS MOPAC pro under CS Chemoffice software package[10]. During MOPAC analysis the wave function was treated as closed shell (restricted).

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

Sl No	R	R_1	R_2	R_3	R_4	Cell Viability (C)	Log (1/C)
1	OCH_2CH_3	H	H	H	H	32	-1.50515
2	OCH_2CH_3	H	OH	H	H	38	-1.57978
3	OCH_2CH_3	H	H	OH	H	39	-1.59106
4	OCH_2CH_3	H	OH	OH	H	85	-1.92942
5	OCH_2CH_3	OH	H	H	OH	50	-1.69897
6	OH	H	H	H	H	72	-1.85733
7	OH	H	H	OH	H	53	-1.72428
8	OH	H	OH	OH	H	94	-1.97313
9	OH	OH	H	H	OH	39	-1.59106
10	$N(CH_2CH_3)_2$	H	H	H	H	67	-1.82607
11	$N(CH_2CH_3)_2$	H	OH	H	H	79	-1.89763
12	$N(CH_2CH_3)_2$	H	OH	OH	H	91	-1.95904
13	$N(CH_2CH_3)_2$	OH	H	H	OH	62	-1.79239

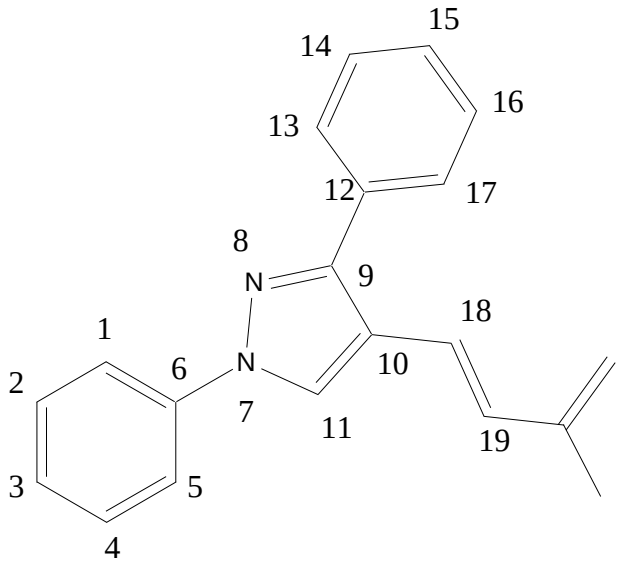
Model development

To begin the model development process, the whole data set (n=26) was divided into training (n=20, 75% of the total number of compounds) and test (n=6, 25% of the total number of compounds) sets by *k*-means clustering technique [11] applied on standardized descriptor matrix of the mulliken's charges and dipole moment. 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.

Table 2: Molecular scaffolds of the compounds along with their activity

Sl No	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Cell Viability (C)	Log (1/C)
14	H	H	H	H	OCH_3	OCH_2CH_3	O	82	-1.91381
15	H	OCH_3	H	H	OCH_3	OCH_2CH_3	O	64	-1.80618
16	H	H	OCH_3	H	OCH_3	OCH_2CH_3	O	81	-1.90849
17	H	OCH_3	OCH_3	H	OCH_3	OCH_2CH_3	O	62	-1.79239
18	OCH_3	H	H	OCH_3	OCH_3	OCH_2CH_3	O	77	-1.88649
19	H	H	H	H	H	OCH_2CH_3	O	80	-1.90309
20	H	H	H	H	OCH_3	OCH_3	O	75	-1.87506
21	H	H	H	H	H	OCH_3	O	78	-1.89209
22	H	H	H	H	OCH_3	OCH_2CH_3	S	94	-1.97313
23	H	H	H	H	H	OCH_2CH_3	S	88	-1.94448
24	H	H	H	H	OCH_3	OCH_3	S	83	-1.91908
25	H	H	H	H	H	OCH_3	S	87	-1.93952
26	H	H	H	H	OH	OCH_3	O	62	-1.79239

Figure 1: Common atom of the molecules



Software used for model development

MINITAB [12] was used for stepwise regression method. STAISTICA [13] was used for the determination of the LOO (leave-one-out) values of the training set compounds.

Stepwise Regression

In stepwise regression [14], a multiple term linear equation was built step-by-step. The basic procedures involve (1) identifying an initial model, (2) iteratively "stepping", i.e., repeatedly altering the model of the previous step by adding or removing a predictor variable in accordance with the "stepping criteria", (F = 3.5 for inclusion; F = 3.4 for exclusion) in our case and (3) terminating the search when stepping is no longer possible given the stepping criteria, or when a specified maximum number steps has been reached. Specifically, at each step all variables are reviewed and evaluated to determine which one will contribute most to the equation. That variable will then be included in the model, and the process started again. A limitation of the stepwise regression search approach is that it presumes that there is a single "best" subset of X variables and seeks to identify it. There is often no unique "best" subset, and all possible regression models with a similar number of X variables as in the stepwise regression solution should be fitted subsequently to study whether some other subsets of X variables might be better.

Statistical parameters

The statistical qualities of various equations were judged by calculating several metrics namely squared correlation variance (R^2), explained variance (R_a^2), standard error of estimate (s) and variance ratio (F) at specified degrees of freedom (df) [15]. Internal validation parameters like Q^2_{int} as well as $r^2_{m(LOO)}$ [16], external validation parameters like $Q^2_{ext(F1)}$, $Q^2_{ext(F2)}$ [17, 18], $r^2_{m(test)}$ [16] and overall validation parameters $r^2_{m(overall)}$ [16] were also reported.

External validation

The statistically internally optimized models were further evaluated for their real predictive power.

$Q^2_{ext(F1)}$ is calculated according to the following formula

$$Q^2_{ext(F1)} = 1 - \frac{\sum (Y_{obs} - Y_{cal})^2}{\sum (Y_{obs} - Y_{training})^2}$$

$Y_{training}$ Means mean activity value of the training set while Y_{obs} and Y_{cal} represent observed and calculated activity values.

$Q^2_{ext(F2)}$ is calculated according to the following formula

$$Q^2_{ext(F2)} = 1 - \frac{\sum (Y_{obs} - Y_{cal})^2}{\sum (Y_{obs} - Y_{test})^2}$$

Y_{test} Means mean activity value of the test set.

An additional parameter which penalizes a model for large differences between observed and predicted values of the prediction set compounds, as well as independent of the mean of training and prediction set, was also calculated for model external predictivity. The expression of r^2_m is defined as:

$$r^2_m = r^2(1 - \sqrt{r^2 - r_0^2})$$

Where r^2 and r_0^2 are determination coefficients of linear relations between the observed and predicted values of the compounds with and without intercept respectively. The r^2_m is applied for test set ($r^2_{m(test)}$), training set ($r^2_{m(LOO)}$) and the overall set ($r^2_{m(overall)}$).

RESULTS AND DISCUSSION

Membership of compounds in different clusters generated using *k*-means clustering technique is shown in **Table 3**. The test set size was set to approximately 25% to the total data set size[11] and the test set members along with their observed and calculated activity are given in **Table 4**.

Table 3: *k*-Means clustering of compounds using standardized descriptor

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

Table 4: Observed and calculated activity from stepwise regression model

Sl. No.	Observed activity against human DU 145 prostate cancer cell line Log(1/C)	Calculated activity
	Training Set	
1	-1.50515	-1.593957
2	-1.57978	-1.603138
3	-1.59106	-1.733048
4	-1.92942	-1.938343
6	-1.85733	-1.748451
7	-1.72428	-1.818671
8	-1.97313	-1.853824
10	-1.82607	-1.859090
12	-1.95904	-1.898520
13	-1.79239	-1.839970
14	-1.91381	-1.900274
15	-1.80618	-1.902685
16	-1.90849	-1.936120
18	-1.88649	-1.880813
19	-1.90309	-1.921988
20	-1.87506	-1.897568
22	-1.97313	-1.911158
23	-1.94448	-1.893506
24	-1.91908	-1.884171
26	-1.79239	-1.637508
	Test Set	
5	-1.69897	-1.6499
9	-1.59106	-1.6452
11	-1.89763	-1.89114
17	-1.79239	-1.93603
21	-1.89209	-1.88472
25	-1.93952	-1.89447

Observed activity (ref. 9); Calculated from eq. (1)

Using stepping criteria based on F value (F = 3.5 for inclusion; F = 3.4 for exclusion), the best fit equation was as follows:

$$\log(1/C) = -2.295(\pm 0.146)+0.51(\pm 0.163)C_2 - 3.07(\pm 0.58)C_4 - 1.05(\pm 0.25)C_5$$
$$+0.43(\pm 0.17)C_{14} + 6.1(\pm 1.6)C_{18}$$
$$R^2 = 0.8423, R_a^2 = 0.786, s = 0.063, PRESS = 0.113, F = 14.96(df = 5,14),$$
$$Q_{\text{int}}^2 = 0.678, n_{training} = 20, r_{m(LOO)}^2 = 0.666, Q_{ext(F1)}^2 = 0.711,\ Q_{ext(F2)}^2 = 0.693,$$
$$n_{\text{test}} = 06,\ r_{m(test)}^2 = 0.637,\ r_{m(\text{overall})}^2 = 0.676$$

The standard error of the respective mulliken's charges is mentioned within parentheses. Eq. (1) could explain 78.6% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 67.8%. The positive coefficients of C_2, C_{14} and C_{18} indicate that activity increases with increase in charge value of atom 2, 14 and 18 respectively. The negative coefficients of C_4 and C_5 indicate that activity increases with decrease in charge value of atoms 4 and 5 respectively. Compounds having highest charge value at position 2 (**compound 1**) showing highest activity against DU145 prostrate cancer cell line. The positive coefficient of atom 14 indicates the importance of hydroxyl group necessary for activity (**compound 2**). It was also observed that presence of hydroxyl group near position 14 also enhances the activity (like in **compounds 5 and 9**, where hydroxyl groups are present at position 13). The positive coefficient of atom 18 indicates the importance of double bond attached pyrazole nucleus and the ester or acid group. Compound having higher charge value for position 18 possesses higher activity (like **compounds 1 and 3**). It was observed that presence of ester or acid group attached to double bond enhances the activity (like **compounds 3 or 9**). Compounds having lower charge value for position 4 showing higher activity (**compounds 5 and 7 etc**), where as compounds having higher charge value showing comparatively lower activity (**compound 23 and 25 etc)**. The negative coefficient of C_5 indicating

that compound with highest charge value showing lowest activity (**compound 8**). The compounds having lower value for atom 5 showing comparatively better activity (**compounds 2 and 3**). The statistical quality of the model showed acceptable internal validation

(Q^2_{int} =0.678 and $r^2_{m(LOO)}$ =0.666), external validation ($Q^2_{ext(F1)}$ =0.711, $Q^2_{ext(F2)}$ =0.693, $r^2_{m(test)}$ =0.637), and overall validation ($r^2_{\text{m(overall)}}$ =0.676).

OVERVIEW AND CONCLUSIONS

The whole dataset (n=26) was divided into a training set (20 compounds) and a test set (06 compounds) based on *k*-means clustering of the standardized descriptor matrix and model was developed from the training set (optimized by Q^2). The predictive ability of the models was judged from the prediction of the activity of the test set compounds. The model indicates the importance of hydroxyl group at various position (like 13, 14 etc) of the moiety. Presence of double bond attached pyrazole nucleus and the ester or acid group is beneficial for antitumor activity. The intercorrelation among the parameters used in equations is shown in **Table 5** and utmost care was exercised to avoid collinearities among the variables. There are diversity in the structures among pyrazole compounds and the structural difference in each compound result from the variation in the aryl substituted groups at position 1 and 3 of the pyrazole structural motif and various functional groups at position 4. These structural variations are responsible for changes in mulliken's charges of the common atoms of various compounds. So the mulliken's charges can be projected as a useful tool in the context of QSAR studies.

Table 5: Intercorrelation among descriptors used in the model from stepwise analysis.

	C_2	C_4	C_5	C_{14}	C_{18}
C_2	1.000	-.360	.049	.015	-.029
C_4	-.360	1.000	-.350	-.316	.534
C_5	.049	-.350	1.000	.512	-.169
C_{14}	.015	-.316	.512	1.000	-.249
C_{18}	-.029	.534	-.169	-.249	1.000

REFERENCES

1. Elguero J. Pyrazoles. In Katritzky AR, Rees CW, Scriven EFV, editors. Comprehensive Heterocyclic Chemistry II. Oxford: Elsevier Science Ltd; 1996. p 1-75.
2. Varvounis G, Fiamegos Y, Pilidis G. Pyrazol-3-ones. Part III: reactivity of the ring substituents. Adv Heterocycl Chem 2001; 80: 75-165.
3. Yamamoto S, Tomita N, Suzuki Y, Suzaki T, Kaku T, Hara T, Yamaoka M, Kanzaki N, Hasuoka A, Baba A, Ito M. Design, synthesis, and biological evaluation of 4-arylmethyl-1-phenylpyrazole and 4-aryloxy-1-phenylpyrazole derivatives as novel androgen receptor antagonists. Bioorg Med Chem 2012; 20: 2338-2352.
4. Zhang D, Wang G, Zhao G, Xu W, Huo L. Synthesis and cytotoxic activity of novel 3-(1H-indol-3-yl)-1H-pyrazole-5-carbohydrazide derivatives. Eur J Med Chem 2011; 46: 5868-5877.
5. Lv PC, Li HQ, Sun J, Zhou Y, Zhu HL. Synthesis and biological evaluation of pyrazole derivatives containing thiourea skeleton as anticancer agents. Bioorg Med Chem 2010; 18: 4606-4614.
6. Vujasinovic I, Paravic-Radicevic A, Mlinaric-Majerski K, Brajsa K, Bertosa B. Synthesis and biological validation of novel pyrazole derivatives with anticancer activity guided by 3D-QSAR analysis. Bioorg Med Chem 2012; 20: 2101-2110.
7. Elgazwy AS. Three-dimensional quantitative structure activity relationship (QSAR) of cytotoxic active 3, 5-diaryl-4,5-dihydropyrazole analogs: a comparative molecular field analysis (CoMFA) revisited study. Chem Cent J 2012; 6: 50.
8. Li CY, Li QS, Yan L, Sun XG, Wei R, Gong HB, Zhu HL. Synthesis, biological evaluation and 3D-QSAR studies of novel 4, 5-dihydro-1H-pyrazole niacinamide derivatives as BRAF inhibitors. Bioorg Med Chem 2012; 20: 3746-3755.
9. Christodoulou MS, Fokialakis N, Nam S, Jove R, Skaltsounisc AL, Haroutouniana SA. Synthesis and *In Vitro* Biological Evaluation of Novel Pyrazole Derivatives as Potential Antitumor Agents. Med Chem 2012; 8: 779-788.
10. CS ChemOffice is software of CambridgeSoft Corporation, USA.
11. Leonard JT, Roy K. On Selection of Training and Test Sets for the Development of Predictive QSAR models. QSAR & Comb Sci 2006; 25: 235-251.
12. MINITAB is statistical software of Minitab Inc, USA, www.minitab.com.
13. STATISTICA is a statistical software of Stat Soft Inc, www.statsoft.com
14. Darlington RB. Regression and linear models. New York: McGraw Hill; 1990.
15. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*, Oxford & IBH Publishing Co. Pvt. Ltd.: New Delhi, **1967.**
16. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. QSAR & Comb Sci 2008; 27: 302-313.
17. Hawkins DM. The problem of overfitting. J Chem Inf Comp Sci 2004; 44, 1-12.
18. 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