International Journal of ChemTech Research
CODEN (USA): IJCRGG ISSN : 0974-4290
Vol.4, No.1, pp 99-104, Jan-Mar 2012

A Theoretical Study of Phenolic Antioxidants at AM1 Semi Empirical Levels

Supratim Ray

Division of Pharmaceutical Chemistry, Dr. B C Roy College of Pharmacy & Allied Health Sciences, Bidhannagar, Durgapur, 713 206, India

Corres.author: supratimray_in@yahoo.co.in
Ph: +919434443946, Telefax: +91343-2532679

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 38 phenol compounds with antioxidant activity. Experimentally determined redox potential reported by Steenken et al (1982) and Jovanovic et al (1998) at pH 7 was used as a direct measure of the antioxidant activity. Different statistical tools used in this communication are stepwise regression analysis and partial least squares analysis (PLS). From the study it was found that hydroxyl group has positive contribution towards antioxidant activity. The best equations are obtained from PLS analysis considering explained variance (R_a^2=0.7188) and stepwise regression analysis considering predictive ability of the model (Q^2=0.6546).
Key words: Phenolic Antioxidants, AM1 Semi Empirical Levels.

Introduction

Reactive oxygen species (ROS) have been recognized as playing an important role in the initiation and/or progression of various diseases such as atherosclerosis, inflammatory injury, cancer and cardiovascular disease (1). Several studies have investigated the potential of plant products to serve as antioxidants against various diseases induced by free radicals (2). The antioxidant effect of plant products is mainly due to phenolic compounds, such as flavonoids, phenolic acids, tannins and phenolic diterpenes (3, 4). There have been numerous studies on the biological activities of phenolics, which are potent antioxidants and free radical scavengers (5, 6). Several authors have performed the QSAR study antioxidants by using different physicochemical parameters (7-9) for better understanding of mechanism of action. In the present paper, a QSAR study was performed for 38 phenol compounds with antioxidant activity reported by Steenken et al (10) and Jovanovic et al (11) to explain the properties of the phenol molecule responsible for antioxidant activity.

Computational

Descriptors

The Mulliken's charges and dipole moment of common atom for 38 phenol compounds were calculated by CS MOPAC pro under CS Chemoffice software package (12). During MOPAC analysis the wave function was treated as closed shell (restricted).

Methods

Stepwise Regression

In stepwise regression (13), 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 = 2 for inclusion; F = 1.9 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.

PLS

PLS is a generalization of regression, which can handle data with strongly correlated and/or noisy or numerous X variables (14-15). It gives a reduced solution, which is statistically more robust than MLR. The linear PLS model finds "new variables" (latent variables or X scores) which are linear combinations of the original variables. To avoid over fitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are non significant. Application of PLS thus allows the construction of larger QSAR equations while still avoiding over fitting and eliminating most variables. PLS is normally used in combination with cross validation to obtain the optimum number of components. This ensures that the QSAR equations are selected based on their ability to predict the data rather than to fit the data. In case of PLS analysis on the present data set, based on the standardized regression coefficients, the variables with smaller coefficients were removed from the PLS regression until there was no further improvement in Q^2 value irrespective of the components.

Data treatment and software

The antioxidant activities of phenol (10-11) compounds (Table 1) were used as such reported (redox potential at pH 7, E_7) for subsequent QSAR analysis as the response variable. All the 34 compounds contain 7 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 atoms for 38 phenol compounds were calculated by CS MOPAC pro under CS ChemOffice software package (12). The stepwise regression PLS analysis were performed using the statistical software MINITAB (16). The statistical qualities of the equations were judged by parameters like explained variance (R^2_a), correlation coefficient (R), standard error of estimate (s) and variance ratio (F) at specified degree of freedom (df). All accepted equations have regression coefficients and F ratio significant at 95 and 99% levels, respectively, if not stated otherwise. The generated QSAR equations were validated by leave-one-out or LOO method (17, 18) using MINITAB software (16) and the calculated parameters are predicted residual sum of squares (PRESS), standard deviation based on PRESS (S_{PRESS}), standard deviation of error of prediction (SDEP) (18) and cross validation R^2 (Q^2). Q^2 is calculated according to the following formula

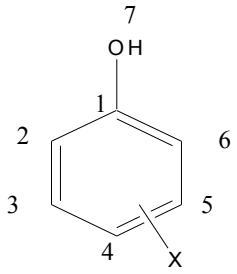
$$Q^2 = 1 - \frac{\sum (Y_{obs} - Y_{cal})^2}{\sum (Y_{obs} - \bar{Y})^2}$$

$\bar{Y}$ means average activity value of the entire data set while Y_{obs} and Y_{cal} represent observed and estimated activity values. Standard deviation of error of prediction (SDEP) is calculated according to the formula

$$SDEP = \sqrt{\frac{PRESS}{n}}$$

PRESS is the predicted residual sum of square using (leave-one-out) statistics and n is the number of component.

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



Sl. No.	X	Redox potential (E_7)
1	4-NO_2	1.23
2	3, 5-Cl_2	1.15
3	4-CF_2	1.13
4	3-NO_2	1.13
5	4-PhCO	1.12
6	3-CN	1.11
7	4-COOH	1.04
8	3-CH_3CO	0.98
9	4-H	0.97
10	4-Br	0.96
11	4-Cl	0.94
12	4-F	0.93
13	3-OH, 4-$COCH_3$	0.89
14	4-CH_3	0.87
15	3, 5-$(CH_3O)_2$	0.85
16	3-CH_3	0.85
17	3-OH, 5-OCH_3	0.84
18	3, 5-$(CH_3)_2$	0.84
19	4-Ph	0.84
20	2-CH_3	0.82
21	3-OH	0.81
22	2-OCH_3	0.77
23	4-OCH_3	0.73
24	3, 4-$(CH_3O)_2$	0.67
25	3, 4, 5-$(CH_3O)_3$	0.66
26	2-OH, 4-COOH	0.6
27	2, 6-$(CH_3O)_2$	0.58
28	2, 3-$(OH)_2$	0.58
29	2, 3-$(OH)_2$, 5-$COOCH_3$	0.56
30	2-OH	0.53
31	2-OH, 4-CH_3	0.52
32	4-OH	0.46
33	4-NH_2	0.41
34	4-CN	1.17
35	4-$COCH_3$	1.06
36	4-t-Bu	0.8
37	2, 6-$(CH_3)2$	0.77
38	2-OCH_3, 4-CH_3	0.68

Figure 1: Common atom of the molecules

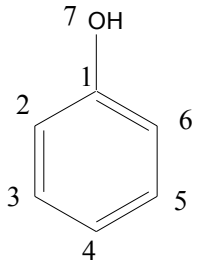


Table 2: Statistical comparison of different models

Type of statistical methods	R^2	R_a^2	Q^2
Stepwise regression	0.7418	0.7014	**0.6546**
PLS	0.7417	**0.7188**	0.6531

*The best values of different parameters are shown in bold

Table 3: Observed and calculated antioxidant activities from different models

Sl.No.	[a] Observed	[b] Calculated	[c] Calculated
1	1.23	1.255596	1.259464
2	1.15	0.9799	0.952901
3	1.13	1.138551	1.119592
4	1.13	1.182112	1.178895
5	1.12	1.000193	0.992163
6	1.11	0.981727	0.966749
7	1.04	1.043627	1.052114
8	0.98	0.943215	0.915381
9	0.97	0.934316	0.897491
10	0.96	1.001537	0.979607
11	0.94	0.887163	0.867446
12	0.93	0.762841	0.750662
13	0.89	0.919505	0.924575
14	0.87	0.847251	0.817025
15	0.85	0.877213	0.870514
16	0.85	0.849337	0.817826
17	0.84	1.061916	1.068617
18	0.84	0.853753	0.827452
19	0.84	0.81981	0.80073
20	0.82	0.851358	0.851854
21	0.81	0.956447	0.943416
22	0.77	0.822699	0.848004
23	0.73	0.751604	0.72079
24	0.67	0.637074	0.634579
25	0.66	0.684195	0.70013
26	0.6	0.574005	0.589838
27	0.58	0.64892	0.650024
28	0.58	0.533486	0.527556
29	0.56	0.532726	0.53503
30	0.53	0.543439	0.516273
31	0.52	0.450524	0.428882
32	0.46	0.755951	0.724171
33	0.41	0.759841	0.719297
34	1.17	0.964389	0.975695
35	1.06	1.009586	1.003922
36	0.8	0.831685	0.803034
37	0.77	0.845928	0.857129
38	0.68	0.7314	0.763029

[a] Observed (ref 10-11), [b] Calculated from Eq. 1, [c] Calculated from Eq. 2.

Results and Discussion

Statistical qualities of all important models are listed in **Table 2**. The observed and calculated activities are given in **Table 3**. The results obtained from different statistical methods are described below and the interpretations of the equations are also depicted.

Stepwise regression

Using stepping criteria based on F value (F = 2.0 for inclusion; F = 1.9 for exclusion), the best fit equation was derived.

$$E_7 = 3.827(\pm 1.399) - 0.69(\pm 0.302)C_2 - 0.64(\pm 0.2235)C_3 - 0.87(\pm 0.2245)C_4 + 11.1(\pm 4.33)C_7 + 0.050(\pm 0.0198)DIP$$

$$n = 38, s = 0.149, R^2 = 0.7418, R^2_a = 0.7014, Q^2 = 0.6546, R = 0.8612$$

$$F = 18.38(df\ 5, 32), PRESS = 0.6039, S_{PRESS} = 0.0243, SDEP = 0.126$$

.....(1)

The standard errors of the respective Mulliken's charges and dipole moment are mentioned within parentheses. Eq. (1) could explain 70.14% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 65.46%. The positive coefficient of C_7 indicates that activity increases with increase in charge value of atom 7. Compounds like **28, 30 and 36** have comparatively higher charges showed comparatively better activity. The positive coefficient of atom 7 indicates the importance of hydroxyl group towards activity. The negative coefficients of C_2, C_3 and C_4 indicate that activity decreases with increase in charge value of atoms 2, 3 and 4 respectively. Compounds with high values of charges for atom 2 (C_2) (like **4, 6**) for atom 3 (C_3) (like **1, 3 and 6**) and for atom 4 (C_4) (like **37**) showed comparatively poor activity. Positive coefficient of dipole moment indicates that compound with higher dipole moment have higher antioxidant activity (like compounds **30** and **31** etc).

PLS

The number of optimum components was 3 to obtain the final equation (optimized by cross validation). Based on the standardized regression coefficients, the following variables were selected for the final equation:

$$E_7 = 3.7538 + 1.1774C_1 - 0.3442C_2 - 0.6503C_3 - 0.7652C_4 + 10.962C_7 + 0.0524DIP$$

$$n = 38, s = 0.186, R^2 = 0.7417, R^2_a = 0.7188, Q^2 = 0.6531, R = 0.8612$$

$$F = 32.55(df\ 3, 34), PRESS = 0.6064, S_{PRESS} = 0.0251, SDEP = 0.126$$

.....(2)

Eq. (2) could explain 71.88% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 65.31%. The positive coefficient of the variable C_1 and C_7 indicate that the activities increases with increase in the charge value of atom 1 and 7 respectively while the negative coefficients of C_2, C_3 and C_4 indicate that the activity decreases with increase in charge values of atoms 2, 3 and 4. Compounds with high values of charges for atom 1 (C_1) (like **25 and 26**) showed comparatively higher activity. Positive coefficient of dipole moment indicates that compound with higher dipole moment have higher antioxidant activity.

Conclusions

For the model, the final equations (1 and 2) obtained from two techniques are of acceptable statistical quality and predictive potential considering the leave-one- out prediction statistics. The best equations are obtained from PLS analysis considering explained variance (R_a^2=0.7188) and stepwise regression analysis considering predictive ability of the model (Q^2=0.6546). From the study it was found that hydroxyl group has positive contribution towards antioxidant activity. The models also show the utility of Mulliken's charges and dipole moment in QSAR study for better understanding about the contribution of atoms or fragments in the molecules towards the antioxidant activity.

References

1. Halliwell B. Antioxidants and human diseases: a general introduction. Nutr. Rev. 1997, 55, S44–S52.
2. Hou W.C., Lin R.D., Cheng K.T., Hung Y.T., Cho C.H.,Chen C.H., Hwang S.Y. and Lee M.H. Free radical-scavenging activity of Taiwanese native plants. Phytomedicine 2003, 10, 170–175.

3. Chung K.T., Wong T.Y., Huang Y.W. and Lin Y. Tannins and human health: a review. Crit. Rev. Food Sci. Nutr. 1998, 38, 421–464.
4. Pietta P.G. Flavonoids as antioxidants. J. Nat. Prod. 2000, 63, 1035–1042.
5. Kahkonen M.P., Hopia A.I., Vuorela H.J., Rauha J.P., Pihlaja K., Kujala T.S. and Heinonen V. Antioxidant activity of plant extracts containing phenolic compounds. J. Agric. Food Chem. 1999, 47, 3954–3962.
6. Sugiura M., Naito Y., Yamaura Y., Fukaya C. and Yokoyama K. Inhibitory activities and inhibition specificities of caffeic acid derivatives and related compounds toward 5-lipoxygenase. Chem. Pharm. Bull. 1989, 4, 1039–1043.
7. Cao G., Sofic E. and Prior R.L. Antioxidant and prooxidant behavior of flavonoids: structure activity relationships. Free Radic. Biol. Med. 1997, 22, 749-760.
8. Wang L.F. and Zang H.Y. Unexpected role of 5-OH in DPPH radical-scavenging activity of 4-thiaflavans. Revealed by theoretical calculations. Bioorg. Med. Chem. Lett. 2004, 14, 2609-2611.
9. Modak B., Contreras M.L., Gonza´lez-Nilo F. and Torres R. Structure-antioxidant activity relationships of flavonoids isolated from the resinous exudate of Heliotropium sinuatum. Bioorg. Med. Chem. Lett. 2005, 15, 309-312
10. Steenken S. and Neta P. One-electron redox potentials of phenols. Hydroxy- and aminophenols and related compounds of biological interest. J. Phys. Chem. 1982, 86, 3661-3667.
11. Jovanovic S.V., Steenken S., Simic M.G. and Hara Y., Antioxidant properties of flavonoids: Reduction potentials and electron transfer reductions of flavonoid radicals, in Rice-Evans C.A. and Packers L., (Eds.), Flavonoids in Health and Disease, Marcel Dekker, Inc., New York, 1998.
12. CS ChemOffice is software of CambridgeSoft Corporation, USA.
13. Darlington R.B., Regression and linear models, McGraw Hill, New York, 1990.
14. Wold S., PLS for multivariate linear modeling in Van de Waterbeemd H., (Ed.), Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry), VCH, Weinheim, 1995, 195-218.
15. Fan Y., Shi L.M., Kohn K.W., Pommier Y. and Weinstein J.N., Quantitative structure-antitumor activity relationships of camptothecin analogs: Cluster analysis and genetic algorithm-based studies, J. Med. Chem., 2001, 44, 3254-3263.
16. MINITAB is statistical software of Minitab Inc, USA, http://www.minitab.com.
17. Word S. and Eriksson L., Statistical validation of QSAR results, in Van de Waterbeemd H., (Ed.), Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry), VCH, Weinheim, 1995, 312-318.
18. Debnath A.K., in Ghose A.K. and Viswanadhan V.N., (Eds) Combinatorial library design and evaluation: Principles, software tools, and applications in drug discovery, Marcel Dekker, New York, 2001, 73–129.

*****