Asian Pub. Corp.

www.asianpubs.org

ARTICLE

# Studies of Polarizability of Phenol and its Derivatives Using Computational Methods

Sameer Dixit[1,✉] and Arun K. Sikarwar[2]

## ABSTRACT

A quantitative structure-property relationship (QSPR) model was developed for prediction of polarizability of phenol derivatives. In this study we have attempted to develop a multiple linear regression (MLR) model with high accuracy and precision. For this first we prepared several models and then validated by statistical parameters like Q factor, LSE, *etc*. and proposed a model which have better prediction power of prediction of polarizability.

**Asian Journal of Organic & Medicinal Chemistry**

Volume: 2 Year: 2017
Issue: 4 Month: October–December
pp: 134–137
DOI: https://doi.org/10.14233/ajomc.2017.AJOMC-P76

Received: 19 June 2017
Accepted: 6 November 2017
Published: 29 December 2017

## KEYWORDS

Polarizability, QSAR, QSPR, Q factor, Predictive error, Predictive square error, Uncertainty of prediction.

## INTRODUCTION

Computational chemistry is applications of computer and computer enable calculations in chemistry for various purposes. One most important scope of computational chemistry is QSAR and QSPR followed by drug designing. Quantitative structure activity relationship (QSAR) provides a way to correlate the effect of structure over activity in terms of mathematical descriptors *viz.* topological indices. In this paper, we have developed models for polarizability using QSPR.

Polarizability is the ability for a molecule to be polarized. It is a property of matter. Polarizabilities determine the dynamical response of a bound system to external fields and provide insight into a molecule's internal structure [1]. Polarizability allows us to better understand the interactions between nonpolar atoms and molecules and other electrically charged species, such as ions or polar molecules with dipole moments. Neutral nonpolar species have spherically symmetric arrangements of electrons in their electron clouds. While in the presence of an electric field, their electron clouds can be distorted. The ease of this distortion is defined as the polarizability of the atom or molecule.

## EXPERIMENTAL

In case of modeling polarizability to build linear relationship and test model, the 49 compound data sets was used as training to build models. With the selected eight different descriptors, we will build linear models using the training data sets and following equations were obtained.

To developing the first model for polarizability of phenol derivatives in we used eight descriptors Mor29p, Mor20e, Mor04m, Mor23m, FDI, RDF045m, MATS5p, R3e. And To

***Author affiliations:***

[1]Department of Chemistry, M.J.P. Government Polytechnic College, Khandwa-450 001, India
[2]Department of Chemistry, Government Home Science P.G. College, Hoshangabad-461 001, India

✉To whom correspondence to be addressed:

E-mail: dixitsameer1@rediffmail.com

Available online at: http://ajomc.asianpubs.org

develop the second model eight descriptors Mor04m, Mor23m, FDI, RDF045m, MATS5p, R3e, eHOMO, eLUMO were used. These models expressed by eqns. 1 and 2, respectively.

Predicted polarizability = (2.10935 × Mor29p) + 5.105064 × Mor20e) + (1.906611 × Mor04m) + (-6.4015 × Mor23m) + (1.973671 × FDI) + (0.771308 × RDF045m) + (-0.16242 × MATS5p) + (5.407675 × R3e) (1)

Predicted polarizability = (1.561645 × Mor04m) + (-14.2767 × Mor23m) + (5.912913 × FDI) + (0.648036 × RDF045m) + (-0.07047 × MATS5p) + (6.622492 × R3e) + (0.407784 × eHOMO) + (0.546949 × eLUMO) (2)

By regression statistics we get correlation coefficient is 0.9958, $r^2$ is 0.9916, adjusted R square is 0.9657 and the standard error is 1.5435 for model-I, which described by eqn. 1. And we get correlation coefficient is 0.9931, $r^2$ is 0.9863, adjusted R square is 0.9596 and standard error is 1.9636 for model-II which described by eqn. 2.

## RESULTS AND DISCUSSION

In order to confirm most powerful predictable model for polarizability we have applied some statistical parameter. It is worthy to mention that a model (regression equation) with excellent statistics may not necessary have excellent predictive power. Thus the next step of regression analysis is to examine predictive power of the proposed model this can be easily done by calculating poglianis quality factor Q. the quality factor [2] Q is define as the ratio of correlation coefficient (r) to the standard error of estimation (sd). Thus the higher the value of r and the lower the standard error of the estimation (sd), are the higher will be the Q values and the better will be the proposed QSPR models. Polarizability model-I (eqn. 1) has better prediction power according to value of Q.

In fact, the first aim of any modeler should be validation for the predictive application of the QSAR model, for both the mechanistic approach and the statistical one. The famous "Kubinyi Paradox" [3,4] emphasized also by Tropsha *et al.* [7] in their famous papers: 'Beware of Q' [5,6] and 'The Importance of being Earnest' [7] is that: The best fit models are not the best ones for prediction! In fact, a QSAR model must, first of all, be a real model, robust and predictive, to be considered a reliable model [8]; only a stable and predictive model can be usefully interpreted for its mechanistic meaning, even so this is not always easy or feasible.

These statistical parameters are support Model-I for polarizability due to low value of LSE and PE is much greater than R for model-I (eqn. 1); is the better model compares to other. The cross-validated PRESS and SSY as recorded in Table-1 indicates model-I (eqn. 1) for polarizability is a better model and will give excellent result. According to SPRESS and PSE values model-I (eqn. 1) is a better model and will also give excellent result.

It is worthy to mention that the QSAR models *viz*. excellent statistics may not have excellent predictive power. There are necessary to investigate predictive power of all the models discussed above. This can be done by cross validation method. The key statistical measures of a QSAR equations predictive ability are the followings.

TABLE-1
CROSS VALIDATION PARAMETERS FOR MODELING OF POLARIZABILITY

| Statistical parameters | Model-I | Model-II |
|---|---|---|
| R | 0.996 | 0.993 |
| $R^2$ | 0.992 | 0.986 |
| SE or Sd | 1.544 | 1.964 |
| N | 49 | 49 |
| no of Descriptors | 8 | 8 |
| PRESS | 97.678 | 158.09 |
| SSY | 433.156 | 369.579 |
| $R^2cv$ | 3.435 | 1.338 |
| SPRESS | 1.563 | 1.988 |
| PSE | 1.412 | 1.796 |
| $R^2A$ | 0.966 | 0.96 |
| LSE | 97.678 | 158.09 |
| PE | 0.572 | 0.573 |
| Q = r/sd | 0.645 | 0.506 |
| PRESS/SSY | 0.226 | 0.428 |

1. Predictive sum of squares (PRESS)
2. Sum of squares of the response values (SSY)
3. Overall predicted ability ($R^2cv$ or $Q^2$)
4. Uncertainty of prediction (SPRESS)
5. Predictive square error (PSE)

If PRESS is smaller than the sum of squares of the response value (SSY), the model indicates better than chance and can be considered "statistically significant". The ratio PRESS/SSY can be used also to calculate approximate confidence intervals of prediction of new observations (compounds). To be reasonable QSAR model PRESS/SSY should be smaller than 0.4 and the value of this ratio smaller than 0.1 indicates an excellent model. The cross-validated PRESS and SSY as recorded in Table-1 indicates model-I (eqn. 1) for polarizability is a better model compare to model-II (eqn. 2) and will give excellent result.

If the PRESS value is transformed in a dimensionless term by relatively to the initial sum of squares one obtain $Q^2$ *i.e.* complement to the fraction of unexplained variance over the total variable. This quantity is also called predictive ability or cross validation correlation coefficient. In such case it is expressed by $r^2cv$ or $R^2cv$. It is observed that $r^2cv < R^2$. This is found to be the case of polarizability model-I (eqn. 1) in present study also.

Generally PRESS and $Q^2$ ($r^2cv$ or $R^2cv$) have good properties, which render than approximate for statistical testing with critical distribution. However, for practical purpose end users the use of square root of PRESS/n seems to be more directly related to the uncertainty of prediction. This ratio is named as Predictive square error and symbolized as PSE. The parameter PSE is particularly used when SPRESS coincides with SE. Obviously like SE and SPRESS the lower value of PSE indicates least uncertainty in prediction. The parameter PSE is important because it has the same unity as that of the activity. Polarizability model-I (eqn. 1) has better prediction power according to value of PSE.

The magnitude of SPRESS indicates uncertainty in prediction. Like the standard error of estimation (SE) the model will be smallest value of SPRESS is considered to have better predictive power. However in our case SPRESS is coincides

with SE and is therefore, not useful to explain the predictive power of models. Polarizability model-I (eqn. 1) have better prediction power according to value of SPRESS.

The probable error of the coefficient of correlation (PE) is an interesting parameter used in QSAR in deciding whether or not the proposed correlation is good or not. This parameter as stated in last chapter V is calculated using the following expression:

$$PE = 2/3\left(1-\frac{r^2}{\sqrt{n}}\right)$$

where r is the correlation coefficient and n is number of observation *i.e.* the number of compound used. It is argued that if (1) r < PE, than correlation is not significant (2) r > PE, than several times greater or at least three times greater than PE, the correlation is indicated and (3) r > 6PE, than correlation is definitely good. Polarizability model-I (eqn. 1) a better prediction power according to value of PE.

This parameter LSE is calculated by summing the square of the residue thus we have:

$$LSE = \Sigma(Residue)^2 = \Sigma(X_{obs}-X_{cal})^2$$

where $X_{obs}$ and $X_{cal}$ are the observed n calculated value (property or activity) respectively. We have also calculated LSE values for proposed models and are recorded in Table-1 which is in favour of the proposed model-I. If LSE value is low for property or activity compare to other models than model is better prediction power. *i.e*. Model has low LSE value has excellent prediction power

## Conclusion

Cross validation parameters for modeling of polarizability support model-I. Observed value of polarizability was plotted against and predicted values (using eqn. 1 shown in Fig. 1). Observed and predicted value of polarizability using eqn. 1 shown in Table-2. Fig. 1 clearly indicates that there is a significant co-relation between observed and predicted values of polarizability. Only BiSPHA [bisphenol-A (4,4'-propane-2,2-diyldiphenol)], 3DMAPH [3-(dimethylamino)phenol] shows deviation. Other molecule shows excellent co-relation for Polarizability. (Correlation coefficient is 0.9931, $r^2$ is 0.9863).

TABLE-2
OBSERVED AND PREDICTED VALUE OF POLARIZABILITY USING EQN. 1

| S. No. | Phenol derivative | Polarizability ± 0.5 $10^{-24}$ cm$^3$ | | Residuals | Standard residuals |
|---|---|---|---|---|---|
| | | Observed | Predicted | | |
| 1 | 4-$OCH_3$ | 13.8 | 11.857 | 1.9428 | 1.0816 |
| 2 | 4-$OC_2H_5$ | 15.63 | 13.902 | 1.7277 | 0.9618 |
| 3 | 4-$OC_3H_7$ | 17.47 | 16.435 | 1.0345 | 0.576 |
| 4 | 4-$OC_4H_9$ | 19.31 | 19.914 | -0.604 | -0.336 |
| 5 | 4-$OC_6H_{13}$ | 22.98 | 21.514 | 1.4656 | 0.8159 |
| 6 | H | 11.15 | 12.849 | -1.699 | -0.946 |
| 7 | 4-$NO_2$ | 13.74 | 13.59 | 0.1501 | 0.0835 |
| 8 | 4-Cl | 13.09 | 14.885 | -1.795 | -1 |
| 9 | 4-I | 16.27 | 14.664 | 1.6057 | 0.8939 |
| 10 | 4-CHO | 13.83 | 11.963 | 1.8672 | 1.0396 |
| 11 | 4-F | 11.15 | 12.713 | -1.563 | -0.87 |
| 12 | 4-$NH_2$ | 12.83 | 13.199 | -0.369 | -0.205 |
| 13 | 4-OH | 11.89 | 12.818 | -0.928 | -0.517 |
| 14 | 4-$CH_3$ | 13.06 | 13.349 | -0.289 | -0.161 |
| 15 | 4-$C_2H_5$ | 14.93 | 16.119 | -1.189 | -0.662 |
| 16 | 4-$NHCOCH_3$ | 16.81 | 15.741 | 1.0689 | 0.5951 |
| 17 | 4-CN | 13.02 | 11.161 | 1.8594 | 1.0352 |
| 18 | 4-$OC_6H_5$ | 21.63 | 19.348 | 2.282 | 1.2704 |
| 19 | Bisphenol-A | 27.02 | 22.726 | 4.2944 | 2.3909 |
| 20 | 4-Br | 14.2 | 14.722 | -0.522 | -0.291 |
| 21 | 4-C $(CH_3)_3$ | 18.44 | 18.82 | -0.38 | -0.212 |
| 22 | 3-$NO_2$ | 13.74 | 15.049 | -1.309 | -0.729 |
| 23 | 3-$NHCOCH_3$ | 16.81 | 15.908 | 0.9016 | 0.502 |
| 24 | 3-Cl | 13.09 | 15.438 | -2.348 | -1.307 |
| 25 | 3-$C(CH_3)_3$ | 18.44 | 17.447 | 0.9926 | 0.5526 |
| 26 | 3-$CH_3$ | 13.06 | 12.736 | 0.3241 | 0.1804 |
| 27 | 3-$OCH_3$ | 13.8 | 12.754 | 1.0457 | 0.5821 |
| 28 | 3-N $(CH_3)_2$ | 16.82 | 11.474 | 5.3459 | 2.9762 |
| 29 | 3-$C_2H_5$ | 14.93 | 15.492 | -0.562 | -0.313 |
| 30 | 3-Br | 14.2 | 12.922 | 1.2783 | 0.7117 |
| 31 | 3-CN | 13.02 | 10.599 | 2.4205 | 1.3476 |
| 32 | 3-F | 11.15 | 13.359 | -2.209 | -1.23 |
| 33 | 3-OH | 11.89 | 13.526 | -1.636 | -0.911 |
| 34 | 3-$NH_2$ | 12.83 | 13.397 | -0.567 | -0.315 |
| 35 | 2-$CH_3$ | 13.06 | 12.885 | 0.1745 | 0.0972 |
| 36 | 2-Cl | 13.09 | 14.448 | -1.358 | -0.756 |
| 37 | 2-F | 11.15 | 13.83 | -2.68 | -1.492 |
| 38 | 2-$OCH_3$ | 13.8 | 16.462 | -2.662 | -1.482 |
| 39 | 2-$C_2H_5$ | 14.93 | 17.555 | -2.625 | -1.462 |
| 40 | 2-OH | 11.89 | 13.79 | -1.9 | -1.058 |
| 41 | 2-OH, 4$CH_3$ | 13.81 | 14.055 | -0.245 | -0.137 |
| 42 | 2-$NH_2$ | 12.83 | 13.65 | -0.82 | -0.456 |
| 43 | 2-CN | 13.02 | 10.642 | 2.3783 | 1.3241 |
| 44 | 2-$NO_2$ | 13.74 | 15.651 | -1.911 | -1.064 |
| 45 | 2-Br | 14.2 | 16.93 | -2.73 | -1.52 |
| 46 | 2-$C(CH_3)_3$ | 18.44 | 18.622 | -0.182 | -0.101 |
| 47 | 4-$C_3H_7$ | 16.77 | 17.5 | -0.73 | -0.407 |
| 48 | 4-$C_4H_9$ | 18.61 | 18.347 | 0.2627 | 0.1463 |
| 49 | 4-$C_5H_{11}$ | 20.44 | 18.801 | 1.6393 | 0.9127 |

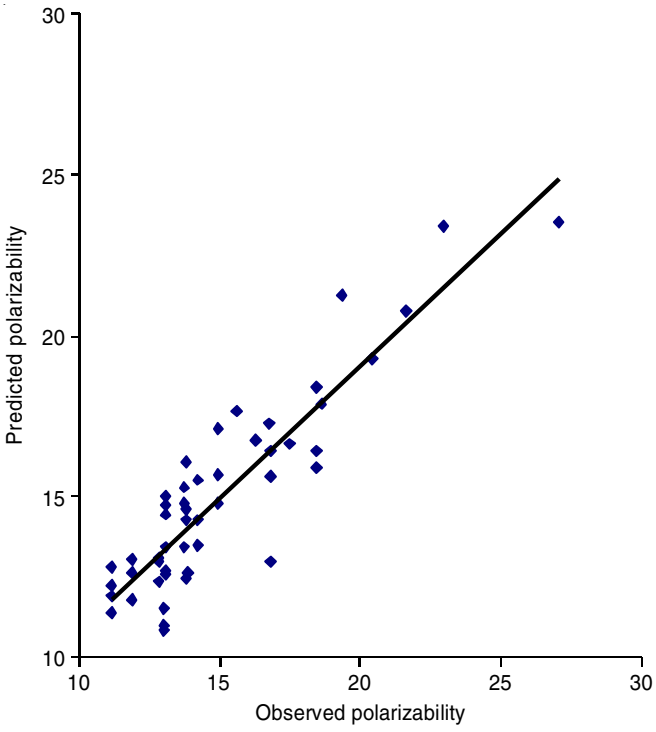


Fig. 1. Correlation of observed and predicted value of polarizability using eqn. 1

## ACKNOWLEDGEMENTS

The authors are thankful to Mr. C.G. Dhabu, Principal M.J.P. Government Polytechnic College, Khandwa, India for providing the necessary research facilities.

## REFERENCES

1. L. Zhou, F.X. Lee, W. Wilcox and J. Christensen, Magnetic Polarizability of Hadrons from Lattice QCD, European Organization for Nuclear Research (CERN), pp. 1-4, 25 May (2010).
2. L. Pogliani, *Amino Acids*, **6**, 141 (1994); https://doi.org/10.1007/BF00805842.
3. J.H. van Drie, *Curr. Pharm. Des.*, **9**, 1649 (2003); https://doi.org/10.2174/1381612033454568.
4. J.H. van Drie, eds.: P. Bultinck, H. de Winter, W. Langenaeker and J.P. Tollenaere, Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker (2004).
5. H. Meyer, *Arch. Exp. Pathol. Pharmakol*, **42**, 109 (1899); https://doi.org/10.1007/BF01834479.
6. A. Golbraikh and A. Tropsha, *J. Mol. Graph. Model.*, **20**, 269 (2002); https://doi.org/10.1016/S1093-3263(01)00123-1.
7. A. Tropsha, P. Gramatica and V.J. Gombar, *QSAR Comb. Sci.*, **22**, 69 (2003); https://doi.org/10.1002/qsar.200390007.
8. P. Gramatica, *QSAR Comb. Sci.*, **26**, 694 (2007); https://doi.org/10.1002/qsar.200610151.