



QSAR Studies on Nitrobenzene Derivatives using Hyperpolarizability and Conductor-like Screening Model as Molecular Descriptors

A. N. Alias* , Z. M. Zabidi 

Faculty of Applied Sciences, Universiti Teknologi MARA Perak Branch Tapah Campus, 35400 Tapah Road Malaysia

Abstract: Nitrobenzene derivatives are organic compounds that have been widely synthesized and used in chemical industries such as the polymer industry, lumber preservatives, textile industry, pesticides, and warlike weapons industry. The rapid growth of nitrobenzene derivatives in the industry requires research into the effects of toxicity in the environment. Quantitative structure-activity relationship (QSAR) models were useful in understanding how chemical structure relates to the toxicology of chemicals. In the present study, we report quantum molecular descriptors using conductor-like screening model (COs) area, the linear polarizability, first and second order hyperpolarizability for modelling the toxicology of the nitro substituent on the benzene ring. All the molecular descriptors were performed using semi-empirical PM6 approaches. The QSAR model was developed using stepwise multiple linear regression. We found that the stable QSAR modelling of toxicology of the benzene derivatives used second order hyper-polarizability and COs area, which satisfied the statistical measures. Second order hyperpolarizability shows the best QSAR model with the value of $R^2 = 89.493\%$, $r^2 = 68.7\%$ and $r_{cv}^2 = 87.52\%$. We also found that the substitution of functional group in the nitrobenzene derivative for second order hyperpolarizability has the same sequence which was the γ ortho < γ meta < γ para. These has made that the second order hyperpolarizability was the best descriptors for QSAR model.

Keywords: Nitrobenzene, toxicity, QSAR, quantum molecular descriptors, high order hyperpolarizability, COs Area

Submitted: March 07, 2022. **Accepted:** June 12, 2022.

Cite this: Alias A, Zabidi Z. QSAR Studies on Nitrobenzene Derivatives using Hyperpolarizability and Conductor-like Screening Model as Molecular Descriptors. JOTCSA. 2022;9(3):953-68.

DOI: <https://doi.org/10.18596/jotcsa.1083840>.

***Corresponding author. E-mail:** ahmadnazib111@uitm.edu.my.

INTRODUCTION

Nitro substituent on the benzene ring (nitrobenzene derivatives) constitutes an organic compound that has been widely synthesized and used in the chemical industry. Nitrobenzene derivatives have been used as depolymerizing agents in the polymer industry to lower molecular weight (1). Nitrobenzene derivatives have also been shown to be good as a lumber preservative

(2). The textile industry uses nitrobenzene derivatives- in the manufacturing and processing of textiles, particularly in the wet processing of textiles (3). Modern pesticides typically contain a nitrobenzene derivative as their active ingredient, which protects plants and crops (4). Nitrobenzene-based explosives, which are often low in sensitivity and great in performance, have been utilized in warlike weapons as one sort of explosive (5).

Nitro group substituents show strong electron-withdrawing effects. It also belongs to the chemical class of compounds with bioactivation (6). Therefore, nitro substituents in benzene are responsible for a number of toxicities and adverse environmental effects (7). The rapid use of these chemicals necessarily requires knowledge of their effects on human and environmental safety and health. For example, o-nitrophenols are toxic to plants, fish, and many other organisms. They can accumulate in the food chain, pose potential risks to both human health and the environment (8). Nitrobenzene derivative has also been linked to carcinogenesis, mutagenesis, skin sensitization, and hepatotoxicity (9). Therefore, the useful understanding of this material on the mechanism and action of toxicity is still going strong (10-12).

The biological properties of organic materials are strongly related to their geometrical and electronic structures. Studies of quantitative structure activity relationships (QSAR) are one of the most important fields in computational chemistry because they can be used to investigate the relationships between biological activity and organic materials. A QSAR is a mathematical representation of biological activity in terms of numerical molecular descriptors. Each numerical molecular descriptor must be invariant to represent the molecular structure. The molecular descriptors are the physical and chemical characteristics of molecules, such as topological index, 3D-molecular geometrical information, thermodynamics descriptors, quantum-chemical descriptors, and constitution descriptors (13). In a wide range of applications, QSAR enables faster and more cost-effective models for new molecular designs. In QSAR, partitioning properties are typically utilized as descriptors. However, the availability and veracity of experimental data limit their application.

The quantum molecular descriptors provide an alternative for molecular descriptors. The quantum molecular descriptors are more accurate and detailed description to relate with biological activity (14). The quantum molecular descriptors using *ab initio* model Hamiltonian calculation such multiconfiguration self-consistent field, Moller-Plesset theory of variation or correlated pair many-electron theory (15). As an alternative to *ab initio* methods, density function theory and semi-empirical quantum chemistry method are practicable for molecular

descriptors. The net atomic charges, the HOMO-LUMO energy gap, chemical hardness, and chemical potential, in particular, have been used to correlate with various biological activities (16,17). Chemical hardness and chemical potential are strongly influenced by HOMO and LUMO energies (18,19). These energies need to consider electronic reorganization in the excited state, which may often lead to conceptually incorrect results for the prediction of QSAR (17).

Semiempirical methods are based on Hartree-Fock formalism omit some molecular integral calculations. For instance, the approximation of zero differential overlap (ZDO) reduces the number of multicenter integrals. Numerous approaches have been introduced to improve the calculation's accuracy, including CNDO, NDDO, INDO, and MINDO (20). Semiempirical methods also have been widely used in the application of QSAR/QSPR because they are significantly faster than *ab initio* and DFT approaches, particularly for large molecule calculations. Several software, such as MOPAC, AMPAC, SPARTAN, and VAMP, has been developed to perform these semi-calculations. Wang et. al. has implemented semi-empirical method using MOPAC software to investigate toxicity endpoints for a wide range of compound (21). Adinin et al. also used the same software to generate quantum molecular descriptors of physicochemical properties of isothiocyanate antimicrobials (22).

The purpose of the present work was to study the polarizability, first and second order hyperpolarizability as molecular descriptors for modeling the toxicology of the nitro substituent on the benzene ring (nitrobenzene derivatives). Dipole moment is an important aspect in determining how a chemical reaction affects in the biological system. Polarizability, hyperpolarizability, and the surface of charge dispersion are both related to dipole moment. A conductor-like screening model (COs) area as a new class of molecular descriptors is also introduced in this paper. To the best of our knowledge, there are no studies using these descriptors for modeling the toxicology of the nitrobenzene derivatives. We performed our calculation using well-established semiempirical electronic structure program known as MOPAC. The semi-empirical quantum chemical calculation was based on the self-consistent field method to calculate the electronic and molecular orbital properties.

These calculations were much faster than ab initio and less time-consuming.

METHOD AND CALCULATION

where μ_{i0} is a component of dipole moment and α_{ij} is a tensor linear polarizability and α_{ij} is a tensor linear polarizability. The parameters β_{ijk} and γ_{ijkl} in Eq. 1 are the components of the first and second order hyperpolarizability. The ijkl suffixes denote as Cartesian component

$$W(E) = W(0) - \mu_{i0} E_i - \frac{1}{2!} \alpha_{ij} E_i E_j - \frac{1}{3!} \beta_{ijk} E_i E_j E_k - \frac{1}{4!} \gamma_{ijkl} E_i E_j E_k E_l - \dots \quad (\text{Eq. 1})$$

Calculation of the Quantum Molecular Descriptors

Molecular hyperpolarizability is primarily related to their role in nonlinear optics. Polarizability is related to a molecule's energy in an external field. The energy as a function of the external field (E) can be expressed as a power series as

implying over the X, Y, and Z axes. The component of the total dipole moment can be obtained using the derivation of energy with respect to the field. The series of molecular dipole moments can be written as (1).

$$\mu_i(E) = \mu_{i0} + \alpha_{ij} E_j + \frac{1}{2!} \beta_{ijk} E_j E_k + \frac{1}{3!} \gamma_{ijkl} E_j E_k E_l + \dots \quad (\text{Eq. 2})$$

The expression in (Eq. 2) represents the response of a dipole moment to an applied electric field of third order in the field. The coefficient on α_{ij} , β_{ijk} and γ_{ijkl} can be obtained using finite field or perturbation method. In this calculation we implemented finite field method by Kurtz et al. (23). The value of linear polarizability was a tensor diagonal vector in x, y or z and it is given by Eq. 3 (23).

$$\alpha = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) / 3 \quad (\text{Eq. 3})$$

The average of first order hyperpolarizability, β and second order (γ) hyperpolarizability for the interest value at 0.25 eV were given by Eq. 2 and 3, respectively.

$$\beta = (\beta_{xxx} + \beta_{yyy} + \beta_{zzz}) \quad (\text{Eq. 4})$$

$$\gamma = 1/5 [(\gamma_{xxxx} + \gamma_{yyyy} + \gamma_{zzzz}) + 2(\gamma_{xxyy} + \gamma_{xxzz} + \gamma_{yyzz})] \quad (\text{Eq. 5})$$

Conductor-like screening model (COs) area is related with the surface charge densities on the adjacent segments. The surface charge of a molecule is distributed on the molecule's interface. The charge distribution causes the localized charge distribution on the surface. The localized charge in the molecule may disrupt the Coulombic interaction, which 'screens' the polarization in the molecule. As a result, the screening is dependent on charge localization and molecular polarizability (16). The COs area is an effective area of the screening surface. We use the Klamt algorithm, which has been modified

from the dielectric polarized continuum mode, to calculate the COs area (24).

The molecular structures were generated using Avogadro version 1.2.0. Next, the geometric optimization was performed using the force field method MMFF94s with a step per update of 4. In this work, we use the parameterization of parametric method 6 (PM6) as the integration method (25). All the calculations were calculated using Semi-empirical MOPAC2016, Version: 21.002 James J. P. Stewart software. MOPAC programs implement the value and components of dipole moments, conductor-like screening model (COs) area, polarizability (α), first order (β) and second order (γ) hyperpolarizability (23). In the QSAR calculation, the linear polarizability constant was calculated using 0.25 eV.

Experimental Data

For the present work, we chose a data set of aquatic toxicity of benzene derivatives of ciliated protozoan *Tetrahymena pyriformis* which was retrieved from Fatemi and Malekzadeh (27). Toxicity is represented in terms of $\log(1/IGC_{50})$, where "IGC₅₀" refers to a concentration that inhibits growth for two days at 50 percent of its normal rate. In this study, a total of 120 nitrobenzene derivatives were investigated. The nitrobenzene derivatives compounds and their corresponding $\log(1/IGC_{50})$ values are listed in Table S1.

Statistical Analysis

Multiple linear Regression: The quantum molecular descriptors (input data) were normalized using Eq. 6.

$$I_i = \frac{I_x - I_{min}}{I_{max} - I_{min}} \tag{Eq. 6}$$

where I_x unnormalized input data, I_{max} was the maximum value of the sample, and I_{min} was the minimum value of value of the sample (28). The structure-toxicity models were developed using multiple linear regression, which was a simple approach for modelling the linear relationship between independent and dependent variables by fitting the linear equation. The Simple Linear Regression equation was stated as in Equation 7:

$$y = b_0 + b_1x_1 + \dots + b_nx_2 \tag{Eq. 7}$$

where b_0 was the intercept; b_1 and b_n was coefficient; x_1 and x_n was the independent descriptors, and y was the predictive (calculated) values. The multiple regression was calculated using the stepwise selection method with an alpha to enter and remove value of 0.25. We perform regression using the statistical software Minitab. The plot of observed vs. calculated toxicity with a 95% prediction index was plotted using Minitab.

Model evaluation and validation: The R^2 was the coefficient of determination, which describes how much of the variability of dependent y was explained by the independent variable of x . In other words, r^2 indicates how well the model predicts the variance of the independent variable explained. The value of R^2 (also called Pearson's r) was computed using equation (8).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \tag{Eq. 8}$$

where y is the y is the observed response variable, \hat{y} was prediction (calculated) value, \bar{y} was the mean value of y . Another indicator we use in this work was r^2 value which was a dimensionless goodness-of-fit indicator between linear graph of $y_{exp.}$ and $y_{cal.}$ The Variance Inflation Factor (VIF) was a measure of the indicator for multi-collinearity, commonly stated as:

$$VIF = \frac{1}{1 - R^2} \tag{Eq. 9}$$

where R^2 was the correlation coefficient calculated form Equation (9). Table 1 shows the rule for interpreting the value of VIF (29).

Table 1: VIF interpretation.

VIF value	Condition
VIF = 1	Not correlated
1 < VIF ≤ 5	The related model is acceptable
5 < VIF < 10	The multicollinearity is substantial
VIF ≥ 10	The related model is unstable, and recheck is necessary

Internal validation was accomplished through the employment of leave-one-out cross validation. The dataset was divided into two sets as training set and test set. This procedure was carried out by leaving one sample from the data set and using the other data samples as the training set, while the testing sample was included in the data set. The procedure was repeated until all data samples had been used as the testing set. The square of cross-validation coefficient (q^2) for LOO-CV is calculated using Eq. 10.

$$q^2 = 1 - \frac{\sum_{i=1}^{training} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{training} (y_i - \bar{y})^2} \tag{Eq. 10}$$

where y_i , \hat{y}_i , and \bar{y} are, respectively, the measured, predicted, and averaged values of the respective testing set.

The developed model was validated further in order to assess its prediction competency using the test and training sets. The k-fold cross validation technique is also useful for

validating QSAR models (30). The k-fold cross validation procedure can give an optimized prediction performance in QSAR (31). The k-fold cross-validation the resampling the data to determine the ability of predictive models. The dataset was divided randomly into k-partition with equal size of segments of fold for training and test set of data. The recommended value for k was 10 (32). In our work, we used 10-fold to holdout with an iteration process of 10 times. The training dataset was commanded to make predictions about the data in the validation fold. Thus, we have used the parameters of the k-fold cross-validated correlation coefficient (r_{cv}^2) for regression validation.

RESULTS

In this work, non-linear optical properties of molecules have been used, that was, polarizability, first and second order hyperpolarizability. The molecular descriptor of polarizability involves three components: α_{xx} , α_{yy} and α_{zz} . The perturbed external electrical field for first order hyperpolarizability involved β_{xxx} , β_{xyy} , β_{xzz} , β_{yxx} , β_{yyy} , β_{yzz} , β_{zxx} , β_{zyy} and β_{zzz} . While the perturbed external electrical field for second order hyperpolarizability involved γ_{xxxx} , γ_{yyyy} , γ_{zzzz} , γ_{xxyy} , γ_{xxzz} , γ_{yyxx} , γ_{yyzz} , γ_{zzxx} , and γ_{zzyy} . Polarizability and hyperpolarizability were related to the coordinate axis of molecular

polarization. The calculated polarizability, first order and second order hyperpolarizability, conductor-like screening model (COs) area, and dipole moment molecule using MOPAC were tabulated in supplementary Table S1-S3. The quantum molecular descriptors were implemented into the QSAR relationship using stepwise selection regression. In the stepwise regression method, polarizability, first and second order hyperpolarizability, and constant reassessment via their partial F (or t) statistics. The linear regression for polarizability was given in Equation (8).

$$\log(1/IGC_{50}) = 1.223 \alpha_{xx} + 1.547 \alpha_{yy} \quad (\text{Eq. 8})$$

$$n = 70, R^2 = 89.21\%, s = 0.3590, F = 280.99, q^2 = 0.677065.$$

The stepwise regression indicates two polarizability descriptors were significant. However, the value of VIF for α_{xx} and α_{yy} was more than 5 (6.70). This modeling equation was too highly correlated, which means multicollinearity was substantial. Figure 1(a) shows the correlation between the experimental and calculated values of $\log(1/IGC_{50})$ obtained from α_{xx} and α_{yy} . The plot shows the r^2 of experimental and calculated $\log(1/IGC_{50})$ was 67.7%. Figure 1(b) shows the residual plot versus the calculated value.

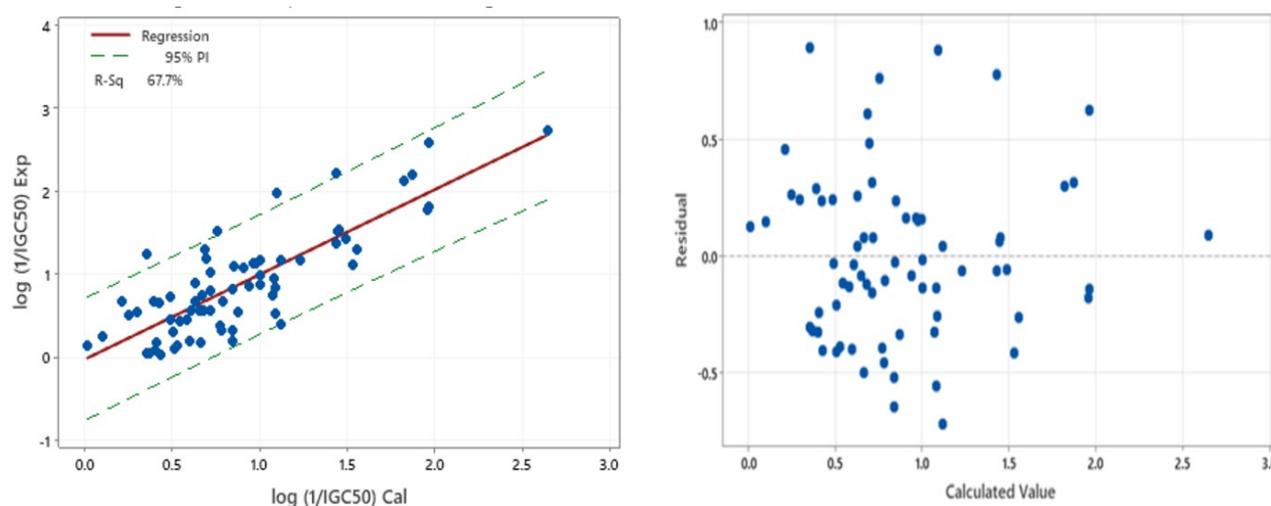


Figure 1: Plot of (left) experiment vs calculated $\log(1/IGC_{50})$ using linear polarizability. (Right) the residual plot with the calculated value.

The molecular descriptor for first order hyperpolarizability calculated from semi-empirical calculation has nine descriptors. The stepwise regression method determined the five values of descriptors that were significant

to the toxicology of benzene derivatives. The QSAR relationship obtained was given by Eq. 9.

$$\log(1/IGC_{50}) = 1.216 \beta_{xxx} + 0.763 \beta_{xzz} + 1.107 \beta_{yyy} + 0.992 \beta_{zyy} \quad (\text{Eq. 9})$$

$n = 70, R^2 = 59.79\%, s = 0.70337, F = 24.53, q^2 = 0.0784.$

The VIF value for $\beta_{xxx}, \beta_{xzz}, \beta_{yyy}, \beta_{zyy}$ were 1.52, 1.27, 1.46 and 1.27 respectively. The plot of experimental vs calculated ($1/IGC_{50}$) using first order hyper-polarizability is shown in Figure 1(b). The r^2 value using first order hyper-polarizability as molecular descriptor was 5.9%.

The second order hyperpolarizability calculated from semi-empirical calculation also has nine descriptors. There were four

values of descriptors that were significant to the toxicology of benzene derivatives. The second order hyperpolarizability relationship yields an equation:

$$\log(1/IGC_{50}) = 1.203 \gamma_{yyyy} + 0.587 \gamma_{zzzz} + 1.247 \gamma_{xxzz} - 1.276 \gamma_{yyxx} \quad (\text{Eq. 10})$$

$n = 70, R^2 = 89.493\%, s = 0.359559, F = 140.52, q^2 = 0.68453$

with the VIF value for $\gamma_{yyyy}, \gamma_{zzzz}, \gamma_{xxzz}$ and γ_{yyxx} were 2.31, 1.29, 4.91 and 3.01 respectively. The plot of experimental vs calculated ($1/IGC_{50}$) using first order hyper-polarizability is shown in Figure 3 and the r^2 value of this plot was 68.7%.

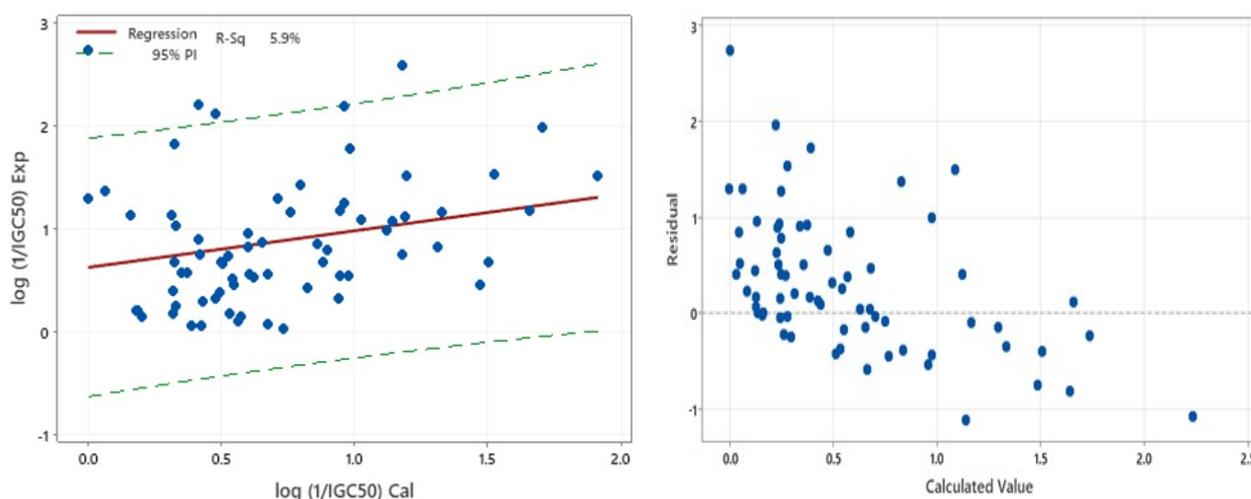


Figure 2: Plot of (left) experiment vs calculated $\log(1/IGC_{50})$ using first order hyperpolarizability. (Right) the residual plot with the calculated value.

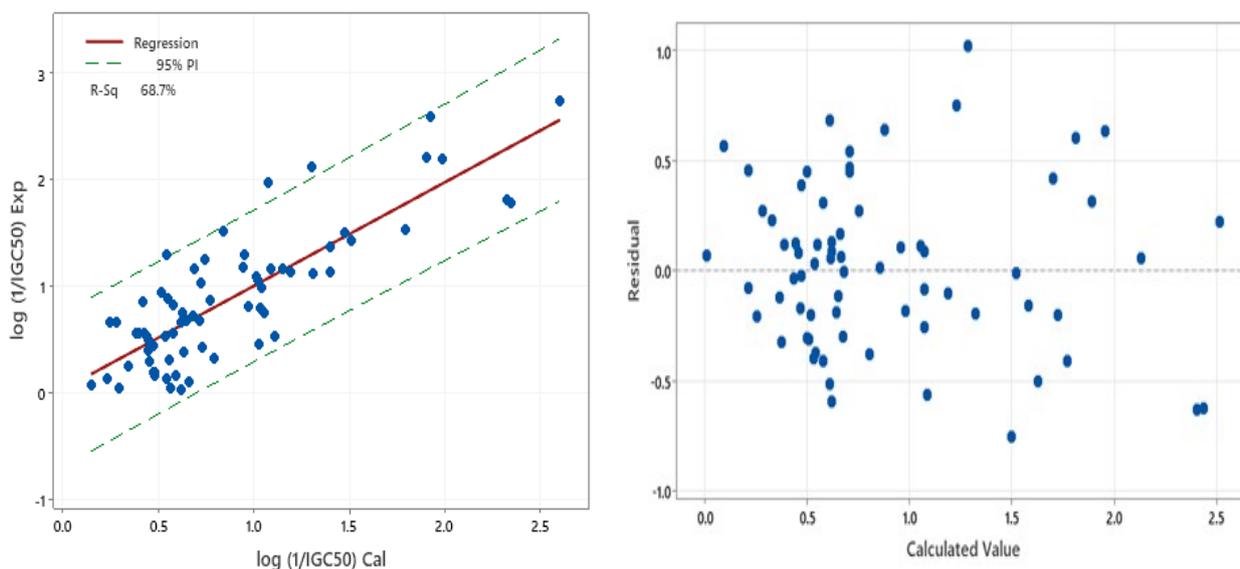


Figure 3. Plot of (left) experiment vs calculated $\log(1/IGC_{50})$ using second order hyperpolarizability. (Right) the residual plot with the calculated value.

The relationship between conductor-like screening model (COs) area and dipole moment with toxicology has been tested. In semiempirical calculation, two types of dipole moment were identified: the dipole moment net charge density and the hybrid. Only the dipole moment net charge density (μ_c) correlated well with COs area, indicating a good correlation with toxicology. The relationship was given by:

$$\log (1/IGC_{50}) = -0.264 \mu_c + 2.335 \text{ Cos} \quad (\text{Eq. 11})$$

$$n = 70, R^2 = 87.04\%, s = 0.3935, F = 228.26, q^2 = 0.615627.$$

Both the Cos area and μ_c have a VIF value of 3.06. The plot of experimental vs calculated ($1/IGC_{50}$) using Cos area and dipole moment net charge density is shown in Figure 4. The r^2 value for this plot was 61.4%.

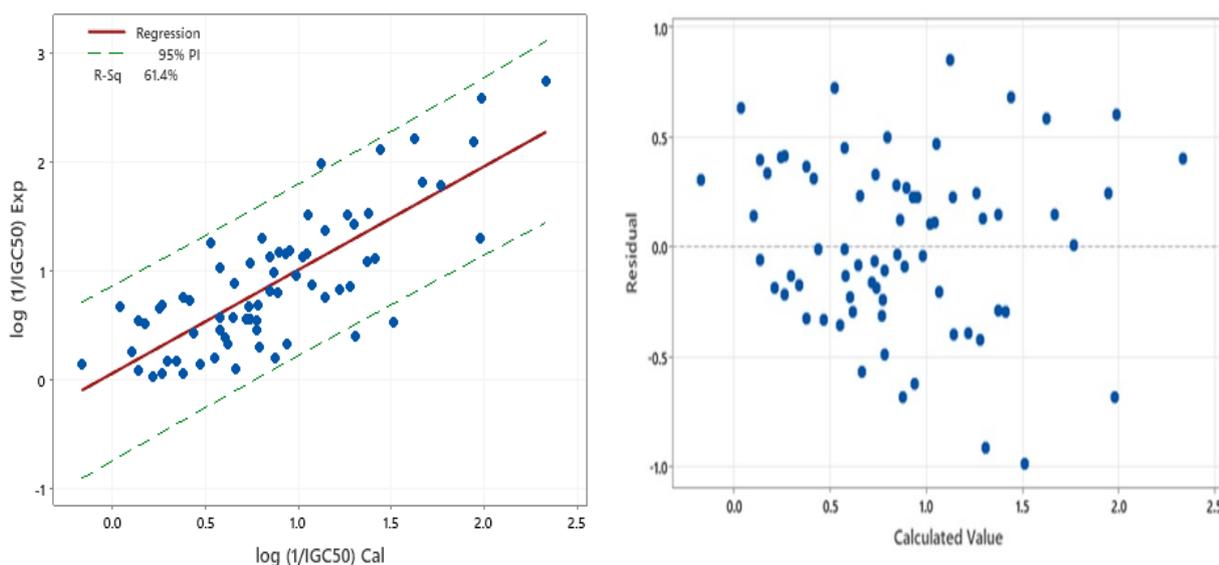


Figure 4: Plot of (left) experiment vs calculated $\log(1/IGC_{50})$ using Cos and dipole moment; (right) the residual plot with the calculated value.

The k-fold cross-validation has been used to check the stability of the QSAR model. An acceptable range of relative deviation of the r^2_{cv} of k-fold cross-validation with r^2 was $\pm 2\%$ (33). The values of r^2_{cv} for polarizability, first and second order hyper-polarizability were 88.78%, 55.2% and 87.52% respectively. While the value of r^2_{cv} for Cos area and dipole moment net charge density was 86.71%. All model relative deviation were in the acceptance range except for second order hyper-polarizability. The F value for polarizability, second order hyper-polarizability, Cos area and dipole moment net charge density has a value greater than 95, which has good levels of significance (34). In addition, the fitting of the experimental and calculated ($1/IGC_{50}$) has an r^2 value for polarizability, second order hyper-polarizability, Cos area, and dipole moment net charge density greater than 0.6 (34). The results also show that the q^2 for all descriptors except first order hyperpolarizability has a value greater than 0.5, which was the acceptance range for the QSAR model. Since

the VIF of α_{xx} and α_{yy} was 6.70 the multicollinearity was substantial. Therefore, the stable QSAR modeling for toxicology of benzene derivatives used second order hyper-polarizability, COs area and dipole moment net charge density as molecular descriptors, which satisfied the statistical measures.

DISCUSSION

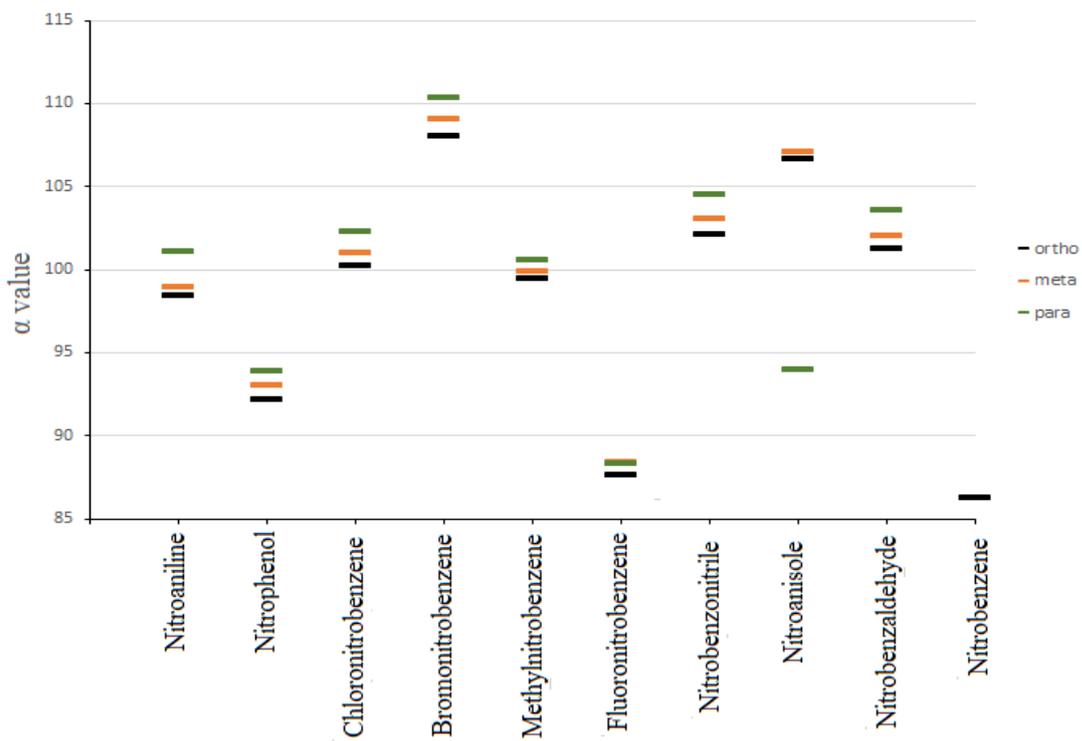
The toxicology of the nitrobenzene derivative elaborated QSAR model reveals that the electronic properties such as polarizability, hyper-polarizability, dipole moment, and conductor-like screening model (COs) area have an impact on this study. In quantum molecular calculations, the polarizability, hyper-polarizability, and dipole moment were calculated based on the non-linear optical properties of molecules whose induced frequency energy that we used to represent the external electric field was 0.25 eV. The interaction dipole moment of a molecule with the field was dependent on the permanent dipole moment, polarizability (α), first order

hyper-polarizability (β) and second order hyper-polarizability (γ). The linear response of electronic charge distribution can be described by linear polarizability. Molecular linear polarizability was related to charge distribution in the molecules. The linear polarizability was very closely linked to intermolecular forces, electronic interaction inside the molecule, and chemical reactivity, among other things (35). Tandon et al. also reported that linear polarizability has a good correlation with chemical-biological activity of anaesthetics and drugs for blocking activities (36). Furthermore, hydrophobicity was found to be directly related to molecular linear polarizability (37).

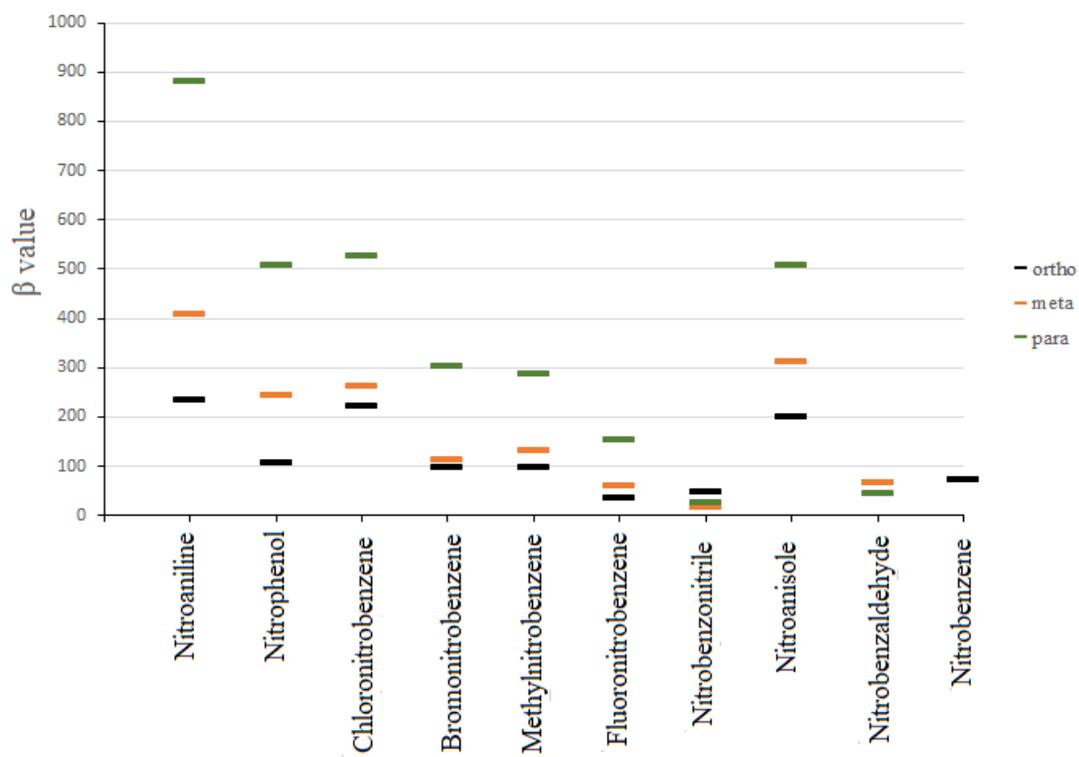
The expected relationship between substituent properties and calculated first order hyperpolarizabilities will be seen to require a more detailed electronic population analysis. The first order hyperpolarizability has showed a low stability model with a value of $r^2 = 59.79\%$. The first order hyperpolarizability also showed low stability and aquatic toxicity of hydrocarbons (EC_{50}) to aquatic organisms (16).

The conductor-like screening model (COs) area with molecular dipole moment net charge density was a surprisingly good fit for toxicity. The COs area was proportional to the surface charge densities of the surrounding segments of molecules. It was the effective area of the screening surface of the charge density on a van der Waals-like surface. The screening surface was correlated to the perturbation of Coulombic interaction in the molecule. The screening surface was dependent on the localization of charge and the polarizability of the molecule (16). While, molecular polarization was contributed by electronic charge, molecular vibration, and rotation, which reflect the molecular dipole moment.

In QSAR, the effect of substitution in the nitrobenzene structure was of great importance to obtain the best molecular descriptor. To confirm the influence of molecular structure on this nitrobenzene, we have performed a comparison of polarizability, hyperpolarizability, and COs area to the meta, ortho, and para-substituents for aniline (NH_2), phenol (OH), chloro, bromo, methyl (CH_3), fluoro, nitrile (CN) and methoxy (OCH_3) as shown in Figure 5. As seen clearly, the effect of meta, ortho, and para-substituents has significant changes to linear polarizability, second order hyperpolarizability, and COs area. The value of linear polarizability has the sequence α ortho < α meta < α para except for fluorobenzene and nitroanisole. For the second order hyperpolarizability, all the nitrobenzene has the same sequence that was the γ ortho < γ meta < γ para. This has made that the second order hyperpolarizability has the highest value of r^2 . The COs area of nitrobenzene has the sequence COs ortho < COs para < COs meta except for nitroaniline, nitroanisole, fluoronitrobenzene and nitrobenzaldehyde. Fluoronitrobenzene and nitrobenzaldehyde have nearly the same value of COs area for meta and para. While the sequence of the substituents from anisole to nitrobenzene was para < ortho < meta. For the first order hyperpolarizability, the substitutional of functional groups was difficult to expect. This might be due to the substitutional of para, meta, or ortho inducing the charge transfer effect in the first order hyperpolarizability calculation (38). This means there was no suitable sequence of the substitutional of para, meta, or ortho. In this case, the first order hyperpolarizability was not a good molecular descriptor for QSAR because it was too high.



(a)



(b)

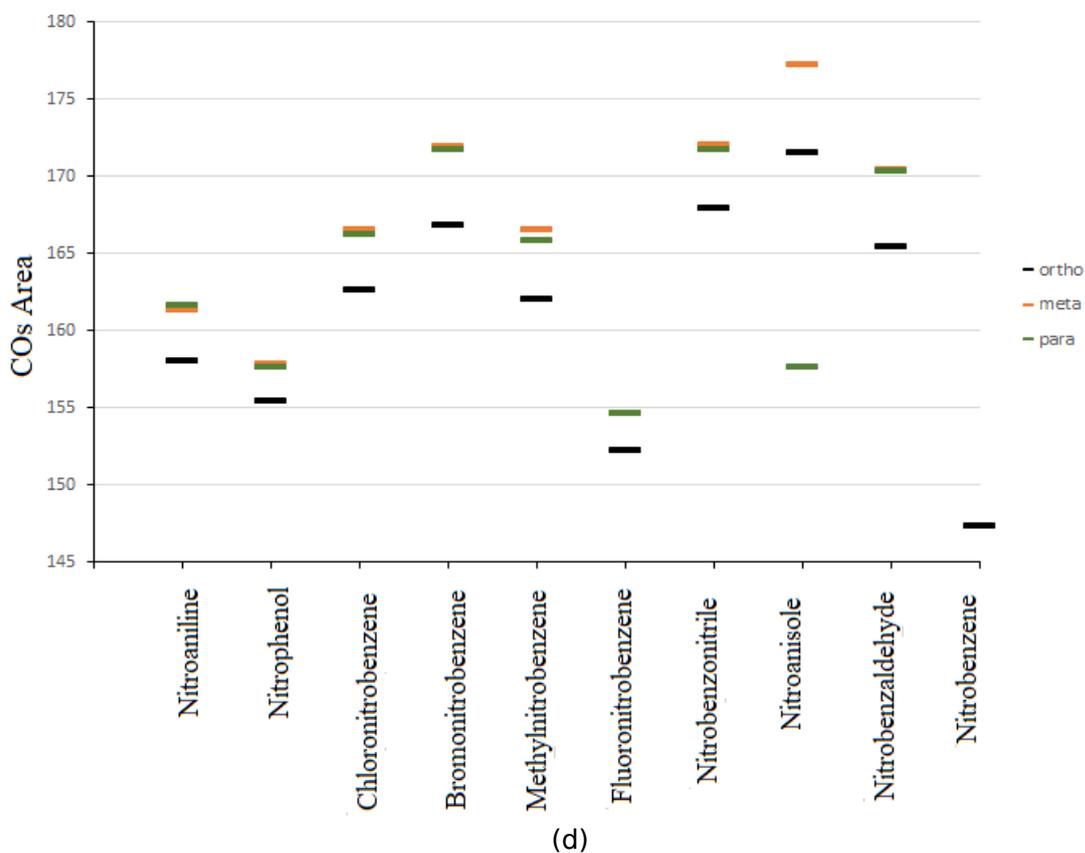
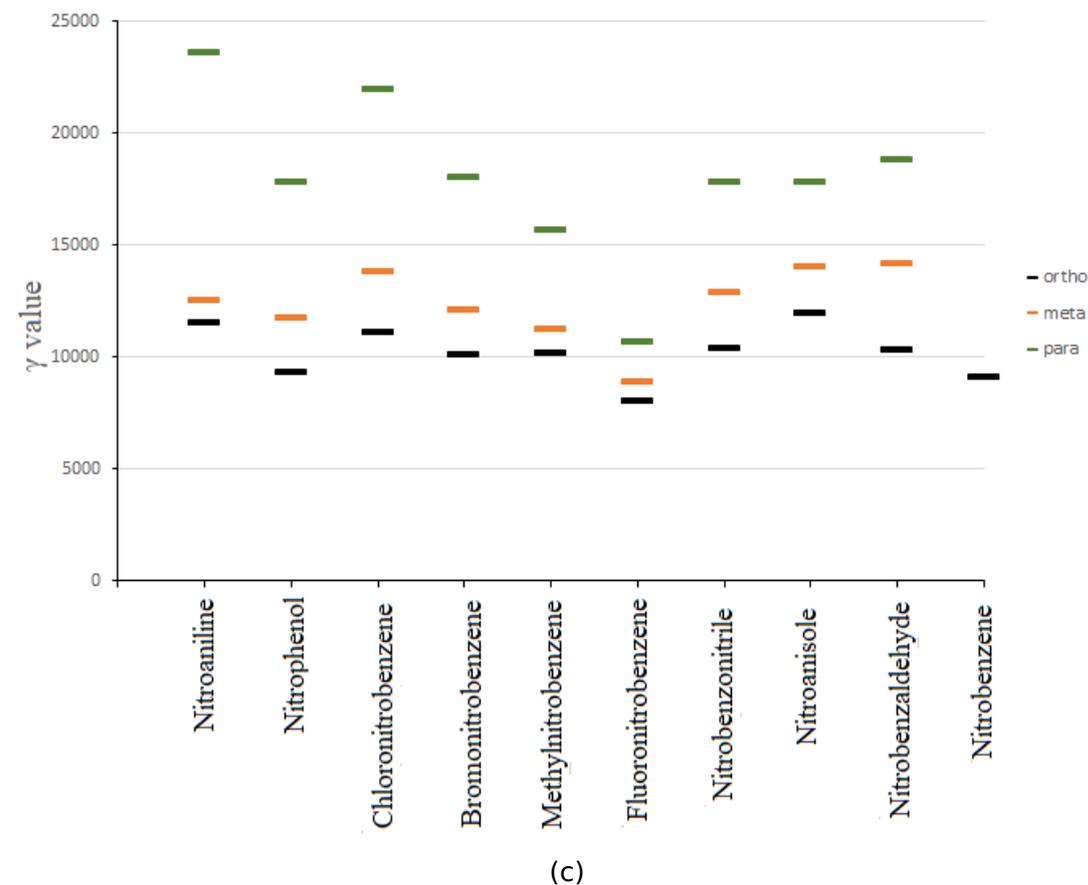


Figure 5 The value of (a) α , (b) β , (c) γ and (d) COs area of ortho, meta and para-substituted nitrobenzenes calculated using PM6 .

CONCLUSION

The work we have used linear polarizability, first and second order hyperpolarizability and conductor-like screening model (COs) area as molecular descriptors for QSAR of nitrobenzene derivative. In this work we use stepwise regression which fit the suitable variables in QSAR model. Second order hyperpolarizability shows the best QSAR model with the value of $R^2 = 89.493\%$, $r^2 = 68.7\%$ and $r_{cv}^2 = 87.52\%$. We also found that the substitutional of functional group in the nitrobenzene derivative for second order hyperpolarizability has the same sequence which was the γ ortho < γ meta < γ para. These has made that the second order hyperpolarizability was the best descriptors for QSAR model.

CONFLICT OF INTEREST

No conflict of interest in this work.

ACKNOWLEDGMENTS

The authors would like to thank Dr James J. P. Stewart from MOPAC Inc. for his permission to use the MOPAC software, and UiTM's Department of Infostructure for the MiniTab software usage permission. Additionally, the authors wish to express gratitude to Universiti Teknologi MARA for financial assistance.

REFERENCES

1. Ayub R, Raheel A. High-Value Chemicals from Electrocatalytic Depolymerization of Lignin: Challenges and Opportunities. *IJMS*. 2022 Mar 29;23(7):3767. [<DOI>](#).
2. Hagiopol C. Chapter 3: The tree fractionation: the extraction of natural polyphenols. CHENG K, Hagiopol C, editors. S.l.: ELSEVIER; 2021. 33-84 p. ISBN: 978-0-12-822205-8.
3. Manickam P, Vijay D. Chapter 2 - Chemical hazards in textiles. In: Muthu S, editor. *Chemical Management in Textiles and Fashion*. Woodhead Publishing; 2021. p. 19-52.
4. Oladele JO, Oyeleke OM, Akindolie BO, Olowookere BD, Oladele OT. Vernonia amygdalina Leaf Extract Abates Oxidative Hepatic Damage and Inflammation Associated with Nitrobenzene in Rats. *Jordan Journal of Biological Sciences*. 2021;14(3):463-9.
5. Siqueira Soldaini Oliveira R, Borges I. Correlation Between Molecular Charge Properties and Impact Sensitivity of Explosives: Nitrobenzene Derivatives. *Prop, Explos, Pyrotech*. 2021 Feb;46(2):309-21. [<DOI>](#).
6. Xavier JA, Silva TL, Torres-Santos EC, de Vasconcelos CC, Boane A, dos Santos RA, et al. Unveiling the relevance of the redox character of nitroaromatic and nitroheteroaromatic compounds as potential medicines. *Current Opinion in Electrochemistry*. 2021 Oct;29:100740. [<DOI>](#).
7. Adhikari C, Mishra B kumar. Quantitative Structure-Activity Relationships of Aquatic Narcosis: A Review. *CAD*. 2018 Mar 21;14(1):7-28. [<DOI>](#).
8. Kadam VV, Balakrishnan RM, Ponnann Ettiappan J, Thomas NS, D Souza SA, Parappan S. Sensing of p-nitrophenol in aqueous solution using zinc oxide quantum dots coated with APTES. *Environmental Nanotechnology, Monitoring & Management*. 2021 Dec;16:100474. [<DOI>](#).
9. Nepali K, Lee HY, Liou JP. Nitro-Group-Containing Drugs. *J Med Chem*. 2019 Mar 28;62(6):2851-93. [<DOI>](#).
10. Ji Z, Ji Y, Ding R, Lin L, Li B, Zhang X. DNA-templated silver nanoclusters as an efficient catalyst for reduction of nitrobenzene derivatives: a systematic study. *Nanotechnology*. 2021 Feb 19;32(19):195705. [<DOI>](#).
11. Rajak SK, others. QSAR study in terms of conceptual density functional theory based descriptors in predicting toxicity of nitrobenzenes towards *Tetrahymena pyriformis*. *Indian Journal of Chemical Technology (IJCT)*. 2022;28(4):467-72.
12. Bilal M, Bagheri AR, Bhatt P, Chen S. Environmental occurrence, toxicity concerns, and remediation of recalcitrant nitroaromatic compounds. *Journal of Environmental Management*. 2021 Aug;291:112685. [<DOI>](#).
13. Shaker B, Ahmad S, Lee J, Jung C, Na D. In silico methods and tools for drug discovery. *Computers in Biology and Medicine*. 2021 Oct;137:104851. [<DOI>](#).
14. Wang L, Ding J, Pan L, Cao D, Jiang H, Ding X. Quantum chemical descriptors in quantitative structure-activity relationship models and their applications. *Chemometrics and Intelligent Laboratory Systems*. 2021 Oct;217:104384. [<DOI>](#).
15. Dahmani R, Manachou M, Belaidi S, Chtita S, Boughdiri S. Structural characterization and QSAR modeling of 1,2,4-triazole derivatives as α -glucosidase inhibitors. *New J Chem*. 2021;45(3):1253-61. [<DOI>](#).
16. Alias AN, Zabidi ZM, Zakaria NA, Mahmud ZS, Ali R. Biological Activity Relationship of Cyclic and

- Noncyclic Alkanes Using Quantum Molecular Descriptors. *OJAppS*. 2021;11(08):966-84. [<DOI>](#).
17. Karelson M, Lobanov VS, Katritzky AR. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem Rev*. 1996 Jan 1;96(3):1027-44. [<DOI>](#).
18. Wady A, Khalid M, Alotaibi M, Ahmed Y. Synthesis, characterization, DFT calculations, and catalytic epoxidation of two oxovanadium(IV) Schiff base complexes. *Journal of the Turkish Chemical Society Section A: Chemistry*. 2022 Jan 17;9(1):163-208. [<DOI>](#).
19. Yunusa U, Umar U, Idriss S, Ibrahim A, Abdullahi T. Experimental and DFT Computational Insights on the Adsorption of Selected Pharmaceuticals of Emerging Concern from Water Systems onto Magnetically Modified Biochar. *Journal of the Turkish Chemical Society Section A: Chemistry*. 2021 Nov 11;8(4):1179-96. [<DOI>](#).
20. Giesekeing RLM. A new release of MOPAC incorporating the INDO /S semiempirical model with CI excited states. *J Comput Chem*. 2021 Feb 15;42(5):365-78. [<DOI>](#).
21. Wang S, Zhang X, Gui B, Xu X, Su L, Zhao YH, et al. Comparison of modes of action between fish, cell and mitochondrial toxicity based on toxicity correlation, excess toxicity and QSAR for class-based compounds. *Toxicology*. 2022 Mar;470:153155. [<DOI>](#).
22. Andini S, Araya-Cloutier C, Lay B, Vreeke G, Hageman J, Vincken JP. QSAR-based physicochemical properties of isothiocyanate antimicrobials against gram-negative and gram-positive bacteria. *LWT*. 2021 Jun;144:111222. [<DOI>](#).
23. Kurtz HA, Stewart JJP, Dieter KM. Calculation of the nonlinear optical properties of molecules. *J Comput Chem*. 1990 Jan;11(1):82-7. [<DOI>](#).
24. Klamt A. The COSMO and COSMO-RS solvation models. *WIREs Comput Mol Sci*. 2011 Sep;1(5):699-709. [<DOI>](#).
25. Hostaš J, Řezáč J, Hobza P. On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions. *Chemical Physics Letters*. 2013 May;568-569:161-6. [<DOI>](#).
26. Stewart J. MOPAC2016. Colorado Springs, CO; 2016.
27. Fatemi MH, Malekzadeh H. Prediction of Log(IGC 50) -1 for Benzene Derivatives to Ciliate *Tetrahymena pyriformis* from Their Molecular Descriptors. *BCSJ*. 2010 Mar 15;83(3):233-45. [<DOI>](#).
28. Osaghi B, Safa F. QSPR study on the boiling points of aliphatic esters using the atom-type-based AI topological indices. *Rev Roum Chim*. 2019;64(2):183-9.
29. Mustapha A, Shallangwa G, Ibrahim MT, Bello AU, Ebuka DA, Uzairu A, et al. QSAR studies on some C14-urea tetrandrine compounds as potent anti-cancer against Leukemia cell line (K562). *Journal of the Turkish Chemical Society, Section A: Chemistry*. 2018 Dec 25;5(3):1391-402. [<DOI>](#).
30. Majumdar S, Basak SC. Editorial: Beware of Naïve q2, use True q2: Some Comments on QSAR Model Building and Cross Validation. *CAD*. 2018 Mar 21;14(1):5-6. [<DOI>](#).
31. Rakhimbekova A, Akhmetshin TN, Minibaeva GI, Nugmanov RI, Gimadiev TR, Madzhidov TI, et al. Cross-validation strategies in QSPR modelling of chemical reactions. *SAR and QSAR in Environmental Research*. 2021 Mar 4;32(3):207-19. [<DOI>](#).
32. Toma C, Manganaro A, Raitano G, Marzo M, Gadaleta D, Baderna D, et al. QSAR Models for Human Carcinogenicity: An Assessment Based on Oral and Inhalation Slope Factors. *Molecules*. 2020 Dec 29;26(1):127. [<DOI>](#).
33. Baumann D, Baumann K. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *J Cheminform*. 2014 Dec;6(1):47. [<DOI>](#).
34. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR models-strategies and importance. *Int J Drug Des Discov*. 2011;3:511-9.
35. Goel H, Yu W, Ustach VD, Aytenfisu AH, Sun D, MacKerell AD. Impact of electronic polarizability on protein-functional group interactions. *Phys Chem Chem Phys*. 2020;22(13):6848-60. [<DOI>](#).
36. Tandon H, Ranjan P, Chakraborty T, Suhag V. Polarizability: a promising descriptor to study chemical-biological interactions. *Mol Divers*. 2021 Feb;25(1):249-62. [<DOI>](#).
37. Zhu T, Chen W, Singh RP, Cui Y. Versatile in silico modeling of partition coefficients of organic compounds in polydimethylsiloxane using linear and nonlinear methods. *Journal of Hazardous Materials*. 2020 Nov;399:123012. [<DOI>](#).
38. Zyss J. Hyperpolarizabilities of substituted conjugated molecules. II. Substituent effects and respective σ - π contributions. *The Journal of Chemical Physics*. 1979 Apr;70(7):3341-9. [<DOI>](#).

QSAR Studies on Nitrobenzene Derivatives using Hyperpolarizability and Conductor-like Screening Model as Molecular Descriptors

A. N. Alias* , Z. M. Zabidi 

Supplementary Material

Name	β_{xxx}	β_{xyy}	β_{xzz}	β_{yxx}	β_{yyy}	β_{yzz}	β_{zxx}	β_{zyy}	β_{zzz}
3-nitrobenzaldehyde	-107.406	-8.13873	6.95055	-194.685	220.2986	4.03478	-0.02997	-0.09594	0.02021
3-Nitroaniline	-277.812	-266.894	2.33934	143.1703	265.0084	-4.05664	0.28304	0.05348	0.0436
2-Nitroaniline	-442.042	237.1784	-0.52997	154.073	174.0243	2.22041	-0.3837	-0.6019	-0.00824
4-NITROBENZAMIDE	164.2175	33.43775	3.63387	-19.7175	-12.1311	13.45701	-0.02028	0.00584	0.1242
4-Nitrobenzyl alcohol	576.203	-119.842	1.09036	48.86409	-57.7258	-1.51291	-4.62689	-0.50518	-5.99972
Methyl 4-nitrobenzoate	157.5613	14.80136	9.99508	-27.0893	43.80339	17.92242	-0.04126	0.27467	0.24997
4-nitrobenzaldehyde	20.61401	-86.3345	10.05155	-19.8487	-26.9678	-2.31037	-0.06159	-0.00429	-0.00595
3-Nitroanisole	322.8077	161.1222	6.96617	81.23778	91.75077	-14.115	-0.42497	0.23712	0.07669
2-nitrobenzaldehyde	13.05212	-24.4508	-6.92444	4.71585	-117.046	5.02317	-0.25362	-0.50811	-0.01777
4-Fluoronitrobenzene	409.4995	-155.675	1.92477	0.00661	-0.09222	-0.00478	0.13113	0.04254	0.00252
3-Nitroacetophenone	-44.2635	-5.72725	2.12282	-210.932	224.3482	10.72745	0.02065	0.0544	-0.00386
4-Nitrophenetole	1618.209	-278.852	5.22649	-135.926	17.04215	-7.79467	0.86974	-0.16206	0.18803
4-Nitroanisole	-1081.34	246.6922	-9.89709	-10.7317	-24.6004	-2.15571	-0.56548	-0.25527	0.03636
4-Nitrobenzyl chloride	348.5093	-155.854	25.27714	-0.74771	0.42059	0.1246	-38.5231	16.44961	33.71734
4-Ethylnitrobenzene	645.8094	-167.319	11.35641	10.68836	-6.49552	-0.08176	-16.753	11.32412	2.18177
2-Nitrobiphenyl	-153.951	15.49075	-7.68834	10.40523	110.5113	41.5736	-4.5394	-12.6303	10.59206
5-Hydroxy-2-nitrobenzaldehyde	-732.208	147.4603	0.38729	-291.476	205.6326	-10.0762	-0.12093	-0.17876	-0.00092
6-Chloro-2,4-dinitroaniline	-1024.71	372.4929	-3.08485	397.0051	201.6174	-9.91456	0.61227	-0.341	0.14031
3-Nitrobenzotrile	40.1177	2.08986	-17.3495	-172.788	169.8454	7.87914	-0.12241	0.00357	-0.00777
4-Nitrobenzotrile	149.9968	-126.313	20.69014	0.04503	0.20171	0.00431	0.00016	0.00033	-0.00001
2-Amino-4-chloro-5-nitrophenol*	-1096.98	-161.256	3.10006	-870.765	231.7681	-5.38898	31.29405	15.44982	5.08301
2,3-Dinitrophenol *	-167.69	116.9766	-46.6554	-191.413	-288.537	-68.9416	37.64876	11.50439	6.6923

Name	β_{xxx}	β_{xyy}	β_{xzz}	β_{yxx}	β_{yyy}	β_{yzz}	β_{zxx}	β_{zyy}	β_{zzz}
3-Nitrophenol *	233.4414	109.0344	6.17516	44.31182	159.3321	-0.52827	-0.52827	0.01679	-0.00208
2,6-Dinitrophenol *	-184.652	-48.288	-4.23249	274.0526	-401.894	-5.62201	0.08114	0.24699	-0.00921
4-Methyl-2-nitrophenol *	21.19537	-7.77141	-11.522	-204.721	119.7415	-6.07916	-0.00979	0.00271	0.00663
2-Nitrophenol*	180.5037	-215.828	-1.38299	-73.9181	-93.1267	-4.23917	-0.03343	-0.02707	0.00405
2-Chloromethyl-4-nitrophenol*	-927.516	-78.3501	-6.02087	-485.373	221.0804	-20.1294	-0.11388	-0.15094	0.08312
2,5-Dinitrophenol *	456.3495	-157.247	3.65367	49.96321	116.1203	12.25693	0.43926	0.08193	-0.02202
2-Nitroresorcinol*	-479.639	288.2737	4.99292	17.79195	-45.3904	-6.34166	0.0459	0.03576	-0.00274
3-Nitro-2-xylene	-259.731	9.47801	-2.34206	-82.7347	205.6168	-10.2414	-0.02697	-0.02249	0.02534
2,6-dimethylnitrobenzene	-69.3849	154.0333	31.85578	1.80146	-2.36372	0.41409	1.33912	0.49623	-0.84623
2,3-dimethylnitrobenzene	-210.093	-0.08592	9.32443	-62.4941	196.671	-9.34246	-11.763	1.83536	-0.80275
2-methyl-3-chloronitrobenzene	-294.951	-89.6077	12.57144	-5.31444	230.5288	-2.03959	-26.6933	3.74238	-0.38834
2-methylnitrobenzene	-280.128	154.0435	-7.43852	43.10892	46.09325	-1.67936	-0.12252	0.39051	-0.27098
2-chloronitrobenzene	-16.2558	191.1657	14.84702	140.4117	175.9841	0.39656	6.2425	-41.7044	0.15244
2-methyl-5-chloronitrobenzene	-252.408	-134.867	-2.99171	64.57506	-119.748	2.66624	9.63111	-6.45379	-1.02694
2,4,5-trichloronitrobenzene	-519.396	-121.814	15.63978	407.1055	-78.3658	-0.59776	10.00321	-34.0882	0.3478
2,5-dichloronitrobenzene	-47.779	-43.399	3.07005	-85.1452	-11.6088	-3.29785	0.65462	-9.21405	-0.10206
6-chloro-1,3-dinitrobenzene	776.1823	-76.3143	20.55704	-316.231	167.4616	7.78888	13.83512	24.09994	0.46451
nitrobenzene	-227.147	107.65	-2.70756	-0.00286	-0.25499	-0.0021	0.01385	-0.02581	-0.00092
3-methylnitrobenzene	-207.598	-13.2319	5.58666	-61.3692	109.5199	-3.35465	0.05627	-0.04426	0.02192
1,3-dinitrobenzene	-0.80821	-0.28553	0.00261	-188.243	222.5985	-1.76682	-0.64097	0.04371	-0.00515
3,4-dichloronitrobenzene	-958.72	-75.853	-3.4705	-182.164	304.0829	-1.56579	0.10785	-0.01365	0.00143
4-methylnitrobenzene	662.188	-180.748	-2.77864	-10.731	-0.06797	3.95765	-0.18072	-0.08166	0.01305
1,4-dinitrobenzene	0.22225	-0.02254	0.00351	0.04947	0.0793	-0.00089	0.0239	0.01609	-0.00304
4-chloronitrobenzene	1024.197	-154.696	4.9467	-0.13816	-0.11031	-0.0016	0.09985	-0.00281	0.00781
2,3,5,6-tetrachloronitrobenzene	0.32491	-0.86776	0.01726	-38.0493	-172.688	1.37661	0.12494	-0.03103	-0.01338
6-methyl-1,3-dinitrobenzene	-275.782	87.76645	0.83589	-338.917	220.994	2.40858	3.00765	-1.02092	-0.48669
3-chloronitrobenzene	293.9259	100.183	1.05459	28.32126	157.9507	-1.33774	0.18775	-0.05016	0.00803
2-bromonitrobenzene	221.647	-65.817	-4.45118	57.10785	-105.418	-3.88547	-0.15725	0.17946	0.0017
3-bromonitrobenzene	112.2158	77.32627	-3.02739	-76.5724	112.0591	-3.18281	0.0472	-0.04491	0.00029

Name	β_{xxx}	β_{xyy}	β_{xzz}	β_{yxx}	β_{yyy}	β_{yzz}	β_{zxx}	β_{zyy}	β_{zzz}
4-bromonitrobenzene	636.1428	-131.777	-1.34632	0.14062	-0.41076	-0.00084	0.02408	0.02996	0.00119
2,4,6-trimethylnitrobenzene	487.2197	-240.514	-41.4805	-9.0686	3.38915	4.35927	-1.02309	-0.50699	0.06539
5-methyl-1,2-dinitrobenzene	182.9819	-122.835	-65.9328	-132.798	-30.9037	-18.1088	-21.2979	10.37502	-2.78741
2,4-dichloronitrobenzene	922.6029	-275.532	4.68841	264.341	196.2572	5.52379	0.23118	0.32871	-0.00996
3,5-dichloronitrobenzene	171.2097	255.2306	2.14095	1.95536	-3.56673	-0.04293	0.02372	-0.07608	-0.00909
6-iodo-1,3-dinitrobenzene	459.143	-162.182	-0.25996	-9.86538	65.73475	-3.92581	-0.07814	0.41036	0.01076
2,3,4,5-tetrachloronitrobenzene	699.5292	78.39559	1.69574	367.8519	230.0996	5.559	-0.12905	0.3356	-0.01229
2,3-dichloronitrobenzene	330.074	121.9904	1.76158	34.03553	465.7443	1.31138	-0.01351	0.12179	0.01527
2,5-dibromonitrobenzene	-62.8901	74.86594	-0.51253	39.59764	-4.23922	1.76134	-0.13096	0.09024	0.00876
1,2-dichloro-4,5-dinitrobenzene	534.0062	289.3468	-0.0491	0.90223	-0.55667	-0.0491	-0.83199	-0.41753	-0.19002
3-methyl-4-bromonitrobenzene	575.4357	-52.5483	-8.91783	-29.4821	121.2631	-3.12227	0.05273	-0.03945	0.00855
2,3,4-trichloronitrobenzene	1011.229	-120.032	3.93728	-26.9388	592.7289	3.07087	-0.24504	0.1254	-0.0093
2,4,6-trichloronitrobenzene	976.7745	-450.077	5.18579	-0.49055	0.95889	-0.01732	-0.04162	-0.18462	0.00464
3,5-dinitrobenzyl alcohol	-234.518	94.46671	-3.67346	116.539	4.2354	-2.2517	-5.65075	16.67305	-5.36324
3,4-dinitrobenzyl alcohol	-136.287	77.06394	66.54098	56.60117	27.28624	10.45835	-6.13512	4.7163	2.21966
1,2-dinitrobenzene	-134.112	-98.0993	-65.3709	-28.3669	20.48465	-2.12874	-0.68789	-3.04461	-2.91564
2,4,6-trichloro-1,3-dinitrobenzene	4.59865	3.75504	0.63119	-678.306	485.1152	33.27589	12.18542	-1.818	-6.60903
2,3,5,6-tetrachloro-1,4-dinitrobenzene	-0.16924	-0.04236	0.00414	0.26336	0.00638	-0.01492	-0.09023	0.1211	-0.00218
2,4,5-trichloro-1,3-dinitrobenzene	691.7179	-6.27857	7.33992	310.2391	-107.273	44.42341	12.45873	-15.707	-1.48189

