# CORROSION INHIBITION OF Q235A STEEL IN ACID MEDIUM USING ISATIN DERIVATIVES: A QSAR STUDY

## (Perencatan Kakisan Keluli Q235A di dalam Medium Asid Menggunakan Terbitan Isatin: Satu Kajian QSAR)

Abdo M. Al-Fakih[1,2], Madzlan Aziz[1]*, Hassan H. Abdallah[3], Hasmerya Maarof[1], Rosmahaida Jamaludin[1], Bishir Usman[1]

[1]*Department of Chemistry, Faculty of Science,*
*University Technology Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*
[2]*Department of Chemistry, Faculty of Science,*
*Sana'a University, Sana'a, Yemen*
[3]*Department of Chemistry, College of Education,*
*Salahaddin University, Erbil, Iraq*

*\*Corresponding author: madzlan@utm.my*

## Abstract

Quantitative Structure-Activity Relationship (QSAR) study was performed on 10 isatin derivatives which were reportedly used as corrosion inhibitors. Dragon software was used to calculate the molecular descriptors. Partial least square (PLS) method was used to run the regression analysis between the descriptors and the corrosion inhibition efficiencies (IE) of the inhibitors. A predictive QSAR model was developed with a correlation coefficient ($r^2_{cal}$) of 0.9676. The model validity was assessed through internal and external validation. The results show that cross-validation regression coefficient ($r^2_{cv}$) and prediction regression coefficient ($r^2_{pred}$) are 0.8163 and 0.9189, respectively. The model was used to predict the IE for ten isatin derivatives. The results confirm a good stability and predictive ability of the model. Dragon-based descriptors provide a very good description of the corrosion inhibition properties of the inhibitors. The results of the QSAR study were found to be consistent with the experimental data.

**Keywords**: QSAR modeling study, corrosion inhibition, isatin derivatives, partial least square

## Abstrak

Kajian hubungan aktiviti-struktur kuantitatif (QSAR) dilaksanakan ke atas 10 terbitan isatin yang dilaporkan diguna sebagai perencat kakisan. Perisian Dragon digunakan untuk mengira pemerihal molekul. Kaedah kuasa dua terkecil separa (PLS) digunakan untuk melaksanakan analisis regresi antara pemerihal dan keberkesanan perencatan kakisan (IE) terhadap perencat kakisan. Ramalan model QSAR dibangunkan dengan pekali kolerasi ($r^2_{cal}$) bernilai 0.9676. Kesahan model dinilai melalui pengesahan dalaman dan luaran. Keputusan menunjukkan bahawa pekali regresi kesahan-silang ($r^2_{cv}$) dan pekali regresi ramalan ($r^2_{pred}$) adalah masing-masingnya 0.8163 dan 9189. Model tersebut digunakan untuk meramal IE bagi 10 terbitan isatin. Keputusan mengesahkan kestabilan dan kemampuan ramalan yang baik terhadap model. Pemerihal berasaskan Dragon memberi penerangan yang baik tentang sifat perencatan kakisan bagi perencat-perencat kakisan. Keputusan kajian QSAR ini adalah konsisten dengan data eksperimen.

**Kata kunci**: kajian pemodelan QSAR, perencatan kakisan, terbitan isatin, kuasa dua terkecil separa

## Introduction

Steel is the main constructional material and has several types with variety of mechanical properties. It has extensive applications in industries because of its low cost and excellent mechanical properties [1-3]. However, it is prone to corrosion and raises a serious concern since it costs a huge loss in properties and safety problems [4]. Therefore, prevention of corrosion or at least reduction in its rate has been widely studied. Heterocyclic organic compounds containing nitrogen, sulfur, oxygen or phosphorous in their molecular structures are the most reported inhibitors of the steel corrosion especially in acidic corrosive media [5, 6]. The performance of the corrosion inhibitors is based on their adsorption onto metal surface. The adsorption of the inhibitors forms a protected film and reduces the corrosion rate [7]. Many experimental and theoretical studies have been done to study the efficiency of many corrosion inhibitors. However, experimental measurements are costly and time consuming. Therefore, theoretical methods have become more desirable since they may overcome disadvantages of running experiments [8, 9]. For example, quantitative structure activity relationship (QSAR) is a modeling technique which is applied in many disciplines of chemistry [10].

In QSAR modeling, molecular descriptors are calculated based on the molecular structures of the compounds. The traditional method is to calculate the descriptors using quantum chemical calculations. In addition, other techniques such as Molconn-Z, CODESSA and Dragon software are used to calculate molecular descriptors [11]. A number of 4885 molecular descriptors can be calculated using Dragon software [12]. Multiple linear regressions (MLR) and partial least square (PLS) are the most applied regression methods in QSAR modeling. Various variable selection methods such as stepwise regression (RS), genetic algorithms (GA), genetic algorithme-partial least square (GA-PLS) [13, 14], and interval partial linear square (IPLS) [15, 16], were proposed to specify the most relevant variables.

This study aims to develop a reliable QSAR model using data of newly synthesized isatin derivatives as corrosion inhibitors for Q235A steel in 3 M HCl. Apart from quantum chemical calculations, Dragon software was used in this work to calculate molecular descriptors based on the molecular structures.

## Materials and Methods

### Dataset

The dataset consists of ten isatin derivatives used as corrosion inhibitors of Q235A steel in acid medium (Table 1) [17]. The corrosion inhibition efficiency sorted in decreasing order from 70.4 to 15.9 %. The dataset was split into 70 % training set and 30 % test set.

Table 1.  Inhibition efficiencies and molecular structures of isatin derivatives collected from Ref. [17]
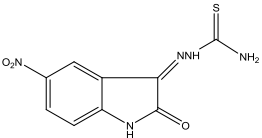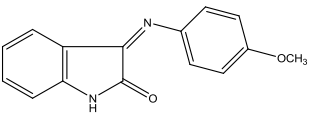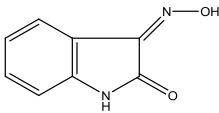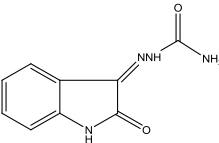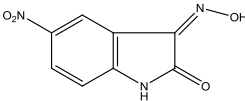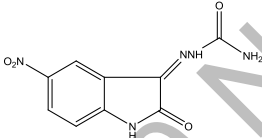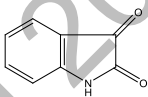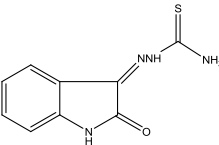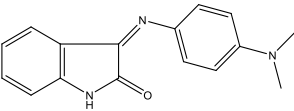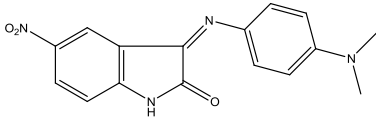
| No. | Structure | Inhibition efficiency, IE (%) |
|---|---|---|
| Training set | | |
| 1 |  | 67.7 |
| 2 |  | 66.8 |

Table 1 (cont'd).  Inhibition efficiencies and molecular structures of isatin derivatives collected from Ref. [17]

| No. | Structure | Inhibition efficiency, IE (%) |
|---|---|---|
| 3 |  | 61.9 |
| 4 |  | 56.6 |
| 5 |  | 35.7 |
| 6 |  | 23.5 |
| 7 |  | 15.9 |
| Test set | | |
| 8 |  | 70.4 |
| 9 |  | 61.0 |
| 10 |  | 18.3 |

**Calculation of descriptors and variable selection**

Prior to descriptors calculation, molecular structures of the inhibitors were drawn with the Chem3D software and optimized using the MM2 followed by the Molecular Orbital Package (MOPAC) module. The Dragon software Version 6.0 [12], was used to calculate the descriptors using the optimized structures. The total number of the calculated descriptors was 2754. A variable selection was applied. One of the highly correlated descriptors pairs with $R \geq 0.90$ was randomly excluded. Descriptors with at least one missing value were then excluded yielding a reduced set of 227. Then descriptors with more than 90% equal values or with zero values were manually eliminated yielding a final set of 165 descriptors. The final set of descriptors was subjected to variable selection using IPLS selection method. The IPLS method was used to select a subset of variables. The selected variables subset could produce reliable models with better prediction ability compared to the use of all dataset variables. The variable selection and models building were performed using PLS Toolbox 6.2 (Eigenvector Research Inc.) in Matlab 7.6.0.324 (R2008a).

**Statistical evaluation of the QSAR models**

The QSAR model was evaluated in terms of its stability and predictive ability using several statistical parameters. The statistical parameters included correlation coefficient ($r^2_{cal}$), cross-validation regression coefficient ($r^2_{cv}$), prediction regression coefficient ($r^2_{pred}$), root-mean-square error of calibration (RMSEC), root-mean-square error of cross-validation (RMSECV), and root-mean-square error of prediction (RMSEP).

<div align="center">

**Results and Discussion**

</div>

**IPLS-based model**

A set of three significant descriptors were selected from the final set (165 descriptors) using IPLS selection method. The descriptors were used to develop QSAR models. The selected descriptors are HATS3u, R3s and De. The HATS3u (Leverage-weighted autocorrelation of lag3/unweighted) descriptor is a type of GETAWAY descriptors. The R3s (R maximal autocorrelation of lag 3/weighted by I-state) descriptor is also one of GETAWAY descriptors. GETAWAY descriptors are 3D geometrical descriptors which calculated from atomic Cartesian coordinate's values. GETAWAY descriptors were designed to match the 3D-molecule geometry, relatedness of atom, and chemical information [18]. The De (D total accessibility index/weighted by Sanderson electronegativity) descriptor is one of WHIM (weighted holistic invariant molecular) descriptors. WHIM descriptors are molecular descriptors based on statistical indices calculated on the projections of the atoms along principal axes. WHIM descriptors are developed to extract 3D molecular information related to molecular size, shape, symmetry, and atom distribution with respect to invariant reference frames [19,20]. Table 2 shows the correlation matrix of the selected descriptors. It is clear that there is no inter-correlation between the descriptors. Several QSAR models were developed using PLS regression and the best model was selected based on statistical parameters. The best IPLS-based model for modeling the inhibition efficiency of isatin derivatives as a function of a set of three significant descriptors is presented by the Equation (1). The correlation coefficients $r^2_{cal}$, $r^2_{cv}$, and $r^2_{pred}$ of the model are 0.9676, 0.8163, and 0.9189, respectively. The values of root-mean-square errors RMSEC, RMSECV, and RMSEP are lower than that of other developed models by the same method. This is an indication of higher prediction ability of the proposed model.

$$IE \% = 61.3368 + 63.4336(De) + 112.0971(HATS3u) - 14.7415 (R3s) \tag{1}$$

whereas $r^2_{cal} = 0.9676$, $r^2_{cv} = 0.8163$, $r^2_{pred} = 0.9189$, RMSEC = 3.5874, RMSECV = 8.8106 and RMSEP = 18.3132

<div align="center">

Table 2. Inter-correlation matrix of descriptors presents the model

</div>

| Descriptor | De | HATS3u | R3s |
|------------|------|--------|-----|
| De | 1 | | |
| HATS3u | -0.249 | 1 | |
| R3s | -0.289 | 0.230 | 1 |

3

**Outliers detection**

Detecting and then removing outliers is to enhance the prediction ability of the model. The possibility of the existence of outliers in the dataset was studied by plotting of standardized residuals against predicted values (Figure 1). A compound was considered as an outlier when the standardized residual was greater than 2 or less than -2 [21]. Figure 1 shows the outlier plot of the model. Only one compound has standardized residual higher than 2 which was excluded as an outlier.
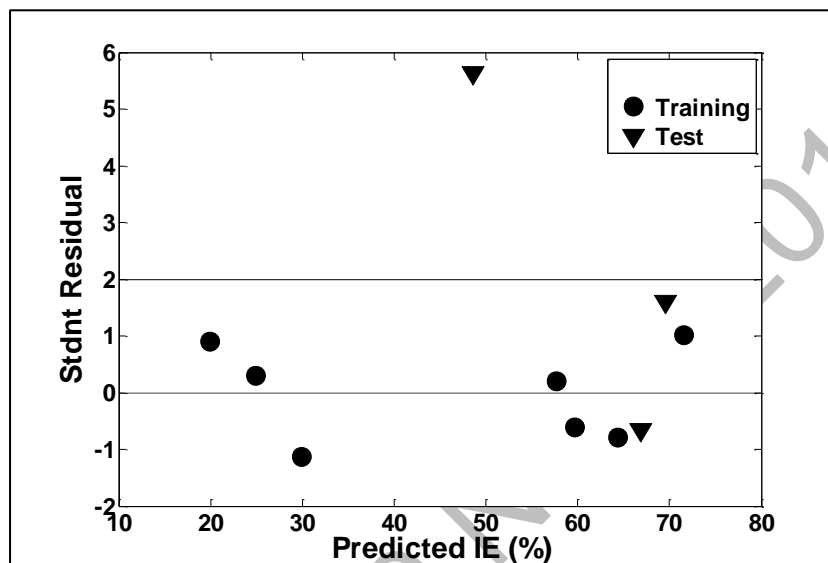


Figure 1.  Plot of standardized residual versus predicted values

**Model interpretation**

Significant descriptors (i.e. HATS3u, R3s and De) were found to be useful to build reliable QSAR models. The proposed QSAR model was used to predict the corrosion inhibition efficiency (IE %) of the dataset compounds. Figure 2 displays a plot of predicted against the experimental values. The plot shows that the QSAR model has a high correlation coefficient. This indicates a good correlation between the experimental and predicted values. According to the statistical parameters, the developed model has high $r^2_{cal}$ and $r^2_{cv}$ of 0.9676 and 0.8163, respectively. Furthermore, the values of RMSEC, RMSECV are lower than that of other developed models by the same method. The result indicates high stability and predictive ability of the proposed model.

**Model validation**

The stability and predictive power of the model were validated using both internal and external validation. Internal validation was carried out using cross-validation leave-one-out. Internal validation provides internal check on the model. The cross-validation regression coefficient $r^2_{cv}$ is 0.8163. The external validation was conducted by the prediction of test set efficiencies. The predicted efficiencies are in a good agreement with the experimental values. The prediction regression coefficient $r^2_{pred}$ is 0.9189. The results indicate a high ability of the model to predict corrosion inhibition efficiency of new structurally relevant compounds. Further validation was performed to check the robustness of the model using a Y-randomization test. The training set was shuffled several times. The PLS regression was performed to each randomized dataset. The obtained values of $r^2_{cal}$ ranged from 0.0418 to 0.2896. The low $r^2_{cal}$ values indicate that the model is not due to a chance correlation.
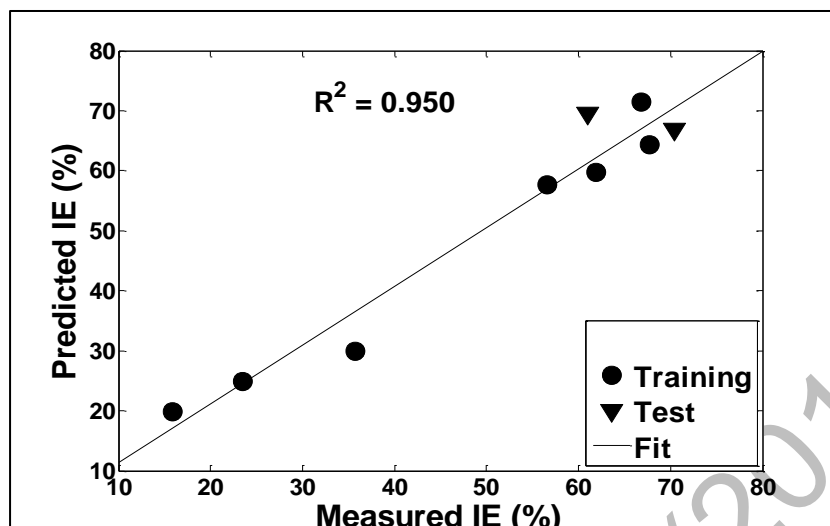
Figure 2. Plot of predicted and measured IE (%)

## Conclusion

Dragon software was used to calculate the molecular descriptors. A QSAR model was built using PLS regression method. According to the statistical assessment and validation results, the proposed model reveals high stability and predictive power. Three significant descriptors (i.e. R3s, HATS3u and De) were used to develop a QSAR model. The descriptors contributed significantly to describe the corrosion inhibition properties of the inhibitors. Therefore, the use of Dragon software provides more opportunities to enrich the corrosion area with new descriptors. The computational methods used in this study are useful for developing predictive QSAR models with potent ability to predict and design novel and potential corrosion inhibitors.

## References

1. Abdallah, M., Helal, E. A. and Fouda, A. S. (2006). Aminopyrimidine derivatives as inhibitors for corrosion of 1018 carbon steel in nitric acid solution. *Corrosion Science,* 48 (7): 1639 - 1654.
2. Badr, G. E. (2009). The role of some thiosemicarbazide derivatives as corrosion inhibitors for C-steel in acidic media. *Corrosion Science,* 51 (11): 2529 – 2536.
3. Pavithra, M. K., Venkatesha, T.V., Punith Kumar, M. K. and Tondan, H. C. (2012). Inhibition of mild steel corrosion by Rabeprazole sulfide. *Corrosion Science,* 60: 104 – 111.
4. Musa, M., Rodhi, M. N. M., Najmiddin Yaakob, N., Hamid, K.H.K. and Idris. J. (2013). Development of bio-based paint by using methyl esters from palm oil for corrosion inhibitor. *Malaysian Journal of Analytical Sciences*, 17 (1): 30 – 37.
5. Raja, P. B., Qureshi, A. K., Abdul Rahim, A., Osman, H. and Awang, K. (2013). Neolamarckia cadamba alkaloids as eco-friendly corrosion inhibitors for mild steel in 1M HCl media. *Corrosion Science,* 69: 292 – 301.
6. Kamal, N. K. M., Fadzil, A. H., Kassim, K., Rashid, S. H. and Mastuli, M. S. (2014). Synthesis, characterization and corrosion inhibition studies of o,m,p-decanoyl thiourea derivatives on mild steel in 0.1 M $H_2SO_4$ solutions. *Malaysian Journal of Analytical Sciences,* 18 (1):21 – 27.

5

7.  Solmaz, R. (2014). Investigation of adsorption and corrosion inhibition of mild steel in hydrochloric acid solution by 5-(4-Dimethylaminobenzylidene)rhodanine. *Corrosion Science*, 79: 169 – 176.
8.  Zhao, H., Zhang, X., Ji, L., Hu, H. and Li, Q. (2014). Quantitative structure–activity relationship model for amino acids as corrosion inhibitors based on the support vector machine and molecular design. *Corrosion Science*, 83: 261 – 271.
9.  Zhang, J., Liu, J., Yu, W., Yan, Y., You, L. and Liu, L. (2010). Molecular modeling of the inhibition mechanism of 1-(2-aminoethyl)-2-alkyl-imidazoline, *Corrosion Science,* 52 (6): 2059 – 2065.
10. Khaled, K.F. (2011). Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model. *Corrosion Science,* 53 (11): 3457 – 3465.
11. Bababdani, B.M. and Mousavi, M. (2013). Gravitational search algorithm: A new feature selection method for QSAR study of anticancer potency of imidazo[4,5-b]pyridine derivatives. *Chemometrics and Intelligent Laboratory Systems,* 122: 1 – 11.
12. Talete srl. (2010), Dragon (software for molecular descriptor calculation), version 6.0. Milano, Italy: Talete. http://www.talete.mi.it/
13. Riahi, S., Ganjali, M.R., Norouzi, P. and Jafari, F. (2008). Application of GA-MLR, GA-PLS and the DFT quantum mechanical (QM) calculations for the prediction of the selectivity coefficients of a histamine-selective electrode. *Sensors and Actuators B: Chemical,* 132 (1): 13 – 19.
14. Soltani, S., Abolhasani, H., Zarghi, A. and Jouyban, A. (2010). QSAR analysis of diaryl COX-2 inhibitors: Comparison of feature selection and train-test data selection methods. *European Journal of Medicinal Chemistry,* 45 (7): 2753 – 2760.
15. Chen, Q., Jiang, P. and Zhao, J. (2010). Measurement of total flavone content in snow lotus (Saussurea involucrate) using near infrared spectroscopy combined with interval PLS and genetic algorithm. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 76 (1): 50 – 55.
16. Abrahamsson, C., Johansson, J., Sparén, A. and Lindgren, F. (2003). Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets. *Chemometrics and Intelligent Laboratory Systems,* 69: 3 – 12.
17. Chen, G., Su, H.-J., Song, Y.-P., Gao, Y., Zhang, J., Hao, X.-J. and Zhao, J.-R. (2013). Synthesis and evaluation of isatin derivatives as corrosion inhibitors for Q235A steel in highly concentrated HCl. *Research on Chemical Intermediates,* 39 (8): 3669 – 3678.
18. Consonni, V., Todeschini, R., Pavan, M. and Gramatica, P. (2002). Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. *Journal of Chemical Information and Computer Sciences,* 42(3): 693 – 705.
19. Pourbasheer, E., Aalizadeh, R., Ganjali, M. R., Norouzi, P. and Shadmanesh, J. (2014). QSAR study of ACK1 inhibitors by genetic algorithm–multiple linear regression (GA–MLR). *Journal of Saudi Chemical Society,* 18(5): 681 – 688.
20. Todeschini, R., Bettiol, C., Giurin, G., Gramatica, P., Miana, P. and Argese, E. (1996). Modeling and prediction by using WHIM descriptors in QSAR studies: submitochondrial particles (SMP) as toxicity blosensors of chlorophenols. *Chemosphere,* 33(1): 71 –79.
21. Melagraki, G. and Afantitis, A. (2013). Enalos KNIME nodes: Exploring corrosion inhibition of steel in acidic medium. *Chemometrics and Intelligent Laboratory Systems,* 123: 9 – 14.