



Prediction of inhibition performance of some benzimidazole derivatives against steel corrosion through QSAR and molecular dynamic simulation

A. U. Bello*, A. Uzairu, and G. A. Shallangwa

Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria

Received 22 Jan 2019,
Revised 12 April 2019,
Accepted 13 April 2019

Keywords

- ✓ Benzimidazole derivatives,
- ✓ DFT,
- ✓ GFA,
- ✓ QSAR,
- ✓ MD simulation.

abdallahbum@yahoo.com;
Phone: +2348033473438;
Fax: (+603) 6196 4053.

Abstract

The corrosion inhibition performance of benzimidazole derivatives against steel corrosion has been investigated by the quantum chemical calculation, QSAR analysis, and molecular dynamics (MD) simulation methods. Density functional theory (B3LYP/6-31G*) quantum chemical parameters such as energy of the highest occupied molecular orbital (E-HOMO), the energy of the lowest unoccupied molecular orbital (E-LUMO), energy band gap ΔE , Dipole Moment, electrophilicity index (ω), chemical softness (σ), chemical hardness (η) and fraction of electron transfer from the inhibitors molecule to the metallic surface (ΔN) have been calculated and well discussed. A linear quantitative structure-activity relationship (QSAR) model was built by multiple linear regression (MLR) method between the computed descriptors and the experimental corrosion inhibition efficiencies. The prediction of corrosion efficiencies of these inhibitors nicely matched the experimental measurements. The validation parameters are the squared correlation coefficients (R^2) of 0.942, adjusted squared correlation coefficients (R^2_{adj}) value of 0.908, cross-validation coefficients (Q^2_{cv}) value of 0.795 and the external validations (R^2_{test}) of 0.983 respectively. These indicate that the generated models were excellent for verifying with internal and external validation parameters. Molecular dynamics (MD) simulation was also applied to search the best adsorption configuration of inhibitor over Fe (1 1 0) surface. It was found that adsorption of the inhibitor molecules on the steel surface mainly occurred by a chemical adsorption phenomenon. Thus, Quantum chemical studies, QSAR analysis along with MD simulation may be a very powerful tool for the rational designing of several promising corrosion inhibitors.

1. Introduction

Corrosion is an unwanted electrochemical process that affects several areas of industrial activity, especially the oil industry, resulting in huge economic losses [1]. The annual cost of corrosion worldwide was estimated at \$ 2.2 Trillion in 2010, which was about 3 % of the world's gross domestic product of \$ 73.33 Trillion [2, 3]. Preventing the corrosion of steel has played an important role in various industries, especially in the chemical and petrochemical processing industries that employ the use of steel [4]. Although it is not possible to completely avoid the corrosion process there are several ways to prevent it or slow down the corrosion rate [5]. Several organic compounds especially those that contain N, O, S, and P heteroatoms, as well as π -electron systems have been previously used as corrosion inhibitors for metals in aqueous solutions [6, 7]. Azoles and particularly their derivatives are known as efficient corrosion inhibitors [8]. For example, some derivatives of benzimidazole have been demonstrated as excellent inhibitors for metals and alloys in acidic solution, and exhibit different inhibition performance with the difference in substituent groups and substituent positions on the imidazole ring [9-11]. Several experimental techniques have been used to study the corrosion and corrosion inhibition in an acid solution such as weight-loss method, potentiodynamic polarization, electrochemical impedance spectroscopy (EIS), and so on [12, 13].

However, experiments are usually time-consuming, expensive, and deficient in elucidating the inhibition mechanism of the system at the Sub-atomic and molecular levels [14]. Therefore, with the advances in computational chemistry and the development of new algorithms, theoretical methods have been used to investigate corrosion inhibition mechanism and was used to design new and environment-friendly corrosion inhibitors [15, 16]. Quantitative structure-activity relationships (QSAR) has been a subject of intense interest in the field of medicinal chemistry in determining the molecular structure as well as elucidating the electronic structure and reactivity [17], but to a less extent in the field of corrosion [18, 19]. However, in recent years, a quantitative structure-activity relationship (QSAR) has aroused many researchers' interest in the studies of corrosion inhibitors [7]. The success of the QSAR approach can be explained by the insights offered for the structural determination of chemical properties, and the possibility of estimating the properties of the new chemical compounds without any need for them to be synthesized and tested [20].

Furthermore, Molecular dynamics (MD) simulation was recently considered as a modern tool to study the adsorption behavior of the inhibitor molecules on the metallic surfaces of interest [14]. These two methods have been proved to be very helpful for predicting the inhibition efficiencies of novel corrosion inhibitors and actual interfacial configuration and adsorption energies of the surface adsorbed inhibitor molecules. This study was aimed to develop QSAR model based on the compounds collected from the literature which can be used to predict the inhibition efficiency of known and hypothetical molecules against steel corrosion and to carry out MD simulation to understand the adsorption mode of the studied inhibitors.

2. Experimental

2.1 Hardware and Softwares

The hardware and software that was used include Dell Intel(R)Core(TM)i7-5500U CPU), 16.00GB RAM @ 2.400GHz 2.400GHz processor on Windows 8.1 Pro 64-bit Operating system, x64-based processor, ChemDraw ultra 12.0, Spartan 14v.1.1.0 software, Material studios 8.0, and Microsoft Office Excel 2013.

2.2 Data collection

Twenty benzimidazoles (BI) derivatives as potential steel corrosion inhibitors were collected from the literature [11,21-23] which was used for this present study. The chemical structure of each compound in the data sets was drawn with ChemDraw ultra V12.0. Their molecular structures and inhibition efficiencies are shown in Table 4.

2.3 Geometry Optimization

The 2D structure of each of the compounds was drawn using ChemDraw Ultra 12.0. They were converted to the 3D structure using Spartan 14. The structures were cleaned by minimizing and checking using a molecular mechanic force field (MM+) option on Spartan 14, so as to remove all strain from the structure of the molecule. Additionally, this will guarantee a well-defined and stable conformer relationship within the compounds in the study [24]. Geometry optimization was set at the ground state utilizing the density functional theory (DFT) at the Becke88 three-parameter hybrid exchange potentials with Lee-Yang-Parr correlation potential (B3LYP) level of theory and for the basis set 6-311G (d) was selected.

2.4 Calculation of Descriptors

In this paper, quantumchemical calculations were carried out with a SPARTAN 14.0 Wave-Function programming Package. Quantum chemical descriptors generated were the energy of the highest occupied molecular orbital (E-HOMO), the energy of the lowest unoccupied molecular orbital (E-LUMO), Dipole Moment, Polarizability and some others were calculated such as energy band gap ($\Delta E = E\text{-LUMO} - E\text{-HOMO}$), global electronic chemical potential (μ), chemical softness (σ), chemical hardness (η) orbital energies (E-HOMO and E-LUMO) using appropriate relations (Equations 1–7) as previously reported in literature [5, 25]. The fully optimized 3D structure in mole file was then imported into Material studios 8.0 software to compute both thermodynamic, topological, autocorrelation constitutional, electronic, and geometric descriptors [26] for further studies [27]:

$$\Delta E = E_{LUMO} - E_{HOMO} \quad (1)$$

$$\eta = -\frac{1}{2}(E_{HOMO} - E_{LUMO}) \quad (2)$$

$$\sigma = \frac{1}{\eta} = -\left(\frac{2}{E_{HOMO}-E_{LUMO}}\right) \quad (3)$$

$$\mu = \frac{1}{2}(E_{HOMO} + E_{LUMO}) \quad (4)$$

$$\omega = \frac{\mu^2}{2\eta} \quad (5)$$

$$\chi = \frac{-(E-HOMO+E-LUMO)}{2} \quad (6)$$

$$\Delta N = \frac{\chi_{Fe}-\chi_{inh}}{2(\eta_{Fe}+\eta_{inh})} \quad (7)$$

Where $\chi_{Fe} = 7eV$, $\eta_{Fe} = 0$ [28]

2.5 Dataset division into modeling and prediction sets

The data set was divided into two sets, the modeling and prediction set. The modeling set is used in developing the model, it contains seventy percent (70%) of the entire data set. While the test set which constitutes the remaining thirty percent (30%) of the whole data set was not used in the construction of the model but to ascertain the predictive ability of the model [29]. This partitioning ensures that a similar principle can be employed for the activity prediction of the test set. Kennard-Stone Algorithm will be applied for dividing Dataset into a modeling and test set [30, 31].

2.6 Model development

In QSAR studies the identification and selection of descriptors which provide maximum information in activity variations and have minimum co-linearity are important. Therefore, a genetic function algorithm (GFA) [32] improves the model accuracy in the selection of proper descriptors. Multiple Linear Regression was used on the modeling set to show the relationship between the dependent variable Y (%IE) and independent variable X (molecular descriptors). In regression analysis, the contingent mean of the dependent variable (%IE) Y relies on (descriptors) X.

2.7 QSAR Model Validation

Validation of a QSAR model's stability and predictive ability is another key step in QSAR modeling. Different statistical parameters have been used for the validation of the suitability of the built model for the prediction of the anti-cancer activity of the studied compounds [20] This includes the square of the correlation coefficient (R^2) this describes the fraction of the total variation attributed to the model. The closer the value of R^2 is to 1.0, the better the regression and equation explain the Y variable. R^2 is the most commonly used internal validation indicator and is expressed as in equation (8):

$$R^2 = 1 - \frac{\sum(Y_{exp}-Y_{pred})^2}{\sum(Y_{exp}-m_{training})^2} \quad (8)$$

Where, Y_{exp} ; Y_{pred} ; $Y_{mtraining}$ are the experimental property, the predicted property and the mean experimental property of the samples in the training set, respectively. The minimum recommended value for this parameter as shown in Table 3 [33].

Adjusted R^2 (R^2_{adj}): R^2 value varies directly with the increase in the number of regressors i.e. descriptors, thus, R^2 cannot be a useful measure for the goodness of model fit. Therefore, R^2 is adjusted for the number of explanatory variables in the model. The adjusted R^2 is defined as in equation (9):

$$R^2_{\text{adj}} = 1 - (1 - R^2) \frac{N-1}{N-P-1} = \frac{(N-1)R^2 - P}{N-P+1} \quad (9)$$

Where p = number of independent variables in the model and N = sample size [34]. The minimum recommended value for this parameter as shown in Table 3.

Cross-validation coefficient parameter (Q^2_{CV}), which is the most commonly used internal validation indicator and is expressed as in equation (10):

$$Q^2_{\text{CV}} = 1 - \frac{\sum(Y_{\text{pred}} - Y_{\text{exp}})^2}{\sum(Y_{\text{exp}} - Y_{\text{mntraining}})^2} \quad (10)$$

where Y_{exp} is the experimental activity, Y_{pred} is the predicted activity, and $Y_{\text{mntraining}}$ is the mean of the experimental activity of the validation set [29]. However, it should be noted that a high Q^2_{CV} does not necessarily mean high predictability of the developed model [20]. In other words, the high value of Q^2_{CV} is a necessary condition, but not sufficient for a developed model to have high predictability.

In order to assess the predictive ability and to check the statistical significance of the developed models, the proposed models were applied to predict the values of %IE of an external set that was not used in the development of the model. The predictive powers of the developed regression models on the training set were evaluated by predicted values of the prediction set. Therefore validation through an external prediction set (R^2_{test}), is a very paramount parameter that is used to test the external predictive ability of a QSAR model. The R^2_{test} value is calculated by equation (11):

$$R^2_{\text{test}} = 1 - \frac{\sum(Y_{\text{pred}} - Y_{\text{exp}})^2}{\sum(Y_{\text{exp}} - Y_{\text{mntraining}})^2} \quad (11)$$

where Y_{exp} is the experimental activity, Y_{pred} is the predicted activity, and $Y_{\text{mntraining}}$ is the mean of the experimental activity of the validation set [29].

Table 3: Minimum recommended values of validated parameters for generally acceptable QSAR

Symbol	Name	Value
R^2	Coefficient of determination	≥ 0.6
Q^2_{cv}	Cross-validation coefficient	< 0.5
R^2_{test}	The coefficient of determination for an external test set	≥ 0.6
$R^2 - Q^2$	Difference between R^2 and Q^2	≤ 0.3
N_{test}	Minimum number of external test set	≥ 5

2.8 Determination of the applicability domain of the model

The applicability domain of the QSAR model is imperative in establishing the model ability to make predictions within the chemical space for which it was developed [29]. The leverage tactic was used in unfolding the applicability domain of the QSAR models [35]. In this regard, Leverage approach is used and is represented as h_i in equation (12):

$$h_i = x_i(X^T X)^{-1} x_i^T \quad (12)$$

Where x refers to the descriptor vector of the considered compound and X Represents the descriptor matrix derived from the training set descriptor values. The warning leverage (h^*) was determined as in equation (13):

$$h^* = \frac{3(p+1)}{N} \quad (13)$$

Where N is the number of training compounds and p is the number of descriptors in the model. The compounds with leverages $h_i < h^*$ are considered to be reliably predicted by the model. The Williams plot, a plot of standardized residuals versus leverage values, is utilized to translate the relevant area of the model in terms of chemical space. The domain of unfailing prediction for external test set molecules' is defined as compounds

which have leverage values within the threshold ($hi < h^*$) and standardized residuals no greater than 3α (3 standard deviation units), hence they are accepted as Y outlier. The test set compounds where ($hi > h^*$) are thought to be unreliably anticipated by the model because of considerable extrapolation. For the training set, the Williams plot is utilized to recognize compounds with the best structural influence ($hi > h^*$ in developing the model).

2.2 Molecular Dynamics Simulation

The molecular dynamics simulation was carried out to describe the interaction between the inhibitor molecules and the metallic surface. The adsorption locator module implemented in the Materials studio 8.0 software from Accelrys was used for the simulation. The inhibitor molecules were modeled and optimized using the Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies (COMPASS) force field. COMPASS is a robust and well-developed force field that was derived based on fitting against a wide range of experimental data for organic and inorganic compounds [36]. This informs its suitability for treating metal and non-metal containing systems. The whole system was performed at 298K controlled by the Andersen thermostat with fixed number-volume-energy (microcanonical) ensemble, with a time step of 1.0 fs, simulation time of 50ps. The MD simulation was carried out in a simulation box (24.82 Å × 24.82 Å × 45.27 Å) with periodic boundary conditions. The box includes a Fe slab, an acid solution layer, and an inhibitor molecule. For the iron surface, Fe (110) surface was selected as the studied surface for that Fe (110) was density packed surface and was the most stable[37]. The iron crystal contained ten layers and seven layers near the bottom were frozen. The density of the acid solution layer was set at 1.0 g/cm³. The adsorption and binding energy values in solution was calculated by equation 14 and 15 respectively[25].

$$E_{adsorption} = E_{Total} - (E_{Fe_{surface+solution}} + E_{Inhibitor}) \quad (14)$$

$$E_{adsorption} = E_{binding} \quad (15)$$

3. Results and Discussion

3.1 Quantum Chemical studies

Quantum chemical calculations were performed on the BI-inhibitors in order to relate their inhibition potentials to molecular structures. According to the frontier orbital approximation, donor-acceptor interactions do occur between frontiers molecular orbitals (HOMO and LUMO) of reacting species [38]. The adsorption process of a corrosion inhibitor molecule onto a metal surface increases with an increase of the HOMO energy (E-HOMO) and a decrease of the LUMO energy (E-LUMO). This is because, from the HOMO orbital the inhibitor molecule will donate the electrons to the *d*-orbital of the metal molecule, and the LUMO orbital of the inhibitor will receive the electrons from the *d*-orbital of the metal molecule, in-electron-donation, and electron-back-donation process. Thus, E-HOMO is often associated with the electron donating ability of a molecule; a high value of E-HOMO indicates the tendency of the inhibitor to donate electrons to the acceptor metal.

The results from Table 2 indicate that the HOMO and LUMO orbitals are largely distributed evenly throughout the molecular structure. It seems that the presence of the benzene ring in the BI molecule has a notable influence on the donor-acceptor interactions during the adsorption process. Frontier molecular orbital energies of the studied imidazole inhibitors (E-HOMO and E-LUMO) are listed in Table 1 According to the frontier molecular orbital theory, a high E-HOMO value for a molecule implies the tendency of the molecule to donate electrons to the appropriate vacant d-orbital of the metal atoms [25]. The higher the HOMO energy the more the probability of a molecule to donate an electron to the metal. Thus, E-HOMO decreases in the order, BI-17 > BI-10 > BI-3 > BI-20 > BI-1 > BI-13 > BI-15 > BI-6 > BI-9 > BI-4 > BI-2 > BI-5 > BI-11 > BI-6 > BI-19 > BI-18 > BI-12 > BI-16 > BI-7 > BI-14. A comparison of the E-HOMO of the studied inhibitors shows that BI-17 has the highest while BI-14 has the lowest which indicates that BI-17 would be the most preferred electron donor while BI-14 would be the least preferred.

On the other hand, a low E-LUMO value for a molecule indicates the relative tendency of the molecule to accept electrons from metallic orbital during back-donation. From Table 1, the E-LUMO value in the descending order are: BI-10 > BI-13 > BI-15 > BI-8 > BI-12 > BI-6 > BI-7 > BI-14 > BI-9 > BI-4 > BI-5 > BI-16 > BI-19 > BI-17 > BI-20 > BI-3 > BI-18 > BI-1 > BI-2, BI-11, indicating that BI-11 has a stronger tendency

to accept electrons from metallic orbital as they adsorb on iron. Therefore, BI-11 can more powerfully adsorb on the metallic surface to achieve better corrosion inhibition effectiveness.

Energy gap ΔE (E-LUMO - E-HOMO) is an important parameter as efficient inhibitor is characterized by a small energy gap[39]. From Table 1, the trend of decreasing the energy gap is consistent with increasing experimental %IEs of the studied BI-inhibitors, *i.e.* the energy gaps decrease as the %IEs increase from BI-1 to BI-20. Thus, BI-13 > BI-7 > BI-14 > BI-12 > BI-8 > BI-15 > BI-10 > BI-6 > BI-9 > BI-16 > BI-5 > BI-4 > BI-19 > BI-18 > BI-20 > BI-1 > BI-2 > BI-11 > BI-3 > BI-17. This suggest that BI-17 with the lowest energy gap of 4.47eV is the best inhibitor in term of this parameter among the series.

Dipole moment (μ) can give information on the polarity of a molecule. Molecules with higher value of the dipole moment have a greater tendency to interact with other molecule through electrostatic interactions [14]. Based on this parameter presented in Table 1, the trend of increase in the μ follow the order: BI-20 > BI-19 > BI-18 > BI-2 > BI-11 > BI-17 > BI-3 > BI-10 > BI-5 > BI-15 > BI-4 > BI-6 > BI-8 > BI-16 > BI-7 > BI-14 > BI-1 > BI-12 > BI-9 > BI-13. BI-20 has the highest and there for the best to be adsorbed on the metallic surface.

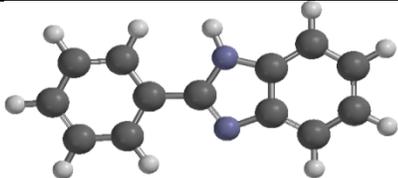
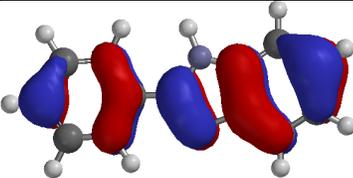
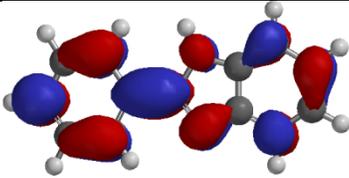
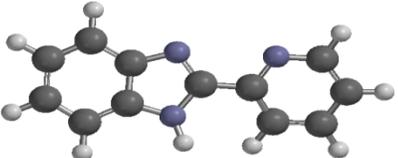
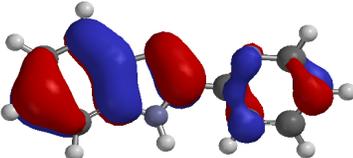
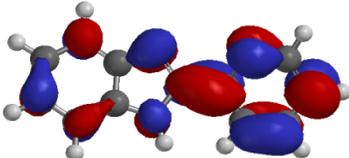
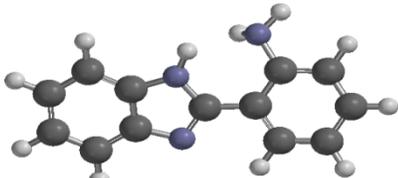
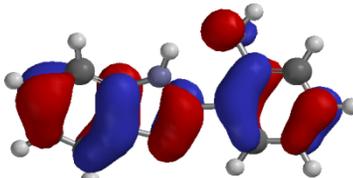
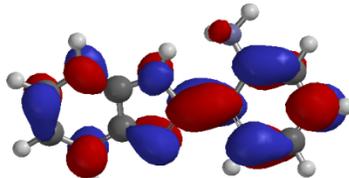
Inhibitor with the large hardness (η) value is expected to be a weaker inhibitor compared to the inhibitors with lower values. From the results in Table 1 the hardness value decreases in the order BI-13 > BI-7 > BI-14 > BI-12 > BI-8 > BI-15 > BI-10 > BI-6 > BI-9 > BI-16 > BI-5 > BI-4 > BI-19 > BI-18 > BI-20 > BI-1 > BI-2 > BI-11 > BI-3 > BI-17, this shows that BI-17 with the lowest value of 2.235 will be the best inhibitor and this inhibitor was found to have relatively high %IE from the experiment.

Table 1: Selected quantum chemical parameters for the studied inhibitors

Inhibitors	E HOMO(eV)	E LUMO(eV)	Dipole Moment	ΔE	η	σ	ω	χ	ΔN
BI-1	-5.70	-1.17	3.09	4.53	2.265	0.4415	13.3626	6.285	0.809738
BI-2	-5.87	-1.37	4.83	4.5	2.25	0.44444	14.7425	6.555	0.500625
BI-3	-5.50	-1.00	4.13	4.5	2.25	0.44444	11.8828	6	1.125
BI-4	-5.86	-0.42	3.6	5.44	2.72	0.36765	13.4091	6.07	1.2648
BI-5	-5.87	-0.43	3.68	5.44	2.72	0.36765	13.4946	6.085	1.2444
BI-6	-5.81	-0.33	3.56	5.48	2.74	0.36496	12.9121	5.975	1.40425
BI-7	-6.07	-0.34	3.44	5.73	2.865	0.34904	14.7147	6.24	1.0887
BI-8	-5.89	-0.22	3.5	5.67	2.835	0.35273	13.2296	6	1.4175
BI-9	-5.83	-0.36	2.26	5.47	2.735	0.36563	13.0993	6.01	1.353825
BI-10	-5.45	0.1	3.75	5.55	2.775	0.36036	9.92843	5.4	2.22
BI-11	-5.87	-1.37	4.83	4.5	2.25	0.44444	14.7425	6.555	0.500625
BI-12	-6.01	-0.31	2.9	5.7	2.85	0.35088	14.2295	6.165	1.189875
BI-13	-5.74	0.00	2.21	5.74	2.87	0.34843	11.82	5.74	1.8081
BI-14	-6.07	-0.34	3.44	5.73	2.865	0.34904	14.7147	6.24	1.0887
BI-15	-5.79	-0.18	3.62	5.61	2.805	0.35651	12.4966	5.88	1.5708
BI-16	-6.04	-0.58	3.47	5.46	2.73	0.3663	14.9551	6.33	0.91455
BI-17	-5.40	-0.93	4.33	4.47	2.235	0.44743	11.1943	5.865	1.268363
BI-18	-5.90	-1.01	5.38	4.89	2.445	0.409	14.593	6.405	0.727388
BI-19	-5.89	-0.92	5.62	4.97	2.485	0.40241	14.4056	6.35	0.807625
BI-20	-5.68	-0.99	7.32	4.69	2.345	0.42644	13.0408	6.175	0.967313

Energy band gap (ΔE), chemical hardness (η), chemical softness (σ), electrophilicity index (ω), global electronegativity (χ), number of transferred electrons (ΔN)

Table 2: Optimized Structures, E-HOMO, and E-LUMO of the studied inhibitors

Inhibitor	Optimized	E-HOMO	E-LUMO
BI-1			
BI-2			
BI-3			

Additionally, Adsorption usually occurs in the region of the molecule where chemical softness (σ) has the highest value. The order across structures in the σ values as reported in Table 1 decreases in the order, BI-17 > BI-2 > BI-11 > BI-3 > BI-1 > BI-20 > BI-18 > BI-19 > BI-5 > BI-4 > BI-16 > BI-9 > BI-6 > BI-10 > BI-15 > BI-18 > BI-12 > BI-7 > BI-14 > BI-10. This suggests that BI-17 has the highest and therefore the most reactive inhibitor molecule. This observation partially agree with the trend in the inhibition efficiencies of the inhibitors.

Moreover, the tendency of the inhibitor molecule to accept electron/s is measured by global electrophilicity index (ω), thus, for the series BI-1 to BI-20, the values for this parameter decreases as BI-16 > BI-2 > BI-11 > BI-7 > BI-14 > BI-18 > BI-19 > BI-12 > BI-5 > BI-4 > BI-1 > BI-8 > BI-9 > BI-20 > BI-6 > BI-15 > BI-3 > BI-13 > BI-17 > BI-10. According to this parameter BI-10 with lower electrophilicity index should exhibits higher inhibition efficiency.

Furthermore, the inhibitor with high electronegativity and hence low electronegativity difference (between the Fe and the inhibitor) will quickly reaches equalization and hence will have a low reactivity which in turn means low inhibition efficiency[25]. From Table 1 it is observed that this value decreases in the order, BI-2 > BI-11 > BI-18 > BI-19 > BI-16 > BI-1 > BI-7 > BI-14 > BI-20 > BI-12 > BI-5 > BI-4 > BI-9 > BI-8 > BI-3 > BI-6 > BI-15 > BI-17 > BI-13 > BI-10. This indicate that BI-10 with the lowest electronegativity value is the best inhibitor among the series.

ΔN value showed inhibition effect resulting from electrons donation. If $\Delta N < 3.6\text{\AA}$, the inhibition efficiency increases with the increasing electron-releasing ability at the metal surface [25]. The calculated ΔN values for the BI-inhibitors are summarized in Table 1. For all BI-inhibitors, ΔN values are smaller than 3.6\AA . It decreases in the order: IM-10 > IM-13 > IM-15 > BI-8 > BI-6 > BI-9 > BI-17 > BI-4 > BI-5 > BI-12 > BI-3 > BI-7 > BI-14 > BI-20 > BI-16 > BI-1 > BI-19 > BI-18 > BI-2 > BI-11. This suggest that of all the twenty inhibitors are electrons donors with varying degrees, and the steel surface is the acceptor thereby binding the inhibitor to the Fe surface resulting in inhibition adsorption layer that prevent the corrosion process.

3.2 Quantitative structure–activity relationship (QSAR) Studies

Genetic function approximation was used on the training set to select the significant descriptors and it was found that among all the computed descriptors, Solvation Energy, Polarizability, E-HOMO, Polar surface area, Ovality, Polarizability, and Log P construct the best model. The selected descriptors were subjected to regression analysis with the experimentally determined inhibition efficiencies as the dependent variable and the selected descriptors as the independent variables using Genetic function approximation (GFA) method in

Material studio software, new GFA-MLR equation as shown in Equation (16) were developed on the basis of the training set.

QSAR model for predicting %IE against Steel corrosion

$$\%IE = 3.250057429 * (\mathbf{E} - \mathbf{HOMO}(\mathbf{eV})) - 1.167830197 * (\mathbf{Polar\ surface\ area}) + 652.158586289 * (\mathbf{Ovality}) - 9.872435953 * (\mathbf{Polarizability}) - 5.645379027 * (\mathbf{Log\ P}) + 126.822103851 \quad (16)$$

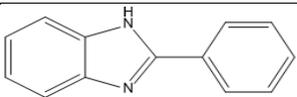
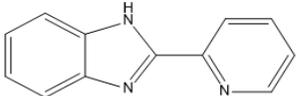
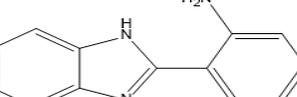
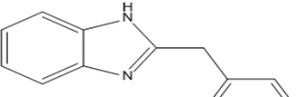
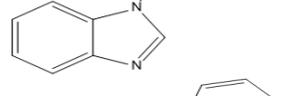
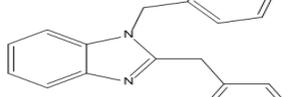
$$N_{train.} = 15, R_{train.}^2 = 0.942, R_{adjusted.}^2 = 0.908, Q_{cv.}^2 = 0.795, N_{test.} = 5, R_{test.}^2 = 0.983$$

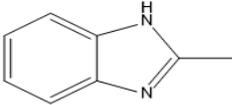
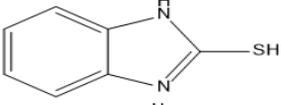
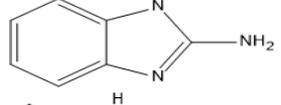
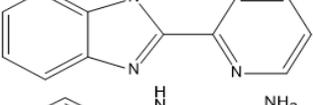
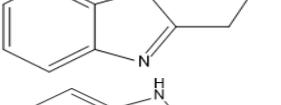
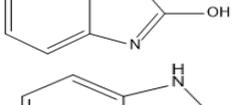
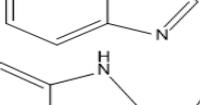
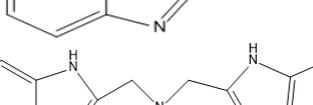
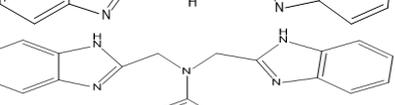
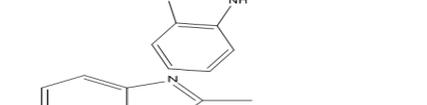
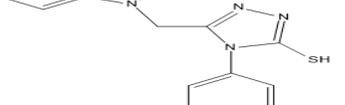
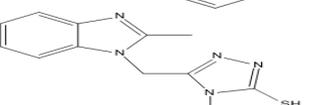
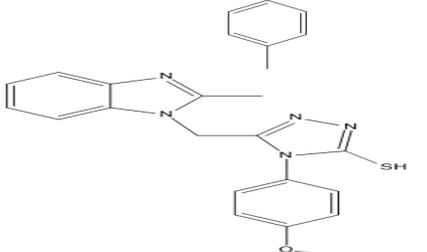
Where N is the number of compounds in the training and test set, $R_{train.}^2$ is the squared correlation coefficient, $R_{adjusted.}^2$ is the Adjusted R-squared, $Q_{cv.}^2$ is the cross-validation coefficients of the training set and $R_{test.}^2$ is the squared correlation coefficient of the prediction (test) set.

3.3 QSAR model validation

In further study, the built model from the training data set was used to evaluate the predictive ability of the model by predicting the %IE values in the prediction set (test set). The results are given in Table 4. The predicted %IE values for the training and test sets were plotted against the experimental %IE as shown in Figure 1. The predicted %IE results obtained for both the training set and test set (Table 4) are in good agreements with the experimental %IE obtained. The residual values obtained between predicted and experimental %IE was very low.

Table 4: Chemical Structures, experimental and predicted %IE of the inhibitors with residuals

S/N	Structure	Experimental (%IE)	Predicted (%IE)	Residuals
1		85	86.75	-1.75
2t*		93	87.05	5.95
3		93	91.24	1.76
4		95	96.1	-1.1
5		96	100.68	-4.68
6		52.2	51.61	0.59
7		88.7	82.49	6.21

8t		75.99	79.33	-3.34
9		72.4	70.65	1.75
10		62.32	63.84	-1.52
11		50.34	51.61	-1.27
12		72.2	77.78	-5.58
13t		91.8	89.92	1.88
14		87.1	82.66	4.44
15		81.1	87.18	-6.08
16		85	86.75	-1.75
17t		93	87.05	5.95
18		93	91.24	1.76
19		95	96.1	-1.1
20t		96	100.68	-4.68

't' represent test sets and * identifies compounds found outside the applicability domain of the QSAR model.

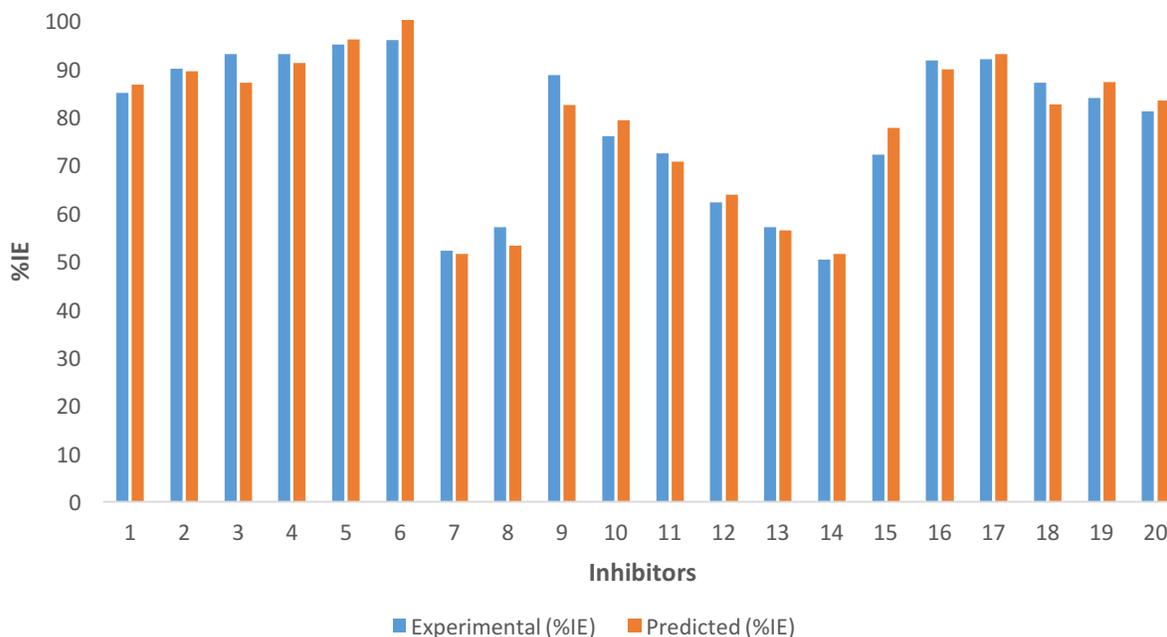


Figure 1: Experimental against predicted inhibition efficiency (%IE)

The result of the QSAR model is in conformity with the standard shown in Table 3 as seen from the built model. The closeness of coefficient of determination (R^2) to its absolute value of 1.0 is an indication that the model explained a very high percentage of the response variable (descriptor) variation, high enough for a robust QSAR model. Its 0.867 value illustrates that 86.7% of variation is residing in the residual meaning that the model is very good.

The high adjusted R^2 (R^2_{adj}) value as seen in the model and its closeness in value to the value of R^2 implies that the model has excellent explanatory power to the descriptors in it. It also illustrates the real influence of applied independent variables on the Dependent Variable. Also, the high and closeness of Q^2_{cv} to R^2_{train} revealed that the model was not over-fitted. The high R^2_{test} as seen in the model is an indication that the model is capable of providing valid predictions for new molecules.

Furthermore, the equation contains five descriptors and each descriptor has a positive or negative coefficient attached to it. These coefficients along with the value of descriptor have a significant role in deciding the overall inhibition efficiency of the inhibitor molecules. Examination of the QSAR-model shows that the coefficients of each descriptor play an important role in deriving the inhibition efficiency. From the point of view of inhibition of the molecules in terms of %IE values, the weight of a positive coefficient is very significant because it contributes towards the increased value of %IE. So the descriptors with high weight positive coefficients are most important followed by descriptors with the low weight negative coefficient and lastly the descriptors with high weight negative coefficients [40]. On the basis of values of the coefficients on the model, the associated descriptors are arranged in a sequence pertaining to their contribution towards overall inhibition efficiency of the inhibitors, in the following increasing order of inhibition efficiency of inhibitors towards corrosion of steel.

Ovality > E – HOMO > Polarizability > Polar surface area > Log P

The leverages for each inhibitor in the dataset were plotted against their standardized residuals, leading to the discovery of outliers and influential inhibitors in the model. Figure 2 shows the Williams plot of standardized residuals against calculated leverages for both the training and test set. The warning leverage (h^*), was found to be 1.2 ($N = 15$ and $p=5$) for the developed QSAR model [41]. The inhibitors that had a standardized residual more than the standard deviation units were considered to be outliers while inhibitors with

a leverage value higher than h^* were considered to be influential or high leverage inhibitors. Based on the leverages one of the inhibitors (IM-2) were found to be outside of the defined AD (Figure 2) of the QSAR model which is considered as an influential compound. In addition, no outlier compounds with standardized residuals $> \pm 3d$ for the dataset were identified.

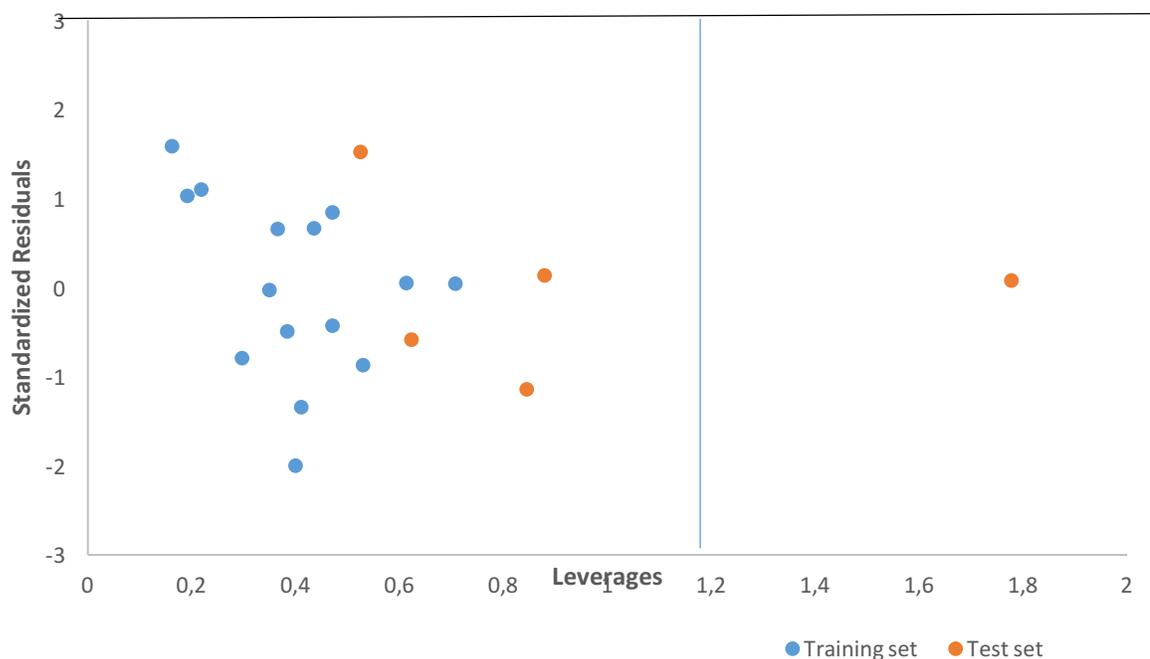


Figure 2: The Williams plot, the plot of standardized Residuals versus the leverages

3.4 Molecular Dynamic Simulation Studies

Geometry optimization of the studied inhibitors, solvent molecules (H_2O), hydronium ions (H_3O^+) and chloride ions (Cl^-) was carried out until the total energy reached a local minimum on the potential energy surface. During the course of the geometry optimization process, atomic coordinates were adjusted based on COMPASS force field until the total energy of the individual structure reaches the minimum energy, afterward, a simulation box was created by the all concerned species. Then the dynamics process was performed, and the whole system reached equilibrium when both the temperature and energy of the system were balanced. The best adsorption configurations of the best inhibitor molecules over the Fe (1 1 0) surface are depicted in Figure 3.

All the other inhibitors were studied in the same way. The adsorption and binding energy values were calculated according to equation 14 and 15 and were listed in Table 5.

It can be observed from the molecular structures of the examined inhibitors that these molecules contain a various lone pair of electrons on N and O atoms, as well as π -aromatic systems. Therefore, the lone pair of electrons on heteroatoms will be donated to the empty d orbitals of iron Metal. It can be seen from Figure 1 that the inhibitor adsorbs almost flatly.

Additionally, it could be seen in Table 5 that the calculated adsorption energy values of the adsorption systems at 298 K are large and negative. These larger negative values of adsorption energies can be ascribed to the strong interaction between the studied inhibitors and iron surfaces. Thus calculated adsorption energy values revealed that inhibitor (BI-6) adsorb on the iron surface more spontaneously than any other inhibitor of the series. These outcomes are in good agreement with the results obtained from wet chemical experimentation as the inhibitor has a higher inhibition efficiency [22].

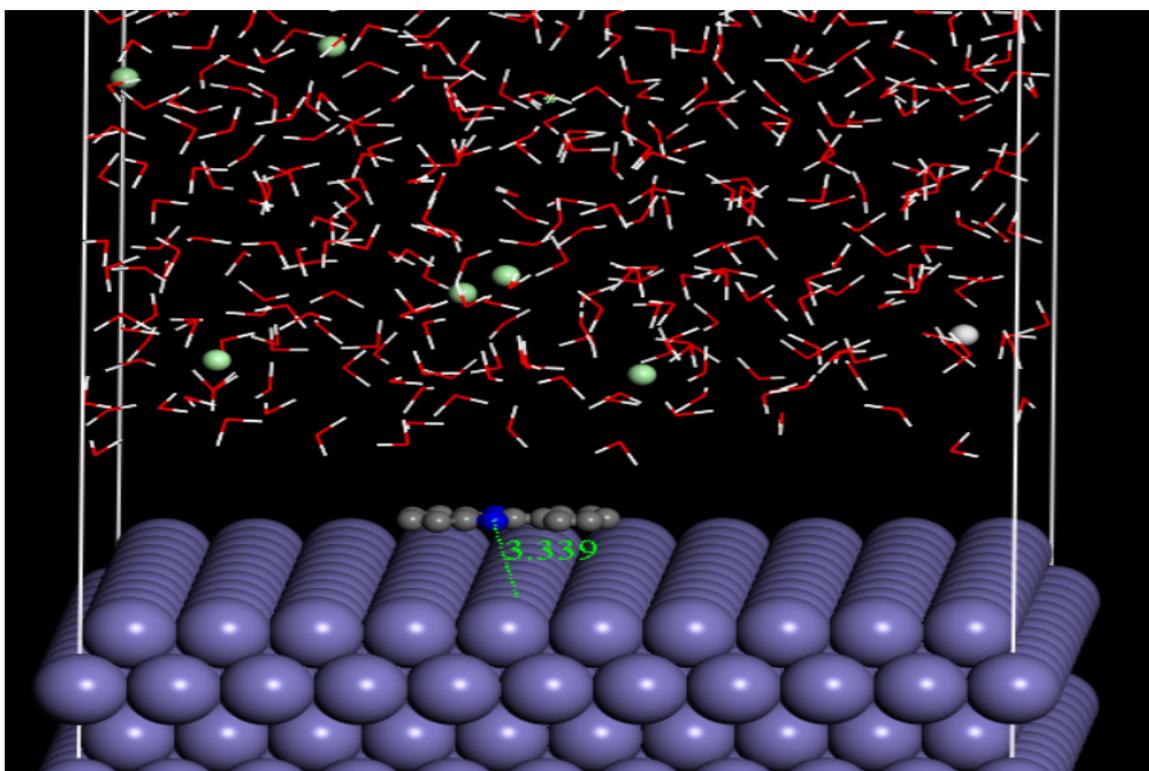


Figure 3: Equilibrium adsorption configurations of inhibitor, BI-1 on Fe (1 1 0) surface

Table 5: Adsorption Energy, Binding Energy, shortest bond distances and inhibition efficiencies (%IE) of the inhibitors

Inhibitors	Adsorption Energy(kcal/mol)	Binding Energy(kcal/mol)	shortest bond distance(Å)	Experimental %IE
BI-1	-142.82	142.82	3.339	85
BI-2	-231.16	231.16	2.78	90
BI-3	-256.72	256.72	2.80	93
BI-4	-214.82	214.82	2.78	93
BI-5	-329.12	329.12	3.03	95
BI-6	-359.23	359.23	2.988	96
BI-7	-144.21	144.21	3.13	52.2
BI-8	-197.50	197.50	3.27	57.1
BI-9	-234.58	234.58	3.10	88.7
BI-10	-227.71	227.71	3.05	75.99
BI-11	-215.70	215.70	2.87	72.4
BI-12	-190.94	190.94	4.39	62.32
BI-13	-157.21	157.21	3.42	57.07
BI-14	-181.67	181.67	3.34	50.34
BI-15	-197.00	197.00	3.17	72.2
BI-16	-280.41	280.41	2.67	91.8
BI-17	-259.21	259.21	2.71	92
BI-18	-228.24	228.24	2.89	87.1
BI-19	-227.37	227.37	3.00	83.9
BI-20	-218.05	218.05	2.86	81.1

The measured short bond distances and the larger binding energy values for the studied inhibitors indicate that a chemical bond was formed between the inhibitor molecule and Fe surface atom. Thus, it is further confirmed from the MD simulations that adsorption of the inhibitor molecules on the metallic surfaces mainly occurred by the chemical adsorption phenomenon since the bond distances ($< 3.5\text{\AA}$) and the magnitude of the binding energy ($> 100\text{kcal/mol}$) are all within the range of chemisorptive interaction [42].

Conclusion

A theoretical modeling approach was performed to study the corrosion inhibition performance of benzimidazole derivatives on the steel surface. Quantum chemical calculation reveals that electron donation and electron acceptance capability of the studied inhibitors is in good agreement with the results obtained from previously performed experimental findings. The prediction of corrosion efficiencies of the studied inhibitors by the built QSAR models gave a very good correlation coefficient. The robustness and applicability of the QSAR models have been established by internal and external validation techniques. The interaction energies (E_{ads} and E_{bind}) between the studied inhibitors and Fe (1 1 0) plane via MD-simulation were large indicating chemical bond formation and that the inhibitors can more tightly adsorb on the iron surface. The order of adsorption energy and binding energy values are in good agreement with the experimental findings and the shortest bond distances measured confirmed chemisorptive interaction.

Acknowledgments-The authors thank Ahmadu Bello University, Zaria for granting us computational soft wares used in this study and other members in the group for their advice in the cause of this study.

References

1. Z. Szklarska-Smialowska, ZS-Smialowska, *NACE International Houston, TX*. (2005).
2. I.B.Obot, *InTech.*, (2014) 123.
3. A. Al Hashem, *Proceedings of the Corrosion UAE, Abu Dhabi, UAE*, (2011).
4. P. Singh, E. E. Ebenso, L. O.Olasunkanmi, I. Obot, M. Quraishi, *The Journal of Physical Chemistry C*, 120(6) (2016) 3408-19.
5. H. U. Nwankwo, C.N. Ateba, L. O. Olasunkanmi, A.S. Adekunle, D.A. Isabirye, D. C.Onwudiwe, *Materials*, 9(2) (2016)107.
6. L.C. Murulana, A. K.Singh, S. K. Shukla, M. M. Kabanda, E. E. Ebenso, *Industrial & Engineering Chemistry Research*, 51(40) (2012) 13282-99.
7. H. Zhao, X. Zhang, L. Ji, H. Hu, Q. Li, *Corrosion Science*, 83 (2014) 261-71.
8. Y.I.Kuznetsov, L. Kazansky, *Russian Chemical Reviews*, 77(3) (2008) 219-232.
9. F. Zhang, Y. Tang, Z. Cao, W. Jing, Z. Wu, Y. Chen, *Corrosion Science*, 61 (2012) 1-9.
10. I. Ahamad, and M. Quraishi, *Corrosion Science*, 51(9) (2009) 2006-2013.
11. Y. Tang, F. Zhang, S. Hu, Z. Cao, Z. Wu, W. Jing, *Corrosion science*, 74 (2013) 271-82.
12. D-Q. Zhang, Q-R. Cai, X-M. He, L-X. Gao, G-D. Zhou, *Materials Chemistry and Physics*, 112(2) (2008) 353-8.
13. M.A. Amin, M.M. Ibrahim, *Corrosion Science*, 53(3) (2011) 873-885.
14. N.A.Wazzan, I. Obot, S. Kaya, *Journal of Molecular Liquids*, 221 (2016) 579-602.
15. Y. Atalay, F. Yakuphanoglu, M. Sekerci, D. Avci, A. Başoğlu, *Spectrochimica Acta Part A*, 64(1) (2006) 68-72.
16. E. E. Ebenso, T. Arslan, F. Kandemirli, N. Caner, I. Love, *International Journal of Quantum Chemistry*. 110(5) (2010) 1003-18.
17. E.Kraka, D. Cremer, *Journal of the American Chemical Society*, 122(34) (2000) 8245-8264.

18. F. Bentiss, M. Traisnel, H. Vezin, M. Lagrenée, *Corrosion science*, 45(2) (2003) 371-80.
19. H. El Sayed, , A. El Nemr, S. Ragab, *Journal of molecular modeling*, 18(3) (2012) 1173-1188.
20. T. Asadollahi, S. Dadfarnia, A. M. H. Shabani, J.B. Ghasemi, M. Sarkhosh, *Molecules*, 16(3)(2011)1928-55.
21. A. Popova, M. Christov, T. Deligeorgiev, *Corrosion*, 59(9) (2003) 756-764.
22. J. Aljourani, K. Raeissi, M. Golozar, *Corrosion science*, 51(8) (2009) 1836-1843.
23. K. Khaled, *Electrochimica Acta*, 48(17) (2003) 2493-2503.
24. V. N. Viswanadhan, A. K. Ghose, G.R. Revankar, R. K. Robins, *Journal of chemical information and computer sciences*. 29(3) (1989) 163-72.
25. S.K.Saha, P. Banerjee, *RSC Advances*, 5(87) (2015) 71120-71130.
26. S.A. Amin, S. Gayen, *Journal of Taibah University for Science*, 10(6) (2016) 896-905.
27. C.W.Yap,. *Journal of computational chemistry*, 32(7) (2011) 1466-1474.
28. A.Y. Musa, R.T. Jalgham, A.B. Mohamad, *Corrosion Science*, 56 (2012) 176-183.
29. A.Tropsha, P. Gramatica, and V.K. Gombur, *Molecular Informatics*, 22(1) (2003) 69-77.
30. K. Rajer-Kanduč, J. Zupan, N. Majcen, *Chemometrics and intelligent laboratory systems*, 65(2) (2003) 221-229.
31. R.W.Kennard, L.A. Stone,. *Technometrics*, 11(1) (1969) 137-148.
32. R. Leardi, *Genetic algorithms in molecular modeling*. (1996), Elsevier. 67-86.
33. W. Wu, C. Zhang, W. Lin, Q. Chen, X. Guo, Y. Qian,. *PloS one*. 10(3) (2015) 119575.
34. U. Abdulfatai, A. Uzairu, S. Uba, *Journal of advanced research*, 8(1) (2017) 33-43.
35. P.Gramatica, E. Giani, E. Papa,. *Journal of Molecular Graphics and Modelling*, 25(6) (2007) 755-766.
36. A. Wymyslowski, N. Iwamoto, M. M. F. Yuen, H. Fan. *Molecular Modeling and Multiscaling Issues for Electronic Material Applications*, 2 (2014).
37. K. Khaled,. *Electrochimica Acta*, 53(9) (2008) 3484-3492.
38. G.Bereket, C. Öğretir, Ç. Özşahin, : *Theochem*, 663(1-3) (2003) 39-46.
39. Ebenso EE, Khaled K, Shukla SK, Singh AK, Eddy N, Saracoglu M,. (2012).
40. K. Y Wong, P. R Duchowicz, A. G Mercader, E. A Castro, *Mini reviews in medicinal chemistry*, 12(10) (2012) 936-46.
41. K. Roy, , S. Kar, and R.N. Das,. *Fundamental concepts*.(2015): Springer.
42. C.O. Akalezi, C.K. Enenebaku, E.E. Oguzie,. *International Journal of Industrial Chemistry*, 3(1) (2012) 13.

(2019) ; <http://www.jmaterenvirosci.com>