Jurnal Teknologi

Experimental Study and High Dimensional QSAR Modelling of Phenylpropanoids of *Alpinia galanga* as Corrosion Inhibitors on Mild Steel

Sunday O. Ajeigbe[a], Norazah Basar[a], Zakariya Y. Algamal[b], Muhammad H. Lee[c], Hasmerya Maarof[a], Madzlan Aziz[a,d]*

[a]Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[b]Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Iraq
[c]Department of Mathematical Sciences, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[d]Advanced Membrane Technology Centre, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Graphical abstract
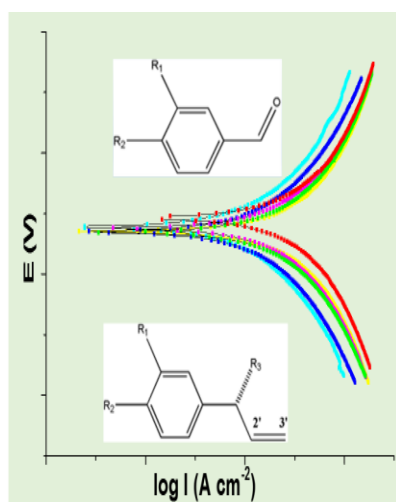
Abstract

Plant extracts as corrosion inhibitors have been extensively investigated and are found as an alternative to synthetic organic compounds. The corrosion inhibition of mild steel in 1 M HCl by 15 compounds comprising of five phenylpropanoids from *Alpinia galanga* and other related compounds was explored experimentally using potentiodynamic polarisation procedures. The inhibition efficiencies determined experimentally for the various inhibitors were used in the Quantitative Structure-Activity Relationship (QSAR) study with their molecular descriptors calculated using Dragon software. Penalised multiple linear regression (PMLR) was adopted as the method of variable selection using elastic net penalty. The elastic net results show low mean-squared error of the training set (MSE_{train}) of 0.121 and test set (MSE_{test}) of 0.131. The model obtained can be applied to predict the corrosion inhibition efficiencies of related organic compounds. Results also reveal that the PMLR based on elastic net penalty is effective in dealing with high dimensional data.

Keywords: Corrosion inhibitor, potentiodynamic polarisation, phenylpropanoids, *Alpinia galanga,* high dimensional QSAR, penalised multiple linear regression (PMLR)

Abstrak

Ekstrak tumbuhan untuk digunakan sebagai perencat kakisan telah dikaji secara meluas kerana ia boleh menggantikan sebatian organik sintetik. Perencatan kakisan terhadap keluli lembut di dalam 1 M HCl oleh 15 sebatian yang mengandungi lima fenilpropanoid daripada Alpina galanga dan sebatian berkaitan lain dikaji secara eksperimen menggunakan kaedah polarisasi potentiodinamik. Kecekapan perencatan yang ditentukan secara eksperimen bagi pelbagai perencat digunakan dalam kajian hubungan kuantitatif struktur-aktiviti (QSAR) dengan pemerihal molekul mereka dikira menggunakan perisian Dragon. Regresi linear berganda terdenda (PMLR) dipilih sebagai kaedah pemilihan pemboleh ubah menggunakan denda jaringan elastik. Keputusan jaringan elastic menunjukkan ralat purata ganda dua set latihan (MSE_{train}) adalah 0.121 dan set ujian (MSE_{test}) 0.131. Model yang diperolehi boleh digunakan untuk menganggar kecekapan perencatan kakisan sebatian organik berkaitan.

Keputusan juga menunjukkan PMLR berdasarkan denda jaringan elestik adalah berkesan apabila berurusan dengan data dimensi tinggi.

Kata kunci: Perencat kakisan, polarisasi potensiodinamik, fenilpropanoid, Alpina galanga, QSAR dimensi tinggi, regresi linear berganda terdenda (PMLR)

1.0 INTRODUCTION

Metals and alloys of metals are extensively used in many engineering applications in various environments including acidic, alkaline and neutral environments because of their excellent combination of properties. But a metal or metal alloy becomes unstable or corrodes even in its normal environments. Thus, corrosion and means to control it is a major concern in the industries and efforts should be geared up to mitigate or minimise it as much as possible. Corrosion is sinister in its behaviour and may not be immediately obvious.

Any attempt to ignore its threats or delay its control will result in production losses, equipment failures, impaired safety and problematic effluents. Corrosion is a ubiquitous and a global technological challenge and a factor in every chemical process plant.

There are quite a few methods adopted in the control and prevention of corrosion, amongst which the usage of corrosion inhibitor is prominent. A corrosion inhibitor is a substance that is added to a corrosive medium to slow down or prevent the metals from reacting with the medium. In the acidic medium, various types of inhibitors have been used for mild steel. Records have shown that most metallic corrosion inhibitors are usually synthetic organic compounds consisting of aromatic rings or heterocyclic atoms such as nitrogen, oxygen, sulphur and phosphorus, or compounds having multiple bonds in their molecule through which adsorption takes place on the metal or mild steel surface [1-7].

In their work, [1] showed that certain physicochemical properties influence inhibitor adsorption unto the metal surface. These inhibitor properties include not only the functional groups and the electron density around the donor atom, but also its π-orbital character as well as its electronic structure [3].

Quantitative structure activity relationship (QSAR) as a computational modelling technique has found an application in many chemistry disciplines [8-10]. Investigations of inhibitor efficiencies of compounds have been carried out both experimentally and theoretically [11]. Tremendous success in the use of QSAR for corrosion inhibition studies has been recorded [12, 13]. However, despite its wide area of application, it is yet to be applied to study corrosion inhibition of natural plant products. Its use will help to identify the roles of plant constituents towards corrosion inhibition with a view to optimising plant constituents as corrosion inhibitors. This work will further help to generate more effective inhibitors within the class of compounds under investigation.

Besides the high prediction power, high prediction reliability is essential to a good QSAR model [13, 14]. Traditionally, the quantum chemical calculations are employed to calculate the molecular descriptors. In addition, different softwares that generate molecular descriptors for QSAR analysis include Molconn-Z, ADAPT, MOLGEN-QSPR, CODESSA and Dragon [15, 16]. Dragon software has extensive applications in QSAR and scientific studies. The Dragon software version 6.0 calculates a number of 4885 molecular descriptors [17].

In corrosion studies, only few number of inhibitor compounds are available in quantities, sufficient enough for corrosion experiments and it is also found that the molecular descriptors calculated are very large and they cause a high dimensional problem. The high dimensionality problem in QSAR modelling in which the number of molecular descriptors, p, surpasses the number of investigated compounds, n, remains one of the challenges facing the researchers [18-20]. Statistical problems associated with modelling high-dimensional QSAR include model overfitting and multicollinearity [21]. The conventional statistical methods such as multiple linear regression (MLR) cannot eliminate the issues of overfitting and multicollinearity. In recent times, penalised regression methods have been employed to perform variable selection and estimation simultaneously.

In any high dimensional data, variable selection is the main objective. The aim of selecting optimal subset of molecular descriptors is to reduce the descriptors to those that contain relevant information, and thereby to improve QSAR modelling. This should be observed in terms of predictive performance (by decreasing the effect of multicollinearity) and in interpretability (to prevent overfitting). A procedure called penalisation, which is always used for variable selection in high dimensional data, attaches a penalty term $P_\lambda(\beta)$ to the ordinary least squares (OLS) to get a better estimate of the prediction error to avoid overfitting and multicollinearity.

In this work, an experimental study of corrosion inhibition of 15 compounds comprising of five phenylpropanoids from *Alpinia galanga* and related compounds was carried out on mild steel in 1 M HCl solution using potentiodynamic polarisation method.

The experimental results obtained were employed to develop QSAR for the inhibitors using the structural based descriptors calculated by the Dragon software. A large number of molecular descriptors with high dimensionality were obtained. The high dimensional data give more information to develop improved models. However, the enormous data pose a big challenge to the classical variable selection methods to deal with. The variable selection procedures that can deal with all descriptors and their multicollinearity are employed. The elastic net penalty was therefore used to develop QSAR of five phenylpropanoids from *A. galanga* and other related compounds derivable from natural products as inhibitors for the control of mild corrosion in acid medium.

The phenylpropanoids are a diverse class of plant-derived organic compounds. The following phenylpropanoids namely; 1'-acetoxychavicol acetate, methyl eugenol, eugenol acetate, eugenol and 4-hydroxycinnamic acid have been previously isolated from the rhizome of A. galanga [22, 23]. Several of the reported biological activities (antioxidative, antibacterial, antifungal, antiallergic, etc.) of A. galanga are attributed to the phenylpropanoids. The plant-derived phenylpropanoids (PPPs) are composed of the largest class of secondary metabolites formed by the higher plants. They constitute the parent molecules for the biosynthesis of several plant polyphenols of structural and functional diversity such as the simple esters and phenolic acids, flavonoids, isoflavonoids, coumarins, curcuminoids, stilbenes, lignans, etc. [24].

The observations portrayed in this work can be used appropriately in optimisation studies involving compounds similar to the ones being investigated aiming at improving their corrosion inhibition capacities.

2.0 METHODOLOGY

2.1 Material Preparation and Inhibitors

Distilled water and pure grades of ethanol, acetone, hydrochloric acid and inhibitor compounds were purchased from Sigma-Aldrich and Merck and were used without further purification for the corrosion experiments. The major compound of *Alpinia galanga*, 1'-acetoxychavicol acetate was isolated and characterised from the rhizome of the plant using a previous method [23]. The other inhibitor compounds were purchased because the pure forms could not be obtained in quantities required for corrosion measurements.

2.2 Metal Specimen Preparation

Pre-cut mild steel specimen classification type, ASTM, A 29/ A 29M-03 [25] of dimension (2.0cm x 2.0cm x 0.25cm) having the composition (wt.%: 0.036 C; 0.172 Mn; 0.0146 P; 0.108 Ni; 0.0538 Cr; 0.082 Cu and balance Fe) were used for the experiments. Prior to every measurement, the samples were polished with successive grades of abrasive paper (grades 180, 400, 800, 1200 and 1500). The specimens were then degreased using acetone, rinsed with distilled water, dried and stored in the desiccator in readiness for polarisation measurements.

2.3 Corrosion Medium

The corrosion solution of 1 M HCl used was prepared by dilution of analytical reagent grade 37% HCl with distilled water. All inhibitors used were dissolved in the 1 M HCl at different inhibitor concentrations ranging from 0.001 M to 0.005 M. The 1 M HCl served as the blank solution in the experiment.

2.4 Potentiodynamic Polarization Tests

Potentiodynamic polarisation tests were performed using the AUTOLAB model PGSTAT30 instrument. The polarisation tests were performed with potential range of ±0.25 V at open circuit (OCP) scanned at the rate of 1 mVs^{-1}. The accurate Tafel portions of both the cathodic and anodic plots were extrapolated to obtain both the corrosion potential (E_{corr}) and corrosion current densities (I_{corr}). A freshly polished mild steel sample was engaged as working electrode in each measurement. Each experiment was performed in triplicate in order to obtain reproducible results. All electrochemical measurements were recorded after mild steel specimen immersion for 30 minutes in the corrosion cell containing the test solution so as to reach steady state.

The GPES electrochemical software was used to analyse the polarisation data. The percentage inhibition efficiencies of each inhibitor at different concentrations were calculated according to Eq. 1 as follows[26, 27].

$$IE\% = \frac{i^{o}_{corr} - i_{corr}}{i^{o}_{corr}} \times 100 \qquad (1)$$

Where i^{o}_{corr} and i_{corr} are corrosion current densities in the absence and presence of the inhibitors respectively.

2.5 QSAR Modeling

The data set consisted of 15 compounds comprising of 5 phenylpropanoids of *Alpinia galanga* and other related compounds. The data set was randomly separated into training and test sets in 70/30 ratio respectively. The data for the training set was used to fit the QSAR model, whereas the test data set was employed to evaluate the QSAR model. Results of potentiodynamic polarisation measurements were

used to investigate the inhibition efficiency of the inhibitors. Chemdraw Ultra version 8 software was used to draw the molecular structures of the inhibitor compounds which were successively optimised adopting Molecular Mechanics (MM2) method and Molecular Orbital Package (MOPAC) module of the software. The Dragon software Version 6.0 was employed to calculate the molecular descriptors of the optimised molecular structures of the inhibitors [17]. Pre-processing was implemented in order to reduce the 2,263 generated descriptors from the whole block of Dragon software to more consistent and useful descriptors. This was achieved by excluding the descriptors with constant values, descriptors having zero values for all compounds, descriptors with relative standard deviation less than 0.001 and descriptors with at least one missing value. A total of 601 calculated descriptors remained to build the QSAR model.

2.5.1 Elastic Net Method

The descriptors having the highest information are required for accurate selection from the entire molecular descriptors calculated. The problems associated with variable selection is finding the descriptors' subset in which the QSAR model built with only this subset would give a better predictive accuracy compared to that built with the complete set of calculated descriptors. Methods of classical variable selection cannot be used for high dimensional QSAR simply because the MLR cannot fit. In recent times, penalisation methods has been used for variable selection in high-dimensional data. In this work, the penalised multiple linear regression (PMLR) was applied using the elastic net penalty.

Generally, OLR assumes that the response variable is $y = \left(y_1, \ldots, y_n \right)'$ a linear combination of p molecular descriptors, $X_1, \ldots, X_p$ in addition to an unknown parameter vector $\beta = (\beta_1, \ldots, \beta_p)'$ as well as an additive error term $e = \left(e_1, \ldots e_2 \right)'$. When $n > p$ the usual estimation procedure for the parameter vector β is the minimisation of the residual sum of squares [28] with respect to β and is shown in Eq. 2 as:

$$\hat{\beta}_{OLS} = argmin_\beta RSS = argmin_\beta (y - X\beta)'(y - X\beta) \qquad (2)$$

Then, the OLS estimator $\hat{\beta}_{OLS} = \left(X'X \right)^{-1} X'y$ is obtained by solving Eq. (2). The OLS estimator is optimal within the class of linear unbiased estimators whenever there is no correlation in the molecular descriptors. Conversely, when there is a high correlation in the molecular descriptors, multicollinearity results in the regression model. This results into complications in the computation of the OLS estimator. In a case of high dimensional data where $n < p$, both the design matrix X and the matrix $X'X$ no longer have full rank p. Thus, $\left(X'X \right)^{-1}$ cannot be calculated and the OLS estimator cannot be resolved.

The penalisation procedures are established on penalty terms which yield unique estimations of the parameter vector β. Enhancement of the prediction accuracy is achievable by the shrinkage of the coefficients. There can also be an enhancement in the interpretability which is achieved by zeroing out some of the coefficients. The final QSAR regression models must contain only the relevant molecular descriptors which are easier to interpret.

The general form of the PMLR is well defined according to Eq. 3 as follows:

$$PMLR = (y - X\beta)'(y - X\beta) + P_\lambda (\beta) \qquad (3)$$

The estimation of the penalised parameter vector are achieved by minimising Eq. 3 with respect to β which results into Eq. 4.

$$\hat{\beta}_{PMLR} = \arg\min_\beta PMLR \qquad (4)$$

The penalty term $P\lambda\left(\beta \right)$ rests on the tuning parameter λ which controls the shrinkage strength. For the tuning parameter $\lambda = 0$, the obtained result is the OLS estimators. In contrast, for large values of λ, the influence of the penalty term on the coefficient estimates increases. Consequently, the penalty region determines the properties of the penalised estimated parameter vector, whereas desirable molecular descriptors are variable selection. Different forms of the penalty terms have been introduced in the literature such as ridge [29], the least absolute shrinkage and selection operator [30] and elastic net penalties.

Elastic net is a penalised method for the selection of variables. It is an introduction was made by Zou and Hastie [31] to deal with the drawbacks of other previous penalised methods by merging LASSO [30] and ridge [28] penalties and is defined by Eq. 5 [19].

$$\hat{\beta}_{elastic} = \arg\min_\beta (SSR + \lambda_1 \sum_{j=1}^{p} \left| \beta_j \right| + \lambda_2 \sum_{j=1}^{p} \beta_j{}^2) \qquad (5)$$

Elastic net is premised on two non-negative tuning parameters λ_1, λ_2.

2.5.2 Prediction of Assessment Criteria

The constructed QSAR model was evaluated and then validated to evaluate its predictive ability modelling 15 compounds comprising of 5 phenylpropanoids of *Alpinia galanga* and other related compounds as corrosion inhibitors. The use of two statistical criteria namely, the mean-squared error [32-34] and the leave-one-out validation (Q^2) (defined by Eqs. 6 and 7) were adopted for the training and the test sets.

$$MSE_{train} = \frac{\sum\limits_{i=1}^{n_{train}} (y_{i,train} - \hat{y}_{i,train})^2}{n_{train}} \qquad (6)$$

$$Q^2_{int} = 1 - \left[\frac{\sum\limits_{i=1}^{n_{train}} (y_{i,train} - \hat{y}_{i,train})^2}{\sum\limits_{i=1}^{n_{train}} (y_{i,train} - \overline{y})^2} \right] \qquad (7)$$

Additionally, the coefficient of determination R^2 were engaged to estimate the predictability power of the QSAR model. The higher the value of Q^2_{int} and the R^2, the higher is the prediction power of the model. Subsequently, validation of the constructed QSAR model was carried out using the test data by computing the MSE_{test} and Q^2_{ext} criteria. The two criteria are defined using the relation given by Algamal, *et al.* [20] as shown in Eqs. 8 and 9 shown as:

$$MSE_{test} = \frac{\sum\limits_{i=1}^{n_{test}} (y_{i,test} - \hat{y}_{i,test})^2}{n_{test}} \qquad (8)$$

$$Q^2_{ext} = 1 - \left[\frac{\sum\limits_{i=1}^{n_{test}} (y_{i,test} - \hat{y}_{i,test})^2}{\sum\limits_{i=1}^{n_{test}} (y_{i,test} - \overline{y}_{train})^2} \right] \qquad (9)$$

where n_{train} and n_{test} represent the training and testing sample sizes, the $y_{i,train}$, $y_{i,test}$, $\hat{y}_{i,train}$, and $\hat{y}_{i,test}$ stand for the IE values of the training data, testing data, and their corresponding predicted IE values, respectively, while $\overline{y}$ and $\overline{y}_{train}$ represent the mean of the all IE values and the mean of all the training IE values, respectively.

3.0 RESULTS AND DISCUSSION

3.1 Potentiodynamic Polarization Measurements

The polarisation plots showing the anodic and cathodic curves for the mild steel specimens in 1 M HCl in the absence and presence of the various concentrations of the inhibitors compounds were obtained. The Tafel plots shown in Figure 1 which compared the efficiencies of the inhibitors at the highest concentration of 0.005 M were analysed to calculate the current density (i_{corr}).
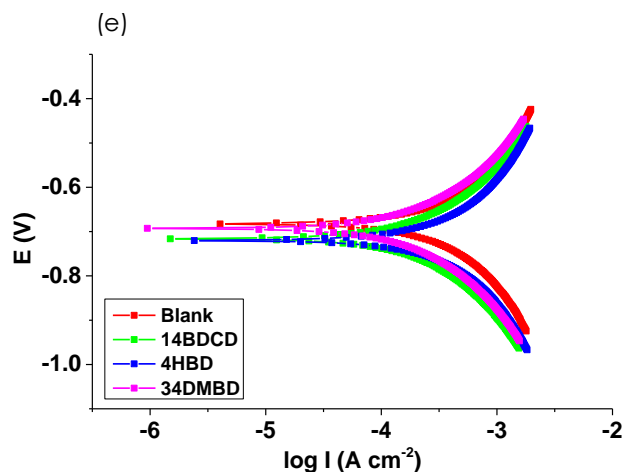
(e)
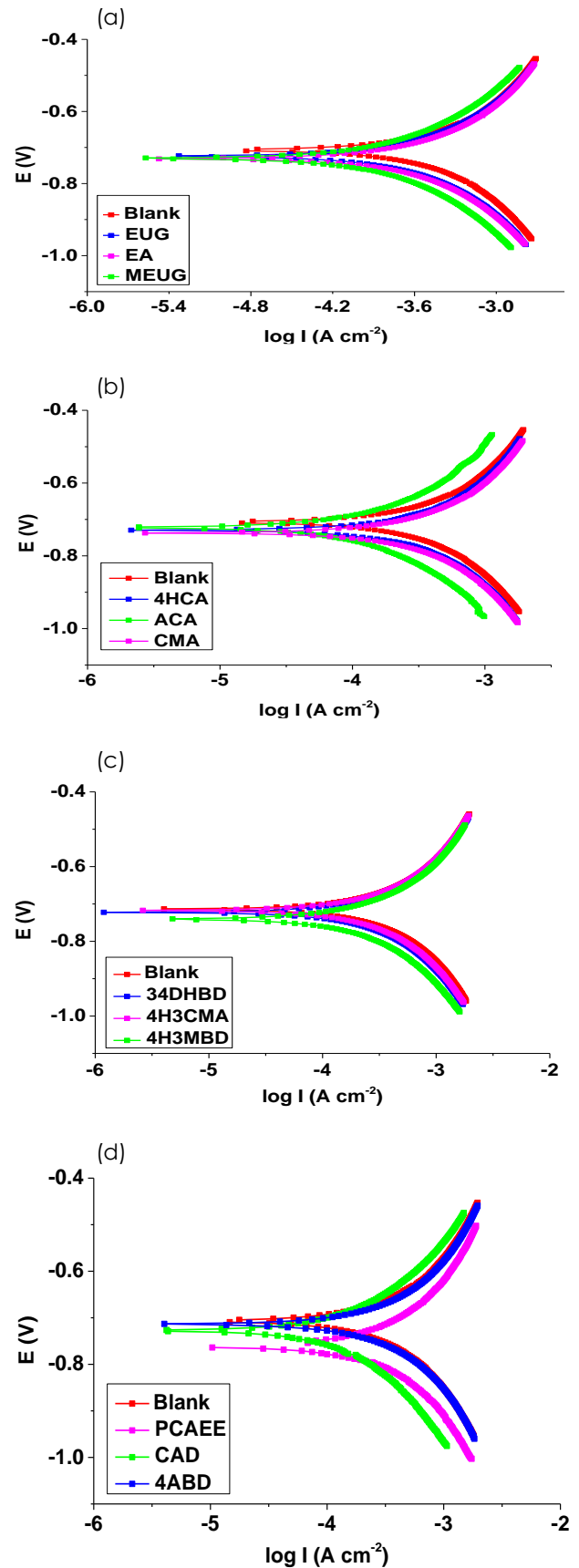
Figure 1 Tafel polarization curves for the inhibitors on mild steel in 1 M HCl with and without inhibitors, (a) EUG, EA, MEUG (b) 4HCA, ACA, CMA (c) 34DHBD, 4H3CMA, 4H3MBD (d) PCAEE, CAD, 4ABD (e) 14BDCD, 4HBD, 34DMBD

Tafel Polarisation Parameters for mild steel in 1 M HCl in the absence and presence of the various concentrations of the inhibitors are shown in Table 1 from where their percentage inhibition efficiencies (IE's) were calculated using Eq. 1.

Table 1 shows the calculated electrochemical parameters associated with the polarisation of mild steel in HCl at different concentrations. It reveals in the table that the corrosion current (I_{corr}) decreases with an increase in the concentration of inhibitor thus resulting into increased efficiencies of the inhibitors. This is as a result of increase in the coverage area of the mild steel by the inhibitor molecules.
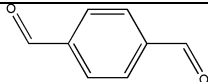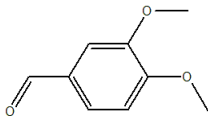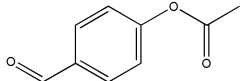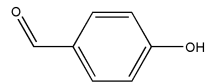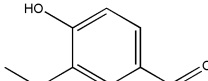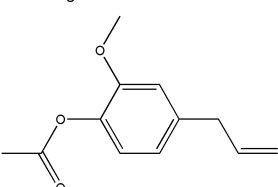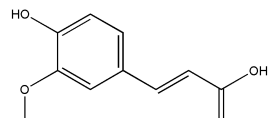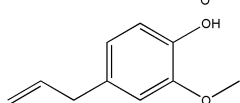
The 2D structures of the 15 compounds used in the study together with their abbreviated names are shown in the Table 2.

Table 1 Tafel Polarization Parameters and percentage inhibition efficiency (%*IE*) of the phenylpropanoids of *Alpinia galanga* and other related compounds calculated from Tafel polarization curves

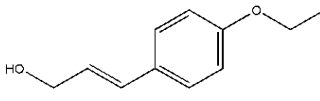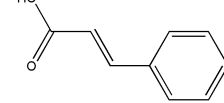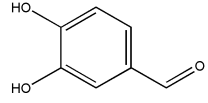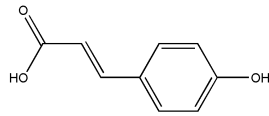| S/No	System	I_{corr} (10⁻⁴Acm⁻²)	b_c (V/dec)	b_a (V/dec)	E_{corr} (V)	R_c (mm/yr)	IE (%)
	Blank 1 M HCl	10.216	0.796	0.588	-0.714	11.871	–
1	ACA	1.601	0.306	0.241	-0.721	1.860	84.33
2	MEUG	1.802	0.244	0.224	-0.730	2.094	82.36
3	CAD	2.061	0.242	0.219	-0.693	2.394	79.83
4	14BDCD	2.241	0.379	0.257	-0.728	2.604	78.06
5	34DMBD	2.609	0.281	0.248	-0.716	3.032	74.46
6	4ABD	2.610	0.335	0.279	-0.711	3.033	74.45
7	4HBD	2.982	0.292	0.239	-0.720	3.465	70.81
8	4H3MBD	3.060	0.207	0.250	-0.741	3.555	70.05
9	EA	3.235	0.292	0.263	-0.724	3.760	68.33
10	4H3MCA	3.551	0.300	0.298	-0.758	4.127	65.24
11	EUG	3.951	0.350	0.304	-0.731	4.591	61.32
12	PCAEE	4.374	0.353	0.302	-0.760	5.083	57.18
13	CMA	4.678	0.377	0.323	-0.734	5.436	54.21
14	34DHBD	5.110	0.421	0.348	-0.723	5.938	49.98
15	4HCA	6.350	0.458	0.400	-0.707	7.381	37.84

Table 2 Names and structures of the inhibitors used as corrosion inhibitors

S/No	Inhibitor	Abbreviation	Structure
1	1'-acetoxychavicol acetate	ACA	
2	Methyl eugenol	MEUG	
3	Cinnamaldehyde	CAD	

4	1,4-benzenedicarboxaldehyde	14BDCD	
5	3,4-dimethoxybenzaldehyde	34DMBD	
6	4-acetoxybenzaldehyde	4ABD	
7	4-hydroxybenzaldehyde	4HBD	
8	4-hydroxy-3-methoxy-benzaldehyde	4H3MBD	
9	Eugenol acetate	EA	
10	4-hydroxy-3-methoxycinnamic acid	4H3MCA	
11	Eugenol	EUG	
12	P-Coumaryl alcohol ethyl Ether	PCAEE	
13	Cinnamic acid	CMA	
14	3,4-dihydroxybenzaldehyde	34DHBD	
15	4-hydroxycinnamic acid	4HCA	

There is a shift in a portion of the cathode along with a minor shift in the anodic portions in the polarisation curves when inhibitors are present. In the presence of the inhibitors, the corrosion potentials, E_{Corr} are more or less constant; therefore, they could be categorised as mixed-type inhibitors predominantly cathodic in effectiveness. These inhibitors behave as proposed by Dariva and Galio [35] in which they act as film forming compounds that cause the formation of precipitates on the metal surface blocking both anodic and cathodic sites.

The corrosion inhibition efficiencies of the inhibitors ranged from 37.8% in 4HCA to 84.3% in ACA. Differences in the inhibition efficiencies result from electronic structures of inhibitor molecules, functional groups, steric factor, aromaticity, electron density at donor atoms, molecular area and molecular weight of the inhibitor. These inhibitors are heterocyclic organic compounds containing oxygen. Heterocyclic compounds have been shown to have corrosion inhibition properties [36]. These inhibitors are highly efficient due to their excellent adsorption unto the mild steel surface. The presence of a lone pair of

electrons of the heteroatoms, the π-electrons and the multiple bonds of the inhibitors influence their adsorption on the mild steel or metal surface.

The higher the number of lone pair and π-electrons in an inhibitor molecule, the greater is the electron density on the molecule and these results into a stronger interaction between the inhibitor and the metal's surface. Considering the ACA molecule, the presence of four oxygen heteroatoms increases the electron density of the inhibitor and hence the enhancement of its inhibition efficiency.

In addition to the oxygen heteroatoms in ACA and MEUG, there is a presence of acetoxyl group suggesting that this group is responsible for their highest corrosion inhibition efficiency. The acetoxyl group has also been shown to be responsible for the biological activities of the compounds of *Alpinia galanga* [37].

The result of highest efficiency recorded for ACA goes further to suggest that the 1'- and 4-acetoxyl groups in the molecule are essential for the corrosion inhibition activity. The presence of the 2'-3' double bond in ACA, MEUG, EA and EUG is also found to enhance corrosion inhibition to certain extent with the efficiency least in EUG. This is as a result of the presence of electron donating substituents on the phenolic moiety. This makes the lone pair of electrons on phenolic oxygen less available for aromatic delocalisation. Eugenol extracted from the nail of giroflier and some of its derivatives have been investigated and found as good corrosion inhibitors [38].

Cinnamaldehyde is seen to have good corrosion inhibition properties. Its molecule contains a single conjugated system of unsaturated bonds (i.e. C=C, C=O and the aromatic ring). The presence of these active centres makes the compound a good corrosion inhibitor because all the active centres can be engaged in interaction with the mild steel surface. In the previous work of Avdeev, *et al*. [39] involving the use of unsaturated aldehydes as corrosion inhibitors, cinnamaldehyde has also been shown to have good corrosion inhibition properties.

The benzaldehydes are also shown to have good anticorrosion properties as a result of the presence of active sites in their structures. As they become more substituted, they tend to diminish in their anticorrosion activities.

The coumaric acid-type phenylpropanoids (PCAEE, CMA and 4HCA) give very low inhibition efficiency. These inhibitors lack the 4-acetoxyl group in their structures. This further confirms that both the acetoxyl group and the double bond 2'-3' double bond enhanced the corrosion inhibition activities.

The low inhibition efficiencies recorded for CMA and 4HCA are also attributable to their acidic nature. It is a known fact that the reaction of acids with metal leads to formation of corrosive products. The lowest inhibition efficiency recorded for 4HCA is due to the presence of a hydroxyl group which makes the compound more acidic than CMA. The hydroxyl group is electron withdrawing resulting into electron deficiency in the compound and hence the least inhibition efficiency (37.8%) is obtained for the compound. Moreover, the molecule is incapable of using its three active sites containing π electrons (the aromatic ring, the C=C bond and the C=O group) to interact with the mild surface, so only the aromatic and carbonyl groups participate in the surface interaction. As a result of involvement of only two active sites for establishing chemisorption process, there is a decline in the adsorption ability of the molecule unto mild steel surface.

3.2 QSAR Model

In order to select the most informatic descriptors with PMLR, the 15 compounds have been randomly divided into a training set of 70% and a test set of 30%. The training set was used to select the descriptors through finding the optimal value for the tuning parameters. The test set was then used to validate the quality of the selected descriptors. To find the optimal values of the pair tuning parameters (λ_1, λ_2) for elastic net, K-fold cross-validation method was used with K=3. Specifically, an earlier value of λ_2 for transforming the training set data into a new augmented training data set was given, where $0 \le \lambda_2 \le 100$. For each value of λ_2 a 3-fold cross-validation was performed to select λ_1. The best value for the pair of both tuning parameters was (0.131, 0.108).

Four molecular descriptors were selected to construct the QSAR model. The names of the selected molecular descriptors and their descriptions were summarised in Table 3. The final QSAR model using Eq. 3 produced Eq. 10 as follows:

$$IE = 171.101 + 34.701 PJI3 + 48.797 Mor11p \\ + 36.961 P\_VSA\_P\_3 \tag{10}$$

The reliability and the external predictive ability of the model is given by the following properties of the model.

$$R^2 = 0.927$$

$$MSE_{train} = 0.121 \text{, } MSE_{test} = 0.131$$

$$Q^2_{\text{int}} = 0.933 \text{, } Q^2_{ext} = 0.925$$

The statistical importance of the generated model was established by the parameters like coefficient of determination (R^2), coefficient of internal validation *(Leave-one-out cross-validation)* (Q^2_{int}), coefficient of external validation (Q^2_{ext}), as well as the mean squared errors (MSE). The coefficient of determination R^2, indicates the relative measure of quality of fit by the regression equation. Consequently, it characterises the part of the variant in the observed data that is depicted by the regression. High values of Q^2_{int} *and* Q^2_{ext} in relation with high value of R^2, is an indicator of a suitable model. Therefore, the obtained high values of the R^2, Q^2_{int} and Q^2_{ext} indicated the reliability and external predictive ability of the constructed QSAR model. In

addition, the low values of MSE_{train} and MSE_{test} indicate that the constructed QSAR model of phenylpropanoids of *Alpinia galanga* and other related compounds as corrosion inhibitors on mild steel in acidic medium is well established without existence of overfitting.

3.3 Interpretation of Descriptors

The interpretation of the descriptors provides an insight into the associated factors of the corrosion inhibition efficiencies of the inhibitor compounds in consideration. The PJI3, Mor11v, Mor11p and P_VSA_p_3 descriptors have been selected by the elastic net method in this work. The Petitjean shape index (PJI3) is a topological anisometric descriptor also termed, graph-theoretical shape coefficient [16] that got its derivation based on the molecular shape using the information of the geometry matrix. The PJI3 descriptor has a positive coefficient in the model which implies that its presence is favourable to the corrosion activity. However, it has been observed that QSAR models obtained using topological indexes are often more difficult to correlate to easily understand chemical concepts in that there is encoding of an indirect relationship between the molecular structure and the descriptor values [13].

The Mor11v (weighted by atomic van der Waals volume) and the Mor11p (weighted by atomic polarizabilities) both belong to the 3D-MoRSE descriptors (3D-Molecule Representation of Structures based on Electron diffraction descriptors). The 3D-MoRSE descriptors are indicators of size, mass and volume of the molecules [40]. The Mor11v has a negative coefficient in the model implying that atomic volumes are unfavourable to the corrosion activity while Mor11p with positive coefficient implies that atomic polarisabilities are favourable to the corrosion activity. Atomic polarisabilities are usually accompanied by high chemical activity and low kinetic stability to produce soft molecules [41] which act as better corrosion inhibitors.

The P_VSA_p_3 descriptors belong to a set of 2D descriptor group having 52 descriptors that describe the electrostatic, steric and pharmacophoric properties in terms of molecular surface [42]. In the model, the P_VSA_p_3 descriptor has a positive coefficient implying that the presence of hydrophilic substituent groups in the aromatic ring is favourable to the corrosion activity. This is supported in the inhibitors (e.g. ACA, MEUG, EA) with the acetoxyl groups having the highest efficiency.

The group type and description of the selected descriptors by elastic net method are summarised in the Table 3.

Table 3 The selected descriptor names and their descriptions

Descriptor	Group type	Description
PJI3	Geometrical descriptors	3D Petitjean shape index
Mor11v	3D-MoRSE descriptors	signal 11 / weighted by van der Waals volume
Mor11p	3D-MoRSE descriptors	signal 11 / weighted by polarizability
P_VSA_p_3	P_VSA-like descriptors	P_VSA-like on polarizability, bin 3

4.0 CONCLUSION

Experimental study was carried out to investigate the inhibition efficiencies of 15 compounds comprising of five phenylpropanoids from *Alpinia galanga* and other related compounds on mild steel corrosion in 1 M HCl. The experimental results showed the effective performance of the phenylpropanoids and related compounds as corrosion inhibitors. Theoretical high dimensional QSAR modeling study was conducted using the obtained experimental data and the molecular descriptors calculated based on the structures of the inhibitors using the Dragon software. Since the classical variable selection methods cannot deal with high dimensional QSAR, the penalised multiple linear regression based on elastic net penalty was applied and four significant descriptors (i.e. PJI3, Mor11v, Mor11p, and P_VSA_p_3) were selected to develop high dimensional QSAR model. These descriptors significantly described the corrosion inhibition behaviour of the studied inhibitors. The molecular structure requirements of the phenylpropanoids as corrosion inhibitors have also been elucidated.

Acknowledgement

References

[1] Singh, A., Ebenso, E. E., and Quraishi, M. A. 2012 Corrosion Inhibition of Carbon Steel in HCl Solution by Some Plant Extracts. *International Journal of Corrosion*. 2012: 1-20.

[2] Li, X. and Deng, S. 2012. Inhibition Effect of Dendrocalamus brandisii Leaves Extract on Aluminum in HCl, H_3PO_4 Solutions. *Corrosion Science*. 65: 299-308.

[3] Rani, B. E. A. and Basu, B. B. J. 2012 Green Inhibitors for Corrosion Protection of Metals and Alloys: An Overview. *International Journal of Corrosion*. 2012: 1-15.

[4] Deng, S. and Li, X. 2012 Inhibition by Ginkgo Leaves Extract of the Corrosion of Steel in HCl and H2SO4 Solutions, *Corrosion Science*. 55: 407-415.

[5] Ji, G., Shukla, S. K., Dwivedi, P., Sundaram, S., and Prakash, R. 2011. Inhibitive Effect ofArgemone mexicanaPlant Extract on Acid Corrosion of Mild Steel. *Industrial & Engineering Chemistry Research*. 50: 11954-11959.

[6] Hooshmand Zaferani, S., Sharifi, M., Zaarei, D., and Shishesaz, M. R. 2013 Application of Eco-friendly Products as Corrosion Inhibitors for Metals in Acid Pickling Processes -A Review. *Journal of Environmental Chemical Engineering*. 1: 652-657.

[7] 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.

[8] Wang, W.-X., Si, H., and Zhang, Z. 2012 Quantitative Structure–activity Relationship Study on Antitumour Activity of a Series of Flavonoids, *Molecular Simulation*. 38: 38-44.

[9] Rasuleva, B. F., Abdullaev, N. D., Syrov, V. N., and Leszczynska, J. 2005 A Quantitative Structure-Activity Relationship (QSAR) Study of the Antioxidant Activity of Flavonoids. *Molecular Informatics*. 24: 1056-1065.

[10] Matsuda, H., Ando, S., Morikawa, T., Kataoka, S., and Yoshikawa, M. 2005 Structure–activity Relationships of 1'S-1'-acetoxychavicol Acetate for inhibitory Effect on NO Production in Lipopolysaccharide-activated Mouse Peritoneal Macrophages. *Bioorganic and Medicinal Chemistry Letters*. 15: 1949-1953.

[11] Mousavi, M., Safarizadeh, H., and Khosravan, A. 2012 A New Cluster Model Based Descriptor for Structure-Inhibition Relationships: A Study of the Effects of Benzimidazole, Aniline and Their Derivatives on Iron Corrosion. *Corrosion Science*. 65: 249-258.

[12] Szeptycka, B. 2005 Quantum-chemical Modeling (QCM) and Quantum Structure-activity relationship Modeling (QSARM) of acid Pickling Inhibition. *Materials and Manufacturing Processes*. 20: 9-21.

[13] Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., *et al.* 2014 QSAR Modeling: Where Have You Been? Where are You Going To? *Journal of Medicinal Chemistry*. 57: 4977-5010

[14] He, L. and Jurs, P. C. 2005 Assessing the Reliability of a QSAR Model's Predictions. *Journal of Molecular Graphics and Modelling*. 23: 503-523.

[15] 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.

[16] Todeschini, R. and Consonni, V. 2009. *Molecular Descriptors for Chemoinformatics*. John Wiley & Sons.

[17] DRAGON. Software for Molecular Descriptor Calculation, version 6.0, by Todeschini R, Consonni V, Mauri A, Pavan M. Milan, Italy: Talete srl. [Online]. Available: http://www.talete.mi.it/.

[18] Filzmoser, P., Gschwandtner, M., and Todorov, V. 2012 Review of Sparse Methods in Regression and Classification with Application to Chemometrics. *Journal of Chemometrics*. 26: 42-51.

[19] Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., and Aziz, M. 2015

[20] Algamal, Z. Y., Lee, M. H., and Al-Fakih, A. M. 2016 High-dimensional Quantitative Structure–activity Relationship Modeling of Influenza Neuraminidase a/PR/8/34 (H1N1) Inhibitors based on a Two-stage Adaptive Penalized Rank Regression. *Journal of Chemometrics*. 30: 50-57.

[21] Al-Fakih, A. M., Aziz, M., Abdallah, H. H., Algamal, Z. Y., Lee, M. H., and Maarof, H. 2015 High Dimensional QSAR Study of Mild Steel Corrosion Inhibition in Acidic Medium by Furan Derivatives. *International Jourmal of Electrochememical Science*. 10: 3568-3583.

[22] Arambewela, L. and Silva, R. 2006. Sri Lankan Medicinal Plant Monographs and Analysis: *Alpinia galanga*. ed Colombo: National Science Foundation.

[23] Matsuda, H., Morikawa, T., Managi, H., and Yoshikawa, M. 2003. Antiallergic Principles from Alpinia galanga: Structural Requirements of Phenylpropanoids for Inhibition of Degranulation and Release of TNF-a and IL-4 in RBL-2H3 cells. *Bioorganic and Medicinal Chemistry Letters*. 13: 3197-3202.

[24] Korkina, L., Kostyuk, V., De Luca, C., and Pastore, S. 2011. Plant Phenylpropanoids as Emerging Anti-inflammatory Agents. *Mini Reviews in Medicinal Chemistry*. 11: 823-835.

[25] Bringas, J. E. 2004. *Handbook of Comparative World Steel Standards*. 3rd ed. USA: ASTM International.

[26] Gunasekaran, G. and Chauhan, L. R. 2004 Eco Friendly Inhibitor for Corrosion Inhibition of Mild Steel in Phosphoric Acid Medium. *Electrochimica Acta*. 49: 4387-4395.

[27] Amira, W. E., Rahim, A., Osman, H., Awang, K., and Raja, P. B. 2011. Corrosion Inhibition of Mild Steel in 1 M HCl Solution by Xylopia Ferruginea Leaves from Different Extract and Partitions. *International Jourmal of Electrochememical Science*. 6: 2998.

[28] Frisch, M. J., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., *et al.* 2009. Gaussian 09, revision A. 1, Gaussian Inc., Wallingford, CT.

[29] Hoerl, A. E. and Kennard, R. W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*. 12: 55-67.

[30] Tibshirani, R. 1996 Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. 58: 267-288.

[31] Zou, H. and Hastie, T. 2005 Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 67: 301-320.

[32] Elmsellem, H., Elyoussfi, A., Steli, H., Sebbar, N. K., Essassi, E. M., Dahmani, M., *et al.* 2016. The Theobromine (chocolate) as Green Inhibitor of Mild Steel Corrosion in hydrochloric Acid: Electrochemical And Theoretical Quantum Studies.*Der Pharma Chemica*. 8: 248-256.

[33] Al-Fakih, A. M., Algamal, Z. Y., Lee, M. H., Abdallah, H. H., Maarof, H., and Aziz, M. 2016. Quantitative Structure–activity Relationship Model For Prediction Study Of Corrosion Inhibition Efficiency Using Two - Stage Sparse Multiple Linear Regression. *Journal of Chemometrics*. 30: 361-368.

[34] Algamal, Z., Lee, M., Al-Fakih, A., and Aziz, M. 2016. High-dimensional QSAR Modelling Using Penalized Linear Regression Model with L 1/2-norm. *SAR and QSAR in Environmental Research*. 27: 703-719.

[35] Dariva, C. G. and Galio, A. F. 2014. Corrosion Inhibitors– Principles, Mechanisms and Applications. *Developments in Corrosion Protection*. ed: In Tech,

[36] Soltani, N., Tavakkoli, N., Khayat Kashani, M., Mosavizadeh, A., Oguzie, E., and Jalali, M. 2013. Silybum Marianum Extract as a Natural Source Inhibitor for 304 Stainless Steel Corrosion in 1.0 M HCl. *Journal of Industrial and Engineering Chemistry*. 20: 3217-3227.

[37] Matsuda, H., Pongpiriyadacha, Y., Morikawa, T., Ochi, M., and Yoshikawa, M. 2003. Gastroprotective Effects of Phenylpropanoids from the Rhizomes of Alpinia Galanga

in Rats: Structural Requirements and Mode of Action. *European Journal of Pharmacology*. 471: 59-67.

[38] Chaieb, E., Bouyanzer, A., Hammouti, B., and Benkaddour, M. 2005. Inhibition of the Corrosion of Steel in 1M Hcl by Eugenol Derivatives. *Applied Surface Science*. 246: 199-206.

[39] Avdeev, Y. G., Kuznetsov, Y. I., and Buryak, A. K. 2013. Inhibition of Steel Corrosion by Unsaturated Aldehydes in Solutions of Mineral Acids. *Corrosion Science*. 69: 50-60.

[40] Soltani, S., Haghaei, H., Shayanfar, A., Vallipour, J., Asadpour Zeynali, K., and Jouyban, A. 2013. QSBR Study of Bitter Taste of Peptides: Application of GA-PLS in Combination with MLR, SVM, and ANN Approaches. *BioMed research International.*

http://dx.doi.org/10.1155/2013/501310.

[41] Udhayakalaa, P., Rajendiranb, T. V., and Gunasekaranc, S. 2012. Theoretical Approach to the Corrosion Inhibition Efficiency of some Pyrimidine Derivatives using DFT Method. *Journal of Computational Methods in Molecular Design*. 2: 1-15.

[42] Balaji, S., Karthikeyan, C., Moorthy, N. H. N., and Trivedi, P. 2004. QSAR Modelling of HIV-1 Reverse Transcriptase Inhibition by benzoxazinones using a Combination of P_VSA and Pharmacophore Feature Descriptors. *Bioorganic and Medicinal Chemistry Letters*. 14: 6089-6094.