QSAR models development for corrosion inhibitor properties of furan derivatives against mild steel.

Saprizal Hadisaputra, Aditya

Chemistry Education Division, University of Mataram. Jalan Majapahit No 62, Mataram, 83125, Indonesia. Email: rizal@unram.ac.id; agus_ap@unram.ac.id; aliefman@unram.ac.id

Abstract

Statistical modeling of the corrosion inhibition properties with furan derivative inhibitors against mild steel has been investigated using the quantitative structure-property relationship QSAR method. This modeling is based on the correlation between corrosion inhibition efficiency (IE%) and several electronic and structural properties of compounds such as EHOMO (Highest Occupied Molecular Orbital energy), ELUMO (Lowest Unoccupied Molecular Orbital energy), EL-H (Gap energy), μ (dipole moment), IP (ionization potential), EA (electron affinity), η (hardness), σ (softness), χ (electronegativity), ΔN (fraction of electron transfer), ω (electrophilicity index), ΔEB-D (back-donation energy), Log P, Vm (critical volume), and Mr (relative molecular mass). These properties were calculated using DFT on B3LYP/6-31G(d). Statistically, they analyzed using four methods: partial least squares regression PLS, principal component regression PCR, multiple linear regression MLR, and principal component analysis PCA. The best QSAR modeling results are by PCR statistical analysis. It is proven by the validation results (R2 = 0.976; R2adj = 0.904) and analysis of collinearity in the data. The predictions of the four furan-derived compounds from PCR modeling gave promising results, especially for the BMOPF (IE% pred = 169.37) and the FMP (IE% pred = 100.81).

Introduction

Corrosion is a natural phenomenon in which metals and their alloys try to return to a more stable thermodynamic state due to reactions with their surroundings [1,2]. Corrosion of metal compounds and their alloys is one of the main problems in the industry today. It is considered a problem because it can be a source of environmental pollution, for example, in the carbon steel industry, where acid is used for pickling and descaling. Researchers are therefore urged to develop methods to protect these minerals and lessen their environmental impact. As a result, corrosion inhibitors are one of the most widely used methods for preventing corrosion and controlling metal deterioration [3,4]. Of course, depending on how corrosive an environment is, different corrosion techniques are used. For instance, organic compounds are the most prevalent class of inhibitors to solve the acidic media issue. These compounds' structural and electrical characteristics play a significant role in their behavior and performance. These substances typically have double and triple bonds in their chemical structure and heteroatoms, including nitrogen, oxygen, sulfur, and phosphorus, enclosed in aromatic rings. It may increase the material's surface adsorption value and increase its reactivity [5-9].

Furan-derived compounds have excellent promise as corrosion inhibitors in acidic conditions [10,11]. These organic compound derivatives have been shown in experiments to take free electrons from the metal surface by utilizing unoccupied orbitals of lower energy levels and donating electrons to the metal surface to create coordinating covalent bonds. It makes it easier

for the inhibitor to stick to the metal surface. Its adsorption enables the blocking of the active sites by decreasing the rate of dissolving of the metal along with the release of protons and increasing the coverage ratio. The enhanced inhibition of the effects of these compounds on corrosion indicates it.

This approach to quantum chemistry has proven to be very useful in determining the molecular structure and the electronic structure and characteristics of the reactive sites to elucidate the reaction mechanism of corrosion inhibition processes. In addition, this approach is also useful for understanding the relationship between corrosion inhibition efficiency and several corrosion inhibitor molecular indices or their quantum parameters. Quantitative Structure-Property Relationship (QSPR) has recently been widely used for quantitative analysis of corrosion inhibition processes to find a consistent relationship between variations in the index of molecular properties and the inhibitory activity of several compounds [12-18].

This study's use of a mathematical approach can produce qualitative and quantitative data that can aid in improving knowledge of the process of corrosion inhibition. Finding a stable structural property link between corrosion inhibition efficiency (IE%) and molecular electronic characteristics is one of the goals of this research. The electronic properties include EHOMO (Highest Occupied Molecular Orbital energy), ELUMO (Lowest Unoccupied Molecular Orbital energy), EL-H (Gap energy), μ (dipole moment), IP (ionization potential), EA (electron affinity), η (hardness), σ (softness), χ (electronegativity), ΔN (fraction of electron transfer), ω (electrophilicity index), ΔEB-D (back-donation energy), Log P, Vm (critical volume), and Mr (relative molecular mass) obtained by computation, DFT B3LYP method and basis set 6-31G (d) for 13 furan derivatives. They have been studied experimentally as a corrosion inhibitor on mild steel in 250 mL of 1M HCl. with an inhibitor of 0.005 M [19,20].

The current study focused on statistical analysis of the data was evaluated through a comparison of 4 mathematical regression models, namely: Partial Least Square (PLS), Principal Component Regression (PCR), Multiple Linear Regression (MLR), and Principal Component Analysis (PCA), to understand depth the mechanism of corrosion inhibition. In addition, this research was also conducted to determine the relevant and mutually influential molecular indices in the variation of corrosion inhibition of the compounds studied. Finally, the results of the established mathematical equations will make it possible to estimate the value of corrosion inhibitors of similar compounds and make it easier for other researchers to synthesize related compounds because of the promising prediction results.

Methodology

2.1 Experimental Data

This study used 13 furan derivative compounds to find the QSPR between the potential corrosion inhibitors and their molecular structures (Figure 1). Furthermore, the corrosion inhibitor of the other four furan derivatives was calculated using the best QSPR equation obtained. This research is a complementary statistical study and another study of previous studies on corrosion inhibitors of furan compounds that have been tested experimentally. The furan-derived compounds whose QSPR will be studied can be seen in Table 1, and the furan-derived compounds whose predicted corrosion inhibitor values will be sought can be seen in Table 2.

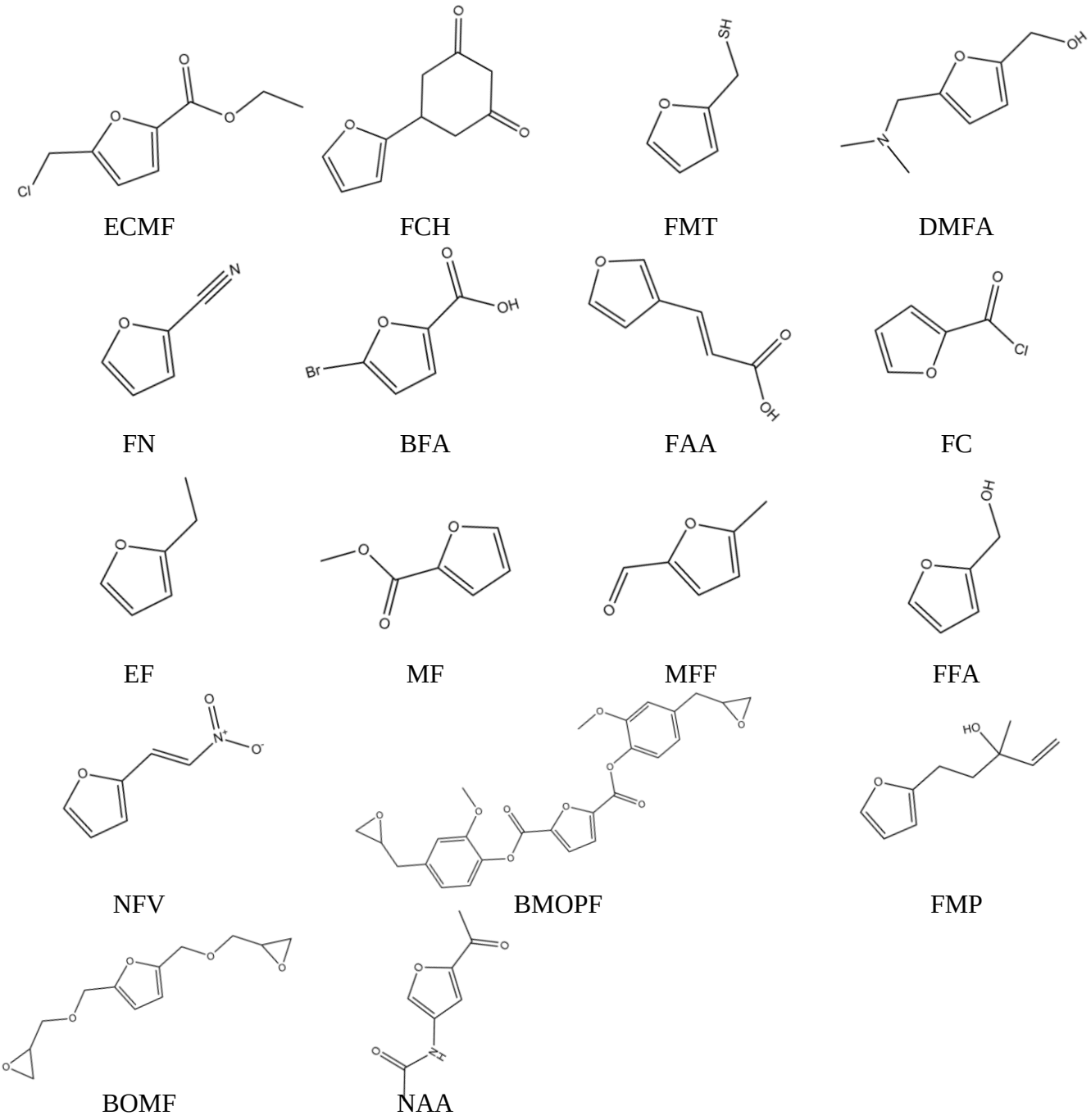


Figure 1. Structure of 13 furan derivatives for QSAR model development and predicted corrosion inhibitor values

Table 1. Furan-derived compounds whose corrosion inhibitor values have been obtained experimentally

No.	IUPAC Name	Abbreviation	IE exp (%) [19]

1.	Ethyl 5-(chloromethyl)-2-furoate	ECMF	96.54
2.	5-(2-furyl)-1,3-cyclohexanedione	FCH	89.93
3.	2-furanmethanethiol	FMT	89.44
4.	2-furonitrile	FN	89.03
5.	5-bromo-2-furoic acid	BFA	88.60
6.	Trans-3-furanacrilyc acid	FAA	78.24
7.	2-ethylfuran	EF	77.34
8.	Methyl 2-furoate	MF	76.75
9.	5-methylfurfural	MFF	76.14
10.	5-(dimethylaminomethyl)furfuryl alcohol	DMFA	71.99
11.	2-furoyl chloride	FC	64.25
12.	Furfuryl alcohol	FFA	53.93
13.	2-(2-nitrovinyl)furan	NVF	35.96

Table 2. Furan-derived compounds whose predicted corrosion inhibitor values

No.	IUPAC Name	Abbreviation	Ref
1.	Bis(2-methoxy-4-(oxiran-2-ylmethyl)phenyl) furan-2,5-dicarboxylate	BMOPF	[21]
2.	2,5-bis((oxiran-2-ylmethoxy)methy)furan	BOMF	[22]
3.	5-(2-furyl)-3-methyl-1-penten-3-ol	FMP	[23]
4.	N-(5-acetylfuran-3-yl)acetamide	NAA	[24]

Computational Calculation

All molecular geometries were optimized by Gaussian 09 Software [25]. Quantum chemical calculations used in geometry optimization are DFT with the B3LYP function and on the 6-31G (d) basis set. It is also used in determining other physical-chemical descriptors that will also be used in QSPR. Monte Carlo simulations of furan derivatives were carried out using the Material Studio application [26]. This study uses iron or Fe crystals with a (1 1 0) surface. The thickness value of the iron used is 8. Furthermore, in the supercell section, the U and V values are used at 20. In the crystal option, the value of the vacuum thickness is 15. Finally, a simulation is carried out by selecting the adsorption location option. The number of inhibitor compounds is set to 1 and water to 100. The simulation in this study also uses atomic targets at the top of the Fe layer so that later the results obtained can be more accurate.

Statistic analysis

A statistical study was carried out using the XLSTAT premium 2021 application [27] on 13 furan-derived compounds to find the QSPR between the corrosion inhibitor value (IE%) and the intrinsic electronic and structural properties of these compounds. In this study, three statistical analyzes were selected to find the QSPR model: Partial Least Square (PLS), Principal Component Regression (PCR), and Multiple Linear Regression (MLR) [13, 28-31]. Furthermore, Principal Component Analysis (PCA) was carried out, which made it possible to see the relationship between descriptors and check redundancy and collinearity between the descriptors studied so that later the best statistical analysis could be determined in the QSPR study of these furan derivatives compounds.

Validation

Validation tests were conducted to evaluate the explanatory quality and prediction level of the QSPR modeling. In this study, only internal validation was carried out. Internal validation is based on several statistical parameters, such as the coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}), prediction coefficient of determination (R^2), PRESS value, and standard deviation (SD). Next, we look for the coefficient of determination cross-validation (R^2_{cv}) to evaluate the relevance of this procedure [32-34]. R^2_{cv} is expressed in the form,

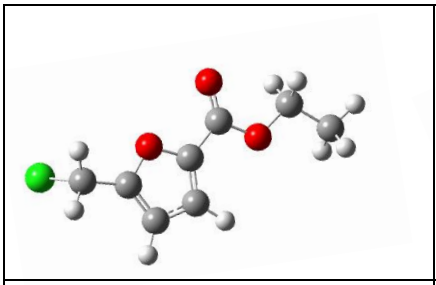
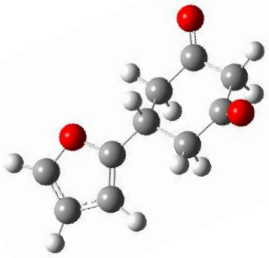
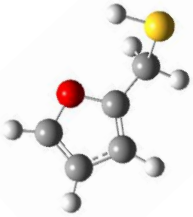
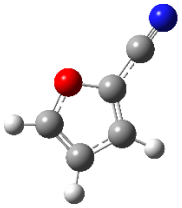
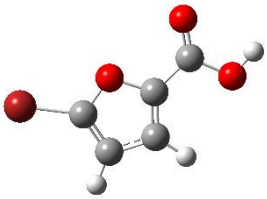
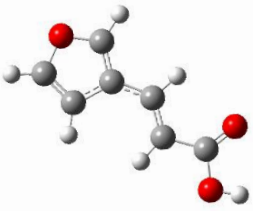
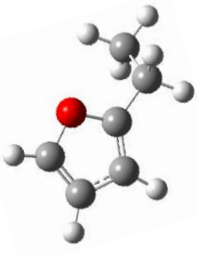
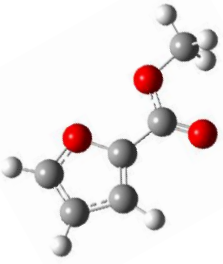
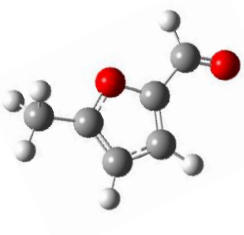
$$R^2_{cv} = 1 - \frac{\sum_{i=1}^{n}(\text{IE exp} - IE\ cal)}{\sum_{i=1}^{n}(\text{IE exp} - IE\ avg)}$$

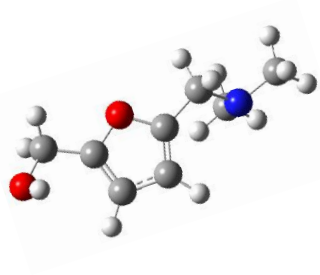
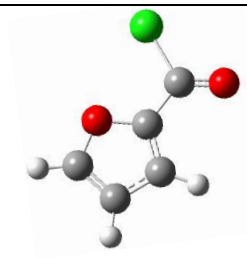
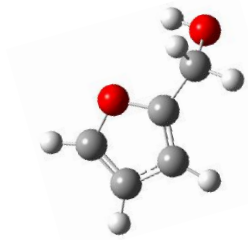
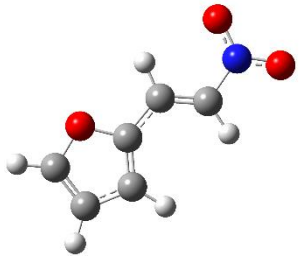
Where IE_{exp} and IE_{cal} are the experimental corrosion inhibitor values and the calculated corrosion inhibitor values, respectively. IE_{avg} is the average IE_{exp} value.

Results and Discussion

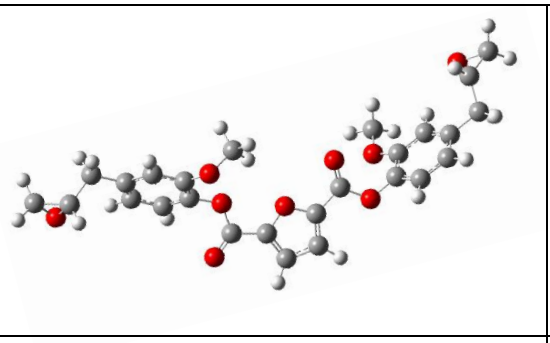
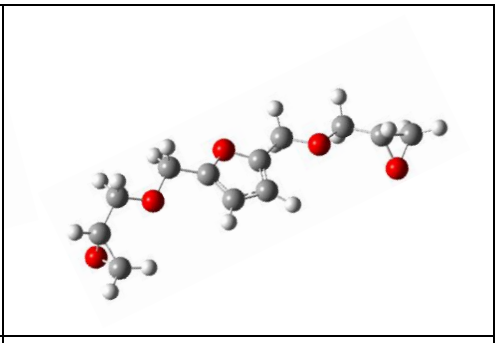
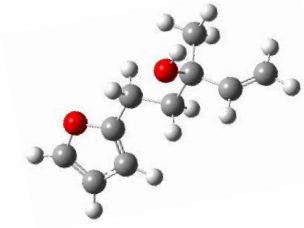
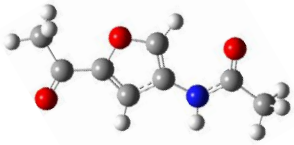
Molecular Geometry

The molecular geometry of the investigated compounds was determined by optimizing all structural parameters using Density Functional Theory (DFT) at the B3LYP/6-31G level (d) [35-37]. The geometry obtained is qualified through the minimum energy of the molecular structure and is proved by the absence of imaginary frequencies. The final optimized geometry can be seen in Figure 2.

		
ECMF	FCH	FMT
		
FN	BFA	FAA
		
EF	MF	MFF

		
DMFA	FC	FFA
		
NFV		

(a)

	
BMOPF	BOMF
	
FMP	NAA

(b)

Figure 2. (a) Furan derivative compounds (experimental) which have been optimized geometry with B3LYP/6-31G (d); (b) Compounds derived from furan (prediction) which have been optimized for geometry with B3LYP/6-31G (d)

Molecular Descriptor Calculations

This corrosion inhibition phenomenon is dependent on various other factors, so experimental studies alone cannot explain the interaction behavior between metals and inhibitors. Therefore, combining experimental and quantum studies such as QSPR is urgently needed. In this study, the QSPR approach focuses on the correlation of the intrinsic properties of each molecule with its corrosion inhibition potential. To explore this relationship, researchers developed many models to link corrosion inhibitors with some of their molecular properties. So, the value of corrosion inhibitors can be predicted through molecular descriptors and can explain the mechanism of corrosion inhibition.

According to the literature, the most relevant descriptors capable of influencing the adsorption of inhibitor molecules onto metal surfaces are electronic, structural, and lipophilic indices. By considering the three index descriptors, this study uses E_{HOMO} (Highest Occupied Molecular Orbital energy), E_{LUMO} (Lowest Unoccupied Molecular Orbital energy), E_{L-H} (Gap energy), μ (dipole moment), IP (ionization potential), EA (electron affinity), η (hardness), σ (softness), χ (electronegativity), ΔN (fraction of electron transfer), ω (electrophilicity index), ΔE_{B-D} (back-donation energy), Log P, Vm (critical volume), and Mr (relative molecular mass) as a descriptor. The descriptor values obtained are illustrated in Table 3

QSPR Study

Partial Least Square (PLS)

Partial least squares is a multivariate statistical technique that can handle many response variables as well as explanatory variables at once. Regression Partial Least Squares (PLS) is an equation that is commonly used, especially when a large number of molecular descriptors for testing a compound are used [38].

PLS is intended to predict quantitatively the corrosion inhibitor activity of the compounds studied . PLS modeling is generally expressed in the following equation:

$$IE\ cal\ (\%) = a_0 + a_1E_{HOMO} + a_2E_{LUMO} + a_3E_{L-H} + a_4\mu + a_5IP + a_6EA + a_7\chi + a_8\sigma + a_9\eta + a_{10}\Delta N + a_{11}\omega + a_{12}Log\ P + a_{13}Mr + a_{14}V_m + a_{15}\Delta E_{B-D} \quad (1)$$

where a_0 is the constant of the regression; a_{1-15} represents the regression coefficient of E_{HOMO}, E_{LUMO}, E_{L-H}, μ, IP, EA, χ, σ, η, ΔN, ω, Log P, Mr, V_m, and ΔE_{B-D}, respectively.

PLS modeling that has been analyzed from descriptor data, accompanied by statistical parameter values, is as follows:

$$IE\ exp\ (\%) = 69.284 + 0.005E_{HOMO} + 0.459E_{LUMO} + 0.913E_{L-H} - 0.377\mu - 0.005IP - 0.459EA - 0.503\chi - 17.790\sigma + 1.826\eta + 0.223\Delta N - 0.005\omega - 0.424Log\ P + 0.001Mr + 0.0002V_m - 7.307\Delta E_{B-D} \quad (2)$$

$N = 13 \quad R^2 = 0.104 \quad SD = 17.392 \quad R^2_{cv} = 667.6667$

Other validation values can be seen in Table 4

Table 4. PLS Validation

Observations	13.000
Sum of weights	12.000
DF	10.000
R^2	0.104
Std. deviation	17.392
MSE	252.080
RMSE	15.877

Table 5. Predicted values and PLS modeling residuals

Observation	Weight	IE exp (%)	Pred(IE exp (%))	Residual	Std. residual	Std. dev. on pred. (Mean)
ECMF	1	96.540	74.683	21.857	1.427	5.169
FCH	1	89.930	75.905	14.025	0.916	5.022
FMT	1	89.440	80.631	8.809	0.575	6.599
FN	1	89.030	73.802	15.228	0.994	5.422
BFA	1	88.600	75.477	13.123	0.857	5.045
FAA	1	78.240	73.736	4.504	0.294	5.446
EF	1	77.340	83.227	-5.887	-0.384	8.365
MF	1	76.750	77.187	-0.437	-0.029	5.138
MFF	1	76.140	71.754	4.386	0.286	6.386
DMFA	1	71.990	83.545	-11.555	-0.755	8.602
FC	1	64.250	71.620	-7.370	-0.481	6.463
FFA	1	53.930	82.121	-28.191	-1.841	7.569
NVF	1	35.960	64.454	-28.494	-1.861	11.832

Based on the PLS modeling in equation 2, the significance of each descriptor vs. the standard regression coefficient is presented in Figure 3.

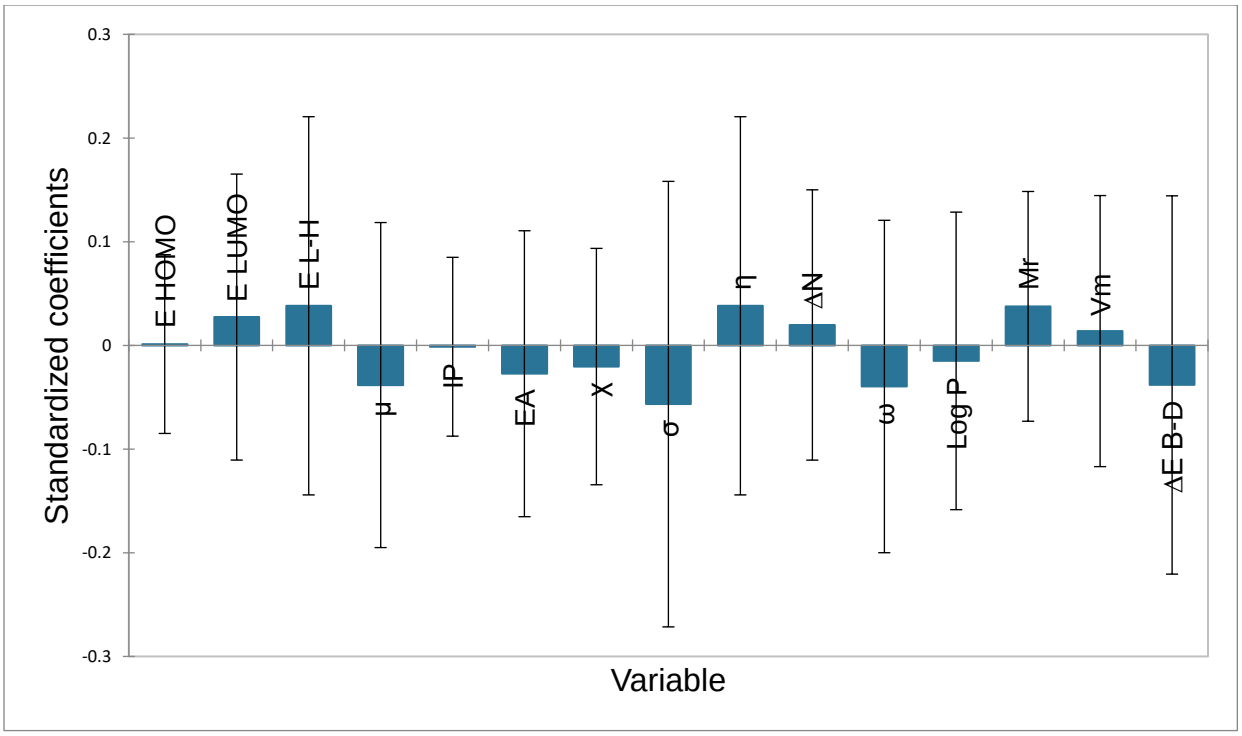


Figure 3. The coefficients of the standard versus the variables in the specified PLS model

Figure 3 shows that the relevance of corrosion inhibitor values based on molecular structure differs from one descriptor to another. However, because the descriptors in the PLS modeling do not have the same units, the standard coefficients obtained are only estimates without a real scale. It implies that these standardized coefficients cannot be used to determine the real relative significance of each descriptor in the regression analysis. Therefore, its usefulness is limited to determining the positive or negative effect of the molecular index on the anticorrosive property under investigation [39-41].

In terms of statistical parameters, the value of the coefficient of determination (R^2 = 0.104), standard deviation (SD = 17.392), dan R^2_{cv} = 667.6667, it appears that PLS does not have a good fit in predicting the value of corrosion inhibitors. The residual error value is still large, and there is a large difference between IE_{exp} and IE_{pred}.

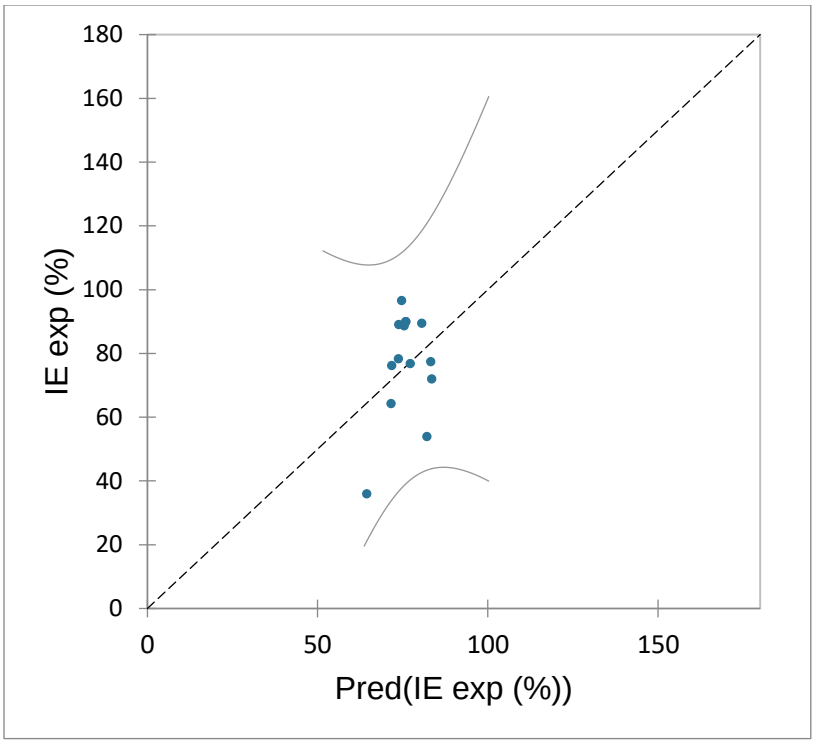


Figure 4. Correlation between IE_{exp} and IE_{Pred} with PLS modeling results

Principal Component Regression (PCR)

The general equations used in PCR are identical to those in PLS regression analysis. PCR modeling results are expressed in the following equation:

$IE\ exp\ (\%) = 1348.903 + 69.943E_{HOMO} + 35.630E_{LUMO} + 47.393E_{L\text{-}H} - 26.791\mu - 69.943IP - 35.630EA - 51.559\chi - 253.194\sigma + 94.787\eta - 173.134\Delta N + 3.647\omega + 12.557Log\ P + 0.184Mr - 0.018V_m - 379.151\Delta E_{B\text{-}D}$.....(3)

N = 13 $R^2 = 0.976$ $R^2_{adj} = 0.904$ PRESS = 2490.589

Other validation values can be seen in Table 6

Table 6. PCR validation

Observations	13
Sum of weights	13
DF	3
R^2	0.976
Adjusted R^2	0.904
MSE	27.033
RMSE	5.199
MAPE	2.700
DW	1.552

Cp	10.000
AIC	43.799
SBC	49.449
PC	0.184
Press	2940.589
Q^2	0.129

Table 7. Predicted values and residuals from PCR modeling

Observation	Weight	IE exp (%)	Pred(IE exp (%))	Residual	Std. residual	Std. dev. on pred. (Mean)
ECMF	1	96.540	95.510	1.030	0.198	5.093
FCH	1	89.930	85.799	4.131	0.795	4.298
FMT	1	89.440	87.374	2.066	0.397	3.040
FN	1	89.030	90.669	-1.639	-0.315	4.867
BFA	1	88.600	87.386	1.214	0.233	4.945
FAA	1	78.240	77.614	0.626	0.120	4.019
EF	1	77.340	76.071	1.269	0.244	4.735
MF	1	76.750	81.167	-4.417	-0.849	3.700
MFF	1	76.140	77.972	-1.832	-0.352	4.928
DMFA	1	71.990	76.893	-4.903	-0.943	4.133
FC	1	64.250	63.549	0.701	0.135	4.950
FFA	1	53.930	51.711	2.219	0.427	4.881
NVF	1	35.960	36.427	-0.467	-0.090	5.154

According to the statistical validation results, PCR has better quality in determining the corrosion inhibitor value of furan derivative compounds compared to PLS. It is indicated by good validation results, such as the coefficient of determination (R^2 = 0.976) and the adjusted coefficient of determination (R^2_{adj} = 0.904). From Table 7, it can also be seen that the residual value is small and stable; nothing is more than 5 or -5. Figure 5 also shows that all calculation results are close to fitting data.

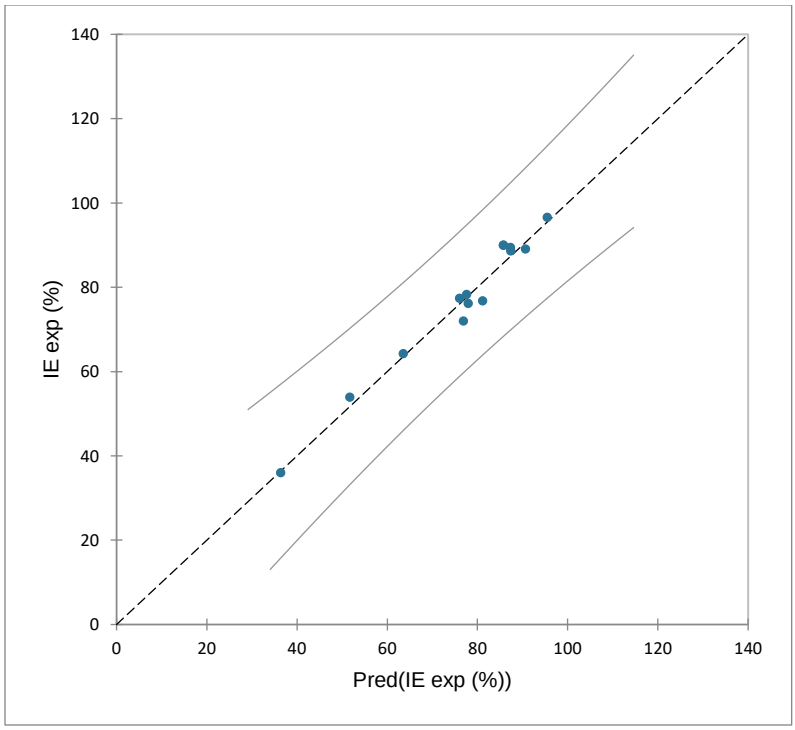


Figure 5. Correlation between IE_{exp} and IE_{Pred} with PCR modeling results

3.2.3 *Multiple Linear Regression* (MLR)

In this research, the MLR model used is the backward model. In MLR, some less influential descriptors are omitted, and only the most influential descriptors are taken. The results of MLR modeling are expressed in the following equation:

$$IE\ exp\ (\%) = 1235.047 + 286.773E_{LUMO} - 27.310\mu - 192.138\Delta N + 3.642\omega + 12.627 Log\ P + 0.163Mr \quad \text{.....(4)}$$

$N = 13 \quad R^2 = 0.973 \quad R^2_{adj} = 0.946 \quad PRESS = 446.163$

Table 8. MLR validation table

Observations	13
Sum of weights	13
DF	6
R^2	0.973
Adjusted R^2	0.946
MSE	15.294
RMSE	3.911
MAPE	3.008
DW	2.220

Cp	4.394
AIC	39.405
SBC	43.360
PC	0.091
Press	446.163
Q^2	0.868

Table 9. Predictive values and residuals of MLR modeling

Observation	Weight	IE exp (%)	Pred(IE exp (%))	Residual	Std. residual	Std. dev. on pred. (Mean)
ECMF	1	96.540	96.570	-0.030	-0.008	2.990
FCH	1	89.930	86.414	3.516	0.899	2.934
FMT	1	89.440	87.482	1.958	0.501	2.098
FN	1	89.030	90.822	-1.792	-0.458	3.348
BFA	1	88.600	85.967	2.633	0.673	2.253
FAA	1	78.240	78.073	0.167	0.043	2.670
EF	1	77.340	75.763	1.577	0.403	3.294
MF	1	76.750	82.280	-5.530	-1.414	2.273
MFF	1	76.140	76.891	-0.751	-0.192	3.117
DMFA	1	71.990	76.499	-4.509	-1.153	3.028
FC	1	64.250	61.808	2.442	0.625	2.316
FFA	1	53.930	52.126	1.804	0.461	3.077
NVF	1	35.960	37.445	-1.485	-0.380	3.457

The results of quantitative structure and property modeling with Multiple linear regression (MLR) analysis show high values on R^2 and R^2_{adj} data. The number is 0.973 and 0.946, respectively. From Table 9, the residual error results are also small and stable. In addition, in Figure 6, it can be seen that all the data are close to the fitting data, which indicates that the results are good.

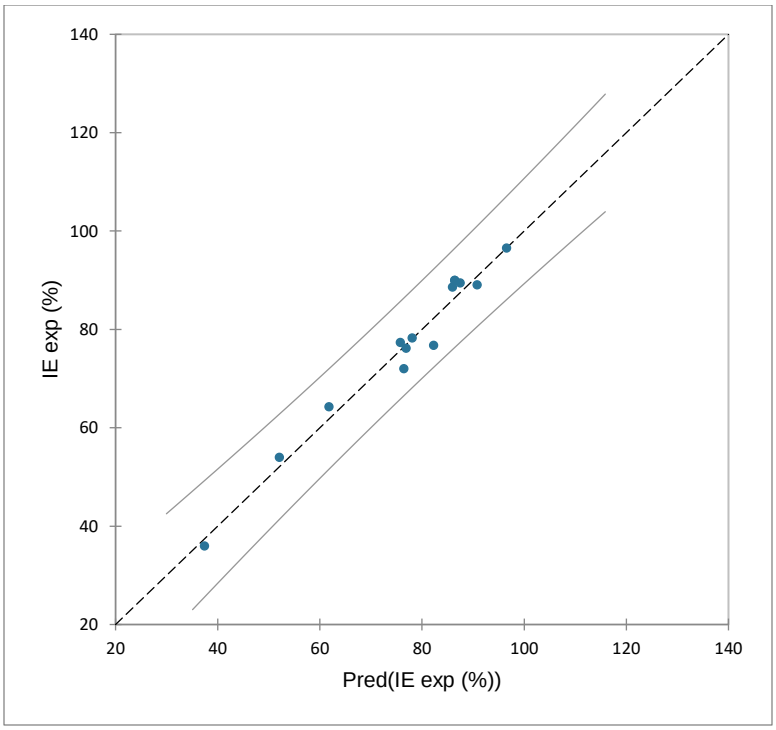


Figure 6. Correlation between IE_{exp} and IE_{Pred} with MLR modeling results

3.2.4 *Principal Component Analysis* (PCA)

A statistical method for qualitative analysis is principal component analysis (PCA) [42]. By condensing a large set of correlated variables into a smaller, uncorrelated set of variables, this descriptive method can be used to reduce the dimensionality of huge data sets. The name of these new variables is main components. It enables the practitioner to cut down on the amount of variables and streamline the information [43]. PCA is performed to reduce the dimensions of a large data set by changing the large set of variables to be smaller and uncorrelated. The new variable is called the principal component or principal axis. It allows the researcher to reduce the number of variables and make the information less overwhelming.

In this study, PCA analysis was carried out to determine the relationships between descriptors. Thus, in the end, it was possible to determine which data analysis technique was most suitable for quantitative structure and property relation modeling. The results of PCA analysis of 13 furan derivatives and their descriptors can be seen in figure 7.

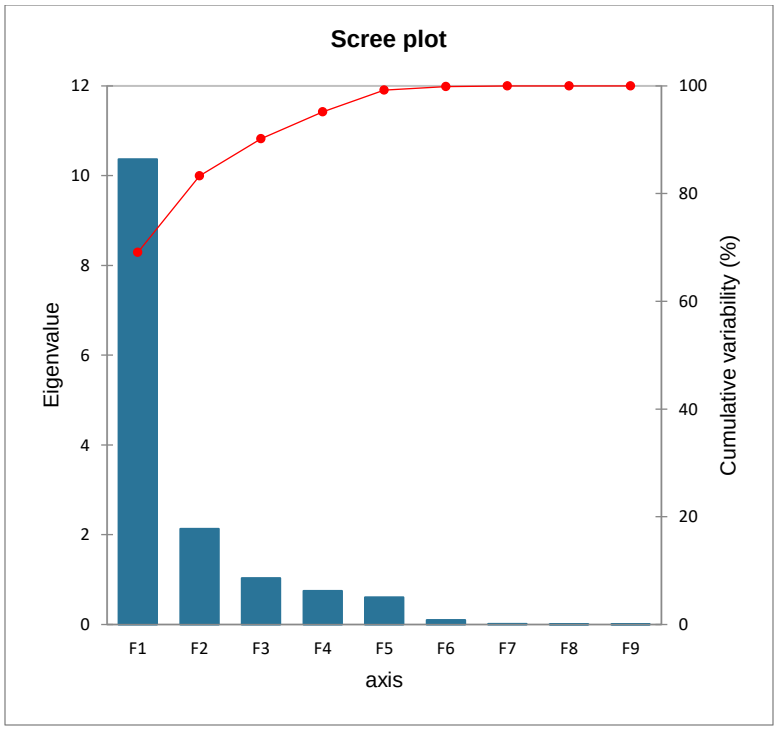


Figure 7. Principal components dan variannya

Table 10. Contribution of descriptors to the principal components F1, F2, and F3

Descriptor	F1		F2		F3	
	Correlation	Contribution (%)	Correlation	Contribution (%)	Correlation	Contribution (%)
E_{HOMO}	0.787	5.970	0.466	10.220	0.329	10.448
E_{LUMO}	0.993	9.513	0.050	0.118	0.023	0.050
E_{L-H}	0.945	8.618	-0.207	2.012	-0.164	2.593
μ	-0.894	7.718	-0.012	0.006	-0.081	0.641
IP	-0.787	5.970	-0.466	10.220	-0.329	10.448
EA	-0.993	9.513	-0.050	0.118	-0.023	0.050
χ	-0.970	9.081	-0.180	1.525	-0.118	1.340
σ	-0.911	8.010	0.219	2.254	0.279	7.498
η	0.945	8.618	-0.207	2.012	-0.164	2.593
ΔN	0.987	9.405	0.042	0.082	0.061	0.362
ω	-0.840	6.802	-0.105	0.520	-0.034	0.112
Log P	-0.310	0.927	-0.207	2.009	0.710	48.650
Mr	-0.216	0.452	0.796	29.801	-0.352	11.968
V_m	-0.285	0.786	0.888	37.092	-0.082	0.652
ΔE_{B-D}	-0.945	8.618	0.207	2.012	0.164	2.593

The contribution of descriptors to the principal components F1, F2, and F3 are summarized in table 10. According to the table, it can be seen that E_{LUMO}, E_{L-H}, μ, EA, η, σ, χ, ΔN, ω, Δand E_{B-D} have significant contributions to F1. For the descriptors, MR and V_m contribute significantly to F2. Whereas E_{HOMO}, IP, and Log P contributed strongly in F3.

The projection of the first three main component variables, namely F1, F2, and F3, based on their percentage contribution in the two correlation graphs is illustrated in Figure 8. The axes take into account as much variability in the data as possible. They represent 69.11%, 14.19%, and 6.91%, respectively, of the total variance, and the total percentage of information is estimated at 90.21%. This value is representative enough to describe the information in the data set.

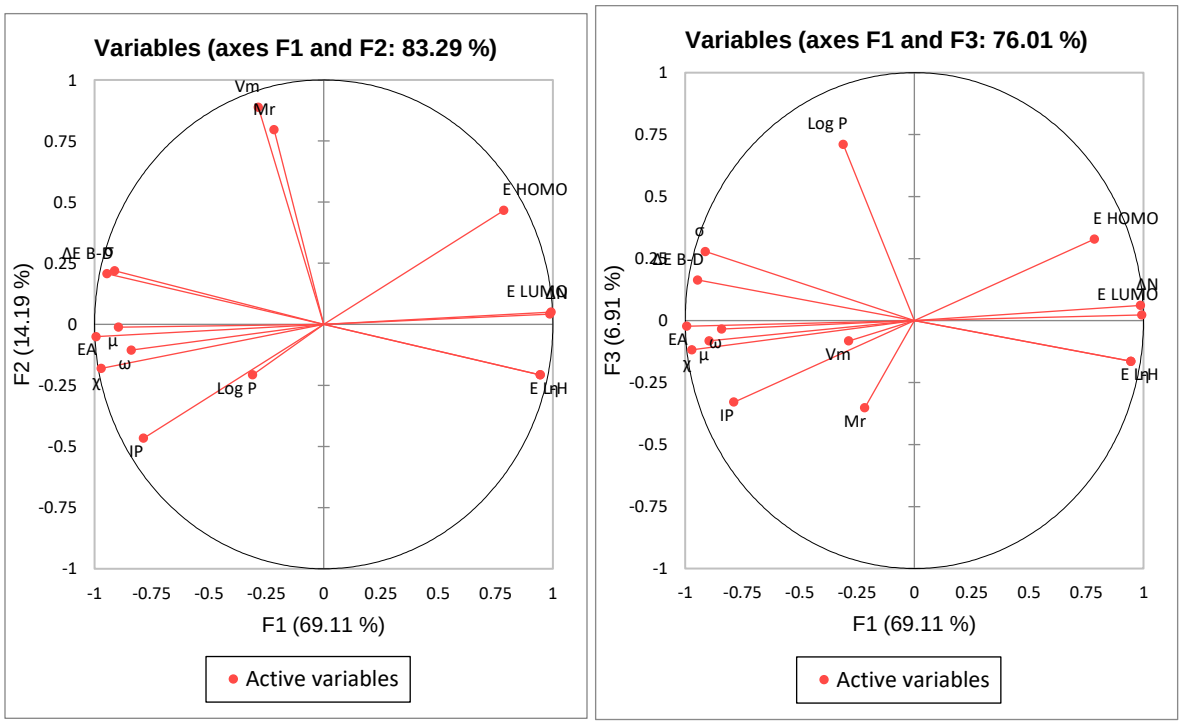


Figure 8. Correlation circles between principal compounds F1-F2 dan F1-F3

The correlation coefficient in the matrix provides information about the high or low relationship between the descriptors. Generally, highly correlated descriptors ($R \geq 0.75$) are not included to reduce the redundancy that exists in the matrix data [44-46]. Table 11 indicates that there is perfect negative collinearity (R = -1) between E_{HOMO} and IP; E_{LUMO} and EA; E_{L-H} and ΔE_{B-D}; as well as η and ΔE_{B-D}. In addition, there are other strong negative collinearities such as E_{HOMO} and EA (R= -0.815), E_{HOMO} and χ (R= -0.905), and so on. Perfect positive collinearity (R=1) can be found in E_{L-H} and η. Other positive collinearities were also identified in E_{HOMO} and E_{LUMO} (R = 0.815), E_{HOMO} and ΔN (0.819), etc. This indicates that these variables are redundant.

According to the results of the PCA descriptive data, there appears to be strong collinearity between descriptors; in other words, some explanatory variables are linear combinations of others. In this case, the matrix (X'X)-1 cannot be inverted because there are many variables with perfect collinearity. Therefore, some statistical analysis cannot be used in this data. Statistical analysis such as MLR cannot be used to predict the value of corrosion

inhibitors because it will lose most of the useful information in making the desired model. Apart from MLR, Multiple Polynomial Regression (MPR) also cannot be used.

The proof regarding the collinearity problem is also emphasized in Table 11. If the tolerance value is less than 0.2 and or the VIF value is more than 10, then it is certain that there is a collinearity problem. Table 11 shows that only a few descriptors meet these requirements, and most do not. It indicates that there is a collinearity problem in the data.

Table 11. Tolerance and VIF descriptors

	Tolerance	VIF
E_{HOMO}	0.000	
E_{LUMO}	0.000	
E_{L-H}	0.000	
μ	0.005	186.335
IP	0.000	
EA	0.000	
χ	0.000	
σ	0.002	490.173
η	0.000	
ΔN	0.000	2610.161
ω	0.007	150.711
Log P	0.466	2.144
Mr	0.149	6.720
V_m	0.165	6.072
ΔE_{B-D}	0.000	

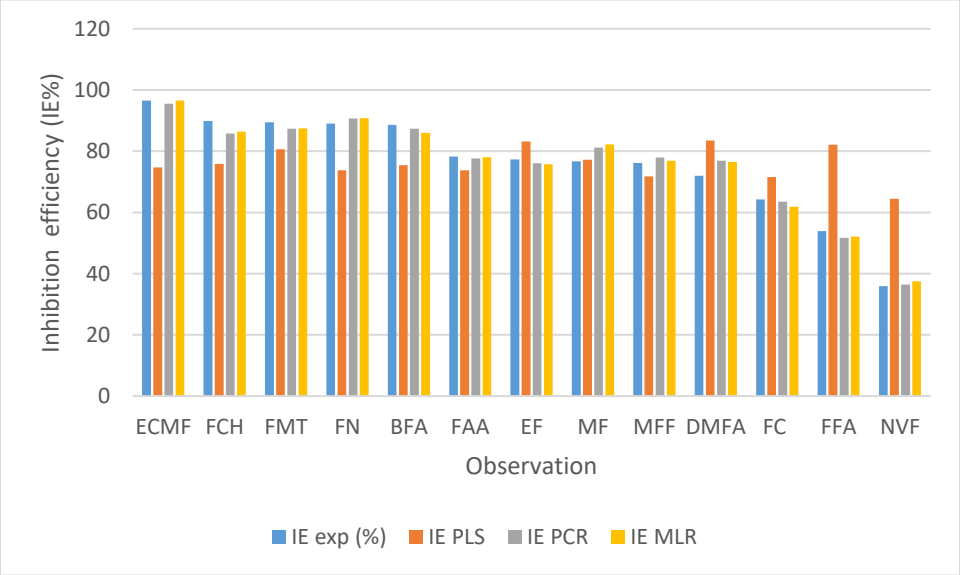


Figure 9. Comparison diagram of IE% values obtained with PLS, PCR and MLR

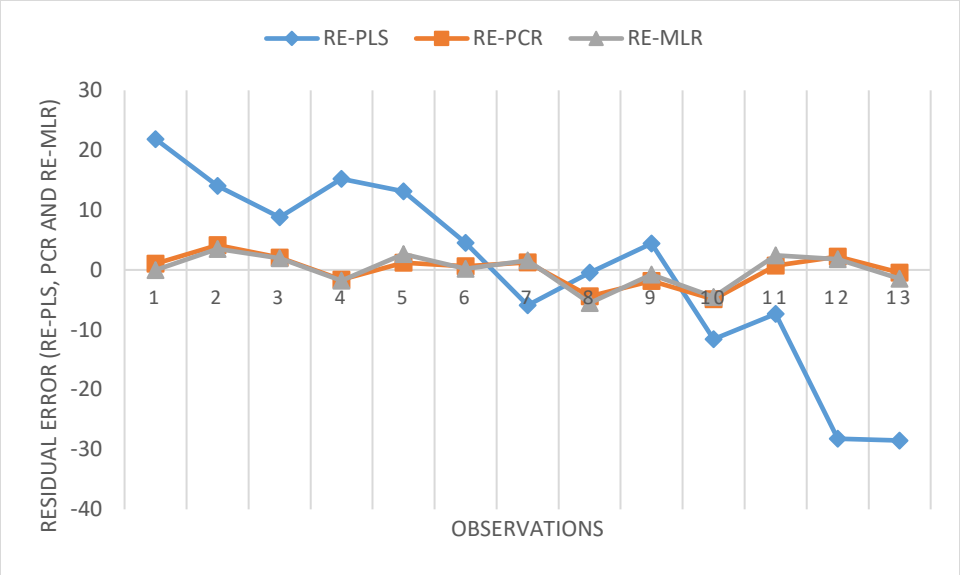


Figure 10. The residual error (RE) between the the experimental and the predicted IE% calculated with PLS, PCR and MLR

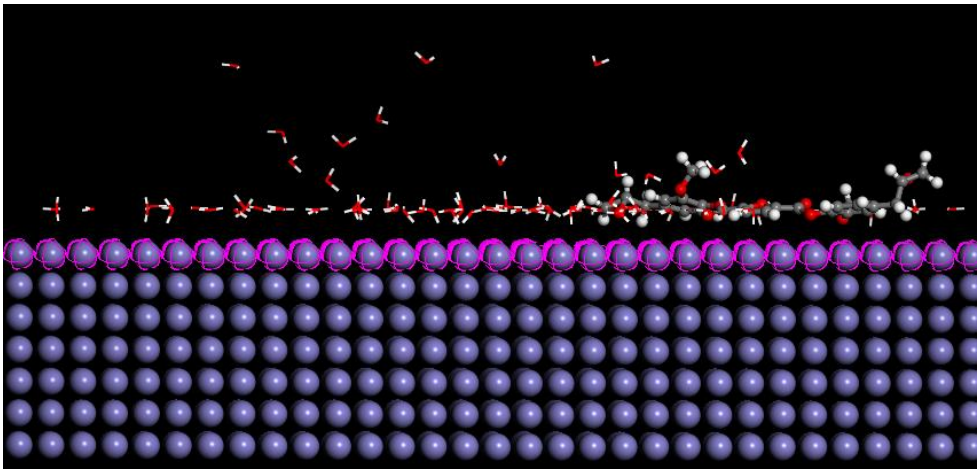
3.3 Monte Carlo Simulation

The previous statistical analysis resulted in a temporary conclusion that the quantitative structure and property relation modeling results with PCR analysis were the best. From these conclusions, the values of prediction inhibitor efficiency ($IE\%_{pred}$) were sought for four compounds with unknown corrosion inhibitor values. $IE\%_{pred}$ value data for the four compounds are presented in table 13

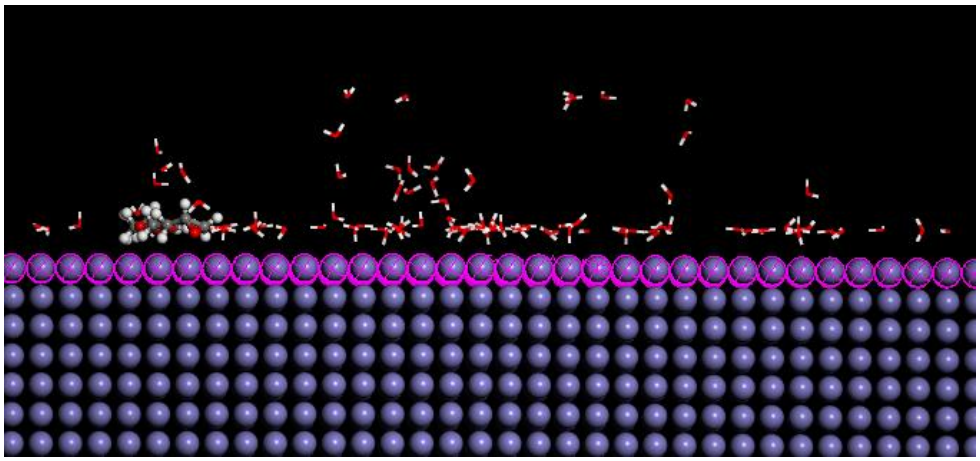
From the table, it can be seen that there are several promising IEpred values, such as the BMOPF compound (IEpred(%) = 169.37) and FMP (IEpred(%) = 100.81). Furthermore, Monte Carlo simulations of inhibitor compounds for iron (Fe) in a water medium were carried out to find the adsorption energy of each compound. Monte Carlo simulation results are shown in table 12 and figure 9 (viewed from the side). Monte Carlo simulation results (viewed above) are illustrated in figure 10

Table 12. Adsorption energy prediction compounds

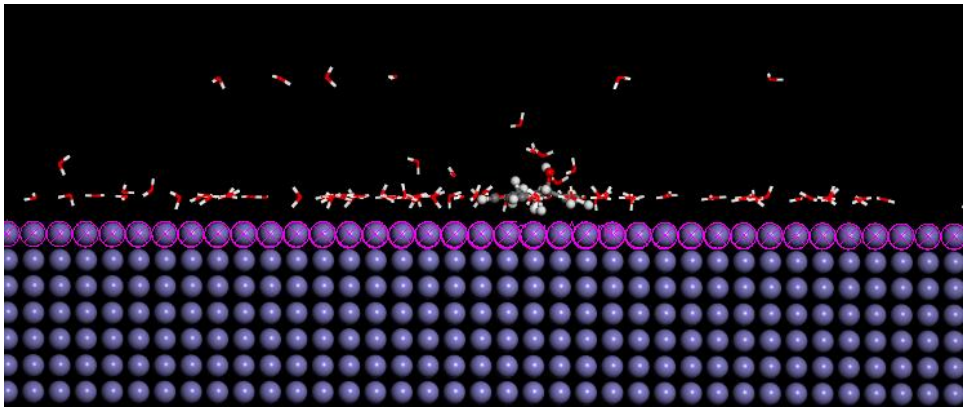
No.	Senyawa	Adsorption energy inhibitor (kJ/mol)	Adsorption energy water (kJ/mol)	IE pred (%)
1.	BMOPF	-272.96906317	-5.03186855	169.3773
2.	BOMF	-160.04338881	-6.78146263	78.68098
3.	FMP	-109.55142600	-6.67838206	100.8158
4.	NAA	-122.62129424	-2.11851366	66.17984



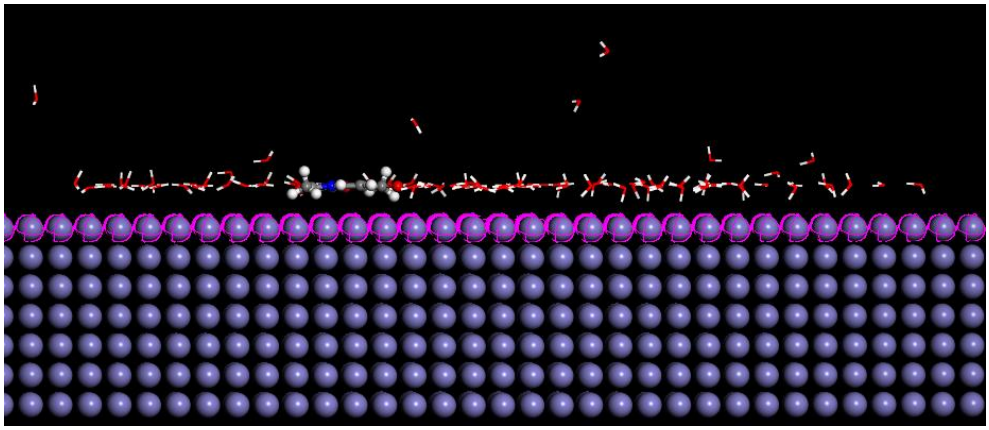
(a)



(b)



(c)



(d)

Figure 11. Visualization from Monte Carlo simulation of 4 prediction compounds, (a) BMOPF; (b) BOMF; (c) FMP; (d) NAA

Conclusion

From the analysis and discussion above, the following conclusions are obtained:

- 17 compounds of furan derivatives divided into 13 main compounds and 4 predictive compounds have been investigated theoretically and statistically as corrosion inhibitors for iron (Fe) metal.
- Calculating molecular descriptors using the DFT quantum method makes it possible to correlate the corrosion inhibition efficiency (IE%) with molecular structure using the QSPR approach.
- Qualitative analysis by PCA allows researchers to examine redundancy and collinearity between proposed descriptors.
- The established regression analysis results show that the studied molecules' anticorrosive activity can be explained based on their electronic and structural properties.
- Examination of the quantitative analysis results showed that PCR analysis was the best statistical method compared to PLS and MLR. It is proven through the data's validation results ($R^2 = 0.976$; $R^2_{adj} = 0.904$) and collinearity analysis.
- The results of QSPR modeling analysis with PCR are proposed to predict the corrosion inhibitor value of new furan derivative compounds.
- The prediction results for this study's 4 new derivative compounds are very promising, especially for the BMOPF compound (IE%pred = 169.37) and the FMP compound (IE%pred = 100.81).

References.

1. Zakeri, A., Bahmani, E., & Aghdam, A. S. R. (2022). Plant extracts as sustainable and green corrosion inhibitors for protection of ferrous metals in corrosive media: A mini review. *Corrosion Communications*.
2. Chugh, B., Taraphdar, P. K., Biswal, H. J., Devi, N. R., Dorothy, R., Manimaran, N., & Rajendran, S. (2022). Corrosion inhibition by aluminum oxide. In *Inorganic Anticorrosive Materials* (pp. 231-249). Elsevier.

3. Nazari, M. H., Zhang, Y., Mahmoodi, A., Xu, G., Yu, J., Wu, J., & Shi, X. (2022). Nanocomposite organic coatings for corrosion protection of metals: A review of recent advances. *Progress in Organic Coatings*, *162*, 106573.
4. Assad, H., & Kumar, A. (2021). Understanding functional group effect on corrosion inhibition efficiency of selected organic compounds. *Journal of Molecular Liquids*, *344*, 117755.
5. Jafar Mazumder, M. A. (2020). A review of green scale inhibitors: Process, types, mechanism and properties. *Coatings*, *10*(10), 928.
6. Alao, A. O., Popoola, A. P., & Sanni, O. (2022). The influence of nanoparticle inhibitors on the corrosion protection of some industrial metals: a review. *Journal of Bio-and Tribo-Corrosion*, *8*(3), 68.
7. Tamalmani, K., & Husin, H. (2020). Review on corrosion inhibitors for oil and gas corrosion issues. *Applied Sciences*, *10*(10), 3389.
8. Wei, H., Heidarshenas, B., Zhou, L., Hussain, G., Li, Q., & Ostrikov, K. K. (2020). Green inhibitors for steel corrosion in acidic environment: state of art. *Materials Today Sustainability*, *10*, 100044.
9. Avdeev, Y. G., & Kuznetsov, Y. I. (2021). Nitrogen-containing five-membered heterocyclic compounds as corrosion inhibitors for metals in solutions of mineral acids-An overview. International Journal of Corrosion and Scale Inhibition, 10(2), 480-540.
10. Issaadi, S., Douadi, T., & Chafaa, S. (2014). Adsorption and inhibitive properties of a new heterocyclic furan Schiff base on corrosion of copper in HCl 1 M: experimental and theoretical investigation. *Applied surface science*, *316*, 582-589.
11. Kognou, A. L. M., Shrestha, S., Jiang, Z. H., Xu, C. C., Sun, F., & Qin, W. (2022). High-fructose corn syrup production and its new applications for 5-hydroxymethylfurfural and value-added furan derivatives: Promises and challenges. *Journal of Bioresources and Bioproducts*.
12. Ser, C. T., Žuvela, P., & Wong, M. W. (2020). Prediction of corrosion inhibition efficiency of pyridines and quinolines on an iron surface using machine learning-powered quantitative structure-property relationships. *Applied Surface Science*, *512*, 145612.
13. Driouch, M., Lazrak, J., Bensouda, Z., Elhaloui, A., Sfaira, M., Saffaj, T., & Taleb, M. (2020). Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium. *Heliyon*, *6*(10), e05067.
14. Quadri, T. W., Olasunkanmi, L. O., Fayemi, O. E., Lgaz, H., Dagdag, O., Sherif, E. S. M., ... & Ebenso, E. E. (2022). Predicting protection capacities of pyrimidine-based corrosion inhibitors for mild steel/HCl interface using linear and nonlinear QSPR models. *Journal of Molecular Modeling*, *28*(9), 254.
15. Awfa, D., Ateia, M., Mendoza, D., & Yoshimura, C. (2021). Application of quantitative structure–property relationship predictive models to water treatment: a critical review. *ACS ES&T Water*, *1*(3), 498-517.
16. Quadri, T. W., Olasunkanmi, L. O., Fayemi, O. E., Lgaz, H., Dagdag, O., Sherif, E. S. M., ... & Ebenso, E. E. (2022). Computational insights into quinoxaline-based corrosion inhibitors of steel in HCl: Quantum chemical analysis and QSPR-ANN studies. *Arabian Journal of Chemistry*, *15*(7), 103870.
17. Quadri, T. W., Olasunkanmi, L. O., Akpan, E. D., Fayemi, O. E., Lee, H. S., Lgaz, H., ... & Ebenso, E. E. (2022). Development of QSAR-based (MLR/ANN) predictive models

for effective design of pyridazine corrosion inhibitors. *Materials Today Communications*, *30*, 103163.

18. Camacho-Mendoza, R. L., Feria, L., Zárate-Hernández, L. Á., Alvarado-Rodríguez, J. G., & Cruz-Borbolla, J. (2022). New QSPR model for prediction of corrosion inhibition using conceptual density functional theory. *Journal of Molecular Modeling*, *28*(8), 238.
19. Al-Fakih, A. M., Aziz, M., Abdallah, H. H., Algamal, Z. Y., Lee, M. H., & Maarof, H. (2015). High dimensional QSAR study of mild steel corrosion inhibition in acidic medium by furan derivatives. *International Journal of Electrochemical Science*, *10*(4), 3568-3583.
20. Al-Fakih, A. M., Abdallah, H. H., & Aziz, M. (2019). Experimental and theoretical studies of the inhibition performance of two furan derivatives on mild steel corrosion in acidic medium. *Materials and Corrosion*, *70*(1), 135-148.
21. Miao, J. T., Yuan, L., Guan, Q., Liang, G., & Gu, A. (2017). Biobased heat resistant epoxy resin with extremely high biomass content from 2, 5-furandicarboxylic acid and eugenol. *ACS Sustainable Chemistry & Engineering*, *5*(8), 7003-7011.
22. Meng, J., Zeng, Y., Chen, P., Zhang, J., Yao, C., Fang, Z., ... & Guo, K. (2020). Flame retardancy and mechanical properties of bio-based furan epoxy resins with high crosslink density. *Macromolecular Materials and Engineering*, *305*(1), 1900587.
23. Nowicki, J., Kula, J., & Hammad, D. (2002). Synthesis of new furan-type terpenoids. *Flavour and fragrance journal*, *17*(3), 203-206.
24. Pham, T. T., Chen, X., Yan, N., & Sperry, J. (2018). A novel dihydrodifuropyridine scaffold derived from ketones and the chitin-derived heterocycle 3-acetamido-5-acetylfuran. *Monatshefte für Chemie-Chemical Monthly*, *149*, 857-861.
25. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... & Fox, D. J. (2009). Gaussian 09 program. *Gaussian Inc., Wallingford, CT*.
26. Meunier, M., & Robertson, S. (2021). Materials Studio 20th Anniversary. *Molecular Simulation*, *47*(7), 537-539.
27. Alamri, A. H., & Alhazmi, N. (2022). Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors. *Journal of Saudi Chemical Society*, *26*(6), 101536.
28. Quadri, T. W., Olasunkanmi, L. O., Fayemi, O. E., Akpan, E. D., Lee, H. S., Lgaz, H., ... & Ebenso, E. E. (2022). Multilayer perceptron neural network-based QSAR models for the assessment and prediction of corrosion inhibition performances of ionic liquids. *Computational Materials Science*, *214*, 111753.
29. Belghiti, M. E., Benhiba, F., Benzbiria, N., Lai, C. H., Echihi, S., Salah, M., ... & Naimi, Y. (2022). Performance of triazole derivatives as potential corrosion in-hibitors for mild steel in a strong phosphoric acid medium: Combining experimental and computational (DFT, MDs & QSAR) approaches. Journal of Molecular Structure, 1256, 132515.
30. Xiong, S., Wu, H., Liu, Z., & Zhang, B. (2022). QSAR Models for the Prediction of the Relationship Among Corrosion Inhibition Efficiency, Friction Coefficient and Oil Film Strength of Lubricants. *Polycyclic Aromatic Compounds*, *42*(6), 3780-3791.
31. Lahyaoui, M., Diane, A., El-Idrissi, H., Saffaj, T., Rodi, Y. K., & Ihssane, B. (2023). QSAR modeling and molecular docking studies of 2-oxo-1, 2-dihydroquinoline-4-carboxylic acid derivatives as p-glycoprotein inhibitors for combating cancer multidrug resistance. *Heliyon*, e13020.

32. Golbraikh, A., & Tropsha, A. (2002). Beware of q2!. *Journal of molecular graphics and modelling*, *20*(4), 269-276.
33. Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular informatics*, *29*(6-7), 476-488.
34. Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR & combinatorial science*, *26*(5), 694-701.
35. Becke, A. D. (1992). Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *The Journal of chemical physics*, *96*(3), 2155-2160.
36. Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, *38*(6), 3098.
37. Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, *37*(2), 785.
38. Consonni, V., & Todeschini, R. (2009). Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing/Volume II: Appendices, References. John Wiley & Sons.
39. Dunn, W. J., & Rogers, D. (1996). Genetic partial least squares in QSAR. In *Genetic algorithms in molecular modeling* (pp. 109-130). Academic Press.
40. Cramer III, R. D. (1993). Partial least squares (PLS): its strengths and limitations. *Perspectives in Drug Discovery and Design*, *1*(2), 269-278.
41. Lindgren, F., & Rännar, S. (1998). Alternative partial least-squares (PLS) algorithms. *3D QSAR in Drug Design: Recent Advances*, 105-113.
42. Ebenso, E. E., Verma, C., Olasunkanmi, L. O., Akpan, E. D., Verma, D. K., Lgaz, H., ... & Quraishi, M. A. (2021). Molecular modelling of compounds used for corrosion inhibition studies: a review. *Physical Chemistry Chemical Physics*, *23*(36), 19987-20027.
43. Hmamouchi, R., Larif, M., Adad, A., Bouachrine, M. O. H. A. M. M. E. D., & Lakhlifi, T. (2014). Structure activity and prediction of biological activities of compound (2-methyl-6-phenylethynylpyridine) derivatives relationships rely on electronic and topological descriptors. *J. Comput. Methods Mol*, *4*, 61-71.
44. Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long range planning*, *46*(1-2), 1-12.
45. Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In *New challenges to international marketing*. Emerald Group Publishing Limited.
46. Moore, D. S., & Kirkland, S. (2007). *The basic practice of statistics* (Vol. 2). New York: WH Freeman.