International Journal of Electrochemical Science 18 (2023) 100207

Contents lists available at ScienceDirect

International Journal of Electrochemical Science

journal homepage: www.journals.elsevier.com/ijoes

Development of QSPR models for furan derivatives as corrosion inhibitors for mild steel

Saprizal Hadisaputra*, Aditya Dawanta Irham, Agus Abhi Purwoko, Eka Junaidi, Aliefman Hakim

Chemistry Education Division, University of Mataram, Jalan Majapahit No 62, Mataram 83125, Indonesia

ARTICLE INFO

Keywords:
QSPR
Inhibition efficiency
Furan
Corrosion inhibition
Mild steel

ABSTRACT

The quantitative structure-property relationship QSPR method has been used to analyze the corrosion inhibition properties of furan derivative inhibitors against mild steel. This modeling is based on the correlation between corrosion inhibition efficiency (IE%) and several electronic 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), $\Delta E_{B\text{-}D}$ (back-donation energy), Log P, Vm (critical volume), and Mr (relative molecular mass). These properties were calculated using DFT at B3LYP/6–31 G(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 QSPR modeling results are by PCR statistical analysis. It is proven by the validation results (R^2 = 0.976; R^2_{adj} = 0.90) and analysis of collinearity in the data. The predictions of the four furan-derived compounds from PCR modeling gave promising results.

1. Introduction

To achieve a more stable thermodynamic state, metals and their alloys undergo a natural process known as corrosion [1,2]. Corrosion of metal compounds and their alloys is one of the key problems the industry sector now dealing with. It is considered a problem because it can be a source of environmental pollution. Researchers are therefore urged to develop methods to protect these minerals and lessen their environmental impact. Corrosion inhibitors are, therefore, one of the most often employed strategies for stopping corrosion and reining in 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 inhibitors to solve the acidic media issue. These compounds' structural and electrical characteristics significantly influence 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 reactivity [5–9].

Furan-derived compounds have excellent promise as corrosion inhibitors in acidic conditions [10,11]. It has been demonstrated that these derivatives of organic compounds can remove unbound electrons from a metal surface by using of vacant orbitals at lower energy levels. In order to form covalent bonds that coordinate, it gives electrons to the metal surface. The inhibitor can adhere to the metal surface more readily as a result. By reducing the metal's dissolution rate, the release of protons, and raising the coverage ratio, its adsorption makes it possible to block the active sites. The higher inhibition of these compounds' effects on corrosion is evidence of it.

It has been found that this approach to quantum chemistry is highly helpful in understanding the molecular structure, electronic structure, and features of the reactive sites. This method is also helpful for comprehending the relationship between the effectiveness of corrosion inhibition and some corrosion inhibitor molecular indices or their quantum characteristics. To establish a consistent link between changes in the index of molecular characteristics and the inhibitory action of various drugs, QSPR has lately become a popular tool for the quantitative investigation of corrosion inhibition processes [12–18].

The application of mathematics can result in qualitative and quantitative data that could help us comprehend the process of corrosion inhibition better. One of the objectives of this study is to identify a stable structural property that links the molecular descriptors of

* Corresponding author.
E-mail address: rizal@unram.ac.id (S. Hadisaputra).

https://doi.org/10.1016/j.ijoes.2023.100207
Received 20 February 2023; Received in revised form 20 May 2023; Accepted 21 May 2023
Available online 24 May 2023

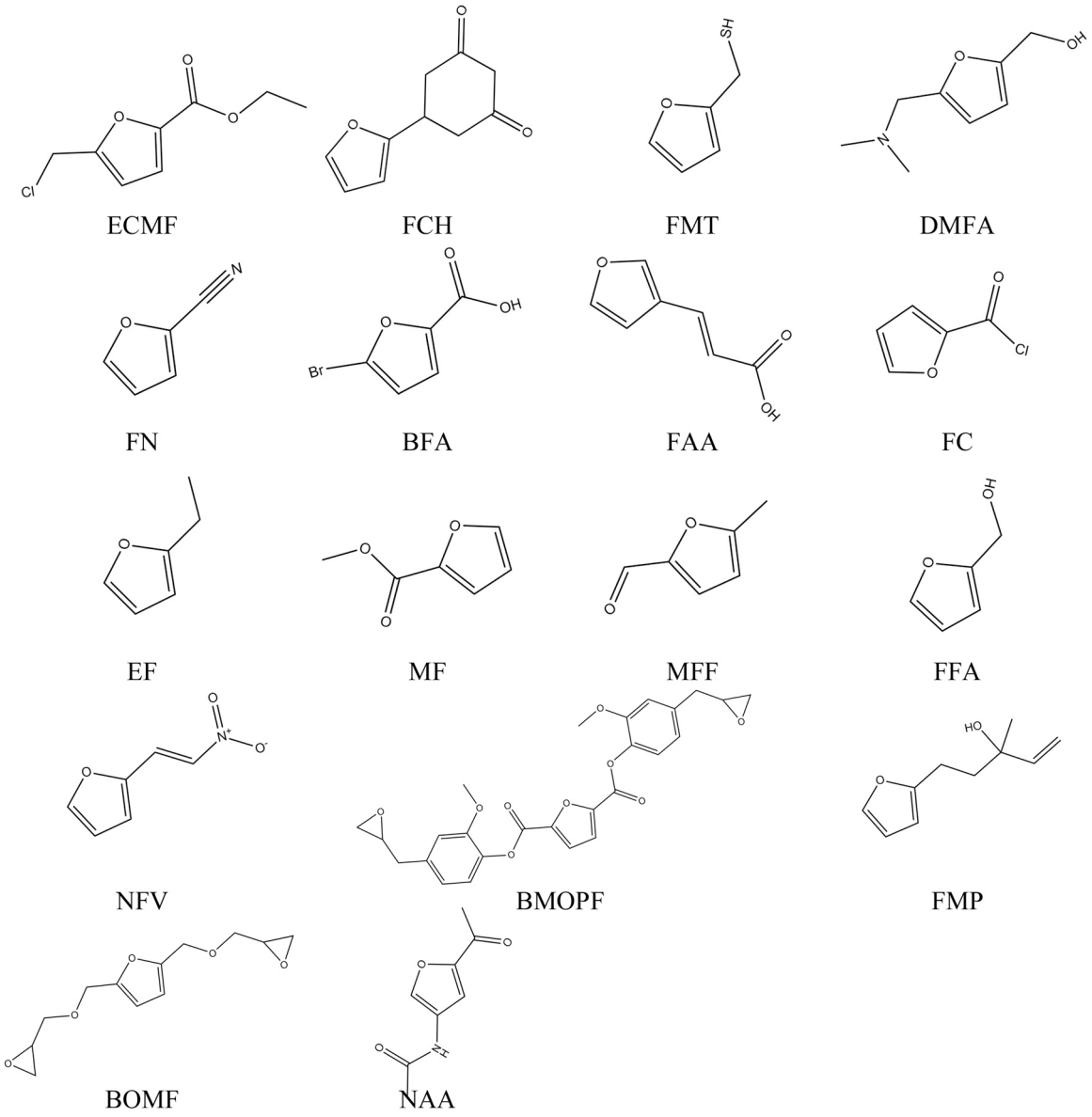

Fig. 1. Structure of 13 furan derivatives for QSPR model development and predicted corrosion inhibitor values.

molecules with their ability to inhibit corrosion (IE%). Experimental research has been done on them in 250 mL of 1 M HCl with an inhibitor of 0.005 M as a corrosion inhibitor on mild steel [19,20].

The current study focused on a statistical analysis of the data evaluated by comparing four mathematical regression models, namely: PLS, PCR, MLR, and PCA, to understand the depth of the corrosion inhibition mechanism. Additionally, this study aimed to identify the relevant and mutually influential chemical descriptors in the variation of corrosion inhibition of the compounds investigated. Finally, due to the accurate prediction outcomes, the outcomes of the constructed mathematical equations will enable the estimation of the value of corrosion inhibitors of related compounds and facilitate other researchers' synthesis of related compounds.

2. 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 (Fig. 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.

2.2. Computational calculation

All molecular geometries were optimized by Gaussian 09 Software [25]. Quantum chemical calculations used in geometry optimization at DFT/B3LYP/6–31 G(d).

Koopman's theorem [26] states that the EHOMO and the ELUMO can be used to express the IP, EA, χ and η. Ionization potential is the quantity of energy required to expel one electron from a molecule [27]. Through eq. 1, it is associated with the energy of the EHOMO:

$$IP = -EHOMO \quad (1)$$

The term electron affinity (EA) refers to the energy released when a proton is introduced to a system [27]. It is connected to ELUMO via

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]

equation 2:

$$A = -ELUMO \quad (2)$$

The ability of an atom or group of atoms to draw electrons toward itself is measured as electronegativity (χ) [28]. It can be calculated using the following equation:

$$\chi = \frac{IP + EA}{2} \quad (3)$$

Chemical hardness (η) is an indicator of how well an atom resists a charge transfer [29]; it is calculated using the following equation:

$$\eta = \frac{IP - A}{2} \quad (4)$$

The fraction of transferred electrons (ΔN) from the furan derivatives to the ferric atom can be determined using Pearson theory [30] (Eq. 5):

$$\Delta N = \frac{\chi Fe - \chi Inh}{2(\eta Fe + \eta Inh)} \quad (5)$$

where Fe and χinh stand for, respectively, the absolute electronegativity of iron and the inhibitor molecule, and the ηFe and ηinh indicate iron and the inhibitor molecule's absolute hardness. The fraction of electrons transported was calculated using the theoretical value of bulk iron's electronegativity, Fe = 7.00 eV [31], and a global hardness of Fe = 0 by assuming that for a metallic bulk IP = EA [32].

The total electrophilic character of a molecule can be quantified on a relative scale using a reactivity descriptor termed electrophilicity. Parr et al. [27] suggested the electrophilicity index measure the energy loss brought on by the maximal electron flow between the donor and acceptor. It was described as follows (Eq. 6).

$$\omega = \frac{\mu^2}{2\eta} \quad (6)$$

It is also used in determining other physical-chemical descriptors that will also be used in QSPR. Monte Carlo simulations of furan

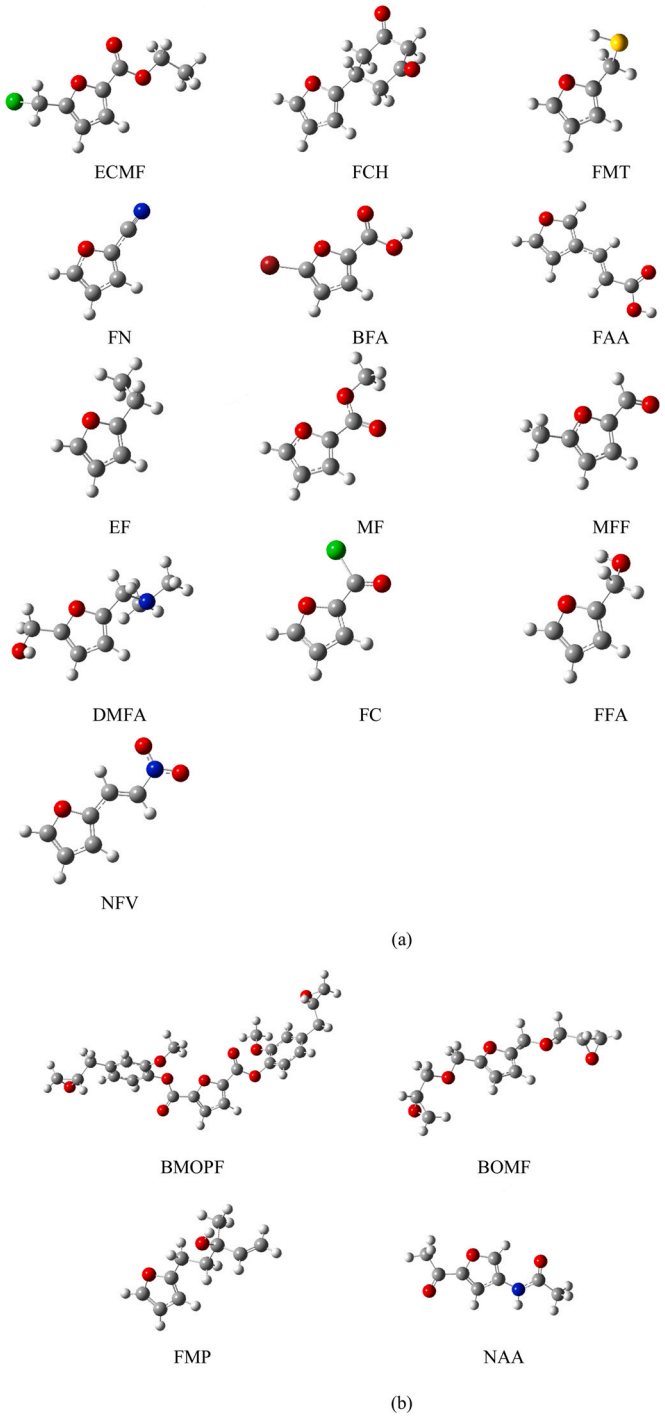

Fig. 2. (a) Furan derivative compounds (experimental) which have been optimized geometry with B3LYP/6–31 G (d); (b) Compounds derived from furan (prediction) which have been optimized for geometry with B3LYP/6–31 G (d).

derivatives were carried out using the Material Studio application [33]. This study uses iron or Fe crystals with a (110) 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.

Table 3
The studied furan descriptors.

Compounds	IE exp (%)	EHOMO	ELUMO	E L-H	μ	IP	EA	χ	σ	η	ΔN	ω	Log P	Mr	Vm	ΔE B-D
ECMF	96.54	-6.6790	-1.5096	5.1693	3.9891	6.6790	1.5096	4.0944	0.3869	2.5846	3.7550	20.5649	1.52	188.61	493.5	-0.6461
FCH	89.93	-6.2417	-1.1061	5.1356	3.9827	6.2417	1.1061	3.6739	0.3894	2.5678	4.2703	20.3651	-0.43	178.18	477.5	-0.6419
FMT	89.44	-5.9982	-0.0353	5.9628	1.3170	5.9982	0.0353	3.0168	0.3354	2.9814	5.9378	2.5856	0.9	114.17	304.5	-0.7453
FN	89.03	-6.8738	-1.2960	5.5777	4.7401	6.8738	1.2960	4.085	0.3585	2.7889	4.0648	31.3431	0.68	93.08	276.5	-0.6972
BFA	88.60	-6.6172	-1.4558	5.1614	2.9341	6.6172	1.4558	4.0365	0.3874	2.5807	3.8239	11.1087	0.92	190.98	336.5	-0.6451
FAA	78.24	-6.3261	-1.5624	4.7636	2.1237	6.3261	1.5624	3.9443	0.4198	2.3818	3.6390	5.3711	0.49	138.12	366.5	-0.5954
EF	77.34	-5.7816	0.6163	6.3979	0.4958	5.7816	-0.6163	2.5826	0.3126	3.1989	7.0655	0.3931	1.42	96.13	306.5	-0.7997
MF	76.75	-6.5473	-1.0756	5.4716	1.7764	6.5473	1.0756	3.8115	0.3655	2.7358	4.3616	4.3165	0.47	126.11	332.5	-0.6839
MFF	76.14	-6.4167	-1.5496	4.8670	4.2841	6.4167	1.5496	3.9832	0.4109	2.4335	3.6707	22.331	0.73	110.11	323.5	-0.6083
DMFA	71.99	-5.7163	0.3698	6.0861	1.8995	5.7163	-0.3698	2.6732	0.3286	3.0430	6.5832	5.4898	0.07	191.65	455.5	-0.7607
FC	64.25	-7.1685	-2.0710	5.0975	4.5827	7.1685	2.0710	4.6198	0.3923	2.5487	3.0332	26.7634	0.78	130.53	305.5	-0.6371
FFA	53.93	-6.1329	0.2122	6.3451	1.6807	6.1329	-0.2122	2.9603	0.3152	3.1725	6.4080	4.4808	0.08	98.10	269.5	-0.7931
NVF	35.96	-6.5631	-2.6555	3.9075	6.3610	6.5631	2.6555	4.6093	0.5118	1.9537	2.3354	39.5272	1.56	139.11	432.3	-0.4884

Table 4
PLS Validation.

Observations	13.00
Sum of weights	12.00
DF	10.00
R^2	0.10
Std. deviation	17.39
MSE	252.08
RMSE	15.87

2.3. Statistic analysis

A statistical study was carried out using the XLSTAT premium 2021 application [34] on 13 furan-derived compounds to find the QSPR between the corrosion inhibitor value and the molecular descriptors of these compounds. This study selected three statistical analyses to find the QSPR model: PLS, PCR, and MLR [13,35–38]. Furthermore, PCA was carried out to see the relationship between descriptors. It also checks redundancy and collinearity between the studied descriptors so that later the best statistical analysis could be determined in the QSPR study of these furan derivatives compounds.

2.4. Validation

The explanatory potency and prediction level of the QSPR modeling were assessed through validation testing. The only form of validation used in this study was internal. Internal validation is based on a number of statistical variables, including 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). The importance of this technique is then assessed by examining the coefficient of determination cross-validation (R^2_{cv}) [39,40]. R^2_{cv} is expressed in the form (Eq. 7),

$$R^2_{cv} = 1 - \frac{\sum_{i=1}^{n} (\text{IEexp} - IE \quad cal)}{\sum_{i=1}^{n} (\text{IEexp} - IE \quad avg)} \tag{7}$$

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.

3. Results and discussion

3.1. Geometry parameters

The geometry parameters of the studied molecules was optimized by DFT/B3LYP/6–31 G(d) [41,42]. The molecular structure's minimal energy qualifies the geometry as being obtained, and the lack of imaginary frequencies supports this claim. The final optimized geometry can be seen in Fig. 2.

3.2. Molecular descriptor calculations

The interaction behavior between metals and inhibitors cannot be fully explained by experimental research alone because this corrosion inhibition phenomenon depends on a number of other parameters. Therefore, combining experimental and quantum studies such as QSPR is urgently needed. The QSPR technique used in this study focuses on the relationship between each molecule's intrinsic characteristics and corrosion inhibition ability. Researchers created various models to connect corrosion inhibitors with some of their chemical characteristics to investigate this relationship. Therefore, molecular descriptors can be used to forecast the value of corrosion inhibitors and explain how they work. The most relevant descriptors capable of influencing the adsorption of inhibitor molecules onto metal surfaces are electronic, structural, and lipophilic indices. The descriptor values obtained are illustrated in Table 3.

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	96.54	74.68	21.85	1.42	5.16
FCH	1	89.93	75.90	14.02	0.91	5.02
FMT	1	89.44	80.63	8.81	0.57	6.59
FN	1	89.03	73.80	15.23	0.99	5.42
BFA	1	88.60	75.47	13.12	0.85	5.04
FAA	1	78.24	73.73	4.50	0.29	5.44
EF	1	77.34	83.22	-5.88	-0.38	8.36
MF	1	76.75	77.18	-0.43	-0.029	5.13
MFF	1	76.14	71.75	4.38	0.28	6.38
DMFA	1	71.99	83.54	-11.55	-0.75	8.60
FC	1	64.25	71.62	-7.37	-0.48	6.46
FFA	1	53.93	82.12	-28.19	-1.84	7.56
NVF	1	35.96	64.45	-28.49	-1.86	11.83

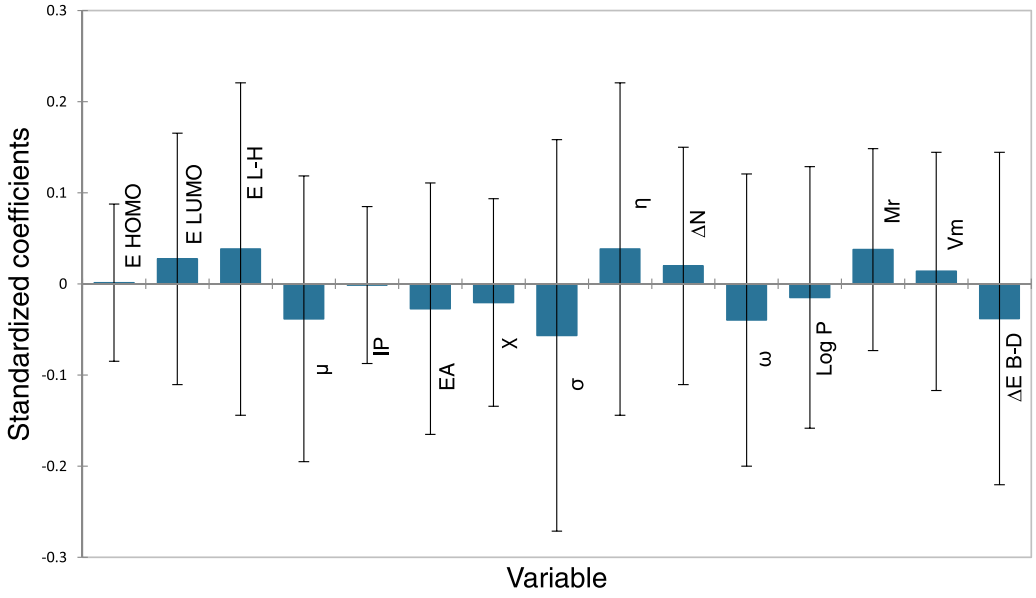

Fig. 3. The coefficients of the standard versus the variables in the specified PLS model.

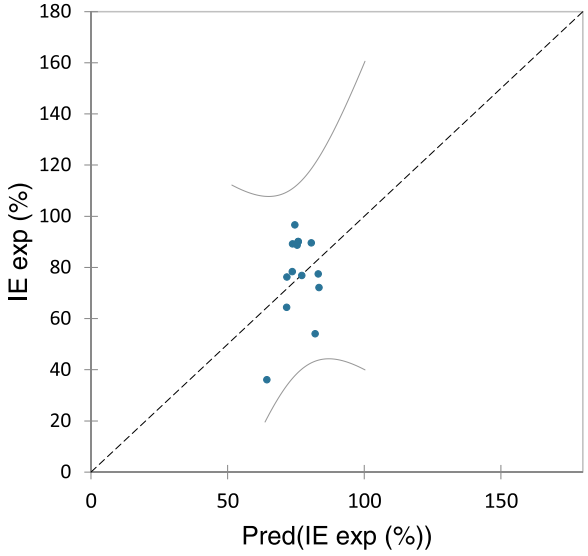

Fig. 4. Correlation between IE_{exp} and IE_{Pred} with PLS modeling results.

Table 6
PCR validation.

Observations	13.00
Sum of weights	13.00
DF	3.00
R^2	0.97
Adjusted R^2	0.90
MSE	27.03
RMSE	5.19
MAPE	2.70
DW	1.55
Cp	10.00
AIC	43.79
SBC	49.44
PC	0.18
Press	2940.58
Q^2	0.12

3.3. QSPR study

3.3.1. Partial Least Square (PLS)

PLS can simultaneously handle a large number of explanatory and response variables. This formula is frequently employed, particularly when a lot of molecular descriptors are being used to test a molecule [43].

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	96.54	95.51	1.03	0.19	5.09
FCH	1	89.93	85.79	4.13	0.79	4.29
FMT	1	89.44	87.37	2.06	0.39	3.04
FN	1	89.03	90.66	-1.63	-0.31	4.86
BFA	1	88.60	87.38	1.21	0.23	4.94
FAA	1	78.24	77.61	0.62	0.12	4.02
EF	1	77.34	76.07	1.27	0.24	4.73
MF	1	76.75	81.16	-4.42	-0.85	3.70
MFF	1	76.14	77.97	-1.83	-0.35	4.92
DMFA	1	71.99	76.89	-4.90	-0.94	4.13
FC	1	64.25	63.54	0.70	0.13	4.95
FFA	1	53.93	51.71	2.22	0.43	4.88
NVF	1	35.96	36.42	-0.46	-0.09	5.15

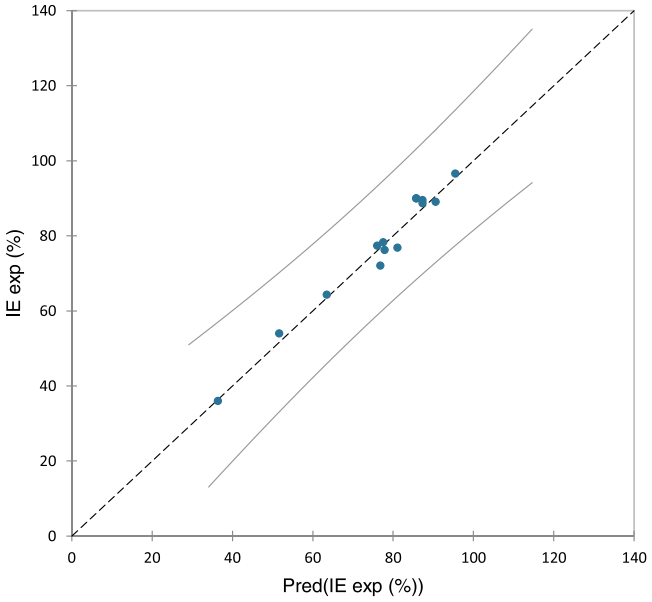

Fig. 5. Correlation between IE_{exp} and IE_{Pred} with PCR modeling results.

Table 8
MLR validation.

Observations	13
Sum of weights	13
DF	6
R^2	0.97
Adjusted R^2	0.94
MSE	15.29
RMSE	3.911
MAPE	3.008
DW	2.22
Cp	4.39
AIC	39.40
SBC	43.36
PC	0.09
Press	446.16
Q^2	0.86

PLS aims to quantitatively forecast the corrosion inhibitor activity of the chemicals under investigation. PLS modeling is typically represented by the equation below:

$$IE\ cal\ (\%) = a_0 + a_1E_{HOMO} + a_2E_{LUMO} + a_3E_{L\text{-}H} + a_4\mu + a_5IP + a_6EA + a_7\chi + a_8\sigma + a_9\eta + a_{10}\Delta N + a_{11}\omega + a_{12}Log\ P + a_{13}Mr + a_{14}V_m + a_{15}\Delta E_{B\text{-}D} \quad (8)$$

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

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

$$IE\ exp\ (\%) = 69{\cdot}284 + 0{\cdot}005E_{HOMO} + 0{\cdot}459E_{LUMO} + 0{\cdot}913E_{L\text{-}H} - 0{\cdot}377\mu - 0{\cdot}005IP - 0{\cdot}459EA - 0{\cdot}503\chi - 17{\cdot}790\sigma + 1{\cdot}826\eta + 0{\cdot}223\Delta N - 0{\cdot}005\omega - 0{\cdot}424Log\ P + 0{\cdot}001Mr + 0{\cdot}0002\,V_m - 7{\cdot}307\Delta E_{B\text{-}D} \quad (9)$$

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

Other validation values can be seen in Table 4. Predicted values and PLS modeling residuals are depicted in Table 5.

Based on the PLS modeling in Equation 2, the significance of each descriptor vs. the standard regression coefficient is depicted in Fig. 3.

Fig. 3 demonstrates how the applicability of corrosion inhibitor values based on molecular structure varies depending on the descriptor. Because the PLS modeling's descriptors don't use the same units, the standard coefficients that are obtained are estimates rather than genuine scales. It suggests that employing these standardized coefficients will not allow for the accurate determination of each descriptor's relative significance in the regression analysis. As a result, its applicability is restricted to evaluating the molecular index's impact on the anticorrosive characteristic under study [44–46].

Regarding statistical parameters, the value of the coefficient of determination (R^2 = 0.10), standard deviation (SD = 17.39), and R^2_{cv} = 667.66, PLS does not seem to be able to accurately anticipate the value of corrosion inhibitors. The residual error value is still large, and there is a large difference between IE_{exp} and IE_{pred} (Fig. 4).

3.4. 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:

$$IE\ exp\ (\%) = 1348{\cdot}903 + 69{\cdot}943E_{HOMO} + 35{\cdot}630E_{LUMO} + 47{\cdot}393E_{L\text{-}H} - 26{\cdot}791\mu - 69{\cdot}943IP - 35{\cdot}630EA - 51{\cdot}559\chi - 253{\cdot}194\sigma + 94{\cdot}787\eta - 173{\cdot}134\Delta N + 3{\cdot}647\omega + 12{\cdot}557Log\ P + 0{\cdot}184Mr - 0{\cdot}018\,V_m - 379{\cdot}151\Delta E_{B\text{-}D} \quad (10)$$

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

Other validation values can be seen in Table 6.

According to the statistical validation results, PCR has better quality in determining furan derivative compounds' corrosion inhibitor value than PLS. It is indicated by good validation results, such as the R^2 = 0.976 and the 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. Fig. 5 also shows that all calculation results are close to fitting data.

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	96.54	96.57	-0.03	-0.008	2.99
FCH	1	89.93	86.41	3.51	0.89	2.93
FMT	1	89.44	87.48	1.95	0.50	2.09
FN	1	89.03	90.82	-1.79	-0.45	3.34
BFA	1	88.60	85.96	2.63	0.67	2.25
FAA	1	78.20	78.07	0.16	0.04	2.67
EF	1	77.34	75.76	1.57	0.40	3.29
MF	1	76.75	82.28	-5.53	-1.41	2.27
MFF	1	76.14	76.89	-0.75	-0.19	3.11
DMFA	1	71.99	76.49	-4.51	-1.15	3.02
FC	1	64.25	61.80	2.44	0.62	2.31
FFA	1	53.93	52.12	1.80	0.46	3.07
NVF	1	35.96	37.44	-1.48	-0.38	3.45

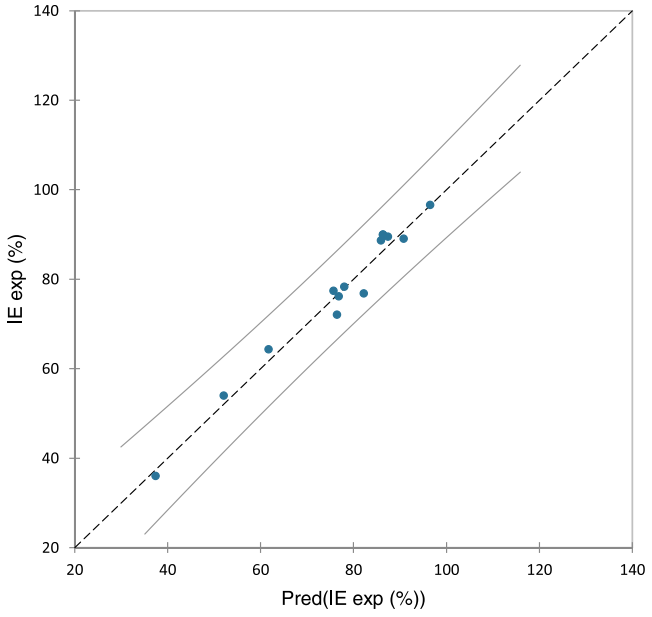

Fig. 6. Correlation between IE_{exp} and IE_{Pred} with MLR modeling results.

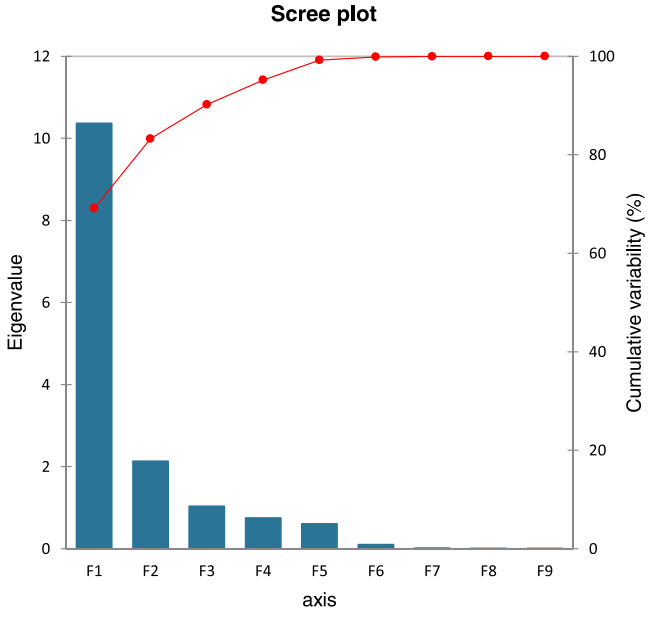

Fig. 7. Principal components dan variannya.

3.5. 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 [47]. The results of MLR modeling are expressed in the following equation:

$$IE\ exp\ (\%) = 1235{\cdot}047 + 286{\cdot}773E_{LUMO} - 27{\cdot}310\mu - 192{\cdot}138\Delta N + 3{\cdot}642\omega + 12{\cdot}627Log\ P + 0{\cdot}163Mr \quad (11)$$

$N = 13\ R^2 = 0.973\ R^2_{adj} = 0.946\ PRESS = 446.163$. Table 8 shows MLR validation.

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

3.6. Principal Component Analysis (PCA)

By reducing a high number of correlated variables to a smaller, uncorrelated collection of variables, PCA can be used to minimize the dimensionality of massive data sets [48]. Primary components are the name for these new variables. The practitioner can streamline the data and minimize the number of variables [49]. The principal component or principal axis is the name of the new variable. It enables the researcher to lower the number of variables and simplify the dense information.

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 Fig. 7.

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

Fig. 8 shows the projection of the first three main component variables, F1, F2, and F3, according to the percentage contribution of each variable in the two correlation graphs. The axes account for as much variation in the data as they can. They make up, respectively, 69.11%, 14.19%, and 6.91% of the overall variance, with an estimated total information percentage of 90.21%. This number is adequate for describing the data set's information.

The matrix's correlation coefficient tells us whether the descriptors have a strong or weak relationship. Generally, highly correlated

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}	0.78	5.97	0.46	10.22	0.32	10.44
E_{LUMO}	0.99	9.51	0.05	0.11	0.02	0.05
E_{L-H}	0.94	8.61	-0.20	2.01	-0.16	2.59
μ	-0.89	7.71	-0.01	0.006	-0.08	0.64
IP	-0.78	5.97	-0.46	10.22	-0.32	10.44
EA	-0.99	9.51	-0.05	0.12	-0.02	0.05
χ	-0.97	9.08	-0.18	1.52	-0.11	1.34
σ	-0.91	8.01	0.21	2.25	0.27	7.49
η	0.94	8.61	-0.20	2.01	-0.16	2.59
ΔN	0.98	9.40	0.04	0.08	0.06	0.36
ω	-0.84	6.80	-0.10	0.52	-0.03	0.11
Log P	-0.31	0.92	-0.20	2.01	0.71	48.65
Mr	-0.21	0.45	0.79	29.80	-0.35	11.96
V_m	-0.285	0.78	0.88	37.09	-0.08	0.65
ΔE_{B-D}	-0.94	8.61	0.20	2.01	0.16	2.59

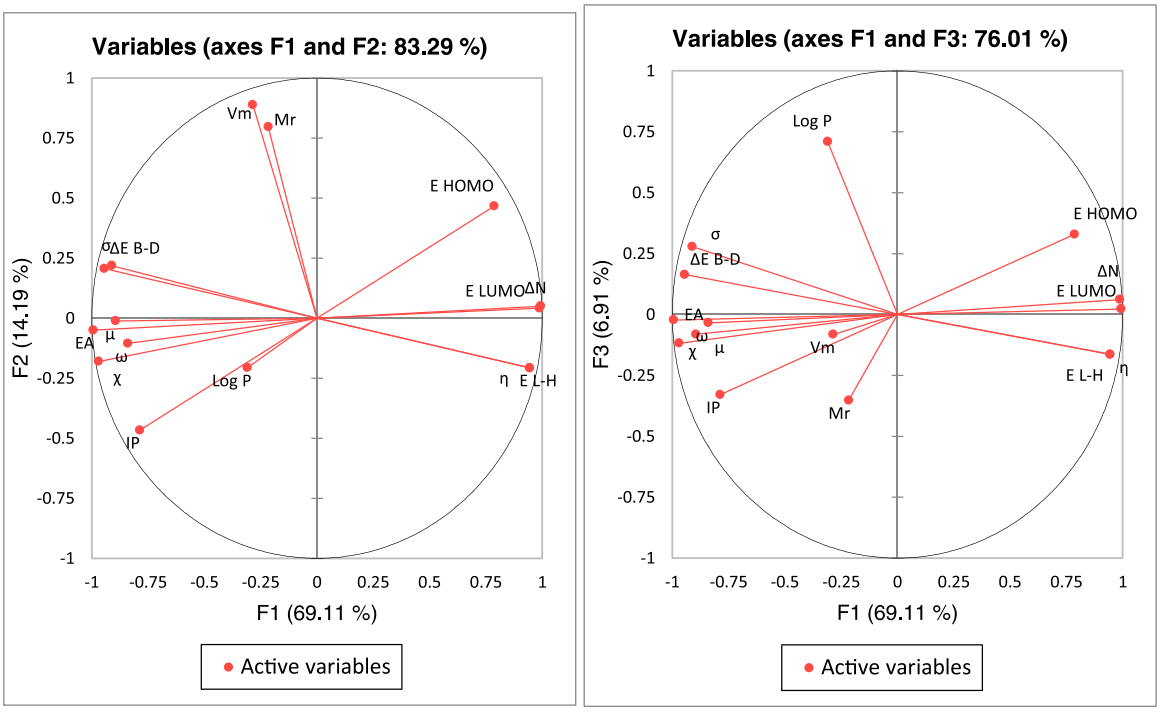

Fig. 8. Principal compound F1-F2 and F1-F3 correlation circles.

Table 11
Tolerance and VIF descriptors.

	Tolerance	VIF
E_{HOMO}	0.000	
E_{LUMO}	0.000	
E_{L-H}	0.000	
μ	0.005	186.33
IP	0.000	
EA	0.000	
χ	0.000	
σ	0.002	490.17
η	0.000	
ΔN	0.000	2610.16
ω	0.007	150.711
Log P	0.466	2.14
Mr	0.149	6.72
V_m	0.165	6.07
ΔE_{B-D}	0.000	

descriptors ($R \geq 0.75$) are not included to reduce the redundancy that exists in the matrix data [49,50]. Table 11 indicates perfect negative collinearity ($R = -1$) between E_{HOMO} and IP; E_{LUMO} and EA; E_{L-H} and ΔE_{B-D}; and η and ΔE_{B-D}. In addition, there are other strong negative collinearities such as E_{HOMO} and EA ($R = -0.81$), E_{HOMO} and χ ($R = -0.90$), 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} ($R = 0.81$), E_{HOMO} and ΔN (0.819), etc. It indicates that these variables are redundant.

According to the PCA descriptive data results, there appears to be strong collinearity between descriptors. In this case, the matrix $(X'X) - 1$ cannot be inverted because many variables have 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 the most useful information in making the desired model. Apart from MLR, Multiple Polynomial Regression (MPR) also cannot be used.

Table 12
Adsorption energy prediction compounds.

No.	Predicted compounds	Adsorption energy inhibitor (kJ/mol)	Adsorption energy water (kJ/mol)	IE pred (%)
1.	BMOPF	-272.96	-5.03	169.37
2.	BOMF	-160.04	-6.78	78.68
3.	FMP	-109.55	-6.67	100.81
4.	NAA	-122.62	-2.19	66.18

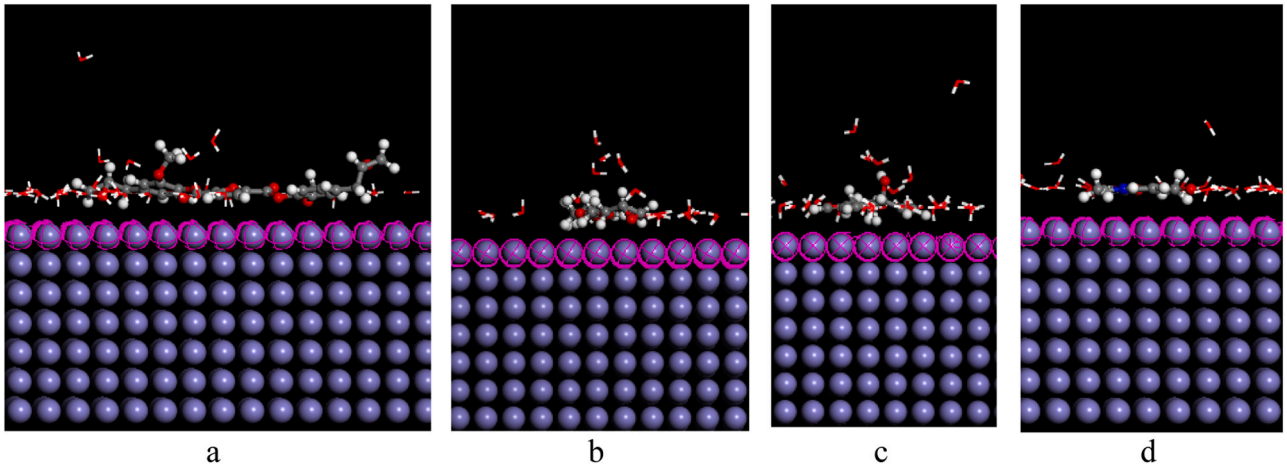

Fig. 9. Visualization from Monte Carlo simulation of 4 prediction compounds, (a) BMOPF; (b) BOMF; (c) FMP; (d) NAA.

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.

3.7. 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 12.

Table 12 shows 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 [51]. Monte Carlo simulation results are shown in Table 12 and Fig. 9 (viewed from the side). Monte Carlo simulation results (viewed above) are illustrated in Fig. 9.

4. Conclusion

Calculating molecular descriptors using the DFT quantum method makes it possible to correlate the IE% with molecular descriptors using the QSPR approach. 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 = 0.97; R^2_{adj} = 0.90) 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 four new derivative compounds are very promising, especially for the BMOPF compound.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Saprizal Hadisaputra reports financial support was provided by Kemenristekdikti Republic of Indonesia through Penelitian Dasar Kompetitif Nasional Grant number 2360/UN 18. L1/PP/2023.

Acknowledgments

Financial assistance was provided by the Kemenristekdikti Republic of Indonesia and is warmly acknowledged (grant number 2360/UN 18. L1/PP/2023).

References

[1] A. Zakeri, E. Bahmani, A.S.R. Aghdam, Plant extracts as sustainable and green corrosion inhibitors for protection of ferrous metals in corrosive media: A mini review, Corros. Commun. 5 (2022) 25–38, https://doi.org/10.1016/j.corcom.2022.03.002
[2] P. Satyabama, S. Rajendran, T.A. Nguyen, Corrosion inhibition of aluminum by oxalate self-assembling monolayer, Anti Corros. Methods Mater. 66 (6) (2019) 768–773, https://doi.org/10.1108/acmm-01-2019-2061
[3] M. Honarvar Nazari, et al., Nanocomposite organic coatings for corrosion protection of metals: A review of recent advances, Prog. Org. Coat. 162 (2022) 106573, , https://doi.org/10.1016/j.porgcoat.2021.106573
[4] H. Assad, A. Kumar, Understanding functional group effect on corrosion inhibition efficiency of selected organic compounds, J. Mol. Liq. 344 (2021) 117755, , https://doi.org/10.1016/j.molliq.2021.117755
[5] M.A. Jafar Mazumder, A, Review of green scale inhibitors: process, types, mechanism and properties, Coatings 10 (10) (2020) 928, https://doi.org/10.3390/coatings10100928
[6] A.O. Alao, A.P. Popoola, O. Sanni, The influence of nanoparticle inhibitors on the corrosion protection of some industrial metals: a review, J. Bio Tribo Corros. 8 (2022) 3, https://doi.org/10.1007/s40735-022-00665-1
[7] K. Tamalmani, H. Husin, Review on corrosion inhibitors for oil and gas corrosion issues, Appl. Sci. 10 (10) (2020) 3389, https://doi.org/10.3390/app10103389
[8] H. Wei, B. Heidarshenas, L. Zhou, G. Hussain, Q. Li, K. (Ken) Ostrikov, Green inhibitors for steel corrosion in acidic environment: state of art, Mater. Today Sustain. 10 (2020) 100044, , https://doi.org/10.1016/j.mtsust.2020.100044
[9] Y.G. Avdeev, Y.I. Kuznetsov, Nitrogen-containing five-membered heterocyclic compounds as corrosion inhibitors for metals in solutions of mineral acids-An overview, Int. J. Corros. Scale Inhib. 10 (2) (2021) 480–540, https://doi.org/10.17675/2305-6894-2020-10-2-2
[10] S. Issaadi, T. Douadi, S. Chafaa, Adsorption and inhibitive properties of a new

heterocyclic Furan Schiff base on corrosion of copper in hcl 1 m: Experimental and theoretical investigation, Appl. Surf. Sci. 316 (2014) 582–589, https://doi.org/10.1016/j.apsusc.2014.08.050
[11] A.L. Mokale Kognou, et al., High-fructose corn syrup production and its new applications for 5-hydroxymethylfurfural and value-added furan derivatives: Promises and challenges, J. Bioresour. Bioprod. 7 (3) (2022) 148–160, https://doi.org/10.1016/j.jobab.2022.03.004
[12] C.T. Ser, P. Žuvela, M.W. Wong, Prediction of corrosion inhibition efficiency of pyridines and quinolines on an iron surface using machine learning-powered quantitative structure-property relationships, Appl. Surf. Sci. 512 (2020) 145612, , https://doi.org/10.1016/j.apsusc.2020.145612
[13] E.H. El Assiri, et al., Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium, Heliyon 6 (10) (2020) e05067, , https://doi.org/10.1016/j.heliyon.2020.e05067
[14] T.W. Quadri, et al., Predicting protection capacities of pyrimidine-based corrosion inhibitors for mild steel/HCl interface using linear and nonlinear QSPR models, J. Mol. Model. 28 (2022) 9, https://doi.org/10.1007/s00894-022-05245-1
[15] D. Awfa, M. Ateia, D. Mendoza, C. Yoshimura, Application of quantitative structure–property relationship predictive models to water treatment: a critical review, ACS EST Water 1 (3) (2021) 498–517, https://doi.org/10.1021/acsestwater.0c00206
[16] T.W. Quadri, et al., Computational insights into quinoxaline-based corrosion inhibitors of steel in HCl: Quantum chemical analysis and QSPR-ANN studies, Arab. J. Chem. 15 (7) (2021) 103870, , https://doi.org/10.1016/j.arabjc.2022.103870
[17] T.W. Quadri, et al., Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors, Mater. Today Commun. 30 (2022) 103163, , https://doi.org/10.1016/j.mtcomm.2022.103163
[18] R.L. Camacho-Mendoza, L. Feria, L.Á. Zárate-Hernández, J.G. Alvarado-Rodríguez, J. Cruz-Borbolla, New QSPR model for prediction of corrosion inhibition using conceptual density functional theory, J. Mol. Model. 28 (2022) 8, https://doi.org/10.1007/s00894-022-05240-6
[19] A.M. Al-Fakih, H.H. Abdallah, M. Aziz, Experimental and theoretical studies of the inhibition performance of two furan derivatives on mild steel corrosion in acidic medium, Mater. Corros. 70 (1) (2018) 135–148, https://doi.org/10.1002/maco.201810221
[20] J.-T. Miao, L. Yuan, Q. Guan, G. Liang, A. Gu, Biobased heat resistant epoxy resin with extremely high biomass content from 2,5-Furandicarboxylic acid and eugenol, ACS Sustain. Chem. Eng. 5 (8) (2017) 7003–7011, https://doi.org/10.1021/acssuschemeng.7b01222
[21] J. Meng, et al., Flame retardancy and mechanical properties of bio-based furan epoxy resins with high crosslink density, Macromol. Mater. Eng. 305 (1) (2019) 1900587, https://doi.org/10.1002/mame.201900587
[22] J. Nowicki, J. Kula, D. Hammad, Synthesis of new furan-type terpenoids, Flavour Fragr. J. 17 (3) (2002) 203–206, https://doi.org/10.1002/ffj.1090
[23] T.T. Pham, X. Chen, N. Yan, J. Sperry, A novel dihydrodifuropyridine scaffold derived from ketones and the chitin-derived heterocycle 3-acetamido-5-acetylfuran, Mon. Chem. 149 (4) (2017) 857–861, https://doi.org/10.1007/s00706-017-2112-8
[24] A. Frisch, (2009). gaussian 09W Reference. Wallingford, USA, 25p, 470.
[25] T. Koopmans, Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen eines atoms, Physica 1 (1–6) (1934) 104–113, https://doi.org/10.1016/s0031-8914(34)90011-2
[26] N. Islam, D. Chandra Ghosh, A new algorithm for the evaluation of the global hardness of polyatomic molecules, Mol. Phys. 109 (6) (2011) 917–931, https://doi.org/10.1080/00268976.2011.558856
[27] R.G. Parr, L. v Szentpály, S. Liu, Electrophilicity index, J. Am. Chem. Soc. 121 (9) (1999) 1922–1924, https://doi.org/10.1021/ja983494x
[28] R.G. Pearson, “Hard and soft acids and bases—the evolution of a chemical concept, Coord. Chem. Rev vol. 100, (1990), pp. 403–425, https://doi.org/10.1016/0010-8545(90)85016-l
[29] R.G. Pearson, Absolute electronegativity and hardness: Application to inorganic chemistry, Inorg. Chem. 27 (4) (1988) 734–740, https://doi.org/10.1021/ic00277a030
[30] S. Issaadi, T. Douadi, S. Chafaa, Adsorption and inhibitive properties of a new heterocyclic Furan Schiff base on corrosion of copper in hcl 1 m: Experimental and theoretical investigation, Appl. Surf. Sci. 316 (2014) 582–589, https://doi.org/10.1016/j.apsusc.2014.08.050
[31] M. Meunier, S. Robertson, Materials studio 20th anniversary, Mol. Simul. 47 (7) (2021) 537–539, https://doi.org/10.1080/08927022.2021.1892093
[32] A.H. Alamri, N. Alhazmi, Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors, J. Saudi Chem. Soc. 26 (6) (2022) 101536, , https://doi.org/10.1016/j.jscs.2022.101536
[33] T.W. Quadri, et al., Multilayer perceptron neural network-based QSAR models for the assessment and prediction of corrosion inhibition performances of ionic liquids, Comput. Mater. Sci. 214 (2022) 111753, , https://doi.org/10.1016/j.commatsci.2022.111753
[34] S. Hadisaputra, A.A. Purwoko, L.R.T. Savalas, N. Prasetyo, E. Yuanita, S. Hamdiani, Quantum chemical and monte carlo simulation studies on inhibition performance of caffeine and its derivatives against corrosion of copper, Coatings 10 (11) (2020) 1086, https://doi.org/10.3390/coatings10111086
[35] S. Xiong, H. Wu, Z. Liu, B. Zhang, QSAR models for the prediction of the relationship among corrosion inhibition efficiency, friction coefficient and oil film strength of lubricants, Polycycl. Aromat. Compd. 42 (6) (2021) 3780–3791, https://doi.org/10.1080/10406638.2021.1873806
[36] M. Lahyaoui, A. Diane, H. El-Idrissi, T. Saffaj, Y.K. Rodi, B. Ihssane, 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 9 (1) (2023) e13020, , https://doi.org/10.1016/j.heliyon.2023.e13020
[37] A. Golbraikh, A. Tropsha, Beware of q2!. 2002, J. Mol. Graph. Model 20 (4) (2002) 269–276, https://doi.org/10.1016/s1093-3263(01)00123-1
[38] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inf. 29 (6–7) (2010) 476–488, https://doi.org/10.1002/minf.201000061
[39] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (5) (2007) 694–701, https://doi.org/10.1002/qsar.200610151
[40] A.D. Becke, Density-functional thermochemistry. I. The effect of the exchange-only gradient correction, J. Chem. Phys. 96 (3) (1992) 2155–2160, https://doi.org/10.1063/1.462066
[41] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A 38 (6) (1988) 3098–3100, https://doi.org/10.1103/physreva.38.3098
[42] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B 37 (2) (1988) 785–789, https://doi.org/10.1103/physrevb.37.785
[43] M. Randić, Book review of molecular descriptors for chemoinformatics—second, revised and enlarged edition (Volume I: Alphabetical Listing; Volume II: Appendices, Bibliography,), J. Chem. Inf. Model vol. 50, (2) (2010) 326, https://doi.org/10.1021/ci900493b
[44] K. Hasegawa, K. Funatsu, Partial least squares modeling and genetic algorithm optimization in quantitative structure-activity relationships, SAR QSAR Environ. Res. 11 (3–4) (2000) 189–209, https://doi.org/10.1080/10629360008033231
[45] R.D. Cramer, Partial least squares (PLS): its strengths and limitations, Perspect. Drug Discov. Des. 1 (2) (1993) 269-27, https://doi.org/10.1007/bf02174528
[46] D.T. Stanton, QSAR and QSPR model interpretation using partial least squares (PLS) analysis, Curr. Comput. -Aided Drug Des. 8 (2) (2012) 107–127, https://doi.org/10.2174/157340912800492357
[47] H. Nasution, N. Enizan, N. Nurlaili, J. Syahri, Design of trolox compounds as antioxidant and their analysis using quantitative structure activity relationship, Acta Chim. Asian 3 (2) (2020) 181–186, https://doi.org/10.29303/aca.v3i2.40
[48] R. Hmamouchi, M. Larif, S. Chtita, A. Adad, M. Bouachrine, T. Lakhlifi, Predictive modelling of the LD50 activities of coumarin derivatives using neural statistical approaches: Electronic descriptor-based DFT, J. Taibah Univ. Sci. 10 (4) (2016) 451–461, https://doi.org/10.1016/j.jtusci.2015.06.013
[49] J.F. Hair, C.M. Ringle, M. Sarstedt, Partial least squares structural equation modeling: rigorous applications, better results and higher acceptance, Long. Range Plan. 46 (1–2) (2013) 1–12, https://doi.org/10.1016/j.lrp.2013.01.001
[50] J. Henseler, C.M. Ringle, M. Sarstedt, Testing measurement invariance of composites using partial least squares, Int. Mark. Rev. 33 (3) (2016) 405–431, https://doi.org/10.1108/imr-09-2014-0304
[51] S. Hadisaputra, A.A. Purwoko, A. Hakim, N. Prasetyo, S. Hamdiani, Corrosion inhibition properties of phenyl phthalimide derivatives against carbon steel in the acidic medium: DFT, MP2, and monte carlo simulation studies, ACS Omega 7 (37) (2022) 33054–33066, https://doi.org/10.1021/acsomega.2c03091