Prediction aluminum corrosion inhibitor efficiency using artificial neural network (ANN)

Sh Ebrahimi[1], E G Kalhor[2], S R Nabavi[3], L Alamiparvin[2], R Pogaku[1*]
[1] Faculty of Engineering, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia
[2] Department of Chemistry, Faculty of Science, Tabriz branch, Islamic Azad University, Tabriz, Iran
[3] Department of Applied Chemistry, University of Mazandaran, Babolsar, Iran

[1]E-mail: [1]ebrahimi_shamsi63@yahoo.com, [1]ravindra@ums.edu.my

Abstract. In this study, activity of some Schiff bases as aluminum corrosion inhibitor was investigated using artificial neural network (ANN). Hence, corrosion inhibition efficiency of Schiff bases (in any type) were gathered from different references. Then these molecules were drawn and optimized in Hyperchem software. Molecular descriptors generating and descriptors selection were fulfilled by Dragon software and principal component analysis (PCA) method, respectively. These structural descriptors along with environmental descriptors (ambient temperature, time of exposure, pH and the concentration of inhibitor) were used as input variables. Furthermore, aluminum corrosion inhibition efficiency was used as output variable. Experimental data were split into three sets: training set (for model building) and test set (for model validation) and simulation (for general model). Modeling was performed by Multiple linear regression (MLR) methods and artificial neural network (ANN). The results obtained in linear models showed poor correlation between experimental and theoretical data. However nonlinear model presented satisfactory results. Higher correlation coefficient of ANN (R>0.9) revealed that ANN can be successfully applied for prediction of aluminum corrosion inhibitor efficiency of Schiff bases in different environmental conditions.

1. Introduction

Corrosion is devastating effect of environment on metal. In another word, corrosion is destruction of metal due to chemical or electrochemical reaction between the metal and its surroundings. Corrosion may be the most common electrochemical phenomenon experienced in everyday life. Common examples of corrosion include rusting of iron and steel, darkening of silver and copper, breaking out and blowing of chromium coating and paint on the body of automobile [1]. Corrosion is a serious problem, since spends natural resources, weakens steel structures, increases the cost of maintenance, deprive human comfort and lead to human mortality [2]. Aluminum is soft and light but strong metal, with a silver – gray appearance, opaque thin oxidation layer that is formed on the surface due to exposure to air, that prevents further rusting. The weight of aluminum is almost one-third of the weight of steel or copper. It is very durable and resistant to rust. Due to the high technological importance of aluminum, methods are provided to prevent aluminum corrosion. One of these methods

[1] To whom any correspondence should be addressed.

is application of the organic inhibitors of Schiff Bases. Schiff bases are organic inhibitors with considerable importance in organometallic chemistry. They are usually used as ligands because of easy connection of their nitrogen to transition metals. Schiff bases are compound with a functional group that contains carbon-nitrogen double bond with the nitrogen atom connected to an aryl or alkyl group. Schiff base is formed from the condensation reaction of an amine group with the carbonyl group of an aldehyde or ketone. For the first time in 1864 it was reported by Professor Schiff, and named in his honor [3]. The inhibitory effect of some Schiff bases on the corrosion of aluminum has been studied by polarization, electrochemical impedance spectroscopy (EIS) and weight loss measurements. According to the results of all three methods, Schiff bases are excellent inhibitors and their efficiency increases with temperature reduction and increase of their concentration [4].

Quantitative Structure Activity Relationship (QSAR) is one of the major application fields of chemometrics in the studies linking molecule properties to their structural characteristics. QSAR is the logical relationship between the properties of compounds (active) and their chemical structure. QSAR is a theoretical method beneficial in relating structural based parameters to corrosion inhibition efficiencies of a group of compounds. Reviewing the relationship between chemical structure and activity of compounds enables predicting the activity of new compounds based on information about their chemical structure. The main objective of the present study is application of QSAR to predict inhibition rate of Schiff Bases for aluminum corrosion [5-6]. There is variety of models in QSAR studies; in this study the focus is on multiple linear regression models (MLR) and Artificial Neural Network (ANN).The regression model is linear, while artificial neural network is nonlinear. In this study, artificial neural network modeling indicates the best correlation between the actual data of corrosion inhibition of Schiff bases for aluminum corrosion and data from the model.

2. Methodology

The aim of the present study is establishment of quantitative structure-activity relationship to acquire the inhibitor efficiency of Schiff bases for aluminum corrosion based on theoretical method.

Application of this method in estimation of the inhibitor efficiency of Schiff bases for aluminum corrosion includes:

2.1. Selection of Schiff base molecule from resources and collection of data for the inhibitory efficiency

In this study the Schiff base molecules and required data, including corrosion inhibitory activity of different molecules for aluminum element, were extracted from scientific resources where the inhibitory efficiency of them is examined in different conditions such as temperature, time, pH, and concentration of inhibitors. Temperature, time, pH, concentration are independent variables, and the inhibitory efficiency is considered as the dependent variable. Number of Schiff base molecules is 42 extracted from references [3-4] and [7-24].

2.2. Drawing the form of molecules by Hyperchem

After extracting Schiff base molecules from resources, the structure of organic compounds (Schiff bases) is drawn by Hyperchem software.

2.3. Optimization of molecules by Hyperchem

To optimize Schiff base molecules the Hyperchem is used. Inhibitory molecules are in the most volatile state regarding energy and spatial formulation. In order to the inhibitory molecules to reach the most appropriate conformation and stable energy state, they can be optimized by the software.

2.4. Calculation of descriptors by Dragon3

Molecular descriptors are the final result of mathematical reasoning and methods that codify chemical information and indicate them as symbols providing the molecule as a valuable number. Descriptors are numerical values representing the properties of molecules. Each of the descriptors provides

specific information about the molecule. For example, descriptors may be physical and chemical properties of a molecule such as molecular weight or the number of hydrogen bonds, the number of non-hydrogen bonds, the number of connected halogen, etc.

After determination of the most appropriate conformation, molecule descriptors are derived and selected by the software; here the Dragon3 has been used. Dragon3 Software is used to calculate descriptors simultaneously. The software can calculate 1497 descriptors for the molecules with up to 150 atoms. If the inhibitory molecule (Schiff bases) is anionic or cationic, after calculation of anion and cation descriptors, to calculate Schiff base descriptors, anion and action descriptors are collected and general molecule descriptor obtained.

2.5. Reduction of descriptors using PCA

As mentioned above, 1497 descriptors were calculated for each Schiff base by Dragon3 software. The number of independent variables to create QSAR model is very high and not all of these descriptors contain information. Because of high number of descriptors, to put them in neural network meaningful descriptors should be used. In this study, to further reduce the descriptors to meaningful descriptors, the principal component analysis (PCA) is used.

To perform the PCA on a series of data, the coordination axes are rotated; most of data changes are in the first new axis and from remaining changes, most of the changes are on the second new axis and from the remaining changes, most of the changes are on the third axis, etc. Data coordination in the new axis is called score. Applying the PCA, primary data converted to score. To reduce the dimensions of data, scores with the maximum changes (information) are selected and scores with changes around the system noise are removed [25].

2.6. Transferring descriptors (structural parameters) to Microsoft Excel and addition of environmental parameters (temperature, concentration, pH)

2.7. Data Matrix

Since the total number of molecules is 42, 5 descriptors for each molecule were obtained from PCA method. Therefore, molecular descriptors, as well as T (temperature), C (inhibitory concentration), t (time), and pH are independent and dependent variables of corrosion inhibitory activity of Schiff base for aluminum, respectively. Adding these environmental variables to structural descriptors, the number of independent variables will be 9. The total number of data is 241, hence, dimension of independent data matrix will be 241 × 9 and dependent data matrix will be 241 × 9. Then, for randomization of data, they were codified as 0, 0, 1, 2 and classified in three sets: training data matrix with code (0), test set with Code (1) and simulation set with code (2); training set is used to create a model between dependent and independent variables, test set to predict and evaluate the performance of the model in education, and simulation set is used for general training model. Training series cover most of the data because the probability for this series is double other series. From the total 241 data, 140 data belong to training set; 49 data for test set, and 52 data for the simulation set, regarding that training data matrix is 140 × 9, test data matrix is 49 × 9 and simulation data matrix is 52 × 9.

2.8. Multi-stage linear regression modeling and MLR model

In the simple linear regression, a linear relationship is obtained between the independent and the dependent variables. In fact, in the multiple linear regressions there is a relationship between several independent variables and the dependent variable. The goal is to study first-order linear model as Eq. (1).

$$y=\beta 0+\beta 1X1+\beta 2X2+\beta 3X3+\beta 4X4+\beta 5X5+\beta 6T+\beta 7C+\beta 8t+\beta 9pH \quad (1)$$

The confidence level has selected 95%. Therefore, independent variables with meaningfulness (p-value) of more than 0.05 will be removed in the final stage equation [26].

2.9. Modeling with artificial neural network (ANN)

Since the relationship between dependent and independent variables may be non-linear, it's normal for linear regression models to show high error for non-linear functions. Therefore, it is better to use a method for approximation of nonlinear functions. One of these methods is modeling using artificial neural networks. Artificial neural networks are nonlinear mapping structures based on the performance of the human brain. Interest to neural networks is due to the ability of the network to mimic the human brain, in addition to his ability to learn and respond. Therefore, neural networks, with its remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and identify the ways very complex and difficult to humans and other computer techniques. A trained neural network can be considered a specialist in analyzing the given information [27]. Modeling by artificial neural network, parameters should be optimized.

2.9.1. Optimization of artificial neural network parameters and modeling. At this stage, neural network is designed and optimized using Neural Power software to predict corrosion inhibition of Schiff bases for aluminum corrosion. To create a model by neural network, network parameters such as the number of neurons, epoch, learning rate and momentum, learning function and learning algorithms, are optimized. Optimization is done one at a time.

2.10. Calculation of error to assess the accuracy of the model using the following formula:

$$MS = \left(\frac{\sum_{s=1}^{ns}\sum_{i=1}^{ni}(y_{si}-\mathrm{out_{si}})^2}{n_i n_s}\right)^{1/2} \quad (2)$$

$$MPE = \frac{1}{N}\sum_{i=1}^{N}|r_i - y_i| \quad (3)$$

$$.S.E(\%) = \left[\frac{\sum_{i=1}^{N}(r_i-y_i)^2}{\sum_{i=1}^{N}(r_i)^2}\right]^{1/2} \times 100 \quad (4)$$

Where, N, ri and yi are the number of data, the real response and the network response or output to the desired parameters, respectively. Ni and n are the number of inputs and outputs of variables. Ysi, is the i th component of expected response ys and outsi, i- th output component produced by the network for the s- th input vector.

3. Results and discussion

3.1. Multiple linear regression model and its results

After creating the appropriate descriptors by PCA and Dragon3, Schiff base divided into three sets of training and testing, and simulation. Training set moved to MINITAB software to create multiple linear regression models. After two linear steps, the equation calculated by the training set. To check if the resulting equation is acceptable, it is studied by test and simulation set. Equation 5 shows first-order linear equation fitted in the first phase for training set.

$$y=89.5-0.000011x1-0.000124x2+0.000382x3+0.00044x4+0.000203x5-0.780T+33.5C-0.00683t-13.4 \quad (5)$$

Coefficient of multiple linear regression model is presented in table 1.

Table 1. P- Values for the independent variables in linear equation (first stage of) with multiple linear regression method

Independent variables	X_1	X_2	X_3	X_4	X_5	T	C	t	pH
Values P	0.210	0.412	0.269	0.26	0.769	0.000	0.022	0.001	0.001

As indicated in Table 1, regarding the selected level of confidence (95%) and in some descriptors P-value is greater than 0.05, these coefficients are not meaningful and should be removed. After removal of these coefficients training equation will be fit and change to Eq. 6 in the second level.

$$y=91.5-0.757T+27.9C-0.00784t-12.3pH \quad (6)$$

3.2. Modeling with multi-layer neural network

Data of neural network design, like multiple linear regressions, is randomly divided into three groups of training, test, and simulation. Training set creates model, test set predicts network performance and simulation generalizes the created network. In desired network design, initial weight values were randomly from [1, -1] and before the training input values were between [1, 1], and output regulated from [2.0 to 8.0]. The number of entries in the input layer is selected equal to the number of variables such as temperature, concentration, pH, time, and the number of descriptors (Scores) selected by the PCA method. Latent layer neurons are determined during optimization and the output layer neurons are equal to the number output variables, with the same inhibition percentage of corrosion. Neural Power is used in modeling the artificial neural network and normalization of the data is done by the software.

3.2.1. Optimization of neural network parameters in Neural Power software. At this stage parameters of the neural network, including the number of latent layer neurons, momentum, learning rate, repetition, learning algorithms and functions, are optimized one at a time. The optimization results of neural network parameters are shown in Table 2.

Table 2. shows the results from the optimization of neural network parameters

Optimized parameters	**Results**
The number of hidden layer neurons	20
Learning rate	0.32
Momentum coefficient	0.68
The number of repetitions	64
Hidden layer and output layer transfer function	Sigmoid function
Learning algorithm	Levenberg Marquardt
The number of hidden layers	1

3.2.2. Calculated errors for optimized neural network. The neural network model was evaluated after optimization. According to the results of neural network model and correlation coefficients of training, test and simulation, the output values of the model were compared and evaluated to experimental values. The error between the predicted and actual values for the training and test of designed neural network is calculated in three forms of RMSE, MPE, RSE and the correlation coefficient (R). Error and correlation coefficients are given in Table 3. According to table 3, the errors of ANN are less, and correlation coefficient is higher than regression models.

Table 3. The correlation coefficient values and errors for neural network optimized.

set	**R**	**RMSE**	**MPE**	**RSE**
training	0.908	8.2799	6.2674	12.8557
test	0.906	9.6604	7.2995	13.2178
simulation	0.878	14.9288	12.0745	19.8114

Results of comparing errors of test set, for three models of multiple linear regression, and ANN (nonlinear method) are indicated in table 4.

Table 4. Comparison of the errors for linear regression models and artificial neural network for test set

models	RMSE	MPE	RSE	R
multiple linear regression	16.9626	14.1420	23.2089	**0.5193**
ANN	9.66049	7.2995	13.2178	**0.9069**

Correlation coefficient between output of neural network and experimental amount of corrosion inhibition for evaluation is reviewed according to figure1.

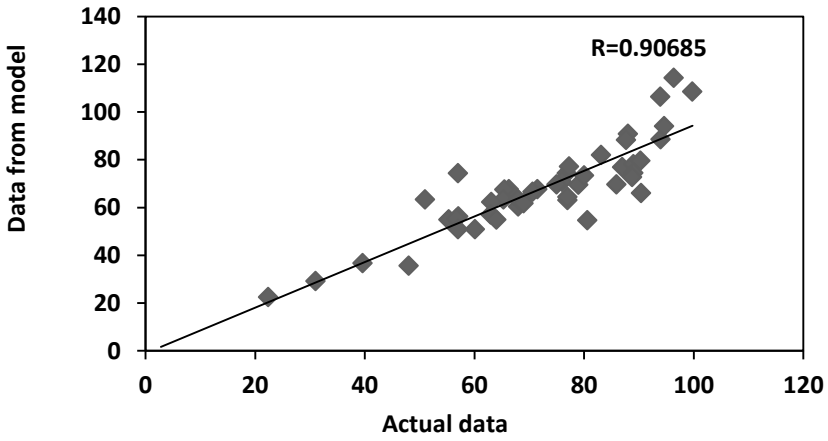

Figure 1. The correlation coefficient (R) show good correlation for evaluation (test) of data between the real data and the data from artificial neural network model.

4. Conclusion

RSE error of test set is high in the multiple linear regression models (23.2089), while the value of R is low that it is the reason for the rejection of this model for corrosion inhibition aluminum by Schiff Bases. R (0.5193) does not indicate good correlation between the model and the real value.

Due to the reduction of RSE error rate for test set in multilayer neural network model, and increase of R (~0.9069), it was concluded that this model, as a non-linear model, can predict the corrosion inhibition of aluminum by Schiff bases. Due to the increase of R in this model, there is good correlation between the actual value and the value of model.

Modeling in this study is through variable descriptors along with the high number of environmental variables, including temperature, concentration, time and pH; high R value (0.9069) for test set of multilayer neural networks model shows the importance of this modeling. It worth to note that, so far in no studies the modeling has been through these environmental variables (Figure 1). The figure shows the correlation coefficient between the actual data of inhibitory Schiff bases for aluminum corrosion and the data of artificial neural network for test set.

5. References

[1] Faran Hofar 2012 *Principals of Corrosion* Jahad Daneshgahi, Sharif Science and Technology University.
[2] Razi S M 1997 *Corrosion Control in Industries* vol. **1** Corrosion Association of Iran

[3] Gomma G K and Wahdan M H 1995 *Mater. Chem. Phys*., **39** 209-13.
[4] Ashassi-Sorkhabi H b B S, Aligholipour B and Seifzadeh D 2006 Appl. Surf .Sci **252**, 4039-47.
[5] Sayed El, Ashry E and Senior S 2011 *Corros.Sci.,* **53**, 1025-34.
[6] Mousavi M, Mohamadalizadeh M and Khosravan A 2011 *Corros,Sci.,* **53**, 3086-91.
[7] Khaled K F 2010 *Corros.Sci.,* **52**, 2905-16.
[8] Obot1 I B, Obi-Egbedi N O and Umoren S A 2009 *Int. J. Electrochem. Sci*., **4**, 863-77.
[9] Foda A S, Elewady G Y, EL-Askalany A and Shalabi K 2010 *ZAŠT. MAT.,* **51**.
[10] Fakrudeen S P, B L H, C A M H and B B R V 2012 *I.J.E.R.A*, **5**, 2049-61.
[11] Naik U J, Panchal V A, Patel A S and Shah N K 2012 *J. Mater. Environ,* **3**, 935-46.
[12] Halambek J, Jukić M, Berković K and Vorkapić-Furač J 2012 *Int. J. Electrochem. Sci*, **7**, 1580-1601.
[13] Yurt A and Aykın Ö 2011 *Corros.Sci*., **53**, 3725-32.
[14] A Y, S U and Dal H 2006 *Appl. Surf.Sci*., **253**, 919- 25.
[15] Safak S, Duran B, Yurt A and Türkog˘lu G 2012 *Corros.Sci*., **54**, 251-59.
[16] Pinto G M, Nayak J and Shetty A N 2011 *Nano. Metal. Chem.*
[17] Noor E A 2009 *Mater.Chem.Phys*., **114**, 533-541.
[18] Li X, Deng S and Fu H 2011 *Corros.Sci*., **53**, 1529-36.
[19] Bhat J I and Alva V D P 2012 *Nano. Metal. Chem.,* **42**, 951-57.
[20] Oguzie E E, Akalezi C O, Enenebeaku C K and Aneke A J N 2011 *Chem. Eng. Comm,* **198**, 46-60.
[21] Aytac A, zmen U O and Kabasakalog˘lu M 2005 *Mater.Chem.Phys.,* **89**, 176-81.
[22] Obot, I. B. and Obi-Egbedi, N. O 2008 *Phy. chem. Eng.,* **330**, 207-12.
[23] Shabab R S A 1999 Retrieved from www.alazhar.edu.ps/Library/aattachedFile.asp?id_no=0044099
[24] Patel A S, Panchal V A and Shah N K 2012 *Bull. Mater.Sci,* **35**, 283-90.
[25] Esbensen K and Geladi P 1987 *Chemometr. Intell. Lab*., **2**, 36-52.
[26] Soleimanian M H 2009 *M.A. thesis,* Islamic Azad University, Tabriz branch, Tabriz, Iran
[27] ilouchian A, and Gamshidi M O 2001 *CRC. press.,* **65** pp. 42-49