Using Neural Networks for Corrosion Inhibition Efficiency Prediction during Corrosion of Steel in Chloride Solutions

K.F. Khaled[1,2,] and Abdelmounam Sherik[3]*

[1]Materials and Corrosion Laboratory, Chemistry Department, Faculty of Science, Taif University, Saudi Arabia
[2]Electrochemistry Research Laboratory, Chemistry Department, Faculty of Education, Ain Shams Univ., Roxy, Cairo, Egypt
[3]Research and Development Center, Saudi Aramco, Dhahran, Saudi Arabia 31311
[*]E-mail: khaledrice2003@yahoo.com

In spite of the huge success that has been attributed to the use of computational chemistry in corrosion studies, most of the ongoing research on the inhibition potential of organic inhibitors is restricted to laboratory work. The quantitative structure inhibition (activity) relationship (QSAR) approach is an effective method that can be used together with experimental techniques to predict inhibitor candidates for corrosion processes. The study has demonstrated that the neural network can effectively generalize correct responses that only broadly resemble the data in the training set. The neural network can now be put to use with the actual data, this involves feeding the neural network with several quantum chemical descriptors as dipole moment, highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbital energy, energy gap, molecular area and volume. The neural network will produce almost instantaneous results of corrosion inhibition efficiency.

1. INTRODUCTION

Corrosion is one of the main problems in the oil and gas production and transportation industries. Oil field corrosion manifests itself in several forms, which include the "sweet corrosion" caused by carbon dioxide (CO_2) gas and/or the "sour corrosion" caused by hydrogen sulfide (H_2S), in various oil and gas operations [1].

Many experimental and theoretical investigations have been undertaken with the common goal of revealing inhibitive action of different series of chemical compounds [2-6]. Although experimental

studies are straightforward, they are often expensive and time-consuming. Alternatively, theoretical and computational chemistry are useful and a powerful means in choosing the appropriate and effective inhibitors by understanding the inhibition mechanism prior to experimentation.

It is a fundamental principle of chemistry that the structural formula of any compound contains coded within it all that compound's chemical, physical, and biological properties. Physical organic chemistry in the 21st century is believed to become increasingly oriented towards elucidating in detail how these properties are determined by the structure. This knowledge will help in developing the most appropriate experimental design. [7].

It is obvious that for many reasons, including the limitations of computer technology and the absence of a proper theoretical basis, that quantitative calculation of chemical and physical properties of chemical compounds from first principles will not be achievable in the near future. Therefore, the development of alternative approaches to find quantitative mathematical relationships between the intrinsic molecular structure and observable properties of chemical compounds will be of increasing importance in the chemistry of the 21st century [7].

Quantum chemical calculations have long been concerned with the correlation of chemical properties in terms of structure and the inhibition efficiency of corrosion inhibitors. Subsequently, most such work in the 20th century has been carried out with co-generic sets of compounds in which just one structural feature is changing at any one time [8-12]. Numerous linear free energy relationships starting from those of Hammett [13] thereby resulted and have given considerable insights into organic chemical mechanisms. It is believed that the quantitative structure-property relationship (QSPR) approach will become the tool of choice for many academic and industrial chemists [7].

Quantitative structure inhibition (activity) relationship (QSIR or QSAR) is a theoretical method, which is useful in relating structural based parameters to corrosion inhibition efficiencies [14-20]. In the development of QSAR models for corrosion inhibitors, attempts have been made to predict corrosion inhibition efficiency with a number of individual structural parameters (descriptors) obtained via various quantum chemical calculation methods as a tool for studying corrosion inhibition. The goal of these trials were to find possible correlations between corrosion inhibition efficiency and a number of quantum molecular properties such as dipole moment (μ), highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals, the gap between HOMO and LUMO, charge density, polarizability and molecular volume, as well as some structural parameters [12, 21-42].

The purpose of a QSIR is to highlight the relationship between inhibition efficiency (or any activity) and structural features (descriptors). In this method, it is to find one or more structural parameters, which relate these descriptors to their inhibition efficiencies through a mathematical equation.

The goal of this study is to present a predictive model for corrosion inhibition of steel by 28 amino acids' molecules and their related compounds using an artificial neural network. This work extends the study presented in a previous work [22] with applications of neural networks to electrochemical techniques. The proposed model obtains predictions of inhibition efficiencies based on several quantum chemical variables and comparing these predicted values with the experimental inhibition efficiencies reported for the amino acids in the literature.

2. COMPUTATIONAL METHOD

Geometrical parameters of all stationary points for the investigated amino acids and their derivatives are optimized by employing analytical energy gradients. The generalized gradient approximation (GGA) within the density functional theory was conducted with the software package DMol^3 in Materials Studio of Accelrys Inc. [43]. All calculations were performed using the Becke-Lee-Yang-Parr (BLYP) exchange correlation functional and the double numerical with polarization (DNP) basis set [44-46], since this was the best set available in DMol^3. A Fermi smearing of 0.005 hartree and a real space cutoff of 3.7 Å was chosen to improve the computational performance. All computations were performed with spin polarization.

The phenomenon of electrochemical corrosion takes place in the liquid phase, so it is relevant to include the effect of solvent in the computations. The Self-Consistent Reaction Field (SCRF) theory [47], with Tomasi's polarized continuum model (PCM) was used to perform the calculations in solution. These methods model the solvent as a continuum of uniform dielectric constant (ϵ=78.5) and define the cavity where the solute is placed as a uniform series of interlocking atomic spheres. Frontier orbital distribution was obtained, at the same basis set level, to analyze the reactivity of inhibitor molecules.

The molecular dynamics (MD) simulation of the interaction between the 28 compounds and Fe (1 1 1) plane surface was carried out in a simulation box (17.38 Å × 17.38 Å × 44.57 Å) with periodic boundary conditions to model a representative part of the interface devoid of any arbitrary boundary effects. The Fe (1 1 1) plane surface was first built and relaxed by minimizing its energy using molecular mechanics, then the surface area of Fe (1 1 1) was increased and its periodicity is changed by constructing a super cell, and then a vacuum slab with 15 Å thicknesses was built on the Fe (1 1 1) surface. The number of layers in the structure was chosen so that the depth of the surface is greater than the non-bond cutoff used in the calculation. Using four layers of iron atoms gives a sufficient depth that the amino acids and their derivatives will only be involved in non-bond interactions with iron atoms in the layers of the surface, without unreasonably increasing the calculation time. This structure then converted to have 3D periodicity [48, 49]. As 3D periodic boundary conditions are used, it is important that the size of the vacuum slab is great enough (15 Å) that the non-bond calculations for the adsorbate (amino acids and their related compounds) do not interact with the periodic image of the bottom layer of atoms in the surface. After minimizing the Fe (1 1 1) surface and the 28 inhibitor molecules, the corrosion system will be built by layer builder to place the amino acids and their related compounds on the Fe (1 1 1) surface. The adsorption of these molecules on the Fe (1 1 1) surface was simulated using the condensed phase optimized molecular potentials for atomistic simulation studies (COMPASS) force field. The adsorption of amino acids and their related compounds onto the Fe (1 1 1) surface provides access to the energy of the adsorption and its effects on the inhibition efficiencies of amino acid molecules [48, 49]. The binding energy between the studied molecules and Fe (111) surface were calculated using the following equation [50, 51]:

$$E_{binding}=E_{total}-(E_{surface}+E_{inhibitor}) \tag{1}$$

Where E_{total} is the total energy of the surface and inhibitor, $E_{surface}$ is the energy of the surface without the inhibitor, and $E_{inhibitor}$ is the energy of the inhibitor without the surface.

2.1 Artificial Neural Networks

Neural network analysis is an artificial intelligence (AI) approach to mathematical modeling. It is a sophisticated model building technique capable of modeling data such that it may be better represented by nonlinear functions. Corrosion is a complex and nonlinear phenomenon that is too complex to be described by analytical methods or empirical rules, which make it an ideal pheromone to be studied using artificial neural networks [22].

The neurons are grouped into discrete layers and interconnected according to a given design. As in nature, the network's function is determined largely by the connections between elements (neurons). Each connection between two neurons has a weight coefficient connected to it. The standard network structure for an approximation function is the multiple layer perception (or feed forward network). The feed forward network often has one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons [52-55]. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors. The linear output layer lets the network produce values outside the −1 to +1 range [54, 56]. For the network, the appropriate notation is used in two-layer networks [52-55] .

2.2 The structure of a neural network
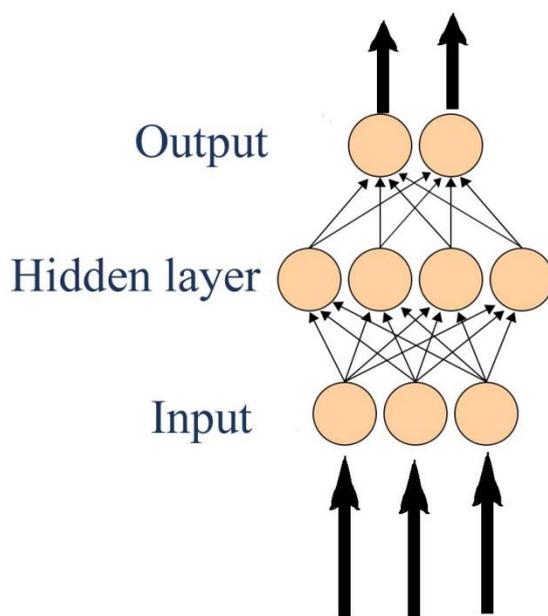
Figure 1. Structure of artificial neural network.

Figure 1 shows the structure of the artificial neural network. It consists of two main layers the input layer (predictor layer) and the output layer. The input layer is used to introduce the input (predictor) variables to the network. The output of the nodes in this layer represents the predictions made by the network for the response variables. This network also contains hidden layers. The optimal number of neurons in the hidden layer(s), n_s, is difficult to specify and depends on the type and complexity of the process or experimentation [1]. This number is usually iteratively determined. In our study we have only a single hidden layer with four nodes. Each node (other than those in the input layer) takes as its input a transformed linear combination of the outputs from the nodes in the layer below it. This input is then passed through a transfer function to calculate the output of the node. The transfer function used by QSAR is an s-shaped sigmoid function. This function is chosen because it is smooth and has easily differentiable features that help the algorithm that is used to train the network [46].

2.3 Training process and topology of the neural network

The neural network is built to train multilayer perceptrons in the context of regression analyses, i.e., to approximate functional relationships between covariates and response variables. Therefore, neural networks are used as extensions of generalized linear models [57].

Training is the process whereby the connection weights and biases are set so as to minimize the prediction error for the network. For a particular set of weights and biases, each of the training cases are introduced to the network and an error function is used to determine how well the calculated outputs match the expected output values [22, 58, 59]. A training algorithm is defined as a procedure that consists of adjusting the coefficients (weights and biases) of a network to minimize an error function (usually a quadratic one) between the network outputs for a given set of inputs and the correct (already known) outputs. If smooth, nonlinearities are used, the gradient of the error function can be computed by the classical back propagation procedure [1].

It was found empirically that a single hidden layer is sufficient for modeling most data sets and it is recommended that we first try to model our data with a single hidden layer. Additional hidden layers allow the neural network to model more complex functions [58]. There is a formal proof, the Kolmogorov theorem, which states that two hidden layers are theoretically sufficient to model any problem, though it is possible that, for some data sets, a network with more hidden layers might be able to find a good model more easily [46].

Each connection weight and node bias is a parameter that can be adjusted during network training. Therefore, each connection and node corresponds to one degree of freedom of the model represented by the neural network. As a general rule, we should strive to have at least twice as many observations as there are degrees of freedom. If we have too many nodes in the hidden layer(s), then the model will tend to over fit our data. If we have too few, then the model may not have sufficient power to fit our data [58].

3. INHIBITORS

Table 1. Inhibition efficiencies and molecular structures of the studied inhibitor series

	Inhibitor Name	Structure	Inhibition Efficiency [27, 58]
1	Glycine		50
2	Alanine		51
3	Valine		47
4	Leucine		63
5	Isoleucine		59
6	Threonine		59
7	Serine		63
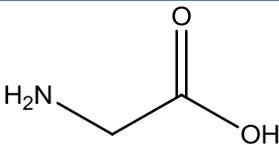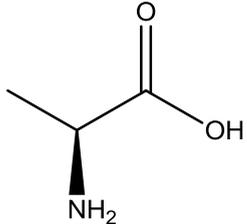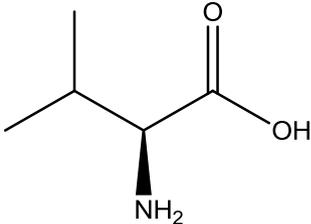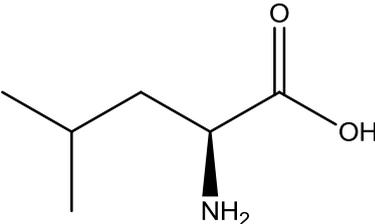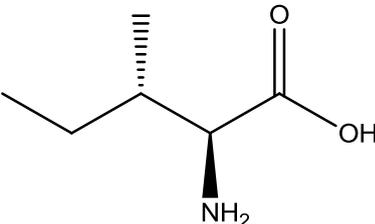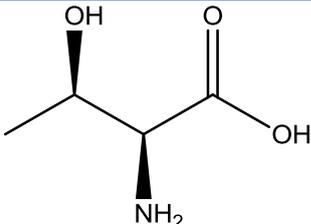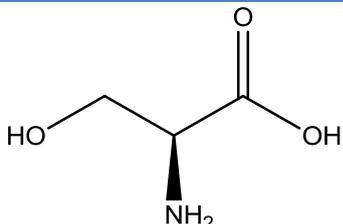
	Inhibitor Name	Structure	Inhibition Efficiency [27, 58]
8	Phenylalanine		0
9	Tyrosine		39
10	**-3,5** Diiodotyrosine		87
11	Tryptophane		80
12	Aspartic acid		52
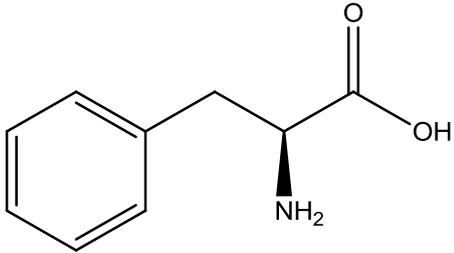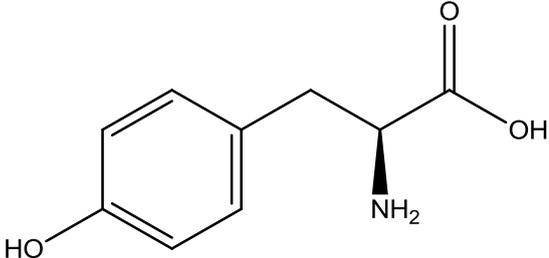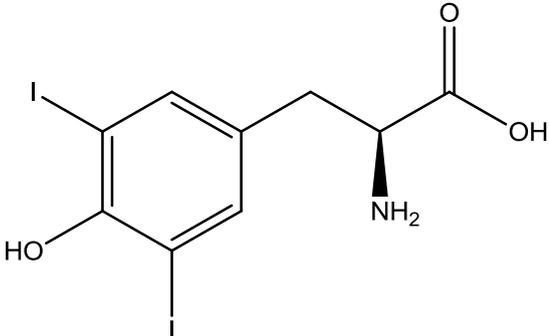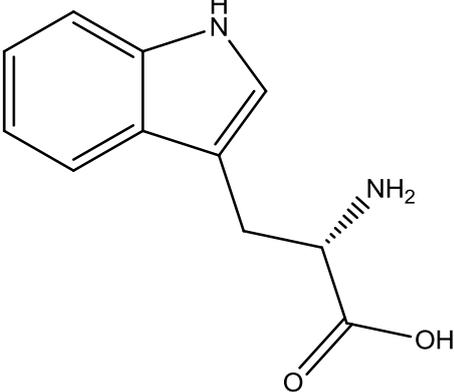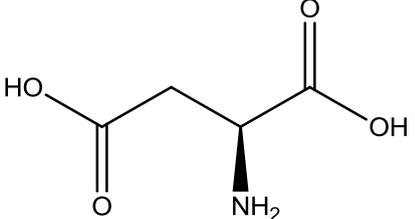
	Inhibitor Name	Structure	Inhibition Efficiency [27, 58]
13	Asparagine		73
14	Glutamic acid		53
15	Glutamine		75
16	Proline		34
17	Hydroxyproline		-140
18	Histamine		67
19	Creatinine		43
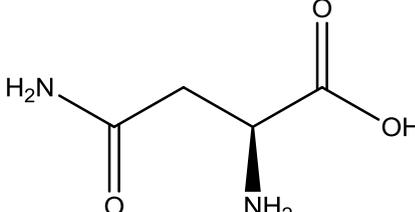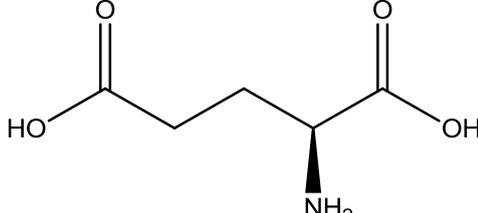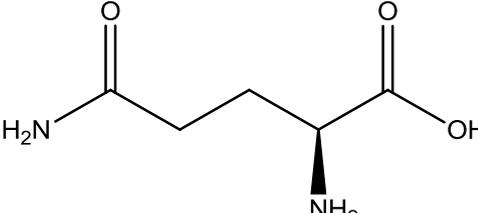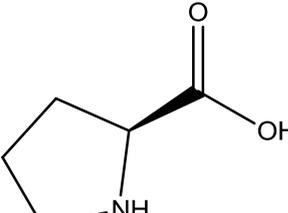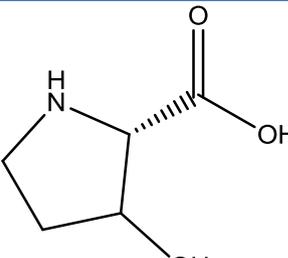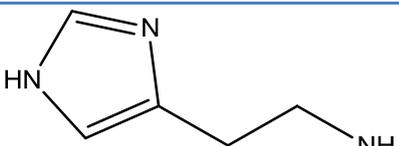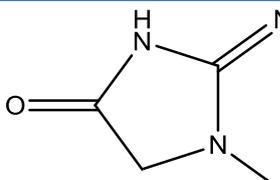
	Inhibitor Name	Structure	Inhibition Efficiency [27, 58]
20	4-nitropyrazole		77.4
21	4-Sulfopyrazole		75.1
22	Cystine		-55
23	Cysteine		-179
24	Methionine		59
25	Lysine		71
26	Creatine		-10
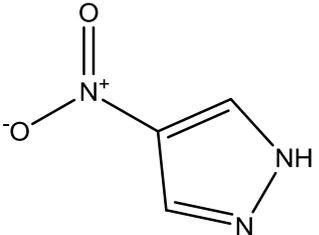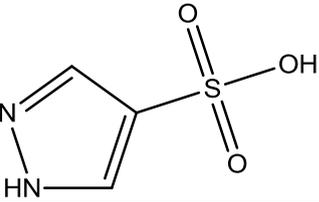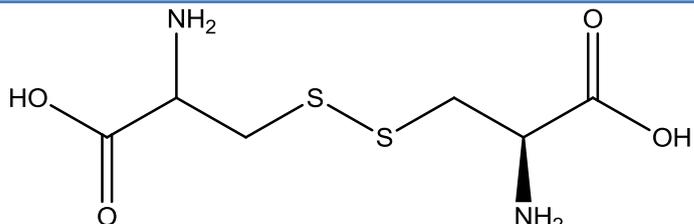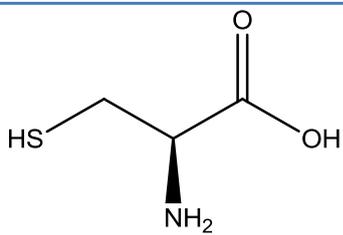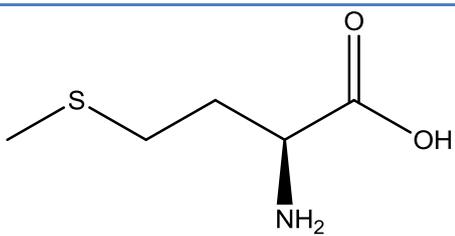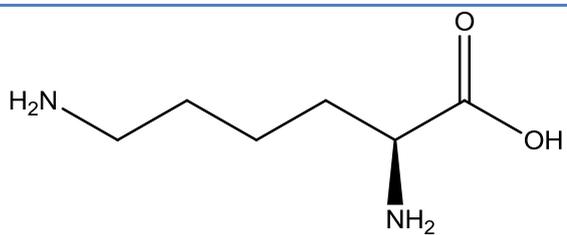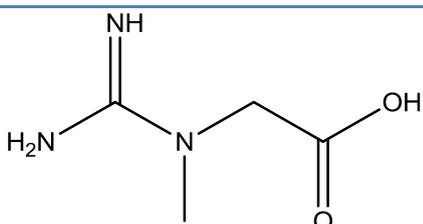
	Inhibitor Name	Structure	Inhibition Efficiency [27, 58]
27	Histidine		41
28	Arginine		16

Corrosion inhibition experimental studies for the 28 amino acids and their related compounds presented in Table 1 were conducted in our laboratory at Rice University, Houston, Texas [60, 61]. The experimental data was collected from Professor Hackerman's literature [60, 61], but the experimental details are outlined briefly here as indicated in the literature [60, 61]. Measurements were performed with a Gamry Instrument Potentiostat/Galvanostat/ZRA. This includes a Gamry Framework system based on the ESA400 and the VFP600 and Gamry applications that include DC105 corrosion and EIS300 electrochemical impedance spectroscopy measurements. A computer collected the data, and Echem Analyst 4.0 software was used for plotting, graphing, and fitting data. Tafel curves were obtained by changing the electrode potential automatically from -250 to +250 mV vs. open circuit potential (E_{oc}) at a scan rate of 1 mV/s. The inhibitor concentration was 10^{-2} M. Corrosion inhibition efficiency of the studied amino acids was measured in hydrochloric acid (1 M) solutions in the presence of the 28 amino acids and related compounds at 10^{-2} M concentration. The temperature of the solutions was maintained at 25 °C. The corrosion rate was determined by using the Tafel polarization method [60, 61].

4. RESULTS AND DISCUSSION

4.1 QSAR study using the artificial neural network

In this work, a neural network training set was used to predict the corrosion inhibition efficiencies for 28 amino acids and their related compounds used to inhibit the corrosion of steel in chloride solutions.

Before searching for potential QSAR, it is worth assessing the quality and distribution of data presented in the study table (Table 2). Most forms of multivariate analysis assume that the input variables have a normal distribution, and are a representative sample. To examine data in Table 2 a univariate analysis, which is a technique used for generating statistics independently for the

experimental inhibition efficiencies. Table 3 shows accepted normal distribution, which enables us to start building a correlation matrix. The normal distribution behavior of the studied data was confirmed by the values of standard deviation, mean absolute deviation, variance, skewness and kurtosis.

Table 3 shows a univariate analysis for the inhibition data presented in Table 2. Table 3 contains several statistical measures that describe the corrosion inhibition data. The most important parameters in Table 3 are the skewness and kurtosis. Skewness is the third moment of the distribution, which indicates the symmetry of the distribution. As the skewness is negative, the distribution of data values within the column is skewed toward negative values. For a symmetrical distribution, the skewness is zero. Kurtosis is the fourth moment of the distribution, which indicates the profile of the column of data relative to a normal distribution. Univariate analysis calculates Fisher kurtosis, which subtracts 3.0 from the definition above. For a normally distributed data set, it gives a value of 0.0. If the kurtosis is positive, the distribution of data in the column is more sharply peaked than a normal distribution. If the kurtosis is negative, the distribution is flatter than a normal distribution [27, 58].

A study table presented in Table 2 contains the calculated descriptors and properties for the studied 28 amino acids and their related compounds used to inhibit the corrosion of steel in chloride solutions. These compound's descriptors and properties are used for developing quantitative structure activity relationships and property prediction.

Table 2. Descriptors for the studied 28 inhibitor molecules calculated using quantum chemical and molecular dynamics, MDs simulation methods

Structure	Inhibition Efficiency [27, 58]	E(HOMO) (Ha)	E (LUMO)(Ha)	[E(LUMO)-E(LUMO)](Ha)	Binding Energy (Kcal/mol)	Adsorption Energy (Kcal/mol)	[Total Energy](Kcal/mol)	Total dipole (VAMP Electrostatics)	Dipole x (VAMP Electrostatics)	Dipole y (VAMP Electrostatics)	Dipole z (VAMP Electrostatics)	Molecular area (vdW area) (Spatial Descriptors)	Molecular volume (vdW volume) (Spatial Descriptors)	Neural Network Prediction for Inhibition Efficiency
glycine	50	-0.2211	-0.0655	0.1556	144.1	-43.85	27.16403	7.99	0.319	-2.017	1.161	107.4517	73.08615	49.95857
4-nitropyrazole	77.48	-0.1954	-0.0563	0.1391	254.2	-47.82	34.90011	3.12	6.93	-1.345	-0.006	132.2922	96.31805	76.82211
alanine	51	-0.2202	-0.0618	0.1584	148.67	-51.311	20.30506	7.91	0.214	-1.837	0.498	128.4906	89.83167	50.66061
serine	63	-0.2086	-0.0547	0.1539	193.2	-59.028	35.88107	6.43	-0.562	-1.283	2.187	138.6798	99.62781	62.37816
threonine	59	-0.212	-0.0528	0.1591	186.3	-62.847	22.80966	7.3	-1.586	-3.518	4.065	157.0477	115.7812	59.52329
proline	34	-0.2158	-0.054	0.1618	117.2	-63.87	10.4539	8.2	-1.006	-2.792	2.249	153.3667	114.0387	35.04029
valine	47	-0.2174	-0.0582	0.1591	141.12	-64.239	26.4814	6.99	-1.313	-2.341	3.074	161.9411	122.4088	45.34483
cysteine	-179	-0.2052	-0.5656	-0.3604	9.87	-64.89	30.29514	9.4	-0.408	-1.058	1.882	147.0621	108.3996	-178.84
creatinine	43	-0.2179	-0.0547	0.1632	137.67	-65.49	-68.3172	7.97	1.659	-1.444	0.388	148.3478	109.073	44.99693
histamine	67	-0.215	-0.0602	0.1548	203.12	-66.62	-10.7584	6.01	-3.433	-0.335	-1.444	157.6746	116.8251	65.3589
leucine	63	-0.2086	-0.0542	0.1545	194.56	-67.46	11.47275	6.32	-1.314	-3.273	0.487	185.1023	139.8994	64.09232
creatine	-10	-0.2134	-0.029	0.1844	29.6	-69.25	-23.7776	8.41	-0.084	0.729	1.092	171.6891	126.3599	-9.89665
isoleucine	59	-0.2088	-0.0526	0.1562	188.76	-71.47	25.71603	6.51	-1.174	-2.39	3.256	180.5596	138.8267	61.36808
hydroxyproline	-140	-0.2677	-0.0654	0.2023	15.8	-71.73	12.79202	9	-0.948	-1.224	-3.271	159.7079	122.4847	-140.043
asparagine	73	-0.2082	-0.0589	0.1493	241.23	-75.06	-46.3684	5.98	-1.389	2.138	2.141	165.4502	123.3463	72.93619
aspartic acid	52	-0.2063	-0.0516	0.1547	158.68	-75.149	-23.7776	7.84	1.739	0.304	0.809	165.3701	120.8097	51.30428

Structure	Inhibition Efficiency [27, 58]	E(HOMO) (Ha)	E (LUMO)(Ha)	[E(LUMO)-E(LUMO)](Ha)	Binding Energy (Kcal/mol)	Adsorption Energy (Kcal/mol)	[Total Energy](Kcal/mol)	Total dipole (VAMP Electrostatics)	Dipole x (VAMP Electrostatics)	Dipole y (VAMP Electrostatics)	Dipole z (VAMP Electrostatics)	Molecular area (vdW area) (Spatial Descriptors)	Molecular volume (vdW volume) (Spatial Descriptors)	Neural Network Prediction for Inhibition Efficiency
histidine	41	-0.22	-0.0532	0.1668	128.98	-75.33	21.99655	7.98	-2.211	-5.769	3.547	187.1373	146.3167	30.64679
phenylalanine	0	-0.2215	-0.0486	0.1729	50.3	-75.41	25.26654	8.32	0.785	-1.999	0.027	213.33	167.868	-24.7506
lysine	71	-0.2015	-0.0567	0.1447	221.56	-76.17	9.344533	5.99	-0.893	-0.892	1.053	207.6746	153.7672	71.00215
glutamic acid	53	-0.2124	-0.0545	0.1579	161.21	-77.05	-7.44533	7.863	0.044	-5.325	5.785	185.522	137.8673	51.45931
4-sulfopyrazole	75.14	-0.1943	-0.0572	0.1371	250.1	-77.78	-9.91027	4.34	-2.347	-0.504	-4.926	159.8533	116.9821	75.517
3,5-diiodotyrosine	87	-0.2639	-0.1377	0.1262	287.45	-77.86	27.23897	2.12	3.424	-0.528	1.419	280.1272	226.3173	87.24676
methionine	59	-0.216	-0.0602	0.1558	187.17	-77.968	14.87045	7.12	-1.013	-4.581	2.911	191.3474	142.7116	59.4929
glutamine	75	-0.2038	-0.0609	0.1428	249.7	-78.91	-27.3518	5.43	-1.093	-7.204	3.419	189.6844	140.8272	73.47794
tyrosine	39	-0.2151	-0.0527	0.1623	125.31	-82.29	9.303199	8.01	1.476	-0.579	0.075	225.5877	177.6149	38.37076
tryptophane	80	-0.2205	-0.0955	0.125	261.98	-82.33	169.8545	2.43	1.331	3.922	-0.975	245.2549	199.7368	80.50289
Arginine	16	-0.2147	-0.052	0.1627	88.9	-85.64	-111.101	8.3	0.568	-6.462	2.554	235.28	174.4835	15.61438
cystine	-55	-0.2065	-0.0174	0.1891	29.11	-92.84	32.15046	8.65	-1.57	-0.603	2.696	271.1958	205.521	-54.9151

Table 2 presents the quantum chemical descriptors as well as the molecular dynamics , MD's calculations that will be used to build a correlation matrix. These descriptors include the HOMOs, LUMOs, total dipole moment, binding energy and adsorption energy.

Table 3. Univariate analysis of the experimental inhibition efficiencies data

	B: Inhibition Efficiency
Number of Sample Points	28
Range	266
Maximum	87
Minimum	-179
Mean	33.95071
Median	52.5
Variance	3.80E+03
Standard Deviation	62.7343
Mean Absolute Deviation	40.836
Skewness	-2.17234
Kurtosis	4.07874

Table 4. Correlation matrix of the studied variables

	B: Inhibition Efficiency [20, 21]	C: E(HOMO) (Ha)	D: E (LUMO)(Ha)	E: [E(LUMO)-E(LUMO)](Ha)	F: Binding Energy (Kcal/mol)	G: Adsorption Energy (Kcal/mol)	H: [Total Energy](Kcal/mol)	I: Total dipole (VAMP Electrostatics)	M: Molecular area (vdW area) (Spatial Descriptors)	N: Molecular volume (vdW volume) (Spatial Descriptors)
B: Inhibition Efficiency	1	0.260799	0.594229	0.544128	0.837994	0.080437	0.015057	-0.62235	0.034239	0.037099
C: E(HOMO) (Ha)	0.260799	1	-0.00351	-0.16446	0.173008	0.054463	-0.12415	-0.00149	-0.24957	-0.29717
D: E (LUMO)(Ha)	0.594229	-0.00351	1	0.986954	0.245998	-0.1067	-0.15711	-0.12138	0.101447	0.084384
E: [E(LUMO)-E(LUMO)](Ha)	0.544128	-0.16446	0.986954	1	0.214764	-0.11397	-0.13499	-0.11949	0.140207	0.13104
F: Binding Energy (Kcal/mol)	0.837994	0.173008	0.245998	0.214764	1	0.066709	0.177904	-0.86751	0.041178	0.058703
G: Adsorption Energy (Kcal/mol)	0.080437	0.054463	-0.1067	-0.11397	0.066709	1	0.103563	0.023798	-0.83649	-0.81538
H: [Total Energy](Kcal/mol)	1.51E-02	-0.12415	-0.15711	-0.13499	0.177904	0.103563	1	-0.38487	0.137823	0.198367
I: Total dipole (VAMP Electrostatics)	-0.62235	-0.00149	-0.12138	-0.11949	-0.86751	0.023798	-0.38487	1	-0.24984	-0.28421
M: Molecular area (vdW area) (Spatial Descriptors)	0.034239	-0.24957	0.101447	0.140207	0.041178	-0.83649	0.137823	-0.24984	1	0.994539
N: Molecular volume (vdW volume) (Spatial Descriptors)	0.037099	-0.29717	0.084384	0.13104	0.058703	-0.81538	0.198367	-0.28421	0.994539	1

Table 4 shows a correlation matrix that illustrates all possible pairwise correlation coefficients for a set of variables. It can be helpful to identify highly correlated pairs of variables, and therefore identify redundancy in the data set [22]. Each cell of the matrix corresponds to the correlation between two columns of study table data (Table 2). The correlation coefficients lie between -1.0 and +1.0. A value approaching +1.0 indicates that the two columns are highly correlated and a value approaching -1.0 also indicates a high degree of correlation, except that the data changes values in opposite directions. A correlation coefficient close to 0.0 indicates very little correlation between the two columns. The diagonal of the matrix always has the value of 1.0. To aid in visualizing the results, the cells in the correlation matrix grid are colored according to the correlation value in each cell. A standard color scheme is used when the correlation matrix is generated: $+0.9 \leq X \leq +1.0$ (orange), $+0.7 \leq X < +0.9$ (yellow), $-0.7 < x > +0.7$ (white), $-0.9 < x > -0.7$ (yellow) and $-1.0 \leq x \leq -0.9$ (orange) [22, 58].

Now it is ready to perform a regression analysis of the descriptor variables presented in Table 4 compared against the measured corrosion inhibition values [60, 61]. There are many more descriptor variables than measured inhibition values, so we should reduce the number of descriptors. Typically, a ratio between two and five measured values for every descriptor should be sought to prevent over fitting. Tables 5 and 6 show a summary of the input data for the neural network training set and the cross validation of the descriptor values used in the neural network analysis. Table 5 shows that all the 28 corrosion inhibitor's data was used in building the QSAR model and these data are validated against the descriptor values calculated from quantum chemical and the MDs calculations [22, 58].

The cross validation data for the neural network model (Table 6) operates by repeating the calculation several times using subset of the original data to obtain a prediction model and then comparing the predicted values with the actual values for the omitted data.

Table 5. Summary of input data for neural network training

Summary of input data for Neural Network Training	
Number of rows requested	28
Number of rows used	28
Number of rows omitted due to invalid row description	0
Number of rows omitted due to invalid data	0
Number of columns requested	13
Number of columns used	13
Number of columns omitted due to invalid column description	0
Number of columns omitted due to invalid data	0
Number of cells omitted due to invalid data	0
Number of cells with missing data	0

Table 6. Cross validation of the input data for neural network training

r^2	0.999761
r^2(CV)	0.289594
Residual Sum of Squares	0.00645
Predictive Sum of Squares	19.181

The key measure of the predictive power of the model is the correlation coefficient r^2. The closer the value is to 1.0 the better the predictive power. For a good model, the r^2 value should be fairly close to 1.0. The correlation coefficient r^2 for this study is equal to 0.9997, which is reasonably high and indicates the predictive power of the model [22, 58]. A closer look at the last column in Table 2 and comparing the predicted values with the experimental values proves the efficiency of the suggested model.

Investigation of the neural network analysis in the QSAR study shows that the network has too many degrees of freedom (usually the number of network connections between nodes) for the number of observations (rows of data) for which the network is being trained. In this study there is one hidden layer with three nodes [22, 58].

Applying the neural network prediction model generates a model containing predictions corresponding to each output of the neural network. The neural network model adds a new column containing a calculation of the model to the study table (Table 2). Also, residual values of the predictions correspond to each output of the neural network.
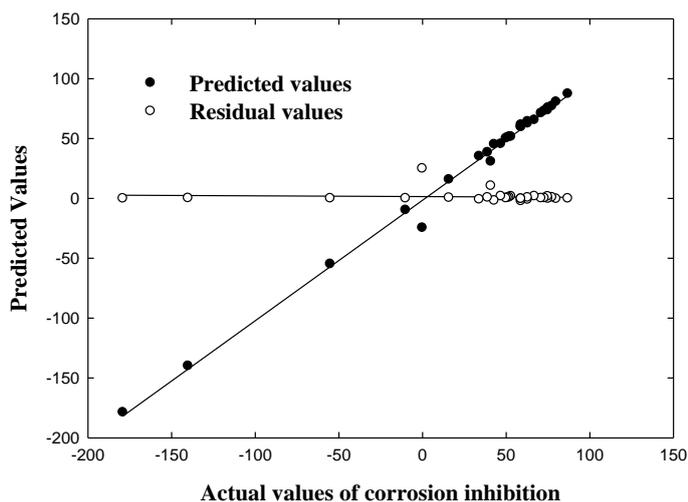
Figure 2. Plot of predicted inhibition and residuals vs. measured corrosion inhibition [20, 21] using NNA.
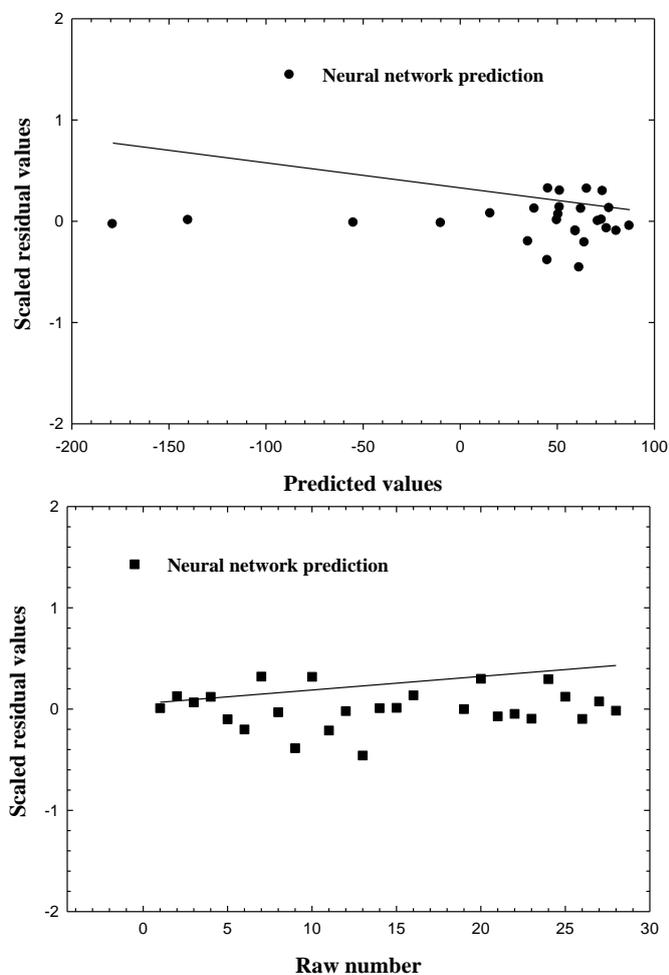
Figure 3. Outlier analysis for inhibition efficiency.

Figure 2 shows a relation between the predicted values, residual values and the experimental data in Table 2. A residual can be defined as the difference between the predicted value in the generated model and the measured value for corrosion inhibition. To test the constructed QSAR model, potential outliers have been identified in Figure 3. An outlier can be defined as a data point whose residual value is not within cross validated r^2 values, is also high, even though the regression is significant according to the F-test.

Figure 3 contains two charts. One contains the residual values plotted against the corrosion inhibition measurements and the other displays the residual values plotted against Table 3 row number. Each chart contains a dotted line that indicates the critical threshold of two standard deviations beyond which a value may be considered to an outlier. Inspection of Figure 4 shows that there are no points outside the dotted lines, which make the QSAR model acceptable.

5. CONCLUSIONS

In spite of the huge success that has been attributed to the use of computational chemistry in corrosion studies, most of the ongoing research on the inhibition potentials of organic inhibitors is restricted to laboratory work. The QSAR approach is still an effective method that can be used together with the experimental techniques to predict inhibitor candidates for inhibiting corrosions. The study has demonstrated that the neural network can effectively generalize correct responses that only broadly resemble the data in the training set. The neural network can now be put to use with the actual data. This involves feeding the neural network the values for the Hammett constants, dipole moment, HOMO energy, LUMO energy, energy gap, molecular area and volume. The neural network will produce almost instantaneous results of corrosion inhibitor efficiency. The predictions should be reliable, provided the input values are within the range used in the training set.

References

1. D. Colorado-Garrido, D.M. Ortega-Toledo, J.A. Hernández, J.G. González-Rodríguez, J. Uruchurtu, *J. Solid State Electrochem.*, 13 (2009) 1715-1722.
2. Z. Zhang, S. Chen, Y. Li, S. Li, L. Wang, *Corros. Sci.*, 51 (2009) 291-300.
3. E. Jamalizadeh, A.H. Jafari, S.M.A. Hosseini, *Journal of Molecular Structure*: THEOCHEM, 870 (2008) 23-30.
4. Y. Yan, W. Li, L. Cai, B. Hou, *Electrochim. Acta*, 53 (2008) 5953-5960.
5. H. Otmacic Curkovic, E. Stupnisek-Lisac, H. Takenouti, *Corros. Sci.*, 51 (2009) 2342-2348.
6. J. Aljourani, K. Raeissi, M.A. Golozar, *Corros. Sci.*, 51 (2009) 1836-1843.
7. A.R. Katritzkya, M. Karelsonb, V.S. Lobanova, *Pure & Appl. Chem.*, 69 (1997) 245-248.
8. K.F. Khaled, *Electrochim. Acta*, 48 (2003) 2493-2503.
9. K. Babić-Samardžija, K.F. Khaled, N. Hackerman, *Appl. Surf. Sci.*, 240 (2005) 327-340.
10. K.F. Khaled, K. Babić-Samardžija, N. Hackerman, *J. Appl. Electrochem.*, 34 (2004) 697-704.
11. K.F. Khaled, N. Hackerman, *Mater. Chem. Phys.*, 82 (2003) 949-960.

12. K.F. Khaled, N. Hackerman, *Electrochim. Acta*, 48 (2003) 2715-2723.
13. A.S. Fouda, H.A. Mostafa, M.N. Moussa, *Portugaliae Electrochimica Acta*, 23 (2005) 275-287.
14. K. Khaled, N. Abdel-Shafi, *Int. J. Electrochem. Sci*, 6 (2011) 4077-4094.
15. K.F. Khaled, *J. Appl. Electrochem.*, 41 (2011) 423-433.
16. E.S.H. El Ashry, S.A. Senior, *Corros. Sci.*, 53 (2011) 1025-1034.
17. S. Deng, X. Li, *Corros. Sci.*, 55 (2012) 407-415.
18. S.G. Zhang, W. Lei, M.Z. Xia, F.Y. Wang, *Journal of Molecular Structure: THEOCHEM,* 732 (2005) 173-182.
19. G.C. Zhang, T. Ma, J.J. Ge, N. Qi, Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/*Journal of Xi'an Shiyou University, Natural Sciences Edition,* 20 (2005) 55-57+76.
20. L. Niu, H. Zhang, F. Wei, S. Wu, X. Cao, P. Liu, *Appl. Surf. Sci.*, 252 (2005) 1634-1642.
21. E.E. Ebenso, M.M. Kabanda, T. Arslan, M. Saracoglu, F. Kandemirli, L.C. Murulana, A.K. Singh, S.K. Shukla, B. Hammouti, K. Khaled, *Int. J. Electrochem. Sci,* 7 (2012) 5643-5676.
22. K. Khaled, N. Al-Mobarak, *Int. J. Electrochem. Sc*i, 7 (2012) 1045-1059.
23. K. Khaled, N. Abdel-Shafi, N. Al-Mobarak, *Int. J. Electrochem. Sci*, 7 (2012) 1027-1044.
24. S.M.A. Hosseini, M. Salari, E. Jamalizadeh, A.H. Jafari, *Corrosion,* 68 (2012) 600-609.
25. K.F. Khaled, N.S. Abdel-Shafi, N.A. Al-Mobarak, *Int. J. Electrochem. Sci*, 7 (2012) 1027-1044.
26. K.F. Khaled, N.S. Abdelshafi, A.A. Elmaghraby, A. Aouniti, N.A. Almobarak, B. Hammouti, *Int. J. Electrochem. Sci*, 7 (2012) in press.
27. K. Khaled, *Corros. Sci.*, (2011).
28. K. Khaled, M.N.H. Hamed, K. Abdel-Azim, N. Abdelshafi, *J. Solid State Electrochem.*, 15 (2011) 663-673.
29. K.F. Khaled, *Mater. Chem. Phys.*, 130 (2011) 1394-1395.
30. K.F. Khaled, *J. Appl. Electrochem.*, 41 (2011) 277-287.
31. K.F. Khaled, A. El-Maghraby, *J. Mater. Envron. Sci.*, 4 (2013) 193-198.
32. K.F. Khaled, *Mater. Chem. Phys.*, 112 (2008) 104-111.
33. K.F. Khaled, *Appl. Surf. Sci.*, 255 (2008) 1811-1818.
34. K.F. Khaled, *Electrochim. Acta*, 53 (2008) 3484-3492.
35. K.F. Khaled, *Int. J. Electrochem. Sci*, 3 (2008) 462-475.
36. K.F. Khaled, K. Babić-Samardžija, N. Hackerman, *Corros. Sci.*, 48 (2006) 3014-3034.
37. K.F. Khaled, *Appl. Surf. Sci.*, 252 (2006) 4120-4128.
38. K. Khaled, K. Babic-Samardzija, N. Hackerman, *Electrochim. Acta*, 50 (2005) 2515-2520.
39. K.F. Khaled, K. Babic-Samardzija, N. Hackerman, *Electrochim. Acta*, 50 (2005) 2515-2520.
40. K. Babić-Samardžija, K.F. Khaled, N. Hackerman, *Appl. Surf. Sci.*, 240 (2005) 327-340.
41. K. Khaled, N. Hackerman, *Electrochim. Acta*, 48 (2003) 2715-2723.
42. K. Khaled, N. Hackerman, *Mater. Chem. Phys.*, 82 (2003) 949-960.
43. J. Zhang, G. Qiao, S. Hu, Y. Yan, Z. Ren, L. Yu, *Corros. Sci.*, 53 (2011) 147-152.
44. J.R. Mohallem, T.O. De Coura, L.G. Diniz, G. De Castro, D. Assafrão, T. Heine, *J. Phys. Chem.* A, 112 (2008) 8896-8901.
45. J.A. Ciezak, J.B. Leão, *The Journal of Physical Chemistry* A, 110 (2006) 3759-3769.
46. J.A. Ciezak, S.F. Trevino, *J. Phys. Chem.* A, 110 (2006) 5149-5155.
47. M.W. Wong, M.J. Frisch, K.B. Wiberg, *J. Am. Chem. Soc.*, 113 (1991) 4776-4782.
48. K. Khaled, *Appl. Surf. Sci.*, 256 (2010) 6753-6763.
49. K.F. Khaled, *J. Solid State Electrochem.*, 13 (2009) 1743-1756.
50. K. Khaled, *Electrochim. Acta*, 53 (2008) 3484-3492.
51. K. Khaled, *Appl. Surf. Sci.*, 255 (2008) 1811-1818.
52. G.N. Vanderplaats, Numerical optimisation techniques for engineering design, McGraw-Hill, New York, 1984.
53. D. Colorado-Garrido, S. Serna, M. Cruz-Chávez, J.A. Hernández, B. Campillo, 2010.

54. D. Colorado-Garrido, D.M. Ortega-Toledo, J.A. Hernandez, J.G. Gonzalez-Rodriguez, J. Uruchurtu, *J. Solid State Electrochem.*, 13 (2009) 1715-1722.
55. D. Colorado-Garrido, D.M. Ortega-Toledo, J.A. Hernández, J.G. González-Rodríguez, Proc. CERMA '07 Proceedings of the Electronics, Robotics and Automotive Mechanics Conference, IEEE Computer Society Washington, DC, USA, 2007.
56. K. Hornik, M. Stinchcombe, H. White, Neural Networks, 2 (1989) 359.
57. F. Günther, S. Fritsch, *The R Journal* 2 (2010) 30-38.
58. Accelrys Materials Studio 6.0 Manual, (2011).
59. E.B. De Melo, *Sci. Pharm.*, 80 (2012) 265-281.
60. V. Hluchan, L. Wheeler, N. Hackerman, *Werkstoffe und Korrosion*, 39 (1988) 512-517.
61. K. Babić-Samardžija, C. Lupu, N. Hackerman, A.R. Barron, A. Luttge, *Langmuir,* 21 (2003) 12187-12196.