

Quantitative Structure and Activity Relationship Modeling Study of Corrosion Inhibitors: Genetic Function Approximation and Molecular Dynamics Simulation Methods

K.F.Khaled^{1,2,*} and N. S. Abdel-Shafir¹

¹ Electrochemistry Research Lab., Chemistry Department, Faculty of Education. Ain Shams University, Roxy, Cairo, Egypt

² Materials and Corrosion Lab., Chemistry Department, Faculty of Science. Taif University, Kingdom of Saudi Arabia

*E-mail: khaledrice2003@yahoo.com

Received: 2 July 2011 / Accepted: 15 August 2011 / Published: 1 September 2011

QSAR studies on the inhibition corrosion efficiencies of twenty three organic compounds on steel surface in hydrochloric acid solutions using several physicochemical descriptors and investigation of the adsorption of these compounds on the steel surface by Monte Carlo simulation method were studied. Topological indices as well as several structural descriptors are used in the development of quantitative structure-activity relationships (QSARs) using genetic function approximation statistical analysis method. From our studies it is clear that quantum descriptors are a better choice when predictivity is the main issue. Among the descriptors with major contribution we should point out that lowest unoccupied molecular orbital energy (E_{LUMO}) and molecular volume are important predictive descriptors. Computational studies have been used to find the most stable adsorption sites for tributylamine inhibitor on steel surface.

Keywords: Corrosion inhibitor, Genetic Function Approximation, QSAR , Monte Carlo

1. INTRODUCTION

Despite the intense empirical searches for new commercial inhibitors, few articles address chemometric analysis of the inhibition corrosion efficiencies. Such a procedure represents a challenge to the application of regular structure-activity chemometric thinking applied in biological fields, since the physical adsorption is nonspecific, in opposition to the key-lock mechanism found in molecular biology [1]. Although under such circumstance we should expect lower statistical correlations than those found on biological studies, early corrosion studies, on the contrary, showed several successful results [2-6] correlating small number of inhibitors and quantum descriptors.

Early attempts were made in the mid-fifties employing Huckel calculations. For a large number of molecules, Bergman [7] obtained excellent correlations between standard reduction potentials with the LUMO and HOMO energies. During the 1960s, Donahue and Vosta [8,9] employed an ab initio calculations to establish certain correlations. Vosta [9] and collaborators considered the correlation of eight gamma-substituted pyridine N-oxides with several ab initio quantum descriptors. Growcock et al [10], studied a general multivariate analysis for chemisorption and corrosion inhibition. They used such physicochemical descriptors as HOMO and LUMO energies, logP, Hammett and Tafel constants, in an investigation of the inhibition of corrosion of mild steel by derivatives of cinnamaldehyde. This was the first work to recognize the importance of the Langmuir constant for obtaining the best linear relationships. Using the CNDO/2 methodology, Abdul-Ahad [11] extended this work to aniline derivatives. Dupin et al [12], carried out an important study with a large set of corrosion inhibitors. The corrosion inhibition by forty-two compounds, including aliphatic amines, imidazolines and related compounds was correlated with some Hansch and Free-Willson parameters. In this study, many non-linear descriptors were tested. Sastri et al [13], in a set of univariate experiments, correlated the inhibition efficiencies of several methyl substituted pyridines and substituted ethane derivatives with MNDO descriptors. In the mid- 1990s, Kutej et al [14], studied dibenzyl sulfoxide adsorption on iron using ab initio calculations to identify the attachment points of the corrosion inhibitors on the iron surface. Öğretir et al [15] employed several semi-empirical descriptors in attempts to correlate the efficiency of pyridine-based inhibitors for mild steel. Several descriptors showed excellent univariate correlations. However, Sastri et al [13], did not use multivariate methods. A related article [16], was concerned with the corrosion inhibition of imidazole derivatives for iron exposed to acidic media.

Lukovits et al [2], used a polynomial regression analysis for the Langmuir adsorption constant for seven thiourea derivatives and obtained acceptable correlation values. Bentiss et al [6], correlated inhibition corrosion efficiencies, determined through charge transfer resistance, of triazole and oxadiazole derivatives, with AM1 quantum descriptors: correlation coefficient values of 0.91-0.96 were obtained. Khalil [17] extended this study and correlated the inhibition by several thiosemicarbazone and thiosemicarbazide derivatives with five quantum MNDO/PM3 descriptors. All these previous studies were conducted on carbon steel.

This work is part of our continuous effort [18-22] aiming the efficient prediction of corrosion inhibition efficiencies from the molecular properties based on the QSAR (Quantitative Structure and Activity Relationship) methodology. Although it is possible to recognize, in the literature, many articles and authors employing common-sense descriptors (HOMO/LUMO energies, the energy gap, the dipole moment, polarizability and others), it is clear that remains a lack of studies searching for efficient quantum and group contribution molecular descriptors for general use in inhibition corrosion prediction. Once defined, these variables could be used to calculate corrosion efficiencies as well as being useful in a general methodology to generate molecular random structures searching for new molecular structures with an optimum inhibition efficiency values. This particular condition requires an intense effort towards finding molecular descriptors useful for predicting correct inhibition efficiencies in cross-validation calculations. These “universal” variables should be very important in

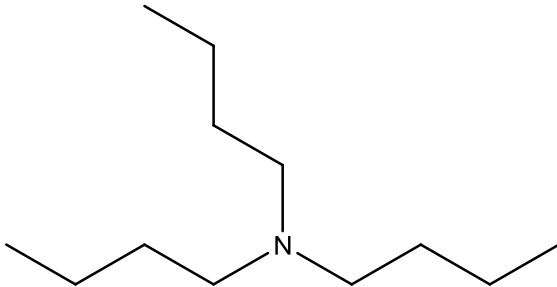
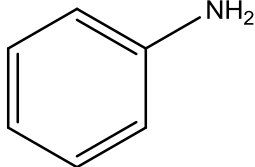
the search for new structures as it is now for the recognition of physical processes occurring at the corrosion metal/solution interface [1].

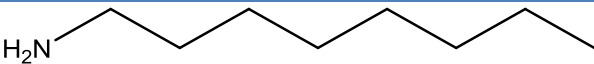
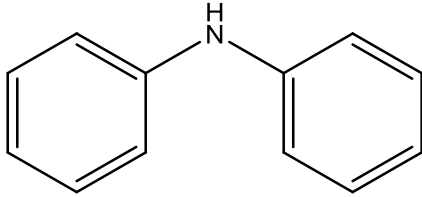
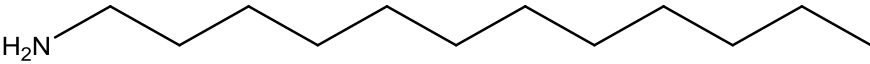
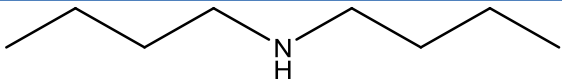
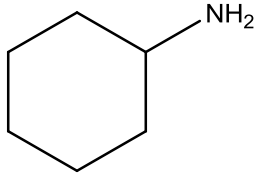
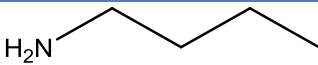
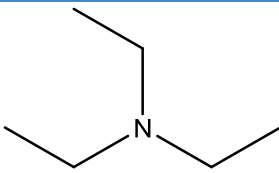
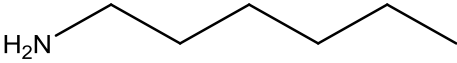
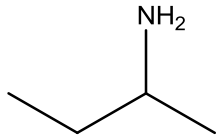
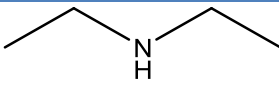
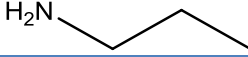
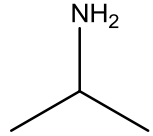
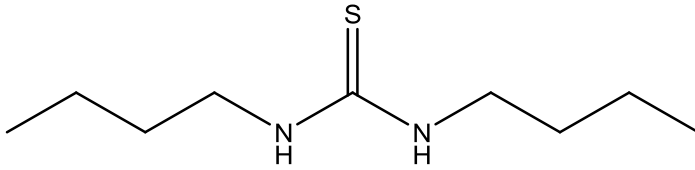
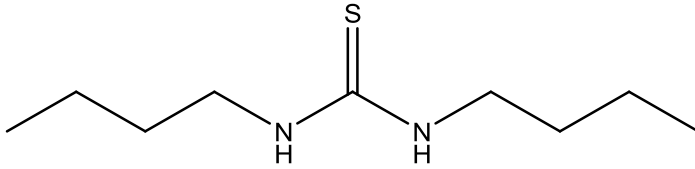
In the present work we carried out detailed theoretical investigations of 23 different corrosion inhibitors including amines, thiourea derivatives, and acetylenic alcohols to estimate its corrosion inhibition efficiencies on 22% Cr stainless steel in hydrochloric acid (15% w/v) solutions at 60 °C [1] using genetic function approximation (GFA) method. The systematic obtaining of these data, for such a large set of molecules, offered the unique opportunity for searching possible correlations between the inhibitors efficiency and molecular properties. It is also the aim of this study to theoretically investigate the adsorption density of one of the studied inhibitors on steel surface at 60 °C in the presence of aqueous solution.

2. EXPERIMENTAL

All corrosion inhibition data were obtained from the literature [1]. The experimental details are outlined briefly here as indicated in reference [1]. Weight loss experiments based on rectangular steel specimens with $2 \times 0.5 \times 0.5$ cm dimensions and a central hole. The experiments were carried out in cylindrical autoclaves internally coated with teflon. The autoclaves were placed in a rolling oven at 60 °C for 3 h. All solutions employed 300 ml of HCl (15% w/v), 2% w/v of the chemical inhibitor and 0.6% w/v of formaldehyde. The experimental conditions were designed to avoid complete dissolution of the metal plates and to strictly adhere to industrial recommendations, by which no more than 2% w/v of active components are allowed for matrix acidification operations [1]. Formaldehyde was employed to minimize hydrogen penetration. These conditions strictly followed those previously reported. The steel specimens were cleaned with acetone, washed with water, dried and weighed with a 0.0001 g precision. Two results were averaged for each inhibitor and presented in Table 1[1].

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

Inhibitor name	Structure	Inhibition Efficiency[1]
1 Tributylamine		97.76
2 Aniline		97.76

Inhibitor name	Structure	Inhibition Efficiency[1]
3 n-Octylamine		88.62
4 Diphenylamine		92.08
5 Dodecylamine		88.41
6 di-n-Butylamine		86.41
7 Cyclohexylamine		86.79
8 n-butylamine		78.19
9 Triethylamine		69.12
10 Hexylamine		74.67
11 sec-Butylamine		83.14
12 Diethylamine		67.54
13 Propylamine		68.18
14 Isopropylamine		63.61
15 1,3-Dibutyl-2-thiourea		97.30
16 1,3-Diethyl-2-thiourea		96.36

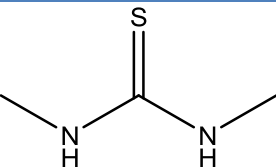
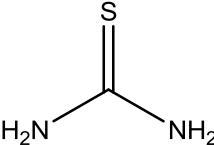
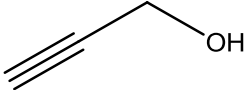
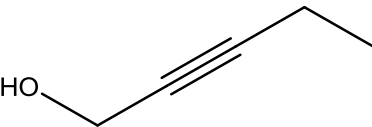
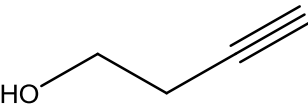
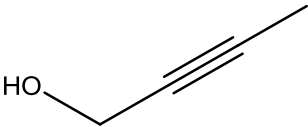
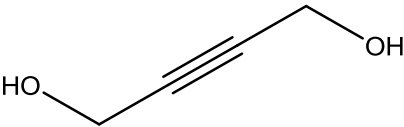
Inhibitor name	Structure	Inhibition Efficiency[1]
17 1,3-Dimethyl-2-thiourea		95.96
18 Thiourea		90.04
19 Propargyl alc.		95.80
20 2-Pentyn-1-ol		87.67
21 3-Butyn-1-ol		89.63
22 2-Butyn-1-ol		74.07
23 2-Butyn-1,4-diol		68.59

Table 1 lists the twenty three inhibitors employed in the study and their experimental inhibition efficiencies for stainless steel [1]. Tributylamine, aniline and thiourea derivatives as 3-dibutylthiourea, 1,3 diethylthiourea and 1,3-dimethylthiourea are among the most efficient inhibitors are followed by propargylic alcohol, diphenylamine, thiourea and some amines. On the other hand, the aliphatic amines, isopropylamine, secbutylamine, propylamine, diethylamine and nbutylamine, are among the less efficient inhibitors tested for stainless steel [1].

3. COMPUTATIONAL AND STATISTICAL DETAILS

Quantum mechanical calculations are conducted using the spin polarized density functional theory (DFT) within the generalized gradient approximation (GGA) using the Perdew–Wang exchange-correlation functional (PW91) [23] as implemented in the DMOI³ package [24]. This method can conduct an accurate and efficient self-consistent calculations using a convergent three-dimensional numerical integration scheme. DFT methods account for exchange-correlation in many electron

systems by the GGA, and they are suitable for studying transition metal clusters [25]. The double numerical basis set including *d*-polarization function (DND) [24] is chosen in this work.

DMol³ is a density functional theory (DFT) quantum mechanical code that enables users to study problems in different environments include, gas phase, solvent, surface, and solid. Owing to its unique approach to electrostatics, DMol³ has long been one of the fastest methods for molecular DFT calculations and can rapidly perform structure optimizations of molecular systems using delocalized internal coordinates. DMol³ can also be used to search very efficiently for transition states using a combination of LST/QST algorithms with conjugate gradient refinement, thereby avoiding the computationally expensive calculation of the Hessian matrix [24].

The genetic function approximation (GFA) algorithm offers a new approach to the problem of building quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) models. Replacing regression analysis with the GFA algorithm enables the construction of models competitive with or superior to those produced by standard techniques and makes available additional information not provided by other techniques. Unlike most other analysis algorithms, GFA gives multiple models, where the populations of the models are created by evolving random initial models using a genetic algorithm. GFA can build models using not only linear polynomials but also higher-order polynomials, splines, and other nonlinear functions.

The genetic function approximation algorithm was initially conceived by taking inspiration from two seemingly disparate algorithms: Holland's genetic algorithm (1975) and Friedman's (1990) multivariate adaptive regression splines (MARS) algorithm [26].

Friedman's MARS algorithm is a statistical technique for modeling data. It provides an error measure, called the lack-of-fit (LOF) score that automatically penalizes models with too many features. It also inspired the use of splines as a powerful tool for nonlinear modeling [26].

The GFA algorithm uses a genetic algorithm to perform a search over the space of possible QSAR/QSPR models using the LOF score to estimate the fitness of each model. Such evolution of a population of randomly constructed models leads to the discovery of highly predictive QSARs/QSPRs [26].

The GFA algorithm approach has a number of important advantages over other techniques. It builds multiple models rather than a single model. It automatically selects which features are to be used in the models. It is better at discovering combinations of features that take advantage of correlations between multiple features. GFA incorporates Friedman's LOF error measure, which estimates the most appropriate number of features, resists overfitting, and allows control over the smoothness of fit. Also, it can use a larger variety of equation term types in construction of its models and finally, it provides, through study of the evolving models, additional information not available from standard regression analysis [26].

In the current study, the studied inhibitors have been simulated as adsorbate on iron surface (111) substrate to find the low energy adsorption sites on the iron surface and to investigate the preferential adsorption of the studied inhibitors. To calculate the adsorption density as well as the binding energy of the studied inhibitors, Monte Carlo method has been used. In this computational work, possible adsorption configurations have been identified by carrying out Monte Carlo searches of the configurational space of the steel/inhibitor system as the temperature is slowly decreased. The

adsorbates were the studied inhibitors constructed and their energy was optimized using Forcite classical simulation engine [27]. The geometry optimization process is carried out using an iterative process, in which the atomic coordinates are adjusted until the total energy of a structure is minimized, i.e., it corresponds to a local minimum in the potential energy surface. Geometry optimization is based on reducing the magnitude of calculated forces until they become smaller than defined convergence tolerances. The forces on the atoms in the studied inhibitors are calculated from the potential energy expression and will, therefore, depend on the force field that is selected.

The molecular dynamics (MD) simulations were performed using Materials Studio software [24]. The MD simulation of the interaction between the studied inhibitor molecule and iron (111) surface was carried out in a simulation box ($17.38 \text{ \AA} \times 17.38 \text{ \AA} \times 44.57 \text{ \AA}$) with periodic boundary conditions to model a representative part of the interface devoid of any arbitrary boundary effects. The Fe (111) was first built and relaxed by minimizing its energy using molecular mechanics, then the surface area of Fe (111) was increased and its periodicity is changed by constructing a super cell, and then a vacuum slab with 15 \AA thicknesses was built on the Fe (111) 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 calculation. Using 6 layers of iron atoms gives a sufficient depth that the inhibitor molecules will only be involved in non-bond interactions with iron atoms in the layers of the surface, without much increasing the calculation time. This structure then converted to have three dimension periodicity. As the three dimension periodic boundary conditions are used, it is important that the size of the vacuum slab is enough (15 \AA) that the non-bond calculation for the adsorbate does not interact with the periodic image of the bottom layer of atoms in the surface. After minimizing the Fe (111) surface and the tributylamine inhibitor molecules, the corrosion system will be built by layer builder to place the inhibitor molecules on Fe (111) surface, and the behaviors of these molecules on the Fe (111) surface were simulated using the COMPASS (condensed phase optimized molecular potentials for atomistic simulation studies) force field. Adsorption locator module in Materials Studio 5.0 [28] have been used to model the adsorption of the inhibitor molecules onto Fe (111) surface and thus provide access to the energetic of the adsorption and its effects on the inhibition efficiencies of the studied [29]. The binding energy between the studied inhibitors and Fe (111) surface were calculated using the following equation [30]:

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

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

4. RESULTS AND DISCUSSION

The main problem for QSAR resides not in performing the correlation itself but setting the variable selection for it [31]; the mathematical counterpart for such problem is known as the “factor indeterminacy” [32-36] and affirms that the same degree of correlation may be reached with in principle an infinity of latent variable combinations. Fortunately, in chemical-physics there are a

limited (although many enough) indicators to be considered with a clear-cut meaning in molecular structure that allows for rational of reactivity and bindings [37,38]. However, the main point is that given a set of N-molecules, one can choose to correlate their observed activities $A_{i=1,N}$ with M-selected structural indicators in as many combinations as [31]:

$$C = \sum_{k=1}^M C_M^k, C_M^k = \frac{M!}{k!(M-k)!} \quad (2)$$

linked by different endpoint paths, as many as [31]:

$$K = \prod_{k=1}^M C_M^k \quad (3)$$

indexing the numbers of paths built from connected distinct models with orders (dimension of correlation) from k=1 to k=M [31].

Table 2. Descriptors for the studied 23 inhibitor molecules calculated using quantum chemical methods

Structures	Inhibition Efficiency [1]	Balaban index JX	Balaban index JY	Wiener index	Total energy	HOMO eigenvalue	LUMO eigenvalue	Energy Gap	Total dipole	Molecular area (vdW area)	Molecular volume (vdW)	GFA prediction for Inhibition Efficiency
Tributylamine	97.76	3.39	3.49	300.00	-2116.93	-9.12	2.83	11.95	0.78	294.53	224.15	91.93
Aniline	97.76	3.00	3.03	42.00	-1071.22	-8.23	0.79	9.02	1.69	125.31	94.90	88.73
n-Octylamine	88.62	2.58	2.60	120.00	-1494.87	-10.01	3.35	13.37	1.82	215.72	157.87	82.58
Diphenylamine	92.08	2.14	2.17	264.00	-1893.45	-8.08	0.19	8.27	0.74	199.19	166.37	98.97
Dodecylamine	88.41	2.75	2.76	364.00	-2118.05	-9.86	3.29	13.15	1.44	301.45	224.52	89.70
di-n-Butylamine	86.41	2.54	2.61	120.00	-1494.37	-9.38	3.12	12.50	1.07	215.80	157.47	83.69
Cyclohexylamine	86.79	2.10	2.13	42.00	-1155.31	-9.91	3.49	13.40	1.63	149.05	112.89	77.34
n-butylamine	78.19	2.16	2.20	20.00	-871.49	-9.85	3.46	13.31	1.48	129.68	90.62	75.18
Triethylamine	69.12	2.88	3.03	48.00	-1182.04	-9.17	3.04	12.21	0.75	165.15	123.48	80.62
Hexylamine	74.67	2.43	2.45	56.00	-1183.13	-9.86	3.38	13.24	1.48	172.61	123.77	78.95
sec-Butylamine	83.14	2.50	2.55	18.00	-871.31	-9.77	3.54	13.30	1.39	126.25	90.66	74.83
Diethylamine	67.54	2.11	2.21	20.00	-871.09	-9.38	3.28	12.66	1.09	130.36	90.36	76.06
Propylamine	68.18	1.93	1.99	10.00	-715.76	-10.00	3.68	13.68	1.74	107.92	73.66	72.37
Isopropylamine	63.61	2.27	2.34	9.00	-715.61	-9.99	3.59	13.58	1.67	106.74	73.88	72.83
1,3-Dibutyl-2-thiourea	97.30	2.98	3.12	261.00	2036.95-	8.18-	0.25	8.43	4.61	262.68	197.74	101.88
1,3-Diethyl-2-thiourea	96.36	2.88	3.08	75.00	1413.72-	8.10-	0.65	8.75	6.44	176.23	130.84	93.08
1,3-Dimethyl-2-thiourea	95.96	2.77	3.02	31.00	1102.21-	8.29-	0.49	8.77	6.12	132.68	97.21	90.44
Thiourea	90.04	2.58	2.81	9.00	-791.42	-8.59	-0.06	8.53	5.15	91.98	63.75	89.70
Propagyl alc.	95.80	2.35	2.46	10.00	-757.9	10.73-	1.48	12.21	1.61	86.50	59.52	81.73
2-Pentyn-1-ol	87.67	2.84	2.92	35.00	-1069.76	-10.26	1.40	11.66	2.18	129.72	92.67	85.48
3-Butyn-1-ol	89.63	2.42	2.50	20.00	-913.79	-10.64	1.75	12.39	1.41	108.55	76.20	82.12
2-Butyn-1-ol	74.07	2.74	2.83	20.00	-913.97	-10.31	1.38	11.69	2.08	107.91	76.05	83.90
2-Butyn-1,4-diol	68.59	2.79	2.94	35.00	1234.41-	10.32-	1.22	11.54	2.35	119.85	84.68	85.57

Table 1 shows the molecular structures for the studied inhibitors with their inhibition efficiencies calculated from weight loss method as presented in the literature [1].

Table 2 shows the structural descriptors for the 23 inhibitors. It also records their inhibition efficiencies. Unless otherwise specified, the following units are used for quantities calculated by QSAR descriptors and properties; area (\AA^2), volume (\AA^3), dipole moment (e \AA), HOMO and LUMO (eV). The atom volumes and surfaces model calculates surface areas and volumes of surfaces around atomistic structures using the atom volumes and surfaces functionality of the Materials Studio software [24].

Table 3. Correlation matrix of the studied variables

	B : Inhibition Efficiency	C : Balaban index JX	D : Balaban index JY	E : Wiener index	F : Total energy	H : Heat of formation	I : HOMO eigenvalue	J : LUMO eigenvalue	K : Energy Gap	L : Total dipole	T : Molecular area (vdW area)	U : Molecular volume (vdW volume)
B : Inhibition Efficiency	1.0000	0.4533	0.4485	0.4299	0.4751	0.2924	0.4605	-0.5481	-0.5824	0.3818	0.3740	0.4029
C : Balaban index JX	0.4533	1.0000	0.9867	0.3873	-0.5054	-0.1503	0.3321	-0.3790	-0.4088	0.2850	0.4588	0.4522
D : Balaban index JY	0.4485	0.9867	1.0000	0.3198	0.4459-	-0.0794	0.3780	0.4553-	-0.4818	0.3985	0.3864	0.3801
E : Wiener index	0.4299	0.3873	0.3198	1.0000	0.9548-	-0.2329	0.3378	0.0657-	-0.1944	0.1077-	0.9297	0.9457
F : Total energy	-0.4751	-0.5054	-0.4459	0.9548-	1.0000	0.2317	-0.4106	0.1374	0.2761	7.0732e- 3	0.9496-	-0.9668
H : Heat of formation	0.2924	-0.1503	-0.0794	0.2329-	0.2317	1.0000	0.5554	-0.6689	-0.7078	0.3293	0.4197-	-0.3488
I : HOMO eigenvalue	0.4605	0.3321	0.3780	0.3378	0.4106-	0.5554	1.0000	-0.5350	-0.8112	0.5145	0.3173	0.3615
J : LUMO eigenvalue	-0.5481	-0.3790	-0.4553	0.0657-	0.1374	-0.6689	-0.5350	1.0000	0.9280	-0.6281	0.1199	0.0739
K : Energy Gap	-0.5824	-0.4088	-0.4818	0.1944-	0.2761	-0.7078	-0.8112	0.9280	1.0000	-0.6616	-0.0569	-0.1082
L : Total dipole	0.3818	0.2850	0.3985	0.1077-	7.0732e- 3	0.3293	0.5145	-0.6281	-0.6616	1.0000	-0.0894	-0.0938
T : Molecular area (vdW area)	0.3740	0.4588	0.3864	0.9297	0.9496-	-0.4197	0.3173	0.1199	-0.0569	-0.0894	1.0000	0.9956
U : Molecular volume (vdW volume)	0.4029	0.4522	0.3801	0.9457	0.9668-	-0.3488	0.3615	0.0739	-0.1082	-0.0938	0.9956	1.0000

Molecular area (vdW area) in Table 2, describes the volume inside the van der waals area of the molecular surface area determines the extent to which a molecule exposes to the external environment. This descriptor is related to binding, transport, and solubility. Molecular volume (vdW volume) in Table 2, describes the volume inside the van der waals area of a molecule. Total molecular dipole moment, this descriptor calculates the molecule dipole moments from partial charges defined on the atoms of the molecule. If no partial charges defined, the molecular dipole moment will be zero. Total energy, HOMO and LUMO energy have been described in our previous study in details [39].

For understanding the quantitative structure and activity relationships, statistical analysis using genetic function approximation (GFA) method, first a study table was belt and presented in Table 2. Second, a correlation matrix, Table 3 was derived, and then regression parameters were obtained.

Table 2 shows the structural descriptors for the 23 inhibitor molecules used in this study (training set). The structure descriptors presented in Table 2 include total energy, HOMO and LUMO energy as well as the area and volume of the studied molecules. Also, distance and connectivity based topological indices and their applications are used in understanding the quantitative structure-activity relationship (QSAR). Use of topological indices has been well illustrated in the literature [40]. The topological indices used in this study and presented in Table 2, namely Balaban (J) and Wiener indices both are well presented in the literature [41]. In the fields of chemical graph theory, molecular topology, and mathematical chemistry, a topological index also known as a connectivity index which is a type of a molecular descriptor that is calculated based on the molecular graph of a chemical compound [42]. Topological indices are used in the development of quantitative structure-activity relationships (QSARs) in which the activity or other properties of molecules are correlated with their chemical structure [43]. Topological indices are 2D descriptors based on graph theory concept. These indices have been widely used in QSAR studies. They help to differentiate molecules according to their size, degree of branching, flexibility, and overall shape.

Table 4. Univariate analysis of the inhibition data

Statistical parameters	
Number of sample points	23.0
Range	34.15
Maximum	97.76
Minimum	63.61
Mean	84.25
Median	87.67
Variance	122.02
Standard deviation	11.29
Mean absolute deviation	9.66
Skewness	-0.42
Kurtosis	-1.34

A univariate analysis is performed on the inhibition efficiency data in Table 1 as a tool to assess the quality of the data available and its suitability for next statistical analysis. Data in Table 1 show acceptable normal distribution. Statistical parameters presented in Table 4 have been discussed in details in our previous study [39].

Table 3 contains a correlation matrix which gives the correlation coefficients between each pair of columns included in the analysis in Table 2. Correlation coefficients between a pair of columns approaching +1.0 or -1.0 suggest that the two columns of data are not independent of each other. The cells in the correlation matrix, Table 3 are colored according to the correlation value of each cell. Inspection of Table 3 shows that the descriptors most highly correlated with corrosion inhibition efficiency include: LUMO energy and volume of the inhibitor molecule. After constructing the correlation matrix the genetic function approximation algorithm will be used to perform a regression

analysis. The GFA algorithm works with a set of strings, called a population. This population is evolved in a manner that leads it toward the objective of the search. Following this, three operations are performed iteratively in succession: selection, crossover, and mutation. Newly added members are scored according to a fitness criterion. In the GFA, the scoring criteria for models are all related to the quality of the regression fit to the data. The selection probabilities must be re-evaluated each time a new member is added to the population [26]. The procedure continues for a user-specified number of generations, unless convergence occurs in the interim. Convergence is triggered by lack of progress in the highest and average scores of the population.

4.1 Friedman LOF measure

Various statistical measures can be adapted to measure the fitness of a GFA model during the evolution process.

Table 5. Validation Table of the Genetic Function Approximation

Predicted Inhibition Efficiency = -4.90 (LUMO) + 0.102 (Molecular Volume) +82.88	
Friedman LOF	83.18
R-squared	0.498
Adjusted R-squared	0.448
Cross validated R-squared	0.368
Significant Regression	Yes
Significance-of-regression F-value	9.92
Critical SOR F-value (95%)	4.640
Lack-of-fit points	20.00
Min expt. error for non-significant LOF (95%)	6.68

Use of the Friedman lack-of-fit (LOF) measure has several advantages over the regular least square error measure. In Materials Studio [24], LOF is measured using a slight variation of the original Friedman formula[44]. The revised formula is:

$$LOF = \frac{SSE}{\left(1 - \frac{c + dp}{M}\right)^2} \quad (4)$$

Where SSE is the sum of squares of errors, c is the number of terms in the model, other than the constant term, d is a user-defined smoothing parameter, p is the total number of descriptors contained in all model terms (again ignoring the constant term) and M is the number of samples in the training set. Unlike the commonly used least squares measure, the LOF measure cannot always be reduced by adding more terms to the regression model. While the new term may reduce the SSE, it also increases the values of c and p, which tends to increase the LOF score. Thus, adding a new term

may reduce the SSE, but actually increases the LOF score. By limiting the tendency to simply add more terms, the LOF measure resists overfitting better than the SSE measure [26].

Table 5 shows the GFA analysis which gives summary of the input parameters used for the calculation. Also, it reports whether the GFA algorithm converged in specified number of generations. The GFA algorithm is assumed to have converged when no improvement is seen in the score of the population over a significant length of time, either that of the best model in each population or the average of all the models in each population. When this criterion has been satisfied, no further generations are calculated.

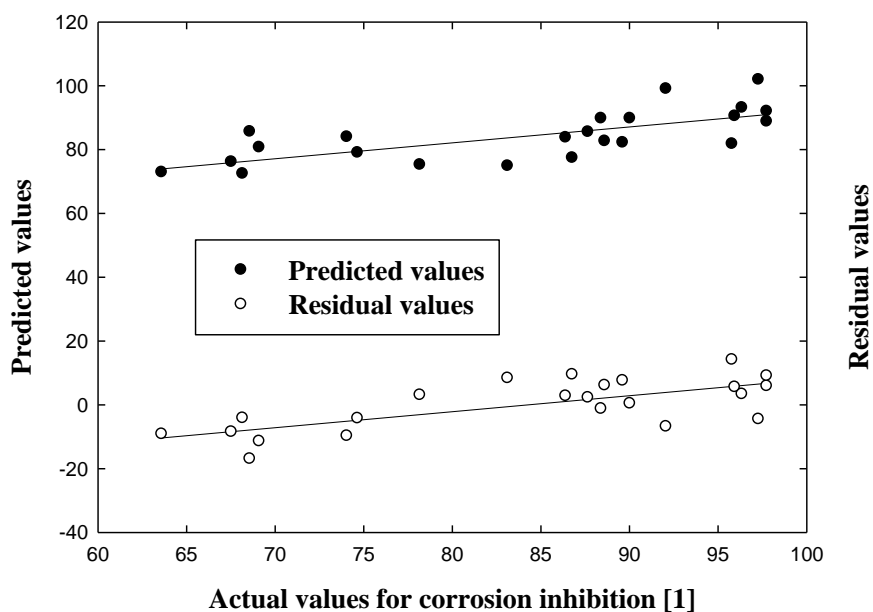


Figure 1. Plot of predicted inhibition and residuals versus measured corrosion inhibition [1]

The Friedman's lack-of-fit (LOF) score in Table 5 evaluates the QSAR model. The lower the LOF, the less likely it is that GFA model will fit the data. The significant regression is given by F-test, and the higher the value, the better the model.

Figure 1 shows the relationship between the measured corrosion inhibition efficiencies of the studied inhibitors presented in Table 2 and the predicted efficiencies calculated by the following equation:

$$\text{Predicted Inhibition Efficiency} = -4.90 (\text{LUMO}) + 0.102 (\text{Molecular Volume}) + 82.88 \quad (5)$$

The distribution of the residual values against the measured corrosion inhibition efficiencies values are presented in Figure 1. The residual values can be defined as the difference between the predicted value generated by the model and the measured values of corrosion inhibition efficiencies.

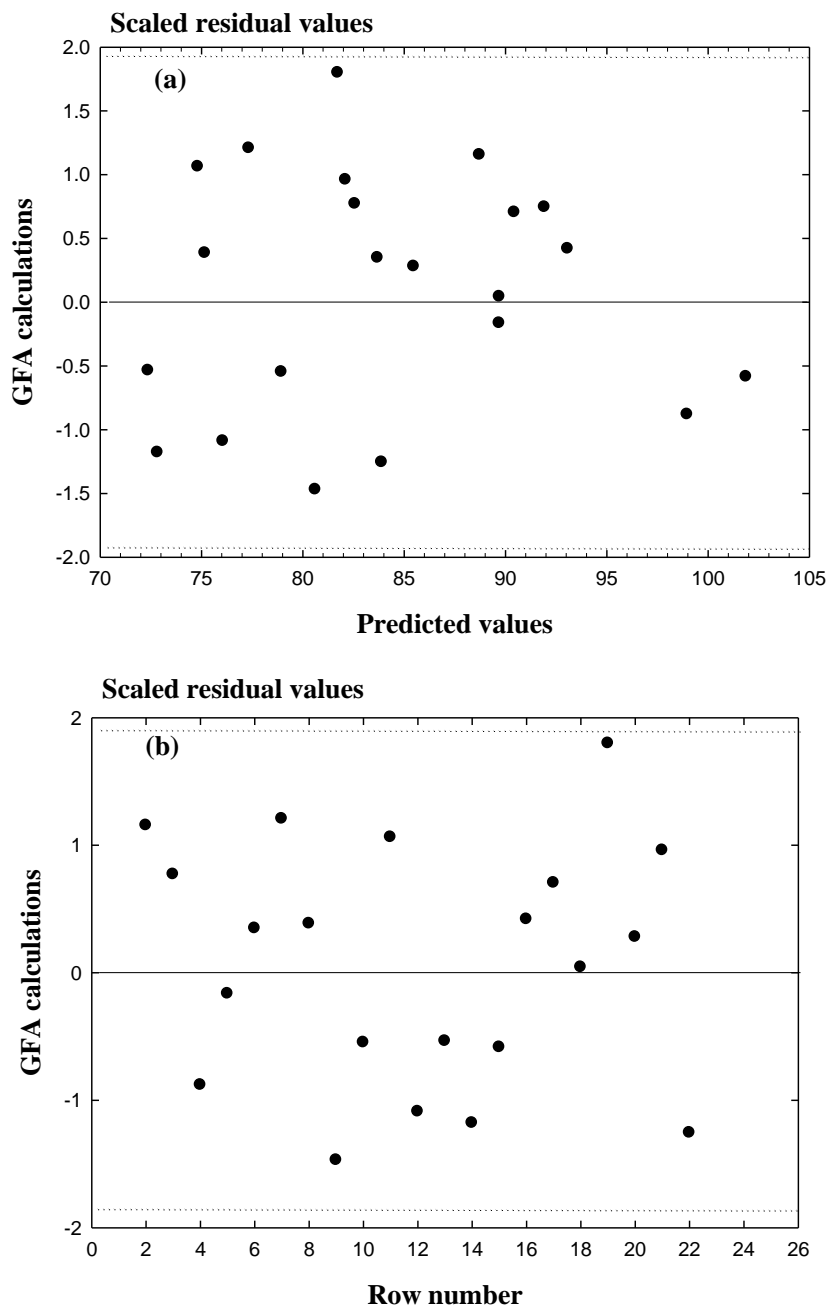


Figure 2. Outlier analysis for inhibition efficiency

Figure 2a-b represents the potential outlier that used to test the constructed QSAR model. An outlier can be defined as a data point whose residual value is not within two standard deviations of the mean of the residual values. Figure 2a represents the residual values plotted against the measured corrosion inhibition efficiencies. Figure 2b shows the residual values plotted against Table 2 row number. Figure 2a-b 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 Fig. 2a-b shows that there is no points appeared outside the dotted lines which make the QSAR model acceptable.

4.2 Molecular dynamic simulations

To investigate the adsorption mechanism of the studied compounds on the steel surface, the adsorption of tributylamine was studied theoretically by using molecular dynamics simulation methods.

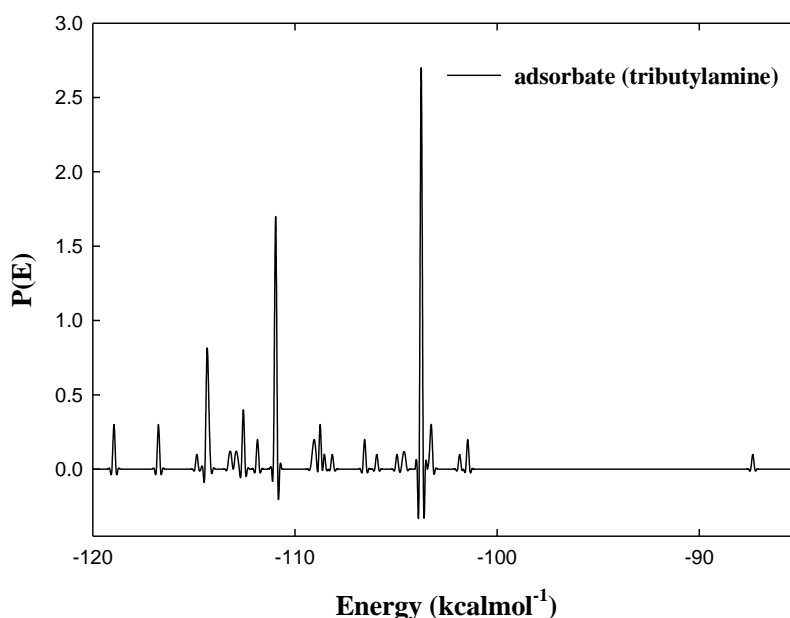


Figure 3. Adsorption energy distribution of the adsorbate (tributylamine) on Fe (111)

Figure 3, shows the distribution of adsorption energy of the adsorbate (tributylamine) Fe (111). As can be seen in Fig. 3, the adsorption energy of tributylamine is more than $-100 \text{ kcal mol}^{-1}$ which explain its highest inhibition efficiency compared to the other series of inhibitors. Figure 4 shows the most suitable configurations for adsorption of tributylamine on Fe (111) substrate obtained by adsorption locator module [45,46] in Materials studio [24]. The outputs and descriptors calculated by the Monte Carlo simulation are presented in Table 6. The parameters presented in Table 6 include total energy, in kcal mol^{-1} , of the Fe(111)– tributylamine configuration. The total energy is defined as the sum of the energies of the adsorbate components, the rigid adsorption energy and the deformation energy. In this study, the substrate energy (iron surface) is taken as zero. In addition, adsorption energy in kcal mol^{-1} , reports energy released (or required) when the relaxed adsorbate component (tributylamine) is adsorbed on the substrate. The adsorption energy is defined as the sum of the rigid adsorption energy and the deformation energy for the adsorbate components. The rigid adsorption energy reports the energy, in kcal mol^{-1} , released (or required) when the unrelaxed adsorbate components (i.e., before the geometry optimization step) are adsorbed on the substrate.

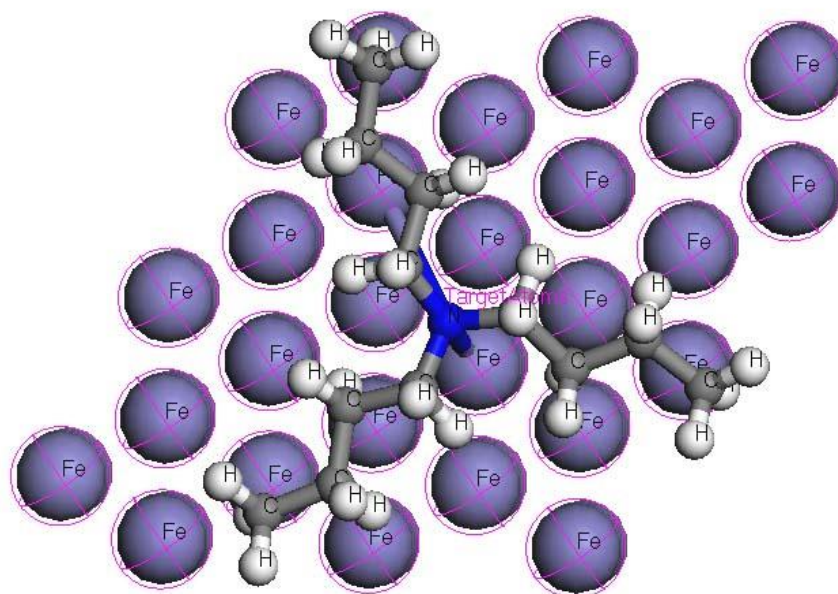


Figure 4. The most suitable configuration for adsorption of tributylamine on Fe (111) substrate obtained by adsorption locator module

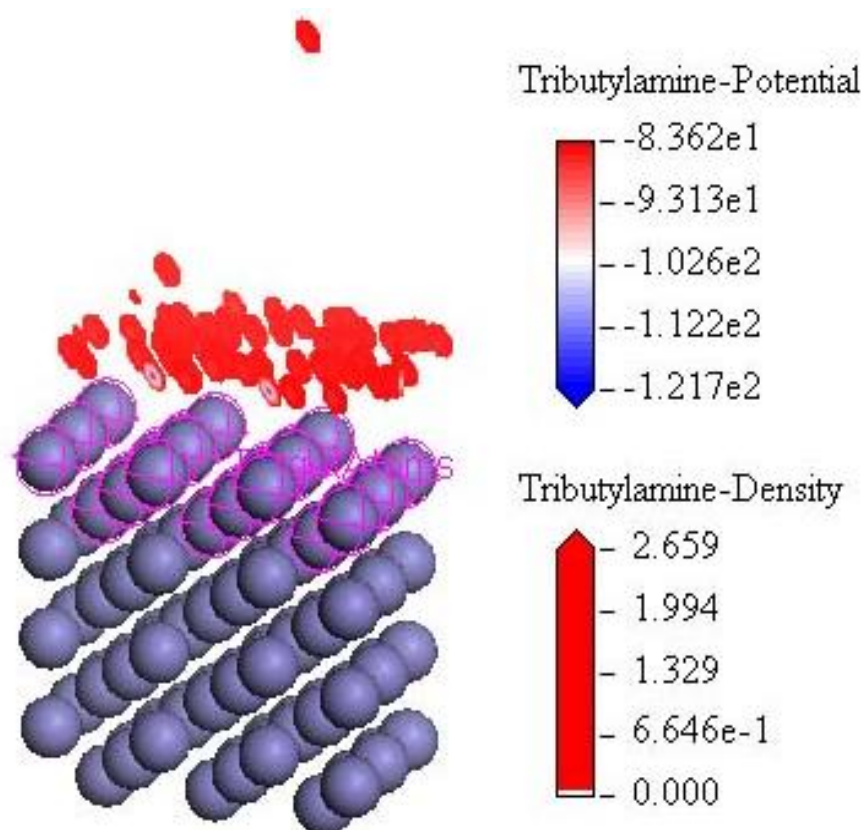


Figure 5. The adsorption density field of tributylamine on the Fe (111) substrate.

Table 6. Outputs and descriptors calculated by the Mont Carlo simulation for adsorption of tributylamine on iron (111)

Inhibitor	Total energy kcal mol ⁻¹	Adsorption energy/ kcal mol ⁻¹	Rigid adsorption / energy kcal mol ⁻¹	Deformation energy/ kcal mol ⁻¹	dE _{ad} /dN _i kcal mol ⁻¹	Calculated binding energy / kcal mol ⁻¹
tributylamine	156.75	-121.66	-122.37	0.703	-121.66	132.45

The deformation energy reports the energy, in kcal mol⁻¹, released when the adsorbed adsorbate components are relaxed on the substrate surface. Table 6 shows also (dE_{ads}/dN_i), which reports the energy, in kcal mol⁻¹, of substrate–adsorbate configurations where one of the adsorbate components has been removed. The binding energy introduced in Table 6 calculated from equation (1). As can be seen from Table 6, tributylamine gives high adsorption energy during the simulation process. High values of adsorption density presented in Fig. 5 indicates that tributylamine is likely to adsorb on the iron surface to form stable ad layers and protect iron from corrosion.

5. CONCLUSIONS

QSAR methods are now being used for predicting the inhibition efficiencies for corrosion inhibitors in dry laboratories. Success depends upon having data on a large number of compounds available. The computational method has proved satisfactory for the inhibition efficiency estimations. High correlation was obtained with the multivariate correlation, i.e. all the indices combined together, where the prediction power was very high for GFA. Although GFA proved to be efficient in predicting ability, more work is still required toward understanding structure-property correlation on inhibition corrosion studies, particularly concerning the analysis of different structural chemical descriptors. Understanding adsorption phenomena is of key importance in corrosion problems. Computational studies helps to find the most stable adsorption sites for a broad range of materials. This information can help to gain further insight about corrosion system, such as the most likely point of attack for corrosion on a surface, the most stable site for inhibitor adsorption and the binding energy of the adsorbed layer.

ACKNOWLEDGMENT

One of the authors is grateful for the financial support provided by the centre of research excellence in corrosion located at King Fahd University, Project # CR-03-2010 titled "Designing new corrosion inhibitors by QSAR and Molecular Dynamics Approaches.

References

1. S. P. Cardoso, E. Hollauer, L. E. P. Borgesc and J. A. da C. P. Gomes *J. Braz. Chem. Soc.*, 17 (2007) 1241.
2. I. Lukovits, I. Bakó, A. Shaban, E. Kálmán, *Electrochim. Acta*, 43 (1998) 131.
3. A. E. Stoyanova; E. I. Sokolova; S. N. Raicheva; *Corros. Sci.*, 39 (1997) 1595.
4. A. A. El-Shafei, M. N. H. Moussa, A. A. El-Far; *Mater. Chem. Phys.*, 70 (2001) 175.
5. M. A. Quraishi, F.A. Ansari, D. Jamal, *Mater. Chem. Phys.*, 77 (2002) 687.
6. F. Bentiss, M. Traisnel, H. Vezin, M. Lagrené, *Corros. Sci.*, 45 (2003) 371.
7. I. Bergman, *Trans. Faraday Soc.*, 50 (1954) 829.
8. F. M. Donahue, K. Nobe, *J. Electrochem. Soc.*, 112 (1966) 886.
9. J. Vosta, J. Eliášek, *Corros. Sci.*, 11 (1971) 223.
10. F. B. Growcock, *Corrosion* 45 (1989) 1003.
11. P. G. Abdul-Ahad, S. H. F. Al-Madfa'i, *Corrosion* 45 (1989) 979.
12. P. Dupin, D. A. Vilorio-Vera, A. de Savignac, A. Lattes, B. Sutter, Ph. Haicour, Correlations Between the Molecular Structure of Some Organic Compounds and Their Corrosion Inhibiting Properties in Deaerated Media Containing Hydrogen Sulfide, 5th European Symposium on Corrosion Inhibitors, 2 (1980) 301.
13. V. S. Sastri, J. R. Perumareddi, *Corrosion*, 53 (1997) 617.
14. P. Kutej, J. Vosta, J. Pancir, J. Macak, N. Hackerman, *J. Electrochem. Soc.*, 142 (1995) 829..
15. C. Öğretir, B. Mihçi, G. Bereket, *J. Mol. Str.*, 488 (1999) 223.
16. G. Bereket, E. Hür, C. Öğretir, *J. Mol. Str.*, 578 (2002) 79.
17. N. Khalil, *Electrochim. Acta*, 48 (2003) 2635.
18. K.F. Khaled, K. Babić-Samardžija, N. Hackerman, *Electrochim. Acta*, 50 (2005) 2515.
19. K.F.Khaled, *Appl. Surf. Sci.*, 252 (2006) 4120.
20. K.F.Khaled, *Electrochim. Acta*, 48 (2003) 2493.
21. K.F.Khaled, *Mat. Chem. and Phys.*, 112 (2008) 290.
22. K.F.Khaled, *J. Appl. Electrochem.*, 41 (2011) 423.
23. J.P. Perdew and Y. Wang, *Phys. Rev. B*, 45 (1992) 13244.
24. B. Delley, *J. Chem. Phys.*, 92 (1990) 508..
25. M. Pereiro, D. Baldomir, M. Iglesias, C. Rosales and M. Castro, *Int. J. Quantum Chem.*, 81 (2001) 422.
26. Materials Studio 5.0 Manual
27. O. Ermer, Calculation of molecular properties using force fields, Applications in organic chemistry", *Structure and Bonding*, 27 (1976)161.
28. J. Barriga, B. Coto and B. Fernandez, *Tribol. Int.*, 40 (2007) 960.
29. K.F. Khaled, *J. Solid State Electrochem.*, 13 (2009) 1743.
30. K.F.Khaled, Sahar A. Fadel-Allah and B. Hammouti *Mater. Chem and Phys.*, 117 (2009) 148.
31. M. V. Putz, A. Putz, M. Lazea, L. Ienciu, A. Chiria, *Int. J. Mol. Sci.*, 10 (2009) 1193.
32. J. H. Steiger, P. H. Schonemann, A history of factor indeterminacy. In Theory Construction and Data Analysis in the Behavioural Science, Shye, S., Ed.; Jossey-Bass Publishers: San Francisco, CA, USA, 1978.
33. C. Spearman, The Abilities of Man; MacMillan: London, UK, 1927.
34. Wilson, E.B. Review of the abilities of man, their nature and measurement, by Spearman, C. *Science* 67 (1928) 244.
35. E. B. Wilson, M. M. Hilferty, The distribution of chi-square. In Proc. Nat. Acad. Sci. USA 1931, 17, 684.
36. E. B. Wilson, J. A. Worcester, A note on factor analysis. *Psychometrika* 4 (1939) 133.
37. J. G. Topliss, R. J. Costello, *J. Med. Chem.*, 15 (1972) 1066.
38. J. G. Topliss, R. P. Edwards, *J. Med. Chem.*, 22 (1979) 1238.

39. K.F.Khaled, Corros. Sci. 2011, doi:10.1016/j.corsci.2011.01.035 in press
40. OECD, Report on the regulatory uses and applications in OECD member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. Organization of Economic Cooperation and Development: Paris, France, 2006; Available online: <http://www.oecd.org/>, accessed January 2009.
41. A. P. Worth, A. Bassan, E. Fabjan, A. Gallegos Saliner, T. I. Netzeva, G. Patlewicz, M. Pavan, I. Tsakovska, The characterization of quantitative structure-activity relationships: Preliminary guidance. European Commission - Joint Research Centre: Ispra, Italy, 2005; Available online: <http://ecb.jrc.it/qsar/publications/>, accessed January 2009.
42. H. Timmerman; T. Roberto; V. Consonni; R. Mannhold; H. Kubinyi (2002). *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH. ISBN 3-527-29913-0.
43. Hall, Lowell H.; Kier, Lemont B. (1976). *Molecular connectivity in chemistry and drug research*. Boston: Academic Press. ISBN 0-12-406560-0.
44. J. H. Friedman, *Multivariate Adaptive Regression Splines, Technical Report No. 102*, Laboratory for Computational Statistics, Department of Statistics, Stanford University: Stanford (November 1988, rev. August 1990).
45. V. Cerný, *J. Optim. Theor. Appl.*, 45 (1985) 41.
46. S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *Science*, 220 (1983) 671.