

QSPR Prediction Analysis of Corrosion Inhibitors in Hydrochloric Acid on 22%-Cr Stainless Steel

S. P. Cardoso,^a E. Hollauer,^{*,b} L. E. P. Borges^c and J. A. da C. P. Gomes^d

^aCentro Federal de Educação Tecnológica de Química de Nilópolis, Rua Lúcio Tavares 1045, Centro, Nilópolis-RJ, Brazil

^bPrograma de Pós-Graduação em Química, Instituto de Química, Universidade Federal Fluminense, Outeiro de São João Batista s/n, 24210-130 Niterói-RJ, Brazil

^cDepartamento de Engenharia Química, Instituto Militar de Engenharia, Praça General Tibúrcio, 80, Praia Vermelha, 22290-270 Rio de Janeiro-RJ, Brazil

^dLaboratório de Corrosão, Programa de Engenharia Metalúrgica e de Materiais, COPPE, Universidade Federal do Rio de Janeiro, Ilha do Fundão, Centro de Tecnologia, Rio de Janeiro-RJ, Brazil

Inibidores de corrosão são largamente utilizados na prevenção de corrosão em operações de estimulação em poços de petróleo. Investigações teóricas e experimentais detalhadas de vinte e três compostos incluindo aminas, derivados de tiouréia e álcoois acetilênicos foram realizadas para estimar a sua eficiência na inibição de corrosão do aço 22% Cr (de estrutura austenítico-ferrítica; duplex) em soluções de ácido clorídrico (15% m/v). Os dados obtidos foram interpretados teoricamente com respeito à análise de regressão (OLS), análise de componentes principais (PCA), predição e análise de regressão parcial (PLS) empregando descritores quânticos e baseados na contribuição de grupos. Em nosso estudo vemos vantagem no uso da função de adsorção de Langmuir isoestérica em peso (WILA), $\ln(\theta M/(1-\theta))$ or $\ln(K_{ads})$. Excelentes correlações foram obtidas para a maioria dos modelos e algumas equações, resultados, curvas de calibração e validação estão descritas no texto. O resultado do presente trabalho representa um esforço preliminar no sentido de prover um eficiente método para a estimativa da eficiência da inibição de corrosão, de inibidores arbitrários em vários metais, ligas e tipos de aço.

Corrosion inhibitors have been widely used to prevent corrosion on stimulation operations on petroleum wells. Detailed experimental and theoretical investigation of twenty three different compounds including amines, thiourea derivatives, and acetylenic alcohols were carried out to estimate their inhibition corrosion efficiency on 22% Cr stainless steel (austenitic-ferritic, duplex) in hydrochloric acid (15% m/v) solutions. The obtained data were theoretically interpreted with respect to prediction, regression analysis (OLS), principal analysis component (PCA) and partial regression analysis (PLS) employing quantum and group contribution descriptors. In our study we found advantage in the use of the weight isoesteric Langmuir adsorption function (WILA), $\ln(\theta M/(1-\theta))$ or $\ln K_{ads}$. Excellent correlations were obtained for most models and few equations, results, calibration and validation plots were discussed in the text. The content of the present work represents a first step toward an efficient estimation of the inhibition corrosion efficiency (ICE) of arbitrary inhibitors towards several metals, alloys and steel types.

Keywords: corrosion inhibitor, regression analysis, QSPR (Quantitative Structure Properties Relationship), interfaces

Introduction

Corrosion inhibitors have been widely used in stimulation operations on petroleum wells.^{1,2} Due to the

ready removal of iron oxides and carbonated minerals, these operations make extensive use of hydrochloric acid solutions (15% m/v) at temperatures up to 80 °C exposed to several steel types. Under such aggressive media the use of corrosion inhibitors (CI) are mandatory, whether used singly or as mixtures. Among the well known CI¹

*e-mail: edu@kabir.gqt.uff.br

currently used to prevent corrosion on HCl media are amines, amides, nitriles, imidazolines, triazoles, pyridine and quinoline derivatives, thiourea derivatives, thiosemicarbazide and thiocyanates as some of the molecular inhibitors currently used to prevent corrosion in hydrochloric acid medium.

The 22% Cr stainless steel (austenitic-ferritic), hereafter simply the duplex steel, is extensively used in the petroleum industry due to excellent mechanical properties and a better corrosion resistance for co-produced fluids in the oil/gas production, that is credited to its high chromium, molybdenum and nitrogen content.¹⁻³ However many authors pointed out that under acid solutions the corrosion rate is significantly larger, a condition that requires the use of CI mixtures. Besides the corrosion inhibitors commonly used to protect low alloy steels showed very poor efficiencies to protect high chromium steels in acid medium, especially when the chloride ion is present. Therefore intense work is in progress in order to develop high efficiency inhibitor mixtures to be used for high chromium alloys, like duplex steel, in acid solution at high temperatures, a common condition found during acidification process of oil wells.

Despite the intense empirical searches for new commercial inhibitors, few articles address chemometric analysis of the inhibition corrosion efficiency (ICE). Such a procedure represents a challenge to the application of regular structure-activity chemometric thinking applied in biological fields, since the physical adsorption is non-specific, in opposition to the key-lock mechanism found in molecular biology. 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 results⁴⁻⁹ correlating small number of inhibitors and quantum descriptors.

Lukovits *et al.*⁴ employed a polynomial regression analysis for the Langmuir adsorption constant for a set of seven thiourea derivatives obtaining correlations within 0.969-0.982 for R-values with few quantum chemical descriptors. Bentiss *et al.*,⁸ based on inhibition efficiency found an excellent correlation between the charge transfer resistance of six triazole and oxadiazole derivatives to three AM1 quantum descriptors. Bentiss⁸ obtained correlations within 0.91-0.96 for the R-values. Recently, Khalil⁹ extended this study by correlating twelve thiosemicarbazone and thiosemicarbazide derivatives to five quantum MNDO/PM3 descriptors. All these studies concern carbon steel and up to our knowledge no similar study has been carried out yet for duplex steel. Clearly much work must be done to improve predictability in this field.

This work is part of a continuous effort aiming the efficient prediction of molecular properties based on the QSPR (Quantitative Structure Properties Relationship) methodology. Although 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 property value. This procedure, however, has a bottleneck related to the ability to perform ICE predictions for compounds, which do not belong to the calibration family information. This particular condition requires an intense effort toward finding molecular descriptors useful for predicting correct ICE 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.

In the present work we carried out detailed experimental and theoretical investigations of 23 different IC compounds including amines, thiourea derivatives, and acetylenic alcohols to estimate its ICE on 22% Cr stainless steel in hydrochloric acid (15% m/v) solutions at 60 °C. These inhibition corrosion efficiencies are then used to build the WILA function, the weight isoesteric Langmuir adsorption function, defined as $\ln(\theta M/(1-\theta))$ or $\ln(K_{ads})$. This WILA function is then correlated to quantum and group contribution parameters. 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. In the next sections we shall present experimental details, followed by the principal component analysis (PCA), a simple model based on the minimization of the second-order cross-validation error with ordinary least squares method (OLS) and a partial least squares (PLS) analysis.

Principal component analysis (PCA) is a classic statistical method well described in the literature.¹⁴ Briefly, it attempts to describe the data variation in multidimensional analysis by employing a reduced number of new orthogonal variables, named principal components or latent variables.

The loadings are the weights that link the original variables to each latent variable. This means that their values may be used as an indication of the importance of each original variable to a given principal component. It is often used searching for linear dependencies or as classification tools.

Ordinary Least Squares (OLS) and Partial Least Squares Regression (PLS) are widely used methods for correlating the variations of a response function to the variations of several descriptors. The OLS is based on the minimization of the sum of the squared error functional. When the data present intense correlation PLS becomes widely used in order to simplify this relationship into a small number of components, known as latent variables. We complement this work with the PLS methodology aiming to offer a balanced analysis of the calibration and prediction problem in the IC field.

Experimental

All inhibition corrosion data here reported have been obtained through weight loss experiments of rectangular steel specimens of $2 \times 1 \times 0.5$ cm dimensions with a central hole. These have been cleaned with acetone, washed with water, dried and then weighed with a 0.1 mg precision. The exposed surface represents the active state of the metal. Two results were averaged for each inhibitor. The experiments were carried out in cylindrical autoclaves internally covered with teflon. These autoclaves have been placed on a rolling oven at 60 °C for 3 h. All solutions were made up with 300 mL of HCl (15% m/v), 2% m/v of the chemical inhibitor and 0.6% m/v of formaldehyde to minimize hydrogen penetration. The experimental conditions were set up to avoid complete dissolution of the reference electrodes and strictly followed industrial recommendations, for which no more than 2% m/v on active components are allowed for matrix acidification operations.

Usually the inhibition corrosion efficiency (ICE) is employed as the response property, however, since in our study all experiments are carried out with equal inhibitor weight (2%) a second function is more adequate to QSPR correlation studies. Since ΔG_{ads} is a thermodynamic property that shows linear dependence to energies, volumes and the inhibitor polarizability we propose the use of the weight isoesteric Langmuir adsorption function (WILA), defined as $\ln(\theta M/(1-\theta))$ (hereafter simply $\ln(K_{\text{ads}})$) as the response property in QSPR runnings. Table 1 lists the 23 inhibitors employed in our study, its name, and the WILA function value. It is important to remark that is not required that any specific inhibitor fit the Langmuir isotherm to obtain good results on regular OLS calculations, actually we are not aware of any inhibitor/

Table 1. The twenty-three inhibitors employed and its WILA value for the function, $\ln(\theta M/(1-\theta))$

	Compound	$\ln(\theta M/(1-\theta))$
1	Tributylamine	7.45
2	Aniline	7.96
3	<i>n</i> -Octylamine	6.44
4	Diphenylamine	8.46
5	Dodecylamine	7.03
6	di- <i>n</i> -Butylamine	5.99
7	Cyclohexylamine	5.32
8	<i>n</i> -butylamine	4.02
9	Triethylamine	5.06
10	Hexylamine	4.64
11	<i>sec</i> -Butylamine	3.46
12	Diethylamine	3.73
13	Propylamine	3.47
14	Isopropylamine	2.95
15	1,3-Dibutyl-2-thiourea	9.60
16	1,3-Diethyl-2-thiourea	9.17
17	1,3-Dimethyl-2-thiourea	8.99
18	Thiourea	7.80
19	Propargyl alcohol	7.63
20	2-Pentyn-1-ol	4.68
21	3-Butyn-1-ol	7.60
22	2-Butyn-1-ol	4.28
23	2-Butyne-1,4-diol	4.56

metal system following Langmuir isotherms. It is enough that the WILA function works as the logarithm of a true adsorption equilibrium constant, and we obtain better results than those employing the logarithm of inhibitor efficiencies.

Among the most efficient inhibitors are many thiourea derivatives, 1,3-dibutyl, 1,3 dimethyl and 1,3 diethyl thiourea, followed by few amines like diphenyl amine, aniline, thiourea itself and 3-butyne-1-ol alcohol. On the other side we point out the aliphatic amines among the less efficient inhibitors isopropylamine, *sec*-butylamine, propylamine and diethylamines. Such data provide important chemometric information related to the absence of any inhibitor efficiency.

Calculations

The AM1 methodology coded on Mopac 6.0¹⁰ was employed for the quantum descriptor calculations, except for the volume calculation that was carried out with Pcmol¹¹ program. For the QSPR calculations we used the QSAR program, written by Fedders and Ponder,¹² and slightly modified in our laboratory. The PCA and PLS were carried out with the Unscrambler 6.11 software.¹³ Additionally to the quantum descriptors, a group contribution descriptors have been employed to offer a well-balanced descriptor set. Implicitly this is related to the balance between the classical theories, which have local

and group contribution and the quantum theories, collective by the nature of the chemical theories.

The group descriptors used were the following: A1 is the number of RNH₂ groups (primary amines); A2 is the number of R₁R₂NH groups (secondary amines); A3 is the number of R₁R₂R₃N groups (tertiary amines); NB is the number of phenyl groups; NC is the number of cyclic carbon rings; NCS is the number of CS bonds; NT is the number of triple CC bonds; NOH is the number of OH groups; NCR is the average number of carbon atoms within a branch; NR is the branching number; while N is the inhibitor number of moles present in the vessel. These group contribution descriptors are listed on Table 2, and totalize eleven descriptors of simple and ready evaluation.

The quantum set of descriptors concerns the following molecular properties: ED is the dimerization energy; M is the molecular IC mass; P is the polarizability given in atomic units; C is the charge of the polar group, C1 is the charge of the S, N and triple CC adsorption site; C2 is the charge of the aromatic ring (or C in presence of a polar group), C12 is the charge of two neighbor atoms to the polar group; C13 is the charge of the three neighbor atoms to the polar group; C14 is the charge of the four neighbor atoms to the polar group; EH is the HOMO energy; EL is the LUMO energy; DIF is the energy gap, DP is the dipole; V is the calculated volume. The fourteen quantum descriptors are listed on Table 3 while the whole set employs twenty-five molecular descriptors. The polar group is defined as the amine, alcoholic or amide group present.

Results and Discussion

The calculations were carried out with centered and self-scaled descriptors and the response function. The calculations employed up to the 25 descriptors previously described.

Principal Component Analysis (PCA)^{14,15}

A preliminary principal component analysis (PCA) has been carried out in order to identify possible linear dependencies and the descriptor variance. The energy gap have been identified as an obvious linear dependency to the HOMO and LUMO energy. The main component, PC1, is a mixture of volume and molecular masses with minor contributions from dimerization energies, polarizability, the energy gap and the LUMO energy. The second component, PC2, showed a small participation of dipole moment. These results are expected since the whole set has masses within the 56-188 au range and therefore masses, volume and polarizability, three strongly correlated descriptors, are dominant terms in the main components.

A cluster analysis was conducted to identify possible anomalous molecules. Figure 1 presents the score cluster plot for PC2/PC1. Other scoring functions are available upon request. The whole family of alcohols (**19-23**) has been grouped with negative values of PC1 while the aliphatic amine family (**1,3,5-14**) and the

Table 2. Table with all the groups contributing descriptors values for each corrosion inhibitor

Compound	A1	A2	A3	NB	NC	NCS	NT	NOH	NCR	NR	N
1	0	0	1	0	0	0	0	0	4	3	0.032
2	1	0	0	1	0	0	0	0	0	0	0.064
3	1	0	0	0	0	0	0	0	8	1	0.046
4	0	1	0	2	0	0	0	0	0	0	0.035
5	1	0	0	0	0	0	0	0	12	1	0.032
6	0	1	0	0	0	0	0	0	4	2	0.046
7	1	0	0	0	1	0	0	0	0	0	0.060
8	1	0	0	0	0	0	0	0	4	1	0.082
9	0	0	1	0	0	0	0	0	2	3	0.059
10	1	0	0	0	0	0	0	0	6	1	0.057
11	1	0	0	0	0	0	0	0	3	2	0.082
12	0	1	0	0	0	0	0	0	2	2	0.082
13	1	0	0	0	0	0	0	0	3	1	0.101
14	1	0	0	0	0	0	0	0	1	2	0.101
15	0	2	0	0	0	1	0	0	4	2	0.031
16	0	2	0	0	0	1	0	0	2	2	0.045
17	0	2	0	0	0	1	0	0	1	2	0.057
18	2	0	0	0	0	1	0	0	0	2	0.078
19	0	0	0	0	0	0	1	1	3	1	0.108
20	0	0	0	0	0	0	1	1	5	1	0.059
21	0	0	0	0	0	0	1	1	4	1	0.071
22	0	0	0	0	0	0	1	1	4	1	0.071
23	0	0	0	0	0	0	1	2	4	1	0.069

Table 3. Table with all quantum descriptors values for each corrosion inhibitor

Compound	ED	M	P	C	C1	C2	C12	C13	C14	EH	EL	DIF	DP	V
1	-11.10	185.35	15.054	-0.447	-0.281	-0.447	-0.338	-0.333	-0.338	-8.900	2.852	11.75	0.805	397
2	-1.78	93.13	8.643	0.091	-0.327	-0.685	0.272	-0.145	-0.145	-8.523	0.635	9.16	1.538	139
3	-1.55	129.24	10.540	-0.140	-0.348	-0.140	-0.426	-0.203	-0.207	-9.703	3.435	13.14	1.422	270
4	0.90	169.23	17.724	0.086	-0.263	-0.643	-0.198	-0.048	-0.193	-8.251	0.204	8.46	0.833	257
5	-10.86	185.35	15.173	-0.423	-0.336	-0.423	-0.423	-0.197	-0.198	-9.798	3.614	13.41	1.582	379
6	-3.36	129.24	10.691	-0.294	-0.299	-0.294	-0.374	-0.146	-0.373	-9.248	3.352	12.60	1.203	274
7	2.165	99.18	7.938	-0.091	-0.334	-0.091	-0.369	-0.197	-0.192	-9.726	3.463	13.19	1.485	190
8	-1.570	73.13	5.731	-0.144	-0.338	-0.144	-0.423	-0.198	-0.199	-9.788	3.666	13.45	1.550	151
9	-5.360	101.19	8.454	-0.451	-0.275	-0.451	-0.335	-0.331	-0.335	-8.960	2.890	11.85	0.902	224
10	-3.420	87.00	8.147	-0.140	-0.348	-0.140	-0.426	-0.203	-0.207	-9.701	3.467	13.17	1.424	211
11	1.060	73.13	5.721	-0.093	-0.333	-0.093	-0.372	-0.197	-0.192	-9.802	3.603	13.40	1.469	155
12	-0.770	73.13	5.973	-0.299	-0.305	-0.299	-0.378	-0.155	-0.378	-9.286	3.226	12.51	1.164	163
13	-1.580	59.00	4.565	-0.141	-0.349	-0.141	-0.427	-0.208	-0.204	-9.695	3.586	13.28	1.439	124
14	3.970	59.11	4.530	-0.094	-0.335	-0.094	-0.373	-0.193	-0.197	-9.838	3.619	13.46	1.502	123
15	-5.650	188.33	15.822	-0.810	-0.338	-0.810	-0.186	-0.154	-0.165	-8.450	0.151	8.60	5.844	328
16	-6.150	132.23	10.960	-0.811	-0.336	-0.811	-0.184	-0.155	-0.166	-8.463	0.140	8.60	5.786	211
17	-8.530	104.18	8.083	-0.834	-0.319	-0.834	-0.160	-0.177	-0.178	-8.292	0.314	8.61	5.039	153
18	-7.670	76.12	5.132	-0.944	-0.328	-0.944	-0.154	-0.221	-0.221	-8.536	0.326	8.86	5.566	69
19	-4.450	56.06	3.833	-0.582	-0.356	-0.582	-0.356	-0.082	-0.488	-10.633	1.696	12.33	1.716	92
20	1.630	84.12	6.741	-0.577	-0.350	-0.577	-0.350	-0.129	-0.447	-10.304	1.503	11.81	1.943	157
21	-3.440	70.09	5.073	-0.481	-0.373	-0.481	-0.373	-0.259	-0.186	-10.675	1.839	12.51	1.724	126
22	-3.440	70.09	5.500	-0.580	-0.353	-0.580	-0.357	-0.130	-0.187	-10.339	1.480	11.82	1.950	125
23	0.140	86.09	5.947	-0.598	-0.368	-0.598	-0.368	-0.099	-0.499	-10.562	1.149	11.71	0.938	140

aromatic amine family (**2,4**) showed a preference for extreme values of PC2, *i.e.* a superior and inferior stripes of molecules with large positive and negative (absolute) values of PC2. The thiourea derivatives (**15-18**) can be seen with positive values for PC2 within the 19-30 range forming a stripe in the superior part of the score plot. The stripping profile of the score plot shows that each family interacts with the duplex steel in a different way, with a major distinction for the aromatic amine molecules. It is also interesting to note that the largest aliphatic amines (**1,5,6,9**) are slightly apart from the main group, especially the larger ones like tributyl amine (**1**) and dodecylamine (**5**). This might be an

indication of a different interaction mechanism with amines having large number of carbon atoms.

Ordinary Regression Analysis (OLS)^{14,15}

In order to assess the physical/chemical most relevant descriptors to the prediction of inhibition efficiency we shall present results for the use of the weigh isoestic Langmuir adsorption function, $\ln(M\theta/1-\theta)$, correlated to the previous twenty-five molecular descriptors. To search for the most representative set among all descriptors for inhibition prediction we shall introduce the average error function, defined as the squared deviations sum of the L corrosion inhibitors WILA functions to the fitted results as shown below:

$$\varepsilon_0 = \frac{\sum_{i=1}^L (y_i - [\sum_{j=1}^N a_j x_{ij}]_{all})^2}{L} \quad (1)$$

where the a_j coefficient was obtained through a OLS calculation employing all molecular IC available as the calibration ensemble. Unfortunately such a type of model is well suited to reproduce the calibration data, especially when using a large descriptor set, but is not adequate to predict ICE of molecules that are not present on the calibration ensemble. In order to improve the predictability of our model we shall present a model based on the minimization of an error based on the

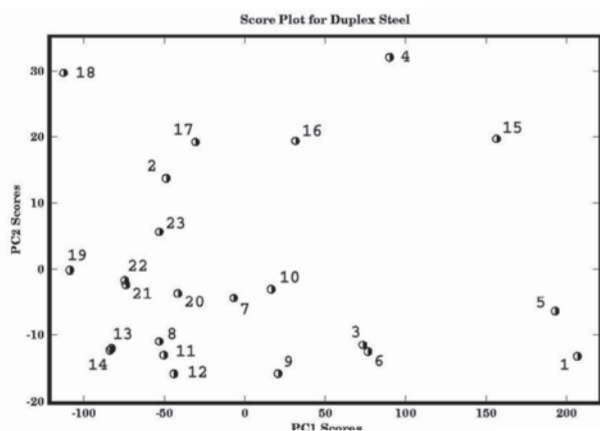


Figure 1. PCA scorings for the molecular descriptor variance. The X axis represents the PC1 while Y axis represents PC2.

cross-validation of a large ensemble of molecules. In this procedure a single or a pair of molecules is excluded from the OLS procedure defining the model, and then the squared deviations is summed for the excluded molecules. In the case of pairs, considering the existence of $L(L-1)/2$ different pairs of possible exclusions the error is summed over all possibilities.

Therefore the first order cross-validation error is shown on equation 2 below, defined by the calculation of a single molecule through an OLS model calibrated with all but this particular inhibitor. The overall error is divided by L , the number of inhibitor corrosion molecules.

$$q_1 = \frac{\sum_{i=1}^L (y_i - [\sum_{j=1}^N a_j x_{ij}]_{\text{not}(i)})^2}{L} \quad (2)$$

Our results rely on a model based on the second-order cross-validation, which is defined by a large number of predictive OLS calculations including all molecule pairs. In this case a particular pair is chosen and the OLS is determined without this information. The response function is then evaluated for these pair of molecules and the errors is summed up to include all possible pairs on the molecular set. In our case, considering the original 23 molecules, there exists 253 different molecular pairs and the second-order cross-validation error sums up all these 253 bootstraps. The average second order cross-validation error is shown below

$$q_2 = \frac{2 \sum_{i=1}^L (y_i - [\sum_{j=1}^N a_j x_{ij}]_{\text{not}(i,j)})^2}{L(L-1)} \quad (3)$$

It is well known the effect of successive new variable additions to the descriptor set on the OLS calculation. Usually the calibration error lowers together with the first order cross-validation error, while the second order cross-validation error shows an irregular behaviour with an initial diminishing followed by a clear divergence with large descriptor number.

In order to seek for the most representative set of descriptors for inhibition efficiency corrosion to be used we developed a simple model based on additions of a single descriptor to a previous set. In this procedure we start with a single descriptor, chosen as the one with the lowest second-order cross-validation error. On a particular iteration the second order cross-validation error is calculated for each descriptor addition, and the model decides to employ the one, which proved to show the smallest second order cross-validation error. For each variable selection the model carries out 253×25

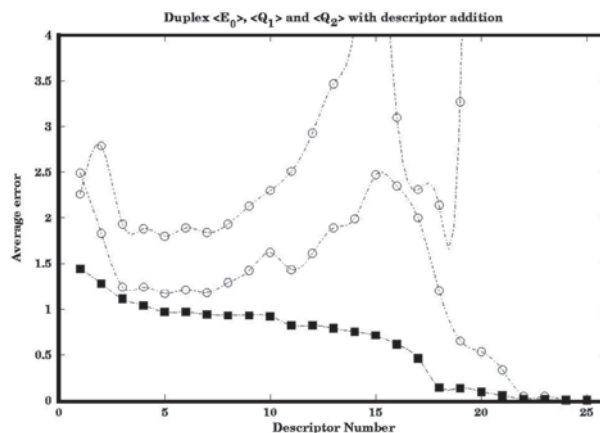


Figure 2. Variation of $\langle E_0 \rangle$, lower curve, $\langle Q_1 \rangle$, intermediate curve, and $\langle Q_2 \rangle$ upper curve for the OLS results with the number of most representative descriptors.

OLS calculations, *i.e.* 6325 bootstrap calculations, choosing the set with smallest predictive error. The procedure is then continued with successive single additions of several descriptors until the original set with 25 descriptors is re-obtained. Figure 2 shows the variation of the calibration, first and second order cross-validation errors plotted against the employed number of descriptors.

Among the most relevant predictive descriptors our model points out the energy gap (DIF), the molecular mass (M), the charge between four close atoms (C14), the dipole and the ramification number (NR) as the five main descriptors. The number of tertiary amines (A3), the number of secondary amines (A2), the average number of carbon atoms within a branch (NCR), the number of phenyl groups (NB), and the number of moles of the inhibitor (N) are among the ten most relevant descriptors for inhibition prediction. On the contrary the number of primary amines (A1), the number of sulfur-carbon groups (NCS), the charge of the three neighbor atoms to the polar group (C13) and the charge of the S,N an C molecular sites (C1) are the descriptor which showed the minor relevance to predict ICE values.

Comparing the most representative descriptors with those previously reported in the literature its noticeable the presence of the HOMO-LUMO energy gap, *i.e.* the dipole moment and the molecular mass in agreement to previous articles present in the literature. Surprisingly we find the branching ratio (NR) and the charge between the three neighbor atoms (C13) as unusual very important predictive descriptors, in contrast to our previous experience that found these descriptors not relevant to calibration purposes. On the contrary the NCS and A1 has been found surprisingly as not relevant

for predicting variables, even though experiences based on calibration optimizations showed these variables to be very important in many cases.

A detailed inspection of Figure 2 shows a minimum in the second order cross-validation error with seven variables. We present this particular model, Y_7 , the descriptors and its respective weights below. It must be pointed out that the symbols M, C14, DIF, D, A2, A3 and NR stands for the standard deviation of these molecular properties values. This particular model showed $R^2 = 0.7961$ and $Q^2 = 0.6755$ with a fairly good results for the correlation of 23 IC molecules.

$$(Y - Y_m) / \sigma_Y = \sum c_i (X_i - X_{im}) / \sigma_{X_i} = 0.358 M - 0.438 C14 + 0.0544 DIF + 0.337 D + 0.1572 A2 + 0.2535 A3 - 0.3764 NR$$

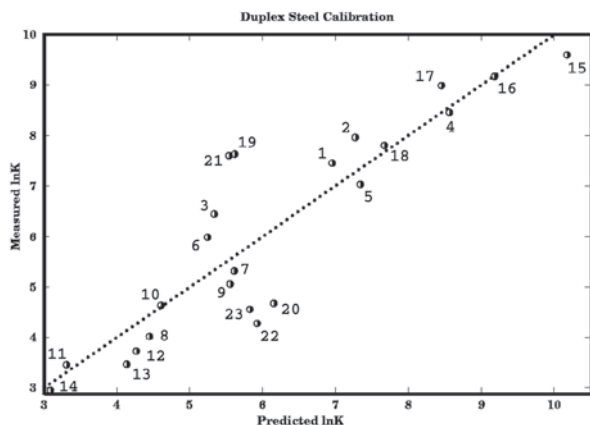


Figure 3. OLS calibration correlation graph for the duplex steel.

Figure 3 shows the calibration results for the Y_7 model while Figure 4 presents the simple cross-validation results. The cross-validation results showed a similar performance to the calibration plot (Figure 3) with a few ICs exhibiting larger errors, e.g. dodecylamine, di-*n*-butylamine and 1,3-diethyl-2-thiourea.

Although the reported values for $R^2 = 0.7961$ and $Q^2 = 0.6755$ are somewhat lower than those observed in traditional biological studies, we must point out that these values results from a major concern regarding predictability on second-order cross-validation procedures. Results with larger values for R^2 and Q^2 could have been found with a different model choice, for instance one based on the maximization of R^2 or Q^2 . Actually we obtained previous results in this study with regression coefficients as large as $R^2 = 0.9323$ and $Q^2 = 0.8037$ with the nine best descriptors by the maximizing the R^2 value. These values, however, are excellent for calibration, while the Y_7 reported values are the best possible for cross-validation procedures of

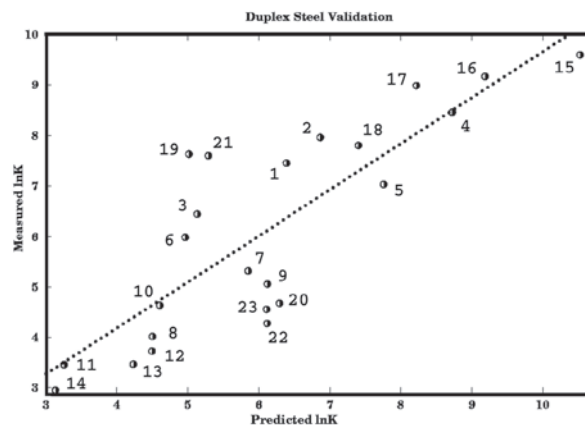


Figure 4. OLS validation correlation graph for the duplex steel.

second order. We expect that our results should be a better way of finding the most representative descriptors for predicting inhibition corrosion efficiencies in the future.

Partial Least-Squares (PLS)^{14,15}

In order to assess the physical/chemical relevant descriptors of the adsorption and corrosion inhibition process we shall present results for the partial least-squares analysis carried out. Picking out two latent components and carrying a PLS for the WILA function as the response property, we obtain a value for R of 0.874. Clearly, much better results were obtained for the WILA function.

Table 4. The PLS loadings for the two main components

Descriptors	PC1 loadings	PC2 loadings
A1	-0.003547	-0.007547
A2	0.010750	0.029840
A3	0.000177	-0.003600
NB	0.003762	0.010850
NC	-0.000446	-0.001446
NCS	0.006380	0.021710
NT	-0.000991	0.003010
NOH	-0.001879	0.000886
NCR	-0.006234	-0.072410
NR	0.000042	-0.006614
N	-0.000328	-0.000414
ED	-0.06186	-0.145000
M	0.646000	0.71800
P	0.060780	0.007262
C	-0.003507	-0.012530
C1	0.002324	0.008491
C2	0.000191	0.001208
C12	0.000639	0.001806
C13	0.000132	-0.000031
C14	-0.005315	-0.018410
EH	0.001164	0.029530
EL	-0.026010	-0.09535
Diff	-0.003766	-0.125000
DP	0.025670	0.09117
V	0.757000	-0.645000

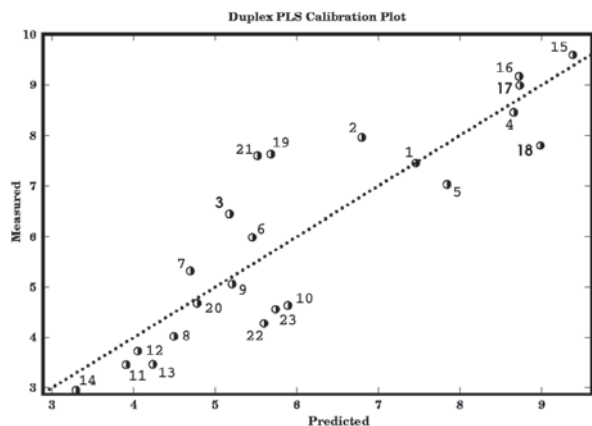


Figure 5. PLS calibration correlation graph.

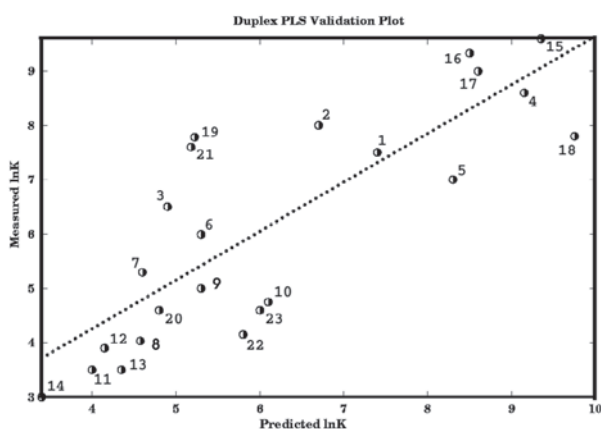


Figure 6. PLS validation correlation graph.

Concerning the validation coefficient correlation (Q) we shall point out the value of 0.816.

The PLS results show similar results and points to an intense mixture with no dominant effect. Table 4 presents the regression coefficient loadings for the two-component PLS obtained with the WILA response function. The largest contributions, in descending order of importance are V, M, ED, DIFF, DP, P, EL, NCR, EH and NCS. Table 4 presents the main components loadings and the remaining from a tenth to a hundredth of the major descriptor contributions. Four (M, DP, NCS and EL) of the nine also appear in the OLS model as the most important descriptors. Figure 5 shows the calibration while Figure 6 shows the validation results for the measured-predicted plot for the WILA function.

It is very informative to investigate the molecular systems with the poorest predictions by PLS methodology. Clearly the alcohol family, 3-butene-1-ol, propargylic alcohol, 2-butene-1-ol and 2-butene-1,4-ol are the ones with the worse prediction results. All these systems have triple bonds conjugated to alcohol groups and our results suggest that a different mechanism

should be taking place within this set. Several authors obtained spectroscopic evidence of a IC polymerization over carbon-steel, however no information yet is available for the duplex steel. Alternatively the ICE for thiourea and its derivatives are much smaller for the duplex steel than it is for other steel types. Similarly other alcohols show very disappointing inhibition corrosion efficiencies, another point favoring a different inhibition corrosion mechanism between amines and alcohols. The referee kindly suggested the use of few alcohols conjugated to double bonds in the inhibitor molecule set in order to investigate this particular structure in the process.

Conclusions

Many chemometric studies have been reported in the literature concerning inhibition corrosion and quantum descriptors. Most of these studies employed six to eight molecules and few (four to six) descriptors. Our ICE results reports were obtained with a large number of molecules and correlate to few relevant descriptors. The final equation obtained with $R^2 = 0.7961$ and $Q^2 = 0.6755$ with 7 descriptors within the OLS methodology points the most relevant descriptors for predicting these ICEs while the R-value of 0.8872 and Q-value of 0.8310 obtained for the PLS procedure points the performance of a three component fit. Both results compare well, especially if we consider how scarce are the studies with such large number of molecules in the literature.

Many activities within this article dealt with descriptor selection through calibration, validation or combination and very reliable results were obtained with selected descriptors. Although the selection of these descriptors might indicate mechanistic information care must be taken when using this information due to its statistical nature. Actually most of the descriptors show strong correlations between themselves and it will always be not clear if the elimination of a single variable should be credited to its specific (chemical) role, to its statistical role or to the correlations with other descriptors. So the indication of descriptors selection sets should be interpreted only as a slight indication of mechanistic value, that should be complemented and cross-checked. This imprecise pattern on the descriptor set comes with a model of very accurate predictive power, and most predictions show errors never superior to 5%.

Quantum and group contribution descriptors were used, and the results show that the use of a mixed

character descriptors offer a well-balanced description between quantum and group contributions. 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 the molecular dipole, the energy gap, the branching ratio (NR) and the charge between the four neighbor atoms (C14) are important predictive descriptors.

Finally we should report that no previous work have been found with QSPR study for the inhibition corrosion efficiency on duplex steel. Therefore, more work is still required toward understanding structure-property correlation on inhibition corrosion studies, particularly concerning the analysis of different steel types. Work is in progress dealing with this point.

Acknowledgments

Two authors (JACP and EH) acknowledge CNPq for financial support, and FAPERJ for grants for computer facilities. This work is part of the PhD Thesis of one of the authors (SPC).

References

1. TrabANELLI, G.; *Inhibitors for Chemical Cleaning Treatments, Working Party (WP) Report 11*, The Institute of Materials: London, 1994, p. 92.
2. JÓIA, C. J. B. M.; BARBOSA, B. C.; *Corrosion 2001*, NACE: Houston, 2001, p.1007.
3. Walker, M. L.; Cassidy, J. M.; Lancaster, K. R.; McCoy, T. H.; *Corrosion 94: The Annual Conference and Corrosion Show Sponsored by NACE International*, Paper 19, 1994.
4. Lukovits, I.; Bakó, I.; Shaban, A.; Kálmán, E.; *Electrochim. Acta* **1998**, *43*, 131.
5. Stoyanova, A. E.; Sokolova E. I.; Raicheva S. N.; *Corros. Sci.* **1997**, *39*, 1595.
6. El-Shafei, A. A.; Moussa, M. N. H.; El-Far, A. A.; *Mater. Chem. Phys.* **2001**, *70*, 175.
7. Quraishi, M. A.; Ansari, F.A.; Jamal, D.; *Mater. Chem. Phys.* **2002**, *77*, 687.
8. Bentiss, F.; Traisnel, M.; Vezin, H.; Lagrené, M.; *Corros. Sci.* **2003**, *45*, 371.
9. Khalil, N.; *Electrochim. Acta* **2003**, *48*, 2635.
10. Stewart, J. J. P.; *MOPAC 6.0, QCPE program*, Bloomington, IN, 1989.
11. Gajewski, J. J.; Gilbert, K. E.; *PCMODEL, Serena Software*, <http://www.serenasoft.com>.
12. Fedders, M.; Ponder, J. W.; *QSPR Program*, 1996. Available at <http://dasher.wustl.edu>, accessed in April 2005.
13. *Unscrambler for Windows, Version 6.11*; CAMO ASA: Trondheim, Norway, 1977.
14. Miller, J. N.; Miller, J. C.; *Statistics and Chemometric Chemistry*, 4th ed., Prentice Hall: London, 2000.
15. Brereton, R. G.; *Chemometrics – Data Analysis for the Chemical Plant*, John Wiley & Sons: New York, 2003.

Received: November 25, 2005

Published on the web: August 11, 2006