Brazilian Journal
of Chemical
Engineering

ISSN 0104-6632
Printed in Brazil
www.abeq.org.br/bjche

Vol. 24, No. 04, pp. 547 - 559, October - December, 2007

PREDICTIVE QSPR ANALYSIS OF CORROSION INHIBITORS FOR SUPER 13% Cr STEEL IN HYDROCHLORIC ACID

S. P. Cardoso[1,2], J. A. C. P. Gomes[2], L. E. P. Borges[3] and E. Hollauer[4*]

[1]Centro Federal de Educação Tecnológica de Química de Nilópolis, CEFET Química,
Rua Lúcio Tavares 1045, Centro, Nilópolis, Rio de Janeiro - RJ, Brazil.
[2]Laboratório 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.
[3]Departamento de Engenharia Química, Instituto Militar de Engenharia, Praça General Tibúrcio 80,
Praia Vermelha, CEP 22290-270, Rio de Janeiro - RJ, Brazil
[4]Programa de Pós-Graduação em Química, Instituto de Química, Universidade Federal Fluminense,
Outeiro de S. João Batista s/num, CEP 24210-130, Centro, Niterói - RJ, Brazil
E-mail: edu@kabir.gqt.uff.br

(Received: March 8, 2006 ; Accepted: April 10, 2007)

Abstract - An experimental and theoretical study on the inhibition corrosion efficiencies of twenty three compounds in hydrochloric acid (15% w/v) on 13% Cr modified stainless steel (martensitic) has been carried out. This inhibitor set includes amines, thiourea derivatives and acetylenic alcohols. Experimental weight losses at 60°C were correlated with group and quantum AM1 descriptors obtained from QSPR analysis. Such data, for a large set of molecules, offer a unique opportunity for searching for correlations between inhibition corrosion efficiency and molecular properties. Calculations based on three different statistical methodologies were carried out. The first method, using calibration procedures, employs an ordinary least squares (OLS) methodology with a simple descriptor selection based on R^2 values. From this procedure, we obtained a model, Y_{15}, having a R^2 value of 0,979 and a Q^2 value of 0.786. The second method employs a descriptor selection based on the second-order cross-validation OLS procedure (SOCV-OLS). In this process, the variables are chosen according to their ability to predict molecular inhibition efficiencies. The best model obtained using this methodology, Q_5, had R^2 and Q^2 values of 0.859 and 0.785, respectively. The third method, based on regular partial least squares (PLS), resulted in R^2 and Q^2 values of 0.859 and 0.754, respectively. All calculations were carried out for the weight isoesteric Langmuir adsorption function (WILA function), $\ln(\theta M/(1-\theta))$ or $\ln K_{ads}$. A careful comparison between the calibration and the cross-validation descriptor selection indicated that they had very few descriptors in common. This article presents some key equations and the most relevant descriptors. We are unaware of any similar QSPR study on super 13% Cr stainless steel in the literature.

Keywords: Corrosion inhibitors; QSPR; Super-13 steel; Stainless steel; AM1.

INTRODUCTION

Corrosion inhibitors have been widely used in stimulation operations in petroleum wells. In these operations, hydrochloric acid solutions (15% w/v) at temperatures up to 60°C are employed to remove iron oxides and carbonated minerals [Mack, 1995]. In such an aggressive medium, the use of corrosion inhibitors (CIs), whether used singularly or as mixtures of different CI´s [Trabanelli, 1984; Santana, 2002; Mack, 1995] is mandatory. Among the well-known CIs employed in hydrochloric acid media are amines, amides, nitriles, imidazolines, triazoles, pyridine, quinoline derivatives, thiourea derivatives, thiosemicarbazide and thiocyanates, among others.

*To whom correspondence should be addressed

Modified 13% Cr (martensitic) stainless steel, 13Cr-5Ni-2Mo, hereafter referred simply as super 13, and many other corrosion resistant alloys (CRAs) containing high chromium and molybdenium concentrations [Walker, 1994; Cassidy, 1995; Bergman, 1954] are intensively used in the petroleum industry. The use of super 13 steel offers superior corrosion resistance for mildly acidic environments, combined with high strength and good low-temperature toughness. It has particular advantages in HCl and HF media. Under extreme conditions, this alloy has an excellent corrosion resistance for oil/gas co-production fluids due to its extended pH range of passivation. Some authors have claimed that the super 13 stainless steel is more difficult to protect against localized attack than the 22% Cr full-duplex alloy. However, the general corrosion and pitting of CRAs can be inhibited during the stimulation process. Extensive work is in progress to develop higher inhibition corrosion efficiencies (ICEs) and to find optimal IC mixtures for use with alloys having a high chromium concentration in hydrochloric acid solutions [Cassidy, 1995; Bergman, 1954] at high temperatures.

The use of chemometric analysis in inhibition corrosion studies is not new. Early attempts were made in the mid-1950s employing Huckel calculations. For a large number of molecules, Bergman [Bergman, 1954] obtained excellent correlations between standard reduction potentials with the lumo and homo energies. During the 1960s, Donahue [Donahue, 1966; Vosta, 1971] employed ab initio calculations to establish certain correlations. Vosta and collaborators studied the correlation of eight gamma-substituted pyridine N-oxides with several ab initio quantum descriptors.

In an earlier series of papers, Growcock et al. [Growcock, 1989, 1989], elaborated a general multivariate analysis for chemisorption and corrosion inhibition. They employed such physico chemical descriptors as homo and lumo energies, logP and the Hammett and Taft constants, in an investigation of the inhibition of corrosion of mild steel by derivatives of cinnamaldehyde. Interestingly, this was the first work to recognise the importance of the Langmuir constant for obtaining the best linear relationships.

Using the CNDO/2 methodology, Abdul-Ahad [Abdul-Ahad, 1989], extended this work to aniline derivatives. Dupin et al. [Dupin, 1980] 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. This is the only work, besides ours, in which more than ten molecules were involved.

In a set of univariate experiments Sastri et al. [Sastri, 1997], correlated the ICEs of several methyl substituted pyridines and substituted ethane derivatives with MNDO descriptors. Studying dibenzyl sulfoxide adsorption on iron in the mid-1990s, Kutej et al. [Kutej, 1995], studying dibenzyl sulfoxide adsorption on iron in the mid-1990s, employed ab initio calculations to recognise the attachment points of CIs on the iron surface. Öğretir et al. [Ogretir, 1999] employed several AM1, PM3, MINDO/3 and MNDO descriptors in attempts to correlate the efficiency of pyridine-based inhibitors for mild steel. Several descriptors showed excellent univariate correlations. However, Sastri et al.[Sastri, 1997], did not use multivariate methods. A related article [Bereket, 2002], published recently, was concerned with the corrosion inhibition of imidazole derivatives for iron exposed to acidic media.

Lukovits et al. [Lukovits, 1998] employed a polynomial regression analysis for the Langmuir adsorption constant for a set of seven thiourea derivatives and obtained good correlation values. Bentiss et al. [Bentiss, 2003] successfully correlated ICEs, determined through charge transfer resistance, of six triazole and oxadiazole derivatives, with AM1 quantum descriptors: R-values of 0.91-0.96 were obtained. Recently Khalil [Khalil, 2003] extended this study and correlated the inhibition by twelve thiosemicarbazone and thiosemicarbazide derivatives with five quantum MNDO/PM3 descriptors. All these previous studies were merely concerned with carbon steel. The lack of data on other types of steel in the literature is apparent and needs to be addressed. Furthermore, the ability to use and transfer data between various steel types would be a most useful and significant advance.

Recently, the field has experienced a revolutionary change with several molecular modeling techniques being used to design new CIs. Articles by Wang et al. [Wang, 1999] and Pradip et al. [Pradip, 2002] are among the early contributions to this revolution. Wang calibrated the ICEs of three imidazole derivatives and used these values to correctly predict the ICEs of another three imidazole derivatives. Pradip et al. elaborated a procedure for the evaluation of the interaction energies of

surfactants used in industrial cleaning. The advantage of their proposal is that experimental data are not necessary. Affrosman et al. [Affrosman, 2001] predicted, on the basis of experimental and computational experience, that the inhibition time would be maximal around C_{10} for physical adsorption on titanium surfaces.

Despite the intense empirical work searching for new commercial inhibitors, very few articles address ICE analysis. For the quantum statistical property relationship (QSPR) studies, no more than fifteen CI molecular systems and descriptors were usually employed. Two major exceptions are the Dupin et al. [Dupin, 1980] work with forty-two CIs and our previous study [Cardoso, 2006].

In the work we describe herein, a detailed experimental and theoretical investigation was carried out on the twenty three different CI compounds, namely amines, thiourea derivatives and acetylenic alcohols, in order to correlate experimental weight loss and the ICE for super 13 steel in hydrochloric acid (15% w/v) solutions at 60°C using quantum and group contribution molecular parameters.

A previous study [Cardoso, 2006] by our group analysed the same set and their ICEs for the 22% Cr stainless steel surface. Principal component analysis (PCA), second-order cross-validation ordinary least-squares analysis (SOCV-OLS) and partial least-squares analysis (PLS) were carried out with excellent results. Correlations obtained for OLS were typically in the range of 0.99-0.92 for R^2, while a simple PLS with three components produced values of 0.86-0.88 R^2. The SOCV-OLS results showed that the best descriptors employed to predict WILA functions were rather different from the best descriptor set selected to describe ICE through the OLS models. All molecular systems showed acceptable fits for these models and usually less than ten descriptors were actually required to obtain good results in the calibration and validation steps. We believe that this is the first study employing the same CI set with different types of steel.

Since the descriptors used here are the same as the ones employed in our previous study, no discussion will be presented for the principal component analysis (PCA). In this study, we shall present experimental details, followed by ordinary least-squares analysis (OLS), the predictive second-order cross-validation (SOCV-OLS) and our partial least squares (PLS) results. It must be pointed that the systematic analysis of data for large set of molecules offers a unique opportunity for determining predictability of ICEs based on intrinsic molecular properties.

EXPERIMENTAL

All corrosion inhibition data were obtained through weight loss experiments based on rectangular steel specimens with 2 × 0.5 × 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. Formaldehyde was employed to minimise 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.

Table 1 lists the twenty three inhibitors employed in our study and their ICE for super 13. Since all experiments were carried out with the same inhibitor weights, a different procedure was more adequate for the QSPR studies. Since ΔG_{ads} is a thermodynamic property, which shows a strong correlation with the energy, volume and with the inhibition polarisability, we used the weight isoesteric Langmuir adsorption function, the WILA function, defined as $\ln(\theta M/(1-\theta)) = \ln K_{ads}$, as the response property in the QSPR calculations.

Tributylamine, aniline and such 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, sec-butylamine, propylamine, diethylamine and n-butylamine, are among the less efficient inhibitors tested for super 13. It is important to point out that the greater corrosion rates for 22% Cr are higher than these for 13% Cr steel in hydrochloric acid media should be credited to the reaction of the chloride ions at the molecular interface.

Table 1: The twenty-three inhibitors employed, the ICE (θ) values, and the WILA function, ln(θM/(1-θ)), values.

	Compound	θ	ln(θM/1-θ)
1	Tributylamine	0.9776	8.99
2	Aniline	0.9776	8.31
3	n-octylamine	0.8862	6.91
4	Diphenylamine	0.9208	7.58
5	Dodecylamine	0.8841	7.25
6	di-n-butylamine	0.8679	6.74
7	Cyclohexylamine	0.7819	5.87
8	n-butylamine	0.6912	5.09
9	Triethylamine	0.7467	5.69
10	Hexylamine	0.8314	6.06
11	Sec-butylamine	0.6754	5.02
12	Diethylamine	0.6876	5.08
13	Propylamine	0.6818	4.84
14	Isopropylamine	0.6361	4.64
15	1,3-Dibutyl-2-thiourea	0.9730	8.82
16	1,3-Diethyl-2-thiourea	0.9636	8.16
17	1,3-Dimethyl-2-thiourea	0.9596	7.81
18	Thiourea	0.9004	6.53
19	Propargyl alcohol	0.9580	7.15
20	2-pentyn-1-ol	0.8767	6.39
21	3-butyn-1-ol	0.8963	6.41
22	2-butyn-1-ol	0.7407	5.29
23	2-butyne-1,4-diol	0.6859	5.23

Theoretical Calculations

All calculations employed the Austin Model 1 (AM1) methodology as coded in Mopac 6.0[25] for most of the descriptors, Pcmodel [Serena Software, 2005] for the volume calculations, and the QSPR programme, coded by Fedders and co-workers, and obtained from the internet [Fedders, 2005]. This last programme, has been adapted in our laboratory for SOCV-OLS analysis [Fedders, 2005]. In this work, we employed the same descriptor set as that used in our previous study [Cardoso, 2006], and this includes the following group contribution descriptors: A1 is the number of RNH_2 groups; A2 is the number of R_1R_2NH groups; A3 is the number of $R_1R_2R_3N$ groups; NB is the number of phenyl groups (structurally isolated); 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 and NR is the branching number, while N is the number of moles of inhibitor present in the vessel.

For the quantum descriptors, we employed the same descriptors as those used in our previous study [Cardoso, 2006]: ED is the dimerisation energy; M is the CI molecular mass; P is the polarisability 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 its absence the polar group charge); C12 is the charge of two atoms of the polar group; C13 is the charge of the three atoms of the polar group; C14 is the charge of the four adjacent atoms to the polar group; EH is the homo energy; EL is the lumo energy; Dif is the difference EL-EH; DP is the dipole and V is the calculated volume. The quantum descriptors add up to fourteen descriptors, while the whole set employs twenty five molecular descriptors. The values for the overall quantum and group contribution descriptors have been already published [Cardoso, 2006] and will not be reported here.

The choice of this descriptor set was based on several articles found in the literature, which indicate the most usual descriptors (dipole, polarisability, lumo and homo energy, etc). We added several others based on common sense and evaluation ease of in order to explore new possibilities.

Ordinary Regression Analysis (OLS)

In order to assess the relevant physico-chemical

relevant descriptors of the adsorption and corrosion inhibition process, we shall present results for a naive, although informative, ordinary least-squares analysis coupled to a simple descriptor selection procedure. All OLS calculations were carried out with centered and self-scaled descriptors and the response function as the isoesteric weight Langmuir adsorption function, $\ln(M\theta/(1-\theta))$, the WILA function. All twenty three molecules were employed.

In order to describe the elimination descriptor algorithm, calculations started with simple OLS and all the twenty five descriptors previously described. In order to identify the most representative descriptors, we carried out an elimination algorithm by which the descriptor with the least absolute contribution to the response functions is eliminated from the descriptor set. Then a new OLS carried out with the remaining ones and the new descriptor was successively eliminated until a minimum number of descriptors was achieved.

The difference between homo and lumo energies (Dif), number of cycles (NC), lumo energy (EL), branching number (NR), number of benzyl groups (NB), dimerisation energy (ED) and charge of the four polar atoms (C14) are the early eliminated descriptors. By the major contribution criteria, in descending order, the most important contributing descriptors were polarisability (P), volume (V), ring charge (C2), total charge (C), number of CS bonds (NCS), charge between two atoms (C12), homo energy (EH), number of alchool groups (NOH), number of secondary (A2) and primary (A1) amines and molecular dipole (DP).

In Figure 1 is shown the variation of R^2 and Q^2 during the variable elimination process. In the Figure, it is clear, that with a large number of descriptors there is an overfit pattern with R^2 and Q^2 values very close to unity. Due to the successive eliminations, R^2 decreases slowly while Q^2 shows an irregular behaviour. Therefore our choice for the best model should have not less than fifteen descriptors in order to obtain reproducible validation coefficients. The final equation for $\ln K_{ads}$ with fifteen descriptors, hereafter referred to Y_{15}; show the OLS descriptors and its loadings below. For the sake of simplicity, the variables and regression coefficients are related to centered and self-scaled values. This model had a regression coefficient of $R^2 = 0.979$ and $Q^2 = 0.786$, an excellent result for the correlation of twenty three response inhibition corrosion efficiencies.

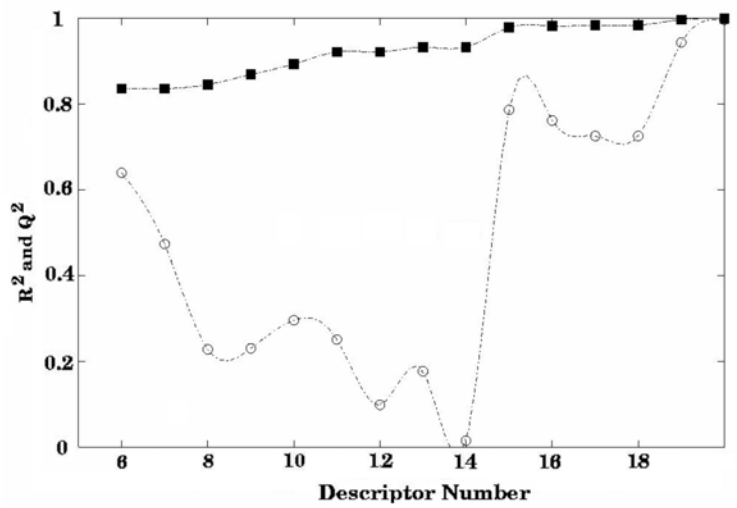

Figure 1: Variations of R^2 (-■-) and Q^2 (-○-) with the number of descriptors employed.

$$Y_{15} = -0.93\,N - 7.64\,P + 6.74\,C - 2.27\,C12 + \\ + 0.94\,C13 - 1.06\,C1 - 7.22\,C2 - 2.17\,EH - \\ -1.17\,DP + 7.45\,V - 1.15\,A1 - 1.81\,A2 + 7.12\,NCS - \\ - 1.869\,NROH - 1.085\,NCR \qquad (1)$$

Figure 2 shows the measured-predicted plot for the calibration procedure while in Figure 3 is shown the measured-predicted plot for the validation procedure. An analysis of Figure 2 shows excellent calibration behaviour, with most of the molecular systems showing acceptable deviations. In Figure 3, however, are shown some outliers: 3-butyn-1-ol, thiourea and propargyl alcohol are underestimated, while 2-butyne-1,4-diol, 2-butyn-1-ol, dodecylamine, isopropylamine and n-butylamine are overestimated in this model. In conclusion, alcohols showed erratic behaviour with large errors. It is important to point out that the identification of related inhibitors showing erratic behaviour has not been reported previously due to the small number of molecules employed in previous

calculations. At this point many possible causes can be suggested like a different adsorption mechanism or a descriptor inadequacy, however, due to the scarcity data on this particular steel surface, no conclusions can be drawn at this point.

A similar analysis on protonated amines has been carried out. The results invariably showed poorer regression coefficients, R^2, and validation coefficients, Q^2. Concerning our space limitations, we shall not present the numerical descriptor table or any graphical results in this article.

An analysis of the descriptor elimination order indicates significant differences from the previously reported analysis for 22% Cr stainless steel [Cardoso, 2006]. Volume and polarisability, are reported here as important descriptors, whereas they were of little importance in the inhibition of 22% Cr stainless steel corrosion. On the contrary, the lumo energy, which indicates the ease of reduction of the inhibitor, was a very important descriptor in duplex stainless steel, but was one of the early eliminated descriptors for super 13 steel. As there are significant differences between these results for the two types of steel, it follows that different corrosion protection mechanisms could arise. For duplex steel, the dimerisation energy was considered an important descriptor, while for super 13 it was eliminated for being a non-representative descriptor. The presence of polarisability and volume among the major contribution descriptors suggests that physical adsorption should be an important physical process taking place on the super 13 steel surface. That was expected and had been reported by several authors in the field. We should always remember that adsorption energies for commertial inhibitors usually have values placed between physical and chemical typical adsorptions energies and, therefore, most of these processes have been classified as strong physical adsorption.

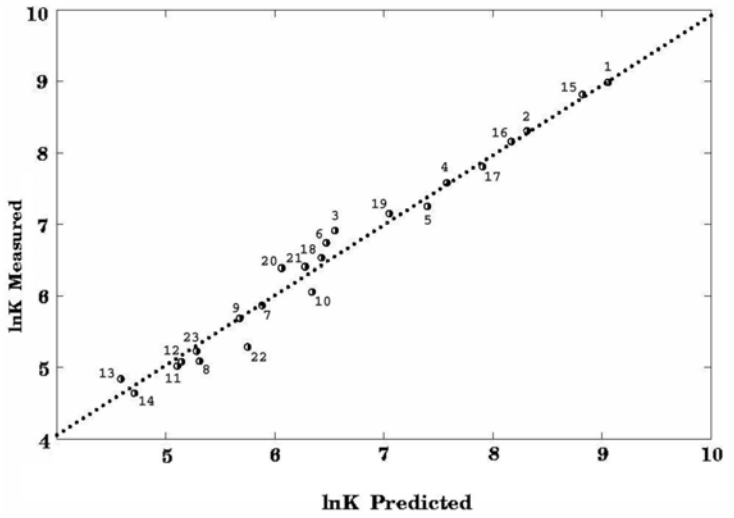

Figure 2: Predicted-measured calibration plot for the Y_{15} OLS model.

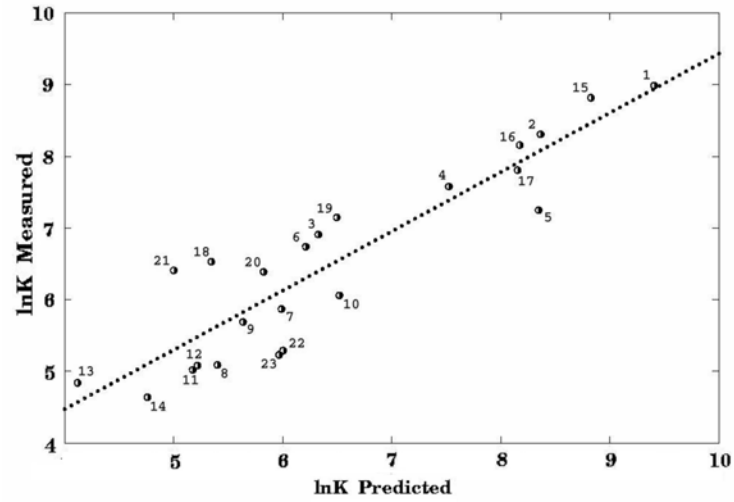

Figure 3: Predicted-measured validation plot for the Y_{15} OLS model.

Second-Order Cross-Validation Analysis (SOCV-OLS)

The selection procedure employed in the previous section, although successful in achieving of valid correlations, is based on calibration procedures. In this section, we investigate which of the descriptors are the most appropriate to predict ICE based on chemometric methods. To search for the most representative set of all descriptors for inhibition prediction, we shall introduce the average error function, defined as the sum of the squared deviations of the L WILA functions fitted result to the result as shown below:

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

where the a_j coefficient was obtained through an OLS calculation employing all molecular CI available as the calibration ensemble. Such a model is well-suited to reproduce the calibration data, especially when using a large molecular set, but is not adequate in predicting or validating molecular ICEs. In order to improve the predictability of our model, we devised a model based on the minimization of the cross-validation error of a large molecular ensemble. In this procedure a single, or a pair of molecules, is excluded from the OLS procedure defining the model, and then the ICE and its squared deviation is summed for the excluded molecules. In the case of pairs, the model considers the existence of L(L-1)/2 different pairs of possible exclusions and the error is summed over all possibilities.

The first order cross-validation error, shown in Equation 3 below, is 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 CI molecules.

$$q_1 = \frac{\sum_{i=1}^{L} (y_i - [\sum_{j=1}^{N} a_j x_{ij}]_{not(i)})^2}{L} \quad (3)$$

Our results rely on a model based on the second-order cross-validation, Equation 3, which is summed for all pairs of molecules, which are excluded from the original molecular set in each calculation. In our case, considering the original twenty three molecules, there exist 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

$$q_2 = \frac{2\sum_{i=1}^{L} (y_i - [\sum_{j=1}^{N} a_j x_{ij}]_{not(i,j)})^2}{L(L-1)} \quad (4)$$

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

In order to determine the most representative set of descriptors we developed a simple model based on single descriptor addition to a previous set. In this procedure, we start with the best correlated descriptor with the response function, and then a single descriptor is added, each iteration, to the previous descriptor list. In a particular iteration, the second order cross-validation error is calculated for each descriptor addition, and the model employs the one which shows the smallest second order cross-validation error. For each variable selection, the model carries out 253 × 25 OLS calculations, i.e. 6325 bootstrap calculations, choosing the set with the smallest predictive error. The procedure is then continued with successive single additions of several descriptors until the whole set of twenty five descriptors is obtained. Figure 4 shows the variation of calibration, the first and second order cross-validation errors plotted against the number of descriptors.

In Figure 4 the behaviour of the calibration, the first and second order cross-validation errors are shown. The result is expected. The calibration error ε_0 decreases with the increasing number of descriptors. Similarly, but somehow irregularly, the first order cross-validation shows a slight decrease with the increase in the number of descriptors. Alternatively the second-order cross-validation error tends to show a large increase with the number of descriptor. This result is due to the great flexibility of functions with large number of descriptors provides to fit the calibration ensemble. Usually this pattern presents optimal results for the correlation with the lack of sensitivity to average values, as in the predictive procedure.

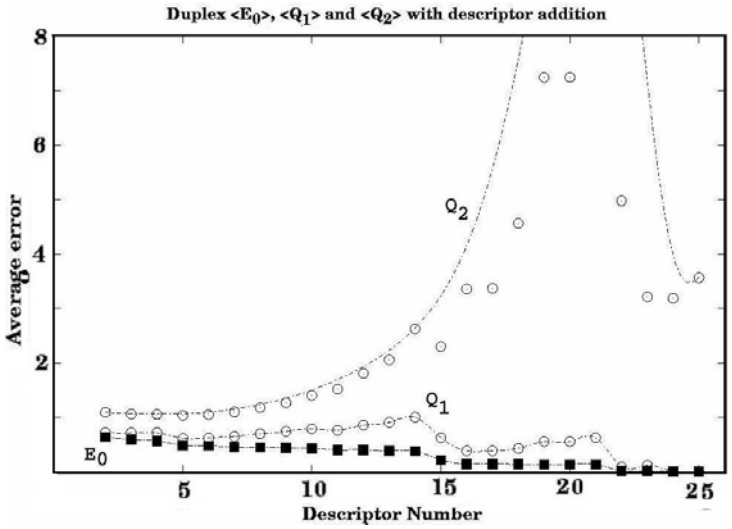

Figure 4: The evolution of the calibration, first and second-order cross-validation error.

The difference from the calibration descriptor set determined in the previous section is noticeable. Of the five most important descriptors associated with the second-order cross-validation procedure, we find only a single common descriptor. Of the ten best predictive and correlation descriptors, are only three common descriptors. Therefore we might conclude that the best predictive descriptors should be obtained in specific procedures with an emphasis on variable selection based on prediction evaluations.

The five best predictive descriptors are molecular mass (M), charge between two neighbouring groups (C12), total number of triple bonds (NT), polarisability (P) and the number of hydroxyl groups (NOH). Surprisingly the charge C12 and the number of hydroxyl groups were irrelevant descriptors in previous studies. The lumo and homo energies and the energy difference, although recognised as very important in many previous studies in the literature, were not of any special value when applied to super 13 steel. Actually this information was confirmed when the selection criteria employed calibration in the OLS previous study. Are insignificant variables the average size of each branch (NCR), charge of the adsorption site (C1), volume (V) and the number of fused aromatic rings (NB). The lumo and homo energies and the energy difference are the next lowest relevant descriptors.

At this point it is important to compare the differences between super 13 steel and duplex stainless steel [Cardoso, 2006]. For the SOCV-OLS descriptor selection the lumo energy, the lumo-homo difference, and the homo energies were significant descriptors for duplex steel but were of little importance for super 13 steel. Similarly, the dimerisation energy was found to be very important for full-duplex steel, but not for super 13 steel. Further differences were found for amines: the number of secondary amine groups was the most important descriptor for super 13 steel, while all amines (A1, A2 and A3) had been considered irrelevant in the previous study. This information shows that similar molecules can provide correlations that are dependent on steel type. However, similar factors can play important roles for both steel types, e.g. the number of alcohol groups and the number of CS bonds are among the most important descriptors for both duplex and super 13 stainless steel. These results suggest that the CS and alcohol groups form strong attachment to the super 13 steel surface. We are not aware of any other study reporting a similar conclusion.

The model with the smallest second order cross-validation error, Q_5, was obtained using five descriptors and is shown below.

$$\frac{(Y-Y_m)}{\sigma_Y} = \sum \frac{(X_i - X_i^m)}{\sigma_{Xi}} = 1,52M - 0,79P + 0,53C12 + 0,80N - 0,66NOH \quad (5)$$

In this formula, the symbols M, C12, N, P, and NOH stand for the standard deviation values of these molecular properties. This particular model showed values of R^2 =0.859 and Q^2 =0.785 with a good regression for all twenty three molecules. Figure 5 contains the calibration results, while Figure 6 presents the cross-validation predicted-measured plots.

Although the R^2 and Q^2 values are acceptable for

corrosion studies, these are somewhat lower than the results usually observed in traditional biological studies. However, since our ICE approach is entirely based on predictive procedures and biological studies rely on calibration procedures, they are strictly not comparable.

The results of the cross-validation for model Q_5 show several outlying molecules, all previously reported in the analysis of the calibration plot. Inhibitors such as hexylamine, propargyl alcohol, 1,3-dimethyl-2-thiourea, 1,3-diethyl-2-thiourea, are undervalued, in contrast to the overestimation of 2-butyn-1-ol, isopropylamine, sec-butylamine and dodecylamine in model Q_5.

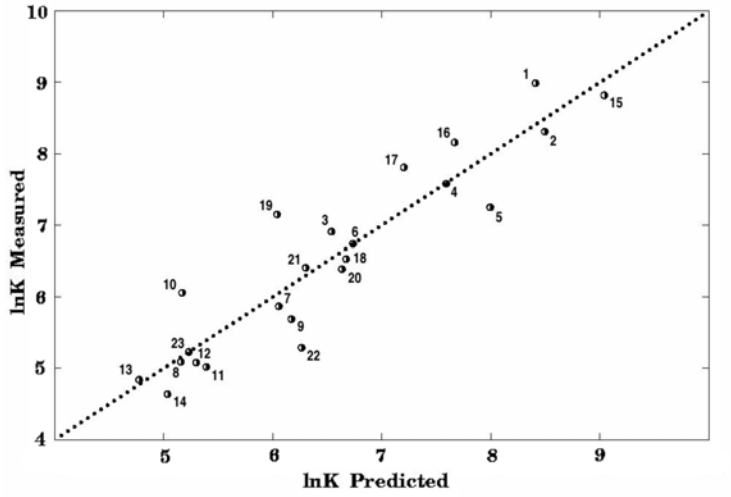

Figure 5: Predicted-measured calibration plot for the Q_5 prediction OLS model.

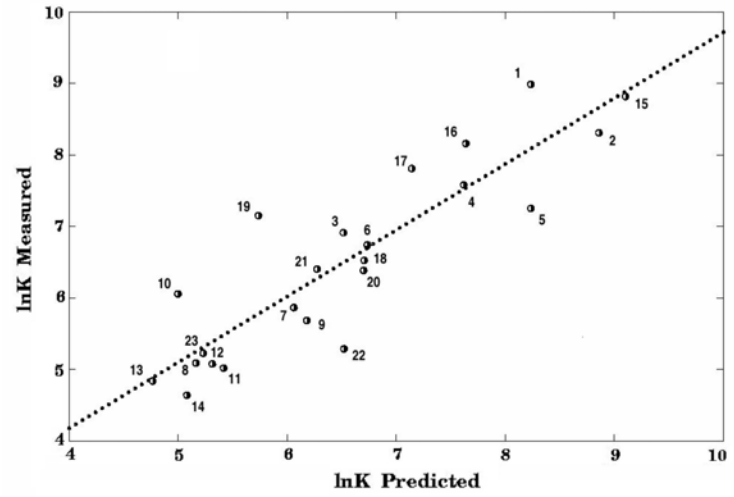

Figure 6: Predicted-measured first-order validation plot for the Q_5 prediction OLS model.

Partial Least-Squares (PLS)

In order to assess the most relevant physical/chemical descriptors, we present the results of a partial least-squares analysis. Choosing the two main components and carrying out a PLS for the two sets of response properties, the ICE and the WILA function, $\ln(M\theta/(1-\theta))$, we obtained 0.759 and 0.854 for the regression coefficient (R), respectively. Clearly the results are better for the WILA function. For the predictive correlation (Q), values of 0.608 and 0.754 respectively for the ICE and the WILA function were obtained.

The contributions of the descriptors to the WILA function are, in descending order of importance: M, P, C1, C2, V, EH, A2, ED, while NCR, Dif, C14, A1, NT and A3 were insignificant. Remarkably, primary and tertiary amines were among the least significant descriptors, while the secondary amine A2 showed up as one of the main descriptors. This strongly suggests that the most significant attachment occured via secondary amines, and not with primary or tertiary amines. There is a three orders of magnitude difference between the major and the minor contributors. This implies that no descriptor group can be indicated a priori. Actually,

the difference between the homo and lumo energies, Dif, made only a minor contribution for super 13 steel, while from many other studies, it is one of the most important descriptors. In Figure 7 is shown in the measured-predicted correlation plot, while Figure 8 contains the measured-predicted validation results.

The analysis of the calibration and validation plots shows that the molecules: aniline, tri-butylamine, 1,3-dibutyl-2-thiourea, propargylic alcohol, 2-buthyl-1-ol and 2-butyne-1,4-diol were poorly represented by the treatment. Interestingly in our previous study on 22% Cr stainless steel [Cardoso, 2006], tri-butylamine, triethylamine and 1,3-dibutyl-2-thiourea were well describded, while the other molecules were also found to be outliers. In Table 2 are shown the regression coefficients obtained with the best PLS model.

Table 2: The regression coefficient for the WILA function.

Descriptor	Regression coefficient
A1	-2.688×10^{-2}
A2	0.115
A3	4.530×10^{-2}
NB	8.762×10^{-2}
NC	-1.305×10^{-2}
NCS	0.102
NT	-4.518×10^{-2}
NOH	-5.172×10^{-2}
NCR	-1.099×10^{-3}
NR	4.996×10^{-2}
N	-6.338×10^{-2}
ED	-0.114
M	0.184
P	0.180
C	-6.838×10^{-2}
C12	8.092×10^{-2}
C13	-4.783×10^{-2}
C14	-1.130×10^{-2}
C1	-0.167
C2	-0.152
EH	-0.122
EL	-7.838×10^{-2}
DIF	-7.004×10^{-3}
DP	0.109
V	0.141

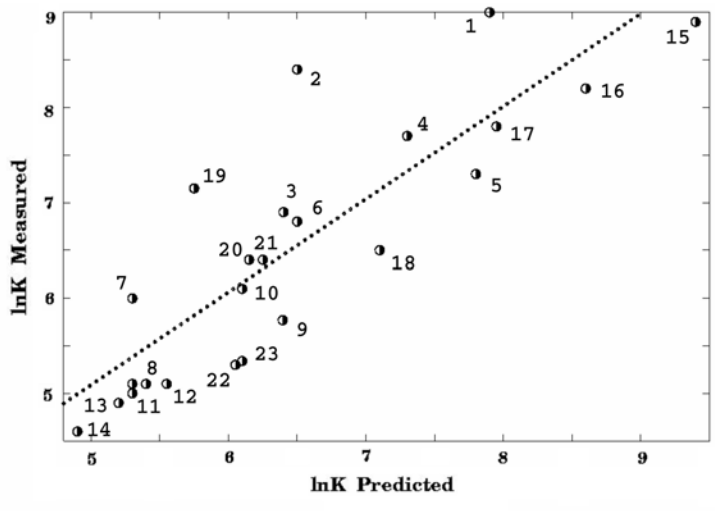

Figure 7: PLS calibration correlation graph.

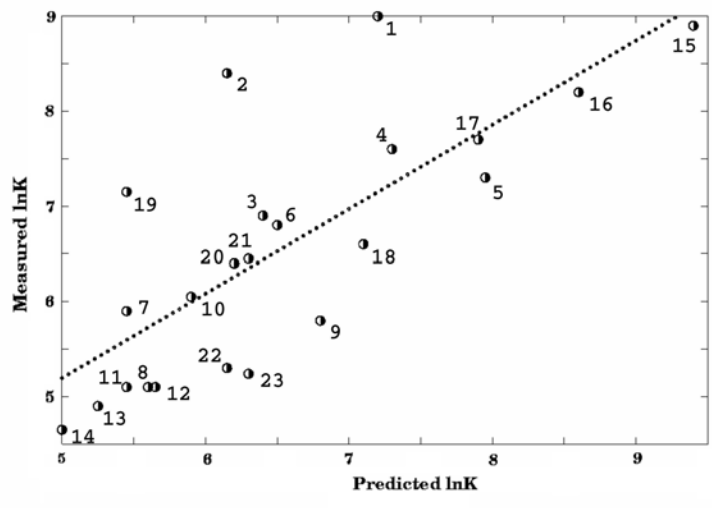

Figure 8: PLS validation correlation graph.

CONCLUSIONS

Many chemometric studies on inhibition corrosion have been reported in the literature. In this particular study we present correlation of a medium size number of molecules (twenty-three) with group contribution and quantum descriptors. Different response were compared for corrosion purposes based on ICE and the WILA functions, with advantages overall work favouring the WILA function.

Four different models were tested based on OLS, descriptor selection based on regular OLS, descriptor selection based on second order cross-validation OLS and the traditional PLS methodology. Two different algorithms regarding descriptor selection were used as well. The first based on the magnitude of the weight coefficients and the second based on the smallest error in the second order cross-validation OLS. Expectations exists for the optimal adequacy of the first method toward the calibration and for the second method toward the model predictability.

Two particular models were selected. The best calibration model showed the best result as Y_{15}, using fifteen descriptors. The R^2 and Q^2 values, respectively of 0.979 and 0.786, compare well with results previously reported in the literature. The second model, which variables was selected by SOCV-OLS, was the Q_5 and showed R^2 and Q^2 values of 0.859 and 0.785, respectively. Finally the PLS method generated R^2 and Q^2 values of 0.854 and 0.754. The PLS results, although using all twenty five descriptors, show similar results to the SOCV-OLS results.

An interesting exercise was the selection of the best descriptors, employing the minimization of the second-order cross-validation error. The selected descriptor set had little similarity with the descriptors selected by OLS, based on calibration procedures. This is an important result and points that an specific search for descriptors, depends on whether the study is focused on calibration or validation. Quantum and group contribution descriptors were used, and the results show that the use of mixed-character descriptors offer a well-balanced descriptor set. Among the descriptors with the greatest contributions, we point out polarisability, molecular dipole, molecular mass, volume, charges, secondary amines which showed relevance in both models employed (OLS and PLS). The descriptors showing unimportant contributions were NCR, surprisingly the homo/lumo difference, charge C14, branching number (NR), number of primary and the number of tertiary amines. Noticeably, the difference between the homo and lumo energies and the number of primary and tertiary amine group descriptors were indicated in some previous corrosion studies as being very important, while they were found here to be only of secondary importance.

Finally, we point out that no previous application of the QSPR methodology has been reported for super 13 steel. Further work is imperative to gain a better general understanding of structure-property correlation in corrosion inhibition. With this in

mind, we are now carrying out studies on carbon-steel. We are including the same descriptor set and we shall be searching for a fundamental corrosion inhibition equation, which would systematise all the collected data.

ACKNOWLEDGMENTS

One of the authors (JACP) acknowledges financial support received from CNPq. EH acknowledges FAPERJ for financing computer facilities and CNPq for research fellowship. This work is part of the doctoral dissertation thesis of one of the authors (SPC).

REFERENCES

Abdul-Ahad, P. G., and Al-Madfai, S. H. F., Elucidation of Corrosion Inhibition Mechanism by Means of Calculated Electronic Indexes, Corrosion, 45, No. 12, 979 (1989).

Affrosman, S., Daviot, J., Holmes, D., Pethrick, R. A. and Wilson, M., Molecular Design for Inhibition of Titanium Corrosion in Resist Cleaner System, Corrosion Science 43, 939 (2001).

Bentiss, F., Traisnel, M., Vezin, H. and Lagrené, M., Linear Resistance Model of the Inhibition Mechanismo of Steel in HCl by Triazole and Oxadiazole Derivatives: Structure-Activity Correlations, Corrosion Science, 45, 371 (2003).

Bergman, I., The Polarography of Polycyclic Aromatic Hydrocarbons and the Relationship Between Their Half-Wave Potentials and Absorption Spectra, Trans. Faraday Soc., 50, 829 (1954).

Bereket, G., Hür, E., Öğretir, C., Quantum Chemical Studies on some Imidazole Derivatives as Corrosion Inhibitors for Iron in Acidic Medium, Journal of Molecular Structure, 578, 79 (2002).

Cardoso, S. P., Hollauer, E., Ponciano G., L. E., Ponciano G., J. A. C., Predictive QSPR Analysis of Corrosion Inhibitors for 22% Cr Steel in Hydrochloric Acid, Journal of Brazilian Chemistry Society, 17(7), 1241 (2006).

Cassidy, J. M., Lancaster, K. R. and Walker, M. L., Electrochemical Study of the Passivation of Some Chromium-Containing Alloys, Corrosion 93: NACE International, Paper 95.

Donahue, F. M. and Nobe, K., Theory of Organic Corrosion Inhibitors, J. Electrochem. Soc. 112, 886 (1966).

Dupin, P., Vilori-Vera, D. A., Savignac, A. de, Lattes, A., Sutter, B., Ph. Haicour, Correlations Between the Molecular Structure of Some Organic Compounds and Their Corrosion Inhibiting Properties in Deareted Media Containing Hydrogen Sulfide, 5th European Symposium on Corrosion Inhibitors, 2, 301 (1980).

Fedders, M., Ponder, J. W., QSAR Programme, 1996. Available at http//dasher.wustl.edu. Access on 06/04/2005. The programme of descriptor selection, based on second-order cross-validation error (SOCV-OLS), represents minor modifications from this original programme.

Growcock, F. B., Inhibition of Steel Corrosion in HCl by Derivatives of Cinnamaldehyde. 1. Corrosion Inhibition Model, Corrosion 45, 12, 1003 (1989).

Growcock, F. B., Frenier, W. W. and Andreozzi, P. A., Inhibition of Steel Corrosion in HCl by Derivatives of Cinnamaldehyde. 1. Structure Activity Correlations, Corrosion 45, 12, 1007 (1989).

Khalil, N., Quantum Chemical approach of Corrosion Inhibition, Electrochimica Acta, 48, No. 18, 2635-2640 (2003).

Kutej, P., Vosta, J., Pancir, J, Macak, J. and Hackerman, N., Electrochemical and Quantum Chemical Study of Dibenzylsulfoxide Adsorption on Iron, J. Electrochem. Soc., 142, No. 3, 829 (1995).

Lukovits, I., Bakó, I., Shaban, A. and Kálmán, E., Polynomial Model of the Inhibition Mechanism of Thiourea Derivatives, Electrochimica Acta, 43, 131 (1998).

Mack, R. D., Corrosion Inhibition of 13Cr, and 15Cr Stainless Steels in HCl-HF Acidizing Fluids, Corrosion 1995, NACE, Paper N. 92/1-9.

MOPAC 6.0, QCPE programme, Bloominton, IN, 1989.

Öğretir, C., Mihçi, B., Bereket, G., Quantum Chemical Studies of Some Pyridine Derivatives as Corrosion Inhibitors, Journal of Molecular Structure, 488, 223 (1999).

PCMODEL, Serena Software, http://www.serenasoft.com

Pradip, Rai, B. Design of Tailor-made Surfactants for Industrial Applications Using a Molecular Modelling Approach, Colloids and Surfaces, 205, No. 1-2, 139-148 (2002).

Santana, G. O., Sathler, L., Gomes, J. A. C. P., Evaluation of Corrosion Inhibitors for Acidizing

Operations, International Corrosion Council, Granada, 2002.

Sastri, V. S. and Perumareddi, J. R., Pitting Corrosion Behaviour of Some Organic Corrosion Inhibitors, Corrosion, 53, No. 8, 617 (1997).

SCI LAB 2.6, http://www.rocq.inria.fr/scilab/

Trabanelli, G., Inhibition of Corrosion-Resistant Alloys in Hot Hydrochloric-Acid Solutions, Werskst Korros 39, No. 12, 589 (1988); Schmitt, G., Research Work on Corrosion Inhibition at the Dacco Corrosion Study Center of the University-OF-Ferrara, Brit. Corros. J., 19, No. 4, 150 (1984).

Vosta, J. and Eliásek, J., Study on Corrosion Inhibition from Aspect of Quantum Chemistry, Corrosion Science, 11, 223-229 (1971).

Wang, D., Li, D., Ying, Y., Wang, M., Xiao, H. and Chen, Z., Theoretical and Experimental Studies of Structure and Inhibition Efficiency of Imidazoline Derivatives, Corrosion Science, 41, 1911 (1999).

Walker, M. L., Cassidy, J. M. and Lancaster, K. R., Electrochemical Study of the Repassivation of Some Chromium-Containing Alloys, Corrosion 93: NACE International, Paper 95.