Chemosphere 160 (2016) 80–88

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Modeling corrosion inhibition efficacy of small organic molecules as non-toxic chromate alternatives using comparative molecular surface analysis (CoMSA)

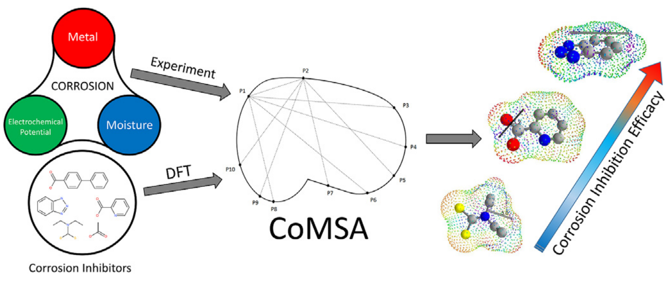
Michael Fernandez [a], [*], Michael Breedon [b], Ivan S. Cole [b], Amanda S. Barnard [a]

[a] *CSIRO Virtual Nanoscience Laboratory, Parkville, VIC, 3052, Australia*
[b] *Electrochemical Systems, CSIRO, Clayton, 3168, Australia*

HIGHLIGHTS

- Novel 3D-QSPR computational method for the rational design of novel corrosion inhibitors.
- Multivariate statistical analysis to identify the most relevant organic corrosion inhibits and its features.
- Computational model of corrosion inhibition from the tridimensional distributions of atomic properties on molecular surfaces.
- Tridimensional structural patterns are revealed which may augment corrosion inhibition efficacy.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 8 January 2016
Received in revised form
12 May 2016
Accepted 10 June 2016
Available online 27 June 2016

Handling Editor: Prof. I. Cousins

Keywords:
Corrosion inhibition
3D-QSAR
DFT methods
Data analytics

ABSTRACT

Traditionally many structural alloys are protected by primer coatings loaded with corrosion inhibiting additives. Strontium Chromate (or other chromates) have been shown to be extremely effectively inhibitors, and find extensive use in protective primer formulations. Unfortunately, hexavalent chromium which imbues these coatings with their corrosion inhibiting properties is also highly toxic, and their use is being increasingly restricted by legislation. In this work we explore a novel tridimensional Quantitative-Structure Property Relationship (3D-QSPR) approach, comparative molecular surface analysis (CoMSA), which was developed to recognize "high-performing" corrosion inhibitor candidates from the distributions of electronegativity, polarizability and van der Waals volume on the molecular surfaces of 28 small organic molecules. Multivariate statistical analysis identified five prototypes molecules, which are capable of explaining 71% of the variance within the inhibitor data set; whilst a further five molecules were also identified as archetypes, describing 75% of data variance. All active corrosion inhibitors, at a 80% threshold, were successfully recognized by the CoMSA model with adequate specificity and precision higher than 70% and 60%, respectively. The model was also capable of identifying structural patterns, that revealed reasonable starting points for where structural changes may augment corrosion inhibition efficacy. The presented methodology can be applied to other functional molecules and extended to cover structure-activity studies in a diverse range of areas such as drug design and novel material discovery.

---

[*] Corresponding author.
E-mail address: michael.fernandezllamosa@csiro.au (M. Fernandez).

http://dx.doi.org/10.1016/j.chemosphere.2016.06.044
0045-6535/

1. Introduction

Many structural alloys are susceptible to corrosion and are protected by inhibitor containing primers, with chromates being particularly effective inhibitors and thus chromate bearing primers being the protective system of choice for a number of industries, particularly the aerospace industry (Kendig and Buchheit, 2003). However chromates are a human toxin and thus pose a risk in their production or when removed from the aircraft during repair or repainting. A recent study of workers active in the chrome-chemical production industry found a massive increase in the risk of lung cancer (Park et al., 2004), so their systematic elimination from service is now being enforced by new legislative guidelines (Costa and Klein, 2008). More than a decade of global research has led to a variety of solutions (Hughes et al., 2010), with one of the most promising being the use of small organic molecules. There exists a rich diversity of potential candidate structures, leading to significant complexity in the nature and range of possible inhibiting mechanism of these small molecules (Gece, 2008).

We have previously investigated the fundamental molecular properties of the Harvey data set of organic corrosion inhibitors (Harvey et al., 2011), which contains 28 different molecules and found that there was no apparent correlation between experimentally determined inhibition and *in vacuo* density functional theory (DFT) calculated molecular properties: ionization potential, electron affinity, fundamental gap and deprotonation energy (Breedon et al., 2014). Further analysis of the Harvey data set using quantitative structure-property relationships (QSPR) techniques revealed that from over 350 different descriptors, the most structurally pertinent were those relating to the number of S or N atoms, their hybridisation; the number of thiol groups; and partial charges on H and C atoms (Winkler et al., 2014).

In this paper we extend our examination of molecular structure from predominately $\leq$ 2D descriptors, to investigate the distribution of electronegativity, polarizability and van der Waals volumes on the molecular surfaces of the Harvey data set of small organic molecules to identify features which correspond with “high-performing” corrosion inhibitors. Archetypal analysis (AA) and *k*-means clustering techniques will be used to identify the prototypes and archetypes within the data set, which will be further investigated using comparative molecular surface analysis (CoMSA), and a novel tridimensional QSPR (3D-QSPR) approach will be used to establish correlations between structure and performance; guiding future experimental design of the next-generation of corrosion inhibitor molecules.

2. Materials and methods

The CoMSA model was developed in four main steps: (1) structural optimization of molecular structures with density functional theory (DFT); (2) generation of the molecular surface and calculation of the smoothed and averaged values of the following atomic properties: electronegativity, polarizability and van der Waals volume at the surface; (3) computation of molecular fingerprints using autocorrelogram transforms and (4) a genetic algorithm (GA) feature selection coupled to multivariate correlation using multiple linear regression (MLR).

2.1. Data set and prior preparation

In this study we utilised a data set of 28 organic corrosion inhibitors known as the Harvey data set (Harvey et al., 2011) with structures and corrosion inhibition activities on AA2024-T3 presented in Fig. S1 in the Supporting Information, where percent inhibition (or acceleration) of an organic molecule is assessed *via* corrosion induced mass loss of an aluminium alloy coupon immersed in an inhibited saline solution after 28 days. The mass loss corrosion inhibition utilised by Harvey et al. (Harvey et al., 2011). which were set against the chromate standard have been classified in this study so that organic molecules with an experimentally determined corrosion inhibition efficacy <80% are “inactive” or corrosion accelerators, while molecules with efficacy >80% are defined as “active” or corrosion inhibitors, which yields 15 active and 13 inactive corrosion inhibitors.

The Harvey data set was selected as a representative test set as it contains molecules which were determined to be either corrosion accelerators or corrosion inhibitors, rather than some test sets which contain only molecules known to be corrosion inhibitors. The Harvey data set was built around the philosophy of utilising several parent molecules which were systematically modified to highlight structure/function relationships, i.e. repositioning of functional group, substitution of a functional group, or heterocyclic modification of an aromatic ring structure; making it an ideal test set for CoMSA.

2.2. Comparative molecular surface analysis (CoMSA)

Organic corrosion inhibitor molecules were structurally relaxed in the Gaussian09 program (Frisch et al., 2009) using DFT method, with default energy and force convergence criteria; the B3LYP functional and a 6-311G(d,p) basis set was used throughout. In the case of highly flexible molecules, sampling from a conformational ensemble of structures could be more suitable, as in the 4D-QSAR approach (Hopfinger et al., 1997). But for the small organic molecules in the Harvey data set, which are rather rigid, it is perfectly valid using the structures at the minimum of the geometrical optimization at the DFT level.

Following this, the molecular surface of the optimized structures was approximated to a solvent-excluded surface or Connolly surface (Connolly, 1983), which is imagined as a cavity in bulk solvent (effectively the inverse of the solvent-accessible surface). Using the MSMS software (Sanner et al., 1996), molecular surfaces are obtained as sets of vertexes in three-dimensional space arranged in a triangular mesh at a density of ~10 vertexes per Å^2. The smoothed and averaged values of three atomic properties namely electronegativity, polarizability and van der Waals volume are then computed at each vertex according to the equation below.

$$\mathrm{P}_{\mathrm{i}}^{\mathrm{k}} = \sum_{\mathrm{j}=1}^{\mathrm{N}} \mathrm{p}_{\mathrm{j}}^{\mathrm{k}} \exp^{-(\mathrm{r}_{\mathrm{ij}})^{2}} \tag{1}$$

where P_i^k is the smoothed and averaged value of atomic property k at vertex i, p_j^k is the atomic property k of atoms j and r_{ij} is the distance between vertex i in the molecular surface and atom j.

This yields three molecular surfaces for each structure to perform the 3D-QSPR analysis of the corrosion inhibition. To correlate the information in the atomic property-smoothed molecular surfaces to the corrosion inhibition we used a grid-independent approach (GRIND). To mitigate alignment mismatching between molecules in the Harvey data set, the molecular surface were transformed to maximum autocorrelation scores (MATS) s which generate alignment-independent descriptors as previously described by Pastor et al. (Pastor et al., 2000). Alignment independence was achieved by computing the pair products of the smoothed atomic properties for vertexes in the molecular surface located at certain relative distances (Fig. 1) according to the equations below.