Corrosion Science xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Corrosion Science

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

Using high throughput experimental data and *in silico* models to discover alternatives to toxic chromate corrosion inhibitors

D.A. Winkler[a,b,c,d,*], M. Breedon[a], P. White[a], A.E. Hughes[a,c], E.D. Sapper[e], I. Cole[a]

[a] CSIRO Manufacturing, Clayton 3168, Australia
[b] Monash Institute of Pharmaceutical Sciences, 399 Royal Parade, Parkville 3052, Australia
[c] Latrobe Institute for Molecular Science, Bundoora 3086, Australia
[d] School of Chemical and Physical Sciences Flinders University, Bedford Park, 5042, Australia
[e] Boeing Research & Technology, Seattle, WA 98124-2207, USA

ARTICLE INFO

Article history:
Received 11 August 2015
Received in revised form 12 January 2016
Accepted 5 February 2016
Available online xxx

Keywords:
A. alloy
A. aluminium
B. modelling studies
C. neutral inhibition

ABSTRACT

Restrictions on the use of toxic chromate-based corrosion inhibitors have created important issues for the aerospace and other industries. Benign alternatives that offer similar or superior performance are needed. We used high throughput experiments to assess 100 small organic molecules as potential inhibitors of corrosion in aerospace aluminium alloys AA2024 and AA7075. We generated robust, predictive, quantitative computational models of inhibitor efficiency at two pH values using these data. The models identified molecular features of inhibitor molecules that had the greatest impact on corrosion inhibition. Models can be used to discover better corrosion inhibitors by screening libraries of organic compounds for candidates with high corrosion inhibition.

1. Introduction

Corrosion is clearly a very important and expensive environmental issue for large number of metallic materials used for a myriad of engineering purposes. In the aerospace industry, problems are particularly acute due to the high cost of the infrastructure and the catastrophic consequences of materials failure due to corrosion. Chromates are very effective corrosion inhibitors and have been the treatment of choice for the aerospace industry to date [1]. However, they pose an unacceptable risk to workers and the environment, particularly during production, because of their toxicity and carcinogenicity. Studies have shown that they are 'hot spot' pollutants. Recent epidemiological data from a study of chrome chemical production workers found the excess lifetime risk of death from lung cancer due to occupational exposures to be 255,000 per million workers [2], massively larger than the 'acceptable' risk of 1 death per million. Consequently, chromates are progressively being restricted or removed from service by legislation [3].

There has been an intense search for more benign replacements for chromate corrosion inhibitors that have equivalent or superior performance [4]. Small organic compounds are showing considerable promise as replacement inhibitors, and the recent literature has identified a number of them that show useful corrosion inhibition properties and potentially lower toxicity than exiting inhibitors. Finsgar has reviewed the efficacy and mechanism of action of benzotriazoles in preventing corrosion of copper [5]. Gece has likewise reviewed the potential of accessible small molecule drugs as corrosion inhibitors [6]. Very recently Kuznetsov reviewed the physicochemical aspects of protection of metals by organic corrosion inhibitors [7]. It has been estimated [8] that there are an immense number ($\sim 10^{80}$) of small organic molecules that could be synthesized, providing an essentially infinite source of potentially high performance, relatively benign corrosion inhibitors. Preliminary computational modelling studies by Winkler et al. have shown that the potential of a range of small organic compounds to inhibit corrosion in aerospace alloys can be predicted reliably [9]. However, this work demonstrated that small changes to the molecular structure of such organic molecules could result in dramatic changes in the performance of inhibitors. Considering the number of families of possible inhibitors and their isomers, there are tens of thousands of candidate inhibitor molecules.

Materials science is increasingly using automated methods of synthesis and characterization to accelerate the discovery and optimization of new materials. Modern machine learning methods using these large and rich data sets are showing have proven useful for the quantitative prediction of a wide range of materials prop-

* Corresponding author at: CSIRO Manufacturing, Bag 10, Clayton South MDC 3169, Australia. Fax: +61 395452446.
E-mail address: dave.winkler@csiro.au (D.A. Winkler).

http://dx.doi.org/10.1016/j.corsci.2016.02.008
0010-938X/

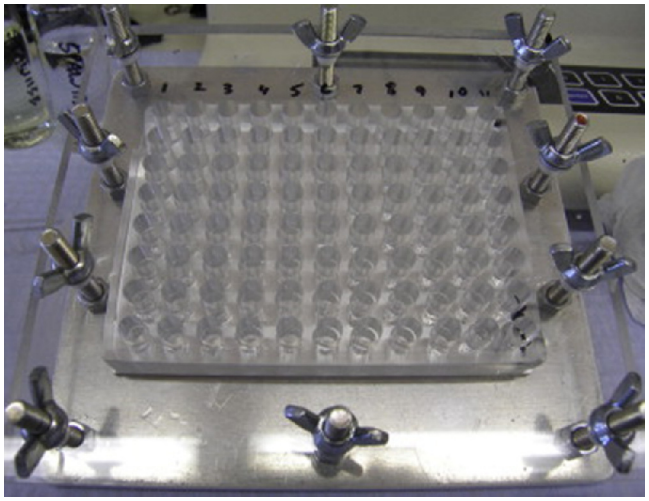

Fig. 1. High throughput corrosion inhibition rig consisting of a 10 mm thick polycarbonate clamped to a 10 mm thick block of polydimethylsiloxane rubber, an abraded plate of alloy and a 5 mm thick metal baseplate. The polycarbonate and polydimethylsiloxane sheets have an 8×11 grid of 6 mm diameter holes for test solutions.

erties [10]. Small organic molecules are particularly amenable to a high throughput synthesis approach as the technology to generate large libraries of these materials was developed for the pharmaceutical industry more than two decades ago. Here we combine novel high throughput corrosion inhibition testing experiments with machine learning methods to model and predict the corrosion inhibition performance of a large library of small organic candidate corrosion inhibitors at two initial pH values.

2. Materials and methods

2.1. Corrosion data

Two aluminium alloys, AA2024 and AA7075 were used for the inhibition study. The compositions (weight%) were determined by ICP as AA2024 (Cu 5.3, Mg 1.6, Mn 0.6, Fe 0.2, Zn < 0.1) and AA7075 (Cu 1.4, Mg 2.4, Mn < 0.1, Fe 0.2, Zn 5.4). While iron and magnesium levels were similar, AA2024 contained significantly more copper than AA7075, and AA7075 had substantially higher zinc content than the AA2024 alloy.

Corrosion inhibition data were generated by a high-throughput testing rig previously described [11] (see also Fig. 1). The two alloys were exposed to 100 different organic candidate corrosion inhibitors at a concentration of 10^{-3} M in 0.1 M NaCl solutions with initial pH values of 3, 4, 5, 6, 8, 10. As the components of most buffers are also potential corrosion inhibitors they were not used to control the pH of the experiments after initial adjustment of the pH. The inhibition of corrosion by the test compounds was assessed by image processing after a 24-h period. The degree of corrosion could be accurately estimated from the brightness of the images of the circular regions on the alloy exposed to the solution in each well. The image based corrosion scores were highly correlated with results of mass loss corrosion experiments. Potassium dichromate was used as a positive control, and 0.1 M saline (added to all wells) was the negative control. The data at initial pH 4 of and 10 were the most complete and were used in the modelling study. Note that the scoring scheme used in the previous publication (0 = no corrosion, and 100 = most corrosion)[11] were rescaled and reversed for the current inhibition study. Consequently, the scores used in this study ranged from zero (no inhibition) to 10 (maximum inhibition). The estimated measurement errors were ±1 in the 10 point scale. The scores are unitless.

2.2. Speciation

The data set consists of 100 small organic molecules with substantial chemical diversity (Supplementary Table 1). In some cases the identity of the organic species in solution (the various ionized forms of the molecules containing acidic or basic heteroatoms) is unambiguous at the pH of the experiment. However, for the heteroaromatic compounds, and indeed, for some of the inhibitors that contain, for example, carboxylate and thiol acidic moieties, the number of ionic species in solution can be quite high. As these ionic forms may have different binding affinities for metals and they will be chemically quite distinct, it might be relevant to understand which ionic species coexist at the experimental pH. However, the pH in the corrosion pits on the alloys is likely to be quite different to that in the bulk. It was also clear from inspection of the inhibition data that the initial pH had a relatively small effect on inhibition in the majority of cases over a 24-h exposure (Supplementary Table 2). Consequently, the compounds were modelled in the neutral form.

Supplementry material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.corsci.2016.02.008.

2.3. Quantum mechanical and modelling studies

The 100 molecules in the data set were built and their geometries optimized using the Sybyl x2.0 molecular modelling package (Certera Limited). These structures were used to generate a range of molecular properties (descriptors) such as the molecular surface area, molecular volume, molar refractivity (molecular size and polarizability), polar surface area, numbers of hydrogen bond donors and acceptors, logP (log of the octanol-water partition coefficient), HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) energies and band gap (calculated using the semiempirical molecular orbital package MOPAC/AM1 and density functional theory (DFT) (see below)). A large pool of computed molecular descriptors was calculated using the DRAGON program [12] and our in-house modelling package, BioModeller [13–15]. Where possible, to aid interpretation of models, we used functional group descriptors that describe to the existence of certain chemical moieties or fragments in molecules (e.g. number of ionized carboxylic acid groups, number of heterocyclic nitrogen atoms, number of thiol groups etc.). Such descriptors make interpretation of models easier for organic chemists as they serve as design rules for new compounds to have good corrosion inhibitory properties. Descriptors such as total molecular charge, and several of the DRAGON descriptors that describe either the existence or frequency of molecular fragments containing atoms a specified number of bonds apart (topological distance) can also be interpreted relatively easily.

We also used quantum chemical properties for inhibitors calculated by DFT using the Spanish Initiative for Electronic Simulations with Thousands of Atoms (SIESTA) [16]. The exchange correlation functional of Perdew–Burke–Ernzerhof (PBE) [17] within the generalized gradient approximation (GGA), and a double zeta plus polarization (DZP) basis set was employed throughout, as per our earlier study [18]. The following molecular properties were calculated *in vacuo*: electron affinity, ionization potential, Mulliken electronegativity, chemical potential, chemical hardness, and the fundamental gap/HOMO-LUMO gap.

We generated machine-learning models that related molecular properties of inhibitors to corrosion inhibition using the BioModeller software. This software uses novel sparse Bayesian feature selection methods to identify relevant molecular properties from large pools of molecular descriptors, and sparse modelling methods to generate optimal structure-inhibition models, both linear and nonlinear. Linear models were simply multiple linear regressions (MLR, weighted combinations of molecular descriptors) to which

Please cite this article in press as: D.A. Winkler, et al., Using high throughput experimental data and *in silico* models to discover alternatives to toxic chromate corrosion inhibitors, *Corros. Sci.* (2016), http://dx.doi.org/10.1016/j.corsci.2016.02.008