Predicting the Performance of Organic Corrosion Inhibitors

David A. Winkler 1,2,3,4,5

1 La Trobe Institute for Molecular Science, La Trobe University, Kingsbury Drive, Bundoora 3086, Australia; drdavewinkler@gmail.com; Tel.: +61-466-258-600
2 Monash Institute of Pharmaceutical Sciences, Monash University, 392 Royal Parade, Parkville 3052, Australia
3 School of Chemical and Physical Sciences, Flinders University, Bedford Park 5046, Australia
4 CSIRO Manufacturing, Bayview Avenue, Clayton 3168, Australia
5 School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, UK

Received: 20 November 2017; Accepted: 5 December 2017; Published: 8 December 2017

Abstract: The withdrawal of effective but toxic corrosion inhibitors has provided an impetus for the discovery of new, benign organic compounds to fill that role. Concurrently, developments in the high-throughput synthesis of organic compounds, the establishment of large libraries of available chemicals, accelerated corrosion inhibition testing technologies, and the increased capability of machine learning methods have made discovery of new corrosion inhibitors much faster and cheaper than it used to be. We summarize these technical developments in the corrosion inhibition field and describe how data-driven machine learning methods can generate models linking molecular properties to corrosion inhibition that can be used to predict the performance of materials not yet synthesized or tested. We briefly summarize the literature on quantitative structure–property relationships models of small organic molecule corrosion inhibitors. The success of these models provides a paradigm for rapid discovery of novel, effective corrosion inhibitors for a range of metals and alloys in diverse environments.

Keywords: corrosion inhibitors; high-throughput corrosion inhibition testing; machine learning; quantitative structure–property relationships (QSPR); organic molecules; molecular design; data-driven models

1. Introduction

Corrosion is responsible for an excessive amount of catastrophic failure in many different industries, causing death, injury, and capital loss. Corrosion prevention and treatment has thus become a multi-trillion-dollar imposition on industry. Corrosion may be inhibited by a range of surface treatments and coatings, many of which are highly effective. For example, in the aerospace industries, chromates have been the mainstay of corrosion treatment of alloys used in aircraft. However, health and safety and environmental concerns have led to the ban or restricted use of corrosion inhibitors containing such elements as tin, chromium, and lead. For example, the use of chromate inhibitors is being phased out in most countries. It has been estimated that workers exposed to chromate residues when aircraft is sanded and recoated have up to a 250,000-fold higher risk of cancer than the general public [1].

Consequently, there has been a large research effort to generate inhibitors and coatings that are very effective and more benign to workers, the public, and the environment. Small organic compounds show promise as corrosion inhibitors, although their mechanisms of action are far from clear. Recent developments in automated high-throughput chemical synthesis [2,3], the availability of large libraries of organic compounds, high-throughput methods of assessing corrosion inhibition by mass loss, electrochemistry or other means, and the rapid growth in the capabilities of machine learning methods such as deep learning provide an unprecedented opportunity to discover or design new,
more effective corrosion inhibitors \[4,5\]. Small molecule, synthetically accessible space is also vast, estimated to be \(-10^{100}\), providing rich opportunities for discovery of novel, effective small molecule corrosion inhibitors \[6\]. However, the size of chemistry space also means that high-throughput experimentation cannot explore a significant fraction of it alone, unless other data-driven methods like design of experiments \[7\], structure–property relationship modelling, and evolutionary methods are used to extensively leverage the data that high throughput screening (HTS) methods can generate.

This paper briefly reviews some of the more important developments in the high-throughput synthesis, design, and modelling of organic corrosion inhibitors. The main focus of the review is the use of the quantitative structure–property relationship (QSPR) method to predict the corrosion inhibitory properties of organic compounds \[8–10\]. Such models can be used to understand the relationships between the chemical structure of inhibitors and their efficacy and to allow the inhibition of compounds not yet synthesized or tested to be predicted. They can also be used as surrogate fitness functions for the evolutionary design of new corrosion inhibitors with multiple desirable properties \[11–13\].

2. High-Throughput Synthesis and Testing of Organic Corrosion Inhibitors

Although high-throughput and combinatorial synthesis of small organic compounds is well established in the pharmaceutical industry, it is rarely used in corrosion inhibition research. This is likely due to the lack of high-throughput corrosion inhibition testing methods, and the existence of very large libraries of commercially available organic compounds that can be tested for corrosion inhibition. Development of high-throughput, direct, or surrogate methods of assessing corrosion inhibition performance is therefore critical to progress in this field.

Significant progress has been made recently in high-throughput corrosion inhibition testing. The most important factor is how the fast, surrogate testing methods used in high-throughput testing correlate with “real world” corrosion inhibition. Currently, electrochemical methods, mass loss, and photometric methods have been employed to assess corrosion inhibition. Seminal work has been reported by Chambers et al., who used direct current polarization between two aerospace aluminium alloy (AA2024) wire electrodes and a multiple-electrode testing system to assess the corrosion inhibition of fifty chemistries in just 9 h. The results correlated highly with those of extended testing over 10 days \[14\]. This research team extended their work by scoring corrosion using fluorometric detection of Al\(^{3+}\) concentrations \[15\]. They measured corrosion inhibition of 14 corrosion inhibitors over a wide range of initial pH values for a period between 1 and 7 days \[16\]. Their later work developed a system for rapidly assessing inhibition characteristics of 100 separate chemistries using direct current (DC) polarization, cyclic voltammetry of re-deposited copper, and fluorometric detection of Al\(^{3+}\) \[17,18\]. Kallip and coworkers reported a novel scanning vibrating electrode technique for assessing corrosion inhibitors \[19\]. They were able to accurately determine percent corrosion inhibition efficiencies of Fe and Zn for four inhibitors. He et al. described a high-throughput electrochemical impedance spectroscopy (HT-EIS) method for the rapid evaluation of corrosion coatings. They developed a 12-element, spatially addressable electrochemical platform interfaced to a commercial EIS instrument \[20\]. Recently, White et al. described a novel method for assessing corrosion inhibition via a high-throughput testing rig (Figure 1) \[1,21\]. Up to 88 simultaneous corrosion inhibition tests can be carried out on a single plate, with positive and negative controls, in approximately one day. Corrosion inhibition was assessed using a novel, robust computerized image processing method. More recently, Shi et al. reported a similar automated system for corrosion assessment using optical imaging that also showed a linear relationship between an apparent grey scale value of the image and the depth of corrosive pitting in the specimen \[22\].
3. Machine Learning Modelling Methods

Clearly, the structures of organic compounds play a major role in how effectively they inhibit metal corrosion. In essence, changes to chemical structure directly modulate corrosion inhibition, a pattern recognition problem. The pharmaceutical industry has a long history of developing methods for generating quantitative models describing similar effects (on biological targets in this case). Recently, this large body of knowledge has been leveraged by materials science to design and discover new bespoke materials with hitherto inaccessible properties. As these methods are data-driven, the emergence of reliable high-throughput methods for screening potential corrosion inhibitors described above means that quantitative modelling methods will be used more frequently in corrosion inhibition research.

Quantitative models linking chemical structure to properties can be developed using machine learning or other statistical methods. These methods—quantitative structure–activity relationships (QSAR) and quantitative structure–property relationships (QSPR) modelling—are essentially the same. The main difference is in the types of properties being modelled (biological versus non-biological) and the ways in which the molecular or material is represented mathematically in the modelling process [8,10,23].

QSAR and QSPR methods involve several processes [24,25]. A set of molecules or materials are collected and tested for the desired property. The structural and physicochemical properties of the molecule or material are encoded mathematically (these are called molecular descriptors) in a way this is relevant to the property being modelled, and a subset selected in a context-dependent way. A mathematical relationship, often nonlinear, is found between the descriptors and the property being modelled, and the predictive power of the model assessed using an independent test set of compounds not used to generate the model [26]. Finally, the model can be interrogated to understand what features of the molecules improve or degrade their performance, and the model can be used to predict the properties of compounds not yet synthesized or tested. Very large real or virtual libraries of organic compounds can be screened very quickly this way, so long as the screened molecules lie close to the domain of applicability of the models (the region of chemical and property space in which the model was trained).

Mathematical relationships between the selected descriptors and the property being modelled can be generated by linear (e.g., multiple linear regression) or nonlinear (e.g., neural network, recursive partitioning, kernel or polynomial regression) methods. Neural network and other machine learning
methods can generate models quickly and effectively and require few or no assumptions to be made about the form of the mathematical relationship, as they are universal approximators [27,28]. The rise of deep learning algorithms within the last five years has stimulated the use of neural network and machine learning approaches substantially [4,5,29].


There have been relatively few reports of QSPR models for corrosion inhibition by small organic molecules, and fewer still that are robust. Many early reported models were based on very small data sets with limited chemical diversity. This increased the probability of chance correlations that were not indicative of causative relationships between the compounds and the inhibition. A relatively large number of studies with small number of inhibitors reported a relationship between the frontier orbital properties of small organic molecule (principally the energies of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO): \(E_{\text{HOMO}}, E_{\text{LUMO}}\) and band gap energies \(E_{\text{LUMO}} - E_{\text{HOMO}}\), calculated by various quantum mechanical (QM) methods and corrosion inhibition. For example, Sastri and Perumareddi reported correlations of corrosion rates with \(E_{\text{HOMO}}\), the energy gap, and Hammett’s parameter (\(\sigma\)) for a small set of organic compounds [30]. Subsequently, Ozcan and Dehri reported a similar correlation between frontier orbital properties of three organic inhibitors, thioacetamide, thiourea, and thiobenzamide [31]. A second study by these authors reported apparent correlations between corrosion inhibition and frontier orbital properties for three other organic compounds—thiourea, methylthiourea, and phenylthiourea—with similar structures and high corrosion inhibition (93%, 96%, and 97%, respectively). The small changes in structure and the small differences in inhibition values relative to experimental error suggest that these correlations were not statistically valid [32]. Sastri et al. subsequently reported that corrosion inhibition correlated with \(E_{\text{HOMO}}\), the HOMO–LUMO gap, the chemical softness of the inhibitor, chemical potential, and the fraction of charge transferred from the inhibitor to the metal [33]. However, this result was based on few compounds and, like all of the above studies, did not include allowance for solvent effects, ionization, and speciation at testing pH. Recently, Bedair reported a quantum chemical study of corrosion inhibition of pyridine, quinolone, acridine, and their n-hexadecyl derivatives on steel. He used density functional theory (DFT), ab initio calculations, and semi-empirical methods to find correlations between molecular structure and corrosion inhibition efficiency. He reported that calculated dipole moments, \(E_{\text{HOMO}}, E_{\text{LUMO}}\), and the HOMO–LUMO gap correlated with corrosion inhibition, although the small number of compounds used and their similarities in structure throw doubt on the utility of these correlations [34]. Saha et al. similarly investigated 2-aminopyrazine, 2-amino-5-bromopyrazine, 3-amino pyrazine-2-thiol, and 3-amino-6-bromopyrazine-2-thiol as corrosion inhibitors using density functional theory (DFT) and molecular dynamics (MD) simulations [35]. In these studies, a solvent model, conductor-like screening model (COSMO), and neutral and protonated species were used in the correlations. However, only qualitative comparisons between corrosion inhibition and quantum chemical parameters were found. In all of the above studies the number of organic compound studies were too few to allow a test set to be used to quantify the predictivity of the models.

Considerable doubt has been thrown on the usefulness of molecular properties calculated by quantum chemical methods by two recent studies that used a much larger and chemically diverse set of organic corrosion inhibitors. Winkler et al. showed that robust, predictive models linking computed molecular properties to corrosion inhibition can be generated using neural networks [36]. However, the molecular properties that most influenced the inhibition were quite arcane and hard to interpret. The first study used mass loss as a measure of corrosion of aerospace aluminium alloys and employed 28 diverse small organic molecule inhibitors. They found that, whether or not speciation was included, there was essentially no correlation between ionization potential, HOMO or LUMO energies, or any other quantum chemically derived descriptors and corrosion efficiency. A subsequent and larger study using 100 diverse inhibitors and a novel high-throughput corrosion inhibition experimental methods also showed negligible correlation between quantum chemically derived parameters and
corrosion inhibition. However, in this case, good, robust models were developed that linked molecular properties calculated by non-quantum chemical methods and corrosion inhibition. These models could make reliable, quantitative predictions of corrosion inhibition for compounds not used to train the model (Figure 2) [1].

![Figure 2](image_url)

**Figure 2.** Quantitative prediction of corrosion inhibitory properties of organic compounds for aerospace alloys AA-2024 generated using molecular descriptors (not derived from quantum mechanical calculations). Inhibition is scaled between 0 (no inhibition) and 10 (highest inhibition). The circles represent the performance of the model in predicting training set, and the triangles represent the performance of the model in predicting the test set. Used with permission from Winkler et al. [1].

There is potential for more sophisticated quantum chemical methods to contribute to modelling of corrosion inhibition. Recently, Breedon et al. reported a novel 3D-QSPR method—comparative molecular surface analysis (CoMSA)—that employed 3D distributions of electronegativity, polarizability, and van der Waals volume on the molecular surfaces of 28 small organic molecules as descriptors [37]. This method could make a qualitative prediction of corrosion inhibition efficiency, identifying high-performing corrosion inhibitor candidates. This approach may be elaborated to improve the accuracy of prediction, take into account solvents and speciation, and allow quantitative, or at least semi-quantitative predictions to be made about corrosion inhibition.

When other types of measured or computed molecular descriptors are used, it is possible to build robust and predictive models of corrosion inhibition from large data sets derived from high-throughput testing of large chemical libraries. These descriptors relate to the types of atoms and their valence in compounds, chemical graph properties, physicochemical properties such as dipole moment and lipophilicity (polarity), and many other molecular characteristics. More than 3000 of these mathematical representations of molecules can be computed using commercially available packages such as Dragon [38]. One of the first studies of the structure–activity relationships in 400 organic compounds of corrosion inhibitors was reported by Horner and Meisel in 1978 [39]. Subsequently, corrosion inhibition for iron and nickel in acidic solution was measured for four organic inhibitors by Jayalakshmi and Muralidharan [40]. They generated qualitative structure–property relationships, showing that the concentration of the inhibitor, hydrophobicity, and π electron density played a role in inhibitor efficiency.

Recently, Keshavarz et al. [41] reported a simple method for predicting corrosion inhibition efficiency for steel of small organic molecules, principally imidazoles and benzimidazoles. They studied 34 diverse chemicals and derived a simple linear regression model for inhibition:

\[
\eta = 38.47 + 20.21n(N) - 7.98n(O + NH_2) + 14.94\eta^+ - 17.93\eta^-
\] (1)
where $\eta$ is the corrosion inhibition efficiency; $n(N)$ is the number of nitrogen atoms; $n(O + NH_2)$ is the sum of the number oxygen atoms and amino groups; $\eta^+$ and $\eta^-$ are the positive and negative effects of structural parameters on the inhibition efficiency. Their model was more accurate in predicting the corrosion efficiency of compounds in a test set of 11 compounds not used to generate the model than two QSPR models reported by Zhang et al. [42] derived from quantum chemically derived parameters. However, the range of inhibition values for the training and test sets was not high (60–95% for the test set) relative to the prediction errors, and the way in which the $\eta^+$ and $\eta^-$ parameters correct the predictions of inhibition is not clear.

The modelling study reported by Zhang et al. involved 34 imidazole and benzimidazoles corrosion inhibitors and used quantum chemically derived descriptors augmented by topographical descriptors called Molecular Connectivity Indices (MCI) [42]. They stated that a previous paper by Zhao et al. (which they did not reference) could not find strong QSPR models for organic corrosion inhibitors using quantum chemical descriptors alone. They augmented the QM descriptors by MCI descriptors and dramatically improved the models. This provided the rationale for the use of MCI descriptors in Zhang et al.’s study. Using this composite descriptor set they could generate models of the relationship between molecular properties and corrosion inhibition. Inhibition was assessed by weight loss. The QM descriptors used included $E_{\text{HOMO}}$, $E_{\text{LUMO}}$, partial charges, electron densities, frontier orbital properties, and polarizability. These were augmented by log octanol-water partition coefficients (a measure of lipid solubility), nX and $\delta_i$’ topological indices derived from the chemical graph of each inhibitor. The data set of 34 inhibitors used 16 of the 18 descriptors, making overfitting likely unless care is taken. Stepwise linear regression generated two models with six parameters that reproduced the training set of 34 compounds with $r^2 = 0.81$ and a standard error of 10%. One model included the effect of electron-donating and -accepting properties of a metal surface. Unfortunately, these authors did not use an independent test set to test the predictivity of the model for new data.

5. Conclusions and Perspective

The need for new, safe and environmentally benign corrosion inhibitors has been strengthened by the unacceptable toxicity of existing, albeit highly effective, methods for corrosion control. There is a very fortunate juxtaposition of this need with the large increase in capabilities of technologies relevant to the design of new organic corrosion inhibitors. Automated, high-throughput methods of screening large numbers of potential inhibitors will accelerate the discovery of new inhibitors and ultimately allow for inhibitors with multiple functions (e.g., corrosion inhibition and the inhibition of methane hydrates in undersea gas pipelines [43]).

These automated corrosion inhibition testing methods can potentially generate large data sets for a diverse range of organic chemotypes that are very well matched to analysis by machine learning and other statistical modelling methods. Our review shows that it is possible to generate robust and predictive computational models that make accurate, quantitative predictions of the degree of corrosion inhibition for organic compounds not yet tested. The most successful models reported in the literature employ empirical molecular descriptors as models that use molecular descriptors derived from quantum chemical calculations that are not statistically significant enough to be useful. We anticipate that quantum chemical calculations based on model systems much closer to real world metal/water/inhibitors systems may play an important role in the future. Although there are relatively few reported studies of high-throughput testing and modelling of corrosion inhibition properties of small organic molecules, we are at the bottom of the S-curve. Improvements in robotics and machine learning will be autocatalytic, leading to a massive increase in the capabilities and reliability of methods for the design of organic corrosion inhibitors in the short to medium term.

Conflicts of Interest: The author declares no conflict of interest.
References


32. Ozcan, M.; Dehri, I.; Erbil, M. Organic sulphur-containing compounds as corrosion inhibitors for mild steel in acidic media: Correlation between inhibition efficiency and chemical structure. Appl. Surf. Sci. 2004, 236, 155–164. [CrossRef]