



In silico screening of modulators of magnesium dissolution

Christian Feiler^{a,*}, Di Mei^a, Bahram Vaghefinazari^a, Tim Würger^{a,b}, Robert H. Meißner^{a,b}, Bérengère J.C. Luthringer-Feyerabend^{a,f}, David A. Winkler^{c,d,e}, Mikhail L. Zheludkevich^{a,f}, Sviatlana V. Lamaka^a

^a Magnesium Innovation Centre - MagIC, Institute of Materials Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany

^b Institute of Polymer and Composites, Hamburg University of Technology, Hamburg, Germany

^c La Trobe Institute for Molecular Science, La Trobe University, Kingsbury Drive, Bundoora, Australia

^d CSIRO Manufacturing, Clayton, Australia

^e Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Australia

^f Institute for Materials Science, Faculty of Engineering, Kiel University, Kiel, Germany



ARTICLE INFO

Keywords:

Magnesium
Corrosion modulators
Density functional theory
QSPR

ABSTRACT

The vast number of small molecules with potentially useful dissolution modulating properties (inhibitors or accelerators) renders currently used experimental discovery methods time- and resource-consuming. Fortunately, emerging computer-assisted methods can explore large areas of chemical space with less effort. Here we show how density functional theory calculations and machine learning methods can work synergistically to generate robust and predictive models that recapitulate experimentally-derived corrosion inhibition efficiencies of small organic compounds for pure magnesium. We further validate our methods by predicting *a priori* the corrosion modulation properties of seven hitherto untested small molecules and confirm the prediction in subsequent experiments.

1. Introduction

Magnesium (Mg) is one of the lightest, most abundant and versatile engineering materials. When Mg-based parts are designed for structural applications in automotive, aerospace and consumer goods industries, corrosion protection strategies need to be applied to extend the service life of Mg structural elements. Conversely, Mg dissolution promoters are used to assure the constant dissolution of Mg anode materials in Mg-air primary batteries. Furthermore, application of either Mg corrosion inhibitors or accelerators (hereafter jointly called modulators) is necessary for the application of Mg-based bio-resorbable implants to achieve specific degradation rates for the treatment of different injuries. Small organic molecules are extremely useful modulators of a wide variety of processes. These include many important biomedical applications (e.g. novel drugs and control of protein-protein interactions [1–3]) and non-biological commercial applications such as increased conversion efficiencies in solar cells [4] as well as corrosion inhibition of light metals [5–7]. Discovery of effective corrosion inhibitors is particularly critical due to the imminent ban of highly effective but toxic chromates [5,8–10]. The search for green corrosion inhibitors is complex. It is becoming apparent that a new universal corrosion

inhibitor matching the performance of chromate will not be found. Most likely, each application will require bespoke corrosion inhibitor molecules and specialized methods of applying these agents (e.g. incorporation in a polymeric coating or porous oxidation treatment [11–14]). Reactive, magnesium-based lightweight materials [15,16] used for automotive [17] and aerospace [18] applications are particularly prone to corrosion, so effective inhibitors are needed to avoid material failure. Bio-resorbable medical implants can also be loaded with modulators that provide control over the degradation rate as well as drugs that improve implant survival and prevent inflammation. [19–23] Furthermore, the next generation of primary Mg-air batteries [24–26] requires efficient dissolution modulators to control self-corrosion (increasing the utilization efficiency) and reduce precipitation of corrosion products on the surface of magnesium anodes (assuring higher output voltage).

The vast majority of the roughly 150 million compounds listed in the Chemical Abstracts Service database are organic. The number of organic compounds reported in the literature is increasing exponentially, with ~120 million compounds being added in the last decade. [27] It has been estimated that the number of small organic compounds that could be synthesized using the rules of organic

* Corresponding author.

E-mail address: christian.feiler@hzg.de (C. Feiler).

<https://doi.org/10.1016/j.corsci.2019.108245>

Received 3 June 2019; Received in revised form 18 September 2019; Accepted 25 September 2019

Available online 27 September 2019

0010-938X/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 1

The elemental composition of commercially pure Mg (CP Mg 220 and CP Mg 342, numbers indicate the contained iron impurities) used in this study. Balance is abbreviated as bal.

| Material | Element content/ wt. % | | | | | | | | | | |
|-----------|------------------------|---------|----------|---------|---------|----------|---------|---------|----------|----------|------|
| | Fe | Cu | Ni | Al | Mn | Ce | Zn | Si | Ca | Zr | Mg |
| CP Mg 220 | 0.0220 | 0.00050 | < 0.0002 | 0.01300 | 0.01500 | < 0.0004 | 0.00040 | 0.00530 | 0.0004 | 0.0005 | bal. |
| CP Mg 342 | 0.0342 | 0.00037 | < 0.0002 | 0.00402 | 0.00237 | 0.0007 | 0.00046 | 0.00071 | < 0.0001 | < 0.0005 | bal. |

chemistry is $\sim 10^{80}$ [28,29]. High-throughput synthesis techniques and computer-assisted synthesis will, no doubt, extend this exponential rise [30,31]. As there are effectively an infinite number of potentially useful small molecules available to solve materials problems now and in the future, the challenge is how to find these 'islands of utility' in a vast sea of possibilities. *In silico* methods employing artificial intelligence (machine learning and evolutionary algorithms) are essential to discover these new small molecule materials. Specifically, for small molecules that modulate corrosion behaviour in magnesium-based materials, chemical intuition alone is inadequate to find the best modulators and to predict their effect on kinetics of the corrosion process. To address this issue, here we show how machine learning-based quantitative structure property relationship (QSPR) methods, combined with the results of quantum chemical methods (density functional theory (DFT) calculations), can predict the corrosion inhibition efficiency of small organic molecules. [29,32–35]

2. Materials & methods

We used a recently published database of inhibition efficiencies (IE) for more than 150 compounds, measured on nine different magnesium-based materials (six Mg alloys and three grades of pure Mg), based on hydrogen evolution experiments. [6] Note that no experimental errors are available for many of the data points as the aim of the initial experimental study was to perform a high-throughput screening to identify promising inhibitors for these Mg-based materials. Promising candidates were subsequently investigated in more detail. Our study on *in silico* screening methods to model the performance of promising corrosion modulators used data for commercially pure magnesium (CP Mg 220, Mg containing 220 ppm iron impurities). As the majority of QSPR studies of corrosion inhibition use datasets that were measured at a fixed concentration, we also modelled percent inhibition efficiencies measured at inhibitor concentrations of 0.05 M in the computational study. This reduced the size of our training dataset to 109 molecules. We also excluded all inorganic substances (26) and molecules with molecular weights > 350 Da (10), as we were seeking small-molecule organic dissolution modulators. Furthermore, dissolution modulators in the database that contained under-represented molecular features (e.g. phenylphosphoric acid) as well as compounds that were investigated in form of their corresponding acid salts (e.g. L-Ornithine hydrochloride) were excluded from the model. The 71 compounds that remained were used to generate a model describing the relationship between molecular properties and the corrosion inhibition performance. *In vacuo* DFT calculations were performed at the TPSSH/def2SVP level of theory using Turbomole 7.2. [36] The hydrogen evolution tests were performed using eudiometers (Art. Nr. 2591-10-500 from Neubert-Glas, Germany) analogous to those employed in our previous studies. [6,7] Water displaced by evolved hydrogen was automatically weighted (SKX series from OHAUS coupled with USB data logger OHAUS 30268984) and the data recorded for further processing using an in-house Python script [7]. A ground sample of CP Mg was placed in the eudiometer container with 500 mL of electrolyte (0.5 wt.% NaCl). A reference value was determined from the normalized volume of hydrogen evolved after 20 h of immersion in the absence of a dissolution modulator ($V_{\text{Reference}}$). Subsequently, the specimen was exposed to an electrolyte solution containing 0.05 M of dissolution modulator for 20 h with the initial pH

being adjusted by NaOH/HCl to 6.8 ± 0.5 (V_{Additive}). The impact of the investigated modulator on the corrosion of CP Mg is described by the IE, which is calculated as follows:

$$\text{IE} = \frac{V_{\text{Reference}} - V_{\text{Additive}}}{V_{\text{Reference}}} \times 100$$

To elucidate the applicability of the artificial neural network (ANN) trained on data for CP Mg 220 to predict IE for a similar material, we additionally performed hydrogen evolution measurements for CP Mg containing 342 ppm iron impurity (CP Mg 342). The elemental composition of CP Mg materials that were used in this work is shown in Table 1. The values were determined by Optical Discharge Emission Spectroscopy (SPECTROLAB with Spark Analyser Vision software, Germany).

3. Results & discussion

The ability of computational modeling and simulation methods, such as density functional theory (DFT) and machine learning (ML), to predict corrosion inhibiting properties of organic molecules has been reviewed recently. [37] The use of molecular descriptors derived from DFT calculations in ANN models of corrosion inhibitor performance has been quite controversial. Some researchers [38–41] claim there is a correlation between DFT-calculated parameters (e.g. the energy differences ($\Delta E_{\text{H/L}}$) between the highest occupied (HOMO) and the lowest unoccupied molecular orbital (LUMO)) and corrosion inhibition efficiencies of small organic molecules. However, several other studies could not identify any relationship between experiment and theory. [42–45] In aluminium alloys, DFT-derived molecular descriptors of small organic compounds were reported to have almost zero correlation with corrosion inhibition. [5,43,46] Furthermore, it was concluded that the use of datasets containing too few or poorly chosen data points can lead to erroneous conclusions about correlations between DFT-derived properties and experimental corrosion inhibition performance [47]. However, it has not been determined whether DFT-derived parameters are useful for generating predictive models of corrosion inhibitor performance for other metals and alloys, such as Mg.

Hence, we initially investigated whether a correlation existed between DFT-derived parameters ($\Delta E_{\text{H/L}}$) and the IE of small organic molecules for CP Mg 220. Because impurities like iron, nickel and copper significantly alter the corrosion rate of Mg, [48–52] the HOMO-LUMO gap energy may be an important parameter as it indicates the affinity of the corrosion modulators for transition metal ions. The kinetic stability of an organic molecule [53,54] increases and hence, its reactivity decreases with increasing HOMO-LUMO gap energy. This parameter is a reliable indicator of the propensity of organic compounds to form complexes with metallic impurities in CP Mg. This is also consistent with the experimental observation [6] that compounds with small HOMO-LUMO gaps show higher IEs, while dissolution accelerating agents have larger HOMO-LUMO gaps (see Fig. 1 left). The RMSD for this trend improves after the dataset is split into aliphatic (39) and aromatic compounds (32) while R^2 decreases (see Fig. 1 right). Although the R^2 value is modest for this single parameter model, its p value is highly significant (2×10^{-11}).

Given the promising correlation between HOMO-LUMO gap and the IE for the 71 small organic molecules we employed this parameter as a

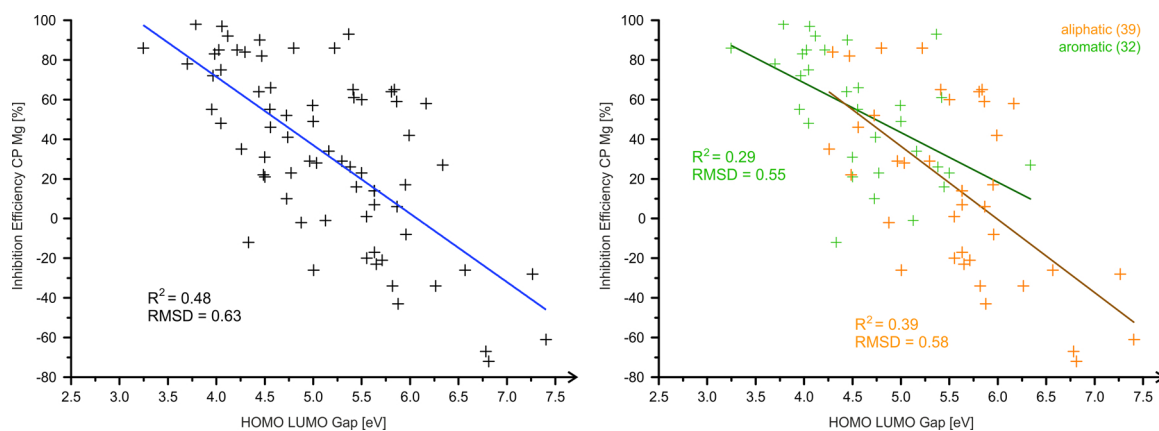


Fig. 1. Correlation of calculated (TPSSH/def2SVP) HL gaps with the experimentally measured [6] inhibition efficiencies for CP Mg 220 (left). Inhibition efficiencies are higher with smaller HOMO-LUMO gaps. Partitioning of the dataset into aromatic and aliphatic compounds revealed that aromatic compounds tend to have lower HOMO-LUMO gaps than aliphatic compounds while exhibiting equal values for IE (right).

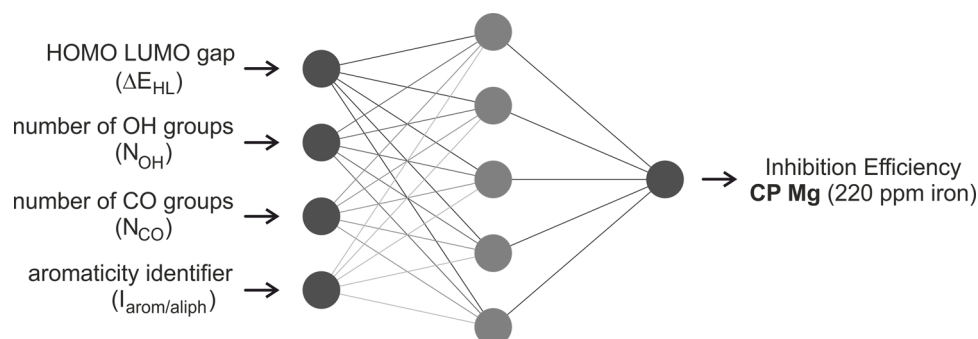


Fig. 2. ANN architecture employed for the prediction of the IE of magnesium dissolution modulators for CP Mg 220. The input layer accepts the four molecular descriptors (ΔE_{HL} , N_{OH} , N_{CO} , $I_{arom/aliph}$), the hidden layer generates the model, and the output layer node provides the calculated IE value for CP Mg 220.

molecular descriptor used to train the ANN model. Calculation of the HOMO-LUMO gap is computationally cheap and is relatively insensitive to basis set size. We found that different basis sets (def2SVP, def2TZVP, and def2QZVP) did not substantially influence the HOMO-LUMO gap trend in a data set consisting of three dissolution accelerators, three moderate inhibitors and three effective corrosion inhibitors. However, larger basis sets increased computational times ~ 100 -fold (for details see SI, section II). This analysis suggested that compounds with small HOMO-LUMO gap are generally better corrosion inhibitors for CP Mg. The correlation of the energy gaps with the inhibition efficiencies is quite high for this material in comparison to studies on Al based alloys where the correlation was essentially zero and the frontier orbital energies may already account for 50% of variance in the data. However, complementary features are needed to further improve the R^2 value of our model. Hence, we used two additional descriptors to train the ANN model (see Fig. 2), the number of hydroxyl and carbonyl groups in the molecules. These were considered important because they are generating chelating moieties (carboxylates and phenols) that are involved in binding the compounds to metals. A fourth molecular descriptor that distinguishes between aromatic and aliphatic molecules was also used because aliphatic compounds generally have higher HOMO-LUMO gaps than aromatic molecules for a given level of corrosion inhibition (see Fig. 1). A three layer feed forward network with one hidden layer was trained to predict the experimental IE values for CP Mg 220 [6] using the resilient back propagation (rprop+) algorithm in RStudio [55]. The source code for the ANN is provided in the supporting information (SI, section III).

To assess the ability of the ANN model to predict IE for new candidate modulators, the dataset was divided in a training (85%, 60 modulators) and test set (15%, 11 modulators). Partitioning of the test

sets was based on random selection. However, random test set selection from small heterogeneous data sets often creates large differences in the prediction performance of an ANN model. Consequently, a cross-validation approach was adopted where the test sets were randomly chosen six times to minimize selection biases for the model (for details see SI, section I). Fig. 3 shows the quality of the model prediction of modulators in the training and test sets. The RMSD values for the predictions of IE for the six cross-validation sets lie between 19% and 34%, consistent with the experimental errors that were observed.

The complete dataset was subsequently used to train an ANN model to make an *a priori* prediction of the performance of seven

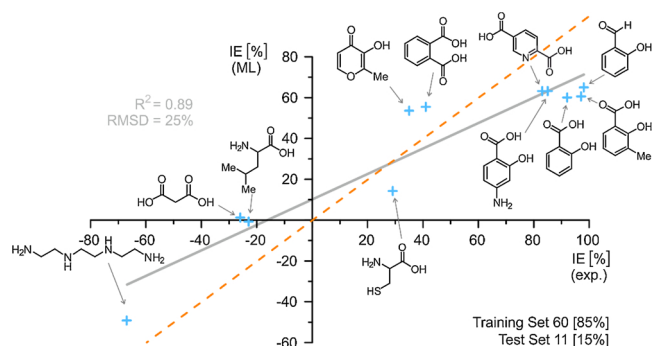


Fig. 3. Correlation of the predicted consensus test set IE with the experimentally derived IE for CP Mg 220. The dashed orange line constitutes perfect correlation between training and test set. Data points are depicted as light blue crosses. The grey line represents the linear least squares fit of the predicted and measured values in the used test set including the fits R^2 and RMSD values. The depicted RMSD value represents the mean deviation in absolute percent.

(corresponding to 10% of the dataset size used to train the model) hitherto untested organic corrosion modulators on CP Mg 220. For this sake, we were looking for compounds that have a similar composition (e.g. functional moieties, atom types) as the molecules used for training of the model since it will not correctly predict the performance of compounds outside of its domain. We excluded highly toxic chemicals and ordered only compounds that are low in price as expensive additives are not likely to be employed in applications since the cost of the protective coating would become too high. The set for blind validation of our ANN model consisted of three aliphatic compounds (1-(2-hydroxyethyl)piperazine (1), acetamide (2) and dimethylolpropionic acid (3)) and four aromatic compounds (pyridazine (4), trimesic acid (5), *n*-octyl gallate (6) and 2,3-pyrazinedicarboxylic acid (7)). This ratio of aromatic and aliphatic compounds was similar to that in the training set of the ANN. To estimate prediction errors and the robustness of the ML model, the ANN was retrained five times and each neural network used to predict the IE of the seven untested compounds. The ML predictions of properties of these seven untested compounds were validated experimentally using the same hydrogen evolution protocols used to generate the training set. The IE of 0.05 M aqueous solution of carboxylic acids was determined for their sodium salts (initial pH 6.8 ± 0.5) in the hydrogen evolution experiments. The errors of the ML prediction and the experimental investigation are summarized in Table 2 and in Fig. 4. In all of our experiments we assumed that the active surface area of the specimen does not change during the hydrogen evolution experiment and that the material composition is homogeneous in all samples. The experimental error was determined by quadratic error propagation based on the error of the reference measurement and the reproducibility of the hydrogen evolution experiments employing one of the seven potential corrosion modulators. Each experiment was performed three times. The experimental uncertainties are most likely caused by inhomogeneities in structure and composition of the as-cast material samples of CP Mg 220.

Table 2

Comparison of inhibition efficiencies (experimental vs. predicted for CP Mg 220) of seven compounds not used to train the ANN model. Calculated (TPSSh/def2SVP) ΔE_{HL} values are given in eV. Note that the IE of the compounds bearing a carboxylic acid moiety were determined for their corresponding sodium salt in the hydrogen evolution experiments. The experimental uncertainties were calculated from the errors of the reference experiment and the subsequent hydrogen evolution measurements containing dissolution modulators. Each hydrogen evolution value is the mean of three experiments with the exception of the value for piperazine which is the mean of four experiments. Averaged values for the pH value and the surface area are provided (see SI for details, section IV.1). The uncertainty of the IE values predicted by the ML model is based on the predicted values of six independently trained ANNs.

| Compound | ΔE_{HL} [eV] | IE ML [%] | IE exp. [%] | H ₂ volume [mL · cm ⁻²] | Final pH | Surface area [cm ²] |
|---------------------------------|-----------------------------|-----------|-------------|--|----------|---------------------------------|
| Reference: 0.5% Sodium chloride | — | — | 0 | 29.3 ± 1.3 | 10.7 | 5.5 |
| 1 1-(2-Hydroxyethyl)piperazine | 6.43 | -35 ± 7 | -96 ± 43 | 57.5 ± 12.7 | 9.7 | 5.5 |
| 2 Acetamide | 6.64 | -5 ± 17 | 2 ± 19 | 28.7 ± 5.7 | 10.8 | 5.3 |
| 3 Dimethylolpropionic acid | 6.50 | 33 ± 33 | 54 ± 23 | 13.5 ± 6.7 | 10.9 | 5.1 |
| 4 Pyridazine | 3.91 | 65 ± 10 | 25 ± 14 | 22.0 ± 4.0 | 10.7 | 4.6 |
| 5 Trimesic acid | 4.97 | 62 ± 8 | 90 ± 9 | 2.9 ± 2.5 | 10.7 | 5.1 |
| 6 <i>n</i> -Octyl gallate | 4.31 | 61 ± 7 | 89 ± 8 | 3.3 ± 2.3 | 8.5 | 5.3 |
| 7 2,3-Pyrazinedicarboxylic acid | 3.93 | 64 ± 6 | 59 ± 13 | 11.9 ± 3.8 | 11.0 | 5.4 |

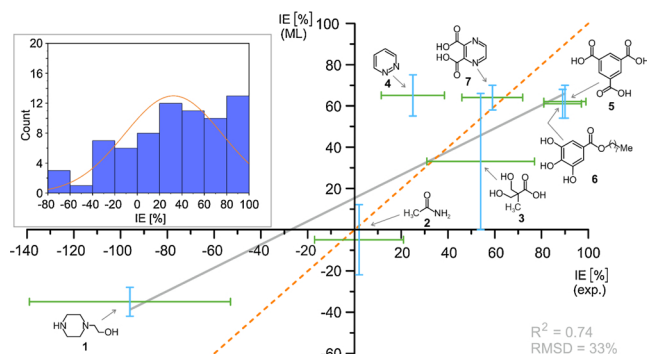


Fig. 4. Correlation of the predicted IE with the experimentally derived IE for CP Mg 220. The orange, dashed line represents perfect correlation. The green error bars represent the error of the experimentally derived IE and the light blue bars represent the variation of the IE value predicted by the ANN. The grey line represents the linear least squares fit of the predicted and measured values in the used test set including the fits R^2 and RMSD values. The RMSD value represents the mean deviation in absolute percent. The inset (top left) shows the IE distribution of the training dataset. The orange line represents the corresponding distribution curve.

For all seven compounds, the ML model correctly predicted whether the compounds inhibited or accelerated Mg dissolution. The predicted values for two out of seven candidates (acetamide (2) and the pyrazine derivative (7)) matched the experimentally determined performance (within the error margin of the methods). Furthermore, the predicted values of trimesic acid (5), the benzoic acid ester derivative (6) as well as pyridazine (4) are in the same range as the experimentally derived values. The performance of the heterocycle (4) is overestimated whereas the performance of the corrosion modulators (5) and (6) are underestimated by roughly 30% compared to the experimentally derived

inhibition efficiencies. The propionic acid derivative **3** shows the largest prediction deviation indicating that its composition is not well represented in the used training set. The piperazine derivative **1** displayed the lowest (negative) IE value experimentally although the predicted IE was less negative. This is due to the small number of compounds in the training set that accelerate dissolution (negative IE values, only four modulators have IE values < -40% and none have an IE < -75% (see Fig. 4, top left)). The ANN is unable to accurately predict IE values outside the range of the training set. However, the model correctly predicts that the piperazine derivative **1** will act as a corrosion-accelerating agent for CP Mg 220, which is consistent with the hydrogen evolution experiments. To quantify the trend between the measured and predicted values we performed a Pearson correlation test using the means of the values predicted by the model and the values determined by the subsequent experiments. The test resulted in a correlation coefficient of 0.88, a good positive correlation. The p-value was 0.009, indicating acceptable statistical significance. The conducted correlation test confirms that there is indeed a correlation between the measured and predicted values.

In summary, the IE predictions of the used ANN are in good agreement with the subsequent experiments as the ordering of the investigated compounds is correct although the absolute values are not predicted accurately. Clearly, a quantitatively accurate forecast requires a larger training data set, and use of improved descriptors as these are the largest factor in model quality. It is very difficult to put a number on the size of training database that is necessary to predict highly accurate IEs of unspecified new compounds. Additionally, a deeper understanding of the underlying degradation mechanisms (e.g. complexation of impurities or Mg^{2+} , blocking of the Mg surface by adsorbed inhibitors or precipitation of coordination polymers) would also allow a more tailored preselection of the candidate modulators for experimental and computational testing. Large scale testing, preferably with high-throughput experimental techniques [56–58] needs to be implemented for the industrially most relevant alloys, especially Mg. If ML models can be trained on experimental data with considerably larger molecular diversity (a wider range of molecular features) the





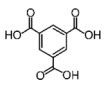
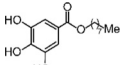

domain of applicability of the model will be larger, the predictions more accurate and it will extend further into accessible chemical space. In principle similar models could recapitulate the properties of formulated inhibitors as well, provided that sufficient experimental data were available to train the models.

To estimate how sensitive the trained ML corrosion inhibitor models are to changes in the metal or alloy composition, we performed hydrogen evolution tests for CP Mg 342. The experiments were performed under the same conditions as those used to generate the data for commercial pure Mg (220 ppm iron impurities) that provided the training set for the ML model. The results are presented in Table 3 and Fig. 5.

In general, CP Mg 342 corrodes slower than CP Mg 220. The amount of hydrogen that is evolved in the reference measurements (0.5 wt.% NaCl in 500 mL electrolyte) is lower for CP Mg 342 (9.10 mL/cm²) than for CP Mg 220 (29.3 mL/cm²). This is slightly counterintuitive, as one might expect faster degradation rates from the material of lower purity. However, corrosion susceptibility of Mg depends not only on the total amount of iron impurity but also on its distribution in the Mg matrix. Lower corrosion rate displayed by CP Mg 342 is caused by a significantly lower amount of Si in its composition. Silicon is known to promote nucleation and growth of Fe-rich particles by introducing a solidification interval. [59] There is a strong correlation between the IEs for the Mg with 220 ppm and 342 ppm Fe (R^2 value of 0.87). The measured IE values for the CP Mg material containing the higher iron content are ~20% larger on average for compounds **2**, **3**, **4**, **6** and **7**. This is in good agreement with the slower corrosion observed for CP Mg 342. Trimesic acid (**5**) has a similar IE for both Mg materials. However, piperazine derivative **1** greatly accelerates dissolution of the more iron rich material resulting in a high RMSD value of 97% for the linear least squares fit (see Fig. 5a). The fit for the correlation of compounds **2–7**, excluding the piperazine derivative **1**, exhibits a slightly lower R^2 value of 0.80 with a highly improved RMSD value of 25% (see Fig. 5b). [60] To confirm the trend displayed by compound **1** we selected six dissolution accelerators (*Tris*(hydroxymethyl)aminomethane (TRIS), ethylenediaminetetraacetic acid (EDTA), nitrilotriacetic acid (NTA),

Table 3

Comparison of inhibition efficiencies of the seven investigated compounds for CP Mg 220 and CP Mg 342. Experimental uncertainties were calculated analogues to the procedure described in Table 1. Each hydrogen evolution value is the mean of three replicates. Averaged values for the pH value and the surface area are provided (see SI section IV.2 for more details).

| Compound | | IE exp. CP Mg 220 ppm [%] | H ₂ volume [mL·cm ⁻²] | Final pH | Surface area [cm ²] | IE exp. CP Mg 342 ppm [%] | H ₂ volume [mL·cm ⁻²] | Final pH | Surface area [cm ²] |
|---|---|---------------------------------|--|----------|------------------------------------|---------------------------------|--|----------|------------------------------------|
| Reference: 0.5% Sodium chloride | NaCl | 0 | 29.3 ± 1.3 | 10.7 | 5.5 | 0 | 9.1 ± 1.1 | 10.8 | 6.1 |
| 1 1-(2-Hydroxyethyl)piperazine |  | -96 ± 43 | 57.5 ± 12.7 | 9.7 | 5.5 | -360 ± 33 | 42.0 ± 2.0 | 9.5 | 5.6 |
| 2 Acetamide |  | 2 ± 19 | 28.7 ± 5.7 | 10.8 | 5.3 | 21 ± 20 | 7.2 ± 0.9 | 10.8 | 5.4 |
| 3 Dimethylolpro-pionic acid |  | 54 ± 23 | 13.5 ± 6.7 | 10.9 | 5.1 | 64 ± 18 | 3.2 ± 0.2 | 10.7 | 5.6 |
| 4 Pyridazine |  | 25 ± 14 | 22.0 ± 4.0 | 10.7 | 4.6 | 48 ± 20 | 4.7 ± 0.8 | 10.5 | 5.5 |
| 5 Trimesic acid |  | 90 ± 9 | 2.9 ± 2.5 | 10.7 | 5.1 | 87 ± 17 | 1.2 ± 0.1 | 10.6 | 5.4 |
| 6 n-Octyl gallate |  | 89 ± 8 | 3.3 ± 2.3 | 8.5 | 5.3 | 96 ± 17 | 0.4 ± 0.2 | 7.8 | 5.4 |
| 7 2,3-Pyrazinedi-carboxylic acid |  | 59 ± 13 | 11.9 ± 3.8 | 11.0 | 5.4 | 98 ± 17 | 0.2 ± 0.04 | 10.8 | 5.5 |

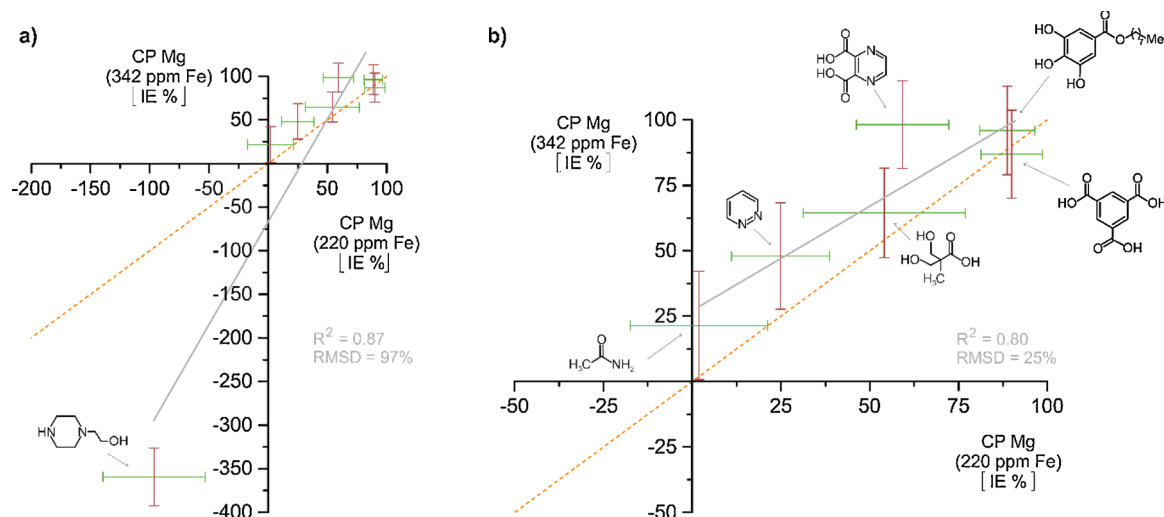


Fig. 5. Correlation between inhibiting performance of seven investigated compounds for two types of magnesium: CP Mg 220 and CP Mg 342 (a). The correlation improves after exclusion of the piperazine derivate (b). The orange, dashed line indicates perfect correlation. The grey line represents the linear least squares fit of the measured values for the performance correlation of the two investigated materials. The RMSD values represent the mean deviation in absolute percent.

3,4-Dihydroxybenzoic acid, glycine and piperazine) that were used to train the ANN and measured the corresponding IEs for CP Mg 342. The IEs of all seven accelerators were significantly more negative for CP Mg 342 compared to that of CP Mg 220 (for details see SI, section IV.3). This is reasonable as inhibitors are mostly targeting the active iron-rich particles (constituting less than 0.05% of the material) while accelerators are chelating Mg that constitutes 99.95% of the investigated material. The influence of minor impurities becomes less significant in the presence of corrosion accelerating agents, so that the rate of magnesium dissolution becomes similar for both magnesium materials of commercial purity. Consequently, the ML model trained on data for CP Mg 220 may be applied to predict the performance of corrosion inhibiting agents for a Mg material with similar properties, as compounds 2-7 had a similar effect on the degradation rate of CP Mg 342. Hence, no additional time-consuming testing campaign would be required. However, in spite of similar dissolution rates of CP Mg 220 and CP Mg 342 in presence of accelerating agents, CP Mg 220 and CP Mg 342 do not display the same tendency in terms of IE. Hence, the trained ANN cannot be applied to predict the inhibition efficiencies of corrosion accelerators for a material with similar properties as used for training of the ML model based on IE values. For example, the calculated IE values of piperazine derivative **1** are significantly more negative for CP Mg 342 albeit the observation that the amount of evolved hydrogen in its presence is lower for CP Mg 342 (42 mL/cm²) than for CP Mg 220 (57.5 mL/cm²). Consequently, corrosion acceleration seems to need a more precise way for quantification by putting more weight to the factor reflecting the amount of actually consumed Mg rather than the IE, as the latter is largely dependent on the performance of the investigated material during reference measurement.

Next, we aim to use a wider range of molecular descriptors, such as chemical potentials, electron density distributions, molecular polarizabilities [61] and molecular signatures [62], together with sparse feature selection methods, (e.g. LASSO [63,64]) to improve the predictive power and applicability of our ML models. Additionally, larger and more chemically diverse data sets would be required to broaden the domain of applicability of IE models in chemical space (as was demonstrated for ML-based QSPR models of aqueous solubility and flash point predictions for small organic compounds) [65,66]. For example, molecular descriptors generated by the software package alvaDesc show promising initial results in our preliminary investigations. [67] We will also expand the size and chemical diversity of our experimental database to allow generation of more broadly applicable and robust ML models as compounds contained in the used dataset were selected based

on their ability to bind Fe ions. As this might be one of the reasons for the strong correlation between HOMO-LUMO gap values and IE we are planning to elucidate if the correlation remains valid for a representative set of randomly chosen dissolution modulators that are not targeting contained impurities. We will also investigate formation of complexes between organic modulators with Mg²⁺, Ni²⁺ and Fe²⁺/Fe³⁺ using quantum chemical and molecular dynamics techniques to better understand the interactions of the small organic molecules with metallic impurities. Recalculation of HOMO-LUMO gap energies of organic inhibitors in an aqueous environment, with full speciation (all possible neutral and ionized forms of the corrosion modulators) may also result in more accurate *in silico* predictions of the IE values. Finally, as most applications employ Mg-based alloys rather than pure Mg materials we aim to establish a predictive model for the corrosion modulation of alloy series, like Mg-Al, Mg-RE, Mg-Ca for structural engineering, bio-resorbable implants and Mg-air battery applications.

4. Conclusion

An important finding of this study is that frontier orbital energies can be important in modelling corrosion inhibition of organic compounds for CP Mg, although this is not the case for Al [5] and Cu alloys. [42,44] It is therefore likely that the importance of this parameter is metal-dependent and not necessarily transferable to other materials as there are different corrosion mechanisms in play. We found that ML models trained on HOMO-LUMO gap energies combined with simple molecular descriptors could recapitulate the IE of small organic compounds for commercially pure Mg containing 220 ppm iron impurities (CP Mg 220). However, it is not clear whether the significant contribution of frontier orbital energy gaps is a consequence of an intrinsic property of CP Mg such as reactivity, or the ability of modulators to form complexes with transition metal impurities like iron, that are important concerning the degradation of CP Mg. Re-plating of transition metal impurities released during Mg dissolution, thought to play a key role in the kinetics of the corrosion of Mg-based materials [48–51], may be prevented by the organic compounds rendering these molecules efficient corrosion inhibitors. In spite of diverse and not completely unravelled inhibition mechanisms in play, we obtained a robust and qualitatively predictive model for corrosion modulators for CP Mg based on the used training set. The model predicted the corrosion modulation performance of seven hitherto untested small organic molecules in good agreement with subsequently experimental measurements. We also showed that the trained model can predict inhibition

efficiencies of corrosion inhibitors for a metal (CP Mg 342) with composition similar to that of the material used to generate it (CP Mg 220). These results strongly suggest that computational ML models will be useful for the rapid discovery and optimization of new organic corrosion inhibitors or dissolution modulators. Although experimental validation will remain an important final step, this ML approach will allow for preselection of the best performing substances for a specific target application, saving a significant amount of time and resources.

Funding

Funding by HZG MMDi IDEA project is gratefully acknowledged. DM thanks China Scholarship Council for the award of fellowship and funding (No. 201607040051). RM gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 192346071 - SFB 986. The authors declare no competing financial interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.corsci.2019.108245>.

References

- S. Schreiber, The small-molecule approach to biology, *C&EN* 81 (2003) 51–61.
- A.R. Horswill, S.N. Savinov, S.J. Benkovic, A systematic method for identifying small-molecule modulators of protein–protein interactions, *Proc. Natl. Acad. Sci. U. S. A.* 44 (2004) 15591–15596.
- C.J. Gerry, S. Schreiber, Chemical probes and drug leads from advances in synthetic planning and methodology, *Nat. Rev. Drug Discov.* 17 (2018) 333–352.
- D. Bi, X. Li, J.V. Milić, D.J. Kubicki, N. Pellet, J. Luo, T. LaGrange, P. Mettraux, L. Emsley, S.M. Zakeeruddin, M. Grätzel, Multifunctional molecular modulators for perovskite solar cells with over 20% efficiency and high operational stability, *Nat. Comm.* 9 (2018) 4482–4492.
- D.A. Winkler, M. Breedon, P. White, A.E. Hughes, E.D. Sapper, I. Cole, Using high throughput experimental data and in silico models to discover alternatives to toxic chromate corrosion inhibitors, *Corros. Sci.* 106 (2016) 229–235.
- S.V. Lamaka, B. Vaghefinazari, D. Mei, R.P. Petruskas, D. Höche, M.L. Zheludkevich, Comprehensive screening of Mg corrosion inhibitors, *Corros. Sci.* 128 (2017) 224–240.
- D. Mei, S.V. Lamaka, C. Feiler, M.L. Zheludkevich, The effect of small-molecule bio-relevant organic components at low concentration on the corrosion of commercially pure Mg and Mg-0.8Ca alloy: an overall perspective, *Corros. Sci.* 153 (2019) 258–271.
- M.S. Reisch, Confronting the looming hexavalent chromium ban, *C&EN* 95 (2017) 28–29.
- T.H. Muster, A.E. Hughes, S.A. Furman, T. Harvey, N. Sherman, S. Hardin, P. Corrigan, D. Lau, F.H. Scholes, P.A. White, M. Glenn, J. Mardel, S.J. Garcia, J.M.C. Mol, A rapid screening multi-electrode method for the evaluation of corrosion inhibitors, *Electrochim. Acta* 54 (2009) 3402–3411.
- T.G. Harvey, S.G. Hardin, A.E. Hughes, T.H. Muster, P.A. White, T.A. Markley, P.A. Corrigan, J. Mardel, S.J. Garcia, J.M.C. Mol, A.M. Glenn, The effect of inhibitor structure on the corrosion of AA2024 and AA7075, *Corros. Sci.* 53 (2011) 2184–2190.
- C. Blawert, W. Dietzel, E. Ghali, G. Song, Anodizing treatments for magnesium alloys and their effect on corrosion resistance in various environments, *Adv. Eng. Mater.* 8 (2006) 511–533.
- L. Wang, J. Zhou, J. Liang, J. Chen, Thermal control coatings on magnesium alloys prepared by plasma electrolytic oxidation, *Appl. Surf. Sci.* 280 (2013) 151–155.
- A. Němcová, P. Skeldon, G.E. Thompson, B. Pacal, Effect of fluoride on plasma electrolytic oxidation of AZ61 magnesium alloy, *Surf. Coat. Technol.* 232 (2013) 827–838.
- S.V. Lamaka, G. Knörnschild, D.V. Snihirova, M.G. Taryba, M.L. Zheludkevich, M.G.S. Ferreira, Complex anticorrosion coating for ZK30 magnesium alloy, *Electrochim. Acta* 55 (2009) 131–141.
- F. Cao, G.-L. Song, A. Atrens, Corrosion and passivation of magnesium alloys, *Corros. Sci.* 111 (2016) 835–845.
- M. Esmaily, J.E. Svensson, S. Fajardo, N. Biribilis, G.S. Frankel, S. Virtanen, R. Arrabal, S. Thomas, L.G. Johansson, Fundamentals and advances in magnesium alloy corrosion, *Prog. Mater. Sci.* 89 (2017) 92–193.
- W.J. Joost, P.E. Krajewski, Towards magnesium alloys for high-volume automotive applications, *Scr. Mater.* 128 (2017) 107–112.
- A.A. Luo, Magnesium casting technology for structural applications, *J. Magnes. Alloy.* 1 (2013) 2–22.
- G. Song, Control of biodegradation of biocompatible magnesium alloys, *Corros. Sci.* 49 (2007) 1696–1701.
- R. Willumeit, F. Feyerabend, N. Huber, Magnesium degradation as determined by artificial neural networks, *Acta Biomater.* 9 (2013) 8722–8729.
- B.J.C. Luthringer, F. Feyerabend, R. Willumeit-Römer, *Magnes. Res.* 27 (2014) 142–154.
- I. Morgenthal, O. Andersen, C. Kostmann, G. Stephani, T. Studnitzky, F. Witte, B. Kieback, Highly porous magnesium alloy structures and their properties regarding degradable implant application, *Adv. Eng. Mater.* 16 (2014) 309–318.
- C. Shuai, S. Li, S. Peng, P. Peng, Y. Laid, C. Gao, Biodegradable metallic bone implants, *Mater. Chem. Front.* (2019) Advance Article.
- T. Zhang, Z. Tao, J. Chen, Magnesium–air batteries: from principle to application, *Mater. Horiz.* 1 (2014) 196–206.
- D. Höche, S.V. Lamaka, B. Vaghefinazari, T. Braun, R. Petruskas, M. Fichtner, M.L. Zheludkevich, Performance boost for primary magnesium cells using iron complexing agents as electrolyte additives, *Sci. Rep.* 8 (2018) 7578.
- F. Cheng, J. Chen, Metal–air batteries: from oxygen reduction electrochemistry to cathode catalysts, *Chem. Soc. Rev.* 41 (2012) 2172–2192.
- A.H. Lipkus, Q. Yuan, K.A. Lucas, S.A. Funk, W.F. Bartelt III, R.J. Schenck, A.J. Trippe, Structural diversity of organic chemistry. A scaffold analysis of the CAS registry, *J. Org. Chem.* 73 (2008) 4443–4451.
- D.A. Erlanson, S.W. Pesik, R.E. Hubbard, W. Jahnke, H. Jhoti, Twenty years on: the impact of fragments on drug discovery, *Nat. Rev. Drug Discov.* 15 (2016) 605–619.
- D.A. Winkler, Predicting the performance of organic corrosion inhibitors, *Metals* 7 (2017) 553–561.
- J. Li, S.G. Ballmer, E.P. Gillis, S. Fujii, M.J. Schmidt, A.M.E. Palazzolo, J.W. Lehmann, G.F. Morehouse, M.D. Burke, Synthesis of many different types of organic small molecules using one automated process, *Science* 347 (2015) 1221–1226.
- M.H.S. Segler, M. Preuss, M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI, *Nature* 555 (2018) 604–610.
- T.C. Le, D.A. Winkler, Discovery and optimization of materials using evolutionary approaches, *Chem. Rev.* 116 (2016) 6107–6132.
- F.F. Chen, M. Breedon, P. White, C. Chu, D. Mallick, S. Thomas, E. Sapper, I. Cole, Correlation between molecular features and electrochemical properties using an artificial neural network, *Mater. Des.* 112 (2016) 410–418.
- J. Ma, R.P. Sheridan, A. Liaw, G.E. Dahl, V. Svetnik, Deep neural nets as a method for quantitative structure–activity relationships, *J. Chem. Inf. Model.* 55 (2015) 263–274.
- T. Fujita, D.A. Winkler, Understanding the roles of the “Two QSARs”, *J. Chem. Inf. Model.* 56 (2016) 269–274.
- TURBOMOLE V7.2, A Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2019, since 2007; available from TURBOMOLE GmbH, 2018, <http://www.turbomole.com>.
- I.B. Obot, D.D. Macdonald, Z.M. Gaseem, Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors. Part I: an overview, *Corros. Sci.* 99 (2015) 1–30.
- H. Ju, Z.-P. Kai, Y. Li, Aminic nitrogen-bearing polydentate Schiff ; base compounds as corrosion inhibitors for iron in acidic media: a quantum chemical calculation, *Corros. Sci.* 50 (2008) 865–871.
- K.F. Khaled, Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: a QSAR model, *Corros. Sci.* 53 (2011) 3457–3465.
- E.E. Ebenso, T. Arslan, F. Kandemirli, N. Caner, I. Love, Quantum Chemical Studies of Some Rhodanine Azosulpha Drugs as Corrosion Inhibitors for Mild Steel in Acidic Medium, *Int. J. Quantum Chem.* 110 (2010) 1003–1018.
- K. Barouni, A. Kassale, A. Albourine, O. Jbara, B. Hammouti, L. Bazzi, Amino acids as corrosion inhibitors for copper in nitric acid medium: experimental and theoretical study, *J. Mater. Environ. Sci.* 5 (2014) 456–463.
- A. Kokalj, *Electrochim. Acta* 56 (2010) 745–755.
- D.A. Winkler, M. Breedon, A.E. Hughes, F.R. Burden, A.S. Barnard, T.G. Harvey, I. Cole, Towards chromate-free corrosion inhibitors: Structure–property models for organic alternatives, *Green Chem.* 16 (2014) 3349–3357.
- N. Kovačević, I. Milošev, A. Kokalj, How relevant is the adsorption bonding of imidazoles and triazoles for their corrosion inhibition of copper? *Corros. Sci.* 124 (2017) 25–34.
- A. Kokalj, D. Costa, Molecular modeling of corrosion inhibitors, *Encyclopedia of Interfacial Chemistry: Surface Science and Electrochemistry*, (2018), pp. 332–345.
- D.A. Winkler, Biomimetic molecular design tools that learn, evolve, and adapt, *Beilstein J. Org. Chem.* 13 (2017) 1288–1302.
- G. Gece, The use of quantum chemical methods in corrosion inhibitor studies, *Corros. Sci.* 50 (2008) 2981–2992.
- S.V. Lamaka, D. Höche, R.P. Petruskas, C. Blawert, M.L. Zheludkevich, A new concept for corrosion inhibition of magnesium: Suppression of iron re-deposition, *Electrochem. Commun.* 62 (2016) 5–8.
- D. Höche, C. Blawert, S.V. Lamaka, N. Scharnagl, C. Mendis, M.L. Zheludkevich, The effect of iron re-deposition on the corrosion of impurity-containing magnesium, *Phys. Chem. Chem. Phys.* 18 (2016) 1279–1291.
- D. Mercier, J. Swiatowska, S. Zanna, A. Seyeux, P. Marcus, Role of segregated Iron at grain boundaries on Mg corrosion, *J. Electrochem. Soc.* 165 (2018) C42–C49.
- E. Michailidou, H.N. McMurray, G. Williams, Quantifying the role of transition metal electrodeposition in the cathodic activation of corroding magnesium, *J. Electrochem. Soc.* 165 (2018) C195–C205.
- J. Li, W. Sun, B. Hurley, A.A. Luo, R.G. Buchheit, Cu redistribution study during the corrosion of AZ91 using a rotating ring-disk collection experiment, *Corros. Sci.* 112 (2016) 760–764.
- J.-I. Aihara, Correlation found between the HOMO–LUMO energy separation and the chemical reactivity at the most reactive site for isolated-pentagon isomers of fullerenes, *Phys. Chem. Chem. Phys.* 2 (2000) 3121–3125.
- T. Würger, C. Feiler, F. Musil, G.B.V. Feldbauer, D. Höche, S.V. Lamaka,

- M.L. Zheludkevich, R.H. Meißner, Data Science Based Mg Corrosion Engineering, *Front. Mater.* 6 (2019) 53.
- [55] RStudio Team, *RStudio: Integrated Development for R*, URL RStudio, Inc, Boston, MA, 2015 <http://www.rstudio.com>.
- [56] P.A. White, G.B. Smith, T.G. Harvey, P.A. Corrigan, M.A. Glenn, D. Lau, S.G. Hardin, J. Mardel, T.A. Markley, T.H. Muster, N. Sherman, S.J. García, J.M.C. Mol, A.E. Hughes, A new high-throughput method for corrosion testing, *Corros. Sci.* 58 (2012) 327–331.
- [57] S.J. García, T.H. Muster, Ö. Özkanat, N. Sherman, A.E. Hughes, H. Terryn, J.H.W. de Wit, J.M.C. Mol, The influence of pH on corrosion inhibitor selection for 2024-T3 aluminium alloy assessed by high-throughput multielectrode and potentiodynamic testing, *Electrochim. Acta* 55 (2010) 2457–2465.
- [58] P.A. White, A.E. Hughes, S.A. Furman, N. Sherman, P.A. Corrigan, M.A. Glenn, D. Lau, S.G. Hardin, T.G. Harvey, J. Mardel, T.H. Muster, S.J. García, C. Kwakernaak, J.M.C. Mol, High-throughput channel arrays for inhibitor testing: proof of concept for AA2024-T3, *Corros. Sci.* 51 (2009) 2279–2290.
- [59] L. Yang, X. Zhou, S.-M. Liang, R. Schmid-Fetzer, Z. Fan, G. Scamans, J. Robson, G. Thompson, Effect of traces of silicon on the formation of Fe-rich particles in pure magnesium and the corrosion susceptibility of magnesium, *J. Alloys Compd.* 619 (2015) 396–400.
- [60] D.L.J. Alexander, A. Tropsha, David A. Winkler, Beware of R2: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models, *J. Chem. Inf. Model.* 55 (2015) 1316–1322.
- [61] D.M. Wilkins, A. Grisafi, Y. Yang, K.U. Lao, R.A. DiStasio Jr., M. Ceriotti, Accurate Molecular Polarizabilities by Coupled-cluster Theory and Machine Learning (submitted), (2019).
- [62] H.M. Kayello, N.K.R. Tadisina, N. Shlonimskaya, J.J. Biernacki, D.P. Visco Jr., An Application of Computer-Aided Molecular Design (CAMD) using the signature molecular descriptor—part 1. Identification of surface tension reducing agents and the search for shrinkage reducing admixtures, *J. Am. Ceram. Soc.* 97 (2014) 365–377.
- [63] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Statist. Soc. B* 58 (1996) 267–288.
- [64] N. Gauraha, Introduction to the LASSO: a convex optimization approach for high-dimensional problems, *Resonance* 23 (2018) 439–464.
- [65] T.C. Le, M. Ballard, P. Casey, M. Liu, D.A. Winkler, Illuminating flash point: comprehensive prediction models, *Mol. Info.* 34 (2015) 18–27.
- [66] M. Salahinejad, T.C. Le, D.A. Winkler, Aqueous Solubility Prediction: Do Crystal Lattice Interactions Help? *Mol. Pharm.* 10 (2013) 2757–2766.
- [67] alvaDesc 1.0, A Development of Alvascience, available from (2019) <http://www.alvascience.com>.