BRIEF COMMUNICATION OPEN

Check for updates

CORDATA: an open data management web application to select corrosion inhibitors

Tiago L. P. Galvão[1]✉, Inês Ferreira[1], Alena Kuznetsova[1,2], Gerard Novell-Leruth[1,3], Ci Song[4], Christian Feiler[4], Sviatlana V. Lamaka[4], Claúdia Rocha[2], Frederico Maia[2], Mikhail L. Zheludkevich[4,5], José R. B. Gomes[3] and João Tedim[1]

The large amount of corrosion inhibition efficiencies in literature, calls for a more efficient way to organize, access and compare this information. The CORDATA open data management application (https://datacor.shinyapps.io/cordata/) can help select appropriate corrosion inhibitors for application specific challenges.

npj Materials Degradation (2022)6:48 ; https://doi.org/10.1038/s41529-022-00259-9

INTRUDUCTION

The ban of one of industry's main solutions against corrosion, i.e. compounds based on hexavalent chromium[1], has recently started in Europe due to health and environmental issues, resulting in a need to find effective replacements[2]. The large and growing amounts of reported corrosion inhibition efficiencies existing in literature and obtained over the years as a result of this event is expected to keep increasing in volume. Moreover, the development of high-throughput testing methodologies[3–8], has allowed to obtain significant databases in shorter timeframes for different substrates, application conditions and molecular structures[3,9,10]. This has called for the development of a data driven application, such as the one developed in this work, which will allow academic and industry researchers to swiftly select the most adequate condition specific corrosion inhibitor to be embedded directly into protective coating systems or through smart nanocontainers. We envision the CORDATA app to be the first step in the corrosion inhibitor selection process before going to the laboratory to perform further research and development activities. Although there are many accounts in literature focusing on corrosion inhibition efficiencies, to the best of our knowledge, this is the first web application dealing with data management for this particular issue. It allows to more efficiently compare many different data sources at the same time, thus making it easier to find appropriate solutions that were already tested experimentally, but that might be lost in the middle of a large volume of experimental data obtained in the past. Moreover, the dynamic nature of a data management web application, will allow it to grow in size and evolve in functionality throughout the years, adapting to the needs of the corrosion science community to better solve societal challenges through open data.

RESULTS AND DISCUSSION

Data driven technologies and machine leaning are among the latest developments and most promising approaches in corrosion science to guide the discovery and design of more effective and environmentally benign corrosion inhibitors and protective coating systems[3,11–25]. However, one of the main challenges dealing with the application of machine learning to understand and design protective systems is building the datasets required for training the predictive models[26,27]. The collection of experimental data, as well as data management and curation, are among the most time-consuming tasks in the machine learning workflow. Therefore, a web application like the one presented herein, will fulfill two main purposes: (1) it can be used by scientists and engineers working in academia and industry to quickly compare the performance of different corrosion inhibitors and select the most appropriate condition specific corrosion inhibitor for each application; and (2) it will provide a framework to organize cured datasets for different substrates which will trigger further machine learning and data driven developments to design corrosion inhibitors.

A general view of the CORDATA application can be seen in Fig. 1 and accessed free of charge through the following url: https://datacor.shinyapps.io/cordata/. The web application was designed to work on personal computers, tablets and mobile phones, and includes several different functionalities (Fig. 2), such as: (1) search for the appropriate application conditions, such as the type of metal and alloy, the possible synergistic combination of inhibitors, the minimum efficiency, select a range of temperature and pH, and a minimum aggressive salt concentration; (2) quickly check the inhibitor structure and the reference used to obtain its corrosion inhibition efficiency; (3) search specific corrosion inhibitors through an internal search engine; (4) select and compare other properties and aspects of the data, such as the molecular weight, SMILES notation, measurement time, corrosion inhibitor concentration, synergistic inhibitor concentration, experimental methodology, literature reference, and name and institution of the contributor that added each specific data entry; and (5) a user interface with detailed instructions is available for users to submit additional data, request the whole dataset or provide their feedback. A spreadsheet template file can be downloaded for users to include their own data, while the whole updated dataset will be available to contributors, to be used in their own machine learning and data driven research.

At the time of this publication nearly five thousand corrosion inhibition efficiencies and almost four hundred compounds have already been added to the database. The data originates from

[1]CICECO-Aveiro Institute of Materials, Department of Materials and Ceramic Engineering, University of Aveiro, 3810-193 Aveiro, Portugal. [2]Smallmatek - Small Materials and Technologies, Lda., Rua Canhas, 3810-075 Aveiro, Portugal. [3]CICECO-Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal. [4]Institute of Surface Science, Helmholtz-Zentrum Hereon, 21502 Geesthacht, Germany. [5]Institute for Materials Science, Faculty of Engineering, Kiel University, 24103 Kiel, Germany. ✉email: tlpgalvao@ua.pt

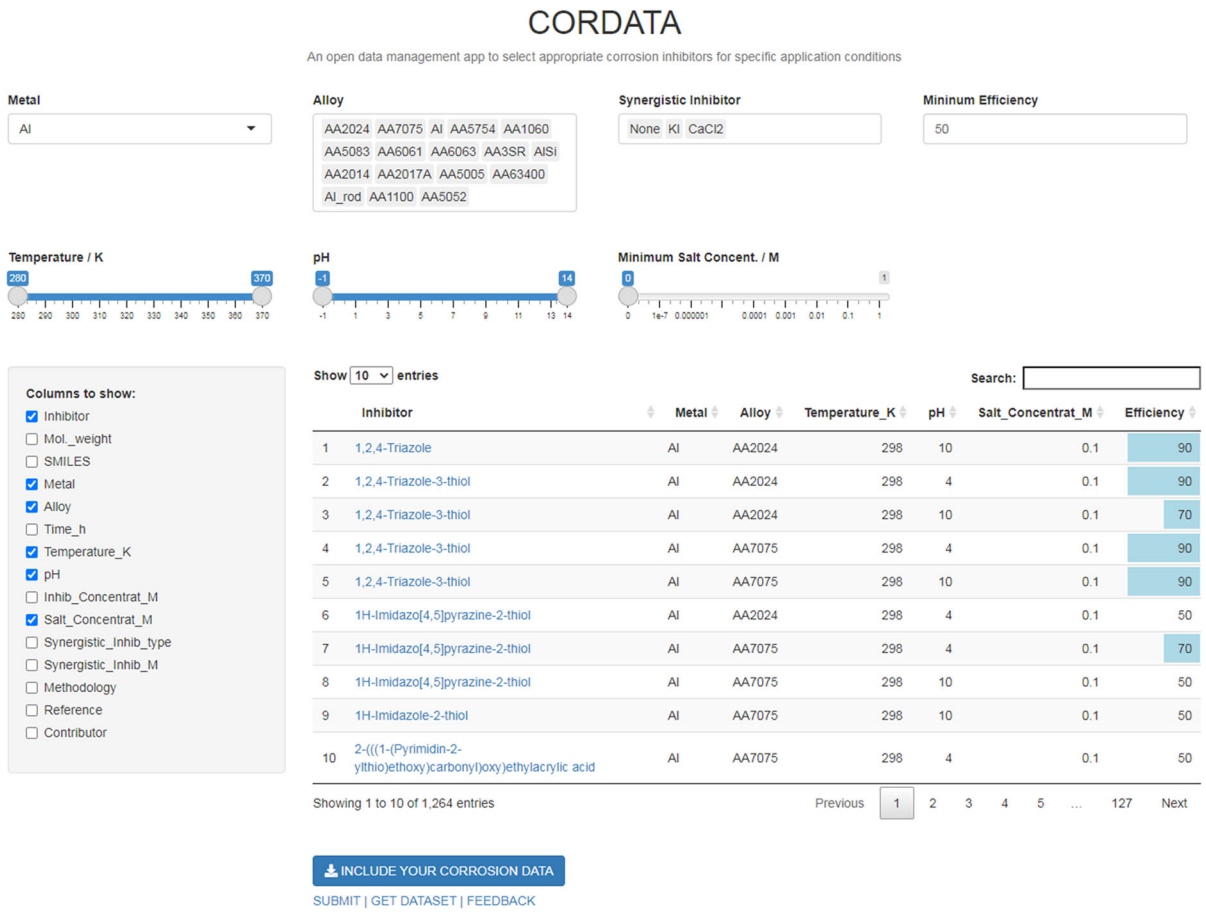

Fig. 1 Graphical user interface (GUI) of the CORDATA web application. The CORDATA GUI (https://datacor.shinyapps.io/cordata/) includes digital features to search and select the data, according to intended application conditions of the corrosion inhibitors, as well as the main table with corrosion inhibition efficiencies and corresponding measurement properties.

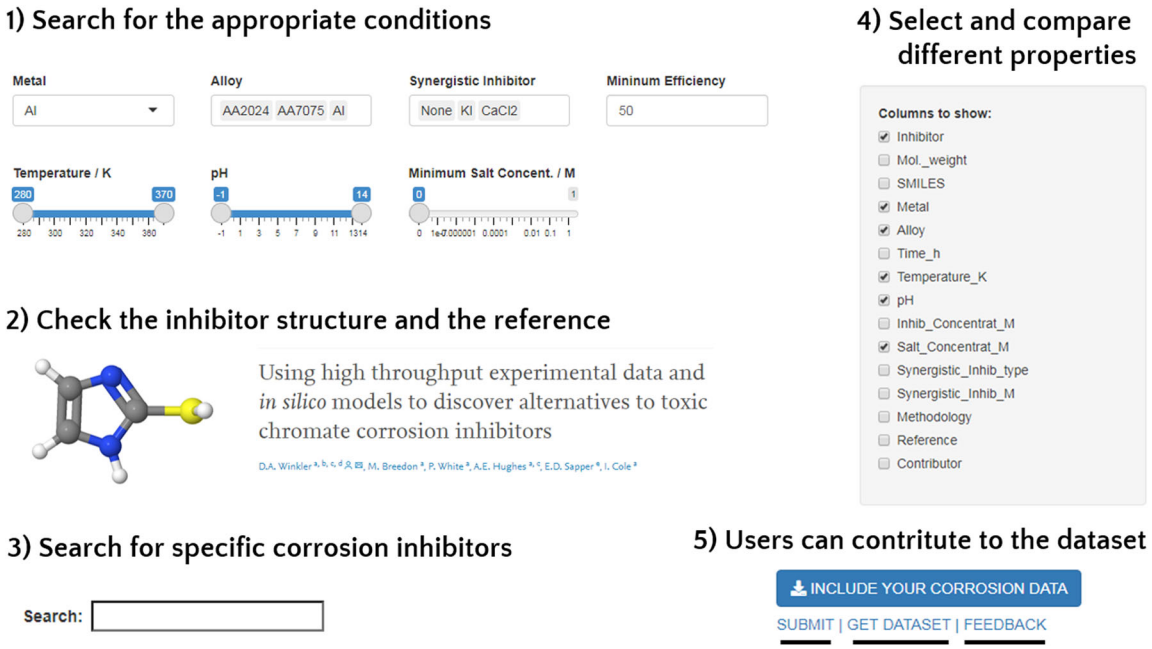

Fig. 2 Main features of the CORDATA web application. The CORDATA application includes (1) digital features to search the appropriate application conditions of the corrosion inhibitors, (2) the possibility to check the inhibitor structure and the respective literature reference, (3) an embedded search engine to search for specific corrosion inhibitors, (4) an option to visualize additional properties, and (5) instructions for users to contribute with data and feedback.

Table 1. Overview of the data included in the CORDATA database for different metals and respective alloys.

Property	Aluminum	Copper	Magnesium	Iron	All
Efficiencies	2011	272	1470	1176	4934
Compounds	173	24	198	57	374
Metals/Alloys	17	2	13	6	39
pH range	[−0.6,14]	[0,13]	[6.8,8.2]	[−0.5,9.6]	[−0.6,14]
Salt concentration range / M	[0,2]	[0,0.6]	[0.001,0.15]	[0,1.2]	[0,2]
Synergistic inhibitors	Yes	Yes	No	Yes	Yes
References	47	25	7	47	121

more than one hundred and twenty publications, mainly for aluminum, copper, magnesium, iron and their main alloys. More specific information about the data included in the database can be found in Table 1.

The total number of efficiency values and compounds are already in a sufficient amount to find efficient corrosion inhibitor solutions for a broad number of application cases and conditions, thus it is expected to be immediately helpful for corrosion scientists and engineers working on the design of more efficient corrosion protective systems. Nevertheless, the data currently included in the application is still only a small part of all the information existing in literature. This number will increase over the years, as more data will be added by the authors and by other research groups that see value contributing to the database, while the web application gains traction among the corrosion science community.

METHODS

The open data management web application developed in this work was built using the R programming language[28], which is a free coding framework for statistical computing and graphical representation. In particular, it employed mainly the Shiny package[29], which makes it easy to build interactive web applications from R.

DATA AVAILABILITY

The authors declare that the data used to perform this work is available within the application described in the paper (https://datacor.shinyapps.io/cordata/). Moreover, the data in other formats is also available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The authors declare that the code of the CORDATA application is available from the corresponding author upon reasonable request.

Received: 11 April 2022; Accepted: 26 May 2022;
Published online: 17 June 2022

REFERENCES

1. Park, R. M. et al. Hexavalent chromium and lung cancer in the chromate industry: a quantitative risk assessment. *Risk Anal.* **24**, 1099 (2004).
2. Still, C. Boeing names CSIRO a supplier of the year. *CSIRO* https://www.csiro.au/en/News/News-releases/2017/Boeing-names-CSIRO-a-supplier-of-the-year (2017).
3. Winkler, D. A. et al. Using high throughput experimental data and in silico models to discover alternatives to toxic chromate corrosion inhibitors. *Corros. Sci.* **106**, 229 (2016).
4. Kallip, S., Bastos, A. C., Zheludkevich, M. L. & Ferreira, M. G. S. A multi-electrode cell for high-throughput SVET screening of corrosion inhibitors. *Corros. Sci.* **52**, 3146 (2010).
5. García, S. J. et al. The influence of pH on corrosion inhibitor selection for 2024-T3 aluminium alloy assessed by high-throughput multielectrode and potentiodynamic testing. *Electrochim. Acta* **55**, 2457 (2010).
6. White, P. A. et al. A new high-throughput method for corrosion testing. *Corros. Sci.* **58**, 327 (2012).
7. White, P. A. et al. Towards materials discovery: assays for screening and study of chemical interactions of novel corrosion inhibitors in solution and coatings. *N. J. Chem.* **44**, 7647 (2020).
8. Zabula, A. V. et al. Screening of molecular lanthanide corrosion inhibitors by a high-throughput method. *Corros. Sci.* **165**, 108377 (2020).
9. Harvey, T. G. et al. The effect of inhibitor structure on the corrosion of AA2024 and AA7075. *Corros. Sci.* **53**, 2184 (2011).
10. Lamaka, S. V. et al. Comprehensive screening of Mg corrosion inhibitors. *Corros. Sci.* **128**, 224 (2017).
11. Winkler, D. A. Predicting the performance of organic corrosion inhibitors. *Metals* **7**, 553 (2017).
12. Winkler, D. A. et al. Towards chromate-free corrosion inhibitors: structure–property models for organic alternatives. *Green. Chem.* **16**, 3349 (2014).
13. Galvão, T. L. P., Novell-Leruth, G., Kuznetsova, A., Tedim, J. & Gomes, J. R. B. Elucidating structure-property relationships in aluminum alloy corrosion inhibitors by machine learning. *J. Phys. Chem. C.* **124**, 5624 (2020).
14. Kokalj, A. et al. Simplistic correlations between molecular electronic properties and inhibition efficiencies: Do they really exist? *Corros. Sci.* **179**, 108856 (2021).
15. Breedon, M., Per, M. C., Cole, I. S. & Barnard, A. S. Molecular ionization and deprotonation energies as indicators of functional coating performance. *J. Mater. Chem. A* **2**, 16660 (2014).
16. Fernandez, M., Breedon, M., Cole, I. S. & Barnard, A. S. Modeling corrosion inhibition efficacy of small organic molecules as non-toxic chromate alternatives using comparative molecular surface analysis (CoMSA). *Chemosphere* **160**, 80 (2016).
17. Chen, F. F. et al. Correlation between molecular features and electrochemical properties using an artificial neural network. *Mater. Des.* **112**, 410 (2016).
18. Würger, T. et al. Data science based Mg corrosion engineering. *Front. Mater.* **6**, 53 (2019).
19. Feiler, C. et al. In silico screening of modulators of magnesium dissolution. *Corros. Sci.* **163**, 108245 (2020).
20. Würger, T. et al. Exploring structure-property relationships in magnesium dissolution modulators. *npj Mater. Degrad.* **5**, 1 (2021).
21. Schiessler, E. J. et al. Predicting the inhibition efficiencies of magnesium dissolution modulators using sparse machine learning models. *npj Comput. Mater.* **7**, 1 (2021).
22. Somers, A. E. et al. Advances in the development of rare earth metal and carboxylate compounds as corrosion inhibitors for steel. *Corros. Eng. Sci. Technol.* **55**, 311 (2020).
23. Milošev, I. et al. Electrochemical, surface-analytical, and computational DFT study of alkaline etched aluminum modified by carboxylic acids for corrosion protection and hydrophobicity. *J. Electrochem. Soc.* **166**, C3131 (2019).
24. Milošev, I. et al. The effect of anchor group and alkyl backbone chain on performance of organic compounds as corrosion inhibitors for aluminum investigated using an integrative experimental-modeling approach. *J. Electrochem. Soc.* **167**, 061509 (2020).
25. Milošev, I. et al. The effects of perfluoroalkyl and alkyl backbone chains, spacers, and anchor groups on the performance of organic compounds as corrosion inhibitors for aluminum investigated using an integrative experimental-modeling approach. *J. Electrochem. Soc.* **168**, 071506 (2021).
26. Moosavi, S. M., Jablonka, K. M. & Smit, B. The role of machine learning in the understanding and design of materials. *J. Am. Chem. Soc.* **142**, 20273 (2020).
27. Coelho, L. B. et al. Reviewing machine learning of corrosion prediction in a data-oriented perspective. *npj Mater. Degrad.* **6**, 1 (2022).
28. R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing* https://www.r-project.org/ (2021).
29. Chang, W. et al. Shiny: Web Application Framework for R. *R Studio* https://shiny.rstudio.com/ (2021).

ACKNOWLEDGEMENTS

This work was developed within the scope of the project CICECO-Aveiro Institute of Materials, UIDB/50011/2020 & UIDP/50011/2020, financed by national funds through the FCT/MEC and when appropriate co-financed by FEDER under the PT2020 Partnership Agreement. It was also financed in the framework of the project DataCor (refs. POCI-01-0145-FEDER-030256, PTDC/QUI-QFI/30256/2017 and https://datacorproject.wixsite.com/datacor) and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement ID 101007430 (COAT4LIFE).

AUTHOR CONTRIBUTIONS

T.L.P.G built the application. T.L.P.G, I.F., A.K., G.N.L., C.S., C.F., S.V.L., C.R. and F.M. collected information and organized the whole dataset for different metals. T.L.P.G, M.L.Z., J.R.B.G and J.T. contributed to develop the application concept, data organization strategy, and supervised the work. All co-authors read and edited the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Tiago L. P. Galvão.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.