Eyesight to AI: Discovery of effective corrosion modulators via predictive machine learning models

T. Würger, C. Song, B. Vaghefinazari, A. Lisitsyna, P. Fischer, G. Wiese, M. Zheludkevich, S. Albarqouni, S. V. Lamaka, C. Feiler

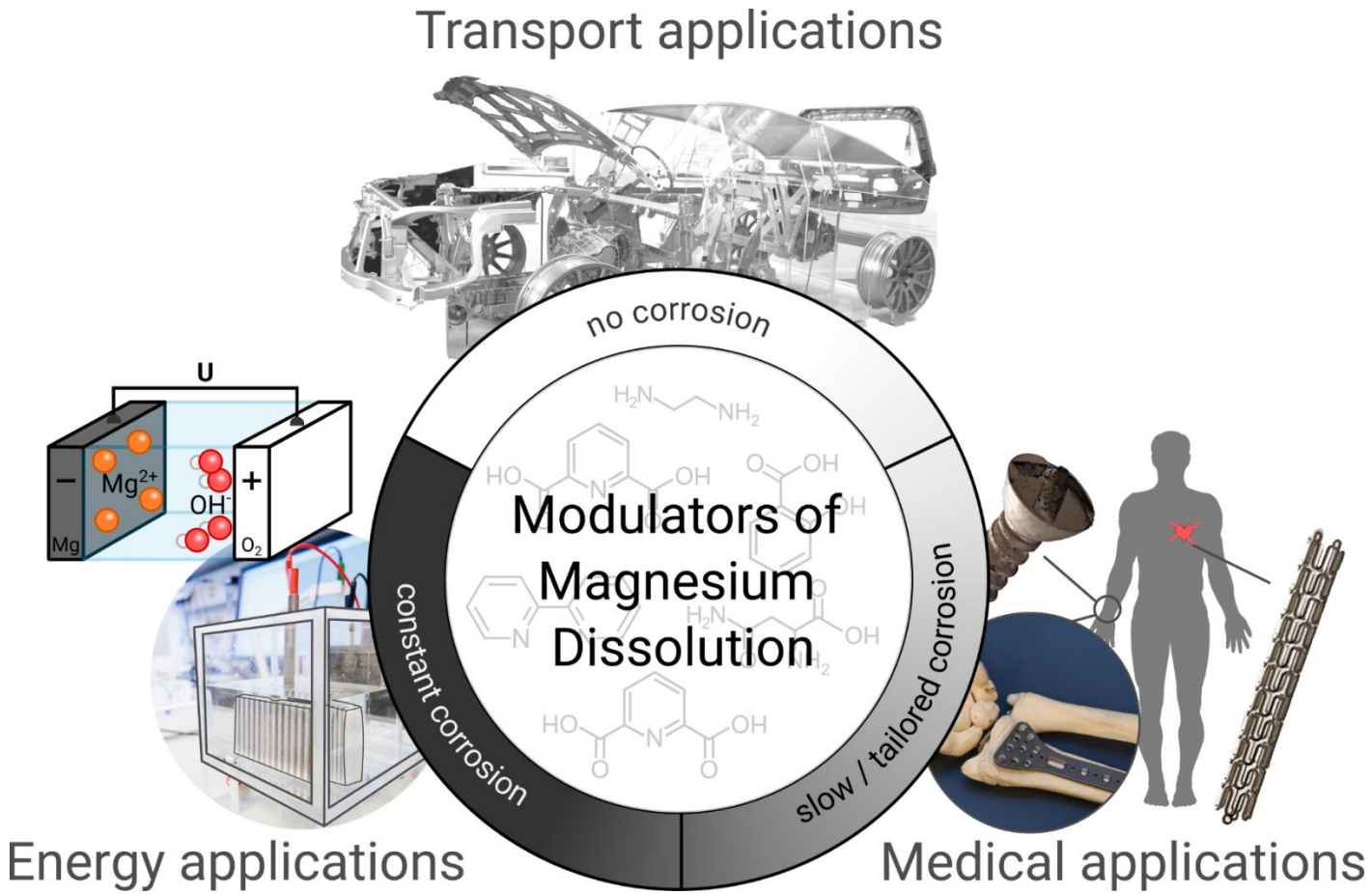
Helmholtz-Zentrum Hereon

MSE Day

14.11.2023

Helmholtz-Zentrum hereon

Dissolution Modulators



- Magnesium is the lightest structural metal
- Highly versatile engineering material
- Degradation control is vital for application

- Small organic molecules can be employed as dissolution modulators

→ Chemical Abstracts Service database: $1.5 \cdot 10^{8}$

→ Estimated number of synthezisable compounds: $1 \cdot 10^{63}$

- Data-driven approaches depict great tools to screen this vast chemical space

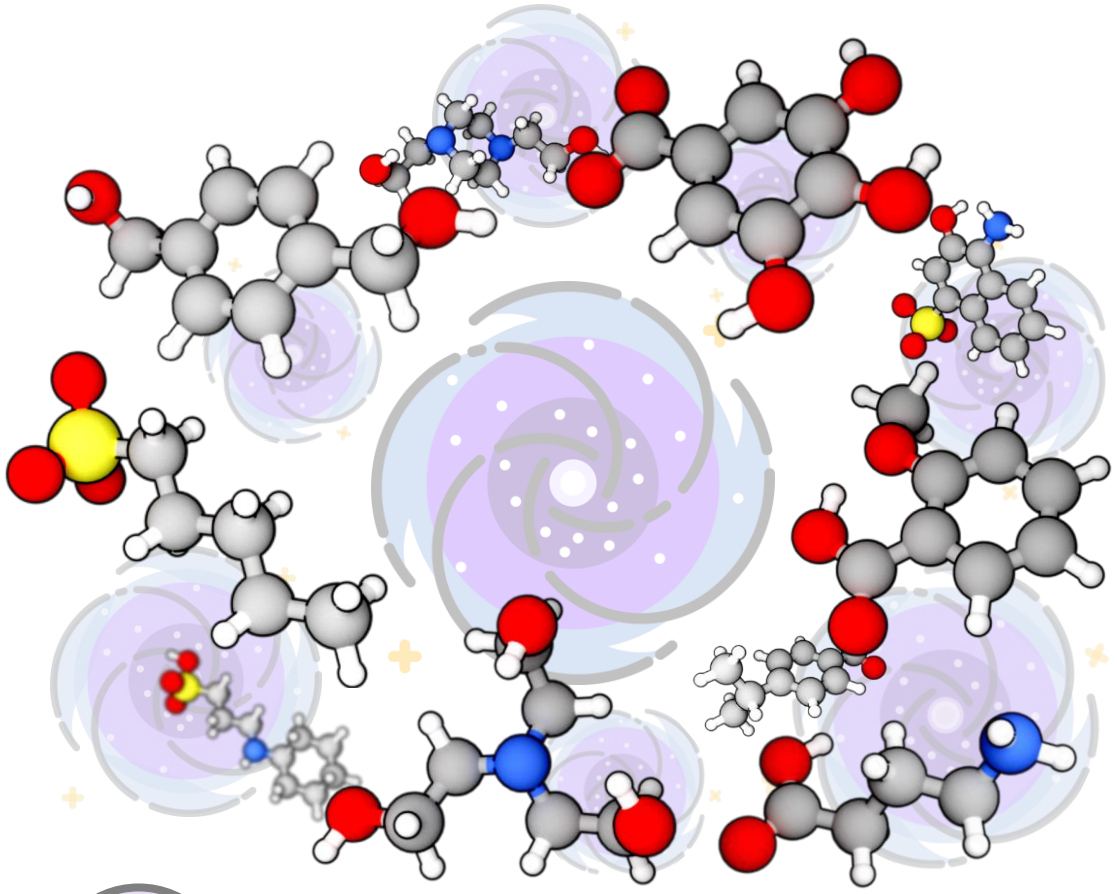
Images by courtesy of S. V. Lamaka, C. Blawert, B. Vaghefinazari and T. Würger.

A. Mullard, *Nature* **2017**, *549*, 445-447.
D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, *Nature Reviews Drug Discovery* **2016**, *15*, 605-619.
D. Höche, S. V. Lamaka, B. Vaghefinazari, T. Braun, R. Petrauskas, M. Fichtner, M.L. Zheludkevich, *Scientific Reports* **2018**, *8*, 7578.
S. V. Lamaka, B. Vaghefinazari, D. Mei, R. P. Petrauskas, D. Höche, M. L. Zheludkevich, *Corrosion Science* **2017**, *128*, 224-240.

Helmholtz-Zentrum hereon

Exploring the Chemical Space

Thought Experiment: “Just” a Matter of Time



Approximately **$1.0 \cdot 10^{21}$** stars in the observable universe

- Automated setup to quantify influence on corrosion rate
- Ten experiments in parallel (duration 24 h)

Chemical Abstracts Service database lists **$1.5 \cdot 10^{8}$** organic chemicals

~ 41095 years

Estimated number of synthesisable compounds amounts to **$1.0 \cdot 10^{63}$**

~ $2.73 \cdot 10^{59}$ years

Helmholtz-Zentrum hereon

The AI² project

HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

Phase 1

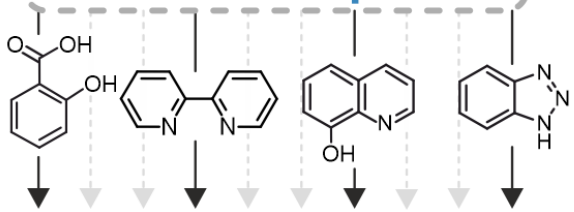
Phase 2

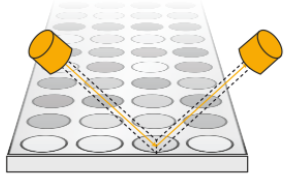
Sviatlana Lamaka

Shadi Albarqouni

Chemical Space

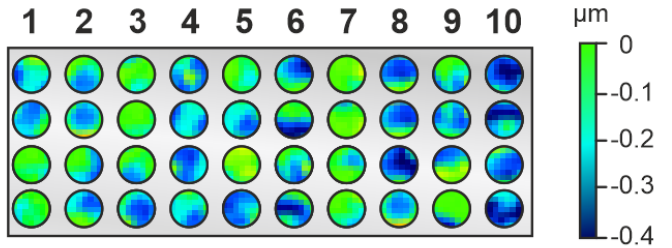
Multi-well experiment





Topograhpy assessment & Image aquisition

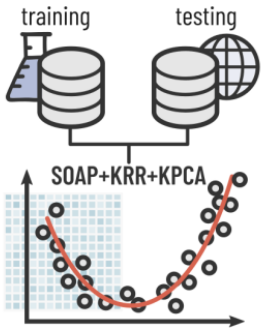
Ground truth generation

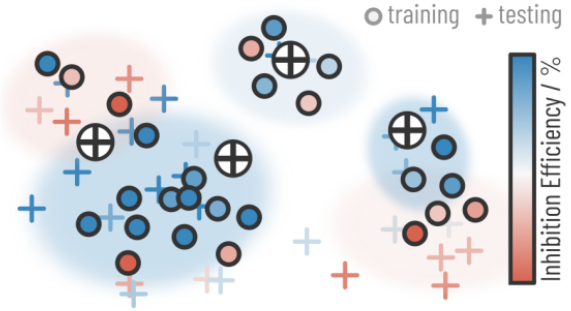


Training data for image recognition

Prediction of untested compouds

image acquisition





Training database for QSPR

Input

Output

Optical image

Volume loss from virtual heatmap

Generation of Quantitative Structure-Property Relationship models

UNet-based Image-to-Image translation

HELMHOLTZ MUNICH

Helmholtz-Zentrum hereon

A. Lisitsyna, C. Song, T. Würger, B. Vaghefinazari, M. L. Zheludkevich, S. Albarqouni, S. V. Lamaka, C. Feiler, manuscript in preparation.

Slide 9 3/4

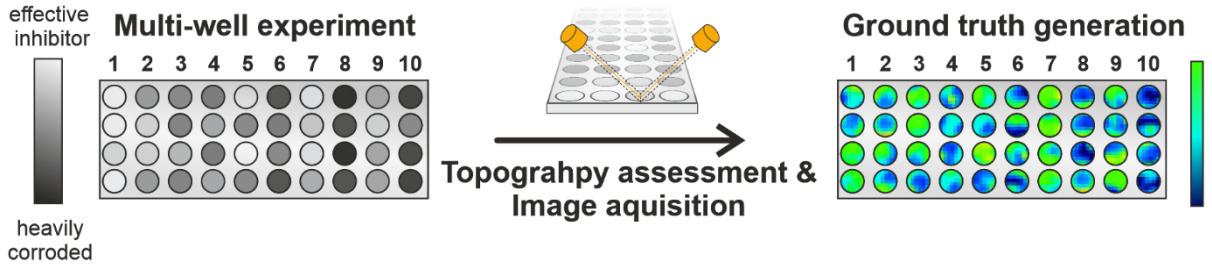
- The project team asked me to remove all the technical content from the slides as we are on the verge of submitting it to a peer-reviewed journal

- However, we are happy to discuss in more detail in personal meetings and are also very open to potential collaborations!

Summary

Mirror mirror on the wall who is the best inhibitor of them all?

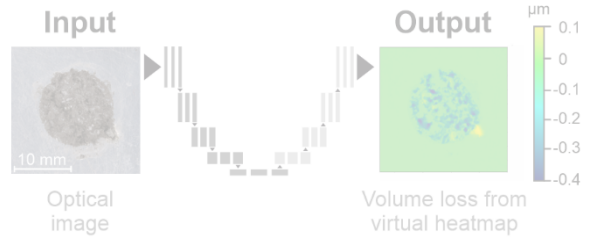
Determination of weight loss from optical appearance of corrosion imprints



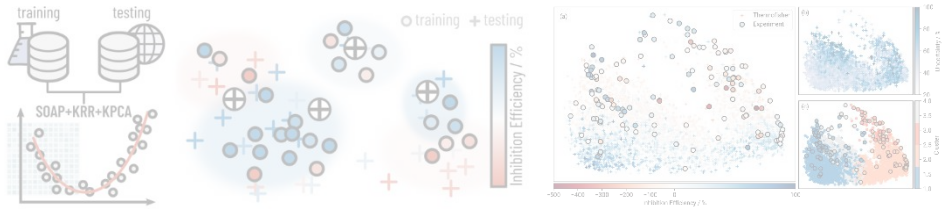
- Data generation using a multi-well setup
- Generation of heatmaps and determination of volume loss in the well by profilometer
- 200 distinct compounds have been tested

- Determination of volume loss based on optical appearance
- Elimates need to perform manual (time-consuming) analysis of imprints
- Correlation between volume loss from experiment and virtual heatmap
- Highly sensitive to minimal offsets during annotation step

UNet-based generation of virtual heatmaps



Performance prediction of untested chemicals by QSPR



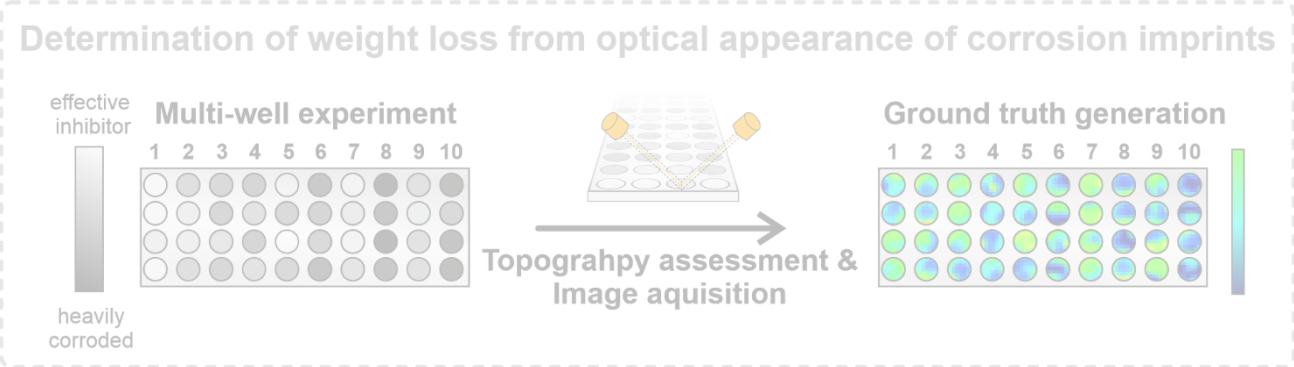
- Training of SOAP-based QSPR model based on ground truth
- Use of additional factors for compound selection (toxicity, solubility, price, uncertainty) rather than only based on molecular similarites
- First blind testing conducted, Next step: active learning and use of volume loss derived from virtual heatmaps as target for training of QSPR models

Helmholtz-Zentrum hereon

A. Lisitsyna, C. Song, T. Würger, B. Vaghefinazari, M. L. Zheludkevich, S. Albarqouni, S. V. Lamaka, C. Feiler, manuscript in preparation.

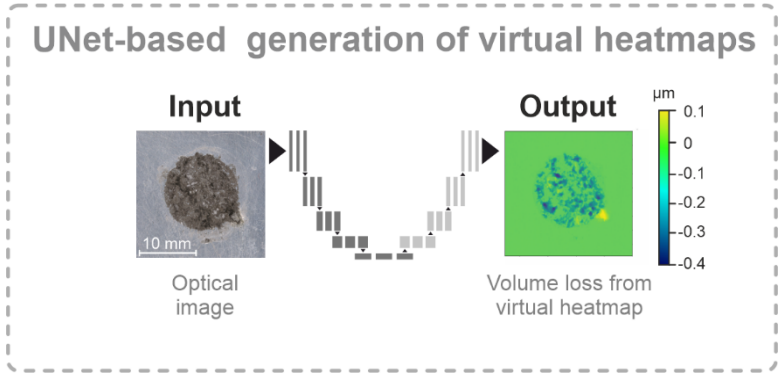
Summary

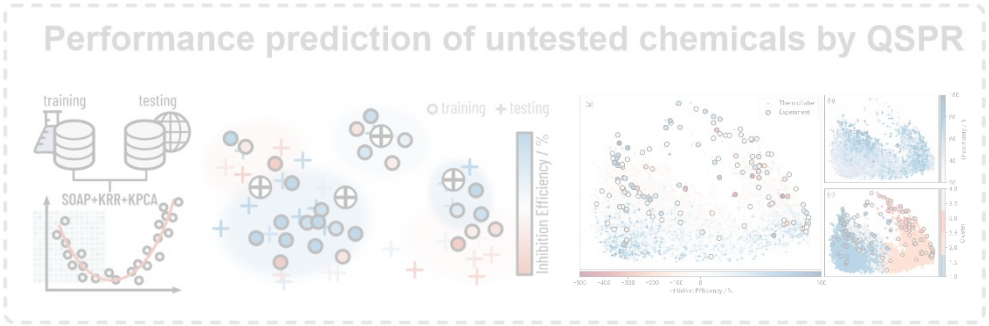
Mirror mirror on the wall who is the best inhibitor of them all?



- Data generation using a multi-well setup
- Generation of heatmaps and determination of volume loss in the well by profilometer
- 200 distinct compounds have been tested

- Determination of volume loss based on optical appearance
- Elimates need to perform manual (time-consuming) analysis of imprints
- Correlation between volume loss from experiment and virtual heatmap
- Highly sensitive to minimal offsets during annotation step





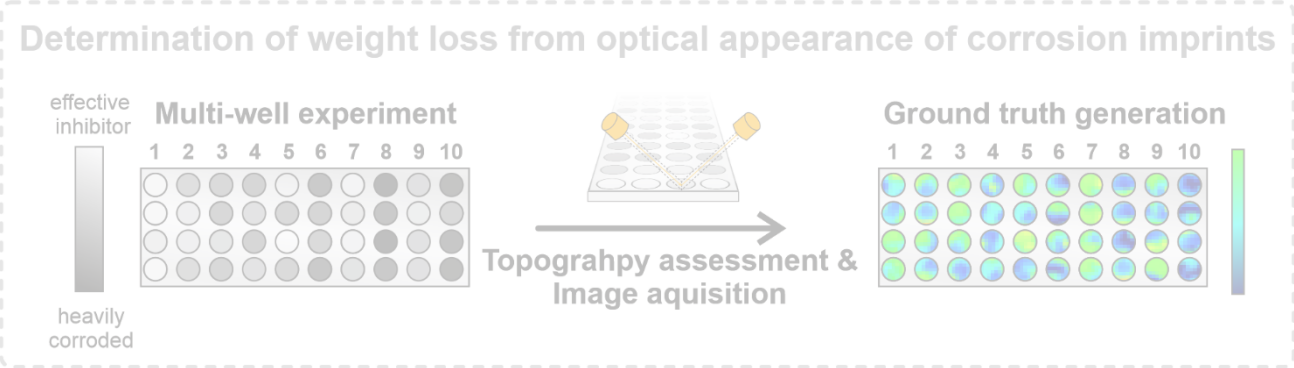
- Training of SOAP-based QSPR model based on ground truth
- Use of additional factors for compound selection (toxicity, solubility, price, uncertainty) rather than only based on molecular similarites
- First blind testing conducted, Next step: active learning and use of volume loss derived from virtual heatmaps as target for training of QSPR models

Helmholtz-Zentrum hereon

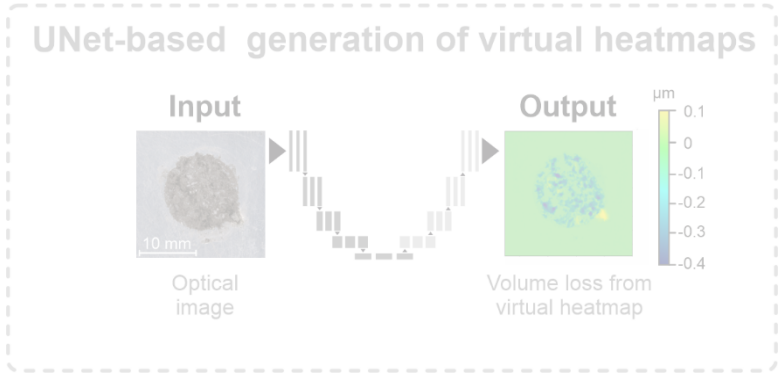
A. Lisitsyna, C. Song, T. Würger, B. Vaghefinazari, M. L. Zheludkevich, S. Albarqouni, S. V. Lamaka, C. Feiler, manuscript in preparation.

Summary

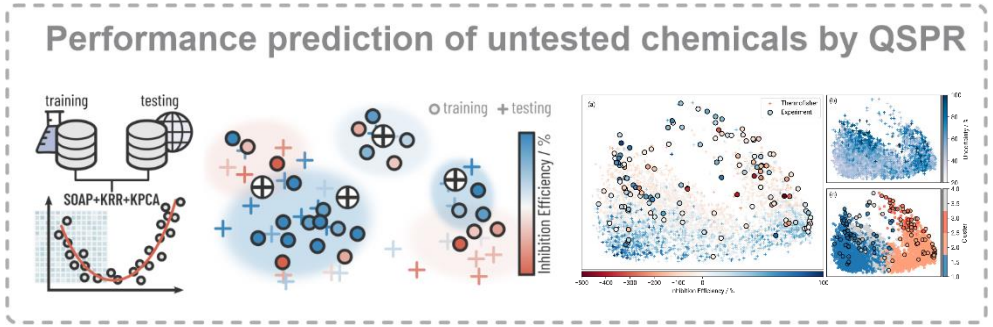
Mirror mirror on the wall who is the best inhibitor of them all?



- Data generation using a multi-well setup
- Generation of heatmaps and determination of volume loss in the well by profilometer
- 200 distinct compounds have been tested



- Determination of volume loss based on optical appearance
- Elimates need to perform manual (time-consuming) analysis of imprints
- Correlation between volume loss from experiment and virtual heatmap
- Highly sensitive to minimal offsets during annotation step

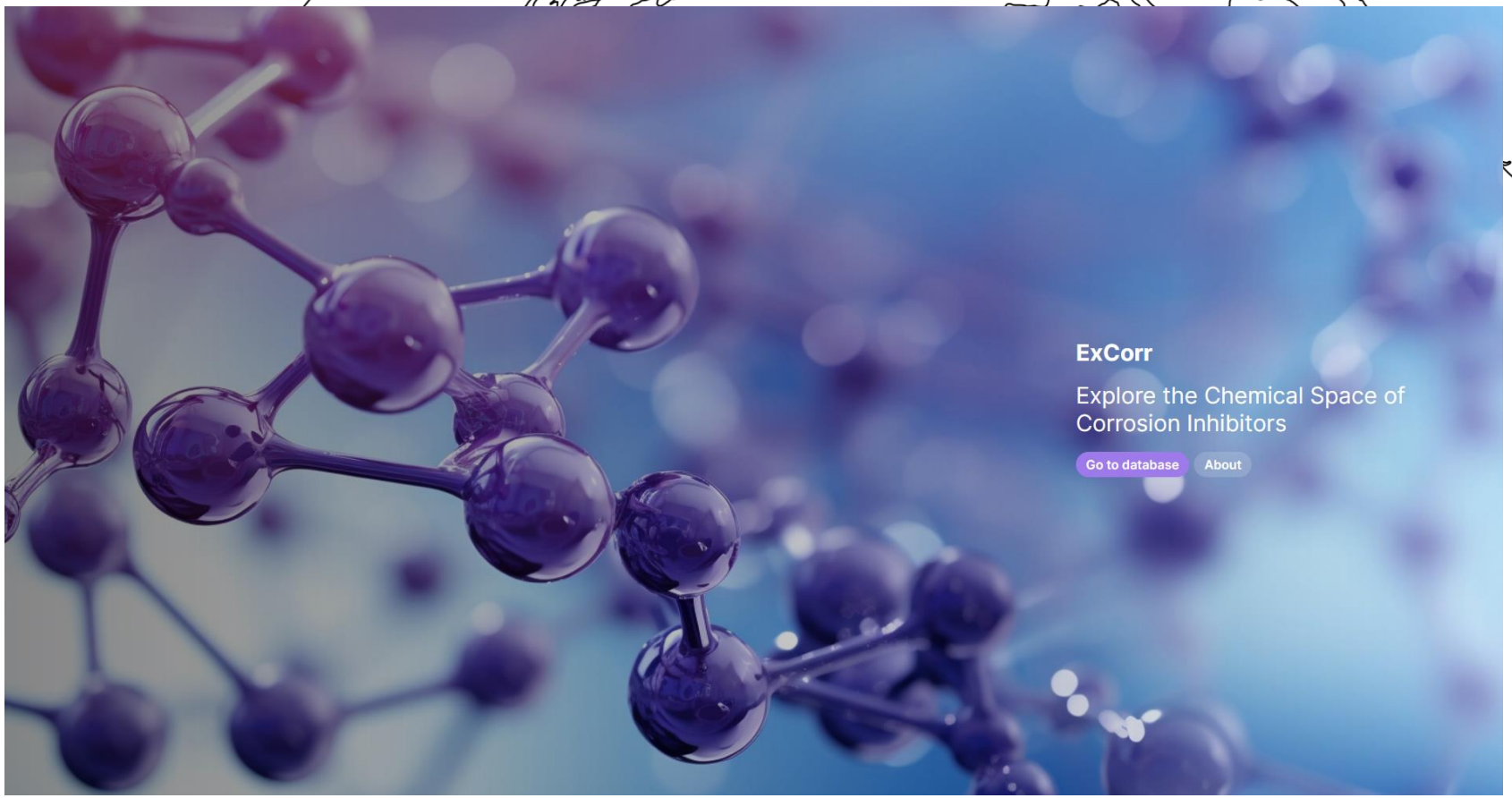


- Training of SOAP-based QSPR model based on ground truth
- Use of additional factors for compound selection (toxicity, solubility, price, uncertainty) rather than only based on molecular similarites
- First blind testing conducted, Next step: active learning and use of volume loss derived from virtual heatmaps as target for training of QSPR models

Helmholtz-Zentrum hereon

A. Lisitsyna, C. Song, T. Würger, B. Vaghefinazari, M. L. Zheludkevich, S. Albarqouni, S. V. Lamaka, C. Feiler, manuscript in preparation.

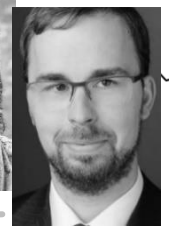
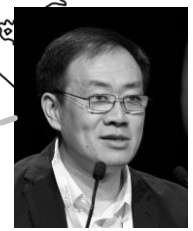
Inhibitor Database
ExCorr



D. A. Winkler, A. E. Hughes, C. Özkan, A. Mol, T. Würger, C. Feiler. S. V. Lamaka, in preparation.

Inhibitor Database
ExCorr



ExCorr

Database About

Search database

Filter database

IUPAC	CAS	SMILES	MW / Da
quinolin-8-ol (8-hydroxyquinc	148-24-3	OC1=C2N=CC=CC2=CC=C1	145.16
quinolin-8-ol(8-hydroxyquir	148-24-3	OC1=C2N=CC=CC2=CC=C1	145.16
quinoline-2-carboxylic acid ((	93-10-7	O=C(O)C1=NC2=CC=CC=C2	173.17

Base	Alloy	Method	IE / %	pH	Aggressive Component	Concentration / mM	References
Al	AA2024-T3		75		0.05M NaCl		(Lamaka et al.;2007)Balaskas e
Al	AA2024-T3		94		0.05M NaCl		(Lamaka et al.;2007)Balaskas e
Al	AA2024-T3	impedance	89.8		3.5% NaCl	5	(Lamaka et al.;2007)Balaskas e

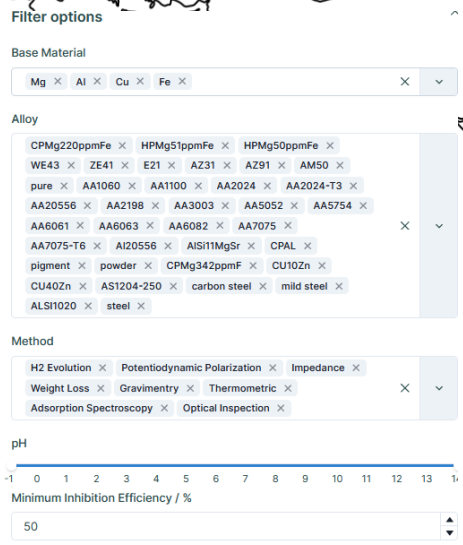
Show similar structures

IUPAC	CAS	SMILES	MW / Da
2-phenylchromen-4-one (2-p	525-82-6	O=C1C=C(C2=CC=CC=C2)O	222.24
2H-Chromen-2-one(1;2-Ben	91-64-5	C1=CC=C2C(=C1)C=CC(=O	146.14
3';6'-dihydroxyspiro[2-benzo	2321-07-5	O=C1OC2(C3=C(OC4=C2C=	332.31

Base	Alloy	Method	IE / %	pH	Aggressive Component	Concentration / mM	References
Al	powder	H2 evolution	45		1M HCl, 2.5% NaCl	10	(Horner and Meisel;1978)

D. A. Winkler, A. E. Hughes, C. Özkan, A. Mol, T. Würger, C. Feiler. S. V. Lamaka, in preparation.

Inhibitor Database
ExCorr

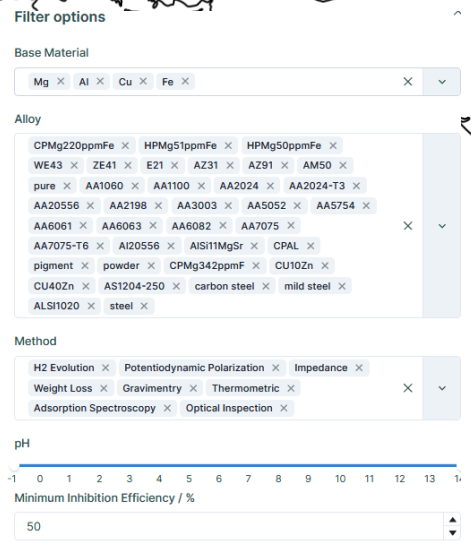


D. A. Winkler, A. E. Hughes, C. Özkan, A. Mol, T. Würger, C. Feiler. S. V. Lamaka, in preparation.

Inhibitor Database
ExCorr

ExCorr

- Review contains experimental information of 2400 compounds for Al-, Cu-, Fe- and Mg-based materials tested at various operating concentrations and with different techniques
- >1000 are already in a machine-readable format (roughly 3000 entries)
- Experimentalists can submit new data via a mask and ExCorr will be updated on a regular basis
- Will be released together with the review.
- https://excorr.web.app/



D. A. Winkler, A. E. Hughes, C. Özkan, A. Mol, T. Würger, C. Feiler. S. V. Lamaka, in preparation.

Thank you!

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Dr. Christian Feiler
Department of Interface Modelling
Institute of Surface Science

christian.feiler@hereon.de

HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

ThermoFisher SCIENTIFIC

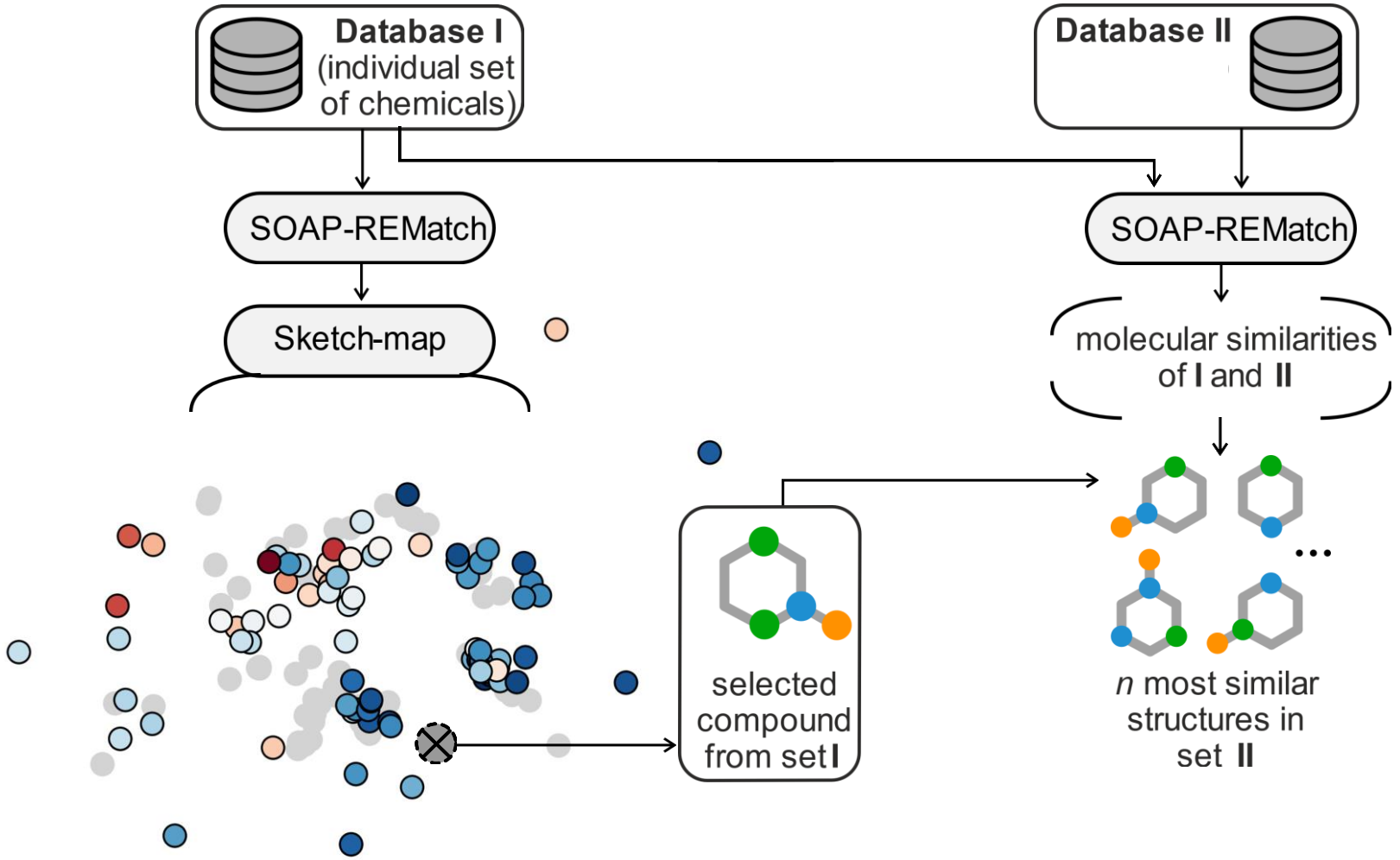
HELMHOLTZ MUNICH

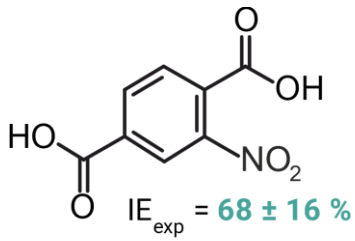
Helmholtz-Zentrum hereon

www.hereon.de

ExChem

Similarity-Based Selection of Compounds of Interest





Selected inhibitor from individual data set

DAAD
Deutscher Akademischer Austauschdienst
German Academic Exchange Service

ThermoFisher SCIENTIFIC

Helmholtz-Zentrum hereon

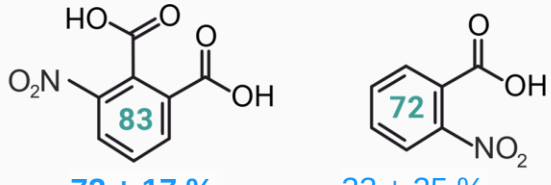
Exploring structure-property relationships in magnesium dissolution modulators. *npj Mater Degrad* **5,** 2 (2021). https://doi.org/10.1038/s41529-020-00148-z

ExChem

Similarity-Based Selection of Compounds of Interest

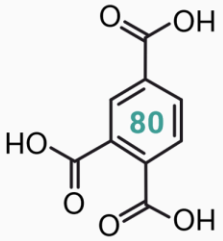
www.ExChem.de

High similarities out of 7094 molecules

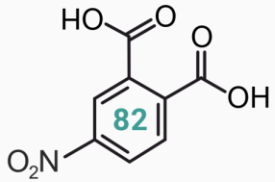


73 ± 17 %

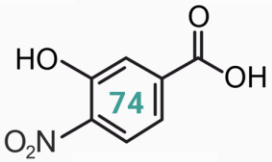
23 ± 25 %



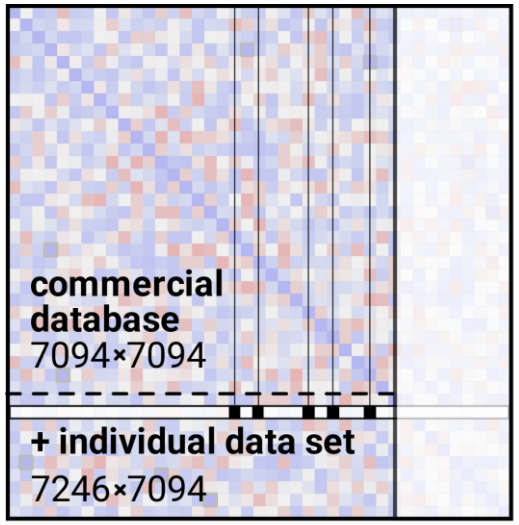
80 ± 16 %



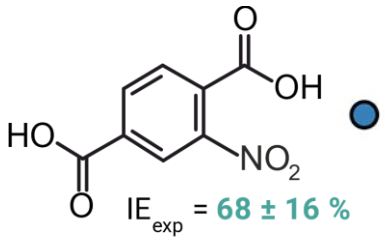
70 ± 20 %



13 ± 31 %

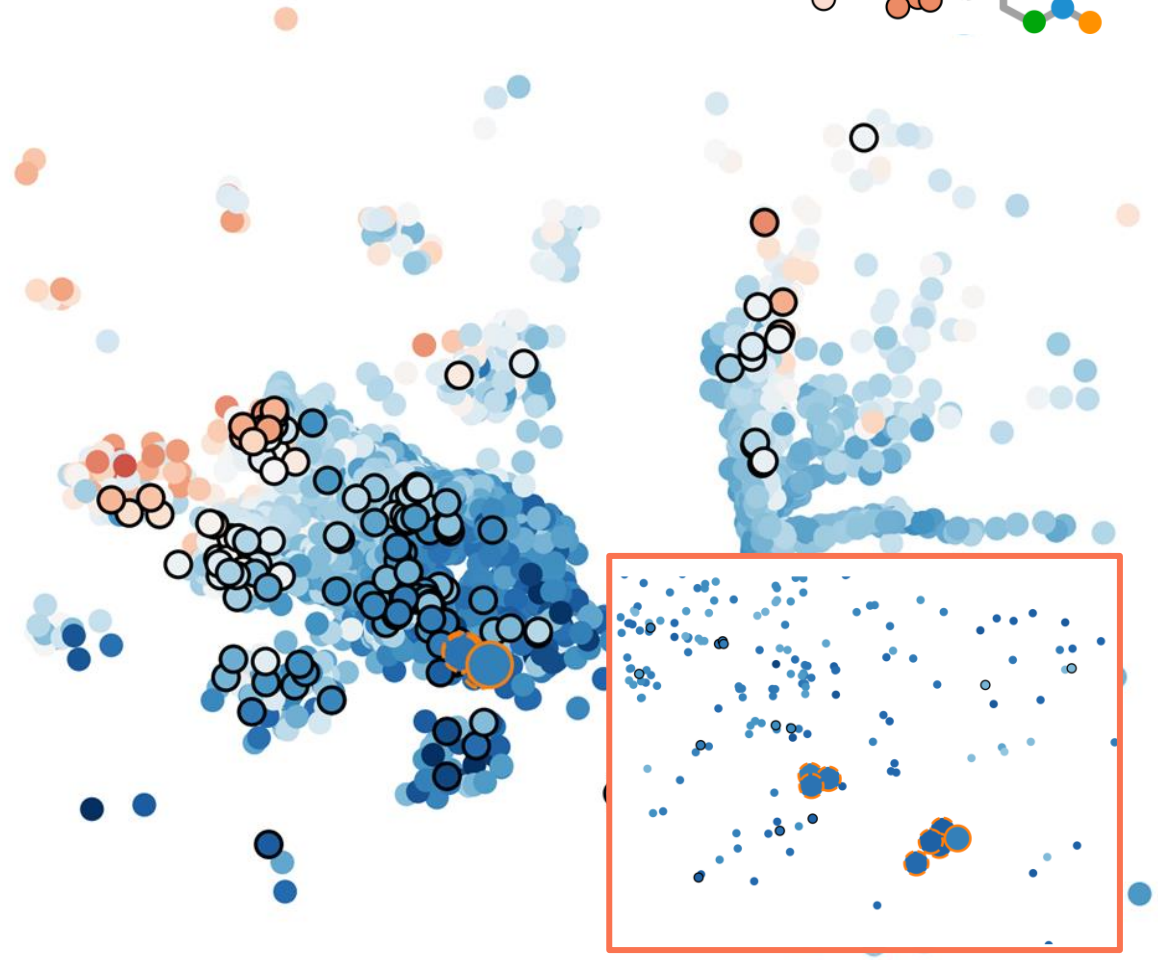


Select index in **global similarity matrix** (7246×7246)



IE_{exp} = 68 ± 16 %

Selected inhibitor from individual data set



DAAD
Deutscher Akademischer Austauschdienst
German Academic Exchange Service

ThermoFisher SCIENTIFIC

Helmholtz-Zentrum hereon

Exploring structure-property relationships in magnesium dissolution modulators.
npj Mater Degrad **5,** 2 (2021). https://doi.org/10.1038/s41529-020-00148-z