A Computational Approach to Magnesium Corrosion Engineering

Vom Promotionsausschuss der
Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation

von
Tim Würger

aus
Stade

2023

Vorsitzender des Prüfungsausschusses:	Prof. Dr.-Ing. Stefan Heinrich (Technische Universität Hamburg)
Gutachter:	Prof. Dr.-Ing. Robert H. Meißner (Technische Universität Hamburg)
	Prof. Dr. Mikhail L. Zheludkevich (Helmholtz-Zentrum Hereon)
Tag der mündlichen Prüfung:	12.09.2022

DOI: https://doi.org/10.15480/882.4920
ORCID: https://orcid.org/0000-0003-0346-0371

A Computational Approach to Magnesium Corrosion Engineering

Abstract

Magnesium is the lightest metallic construction material and bears high potential for automotive components, medical implants and energy storage systems. However, the practical use of untreated magnesium alloys is restricted as they are prone to corrosion. An essential prerequisite for the inhibition or control of the degradation process is a deeper understanding of the underlying corrosion mechanisms. For instance, identifying the reaction pathways most relevant for the release of gaseous hydrogen—one of the main products during magnesium corrosion—builds the foundation for adapting the hydrogen evolution rate in general. A possible approach to modulate the hydrogen evolution and magnesium dissolution rate is the introduction of small organic molecules that either stabilize the corroding surface or capture corrosive species. Such molecules constitute benign and useful materials to modify the service environment of magnesium for applications that require tailored degradation properties. This thesis illuminates different aspects of magnesium degradation, ranging from atomistic insights into the corrosion process itself to further developing strategies for effective corrosion control. By combining atomistic simulations and machine learning techniques, a holistic framework is developed to improve understanding and enable more efficient magnesium corrosion engineering in the future.

For Papa, Simon and Micha.

Acknowledgments

In the last four years, preparing and writing this thesis have been a huge part of my life. Naturally, it would have been impossible for me to stand where I stand now without the help, friendship and love of many wonderful people. For that I am truly grateful. Some of these people, however, deserve special thanks, which I would like to extend in the following.

Robert, I had the pleasure of being one of your first PhD students. Thank you for your continuous guidance, but also the freedom and trust you gave me for my research. I have always been fascinated and inspired by your enthusiasm, curiosity and the fact that you never rested until you fully understood something—basically the perfect traits I could imagine for any scientist.

Mikhail, when I think about magnesium corrosion, you are the first person that comes to my mind. I greatly enjoyed working with and learning from you. Your sense of the big picture and your ability to grasp the essence of a problem have always impressed me, and I hope that one day I will acquire a similarly clear view of things as you.

Christian, I admire your contagious love for science that also helped me to grow. Each project we worked on together became something I am proud of today and there are now countless words I will always link to you. Thanks for all the fruitful and fun talk, writing sessions, our research stay in Melbourne, and our internal visualization competitions. It has been a fun ride and I am really looking forward to the future.

Wolfgang and Gregor, you basically lay the foundation for my fascination of the atomistic world. Without you, I would still think of DFT as discrete Fourier transform and I would definitely not have written this thesis. Thank you for your patience and your never-ending support.

My friends, without whom my studies would have been a boring, stale and way more difficult experience. There are few people I spent so much time with and I am looking forward to all the fun times in the future. You are the salt to my soup and I am glad that you are always by my side.

My mother, for helping me to get where I am now. Every achievement of mine is also partially yours. You support me no matter what I do and I feel truly privileged for that. You are an amazing mother and I hope I will be one day a parent as loving and caring as you.

Last but not least, I want to thank my wife and best friend Wiebi. I truly believe the world would be a better place if everybody—and I include myself— was a little more like you. Thanks for always being there for me, giving me your unconditional, unwavering support, and showing me what is important when I need to see it the most. You are constantly helping me to push my own boundaries to become my better self. I love you.

PREFACE

Throughout my studies in mechanical engineering, I developed a passion for medical engineering, which finally lead to a respective Master's degree in 2018. However, this thesis—conducted under the supervision of Prof. Dr.-Ing. Robert Meißner at Hamburg University of Technology between February 2018 and January 2022—is embedded more in the context of computational materials science. To better understand this not directly apparent transition, it might be worthy to know that I worked part-time as a student research assistant in the latter field, where I discovered my fascination for the atomistic world under supervision of Dr. Gregor Vonbun-Feldbauer and Dr. Wolfgang Heckel. After my studies, Dr. Gregor Vonbun-Feldbauer introduced me to Prof. Dr.-Ing. Robert Meißner, who then offered me a PhD position conflating my interests.

Anybody I talked to about my PhD topic probably assumes I am deeply involved in the development of biodegradable magnesium implants or lightweight structures to fight climate change. Although this is not entirely untrue, these assumptions rather reflect my huge fascination with these particular subjects and the amazing versatile material magnesium actually is, and not my everyday work life. A little closer to the truth is that I worked for four years on a topic which could put my affection for atomistic simulations to use in fundamental research; to pave the way for the implementation of a material, which has the potential to revolutionize light metal engineering. Density functional theory and molecular dynamics became indispensable tools in the course of my doctorate. A major part and a personal pinnacle of this work, however, originates from my encounter with machine learning techniques.

Naturally, the studies present in this dissertation are not the product of me alone, but derive from collaborations with various amazing scientists from Helmholtz-Zentrum Hereon and Hamburg University of Technology. The result can be seen here, on the following pages, which hopefully convey some of my enthusiasm for magnesium. I, for my part, will never look at this material the same way again; and I will probably never stop telling people about the concept of self-dissolving magnesium implants—simply because I think they are brilliant.

Contents

Listing of Figures

LISTING OF TABLES

Chapter 1

Advanced Materials for a Brighter Future

> The beginning of knowledge is the discovery of something we do not understand.
>
> — Frank Herbert

Humanity today faces unprecedented challenges that threaten our planet and thus our species itself. Increasing temperatures and changing climate due to greenhouse gas emissions (such as CO_2) implicate fatal changes to our environment, leading to drastically shrinking habitats, limited food sources and water shortage. As the polar ice caps are melting, rising sea levels result in flooding and erosion of coastal areas. Heavy storms, shifting rainfall and other extreme weather phenomena will keep on occurring more often. The manifold of consequences include impaired water quality, but also decreasing availability of water resources in many areas on the planet. Additionally, more frequent heat waves propagate droughts and wildfires and many plant and animal species will be brought to the brink of extinction, if global average temperatures continue to increase uncontrolled. In particular developing countries are affected as people living there often heavily depend on their natural environment, however, also for countries of the global north the effects of climate change are severe. Impacts on human health as well as damage to infrastructure and property impose heavy costs on society and the economy.[1]

The Paris Agreement from 2015 defines clear targets to combat the negative effects of man-made climate change: limiting the global temperature rise to below 2°C and pursuing efforts to minimize warming to no more than 1.5°C above pre-industrial levels. Without rigorous measures to reduce CO_2 emissions, humanity approaches a tipping point, beyond which irreparable damage is inflicted to the natural environment. Close to 50% of all greenhouse gases originate from the transport and energy sector, rendering sustainable solutions in these fields indispensable.[1] By continuing to increase efficiency and shifting from fossil fuels to electric-drive vehicles, road transport could contribute a major mitigation share.[2] Promising

approaches to reduce emissions can be found in the interdisciplinary field of materials science, which involves research in light-weight materials and novel anode materials for energy storage systems.[3,4] Modern engineering materials, however, are required to not only excel in a single discipline but are rather expected to stand out in a multitude of properties in order to meet the high standards of pioneering industrial applications. Possible examples are light-weight materials that also show high mechanical performance and corrosion resistance or materials with multiple hierarchical levels that combine both hardness and ductility.[5] Of course, these materials should also meet the increasing requirements of a sustainable industry.

1.1 Materials Solutions: The Case of Magnesium

Magnesium (Mg) is among the most abundant elements on our planet and exhibits high potential to revolutionize light metal engineering in a large number of application fields. Treated magnesium-based alloys, for example, are promising candidates to satisfy a variety of current demands in the aeronautic and automotive industry[6], either as structural elements or in battery applications.[7–9] Furthermore, its biocompatibility enables Mg as dissolvable biomedical implant material. As corrosion plays a key role in these areas, the comparably high electrochemical activity of magnesium renders measures to control degradation processes in the employed material vital for its implementation. Each application field imposes unique challenges that need to be addressed to unravel the full potential of Mg-based materials. In transport applications, such as aeronautics and automotive, corrosion needs to be completely prevented to ensure the integrity of the material and avoid material failure. For energy applications, batteries with a Mg anode need a steady dissolution rate to keep the output voltage constant. Medical applications, where Mg is used in stents or temporary biodegradable bone implants, require a degradation rate tailored to a patient-specific injury to support recovery.

There are several approaches to protect magnesium from corrosion, including alloying and surface coatings.[10–12] Furthermore, the introduction of magnesium dissolution modulators[13–17], small organic molecules that inhibit or promote magnesium corrosion, facilitates tuning of the corrosion rate to meet application-specific demands (Figure 1.1). The latter approach particularly allows for tailored degradation rates of resolvable medical implants (e.g., stents) or adjustment of the anode material properties in magnesium-air batteries to boost the battery performance. Implementation of these schemes in magnesium engineering could be greatly enhanced by profound knowledge of the underlying corrosion mechanisms, which are still not fully understood to date. This includes the evolution of hydrogen but also

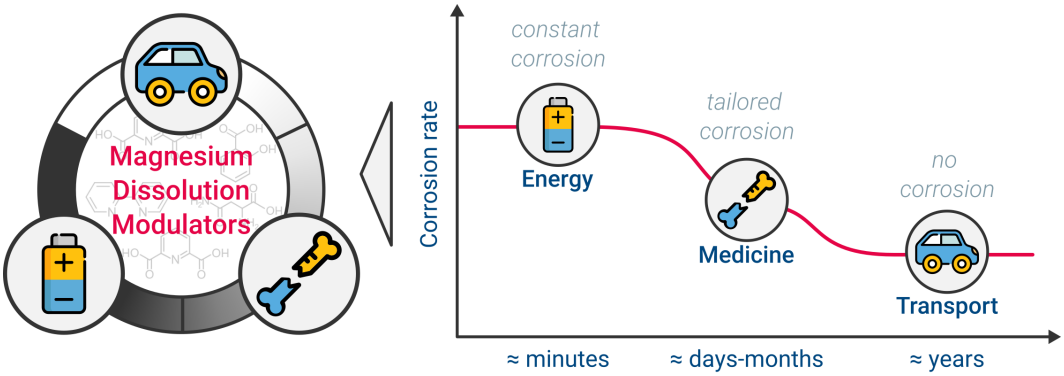


Figure 1.1: Domain-specific magnesium corrosion rates and their timescales. In magnesium corrosion engineering, degradation rates should be constant and, ideally, adaptable to the target application. The corrosion rate can be influenced following various strategies, such as employing small organic molecules as magnesium dissolution modulators.

the quantification of potential inhibition mechanisms by magnesium dissolution modulators. Lastly, efficient high-throughput and screening strategies are crucial to search for potential magnesium dissolution modulators in the vast chemical space.

1.2 Open Questions and Loose Ends in Magnesium Corrosion Research

This thesis sets out to address a number of open questions and loose ends in the realm of magnesium corrosion research. Employing several computational techniques, three regions of interest are deeply focused (see Figure 1.2). At first, potential reaction pathways for hydrogen evolution and magnesium dissolution at the magnesium-water interface are further investigated. In aqueous environments, the overall corrosion reaction is given as

$$\mathrm{Mg} + 2\,\mathrm{H_2O} \rightleftharpoons \mathrm{Mg(OH)_2} + \mathrm{H_2} \tag{1.1}$$

A peculiarity of magnesium is that for anodic overpotentials, both the anodic and the cathodic reaction rate are increasing, against the common conception of electrochemistry. Several theories exist that try to explain this phenomenon also called anomalous hydrogen evolution. In Chapter 3, the major reaction pathways during Mg corrosion are further investigated and discussed with respect to some of these theories. By stabilizing the corroding surface, or trapping corrosive species in complexes, magnesium dissolution modulators can strongly influence the corrosion rate. In Chapter 4, various dissolution modulating mechanisms are highlighted and discussed. Using atomistic simulations, the interaction between these small

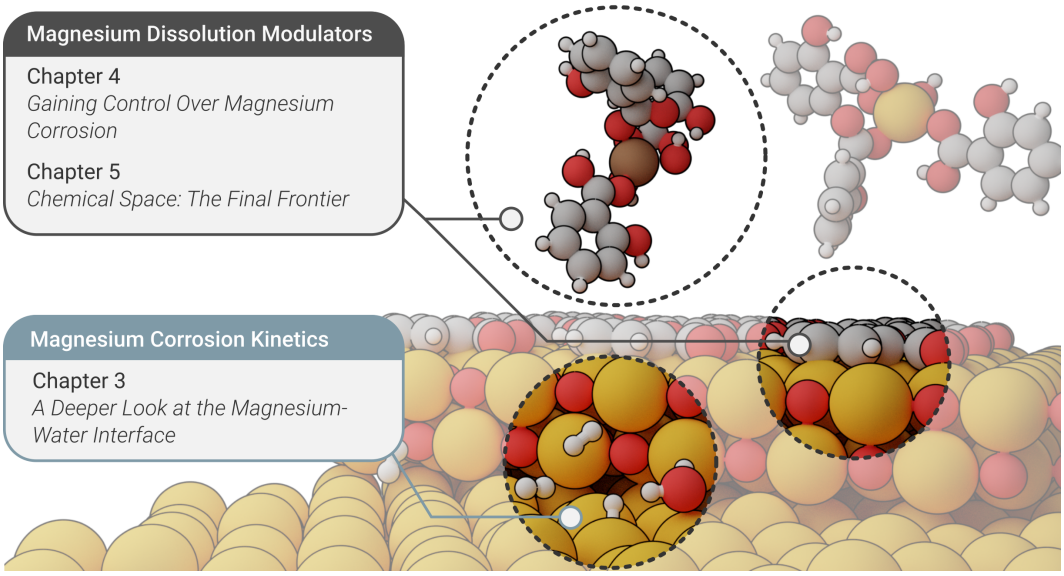


Figure 1.2: Overview of chapters. Different sites of the corroding magnesium surface (schematic representation) are highlighted and addressed in the according chapters.

organic molecules and the magnesium surface is quantified to learn more about what defines a good modulator. Finally, suitable compounds for corrosion control are a key factor to fully unravel the potential of Mg with respect to degradation-dependent applications. Chapter 5 deals with the question how potential magnesium dissolution modulators can be identified in the vast chemical space.

To get a feeling for the occurring errors and assumptions, a deep understanding of the underlying methods is vital to interpret, understand and transfer the obtained results. Hence, before diving into the results, the following Chapter 2 deals with the theory behind the employed methods.

Chapter 2

Modeling the Real World: Fundamentals

> Those who are not shocked when they first come across quantum theory cannot possibly have understood it.
>
> — Niels Bohr

Various computational methods have been applied within the scope of this thesis to approach current challenges in magnesium corrosion engineering, giving rise to a number of open questions: How can computer simulations derive materials properties from first principles? How can corrosion events be modeled and characterized? How can an additive's effect on the corrosion rate be predicted? To be able to answer these questions, the theoretical background behind these methods is briefly explained.

Throughout this thesis, the following notations apply when describing governing equations, unless otherwise indicated. Within quantum theory, a hat above a symbol ($\hat{\square}$) denotes a quantum mechanical operator and quantum states are described within the bra-ket notation ($\langle\square|\square\rangle$). Capital symbols in bold font ($\mathbf{\square}$) denote matrices and vectors, while a bar above a symbol ($\bar{\square}$) generally denotes an averaged quantity.

This chapter begins with an introduction to electronic structure calculations and particularly density functional theory (DFT), including its origin, approximations, and implementation. A basic understanding of the underlying theoretical framework is crucial to select a suitable ab initio method, correctly apply it and finally estimate the simulation accuracy. Within the DFT framework, methods to simulate surface interactions and reactions are briefly described. However, surface events (e.g., dissolution or adsorption) are often affected by thermal motion. By incorporating factors such as temperature and time, molecular dynamics (MD) simulations allow a more accurate description of such dynamic events. Hence, the chapter proceeds with the fundamentals of classical and ab initio MD. Finally, the chapter concludes with how atomic structures derived from ab initio calculations can be encoded to

identify quantitative structure-activity relationships (QSAR).

Generally, fundamental knowledge of chemistry, physics, and materials science is assumed, since a comprehensive guide to these fields would be beyond the scope of this work. The interested reader may find great joy in the works of Kittel and Parr and Yang for deeper insights into solid state physics and quantum chemical calculations, respectively.[18,19] To fully grasp the potential of atomic structure encoding and molecular machine learning techniques, respective works by Ceriotti et al. are highly recommended.[20–22]

2.1 Materials Properties from First Principles

Any material on the planet, whether it is in the shape of crystals, amorphous solids, molecules, or the person currently reading this thesis, is nothing other than a collection of atoms bound together by electromagnetic forces. Information about the electronic structure of an atom thus provides a direct link to many associated materials properties. To predict the electronic structure and find the ground state energy E_0 of a system, an approximate solution for the time-independent, non-relativistic Schrödinger equation has to be found:

$$\hat{H}|\psi\rangle = E|\psi\rangle, \tag{2.1}$$

where $\hat{H}$ is the Hamilton operator or *Hamiltonian* for a molecular system. Note that the time-independent Schrödinger equation only applies when the Hamiltonian does not depend on time explicitly. The operator $\hat{H}$ corresponds to the total energy of a quantum mechanical system:

$$\hat{H} = \hat{T}_\mathrm{e} + \hat{T}_\mathrm{N} + \hat{V}_\mathrm{ee} + \hat{V}_\mathrm{NN} + \hat{V}_\mathrm{Ne}, \tag{2.2}$$

where $\hat{T}_\mathrm{e}$ and $\hat{T}_\mathrm{N}$ are the operators corresponding to the kinetic energy of the electrons and nuclei, respectively. $\hat{V}_\mathrm{ee}$,$\hat{V}_\mathrm{NN}$ and $\hat{V}_\mathrm{Ne}$ describe the potential part of the Hamiltonian, representing the potential due to the electron-electron and nucleus-nucleus interactions and the attractive electrostatic interaction between the electrons and nuclei, respectively. Despite its simple form, Equation (2.1) is notoriously challenging to solve. Fortunately, using a set of reasonable physical approximations, the many-particle Schrödinger equation can be simplified so it can be solved numerically.

The first approximation arises from the immense weight difference between electrons and nuclei. For example, the nucleus of a Mg atom weighs roughly 40,000 times more than an electron and therefore moves much slower. Hence, to a good approximation, the electrons can be considered to move in the field of fixed nuclei, also known as the *Born-Oppenheimer approximation*. As a result, the Hamiltonian $\hat{H}$ can be separated into a nuclear, $\hat{H}_\mathrm{N}$, and

electronic, $\hat{H}_e$, part. After solving the electronic part of a problem for a fixed configuration of atoms, the atomic positions can be adjusted according to the calculated forces acting on them. The electronic Schrödinger equation is given by

$$\hat{H}_e|\psi\rangle = (\hat{T}_e + \hat{V}_{ee} + \hat{V}_{Ne})|\psi\rangle = E_e|\psi\rangle. \tag{2.3}$$

However, the electronic wave function becomes an informational challenge. Each electron adds three degrees of freedom (x, y and z) to the wave function. Even for a rather crude sampling of the wave function at 10 points in each dimension, a total of 10^{36} points would have to be sampled for a Mg atom with 12 electrons, well exceeding the number of stars in the observable universe. Fortunately—at least from a materials scientist's point of view—a popular solution to this problem called density functional theory (DFT) saw the light of day in 1964.[i]

2.1.1 Density Functional Theory

A landmark paper by Hohenberg and Kohn stated two theorems representing the major theoretical cornerstones of modern day DFT.[23]

Hohenberg-Kohn Theorem 1: The potential, and hence also the total energy, is a unique functional of the electron density ρ.

In other words, in a system of interacting electrons in the presence of an external potential V_{ext}, the ground state electron density ρ_0 uniquely specifies V_{ext} and thus the Hamiltonian and can be used to determine all properties of the system:

$$\rho_0 \rightarrow \hat{H} \rightarrow \psi_0 \rightarrow E_0. \tag{2.4}$$

As part of the Born-Oppenheimer approximation, the attractive electrostatic interaction between the electrons and nuclei, V_{Ne}, defines the external potential V_{ext}. As the total ground state energy E_0 is a functional of the ground state density ρ_0, so are its individual components:

$$E_0[\rho_0] = T_e[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0]. \tag{2.5}$$

From the variational principle it follows that a universal functional $F_{HK}[\rho]$ exists for any given external potential V_{ext}, leading to the ground state energy E_0.

[i] For his contributions to density functional theory, Walter Kohn won the Nobel Prize in Chemistry in 1998.

Hohenberg-Kohn Theorem 2: The functional that delivers the ground state energy of the system, delivers the lowest energy if and only if the input density is the true ground state density ρ_0.

Consequently, any electron density ρ associated with some external potential V_{ext} implies an energy $E[\rho]$, working as an upper bond to the true ground state energy $E_0[\rho_0]$:

$$E[\rho] \geq E_0[\rho_0]. \tag{2.6}$$

While $E_{\text{Ne}} = \int \rho_0(\mathbf{r}) V_{\text{Ne}} \, \mathrm{d}^3\mathbf{r}$ depends on the system, the form of the remaining terms in Equation (2.5) is universal and independent of the number of particles, although unknown. The system-independent parts are collected into a new quantity, the Hohenberg-Kohn functional F_{HK}:

$$\begin{aligned} E_0[\rho_0] &= \underbrace{T_{\text{e}}[\rho_0] + E_{\text{ee}}[\rho_0]}_{\text{universally valid}} + \underbrace{\int \rho_0(\mathbf{r}) V_{\text{Ne}} \, \mathrm{d}^3\mathbf{r}}_{\text{system dependent}} \\ &= F_{\text{HK}}[\rho_0] + \int \rho_0(\mathbf{r}) V_{\text{Ne}} \, \mathrm{d}^3\vec{\mathrm{r}}. \end{aligned} \tag{2.7}$$

If $F_{\text{HK}}[\rho] = T_{\text{e}}[\rho] + E_{\text{ee}}[\rho]$ was known exactly, the Schrödinger equation could also be solved—not merely approximately—but exactly. Unfortunately, the explicit form of the functionals for the kinetic energy $T_{\text{e}}[\rho]$ and the electron-electron interaction $E_{\text{ee}}[\rho]$ remain unknown. However, at least the known classical Hartree repulsion part E_{H} can be extracted from $E_{\text{ee}}[\rho]$

$$E_{\text{ee}}[\rho] = E_{\text{H}}[\rho] + E_{\text{ncl}}[\rho], \tag{2.8}$$

leaving E_{ncl} as the non-classical contribution to the electron-electron interaction, containing all effects of self-interaction correction, exchange and Coulomb correlation.

In 1965, Kohn and Sham proposed a strategy to approach the yet unknown functional F_{HK}.[24] Introducing the concept of a non-interacting reference system built from a set of one-electron wave functions φ_n (Kohn-Sham orbitals), the major part of the kinetic energy, T_{s}, can be computed to good accuracy:

$$T_{\text{s}} = -\frac{1}{2} \sum_{i}^{N} \langle \varphi_n | \nabla^2 | \varphi_n \rangle \, . \tag{2.9}$$

Hence, as much information as possible is calculated exactly and only a part of the total

energy is determined by an approximate functional E_{XC}:

$$F[\rho(\mathbf{r})] = T_{\mathrm{s}}[\rho(\mathbf{r})] + E_{\mathrm{H}}[\rho(\mathbf{r})] + E_{\mathrm{XC}}[\rho(\mathbf{r})]. \tag{2.10}$$

The exchange-correlation functional E_{XC} contains the residual part of the kinetic energy, T_{C}, and the non-classical contribution to the electron-electron interaction, E_{ncl}, making it effectively the functional of everything that is unknown:

$$E_{\mathrm{XC}}[\rho] = (T[\rho] - T_{\mathrm{s}}[\rho]) + (E_{\mathrm{ee}}[\rho] - E_{\mathrm{H}}[\rho]) = T_{\mathrm{C}}[\rho] + E_{\mathrm{ncl}}[\rho]. \tag{2.11}$$

Altogether, the total Kohn-Sham energy E_{KS} reads as

$$E_{\mathrm{KS}}[\rho] = T_{\mathrm{s}}[\rho] + E_{\mathrm{H}}[\rho] + E_{\mathrm{XC}}[\rho] + E_{\mathrm{Ne}}[\rho]. \tag{2.12}$$

Applying the variational principle, minimization of this expression results in the Kohn-Sham equations

$$\begin{aligned} \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{Ne}}(\mathbf{r}) + V_{\mathrm{XC}}(\mathbf{r})\right)\varphi_n = \varepsilon_n\varphi_n \qquad \text{with } V_{\mathrm{XC}}(\mathbf{r}) = \frac{\partial E_{\mathrm{XC}}}{\partial \rho(\mathbf{r})} \\ \Rightarrow \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{s}}(\mathbf{r})\right)\varphi_n = \varepsilon_n\varphi_n, \end{aligned} \tag{2.13}$$

which yield the Kohn-Sham orbitals φ_n that reproduce the density $\rho(\mathbf{r})$ of the original many-body system

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} |\varphi_n(\mathbf{r})|^2. \tag{2.14}$$

Once the various contributions of the potential V_{s} are known, V_{s} can be inserted into the Kohn-Sham equations to obtain the orbitals, which determine the ground state density and thus the ground state energy. As V_{s} already depends on the electron density, the Kohn-Sham strategy requires a self-consistent approach where the starting electron density ρ_{init} is guessed. It is noteworthy that if E_{XC} (and V_{XC}) were exactly known, this procedure would lead to the exact energy. Hence, a central goal of DFT is to find better approximations of these two quantities.

2.1.2 Approximations of the Density Functional

The simplest approach to approximate E_{XC} is the local density approximation (LDA). For a given electron density, the exchange-correlation energy density $\varepsilon_{\mathrm{XC}}$ is approximated by that

of the well-studied homogeneous electron gas evaluated locally:

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[\rho] = \int \rho(\mathbf{r})\varepsilon_{\mathrm{XC}}(\rho)\,\mathrm{d}^3\mathbf{r}. \tag{2.15}$$

Since the LDA is based on the assumption that the density $\rho(\mathbf{r})$ is constant throughout space, it works well in the case of solids with a rather steady density distribution. However, a more sophisticated approach is required to accurately describe atoms and molecules, usually yielding high density gradients.

The generalized gradient approximation (GGA) accounts for the non-homogeneity of the true electron density by including not only the local density $\rho(\mathbf{r})$ but also information about the gradient of the charge density $\nabla\rho$ at this position:

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[\rho] = \int \rho(\mathbf{r})\varepsilon_{\mathrm{XC}}(\rho, \nabla\rho)\,\mathrm{d}^3\mathbf{r}. \tag{2.16}$$

Within the GGA, various functionals were developed that brought a general improvement to describing ground state properties of solids and molecular systems. A widely adopted example is the PBE functional[25] but also other non-local correlation functionals exist that approximately account for dispersion interactions.[26,27] The Jacob's ladder by Perdew shows more advanced formulations of the exchange-correlation functional ranked by their chemical accuracy, however, more accurate functionals often come with a significantly higher computational effort.[28] An example is the recently proposed Strongly Constrained and Appropriately Normed (SCAN) functional, which is the first meta-GGA functional constructed such that it satisfies all known constraints a semi-local functional can satisfy, while the remaining free parameters are fitted to reproduce exact or accurate reference values, or norms, of exchange and correlation energies.[29]

2.1.3 Plane-Wave Implementation

In realistic systems, a cubic millimeter contains around 10^{20} atoms. Clearly, solving the Kohn-Sham equations numerically for all of these atoms is not feasible. Moreover, the one-electron functions only play a secondary role within the Kohn-Sham scheme and are primarily used to construct the electron density $\rho = \sum_i^N |\varphi_i(\mathbf{r})|^2$. Therefore, a popular approach is to linearly expand the Kohn-Sham orbitals in a set of L predefined basis functions η_μ multiplied with the expansion coefficients $c_{\mu\mathrm{i}}$:

$$\varphi_n(\mathbf{r}) = \sum_{\mu=1}^{L} c_{\mu\mathrm{i}}\eta_\mu. \tag{2.17}$$

For $L \to \infty$, the basis set would be complete and $\varphi_n(\mathbf{r})$ could be expressed exactly. In real applications, however, L is finite and has to be chosen such that the Kohn-Sham orbitals are approximated as accurately as possible.

Suitable basis sets can be generally divided into two types: atom-centered and non-atom-centered. The periodicity in large-scale systems (crystals) can be naturally addressed exploiting Bloch's theorem.

Bloch's Theorem: For a periodic potential, solutions to the Schrödinger equation can be expressed by a plane wave modulated by a periodic function.

In contrast to atom-centered basis functions, plane waves are not centered at the nucleus but extend through the entire space and implicitly involve the concept of periodic boundary conditions. Multiplied with a periodic function u, they approximate the Kohn-Sham orbitals φ_n to

$$\varphi_{n,\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r})\, \mathrm{e}^{i\mathbf{k}\mathbf{r}}, \tag{2.18}$$

where $u_{n,\mathbf{k}}(\mathbf{r})$ is of the same periodicity as the lattice and $\mathbf{k}$ is a reciprocal vector of real numbers within the first Brillouin zone, the primitive unit cell in reciprocal space. Here, n corresponds to the band index defining the number of independent eigenvalues for each value of $\mathbf{k}$.

According to Bloch's theorem, the wave vector $\mathbf{k}$ and its associated eigenvalues ε_n are periodic functions in the reciprocal space, implying $\varepsilon_{n,\mathbf{k}+\mathbf{G}} = \varepsilon_{n,\mathbf{k}}$ for a translation by the reciprocal lattice vector $\mathbf{G}$. Thus, the summation over an infinite number of translations $\mathbf{R}$ in real space can be replaced by an integral over the (finite) first Brillouin zone (BZ). As the one-electron wave functions vary slowly with respect to the wave vector $\mathbf{k}$, the integral is replaced by a weighted sum over discrete points, called k-points. Choosing the number of k-points usually implies a trade-off between accuracy and computational effort. Following this scheme, quantities such as the electron density ρ can be calculated:

$$\sum_{\mathbf{R}}^{\infty} \Rightarrow \int_{\mathbf{k}\in 1\mathrm{st\,BZ}} \mathrm{d}\mathbf{k} \approx \sum_{\mathbf{k}} w_{\mathbf{k}} \quad \Rightarrow \quad \rho(\mathbf{r}) \approx \sum_{\mathbf{k}} w_{\mathbf{k}} \sum_{n} |\varphi_{n,\mathbf{k}}(\mathbf{r})|^2, \tag{2.19}$$

where the plane-wave expansion of the Kohn-Sham orbitals in reciprocal space for the simulation cell volume Ω is given by

$$\varphi_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{n,\mathbf{k}}(\mathbf{G})\, \mathrm{e}^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}. \tag{2.20}$$

Including an infinite number of plane waves would exceed any practical limits. Hence, as

plane waves at the lower end of the kinetic energy range are most relevant, the basis set is truncated to a finite number of plane waves according to

$$\frac{\hbar^2}{2m_\mathrm{e}}|\mathbf{k}+\mathbf{G}|^2 < E_\mathrm{cut}, \tag{2.21}$$

where E_cut is the kinetic cutoff energy. As a result, increasing E_cut or the simulation cell volume Ω also increases the basis set size, thus systematically improving the accuracy.

2.1.4 Pseudopotentials

A substantial disadvantage of plane waves is their inefficiency in describing wave functions with large curvatures and high kinetic energies, as is the case in the core region of atoms. Properly treating core electrons may become increasingly important for calculating spectral properties of the core or GW calculations.[30,31] An accurate description of such regions with plane waves would require an unreasonably large basis set, thus dominating the convergence of E_cut and skyrocketing the computational effort. Fortunately, the plane-wave functions of the core electrons barely overlap, leading to similar behavior in different chemical environments. Furthermore, the core electrons take almost no part in chemical bonding, giving the foundation for the *frozen core approximation*. Consequently, the complex effects of the core can be substituted with a precalculated potential, the *pseudopotential*, and only the valence electrons are treated explicitly (Figure 2.1). The idea of the pseudopotential approximation is to eliminate the electronic degrees of freedom of the core electrons and to describe the valence electrons with plane nodeless wave functions without core oscillation using a non-singular pseudopotential. The corresponding set of pseudo wave functions become the all-electron wave functions outside a chosen cutoff radius r_cutoff. Due to the nodeless shape of the pseudo wave functions, the basis set to approximate the one-electron wave function can be reduced to a reasonable size, thus reducing the cost of the calculation.

One practical realization of the pseudopotential approximation is the projector augmented wave method. Here, the rapidly oscillating wave functions in the core region are first replaced by smooth pseudo wave functions. These smooth pseudo wave functions are then projected onto the real, precalculated all-electron wave functions, thus reproducing the correct eigenvalue spectrum.

2.1.5 Atomic Structure Optimization

When determining the electronic structure of a given atomic configuration with DFT, the considered configuration might be far from the optimum. As a result, the properties as

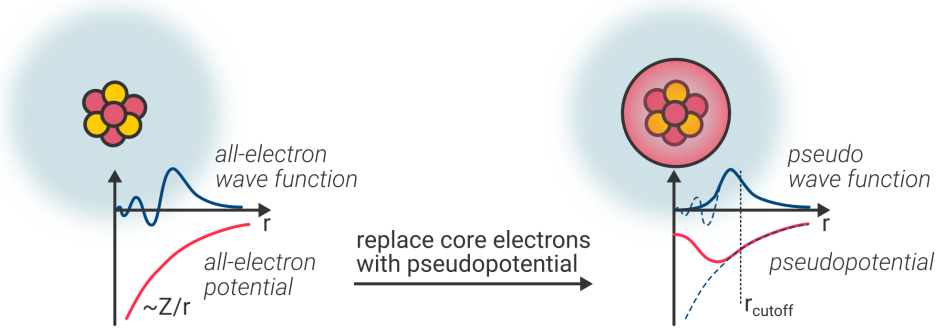


Figure 2.1: Pseudopotential. The pseudopotential represents the effective core charge. Outside the cutoff radius r_{cutoff}, the pseudo wave function and pseudopotential match the all-electron wave function and potential, respectively.

derived from the DFT calculations also would become less relevant. Hence, it is crucial to not only optimize the electronic structure but also the atomic positions (geometric relaxation) to obtain the true ground state of the considered system. Therefore, the atomic positions $\mathbf{R}_I$ are optimized until the lowest-energy equilibrium configuration is achieved and the forces $\mathbf{F}_I$ on the atoms vanish. Hence, the equilibrium state for M atoms can be expressed as

$$\mathbf{F}_I = -\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}_I} = 0 \qquad \text{for } I = 1, 2, ..., M, \tag{2.22}$$

where $\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}_I}$ is a gradient. To solve this minimization problem, the Hellmann-Feynman theorem is applied, which allows to derive the atomic forces from the full Hamiltonian $\hat{H}$ with respect to the atomic positions.

Hellmann-Feynman Theorem: The derivative of the total energy with respect to a parameter directly relates to the expectation value of the derivative of the Hamiltonian with respect to that same parameter.

As only the repulsive nucleus-nucleus and the attractive nucleus-electron interactions depend on the atomic positions, the force $\mathbf{F}_I$ can be calculated as

$$\mathbf{F}_I = -\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}_I} = -\langle\psi|\frac{\partial H}{\partial \mathbf{R}_I}|\psi\rangle = -\int \rho(\mathbf{r})\frac{\partial V_{\text{Ne}}}{\partial \mathbf{R}_I}\,\mathrm{d}^3\mathbf{r} - \frac{\partial E_{\text{NN}}(\mathbf{R})}{\partial \mathbf{R}_I}. \tag{2.23}$$

Aside from geometric relaxations of atomic configurations, the calculated forces are also essential for calculating vibrational frequencies and for molecular dynamics simulations.

Various optimization algorithms, such as Newton-Raphson or Steepest Descent, are suitable to implement a minimization scheme for Equation (2.22), consisting of an inner

self-consistent loop for electronic structure optimization and an outer loop for atomic structure optimization. The minimization scheme continues until the calculated forces fall below a predefined threshold. Finally, the found equilibrium geometry defines a stationary point on the potential energy surface (PES). Depending on the second-order derivative of the energy, this stationary point represents a local minimum for $\frac{\partial^2 E(\mathbf{R})}{\partial \mathbf{R}_I^2} > 0$ or a saddle point for $\frac{\partial^2 E(\mathbf{R})}{\partial \mathbf{R}_I^2} = 0$.

2.1.6 Molecular Dynamics

Extending the concept of atomic structure optimizations, molecular dynamics (MD) simulations are a powerful tool for studying the motion of atoms in molecules, liquids, and solids. In a typical sense, the term *molecular dynamics* refers to the propagation of point particles—atomic nuclei or effective particles, combining multiple nuclei—according to the laws of Newtonian mechanics.[32] The forces acting on the particles are calculated during the simulation at discrete points along the trajectory. A trajectory describes the time-dependent evolution of a system, given as an ensemble of frames that share the same macroscopic (thermodynamic) state but differ in the microscopic states. Such ensembles are called microcanonical (NVE), canonical (NVT) or isothermal-isobaric ensemble (NPT), depending on whether the number of atoms N, the volume V, the energy E, the pressure P or the temperature T are conserved throughout the MD simulation. In the NVT ensemble, for example, a system has the absolute temperature T but may exchange energy with an external heat bath.

Depending on how the acting forces are calculated—using electronic structure methods such as DFT or empirical interaction potentials—one can distinguish between ab initio and semi-empirical MD. Ab initio molecular dynamics (AIMD) simulations treat the electronic degrees of freedom explicitly in quantum mechanical calculations and thus usually involve a significantly higher computational effort, limiting respective system sizes to a few hundred atoms and time scales to the order of picoseconds. Particularly for amorphous structures such as liquids, the finite size effects can introduce anisotropic behavior and hamper the simulation through implied interactions between periodic images due to the periodic boundary conditions. Another challenge is imposed by the rather short time scales and one of the most fundamental axioms of statistical mechanics. While time averages can be calculated by MD simulations, experimental observables are assumed to be ensemble averages. The *ergodic hypothesis* resolves this dilemma by stating that the time average equals the ensemble average. Hence, AIMD simulations have to be designed in a way—or run long enough—to sample a sufficient amount of phase space to ensure this equality is satisfied. Aside from the

implied challenges using AIMD simulations, however, they generate the most versatile and accurate modeling of interatomic interactions as the electronic structure is explicitly taken into account. This is of particular importance for modeling chemical reactions.

Within the Born-Oppenheimer approximation, the atomic forces can be directly derived from the Hellmann-Feyman theorem according to Equation (2.23). The nuclei are then propagated by integrating Newton's equations of motion, given as

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}_I} = \mathbf{F}_I \,. \tag{2.24}$$

Most integrators, such as *Verlet*, *Leapfrog* and *Velocity Verlet*, assume that the acceleration $\ddot{\mathbf{R}}$ and velocity $\dot{\mathbf{R}}$ can be approximated using a Taylor series.[33]

In the past, AIMD has mainly relied on two different approaches.[34] In Born-Oppenheimer molecular dynamics (BOMD), the total energy of a system is fully minimized at every MD step using an electronic structure method, rendering this scheme computationally rather demanding. Instead of a relatively time-consuming iterative energy minimization, original Car-Parrinello molecular dynamics (CPMD) treats the electronic degrees of freedom as classical time-dependent fields with a fictitious mass μ that are propagated using Newtonian dynamics. To keep the electronic and nuclear subsystems adiabatically separated, resulting in electrons following the nuclei very close to their immediate electronic ground state, μ must be chosen sufficiently small. Since the acceptable integration time step scales with approximately $\sqrt{\mu}$, the step size becomes significantly smaller than for BOMD, thus limiting the achievable simulation timescales. A Car–Parrinello-like approach to BOMD, called second-generation CPMD, attempts to unify the best of both schemes, realizing the large integration time steps of BOMD while preserving the efficiency of CPMD.[35,36] For that purpose, the fictitious Newtonian dynamics for the electronic degrees of freedom is replaced with an enhanced electron-ion dynamics that allows to refrain from an additional fictitious mass parameter. The increased efficiency of this scheme stems from the fact that the instantaneous electronic ground state is very closely approached within just one electronic gradient calculation. For the short-term integration of the electronic degrees of freedom, a highly accurate, yet efficient, propagation scheme based on the predictor–corrector integrator of Kolafa[37] is employed. Since only one electronic force calculation is required, the self-consistent field (SCF) cycle and any expensive diagonalization steps can be completely avoided while keeping very close to the potential energy surface and, at the same time, choosing Δt to be as large as in BOMD. However, due to the nature of the predictor-corrector scheme, the introduced nuclear dynamics is slightly dissipative. In order to account for the dissipation, the acting forces in second-order CPMD are finally corrected using a friction term, described by the friction

coefficient γ_D. The canonical sampling of the Boltzmann distribution is then accomplished by means of a modified Langevin-type equation that includes an additive white noise term that obeys the fluctuation-dissipation theorem.[34] Hence, if γ_D was known, an exact canonical sampling of the Boltzmann distribution could be ensured despite the dissipation. Fortunately, the inherent value of γ_D can be bootstrapped so as to yield the correct average temperature as defined by the equipartition theorem.[34]

2.2 Computational Surface Science

Modern surface science addresses physical and chemical processes occurring at the interface of two phases (e.g., solid-solid, solid–liquid, solid–gas, solid–vacuum, and liquid–gas). Popular topics in this field are heterogeneous catalysis[38], semiconductor technology[39] and electrochemistry[40–43]. The study of surfaces can be approached either experimentally or computationally. While experimental surface science explicitly reflects reality, a deep understanding of the underlying processes on an atomistic level can often not be obtained. Hence, the ability to extract understanding purely from the basic rules of physics is an attractive and significant benefit of computational surface science. Moreover, the exponential expansion of accessible computational power and methodological innovations continuously increases the potential of simulations, as models can become more complex, thus more accurately resembling reality.

DFT is a valuable tool to better understand chemical occurrences at the surface. Particularly in corrosion engineering, a deep insight into occurring surface interactions is crucial, since the governing corrosion reactions happen at the solid-liquid or solid-(humid) air interface. Here, DFT calculations can help, for example, to predict the impact of small organic molecules on the surface stability or to determine the energy barrier of a surface reaction.

However, many recurring concerns in computational science with DFT are also relevant in computational surface science. How accurate is a selected exchange-correlation functional? How can the computational efficiency of the chosen method be increased? How can the obtained results be verified (is an experimental reference available)? How to find a trade-off between model complexity and computational effort? All of these concerns address the very fundamental question: how reliable are the conducted simulations? To fully address such questions, a deeper understanding of the model building process and the employed methods is essential. The following sections provide insights into the preparation of surface models for DFT calculations and relevant methods for the simulation of corrosion events.

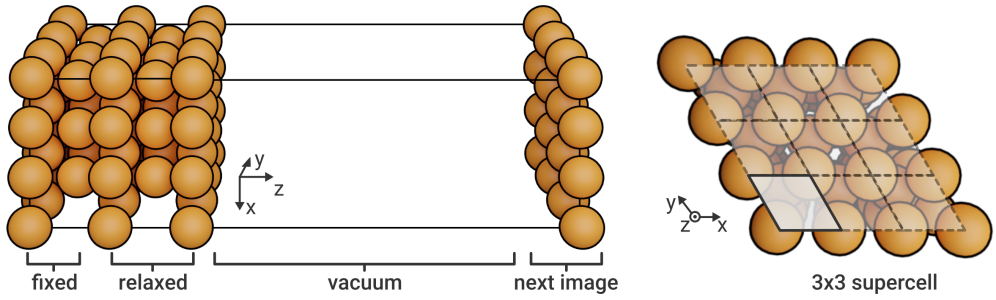


Figure 2.2: Periodic surface slab of Mg(0001). The Mg unit cell is repeated in all room directions according to the desired slab size, creating a 3×3 supercell with 5 layers. Adding a vacuum in z-direction results in the final surface model. The lower part of the surface slab is fixed to represent the bulk phase. (left) Side view, perpendicular to surface normal. (right) Top view, parallel to surface normal.

2.2.1 Periodic Slabs

In crystalline solids such as magnesium, the atoms are arranged in a periodic lattice. The smallest repeating unit of such a lattice is the unit cell, which completely represents the symmetry and structure of the entire crystal. Stringing an infinite number of unit cells together would fill up the space completely, with an invariance for translations by a multiple of the lattice parameters. Introducing periodic boundary conditions according to Bloch's theorem (Section 2.1.3), the infinite crystal can be modeled by a finite cell. Naturally, an infinite crystal is unrealistic due to surface or lattice defects, which interrupt the periodicity in all room directions—and plainly because every solid has boundaries. However, these defects can be simulated with supercells that include a sufficiently large volume of the lattice (Figure 2.2). A supercell consists of multiple primitive unit cells that are treated explicitly in the simulation. Moreover, the infinitely repeated atomistic system approximates quite well the real state of a bulk material.

Supercells become particularly important when dealing with intermetallic impurities, as they are quite common in manufactured magnesium. For example, commercial purity magnesium contains around 220 ppm iron (Fe) and 5 ppm copper (Cu). As the hexagonal primitive unit cell of magnesium contains only two atoms, considering more than two elemental species is already impossible. Even when taking into account only iron impurities, replacing one Mg atom with Fe would already lead to an impurity concentration of nearly 700,000 mass ppm, far too large for any realistic magnesium alloy. Assuming a bulk supercell built from $3 \times 3 \times 3$ unit cells with $2 \cdot 3^3 = 54$ atoms, substituting one Mg with an Fe atom would still lead to an impurity concentration of over 40,000 mass ppm. In fact, replicating an iron content of 220 ppm would demand explicit modeling of over 10,000 atoms, implying an immense computational effort. Therefore, building a (surface) model always implies a

trade-off between model accuracy and computational resources.

Cleaving the bulk supercell and adding vacuum space in the desired surface direction creates a surface slab, breaking the translational symmetry accordingly (Figure 2.2). As there is no periodicity perpendicular to a surface, only one k-point is required for sampling in this direction. Nevertheless, the plane-wave approach still implies periodic boundary conditions in all room directions, so the vacuum region has to be chosen sufficiently large to prevent surfaces of consecutive slabs from affecting each other. Furthermore, the surface slab has to be thick enough, meaning that enough surface layers are included in the slab model to rule out any significant impact on the energetics. Both parameters, the size of the vacuum region and the number of surface layers, should be carefully optimized in convergence tests. As a rule of thumb, 20 ± 5 Å is usually sufficiently large for the vacuum region, however, the number of surface layers strongly depends on the surface configuration. Typically, more layers are added until the center of the slab reflects bulk properties. Since an increasing number of surface layers also raises the computational effort, strategies to keep them to a minimum are preferable. An effective approach is to create an asymmetric slab by freezing the position of the bottom layers, thus imposing bulk boundary conditions on the adjacent layers free to relax.

2.2.2 Adsorption Energies

In many cases, the adsorption energy of a molecular compound upon a surface is an insightful property. In magnesium corrosion, for example, it can describe the impact on the surface stability, thus indicating if a specific compound inhibits or accelerates the corrosion rate. An adsorption reaction involves three components that all have to be treated independently in DFT calculations: the isolated molecule in vacuum, the clean substrate and the combined system, including both the adsorbed molecule and the substrate. Accordingly, the adsorption energy E_{ad} per adsorbed molecule released in or required for the adsorption reaction can be calculated using the system energies E_{mol}, E_{sub} and $E_{\mathrm{mol,sub}}$, respectively:

$$E_{\mathrm{ad}} = \frac{1}{N}\left(E_{\mathrm{mol,sub}} - E_{\mathrm{sub}} - N \cdot E_{\mathrm{mol}}\right), \tag{2.25}$$

where N denotes the number of adsorbed molecules. Schematic simulation cells for all three components are illustrated in Figure 2.3, using the example of a water molecule adsorbed upon a Mg surface.

In a DFT calculation, the energies for the combined system (a) and the clean magnesium surface (b) have to be determined using the same k-point mesh and cell size. For the isolated water molecule (c), the simulation cell is chosen large enough to eliminate any interaction

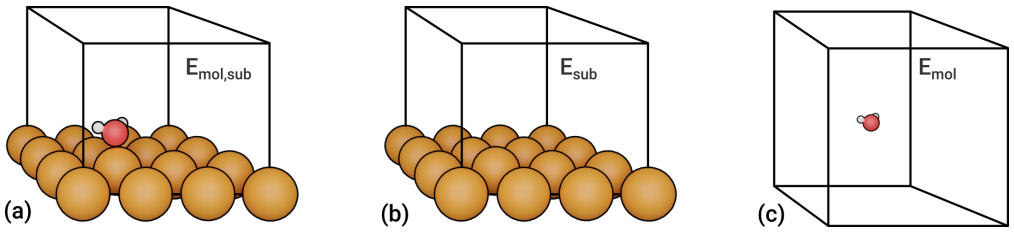


Figure 2.3: Components for calculating the adsorption energy. (a) The combined molecule-substrate system. (b) The clean surface. (c) The isolated molecule. For the sake of visualization, only one surface layer is shown in the schematics. However, the entire surface slab is included in the DFT calculation.

between periodic images. Since periodicity plays no active role in this case, typically one k-point is considered enough for sampling the Kohn-Sham equations. Finally, all individual energies can be inserted in Equation (2.25) to obtain the final adsorption energy E_{ad}.

According to the laws of thermodynamics, a system tends to strive for the state with the lowest energy. This has a direct influence on the sign of E_{ad} and thus its interpretation. Assuming the adsorption of a water molecule upon the Mg(0001) surface is energetically more favorable than for it to remain in the gas phase, then

$$E_{\mathrm{mol,sub}} < E_{\mathrm{sub}} + E_{\mathrm{mol}}. \tag{2.26}$$

As a result, the adsorption energy is negative ($E_{\mathrm{ad}} < 0$) for favorable adsorption and vice versa. It is noteworthy, however, that this condition of favorable adsorption first of all applies for 0 K and can be influenced by entropic as well as finite temperature effects.

2.2.3 Surface Reaction Pathways

In a general sense, corrosion is the electrochemical oxidation of a metal in reaction with an oxidant. Depending on the type of oxidant, the metal is converted into a chemically more stable form, such as an oxide or hydroxide. As corrosion is a diffusion-controlled process, it occurs on exposed surfaces and is thus defined by (electro)chemical surface reactions. Reactions that are diffusion-controlled have a reaction rate equal to the rate of transport of the reactants through the reaction medium, usually a solution.[44] In other words, the reactants diffuse through the reaction medium until they collide in the right stoichiometry and form an activated complex which can form the product species. In diffusion-controlled reactions, the formation of products from the activated complex is significantly faster than the diffusion of reactants. In case a reaction consists of multiple steps, the step with the lowest reaction rate is considered as rate-determining.

Whether a reaction takes place or not, strongly depends on the thermodynamic driving

force and the associated energy barrier. In general, a reaction will be energetically favorable, if the Gibbs free energy ΔG for the process, given as

$$\Delta G = \Delta H - T\Delta S \tag{2.27}$$

is negative, where H, T and S denote the enthalpy, temperature and entropy of the reaction, respectively. Often, the process is dominated by either the enthalpic or the entropic contribution. For example, the reaction is likely to occur when stronger bonds are formed than are broken, provided the energy barrier is not too high (primarily a ΔH effect). However, when large molecules with a higher energy (chemical potential) react to a greater number of smaller molecules, primarily the entropic change ΔS defines the thermodynamic driving force. According to whether the final energy H of the products is lower or higher than the initial energy, the chemical reaction is defined as exothermic or endothermic, respectively.

When transitioning from reactants to products along a certain reaction coordinate, the configuration corresponding to the highest potential energy is called transition state, which is a first-order saddle point on the PES. The transition state is essential to determine the energy barriers in a chemical reaction, and thus the activation energy and reaction kinetics. Both, transition state and reaction rate, can be derived from the minimum energy path (MEP). Common strategies to finding the MEP of a chemical reaction are chain-of-states methods.[45,46] A chain of images of the system is generated between the final product state and the initial reactants. Then, all of these images are optimized in a concerted manner, approaching the MEP.

A widely adopted chain-of-state method is the nudged elastic band (NEB) method (Figure 2.4).[47–49] First, a number i of intermediate images between experimentally known or realistic reactants and products are generated along the reaction path. It is crucial that the start and endpoint of this path are well relaxed with respect to the atomic forces. Typically, a linearly interpolated path is adequate as an initial pathway to start the NEB calculation. During relaxation, each image moves to the lowest energy state possible while keeping equal spacing between adjacent images. This constrained optimization is accomplished by a force projection scheme, where the total force $\mathbf{F}_i^{\mathrm{NEB}}$ acting on an image i is the sum of two independent components:

$$\mathbf{F}_i^{\mathrm{NEB}} = \mathbf{F}_i^{\perp} + \mathbf{F}_i^{\mathrm{S}\parallel}. \tag{2.28}$$

The spring force $\mathbf{F}_i^{\mathrm{S}\parallel}$ and the true force $\mathbf{F}_i^{\perp}$ act along and perpendicular to the band, respectively. To make the projections, the local reaction path tangent $\hat{\boldsymbol{\tau}}_i$ is determined as the unit vector to the adjacent, higher energy image.[49] With respect to $\hat{\boldsymbol{\tau}}_i$, the component of the force due to

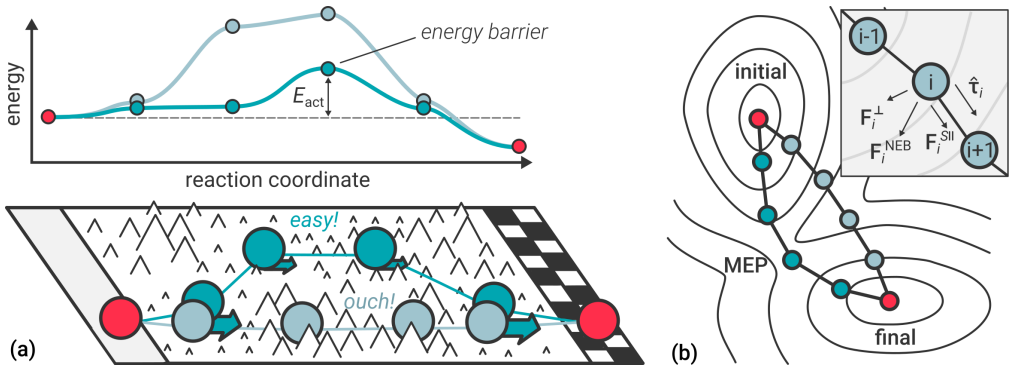


Figure 2.4: Nudged elastic band method. A number of images are generated by linear interpolation (gray) to sample the reaction path between known reactants and its products (red). During optimization to the minimum energy path (turquoise), the images are kept equidistant by spring forces along the reaction coordinate and move according to the force due to the potential perpendicular to the band. The energy difference between saddle point and starting point defines the activation energy E_{act}. (a) The optimization can be imagined as balls in mountain terrain. The direct is the shortest way to the destination but is paved with high mountains. A slightly longer way, however, poses a less exhausting alternative. (b) During the optimization, the initial band of images relaxes to the minimum energy path on the potential energy surface. Figure (b) was adapted from [47].

the potential perpendicular to the band, $\mathbf{F}_i^{\perp}$, can be calculated as

$$\mathbf{F}_i^{\perp} = -\nabla E(\mathbf{R}_i) + \nabla E(\mathbf{R}_i)\,\hat{\boldsymbol{\tau}}_i\,\hat{\boldsymbol{\tau}}_i, \tag{2.29}$$

where E is the energy of the system and $\mathbf{R}_i$ is the position of the ith image. The spring force $\mathbf{F}_i^{S\parallel}$ parallel to the band results from

$$\mathbf{F}_i^{S\parallel} = k\,(|\mathbf{R}_{i+1} - \mathbf{R}_i| - |\mathbf{R}_i - \mathbf{R}_{i-1}|)\,\hat{\boldsymbol{\tau}}_i \tag{2.30}$$

with k as the spring constant. Due to the equidistant spacing of the NEB images, the true transition state is most likely not defined by, but lies close to, the point with the highest energy. On the search for the transition state, the highest energy image can be used as basis for an ensuing refinement in a local search. One approach, which avoids having to run multiple optimizations, is the climbing-image NEB (CI-NEB)[50] method. It allows to simultaneously gain information about the MEP and achieve convergence to a saddle point (transition state) of the highest energy image, without adding significant computational effort. At first, the highest energy image $i_{\max}$ is identified in a few iterations with the regular NEB. Then, the force on this image is replaced by

$$\mathbf{F}_{i_{\max}}^{\text{CI}} = -\nabla E(\mathbf{R}_{i_{\max}}) + 2\nabla E(\mathbf{R}_{i_{\max}})\,\hat{\boldsymbol{\tau}}_{i_{\max}}\,\hat{\boldsymbol{\tau}}_{i_{\max}}, \tag{2.31}$$

which is the full force due to the potential, with the component along the elastic band reverted.

Hence, the spring force does not affect $i_{\max}$ at all. In other words, the climbing image moves up the PES (maximizing it's energy) along the elastic band, and down the PES (minimizing it's energy) perpendicular to the band. The other images define the one degree of freedom for maximizing the energy. Once the band of images converged to the MEP, it gives a good approximation to the reaction coordinate around the saddle point, which is exactly defined by the climbing image. However, a final vibrational analysis is mandatory to confirm whether the climbing image is a first-order saddle point on the PES, with exactly one negative frequency.

Once confirmed, the converged climbing image $i_{\max}$ plays a central role within harmonic transition state theory (TST)[51–53] to determine the reaction rate k, which can be calculated as

$$k = \frac{\prod_j^{3N} \nu_j^{i_{\text{ini}}}}{\prod_j^{3N-1} \nu_j^{i_{\max}}} \exp\left(\frac{-(E_{i_{\max}} - E_{i_{\text{ini}}})}{k_{\text{B}}\, T}\right). \tag{2.32}$$

For a number of atoms N, $\nu_j^{i_{\text{ini}}}$ and $\nu_j^{i_{\max}}$ are the stable normal mode frequencies at the initial and transition state, respectively. $E_{i_{\text{ini}}}$ and $E_{i_{\max}}$ are the corresponding energies, k_{B} is the Boltzmann constant and T is the temperature. The form of Equation (2.32) is analogous to the Arrhenius equation

$$k = \nu \exp\left(\frac{-E_{\text{act}}}{k_{\text{B}}\, T}\right), \tag{2.33}$$

where the activation energy E_{act} is calculated as the difference between the energy of the climbing image, $E_{i_{\max}}$, and that of the initial state, $E_{i_{\text{ini}}}$. The entire exponential term resembles the Boltzmann probability of a reactant system along the reaction coordinate overcoming the energy barrier and transitioning into the product state. The pre-exponential factor ν corresponds to the ratio of the product of the vibrational modes at the initial state over the product of the real-value vibrational modes at the converged climbing image (transition state). Hence, the pre-exponential factor ν can be understood as "attempt frequency", meaning the number of times per second the system vibrates along the reaction coordinate. In other words, the reaction rate k can be interpreted as the product between how often it is tried to overcome the energy barrier with how likely the success is for overcoming the energy barrier.

Efficiently employing the (CI-)NEB method is afflicted with some challenges. First of all, the stability of a NEB calculation strongly depends on the number of images to sample the reaction coordinate. For too few images, the resolution is too low whereas too many imply unstable convergence and high computational effort. Aside from that, the initial NEB band built from interpolation between the reactant and product states might be a bad guess, thus slowing down convergence or biasing the final solution. This is particular true when dealing with amorphous structures, such as liquid water. Therefore, the convergence behavior often has to be carefully tested with respect to the input parameters. Finally, it may be noted again

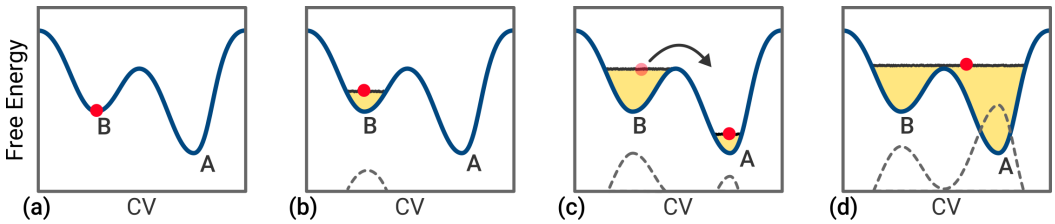


Figure 2.5: Schematic illustration of a metadynamics simulation. The current system state is marked by a red circle within an unknown free energy surface along a predefined collective variable (CV). (a) At first, the simulation reaches a local minimum (basin B). (b) History-dependent Gaussian-type bias potentials are added to that part of the potential energy surface. The summed up bias potentials are marked by a gray dashed line. (c) Once the biased potential is large enough to overcome the energy barrier, the system falls into the global minimum in basin A. Again, Gaussian-like bias potentials are added to that part of the potential energy surface. (d) Finally, the biased potential is larger than the free energy barrier, the free energy landscape is flattened and the system can easily move from one minimum to the other. Summation of the deposited Gaussian-like bias potentials gives an estimate of the negative free energy landscape.

how important the careful equilibration of the considered initial and final configuration is. When the reactants and products are unrealistic or not well-relaxed with respect to the atomic forces, the NEB simulation is bound to yield inaccurate or even wrong results.

2.2.4 Metadynamics

In contrast to obtaining adsorption energies or reaction barriers from static DFT calculations, MD simulations allow a more comprehensive coverage of the phase space and and the consideration of temperature effects. However, high energy barriers, often separating reactant and product states in such processes, are unlikely to be overcome in conventional MD timescales. According to transition state theory, there is an exponential relationship between the height of an energy barrier and the time scale of state transition, as is shown by Equation (2.33). Hence, rare dynamic events of interest often happen on rather large timescales, which are currently unfeasible for AIMD simulations assuming reasonable computational effort. Additionally, local energy minima have to be passed multiple times to obtain statistical meaningful reaction free energies for reasons of ergodicity and entropy, giving further emphasis to long simulation times.

A number of enhanced sampling methods have been developed to address these challenges, one of them being *metadynamics*.[33,54] Metadynamics sampling reconstructs the probability distribution $P(s)$ along one or a few predefined collective variables (CVs), $s_1 \dots s_n$, using Gaussian bias potentials to bypass high energy barriers in the free energy surface. The process is schematically illustrated in Figure 2.5.

It is noteworthy that a constant Gaussian height leads to the free energy not converging to a definite value but rather oscillating around the correct result, giving rise to an average error

which is proportional to the square root of the bias potential deposition rate.[55] Reducing this rate implies an increase of computational time required to sample the free energy surface. Moreover, continuing a run increases the risk that the system is irreversibly pushed into regions of the configurational space which are physically irrelevant. Therefore, the original metadynamics algorithm from Laio and Parrinello[54] has been modified multiple times to address these concerns.[33] In well-tempered metadynamics[55], for example, which is further described in the following, the parameters of the simulation can be tuned such that the computational effort focuses only on the physically relevant regions of the parameter space.[55] In this context, the Gaussian height is continuously adapted, eventually becoming so small after finite time that the free energy completely converges.

For an unbiased probability distribution P, given by

$$P(\mathbf{R}, t) = \frac{\mathrm{e}^{-\beta U(\mathbf{R},t)}}{\int \mathrm{d}\mathbf{R}\, \mathrm{e}^{-\beta U(\mathbf{R},t)}}, \tag{2.34}$$

where $U(\mathbf{R}, t)$ is the potential of a microscopic state for all atomic coordinates $\mathbf{R}$ and $\beta = \frac{1}{k_\mathrm{B}T}$, the free energy surface can be obtained from

$$F(s) = -\frac{1}{\beta} \log P(s) \tag{2.35}$$

along (a set of) predefined CV(s) $s(\mathbf{R})$. The sampling process can be accelerated to better capture rare dynamic events and to estimate the associated free energy by adding history-dependent bias potentials $V(s, t)$ that disfavor frequently-visited space:

$$V(s, t) = \Delta T \ln\left(1 + \frac{\omega N(s, t)}{\Delta T}\right). \tag{2.36}$$

Here, $N(s, t) = \int_0^t \delta_{s,s(t')}\, \mathrm{d}t'$ is the histogram of the CV $s(\mathbf{R})$, ω is an energy scaling term with the dimension of an energy rate and ΔT is a constant bias temperature. The rate with which $V(s, t)$ changes is given by

$$\dot{V}(s, t) = \frac{\omega \Delta T \delta_{s,s(t)}}{\Delta T + \omega N(s, t)} = \omega \exp\left(-\frac{V(s, t)}{\Delta T}\right) \delta_{s,s(t)}. \tag{2.37}$$

By replacing $\delta_{s,s(t)}$ with τ_G—the Gaussian height deposit time step—standard metadynamics (where the height of deposited Gaussian-type bias potentials is constant) can be modulated to include a history-dependent bias potential, which height is determined by $w = \omega \exp\left(-\frac{V(s,t)}{\Delta T}\right) \tau_\mathrm{G}$. For large times, $\dot{V}(s, t)$ approaches zero, allowing to assume a

thermodynamic equilibrium. As a result, the probability distribution eventually converges to $P(s,t)\,\mathrm{d}s \propto \exp\left(\frac{-F(s)-V(s,t)}{T}\right)\mathrm{d}s$, leading to the following relationship between $\dot{V}(s,t)$ and $F(s)$:

$$\dot{V}(s,t) = \omega \exp\left(-\frac{V(s,t)}{\Delta T}\right) P(s,t) = \omega \exp\left(-\frac{V(s,t)}{\Delta T}\right) \frac{\exp\left(\frac{-F(s)-V(s,t)}{T}\right)}{\int \mathrm{d}s \, \exp\left(\frac{-F(s)-V(s,t)}{T}\right)}. \quad (2.38)$$

For $t \to \infty$, the bias potential is given by

$$V(s) = -\frac{\Delta T}{T + \Delta T} F(s), \quad (2.39)$$

showing that the bias does not fully compensate $F(s)$. In practice, however, Equations (2.36) and (2.39) can be used to estimate the free energy surface as

$$\tilde{F}(s,t) = -\frac{T + \Delta T}{\Delta T} V(s,t) = -(T + \Delta T) \ln\left(1 + \frac{\omega N(s,t)}{\Delta T}\right). \quad (2.40)$$

Modulating the constant ΔT allows to define different behavior of the Gaussian deposition. While the added bias potential is zero for $\Delta T = 0$, $\Delta T \to \infty$ results in $-\frac{T+\Delta T}{\Delta T} \to -1$, thus giving $\tilde{F}(s,t) = -V(s,t)$ (as is the case for standard metadynamics). However, when $\Delta T \to \infty$, then the convergence of $V(s, t \to \infty)$ cannot be demonstrated by means of Equation (2.38). This directly reflects the above-mentioned drawback of metadynamics: in a single simulation, the bias does not converge but oscillates around the correct free energy $F(s)$. For finite values of ΔT, the calculated free energy surface corresponds to the target temperature T, while the probability distribution along the CV is altered and corresponds to an enhance temperature $T + \Delta T$. The estimated free energy surface $\tilde{F}(s,t)$ can be recovered from the biased probability distribution $s(\mathbf{R})$ by calculating $V(s,t)$ according to

$$V(s,t) = w \sum_{t'} \exp\left(\frac{-s(\mathbf{R}) + s(\mathbf{R}_G(t'))}{2\delta_s^2}\right)^2 \qquad \text{with } t' = \tau_\mathrm{G}, 2\tau_\mathrm{G}, 3\tau_\mathrm{G}, \ldots \quad (2.41)$$

By tuning ΔT, the exploration in the CV space is facilitated and automatically limited to physically interesting regions, with energies ranging in the order of $T + \Delta T$. In open-source software, such as `PLUMED`[56], ΔT is set up using a bias factor $\gamma = \frac{T+\Delta T}{T}$, defined as the ratio between the (fictional) temperature of the CVs and the temperature of the system. Other input parameters are the Gaussian deposition stride τ_G, the initial Gaussian height w and the Gaussian width σ_G.

2.3 Molecular Machine Learning

Over the past decade, machine learning-based techniques witnessed a tremendous upsurge in numerous industries, including transportation, healthcare and energy, to name a few. Similar to what computers have been in the 1980s and 1990s, machine learning is largely conceived as one of the major disruptive technologies of our time.[57] Machine learning aims at recognizing patterns in data, thus informing ways of treating yet unseen problems. A popular example are self-driving cars. Here, the vast amount of incoming sensor data has to be turned into a decision of how to control the car, made by a computer that "learned" to identify the pattern of a dangerous situation.[57]

After encoding the molecular structures obtained in atomistic simulations into a machine-readable form, they can serve as direct input to train a machine learning model, i.e, to recognize patterns in the data with respect to a target property of interest, the response variable. This property can be of a more theoretical nature, such as the chemical potential or dipole moment, however, also quantities indicating the chemical activity of a compound in the experiment can be predicted, provided a connection to the molecular structure exists. After summarizing a supposed relationship between molecular structure and a response variable in a given dataset of compounds, the machine learning model can predict the properties or activity of yet untested chemicals. Depending on the chosen response variable, the model is then termed quantitative structure-property relationship (QSPR) or quantitative structure-activity relationship (QSAR) model, respectively. Both terms are closely related and describe the quantitative relationship between the structure of a molecule and its, e.g., chemical or biological effect (QSAR) or chemical and physical properties (QSPR).[58,59] Particularly QSAR models exhibit great potential in the field of magnesium corrosion engineering. Instead of identifying potential magnesium dissolution modulators based on merely educated guesses and exhaustive experiments, a given set of tested compounds can be utilized to, for instance, find a corrosion inhibitor with significantly less time investment or generally elucidate relationships between molecular structure and chemical activity.

This can be approached following different strategies. On the one hand, molecular similarities can serve as basis for the qualitative prediction of a target property. Assuming a molecule A yields a chemical activity $f(\mathrm{A})$, then a molecule B is expected to bear a similar chemical activity $f(\mathrm{B})$, when B is structurally similar to A. As a result, it can be effective to search for molecules similar to an already investigated compound with desired behavior in the experiment. This may be desired in order to find a cheaper or more environmentally friendly alternative with the same performance. On the other hand, the response variable can also be quantitatively predicted using regression models.

The following sections further discuss how the molecular structure can be encoded to be used as predictor in a machine learning model. Additionally, qualitative and quantitative methods are described, which allow to reveal and exploit structure-activity relationships to accelerate the search for dissolution modulators in magnesium corrosion engineering.

2.3.1 Encoding the Atomic Structure

In order to apply QSAR (and QSPR) techniques to establish a relationship between molecular structure and a response variables of interest, it is necessary to define what a molecular structure is. The molecular structure can be understood as a set of properties that may be used as numerical representation to quantitatively characterize a molecule.[58] Unfortunately, this collection of parameters is not unique, as various mathematical representations exist. A molecule can be described by means of experimental or theoretical descriptors. While the former are—as the name suggests—obtained in experiments, the latter are calculated using well-specified chemoinformatic algorithms or electronic structure calculations. Thousands of theoretical descriptors have been proposed in literature, which are potentially useful to find a connection to a considered experimental property.[60]

Depending on the rules and criteria used, chemical compounds can be represented in a variety of ways. The molecular representation strongly influences which kind of information, encoded in molecular descriptors, can be collected. By mathematical manipulation of the structure representation, molecular descriptors can be directly derived using a set of mathematical operators. Theoretical descriptors can be subdivided into five classes[58,61]: 0D descriptors do not provide any information about the molecular structure or connectivity of atoms; 1D descriptors regroup structural fragments such as functional groups; 2D descriptors provide information on molecular topology based on the graph representation of the molecule(s); 3D descriptors provide information about the spatial coordinates of atoms in a molecule.; 4D descriptors characterize the interactions between the molecule(s) and the active site(s) of a receptor or the multiple conformational states of the molecule(s). All five classes of molecular representations are listed in Table 2.1 along with examples for associated molecular descriptors.

In contrast to molecular descriptors, molecular fingerprints try to capture different local features of a molecular structure, thus identifying structural fragments participating in the molecule composition. Depending on the set of rules applied to explore a molecule in all possible substructure patterns, different molecular fingerprints can be defined.[58] A highly robust approach is to calculate the smooth overlap of atomic positions (SOAP).[68–70] SOAP is a fingerprint that encodes local atomic environments by representing the local atomic density

Table 2.1: Classes of molecular representations and associated examples.

Class	Examples of molecular descriptors
0D	Atom types, molecular weight, bond types
1D	Counts of atom and bond types, counts of hydrogen bond donors or acceptors, cyclomatic number, number of functional groups by type
2D	Molecular graphs, calculated values such as lipophilicity or topological polar surface area
3D	Geometrical descriptors, such as WHIM[62,63] and GETAWAY[64,65]
4D	Descriptors considering a molecule's electronic distribution or interactive regions, such as or Volsurf[66] or CoMFA[67] methods

ρ_χ within the environment χ as a sum of Gaussian functions with width ξ, centered on each atom of species α:

$$\rho^{\alpha}_{\chi_i}(\mathbf{r}) = \sum_{k \in \chi^\alpha} \exp\left[\frac{(\mathbf{r}-\mathbf{r}_k)^2}{2\xi^2}\right] f_c(|\mathbf{r}_{ik}|). \tag{2.42}$$

The local atomic density is calculated for all atoms in a molecule, with positions $\mathbf{r}_{ik}$ relative to the central ith atom and the cutoff function f_c to smoothly select the particles within a cutoff radius r_c from the central atom (Figure 2.6). The structural information surrounding an atom directly correlates with the size of r_c. In order to compare two molecules A and B, the SOAP kernel can then be defined as the rotationally-averaged squared overlap of all local atomic neighbor densities:

$$k(\chi^{\mathrm{A}}_i, \chi^{\mathrm{B}}_j) = \int_{\mathrm{SO}(3)} \left| \sum_\alpha \int_{\mathbb{R}^3} \rho^{\alpha}_{\chi^{\mathrm{A}}_i}(\mathbf{r})\, \rho^{\alpha}_{\chi^{\mathrm{B}}_j}(\mathbf{r})\, \mathrm{d}\mathbf{r} \right|^2 \mathrm{d}\hat{R}. \tag{2.43}$$

A notable property of the SOAP kernel is its rotational and translational invariance. The integration over all rotations can be carried out analytically, based on a spherical harmonics decomposition of the two densities ρ_χ.[68,69] For more details, please check the original publication by Bartók et al..[69] Normalizing the SOAP kernel results in the environmental covariance matrix $\mathbf{C}$ that contains all possible pairings of environments, and thus all the information on the pair-wise similarities between the molecules A and B:

$$C_{ij}(\mathrm{A},\mathrm{B}) = \frac{k(\chi^{\mathrm{A}}_i, \chi^{\mathrm{B}}_j)}{\sqrt{k(\chi^{\mathrm{A}}_i, \chi^{\mathrm{A}}_i)\, k(\chi^{\mathrm{B}}_j, \chi^{\mathrm{B}}_j)}}. \tag{2.44}$$

Based on this matrix, a global kernel $K(\mathrm{A},\mathrm{B})$ can be introduced, which allows to directly compare two compounds.

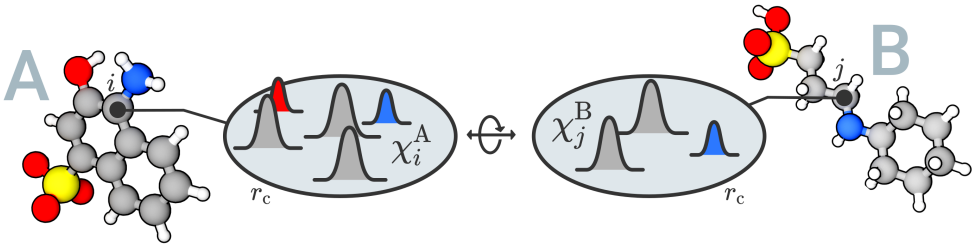


Figure 2.6: The SOAP kernel. For comparing two molecules A and B, the rotationally-averaged squared overlap between all local atomic environments, χ_i^{A} and χ_j^{B}, within a cutoff radius r_c is calculated and summarized in a global similarity measure $K(\mathrm{A},\mathrm{B})$. The SOAP kernel is rotationally and translationally invariant. For more details, please check the original publication by Bartók et al..[69]

In order to extract a single similarity measure from $\mathbf{C}$, several strategies can be pursued, of which two are further described in the following. An inexpensive approach is to simply store the average SOAP fingerprint between all environments of two structures, resulting in the average structural kernel

$$\bar{K}(\mathrm{A},\mathrm{B}) = \frac{1}{N^2}\sum_{ij} C_{ij}(\mathrm{A},\mathrm{B}). \tag{2.45}$$

However, two significantly different structures can consist of environments that yield the same fingerprint upon averaging, thus making these structures seem to be the same and rendering this approach potentially insensitive. An alternative approach is to build the regularized entropy match (REMatch) kernel. Here, the pairs of local environments that exhibit the highest degree of structural similarity are used to build the global similarity metric. For this reason, the similarity between the compounds A and B is given by the weighted sum over the elements of $\mathbf{C}(\mathrm{A},\mathrm{B})$:

$$\begin{aligned}\hat{K}_\gamma^\zeta(\mathrm{A},\mathrm{B}) &= \left[\mathrm{Tr}\,\mathbf{P}_\gamma \mathbf{C}(\mathrm{A},\mathrm{B})\right]^\zeta \\ \mathbf{P}_\gamma &= \underset{\mathbf{P}\in\mathcal{U}(N,M)}{\mathrm{argmin}} \sum_{ij} P_{ij}(1 - C_{ij}(\mathrm{A},\mathrm{B}) + \gamma \ln P_{ij}).\end{aligned} \tag{2.46}$$

The optimal pairing is found by exploring the space of doubly stochastic matrices $\mathcal{U}(N,M)$, where the discrepancy between matching pairs of environments is minimized and then regularized with the information entropy of the weight matrix $E(\mathbf{P}) = -\sum_{ij} P_{ij}\ln P_{ij}$.[68] While the sensitivity of the kernel is controlled by the parameter ζ, switching between a strict or broad selection of best-matching pairs of local environments is achieved by adapting the parameter γ. Since both kernels, $\bar{\mathbf{K}}$ and $\hat{\mathbf{K}}_\gamma^\zeta$, are positive-definite, it is possible to define

a kernel-induced distance, which can be used for dimensionality reduction and clustering:

$$D(\mathrm{A},\mathrm{B})^2 = K(\mathrm{A},\mathrm{A}) + K(\mathrm{B},\mathrm{B}) - 2K(\mathrm{A},\mathrm{B}). \tag{2.47}$$

2.3.2 Molecular Similarity Maps

The calculated similarities and distances respectively contained in the SOAP kernel $\mathbf{K}$ and the distance matrix $\mathbf{D}$ describe relations between different molecular structures. In order to visualize and better understand these relations, they have to be transformed to something that is tangible by the human mind. Various dimensionality reduction algorithms exist that allow to map information from a higher-dimensional space to two or three dimensions. Particularly two-dimensional mappings of the computed SOAP kernel allow a highly intuitive interpretation of molecular similarities between the considered compounds.

A popular strategy to reduce dimensions is to find the low-dimensional Cartesian projection that best reproduces the pairwise distances in the high-dimensional space. This approach, called multi-dimensional scaling, forms the foundation of the non-linear multi-dimensional scaling algorithm *sketch-map*[20,21,71] that is based on optimizing a non-linear objective function[68]

$$S^2 = \sum_{ij} \left(F\left(D(\mathbf{X}_i, \mathbf{X}_j)\right) - f\left(d(\mathbf{x}_i, \mathbf{x}_j)\right) \right)^2 \tag{2.48}$$

and describes the mismatch S between molecular (dis)similarities $D(\mathbf{X}_i, \mathbf{X}_j)$ in high-dimensional and $d(\mathbf{x}_i, \mathbf{x}_j)$ in low-dimensional space. While D is a kernel-induced distance metric between the abstract molecular input structures $\mathbf{X}_i$, as shown in Equation (2.47), d is typically the Euclidean distance between the corresponding projections $\mathbf{x}$ in low dimension. The transformations F and f are non-linear sigmoid functions of the form

$$f(r) = 1 - (1 + (2^{a/b} - 1)(r/\sigma)^a)^{-b/a} \tag{2.49}$$

that focus the optimization problem in Equation 2.48 on the most significant distances. By tuning the involved switching distance σ and the exponents a and b (which are denoted as A and B for the high-dimensional function F), close/distant (similar/dissimilar) structures in high-dimensional space $\mathbb{R}^n$ are projected to lower dimensions $\mathbb{R}^m$ ($m < n$), keeping their relation. Due to the form of the sigmoid function, mapped molecular similarities or dissimilarities appear arbitrarily close together or far apart, respectively, thus preventing a clear physical interpretation of the resulting two-dimensional distances. However, although the accuracy in reproducing the distances is low, the quality of reproducing the connectivity information is high. While a and b (or A and B in high dimension) only have a small effect

on the projection, the resulting sketch-map is highly sensitive to the switching distance σ. A typical approach for tuning σ is to first compute the histogram of distances in the similarity matrix $\mathbf{D}$, detecting the peak value D_{max} and then setting $\sigma \approx 0.8 D_{\text{max}}$.

Another dimensionality reduction technique involving significantly less hyperparameters is kernel principal component analysis (KPCA). KPCA is an extension of principal component analysis (PCA), which is capable of building non-linear mappings that maximize the variance in the data. Structural points are projected to the eigenvectors corresponding to the highest eigenvalues, the principal components, of the covariance matrix constructed from the input data. In KPCA, the originally linear operations of PCA are performed in a reproducing kernel Hilbert space.[ii]

Apart from sketch-map and KPCA, various other dimensionality reduction techniques exist that differ in the implemented projection scheme.[72] Applying a dimensionality reduction algorithm to a calculated SOAP kernel $\mathbf{K}$ or distance matrix $\mathbf{D}$ of size $n \times n$, where n denotes the number of compounds in a given dataset, the high-dimensional similarity data can be effectively reduced to two dimensions, resulting in a similarity map of size $n \times 2$ (Figure 2.7). Each point in this map corresponds to a molecular compound in the original dataset. Compounds lying close together show high structural similarity, whereas structures lying far apart can be considered as rather dissimilar. Coloring the map according to a property of interest, such as some quantity representing the chemical activity, adds again a dimension that allows to reveal structure-activity (structure-property) relationships, indicated by formed clusters. When a cluster predominantly contains structures with mutual properties, a yet unknown compound mapped to this cluster usually behaves accordingly. Here, the strong capacity of KPCA and sketch-map may be emphasized, which allows for an out-of-sample embedding of new datapoints in an existing similarity map.

Depending on the input data, clusters might not always be obvious to detect in the resulting structure-property landscape. Although visual inspection often gives a first idea, more sophisticated methods promise more reliable results. A number of algorithms for cluster detection exist that can generally be divided into four categories: centroid-based clustering, density-based clustering, distribution-based clustering and hierarchical clustering. One of the most well-known clustering algorithms stems from the first category and is called k-means clustering.[73] The algorithm aims to partition the data into k clusters, where each of the samples n belongs to the cluster with the nearest mean (cluster center). At first, k initial means are randomly generated within the data domain. Accordingly, k clusters are

[ii] Hilbert spaces allow to generalize linear algebra and calculus methods from two- and three-dimensional Euclidean spaces to spaces with infinite dimension. In reproducing kernel Hilbert spaces, point evaluation is a continuous linear functional.

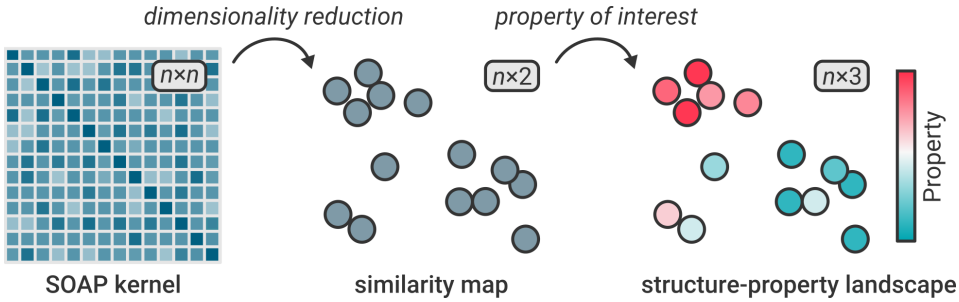


Figure 2.7: From SOAP to structure-property landscapes. Using dimensionality reduction techniques, the high-dimensional SOAP kernel can be reduced to a two-dimensional representation of the molecular similarities, where points that are close/distant correspond to molecular compounds that are similar/dissimilar. Coloring the resulting similarity map according to a property of interest results in a structure-property landscape, highlighting structure-property relationships.

created by assigning each sample to the nearest mean (Voronoi diagram), then the centroid of each of the k clusters becomes the new mean. The procedure is repeated until the sum of squared distances between samples and means converges. To enhance the cluster detection performance, the number of clusters has to be fine-tuned with respect to the sum of squared errors or the silhouette coefficient.[74]

2.3.3 Experimental Target Prediction

Obtaining experimental data is often time- and resource-consuming, or even potentially perilous when dealing with toxic chemicals. Hence, building machine learning models that predict an experimental property of interest can constitute an effective way of strongly limiting the number of required experiments. Some examples for potential target properties are the biological or chemical activity of the considered chemicals, their toxicity or their reactivity. In the realm of magnesium corrosion engineering, a well adopted parameter to describe the impact of a dissolution modulator on the corrosion rate is the inhibition efficiency (IE):

$$\mathrm{IE} = \frac{V_{\mathrm{H_2}}^{0} - V_{\mathrm{H_2}}^{\mathrm{mod}}}{V_{\mathrm{H_2}}^{0}} \cdot 100\%, \tag{2.50}$$

where $\mathrm{V}_{\mathrm{H_2}}^{0}$ and $\mathrm{V}_{\mathrm{H_2}}^{\mathrm{mod}}$ denote the volume of hydrogen released when exposed to sodium chloride solution, in the absence or presence of the investigated dissolution modulator, respectively. Naturally, for QSAR/QSPR models the chosen target property has to correlate with the molecular structure at least to a certain extent to achieve good predictive performance.

When linked to an experimental dataset, molecular descriptors or SOAP fingerprints can be used as direct input (features) of a regression model to quantitatively predict the

experimental target property (e.g., IE) of yet untested compounds. Depending on the type of regression model, it can happen that the fitted parameters lead to highly accurate descriptions of the given data but fail to describe additional data or predict future experiments dependably. An approach to address this phenomenon called overfitting and to reduce the variance of the regression model is to introduce a regularization term that prevents the regression weights from becoming too large. For ridge regression, which includes an $\mathcal{L}^2$ regularization term, the loss function to be minimized is defined as

$$\ell = ||\mathbf{Y} - \mathbf{X}\mathbf{P}_{XY}||^2 + \lambda||\mathbf{P}_{XY}||^2, \tag{2.51}$$

where $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{P}_{XY}$ are the response variable(s), predictors (features) and the weights for the regression model, respectively. The optimal regularization parameter λ has to be determined in a hyperparameter tuning.

Although linear regression methods are appealing for their simplicity and interpretability, they do not generate robust and predictive models when the underlying structure-activity relationship is significantly non-linear. Kernel methods, as in the case of the SOAP kernel, introduce a non-linear relationship between samples in the form of a positive-definite kernel function $k(\mathbf{X}, \mathbf{X}')$ and are thus suitable for modeling non-linear relationships. Generally, they lead to a more efficient determination of low-dimensional latent space and increased regression performance.[75] Kernel ridge regression (KRR) is the kernel equivalent to ridge regression, using the kernel (feature) matrix as input variable.

The synergistic combination of KPCA and KRR represents a powerful approach for efficient in silico design or discovery of novel magnesium dissolution modulators. However, methods that unify the pattern recognition capabilities of unsupervised learning[iii] and the predictive power of supervised learning may have increased accuracy and efficiency and may be better at identifying the most important chemical features in structure-activity relations. Kernel Principal Covariates Regression (KPCovR) achieves that by a combination of a KPCA-like and a KRR-like loss function, attempting to find a low-dimensional projection of the feature vectors that simultaneously minimizes information loss and error in predicting the target properties.[75] Applications of KPCovR to other datasets of molecules and materials[75] clearly demonstrate its utility in deciphering structure-activity relationships.

[iii] Type of algorithm that learns patterns from unlabeled data. In contrast, supervised learning describes a type of algorithm that learns from labeled data.

Chapter 3

A Deeper Look at the Magnesium-Water Interface

All things are difficult before they are easy.

— Thomas Fuller

Magnesium corrosion is an electrochemical process in which the metal's oxidation state changes as a result of interaction with the environment. As this process is diffusion-controlled[i], it occurs on exposed surfaces and is thus mainly defined by the interface between the solid and its surroundings. Particularly in aqueous environments, Mg and its alloys possess comparatively poor corrosion resistance as compared to other metals, which can be mainly attributed to two key factors.[76,77] Firstly, the highly electronegative potential of Mg allows corrosion to proceed even in the absence of molecular oxygen.[76,77] The cathodic water reduction reaction predominates at such negative potentials, thus promoting hydrogen evolution and implicitly magnesium dissolution. Secondly, any surface film formed upon Mg yields poorly protective properties[76,77] as oxide or hydroxide layers forming upon Mg are soluble in most aqueous environments or in the presence of humidity. Moreover, Mg surface layers merely incompletely cover the underlying Mg metal surface, are highly defective and are easily dismantled by interaction with chloride.[76]

A sustainable implementation of magnesium-based materials in aqueous environments demands effective measures for corrosion protection or control, specific to the given target application. This, however, requires profound knowledge of the underlying corrosion mechanisms. As experimental approaches often require complementary information to explain the occurring corrosion mechanisms on an atomistic level in detail, quantum mechanical computations constitute a highly promising strategy to gain deeper insight into the fundamental electrochemical reactions of Mg corrosion in an aqueous environment.[78–80]

[i] The reactants diffuse through the reaction medium until they collide in the right stoichiometry and form an activated complex which can form the product species. The rate of transport of the reactants through the reaction medium, usually a solution, defines the reaction rate.[44]

However, the physical concept behind magnesium corrosion is not determined by a single mechanism, but rather resembles a complex interplay of different aspects, such as corrosion product formation[81–83], anodic regions[84] and noble impurities[85]. Also the formation of a Mg^+ species which promotes water reduction in the context of anomalous hydrogen evolution is a constant topic of discussion.[76,86–88] Consequently, establishing an accurate computational model that incorporates all, or at least most, of the inherent factors that influence the corrosion of Mg is highly complex. In particular, the atomistic simulation of electrified solid/liquid interfaces is still highly challenging and lacks simple and intuitive approaches. Although recently developed techniques to combine electrochemistry and thermodynamics in atomistic simulations allow to actively relate quantum mechanical results with electrochemical parameters, such as voltage and pH, this still remains a challenging task.[40,89–93] Moreover, sophisticated quantum mechanical calculations are typically limited to rather small system sizes of a few hundred atoms and a few picoseconds of time due to high computational cost. Finally, such calculations must be validated by experimental results, since the computer will eventually produce some kind of output, whether it makes physical sense or not.

Several methods exist in the DFT framework that allow to calculate reaction kinetics, observe the charge evolution or generally identify favorable reaction mechanisms. These methods require some prior knowledge, or at least intuition, of potential corrosion mechanisms that has to be incorporated in the model building process to ensure chemical soundness. A peculiarity of Mg is that the hydrogen evolution rate increases with increasing applied anodic potential or current for anodic polarization, which is opposite of the expected behavior based on standard electrochemical kinetics.[94] This so-called anomalous hydrogen evolution is a controversial topic in magnesium corrosion research.[76,77] An overview of the governing Mg corrosion reactions and popular theories for anomalous hydrogen evolution is given in the beginning of this chapter. Subsequently, further investigations of the energetics, kinetics and charge states are performed to obtain deeper insights into the reaction pathways responsible for hydrogen evolution and magnesium dissolution. Parts of the work and figures presented hereafter have been published in

---

[95] **T. Würger**, C. Feiler, G. B. Vonbun-Feldbauer, M. L. Zheludkevich, and R. H. Meißner. A first-principles analysis of the charge transfer in magnesium corrosion. *Scientific Reports*, 10(1):15006, 12 (2020).

[96] B. Zeller-Plumhoff, M. Gile, M. Priebe, H. Slominska, B. Boll, B. Wiese, **T. Würger**, R. Willumeit-Römer, & R. H. Meißner. Exploring key ionic interactions for magnesium degradation in simulated body fluid – a data-driven approach. *Corrosion Science*, 109272 (2021).

3.1 Magnesium Corrosion Mechanisms

In aqueous environments, the two main Mg corrosion products are magnesium hydroxide and hydrogen. The overall corrosion reaction[76] is a redox reaction given as

$$\mathrm{Mg} + 2\,\mathrm{H_2O} \rightleftharpoons \mathrm{Mg(OH)_2} + \mathrm{H_2} \tag{3.1}$$

and is reversible, although strongly favoring the formation of products. This process occurs by electron transfer at the magnesium-water interface and involves the oxidation of Mg atoms to form ionic species, accompanied by the release of electrons. Electrochemistry classifies corrosion reactions as either anodic or cathodic processes. The oxidation of Mg, the reducing agent, is an anodic half reaction and defined as

$$\mathrm{Mg} \rightleftharpoons \mathrm{Mg^{2+}} + 2\,\mathrm{e^-} \quad \text{(anodic)}. \tag{3.2}$$

The liberated electrons must be consumed by other species to maintain electroneutrality. Hence, the anodic process must be accompanied by a cathodic half reaction in which an oxidizing agent (a molecule, atom, or ion) receives electrons, thus decreasing its oxidation state (reduction). Dependent on the environment, several different oxidizing species can act as electron acceptors. However, the primary cathodic reactions in corrosion are the hydrogen evolution reaction (HER) and the oxygen reduction reaction (ORR). For the HER, water reduction leads to the evolution of hydrogen and the formation of hydroxide ions:

$$2\mathrm{H_2O} + 2\,\mathrm{e^-} \rightleftharpoons \mathrm{H_2} + 2\,\mathrm{OH^-} \quad \text{(cathodic)}. \tag{3.3}$$

In contrast to the HER, in ORR oxygen is reduced instead of water, without the release of molecular hydrogen:

$$\mathrm{O_2} + 2\mathrm{H_2O} + 4\,\mathrm{e^-} \rightleftharpoons 4\,\mathrm{OH^-} \quad \text{(cathodic)}. \tag{3.4}$$

Although ORR also occurs as a cathodic process and contributes to the overall corrosion reaction in Equation (3.1), its contribution is comparably small for high Mg corrosion rates

and is thus neglectable. Dissolved magnesium ions from the anodic half reaction recombine with hydroxide ions formed in the cathodic reactions, yielding magnesium hydroxide in the process:

$$\mathrm{Mg^{2+}} + 2\,\mathrm{OH^-} \rightleftharpoons \mathrm{Mg(OH)_2} \tag{3.5}$$

When both half reactions are in equilibrium, the associated equilibrium potential, also called open circuit potential, is given by the *Nernst equation*:

$$E = E_0 + \frac{RT}{n_\mathrm{e} F} \ln\left(\frac{a_\mathrm{Ox}}{a_\mathrm{Red}}\right), \tag{3.6}$$

where E_0 is the standard potential, R the universal gas constant, T the temperature, n_e the number of transferred electrons, F the Faraday constant and a the activity of the according redox species.[97] The Nernst equation describes the dependency of the electrode potential on the concentration of the considered oxidized and reduced species. By considering different occurring surface reactions and coverages of involved species in a DFT simulation, Equation (3.6) can be also used to calculate phase and Pourbaix diagrams employing ab initio methods.[91,98] As both half reactions are in equilibrium, the net current i is zero, since the system contains a cathodic current i_c that is balanced by an equal and opposite anodic current i_a:

$$i_0 = i_\mathrm{a} = i_\mathrm{c}, \tag{3.7}$$

where i_0 is the exchange current density.

Under physiological conditions—as is the case for magnesium implants—various additional species contribute to the degradation of magnesium, leading to increasingly complex corrosion mechanisms. Studying the degradation of pure magnesium wire in simulated body fluid, Zeller-Plumhoff et al. investigated the interaction of individual ionic components ($\mathrm{HCO_3^-}$, $\mathrm{HPO_4^{2-}}$ and $\mathrm{Ca^{2+}}$) with respect to the degradation of Mg and the formation of precipitates as a protective layer.[96] Although the formation of Ca-containing precipitates was kinetically difficult, the presence of $\mathrm{Ca^{2+}}$ obstructs the interaction between $\mathrm{Mg^{2+}}$ and $\mathrm{HCO_3^-}$ or $\mathrm{HPO_4^{2-}}$ to such an extent that a formation of Mg precipitates was impossible and therefore little to no protective layer was formed. Correlating precipitation and degradation rates suggests that a Mg implant in the human body would require a degradation rate at levels below 2 mm per year to enable the formation of a stable degradation layer to which cells may adhere. No apparent quantitative relationship between the interaction of ionic components and other experimental parameters could be identified that would allow for accurate estimations of the Mg degradation rate. Atomistic modeling techniques could serve as valuable tool in an attempt to better understand the occurring degradation process under physiological

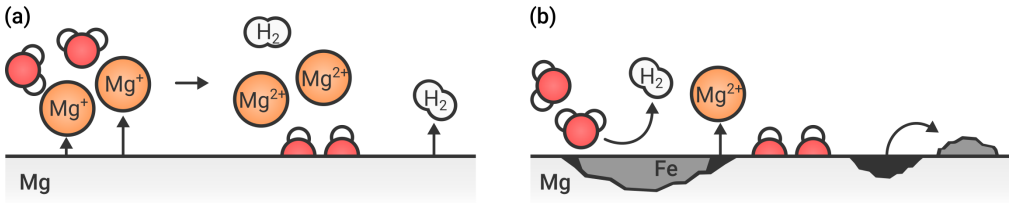


Figure 3.1: Theories for anomalous hydrogen evolution. (a) *Monovalent magnesium ion*. Assuming Mg dissolves as intermediate monovalent Mg^+, a reaction between the released Mg^+ species and water would lead to additional hydrogen gas formation. Hence, the more magnesium dissolves, the more hydrogen evolves, thus explaining the increased cathodic reaction rates for anodic overpotentials. (b) *Particle undermining*. Noble impurities induce local galvanic cells that promote local corrosion, continuously undermining the enclosed particles. Once these particles lose their cathodic protection due to dissolution of the surrounding Mg matrix, they dissolve and can redeposit upon the Mg surface, increasing the cathodic area and thus the cathodic reaction rate.

conditions, however, finding a trade-off between model complexity and feasibility remains a yet unsolved and challenging task.

3.1.1 Anomalous Hydrogen Evolution

A peculiarity of magnesium is an electrochemical phenomenon known as anomalous hydrogen evolution.[76,77] Usually, the anodic reaction rate increases and the cathodic reaction rate decreases with an increase of the current density or applied potential. As a result, for most metals, anodic overpotentials evoke an increase of the anodic dissolution rate, whereas the cathodic hydrogen evolution rate decreases simultaneously. In case of Mg, however, the hydrogen evolution behavior seems to contradict the very basics of electrochemical fundamentals. When increasing the applied anodic potential, the cathodic reaction rate does not decrease, but rather increases further. Moreover, the anodic dissolution current increases faster than expected. For pure magnesium and magnesium alloys, different mechanisms have been proposed to explain this phenomenon.[76,77] It is striking that nearly 20 years lie between two comprehensive review articles published by Song and Atrens and Esmaily et al., summarizing these theories, and there is still no consensus on the main driving force behind anomalous hydrogen evolution.

Two popular explanations are the *monovalent magnesium ion* theory[86–88,99,100] and the *particle undermining* model[85,101] (Figure 3.1). The former theory proposes that Mg could also be dissolved as intermediate monovalent Mg^+. In a secondary reaction between the released Mg^+ species and water, additional hydrogen gas would be formed. Hence, the more magnesium dissolves, the more hydrogen evolves, thus explaining the increased cathodic reaction rates for anodic overpotentials. Although many observations can be explained by assuming the production of Mg^+, there is no persuasive experimental evidence. Furthermore,

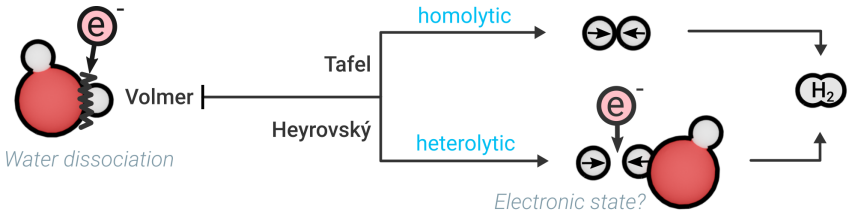


Figure 3.2: Hydrogen evolution reaction (HER) pathways. Initially, water is reduced in the Volmer step and a dissociated H atom adsorbs on the surface. Subsequently, one out of two possible mechanisms leads to hydrogen evolution: either two adsorbed H atoms recombine in the homolytic Tafel step, or an adsorbed H atom reacts directly with a water molecule in the heterolytic Heyrovský step.

if Mg^+ indeed could exist in solution, it would be extremely reactive and would thus be inflicted with an extremely short lifetime.[86] According to Petty et al., however, it must be stable for a considerable amount of time without being oxidized by water.

During Mg corrosion, a dark corrosion film forms at open circuit potential. Although it could be observed that this film promotes the evolution of hydrogen at a faster rate than uncorroded regions on the Mg surface[102], it can not be held solely responsible for the increased catalytic activity towards the HER. The hydrogen evolution rate increases with respect to the applied anodic current density, not the amount of corroded area.[76] The *particle undermining* model explains anomalous hydrogen evolution in terms of noble impurities, such as iron, that accumulate in or under the corrosion product film.[103] Due to the low electronegativity of magnesium, essentially all transition metals potentially enclosed in the Mg matrix are more noble, inducing local galvanic cells[ii] and thus accelerated local corrosion.[104] Corrosion of the adjoining magnesium matrix continuously undermines the enclosed particles until they eventually leave the Mg matrix by non-Faradaic release. According to a scheme proposed by Höche et al., the released particles can redeposit on the Mg surface after dissolution, thus extending the cathodic area and further increasing the hydrogen evolution rate.[85] For anodic polarization, the increased Mg dissolution rate would then entail an enhanced release of enclosed impurities, which would further increase the hydrogen evolution rate through redeposition on the magnesium surface.

3.1.2 Reaction Pathways

The hydrogen evolution process in Equation (3.3) can be divided into two separate stages (Figure 3.2). Initially, water is reduced in the Volmer step, followed by adsorption of the

[ii] Local galvanic cells are electrochemical cells in which a Faradaic current is generated in spontaneous redox reactions due to physical or chemical differences between adjacent areas on a metal surface in an electrolyte.[97]

dissociated hydrogen atom upon the magnesium surface:

$$H_2O + e^- \rightleftharpoons OH^- + H_{ad} \quad \text{(Volmer)}. \tag{3.8}$$

In the second stage, hydrogen evolves following one out of two possible mechanisms.[89,90,105,106] The adsorbed hydrogen atom either recombines with another adsorbed hydrogen atom in the homolytic Tafel step, or reacts directly with the partially positive hydrogen atom of a close water molecule in the heterolytic Heyrovský step:

$$H_{ad} + H_{ad} \rightleftharpoons H_2 \quad \text{(Tafel)} \tag{3.9}$$

$$H_2O + e^- + H_{ad} \rightleftharpoons OH^- + H_2 \quad \text{(Heyrovský)}. \tag{3.10}$$

The energetics and kinetics of the proposed hydrogen evolution reactions occurring on a magnesium surface during corrosion were already discussed in the literature.[40,78,79,107,108] In a theoretical study by Yuwono et al., the Volmer-Heyrovský pathway was found to be energetically more favorable than the Volmer-Tafel pathway, and thus, to largely contribute to the hydrogen evolution at cathodic and anodic overpotentials.[78] However, the reaction in Equation 3.10 requires excess electrons. In a study by Surendralal et al., the authors found that the adsorbed hydrogen atom is a hydride (H^-)[109] and not in a charge neutral state as has been previously assumed.[43] Thus, they proposed a "Heyrovský-like" reaction

$$H_2O + H^-_{ad} \rightleftharpoons OH^- + H_2, \tag{3.11}$$

where the excess electron required for Equation (3.10) is provided by the adsorbed hydride. This is in good agreement with studies on Mg-based alloys and hydrogen storage systems[110–112] that also observed adsorption of hydrides on the used substrate. In the course of this chapter, the magnesium corrosion mechanisms are further investigated with respect to the potential HER pathways, also taking into account the above-mentioned theories for anomalous hydrogen evolution.

3.2 Hydrogen Evolution at the Magnesium Surface

Quantum mechanical methods have high potential for modeling parts of the Mg corrosion process, as they provide insights into the underlying electrochemical mechanisms, particularly addressing reaction kinetics as well as energy states and the electronic structure. Although many aspects of the HER mechanism upon magnesium surfaces have been addressed before in the literature, an explicit investigation of the charge transfer during hydrogen evolution

has not been conducted yet. Detailed studies on an atomistic level will facilitate to fully unravel the reaction mechanisms accompanying Mg corrosion and may help to identify key factors of anomalous hydrogen evolution. In the following, more information on the surface model and the employed methods is given that form the basis of the subsequent calculations. The calculated reaction kinetics are compared to experimental results to identify favorable mechanisms.

3.2.1 Surface Model System

All results presented in this section were obtained from DFT calculations using the plane-wave code Vienna Ab Initio Simulation Package (VASP)[113–116] with the projector augmented-wave method[117,118] (cf. Section 2.1). Since van der Waals (vdW) interactions are a key factor in hydrogen evolution reactions[119], the exchange-correlation functional optB88-vdW[27,120–124] was employed for all DFT computations. This functional accounts for dispersion interactions in an approximate fashion by including a non-local correlation part in the exchange-correlation energy and is expected to capture well the subtle energetic contributions of hydrogen bonded systems in solid-liquid interfaces.[119,125,126]

All simulations were performed using a $6 \times 6 \times 1$ gamma-centered grid of k-points.[127] The plane-wave expansion was limited by a cutoff energy of 520 eV. During the relaxation process, the atom positions were allowed to adjust until the atomic forces were less than 5 meV Å^{-1}. Subsequently, the structure was investigated statically to obtain more accurate total energies via the tetrahedron method with Blöchl corrections.[128] Convergence tests and chosen parameters for the DFT calculations described in the following are presented in Appendix A.1.

In order to find a good trade-off between simplicity and accuracy in the model building process, profound knowledge of the magnesium-water interface is crucial. At room temperature and in the absence of water, Mg instantaneously forms a Mg oxide layer in an exothermic reaction with air. In aqueous immersion, this MgO film is hydroxylated, thus forming a predominant $Mg(OH)_2$ (brucite) layer.[76] Since both MgO and $Mg(OH)_2$ are soluble in water, acidic and neutral pH conditions tend to dissolve the passivating MgO/$Mg(OH)_2$ film, exposing the metal to the solution. Furthermore, deviations in the crystal structures between the MgO/$Mg(OH)_2$ film and the substrate may induce internal stresses leading to cracks in the passivation layer. In addition to that, noble impurities enclosed in the Mg matrix might not be covered by the MgO/$Mg(OH)_2$ passivation film from the start, such that local corrosion is readily triggered. Accordingly, the corrosion reactions, i.e., hydrogen evolution and Mg dissolution, take place at the exposed magnesium surface, where the passivation

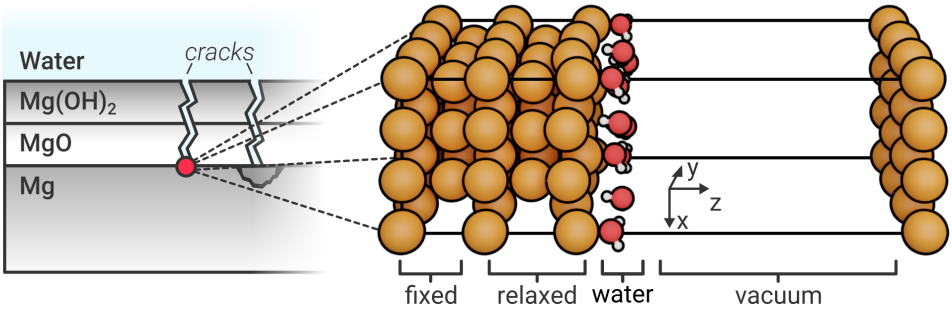


Figure 3.3: Surface model for simulating hydrogen evolution upon Mg(0001). The exposed corroding magnesium surface is represented by a 3×3 Mg(0001) slab consisting of five atomic layers. The two lowermost layers are fixed, the remaining are allowed to relax. The solid-liquid interface is modeled by explicit water molecules. A vacuum region of around 20 Å above the water layer is added to avoid interaction between the periodic images in z-direction.

layer is dissolved, damaged or non-existent.

The exposed magnesium surface was modeled using a Mg(0001) slab of five Mg layers in a 3×3 supercell (cf. Section 2.2.1). Here, the Mg(0001) surface was considered as it is energetically most favorable when compared to other magnesium surfaces.[129] Atoms in the two lowermost layers were kept fixed in their bulk-like positions, whereas the remaining layers were free to relax, thus being able to respond to occurring forces due to surface effects or adsorption. Relaxations of the bulk resulted in lattice constants $a = 3.19$ Å and $c = 5.15$ Å, which are in good agreement with theoretical and experimental values stated in the literature.[40,107,130] The solid-liquid interface was modeled by adding (at least) one bilayer of water consisting of six explicit water molecules (see Figure 3.3). A vacuum region of around 20 Å was added above the water bilayer to avoid interaction between the periodic images. Additionally, a dipole correction[131] was applied to compensate for the slab asymmetry.

Several structural aspects have to be elucidated prior to developing a representative system for the hydrogen evolution reactions during aqueous Mg corrosion. Firstly, water plays a major role in the primary reaction mechanisms, i.e., water dissociation (Volmer pathway), hydrogen evolution (via the Tafel or Heyrovský pathway) and magnesium dissolution. The aqueous environment can be modeled in different ways, including implicit solvent models[108] as well as the addition of explicit water molecules[78]. In accordance with the literature[78,132,133], ice-like hexagonal water bilayers were found to be the most stable configuration at the magnesium-water interface in DFT calculations, for a surface saturated with H_2O. Here, every second water molecule binds with its oxygen molecule to the top site of the magnesium surface at a

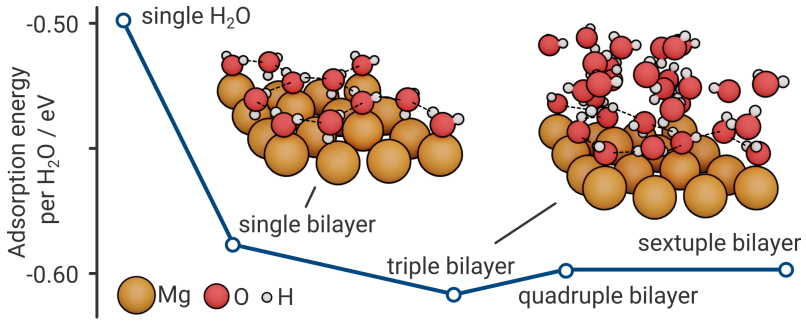


Figure 3.4: Adsorption energy per H_2O molecule with respect to the amount of stacked water bilayers. The ice-like hexagonal water bilayer structure at the solid–liquid interface is indicated by dashed lines. The results indicate that a single water bilayer describes the adsorption energy sufficiently as additional bilayers exhibit an insignificant impact on the calculated energies. Adapted from Figure 1 in [95].

distance of 2.16 Å. The other water molecules may adopt two possible orientations, referred to as H-down or H-up structures, with one hydrogen atom either pointing towards or away from the surface, respectively.[132] With a calculated energy difference $\Delta E_{\text{up-down}} = 0.05\,\text{eV}$, the H-down configuration is energetically slightly more favorable for a single bilayer (consisting of six explicit water molecules). For multiple water bilayers, the system relaxed to an unstructured mixture of both configurations due to interlayer hydrogen bonds. Despite a growing hydrogen bond network for additional water bilayers, which also influences molecular orientations at the magnesium-water interface, the adsorption energies per H_2O molecule remain mostly constant ($E_{\text{ad},H_2O} = 0.60 \pm 0.01\,\text{eV}$) after full saturation of the Mg(0001) surface (Figure 3.4). As shifts in the water structure can evoke dominating changes in the energetic landscapes of the analyzed reaction pathways, the following calculations include only a single bilayer to avoid such effects, assure comparability and focus the reaction mechanism. However, DFT calculations and NEB simulations of the Volmer-Tafel and Volmer-Heyrovský pathways revealed that a varying number of water bilayers has an influence on the energy of formation ΔE_f as well as the reaction barriers (Table 3.2). Hence, calculations for the entire HER cascade have also been conducted for three water bilayers to confirm any observed trends. For models concerning the magnesium dissolution reaction pathway, four bilayers were used to ensure sufficient amount of water for a full solvation shell around the magnesium ion.

3.2.2 Charge Evolution Sampling

Identifying energy barriers in potential corrosion mechanisms increases the understanding of reaction kinetics and enables finding more favorable reaction pathways for corrosion models. Furthermore, electron transfers play a significant role and thus also contributes to the reaction

Table 3.2: Reaction energies and barriers with respect to included water bilayers. Values are shown for the Volmer, Tafel and Heyrovský step in the hydrogen evolution process upon the Mg(0001) surface with respect to the number of explicitly modeled water bilayers. For the in vacuo case, no energy barriers were determined.

	Volmer			Tafel			Heyrovský		
# Water bilayers	0	1	3	0	1	3	0	1	3
Reaction energy E_r	-1.68	-1.45	-1.82	0.20	0.23	0.76	-1.36	-1.17	-1.05
Energy barrier E_b	—	0.66	0.55	—	1.27	1.63	—	0.28	0.41

energetics, as can be seen in the contemplable reaction pathways in Mg corrosion, e.g., Equation (3.2) and (3.3). Coupling chain-of-states methods with charge analysis algorithms allows to simulate the minimum energy pathways (MEPs) and associated charge transfers in potential hydrogen evolution pathways as well as Mg dissolution reactions.

The CI-NEB method (cf. Section 2.2.3), in the following referred to as NEB, allows finding saddle points and MEPs between known reactants and products. Interpolation between the initial and final states of the expected reaction gives a number of intermediate images along the reaction coordinate which are then being optimized by minimizing the acting forces. Each image is relaxed to the lowest energy possible while keeping equal spacing to adjacent images. Convergence of the band was assumed when the computed forces fell below $0.05\,\mathrm{eV\,Å^{-1}}$. Based on the complexity of the reaction path, also the chosen number of images within the NEB method plays an important role. Too few images may miss local reaction barriers whereas too many images will significantly increase the computational effort and can slow down the convergence. In order to obtain detailed information on the charge transfer within the HERs, here a rather high number of images was chosen to accurately illustrate the relationship between the reaction energetics and the corresponding electronic state.

The Bader charge analysis[134–138] allows dividing a molecule into non-overlapping atomic domains with well-defined boundaries (Bader volumes) by analyzing the electronic charge density. Henkelman et al. developed an efficient way to partition a charge density grid derived from plane-wave DFT calculations into Bader volumes.[137] Their implementation uses steepest ascent paths along the charge density gradient from grid point to grid point until a charge density minimum is reached. The charges encased in the resulting Bader volumes are a good approximation for the actual charge state of an atom.

As part of this thesis, the tool `baderVis` was developed. By combining the two aforementioned methods, it is possible to simulate and visualize the charge evolution along a reaction pathway, thus indicating potential relationships between occurring charge transfers and reaction energetics (Figure 3.5). On the basis of a converged NEB calculation with a

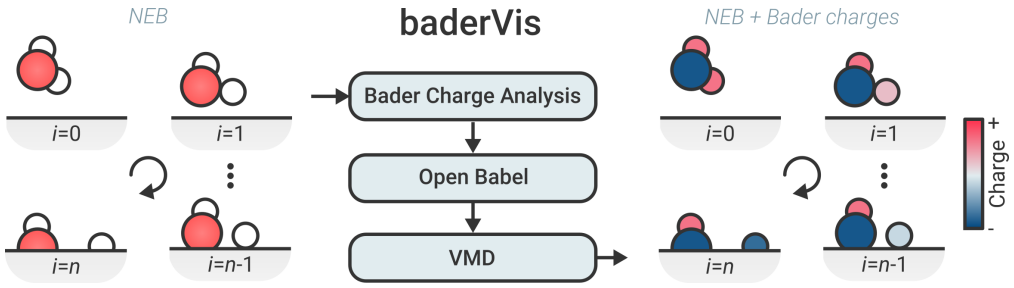


Figure 3.5: Charge evolution sampling with `baderVis`. For each image of a nudged elastic band (NEB) calculation, the individual charges of the atoms are determined in a Bader charge analysis. The atomic coordinates and charges of all images are then converted to PQR files, before they are finally processed using `Open Babel` for visualization via `VMD`.

number of images n, for each image a Bader charge analysis is performed. The output is converted via the `PQR` file format to the `mol2` file format using `Open Babel` [139], combining the atomic coordinates and obtained charges. For visualization with `VMD` [140], the `mol2` files serve as input for a `Tcl` script, which colors the visualized atoms according to their associated atomic charge. Stringing all adapted NEB images together provides a comprehensible insight into the charge evolution during the investigated reaction pathway.

3.2.3 Identifying Favorable Adsorption Sites

A prerequisite for modeling an electrochemical reaction using the NEB method are that products and reactants are known and an intuition for the potential reaction pathway is given. Hence, energetically favorable adsorption sites for the reactants during the HER have to be identified. Although the adsorption energetics of the considered reactants (H_2O, OH^- and H^+) have already been investigated in detail [107], the computations were replicated using the optB88-vdW functional as the impact of dispersion effects was neglected in the study of Williams et al. and other binding sites might become more favorable taking dispersion into account. As magnesium has a hexagonal close-packed (hcp) crystal structure, four potential adsorption sites have to be taken into account: the *top*, *bridge*, *hollow-fcc* and *hollow-hcp* position (Figure 3.6a). Accordingly, for each reactant, four different adsorption energies were calculated according to Equation (2.25). The results are summarized in Table 3.3.

It is noteworthy that an energetically favorable adsorption of H_2O is only given at top positions, whereas OH and H only stably adsorbed at hollow-fcc and hollow-hcp positions, respectively. As atomic relaxation of the reactants always ended up in either the top or hollow position, no adsorption energies are given for the bridge configuration. Adsorption of OH and H is energetically slightly more favorable at hollow-hcp and hollow-fcc positions, respectively. Comparing results for the exchange-correlation functionals PBE and optB88-vdW shows

Table 3.3: Adsorption energies of H_2O, OH and H at potential adsorption sites upon Mg(0001) with respect to the employed exchange-correlation functional. Starred values are derived from computations involving one water bilayer.

Functional	Adatom	top	hollow-fcc	hollow-hcp
PBE	H_2O_{ad}	-0.42 eV	—	—
	OH_{ad}	—	-5.32 eV	-5.35 eV
	H_{ad}	—	-0.05 eV	-0.04 eV
optB88-vdW	H_2O_{ad}	-0.50 / -0.59* eV	—	—
	OH_{ad}	—	-5.50 / -5.44* eV	-5.53 / -5.45* eV
	H_{ad}	—	-0.10 / -0.15* eV	-0.09 / -0.12* eV

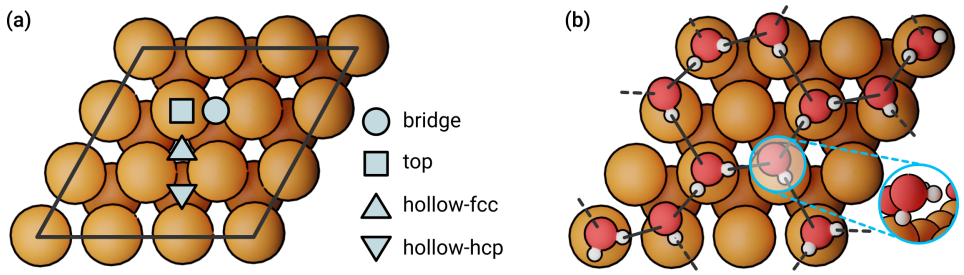


Figure 3.6: Magnesium surface adsorption. (a) Potential adsorption sites upon the Mg(0001) surface. (b) Water bilayer adsorption.

that the inclusion of dispersion effects can have a significant influence on the calculated adsorption energy. For H_2O, the calculated adsorption energy is 16% higher, for H even 50% higher, when including dispersion effects. For OH, the change in adsorption energies of 3% is negligibly small. It may me noted that favorable adsorption sites do not change when including dispersion effects in the calculation.

When including explicit water molecules in the surface model (Figure 3.6b), the resulting change in the electronic structure of the surface implies a general reduction of adsorption energies for OH, and an increase for H_2O and H. This energetic change has a direct impact on the hydrogen evolution reaction energies. In Table 3.4, the reaction energies for the Volmer and Tafel step are presented with respect to different adsorption site combinations and the inclusion of explicit water molecules. The reaction energy of the Heyrovský step is not further considered in this context, since it merely depends on the adsorption site of the hydrogen atom, which was already analyzed in Table 3.3. For both the Volmer and Tafel step, the adsorption of OH and H after water dissociation is energetically most favorable at hollow-hcp sites when including explicit water. As a reference, the Bader charges of the adsorbed species is given in Figure B.1. It is noteworthy that adsorption at equal adsorption

Table 3.4: Hydrogen evolution reaction energies for different OH and H adsorption sites upon Mg(0001). Results were obtained in vacuo and with respect to the inclusion of solvent (one water bilayer) in the simulation.

	Reaction	fcc+fcc	fcc+hcp	hcp+fcc	hcp+hcp
Vacuum	$H_2O \rightarrow OH_{ad}+H_{ad}$	-1.67 eV	-1.60 eV	-1.49 eV	-1.68 eV
	$H_{ad}+H_{ad} \rightarrow H_2$	0.25 eV	0.18 eV	0.18 eV	0.20 eV
Water bilayer	$H_2O \rightarrow OH_{ad}+H_{ad}$	−1.43 eV	−1.22 eV	−1.40 eV	−1.45 eV
	$H_{ad}+H_{ad} \rightarrow H_2$	0.28 eV	0.25 eV	0.25 eV	0.23 eV

sites seems to be energetically more favorable (fcc+fcc and hcp+hcp). However, since only configurations with minimal distance between the reactants (Tafel) and products (Volmer) were considered, repulsion effects for fcc+hcp or hcp+fcc can play a role due to the smaller distance between these sites.

3.2.4 Minimum Energy Pathways

In the beginning of the hydrogen evolution process, water dissociates in the **Volmer step** according to Equation (3.8), forming a hydroxide ion and a hydrogen atom which both adsorb on the magnesium surface at hollow-hcp sites (Table 3.3). Starting with a partially positive charge in the water molecule, the adsorption process leads to a polarity inversion of the hydrogen atom resulting in a hydride (H^-) on the surface. A water molecule in the H-down configuration is considered to dissociate in the NEB simulation as it is expected to preferentially take part in the reaction due to its small distance to the surface. In total, 24 intermediate images derived from a geometric interpolation between the assumed initial and final states sample the reaction pathway. During the NEB run, this band of images is slowly converged until it resembles the MEP, as illustrated in Figure 3.7.

At first, the downwards-pointing water molecule is distanced around 2.3 Å from the surface with an approximate bond angle of $\phi_{H-O-H} = 103°$ (step **a**). The deviation from the reference bond angle for ice I_h (109.5°)[141] indicates interactions between the Mg(0001) surface and hydrogen atoms within the adsorbed water bilayer. In contrast, the bond angle of adjacent water molecules is around $\phi_{H-O-H} = 109.1°$, which is in good agreement with the literature value.

The interactions at the magnesium-water interface can also be observed in the electronic structure, as vividly illustrated in Bader charge analyses for the intermediate images of the NEB simulation (Figure 3.7, bottom). Water molecules bound to top sites act here as electron acceptors, whereas those in H-down orientations act as electron donors.[142] This results

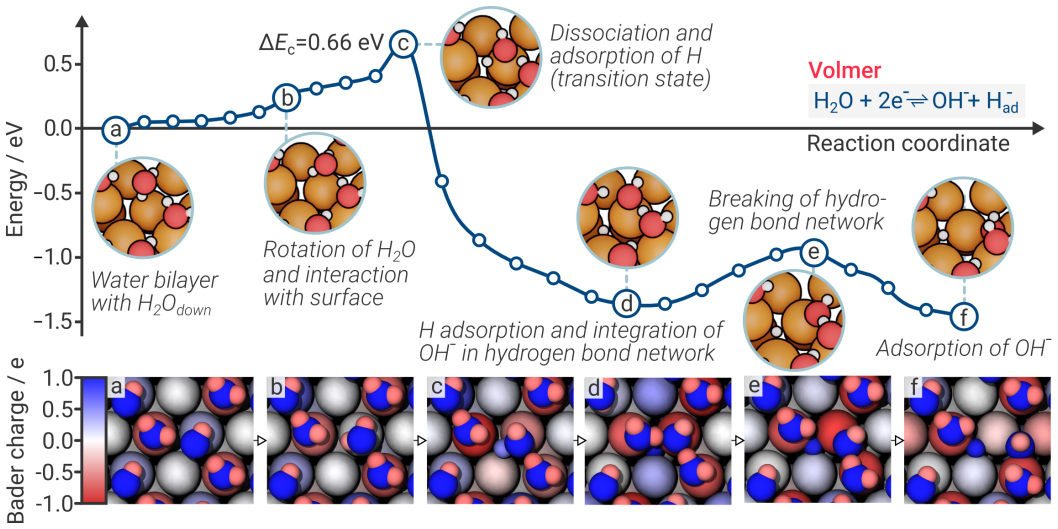


Figure 3.7: Minimum energy path (MEP) of the Volmer reaction as derived from density functional theory (DFT) calculations using the climbing image nudged elastic band (CI-NEB) method on Mg(0001). (top) The NEB computation consists of 24 images; the MEP is presented by a force-based cubic spline. Characteristic states are illustrated along with the energy barrier ΔE_c defined by the transition state c. (bottom) Top view of the images (a–f). Atoms are colored according to their respective Bader charge in elementary charges e. Adapted from Figure 2 in [95].

in alternating negative and positive surface charges of the Mg(0001) surface, displayed as blue and red spheres, respectively. The centers of the hexagonal rings forming the water bilayer are only slightly negatively polarized. Discrete values of the charge evolution in the Volmer (as well as Tafel and Heyrovský) step are presented in Figures B.2 – B.4. Following the MEP, the water molecule slightly rotates, positioning its downwards-pointing hydrogen atom above a hollow-hcp site in preparation of the dissociation (step **b**). In the process, the interaction between its oxygen atom and the magnesium surface significantly increases, indicated by considerable charge fluctuations that are apparent in the surface Mg atom as well as the water molecule in its vicinity. The point of dissociation (step **c**) defines the energy barrier of the reaction with $\Delta E_c = 0.66\,\text{eV}$, which agrees well with previously published computational studies [108]. The water bond angle increases to $\phi_{\text{H–O–H}} = 110.9°$ and electron densities from the surrounding Mg atoms shift to the adsorbing hydrogen atom, giving it a partially negative charge of $-0.56\,e$ that quickly shifts to roughly $-0.95\,e$ after completion of the water dissociation. This further increase in negative charge is caused by the remaining hydroxide ion that binds to a top site, causing an additional charge transfer from the interacting Mg surface atom to the adsorbed hydride in the process. As hydrogen bonds form between the previously adsorbed hydroxide ion and adjacent water molecules, it is integrated into the water bilayer and a local energy minimum in the reaction is reached at

$\Delta E_\mathrm{d} = -1.37\,\mathrm{eV}$ (step **d**). After overcoming a local energy barrier with $E_\mathrm{e} - E_\mathrm{d} = 0.4\,\mathrm{eV}$ (step **e**), the hydroxide ion adsorbs at a neighboring hollow-hcp site and the global energy minimum $\Delta E_\mathrm{f} = -1.45\,\mathrm{eV}$ of the Volmer step is attained (step **f**). The final charge of the adsorbed hydride was determined as $-0.96\,e$.

After the Volmer step the hydrogen evolution process can continue in two ways, either following the Tafel or Heyrovský pathway according to Equation (3.9) or Equation (3.10), respectively. During the **Tafel step**, two adsorbed hydrogen atoms recombine which subsequently causes gas evolution after desorption of the formed H_2. The adsorption of two hydrogen atoms in neighboring fcc sites (fcc-fcc configuration) yields almost isoenergetic ($\Delta E_\mathrm{ad} = 0.05\,\mathrm{eV}$) adsorption energies compared to the hcp-hcp configuration (see Table 3.4). As the hydrogen atom in the Volmer step is preferentially adsorbed at hcp sites, the hcp-hcp configuration is considered as the starting point for finding the MEP. However, as will be later shown, the adsorbed H atoms reorganize into an hcp-fcc configuration before forming molecular hydrogen, so the chosen starting point only plays a minor role. To put emphasis on the determination of the energy barrier of the hydrogen recombination on the Mg(0001) surface, influences of adsorbed hydroxide as well as the formerly reacted water molecule in the Volmer step have been neglected for the NEB simulation of the Tafel reaction pathway. An MEP including co-adsorbed hydroxide species and additional water bilayers is illustrated in Figure 3.10.

The potential reaction pathway is again sampled using 24 intermediate images between the initial and final states, whereas the latter is designated as released H_2 bound to the water bilayer. While the first 17 images illustrate the reaction pathway, subsequent images merely show reorientations of the H_2 molecule and were consequently neglected in the presented MEP (Figure 3.8).

Initially the two hydrogens are adsorbed at adjacent hollow-hcp sites, both exhibiting a charge of -0.95 to $-0.96\,e$ (step **a**). Concomitantly, three out of five adjacent Mg atoms located in the top layer are positively polarized. To initiate the Tafel reaction, one hydrogen atom diffuses from its hollow-hcp site across a Mg-Mg bridging site to the hollow-fcc site ($\Delta E_\mathrm{c} = 0.01\,\mathrm{eV}$) by overcoming a local energy barrier of $\Delta E_\mathrm{b} = 0.09\,\mathrm{eV}$. The distance between the investigated hydrogen atoms is reduced from around 2.98 Å to 2.36 Å (step **b** and **c**) by the diffusion process. Additionally, positive polarizations of adjacent Mg atoms are shifted according to the diffusion path. The hydrogen atom at the hollow-hcp site is pushed to the tetrahedral site, which results in an increase of its negative charge to $-1.02\,e$ due to the Mg atom below.

Along the MEP, the atomic distance of the neighboring hydrides further decreases to 0.95 Å at the bridging site between both atoms, causing a distinct rise in energy until the

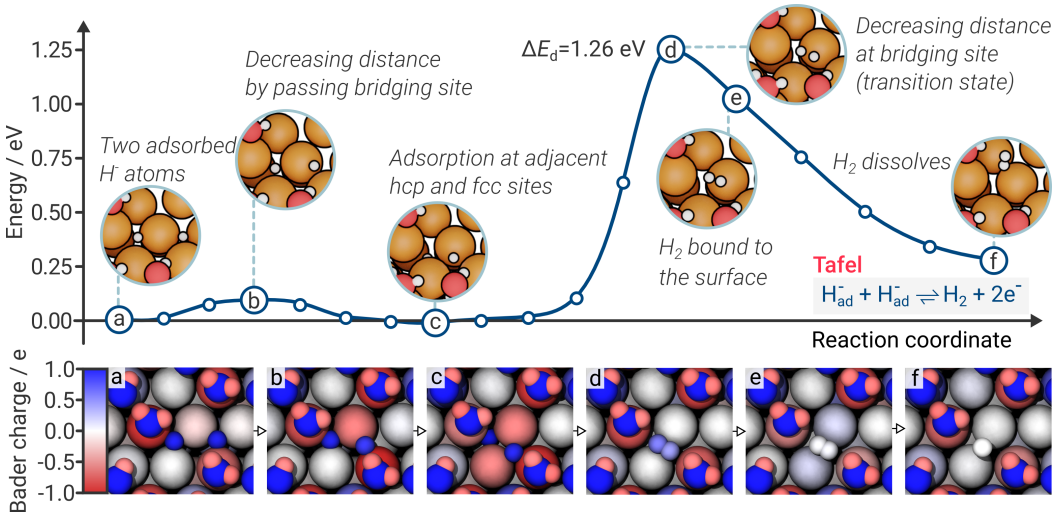


Figure 3.8: Minimum energy path (MEP) of the Tafel reaction as derived from density functional theory (DFT) calculations using the climbing image nudged elastic band (CI-NEB) method on Mg(0001). (top) The NEB computation consists of 24 images of which 16 are shown; the MEP is presented by a force-based cubic spline. Characteristic states are illustrated along with the energy barrier ΔE_d defined by the transition state d. (bottom) Top view of the images (a–f). Atoms are colored according to their respective Bader charge in elementary charges e. Adapted from Figure 3 in [95].

reaction barrier $\Delta E_\mathrm{d} = 1.26\,\mathrm{eV}$ is reached (step **d**) and H_2 is formed. This is accompanied by a simultaneous charge transfer between the dissociating hydrogen and the surface Mg atoms. The computed energy barrier for the Tafel reaction is in good agreement with the literature.[108]

As the H_2 molecule diffuses to the water bilayer (step **e** and **f**), the energy of the MEP decreases until a charge neutral H_2 molecule with a final atomic distance of 0.75 Å is attained. The final image of the MEP exhibits a relative energy of $\Delta E = 0.23\,\mathrm{eV}$ in comparison to the starting point (step **a**). The negative charge density of the hydrogen atoms is transferred to Mg surface atoms that act as binding sites of water molecules.

Characteristic of the **Heyrovský step** is that a water molecule directly reacts with an adsorbed hydrogen atom, according to Equation (3.10). Subsequently, H_2 is released and the remaining hydroxide ion forms a bond with the Mg(0001) surface. Here, an adsorption at a hollow-hcp site is energetically most favorable. The initially relaxed water bilayer is considered as a starting point for finding the MEP of this reaction. Similar to the investigation of the Tafel step, it is assumed that the hydride is adsorbed at a hollow-hcp site at the start of the Heyrovský reaction. Moreover, the corresponding hydroxide ion is neglected here to put emphasis on the isolated reaction between H_2O and H_ad. A NEB simulation using 24 intermediate images results in the final MEP of the Heyrovský reaction (Figure 3.9).

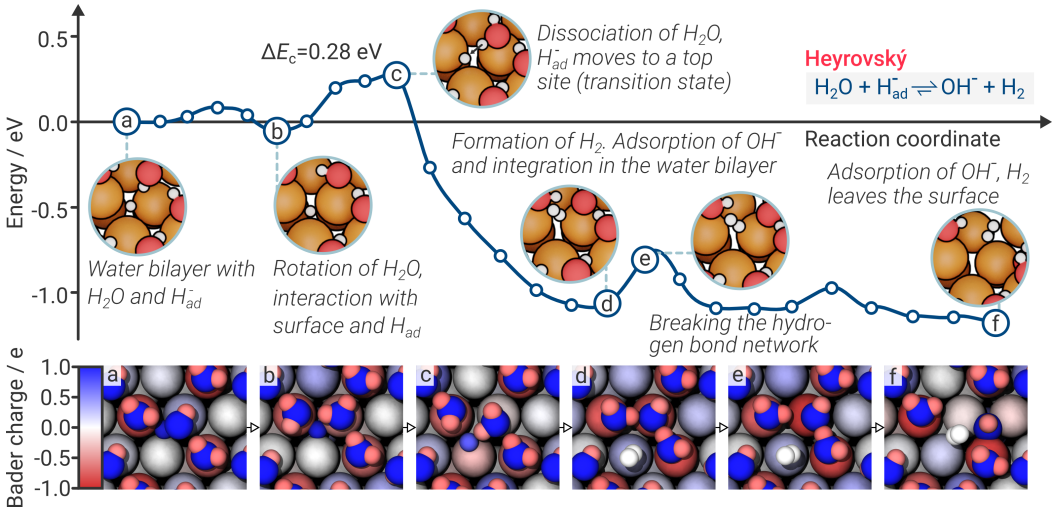


Figure 3.9: Minimum energy path (MEP) of the Heyrovský reaction as derived from density functional theory (DFT) calculations using the climbing image nudged elastic band (CI-NEB) method on Mg(0001). (top) The NEB computation consists of 24 images of which 16 are shown; the MEP is presented by a force-based cubic spline. Characteristic states are illustrated along with the energy barrier ΔE_d defined by the transition state d. (bottom) Top view of the images (a–f). Atoms are colored according to their respective Bader charge in elementary charges e. Adapted from Figure 3 in [95].

The initial distance between H_{ad} and H_{down} of the reacting water molecule is comparably low with roughly 1.6 Å (step **a**). However, before the hydrogen atom dissociates, the water molecule rotates, overcoming a small energy barrier of 0.08 eV. The reorientation results in a stronger interaction with the Mg surface and subsequently in an energetically more favorable restructuring of the water bilayer (step **b**). As the distance between H_{ad} and H_{down} further decreases to 1.3 Å, H_{ad} migrates over a bridging site to a near top site (step **c**). The intensified surface polarization facilitates the deprotonation of the water molecule and a proton transfer occurs at an energy barrier of $\Delta E_\mathrm{c} = 0.28\,\mathrm{eV}$. The considerably lower energy barrier for the Heyrovský step as compared to the Tafel step is a direct consequence of the observed atomic charges. Naturally, recombination of two H^- species implies a large energy barrier due to repulsive forces, whereas attractive forces between the adsorbed hydride and the partially positive H atom of the water molecule entail a significantly smaller energy barrier. As a consequence, applied anodic overpotentials promise to lower the energy barrier of the Tafel step, thus increasing the associated HER rate.[78]

Subsequently, the images of the MEP distinctively decrease in energy to $\Delta E_\mathrm{d} = -1.07\,\mathrm{eV}$ under formation of H_2. Simultaneously, the hydroxide ion is incorporated in the hydrogen bond network of the water bilayer (step **d**). In the last steps of the MEP, the hydroxide ion adopts an energetically slightly more favorable adsorption geometry ($\Delta E_\mathrm{f} = -1.17\,\mathrm{eV}$)

through migration to a nearby hollow-hcp site (step **e** and **f**) by overcoming a small energy barrier of $E_e - E_d = 0.27\,\text{eV}$.

3.2.5 Reaction Kinetics

All energy maxima of the MEP scans for the Volmer, Tafel, and Heyrovský reactions were confirmed as transition states by vibrational analyses, showing exactly one imaginary frequency, respectively. The corresponding normal modes reflect the change in geometry in going from reactants to products. Hence, the reaction rates k of the HER steps can be determined by inserting the corresponding activation energies E_{act} into the Arrhenius equation, defined in Equation (2.33). In the case of the Volmer step, a temperature of $T = 300\,\text{K}$, a calculated prefactor[51] of $\nu = 14.9\,\text{THz}$ (with a reference value of 10.0 THz)[44], and the Boltzmann factor k_B lead to a reaction rate $k = 110\,\text{s}^{-1}$. The exchange current density i_0 can be estimated[90] by using the following relationship:

$$i_0 = k\, e\, N / A, \tag{3.12}$$

where A / N is the surface area per Mg atom. Using the experimental value $i_{0,\text{HER}} = 4 \cdot 10^{-6}\,\text{A cm}^{-2}$, Equation (2.33) and (3.12) can be employed for validation of the computed energy barriers by calculating an estimated energy barrier E_{act}^{max}.[90,143] According to the previously calculated energy barriers, E_{act}^{max} as well as the reaction kinetics for all reaction steps were determined and summarized in Table 3.5.

Table 3.5: Activation energies E_{act} and estimated activation energies E_{act}^{max} for the Volmer, Tafel, and Heyrovský step on the Mg(0001) surface. Values for E_{act}^{max} are based on the experimental exchange current density $i_{0,\text{HER}} = 4 \cdot 10^{-6}\,\text{A cm}^{-2}$, the calculated prefactor ν and the reaction rate k. Adapted from Table 1 in [95].

	E_{act} / eV	ν / THz	k / s^{-1}	E_{act}^{max} / eV
Volmer	0.66	14.9	$1.1 \cdot 10^{2}$	0.90
Tafel	1.26	86.7	$4.8 \cdot 10^{-8}$	0.94
Heyrovský	0.28	0.9	$1.4 \cdot 10^{7}$	0.82

The estimated energy barrier E_{act}^{max} is to be understood as an upper limit for E_{act}. In other words, a reaction step is unlikely to be (mainly) responsible for the considered i_0, when its calculated energy barrier lies above E_{act}^{max}. Since a clean Mg(0001) has been assumed in the present model but in aqueous solution most of the Mg surface is covered with MgO and $Mg(OH)_2$[76]—thus providing fewer active surface sites directly associated with the HER—the estimated barrier is likely to be smaller than anticipated due to a reduced number of reactive

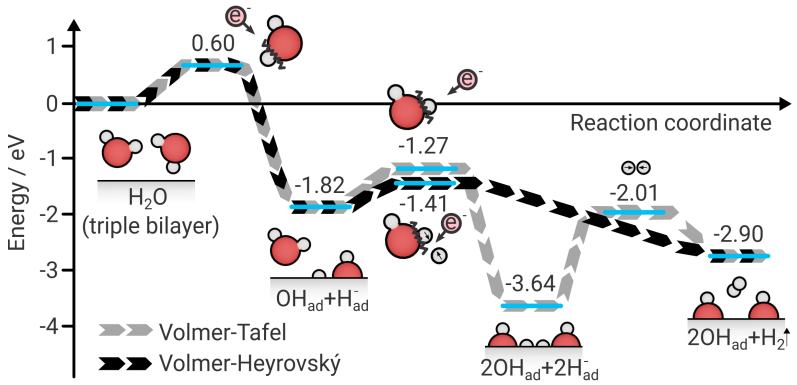


Figure 3.10: Energy barriers for the Volmer–Tafel and Volmer–Heyrovský pathways. Results were derived from a Mg–water interface model with three water bilayers adsorbed upon the clean Mg(0001) surface. Adapted from Figure 4 in [95].

sites.[90] In this respect, the obtained energy barrier for the Volmer reaction appears to match best the estimated value derived from experiments ($\Delta E = E_{\text{act}}^{\text{max}} - E_{\text{act}} = -0.24\,\text{eV}$). The energy barrier for the Tafel step is significantly larger than the upper limit $E_{\text{act}}^{\text{max}}$ ($\Delta E = 0.32\,\text{eV}$) and results in a relatively low reaction rate k in the order of $10^{-8}\,\text{s}^{-1}$. For the Heyrovský step, the calculated energy barrier is significantly lower ($\Delta E = -0.54\,\text{eV}$), resulting in an accordingly high reaction rate k in the order of $10^{7}\,\text{s}^{-1}$. Based on these results, it becomes evident that the Volmer step is the rate-limiting step for the HER at the Mg(0001) surface, rather than the Tafel step as has been previously assumed.[105] In fact, the Tafel step only plays a minor role at open circuit potential and does not contribute significantly to the hydrogen evolution.[95] Adsorbed hydrogen due to the Volmer step would react in high rates with interacting water molecules (Heyrovský step) leading to hydrogen evolution, whereas the contribution of recombining hydrogen atoms (Tafel step) is rather small.

After the isolated investigation of the dominant reaction mechanisms in the hydrogen evolution process, further extensions to the model are supposed to quantify the impact of the simplifications made beforehand. Hence, two additional explicit water bilayers as well as intermediate reaction products (consisting of adsorbed hydroxide ions) are included in order to obtain a comprehensive understanding of the energy barriers in the Volmer-Tafel and Volmer-Heyrovský pathways. The results are illustrated in Figure 3.10.

Although additional water bilayers as well as the inclusion of intermediate reaction products have an impact on the occurring energy barriers, the outcome of the Volmer-Heyrovský pathway being the energetically most favorable reaction pathway for the HER on Mg remains unaltered. While this confirms results by previous studies[40,43,78,79], the model applied here provides valuable insights into the charge migration of the occurring reaction mechanisms.

The shown results indicate how important a thorough investigation of the occurring charge transfers during surface reaction is. The `baderVis` algorithm, developed as part of this thesis, constitutes a valuable tool in these investigations to better understand potential reaction barriers. For instance, the difference in energy barriers between the Tafel and Heyrovský step becomes only obvious when considering the associated charge transfers. By analyzing the inherent transition states, further insights into the reaction kinetics are gained that allow for a direct comparison to experimental results. In this way, theoretical results can be validated and occurring reaction pathways identified.

3.3 Magnesium Dissolution

Analogously to the cathodic HER, the anodic reaction in aqueous Mg corrosion is defined by magnesium dissolution as described in Equation (3.2). In that process, Mg dissociates from the surface, undergoing oxidation to Mg^{2+}. A study by Petty in 1954[99] proposed Mg could also be dissolved as monovalent Mg^{+}, giving rise to a popular theory that hydrogen evolution on dissolving Mg surfaces arises from a secondary reaction between Mg^{+} and water, where additional hydrogen gas is formed. Although there is no persuasive experimental evidence, this theory is largely promoted in particular as an explanation for the increased hydrogen evolution rates during anodic dissolution (anomalous hydrogen evolution).[76,86–88,100] Aside from other popular theories trying to explain this phenomenon[76,77,87,88], recent experimental findings suggest an enhanced catalytic activity of magnesium primarily associated with regions dominated by the anodic dissolution reaction.[84,143–145] These findings are supported by a theoretical study, where hydrogen evolution rates increase with simultaneous dissolution of Mg for increasing potentials.[79] Taylor gives a possible explanation for this behavior by suggesting that parts of the hydroxide (passivation) layers are removed anodically through the dissolution of the Mg atom to which the hydroxide ion is adsorbed. In this way, more clean surface area would be revealed, thus further promoting water dissociation and hydrogen evolution.[146]

Based on the reported experimental and theoretical findings, two potential reaction pathways are revealed, where the dissolving magnesium ion is either exclusively assisted by water or co-adsorbed hydroxide further facilitates the dissolution.[78] Modeling co-adsorbed hydroxide species for a dissolving Mg ion with the NEB method, however, turns out to be more complicated, since the initial reaction pathway has to be sampled between all atomic positions individually. In other words, for a Mg atom dissolving together with co-adsorbed OH, the hydroxide ions in the final image could, e.g., correspond to the hydroxide ions in the initial image. It can happen though that the atoms involved in the coordination of the Mg ion

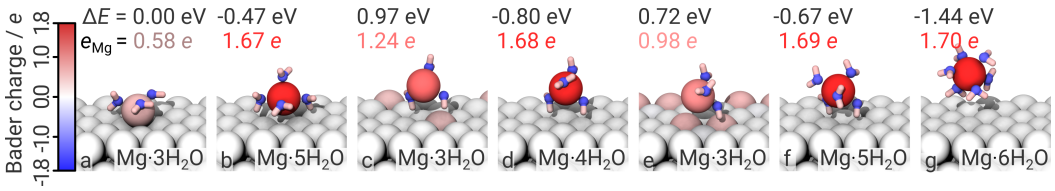


Figure 3.11: Characteristic steps in the reaction pathway of a dissolving magnesium ion. Only the first solvation shell of the Mg ion is illustrated. Values for the relative energies ΔE and Bader charges e_{Mg} are depicted in the upper left corner of each image. The atoms are colored according to their respective Bader charges. Adapted from Figure 5 in [95].

change due to proton jumping along the reaction path (Grotthuss mechanism).[147] Due to limitations of the underlying NEB model, the focus here shall lie on the computationally simpler water-assisted reaction pathway.

A NEB calculation has been set up to gain a deeper understanding of the reaction kinetics as well as the charge evolution of a single dissolving Mg ion. The computational model used so far in this work has been further extended to include an additional Mg atom adsorbed upon the Mg(0001) surface, thus representing a more realistic, rougher Mg surface. Additionally, a total number of four water bilayers have been included to ensure sufficient water molecules for full solvation of the dissolving Mg ion. The described extended model acts as initial state of the NEB calculation. For the final state, the adsorbed Mg atom is moved away from the surface into solution before it is relaxed, resulting in a solvation shell of six water molecules in an octahedral arrangement, agreeing well with the literature.[148] The surface model is illustrated in Figure B.7. As the water structure shifts during dissolution, the NEB path is dominated by relatively large energy fluctuations, hampering its convergence. Hence, for a higher resolution of the reaction pathway the number of intermediate images was further increased to 48. In order to gain a deeper understanding of the dissolution pathway and the involved charge evolution, the relaxed MEP and corresponding Bader charges of the dissolving Mg ion are presented in Figure B.6. Characteristic images are visualized in Figure 3.11 in more detail.

During dissolution, the Mg ion passes a number of favorable coordination states which entail changes in the energetic state and electronic structure. While adsorbed upon the Mg(0001) surface, the Mg atom is coordinated by three water molecules with a Bader charge of 0.58 e (step **a**). After overcoming an initial energy barrier of $\Delta E_{\mathrm{b}} = 0.44$ eV, the Mg atom is ionized (step **b**). More water molecules are bound to the Mg ion, increasing its coordination number to five, its Bader charge to 1.67 e, and its distance from the surface by around 1.55 Å to 3.99 Å. During the remaining observed dissolution process (steps **c**-**g**), the coordination number of the Mg ion varies between three and six water molecules. For a

coordination with three water molecules (step **c** and **e**), the Mg ion shows interaction with the surface and actually exhibits a Bader charge of around 1 e indicating an existing Mg^+ species. However, investigating the MEP reveals that $Mg{\cdot}3H_2O$ complexes can only be found within high energy barriers (0.97 eV) and are thus highly unstable. Hence, they can not serve as explanation for an anomalous hydrogen evolution mechanism that depends on a stable monovalent Mg species. For higher coordination number (four to six water molecules) the Bader charge of the Mg ion becomes stable at a value of around 1.7 e, confirming Mg^{2+} as the energetically most favorable charge state during Mg dissolution. After formation of the $Mg^{2+}{\cdot}\,6H_2O$ complex, the relative energy reaches a minimum at $\Delta E = -1.44$ eV.

Again, an estimate for the energy barrier of Mg dissolution can be derived from experimental results according to Equation (2.33) and (3.12). Assuming a prefactor $\nu = 1 \cdot 10^{13}$ Hz and an anodic exchange current density $i_{0,\mathrm{Mg}} = 1.6 \cdot 10^{-5}\,\mathrm{A\,cm^{-2}}$, the reaction barrier can be estimated to 0.85 eV.[44,149] Comparison with the calculated value of 0.97 eV indicates that the dissolution reaction is not solely water-assisted but rather facilitated by co-adsorbed species, for instance hydroxide ions, as already proposed in the literature.[78,79,146] DFT calculations for the solvated Mg^{2+} ion and $Mg(OH)_2$ molecule provide further evidence for this indication (Figure B.5). The obtained results vividly illustrate the energy barriers and charge evolution of a Mg ion undergoing dissolution and strongly contradict the claim of a energetically stable Mg^+ species which promotes water reduction.

As the employed methods and model system naturally imply errors, studies employing dynamic simulation methods are necessary to gain further insights. Although the combination of the NEB method and Bader charge analysis results in a deep insight into the minimum energy path of a dissolving Mg ion, the continuous movements and entropic effects in the water structure require a high resolution of the reaction pathway, resulting in an immense computational cost. Other methods, such as metadynamics-based ab initio molecular dynamics simulations, are expected to be more appropriate for modeling aqueous Mg corrosion and will be used below to address free energy barriers and transition dynamics more comprehensively, as well as to adequately account for entropic effects of water at a finite temperature.

In an attempt to more accurately describe the degrading Mg surface, the previous model was adapted to include edge defects and iron impurities. The adapted model contained six complete layers plus a seventh defective layer of Mg. A total of 220 water molecules were included in the simulation of which 16 were assumed to be dissociated and adsorbed upon the incomplete layer, corresponding to a surface coverage of $\theta = 1/3$.[43] While the resulting 16 hydroxide ions were adsorbed directly upon the surface, the remaining hydrogen atoms were assumed to adsorb sub-surface. One Mg atom in the first layer of the exposed

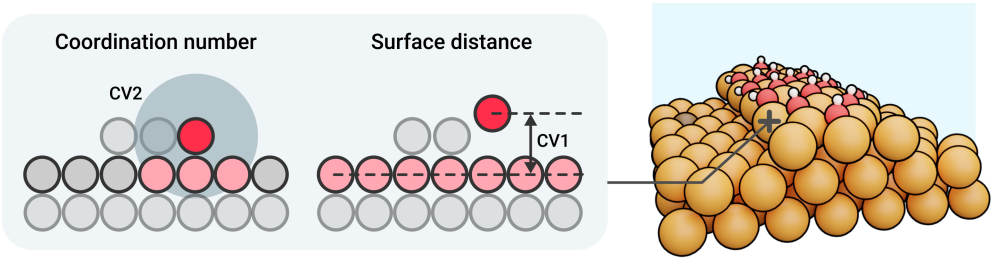


Figure 3.12: Collective variables for sampling Mg dissolution. The dissolution pathway of an edge-site Mg atom is simulated using metadynamics. The collective variables (CVs) are defined as the surface distance and the coordination number with the topmost complete substrate layer. For sake of clarity, the surface model is illustrated without showing explicit water molecules (indicated as blue background).

surface was replaced by an Fe atom. Employing the PBE exchange–correlation functional as implemented in the `CP2K/Quickstep` package [150], the system was first equilibrated for 1.4 ps using BOMD, then for 38 ps using second-generation CPMD. All related parameters are presented in Appendix A.2. The equilibrated model mimics the degrading, hydroxylated Mg surface where dissolved Mg atoms exposed unhydroxylated regions (Figure 3.12, right).

For sampling the Mg dissolution process, two collective variables (CVs) were defined (Figure 3.12, left): the height of the dissolving Mg atom above (CV1) and its coordination number (CN) with the Mg atoms within the topmost complete layer of the substrate (CV2). Based on a study by Liu et al. [151], this combination of CVs is expected to discriminate between preferential dissolution of individual Mg atoms without making assumptions about the reaction mechanism. Calculation of the CN is implemented in `PLUMED` as the number of contacts between two groups of atoms, given as

$$\sum_{i \in A} \sum_{j \in B} s_{ij}, \tag{3.13}$$

where s_{ij} is 1 if there is contact between atoms i and j, and zero otherwise. For practical reasons, s_{ij} is replaced by a switching function to ensure that the calculated CV has continuous derivatives. The default switching function is given as

$$s_{ij} = \frac{1 - (\mathbf{r}_{ij}/r_0)^6}{1 - (\mathbf{r}_{ij}/r_0)^{12}}, \tag{3.14}$$

where $\mathbf{r}$ is the interatomic distance between the ith and jth atoms. For the Mg-Mg bonds between the dissolving Mg atom and the topmost surface layer, r_0 was taken as 3.4 Å.

Starting from the equilibrated system, well-tempered metadynamics was performed biasing the defined variables, thus aiming at determining the lowest free energy pathway for the

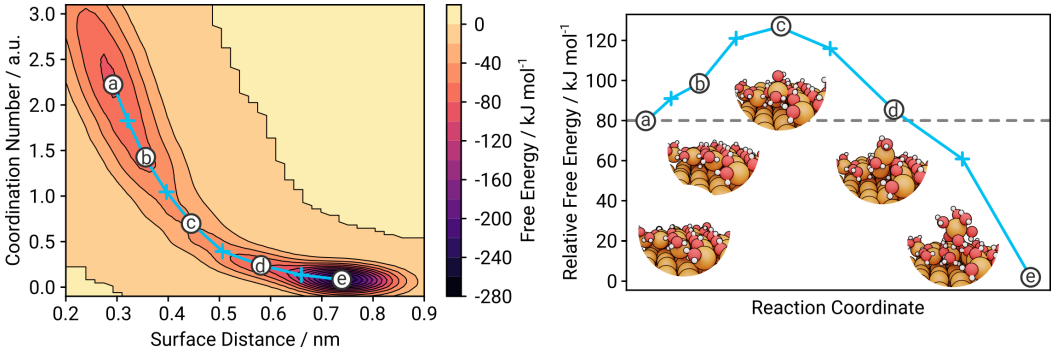


Figure 3.13: Free energy surface and barrier for Mg dissolution. (left) The calculated minimum energy pathway (MEP) is mapped onto the free energy surface, defined by both collective variables (CVs). (right) Characteristic images of the MEP are illustrated. Only water molecules taking part in the solvation of the departing Mg ion are shown. (a) Mg ion is part of edge site. (b) Surrounding water molecules lift the Mg ion out of the Mg matrix. (c) Transition state; water molecules adsorb below the dissolving Mg ion. Co-adsorbed hydroxide forms a bridge between surface and dissolving Mg. (d) Coordination number of the dissolving Mg ion increases to five. (e) Completed dissolution of the Mg ion. Final coordination number of six.

dissolution of Mg. The initial Gaussian height was taken as 3.0 kJ/mol, along with a bias factor of 10. The Gaussian widths were set to 0.05 nm and 0.1 for CV1 and CV2, respectively. A Gaussian bias potential was deposited every 100 MD steps.

All Gaussian bias potentials deposited during the metadynamics simulation were summed up, yielding the free energy surface with respect to the defined CVs (Figure 3.13a). Within the free energy surface, the MEP for Mg dissolution can be obtained using the NEB method as implemented in the `metadynminer` package[152]. The reaction barrier and pathway are illustrated, along with characteristic configurations, in Figure 3.13. Similar to the previous NEB approach, the MEP as derived from metadynamics shows that the dissolving Mg atom passes multiple coordination states. Moreover, it becomes clear that co-adsorbed hydroxide actively contributes to the dissolution mechanism. The calculated energy barrier for the Mg dissolution process is 41.57 kJ/mol, which corresponds to around 0.43 eV, and is thus considerably smaller than the reaction barrier as determined by the NEB method (0.97 eV), and well below the estimated upper limit of 0.85 eV. However, the lower energy barrier can also be attributed, at least in part, to the inclusion of an Fe impurity in the metadynamics simulation, which has not been included in the NEB calculations. For future work, the influence of noble impurities should be further investigated, with respect to different types, amounts and locations.

Successful Mg corrosion engineering should aim at controlling the parameters responsible for the reaction rates of the anodic and cathodic processes, i.e., Mg dissolution and hydrogen evolution. For example, by decreasing the active sites at the exposed surface, the exchange

current density can be decreased. When aiming at accelerating the corrosion, introducing catalysts for the rate-limiting Volmer step can be effective. However, applied potential can have a significant impact on the observed reaction kinetics. Particularity anomalous hydrogen evolution, as observed for Mg, is still not fully understood and aggravates corrosion protection and control. Although parts of this thesis dealt with the hypothesis of a monovalent Mg species, deeper insights into this phenomenon would require applied constant potentials to be explicitly included in the computational models. Fortunately, an increasing number of approaches allow to do so with decreasing complexity and computational effort.[43,78,91]

General means of corrosion protection/control can also be realized by, e.g., stabilizing the passive hydroxide layer in aqueous environments. Additionally, alloying, doping or co-adsorption with other atomic or molecular species can have a significant impact on the corrosion reaction barriers. A comprehensive overview of Mg corrosion control and inhibition strategies is given in the following chapter.

Chapter 4

Gaining Control Over Magnesium Degradation

> There is nothing that living things do that cannot be understood from the point of view that they are made of atoms acting according to the laws of physics.
>
> — Richard P. Feynman

Although the efficacy of a corrosion control strategy can be assessed in experiments, the exact underlying corrosion control mechanisms remain often unknown. Atomistic simulations, such as density functional theory calculations or molecular dynamics, can be valuable tools to obtain a more detailed understanding of these mechanisms, thus revealing potential aspects that need to be addressed in corrosion protection schemes. However, as corrosion control and degradation mechanisms can be exceedingly complicated, finding a good trade-off between model complexity and accuracy is not trivial.

For an effective magnesium corrosion engineering, gaining deeper insights into the governing corrosion mechanisms is one side of the medal. The other one is constituting potent strategies to inhibit or control the surface reactivity characteristics of the material, which is key to unlocking the full potential of Mg. Different fields of application may require specific corrosion modulation strategies, depending on the target implementation and the given service environment. For example, surface coatings that attempt to fully inhibit magnesium degradation are unsuitable for Mg-air battery implementations, where the controlled dissolution of Mg is a key performance factor. Clearly, compatible corrosion modulating strategies are needed for these applications. Fortunately, there are numerous concepts to control or prevent the degradation, which shall be addressed in the following section.

During this chapter, approaches to model corrosion inhibition using atomistic simulations are further discussed. After a brief overview of corrosion control strategies for Mg and

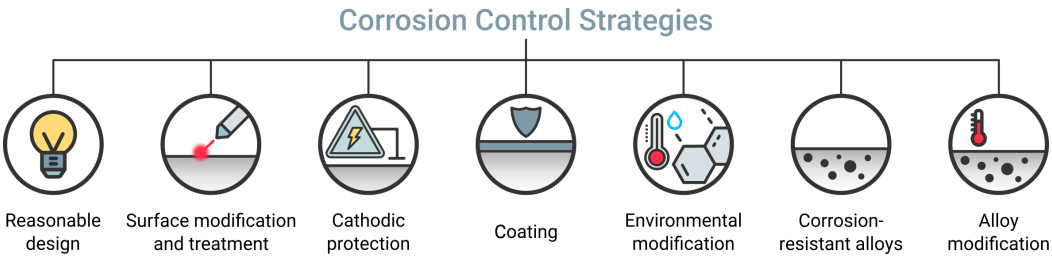


Figure 4.1: Strategies for corrosion control and prevention.[155] Depending on the target application, various effective approaches exist within the shown categories to control the surface activity of magnesium-based materials.

Mg-based materials, two modeling approaches are further elaborated. Parts of the work and figures presented hereafter have already been published in

[153] L. Fockaert, **T. Würger**, R. Unbehau, B. Boelen, R. H. Meißner, S. V. Lamaka, M. L. Zheludkevich, H. Terryn, & J. Mol. ATR-FTIR in Kretschmann configuration integrated with electrochemical cell as in situ interfacial sensitive tool to study corrosion inhibitors for magnesium substrates. *Electrochimica Acta*, 345, 136166 (2020).

[154] X. Li, **T. Würger**, C. Feiler, R. H. Meißner, M. Serdechnova, C. Blawert, & M. L. Zheludkevich. Atomistic Insight into the Hydration States of Layered Double Hydroxides. *ACS Omega*, accepted (2022).

4.1 Strategies for Corrosion Control: An Overview

A comprehensive understanding of the fundamental Mg degradation mechanisms (cf. Chapter 3) is a key aspect in the development of effective corrosion protection and control strategies. Corrosion engineering concepts would need to address the unstable passivation layer and high surface reactivity of Mg that lead to increased hydrogen evolution and Mg dissolution in aqueous environments. Several general approaches exist to mitigate corrosion of magnesium-based materials, not fundamentally differing from protections schemes for other conventional metals.[155] However, due to the unique corrosion properties of Mg, successful implementation requires careful design tailored to the application at hand. Song listed and described seven typical approaches (Figure 4.1) that shall be further discussed in the following.[155]

First of all, choosing a **reasonable design** can help to prevent Mg corrosion in the first place. Geometric changes to a Mg component, for example, can reduce the occurrence of moisture accumulations, and thus associated corrosion damage. The implementation of

magnesium-based components should be as simple, cost-efficient and effective as possible and should always consider the corrosion-related characteristics of the service environment.

By applying a **cathodic protection** potential that is more negative than the equilibrium potential of the Mg component, the corrosion rate can be drastically reduced. However, the highly negative electrode potential of Mg (-2.37 V vs. normal hydrogen electrode) complicates finding an engineering material with a more negative potential, which can act as sacrificial anode. Imposing cathodic currents could theoretically circumvent this issue, although high cathodic current densities would be required, which would then lead to intensive hydrogen evolution and low current efficiency. Still, between corrosion and equilibrium potential the hydrogen evolution rate reaches a minimum and Mg dissolution may nearly vanish, such that limited cathodic protection is practically possible.

The arguably most famous and reliable approach to corrosion mitigation is the **development of corrosion resistant alloys**. As the part of a Mg alloy preferentially suffering to corrosion is its matrix[i], a key approach is to make the matrix phase passive or inert. Alloying with inert elements theoretically reduces oxidation and dissolution, however, due to the high chemical activity of Mg, high amounts of alloying elements are required that can significantly influence the mechanical properties of the alloy. Hence, increasing the passivity of the Mg alloy might be a more practical approach. As all known highly passive elements (e.g., Ti, Al, Cr and Ni) have confined solid solubility in the Mg matrix phase, none of them can effectively passivate Mg in a corrosive environment. Nevertheless, creating passive alloys is generally possible using highly passive alloying elements that form super-saturated solid solutions with Mg.

Alloy modification allows to further adapt the corrosion resistance of an existing Mg alloy. Carefully designed heat-treatments, for example, were shown to influence the corrosion resistance by affecting the grain boundaries, without significantly changing the composition and phase constituent.[156,157] A negative side effect of this approach, however, can be impairment of the mechanical properties.

Surface modification and treatment is a favorable alternative, when increasing the corrosion resistance of a Mg component goes at the expense of (potentially vital) mechanical performance. By modifying the surface layer of a Mg alloy but leaving the bulk alloy unchanged, anodic and/or cathodic polarization resistance can be enhanced, anodic and cathodic phases can be isolated and electrochemical differences between them can be reduced or eliminated. Popular techniques for surface treatment are, e.g., surface machining and cleaning, laser treatment, hot diffusion, surface conversion and anodizing.[11]

[i] *Corrosionpedia*: *Matrix Phase*. "The matrix phase refers to a continuous phase in which a composite or two-phase alloy is formed in a microstructure of metals and completely surrounds the other (or dispersed) phase so as to ensure the stability of the dispersion phase."

In contrast to modifying or treating the Mg alloy surface, a protective **coating** can be applied as independent layer.[158,159] On the one hand, a coating separates the Mg alloy surface from the corrosive environment, thus preventing the surface to take part in corrosive reactions. On the other hand, the polarization resistance of the alloy substrate is considerably increased, dramatically deferring its degradation. Several methods exist to apply coatings to Mg alloys, including plating techniques, plasma electrolytic oxidation (PEO), chemical and physical vapor deposition (C/PVD), hot and cold spray as well as magnetic sputtering. Coatings can be additionally loaded with corrosion inhibitors that actively protect the surface from degradation.[159–161] This can be realized using, for example, layered double hydroxides (LDHs) with intercalated compounds that yield a corrosion inhibiting effect.[162] Through an anion exchange between Cl^- and the intercalated corrosion inhibitor, the corrosive chloride ion is bound in the intergallery of the LDH system and the corrosion mitigating compound is released. Li et al. conducted DFT calculations and MD simulations to investigate the anion-exchange capacity of four LDH systems with respect to the intercalated water content.[154] The simulations suggest that a high self-diffusion coefficient, which is linked to the mobility of the intercalated anions, is vital for their release via the anion-exchange mechanism and thus for successful corrosion protection.

Depending on the target application, corrosion inhibitors can also be directly introduced into the service environment to alter the corrosion rate of a Mg alloy. **Environmental modification** by introducing such magnesium dissolution modulators is an effective approach to adapt the corrosion rate with respect to a given target application. Atomistic modeling of the associated corrosion inhibition or acceleration mechanisms enables calculation of properties that may show a connection to the corrosion inhibition efficiency (cf. Section 4.2). For immersion, other strategies to control Mg degradation are increasing the pH value to stabilize the natural passivation layer or reducing the concentration of aggressive species, such as chloride. Under atmospheric conditions, also more obvious changes to the environment such as sustaining dry air can have an alleviating effect on corrosion.[163] However, modifying an open environment may not always be achievable and environmental impact as well as health hazards are major concerns. Due to their flexibility with respect to the employment in numerous applications, magnesium dissolution modulators shall be further discussed in the following.

4.2 Modeling Magnesium Dissolution Modulators

Depending on the application, environmental and health aspects may also play a large role in closed environments. Naturally, corrosion inhibitors used in active coatings for magnesium-

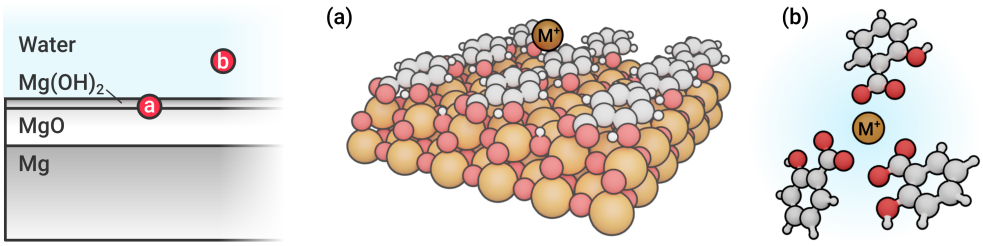


Figure 4.2: Magnesium dissolution modulator mechanisms to prevent iron redeposition. On the left side, the sphere of action of the respective mechanism is illustrated. (a) Surface stabilization and passivation. (b) Entrapment of noble impurities after release from the bulk material.

based medical implants are required to be non-toxic and biocompatible. Furthermore, due to the imminent ban of highly effective but toxic chromates as part of the REACH regulation [164], there is a particularly strong demand for green corrosion inhibitors in industry. [16,165,166] The search for green and cost-efficient corrosion inhibitors, however, is intricate and a new universal corrosion inhibitor matching the performance of chromate will most likely not be found. Computational approaches that facilitate this search are further discussed in Chapter 5.

In a comprehensive screening of Mg corrosion inhibitors, Lamaka et al. determined the corrosion inhibition efficiencies of over 150 iron complexing agents. [14] As described in Section 3.1, noble impurities (such as Fe) are released during Mg dissolution and can redeposit upon the alloy surface, thus enlarging the cathodic area and increasing the corrosion rate. Molecules that form stable complexes with these corrosive species or block them from redepositing upon the Mg alloy surface are a successful approach to mitigate corrosion. [13,14,85] Both mechanisms are illustrated in Figure 4.2. It is noteworthy that molecules building stable complexes with Mg^{2+} ions can also break the protective $Mg(OH)_2$ layer, increasing the Mg dissolution rate and thus accelerating the corrosion.

When using atomistic modeling techniques to investigate the above-mentioned inhibition mechanisms, there are two major quantities of interest, which shall be considered in this section. With respect to the complexation mechanism, i.e., the entrapment of Mg^{2+} ions or corrosive species such as iron particles, the complex formation energy of the employed Mg dissolution modulator can be a measure for the stability of the formed complex. For Mg dissolution modulators that inhibit or accelerate corrosion by interacting with the degrading surface, calculating the adsorption energy can give further insights. In the following, computational approaches based on DFT to determine these quantities are further discussed. Furthermore, potential correlations with the corrosion inhibition efficiency are investigated.

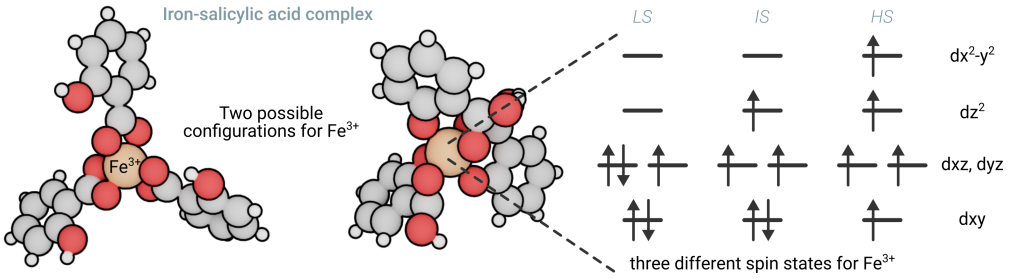


Figure 4.3: Complexation of metal ions. For salicylic acid and Fe^{3+}, 15 potential configurations have to be considered to find the most favorable complex formation energy. On the one hand, five different conformers/isomers of the complex exist, on the other hand, there are three different spin states for Fe^{3+} (LS: low spin; IS: intermediate spin; HS: high spin).

4.2.1 Complexation of Metal Ions

As the complex formation energy of a dissolution modulator complex with respect to a bound metal ion can be simply understood as a binding energy, the calculation of such energy can be performed according to Equation (2.25). For instance, considering the complexation of a solute Fe^{3+} ion by salicylic acid, the complex formation process can be described via

$$Fe^{3+} + 3\ C_7H_5O_3{}^- \rightleftharpoons Fe(C_7H_5O_3)_3. \tag{4.1}$$

Calculating the total energies of the three separate systems—iron(III), salicylic acid and the coordination complex—using DFT, the energy released through complex formation can be obtained. However, as chemical compounds may have more than one active site, the exact coordination complex is not always obvious. Figure 4.3 (left) illustrates two of five potential complexes between salicylic acid and iron(III).

Apart from identifying all possible geometrical configurations of the coordination complex, the spin states of the central ion's d electrons further complicate matters. When dealing with iron(III), these spin states vary between high-, low- and intermediate spin configurations (Figure 4.3, right). In the case of Fe^{3+}, which has five d electrons, thus three calculations per conformer have to be performed to find the correct ground state energy. Particularly for a larger number of compounds with varying numbers of functional groups, automating this process becomes exceedingly difficult and computationally expensive. Furthermore, solvation and temperature effects might have a significant impact on the complex formation energetics. Molecular dynamics simulations and enhanced sampling methods, such as metadynamics, constitute potential solutions to this challenge.

4.2.2 Surface Adsorption

Apart from protecting the surface from redepositing noble impurities, corrosion inhibitors can also stabilize the natural passivation layer of Mg, thus mitigating degradation. Fockaert et al. investigated chemisorption mechanisms of organic inhibitors as well as their corrosion inhibition efficiency towards magnesium-based substrates.[153] Four carboxylic compounds, namely 2,5-pyridinedicarboxylic acid (PDC), 3-methylsalicylic acid (MSA), sodium salicylate (SS) and fumaric acid (FA), were selected based on their promising corrosion inhibiting capabilities. All compounds were shown to chemisorb at the MgO/$Mg(OH)_2$ surface by bond formation between the substrate and carboxylate moieties. Orientation analysis using polarized infrared light revealed that aliphatic carboxylate compounds bond perpendicularly to the magnesium surface, whereas carboxylate bonds of aromatic compounds were oriented in plane with the magnesium surface. The orientation difference can be associated with the involvement of π bonds in the adsorption of aromatic carboxylates upon the MgO/$Mg(OH)_2$ interface.

To obtain a more comprehensive understanding of the adsorption behavior of the investigated inhibitor molecules at the MgO/$Mg(OH)_2$ interface, DFT calculations were performed using `VASP`[113–116] with the projector augmented wave method[117,118], using the exchange-correlation functional optB88-vdW[27,120–124] (cf. Section 2.1). All computations were performed using a $4 \times 4 \times 1$ Γ-centered grid of k-points.[127] The plane-wave expansion was limited by a cutoff energy of 520 eV. During the relaxation process, the atomic positions were allowed to relax until the Hellmann-Feynman forces were less than 5 meV/Å. For the relaxation process, the systems were pre-converged using the conjugate gradient algorithm; subsequently a quasi-Newton algorithm was employed, which is efficient for atomic coordinates close to the minimum. In a third computation step, the structure was investigated statically to obtain more accurate total energies via the tetrahedron method with Blöchl corrections.[128]

The MgO/$Mg(OH)_2$ interface was modeled using a slab of five MgO(100) layers in a 2×2 surface cell. With respect to the number of carboxyl groups within the adsorbed inhibitor molecules, hydrogen atoms were co-adsorbed accordingly on the oxygen sites of the MgO(100) surface, resulting in a partially hydroxylated surface that represents the beginning of the passivation layer formation and leaving the deprotonated inhibitor molecule to adsorb. Atoms in the two bottom layers were kept fixed in their bulk-like positions, whereas the remaining layers were free to relax, thus being able to respond to occurring forces due to surface interactions. A vacuum region of around 20 Å was added above the surface slab in order to avoid interaction between periodic images. A dipole correction was applied

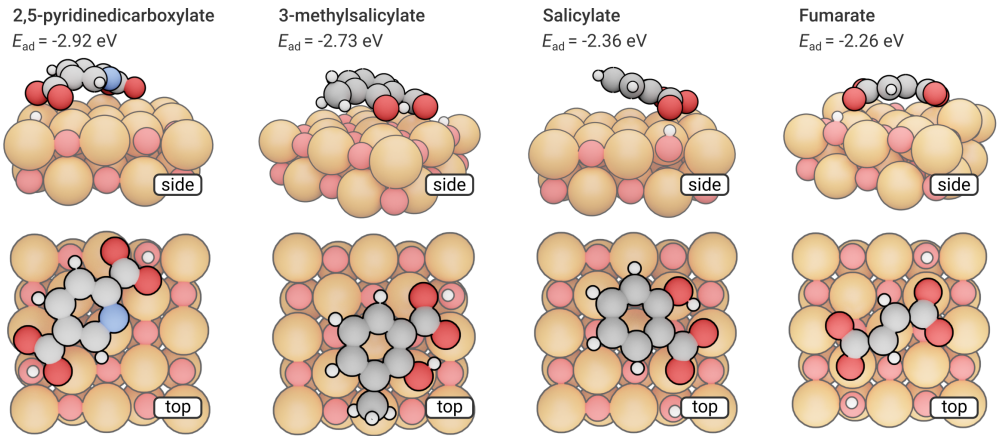


Figure 4.4: Surface adsorption behavior of four small organic molecules. The DFT-calculated adsorption energies of the investigated inhibitor molecules upon the partially hydroxylated MgO(100) surface are given for a coverage $\theta = 0.25$ ML. Adapted from Figure 15 in [153].

to compensate for slab asymmetry.[131] The adsorption energies of the adsorbed inhibitor molecules on the partially hydroxylated MgO(100) surface were determined according to Equation (2.25).

Figure 4.4 shows the preferential adsorption geometries of the compounds upon the partially hydroxylated MgO(100) surface for a surface coverage of $\theta = 0.25$ monolayers.[ii] The DFT calculations confirm the in-plane orientation of aromatic compounds, allowing for π-bonds between the aromatic ring and MgO/$Mg(OH)_2$. The highest adsorption energy was obtained for PDC. However, it is noteworthy that inhibitors containing higher numbers of carboxylate groups tend to yield higher adsorption energies due to a higher number of surface bonds and co-adsorbed hydrogen atoms. Comparing the dicarboxylic acids PDC ($E_{ad} = -2.92$ eV) and FA ($E_{ad} = -2.26$ eV) and the monocarboxylic acids MSA ($E_{ad} = -2.73$ eV) and SS ($E_{ad} = -2.36$ eV) reveals that molecular structures with additional hetero-atoms (nitrogen or hydroxyl groups, respectively) increase the adsorption energy. As a consequence, it is expected that also hetero-atoms are involved in the interaction with magnesium (hydr)oxide. Based on the energetically favorable in-plane orientation, it is inferred that both carboxylic acid groups of the dicarboxylic acids, PDC and FA, are involved in the adsorption, implying a bridging mechanism. The inhibition performance was determined by Fockaert et al. via electrochemical impedance spectroscopy (EIS) and potentiodynamic polarization (PDP) measurements.[153] Inhibition performance increases in the order FA < PDC < MSA and was mainly associated to the formation of a MgO/$Mg(OH)_2$

[ii] The surface coverage θ is here defined as the number of adsorbed molecules on a surface divided by the number of molecules in a filled monolayer on that surface.[167]

layer. In contrast to results for commercial purity Mg by Lamaka et al. [14], SS was shown to act as a corrosion accelerator rather than a corrosion inhibitor for high purity Mg. A possible explanation for this observation is the significantly lower iron content of high purity Mg. Hence, instead of capturing released iron particles, SS would preferentially form stable complexes with Mg^{2+} ions, thus increasing surface dissolution. Considering the dicarboxylic and monocarboxylic acids separately, a relationship between corrosion inhibition efficiency and adsorption energy is indicated, where lower adsorption energies are correlated with lower inhibition efficiency and vice versa. However, more calculations would be necessary to make a statistically significant statement about the connection between adsorption energy and inhibition efficiency.

4.2.3 Estimating the Inhibition Efficiency

The corrosion inhibition efficiency (IE) is a popular metric to assess the efficacy of a selected corrosion control strategy. In an attempt to find a connection between the adsorption energy of a Mg dissolution modulator and its IE, a large number of DFT calculations have been performed. In comparison to the previous partially hydroxylated MgO(100) slab with five layers, the substrate system was changed to a fully hydroxylated MgO(111) slab with six layers using a 4×2 super cell and a $6 \times 6 \times 1$ Γ-centered grid of k-points. Moreover, the computationally cheaper exchange-correlation functional PBE [25] was employed to save time and resources in the high-throughput approach. Here, it should be emphasized that the obtained results have been calculated without including corrections for dispersion effects. Apart from that, all settings for the DFT calculations remained the same as for the previous model system.

For reference purposes, all investigated compounds are based on the corrosion inhibitor database generated by Lamaka et al.. [14] Only compounds with an inhibitor concentration of 0.05 M in the experiment were considered in order to preclude concentration-induced effects. A list of all selected compounds is given in Table B.1. Figure 4.5 shows a correlation plot between adsorption energies and corresponding inhibition efficiencies of 45 compounds. Despite a relatively low coefficient of determination ($R^2 = 0.3$) and comparably high root mean square error (RMSE = 42%), a connection between E_{ad} and IE is visible with some statistical significance. Lower adsorption energies appear to correlate with lower IEs and vice versa. A Pearson test results in a correlation coefficient $r = -0.54$ and a p-value < 0.001, indicating a moderate negative relationship, which supports this claim. However, large deviations from this apparent trend are striking and seem to be connected to the bond type (physisorption or chemisorption) and contained species, such as nitrogen, of the

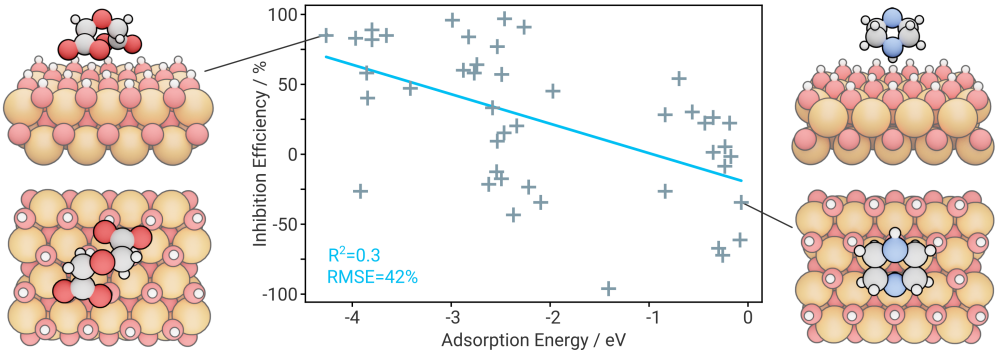


Figure 4.5: Correlation between surface adsorption energy and corrosion inhibition efficiency. The blue line depicts a linear regression fit. The structures corresponding to the highest and lowest calculated adsorption energies are illustrated, respectively. (left) Diglycolic acid. (right) Piperazine.

investigated compounds. Bidentate carboxylate binding can be generally associated with higher adsorption energies than monodentate binding, however, not all compounds containing two carboxylic groups also yield high IEs.

Although the adsorption energy on its own does not allow quantitatively accurate estimates of the inhibition efficiency, its apparent correlation renders it an attractive input parameter for data-driven models. The model accuracy could be further improved with a simulation setup closer to the mode of action of the compounds, albeit at higher computational cost. In this regard, introducing explicit water and simulating the system at realistic temperatures using MD is expected to have a significant impact on the simulation quality. Moreover, employing enhanced sampling techniques like metadynamics would allow to sample the configuration space of the adsorbing compounds, leading to more accurate adsorption energies including entropic effects (free energy of adsorption). The observed trend, however, would not differ necessarily.

Despite ongoing discussions of linking DFT-derived quantities to the corrosion inhibition efficiency[168], Feiler et al. found a relationship between the HOMO-LUMO gap and the IE of molecular compounds for commercial purity magnesium.[16] Using the HOMO-LUMO gap as input parameter for training an artificial neural network (ANN) thereby yielded accurate prediction of the IE, slightly above the experimental error. Hence, developing workflows to calculate quantities such as the free energy of adsorption or complex formation in a high-throughput manner is highly attractive for building predictive ML models. More information on predicting the dissolution modulating performance of molecular structures is given in Chapter 5. Two of the major challenges in pursuing a high-throughput approach using metadynamics are the required computational resources and the collective variables (CVs). The complex nature of metadynamics simulation requires a careful choice of the

CVs, particularly for ab initio MD.

Magnesium dissolution modulators play in important role in modern Mg corrosion engineering with respect to various applications. Although comprehensive corrosion inhibitor databases exist[14], generation of such is resource- and time-consuming. Moreover, identifying chemical compounds with desired properties in the vast chemical space can be compared to finding the needle in the haystack. The presented results in this chapter indicate the high potential of DFT to better understand corrosion inhibition mechanisms and to generate parameters in silico that relate to the corrosion inhibition efficiency. Such parameters can be used to build predictive models, helping in the search for new Mg dissolution modulators. But how exactly can computational methods be used to slowly conquer the chemical space?

Chapter 5

Chemical Space: The Final Frontier

> Machines take me by surprise with great frequency.
>
> — Alan Turing

Small organic molecules that modulate the degradation behavior of Mg constitute benign and useful compounds to modify the service environment of light metal materials for specific applications. The properties of these modulators of magnesium dissolution can be tailored to specific target applications, e.g., as component of an active protective coating or as a part of the electrolyte of a Mg-air battery.[169–172] The massive advantage of organic dissolution modulators is their basically unlimited chemical space, providing countless potential solutions for various applications. The number of available organic compounds is increasing swiftly, with roughly 120 million organic compounds being reported over the last decade alone.[173] Estimates of the number of organic compounds with potentially useful properties are even in the range of 10^{63} and are thus essentially infinite.[174] Automation and robotics technologies are also expanding rapidly and enable modern combinatorial chemistry techniques that can synthesize larger and more diverse chemical libraries. Undoubtedly, combinations with computer-assisted synthesis strategies will further extend this exponential rise in available organic compounds.[175]

As a consequence, the most challenging task is to select compounds with beneficial properties for certain applications from this effectively infinite chemical space of small organic molecules. Experimental approaches alone cannot possibly explore more than a tiny fragment of the vast space of compounds with potentially useful dissolution modulating properties, despite impressive developments in high throughput techniques.[176–179] The immense size of the chemical space is illustrated by a thought experiment in Figure 5.1. The observable universe contains approximately $1.0 \cdot 10^{21}$ stars. A significantly smaller—but still unimaginably large—number of $1.5 \cdot 10^{8}$ organic chemicals is listed in the Chemical Abstract

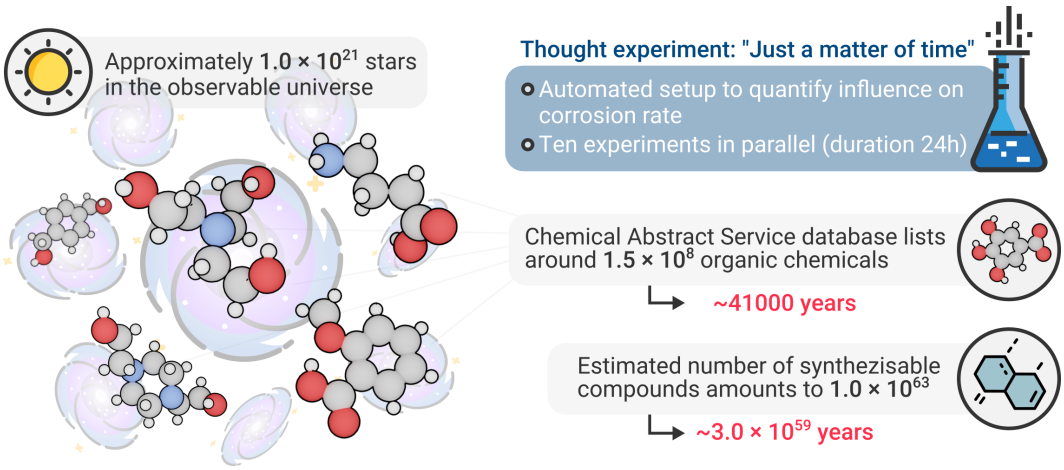


Figure 5.1: Exploring the chemical space. Similar to our universe comprising billions of stars, the vast space of chemical structures contains an immense number of different compounds. Testing all of these in experiments would require an incomprehensible amount of time, thus rendering strategies for faster screening of the chemical space to identify potentially useful structures inevitable.

Service (CAS) database. Assuming an experimental setup, which allows to determine a compound's influence on the corrosion rate in 24 h and to run ten experiments in parallel, it would take around 41000 years to test the entire CAS database. Considering the estimated number of $1.0 \cdot 10^{63}$ synthesizable compounds, the required time would even amount to approximately $3.0 \cdot 10^{59}$ years, exceeding the age of our universe by far. Therefore, there is a great demand for new techniques that facilitate an efficient screening of the chemical space for chemical compounds of interest.

This chapter addresses the potential of machine learning techniques in the search for potent compounds that can contribute to a more effective magnesium corrosion engineering. Training data-driven prediction models on experimental databases allows to estimate different target properties, characteristic to the application at hand. In combination with molecular similarities, these models allow for an enhanced screening of the chemical space and enable a more efficient design of experiments. Parts of the work and figures presented hereafter have been published in

[15] **T. Würger**, C. Feiler, F. Musil, G. B. Vonbun-Feldbauer, D. Höche, S. V. Lamaka, M. L. Zheludkevich, & R. H. Meißner. Data Science Based Mg Corrosion Engineering. *Frontiers in Materials*, 6, 53 (2019).

[16] C. Feiler, D. Mei, B. Vaghefinazari, **T. Würger**, R. H. Meißner, B. J. Luthringer-Feyerabend, D. A. Winkler, M. L. Zheludkevich, & S. V. Lamaka. In silico screening of modulators of magnesium dissolution. *Corrosion Science*, 163, 108245 (2020).

[180] **T. Würger**, D. Mei, B. Vaghefinazari, D. A. Winkler, S. V. Lamaka, M. L. Zheludkevich, R. H. Meißner, & C. Feiler. Exploring structure-property relationships in magnesium dissolution modulators. *npj Materials Degradation*, 5(1), 2 (2021).

[181] E. J. Schiessler, **T. Würger**, S. V. Lamaka, R. H. Meißner, C. J. Cyron, M. L. Zheludkevich, C. Feiler, R.C. Aydin. Predicting the inhibition efficiencies of magnesium dissolution modulators using sparse machine learning models. *npj Computational Materials*, 7, 1-9 (2021).

[182] **T. Würger**, L. Wang, D. Snihirova, M. Deng, S. V. Lamaka, D. A. Winkler, D. Höche, M. L. Zheludkevich, R. H. Meißner, C. Feiler. Data-driven Selection of Electrolyte Additives for Aqueous Magnesium Batteries. *Journal of Materials Chemistry A*, 10, 21672-21682 (2022).

5.1 Data-Driven Prediction Models

Data-driven computational methods[15,16,165,183–187] can efficiently explore larger areas of chemical space with orders of magnitude less time and effort than experimental approaches. Hence, they offer a very efficient way to preselect a short list of promising candidates prior to experimental investigation. Additionally, computational techniques can provide deeper insight into the underlying chemical mechanisms and most important chemical functional moieties.[79,95,153,188–190] A combination of experimental and computational methods constitutes a sound foundation for a data-driven discovery of modulators. Machine learning techniques that model complex quantitative structure-activity or structure-property relationships can predict target properties of hitherto unsynthesized or untested compounds.[15,16,166,183,191]

5.1.1 Database of Magnesium Dissolution Modulators

Such machine learning methods, however, require large, reliable, chemically diverse and balanced training data sets to make accurate predictions that can be generalizable to a broad range of materials. As data-driven models are not capable of reliably predicting the performance of molecules with poorly represented features (e.g., functional groups, elemental species) in the training data, the underlying training set has to reflect the complexity of the

relevant chemical environment. Predictions made by these models are most accurate for compounds that are within the domain of applicability of the model, i.e., the span of values over which each molecular feature in the training set varies.

In an extensive experimental study employing hydrogen evolution experiments, Lamaka et al. measured the corrosion inhibition performance of over 150 organic compounds for nine distinct Mg-based materials.[14] Commercial magnesium processing includes several steps[192], thus preventing the inclusion of metallic impurities, such as iron, is nearly impossible. These impurities induce local galvanic cells into the material that accelerate corrosion and increase hydrogen evolution and Mg dissolution. As Mg dissolution predominantly occurs in intermetallic contact areas, the process releases impurities that re-deposit on the surface, thus increasing the cathodic area and the corrosion rate.[85] Molecules that form stable complexes with the impurities (e.g., $Fe^{2+/3+}$) constitute a promising strategy to modulate the degradation properties of Mg. They also provide the starting point for building an extensive database of magnesium dissolution modulators.[13,14]

5.1.2 Predicting the Corrosion Inhibition Efficiency

Training QSAR models on this database allows to exploit structural features of known compounds to predict the inhibition efficiency of yet untested chemicals with respect to the magnesium-based materials investigated by Lamaka et al.. Depending on which approach is chosen to encode the molecular structures in the database, a number of ML techniques can be used as basis for such models (cf. Section 2.3). Artificial neural networks (ANNs), for example, are valuable tools in the prediction of the corrosion inhibition efficiency (IE), when trained on suitable input parameters. Such parameters could include for instance molecular descriptors representing structural features, such as the number of certain functional groups or atomic species. Moreover, adding DFT-calculated parameters as input can further increase the predictive power of the model.

However, the correlation between the corrosion inhibition efficiency of small organic molecules and DFT-calculated parameters is a controversial topic in the literature. While some researchers claim that there is a clear connection[193–195], e.g., for the HOMO-LUMO gap ΔE_{HL}, which is defined as the difference between the highest occupied (E_{HOMO}) and lowest unoccupied molecular orbital energy (E_{LUMO}), others could not identify any relationship between experiment and theory.[165,196–198] For aluminum alloys, the DFT-derived quantities show almost zero correlation with the IE. In contrast, the HOMO-LUMO gap appears to play a large role for magnesium-based materials, since ΔE_{HL} indicates the affinity of the corrosion modulators for transition metal ions (cf. Section 4.2). As the kinetic stability of an

organic molecule[15,199] increases with increasing ΔE_{HL} and thus its reactivity decreases, the HOMO-LUMO gap is a reliable indicator of the propensity of organic compounds to form complexes with metallic impurities in commercial purity magnesium (CPMg) and magnesium alloys.[15] This is also consistent with the experimental observation[14] that compounds with small HOMO-LUMO gaps yield higher IEs, while dissolution accelerating agents have larger HOMO-LUMO gaps.

This relationship could be highlighted using a subset of the inhibitor database for CPMg with 220 ppm iron impurities (CPMg220).[16] The elemental composition of CPMg220 is provided in Table C.1. Selecting only compounds measured at a constant concentration of 0.05 M and with a weight below 350 Da[i], the database was further reduced to include only small organic molecules. The underlying data distribution with respect to IE is shown in Figure D.1a. Furthermore, chemicals containing underrepresented atomic species (e.g., phosphorous) and compounds that were investigated in form of their corresponding acid salts were omitted. Although the remaining 71 molecular structures show a clear linear relationship between HOMO-LUMO gap and corrosion inhibition efficiency, the rather low R^2 of 0.48 indicates that the IE is not dependent on ΔE_{HL} alone.[16] Combining ΔE_{HL} with molecular descriptors containing further information on the chemical structure, however, Feiler et al. trained an ANN that could predict the IE of seven yet untested compounds with relatively high accuracy (R^2 = 0.74, RMSE = 33%).[190]

Naturally, the versatility of the vast chemical space is associated with a wide range of different functional moieties and molecular features, thus rendering it challenging to identify meaningful input features to develop predictive models with a wide applicability. Cheminformatics software packages like `alvaDesc`[200] facilitate calculations of a large variety of molecular descriptors, ranging from structural and topological features to more complex input features like molecular fingerprints and arcane descriptors derived from chemical graph theory (cf. Section 2.3). On the basis of 1260 potential input features generated accordingly for the given database, Schiessler et al. trained deep learning models to predict the inhibition efficiency of chemical compounds with respect to the magnesium alloy ZE41 (Figure 5.2).[181] After removing all molecular descriptors that exhibited constant values or were essentially zero, 1254 descriptors remained and were extended by six DFT-derived parameters, including the frontier orbitals energies E_{HOMO} and E_{LUMO}. In the resulting set of 1260 molecular descriptors, input features with the greatest impact on the corrosion inhibition responses of 60 small organic molecules on ZE41 were tried to be identified. The underlying distribution corrosion inhibition efficiencies is presented in Figure D.1b. Similar to the previously described approach[16], only dissolution modulators were selected for

[i] Dalton (Da) is a measure of molecular weight or molecular mass; 1 Da ≙ 1 g/mol.

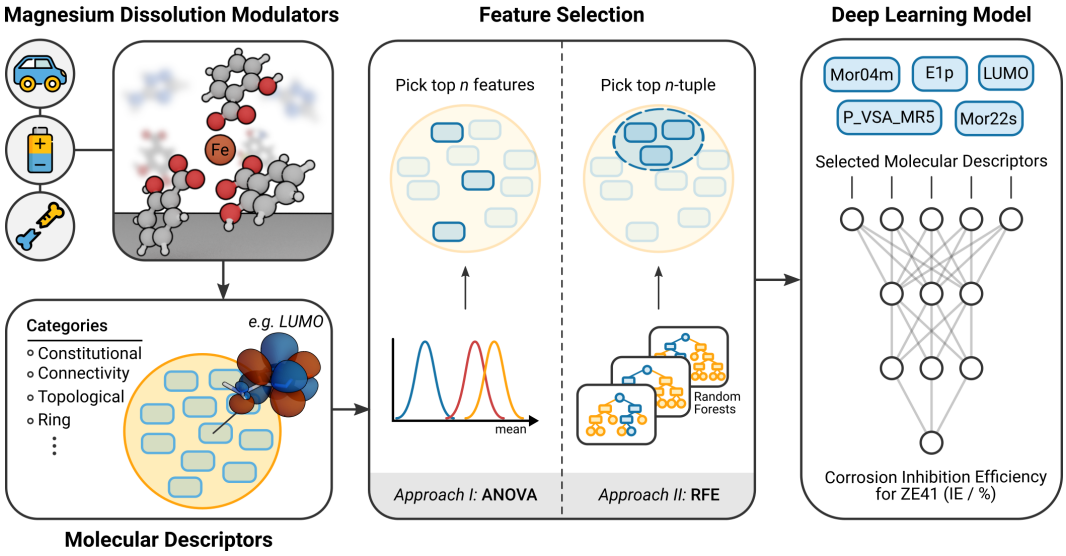


Figure 5.2: Corrosion inhibition efficiency (IE) prediction for ZE41. Utilizing a database of magnesium dissolution modulators, relevant molecular features are selected either by (approach I) analysis of variance (ANOVA) or by (approach II) recursive feature elimination based on random forests (RFE), a type of machine learning. The best-performing feature set defines the input for a deep learning model. This model allows the desired predictions of quantitative structure-activity relationships (QSAR) for the efficiency of magnesium dissolution modulators, here shown for the magnesium alloy ZE41. Adapted from Figure 1 in [181].

training the model that exhibit a molecular weight of less than 250 Da and that were employed at a concentration of 0.05 M. Sparse feature selection has a high impact on the quality of predictive models, which substantially depend on the selected molecular descriptors, as input features with low relevance to the target property can degrade the model. A comparison between ANOVA f-tests and recursive feature elimination based on random forests shows the high potential of machine learning techniques towards statistical analysis to create sparse, predictive QSAR models.[181] For example, training a deep learning model on five molecular descriptors as selected by recursive feature elimination and random forests leads to a root mean square error (RMSE) of 26% when predicting the IE of a pre-defined validation set, whereas ANOVA only achieves an RMSE of 51%. Although ΔE_{HL} was not selected as feature for predicting the IE for ZE41, it is noteworthy that for all trained models, E_{LUMO} was identified to be connected to the corrosion inhibition efficiency. As a result, it is fair to conclude that DFT-calculated molecular properties should not be disregarded when building predictive models. For magnesium dissolution modulators, they can significantly contribute to the predictive power of ML models.

Instead of using distinct molecular descriptors, structural features within the given dataset can also be conflated in molecular fingerprints, such as the SOAP kernel (cf. Section 2.3).

This kernel defines a global matrix containing molecular similarities between the chemical compounds of the underlying dataset. Applying unsupervised learning methods such as kernel principal component analysis (KPCA) or sketch-map, the high-dimensional similarity space can be projected onto two dimensions to help the visualization of potential structure-activity relationships. Distortions of the space are focused so that close/distant, i.e., similar/dissimilar, structures in the high-dimensional space maintain this relationship in lower dimensions. When combining the pattern recognition capabilities of unsupervised learning with corrosion experiments, it is possible to visualize the relationship between the molecular structures and the corresponding inhibition efficiencies in a structure-activity landscape, where all experimentally tested structures act as landmark points. Clusters in the resulting similarity map may indicate correlations between the inhibition efficiency and associated molecular structure, allowing to relate certain structural features to potential dissolution modulating properties. Such clusters can be identified by visual inspection or using a variety of different clustering algorithms in low- and high-dimensional space.[22,201] Subsequently, the inhibition efficiencies of uninvestigated compounds can be qualitatively predicted following the proposed workflow in Figure 5.3. To predict the dissolution modulating properties of an untested compound, its structural relationship to the identified clusters has to be determined. This can be accomplished by out-of-sample embedding, where the new structure is projected into the generated similarity map by reproducing the distances to the previously defined landmark points.[20] Once the structure is projected into the structure-activity landscape, its relative position to previously identified clusters can help to assess the impact on corrosion events. Clearly, an unknown compound projected into, or close to, a cluster dominated by a particular molecular property would be expected to show similar behavior. Concurrently, this approach indicates whether it is reasonable to examine untested additives in further corrosion experiments, hence saving a tremendous amount of time and resources compared to an experimental high-throughput approach.

Although this approach is useful for estimating the impact new additives may have on the corrosion of magnesium-based materials, some applications require quantitative predictions. For labeled data, supervised learning algorithms such as kernel ridge regression (KRR) attempt to find connections between the molecular similarities and a target property of interest, in this case the corrosion inhibition efficiency.[ii] The structure-activity landscape can then be complemented with quantitative predictions of the IE for yet untested compounds. The synergistic combination of both methods provides a powerful approach for the design of magnesium corrosion modulators that exploits the great efficiencies of in silico methods.

[ii] Other molecular properties predicted with high accuracy using SOAP-based KRR models include the frontier orbital energies and molecular polarizability.[68]

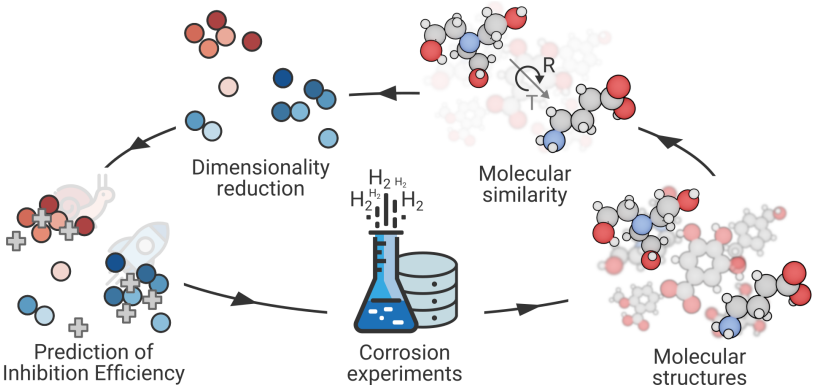


Figure 5.3: Qualitative prediction of inhibition efficiencies (IE) of magnesium dissolution modulators. Molecular similarities are calculated between chemical structures derived from a database of magnesium dissolution modulators. Applying dimensionality reduction and coloring the resulting map according to the IE, two-dimensional structure-activity landscapes are generated that reveal structure-activity relationships. When projecting yet untested compounds onto this map via out-of-sample embedding (gray crosses), the relative position to given clusters can indicate their performance in the corrosion experiment. After validation, the underlying database can be extended, thus further improving the predictive power of the qualitative model. Adapted from Figure 1 in [15].

Thus, a comprehensive molecular similarity map based on a SOAP kernel derived from all structures in a training and test set can be used to virtually screen a large number of potential candidates. Concurrently, their degradation modulating performance can be predicted with KRR, either to validate the inhibition performance of a known modulator, or to predict the degradation modulating properties of unsynthesized and/or untested organic compounds.

To assess the potential of unsupervised and supervised learning approaches with SOAP, the SOAP-REMatch kernel (cf. Section 2.3) has been calculated for the same 71 training and 7 test structures as in the study of Feiler et al. and used for model training and testing, respectively. Hyperparameters were fine-tuned in a grid search with k-fold cross validation (Figure D.2), resulting in a hyperparameter set of $r_c = 3.0\,\text{Å}$, $\xi = 0.3$, $\zeta = 0.6$ and $\gamma = 0.3$ for the SOAP-REMatch kernel, as well as $\lambda_{KRR} = 11$ for the regularization parameter of the KRR model. A sketch-map representation of the 71 training structures is presented in Figure 5.4a, where the 7 test structures are embedded within the structure-activity landscape and highlighted. It is clear that the four stronger corrosion inhibitors on the right-hand side of the map cluster together, while the three compounds with weak inhibition to strong dissolution accelerating properties are clustered near each other on the left-hand side of the sketch-map. Considering the IEs of the surrounding training structures, the relative test set positions allow for a qualitatively accurate prediction.

The KRR model was validated by comparing its key performance indicators to those of the ANN model (Figure D.4).[16] Although the ANN performed well in terms of prediction

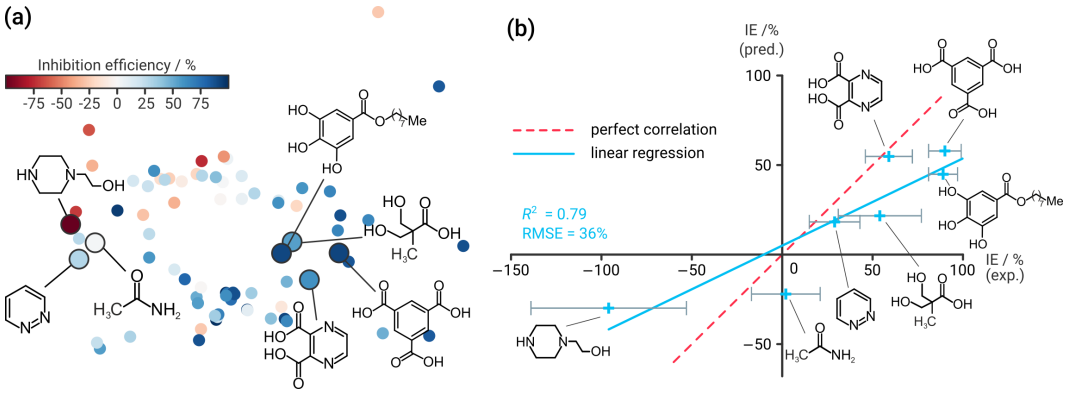


Figure 5.4: Unsupervised vs. supervised learning. (a) A sketch-map representation based on molecular similarities and inhibition efficiencies (IEs) of 71 compounds (colored dots). Clusters concentrating similar dissolution modulating properties indicate a structure-activity relationship. Qualitative prediction of the IE of seven test compounds is obtained using out-of-sample embedding, project them onto the map in relation to the observed clusters. (b) Predicted test set IEs obtained from a kernel ridge regression (KRR) model. The blue line is a linear least square fit of the predicted and measured values. The RMSE value is in absolute percent. The red dashed line represents ideal correlation. The error bars depict the standard deviation of the experimentally derived IE. Adapted from Figures 1 and 2 in [180].

accuracy, its performance depends on the careful selection of molecular descriptors that can strongly influence the prediction outcome and interpretation of the models is often problematic. Hence, combining all structural features in a global similarity measure, as defined by SOAP, provides an attractive approach to reduce the complexity of the model input. Employing the resulting SOAP kernel as input for KRR enables physically interpretable predictions that can be directly ascribed to the molecular structure. The KRR-predicted IEs for the seven test compounds (see Figure 5.4b) had a higher R^2 of 0.79 and a slightly higher RMSE of 36% compared to the ANN model (R^2 = 0.74, RMSE = 33%).[16] However, with only seven tested compounds the difference between the ANN and KRR models is not highly significant. A Pearson rank correlation test was also conducted on the KRR-predicted and experimentally determined values. Similar to the ANN study, there was a strong correlation for the KRR model (Pearson correlation coefficient r = 0.89). The associated p-value of 0.007 indicates acceptable statistical significance for the model.

5.2 On the Search for Magnesium Dissolution Modulators

After benchmarking the robustness of the SOAP-based KRR model against the distinct descriptor-based ANN model, the former is used in a screening approach to identify potent magnesium dissolution modulators. By augmenting the given database of magnesium dissolution modulators with additional compounds of unknown performance, their potential

IE can be at first qualitatively estimated using sketch-map, where already tested dissolution modulators serve as reference. Selecting compounds of interest is further specified by complementing the unknown compounds with quantitative predictions of the IE, using a KRR model trained on the SOAP kernel and IEs of all considered structures. Experimental validation helps to further improve the underlying model by including investigated compounds in the training set. The synergistic combination of both methods provides a powerful approach for the design of Mg dissolution modulators that exploits the efficiencies of in silico methods.

5.2.1 Exploiting Structure-Activity Relationships

The given database was extended by 74 additional compounds with unknown IE that comprise chemicals with similar functional groups to those 71+7 in the original training and test set. For the resulting 152 organic molecules, the SOAP-REMatch kernel was recalculated with the same set of hyperparameters and used to generate a second sketch-map (Figure 5.5). The untested compounds are depicted as gray dots while the already tested training set compounds are color-coded according to their IE. As the augmented compounds lie in or near the domain of applicability of the original model, the sketch-map representation is expected to give reliable qualitative predictions. Only non-toxic, inexpensive, small organic molecules were included in the untested set. The molecular weight of training compounds was < 350 Da while the majority of the test set molecules (64) used for prediction had molecular weights < 200 Da. Small anti-corrosion additives are required to increase the efficiency of protective coatings without impairing their structural integrity. Considering the IEs of the training structures, six clusters were identified and selected by visual inspection. Results of a k-means clustering substantiate the cluster definitions (Figure D.5). In Figure 5.5, the mean IE values of molecules in the training set clusters decrease in the order **b** ($71 \pm 15\%$) > **d** ($65 \pm 6\%$) > **c** ($50 \pm 27\%$) > **e** ($37 \pm 57\%$) > **f** ($3 \pm 24\%$). Clearly, some of the untested compounds mapped onto clusters of modulators with inhibitory effects (**b**, **c**, **d**) while others generated new clusters (**a**) or were located in map regions corresponding to compounds with highly diverse properties, ranging from strong accelerators to effective inhibitors (**f**). The mean KRR-predicted IE values for molecules in the test set in defined clusters decrease in the order **b** ($51 \pm 9\%$) ≈ **c** ($48 \pm 14\%$) > **d** ($39 \pm 10\%$) ≈ **a** ($35 \pm 11\%$) > **e** ($24 \pm 5\%$) > **f** ($-35 \pm 12\%$), indicating a qualitatively accurate prediction.

A blind test is an excellent way to assess the predictive power of a model. A total of 10 chemicals, representative of each cluster, were randomly selected and tested experimentally under the same conditions as the compounds in the original database (cf. Section 5.1).[iii] As

[iii] The experiments (cf. Appendix C.1) were conducted by Di Mei and Bahram Vaghefinazari of Helmholtz-

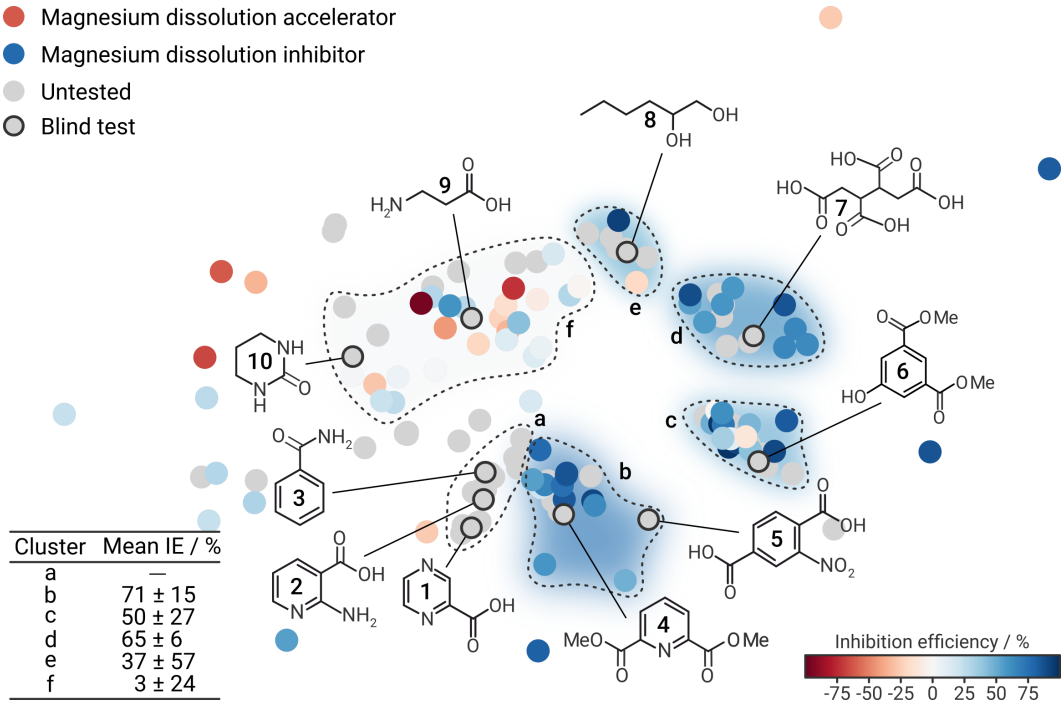
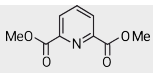
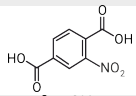
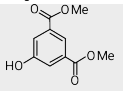
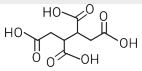
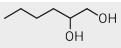
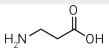


Figure 5.5: Sketch-map representation of 152 molecular structures. The sketch-map is based on molecular similarities of 78 tested (colored according to inhibition efficiency IE) and 74 untested (gray) modulators. Six clusters were highlighted based on results from k-means clustering. The mean IEs of the clusters are shown in the lower left corner. Adapted from Figure 3 in [180].

the experimental performance of compounds located within cluster **a** is an uncharted area of the sketch-map, the number of compounds selected for the blind testing was in proportion to the size of the cluster. The general agreement between predicted and experimental values is good except for benzamide (compound **3**) that is predicted to have a moderate inhibiting effect (44%) whereas the experiment showed it was a dissolution accelerator with an IE of -43 ± 30%, see Table 5.2. The discrepancy can be explained by benzamide precipitating at an inhibitor concentration of 0.01 M while the model was trained on data with a 0.05 M modulator concentration. Due to this difference in inhibitor concentrations, compound **3** was omitted from correlation of experimental and predicted results, see Figure 5.6. The trained KRR-model yields high predictive performance for the remaining nine chemicals with an $R^2 = 0.72$ and an RMSE = 32%. Additional information on the training and test errors of the KRR model are provided in Figure D.3. More detailed information on the experiments are provided in Table C.2. A Pearson rank correlation resulted in a positive correlation coefficient of 0.85 and a p-value of 0.004. As Table 5.2 shows, most of the predicted IE values agree

Zentrum Hereon.

Table 5.2: Experimental and predicted inhibition efficiencies (IEs) of ten compounds not used to train the model. IEs for compounds with a carboxylic acid moiety were determined as sodium salt in the associated hydrogen evolution experiments. Experimental uncertainties were calculated from three experiments, except for **1**, **3** and **10** where four experiments were performed. Values for final pH after immersion tests are provided. The starred value denotes the reference for calculating the inhibition efficiency. Adapted from Table 1 in [180].

		Compound	IE_{KRR} / %	IE_{exp} / %	H_2 volume / mL cm^{-2}	Final pH	Cluster
–	NaCl	0.5% Sodium chloride*	—	0	23.5 ± 3.8	10.5	—
1		Pyrazinecarboxylic acid	18	1 ± 41	23.2 ± 9.0	10.5	**a**
2		2-Aminopyridine-3-carboxylic acid	30	54 ± 16	10.9 ± 0.2	9.4	**a**
3		Benzamide	44	−43 ± 30	33.7 ± 6.1	10.5	**a**
4		Dimethyl 2,6-pyridine-dicarboxylate	63	64 ± 19	8.5 ± 2.4	8.8	**b**
5		2-Nitroterephthalic acid	60	68 ± 16	7.6 ± 0.5	10.9	**b**
6		Dimethyl 5-hydroxy-isophthalate	55	29 ± 24	16.7 ± 4.4	9.2	**c**
7		1,2,3,4-Butane1-tetracarboxylic acid	53	27 ± 17	17.1 ± 1.7	9.1	**d**
8		1,2-Hexanediol	28	2 ± 19	23.1 ± 2.7	10.4	**e**
9		β-Alanine	-7	−86 ± 17	43.6 ± 1.7	10.1	**f**
10		N, N'-Trimethyleneurea	-26	−37 ± 26	32.2 ± 4.9	10.5	**f**

with the experimentally determined values within experimental error. It is noteworthy that the IE values of aliphatic compounds in the blind testing set (**7**, **8**, **9**, **10**) are overestimated. Tetracarboxylic acid **7** is located in one of the more dense clusters **d** which might explain the small IE variation within the cluster (Figure 5.5). Chemicals in the training set that cluster in **d** have a mean inhibition efficiency of 65 ± 6%. Hence, untested compounds mapped close to this cluster should exhibit a similar IE. The in silico model estimates the IE value of **7** to be significantly lower than the mean value of the data points defining the cluster, a trend in agreement with the lower experimental value of 27 ± 17% IE. The complex speciation of

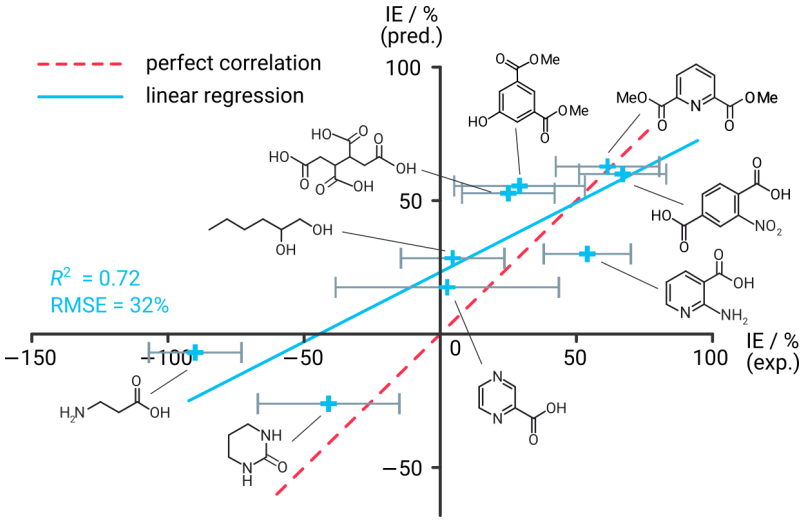


Figure 5.6: Validation of the KRR model. Correlation between the mean of 9 experimental blind testing compounds and their predicted inhibition efficiency (IE) values. The error bars depict the standard deviation of the experimentally derived IE. The light blue line is the linear least squares fit of the predicted and measured values. The RMSE value is in percent. The red, dashed line represents a perfect correlation. Benzamide was omitted because of precipitation during the experiment. Adapted from Figure 4 in [180].

tetracarboxylic acid at pH between 7 and 10 may be responsible for some of the prediction error, which is nonetheless only slightly outside one standard deviation of the experimental error. Aliphatic diol **8** is located in cluster **e** (mean of 37 ± 57% IE). There are only two compounds allocated to this cluster. One acts as an efficient inhibitor, while the other is a moderate corrosion accelerator; explaining the very large standard error around the mean value of this cluster. However, the predicted performance of diol **8** reflects the correct trend with an IE value significantly lower than the mean IE of cluster **e**. The urea **10** and the amino acid **9** are located in cluster **f** that contains modulators with highly diverse IEs, ranging from weak inhibitors to potent accelerators. While the value of the piperidone compound **10** is quite accurate, the model heavily underestimates the dissolution accelerating properties of β-alanine (IE_{pred} = −7%) compared to the measured IE (−86 ± 17%) for CPMg220. Again, at a final pH around 10, the amino acid will be ionized, and the chemotype may not be adequately represented in the training data (a potential issue whenever the chemical diversity is large compared to the size of the data set). The accelerator properties of both compounds were confirmed by subsequent validation experiments.

Contrary to the models' general overestimation of IE values in the four aliphatic compounds, the predicted IE values for the five aromatic molecules **1**, **2**, **4**, **5**, **6** agree with the experimental IE values within experimental error. The isophthalate **6** is associated with the molecular cluster **c** and the predicted value of 55% IE is in good agreement with the mean of the

compounds that were used for training of the model. The pyrazine derivative **1** and the pyridine derivative **2** are located within cluster **a** which represents an uncharted area of the sketch-map. Although both predictions are in good agreement with the experimentally derived values, more experimental data points are necessary to provide robustness for predictions in this region of similarity space. The neighboring cluster **b** includes the pyridine derivative **4** as well as the nitro-substituted modulator **5**. The performance prediction of the two modulators is in good agreement with the experimental IEs. Modulators **4** and **5** have the highest predicted inhibiting effect that was confirmed by the conducted hydrogen evolution experiments. Furthermore, the similarity observed in trends for molecules in clusters **a** and **b** suggest they may in fact be members of a single larger cluster (as already proposed by the k-means clustering in Figure D.5), something that could be confirmed by experimental IE measurements of chemicals lying between the two clusters. In summary, the predictions of the compounds in the blind test set is in qualitative agreement with the experimental IE values determined subsequently.

5.2.2 Uncharted Similarity Space

Of the 64 modulators whose predicted properties were not checked experimentally, 9 are located outside the defined clusters. The complete list of predicted values is provided in Table D.1 as $\mathrm{IE}_{\mathrm{KRR}}$. One of these modulators lies close to cluster **c** and is likely to be a good inhibitor. Two others map to the top left area of the sketch-map close to cluster **f** so they will probably also be effective dissolution accelerators. The remaining six compounds map between clusters **a**, **b** and **f**. As **a** and **b** contain inhibiting agents while **f** contains weak inhibitors and accelerators, no definite statement about the experimental IE of these six compounds can be made.

5.2.3 Enhanced Screening of Chemical Compound Space

Clearly, the search for effective dissolution modulators in the vast chemical space of compounds with potentially useful properties requires very efficient tools. Manual selection of compounds for experimental screening is often biased by the individual chemical intuition, compound availability, cost, toxicity, and experience. Hence, whole regions of chemical space that may contain highly effective modulators (islands of utility) can easily be overlooked. This provided the motivation for the development of the data-driven method *ExChem* for unbiased identification of chemical leads, as illustrated in Figure 5.7. For a molecule x_{A} with a target property y_{A}, a molecule $x_{\mathrm{B}} \sim x_{\mathrm{A}}$ is likely to yield a target property $y_{\mathrm{B}} \sim y_{\mathrm{A}}$, assuming a structure-activity relationship exists. Once a cluster containing, for example, potent corrosion

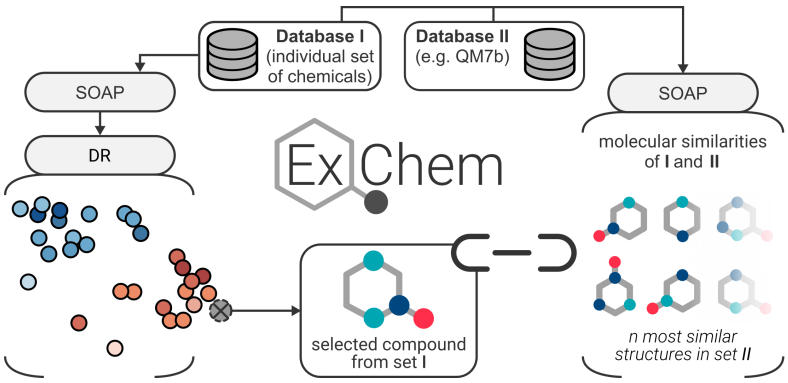


Figure 5.7: Workflow of *ExChem*. Molecular compounds of an individual dataset of chemicals (I) are encoded using the smooth overlap of atomic positions (SOAP) kernel. Dimensionality reduction (DR) of associated molecular similarities allows to create a similarity map that vividly illustrates structure-activity relationships. For a given database II, an extended SOAP kernel is calculated that contains the molecular similarities of II with respect to I. Now, for any selected compound of interest in database I, the n most similar structures in dataset II are suggested, thus facilitating the discovery of potent magnesium dissolution modulators. Adapted from Figure 5 in [180].

inhibitors is identified, untested structures that map onto this cluster may have similar or superior corrosion inhibiting properties. Implementation of this approach requires a second, extensive database of potential candidates and a SOAP-REMatch kernel computed from the structures of both databases. Picking a structure of interest in the sketch-map results in the visualization of the most similar compounds of the second database, providing a basis for the automated discovery of corrosion modulators.

The best-performing corrosion inhibitor **5** in the test set maps onto the edge of cluster **b** that is formed from modulators with high IE values. To check the model robustness concerning the periphery of observed clusters, this structure constitutes a promising starting point to screen a larger molecular database for similar compounds. The QM7b database[202,203] contains 7211 compounds, thus providing a pool of compounds not included in the training or test datasets. After computing the SOAP-REMatch kernel ($r_c = 3.0$ Å, $\xi = 0.3$, $\zeta = 0.6$, $\gamma = 1.0$) for all structures from this database and the initial dataset (7211 + 152), the global similarity matrix (7363 × 7363 diagonally symmetric) was used to identify structurally similar compounds from the QM7b database. A sketch-map complemented with KRR-based IE predictions using this kernel is illustrated in Figure D.6.

When searching for similar compounds in the QM7b database, a similarity submatrix (7363 × 7211) was used to eliminate hits in the 152-member training and test sets. This identified five structures with the highest similarity to 2-nitrophthalic acid (**5**) and predicted their IE values (Figure D.6). However, the suggested molecules found by this proof of principle example are biased as the QM7b database only contains molecules with ≤ 7 heavy

atoms[iv], whereas molecule **5** contains 13 heavy atoms. Clearly, the quality of suggestions is highly dependent on the properties of the underlying database. Screening of databases containing larger molecules (e.g., the GDB-13 which contains roughly 1 billion structures) will undoubtedly increase the value of the proposed workflow, albeit with increased computational cost for computation of the SOAP kernel. Moreover, when screening large databases, it is very important to understand how the model's domain of applicability maps onto the database, since molecules well outside the model domain will be poorly predicted. Nevertheless, the suggested compounds in this example still replicate local structural motifs in substantially smaller molecules that may play a key role in corrosion inhibition behavior.

To assess the accuracy of the workflow in the discovery of lead structures, a database provided by Thermo Fisher Scientific, containing 7094 commercially available small organic chemicals, was subsequently screened. The original dataset contained over 80,000 structures but was drastically reduced to only include small organic compounds that roughly lie within the domain of applicability of the underlying model. Again, the SOAP-REMatch kernel (r_c = 3.0 Å, ξ = 0.3, ζ = 0.6, γ = 1.0) was computed for all structures (7094 + 152). A sketch-map colored with KRR-based IE predictions using this kernel is depicted in Figure 5.8. For the selected 2-nitrophthalic acid (**5**), five chemicals yielding high structural similarities are chosen from a similarity submatrix in the kernel (7246×7094). Although these chemicals exhibit high similarities in the high-dimensional space, they are not necessarily mapped to the two-dimensional space in direct proximity due to the nature of the sketch-map algorithm, as is displayed in the inlay in the lower right corner of Figure 5.8. However, all selected structures are mapped close to each other and to already tested structures. The predicted IE values were again validated in hydrogen evolution experiments[v] following the previous test procedure (cf. C.1). A comparison between KRR-predicted and experimental IEs is presented in Table 5.3. More detailed information on the experiments is summarized in Table C.3.

Despite two outliers, the results show good to excellent correlation between prediction and experiment (R^2 = 0.84, RMSE = 36%). All five candidates were predicted to exhibit an inhibiting effect. It is noteworthy, that the prediction for the two phthalic acid derivatives **12** and **13**, as well as the tricarboxylic acid **14** are in very good agreement with the experimental investigation. However, compounds **11** and **15** only exhibit a moderate inhibiting effect on the corrosion of CPMg in contrast to the comparably high IE values that are predicted by the KRR model. An important mechanism of corrosion inhibition for Mg-based materials is the

[iv] The QM7b database is a subset of the GDB-13 database and only includes molecules with up to seven heavy atoms, defined by C, N, O, and S.

[v] The hydrogen evolution experiments were conducted by Di Mei and Bahram Vaghefinazari of Helmholtz-Zentrum Hereon.

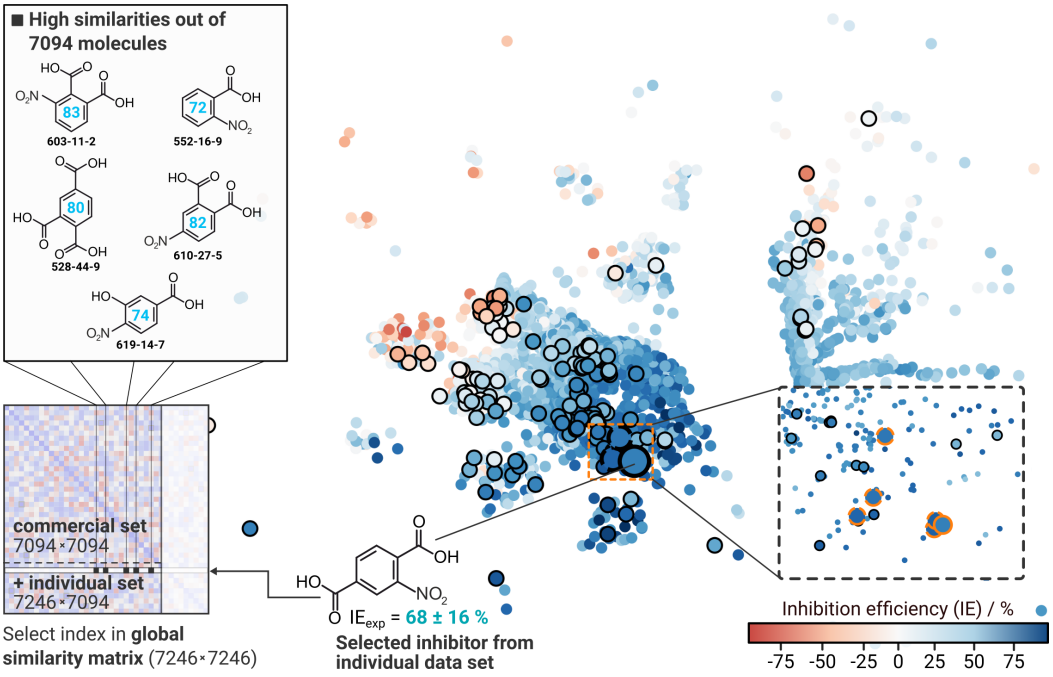
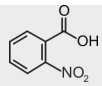
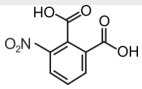
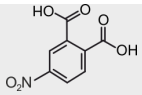
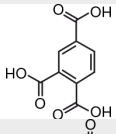
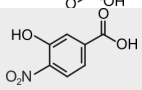


Figure 5.8: Sketch-map based on a commercial database and 152 individual chemicals (7246 compounds in total). The dots are colored according to predicted inhibition efficiencies (IEs) by means of kernel ridge regression (KRR). The KRR model was trained on 78 experimental IEs (black-rimmed dots). By referencing the selected 2-nitrophthalic acid (compound 5, large black-rimmed dot) to the underlying SOAP kernel, five highly similar molecules are determined from the global similarity matrix (dashed orange-rimmed dots) along with their predicted IE values and CAS identifiers. As illustrated in the inlay on the right side, high similarities in the high-dimensional space do not automatically result in direct proximity in two-dimensional space. Adapted from Figure 6 in [180].

capability to form complexes with iron ions (Fe^{2+}/Fe^{3+}) that are released during the corrosion process. Here, substitution of the aromatic system with electron withdrawing groups also affects the interaction strength of the carboxylate with Fe which in turn influences the IE.[204] Although carboxylic acids **11** and **15** contain nitro groups in *ortho* and *para* position with similar electron withdrawing effects, these two compounds also lack the vicinal dicarboxylate moiety present in the other inhibitors that is an important chelating functionality for metals (as in the chelating agent EDTA).[205] This explains the differences in the experimental performance. As mentioned previously, the employed KRR model is based on structural similarities and does not consider electronic properties. Hence, it may only indirectly learn that there is a correlation between electronic effects and structure. It is obvious that the underlying training database does not contain a sufficient amount of structures that exhibit the substitution pattern. Thus, the performance of the two compounds is overestimated but still qualitatively correct. Contrary to this, the predicted values for the two phthalic acid

Table 5.3: KRR-predicted and experimental IEs of five compounds to validate the similarity-based discovery workflow. IEs for compounds with a carboxylic acid moiety were determined as the sodium salt in the hydrogen evolution experiments. Experimental uncertainties were calculated from three experiments. A linear least squares fit of the predicted and measured values yields an $R^2 = 0.84$ and RMSE = 36%. Adapted from Table 2 in [180].

		Compound	IE_{KRR} / %	IE_{exp} / %	H_2 volume / mL cm^{-2}	Final pH
11		2-Nitrobenzoic acid	72	23 ± 25	18.0 ± 5.3	10.9
12		3-Nitrophthalic acid	83	73 ± 17	6.2 ± 1.4	11.2
13		4-Nitrophthalic acid	82	70 ± 20	4.6 ± 0.8	11.1
14		1,2,4-Benzenetricarboxylic acid	80	80 ± 16	7.1 ± 3.5	11.1
15		3-Hydroxy-4-nitrobenzoic acid	74	13 ± 31	20.5 ± 7.8	9.4

derivatives **12** and **13** are highly accurate despite the presence of a nitro moiety. This can be explained by the fact that each of the compounds bears a second vicinal carboxyl group (chelating effect) that is not affected by substitution of the aromatic system with a nitro functionality. Hence, they display higher degrees of inhibition in the experiment and are in excellent agreement with the predicted values. For a larger training dataset, the cutoff radius r_c of the SOAP kernel can be adapted to higher values to better capture the impact of such structural features.

The Python-based workflow *ExChem* for automated screening of large molecular databases using chemical similarities allows unbiased identification of efficient candidates to combat the degradation of Mg-based materials. Particularly for experimentalists, the employed machine learning-based strategies facilitate an intuitive and fast screening of large databases to identify similar compounds, simplifying the search for compounds with potentially useful properties, and dramatically decreasing the time and resources required relative to those for experimental discovery methods. For easier use, the developed approach was implemented in a web application, which is found at www.exchem.de. Naturally, the presented concept is not limited to magnesium and the corrosion inhibition performance. In the next section, the approach is adopted to explore the structure-activity landscape of magnesium-air battery

additives in a similar fashion by employing a corresponding experimental database to train the model.

5.3 Smart Selection of Magnesium Battery Additives

The utilization of Mg as energy material, particularly as high energy density conversion electrode in metallic configuration, seems to be a promising development. When Mg is employed as anode material in a battery, tuning the dissolution rate is desired to ensure continuous discharge behavior, especially when the electrolyte is water-based.[206–208] Such aqueous primary Mg batteries have several advantages in comparison to lithium-air batteries. They operate at any pressure, have greater power storage capacity, are cheaper, safer and have a smaller environmental footprint.[8,209–212] Surrounding air can directly provide the oxygen for the cathodic reaction, enabling a green battery composed of magnesium, water and air. An air cathode is beneficial as it results in a significant reduction in weight and consequently increases the specific energy of the battery. Moreover, the battery can be easily rejuvenated by displacing the depleted Mg-based anode, rendering this technology a promising candidate for remote and portable devices. Three detrimental phenomena fundamentally determine the performance of the anode: self-corrosion of the anode material mainly caused by micro-galvanic reactions with impurities (e.g., Fe, Cu, Ni) or noble secondary phases; passivation of the anode surface caused by secondary reactions, i.e., the formation of corrosion- and discharge products at the interface; and the so-called chunk effect[213–215] that causes detachment of large metallic particles during discharge, implying loss in utilization efficiency (UE) and structural anode integrity. Fortunately, the properties of the anode material can be improved and controlled by micro alloying[206,216,217], or by the use of battery electrolyte additives[8]. In particular, primary Mg-air batteries[169,170,218–220] require effective additives to control self-corrosion, thus increasing the utilization efficiency, and to reduce formation and deposition of corrosion and discharge products on the anode surfaces, thereby increasing the discharge potential (DP). Each application has specific needs, e.g., sensors in sea buoys need low output voltages over a long period of time. Here it is important to maximize the UE of materials for longevity, rather than having high DPs. Alternatively, power sources for autonomous underwater vehicles (AUV) require high output voltages for a shorter period, thus high DP values are more important than high UE to ensure that the AUV is fully operational during its deployment. Clearly, the ideal situation is to create additives for Mg-based batteries that generate high UE and DP to satisfy the needs of most or even all applications, and to avoid the need for bespoke additives for each target application. Achieving this goal is complicated by the fact that a high DP is often accompanied by a low

UE and vice versa.

The dissolution modulating properties of small organic molecules allow to effectively control the key performance parameters of Mg-air batteries.[8,169] Effective electrolyte additives have been reported recently, however, identifying further molecules with a positive impact on the key performance parameters of magnesium batteries is not trivial. Fortunately, data-driven prediction models and enhanced screening methods (cf. Section 5.2) immensely facilitate the search for potent electrolyte additives.

As foundation for developing such a data-driven approach, an experimental database including the DPs and UEs of 49 electrolyte additives was used. The underlying discharge experiments are described in more detail in Appendix C.2. Associated molecular structures were DFT-optimized using the quantum chemical software package Turbomole 7.4[221] at the TPSSh/def2SVP level of DFT. Then, the structures were encoded using distinct molecular descriptors (as in the ANN studies[16,181]) and SOAP fingerprints, respectively. Here, both encoding strategies are employed to compare their performance in predictive models for selecting potent battery additives. Although SOAP yields certain advantages as the entire molecular structure is easily encoded in a single quantity, the rather small size of the given database might now allow to learn molecular properties that only implicitly relate to the molecular structure. Hence, including specific molecular descriptors that, e.g., explicitly describe electronic properties, can significantly improve the predictive performance of a trained model.

The optimized structures from the DFT calculations were used as input for the cheminformatics software package `alvaDesc`[200] to generate roughly 2500 molecular descriptors related to structural features. After omitting molecular descriptors with small variances and/or those that are close to zero, the remaining 852 descriptors were used in combination with the seven DFT descriptors to define a feature matrix $\mathbf{\Phi}_\mathrm{D}$ for all compounds. The considered DFT properties are: the frontier orbital energies (HOMO, LUMO) and the HOMO-LUMO gap (ΔE_HL); the calculated heat capacities (C_p, C_v) and the chemical potential (μ) calculated at 298.71 K; plus the dipole moments of the small organic molecules. As the number of features is too large for a dataset of 49 molecules, and many feature are not informative, the most relevant features with respect to the target properties were determined using the principal covariates adaptation of CUR matrix decomposition, as implemented in `scikit-cosmo`[222]. Both descriptor families, the SOAP kernel $\mathbf{K}_\mathrm{S}$ comprising various structural features, and the feature matrix $\mathbf{\Phi}_\mathrm{D}$ containing multiple descriptors were used to train QSAR models.

Although linear methods are simple and interpretable, they do not generate robust and predictive models when the underlying structure-activity relationship is significantly nonlinear. Kernel methods introduce a nonlinear relation between samples in the form of a positive-

definite kernel function $k(\mathbf{x}, \mathbf{x}')$ and are suitable for modeling nonlinear relationships. This results in a more efficient determination of a low-dimensional latent space and increased regression performance.[75] Therefore, the kernel matrix $\mathbf{K}_\mathrm{D}$ was calculated from the feature matrix $\mathbf{\Phi_D}$ using cosine similarity, analogous to the SOAP kernel implemented in `librascal`. All kernels used in the model training were centered and normalized.

Exploration of structure-activity relationships within a given dataset is facilitated by the use of methods that combine the capabilities of pattern recognition and the predictive power of unsupervised and supervised learning, respectively. Hence, the kernel matrices calculated from both molecular descriptor approaches are used as input for KPCovR models, along with the experimental target values DP and UE. In the following, the effectiveness of molecular descriptors and ML methods is evaluated with respect to reliably selecting potentially effective electrolyte additives for aqueous Mg batteries for further experiments. Two types of KPCovR models are constructed: a SOAP-based S-model and a distinct molecular descriptor-based D-model. All hyperparameters to train the resulting two models were optimized in a 5-fold cross-validated grid search. Further information regarding the models can be found in Appendix D.2.

5.3.1 Active Design of Experiments

A century ago, Fisher laid the foundation[223] for the modern design of experiments (DoE) strategies that optimize processes by systematically screening all independent variables relevant to a specific process. Unfortunately, the experimental efforts to implement a conventional DoE is intractable for a relatively low number of independent processes. A full-factorial DoE requires M^N experiments where N is the number of independent variables and M is the number of values each variable can adopt. Hence, eight independent process variables that can adopt ten values require 100 million experiments. Partial factor designs sample these large spaces with various degrees of sparsity and can be designed to greatly reduce the number of required experiments, albeit with loss of information.

Recent approaches to reduce the experimental workload employ active design of experiments (ADoE) concepts, combining active learning with the original DoE approach.[224,225] An ADoE study is essentially adaptive data collection with four basic steps. In the first step, a small seed dataset is collected and used to build a predictive model for specific target properties. Here, the DP and UE values for 49 small organic molecules were measured when employed as electrolyte additives for primary Mg-air batteries.[vi] On the basis of this initial dataset, models are developed that predict the properties of a new set of potential

[vi] The experiments (cf. Appendix C.2) were conducted by Linqian Wang of Helmholtz-Zentrum Hereon.

additives in the second step. In the third step, additional experimental data are collected for those predicted additives with the highest prediction uncertainty. In this case study, a test set of small organic chemicals not considered for training of the initial model was used, similar to that described in Section 5.2.[16,180] Finally, all datapoints are subsequently used to train a second QSAR model used to predict the properties of a second set of test chemicals to evaluate accuracy and robustness of the augmented model. This process can be iterated to generate larger and larger models and domains of applicability while strongly constraining the numbers of experiments required. Increasing the size and diversity of a training set leads to higher model accuracy and better generalization. Outliers are not detrimental to the development of predictive QSAR models as they are in fact exploited in the ADoE method to incrementally increase the size of the model domain. The *ExChem* approach[180] facilitates rational selection of compounds distant from those in the training set, i.e., selecting chemicals from a new region of chemical space. This is accomplished by exploiting molecular similarities as obtained by SOAP.

Here, the previously employed commercial database comprising 7094 compounds (Thermo Fisher Scientific) and the 49 nominated structures from the experimental dataset were used to generate a SOAP kernel (r_c = 5 Å, ξ = 0.3, ζ = 2), forming the basis of the *ExChem* approach[180]. Since ML models only make reliable predictions for structures in or near the domain of the training data, the kernel was further reduced to include only the 491 molecules with ≥98% similarity to at least one compound in the experimental dataset. Since the molecular similarities between identical compounds, i.e., the diagonal of the SOAP kernel, is normalized to 1, a similarity of 98% would correspond to a value of 0.98 in the kernel matrix. The resulting SOAP kernel is a similarity matrix $\mathbf{K}_S$ with a dimension of 540×540. After dimensionality reduction by means of KPCA, two sets of compounds (I and II), containing five structures each, were selected from the resulting similarity map by the *ExChem* routine for experimental validation (see Figure 5.9). *ExChem* identifies molecules in a commercial database that exhibit structural features similar to a compound of interest in an individual experimental database. Provided a structure-activity relationship exists, the identified structures are expected to perform similarly in the experiment. The first set contains 1,4-benzenedimethanol (I.1), 1,4-bis(2-hydroxyethyl)piperazine (I.2), 4-amino-3-hydroxy-1-naphthalene- sulfonic acid (I.3), 1-pentanesulfonic acid (I.4) and 4-aminobutyric acid (I.5). The second set comprises 4-isopropylbenzoic acid (II.1), CAPS (II.2), gallic acid (II.3), triethanolamine (II.4) and 2-methoxybenzoic acid (II.5). On experimental validation, the previously untested electrolyte additives are added to the training set to increase the predictive power of the model and to broaden the model domain. Experimental validation of both test sets I and II was performed under the same conditions as for the original dataset of

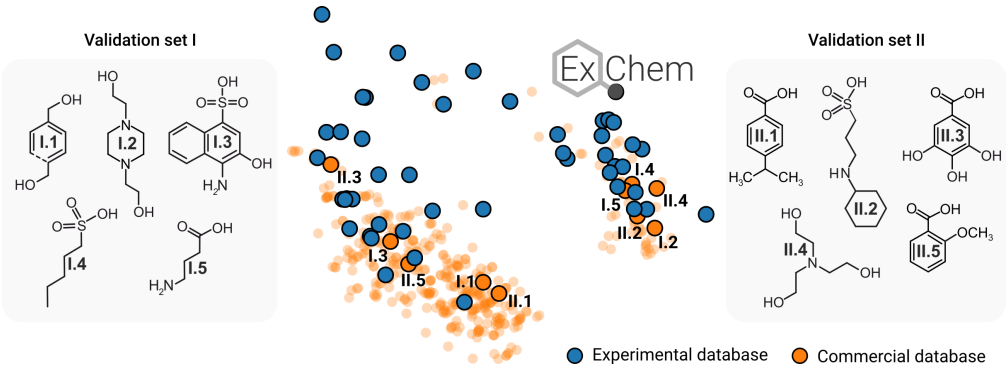


Figure 5.9: KPCA projection of the SOAP kernel containing the molecular similarities between 49 experimentally tested electrolyte additives and 491 compounds of a commercial database. Following the *ExChem* scheme, two sets of compounds (I and II), containing five structures each, were selected for experimental validation to assess the models' fitness for the active design of experiments.

49 electrolyte additives. Highly dissimilar structures may be selected on purpose to increase the domain of applicability of the model, using for example farthest point sampling[222]. Small training sets have a high chance of generating predictions for the electrolyte additives that differ from the experimental performance. Nevertheless, ADoE allows these models to be bootstrapped to screen larger regions of chemical space.

5.3.2 Prediction of Key Battery Parameters

In the following, two models (see Materials and Methods) are employed and referred to as S_K and D_K, according to the underlying molecular descriptors used for training. Both models were trained on the experimental database of 49 electrolyte additives and used to predict the DP and UE of 491 compounds from the commercial database, including the two sets of compounds selected for testing. The SOAP kernel of S_K was calculated based on the found hyperparameters in the associated hyperparameter tuning (see Table D.2). The hyperparameter tuning was performed by a 5-fold cross-validated grid search and resulted in an average validation error of 0.07 V vs. Ag/AgCl (DP) and 7% (UE) for S_K. For the prediction of test set II, test set I was included in the training data and the same set of hyperparameters was employed to recalculate the SOAP kernel. For the D_K model, the number of features n used for training was further reduced to include only the most discriminating molecular descriptors. Therefore, the principal covariates adaptation of CUR matrix decomposition (PCov-CUR) was employed.[222] By observing the training and validation error of 5-fold cross-validated models that were trained on a decreasing number of features, an optimal number of features of $n = 12$ could be identified (see Figure D.7). For DP, the validation error rises again for $n < 12$. In the case of UE, it is noteworthy that the

validation error for $n = 12$ is only slightly lower than for $n = 1$. The hyperparameter tuning for D_K resulted in an average validation error of 0.09 V vs. Ag/AgCl (DP) and 6% (UE). Analogously to S_K, hyperparameters and identified features were kept the same throughout retraining of the model to predict set II.

Aside from the predictive power of supervised ML, KPCovR also incorporates the pattern recognition capabilities of unsupervised ML. For each model, a similarity map is shown in Figure 5.10 and experimental and predicted DP and UE values are correlated. In contrast to pure KPCA, KPCovR naturally provides a two-dimensional projection of the underlying SOAP ($\mathbf{K}_S$) or distinct descriptor kernel ($\mathbf{K}_D$), already incorporating the target property. For the SOAP kernel, dimensionality reduction by KPCovR results in a 2D projection that clearly illustrates molecular similarities (see Figure 5.10a and c). Aromatic and aliphatic compounds are separated into a larger (top left) and smaller (bottom right) cluster, respectively. Although clusters with similar properties form in pure KPCA projections as well, the position of training structures appears more random with respect to the observed target property (see Figure 5.4a). In contrast, the KPCovR projection strongly illustrates structure-activity relationships, as compounds exhibiting a similar target map closely together. This allows exploration of the effects of specific structural motifs, rendering this approach highly interpretable. The advantage of KPCovR over pure KPCA is also evident for D_K, despite less distinct clusters. However, here the rather arcane nature of the features precludes using KPCovR projections for an intuitive interpretation of structure-activity relationships.

For the experimental and predicted target properties, a tolerance area around perfect correlation was defined, indicating a prediction accuracy for the ADoE approach that is considered as acceptable. Based on the experimental errors appearing in the training set, these tolerances were assumed as 0.1 V vs. Ag/AgCl and 10% for DP and UE, respectively. Compound II.1 was omitted from further consideration due to its unusual discharge behavior in the experiment that resulted in a very low mean DP of -0.74 V vs. Ag/AgCl (see discussion in Appendix C.2 and Figure C.2). Independent of the model, the predicted DP of I.3 and I.5, as well as the UE of I.4 were significantly underestimated and lay outside of the defined tolerance area. Compound I.5 was projected into unmapped areas of the respective similarity maps, providing an explanation for mispredictions of the associated DP and UE as the compound lies outside of the model domain. In general, these mispredictions can be ascribed to the fact that only few training samples represent the exceptional activity of these compounds in the experiment.

In the case of S_K, the DP and UE of compound I.2 are significantly over- and underpredicted, respectively. As I.2 is aliphatic and contains nitrogen, it was projected onto a cluster close to compounds exhibiting similar structural features. However, the pyrazine ring of I.2 strongly

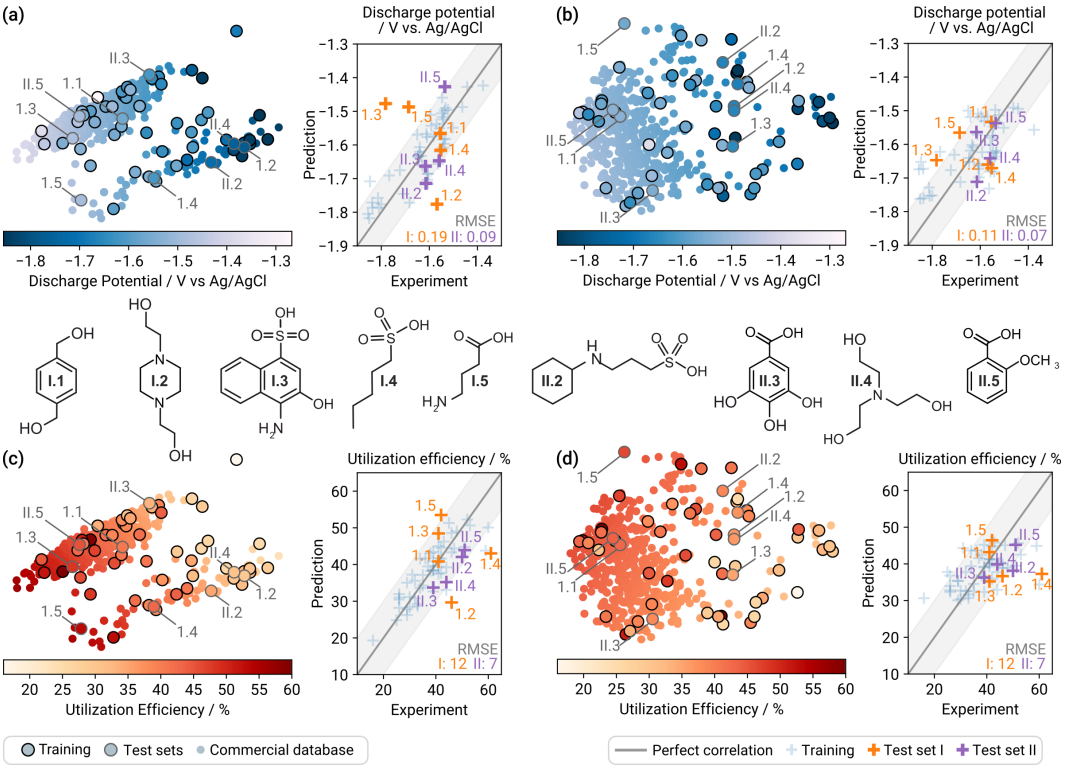


Figure 5.10: QSAR models based on KPCovR. (a) and (c) show the results of the SOAP-based model S_K for DP and UE, respectively. (b) and (d) show the results of the molecular descriptor-based model D_K for DP and UE, respectively. Black-rimmed dots in the scatter plots denote the original training set of 49 experimentally tested electrolyte additives. Gray-rimmed dots mark the predicted test sets (I and II) selected for experimental validation. No edging indicates a predicted structure from the commercial database that has not been further considered. The gray line in the regression parts depicts perfect correlation. The light gray area marks a deviation of 0.1 V vs. Ag/AgCl for DP and 10% for UE, considered acceptable for the predictions. Adapted from Figure 3 in [182].

differs from other structural features of compounds mapped to this region, thus explaining the observed deviation from the prediction.

Although most predictions of both models lie within the defined tolerance for an ADoE approach, S_K performs significantly better in describing the training data. Since the predictions lie in a larger interval compared to D_K, the range in the training data can be reflected more adequately. Unfortunately, the higher spread of the predictions also gives rise to larger deviations from the ground truth, implying a higher RMSE for DP. Although the prediction interval is considerably smaller for D_K, the RMSE does not change for UE and is even significantly lower for DP. This difference in RMSE can be mainly ascribed to the distribution of DP and UE in the measured test sets. As most measured values for I and II lie between -1.55 and -1.65 V vs. Ag/AgCl for DP and 40% and 50% for UE, they fall into the

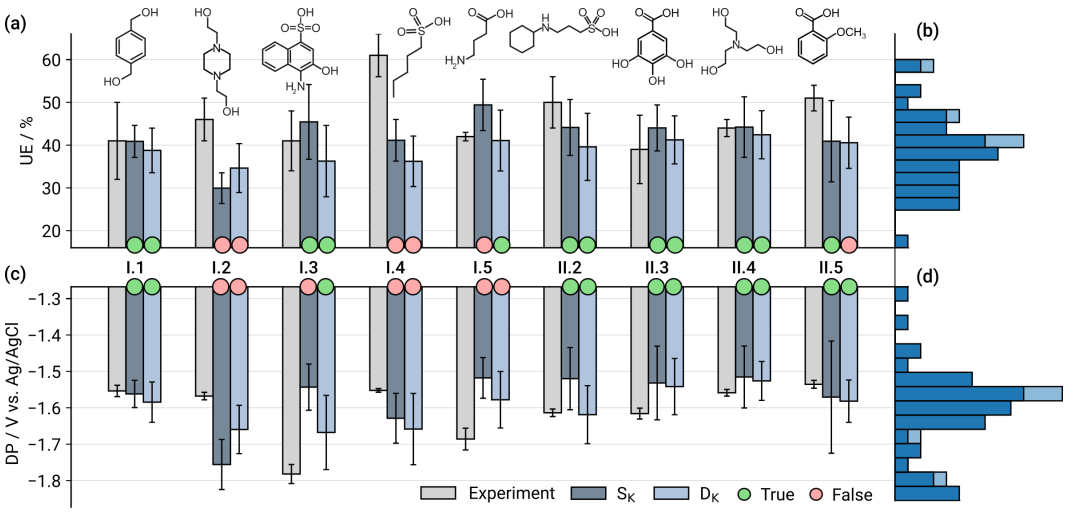


Figure 5.11: Ensemble predictions of S_K and D_K. Results are shown for the (a) utilization efficiency UE and the (c) discharge potential DP of test sets I and II. Respective distributions of underlying training data are shown in (b) and (d) for the according data range. Dark blue bars correspond to the original training dataset of 49 compounds (test set I predictions), whereas light blue bars illustrate the distribution of the training set extended by experimental values of test set I (test set II predictions). The colored circles illustrate if the model predictions overlap with experimental values. Adapted from Figure 4 in [182].

most populated bins in the histograms shown in Figure 5.11b and d. Naturally, a model with a small prediction interval implies a low RMSE, when the test data interval is also small.

5.3.3 Uncertainty Estimation

As the underlying regression technique of both models is inherently interpolative, the predictive performance strongly depends on how well the training set properly represents the local features in the system of interest. Particularly small numbers of training samples render estimates of error and uncertainty inevitable to assess the prediction accuracy and to effectively expand the model domain in active learning schemes. The uncertainty can be estimated based on the spread of predictions of an ensemble (committee) of independently trained ML models, e.g., by subsampling of the full training dataset (see Figure 5.11). Accordingly, a calibrated committee of $M = 50$ subsampled models was trained to obtain an estimate of the model predictions. For a more detailed description of the general procedure and the governing equations, the reader is referred to the Supplementary Information and the according literature.[70,226]

In case of the molecular descriptor-based model D_K, the feature selection process was repeated for each subsampled model. A distribution of molecular descriptors that appeared more than five times during subsampling is shown in Figure S2b. As was found in a previous

study[181], the applied feature selection method indicates that the Chemically Advanced Template Search (CATS) descriptors[227] are highly useful for prediction. These descriptors are generally known to be useful for data-driven drug discovery.[228] Additionally, the class of 3D-MoRSE (Molecular Representation of Structures based on Electron diffraction) were found to be relevant.[229,230] These are abbreviated as "Mor" and are a mathematical representation of XRD patterns weighted by different properties. None of the considered DFT-derived descriptors was found to have significant importance for the prediction of key battery parameters, contrary to the findings for the models that were developed to predict the corrosion inhibition efficiencies of small organic molecules.[16,181]

All committee predictions for the test sets I and II are presented in Figure 5.11a and c along with their associated uncertainties. The distributions of DP and UE in the underlying training data are illustrated in Figure 5.11b and d, respectively. Again, compound II.1 was omitted from further consideration due to issues already briefly discussed previously. Predictions that lie within the experimental error, or for which the error bars of predicted and measured values overlap, are assumed to be correctly predicted. For both models, the DP is overestimated for compounds I.2 and I.4 and underestimated for I.5. Furthermore, S_K significantly underestimates the DP of I.3. While the underestimation of I.3 and I.5 can be ascribed to the scarcity of available training data in this DP range, I.2 and I.4 are poorly predicted although a majority of training data exhibits a DP around -1.6 V vs. Ag/AgCl. However, more than half of the compounds in the training dataset do not contain sulfur nor nitrogen, which may explain errors in the prediction of test structures containing these species. Both S_K and D_K significantly underestimate the UE for compounds I.2 and I.4. Additionally, S_K slightly overestimates and D_K slightly underestimates the UE of I.5 and II.5, respectively. Relatively high UE values generally tend to be underestimated, which can be due to the scarcity of available training data in this region. It is noteworthy that the predictions of II.5 by S_K have high uncertainties.

Predictions for compounds containing only carbon, oxygen, and hydrogen are relatively accurate, particularly for compounds I.1 and II.3. Strikingly, most predictions for test set II are in acceptable agreement with the experimentally measured DP and UE values. In terms of RMSE, both models perform equally well when predicting UE, yielding errors of 12% (I) and 6% (II) for S_K and 12% (I) and 7% (II) for D_K. For predictions of DP, D_K performs significantly better than S_K, with 0.1 (I) and 0.04 (II) V vs. Ag/AgCl in contrast to 0.16 (I) and 0.07 (II) V vs. Ag/AgCl. However, both models generally do not allow quantitatively accurate predictions of the battery anode performance due to the limited dataset size, which leads to rather high uncertainties.

The effect of more training data is illustrated in Figure 5.12. Committee predictions for

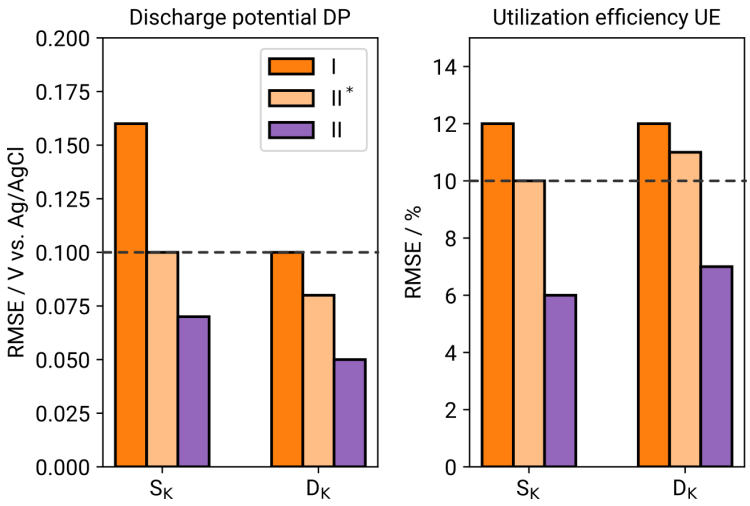


Figure 5.12: Bar plot representing the RMSE of both models S_K and D_K for the committee prediction of DP (left) and UE (right), determined for the blind test sets I (orange) and II (purple). II* (light orange) corresponds to predictions where the active learning step was omitted and blind test set I was not included for training the model. The dashed black line indicates the assumed acceptable tolerance based on the experimental error. Adapted from Figure 5 in [182].

test II improve by 0.03 V vs. Ag/AgCl and 4% for DP and UE, respectively, when test set I is included in the training set. Consequently, both models are expected to improve with more training data and a broader selection of compounds would allow improved mapping of molecular descriptors and local atomic features to the battery anode performance. Although both models lie well within the defined tolerance area, the more intuitive SOAP-based S_K model is expected to better adapt the ADoE principle for a reliable screening of the chemical space for effective battery additives.

5.3.4 Finding the Needle in the Haystack

A key part of the proposed ADoE workflow is the selection of new candidates for testing and retraining of the model after experimental validation. A selected compound should fulfill at least one of the three following requirements: (a) it should maximize the UE and negative DP, (b) it should increase the model's predictive power, and/or (c) it should widen the model's domain of applicability. For the second and third requirement, *ExChem* provides a valuable tool to extend the experimental training database with structures from known or uncharted areas of the chemical space. The trained models further allow the assignment of uncertainties to predicted compounds. Selecting structures with high uncertainties has good potential to significantly increase the predictive power of the model. The model performance can also be generally improved by balancing the number of compounds exhibiting high and low DPs

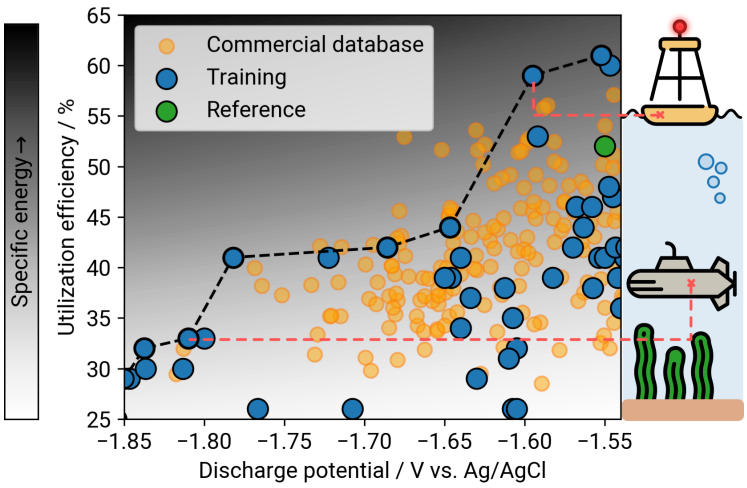


Figure 5.13: Scheme for a multi-objective optimization approach to identify electrolyte additives with a desired effect on the battery anode performance. A Pareto front (dashed line) is defined for the extended experimental training set (blue; original dataset of 49 electrolyte additives plus test set I), depicting the optimal trade-off between large (negative) DP and large UE. On the basis of the experimental data, DPs and UEs for the commercial database are predicted (orange). Compounds predicted close to or beyond the Pareto front are expected to enhance the battery anode performance for, e.g., sea buoys or AUVs. For better orientation, the background is colored according to the specific energy (SE), unifying the performance parameters DP and UE. Adapted from Figure 6 in [182].

and UEs in the training set. Finding compounds that maximize the DP and UE, however, cannot be accomplished by screening the chemical space using molecular similarities alone. Therefore, a multi-objective optimization approach was developed to identify electrolyte additives that are expected to boost the battery anode performance. Figure 5.13 shows the relation between DP and UE for the extended experimental training set (blue; 49 plus 5 electrolyte additives) and associated predictions of the commercial database as determined by the S_K model (orange). Compounds yielding optimal high (negative) DP and high UE constitute a Pareto front (the surface defining solutions of equal value that involve different trade-offs) for the experimental training data. Thus, new structures close to or beyond this Pareto front will provide a good trade-off between DP and UE. To facilitate the selection of potential candidates for further testing, the specific energy (SE) unifying DP and UE is used, which is defined for the given case as

$$\mathrm{SE} = \frac{\mathrm{DP} \cdot I \cdot t}{W_{\mathrm{act}}} = \frac{\mathrm{DP} \cdot \mathrm{UE} \cdot I \cdot t}{W_{\mathrm{theo}} \cdot 100\%} \tag{5.1}$$

Here, W_{theo} and W_{act} represent theoretical and actual weight loss of the anode, respectively, I is the applied current and t the discharge time. In Figure 5.13, the SE was calculated assuming a desired DP of at least -1.55 V vs. Ag/AgCl and a UE of at least 25%. As SE

is a measure of how much energy a battery contains in comparison to its weight, it is large for compounds exhibiting high DPs or UEs and is small for compounds with low values. A reasonable selection for experimental testing would lie close to or beyond the Pareto front in an area with a high SE. After experimental validation and retraining of the underlying ML model, the Pareto front is recalculated in an iterative manner while moving towards new boundaries within the ADoE workflow, thus identifying electrolyte additives with improved properties.

Future work involves increasing the number of iterations in the ADoE routine to further improve the model accuracy. In this regard, the SOAP-based KPCovR model combined with *ExChem* allows searching for specific structures that increase the model's domain of applicability and/or significantly enhance the performance of Mg batteries.

Chapter 6

Magnesium Corrosion Engineering: the Present and the Future

> If I have seen further it is by standing on the shoulders of giants.
>
> — Isaac Newton

Magnesium has high capability of becoming the engineering material of the future due to its light weight and characteristic corrosion properties. However, controlling the degradation behavior of magnesium is crucial to unlocking its full potential with respect to applications in the transport, medical and energy sector. In the past chapters, magnesium corrosion engineering was regarded from different perspectives, focusing a computational point of view.

6.1 Towards an Atomistic Picture of Magnesium Corrosion

First of all, a deeper mechanistic understanding of hydrogen evolution on the magnesium surface was obtained. Based on density functional theory (DFT) calculations, the two predominant hydrogen evolution reaction pathways, i.e., the Tafel and Heyrovský reaction, have been investigated using the climbing image nudged elastic band method (NEB) in combination with concurrent Bader charge analyses of each considered image. Prerequisite to the two considered reaction schemes is the dissociation of a water molecule in the Volmer step, which leads to adsorption of a hydroxide and a hydride (H^-) upon the Mg surface. The results of the NEB calculations indicate that the energy barrier for the formation of hydrogen is lower for the Heyrovský than for the Tafel pathway, confirming the Volmer–Heyrovský reaction cascade as the predominant mechanism that leads to the formation of hydrogen during the degradation of magnesium. Analysis of the occurring charge transfers using the developed tool `baderVis` confirmed the existence of adsorbed hydrides that explain the

observed energy barriers. Furthermore, the comprehensive analysis of the occurring energy barriers revealed that the Volmer reaction is the rate-limiting step in hydrogen evolution. In the same manner, the charge state and energy barriers of a dissolving Mg ion was investigated in detail. The findings strongly contradict the existence of a stable Mg^+ species involved in the water dissociation process, a mechanistic aspect, which has been controversially discussed in the literature over the course of the last decade. Concomitantly, the obtained results indicate involvement of co-adsorbed species that facilitate the dissolution of Mg. Adapting the underlying model to be more realistic, it was found that indeed co-adsorbed hydroxide ions take part in the dissolution process and decrease the dissolution barrier. Noble impurities are expected to have a significant impact on the Mg dissolution barrier and should be object of further investigations in the future. Furthermore, unraveling the mechanism responsible for anomalous hydrogen evolution requires constant potential methods that explicitly treat anodic overpotentials in the simulation. A promising approach was developed by Surendralal et al., where a computational Ne electrode is doped with fractional electronic charges, thus allowing to control the applied potential and constructing an electrolytic electrochemical cell within the DFT framework.[43]

Successful magnesium corrosion engineering depends on a deep understanding of the underlying degradation mechanisms. Identifying major impact factors enables active tuning of corrosion barriers, thus tailoring the degradation rate to the application at hand. The new insights into the Mg corrosion mechanisms obtained in this thesis allow for more effective approaches towards corrosion control. For instance, by knowing more about the exact rate-limiting reaction steps, specific catalysts can be found that, e.g., generally accelerate the corrosion in Mg battery applications. In contrast, associated reaction barriers can be increased by involving co-adsorbed species to significantly lower the corrosion rate.

6.2 *Modus Operandi* of Magnesium Dissolution Modulators

Apart from conservative strategies to prevent magnesium degradation, numerous other approaches exist that enable corrosion inhibition and control. A highly promising approach is the introduction of magnesium dissolution modulators, small organic molecules that influence the Mg degradation rate. These modulators of corrosion either stabilize or destabilize the degrading surface via adsorption, or prevent corrosive species, such as noble impurities released in the corrosion process, from adsorbing upon the surface. The latter case is mainly defined by two mechanisms: blocking the surface through adsorption or capturing corrosive species in solution via complex formation. Atomistic simulations are a valuable tool to better understand these mechanisms on a fundamental level. In combination with experimental

databases of corrosion inhibitors, it is possible to find potential relationships between ab initio calculations and the corrosion inhibition efficiency (IE) of such compounds. In order to identify statistically significant connections, however, a large number of calculations is required.

Investigating the adsorption mechanism of over 40 compounds upon the hydroxylated magnesium oxide surface showed an apparent relationship between adsorption energy and inhibition efficiency. Nevertheless, the rather low correlation value indicates that the dissolution modulating performance also depends on other parameters. The large number of investigated molecules constitutes the beginning of a database that promises to give further insights into the relationship between inhibitor adsorption and corrosion inhibition. The presented results indicate that heteroatoms, such as nitrogen or carboxyl groups, play a significant role for the corrosion inhibition efficiency. By connecting the adsorption structure and inhibition efficiencies of an increasing number of compounds using quantitative structure-activity relationship (QSAR) models, this apparent correlation can be further elucidated in the future to more easily identify effective magnesium dissolution modulators based on simulation results. Modeling complex formation is not as straightforward. A large configuration space aggravates the generation of complex formation energies in large numbers. However, automated workflows based on enhanced sampling methods promise to generate quantities, e.g., complex formation and adsorption energies, that can be correlated with experimental corrosion control metrics in a high-throughput manner. Such workflows enable high-throughput DFT calculations that can be used as input for data-driven predictive models to estimate the corrosion inhibition efficiency.

Figure 6.1 illustrates two design drafts for potential automated workflows that qualitatively show the general simulation process. Both workflows have in common that they require input in form of compounds and/or substrates of interest, as well as water and ion models to resemble the service environment. Using the open-source software `PLUMED`[56], the system can be described with one or more collective variables (CVs) that define the sampling space for metadynamics (cf. Section 2.2). In both cases, the output is given along these CVs as free energy profiles from which the quantity of interest can be derived.

Calculating a free energy profile from first principles, in for example `CP2k`[34], has some advantages, but also disadvantages, as compared to using molecular dynamics (MD) codes such as `GROMACS`[231] with semi-empirical force fields. Although the calculated energies are presumably more accurate, performing metadynamics simulations in the DFT framework demands immense computational resources that may not be available, particularly for high-throughput calculations, where numerous compounds are considered. Also finite size effects may have a negative impact on the simulation. On the other hand, many force fields for

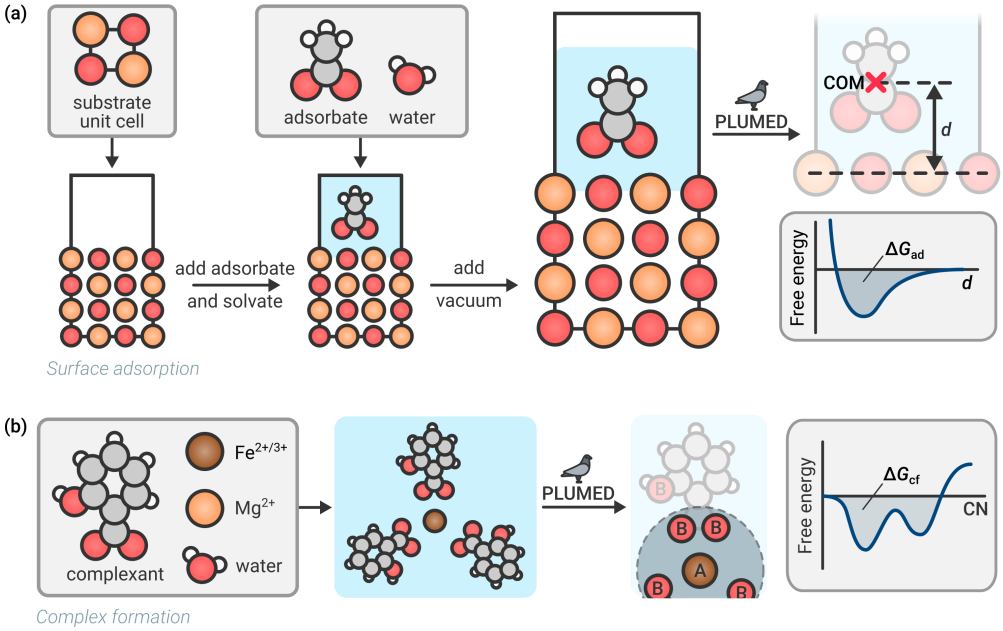


Figure 6.1: Automated workflow schemes to calculate free energies with metadynamics. (a) Free energy of adsorption. (b) Free energy of complex formation.

semi-empirical MD do not account for chemical reactions such as proton transfers, which are likely to occur in adsorption or complex formation processes. Therefore, the user must assess in advance which compromises and simplifications are tolerable for the particular simulation case, and which are not.

When calculating the free energy of adsorption for a substrate and adsorbate of interest, the simulation setup can easily be automated (Figure 6.1a). The substrate unit cell is replicated in all room directions until the resulting supercell is large enough to provide sufficient surface area for adsorption. Additionally, a sufficiently large vacuum region is added in surface direction to create a surface slab. After inserting the adsorbate of interest into this region, the remaining space is solvated with water and around 5 Å of vacuum is added above to prevent interaction with the periodic image. The CV can be easily defined as the distance d between the top surface layer and the center of mass (COM). During the metadynamics simulation, the distance between adsorbate and surface is changed continuously. At passed states, a positive Gaussian potential is added to the real energy landscape of the system, thus discouraging the simulation from passing this state again. The free energy of adsorption ΔG_{ad} can be directly derived from the resulting free energy profile.[232,233] Given enough distance from the surface, the free energy converges to a constant value in the bulk region of the solvent. After shifting

the free energy profile to be zero in the bulk region, ΔG_{ad} can be calculated as

$$\Delta G_{\mathrm{ad}} = -k_{\mathrm{B}}T \ln\left(\frac{1}{\delta}\int_{r_{\mathrm{c}}}^{r_{\mathrm{c}}+\delta} \mathrm{d}z e^{-F(z)/k_{\mathrm{B}}T}\right), \tag{6.1}$$

where r_{c} is the distance of the closest approach of the molecule to the surface and $r_{\mathrm{c}} + \delta$ the point where the solvent can be considered as bulk.[234]

Automated calculation of the free energy of complex formation uses the molecular structures of potential complexants and central ions as initial input (Figure 6.1b). According to the charge of the central ion and the expected charge of the formed complex, the number of complexants considered for the simulation is determined. All molecular structures are inserted into a sufficiently large simulation cell, such that central ion and complexants lie close to each other. Subsequently, the cell size is further increased and the simulation cell solvated with water. The CV is defined as the coordination number (CN) of the central ion. In `PLUMED`, the CN is calculated between two groups, which have to be defined in beforehand. Here, group A corresponds to the central ion, whereas group B contains all atoms within the complexants that potentially take part in the coordination (e.g., oxygen or nitrogen). By using the CN as CV, the simulation passes numerous coordination configurations, yielding the respective free energy profile as output. Assuming no coordination of the central ion as reference (CN = 0), the free energy of complex formation, ΔG_{cf}, can be determined as the area underneath the curve.

Both workflows enable calculation of IE-related properties in a high-throughput manner. The required inputs can be easily provided, however, when using classical MD (e.g., for large systems or limited computational resources) adequate force fields are crucial to accurately describe the interaction between compounds and substrate or central ion. It is noteworthy that the workflows have not been implemented yet and only exist qualitatively so far. Nevertheless, MD codes such as `GROMACS` come with a whole arsenal of useful functions that significantly reduce the implementation effort of the proposed workflows. The molecular topologies required within these workflows can be easily generated automatically using high-throughput frameworks, such as `STaGE`.[235]

6.3 Computer-Aided Design of Experiments

Particularly for data-driven models, quantities such as the adsorption energy can be of high value. Using a multitude of theoretical molecular descriptors, predictive models based on machine learning techniques were built that allow to accurately estimate the inhibition efficiency of yet untested compounds. Such machine learning models can also

be based on structural similarities as determined by smooth overlap of atomic positions (SOAP), a molecular fingerprint that incorporates the structural features of a molecule. Based on comprehensive experimental databases, QSAR models were trained to predict the performance of unknown molecular structures as corrosion inhibitor or electrolyte additive in Mg batteries. Comparison between molecular descriptor-based and SOAP-based models showed similar predictive power. When predicting the inhibition efficiency of corrosion inhibitors for commercial purity Mg, both approaches make qualitatively correct predictions. When predicting the key performance parameters of yet untested electrolyte additives, the predictions of both schemes are in an acceptable range. However, it has to be mentioned that the training database in the latter case was significantly smaller.

A great advantage of the SOAP kernel is that it allows to generate two-dimensional structure-activity landscapes that vividly illustrate the similarity between chemical compounds. Moreover, correlating these maps with properties of interest, such as the corrosion inhibition efficiency or discharge potential of a battery, allows to easily identify structure-activity relationships. Accordingly, such structure-activity relationship maps were created for numerous tested and potential corrosion inhibitors and electrolyte additives. To further assess the predictive power of the trained models, additional untested compounds were selected for blind testing, based on their relative position to observed clusters. The results confirm that the models can qualitatively predict the impact of small organic molecules on the corrosion and discharge performance of certain magnesium-based materials, as the predicted values are in good agreement with the experimental values. Clustering the data already in high-dimensional space, for example using the probabilistic analysis of molecular motifs (PAMM) approach, could increase the accuracy of cluster definitions and thus the qualitative predictive performance of the developed model in the future.[201] Additionally, the validation results indicate that solubility is an additional important factor that needs to be considered in the selection process of target chemicals, where such data is available. Furthermore, using molecular structures closer to their "mode of action", i.e., solvated, at correct pH or in complex formations, could significantly improve the model performance. Here, especially the proposed high-throughput atomistic simulations (Figure 6.1) should provide useful model input.

Elucidating structure-activity relationships within a given dataset is further facilitated by employing methods that combine the pattern recognition capabilities and predictive power of unsupervised and supervised learning, respectively. When used with SOAP, kernel principal covariates regression (KPCovR) generates structure-activity landscapes that—aside from molecular similarities—incorporate the target property in the low-dimensional projection (cf. Sections 2.3 and 5.3). Simultaneously, the information loss and error in predicting the

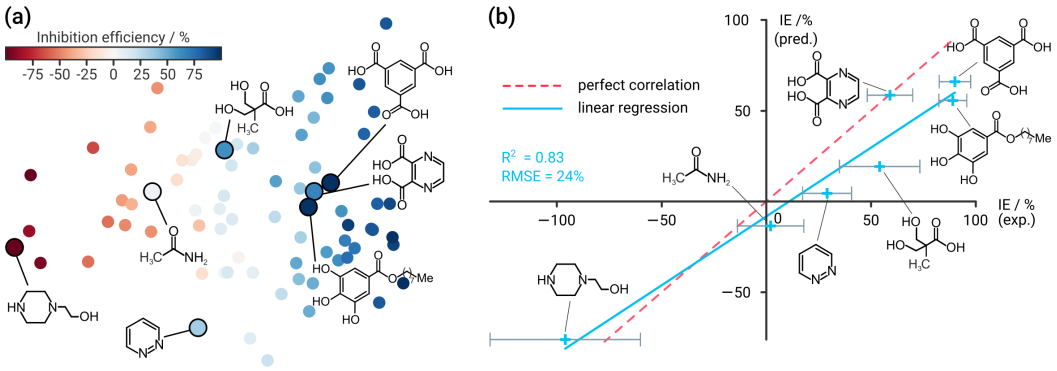


Figure 6.2: Kernel Principal Covariates Regression (KPCovR). KPCovR combines kernel principal components analysis (KPCA) and kernel ridge regression (KRR). (a) The KPCovR projection as derived from molecular similarities and inhibition efficiencies (IEs) of 71 training and 7 test compounds (colored dots). In contrast to pure KPCA, KPCovR illustrates structure-activity relationship more intuitively, since the target property is incorporated in the projection. (b) Predicted test set IEs obtained from the KPCovR model. The blue line is a linear least square fit of the predicted and measured values. The RMSE value is in absolute percent. The red dashed line represents ideal correlation. The error bars depict the standard deviation of the experimentally derived IE.

target property are minimized using only the resulting latent space vectors. The potential of KPCovR in the creation of QSAR models becomes apparent when directly compared to models based on individual unsupervised and supervised learning techniques. Respectively training and testing a KPCovR model on the same 71 training and 7 test structures as in Figure 5.4 yields significantly higher predictive performance than the decoupled approach using sketch-map and kernel ridge regression (KRR). Although the two-dimensional projection does not purely reflect the molecular similarities anymore, incorporating the inhibition efficiency leads to a more intuitive exploration of potential structure-activity relationships (Figure 6.2). For the given case, employing a KPCovR model also leads to a general improvement of the quantitative prediction accuracy for the seven test structures with an $R^2 = 0.83$ (KRR: 0.79) and RMSE = 24% (KRR: 36%). However, it is noteworthy that an improvement of the regression performance using KPCovR is not necessarily the case.[75]

These modeling methods constitute a promising way to rapidly identify the most promising molecules for specific target applications that have reduced toxicity, environmental impact, and cost. However, experimental validation remains an important final step, especially for chemicals in unlabeled regions of the structure-activity landscape. Expanding the size, diversity, and quality of data used to train these models is a key task to improve their robustness and domain of applicability. For that reason, the Python-based workflow *ExChem* (www.exchem.de) has been developed for automated screening of large molecular databases, using molecular similarities. In a combination of active learning and design

of experiments, *ExChem* is a great asset to suggest compounds for further testing that are similar or dissimilar to the given training database, thus either identifying potent structures, improving the model accuracy or effectively increasing the models' domain of applicability. Naturally, the presented concept is not limited to Mg and can be adopted to explore the structure-activity landscape of, e.g., Al-, Cu- and steel-based materials in a similar fashion, given a corresponding experimental database exists to train the model. The developed machine learning-based strategies facilitate an intuitive and fast screening of large databases to identify similar compounds, simplifying the search for compounds with potentially useful properties, and dramatically decreasing the time and resources required relative to those of experimental discovery methods. By implementing the *ExChem* routine in an electronic lab notebook it could become exceedingly attractive for experimentalists. A given experimental database could be continuously extended by new experiments while the prediction accuracy of the underlying machine learning model would increase simultaneously, thus helping to identify, e.g., potent corrosion inhibitors.

The results presented in this thesis provide an insightful contribution to the realm of Mg corrosion engineering and can inspire novel approaches to gain control over the degradation behavior of Mg-based materials. Current magnesium corrosion research, however, still raises unanswered questions, such as the exact mechanism of anodic hydrogen evolution. Although experimental research is indispensable to gain a deeper understanding and control of the underlying corrosion mechanisms, computational schemes proof themselves as necessary addition to obtain a holistic view of Mg degradation. Particularly experimental design benefits from atomistic simulations and data-driven prediction models, as they can significantly decrease time investment, effort, resources and costs.

The continuous increase of available computational resources allows to consider exceedingly complex models. However, it is easy to create a model that consumes immense computational resources but brings no added value. Hence, a careful design of simulations is as important as for experiments. A possible approach to increase model size and simultaneously reduce computational effort is the development of reactive or machine learning force fields that accurately describe the interaction of magnesium, magnesium oxide and magnesium hydroxide with water, organic molecules and noble impurities.[236–238] Particularly for the proposed automated workflows employing enhanced sampling methods, such force fields would enable high-throughput calculations of additional parameters (e.g., free energy of adsorption or complex formation energy) that can be linked to the degradation properties. These parameters promise to further improve the accuracy of data-driven models. Moreover, they can serve as valuable input for multi-scale approaches.[239–241]

The importance of computational approaches to magnesium corrosion engineering is

likely to continue to grow. In a future scenario, the corrosion characteristics of magnesium will have been completely understood experimentally and theoretically, thus unraveling the full potential of magnesium-based materials in modern engineering applications. Once this is achieved, magnesium will become an indispensable material for a sustainable future that will help to combat the ongoing climate change and will hopefully improve the life of all people living on our planet.

BIBLIOGRAPHY

[1] W. H. Organization. *COP24 special report: health and climate change*. World Health Organization, 2018.

[2] F. Creutzig, P. Jochem, O. Y. Edelenbosch, L. Mattauch, D. P. Van Vuuren, D. McCollum, and J. Minx. Transport: A roadblock to climate change mitigation? *Science*, 350 (6263):911–912, 11 2015.

[3] A. Vadiraj, M. Abraham, and A. S. Bharadwaj. Trends in Automotive Light Weighting. In A. Gokhale, N. Prasad, and B. Basu, editors, *Light Weighting for Defense, Aerospace, and Transportation.*, pages 89–102. Indian Institute of Metals Series, Singapore, 2019.

[4] D. Gielen, F. Boshell, and D. Saygin. Climate and energy challenges for materials science. *Nature Materials*, 15(2):117–120, 2 2016.

[5] A. Dreyer, A. Feld, A. Kornowski, E. D. Yilmaz, H. Noei, A. Meyer, T. Krekeler, C. Jiao, A. Stierle, V. Abetz, H. Weller, and G. A. Schneider. Organically linked iron oxide nanoparticle supercrystals with exceptional isotropic mechanical properties. *Nature Materials*, 15(5):522–528, 2 2016.

[6] A. I. Taub and A. A. Luo. Advanced lightweight materials and manufacturing processes for automotive applications. *MRS Bulletin*, 40(12):1045–1053, 11 2015.

[7] Z. Ma, D. R. MacFarlane, and M. Kar. Mg Cathode Materials and Electrolytes for Rechargeable Mg Batteries: A Review. *Batteries & Supercaps*, 2(2):115–127, 2 2019.

[8] D. Höche, S. V. Lamaka, B. Vaghefinazari, T. Braun, R. P. Petrauskas, M. Fichtner, and M. L. Zheludkevich. Performance boost for primary magnesium cells using iron complexing agents as electrolyte additives. *Scientific Reports*, 8(1):1–9, 12 2018.

[9] R. Shah, V. Mittal, E. Matsil, and A. Rosenkranz. Magnesium-ion batteries for electric vehicles: Current trends and future perspectives. *Advances in Mechanical Engineering*, 13(3):1–9, 2021.

[10] J. E. Gray and B. Luan. Protective coatings on magnesium and its alloys - A critical review. *Journal of Alloys and Compounds*, 336(1-2):88–113, 4 2002.

[11] C. Blawert, W. Dietzel, E. Ghali, and G. Song. Anodizing treatments for magnesium alloys and their effect on corrosion resistance in various environments. *Advanced Engineering Materials*, 8(6):511–533, 6 2006.

[12] Z. Jia, P. Xiong, Y. Shi, W. Zhou, Y. Cheng, Y. Zheng, T. Xi, and S. Wei. Inhibitor encapsulated, self-healable and cytocompatible chitosan multilayer coating on biodegradable Mg alloy: A pH-responsive design. *Journal of Materials Chemistry B*, 4(14):2498–2511, 3 2016.

[13] S. V. Lamaka, D. Höche, R. P. Petrauskas, C. Blawert, and M. L. Zheludkevich. A new concept for corrosion inhibition of magnesium: Suppression of iron re-deposition. *Electrochemistry Communications*, 62:5–8, 1 2016.

[14] S. V. Lamaka, B. Vaghefinazari, D. Mei, R. P. Petrauskas, D. Höche, and M. L. Zheludkevich. Comprehensive screening of Mg corrosion inhibitors. *Corrosion Science*, 128:224–240, 2017.

[15] T. Würger, C. Feiler, F. Musil, G. B. V. Feldbauer, D. Höche, S. V. Lamaka, M. L. Zheludkevich, and R. H. Meißner. Data Science Based Mg Corrosion Engineering. *Frontiers in Materials*, 6:53, 4 2019.

[16] C. Feiler, D. Mei, B. Vaghefinazari, T. Würger, R. H. Meißner, B. J. Luthringer-Feyerabend, D. A. Winkler, M. L. Zheludkevich, and S. V. Lamaka. In silico screening of modulators of magnesium dissolution. *Corrosion Science*, 163:108245, 2 2020.

[17] D. Mei, S. V. Lamaka, C. Feiler, and 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. *Corrosion Science*, 153:258–271, 6 2019.

[18] C. Kittel. *Introduction to Solid State Physics*. Wiley, 2004.

[19] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules (International Series of Monographs on Chemistry)*. Oxford University Press, USA, 1994.

[20] M. Ceriotti, G. A. Tribello, and M. Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences*, 108(32):13023–13028, 8 2011.

[21] M. Ceriotti, G. A. Tribello, and M. Parrinello. Demonstrating the transferability and the descriptive power of sketch-map. *Journal of Chemical Theory and Computation*, 9(3):1521–1532, 3 2013.

[22] M. Ceriotti. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *Journal of Chemical Physics*, 150(15):150901, 2019.

[23] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B): B864, 11 1964.

[24] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 11 1965.

[25] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 10 1996.

[26] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der Waals density functional for general geometries. *Physical Review Letters*, 92(24): 246401, 6 2004.

[27] J. Klimeš, D. R. Bowler, and A. Michaelides. Chemical accuracy for the van der Waals density functional. *Journal of Physics Condensed Matter*, 22(2):022201, 1 2010.

[28] J. P. Perdew. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conference Proceedings*, 577(1):1–20, 8 2003.

[29] A. P. Bartók and J. R. Yates. Regularized SCAN functional. *Journal of Chemical Physics*, 150(16):1–6, 2019.

[30] M. Shishkin and G. Kresse. Implementation and performance of the frequency-dependent GW method within the PAW framework. *Physical Review B - Condensed Matter and Materials Physics*, 74(3):035101, 7 2006.

[31] N. A. Besley. Modeling of the spectroscopy of core electrons with density functional theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(6): 1–22, 2021.

[32] J. Leszczynski, A. Kaczmarek-Kedziera, T. Puzyn, M. G. Papadopoulos, H. Reis, and M. K. Shukla. *Handbook of Computational Chemistry*. Springer International Publishing, Cham, 1 2017.

[33] L. Zheng, A. A. Alhossary, C. K. Kwoh, and Y. Mu. Molecular dynamics and simulation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1-3(V):550–566, 2018.

[34] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter. CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations. *Journal of Chemical Physics*, 152(19): 194103, 2020.

[35] T. D. Kühne. Second generation Car-Parrinello molecular dynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(4):391–406, 7 2014.

[36] T. D. Kühne and E. Prodan. Disordered crystals from first principles I: Quantifying the configuration space. *Annals of Physics*, 391:120–149, 2018.

[37] J. Kolafa. Time-Reversible Always Stable Predictor-Corrector Method for Molecular Dynamics of Polarizable Molecules. *Journal of Computational Chemistry*, 25(3): 335–342, 2 2004.

[38] R. Schlögl. Heterogeneous catalysis. *Angewandte Chemie - International Edition*, 54 (11):3465–3520, 3 2015.

[39] D. Seifzadeh, A. Bezaatpour, A. N. Shamkhali, and H. Basharnavaz. Experimental and Theoretical Studies to Examine the Inhibition Effect of a Schiff Base Against Magnesium Corrosion. *Transactions of the Indian Institute of Metals*, 69(8):1545–1555, 10 2016.

[40] J. A. Yuwono, N. Birbilis, K. S. Williams, and N. V. Medhekar. Electrochemical stability of magnesium surfaces in an aqueous environment. *Journal of Physical Chemistry C*, 120(47):26922–26933, 2016.

[41] J. A. Yuwono, N. Birbilis, R. Liu, Q. Ou, Q. Bao, and N. V. Medhekar. Aqueous electrochemical activity of the Mg surface: The role of group 14 and 15 microalloying elements. *Journal of the Electrochemical Society*, 164(13):C918–C929, 11 2017.

[42] S. Surendralal, M. Todorova, and J. Neugebauer. Impact of Water Coadsorption on the Electrode Potential of H-Pt(1 1 1)-Liquid Water Interfaces. *Physical Review Letters*, 126(16):166802, 4 2021.

[43] S. Surendralal, M. Todorova, M. W. Finnis, and J. Neugebauer. First-Principles Approach to Model Electrochemical Reactions: Understanding the Fundamental Mechanisms behind Mg Corrosion. *Physical Review Letters*, 120(24):246801, 6 2018.

[44] P. Atkins, D. P. Julio, and J. Keeler. *Atkins' Physical Chemistry*. Oxford University Press, 2018.

[45] F. Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, Ltd, 2017.

[46] G. Henkelman, G. Jóhannesson, and H. Jónsson. Methods for Finding Saddle Points and Minimum Energy Paths. In *Theoretical Methods in Condensed Phase Chemistry*, pages 269–302. Kluwer Academic Publishers, Dordrecht, 2005.

[47] D. Sheppard, R. Terrell, and G. Henkelman. Optimization methods for finding minimum energy paths. *The Journal of Chemical Physics*, 128(13):134106, 4 2008.

[48] H. Jónnson, G. Mills, and K. W. Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*, chapter 16, pages 385–404. WORLD SCIENTIFIC, 6 1998.

[49] G. Henkelman and H. Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *Journal of Chemical Physics*, 113(22):9978–9985, 2000.

[50] G. Henkelman, B. P. Uberuaga, and H. Jónsson. Climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 113(22):9901–9904, 2000.

[51] G. H. Vineyard. Frequency factors and isotope effects in solid state rate processes. *Journal of Physics and Chemistry of Solids*, 3(1-2):121–127, 1957.

[52] H. Eyring. The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2):63–71, 11 1935.

[53] A. F. Voter and J. D. Doll. Transition state theory description of surface self-diffusion: Comparison with classical trajectory results. *The Journal of Chemical Physics*, 80 (11):5832–5838, 8 1984.

[54] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12562–12566, 10 2002.

[55] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2), 2008.

[56] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180(10):1961–1972, 10 2009.

[57] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 12 2019.

[58] A. Mauri, V. Consonni, and R. Todeschini. Molecular descriptors. *Handbook of Computational Chemistry*, pages 2065–2093, 1 2017.

[59] S. Yousefinejad and B. Hemmateenejad. Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemometrics and Intelligent Laboratory Systems*, 149: 177–204, 2015.

[60] R. Todeschini and V. Consonni. *Molecular Descriptors for Chemoinformatics*, volume 41 of *Methods and Principles in Medicinal Chemistry*. Wiley, 7 2009.

[61] F. Grisoni, D. Ballabio, R. Todeschini, and V. Consonni. Molecular descriptors for structure–activity applications: A hands-on approach. *Methods in Molecular Biology*, 1800:3–53, 2018.

[62] R. Todeschini, M. Lasagni, and E. Marengo. New molecular descriptors for 2D and 3D structures. Theory. *Journal of Chemometrics*, 8(4):263–272, 7 1994.

[63] R. Todeschini and P. Gramatica. The Whim Theory: New 3D Molecular Descriptors for Qsar in Environmental Modelling. *SAR and QSAR in Environmental Research*, 7 (1-4):89–115, 12 1997.

[64] V. Consonni, R. Todeschini, M. Pavan, and P. Gramatica. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *Journal of Chemical Information and Computer Sciences*, 42(3):693–705, 5 2002.

[65] V. Consonni, R. Todeschini, and M. Pavan. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *Journal of Chemical Information and Computer Sciences*, 42 (3):682–692, 5 2002.

[66] G. Cruciani, M. Pastor, and W. Guba. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *European Journal of Pharmaceutical Sciences*, 11 (SUPPL. 2):S29–S39, 10 2000.

[67] R. D. Cramer, D. E. Patterson, and J. D. Bunce. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *Journal of the American Chemical Society*, 110(18):5959–5967, 1988.

[68] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, 18(20): 13754–13769, 5 2016.

[69] A. P. Bartók, R. Kondor, and G. Csányi. On representing chemical environments. *Physical Review B - Condensed Matter and Materials Physics*, 87(18):184115, 5 2013.

[70] F. Musil, S. De, J. Yang, J. E. Campbell, G. M. Day, and M. Ceriotti. Machine learning for the structure-energy-property landscapes of molecular crystals. *Chemical Science*, 9(5):1289–1300, 1 2018.

[71] G. A. Tribello, M. Ceriotti, and M. Parrinello. Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 109(14):5196–5201, 4 2012.

[72] S. Ayesha, M. K. Hanif, and R. Talib. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 59:44–58, 7 2020.

[73] S. P. Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[74] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C):53–65, 11 1987.

[75] B. A. Helfrecht, R. K. Cersonsky, G. Fraux, and M. Ceriotti. Structure-property maps with Kernel principal covariates regression. *Machine Learning: Science and Technology*, 1(4):045021, 10 2020.

[76] M. Esmaily, J. E. Svensson, S. Fajardo, N. Birbilis, G. S. Frankel, S. Virtanen, R. Arrabal, S. Thomas, and L. G. Johansson. Fundamentals and advances in magnesium alloy corrosion. *Progress in Materials Science*, 89:92–193, 8 2017.

[77] G. L. Song and A. Atrens. Corrosion mechanisms of magnesium alloys. *Advanced Engineering Materials*, 1(1):11–33, 1999.

[78] J. A. Yuwono, N. Birbilis, C. D. Taylor, K. S. Williams, A. J. Samin, and N. V. Medhekar. Aqueous electrochemistry of the magnesium surface: Thermodynamic and kinetic profiles. *Corrosion Science*, 147:53–68, 10 2019.

[79] J. A. Yuwono, C. D. Taylor, G. S. Frankel, N. Birbilis, and S. Fajardo. Understanding the enhanced rates of hydrogen evolution on dissolving magnesium. *Electrochemistry Communications*, 104:106482, 6 2019.

[80] H. Ke and C. D. Taylor. Density Functional Theory: An Essential Partner in the Integrated Computational Materials Engineering Approach to Corrosion. *Corrosion*, 75(7):708–726, 5 2019.

[81] P. F. King. The Role of the Anion in the Anodic Dissolution of Magnesium. *Journal of the Electrochemical Society*, 113(6):536–539, 6 1966.

[82] P. F. King. Magnesium as a Passive Metal. *Journal of the Electrochemical Society*, 110(11):1113–1116, 11 1963.

[83] J. L. Robinson and P. F. King. Electrochemical Behavior of the Magnesium Anode. *Journal of The Electrochemical Society*, 108(1):36, 1 1961.

[84] N. Birbilis, A. D. King, S. Thomas, G. S. Frankel, and J. R. Scully. Evidence for enhanced catalytic activity of magnesium arising from anodic dissolution. *Electrochimica Acta*, 132:277–283, 6 2014.

[85] D. Höche, C. Blawert, S. V. Lamaka, N. Scharnagl, C. Mendis, and M. L. Zheludkevich. The effect of iron re-deposition on the corrosion of impurity-containing magnesium. *Physical Chemistry Chemical Physics*, 18(2):1279–1291, 12 2016.

[86] A. Atrens, G. L. Song, M. Liu, Z. Shi, F. Cao, and M. S. Dargusch. Review of recent developments in the field of magnesium corrosion. *Advanced Engineering Materials*, 17(4):400–453, 4 2015.

[87] A. D. Atrens, I. Gentle, and A. Atrens. Possible dissolution pathways participating in the Mg corrosion reaction. *Corrosion Science*, 92:173–181, 3 2015.

[88] A. Atrens, G. L. Song, F. Cao, Z. Shi, and P. K. Bowen. Advances in Mg corrosion and research suggestions. *Journal of Magnesium and Alloys*, 1(3):177–200, 2013.

[89] E. Skúlason. Modeling electrochemical reactions at the solid-liquid interface using density functional calculations. *Procedia Computer Science*, 51(1):1887–1896, 1 2015.

[90] E. Skúlason, G. S. Karlberg, J. Rossmeisl, T. Bligaard, J. Greeley, H. Jónsson, and J. K. Nørskov. Density functional theory calculations for the hydrogen evolution reaction in an electrochemical double layer on the Pt(111) electrode. *Physical Chemistry Chemical Physics*, 9(25):3241–3250, 6 2007.

[91] J. Rossmeisl, J. K. Nørskov, C. D. Taylor, M. J. Janik, and M. Neurock. Calculated phase diagrams for the electrochemical oxidation and reduction of water over Pt(111). *Journal of Physical Chemistry B*, 110(43):21833–21839, 2006.

[92] J. Rossmeisl, A. Logadottir, and J. K. Nørskov. Electrolysis of water on (oxidized) metal surfaces. *Chemical Physics*, 319(1-3):178–184, 12 2005.

[93] S. Sakong, M. Naderian, K. Mathew, R. G. Hennig, and A. Groß. Density functional theory study of the electrochemical interface between a Pt electrode and an aqueous electrolyte using an implicit solvent method. *Journal of Chemical Physics*, 142(23): 234107, 6 2015.

[94] A. Atrens and W. Dietzel. The negative difference effect and unipositive Mg. *Advanced Engineering Materials*, 9(4):292–297, 4 2007.

[95] T. Würger, C. Feiler, G. B. Vonbun-Feldbauer, M. L. Zheludkevich, and R. H. Meißner. A first-principles analysis of the charge transfer in magnesium corrosion. *Scientific Reports*, 10(1):15006, 12 2020.

[96] B. Zeller-Plumhoff, M. Gile, M. Priebe, H. Slominska, B. Boll, B. Wiese, T. Würger, R. Willumeit-Römer, and R. H. Meißner. Exploring key ionic interactions for magnesium degradation insimulated body fluid - a data-driven approach. *Corrosion Science*, page 109272, 1 2021.

[97] C. G. Zoski. *Handbook of Electrochemistry*. Elsevier, 2007.

[98] H. A. Hansen, J. Rossmeisl, and J. K. Nørskov. Surface Pourbaix diagrams and oxygen reduction activity of Pt, Ag and Ni(111) surfaces studied by DFT. *Physical Chemistry Chemical Physics*, 10(25):3722–3730, 6 2008.

[99] R. L. Petty, A. W. Davidson, and J. Kleinberg. The Anodic Oxidation of Magnesium Metal: Evidence for the Existence of Unipositive Magnesium. *Journal of the American Chemical Society*, 76(2):363–366, 1954.

[100] S. Thomas, N. V. Medhekar, G. S. Frankel, and N. Birbilis. Corrosion mechanism and hydrogen evolution on Mg. *Current Opinion in Solid State and Materials Science*, 19 (2):85–94, 2015.

[101] G. L. Makar and J. Kruger. Corrosion of magnesium. *International Materials Reviews*, 38(3):138–153, 1 1993.

[102] M. Curioni. The behaviour of magnesium during free corrosion and potentiodynamic polarization investigated by real-time hydrogen measurement and optical imaging. *Electrochimica Acta*, 120:284–292, 2 2014.

[103] M. Taheri, J. R. Kish, N. Birbilis, M. Danaie, E. A. McNally, and J. R. McDermid. Towards a physical description for the origin of enhanced catalytic activity of corroding magnesium surfaces. *Electrochimica Acta*, 116:396–403, 1 2014.

[104] R. E. McNulty and J. D. Hanawalt. Some Corrosion Characteristics of High Purity Magnesium Alloys. *Transactions of The Electrochemical Society*, 81(1):423, 4 1942.

[105] T. Vegge. Locating the rate-limiting step for the interaction of hydrogen with Mg(0001) using Density-Functional Theory calculations and rate theory. *Physical Review B - Condensed Matter and Materials Physics*, 70(3):1–7, 2004.

[106] E. Skúlason, V. Tripkovic, M. E. Björketun, S. Gudmundsdóttir, G. Karlberg, J. Rossmeisl, T. Bligaard, H. Jónsson, and J. K. Nørskov. Modeling the electrochemical hydrogen oxidation and evolution reactions on the basis of density functional theory calculations. *Journal of Physical Chemistry C*, 114(42):18182–18197, 10 2010.

[107] K. S. Williams, J. P. Labukas, V. Rodriguez-Santiago, and J. W. Andzelm. First principles modeling of water dissociation on Mg(0001) and development of a Mg surface Pourbaix diagram. *Corrosion*, 71(2):209–223, 2015.

[108] K. S. Williams, V. Rodriguez-Santiago, and J. W. Andzelm. Modeling reaction pathways for hydrogen evolution and water dissociation on magnesium. *Electrochimica Acta*, 210:261–270, 2016.

[109] K. Mackay, R. Mackay, and W. Henderson. *Introduction to Modern Inorganic Chemistry*. CRC Press, London, 6th edition, 2017.

[110] W. Binns, F. Zargarzadah, V. Dehnavi, J. Chen, J. Noel, and D. Shoesmith. Physical and Electrochemical Evidence for the Role of a Mg Hydride Species in Mg Alloy Corrosion. *Corrosion*, page 2918, 7 2018.

[111] A. J. Du, S. C. Smith, X. D. Yao, and G. Q. Lu. Ab initio studies of hydrogen desorption from low index magnesium hydride surface. *Surface Science*, 600(9): 1854–1859, 5 2006.

[112] S. Kang, T. Ogitsu, S. A. Bonev, T. W. Heo, M. D. Allendorf, and B. C. Wood. Understanding Charge Transfer at Mg/MgH2 Interfaces for Hydrogen Storage. *ECS Transactions*, 77(10):81–90, 5 2017.

[113] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47(1):558–561, 1 1993.

[114] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B - Condensed Matter and Materials Physics*, 54(16):11169–11186, 10 1996.

[115] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 7 1996.

[116] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metalamorphous- semiconductor transition in germanium. *Physical Review B*, 49(20): 14251–14269, 5 1994.

[117] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, 12 1994.

[118] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758–1775, 1 1999.

[119] J. Carrasco, J. Klimeš, and A. Michaelides. The role of van der Waals forces in water adsorption on metals. *Journal of Chemical Physics*, 138(2):024708, 1 2013.

[120] J. Klimeš, D. R. Bowler, and A. Michaelides. Van der Waals density functionals applied to solids. *Physical Review B - Condensed Matter and Materials Physics*, 83 (19):195131, 5 2011.

[121] G. Román-Pérez and J. M. Soler. Efficient implementation of a van der waals density functional: Application to double-wall carbon nanotubes. *Physical Review Letters*, 103(9):096102, 8 2009.

[122] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Erratum: Van der Waals density functional for general geometries (Physical Review Letters (2004) 92 (246401)). *Physical Review Letters*, 95(10):109902, 9 2005.

[123] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth. Higher-accuracy van der Waals density functional. *Physical Review B - Condensed Matter and Materials Physics*, 82(8):081101, 8 2010.

[124] T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard, and D. C. Langreth. Van der Waals density functional: Self-consistent potential and the nature of the van der Waals bond. *Physical Review B - Condensed Matter and Materials Physics*, 76(12): 125112, 9 2007.

[125] W. Heckel, T. Würger, S. Müller, and G. Feldbauer. Van der Waals Interaction Really Matters: Energetics of Benzoic Acid on TiO 2 Rutile Surfaces. *The Journal of Physical Chemistry C*, 121(32):17207–17214, 8 2017.

[126] B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi, and M. Scheffler. On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters. II. the water hexamer and van der Waals interactions. *Journal of Chemical Physics*, 129(19), 2008.

[127] P. Kratzer and J. Neugebauer. The basics of electronic structure theory for periodic systems. *Frontiers in Chemistry*, 7(MAR):106, 2019.

[128] P. E. Blöchl, O. Jepsen, and O. K. Andersen. Improved tetrahedron method for Brillouin-zone integrations. *Physical Review B*, 49(23):16223–16233, 6 1994.

[129] J. J. Tang, X. B. Yang, L. Ouyang, M. Zhu, and Y. J. Zhao. A systematic first-principles study of surface energies, surface relaxation and Friedel oscillation of magnesium surfaces. *Journal of Physics D: Applied Physics*, 47(11), 2014.

[130] J. Du, J. Yang, M. Kuwabara, W. Li, and J. Peng. Effects of carbon and/or alkaline earth elements on grain refinement and tensile strength of AZ31 alloy. *Materials Transactions*, 49(10):2303–2309, 10 2008.

[131] J. Neugebauer and M. Scheffler. Adsorbate-substrate and adsorbate-adsorbate interactions of Na and K adlayers on Al(111). *Physical Review B*, 46(24):16067–16080, 1992.

[132] S. Schnur and A. Groß. Properties of metal-water interfaces studied from first principles. *New Journal of Physics*, 11(12):125003, 12 2009.

[133] A. Michaelides. Density functional theory simulations of water-metal interfaces: Waltzing waters, a novel 2D ice phase, and more. *Applied Physics A: Materials Science and Processing*, 85(4):415–425, 12 2006.

[134] M. Yu and D. R. Trinkle. Accurate and efficient algorithm for Bader charge integration. *Journal of Chemical Physics*, 134(6):064111, 2 2011.

[135] G. Henkelman, A. Arnaldsson, and H. Jónsson. A fast and robust algorithm for Bader decomposition of charge density. *Computational Materials Science*, 36(3):354–360, 6 2006.

[136] E. Sanville, S. D. Kenny, R. Smith, and G. Henkelman. Improved grid-based algorithm for Bader charge allocation. *Journal of Computational Chemistry*, 28(5):899–908, 4 2007.

[137] W. Tang, E. Sanville, and G. Henkelman. A grid-based Bader analysis algorithm without lattice bias. *Journal of Physics Condensed Matter*, 21(8):084204, 2 2009.

[138] R. F. W. Bader. *Atoms in Molecules: A Quantum Theory, Oxford University Press*. Clarendon Press, 1990.

[139] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. Open Babel: An Open chemical toolbox. *Journal of Cheminformatics*, 3 (10):33, 10 2011.

[140] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 2 1996.

[141] N. Bjerrum. Structure and Properties of Ice. *Science (New York, N.Y.)*, 115(2989): 385–90, 4 1952.

[142] A. Hodgson and S. Haq. Water adsorption and the wetting of metal surfaces. *Surface Science Reports*, 64(9):381–451, 9 2009.

[143] G. S. Frankel, A. Samaniego, and N. Birbilis. Evolution of hydrogen at dissolving magnesium surfaces. *Corrosion Science*, 70:104–111, 5 2013.

[144] S. Fajardo and G. S. Frankel. Effect of impurities on the enhanced catalytic activity for hydrogen evolution in high purity magnesium. *Electrochimica Acta*, 165:255–267, 5 2015.

[145] S. Fajardo, C. F. Glover, G. Williams, and G. S. Frankel. The Source of Anodic Hydrogen Evolution on Ultra High Purity Magnesium. *Electrochimica Acta*, 212: 510–521, 2016.

[146] C. D. Taylor. A First-Principles Surface Reaction Kinetic Model for Hydrogen Evolution under Cathodic and Anodic Conditions on Magnesium. *Journal of The Electrochemical Society*, 163(9):C602–C608, 2016.

[147] T. Miyake and M. Rolandi. Grotthuss mechanisms: From proton transport in proton wires to bioprotonic devices. *Journal of Physics Condensed Matter*, 28(2):023001, 12 2016.

[148] F. C. Lightstone, E. Schwegler, R. Q. Hood, F. Gygi, and G. Galli. A first principles molecular dynamics simulation of the hydrated magnesium ion. *Chemical Physics Letters*, 343(5-6):549–555, 2001.

[149] D. R. Banjade. *Fundamental Investigation of Magnesium Corrosion Using Experiments and Simulation*. PhD thesis, Brigham Young University, 2018.

[150] J. Vandevondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications*, 167(2):103–128, 4 2005.

[151] L.-M. Liu, A. Laio, and A. Michaelides. Initial stages of salt crystal dissolution determined with ab initio molecular dynamics. *Physical Chemistry Chemical Physics*, 13(29):13162, 8 2011.

[152] D. Trapl and V. Spiwok. Analysis of the Results of Metadynamics Simulations by metadynminer and metadynminer3d. *arXiv*, 8 2020.

[153] L. Fockaert, T. Würger, R. Unbehau, B. Boelen, R. Meißner, S. Lamaka, M. Zheludkevich, H. Terryn, and J. Mol. ATR-FTIR in Kretschmann configuration integrated with electrochemical cell as in situ interfacial sensitive tool to study corrosion inhibitors for magnesium substrates. *Electrochimica Acta*, 345:136166, 6 2020.

[154] X. Li, T. Würger, C. Feiler, R. H. Meißner, M. Serdechnova, C. Blawert, and M. L. Zheludkevich. Atomistic Insight into the Hydration States of Layered Double Hydroxides. *ACS Omega*, accepted, 2022.

[155] G. L. Song. *Corrosion behavior and prevention strategies for magnesium (Mg) alloys*. Woodhead Publishing Limited, 2013.

[156] W. Zhou, T. Shen, and N. N. Aung. Effect of heat treatment on corrosion behaviour of magnesium alloy AZ91D in simulated body fluid. *Corrosion Science*, 52(3): 1035–1041, 3 2010.

[157] M. Song and K. Chen. Effects of the enhanced heat treatment on the mechanical properties and stress corrosion behavior of an Al-Zn-Mg alloy. *Journal of Materials Science*, 43(15):5265–5273, 8 2008.

[158] Y. Chen, X. Lu, C. Blawert, M. L. Zheludkevich, T. Zhang, and F. Wang. Formation of self-lubricating PEO coating via in-situ incorporation of PTFE particles. *Surface and Coatings Technology*, 337:379–388, 3 2018.

[159] Y. Chen, X. Lu, S. V. Lamaka, P. Ju, C. Blawert, T. Zhang, F. Wang, and M. L. Zheludkevich. Active protection of Mg alloy by composite PEO coating loaded with corrosion inhibitors. *Applied Surface Science*, 504:144462, 2 2020.

[160] G. Zhang, E. Jiang, L. Wu, W. Ma, H. Yang, A. Tang, and F. Pan. Corrosion protection properties of different inhibitors containing PEO/LDHs composite coating on magnesium alloy AZ31. *Scientific Reports*, 11(1):1–14, 2 2021.

[161] Y. Li, X. Lu, M. Serdechnova, C. Blawert, M. L. Zheludkevich, K. Qian, T. Zhang, and F. Wang. Incorporation of LDH nanocontainers into plasma electrolytic oxidation coatings on Mg alloy. *Journal of Magnesium and Alloys*, 9 2021.

[162] C. Jing, B. Dong, A. Raza, T. Zhang, and Y. Zhang. Corrosion inhibition of layered double hydroxides for metal-based systems. *Nano Materials Science*, 3(1):47–67, 3 2021.

[163] G. Song, S. Hapugoda, and D. St John. Degradation of the surface appearance of magnesium and its alloys in simulated atmospheric environments. *Corrosion Science*, 49(3):1245–1265, 3 2007.

[164] E. S. Williams, J. Panko, and D. J. Paustenbach. The European Union's REACH regulation: a review of its history and requirements. *Critical Reviews in Toxicology*, 39(7):553–575, 8 2009.

[165] D. A. Winkler, M. Breedon, A. E. Hughes, F. R. Burden, A. S. Barnard, T. G. Harvey, and I. Cole. Towards chromate-free corrosion inhibitors: Structure-property models for organic alternatives. *Green Chemistry*, 16(6):3349–3357, 5 2014.

[166] D. A. Winkler, M. Breedon, P. White, A. E. Hughes, E. D. Sapper, and I. Cole. Using high throughput experimental data and in silico models to discover alternatives to toxic chromate corrosion inhibitors. *Corrosion Science*, 106:229–235, 5 2016.

[167] M. Grasserbauer. General aspects of trace analytical methods-IV. recommendations for nomenclature, standard procedures and reporting of experimental data for surface analysis techniques. *Pure and Applied Chemistry*, 51(11):2243–2250, 1979.

[168] A. Kokalj, M. Lozinšsek, B. Kapun, P. Taheri, S. Neupane, P. Losada-Pérez, C. Xie, S. Stavber, D. Crespo, F. U. Renner, A. Mol, and I. Milošev. Simplistic correlations between molecular electronic properties and inhibition efficiencies: Do they really exist? *Corrosion Science*, page 108856, 7 2020.

[169] L. Wang, D. Snihirova, M. Deng, B. Vaghefinazari, S. V. Lamaka, D. Höche, and M. L. Zheludkevich. Tailoring electrolyte additives for controlled Mg-Ca anode activity in aqueous Mg-air batteries. *Journal of Power Sources*, 460:228106, 6 2020.

[170] D. Snihirova, L. Wang, S. V. Lamaka, C. Wang, M. Deng, B. Vaghefinazari, D. Höche, and M. L. Zheludkevich. Synergistic Mixture of Electrolyte Additives: A Route to a High-Efficiency Mg-Air Battery. *Journal of Physical Chemistry Letters*, 11(20): 8790–8798, 2020.

[171] D. Raps, T. Hack, J. Wehr, M. L. Zheludkevich, A. C. Bastos, M. G. Ferreira, and O. Nuyken. Electrochemical study of inhibitor-containing organic-inorganic hybrid coatings on AA2024. *Corrosion Science*, 51(5):1012–1021, 5 2009.

[172] J. Yang, C. Blawert, S. V. Lamaka, D. Snihirova, X. Lu, S. Di, and M. L. Zheludkevich. Corrosion protection properties of inhibitor containing hybrid PEO-epoxy coating on magnesium. *Corrosion Science*, 140:99–110, 8 2018.

[173] A. H. Lipkus, Q. Yuan, K. A. Lucas, S. A. Funk, W. F. Bartelt, R. J. Schenck, and A. J. Trippe. Structural diversity of organic chemistry. A scaffold analysis of the CAS Registry. *Journal of Organic Chemistry*, 73(12):4443–4451, 6 2008.

[174] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, and H. Jhoti. Twenty years on: The impact of fragments on drug discovery. *Nature Reviews Drug Discovery*, 15 (9):605–619, 8 2016.

[175] J. Li, S. G. Ballmer, E. P. Gillis, S. Fujii, M. J. Schmidt, A. M. Palazzolo, J. W. Lehmann, G. F. Morehouse, and M. D. Burke. Synthesis of many different types of organic small molecules using one automated process. *Science*, 347(6227):1221–1226, 3 2015.

[176] S. J. García, T. H. Muster, Ö.. Özkanat, N. Sherman, A. E. Hughes, H. Terryn, J. H. de Wit, and J. M. Mol. The influence of pH on corrosion inhibitor selection for 2024-T3 aluminium alloy assessed by high-throughput multielectrode and potentiodynamic testing. *Electrochimica Acta*, 55(7):2457–2465, 2 2010.

[177] 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. Garcia, J. M. Mol, and A. E. Hughes. A new high-throughput method for corrosion testing. *Corrosion Science*, 58:327–331, 5 2012.

[178] 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, and J. M. Mol. A rapid screening multi-electrode method for the evaluation of corrosion inhibitors. *Electrochimica Acta*, 54(12):3402–3411, 4 2009.

[179] M. Meeusen, L. Zardet, A. M. Homborg, M. Lekka, F. Andreatta, L. Fedrizzi, B. Boelen, H. Terryn, and J. M. C. Mol. A Complementary Electrochemical Approach for Time-Resolved Evaluation of Corrosion Inhibitor Performance. *Journal of The Electrochemical Society*, 166(11):C3220–C3232, 5 2019.

[180] T. Würger, D. Mei, B. Vaghefinazari, D. A. Winkler, S. V. Lamaka, M. L. Zheludkevich, R. H. Meißner, and C. Feiler. Exploring structure-property relationships in magnesium dissolution modulators. *npj Materials Degradation*, 5(1):2, 12 2021.

[181] E. J. Schiessler, T. Würger, S. V. Lamaka, R. H. Meißner, C. J. Cyron, M. L. Zheludkevich, C. Feiler, and R. C. Aydin. Predicting the inhibition efficiencies of magnesium dissolution modulators using sparse machine learning models. *npj Computational Materials*, 7(1):1–9, 12 2021.

[182] T. Würger, L. Wang, D. Snihirova, M. Deng, S. V. Lamaka, D. A. Winkler, D. Höche, M. L. Zheludkevich, R. H. Meißner, and C. Feiler. Data-driven selection of electrolyte additives for aqueous magnesium batteries. *Journal of Materials Chemistry A*, 10: 21672–21682, 2022.

[183] D. A. Winkler. Predicting the performance of organic corrosion inhibitors. *Metals*, 7 (12):553, 12 2017.

[184] M. Fernandez, M. Breedon, I. S. Cole, and A. S. Barnard. Modeling corrosion inhibition efficacy of small organic molecules as non-toxic chromate alternatives using comparative molecular surface analysis (CoMSA). *Chemosphere*, 160:80–88, 10 2016.

[185] F. F. Chen, M. Breedon, P. White, C. Chu, D. Mallick, S. Thomas, E. Sapper, and I. Cole. Correlation between molecular features and electrochemical properties using an artificial neural network. *Materials and Design*, 112:410–418, 12 2016.

[186] T. C. Le and D. A. Winkler. Discovery and Optimization of Materials Using Evolutionary Approaches. *Chemical Reviews*, 116(10):6107–6132, 5 2016.

[187] M. H. Segler, M. Preuss, and M. P. Waller. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, 555(7698):604–610, 3 2018.

[188] I. Milošev, D. Zimerl, C. Carriére, S. Zanna, A. Seyeux, J. Iskra, S. Stavber, F. Chiter, M. Poberžnik, D. Costa, A. Kokalj, and P. Marcus. Editors' Choice-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. *Journal of The Electrochemical Society*, 167:061509, 2020.

[189] M. Poberžnik, F. Chiter, I. Milošev, P. Marcus, D. Costa, and A. Kokalj. DFT study of n-alkyl carboxylic acids on oxidized aluminum surfaces: From standalone molecules to self-assembled-monolayers. *Applied Surface Science*, 525:146156, 9 2020.

[190] C. Feiler, D. Mei, B. Luthringer, S. Lamaka, and M. Zheludkevich. Rational Design of Effective Mg Degradation Modulators. *CORROSION*, page 3597, 7 2020.

[191] T. L. Galvão, G. Novell-Leruth, A. Kuznetsova, J. Tedim, and J. R. Gomes. Elucidating Structure-Property Relationships in Aluminum Alloy Corrosion Inhibitors by Machine Learning. *Journal of Physical Chemistry C*, 124(10):5624–5635, 3 2020.

[192] M. O. Pekguleryuz, K. U. Kainer, and A. A. Kaya. *Fundamentals of magnesium alloy metallurgy*. Woodhead Publishing Limited, Philadelphia, 2013.

[193] E. E. Ebenso, T. Arslan, F. Kandemirli, N. Caner, and I. Love. Quantum chemical studies of some rhodanine azosulpha drugs as corrosion inhibitors for mild steel in acidic medium. *International Journal of Quantum Chemistry*, 110(5):1003–1018, 4 2010.

[194] K. F. Khaled. Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A QSAR model. *Corrosion Science*, 53(11):3457–3465, 11 2011.

[195] H. Ju, Z. P. Kai, and Y. Li. Aminic nitrogen-bearing polydentate Schiff base compounds as corrosion inhibitors for iron in acidic media: A quantum chemical calculation. *Corrosion Science*, 50(3):865–871, 3 2008.

[196] N. Kovačević, I. Milošev, and A. Kokalj. How relevant is the adsorption bonding of imidazoles and triazoles for their corrosion inhibition of copper? *Corrosion Science*, 124:25–34, 2017.

[197] A. Kokalj. Is the analysis of molecular electronic structure of corrosion inhibitors sufficient to predict the trend of their inhibition performance. *Electrochimica Acta*, 56 (2):745–755, 12 2010.

[198] A. Kokalj and D. Costa. Molecular modeling of corrosion inhibitors. *Encyclopedia of Interfacial Chemistry: Surface Science and Electrochemistry*, pages 332–345, 1 2018.

[199] 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. *Physical Chemistry Chemical Physics*, 2(14):3121–3125, 1 2000.

[200] A. Mauri. alvaDesc: A tool to calculate and analyze molecular descriptors and fingerprints. In *Methods in Pharmacology and Toxicology*, pages 801–820. Humana Press Inc., 2020.

[201] P. Gasparotto, R. H. Meißner, and M. Ceriotti. Recognizing Local and Global Structural Motifs at the Atomic Scale. *Journal of Chemical Theory and Computation*, 14(2):486–498, 2 2018.

[202] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K. R. Müller, and O. Anatole Von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, 15:1–9, 2013.

[203] L. C. Blum and J. L. Reymond. 970 Million druglike small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society*, 131(25):8732–8733, 7 2009.

[204] C. Hansch, A. Leo, and R. W. Taft. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chemical Reviews*, 91(2):165–195, 1991.

[205] J. Byegård, G. Skarnemark, and M. Skålberg. The stability of some metal EDTA, DTPA and DOTA complexes: Application as tracers in groundwater studies. *Journal of Radioanalytical and Nuclear Chemistry*, 241(2):281–290, 4 1999.

[206] M. Deng, L. Wang, D. Höche, S. V. Lamaka, P. Jiang, D. Snihirova, N. Scharnagl, and M. L. Zheludkevich. Ca/In micro alloying as a novel strategy to simultaneously enhance power and energy density of primary Mg-air batteries from anode aspect. *Journal of Power Sources*, 472:228528, 10 2020.

[207] T. Zhang, Z. Tao, and J. Chen. Magnesium-air batteries: From principle to application. *Materials Horizons*, 1(2):196–206, 3 2014.

[208] D. Wu, Z. Wen, H. Jiang, H. Li, Y. Zhuang, J. Li, Y. Yang, J. Zeng, J. Cheng, and J. Zhao. Ultralong-Lifespan Magnesium Batteries Enabled by the Synergetic Manipulation of Oxygen Vacancies and Electronic Conduction. *ACS Applied Materials and Interfaces*, 13:12049–12058, 2021.

[209] P. Canepa, G. Sai Gautam, D. C. Hannah, R. Malik, M. Liu, K. G. Gallagher, K. A. Persson, and G. Ceder. Odyssey of Multivalent Cathode Materials: Open Questions and Future Challenges. *Chemical Reviews*, 117(5):4287–4341, 3 2017.

[210] M. Armand and J. M. Tarascon. Building better batteries. *Nature*, 451(7179):652–657, 2 2008.

[211] Q. Liu, Z. Yan, E. Wang, S. Wang, and G. Sun. A high-specific-energy magnesium/water battery for full-depth ocean application. *International Journal of Hydrogen Energy*, 42(36):23045–23053, 9 2017.

[212] M. Deng, L. Wang, B. Vaghefinazari, W. Xu, C. Feiler, S. V. Lamaka, D. Höche, M. L. Zheludkevich, and D. Snihirova. High-energy and durable aqueous magnesium batteries: Recent advances and perspectives. *Energy Storage Materials*, 43:238–247, 12 2021.

[213] J. Song, J. She, D. Chen, and F. Pan. Latest research advances on magnesium and magnesium alloys worldwide. *Journal of Magnesium and Alloys*, 8(1):1–41, 3 2020.

[214] M. Yuasa, X. Huang, K. Suzuki, M. Mabuchi, and Y. Chino. Discharge properties of Mg-Al-Mn-Ca and Mg-Al-Mn alloys as anode materials for primary magnesium-air batteries. *Journal of Power Sources*, 297:449–456, 11 2015.

[215] M. Deng, L. Wang, D. Höche, S. V. Lamaka, D. Snihirova, B. Vaghefinazari, and M. L. Zheludkevich. Clarifying the decisive factors for utilization efficiency of Mg anodes for primary aqueous batteries. *Journal of Power Sources*, 441(October):227201, 11 2019.

[216] M. Deng, L. Wang, D. Höche, S. V. Lamaka, C. Wang, D. Snihirova, Y. Jin, Y. Zhang, and M. L. Zheludkevich. Approaching "stainless magnesium" by Ca micro-alloying. *Materials Horizons*, 8(2):589–596, 2 2021.

[217] M. Deng, L. Wang, D. Höche, S. V. Lamaka, D. Snihirova, P. Jiang, and M. L. Zheludkevich. Corrosion and discharge properties of Ca/Ge micro-alloyed Mg anodes for primary aqueous Mg batteries. *Corrosion Science*, 177:108958, 12 2020.

[218] L. Wang, D. Snihirova, M. Deng, B. Vaghefinazari, D. Höche, S. V. Lamaka, and M. L. Zheludkevich. Enhancement of discharge performance for aqueous Mg-air batteries in 2,6-dihydroxybenzoate-containing electrolyte. *Chemical Engineering Journal*, 429: 132369, 2 2022.

[219] L. Wang, D. Snihirova, M. Deng, C. Wang, D. Höche, S. V. Lamaka, and M. L. Zheludkevich. Indium chloride as an electrolyte additive for primary aqueous Mg batteries. *Electrochimica Acta*, 373:137916, 3 2021.

[220] B. Vaghefinazari, D. Höche, S. V. Lamaka, D. Snihirova, and M. L. Zheludkevich. Tailoring the Mg-air primary battery performance using strong complexing agents as electrolyte additives. *Journal of Power Sources*, 453:227880, 3 2020.

[221] TURBOMOLE. V7.4. A Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2019 since 2007. *Available from TURBOMOLE GmbH*, 2019.

[222] R. K. Cersonsky, B. A. Helfrecht, E. A. Engel, S. Kliavinek, and M. Ceriotti. Improving sample and feature selection with principal covariates regression. *Machine Learning: Science and Technology*, 2(3):2012.12253, 12 2021.

[223] J. F. Box. R. A. Fisher and the Design of Experiments, 1922-1926. *The American Statistician*, 34(1):1, 2 1980.

[224] Y. Sverchkov and M. Craven. A review of active learning approaches to experimental design for uncovering biological networks. *PLOS Computational Biology*, 13(6): e1005466, 6 2017.

[225] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials*, 5(1):1–17, 2 2019.

[226] G. Imbalzano, Y. Zhuang, V. Kapil, K. Rossi, E. A. Engel, F. Grasselli, and M. Ceriotti. Uncertainty estimation for molecular dynamics and sampling. *The Journal of Chemical Physics*, 154(7):074102, 2 2021.

[227] M. Reutlinger, C. P. Koch, D. Reker, N. Todoroff, P. Schneider, T. Rodrigues, and G. Schneider. Chemically advanced template search (CATS) for scaffold-hopping and prospective target prediction for 'orphan' molecules. *Molecular Informatics*, 32(2): 133–138, 2 2013.

[228] L. David, A. Thakkar, R. Mercado, and O. Engkvist. Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, 12(1):1–22, 9 2020.

[229] O. Devinyak, D. Havrylyuk, and R. Lesyk. 3D-MoRSE descriptors explained. *Journal of Molecular Graphics and Modelling*, 54:194–203, 11 2014.

[230] J. H. Schuur, P. Selzer, and J. Gasteiger. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, 36(2):334–344, 1996.

[231] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindah. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 9 2015.

[232] R. H. Meißner, J. Schneider, P. Schiffels, and L. Colombi Ciacchi. Computational prediction of circular dichroism spectra and quantification of helicity loss upon peptide adsorption on silica. *Langmuir*, 30(12):3487–3494, 2014.

[233] J. Schneider and L. Colombi Ciacchi. Specific material recognition by small peptides mediated by the interfacial solvent structure. *Journal of the American Chemical Society*, 134(4):2407–2413, 2012.

[234] E. G. Brandt and A. P. Lyubartsev. Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogues and a Titanium Binding Peptide on the TiO 2 (100) Surface. *The Journal of Physical Chemistry C*, 119(32):18126–18139, 8 2015.

[235] M. Lundborg and E. Lindahl. Automatic GROMACS topology generation and comparisons of force fields for solvation free energy calculations. *Journal of Physical Chemistry B*, 119(3):810–823, 1 2015.

[236] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K. R. Müller. Machine Learning Force Fields. *Chemical Reviews*, 121(16):10142–10186, 8 2021.

[237] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A. Grama, and A. C. Van Duin. The ReaxFF reactive force-field: Development, applications and future directions. *npj Computational Materials*, 2(1):1–14, 3 2016.

[238] P. Yoo, M. Sakano, S. Desai, M. M. Islam, P. Liao, and A. Strachan. Neural network reactive force field for C, H, N, and O systems. *npj Computational Materials*, 7(1): 1–10, 1 2021.

[239] C. Dong, Y. Ji, X. Wei, A. Xu, D. Chen, N. Li, D. Kong, X. Luo, K. Xiao, and X. Li. Integrated computation of corrosion: Modelling, simulation and applications. *Corrosion Communications*, 2:8–23, 6 2021.

[240] B. Mortazavi, M. Silani, E. V. Podryabinkin, T. Rabczuk, X. Zhuang, and A. V. Shapeev. First-Principles Multiscale Modeling of Mechanical Properties in Graphene/Borophene Heterostructures Empowered by Machine-Learning Interatomic Potentials. *Advanced Materials*, 33(35):2102807, 9 2021.

[241] M. F. Horstemeyer. Multiscale modeling: A review. *Practical Aspects of Computational Chemistry: Methods, Concepts and Applications*, pages 87–135, 2010.

Appendix A

Computational Details

To perform DFT calculations with VASP and CP2k, various input parameters have to be defined. Although the pre-defined default parameters often yield reasonable and reliable results, some input parameters have to be fine-tuned in convergence tests with respect to the given simulation case. The general goal is to find a set of parameters that yields an adequate trade-off between accuracy and computational costs. In the following, several parameter sets for VASP and CP2k are provided that have been used in the frame of this thesis. Comprehensive descriptions of the given parameters can be found in the respective documentations online (VASP: https://www.vasp.at/wiki/index.php/The_VASP_Manual, CP2k: https://manual.cp2k.org/).

A.1 VASP Settings

Table A.1: Settings for calculating the Bader charges.

Tag	Value	Description
GGA	BO	optB88-vdW as exchange-correlation functional
PARAM1	0.1833333333	optB88-vdW parameter
PARAM2	0.2200000000	optB88-vdW parameter
LUSE_VDW	T	optB88-vdW parameter
AGGAC	0.0000	optB88-vdW parameter
LASPH	T	optB88-vdW parameter
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	6	minimum number of electronic SCF steps.
ISMEAR	-5	partial occupancy for each orbital
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	50E-7	global break condition for the electronic SC-loop (eV)
LWAVE	F	writing wave functions?
LCHARG	T	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?
IDIPOL	3	monopole/dipole/quadrupole corrections to the energy
LDIPOL	T	corrections to the potential and forces
LAECHG	T	all-electron charge density

Table A.2: Settings for nudged elastic band calculations of hydrogen evolution reactions.

Tag	Value	Description
GGA	BO	optB88-vdW as exchange-correlation functional
PARAM1	0.1833333333	optB88-vdW parameter
PARAM2	0.2200000000	optB88-vdW parameter
LUSE_VDW	T	optB88-vdW parameter
AGGAC	0.0000	optB88-vdW parameter
LASPH	T	optB88-vdW parameter
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	4	minimum number of electronic SCF steps.
ISMEAR	1	partial occupancy for each orbital
SIGMA	0.2	determines the width of the smearing in eV
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	1E-6	global break condition for the electronic SC-loop (eV)
EDIFFG	-0.05	break condition for the ionic relaxation loop (eV/Å)
NSW	600	maximum number of ionic steps
IBRION	3	scheme to update and move ions
POTIM	0.1	step width
SMASS	2	velocity control
ISIF	2	degrees-of-freedom allowed to change
ISYM	0	treating symmetry
LWAVE	F	writing wave functions?
LCHARG	T	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?
IMAGES	24	number of NEB images between the fixed endpoints
SPRING	-5	spring constant in eV/Å^2 between the images
LCLIMB	T	climbing image algorithm
IDIPOL	3	monopole/dipole/quadrupole corrections to the energy
LDIPOL	T	corrections to the potential and forces

Table A.3: Settings to calculate hydrogen evolution reaction steps upon the Mg(0001) surface. A three-step relaxation scheme was followed. In case an input parameter changed within this scheme, three values are given, respectively.

Tag	Value	Description
GGA	BO	optB88-vdW as exchange-correlation functional
PARAM1	0.1833333333	optB88-vdW parameter
PARAM2	0.2200000000	optB88-vdW parameter
LUSE_VDW	T	optB88-vdW parameter
AGGAC	0.0000	optB88-vdW parameter
LASPH	T	optB88-vdW parameter
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	4 / 4 / 6	minimum number of electronic SCF steps.
NELMDL	— / — / -10	number of non-self-consistent steps at the beginning
NELM	— / — / 300	maximum number of electronic SCF steps performed.
ISMEAR	1 / 1 / -5	partial occupancy for each orbital
SIGMA	0.2 / 0.2 / —	determines the width of the smearing in eV
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	50 E-5 / E-5 / E-6	global break condition for the electronic SC-loop (eV)
EDIFFG	-5E-3	break condition for the ionic relaxation loop (eV/Å)
NSW	300 /300 / —	maximum number of ionic steps
IBRION	2 / 1 / —	scheme to update and move ions
ISIF	2	degrees-of-freedom allowed to change
LWAVE	F	writing wave functions?
LCHARG	T	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?
IDIPOL	3	monopole/dipole/quadrupole corrections to the energy
LDIPOL	T	corrections to the potential and forces
DIPOL	0.5 0.5 0.4	center of the cell for calculating dipole-moment

Table A.4: Settings for nudged elastic band calculations of Mg dissolution.

Tag	Value	Description
GGA	BO	optB88-vdW as exchange-correlation functional
PARAM1	0.1833333333	optB88-vdW parameter
PARAM2	0.2200000000	optB88-vdW parameter
LUSE_VDW	T	optB88-vdW parameter
AGGAC	0.0000	optB88-vdW parameter
LASPH	T	optB88-vdW parameter
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	4	minimum number of electronic SCF steps.
ISMEAR	1	partial occupancy for each orbital
SIGMA	0.2	determines the width of the smearing in eV
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	1E-4	global break condition for the electronic SC-loop (eV)
EDIFFG	-0.05	break condition for the ionic relaxation loop (eV/Å)
NSW	400	maximum number of ionic steps
IBRION	3	scheme to update and move ions
POTIM	0	step width
IOPT	7	optimizer
ISIF	2	degrees-of-freedom allowed to change
ISYM	0	treating symmetry
LWAVE	F	writing wave functions?
LCHARG	T	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?
IMAGES	48	number of NEB images between the fixed endpoints
SPRING	-5	spring constant in eV/Å^2 between the images
LCLIMB	T	climbing image algorithm
IDIPOL	3	monopole/dipole/quadrupole corrections to the energy
LDIPOL	T	corrections to the potential and forces

Table A.5: Settings for calculating the dynamical matrix.

Tag	Value	Description
GGA	BO	optB88-vdW as exchange-correlation functional
PARAM1	0.1833333333	optB88-vdW parameter
PARAM2	0.2200000000	optB88-vdW parameter
LUSE_VDW	T	optB88-vdW parameter
AGGAC	0.0000	optB88-vdW parameter
LASPH	T	optB88-vdW parameter
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	4	minimum number of electronic SCF steps.
ISMEAR	1	partial occupancy for each orbital
SIGMA	0.2	determines the width of the smearing in eV
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	1E-8	global break condition for the electronic SC-loop (eV)
EDIFFG	-1E-8	break condition for the ionic relaxation loop (eV/Å)
NSW	19	maximum number of ionic steps
IBRION	3	scheme to update and move ions
POTIM	0	step width
ICHAIN	1	Dynamical Matrix method
ISIF	2	degrees-of-freedom allowed to change
ISYM	0	treating symmetry
LWAVE	F	writing wave functions?
LCHARG	T	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?

Table A.6: Settings to calculate adsorption energies of magnesium dissolution modulators upon the MgO(100) surface. A three-step relaxation scheme was followed. In case an input parameter changed within this scheme, three values are given, respectively.

Tag	Value	Description
GGA	BO	optB88-vdW as exchange-correlation functional
PARAM1	0.1833333333	optB88-vdW parameter
PARAM2	0.2200000000	optB88-vdW parameter
LUSE_VDW	T	optB88-vdW parameter
AGGAC	0.0000	optB88-vdW parameter
LASPH	T	optB88-vdW parameter
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	4 / 4 / 6	minimum number of electronic SCF steps.
NELMDL	— / — / -10	number of non-self-consistent steps at the beginning
NELM	— / — / 300	maximum number of electronic SCF steps performed.
ISMEAR	1 / 1 / -5	partial occupancy for each orbital
SIGMA	0.2 / 0.2 / —	determines the width of the smearing in eV
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	90E-5 / E-5 / E-6	global break condition for the electronic SC-loop (eV)
EDIFFG	-5E-3	break condition for the ionic relaxation loop (eV/Å)
NSW	200 / 500 / —	maximum number of ionic steps
IBRION	2 / 1 / —	scheme to update and move ions
ISIF	2	degrees-of-freedom allowed to change
LWAVE	F	writing wave functions?
LCHARG	T	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?
IDIPOL	3	monopole/dipole/quadrupole corrections to the energy
LDIPOL	T	corrections to the potential and forces
DIPOL	0.5 0.5 0.35	center of the cell for calculating dipole-moment

Table A.7: Settings to calculate adsorption energies of magnesium dissolution modulators upon the MgO(111) surface. A three-step relaxation scheme was followed. In case an input parameter changed within this scheme, three values are given, respectively.

Tag	Value	Description
GGA	PE	PBE as exchange-correlation functional
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	4 / 4 / 6	minimum number of electronic SCF steps.
NELMDL	— / — / -10	number of non-self-consistent steps at the beginning
NELM	— / — / 300	maximum number of electronic SCF steps performed.
ISMEAR	0 / 0 / -5	partial occupancy for each orbital
SIGMA	0.05 / 0.05 / —	determines the width of the smearing in eV
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	1E-4 / 1E-4 / 1E-6	global break condition for the electronic SC-loop (eV)
EDIFFG	-5E-3	break condition for the ionic relaxation loop (eV/Å)
NSW	300 / 300 / —	maximum number of ionic steps
IBRION	2 / 1 / —	scheme to update and move ions
ISIF	2	degrees-of-freedom allowed to change
LWAVE	F	writing wave functions?
LCHARG	F	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?

Table A.8: Settings for isolated molecules. A two-step relaxation scheme was followed. In case an input parameter changed within this scheme, two values are given, respectively.

Tag	Value	Description
GGA	PE	PBE as exchange-correlation functional
ENCUT	520	cutoff energy for the plane-wave-basis set in eV
PREC	Accurate	precision-mode
NELMIN	4	minimum number of electronic SCF steps.
ISMEAR	0	partial occupancy for each orbital
SIGMA	0.05	determines the width of the smearing in eV
ALGO	V	specify the electronic minimisation algorithm
LREAL	AUTO	projection operators evaluation
EDIFF	2E-6	global break condition for the electronic SC-loop (eV)
EDIFFG	-5E-3	break condition for the ionic relaxation loop (eV/Å)
NSW	3000 / 3000	maximum number of ionic steps
IBRION	2 / 1	scheme to update and move ions
ISIF	2	degrees-of-freedom allowed to change
LWAVE	F	writing wave functions?
LCHARG	F	writing charge densities?
LPLANE	T	plane-wise data distribution in real space?
LSCALU	F	parallel LU decomposition?
IDIPOL	4	monopole/dipole/quadrupole corrections to the total energy
LDIPOL	T	corrections to the potential and forces
DIPOL	0.5 0.5 0.5	center of the cell for calculating dipole-moment

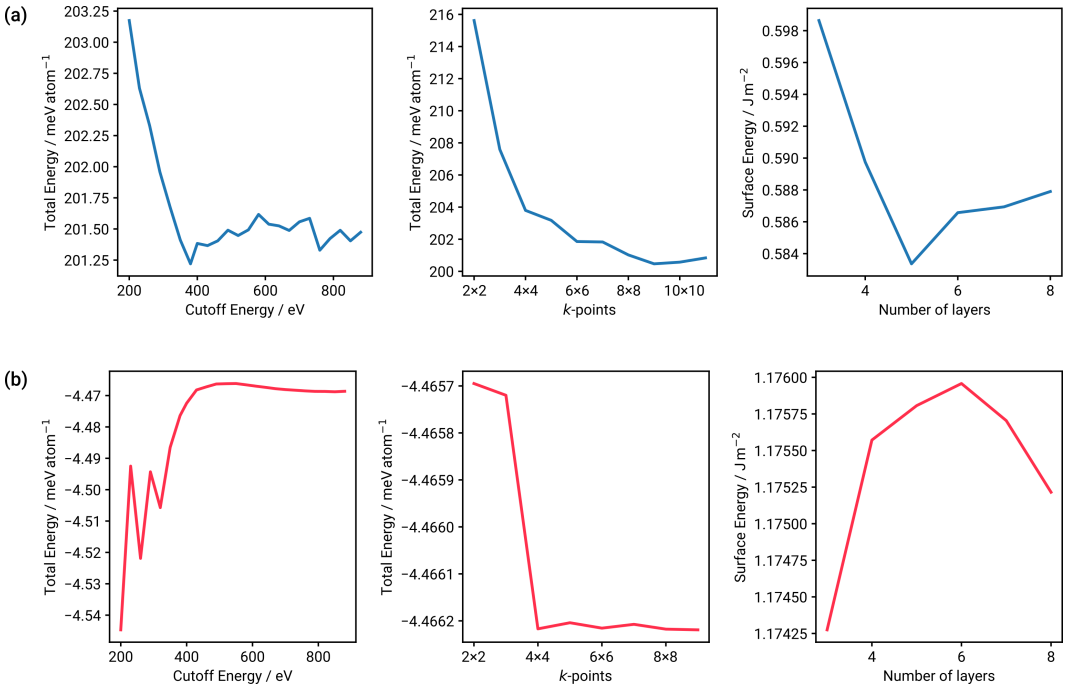


Figure A.1: Convergence tests. System energies with respect to (left) the employed cutoff energy, (middle) the number of k-points and (right) the number of surface layers for (a) Mg(0001) and (b) MgO(100).

A.2 CP2k Settings

A.2.1 Born-Oppenheimer Molecular Dynamics

```
@set data /gpfs/home/wuerger/software/cp2k-2020/data
@set RESTART 1
@set PROJECT_NAME mag-wat-MD
@set COORD_FILE mg_2x2_stripe_16OH_sol.xyz

&GLOBAL
  PROJECT ${PROJECT_NAME}
  RUN_TYPE MD
  PRINT_LEVEL LOW
  PREFERRED_DIAG_LIBRARY  ELPA
&END GLOBAL

&MOTION

  &MD
    ENSEMBLE NVT

      &THERMOSTAT
       TYPE CSVR
       REGION MOLECULE
       &CSVR
         TIMECON 100
       &END CSVR
      &END THERMOSTAT

    &THERMAL_REGION
      &PRINT
        &TEMPERATURE
        &END
      &END PRINT

      &DEFINE_REGION         #Mg
        TEMPERATURE 300
        TEMP_TOL 50
        LIST 1..308
      &END DEFINE_REGION

      &DEFINE_REGION         #OH+H
        TEMPERATURE 300
```

```
      TEMP_TOL 50
      LIST 309..356
    &END DEFINE_REGION

    &DEFINE_REGION      #H2O
      TEMPERATURE 300
      TEMP_TOL 50
      LIST 357..968
    &END DEFINE_REGION

  &END THERMAL_REGION
  TIMESTEP 0.5
  STEPS 30000
  TEMPERATURE 300
  TEMP_KIND
  TEMP_TOL 50

  &PRINT
    &ENERGY
      &EACH
        MD 1
      &END EACH
    &END ENERGY

    &PROGRAM_RUN_INFO
      &EACH
        MD 10
      &END EACH
    &END PROGRAM_RUN_INFO
  &END PRINT
&END MD


&CONSTRAINT
  &FIXED_ATOMS
    LIST 1..48
    COMPONENTS_TO_FIX XYZ
  &END
&END

&PRINT
  &VELOCITIES
    &EACH
      MD 1
```

```
      &END EACH
    &END

    &TRAJECTORY
     &EACH
       MD 1
     &END EACH
    &END

    &FORCES
      &EACH
       MD 1
      &END EACH
    &END

  &END PRINT

&END MOTION

&FORCE_EVAL
  METHOD Quickstep
  &PRINT
    &FORCES ON
      &EACH
        QS_SCF 0
      &END
      FILENAME =${PROJECT_NAME}.forces
      ADD_LAST NUMERIC
    &END
    &DISTRIBUTION
    &END
  &END
  &DFT
    BASIS_SET_FILE_NAME ${data}/BASIS_MOLOPT
    POTENTIAL_FILE_NAME ${data}/GTH_POTENTIALS
    WFN_RESTART_FILE_NAME ./mag-wat-MD-RESTART.wfn

    CHARGE 0

    &MGRID
      CUTOFF 500
      NGRIDS 5
      REL_CUTOFF 50.
    &END MGRID
```

```
128
129   &QS
130     METHOD GPW
131     EXTRAPOLATION USE_PREV_WF
132     EXTRAPOLATION_ORDER 0
133   &END QS
134
135   &SCF
136     MAX_SCF 400
137     ADDED_MOS 700
138     SCF_GUESS RESTART
139     EPS_SCF 5.0E-5
140     CHOLESKY INVERSE_DBCSR
141
142     &SMEAR  ON
143        METHOD FERMI_DIRAC
144        ELECTRONIC_TEMPERATURE [K] 500
145     &END SMEAR
146     &DIAGONALIZATION
147          ALGORITHM STANDARD
148          EPS_ADAPT 0.01
149     &END DIAGONALIZATION
150     &MIXING
151         METHOD BROYDEN_MIXING
152         ALPHA   0.08
153         BETA    1.5
154         NBROYDEN  8
155     &END
156
157     &PRINT
158       &RESTART
159         &EACH
160           QS_SCF 50
161         &END EACH
162       &END RESTART
163     &END PRINT
164   &END SCF
165
166   &XC
167     &XC_FUNCTIONAL PBE
168     &END XC_FUNCTIONAL
169
170     &XC_GRID
171       XC_SMOOTH_RHO NN50
```

```
        XC_DERIV NN50_SMOOTH
      &END
      &VDW_POTENTIAL
        DISPERSION_FUNCTIONAL PAIR_POTENTIAL
        &PAIR_POTENTIAL
          TYPE DFTD3
          PARAMETER_FILE_NAME ${data}/dftd3.dat
          REFERENCE_FUNCTIONAL PBE
          R_CUTOFF [angstrom] 12.0
        &END PAIR_POTENTIAL
      &END VDW_POTENTIAL
    &END XC

    &PRINT
      &MULLIKEN
        FILENAME=mulliken
        COMMON_ITERATION_LEVELS 10
        &EACH
          MD 1
        &END EACH
      &END MULLIKEN
    &END PRINT
  &END DFT
  &SUBSYS
    &CELL
ABC 19.0156192780 22.1237201691 50
      PERIODIC XYZ
    &END CELL

    &TOPOLOGY
      COORD_FILE_NAME ./${COORD_FILE}
      COORD_FILE_FORMAT XYZ
    &END
    &KIND Fe
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q16
    &END KIND

    &KIND Mg
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q10
    &END

    &KIND H
```

```
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q1
    &END KIND

    &KIND O
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q6
    &END KIND

  &END SUBSYS
&END FORCE_EVAL

@if ${RESTART} == 1
&EXT_RESTART
  RESTART_FILE_NAME ./${PROJECT_NAME}-1.restart
  RESTART_DEFAULT F
  RESTART_POS T
&END EXT_RESTART
@endif
```

A.2.2 Second Generation Car-Parrinello Molecular Dynamics

```
@set data /gpfs/home/wuerger/software/cp2k-2020/data
@set RESTART 1
@set PROJECT_NAME mag-wat-MD
@set COORD_FILE mg_2x2_stripe_16OH_sol.xyz

&GLOBAL
  PROJECT ${PROJECT_NAME}
  RUN_TYPE MD
  PRINT_LEVEL LOW
  PREFERRED_DIAG_LIBRARY  ELPA
&END GLOBAL

&MOTION

  &MD
    ENSEMBLE LANGEVIN
    TIMESTEP 0.5
    STEPS 30000
    TEMPERATURE 300
    TEMP_KIND
    TEMP_TOL 50

    &LANGEVIN
      GAMMA       0.005
      NOISY_GAMMA 4.0E-5
    &END LANGEVIN

    &THERMAL_REGION
      DO_LANGEVIN_DEFAULT TRUE
      &PRINT
        &TEMPERATURE
        &END
      &END PRINT

      &DEFINE_REGION         #Mg
        TEMPERATURE 300
        TEMP_TOL 50
        NOISY_GAMMA_REGION  4.2E-5
        LIST 1..308
      &END DEFINE_REGION

      &DEFINE_REGION         #OH+H
```

```
      TEMPERATURE 300
      TEMP_TOL 50
      NOISY_GAMMA_REGION  1.6E-4
      LIST 309..356
    &END DEFINE_REGION

    &DEFINE_REGION      #H2O
      TEMPERATURE 300
      TEMP_TOL 50
      NOISY_GAMMA_REGION 1.6E-4
      LIST 357..968
    &END DEFINE_REGION

  &END THERMAL_REGION

  &PRINT
    &ENERGY
      &EACH
        MD 1
      &END EACH
    &END ENERGY

    &PROGRAM_RUN_INFO
      &EACH
        MD 10
      &END EACH
    &END PROGRAM_RUN_INFO
  &END PRINT
&END MD

&CONSTRAINT
  &FIXED_ATOMS
    LIST 1..48
    COMPONENTS_TO_FIX XYZ
  &END
&END

&PRINT
  &VELOCITIES
    &EACH
      MD 1
    &END EACH
  &END

```

```
    &TRAJECTORY
     &EACH
       MD 1
     &END EACH
    &END

    &FORCES
      &EACH
       MD 1
      &END EACH
    &END

  &END PRINT

&END MOTION

&FORCE_EVAL
  METHOD Quickstep
  &PRINT
    &FORCES ON
      &EACH
        QS_SCF 0
      &END
      FILENAME =${PROJECT_NAME}.forces
      ADD_LAST NUMERIC
    &END
    &DISTRIBUTION
    &END
  &END
  &DFT
    BASIS_SET_FILE_NAME ${data}/BASIS_MOLOPT
    POTENTIAL_FILE_NAME ${data}/GTH_POTENTIALS
    WFN_RESTART_FILE_NAME ./mag-wat-MD-RESTART.wfn

    CHARGE 0

    &MGRID
      CUTOFF 500
      NGRIDS 5
      REL_CUTOFF 50.
    &END MGRID

    &QS
      METHOD GPW
```

```
      EXTRAPOLATION  ASPC
      EXTRAPOLATION_ORDER  0
    &END QS

    &SCF
      MAX_SCF 50
      MAX_SCF_HIST 15
      SCF_GUESS RESTART

      &OT ON
        N_HISTORY_VEC 13
        MINIMIZER DIIS
        PRECONDITIONER FULL_SINGLE_INVERSE
        ENERGY_GAP 0.002
        STEPSIZE 0.01
        CHOLESKY INVERSE_DBCSR
      &END OT
      &OUTER_SCF
        MAX_SCF 80
        EPS_SCF 5.0E-5
      &END OUTER_SCF

      &PRINT
        &RESTART
          &EACH
            QS_SCF 50
          &END EACH
        &END RESTART
      &END PRINT
    &END SCF

    &XC
      &XC_FUNCTIONAL PBE
      &END XC_FUNCTIONAL

      &XC_GRID
        XC_SMOOTH_RHO NN50
        XC_DERIV NN50_SMOOTH
      &END
      &VDW_POTENTIAL
        DISPERSION_FUNCTIONAL PAIR_POTENTIAL
        &PAIR_POTENTIAL
          TYPE DFTD3
          PARAMETER_FILE_NAME ${data}/dftd3.dat
```

```
          REFERENCE_FUNCTIONAL PBE
          R_CUTOFF [angstrom] 12.0
        &END PAIR_POTENTIAL
      &END VDW_POTENTIAL
    &END XC

    &PRINT
      &MULLIKEN
        FILENAME=mulliken
        COMMON_ITERATION_LEVELS 10
        &EACH
          MD 1
        &END EACH
      &END MULLIKEN
    &END PRINT
  &END DFT
  &SUBSYS
    &CELL
ABC 19.0156192780 22.1237201691 50
      PERIODIC XYZ
    &END CELL

    &TOPOLOGY
      COORD_FILE_NAME ./${COORD_FILE}
      COORD_FILE_FORMAT XYZ
    &END
    &KIND Fe
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q16
    &END KIND

    &KIND Mg
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q10
    &END

    &KIND H
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q1
    &END KIND

    &KIND O
      BASIS_SET DZVP-MOLOPT-SR-GTH
      POTENTIAL GTH-PBE-q6
```

```
    &END KIND

  &END SUBSYS
&END FORCE_EVAL

@if ${RESTART} == 1
&EXT_RESTART
  RESTART_FILE_NAME ./${PROJECT_NAME}-1.restart
  RESTART_DEFAULT T
#  RESTART_VEL T
#  RESTART_POS T
&END EXT_RESTART
@endif
```

A.2.3 Well-Tempered Metadynamics with PLUMED

```
RESTART

mg: GROUP ATOMS=296
oatoms: GROUP ATOMS=309-324,357-966:3

mgsur: GROUP ATOMS=241-288
c1: COM ATOMS=241-288

d1: DISTANCE ATOMS=c1,mg COMPONENTS NOPBC
coor: COORDINATION GROUPA=mg GROUPB=mgsur R_0=0.34 NOPBC

# Activate well-tempered metadynamics
metad: METAD ARG=d1.z,coor ...
  PACE=100 HEIGHT=3.0 # Deposit a Gaussian every 100 time steps, with
    initial height equal to 3.0 kJ/mol.
  BIASFACTOR=20.0
  SIGMA=0.05,0.1
  FILE=HILLS GRID_MIN=0,0 GRID_MAX=6,12
  TEMP=300.0
...

# Print both collective variables and the value of the bias potential on
    COLVAR file
PRINT ARG=d1.z,coor,metad.bias FILE=COLVAR STRIDE=1
```

Appendix B

Supplementary Simulation Information

B.1 Charge Evolution in Magnesium Corrosion

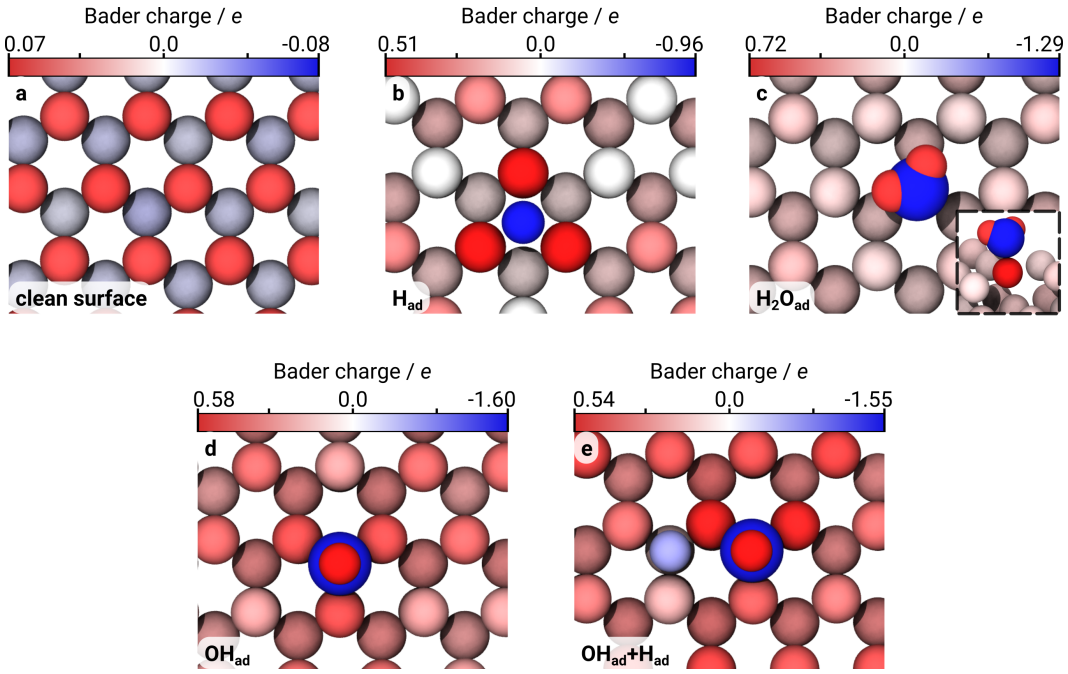


Figure B.1: Surface Bader charges. (a) clean Mg(0001) surface, (b) adsorbed hydrogen, (c) adsorbed water, (d) adsorbed hydroxide and (e) adsorbed hydroxide and hydrogen. Note that the color scales correspond to the respective maximum and minimum values.

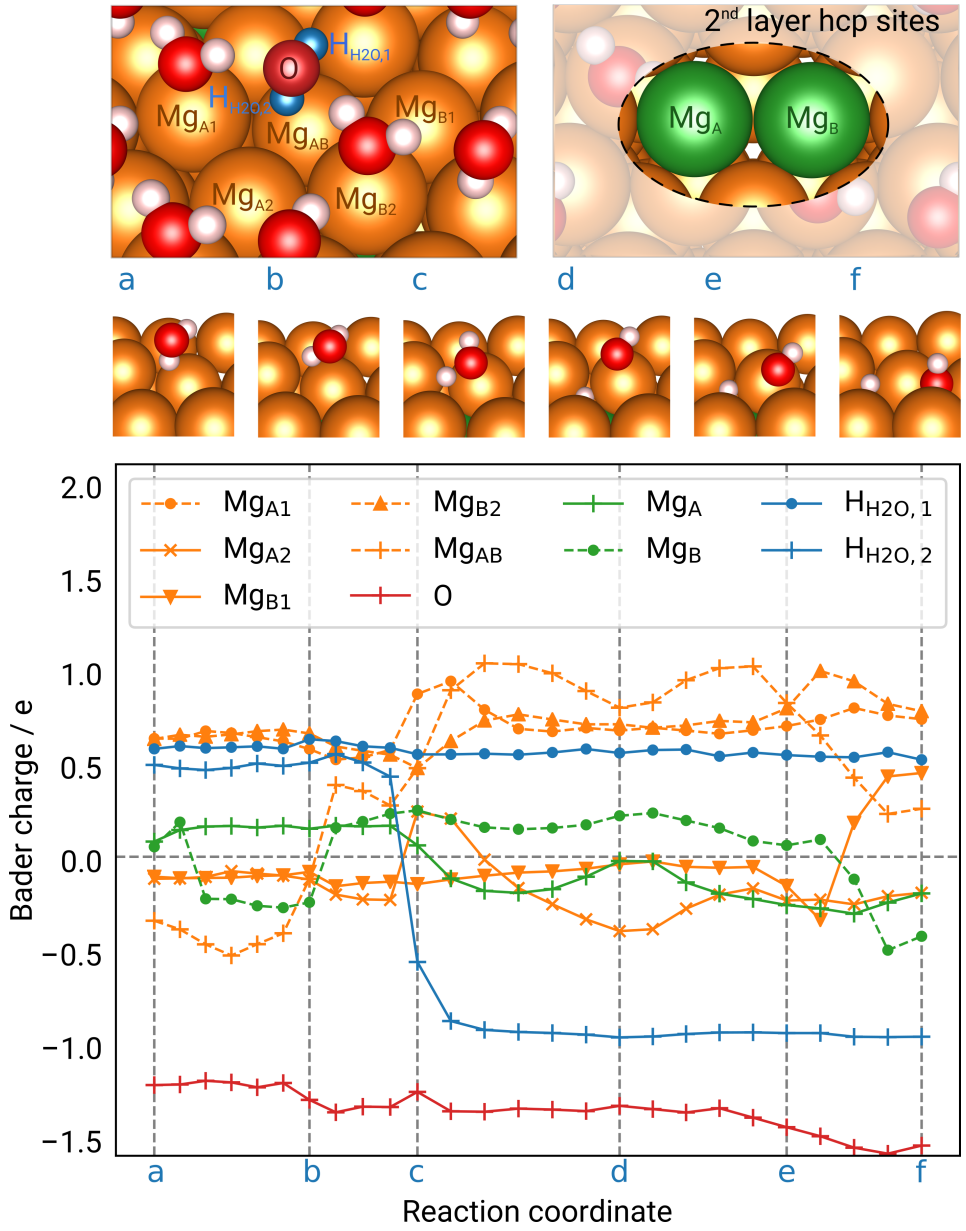


Figure B.2: Bader charge evolution during the NEB simulation of the Volmer step (including one water bilayer). Characteristic reaction steps (a-f) were visualized without the surrounding water bilayer. Atoms which appear to take part in the reaction were labeled, and their bader charge was plotted with respect to the reaction coordinate.

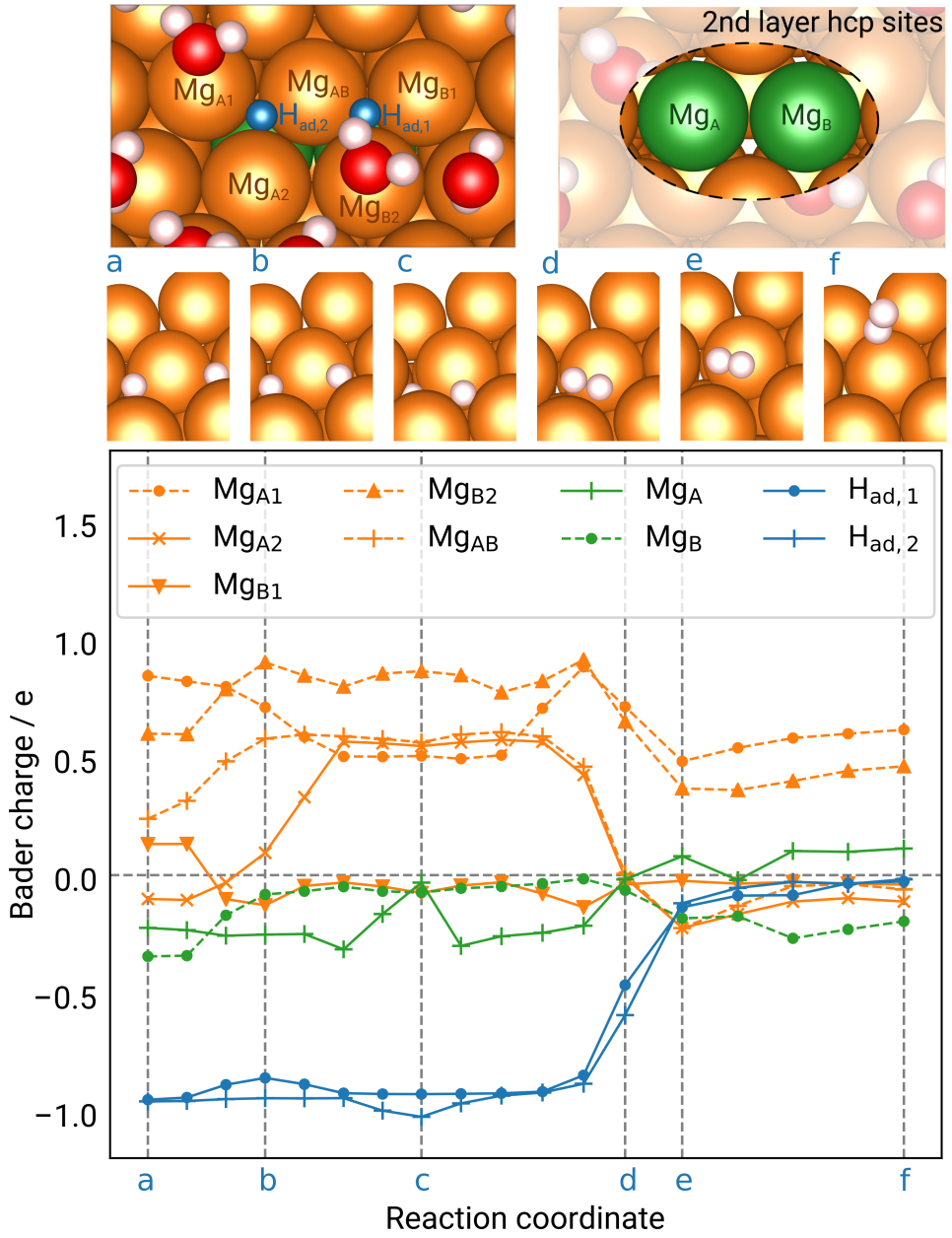


Figure B.3: Bader charge evolution during the NEB simulation of the Tafel step (including one water bilayer). Characteristic reaction steps (a-f) were visualized without the surrounding water bilayer. Atoms which appear to take part in the reaction were labeled, and their bader charge was plotted with respect to the reaction coordinate.

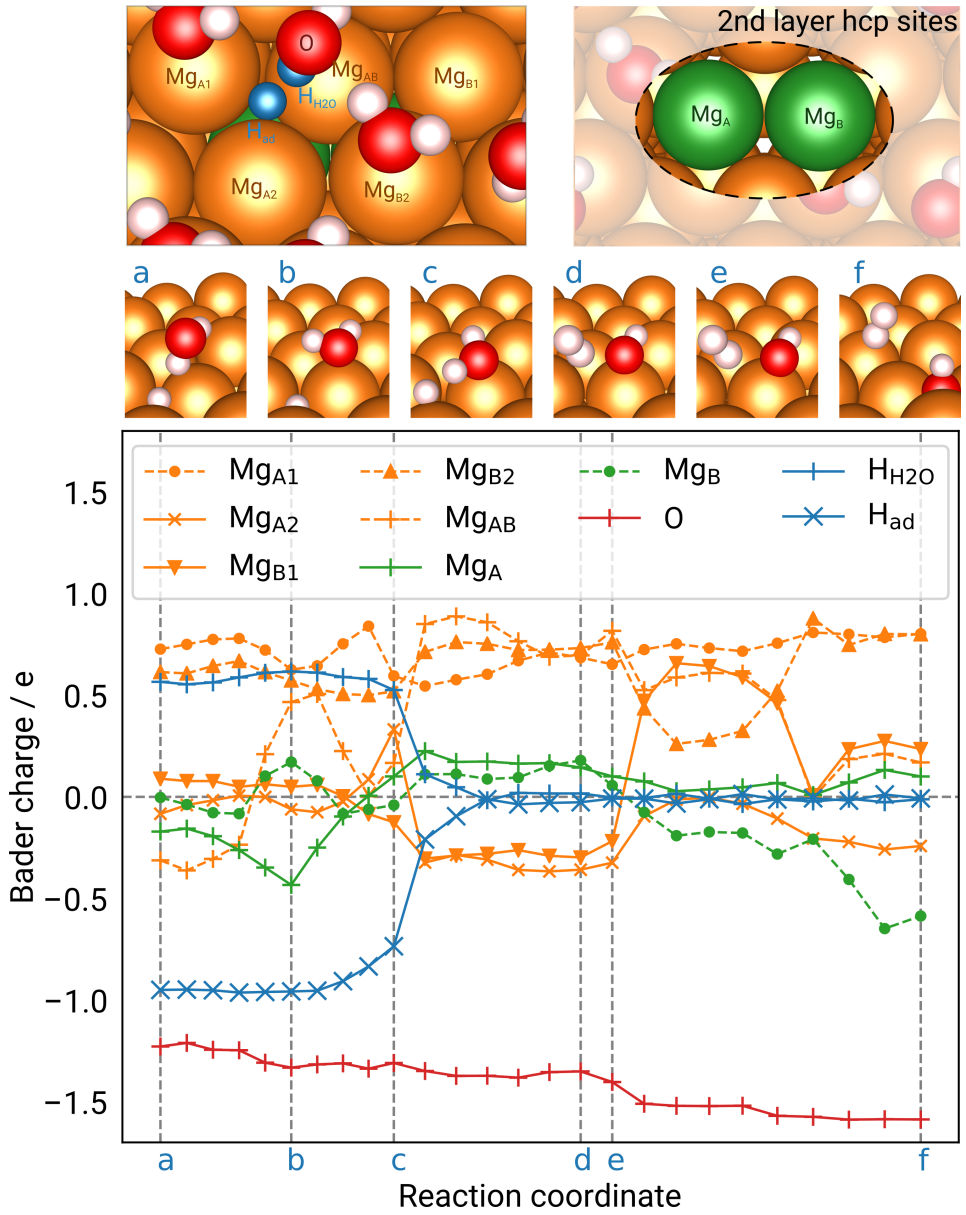


Figure B.4: Bader charge evolution during the NEB simulation of the Heyrovski step (including one water bilayer). Characteristic reaction steps (a-f) were visualized without the surrounding water bilayer. Atoms which appear to take part in the reaction were labeled, and their bader charge was plotted with respect to the reaction coordinate.

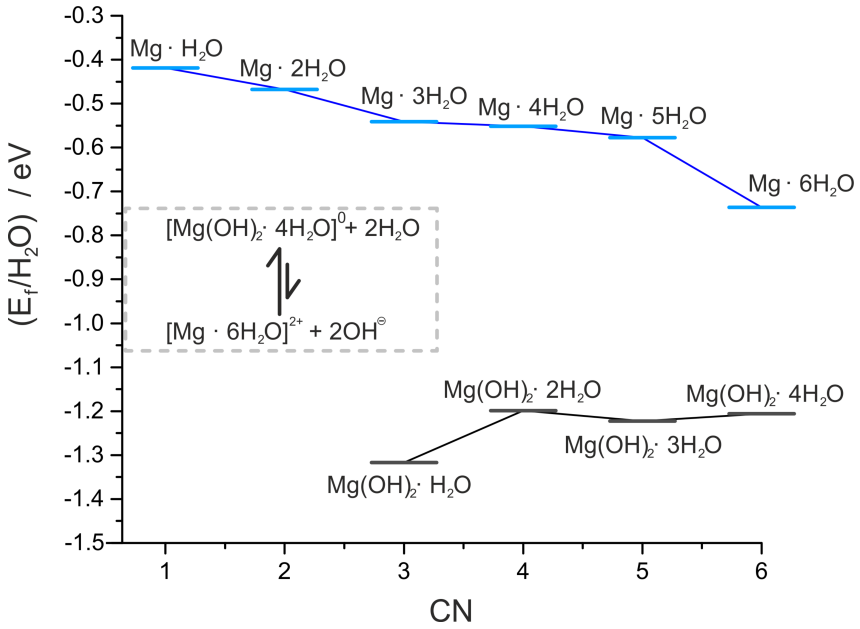


Figure B.5: Complex formation energies of Mg and Mg(OH)2 with water. CN = coordination number. Driving force for the formation of $Mg(OH)_2$ (inlay).

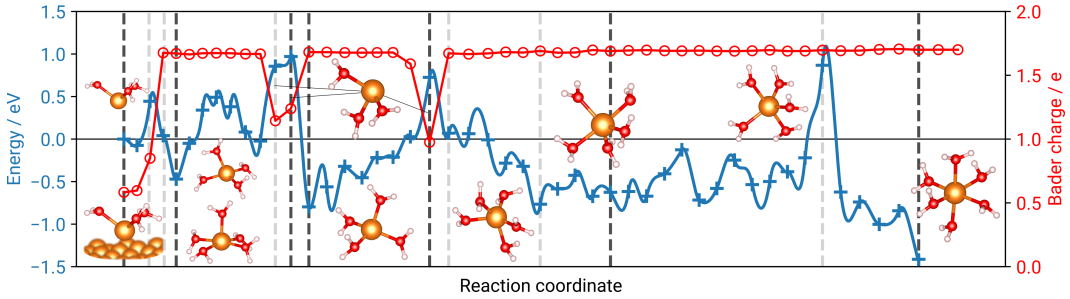


Figure B.6: Minimum energy pathway and bader charge evolution of Mg dissolution (including four water bilayers). Characteristic reaction steps were visualized including the first solvation shell of the Mg ion. The bader charge of the dissolving Mg ion was plotted with respect to the reaction coordinate.

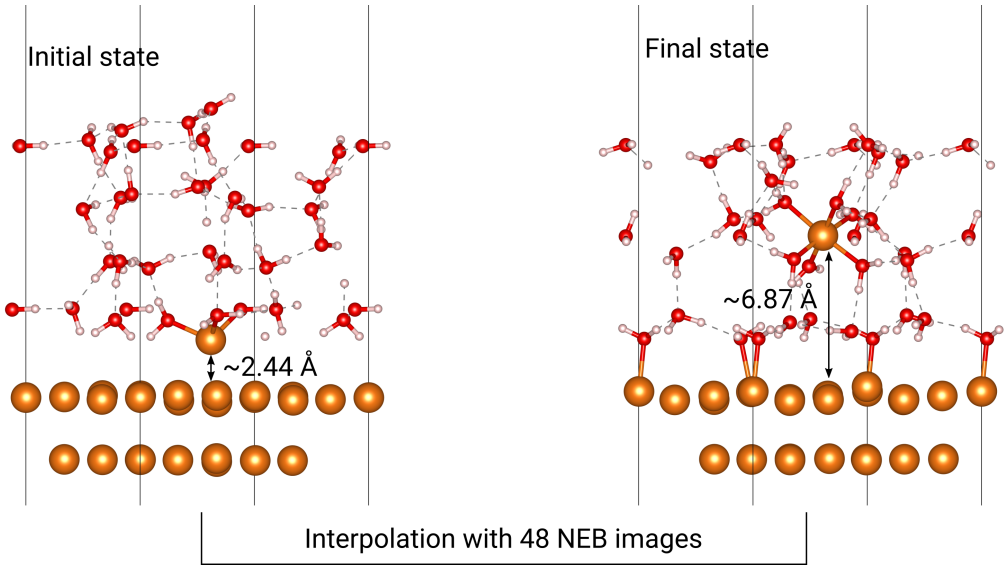


Figure B.7: Mg dissolution model for the NEB calculation. (left) Initial state. (right) Final state. Interpolation with 48 NEB images between both states defines the band for modeling the dissolution pathway.

B.2 Surface Adsorption

Table B.1: Adsorption energies and corrosion inhibition efficiencies of 45 compounds.

Compound	SMILES	IE / %	E_{ad} / eV
1,2-Hydroxyethylpiperazine	C1CN(CCN1)CCO	-96	-1.41
2,3-Pyrazinedicarboxylic acid	c1cnc(c(n1)C(=O)O)C(=O)O	59	-3.85
2-Amino-5-methylbenzoic acid	c1c(c(cc(c1)C)C(=O)O)N	78	-2.53
3,4-Dihydroxybenzoic acid	c1c(ccc(c1O)O)C(=O)O	-12	-2.54
3,5-Dimethylpyrazole	c1(cc(n[nH]1)C)	27	-0.35
3,5-Dinitrosalicylic acid	c1(c(c(cc(c1)N(=O)=O)C(=O)O)O)N(=O)=O	48	-3.41
3-Amino-1,2,4-triazole	c1[nH]nc(n1)N	23	-0.19
3-Methylcatechol	c1(c(c(ccc1)O)O)C	-1	-0.17
3-Methylsalicylic acid	c1(c(c(ccc1)C(=O)O)O)O)C	97	-2.99
4-Aminosalicylic acid	c1c(c(ccc1N)C(=O)O)O	85	-2.82
4-Hydroxybenzoic acid	c1c(ccc(c1)O)C(=O)O	10	-2.53
Acetamide	CC(=O)N	2	-0.35
Acetohydroxamic acid	CC(=O)NO	6	-0.23
Alanine	[C@H](C)(C(=O)O)N	-21	-2.62
Benzoic acid	c1cc(ccc1)C(=O)O	34	-2.58
Benzotriazole	c1ccc2c(c1)nn[nH]2	23	-0.44
Bipyridine	c1(ccccn1)c1ccccn1	31	-0.56
Bismuthiol	c1(nnc(s1)S)S	55	-0.70
Diglycolic acid	C(=O)(C)COCC(=O)C	86	-4.26
Ethylendiamine	C(CN)N	-61	-0.08
Fumaric acid	C(=C\C(=O)O)/C(=O)O	84	-4.00
Gluconic acid	C(C(C(C(C(C(=O)O)O)O)O)O)O)O	58	-2.49
Glycine	C(C(=O)O)N	-43	-2.37
Glycolic acid	C(=O)(C)CO	65	-2.74
Kojic acid	c1(cc(=O)c(co1)O)CO	46	-1.97
Leucine	[C@H](CC(C)C)(C(=O)O)N	-17	-2.49
Malonic acid	C(=O)(O)CC(=O)O	-26	-3.91
Mandelic acid	c1ccc(cc1)[C@@H](C(=O)O)O	61	-2.88
Oxalic acid	C(=O)(O)C(=O)O	86	-3.66

Panthenol	CC(C)(CO)C(C(=O)NCCCO)O	-8	-0.23
Phenylalanine	[C@H](Cc1ccccc1)(C(=O)O)N	16	-2.47
Phtahlic acid	c1c(c(ccc1)C(=O)O)C(=O)O	41	-3.85
Picolinic acid	c1(ccccn1)C(=O)O	21	-2.34
Piperazine	C1CNCCN1	-34	-0.07
Proline	C1CC[C@@H](C(=O)O)N1	-34	-2.10
Pyridazine	c1cnncc1	29	-0.84
Pyridine-2,5-dicarboxylic acid	c1(ccc(C(=O)O)cn1)C(=O)O	83	-3.80
Pyridine-2,6-dicarboxylic acid	c1(cccc(n1)C(=O)O)C(=O)O	90	-3.80
Quinic acid	C1C(C(C(CC1(C(=O)O)O)O)O)O)O	59	-2.77
Salicylaldehyde	c1c(c(ccc1)C=O)O	98	-2.46
Salicylic acid	c1c(c(ccc1)C(=O)O)O	92	-2.26
Triethylentetraamine	C(CNCCNCCN)N	-67	-0.30
Tris	C(CO)(CO)(CO)N	-72	-0.26
Uracil	[nH]1ccc(=O)[nH]c1=O	-26	-0.84
Valine	[C@H](C(C)C)(C(=O)O)N	-23	-2.21

Appendix C

Experimental Details

C.1 Corrosion Inhibition Experiments

In corrosion experiments, hydrogen evolution is measured in the presence of modulators and referenced to the sodium chloride electrolyte in the absence of these compounds. The effect of a modulator is quantified by the corrosion inhibition efficiency, Equation (2.50), which is positive for corrosion inhibitors and negative for corrosion accelerators (cf. Section 2.3):

$$\mathrm{IE} = \frac{V_{\mathrm{H_2}}^{0} - V_{\mathrm{H_2}}^{\mathrm{mod}}}{V_{\mathrm{H_2}}^{0}} \cdot 100\%. \tag{C.1}$$

Eudiometers (Art. No. 259110-500 from Neubert-Glas, Germany) were used for these investigations. A flask below the eudiometer was filled with a piece of the bulk Mg sample and 500 ml of electrolyte (0.5 wt.% NaCl) without (denoted hereafter as reference) and with addition of a dissolution modulator. The reference value was determined from the normalized volume of hydrogen evolved (V_{0,H_2}) after 20 h of immersion. Most Mg samples were 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 and the volume of evolved hydrogen quantified (V_{mod,H_2}). In additional experiments extending the given database, water displaced by evolved hydrogen was automatically quantified (SKX series from OHAUS coupled with USB data logger OHAUS 30268984) and the data recorded for further processing using an in-house Python script.[17] The testing time is considered to be sufficient as the hydrogen evolution rate is in a steady state after roughly 10 h.[14,17]

Table C.2: Experimental results of the performed hydrogen evolution experiments for validation of the KRR model. Inhibition efficiency is abbreviated as IE; a, b and c define the length, width, and thickness of the CPMg220 bulk sample, respectively.

Compound	H_2 / mL	H_2 / mL cm^{-2}	IE / %	final pH	*a* / mm	*b* / mm	*c* / mm	surface / cm^{-2}
0.5% NaCl	83.53	18.25	—	10.38	13.30	11.84	2.84	4.58
	95.74	26.87	—	10.42	10.44	9.71	3.81	3.56
	118.19	25.36	—	10.56	13.33	11.82	3.00	4.66
Reference	—	23.49	—	—	—	—	—	—
Dimethyl-	73.36	14.63	37.70	9.20	13.68	12.41	3.10	5.01
5-hydroxyiso-	45.07	12.78	45.60	9.17	10.30	9.44	4.01	3.53
phthalate	112.31	22.79	3.00	9.29	12.63	13.19	3.09	4.93
1,2,3,4-	67.76	14.68	37.50	9.57	13.25	11.77	2.99	4.62
Butanetetra-	60.93	17.64	24.90	7.71	10.33	9.53	3.74	3.45
carboxylic acid	90.63	18.84	19.80	10.05	13.15	12.37	3.05	4.81
β-Alanine	206.45	44.66	-90.10	10.06	11.76	13.32	2.97	4.62
	142.94	41.28	-75.70	10.01	10.30	9.24	3.99	3.46
	208.62	44.90	-91.10	10.14	11.79	13.30	3.01	4.65
2-Nitro-	38.96	8.07	65.60	10.95	13.07	12.40	3.11	4.83
terephthalic acid	25.48	6.89	70.70	10.88	10.46	9.62	4.20	3.70
	37.17	7.78	66.90	10.89	13.33	11.75	3.28	4.78
2-Aminopyridine-	49.95	10.72	54.40	10.45	13.30	11.80	3.03	4.66
3-carboxylic acid	40.44	11.20	52.30	8.54	10.41	9.62	4.01	3.61
	50.31	10.77	54.20	9.34	13.32	11.83	3.02	4.67
Pyrazine-	64.32	14.53	38.20	10.58	11.64	13.14	2.76	4.43
carboxylic acid	215.72	34.92	-48.60	10.61	15.58	14.40	2.82	6.18
	140.28	29.33	-24.80	10.67	13.21	12.53	2.86	4.78
	63.62	14.49	38.30	10.32	12.08	13.35	2.29	4.39
Benzamide	424.20	37.79	-60.90	10.59	19.13	19.86	4.65	11.22
	244.92	40.05	-70.50	10.56	14.36	15.51	2.78	6.12
	110.46	24.15	-2.80	10.62	13.36	11.87	2.78	4.57
	214.03	32.71	-39.20	10.30	14.50	16.28	2.96	6.54
Dimethyl-2,6-	24.11	5.09	78.30	8.82	12.23	13.37	2.86	4.73
pyridine-	49.00	10.38	55.80	8.80	12.25	13.24	2.90	4.72
dicarboxylate	48.59	9.99	57.50	8.65	12.69	12.60	3.29	4.86

1,2-Hexanediol	99.22	21.27	9.40	10.37	11.72	13.25	3.12	4.66
	155.92	26.84	-14.30	10.35	13.81	15.28	2.73	5.81
	108.5	21.15	10.00	10.38	13.23	12.89	3.29	5.13
N,N′-Trimethyl-	230.61	34.14	-45.30	10.59	14.39	15.56	3.80	6.75
eneurea	337.54	32.26	-37.30	10.56	18.66	20.52	3.58	10.46
	371.35	38.00	-61.80	10.62	19.03	19.05	3.31	9.77
	172.50	24.58	-4.60	10.29	16.44	14.45	3.67	7.02

Table C.1: Elemental composition. The elemental composition of commercially pure Mg containing 220 ppm iron impurities (CPMg220). Balance Mg.

Element content / wt. ‰									
Fe	Cu	Ni	Al	Mn	Ce	Zn	Si	Ca	Zr
0.220	0.005	< 0.002	0.130	0.150	< 0.004	0.004	0.053	0.004	0.005

Table C.3: Experimental results of the performed hydrogen evolution experiments for validation of the similarity-based discovery workflow. Inhibition efficiency is abbreviated as IE; a, b and c define the length, width, and thickness of the CPMg220 bulk sample, respectively.

Compound	H_2 / mL	H_2 / mL cm^{-2}	IE / %	final pH	a / mm	b / mm	c / mm	surface / cm^{-2}
0.5% NaCl	83.53	18.25	—	10.38	13.30	11.84	2.84	4.58
	95.74	26.87	—	10.42	10.44	9.71	3.81	3.56
	118.19	25.36	—	10.56	13.33	11.82	3.00	4.66
Reference	—	23.49	—	—	—	—	—	—
2-nitrobenzoic	95.59	14.80	37.10	10.77	14.72	16.18	2.76	6.47
acid	136.27	24.20	-3.20	10.93	15.57	13.92	2.18	5.62
	84.41	15.10	35.60	11.10	15.55	14.72	1.66	5.58
3-nitrophthalic	43.74	7.80	66.80	11.17	14.47	12.93	3.41	5.61
acid	33.07	5.17	78.00	11.19	15.42	16.03	2.30	6.39
	36.24	5.72	75.70	11.17	16.26	14.96	2.36	6.34
1-2-4-benzene	28.84	5.60	76.20	11.08	13.65	11.83	3.77	5.15
tricarboxylic acid	19.73	4.13	82.40	11.09	13.47	13.07	2.37	4.78
	24.15	4.14	82.40	11.11	15.23	14.83	2.20	5.84
4-nitrophthalic	54.06	11.09	52.80	11.08	13.96	11.29	3.41	4.87
acid	28.66	4.61	80.40	11.11	16.43	14.65	2.25	6.21
	26.67	5.73	75.60	11.10	13.70	12.60	2.29	4.66
3-hydroxy-4-nitro	118.68	17.75	24.40	9.62	19.30	8.85	5.81	6.69
benzoic acid	184.14	29.30	-24.70	9.51	14.89	16.47	2.20	6.28
	95.36	14.48	38.40	9.12	19.53	8.72	5.63	6.59

C.2 Discharge Experiments

The underlying discharge experiments were performed using a multi-channel IVIUM OctoStat in the conventional three-electrode arrangement. The experimental setup is illustrated in Figure C.1. An as-cast Mg-0.15Ca alloy was adopted as working electrode (elemental composition: 21 ppm Al, 14 ppm Fe, 2 ppm Ni, 2 ppm Cu, 160 ppm Mn, 56 ppm Si, and 2 ppm Zn) and embedded into resin. A surface area of 2.25 cm^2 was left exposed and was ground with SiC emery papers of up to 1200 grit. An Ag/AgCl electrode and coiled Pt wire were employed as the reference and the counter electrode, respectively. A solution of 3.5 wt.% NaCl was used as reference electrolyte. For each testing electrolyte, 0.1 M of electrolyte additive was added into NaCl solution, resulting in a total volume of 350 ml. All electrolytes were prepared with deionized water. Almost all the additive-containing electrolytes were neutralized to pH 7.0 ± 0.1 by adding 0.1 M NaOH, except for 1,4-bis(2-hydroxyethyl)piperazine (initial pH 10.17) and triethanolamine (initial pH 9.81). The discharge curves were recorded for 4 h at a constant current density of 5 mA cm^{-2}. During this period, the electrolyte was continuously stirred at 350 rpm. The discharge tests for each additive-containing electrolyte were repeated for three times to ensure reproducibility. Then, DP was determined as the average value of the observed discharge potential and UE was calculated according to

$$\mathrm{UE} = \frac{W_{\mathrm{theo}}}{W_{\mathrm{act}}} \cdot 100\%. \tag{C.2}$$

Here, W_{theo} and W_{act} represent theoretical and actual weight loss of the anode, respectively. W_{theo} is calculated according to Faraday's second law in electrolysis:

$$W_{\mathrm{theo}} = \frac{I \cdot M \cdot t}{F \cdot n} \tag{C.3}$$

where I is the applied current, t the discharge time, and F the Faraday constant. M and n represent the atomic mass and the ionic valence of the anode material, respectively. After the discharge products were removed by immersion in chromic acid for 5 min, W_{act} was obtained as the weight difference of the anode before and after the discharge test.

In the experiment, compound II.1 (4-isopropylbenzoic acid) showed unusual discharge behavior, resulting in an abnormal discharge potential (DP). The results from three parallel discharge tests are shown in Figure C.2. The DP of a Mg-0.15Ca anode in 4-isopropylbenzoate-containing electrolyte continuously jumps between -1.2 and -0.2 V vs. Ag/AgCl. 4-isopropylbenzoate is reported as an efficient corrosion inhibitor for pure Mg and AZ-series Mg alloys.[14] It is very likely that the gradual positive shift of discharge potential is caused

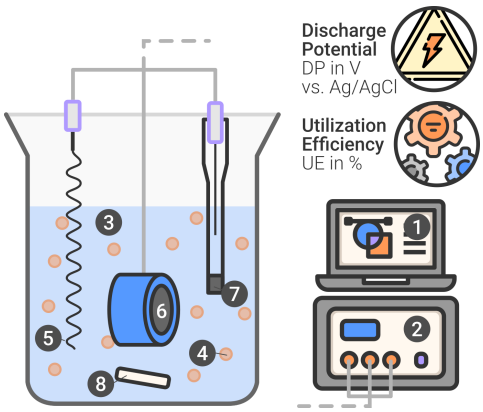


Figure C.1: Experimental setup for half-cell measurements. (1) PC; (2) potentiostat; (3) 3.5 wt.% NaCl; (4) electrolyte additives; (5) counter electrode (Pt wire); (6) working electrode (Mg-Ca alloys); (7) reference electrode (Ag/AgCl); (8) stir bar.

by the formation of a passive layer on the electrode surface. The sudden negative shift of discharge potential is due to the self-peeling of this passive layer after the accumulation of discharge products on the electrode surface. During the experiment, passive layer formation and self-peeling alternate, leading to the observed discharge behavior.

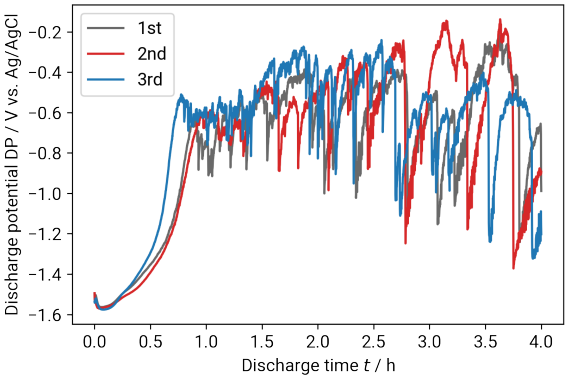


Figure C.2: Discharge curves for 4-isopropylbenzoic acid (compound II.1). Results are shown for three samples.

Table C.4: Experimental results of the performed discharge experiments. The horizontal line denotes the start of the test set.

Additive	Utilization Efficiency / %	Discharge Potential / V vs. Ag/AgCl
8-HQ-5-Sulfonic acid	59.72 ± 3.66	-1.55 ± 0.01
Salicylic acid	59.00 ± 3.22	-1.60 ± 0.01
L-Lysine	52.70 ± 8.44	-1.59 ± 0.02
2-hydroxy-5-nitrobenzoic acid	51.00 ± 1.83	-1.52 ± 0.02
3.5% NaCL	51.78 ± 2.98	-1.55 ± 0.02
Fumaric acid	52.13 ± 2.88	-1.44 ± 0.03
ACES	40.79 ± 5.17	-1.72 ± 0.06
Malonic acid	46.39 ± 3.43	-1.53 ± 0.02
5-Aminosalicylic acid	46.73 ± 1.80	-1.55 ± 0.01
Diglycolic acid	47.94 ± 4.75	-1.55 ± 0.01
5-methyl-2-nitrobenzoic acid	41.08 ± 8.31	-1.54 ± 0.01
Tapso	32.88 ± 3.65	-1.80 ± 0.03
Citric acid	46.48 ± 1.83	-1.56 ± 0.01
Maleic acid	42.73 ± 2.94	-1.46 ± 0.01
Ascorbic acid	41.15 ± 4.09	-1.55 ± 0.01
CHES	31.85 ± 1.88	-1.84 ± 0.01
EDTA	35.39 ± 2.66	-1.61 ± 0.02
L-Cysteine	36.54 ± 3.35	-1.63 ± 0.10
Thiosalicylic acid	41.76 ± 1.90	-1.57 ± 0.01
Oxalic acid	45.22 ± 5.23	-1.38 ± 0.08
5-Sulfosalicylic acid	39.23 ± 1.14	-1.65 ± 0.02
Tricine	32.26 ± 1.65	-1.61 ± 0.01
Triethylenetetramine	38.14 ± 6.52	-1.56 ± 0.01
3,5-DHB	39.35 ± 1.35	-1.65 ± 0.04
Tris	38.92 ± 1.27	-1.54 ± 0.03
3.4-Dimethylbenzoic acid	38.90 ± 4.70	-1.50 ± 0.04
HEPES	28.63 ± 2.61	-1.85 ± 0.03
4-HBA	38.69 ± 0.75	-1.58 ± 0.01
NTA	29.43 ± 4.46	-1.63 ± 0.03
2,6-DHB	38.45 ± 2.85	-1.61 ± 0.01
3,4-DHB	33.75 ± 2.22	-1.64 ± 0.00

2,3-DHB	31.11 ± 3.88	-1.61 ± 0.02
5-Methylsalicylic acid	35.90 ± 3.71	-1.54 ± 0.01
Bicine	26.08 ± 0.92	-1.71 ± 0.05
EGTA	25.93 ± 0.64	-1.61 ± 0.01
DTPA	25.74 ± 1.80	-1.61 ± 0.01
3-Methylsalicylic acid	33.78 ± 2.31	-1.27 ± 0.13
Tiron	16.35 ± 0.10	-1.68 ± 0.03
Taurine	24.52 ± 1.74	-1.86 ± 0.05
1,3,4-Thiadiazole-2,5-dithiol dipotassium salt	40.61 ± 3.71	-1.64 ± 0.22
Sodium-4-hydroxybenzenesulfonate	25.57 ± 1.03	-1.77 ± 0.04
Hydroquinonesulfonic acid	44.05 ± 5.82	-1.65 ± 0.04
3-Amino-4-hydroxybezenesulfonaic acid	28.98 ± 2.26	-1.85 ± 0.04
BES	33.00 ± 4.07	-1.81 ± 0.04
MOPS	29.60 ± 7.27	-1.81 ± 0.05
Sodium-1-butanesulfonate	41.74 ± 5.50	-1.54 ± 0.02
TES	30.12 ± 4.34	-1.84 ± 0.03
MES	44.19 ± 5.67	-1.56 ± 0.01
Sulfonllic acid	42.19 ± 4.17	-1.54 ± 0.01
3-Nitrobezenesulfonic acid	35.06 ± 1.29	-1.53 ± 0.01
1,4-Benzenedimethanol	41.44 ± 8.85	-1.55 ± 0.02
1,4-Bis(2-hydroxyethyl)piperazine	46.19 ± 5.31	-1.57 ± 0.01
4-Amino-3-hydroxy-1-naphthalene-sulfonic acid	40.78 ± 7.46	-1.78 ± 0.03
1-Pentanesulfonic acid	61.13 ± 4.51	-1.55 ± 0.01
4-Aminobutyric acid	42.50 ± 0.85	-1.69 ± 0.03
4-Isopropylbenzoic acid	38.71 ± 3.83	-0.74 ± 0.03
CAPS	49.87 ± 6.40	-1.61 ± 0.01
Gallic acid	39.37 ± 7.67	-1.62 ± 0.01
Triethanolamine	44.43 ± 1.56	-1.56 ± 0.01
2-Methoxybenzoic acid	50.67 ± 2.51	-1.54 ± 0.01

Appendix D

Machine Learning Models

D.1 Corrosion Inhibition Efficiency

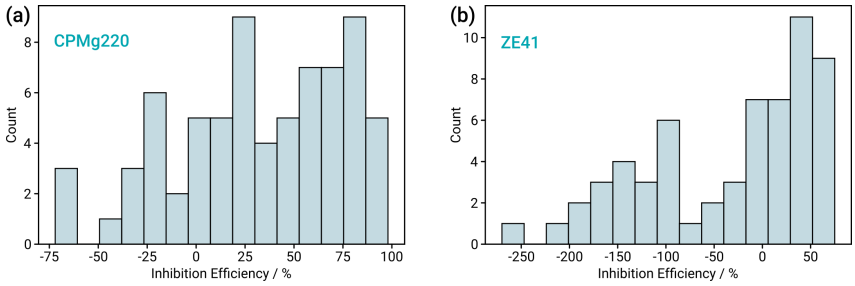


Figure D.1: Experimental data distribution of corrosion inhibitors. (a) Commercial purity magnesium with 220 ppm iron impurities. (b) Magnesium alloy ZE41.

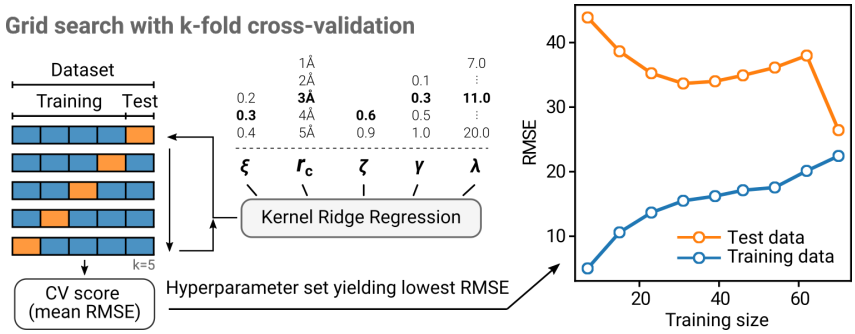


Figure D.2: Hyperparameter tuning of kernel ridge regression model to predict the IE. To minimize the prediction error for the inhibition efficiency, a number of hyperparameters for the SOAP-REMatch kernel and for the kernel ridge regression model are optimized using grid search with k-fold cross-validation.

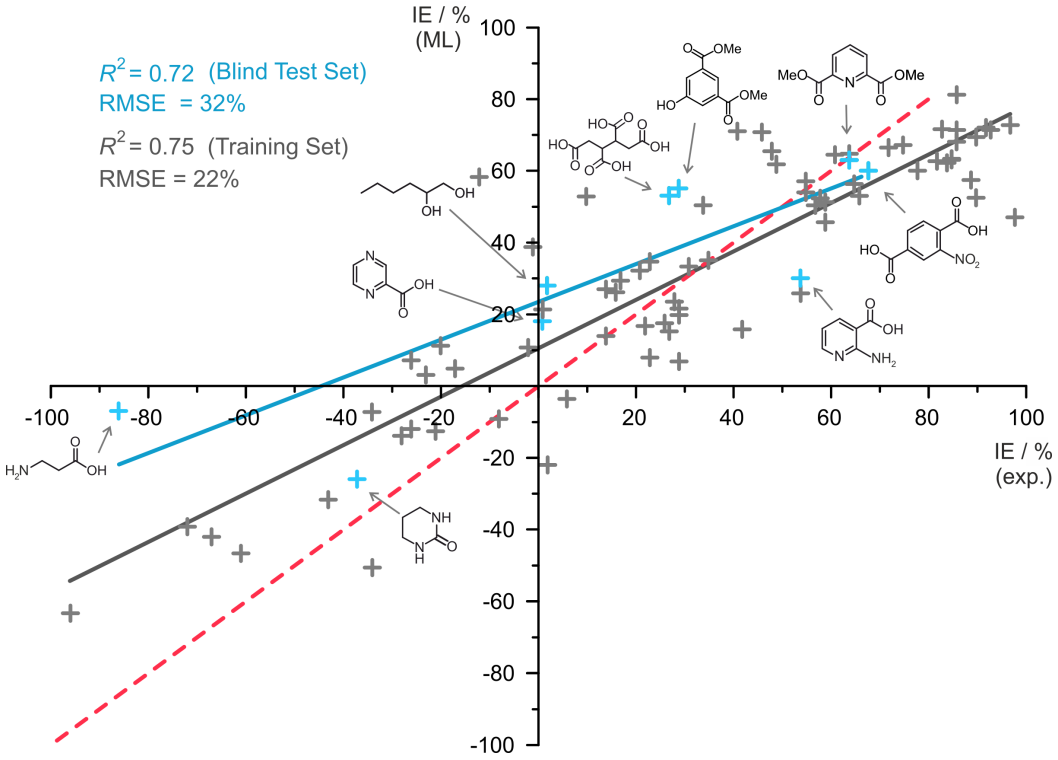


Figure D.3: Errors of the kernel ridge regression model predicting the IE. Training (gray) and test error (blue) by means of a linear least squares fit of the predicted and measured values. The training and test set include 78 and 10 structures, respectively. The red, dashed line represents a perfect correlation. Benzamide was omitted because of precipitation during the experiment.

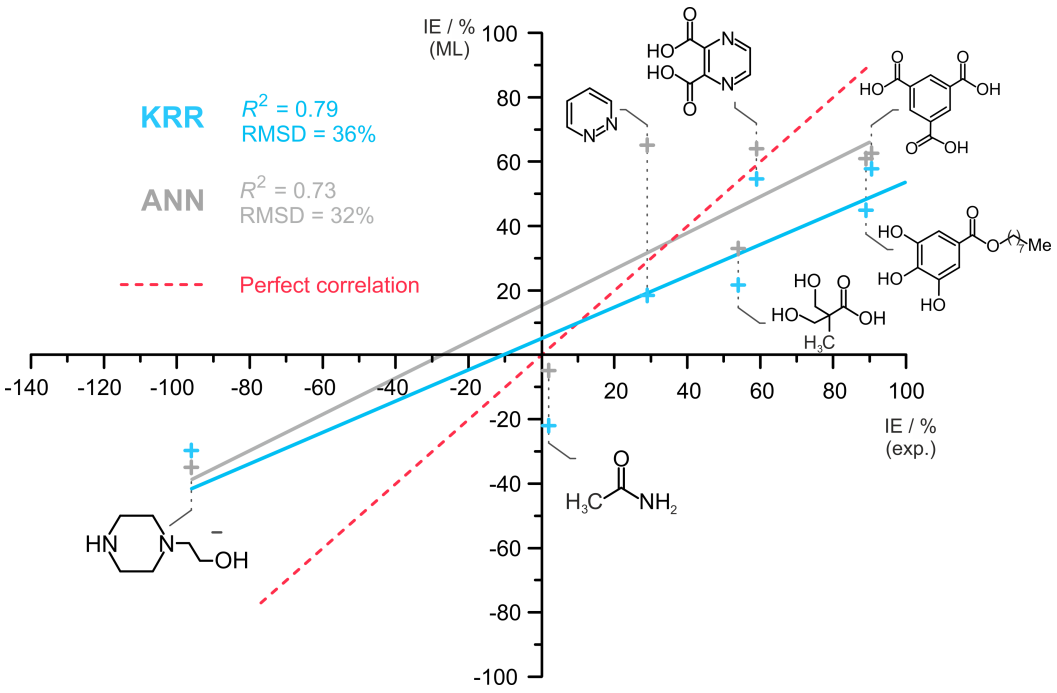


Figure D.4: Performance evaluation of the KRR model. Comparison of the prediction accuracy of the kernel ridge regression (KRR) (depicted in blue), based on the SOAP-REMatch kernel, and the developed artificial neural network (ANN) approach (depicted in gray) by means of a linear least squares fit of the predicted and measured values. The red, dashed line represents a perfect correlation.

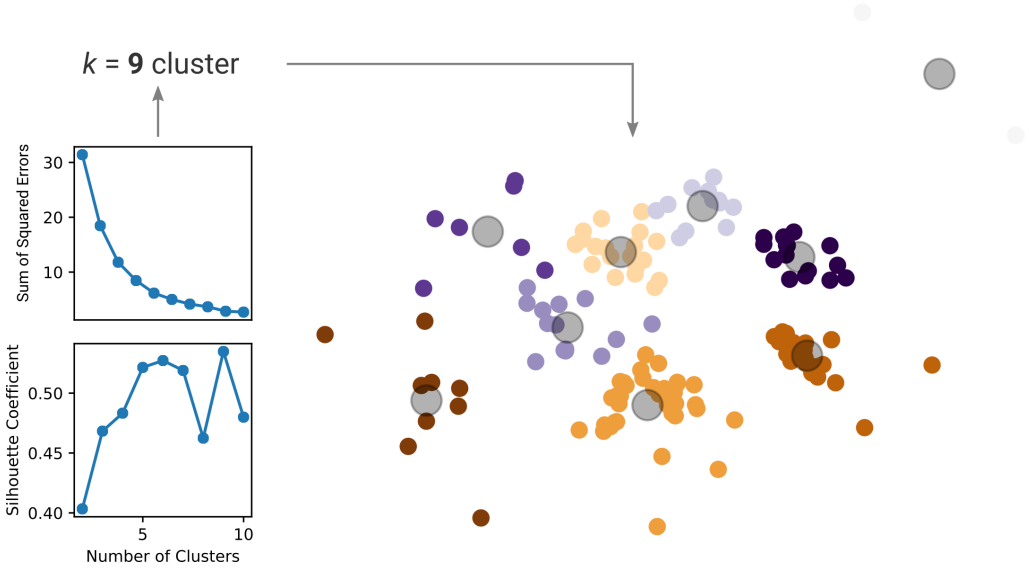


Figure D.5: K-means clustering of a sketch-map comprising 152 structures. After screening the sum of squared error and the silhouette coefficient with respect to the number of clusters, an optimal number of nine clusters was determined (centroids denoted as gray circles). Generally, the determined clusters agree well with the clusters already defined by visual inspection. Differences are that the orange cluster was split into a cluster **a** and **b** as cluster **a** solely consists out of yet untested structures. The *k*-means algorithm assigns four additional structures to cluster **e** that were not considered during the visual inspection due to a noticeable gap. Parts of the remaining clusters identified by the *k*-means algorithm are included in cluster **f** that was loosely defined based on the varying inhibition efficiencies within the cluster. Due to the nature of the *k*-means clustering algorithm, several structures that were not assigned to any cluster during visual inspection, are now assigned to a cluster. It is noteworthy that other clustering algorithms presumably would lead to differing cluster definitions.

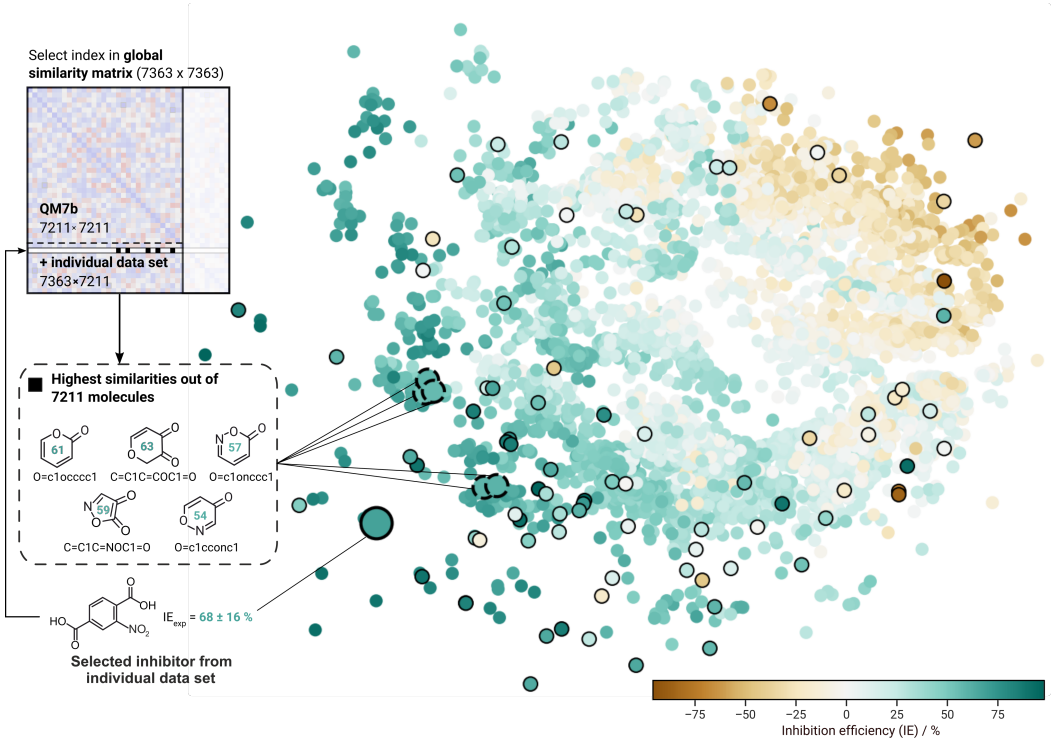


Figure D.6: Sketch-map comprising structures of the QM7b database and 152 individual chemicals (7363 structures in total). The dots are colored according to predicted IEs by means of KRR. The KRR model was trained on 78 experimental IEs (black-rimmed dots). For the sake of clarity, four outliers of the individual dataset and two of the QM7b database have been excluded from the visualization. By referencing the selected compound **5** (large black-rimmed dot) to the underlying SOAP kernel, five highly similar molecules can be determined (dashed black-rimmed dots) along with their predicted IE values.

Table D.1: List of all 152 compounds included in the sketch-map along with their SMILES strings. For structures where no experimental inhibition efficiency IE_{exp} is available a predicted inhibition efficiency IE_{KRR} based on kernel ridge regression (KRR) is presented. For the blind validation test structures, both values are shown.

Compound	SMILES	IE_{exp} / %	IE_{KRR} / %
2-Amino-5-methyl-salicylic acid	c1c(c(cc(c1)C)C(=O)O)N	78	–
3-Amino-1,2,4-triazole	c1[nH]nc(n1)N	23	–
3-Methyl-salicylic acid	c1(c(c(ccc1)C(=O)O)O)C	97	–
3-Methyl-catechol	c1(c(c(ccc1)O)O)C	-1	–
4-Amino-salicylic acid	c1c(c(ccc1N)C(=O)O)O	85	–
4-Hydroxybenzoic acid	c1c(ccc(c1)O)C(=O)O	10	–
5-Amino-salicylic acid	c1c(c(cc(c1)N)C(=O)O)O	86	–
5-Methyl-salicylic acid	c1c(c(cc(c1)C)C(=O)O)O	72	–
5-Nitrobarbituric acid	C1(=O)C(C(=O)NC(=O)N1)N(=O)=O	82	–
3,4-Dipicolinic acid	c1(c(cccn1)C(=O)O)C(=O)O	64	–
2,5-Dipicolinic acid	c1(ccc(C(=O)O)cn1)C(=O)O	83	–
2,6-Dipicolinic acid	c1(cccc(n1)C(=O)O)C(=O)O	90	–
3,4-Hydroxybenzoic acid	c1c(ccc(c1O)O)C(=O)O	-12	–
3,5-Dimethyl-pyrazole	c1(cc(n[nH]1)C)	27	–
5,5-Dimethylhydation	C1(CNC(=O)N1)(C)C	-28	–
3,5-Dinitro-salicylic acid	c1(c(c(cc(c1)N(=O)=O)C(=O)O)O)N(=O)=O	48	–
Acetohydroxamic acid	CC(=O)NO	6	–
Alanine	[C@H](C)(C(=O)O)N	-21	–
Asparagine	[C@H](CC(=O)N)(C(=O)O)N	1	–
Aspartic acid	[C@H](CC(=O)O)(C(=O)O)N	14	–
Benzoic acid	c1cc(ccc1)C(=O)O	34	–
Benzotriazole	c1ccc2c(c1)nn[nH]2	23	–
Bipyridine	c1(ccccn1)c1ccccn1	31	–
Bismuthiol	c1(nnc(s1)S)S	55	–
Citric acid	C(C(CC(=O)O)(C(=O)O)O)C(=O)O	65	–
Cystein	[C@H](CS)(C(=O)O)N	29	–

CyDTA	C1CCC(C(C1)N(CC(=O)O)CC(=O)O)N(CC(=O)O)CC(=O)O	-2	–
Dodecylbenzenesulfonic acid	c1(ccc(cc1)CCCCCCCCCCCC)S(=O)(=O)O	93	–
Dimethylglyoxime	C(=N\O)(/C(=N/O)/C)C	22	–
Diglycolic acid	C(=O)(C)COCC(=O)C	86	–
Ethylenediamine	C(CN)N	-61	–
EDTA	C(CN(CC(=O)O)CC(=O)O)N(CC(=O)O)CC(=O)O	29	–
Fumaric acid	C(=C\C(=O)O)/C(=O)O	84	–
Gluconic acid	[C@H]([C@@H]([C@H]([C@H](CO)O)O)O)(C(=O)O)O	58	–
Glutamic acid	[C@H](CCC(=O)O)(C(=O)O)N	7	–
Glycine	C(C(=O)O)N	-43	–
Glycolic acid	C(=O)(C)CO	65	–
Histidine	[C@H](Cc1[nH]cnc1)(C(=O)O)N	26	–
Kojic acid	c1(cc(=O)c(co1)O)CO	46	–
Lauroylsarcosine	C(=O)(O)CN(C)C(=O) CCCCC-CCCCCC	17	–
Leucine	[C@H](CC(C)C)(C(=O)O)N	-17	–
Lysine	[C@H](CCCCN)(C(=O)O)N	60	–
Maleic acid	C(=C\C(=O)O)\C(=O)O	52	–
Malonic acid	C(=O)(O)CC(=O)O	-26	–
Maltol	c1(c(c(=O)cco1)O)C	35	–
Mandelic acid	c1ccc(cc1)[C@@H](C(=O)O)O	61	–
Methionine	[C@H](CCSC)(C(=O)O)N	28	–
Nicotinic acid	c1c(cccn1)C(=O)O	66	–
Nitrilotriacetic acid	N(CC(=O)O)(CC(=O)O)CC(=O)O	-20	–
Oxalic acid	C(=O)(O)C(=O)O	86	–
para-tert-Butylbenzoic acid	c1c(ccc(c1)C(C)(C)C)C(=O)O	57	–
para-Toluic acid	c1c(ccc(c1)C)C(=O)O	49	–
Panthenol	C(=O)(NCCCO)[C@H](C(C)(CO)C)O	-8	–
Phenylalanine	[C@H](Cc1ccccc1)(C(=O)O)N	16	–
Phthalic acid	c1c(c(ccc1)C(=O)O)C(=O)O	41	–
Picolinic acid	c1(ccccn1)C(=O)O	21	–

Piperazine	C1CNCCN1	-34	–
Proline	C1CC[C@@H](C(=O)O)N1	-34	–
Quinaldic acid	c1(ccc2c(n1)cccc2)C(=O)O	85	–
Quinic acid	[C@@]1(CC[C@H]([C@@H]([C@@H]1O)O)O)(O)C(=O)O	59	–
Salicylaldehyde	c1c(c(ccc1)C=O)O	98	–
Salicylaldoxime	c1c(c(ccc1)/C=N/O)O	55	–
Salicylhydroxamic acid	c1c(c(ccc1)C(=O)NO)O	75	–
Salicylic acid	c1c(c(ccc1)C(=O)O)O	92	–
Serine	[C@H](CO)(C(=O)O)N	14	–
Tartaric acid	[C@H]([C@H](C(=O)O)O)(C(=O)O)O	64	–
Threonine	[C@H]([C@@H](C)O)(C(=O)O)N	42	–
Triethylentetraamine	C(CNCCNCCN)N	-67	–
Tris(hydroxymethyl)aminomethane	C(CO)(CO)(CO)N	-72	–
Uracil	[nH]1ccc(=O)[nH]c1=O	-26	–
Valine	[C@H](C(C)C)(C(=O)O)N	-23	–
Trimesic acid	c1c(cc(cc1C(=O)O)C(=O)O)C(=O)O	90 ± 9	–
n-Octyl gallate	c1(c(c(cc(c1)C(=O)OCCCCCCCCC)O)O)O	89 ± 8	–
Dimethylolpropionic acid	C(C(=O)O)(C)(CO)CO	54 ± 23	–
Acetamide	CC(=O)N	2 ± 19	–
Pyridazine	c1cnncc1	29 ± 8	–
2,3-Pyrazinedicarboxylic acid	c1cnc(c(n1)C(=O)O)C(=O)O	59 ± 13	–
1-(2-Hydroxyethyl)piperazine	C1CN(CCN1)CCO	-96 ± 43	–
Dimethyl-2-5-pyridine-dicarboxylate	COC(=O)c1ccc(nc1)C(=O)OC	–	58
Dimethyl-2-6-pyridinedicarboxylate	COC(=O)c1cccc(n1)C(=O)OC	64±19	63
Methyl-pyridazine-4-carboxylate	COC(=O)c1ccnnc1	–	53
Pyrazinecarboxylic acid	OC(=O)c1cnccn1	1±41	18
Pyrimidine-2-carboxylic acid	OC(=O)c1ncccn1	–	20
4-Pyridazinecarboxylic acid	OC(=O)c1ccnnc1	–	43
Pyridazine-3-carboxylic acid	OC(=O)c1cccnn1	–	42
2-3-Pyrazinedicarbonitrile	N#Cc1nccnc1C#N	–	52
Pyrimidine	c1cncnc1	–	-1

4-4-Dimethoxy-2-2-bipyridine	COc1ccnc(c1)-c2cc(OC)ccn2	–	48
2-2-Bipyridine-4-4-dicarboxylic acid	OC(=O)c1ccnc(c1)-c2cc(ccn2)C(O)=O	–	47
2-2-Bipyridyl	c1ccc(nc1)-c2ccccn2	–	36
2-Ethyl-1-3-hexanediol	CCCC(O)C(CC)CO	–	30
2-4-Diethyl-1-5-pentanediol	CCC(CO)CC(CC)CO	–	22
1-5-Hexanediol	CC(O)CCCCO	–	24
2-Ethyl-1-hexanol	CCCCC(CC)CO	–	29
3-Methyl-1-5-pentanediol	CC(CCO)CCO	–	13
1-2-Hexanediol	CCCCC(O)CO	2±19	28
1-2-6-Hexanetriol	OCCCCC(O)CO	–	23
N-Methyldiethanolamine	CN(CCO)CCO	–	-49
β-Alanine	NCCC(O)=O	-86±17	-7
2-6-Pyridinedimethanol	OCc1cccc(CO)n1	–	18
1-2-Dihydroxybenzene	Oc1ccccc1O	–	6
Isopentylamine	CC(C)CCN	–	-3
2-Methylbutylamine	CCC(C)CN	–	-3
1-4-Benzodioxan-6-amine	Nc1ccc2OCCOc2c1	–	42
3-4-Dimethoxyaniline	COc1ccc(N)cc1OC	–	38
1-4-Benzodioxan-5-carboxylic acid	OC(=O)c1cccc2OCCOc12	–	57
3-5-Dimethoxyaniline	COc1cc(N)cc(OC)c1	–	38
Piperonylamine	NCc1ccc2OCOc2c1	–	41
3-4-Dimethoxybenzylamine	COc1ccc(CN)cc1OC	–	33
6-Aminopyridine-3-carboxylic acid	Nc1ccc(cn1)C(O)=O	–	28
5-Aminopyridine-2-carboxylic acid	Nc1ccc(nc1)C(O)=O	–	22
2-Aminopyridine-3-carboxylic acid	Nc1ncccc1C(O)=O	54±16	30
2-Amino-isonicotinic acid	Nc1cc(ccn1)C(O)=O	–	29
4-Amino-nicotinic acid	Nc1ccncc1C(O)=O	–	25
N,*N*′-Trimethyleneurea	O=C1NCCCN1	-37±26	-26
N,*N*′-Bis(hydroxymethyl)urea	OCNC(=O)NCO	–	-40
2-Hydroxyethylurea	NC(=O)NCCO	–	-34
D-(+)-Mannose	OC[C@H]1OC(O)[C@@H](O)[C@@H](O)[C@@H]1O	–	23
D-(+)-Galactose	OC[C@@H](O)[C@H](O)[C@H](O)[C@@H](O)C=O	–	32
4(5)-Hydroxymethylimidazole	OCc1c[nH]cn1	–	-8

4-Methoxy-pyridine-2-carboxylic acid	COc1ccnc(c1)C(O)=O	–	38
Terephthalic acid	OC(=O)c1ccc(cc1)C(O)=O	–	39
2-5-Dihydroxyterephthalic acid	OC(=O)c1cc(O)c(cc1O)C(O)=O	–	58
Isophthalic acid	OC(=O)c1cccc(c1)C(O)=O	–	40
2-Aminoterephthalic acid	Nc1cc(ccc1C(O)=O)C(O)=O	–	48
2-Nitroterephthalic acid	OC(=O)c1ccc(C(O)=O)c(c1)[N+]([O-])=O	68±16	60
5-Methoxyisophthalic acid	COc1cc(cc(c1)C(O)=O)C(O)=O	–	52
Dimethyl-5-methoxyisophthalate	COC(=O)c1cc(OC)cc(c1)C(=O)OC	–	60
4-Methoxyisophthalic acid	COc1ccc(cc1C(O)=O)C(O)=O	–	58
Dimethyl-5-hydroxyisophthalate	COC(=O)c1cc(O)cc(c1)C(=O)OC	29±24	55
5-Methylisophthalic acid	Cc1cc(cc(c1)C(O)=O)C(O)=O	–	43
5-*tert*-Butylisophthalic acid	CC(C)(C)c1cc(cc(c1)C(O)=O)C(O)=O	–	38
5-Methoxyisophthalic acid	COc1cc(cc(c1)C(O)=O)C(O)=O	–	52
2-Hydroxyterephthalic acid	OC(=O)c1ccc(C(O)=O)c(O)c1	–	42
1-2-4-Benzenetricarboxylic acid	OC(C1=C(C(O)=O)C=C(C(O)=O)C=C1)=O	–	53
Mellitic acid	OC(=O)c1c(C(O)=O)c(C(O)=O)c(C(O)=O)c(C(O)=O)c1C(O)=O	–	79
1-2-3-4-Butanetetracarboxylic acid	OC(=O)CC(C(CC(O)=O)C(O)=O)C(O)=O	27±17	53
tricarballylic acid	OC(=O)CC(CC(O)=O)C(O)=O	–	46
1-2-4-Butanetricarboxylic acid	OC(=O)CCC(CC(O)=O)C(O)=O	–	42
1-2-4-5-Benzenetetracarboxylic acid	OC(=O)c1cc(C(O)=O)c(cc1C(O)=O)C(O)=O	–	66
1-4-Bis(2-hydroxyethyl)piperazine	OCCN1CCN(CCO)CC1	–	-44
N-*N*′-Bis(2-hydroxyethyl)ethylenediamine	OCCNCCNCCO	–	-41
1-[*N*,*N*-Bis(2-hydroxyethyl)amino]-2-propanol	CC(O)CN(CCO)CCO	–	-35
Acetanilide	CC(=O)Nc1ccccc1	–	37
Benzamide	NC(=O)c1ccccc1	-43±30	44
4-Nitroacetanilide	CC(=O)Nc1ccc(cc1)[N+]([O-])=O	–	40

3-Nitroacetanilide	CC(=O)Nc1cccc(c1)[N+]([O-])=O	–	41
3-Aminoacetanilide	CC(=O)Nc1cccc(N)c1	–	26
2-Aminoacetanilide	CC(=O)Nc1ccccc1N	–	21
4-Aminoacetanilide	CC(=O)Nc1ccc(N)cc1	–	24
ortho-Toluic acid	Cc1ccccc1C(O)=O	–	35
2-Methoxybenzoic acid	COc1ccccc1C(O)=O	–	50

D.2 Battery Anode Performance

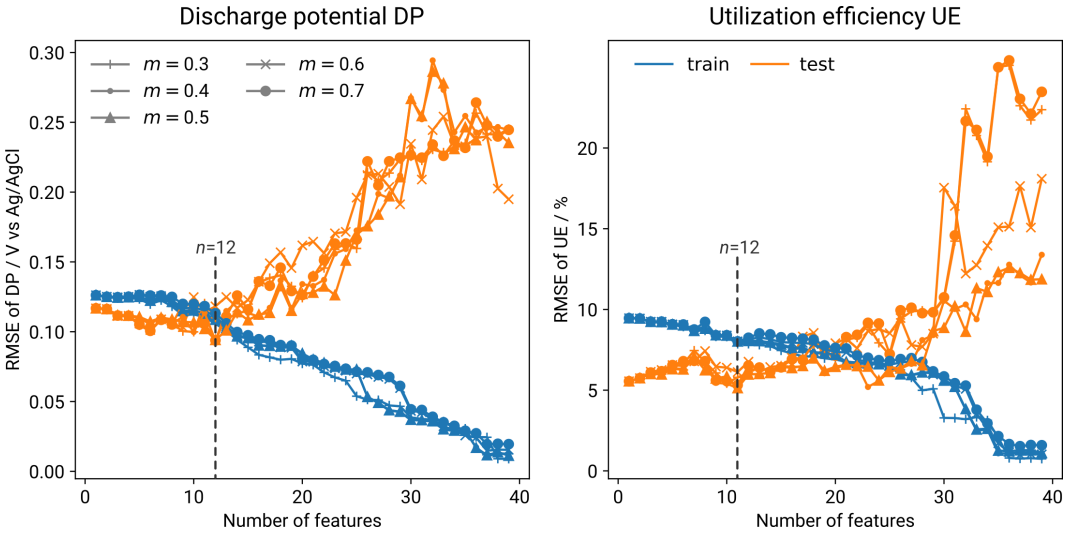


Figure D.7: RMSE of DP and UE predictions with respect to the number of features. The RMSEs were determined by PCov-CUR matrix decomposition in a 5-fold cross validation. Results are shown for different mixing factors used to balance the KPCA and KRR components of the KPCovR model. Training and test errors are marked as blue and orange, respectively. Model training on 12 features leads to a minimal test error for DP and UE.

Table D.2: Hyperparameter tuning results for the battery anode performance models. Hyperparameters are given for the SOAP-based S- and distinct descriptor-based D-model. Parameter definitions and their associated equations are based on Musil et al.[70] and Helfrecht et al.[75]

Parameter	S_K	D_K
cutoff radius r_c	8.0	–
Gaussian width ξ	0.3	–
power ζ	2.0	–
regularization parameter λ	0.05	0.30
mixing parameter α	0.5	0.5
number of features n	–	12

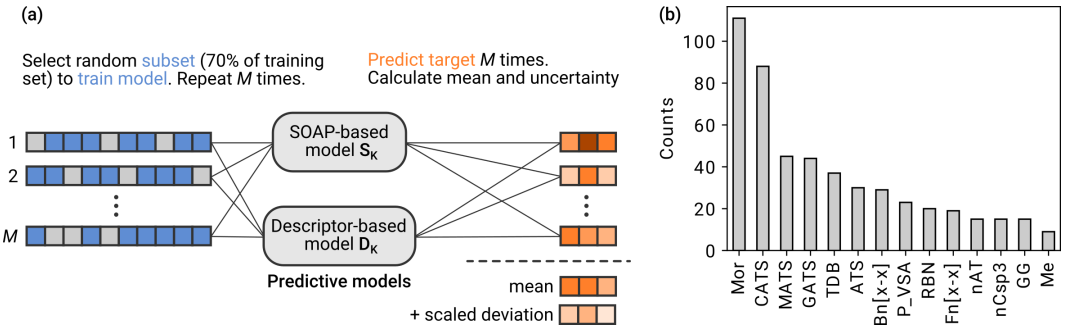


Figure D.8: Uncertainty estimation for the predictions of the battery anode performance models, S_K and D_K. (a) M random subsets are selected from the training dataset and used to train M S- and D-models. The target is predicted by the committee model, defined as the average of all M subset predictions. The single point uncertainty can be calculated and calibrated as described by Musil et al. [70] and Imbalzano et al. [226]. (b) Distribution of features for $M = 50$ subset models during training of D_K. For each subset model, 12 features were selected by PCov-CUR matrix decomposition.