

Multi-scale modeling of layered double hydroxides for active corrosion protection

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Dr.-Ing.

der Christian-Albrechts-Universität

vorgelegt von

Xuejiao Li

aus

Hebei, VR China

2023



Prüfungskommission

Vorsitz: Prof. Dr. Lorenz Kienle

Betreuung: Prof. Dr. Mikhail Zheludkevich

der Zweitgutachter: Prof. Dr. Pascal Friederich

weitere Mitglieder: Prof. Dr.-Ing. Robert Meißner

Prof. Dr.-Ing. Stephan Wulfinghoff

Tag der mündlichen Prüfung: 16th October 2023

Eidesstattliche Erklärung

Hiermit erkläre ich, dass die beigelegte Dissertation, abgesehen von der Beratung durch die Betreuerin, nach Inhalt und Form meine eigene Arbeit ist.

Die Arbeit, ganz oder zum Teil, wurde nie schon einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt und ist abgesehen, von den im Anhang angegebenen Veröffentlichungen, nicht anderweitig zur Veröffentlichung vorgelegt worden.

Außerdem ist die Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden.

Hiermit erkläre ich, dass mir noch nie ein akademischer Grad entzogen wurde.

Unterschrift: Xuejiao Li

Geesthacht, October 30, 2023

Abstract

Corrosion protection is critical in almost all engineering applications whereas chromium(VI)-based solutions have been used to ensure structural integrity in light metal engineering in the last decades. The urgent need to eliminate this highly effective but equally toxic approach remains one of the major challenges that industry faces to-date, despite decades of efforts to identify environmentally benign alternatives. A promising strategy is based on the application of layered double hydroxides (LDHs) intercalated with potent corrosion inhibitors, which have shown a great potential to control the corrosion rates of aluminium and magnesium-based materials. However, the possible compositions to realize an LDH-based corrosion protection systems are essentially infinite and cannot be evaluated by experimental techniques alone because the space of possible materials is too vast to be explored, even by the most optimistic projections of high-throughput experimental methods alone. This thesis lays the foundation for a computational workflow based on atomistic simulation techniques and machine learning to accelerate the development and evaluation of LDH-based corrosion inhibition systems. The atomistic simulation results shed more light on the anion exchange mechanism and demonstrate how the water content in the interlayer region affects the structure and stability of LDHs. The results show that the initially intercalated anions are released first during the anion exchange process, followed by new anion intercalation, and finally the change in water content stabilizes the LDH system. Furthermore, the mobility of the intercalated anions may be an intrinsic factor influencing their release during anion exchange. In addition to the atomistic simulations, machine learning techniques were used to create a machine-learned force field (ML-FF) for LDH system with almost *ab initio* accuracy. Despite being trained on small size LDH structures, this ML-FF has demonstrated the ability to scale up to investigate large LDH systems, providing a powerful and accurate force field for further computer-assisted design of LDH-based systems. Another aspect of the presented work is the development of a quantitative structure-property relationship (QSPR) model for the pre-selection of effective corrosion inhibitors that extends the computer-aided LDH design by another element to render it more versatile. The QSPR-based models provided robust estimates for the majority of the blindly tested compounds and the used techniques provide a framework to increase the domain of applicability of the models by increasing the size and versatility of the underlying training datasets following an active learning approach. This thesis provides a proof of concept on how atomic simulations that provide detailed information on the underlying mechanism can be synergistically combined with machine learning techniques to speed up the development of novel LDHs in future research.

Zusammenfassung

Effektiver Korrosionsschutz ist bei fast allen technischen Anwendungen und insbesondere beim Einsatz von metallischen Leichtbauwerkstoffen von entscheidender Bedeutung. Dabei bleibt der vielseitige Schutz durch Anwendung von Chrom(VI)-haltige Schutzsystemen seit langer Zeit unübertroffen. Die dringende Notwendigkeit, diesen hochwirksamen, aber ebenso toxischen Ansatz zu eliminieren, ist nach wie vor eine der größten Herausforderungen für die Industrie, trotz jahrzehntelanger Bemühungen, umweltfreundliche Alternativen zu finden. Eine vielversprechende Alternative basiert auf der Anwendung von in Schichtdoppelhydroxiden (LDH) intakalierten Korrosionsinhibitoren, welche bereits ein großes Potenzial zur Kontrolle der Korrosionsraten von Aluminium- und Magnesiumwerkstoffen gezeigt haben. Die möglichen Zusammensetzungen zur Realisierung von Korrosionsschutzsystemen auf LDH-Basis sind jedoch im Grunde unendlich und können nicht allein durch experimentelle Ansätze evaluiert werden. Die vorliegende Arbeit legt den Grundstein für einen rechnergestützten Arbeitsablauf, der auf atomistischen Simulationstechniken und maschinellem Lernen basiert, um die Entwicklung und Bewertung von LDH-basierten Systemen nachhaltig zu beschleunigen. Die durchgeführten atomistischen Simulationen konnten dazu eingesetzt werden, bisher unbekannte Aspekte des Anionenaustauschprozesses aufzudecken. Die Ergebnisse zeigen, dass die ursprünglich eingelagerten Anionen während des Anionenaustauschs zuerst freigesetzt werden und das Gesamtsystem nach der Einlagerung eines anderen Anions durch die Veränderung des relativen Wassergehalts stabilisiert wird. Darüber hinaus kann die Mobilität der interkalierten Anionen ein wesentlicher Faktor sein, der ihre Freisetzung beim Anionenaustausch beeinflusst. Zusätzlich zu den atomistischen Simulationen wurden Techniken des maschinellen Lernens eingesetzt, um ein rechenzeiteffizientes Kraftfeld (ML-FF) für LDHs mit nahezu *ab initio* Genauigkeit zu erstellen. Es konnte gezeigt werden, dass das entwickelte ML-FF die akkurate Berechnung großer Systeme erlaubt. Ein weiterer Aspekt der vorliegenden Arbeit ist die Entwicklung eines quantitativen Struktur-Eigenschafts-Beziehungsmodells (QSPR) zur Vorauswahl effektiver Korrosionsinhibitoren, dass das computergestützte LDH-Design um ein weiteres Element erweitert. Die QSPR-basierten Modelle lieferten eine angemessene Vorhersagegenauigkeit. Darüber hinaus schaffen die verwendeten Techniken die Grundlage, um den Anwendungsbereich der Modelle zu erweitern, indem die Größe und Vielseitigkeit der zugrunde liegenden Trainingsdatensätze durch Ansätze aus dem aktiven Lernen erhöht wird. In dieser Arbeit wurde gezeigt, wie atomistische Simulationen, die detaillierte Informationen über den zugrunde liegenden Mechanismen liefern, synergetisch mit Techniken des maschinellen Lernens kombiniert werden können, um die Entwicklung neuartiger LDHs zu

Zusammenfassung

beschleunigen.

Acknowledgements

I would like to thank my thesis director Prof. Mikhail Zheludkevich and my daily supervisor Dr. Christian Feiler for giving me guidance, advice and support on my Ph.D. work. They are very patient, understanding, and open-minded and they are always striving to answer all of my questions, listen to all of my ideas, and share their perspectives and knowledge. None of this work would have been possible without their help.

I also deeply acknowledge Prof. Robert Meißner and Dr. Tim Würger for their generously sharing knowledge in DFT and MD simulations. I thank them for their valuable insights, guidance and fruitful discussions. I also want to show my acknowledgement to Dr. Sviatlana Lamake and Mr. Bahram Vaghefinazari for the discussions and experimental support. I have also received huge support from Prof. Pascal Friederich and his group members at KIT to expand my knowledge of machine learning. Special thanks to Dr. Patrick Reiser for teaching me how to use the KGCNN package and helping me in the development of the machine-learned force field model. I want to gratefully thank the Helmholtz Information & Data Science Academy (HIDA) for providing financial support within the HIDA Trainee Network program to make the collaboration with Prof. Pascal Friederich possible.

Many thanks to former colleagues Dr. Di Mei, Dr. Zahid Mohammad Mir, Anissa Bouali and M. Umer bilal for their support at the beginning of my work, and to Dr. Maria Rosario Silva Campos and Wendy Xu for the enjoyable discussions and the nice moments in Geesthacht. I would also like to thank Dr. Daniel Höche and all the members in the MOM department for the useful discussions and support. Many thanks to Sabine Schrader for her kindness and help with all the administrative related issues. Moreover, funding by the Helmholtz-Zentrum Hereon I2B project MUFfin is greatly acknowledged. Finally, I would not be able to complete this work without the unconditional support and encouragements of my parents and my boyfriend Jeroen.

Geesthacht, October 30, 2023

Xuejiao Li

Contents

Prüfungskommission	i
Eidesstattliche Erklärung	iii
Abstract (English/Deutsch)	v
Acknowledgements	ix
List of Figures	xiii
List of Tables	xvii
Acronyms	xix
1 Introduction	1
1.1 Light metals and corrosion protection	1
1.2 Layered double hydroxides	3
1.2.1 Characterization of LDH	4
1.3 Multi-scale simulation	5
1.3.1 Density functional theory	6
1.3.2 Molecular dynamics	9
1.3.3 Machine learning	11
1.3.3.1 Machine-learned force fields	12
1.3.3.2 Quantitative structure-property relationships	14
2 Motivation	19
3 Understanding the anion exchange process	21
3.1 Methodology	21
3.2 The impact of NaCl concentration on the anion exchange	23
3.3 Conclusion	25
4 The impact of intercalated water on LDH properties	27
4.1 Methodology	28
4.2 The impact of different hydration states on Mg ₄ Al ₂ -LDH systems	31
4.3 Changing the composition of LDH	41
	xi

Contents

4.4	Conclusion	45
5	Machine-learned force field	47
5.1	Methodology	48
5.2	Performance evaluation of the initial ML-FF model	52
5.3	Active learning with unit cell LDH	56
5.4	Active learning with supercell LDHs	61
5.5	Testing the retrained model on an expanded LDH system	66
5.6	Conclusion	66
6	Screening for benign corrosion inhibitors	69
6.1	Methodology	70
6.2	Selecting relevant features for each model	72
6.3	Performance evaluation of each model	76
6.4	Conclusion	81
7	Conclusion & Outlook	83
7.1	Conclusion	83
7.2	Outlook	84
	Contribution statement	87
A	Appendix	89
A.1	The impact of intercalated water on LDH properties	89
A.1.1	The definition of the planes for the intercalated anions	89
A.1.2	MSD at different hydration states for the intercalated anions	90
A.1.3	Angle distribution of the intercalated CO_3^{2-} at different hydration states	90
A.2	Machine-learned force field	91
A.2.1	Averaged O-H atom distance in the MOH layers	91
A.2.2	Averaged O-H atom distances in water molecules and MOH layers	91
A.3	Screening for benign corrosion inhibitors	92
A.3.1	The chemical composition of AZ91	92
A.3.2	Averaged RMSEs and number of features	93
A.3.3	Compounds containing =CHX fragment	94
A.3.4	p-value between features and target	95
A.3.5	Experimental database	96
A.3.6	Learning curves	98
	Bibliography	129

List of Figures

1.1	Schematic representation of an organic coating loaded with LDHs as smart inhibitor carriers for the corrosion protection of light metal alloys such as Mg. .	2
1.2	Schematic representation of the LDH structure and anion exchange between two types of anions.	4
1.3	Schematic representation of a molecule and a crystal that can be represented as a graph.	14
1.4	Schematic of random forest regression.	16
1.5	Schematic diagrams of an SVR with the ε -insensitive loss function and a KRR with the squared error loss function.	17
1.6	Schematic overview of the process to develop a QSPR model.	17
3.1	Immersion of an LDH in a NaCl solution.	22
3.2	The dissolution of a small amount of metal ions from the MOH layer of LDH. .	24
3.3	Interlayer distance and numbers of intercalated NO_3^- , Cl^- , and water molecules at different NaCl concentrations.	25
4.1	Schematic representation of a layered double hydroxide system with the counter anions investigated in this study.	28
4.2	DFT optimized structures for $\text{Mg}_4\text{Al}_2(\text{OH})_{12}\cdot 2\text{NO}_3\cdot 6\text{H}_2\text{O}$ and $\text{Mg}_4\text{Al}_2(\text{OH})_{12}\cdot 2\text{NO}_3\cdot 0\text{H}_2\text{O}$	29
4.3	DDEC6 charges of different types of atoms for each of the investigated LDH systems.	32
4.4	Hydration energy and interlayer distance as function of the water content. . .	34
4.5	Number density of different types of atoms along z -direction for LDH- NO_3^- , LDH- Cl^- , LDH- CO_3^{2-} and LDH- $\text{C}_2\text{O}_4^{2-}$	35
4.6	Angle distribution between anion plane and the MOH layer.	36
4.7	RDFs of water molecules around the intercalated anions and their respective cumulative numbers for the four types of LDHs at two hydration states.	38
4.8	Self-diffusion coefficient of intercalated anions at different hydration states in x - and y -direction.	40
4.9	Average H-bonds per intercalated water at different hydration states for LDH- NO_3^- , LDH- Cl^- , LDH- CO_3^{2-} and LDH- $\text{C}_2\text{O}_4^{2-}$	41

List of Figures

4.10	Schematic representation of the distribution of Al and Mg atoms in the MOH layer of $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ and $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ systems.	42
4.11	DDEC6 charges of different types of atoms for the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system at different hydration states.	43
4.12	Hydration energy and interlayer distance as function of the water content for the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system.	44
4.13	Self-diffusion coefficient of $\text{C}_2\text{O}_4^{2-}$ at different hydration states in x - and y -direction for the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ and $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ systems.	44
4.14	Illustration of the wave structure of the metal hydroxide layers for a $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system.	45
5.1	Illustration of SchNet architectural overview.	48
5.2	Schematic representation of the workflow used for constructing the ML-FF. . .	49
5.3	Edge generation in two dimensions using the k-nearest neighbors algorithm and the radius-based algorithm.	51
5.4	Illustration of the SchNet architecture used for the LDH system.	51
5.5	LDH training curve with respect to epoch variation.	53
5.6	Comparison of ML-FF model predicted values and actual values obtained from DFT calculations for the force and scaled energy.	53
5.7	Comparison of ML-FF model predicted values and actual values obtained from DFT calculations for the energy of three different sizes of LDHs.	54
5.8	ML-FF energy and atom distances varied with simulation time using the initial ML-FF model as a force field.	55
5.9	Comparison of predicted energy and force from the initial ML-FF model and actual energy and force obtained from DFT calculation for the unit cell LDH. . .	56
5.10	ML-FF energy and atom distances varied with simulation time in the iteration process for the unit cell LDH.	57
5.11	Comparison of predicted energy and force from the retrained ML-FFs and actual energy and force obtained from DFT calculations in the iteration process for the unit cell LDH.	59
5.12	The energy and force MAE values varied with the number of iterations for the unit cell LDH.	60
5.13	The C-C atom distances in oxalates varied with simulation time in different iterations for the medium size LDH.	62
5.14	Comparison of predicted energy and force from the retrained ML-FFs and actual energy and force obtained from DFT calculations in the iteration process using the medium supercell LDH.	63
5.15	O-H atom distances in water molecules varied with simulation time for the medium size LDH.	64
5.16	Atom distances varied with the number of iterations for the large size LDH. . .	65

5.17	Comparison of predicted energy and force from the retrained ML-FFs and actual energy and force obtained from DFT calculations in the iteration process using the large supercell LDH.	65
5.18	ML-FF energy and atom distances varied with simulation time for an expanded LDH.	67
6.1	Schematic representation of a layered double hydroxide system with a large number of organic inhibitor candidates.	70
6.2	Schematic representation of the ML workflow used in this study.	71
6.3	Averaged RMSE of the test sets in the cross-validations varied with the number of features in the SVR and KRR models.	73
6.4	Pearson correlation coefficients of the two models.	75
6.5	Performance evaluation of the two models.	78
6.6	Similarity matrix of the 8 blind tested compounds and the 58 compounds in the dataset for the two models.	80
7.1	Schematic representation of a feedback loop between atomistic simulations, inhibitor selection and machine-learned force field development.	85
7.2	Schematic workflow depicting the use of multi-scale simulation techniques to aid in the experimental design of layered double hydroxides.	86
A.1	The definition of the planes for the intercalated anions.	89
A.2	MSD at different hydration states for NO_3^- , Cl^- , CO_3^{2-} and $\text{C}_2\text{O}_4^{2-}$	90
A.3	Angle distribution between CO_3^{2-} plane and the MOH layer at different hydration states.	90
A.4	The O-H atom distance in the metal hydroxide layer of LDH changes over simulation time for the unit cell LDH.	91
A.5	O-H atom distances varied with simulation time in the iteration process for the medium size LDH.	91
A.6	Compounds in the used dataset containing the atom-centred fragment =CHX.	94
A.7	p-value calculation for the two models.	95
A.8	Learning curves of the two models.	98

List of Tables

1.1	Summary of properties of Al and Mg	1
4.1	DDEC6 derived partial charges of different types of atoms used in MD simulations for LDH-NO ₃ ⁻ , LDH-Cl ⁻ , LDH-CO ₃ ²⁻ and LDH-C ₂ O ₄ ²⁻ systems.	33
4.2	DDEC6 derived partial charges of different types of atoms used in MD simulations for Mg ₄ Al ₂ -LDH-C ₂ O ₄ ²⁻ and Mg ₆ Al ₂ -LDH-C ₂ O ₄ ²⁻ systems.	42
5.1	Comparison of energy and force MAEs between the retrained models with the unit cell LDH and the initial model.	60
5.2	Comparison of energy and force MAEs between retrained models using two supercell LDHs and the initial model.	66
6.1	Experimental and predicted values for the blindly tested compounds.	77
6.2	IEs of the extracted top 5 similar structures from experimental literature.	80
A.1	The chemical composition of AZ91 alloy as analysed by spark (optical) emission spectroscopy.	92
A.2	Averaged RMSEs varied with the number of features.	93
A.3	Names and IEs of the 58 chemicals in the dataset extracted from experimental literature.	96

Acronyms

Al Aluminium.

ANN Artificial Neural Network.

ASE Atomic Simulation Environment.

C-C Carbon-Carbon.

cfconv Continuous-Filter Convolutional.

DDEC6 Density-Derived Electrostatic and Chemical.

DFT Density Functional Theory.

DTNN Deep Tensor Neural Network.

FT-IR Fourier Transform Infrared Spectroscopy.

GDML Gradient-Domain Machine Learning.

GGA Generalized Gradient Approximation.

GNN Graph Neural Network.

HOMO Highest Occupied Molecular Orbital.

IE Inhibition Efficiency.

KGCNN Keras Graph Convolution Neural Networks.

KRR Kernel Ridge Regression.

LDA Local Density Approximation.

LDH Layered Double Hydroxide.

LJ Lennard-Jones.

Acronyms

LUMO Lowest Unoccupied Molecular Orbital.

MAE Mean Absolute Error.

MC Monte Carlo.

MD Molecular Dynamics.

Mg Magnesium.

ML Machine Learning.

ML-FF Machine-Learned Force Field.

MOH Metal Hydroxide.

MSD Mean Squared Displacement.

NNP Neural Network Potential.

NPT Isothermal–Isobaric Ensemble.

NVE Microscopic Ensemble.

NVT Canonical Ensemble.

O-H Oxygen-Hydrogen.

QSPR Quantitative Structure-Property Relationship.

RBF Radial Basis Function.

RDF Radial distribution function.

RF Random Forest.

RFE Recursive Feature Elimination.

RMSE Root Mean Squared Error.

SVM Support Vector Machine.

SVR Support Vector Regression.

VASP Vienna Ab Initio Simulation Package.

XRD X-Ray Powder Diffraction.

1 Introduction

1.1 Light metals and corrosion protection

Light metals such as aluminium (Al) and magnesium (Mg) with exceptional strength-to-weight ratios, as demonstrated in Table 1.1, and high processability are widely used in engineering applications such as the aerospace and automotive industries. [1–5] It is estimated that over 60% of aircraft structures are made of aluminum alloys [6], while magnesium alloys amount for approximately 2% [7]. In automotive industry, Mg and Al parts have grown rapidly, with annual growth rates of nearly 15% for Mg in the 1990s and more than 80% for Al from 2007 to 2012. [8] However, due to their low corrosion resistances and highly complex application environments, effective protection strategies are required to maintain their stability and durability during their service life in the respective applications. [3, 9]

Table 1.1: Comparison of various properties of Al and Mg: density, strength and emissions. [10–12]

Properties	Al	Mg
Density ($\text{kg} \cdot \text{m}^{-3}$)	2700	1738
Strength-to-Weight Ratio ($\text{kN} \cdot \text{m} \cdot \text{kg}^{-1}$)	130	158
Emissions ($\text{kg}[\text{CO}_2] \cdot \text{kg}^{-1}$)	22	6.9

Corrosion is commonly defined as the degradation of a material through interaction with its environment, resulting the destruction of its chemical, physical and mechanical properties and is frequently encountered in metallic materials since most metals are easily oxidized by losing their electrons to oxygen. [13] It has direct economic, energy, and security consequences that affect both individuals and society as a whole. [14, 15] For example, the global cost of corrosion in 2013 was estimated to be US\$2.5 trillion, or 3.4% of the global gross domestic product (GDP). [16] Strategies for corrosion protection can extend the service life of a metal structure, improve its service safety and reduce maintenance cost. [14, 15] The application of protective coatings is widely employed to protect light metal alloys from corrosion by isolating them from the corrosive environments. [17–20] The protection efficiency of such coatings

Chapter 1. Introduction

decreases significantly after the formation of defects caused by e.g. ageing or fatigue. [9, 19] The introduction of corrosion inhibitors into the coatings has been applied to overcome this issue and provide active and consistent long-term corrosion protection for the light metal alloys. [9, 21, 22] In the past, chromium conversion coatings as well as polymer coatings with chromate-based inhibiting pigments were widely used for corrosion protection of Al and Mg alloys due to their remarkable corrosion resistance. [23, 24] However, their use is highly restricted due to the high toxicity of hexavalent chromium to humans and the environment. [23, 25, 26] The use of Cr(VI) compounds is prohibited for all applications under European Chemical Agency (ECHA) regulations from 2024 on. [27] There is thus a strong incentive for finding non-toxic 'green' inhibitors with comparable performance, which would be essential in the field of corrosion protection. [28]

Corrosion inhibitors can be embedded directly into the surface coating at the primer stage. However, it must be assured that they can be released by external stimuli (e.g. chemical or mechanical damages) to inhibit corrosion reactions in the damaged areas or to promote a self-healing reaction to reseal the scratch. [21, 22, 29–31] This release is not yet controllable as all corrosion inhibitors are frequently released rapidly once a defect occurs [19] and, in addition, potentially negative interactions between the corrosion inhibitor and the coating are sometimes observed [32]. An improved strategy for controlled release of the corrosion inhibiting species is to load the inhibitors into a smart carrier prior to embedding it in the coating. [9, 33, 34] The smart carrier is responsive to external stimuli (e.g. pH, chemical or mechanical damages, humidity, light and temperature) and will gradually release the corrosion inhibiting species upon being subjected to one or more of these external influences. [9, 33–38] Several materials are currently utilized as smart carriers for loading corrosion inhibitors, e.g. carbon nanotubes [39], zeolites [40, 41], montmorillonite [42–44] and layered double hydroxides (LDHs) [45–47].

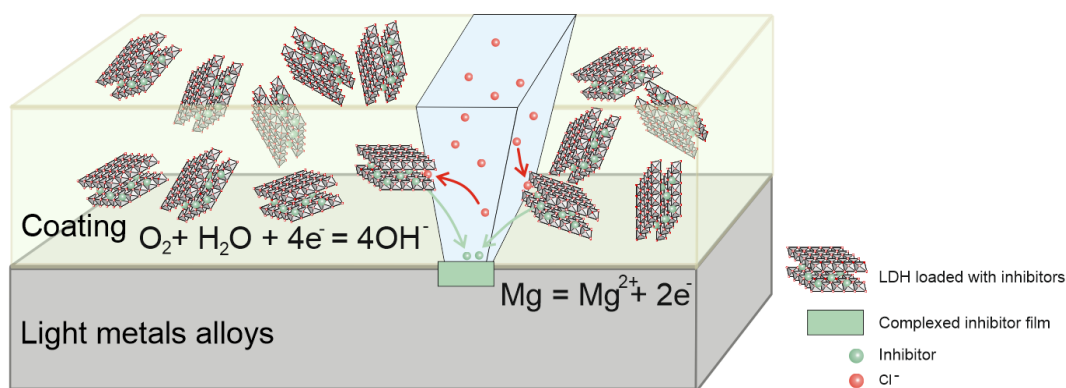


Figure 1.1: Schematic representation of an organic coating loaded with LDHs as smart inhibitor carriers for the corrosion protection of light metal alloys such as Mg.

Compared to other smart carriers, LDHs have received a significant research focus over the last

decades due to their unique anion exchange characteristics and numerous functional applications (e.g. corrosion protection). [9] Figure 1.1 depicts a schematic diagram of the inclusion of inhibitor-loaded LDHs into an organic coating to protect Mg from corrosion. A wide range of inhibitors have been investigated to be intercalated into LDHs, including inorganic anions like vanadate [19], phosphate [48], tungstate [49], molybdate [50] as well as organic anions like quinaldate [51], laurate [52], 8-hydroxyquinoline [45] and 2-mercaptobenzothiazolate [51], with the aim of improving the corrosion resistance of the resulting organic coating. There are essentially an infinite number of potential inhibitors that could be incorporated into the intergallery space of an LDH which cannot be investigated by experimental techniques alone. [30] Hence, the use of computational techniques is indispensable to pre-select a good inhibitor to be intercalated into LDH, to study the structure and stability of novel LDH and to gain a deeper understanding of LDH-based corrosion protection schemes.

1.2 Layered double hydroxides

Layered double hydroxides (LDHs) are a class of layered nano-structures consisting of positively charged metal hydroxide layers, intercalated counter anions and water molecules. [53–55] The metal hydroxide layer is composed of octahedral units connected by edge sharing, as shown in Figure 1.2, with usually one M^{2+} ion in the center of each neutral octahedral unit and six hydroxide ions at the corners. The unit will be positively charged if the M^{2+} is substituted with an M^{3+} . Intercalation of anions is required to balance the positive net charge in the metal hydroxide layer. LDH was originally discovered from nature and existed in the form of hydrotalcite $[Mg_6Al_2(OH)_{16}]CO_3^{2-} \cdot 4H_2O$. [56] Following the natural blueprint, multiple synthesis routes were developed to manufacture synthetic LDH-type compounds as well as derivatives of the parent structure (e.g. different metal ions or different intercalated anions). [57] There are three main routes which are frequently used in the synthesis of LDHs: co-precipitation [58–61], ion-exchange (as demonstrated in Figure 1.2) [59, 62, 63] and rehydration [64–66]. [57, 67–69] Although co-precipitation is the easiest and most regularly utilized LDH synthesis method, the ion exchange process is used when the desired anions are not ready for synthesis using other methods, where the desired anions replace the anions present in a previously prepared LDH. [70, 71] For the anion exchange of LDHs, three main mechanisms are proposed [72]: a single-step mechanism for direct anion exchange, where efflux of precursor anions and entry of new anions occur simultaneously [73], a two-step mechanism involving desorption of precursor anions and adsorption of incoming anions [74], and a two-step mechanism that involves dissolution of precursor LDH and re-precipitation of a new LDH [72, 75]. The general formula for LDHs is $[M_{1-x}^{2+}M_x^{3+}(OH)_2]^{x+}A_{x/n}^{n-} \cdot mH_2O$, where M^{2+} and M^{3+} denote divalent and trivalent cations, respectively, A^{n-} an intercalated anion and x a molar ratio $M^{3+}/(M^{2+} + M^{3+})$, which is usually between 0.17 and 0.33. [17, 76–80] If x remains within the specified range, it will prevent the formation of the undesired side products $M(OH)_2$ or $M(OH)_3$ during the synthesis of LDHs. [57]

LDHs are currently used in a variety of fields, including catalysis [81–84], adsorbent [85–

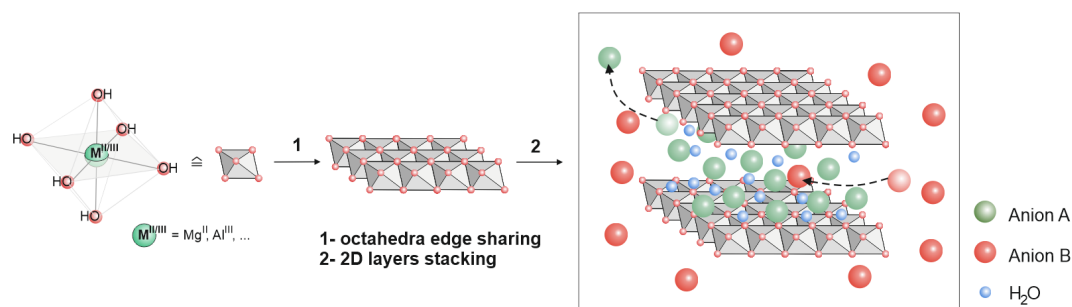


Figure 1.2: Schematic representation of the LDH structure and anion exchange between two types of anions.

87], drug delivery [88–90], origin-of-life study [91, 92] and corrosion protection [19, 48, 51] due to their ability of anion exchange, versatile compositions, chemical stability and biocompatibility. [57] The LDHs are tuned for their desired application by the type of metal ion in the metal hydroxides layers, e.g. a promising composition for the application in photocatalytic H_2 generation from water is a combination of the divalent cation Zn^{2+} with Cr^{3+} [81, 82] while the combination of Mg^{2+} or Zn^{2+} with Al^{3+} is widely applied in corrosion protection [48, 76]. The M^{2+}/M^{3+} atomic ratio can also affect the properties of an LDH [93] and its behavior in the corresponding application. [84] Liu et al. [93] observed that a higher M^{2+}/M^{3+} ratio weakened the binding strength of the intercalated anions to the metal hydroxides layers and Kim et al. [84] demonstrated that a variation of M^{2+}/M^{3+} affected the photocatalytic activity. With regard to the intercalated anions, their properties play a decisive role in the respective applications. Amino acids are widely intercalated into LDHs for the study of the origin of life [91, 92] while pharmaceuticals, such as non-steroidal anti-inflammatory [94], anti-diabetic, cardiovascular, antibiotics or antioxidants [68], have been intercalated into LDHs towards targeted drug delivery. Corrosion inhibitors are frequently intercalated into LDHs for corrosion protection [19, 48, 51]. In addition to the type of metal ions, their atomic ratio and the type of intercalated anions, the amount of intercalated water (hydration state) is known to affect the structural stability and the anion exchange behavior of LDHs [90–92, 95, 96] by rearranging the intercalated compounds, and furthermore influence their performance in the respective applications.

1.2.1 Characterization of LDH

The increasing range of LDH applications demands for a thorough understanding of the material and its underlying mechanisms. Prior studies have mainly focused on experimental characterization, employing a variety of characterization methods.

X-ray powder diffraction (XRD), scanning electron microscopy (SEM) and transmission electron microscopy (TEM) are widely applied to study the structural change of LDHs from

experiments. [19, 80, 97–99] Miyata et al. [100] are the pioneers in studying anion exchange experimentally by calculating the equilibrium constants of anion exchange reactions. More research has been conducted using microcalorimetry [101, 102] and electrochemistry [103] to assess the anion exchange reaction by calculating the enthalpy of formation and the exchange potential, respectively. The time evolution of the XRD patterns has also been used to assist in the observation of the anion exchange by Bouali et al. [104] and Palin et al. [105]. Tedim et al. [106] used ion chromatography to investigate the anion exchange between intercalated nitrate (NO_3^-) in LDH and chloride (Cl^-) in the sodium chloride (NaCl) solution. Due to the fast initial anion exchange rate and the low stability of the intermediate phase formed during the exchange process in most cases, the distinction between the proposed mechanisms and the identification of the exact mechanism remains challenging to-date. [72] Moreover, Fourier transform infrared spectroscopy (FT-IR) and elemental analysis are used to identify the successful intercalation of new anions during the synthesis of LDHs [97, 98, 107] and the amount of intercalated water can be estimated by thermogravimetric analysis. [108–110] For the LDHs in the catalytic applications, Raman spectroscopy is typically used to monitor the chemical catalytic reactions because of its ability to determine chemical properties. [111, 112] Despite the numerous experimental investigations on LDH systems [19, 48, 80, 95, 97–99, 104], obtaining a comprehensive insight into the LDH interlayer by experimental characterization is complicated by the need for high spatial and time resolutions, that are beyond the capability of current characterization techniques. The use of atomistic simulations is therefore foreseen as a strategy to overcome this limitation and thereby improve the fundamental understanding of ongoing phenomena.

1.3 Multi-scale simulation

Computational studies become more and more relevant in scientific research on e.g. materials science and engineering, since it offers advantages such as gaining insights into materials at the atomic scale and interpreting experimental observations. [113–115] Over the last few decades, computational approaches have made tremendous progress due to the exponential growth in computational power, the development of dedicated theoretical simulation methods and highly parallelized algorithms. [113, 114, 116] Multi-scale simulation strategies [117–119] have become a well-established method based on the concept of studying the behavior of a system by modeling it separately for systems at different spatial and temporal scales, achieving a combination of accuracy and efficiency with the ability to accurately characterize phenomena that is difficult to obtain with a single approach. [118, 120]

Atomistic simulations (e.g. density functional theory (DFT) and molecular dynamics (MD)) and continuum modeling techniques (e.g. finite element theory (FET)) are the two main classes of techniques used in multi-scale simulations. [117] Atomistic simulations focus on the study of systems at the atomic scale and are suitable for studying the structural, electronic, chemical and mechanical properties of materials at the nanoscale as the atomic coordinates can be accurately tracked. [121, 122] Different from atomistic simulations, continuum mod-

eling is typically used to study the behavior of materials observed at the macroscopic scale, which is considered to be a continuous medium. [123, 124] Machine learning techniques are powerful tools in identifying correlations in data [125, 126] and have been applied in materials science to accelerate the discovery/design of novel materials and predict material properties [127, 128]. It has been demonstrated that machine learning techniques and multi-scale modeling techniques, such as machine-learned force field development and atomistic simulations [129, 130], can naturally complement each other. As a result, machine learning techniques can be thought of as a separate technique class in multi-scale modeling.

With the rapid development of modeling and simulation techniques, complementary theoretical studies have been applied to carry out a comprehensive investigation [131, 132] and unravel the chemistry within the LDH intergallery. Over the last 20 years, DFT and MD have been the two preferred techniques for studying the structure and stability of LDHs. [14, 133] In LDH research, DFT and MD have already demonstrated to provide deep insights in the property of the investigated material, but become especially powerful when being combined with each other. [134–139] DFT has been used to study chemically induced structural change [93, 138, 140] and electronic properties of LDHs [141–143] thanks to its almost *ab initio* accuracy and ease of numerical implementation [144]. Although studying large LDH structures with DFT alone is computationally expensive, the results of small scale (dozens of atoms) DFT simulations have been used to provide the starting geometry as well as the atomic charges and forces of the included atoms for a subsequent MD simulation, which is more suitable for studying large systems on acceptable time scales. [76, 92, 145, 146] Machine learning is a time- and cost-effective strategy for data analysis and interpretation in materials science, however, pure machine learning usually disregards fundamental physical laws and may result in invalid or non-physical solutions to problems. [126] In this thesis, multi-scale modeling comprised of machine learning and atomistic simulations is proposed to study LDH systems where different techniques can mutually benefit from one another.

1.3.1 Density functional theory

Density functional theory (DFT) is a quantum-mechanical method widely used in chemistry, physics, biology and materials science to calculate the electronic structure of atoms, molecules and solids based on their electron density distribution. [147–149] DFT is based on the Hohenberg-Kohn theorems [150], which states that the total energy of a many-electron system in its ground state is a functional of the electron density (ρ). This function has three terms, as shown in Equation 1.1, where the external potential (E_{external}) depends explicitly on the electron density, while the dependence of the kinetic energy (E_{kinetic}) and Coulomb energy (E_{Coulomb}) on the electron density is implicit [151]:

$$E = E_{\text{external}}(\rho) + E_{\text{kinetic}}(\rho) + E_{\text{Coulomb}}(\rho) \quad (1.1)$$

Shortly afterwards, Kohn and Sham [152] came up with the idea of splitting the terms of the Hohenberg-Kohn theorems that implicitly depend on electron density into kinetic and

Coulomb energies of independent electrons, and adding an exchange and correlation term (E_{XC}) to account for the difference:

$$E = E_{\text{external}}(\rho) + E_{\text{kinetic}}(\mathbf{r}) + E_{\text{Coulomb}}(\mathbf{r}) + E_{XC}(\rho) \quad (1.2)$$

where \mathbf{r} represents the coordinates of electrons in the system. Based on the independent electron approximation [153], the kinetic and Coulomb energies in Equation 1.2 can be calculated, but the exchange and correlation energy, $E_{XC}(\rho)$, is unknown. Applying the Hohenberg-Kohn variational principle [154], minimization of the Equation 1.2 results in the Kohn-Sham equations [152]:

$$[V_{\text{external}}(\mathbf{r}) + V_{\text{kinetic}}(\mathbf{r}) + V_{\text{Coulomb}}(\mathbf{r}) + V_{XC}(\mathbf{r})] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \quad (1.3)$$

where V represents the corresponding potential of each energy term, ϕ_i is the wavefunction and

$$V_{XC}(\mathbf{r}) = \frac{\partial E_{XC}(\rho)}{\partial \rho(\mathbf{r})} \quad (1.4)$$

The wavefunctions ϕ_i , can be used to obtain the density:

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2 \quad (1.5)$$

Although the functional is unknown, there must be a functional $E_{XC}(\rho)$ that provides the exact ground-state energy and density.

After that, an enormous effort has been devoted in the construction of the E_{XC} [151] in order to obtain an exact function between the total energy and the electron density. The simplest approximate functional of the E_{XC} is the local density approximation (LDA) [155], describing the exchange and correlation energy of a homogeneous electron gas system. In the LDA, it is possible to calculate the exact exchange energy and determine the correlation energy using numerical techniques (such as stochastic numerical methods [156]). [151] For a real material, the LDA can still be used to describe the exchange and correlation energy for the regions where the electron density varies slowly. [151, 156] The material is divided into several regions and each region is associated with a homogeneous electron gas with a fixed local density. [151] In the end, the exchange and correlation energy of the entire system can be obtained by adding up the contributions from all regions. [151] Although LDA works well for solids with a relatively constant density distribution, a more sophisticated approach is required to produce accurate exchange and correlation energies for finite systems such as molecules and clusters. [151, 157] According to the study from Perdew et al. [158], the exchange and correlation energy obtained from LDA can be improved by taking into account not only the density but also its gradients, resulting in the generalized gradient approximation (GGA). GGAs have been shown to provide excellent atomic ground state energies [159], and to significantly improve the dissociation energy and bond lengths of small molecules [160], and cohesive properties of solids [161].

The Kohn-Sham equation can be expressed as a linear combination of plane waves, which

Chapter 1. Introduction

is known to be especially effective when combined with periodic boundary conditions and is widely used in condensed matter physics. [151] A plane-wave is defined as the complex exponential:

$$\exp(i\mathbf{G} \cdot \mathbf{r}) \quad (1.6)$$

where \mathbf{G} represents a reciprocal lattice vector of a crystal unit cell and \mathbf{r} represents the electron coordinates in the direct lattice of this unit cell. The single particle electronic wavefunction in a crystal structure can be expressed as the product of a function ($u_{\mathbf{n}\mathbf{k}}(\mathbf{r})$) with the same periodicity as the unit cell and a plane-wave, according to the Bloch theorem [162]:

$$\phi_{\mathbf{n}}(\mathbf{r}) \rightarrow \phi_{\mathbf{n}\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) u_{\mathbf{n}\mathbf{k}}(\mathbf{r}) \quad (1.7)$$

where

$$u_{\mathbf{n}\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{\mathbf{n},\mathbf{k}}(\mathbf{G}) \exp(i\mathbf{G} \cdot \mathbf{r}) \quad (1.8)$$

where n denotes the eigenstate index, \mathbf{k} a reciprocal vector within the first Brillouin zone and Ω the cell volume in the direct lattice. A complete extension of an infinite number of plane waves is computationally impossible, and the finite number of plane waves in the extension is determined by the kinetic energy cutoff E_{cut} . As a result, solving the Kohn-Sham equations becomes a matter of solving the plane-wave coefficients, which can be done with efficient numerical schemes. [163] After obtaining the solutions of the Kohn-Sham equations of an atomic configuration, it is possible to calculate the total energy and also material properties such as the equilibrium structure, band structures of metals and semiconductors and the partial atomic charges. [164]

DFT has been applied in the context of LDH-like structures to study the chemically induced structural change [93, 138, 140] and the electronic properties (e.g. band structure, density of states and band edge placement) [141–143, 165, 166]. In the early stages, DFT was applied to assist the structural investigation for layered hydroxides materials with poor crystallinity. [14, 167] Afterwards, it was used to study the orientation of the intercalated anions with respect to the metal hydroxide layers and to explore the interface between LDH particles and a metal substrate. [14, 168, 169] DFT was applied to study the exfoliation mechanism of LDHs dispersed in formamide to gain a better understanding on how formamide affects layer separation. [166, 170] Moreover, Tavares et al. [141, 166] noted that hydration can induce changes in the electronic structure of an LDH and the formation of defects on an LDH can influence its band gap according to their DFT study. It was observed that dehydration caused a significant structure modification in the exfoliated Zn_2Al -LDH intercalated with CO_3^{2-} , causing CO_3^{2-} to become HCO_3^- . [141] Xu et al. [165] investigated the oxygen evolution reaction in a variety of LDH systems ($\text{M}^{2+} = \text{Mg}, \text{Co}, \text{Ni}$ and Zn ; $\text{M}^{3+} = \text{Al}$ and Ga) using DFT calculations. They studied the electronic properties of these LDHs in details and selected the LDH with the most efficient oxygen evolution photocatalyst. In addition to studying the structural and electronic properties of LDHs, Moraes et al. [142] and Costa et al. [169] elucidated the thermodynamic potentials of anion exchange supported by DFT calculations. Based on their

Gibbs free energy calculations for LDHs intercalated with different anions, Costa et al. [169] obtained an anion exchange selectivity sequence of LDH: $\text{NO}_3^- < \text{Br}^- < \text{Cl}^- < \text{F}^- < \text{OH}^- < \text{CO}_3^{2-}$. It has been demonstrated that it is more energetically favorable for the intercalated CO_3^{2-} to be parallel to the metal hydroxide layers whereas the intercalated NO_3^- tends to be tilted and its orientation depends on the water content in the interlayer. [169] Large-scale studies of LDH with DFT are computationally very expensive, as the corresponding simulations for systems with several thousands atoms require a long computational time to study a multitude of LDH properties, while avoiding finite-size effects. [92, 171]

1.3.2 Molecular dynamics

Compared to DFT calculations, molecular dynamics (MD) simulations are more suitable for studying large systems on acceptable timescales because they treat atoms as classical particles and use various empirical potentials in their calculations. [172] MD has been widely applied to gain a better understanding of LDH structures [92, 173–178], interactions between intercalated compounds and metal hydroxide layers [92, 173–175, 177, 179, 180], and intercalated compound dynamics [90, 177, 179, 181]. MD simulation refers to a deterministic simulation which is used to study the temporal evolution of the interaction and motion of all atoms in a multi-body system, where the trajectory evolution of the system follows Newton's second law [182, 183]:

$$F = m \cdot a = m \cdot \frac{d^2 x}{dt^2} \quad (1.9)$$

with

$$F = - \frac{\partial U}{\partial x} \quad (1.10)$$

where F represents the conservative forces between atoms in the system, U the potential energy of the system and x the position of the atoms. An accurate force field is a central part of MD simulations as it describes the interactions between atoms in a system. [184] A general mathematical expression for a force field is the sum of intramolecular (e.g. bonds (E_{bonds}), angles (E_{angles}), dihedral ($E_{\text{dihedrals}}$) and improper ($E_{\text{impropers}}$) torsions) and intermolecular (Lennard-Jones (E_{LJ}) and electrostatic (E_{elec}) interactions) contributions [184, 185]:

$$U = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}} + E_{\text{impropers}} + E_{\text{LJ}} + E_{\text{elec}} \quad (1.11)$$

where U represents total potential energy. For most parameters in the intramolecular contributions, they are extracted or refined using experimental data (e.g. X-ray diffraction, infrared or Raman or vibrational spectra). [184] Parameters of the Lennard-Jones potential are calculated on the basis of the thermodynamic and crystallographic characteristics. [186] The partial charges used in the electrostatic interactions are usually obtained from a high accuracy quantum-mechanical calculations. [184] The density-derived electrostatic and chemical (DDEC6) approach is one of the most accurate and frequently applied methods to obtain partial charges and the calculation is based on charge densities obtained from quantum-

Chapter 1. Introduction

mechanical calculations. [187] There are a number of popular general force fields which are widely applicable, e.g. GROMOS [188] is often applied to the simulations of biomolecules and CLAYFF is applicable to mineral systems [185]. In addition to the general force fields, there are a number of specific potentials developed for particular systems or compounds, e.g. water potential (SPE/C) [189]. During the simulation, the forces on each atom are calculated and then used to update the position and velocity of each atom at each time step. [190] Periodic boundary conditions are often used to simulate a crystal material to remove the surface effect. [191] Ensembles are the conditions for performing MD simulations and include the microscopic ensemble (NVE), the canonical ensemble (NVT), and the isothermal-isobaric ensemble (NPT), where N represents the number of atoms in the system, V the volume of the simulation box, E the total energy of the system, T the temperature of the system, and P the pressure of the system. [184] For example, in the NVT ensemble, the number of atoms in the system, the volume of the simulation box and the temperature remain constant throughout the simulation. Because most experimental work was done at constant temperature or pressure, the NVT and NPT ensembles are useful for comparing simulation results to experiment results. [184] In MD simulations, temperature can be controlled by modifying the velocities of the atoms in the system or by using a thermostat, and pressure can be controlled by adjusting the size of the unit cell and the atomic coordinates in the system. [184, 192, 193]

After the release of the CLAYFF force field [185], classical MD has been the main computational method for LDHs and other clay-like materials. [14, 181] To gain an insight into the structure of LDH interlayer content, Gregoire et al. [92], Erastova et al. [175] and Chen et al. [178] analysed the distribution of intercalated water and the orientation of intercalated anions by performing MD simulations. Gregoire et al. [92] discovered that the orientation of the intercalated anion was dependent on the water content in the interlayer region, especially at high hydration levels and Chen et al. [178] observed that the variation of the water content affected the configuration of the intercalated NO_3^- from being flat to tilted with the increase of water content while CO_3^{2-} always remains flat. For LDHs intercalated with CO_3^{2-} and Cl^- , a transition from a single monolayer consisting of intercalated anions and water molecules to a coexistence of monolayer and bilayer was observed at high hydration level. [178] The stability and dynamic properties of LDHs can be obtained by calculating the hydrogen bonding network of LDHs and the mobility of the compound in the intercalated region. [177, 179] Hydrogen bonding networks of LDHs have been extensively studied in MD simulations to display the interactions between intercalated anions, water and metal hydroxide layer. [173–175, 177, 178, 194] The mobility of intercalated species can be obtained from MD simulations by calculating the self-diffusion coefficients. [90, 179] Moreover, the relationship between hydration energy and water content was investigated to explore the influence of water content on the structure of montmorillonite [195, 196] and LDH [173, 194, 197, 198], and to search for the energetically well-defined hydration state for each structure. Murthy et al. [179] calculated the hydration energy and used it as an indication of the LDH stability, where LDH has been observed to be stabilized by forming hydrogen bonding networks, resulting in a low mobility of the intercalated anion. With the assistance of MD simulations, it is possible to model LDHs under more realistic

experimental conditions, e.g. Perez-Sanchez et al. [76] and Tavares et al. [166] investigated the stability of LDH particles immersed in a sodium chloride water solution and the anion exchange occurred on the basal surface of exfoliated LDHs, respectively. According to Perez-Sanchez et al. [76], the metal hydroxide structure of LDH was maintained in the NaCl solution despite the dissolution of a small amount of metal atoms and hydroxyl groups.

Unlike MD, Monte Carlo (MC) simulations are a stochastic method and cannot provide time evolution information of a system. [182, 183] Rather, it is applied to generate an ensemble of representative configurations under a specific thermodynamic condition with the application of random perturbations. [182, 183] It has been shown that both methods can provide consistent results considering the thermodynamic and conformational behavior of the same system. [199, 200] MC has been successfully applied to study the adsorption of surfactants on hydrophilic surfaces [201, 202] and in slit nanopores [203]. In addition to surfactants, the adsorption of gases or molecular fluids in porous structures e.g. metal-organic frameworks [204–207], covalent organic frameworks [136] and shale organic carbon slit micropores [208] has been investigated using MC simulations. LDHs can be considered as a porous structure as well due to its large porosity [57] and Amamra et al. [209] demonstrated that LDHs can be used to study the adsorption of sodium salt of drug molecules with the application of MC simulations. However, because traditional MC simulations do not follow Newton's equations, one of the major limitations of using them in LDH research (e.g. the anion exchange process) is the inability to collect dynamic information. [210]

1.3.3 Machine learning

Machine learning (ML), which is based on data-driven models that can learn from existing data [211], has grown rapidly in recent years as a result of algorithmic advancements and technological advances in computing hardware. [212] While influencing our daily life, such as the recently developed ChatGPT [213, 214], ML techniques have also gained an important role in materials science and chemistry [127, 128]. Four main categories of ML techniques exist: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. [215, 216] The available data needs to be labeled for supervised learning, and the goal is to learn a function between the input features (e.g. molecular structure, electronic properties) and the labeled output (e.g. corrosion inhibition efficiency). [217] Unsupervised learning analyzes unlabeled datasets in order to discover hidden patterns in the data. [216] Semi-supervised learning is an intermediate between supervised learning and unsupervised learning that processes both labeled and unlabeled data. [218] It is especially useful in situations with little labeled data and a lot of unlabeled data. [216] The reinforcement learning method aims to take actions that maximize the reward or minimize the risk based on information gathered from interactions with the environment. [216] Each category has a different application based on its learning capability and the nature of the available data, e.g. supervised learning is a typically task-driven approach where models learn from labeled data to generate an input-output function. [215–217] Among all the ML categories, supervised learn-

ing is by far the most widespread category in materials science. [212] In materials science, different supervised learning algorithms including traditional algorithms (e.g. classification analysis, regression analysis) and deep learning methods [215] have been applied to material discovery such as compound prediction [219–221], structure prediction [222, 223] and the prediction of material properties such as the band gap [224], superconductivity [225] and bulk and shear moduli [226]. In addition to the material property predictions mentioned previously, ML algorithms have also been established to construct machine-learned force fields for MD simulations [129, 130], adaptive design process, and active learning, which can guide models to explore interesting material spaces more efficiently compared to random selection. [227]

1.3.3.1 Machine-learned force fields

In recent years, machine-learned force fields (ML-FFs) have been a breakthrough success, promising to achieve the accuracy of quantum mechanical calculations at a significantly reduced computational cost, which previously typically required a trade-off in materials science. [129] A reliable force field is undoubtedly a key component towards accurate MD and MC simulations. [228] The state-of-the-art force field mostly used for LDHs is the CLAYFF force field [185], which is based on simple functional forms with empirical bond and non-bond parameters. The empirical parameters are derived from quantum mechanical calculations (e.g. partial charges) and fits to experimentally determined structures and physical properties. [185] A classical force field-based calculation is typically more computationally efficient than purely quantum mechanical calculations, but determining the best functional form and the corresponding parameters of the potential energy surface can be challenging due to the inability to fully separate the intricate electronic effects into atom-atom contributions [229]. [184, 228, 230] Furthermore, the classical force field can be limited in terms of producing accurate simulation results since significant physical approximations have been employed to describe the molecule interactions in a tractable manner [229]. These limitations are particularly present if the initial force field was not developed specifically for the system, or even impede the modeling of chemical processes. [184, 228, 231] Reactive force fields have been introduced to simulate chemical processes (e.g. heterogeneous catalysis reactions [232]) based on an approximated reaction path but their functional forms are more complex, making parameterization even more difficult, particularly for non-experts. [231, 233, 234]

The ML-FFs have shown great potential to mitigate the limitations of classical MD simulations (e.g. chemical interactions) and to reduce the need for computationally expensive DFT simulations to generate the required force fields. [129] The central idea of force fields derived from machine learning is to build an ML-FF based on a number of completed calculations with sufficiently high accuracy that are currently mainly based on DFT. [235] This type of force field has been developed rapidly over the past two decades [236] with the substantial increase in computing power [237, 238], ML algorithms [211, 239], architectures [240, 241] and the amount of reliable training data [242, 243]. [129] ML-FFs have already been applied to study chemical reactions [244, 245] and adsorbent–surface interactions on solid surfaces [246–249],

the search for new crystal structures [250–252], and to investigate the behavior of complex systems (e.g. metal–organic frameworks) [253].

Artificial neural networks (ANNs) are one of the earliest and most frequently used ML frameworks for building ML-FFs, capable of running large time- and length-scales simulations with almost *ab initio* accuracy and affordable computational cost. [129, 130] An ANN is a data processing system composed of highly interconnected elements inspired by the human brain and vaguely mimicking the way biological neurons communicate with one another. [254, 255] ANNs are typical parametric algorithms for solving nonlinear functions of a fixed number of variables with a high degree of flexibility and are able to provide accurate results for complex and non-linear problems. [256, 257] There are several successful schemes reproducing a force field based on ANNs proposed in the last few years, including neural network potential (NNP) [242, 258], deep tensor neural network (DTNN) [243] and deep potential molecular dynamics (DeePMD) [259, 260]. Smith et al. [242, 258] generated a transferable deep NNP (called ANI-1) for organic molecules based on atomic forces obtained from DFT calculations, and Bag et al. [248] developed an ML-FF by training ANNs to study the adsorption energy of an amino acid on the surface of graphene and gold. Gradient-domain machine learning (GDML) [261] is a kernel-based model trained on the energy of the material structure and the forces acting on the atoms in each structure, where the forces are obtained by applying a gradient operator on the energy. [262, 263] Saucedo et al. [263] noted that although GDML is highly data-efficient, robust, and follows fundamental physical laws, it is only applicable to small molecules due to the drastic increase in the number of structural descriptor calculations (such as dihedral angles) [231], which limits its application. In contrast, DTNN is known to be size extensive and efficient, with a highly uniform accuracy throughout the material structure. [243] DTNN is a graph neural network (GNN) architecture where the material structure can be considered as a graph in which all atoms are represented by nodes and the interactions between atoms are represented by edges, thus being able to maintain the rotational and translational invariance. [243] Complex chemical interactions can be modeled by exchanging information along the edges that connect the atoms in each graph. [231, 243] Sharing the same invariances as DTNN, Schnet [264], another well-known GNN architecture, has demonstrated promising potential in learning from graph-structured data to build a corresponding ML-FF with higher accuracies as compared to DTNN. [265]

Among various machine learning methods, GNNs are one of the fastest growing methods due to their high application rates in the past few years. [266, 267] GNNs can be interpreted as generalizations of convolutional neural networks, which are one of the most common deep learning methods. [215, 266] GNNs operate on graph-structured data, which exists in numerous applications and is associated with nodes and edges attributes. [266, 268] When compared to compositional or fixed-sized vector representations, using the chemical graph as input directly provides greater flexibility and scalability in the structure representation. [266] The nodes and edges attributes can be linked to the atoms and bonds in material structures as seen in Figure 1.3, where GNNs can operate on the material representations to build a model to predict interesting material properties. [266] The application of GNN is able to automatically

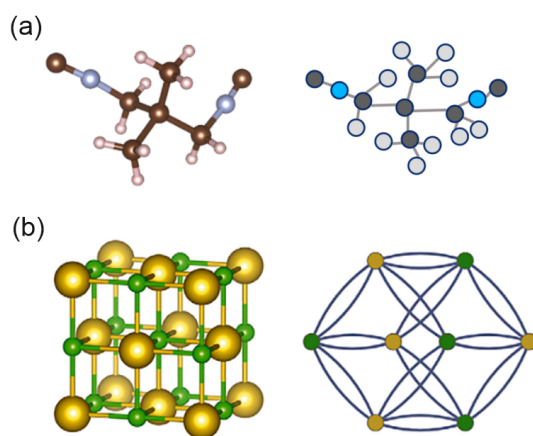


Figure 1.3: Schematic representation of **(a)** a molecule and **(b)** a crystal that can be represented as a graph. The atomic structures are shown on the left and the graphs are depicted on the right side. (Source: [273])

extract structural information from a given material system, bypassing the calculation of atomic fingerprints that are typically used to represent the local environments of the material system. [269] GNNs have been successfully applied to a variety of challenges, e.g. predicting the properties (such as toxicity, reactivity) and the synthetic routes of molecules and materials, and accelerating atomic simulations. [266, 269–271] GNN-based ML-FFs have shown excellent accuracy in predicting atomic forces for various material systems including organic molecules and multi-element solid-state systems. [269] Wang et al. [228] built an ML-FF for large flexible organic molecules by applying GNN, where the GNN model was only applied to replace functional forms of bonding interactions. Park et al. [269] applied the GNN-based force field to achieve a higher accuracy in force prediction in an interfacial system of amorphous $\text{Li}_4\text{P}_2\text{O}_7$ and Al_2O_3 with HF gas compared to another NN architecture. When developing an ML-FF for a crystal structure, periodic boundary conditions need to be taken into account, e.g. the nodes and edges in the periodic images within a given cut-off distance should be included in the calculation. [265] The application of ML-FFs can provide a higher degree of flexibility and accuracy in simulations than the previously mentioned MD simulations using classical force fields. [272]

1.3.3.2 Quantitative structure-property relationships

The ability to predict interesting properties of novel materials prior to synthesis and to find connections between microscopic and macroscopic materials properties is of great benefit to researchers in the fields of chemistry, biology, physics and engineering. [274] Quantitative structure-property relationship (QSPR) methods are based on a hypothesis that there is a relationship between the structure of a molecule and the macroscopic properties of that structure, where changes in the structure can be reflected in the observed properties [274]. The QSPR model has been widely applied to accelerate the discovery of advanced materials [274, 275],

e.g. it has been employed in catalyst design to evaluate the relationship between molecular structure and catalyst activity [276] and to investigate the glass transition temperature of polymers. [277] The relationships in QSPR models can be complex, necessitating the use of machine learning techniques to generate robust and predictive links between molecular structure and macroscopic properties. [274]

For a QSPR model based on machine learning techniques, reliable and versatile training data is an essential component as the accuracy and robustness of the model is highly dependent on the size and quality of the available dataset. [212] Once the dataset has been chosen, the representation of the data is crucial for the model and the data samples, e.g. molecules can be encoded by different approaches. [212,278] The molecular descriptors can be broadly classified into constitutional descriptors (e.g. chemical composition), structural descriptors (e.g. the three-dimensional coordinates of the molecule), topological descriptors (e.g. connectivity of atoms in the molecule), quantum-chemical descriptors which are normally obtained from DFT calculations (e.g. orbital energies) and physicochemical descriptors reflecting the physical and chemical properties of the molecule. [274] Feature selection is an important topic in ML because it aims to select relevant features while removing irrelevant, redundant, and noisy features to achieve the best model performance. [279] Selecting relevant features is an important step in QSPR model generation when the number of available descriptors is higher than the size of the dataset. [280,281] There are three major approaches for feature selection: filter, wrapper, and embedded approaches. [279] The filter approach is frequently used as a pre-processing step, with features chosen based on their performance in various statistical tests (e.g. Pearson correlation [282]) to determine their relevance to the labeled target. [283] Instead of evaluating features individually, the wrapper approach evaluates by building a model for each generated subset of features, taking their interactions into account (e.g. recursive feature elimination (RFE) [284]). [285] RFE searches for a subset of features by beginning with all available features in the training dataset and successfully removing features until the desired number of features remain. [286] In the case of embedded methods, feature selection is conducted and optimized during the learning process (e.g. random forest (RF) [287]). [288] RFs have proven to be a useful algorithm for dealing with feature selection problems due to their ability to calculate the importance of each feature. [289] For each tree in an RF, as shown in Figure 1.4, the feature importance can be calculated based on its ability to decrease the impurity, which is the variance of the RF regression. [290] The presence of correlated features, on the other hand, has been shown to affect its ability to identify important features, potentially lowering their accuracy. [291,292] To address this issue, a combination of RFs and RFE is commonly used [291,292] and its potential to select relevant features to model corrosion inhibition efficiencies of small organic molecules has been demonstrated in a recent study [293].

A variety of ML algorithms have been used to build QSPR models to screen a large number of compounds with potentially interesting properties based on available data in chemical space, accelerating material selection in comparison to traditional investigations based on experimental observations. [28,293,295–297] For regression problems where the property is a

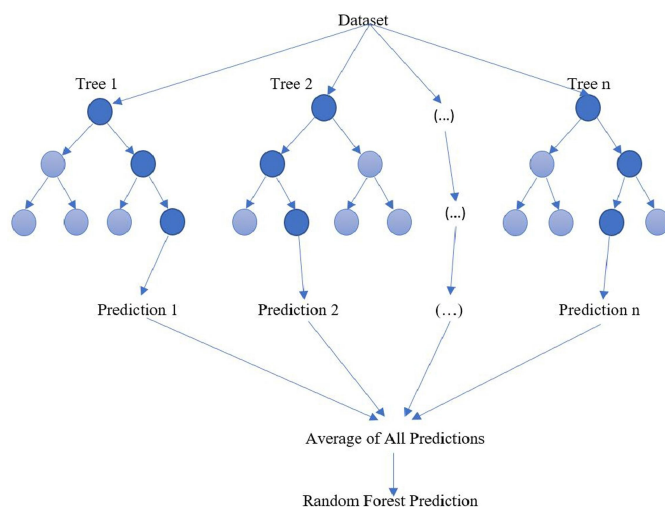


Figure 1.4: Schematic of random forest regression (Source: [294]).

continuous variable, the ML algorithms can be multiple linear regression in the simple case, or a polynomial, kernel-based, or neural network-based regression when the relationship is not linear. [274] Kernel-based learning algorithms have gained popularity over neural network-based methods in some applications, due to their ability in dealing with nonlinear data by introducing linearity. [298] Furthermore, kernel-based learning algorithms are well suited for regression tasks with limited datasets because they rely on feature engineering rather than inferring relevant representations directly from the data, which typically requires more data. [299–301] Among the variety of ML algorithms, support vector machines (SVMs) are one of the most powerful, precise and robust supervised learning methods because of their good theoretical foundations and high generalization capacity. [302, 303] A variety of complex real-world problems have been solved through the application of SVMs, e.g. image classification [304], hand-written character recognition [305] and face detection [306] in the past twenty years. [302] SVMs were originally developed for solving classification problems and by applying the same principle as SVMs, support vector regression (SVR) was developed to solve regression problems with high accuracy. [307] Likewise, kernel ridge regression (KRR) [308] has been recognized by other researchers as being able to obtain accurate results comparable to those of SVR. [309, 310] The difference between SVR and KRR is their loss function, whereas KRR applies a squared error loss, SVR employs an ϵ -insensitive loss as illustrated in Figure 1.5a and b. Both SVR and KRR belong to the kernel-based learning algorithms. [311, 312] The function of a kernel is to implicitly transform the nonlinear input features from the original space to a high-dimensional feature space, where the regression can be accomplished linearly. [298] The most widely used kernel in SVM is the radial basis function (RBF) kernel [313] and Smola et al. [314] pointed out that the RBF kernel is generally a reasonable choice for datasets with little information on their shape.

Following the selection of an ML algorithm, QSPR model generation can proceed in the following steps: selection of a dataset and extraction of structural descriptors, feature selection,

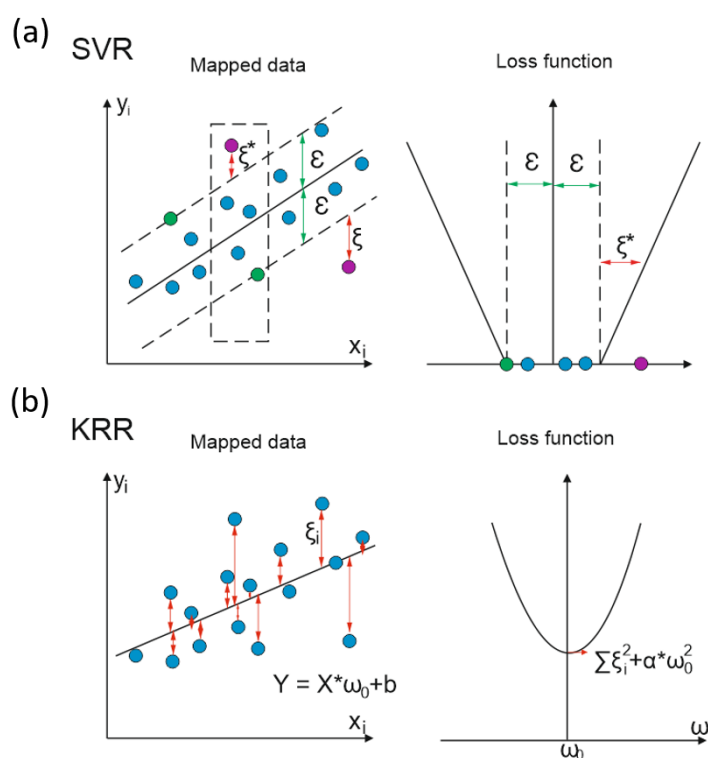


Figure 1.5: Schematic diagrams of (a) an SVR with the ϵ -insensitive loss function and (b) a KRR with the squared error loss function.

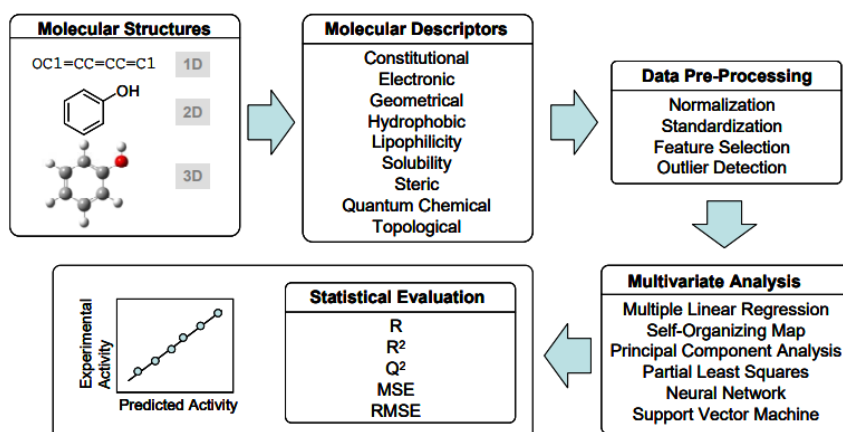


Figure 1.6: Schematic overview of the process to develop a QSPR model (Source: [315]).

Chapter 1. Introduction

model construction, and model evaluation, as shown in Figure 1.6. [280] Due to a lack of data on LDH, QSPR has not yet been applied directly to build a predictive model for an LDH system. On the other hand, QSPR has been widely used to investigate corrosion-related fields such as the influence of the outdoor environment on the corrosion rates [316,317] and the identification of effective corrosion inhibitors for a variety of metal alloys [293,318–320]. Zhi et al. [317] developed a QSPR model based on the RF method to investigate the effect of various environmental factors on the corrosion rate of low alloy steels. In their work, input features include low alloy steel type, environmental factors, and exposure time, which are all their parameters of interest, so there is no feature selection step in their model construction. For the effective corrosion inhibitor identification, a variety of machine learning algorithms (e.g. neural networks, kernel ridge regression, and random forests) have been applied to predict the corrosion inhibiting effect of small organic compounds for different types of Mg alloys [293,295], Aluminum alloys [318,319,321] and Copper-based materials [320]. Würger et al. [295] used an experimental dataset of 78 modulators to build a QSPR model based on the KRR algorithm to predict the corrosion efficiency of small organic compounds for commercially pure magnesium (CPMg220). The robustness of the KRR model was assessed using blind tests on ten new compounds, and the predicted values for these new compounds agree well with the experimental results with a root mean squared error (RMSE) of 36%.

Molecular descriptors generated by cheminformatics tools (such as alvaDes [322] or rdkit [323]) can number in the thousands for each organic compound, necessitating feature selection to select relevant features from a large pool of available features, especially when the training dataset is small. Schiessler et al. [293], Liu et al. [300] and Alamri et al. [324] used a random forest based method to select sparse features. In addition, quantum chemical descriptors (e.g. HOMO, LUMO) have been widely used to build the required QSPR models. [293, 325, 326] Breedon et al. [327] and Winkler et al. [318], on the other hand, noted that HOMO and LUMO did not correlate strongly with the corrosion inhibitions of organic compounds. As a result, before using them as input features, a thorough investigation should be conducted to verify their importance in developing a QSPR for the corrosion prediction of small organic compounds.

2 Motivation

Although LDHs have been extensively characterized by experimental investigations [19, 48, 95], it remains highly challenging to obtain a comprehensive insight into the LDH interlayer from experiments alone. Following Section 1.3, computational techniques have become indispensable in materials design in recent years because they generate significant synergies with experimental approaches to gain a deeper mechanistic understanding of LDH-based corrosion protection schemes. This motivated the development of a framework based on multi-scale simulation techniques to aid in the design of novel multifunctional surfaces based on LDHs to protect lightweight metal alloys from corrosion that is presented in this thesis work.

On the atomistic level, computer simulations were used to investigate anion exchange processes and hydration effects based on DFT and MD simulations to elucidate more details on the LDH intercalation behavior. To date, the most widely used force field for LDH in MD simulations has been the CLAYFF force field [185] which describes the interatomic potentials related to the metal hydroxide layers of an LDH. For LDHs intercalated with novel inhibitors, the force fields of the inhibitors are normally not directly available, requiring additional work to develop them, while achieving force fields with high accuracies is generally considered non-trivial. In order to improve simulation accuracy and accelerate atomistic simulations, a machine-learned force field for LDH systems was developed in this work. Another challenge in realizing effective LDH-based active corrosion protection is the selection of potent corrosion inhibitors for the LDH design. However, there are numerous organic compounds that could potentially be used to inhibit corrosion reaction of Mg and its alloys [30], making the identification of the most potent compounds more difficult. In this work, different supervised machine learning models were trained to identify effective corrosion inhibitors in the vast chemical space of available molecules.

Finally, a workflow consisting of atomistic simulation, machine-learned force field development and inhibitor selection has been developed in frame of this work to assist the design of novel LDH used in corrosion protection.

3 Understanding the anion exchange process

The exchange of intercalated inhibitors with a corrosion-promoting species such as chloride is a fundamental step toward using layered double hydroxides (LDHs) for corrosion protection strategies. [19, 48, 51] Although the exchange mechanisms are still highly debated, particular stimuli (e.g. local change in pH, presence of aggressive species) have been observed to be capable of triggering the exchange. [24] Following the experimental investigations of the anion exchange described in Section 1.2.1, mathematical [328] and kinetic models [329] have been applied to study the anion exchange of LDHs by fitting experimental data to the corresponding model. [330–333] The exchange order of various anions was investigated using atomistic (DFT and MD) simulations to calculate the binding energy of the anion to the metal hydroxide layer of LDHs or the free energy of the anion exchange reaction. [93, 137, 180] The details of the release and rearrangement of intercalated anions during anion exchange, on the other hand, remain unknown. As a result, the anion exchange of LDH particles in a NaCl solution, as well as the effect of NaCl concentration on anion exchange, were investigated in this study.

3.1 Methodology

The purpose of this study is to learn more about the anion exchange mechanism of LDHs and how sensitive it is to chloride concentrations using the MD simulation model proposed by Perez-Sanchez et al. [76], with the LDH particles immersed in a NaCl water solution. The structure of Al-Mg-LDH-NO₃⁻ was extracted from the work of Perez-Sanchez et al. [76], and the LDH was immersed in NaCl solutions with different concentrations (0 M, 0.16 M, 0.64 M, 1.6 M and 3.2 M). The immersion process was carried out in Packmol [334] by placing the LDH in the center of a larger simulation box with a 1 nm shell around it, and then randomly filling the empty space with water and some Na⁺ and Cl⁻ according to the set concentrations. This is illustrated in Figure 3.1. Following the simulation procedure of Perez-Sanchez et al. [76], energy minimization and equilibration steps were carried out while fixing the atomic positions of the immersed LDH. Afterward, the restraints on the immersed LDH were removed and an isothermal-isobaric ensemble (NPT) production was run for 20 ns to study the anion exchange process. Details on the equilibration and production steps can be found in the work

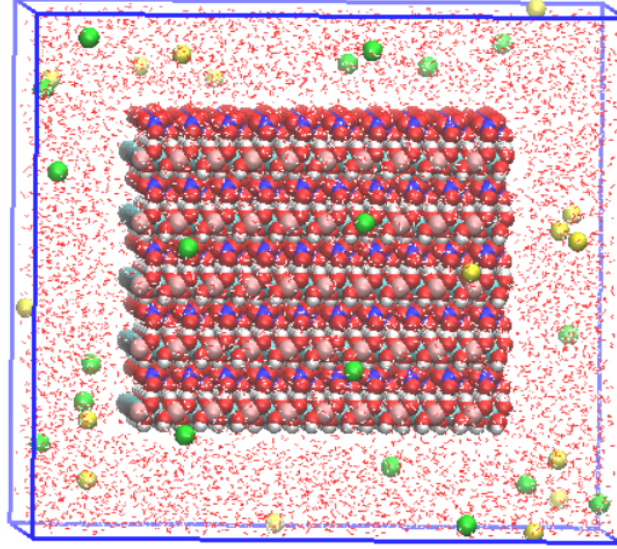


Figure 3.1: An LDH immersed in a NaCl solution is depicted schematically. Pink, cyan, red, blue, green and yellow spheres represent aluminium, magnesium, oxygen, nitrogen, chloride and sodium atoms in the metal hydroxide (MOH) layer and ions in the system, respectively. Water molecules are presented as red dots for improved visualization.

of Perez-Sanchez et al. [76]

Since Perez-Sanchez et al. [76] observed metal ion dissolution in their work, the distance between each metal ion and its nearest metal neighbor in each layer was calculated and used in this study to determine whether the metal ion was dissolved or not. Because the system is already in equilibrium during the majority of the simulation time in the NPT production, this calculation was only performed in the final 1 ns of the MD simulation and was representative to describe the entire time-consuming period at equilibrium, resulting in a significant reduction in data analysis effort. Metal ions with a nearest neighboring distance larger than or equal to 1 nm were assumed to be dissolved (for comparison, the extend of the LDH structure is 5.3 nm in X direction). For the metal ions whose nearest neighboring distance were less than 1 nm, the average and standard deviation of their nearest neighboring distances were calculated. The threshold distance (referred to as d_{limit}) is equal to the average distance plus the standard deviation in determining whether or not the metal ions were dissolved. If the nearest neighboring distance of the metal atoms was less than d_{limit} , the metal atoms were assigned to the LDH structure (referred to as LDH_{Metals}), otherwise they were considered as dissolved.

As the performed MD simulation describes a dynamic process and 1 ns simulation contains 201 frames, each frame has to be analyzed to distinguish between atoms belonging to LDH_{Metals} and those that are dissolved. The probability of each metal ion belonging to the class of LDH_{Metals} was calculated by adding its attribution from 201 frames and then dividing

3.2 The impact of NaCl concentration on the anion exchange

by 201. Metal ions with a probability higher than 95% were considered to be in the LDH system throughout the simulation, while those with a probability below this threshold were considered to be in solution. Following that, the average coordinate position in Z direction of the LDH_{Metals} ions in each layer was calculated separately to determine the interlayer distance of the LDH. Finally, the interlayer distance was calculated by dividing the distance between the bottom and top LDH layers by four. The LDH-interlayer distance (referred to as $d_{interlayer}$) is an important quantity for determining whether there is sufficient space for anion intercalation.

3.2 The impact of NaCl concentration on the anion exchange

Metal ion dissolution occurred during the MD simulation (Figure 3.2a), which is in agreement with the experimental observations from Buchheit et al. [21], Poznyak et al. [51], and Galvão et al. [335]. The concentration of NaCl affects the amount of dissolved metal ions, as shown in Figure 3.2b, whereas a higher concentration of NaCl stabilizes the LDH sheets. This can be attributed to the presence of Na^+ ions in the solution, which impedes the dissolution of metal ions from the LDH due to the repulsion between two positively charged ions.

The influence of NaCl concentration was insignificant for the interlayer distance of LDH (Figure 3.3a) because all determined values are close to the value of the system where the NaCl concentration is zero. When there is no Cl^- existed in the simulation system (the concentration of NaCl is zero), there is no anion exchange between NO_3^- and Cl^- . It was observed that NO_3^- released from LDHs without Cl^- existing in the solution. The release could be triggered by the chemical potential difference of NO_3^- in LDH and in the solution. The metal ion dissolution and the NO_3^- release maintained the neutrality of the remaining LDH.

The shape of each LDH layer became irregular after the metal ions were partially dissolved, compared to its initial parallelogram. After combining five irregular LDH layers as a whole, the shape of the LDHs became more irregular, making it difficult to define the intercalation of anions (such as NO_3^- and Cl^-) and water into the LDHs, especially for the anions and water molecules that were close to the metal ions located at the edges of the LDH layers. To provide a clear definition for intercalation, the LDH_{Metals} ions at each LDH layer were divided into two parts: edge atoms and inside atoms. Anions and water molecules that are within 0.45 nm of a group of inside atoms are defined as intercalated and the selected distance is close to half of the $d_{interlayer}$ for all LDH systems.

For the simulation systems with different NaCl concentrations, the anion exchange process was investigated by calculating the remaining NO_3^- and the newly intercalated Cl^- . The calculation results show that the amount of released NO_3^- did not have a clear correlation with the NaCl concentration (Figure 3.3b), which agrees well with the experimental observation of Tedim et al. [106], where an increase in NaCl concentration from 0.05 M to 0.5 M did not result in an increase in nitrate release. The amount of intercalated Cl^- is shown in Figure 3.3c, and the trend clearly shows an increase in intercalated Cl^- with increasing NaCl concentration in the

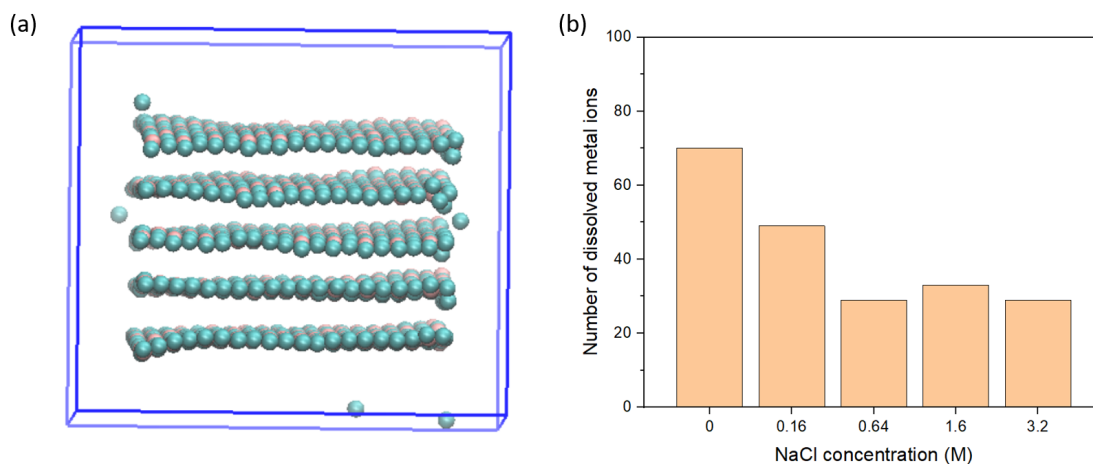


Figure 3.2: **(a)** Schematic representation of the dissolution of a small amount of metal ions from the MOH layer of LDH. **(b)** Number of dissolved metal ions at different NaCl concentrations.

surrounding solution. The effect of NaCl concentration on the amount of intercalated water molecules was insignificant, as shown in Figure 3.3d. Variations in the amount of released NO_3^- and intercalated Cl^- did not result in an effect on the interlayer distance as long as the amount of water remains constant, indicating that the amount of intercalated water is the dominant factor in determining the interlayer distance of the LDH. The steady amount of intercalated water and interlayer distance values observed in this study can be attributed to the slow process of complete NO_3^- - Cl^- exchange. This is consistent with the findings of Tedim et al. [106], as the experiment to exchange 70% of intercalated nitrate ions with chloride took about one month. It should be noted however that the performed simulations were only able to capture the initial stage of the anion exchange, and further nitrate release, chloride intercalation and hydration state variation was out of scope for this work.

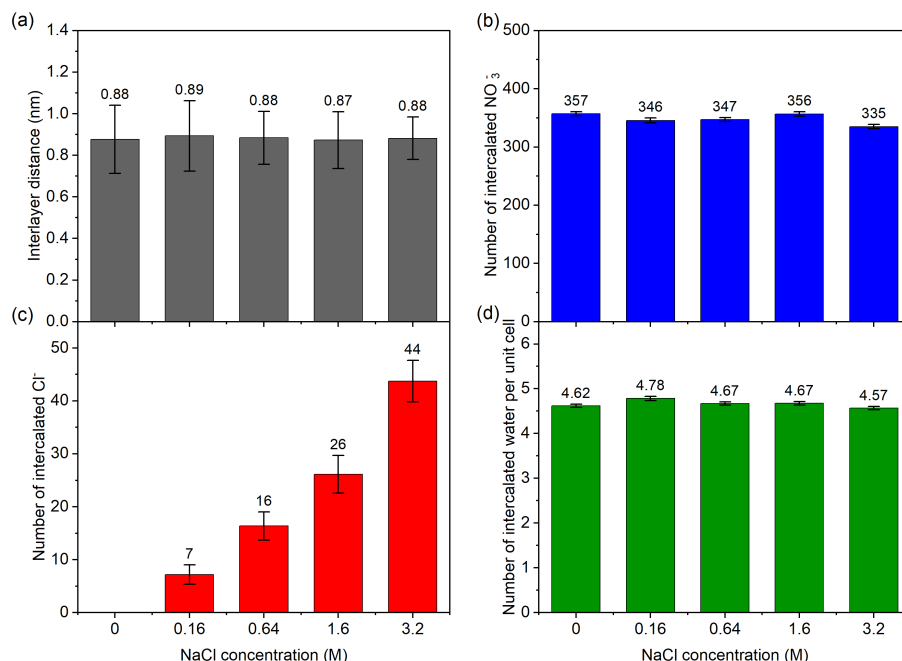


Figure 3.3: (a) Interlayer distances and numbers of intercalated (b) NO_3^- , (c) Cl^- , (d) water molecules at different NaCl concentrations.

3.3 Conclusion

In summary, the anion exchange study of LDH initially intercalated with NO_3^- and then immersed in different concentrations of NaCl solution provides more details on the initial stage of anion exchange between NO_3^- and Cl^- . A small amount of metal ions were dissolved during the MD simulation, which agrees well with the experimental observations from Buchheit et al. [21], Poznyak et al. [51], and Galvão et al. [335]. The metal ion dissolution was believed to be beneficial in the anion exchange process because it increased the contact areas between the LDH and the NaCl solution. The release of the intercalated NO_3^- occurred first at the start of the anion exchange, followed by the intercalation of the Cl^- , which is consistent with the two-step anion exchange mechanism proposed by Mikami et al. [74]. Essentially, the release of NO_3^- can create empty space in the interlayer region, into which the Cl^- from the solution can move and be intercalated. It was moreover observed that this exchange did not have an effect on the LDH interlayer distance as long as the water content remained constant, indicating that the water content in the interlayer region can be an important factor in determining the interlayer distance. Therefore, in the following part of this thesis, the influence of the hydration states on the properties of LDHs will be discussed.

4 The impact of intercalated water on LDH properties

Parts of the work and figures presented in this chapter have been published in ACS omega 7.14 (2022): 12412-12423. 'Atomistic Insight into the Hydration States of Layered Double Hydroxides' by X. Li, T. Würger, C. Feiler, RH. Meißner, M. Serdechnova, C. Blawert, ML. Zheludkevich. [336] <https://doi.org/10.1021/acsomega.2c01115>

The affinity of various anions to a metal hydroxide (MOH) layer of LDHs with identical $\text{Mg}^{2+}/\text{Al}^{3+}$ ratios has been investigated from both experimental [100] and theoretical [93] perspectives. It is noteworthy that both produced the same affinity order: $\text{NO}_3^- < \text{Br}^- < \text{Cl}^- < \text{F}^- < \text{OH}^- < \text{SO}_4^{2-} < \text{CO}_3^{2-}$. As NO_3^- has the lowest affinity and can be easily replaced by other anions, LDHs intercalated with NO_3^- are frequently employed as precursors to obtain novel LDH host-guest systems through anion exchange between NO_3^- and the anion of interest. [337, 338] Hou et al. noted that the hydration state of the LDH can be related to its anion exchange behavior based on an experimental study. [95] Additionally, other studies demonstrated that hydration is related to the anion exchange behavior and plays a significant role in determining the stability of LDHs. [90–92]

The influence of hydration states on the Al-Mg-LDH (referred to as LDH in the following paragraphs) was investigated in this work and four different LDH systems intercalated with nitrate (NO_3^-), carbonate (CO_3^{2-}), oxalate ($\text{C}_2\text{O}_4^{2-}$) and chloride (Cl^-), respectively (see Figure 4.1) were used as basis for the study. The latter is of great interest as Cl^- is an aggressive species that initiates corrosion process in Mg-based materials and may have an impact on the ion exchange process of LDHs. As a result, the corrosion progression can be mitigated by exchange of Cl^- with an intercalated corrosion inhibiting agent. The inhibitor selected for this work is oxalate ($\text{C}_2\text{O}_4^{2-}$), which has demonstrated a remarkable corrosion inhibiting effect for several magnesium alloys. [30] Although LDHs intercalated with NO_3^- , Cl^- and CO_3^{2-} have been studied in previous simulations, the details on the impact of hydration state are still lacking. The study of these three well-known LDHs can serve as a benchmark for the combination of DFT calculations and MD simulations method used in this work, in addition to providing details on the influence of hydration state. Following the benchmark, the method developed in the work can be used to investigate LDHs intercalated with new anions (e.g. $\text{C}_2\text{O}_4^{2-}$), which

can assist the design of new type of LDHs.

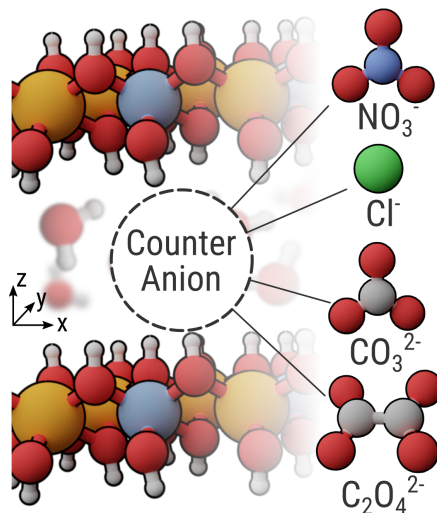


Figure 4.1: Schematic representation of a layered double hydroxide system with the counter anions investigated in this study.

For the LDH intercalated with NO_3^- , Cl^- and CO_3^{2-} , the ratio of $\text{Mg}^{2+}/\text{Al}^{3+}$ was set to two since it is one of the commonly found LDHs in experiments [76, 93]. Because there is limited information available in literature on the LDH intercalated with oxalate ($\text{C}_2\text{O}_4^{2-}$), two ratios ($\text{Mg}^{2+}/\text{Al}^{3+}$ equals to 2 and 3) were investigated to demonstrate the influence of the MOH layer charge density on the LDH stability.

4.1 Methodology

For all LDH systems, a dehydration process was carried out by decreasing the number of intercalated water. Initially, each unit cell contained six water molecules ($\text{Mg}_4\text{Al}_2(\text{OH})_{12}-2\text{NO}_3-6\text{H}_2\text{O}$ (shown in Figure 4.2a), $\text{Mg}_4\text{Al}_2(\text{OH})_{12}-2\text{Cl}-6\text{H}_2\text{O}$, $\text{Mg}_4\text{Al}_2(\text{OH})_{12}-\text{CO}_3-6\text{H}_2\text{O}$, $\text{Mg}_4\text{Al}_2(\text{OH})_{12}-\text{C}_2\text{O}_4-6\text{H}_2\text{O}$ and $\text{Mg}_6\text{Al}_2(\text{OH})_{16}-\text{C}_2\text{O}_4-6\text{H}_2\text{O}$). At each dehydration step, one water molecule per unit cell was removed from the system until no water molecules remained (as shown in Figure 4.2b). For each dehydration step, the geometry of each unit cell was optimized by employing DFT calculation using the plane-wave (PW) DFT code Vienna Ab Initio Simulation Package (VASP) [339–342] with the projector augmented wave (PAW) method [343, 344]. The exchange-correlation (XC) function optB88-vdW [345–350] was applied in all geometry optimization simulations to take the van der Waals (vdW) interactions into account for LDH. The gamma-centered grid of k -points [351] was applied to sample the first Brillouin zone. [76] A cutoff energy of 520 eV for the plane-wave expansion and an electronic convergence criterion of 1×10^{-6} eV were applied. Based on the charge densities obtained from the DFT calculations, the density-derived electrostatic and chemical (DDEC6) charges were calculated using the Chargemol program [352, 353]. These charges were calculated for each type of atom at different hydration states. The averaged value of the DDEC6 charges across different hydration states

was used for the subsequent MD simulations.

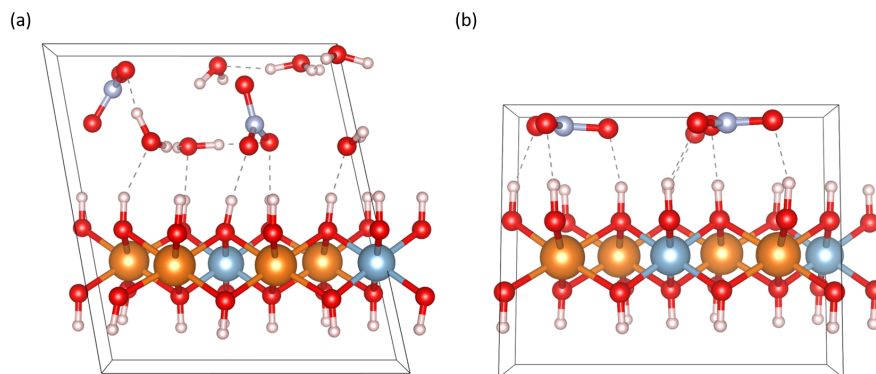


Figure 4.2: DFT optimized structures for (a) $\text{Mg}_4\text{Al}_2(\text{OH})_{12}-2\text{NO}_3-6\text{H}_2\text{O}$ and (b) $\text{Mg}_4\text{Al}_2(\text{OH})_{12}-2\text{NO}_3-0\text{H}_2\text{O}$.

The optimized unit cell with six water molecules obtained from DFT was replicated $5 \times 10 \times 5$ for $\text{Mg}_4\text{Al}_2\text{-LDH-NO}_3^-$ and $\text{Mg}_4\text{Al}_2\text{-LDH-Cl}^-$ and $10 \times 5 \times 5$ for $\text{Mg}_4\text{Al}_2\text{-LDH-CO}_3^{2-}$ and both $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ and $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ to obtain a larger system as input for the MD simulations. Periodic boundary conditions were applied in all MD simulations to avoid finite size effects and all presented MD simulations were performed using the GROMACS 5.1.5 package [354]. The leapfrog algorithm [355] was used for the integration of the equations of motion. The total potential energy in MD is a sum of non-bonded interactions like Lennard-Jones (LJ), electrostatic interactions and bonded interactions like bond stretching, angle bending, and dihedral torsion. A cutoff was set to 1.4 nm for the non-bonded interactions, with a potential force-switch modifier function for LJ and a combination of the Particle Mesh Ewald (PME) method [356] and the Coulomb potential-shift function for the long-range electrostatic interactions. More details on how to apply these potential switch methods can be found in the work of Perez-Sanchez et al. [76] The LJ parameters for the MOH layers were taken from the CLAYFF force field [185]. The SPC/E model [357] was selected for the intercalated water. The required parameters of the anions were adopted from the work of Cadena et al. [358] for NO_3^- , from Smith et al. [359] for Cl^- and from Schmid et al. [360] for CO_3^{2-} . To obtain the parameters for $\text{C}_2\text{O}_4^{2-}$, the Automated Topology Builder repository [361] based on the GROMOS force field [360] was utilized.

Each system was subjected to a step-wise optimization process for the MD simulations, beginning with an energy minimization step. Following that, simulation steps in the canonical ensemble (NVT) and isothermal–isobaric ensemble (NPT) were performed to reach the equilibrium state. After that, an NPT production run was carried out to obtain the input for the data analysis. The thermostat and barostat used at different steps were identical to those used by Perez-Sanchez et al. [76] In the equilibrium steps, the temperature was fixed at 298 K using velocity-rescaling [362] and the pressure was fixed at 1 bar with the Berendsen pressure-coupling method [363]. The Nosé-Hoover thermostat [364, 365] and the Parrinello-Rahman barostat [366] were used to fix the temperature and pressure during the production run. After

The impact of intercalated water on LDH properties

the production simulation, a fixed amount of water (one water per unit cell) was removed and the production step was re-run in order to achieve a new equilibrium state for the new system. Following the same dehydration procedure as in the DFT calculations, one water per unit cell was removed step by step until no intercalated water remained in the system. Initially, seven hydration states for each type of LDH were considered, ranging from six to zero water molecules in the LDH intergallery. Furthermore, two extra hydration states with the number of water per unit cell below one were carried out for the $\text{Mg}_4\text{Al}_2\text{-LDH-Cl}^-$, $\text{Mg}_4\text{Al}_2\text{-LDH-CO}_3^{2-}$ and $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ systems to add more data points to the definition of their stable hydration, respectively. These two states were created by removing respectively a half and three quarters of the total amount of intercalated water from the state with one water per unit cell.

After the MD simulations, the structure properties including number density distribution, radial distribution function (RDF), orientation of the intercalated anions (angle calculation between anion plane and the MOH layer) as well as the self-diffusion coefficients of the intercalated anions were obtained from the last 10 ns of the production simulations using the `gmx density`, `gmx rdf`, `gmx gangle` and `gmx msd` tools in GROMACS. The total potential energy was obtained by the `gmx energy` tool. The hydration energy was defined as:

$$E_{\text{Hydration}} = \frac{1}{N} (U(N) - U(0)) \quad (4.1)$$

where $U(N) - U(0)$ is the total potential energy difference between the state with N intercalated water molecules and the state without intercalated water. The interlayer distance of the LDH was estimated by dividing the simulation box size in z -direction by the number of explicitly modeled LDH layers. The number density of any type of atoms was calculated along z -direction to provide the cross-section information of the LDH. In this calculation, the whole system was divided into 500 cells along z -direction and the number of the atoms was accordingly calculated at each cell and subsequently divided by its volume. As there were five LDH layers in z -direction, five repeated patterns were obtained for the number density distribution. To get a good overview of the whole LDH, the five repeated patterns were averaged. All angles between the anion plane and the MOH layer were collected in the last 10 ns. The probability for different angles was calculated to show the angle distribution.

The radial distribution function $g(r)$ was applied to determine the relative position between intercalated water and anions. Furthermore, the coordination number of water molecules around the anions could be obtained by integration of $g(r)$ over distance (r). After the space around the anions was divided into equally spaced shells, the $g(r)$ was calculated with the densities (ρ) of the water molecules in the shells divided by the averaged density of the water in the system:

$$g(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{\text{local}}} = \frac{1}{\langle \rho_B \rangle_{\text{local}}} \frac{1}{N_A} \sum_{i \in N_A} \sum_{j \in N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2} \quad (4.2)$$

where N denotes the number of atoms, B the intercalated water molecules and A the intercalated anions. In the RDF calculation, the center of mass was used for the intercalated anions and the oxygen atom for H_2O . The self-diffusion coefficient (D) was calculated by the Einstein

4.2 The impact of different hydration states on Mg₄Al₂-LDH systems

equation [367], relating the mean squared displacement (MSD) to the time:

$$\lim_{t \rightarrow \infty} (r_i(t) - r_i(0))_{i \in A}^2 = 4 D_A t \quad (4.3)$$

where $r_i(t) - r_i(0)$ represents the displacement of the anions (A) from their reference position and t denotes time. Analogously to the RDF calculation, the center of mass was used to define the position of anions and to calculate their displacement. The calculation of the self-diffusion coefficient focused only on the x - and y -direction since the diffusion in z -direction can be neglected after the system reached equilibrium. The hydrogen bonds were calculated based on the cutoffs of the angle (hydrogen-donor-acceptor) and the distance (donor-acceptor) in GROMACS with command `gmx hbond`. The cutoff of the angle is $\approx 30^\circ$ and the distance cutoff is 0.35 nm.

The simulation results are divided into results and discussion on four Mg₄Al₂-LDH systems (Mg₄Al₂-LDH-NO₃⁻, Mg₄Al₂-LDH-Cl⁻, Mg₄Al₂-LDH-CO₃²⁻ and Mg₄Al₂-LDH-C₂O₄²⁻) and on one Mg₆Al₂-LDH system (Mg₆Al₂-LDH-C₂O₄²⁻). According to Tavares et al. [166], hydration can induce changes in the electronic structure of an LDH, which in turn affects the partial charges of the atoms in the LDH. Furthermore, the partial charge parameters of the force field can have a significant impact on the calculation of electrostatic interactions and the outcome of MD simulations, particularly the stability of the LDH. [76] As a result, it is critical to ensure that the charge values used as the basis for parameterizing the employed force field are reasonable, reliable, and, ideally, compatible with other commonly used force fields. For this reason, the partial charges were calculated at different hydration states for the atoms in the MOH layer as well as for the intercalated molecules in each unit cell of the four Mg₄Al₂-LDH systems and one Mg₆Al₂-LDH, using DFT calculations in combination with the atomic population analysis method DDEC6. The structure and stability properties of different LDHs with varying hydration states were investigated based on MD simulations and the results are discussed in more details in the following paragraphs.

4.2 The impact of different hydration states on Mg₄Al₂-LDH systems

For the four Mg₄Al₂-LDH systems (Mg₄Al₂-LDH-NO₃⁻, Mg₄Al₂-LDH-Cl⁻, Mg₄Al₂-LDH-CO₃²⁻ and Mg₄Al₂-LDH-C₂O₄²⁻, referred to as LDH-NO₃⁻, LDH-Cl⁻, LDH-CO₃²⁻ and LDH-C₂O₄²⁻ in the following paragraphs), the partial charges were calculated at different hydration states for the atoms in the MOH layer and the intercalated molecules (see Figure 4.3) of each unit cell. The DDEC6 results show that the influence of the water content on the partial charges can be neglected because they are unaffected by the number of intercalated water molecules. It should be noted that the partial charges of different types of atoms were averaged over different hydration states. The partial charges of the atoms in the MOH layer were then averaged across the four Mg₄Al₂-LDH systems, and partial charges of different intercalated anions were streamlined to approximate, homogenize, and account for the various oxidation states in the system. For the intercalated water molecules, the partial charges from the SPC/E model [357]

The impact of intercalated water on LDH properties

were employed. Finally, the values listed in Table 4.1 were subsequently used as basis for the MD simulations.

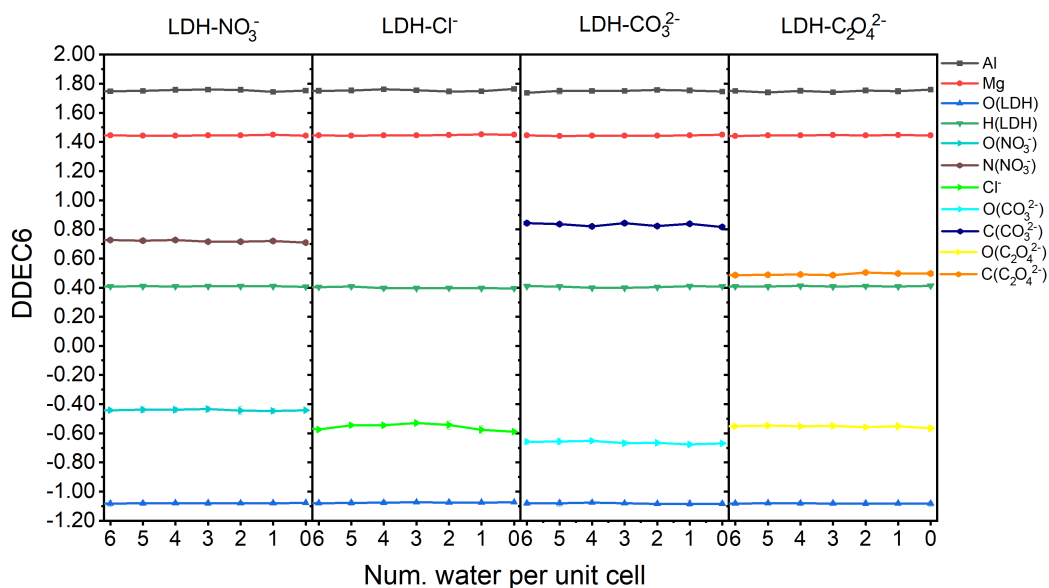


Figure 4.3: DDEC6 charges of different types of atoms for each of the investigated LDH systems (from left to right are LDH- NO_3^- , LDH- Cl^- , LDH- CO_3^{2-} and LDH- $\text{C}_2\text{O}_4^{2-}$) at different hydration states. The color code is shown in the legend and as follows: black for aluminium, red for magnesium, blue for hydroxide oxygen, dark green for hydroxide hydrogen, turquoise for nitrate oxygen, brown for nitrate nitrogen, green for chloride, cyan for carbonate oxygen, navy for carbonate carbon, yellow for oxalate oxygen and orange for oxalate carbon.

As shown by the red lines in Figure 4.4, the interlayer distance of each type of LDH decreases with decreasing water content in the MD simulations. The black lines depict the LDH hydration energy with respect to different water contents. The most stable hydration state is defined by the water content corresponding to the minimum hydration energy for LDH- NO_3^- (three H_2O per unit cell) or the water content initiating a plateau for LDH- Cl^- (two H_2O per unit cell), LDH- CO_3^{2-} (half H_2O per unit cell, which equals one H_2O per two unit cells) and LDH- $\text{C}_2\text{O}_4^{2-}$ (two H_2O per unit cell). It is noteworthy that there is no further decrease in hydration energy when the number of water molecules per unit cell is less than one for the LDH- Cl^- and LDH- $\text{C}_2\text{O}_4^{2-}$ systems. The hydration energy for the LDH- CO_3^{2-} system is clearly lower than for the other three systems investigated in this work, indicating the extremely high stability of the intercalated CO_3^{2-} as mentioned by Sasai et al. [368] It achieves its stable hydration state by removing half of the total amount of water at the state with one water molecule per unit cell. The interlayer distance at the most stable hydration state is the same as the experimental

4.2 The impact of different hydration states on Mg_4Al_2 -LDH systems

Table 4.1: DDEC6 derived partial charges (q , in a.u.) of different types of atoms used in MD simulations for LDH-NO_3^- , LDH-Cl^- , LDH-CO_3^{2-} and $\text{LDH-C}_2\text{O}_4^{2-}$ systems.

	Element	LDH-NO_3^-	LDH-Cl^-	LDH-CO_3^{2-}	$\text{LDH-C}_2\text{O}_4^{2-}$
MOH layer	Al		1.7700		
	Mg		1.4400		
	O(LDH)		-1.0800		
	H(LDH)		0.4000		
Intergallery	O(NO_3^-)	-0.4300	-	-	-
	N(NO_3^-)	0.7200	-	-	-
	Cl^-	-	-0.5700	-	-
	O(CO_3^{2-})	-	-	-0.6600	-
	C(CO_3^{2-})	-	-	0.8400	-
	O($\text{C}_2\text{O}_4^{2-}$)	-	-	-	-0.5340
	C($\text{C}_2\text{O}_4^{2-}$)	-	-	-	0.4980
	O(H_2O)	-0.8476	-0.8476	-0.8476	-0.8476
	H(H_2O)	0.4238	0.4238	0.4238	0.4238

reference for the LDH-NO_3^- system [369], which is 0.89 nm. For the LDH-Cl^- system, the interlayer distance is 0.84 nm, which is slightly higher than the experimentally determined value of 0.79 nm [370]. The interlayer distance of LDH-CO_3^{2-} is 0.73 nm at its most stable hydration state, which is slightly lower than the experimental value of 0.76 nm. The interlayer distance of $\text{LDH-C}_2\text{O}_4^{2-}$ is 0.75 nm, which is smaller than that of the LDH-Cl^- system and very close to the experimentally known interlayer distance of the LDH-CO_3^{2-} of 0.76 nm.

To elucidate the orientation and position of the intercalated anions relative to the MOH layer, the number density of the principal atoms, listed in Table 4.1, was investigated along z -direction at the most stable hydration state (denoted as stable hydration state) and the higher hydration state next to it (named as high hydration state) for the four Mg_4Al_2 -LDH systems investigated in this work. The number densities of the atoms belonging to the MOH layers (Al, Mg, H (OH^-)) of LDH-NO_3^- and LDH-Cl^- appear to be less stable at the high hydration state compared to the stable hydration state, resulting in the broader peaks in Figure 4.5a and the formation of a wave-like pattern as shown in Figure 4.5c. The peak broadening and wave-like pattern formation are both caused by the fact that the atoms in the MOH layers are not perfectly in-plane at the high hydration state. When the corresponding LDH is at the stable hydration state (Figure 4.5b and d), the peak broadening and wave-like pattern disappear. The number densities of the atoms in the MOH layers of LDH-CO_3^{2-} and $\text{LDH-C}_2\text{O}_4^{2-}$ are analogous for the high and the stable hydration states as shown in Figures 4.5e, f, g and h. This indicates that the MOH layers are stable in both states and are unaffected by the changes in intercalated water content for these two types of LDHs. At the stable hydration state of LDH-NO_3^- , there are two distinct O peaks of NO_3^- and H_2O (Figure 4.5b), respectively, indicating that most of the O atoms (of both NO_3^- and H_2O) are close to the MOH layers. In contrast, the distribution of O (for both NO_3^- and H_2O) at the high hydration state is broad across the entire interlayer

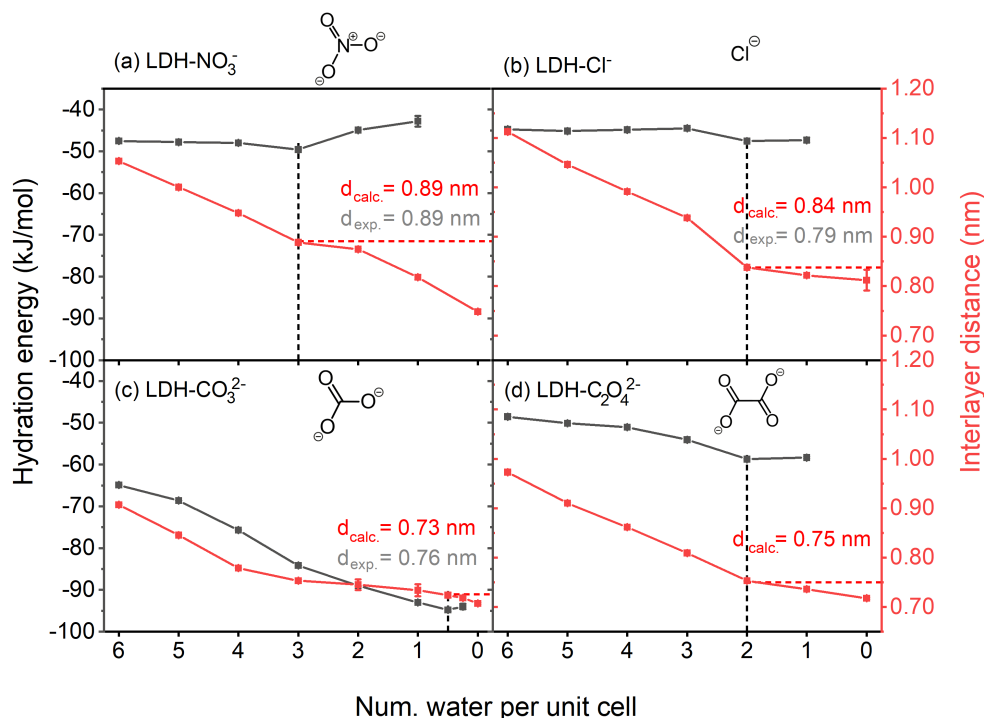


Figure 4.4: Hydration energy (black) and interlayer distance (red) as function of the water content. The dotted black lines are used as guide for the eye to highlight the most stable hydration states for each of the investigated LDH systems (a) LDH-NO₃⁻, (b) LDH-Cl⁻, (c) LDH-CO₃²⁻ and (d) LDH-C₂O₄²⁻. In a similar fashion, the dotted red lines were included to mark the corresponding interlayer distance. The values of the interlayer distance obtained from simulations ($d_{\text{calc.}}$) are written in red while the corresponding experimental reference ($d_{\text{exp.}}$) is shown in grey. The experimental reference for the LDH-C₂O₄²⁻ system cannot be provided in this work as the exchange procedure needs to be further optimized.

region (Figure 4.5a). The difference of the O distribution of NO₃⁻ between stable and high hydration states can be partially attributed to the rotation of the trigonal planar structure of the NO₃⁻ anion, as shown in the inlays in Figures 4.5a and b. This rotation leads to a smaller interlayer distance and a stronger interaction between the intercalated compounds and the MOH layers at the stable hydration state. For the LDH-Cl⁻, the distribution of O (from H₂O) at the high hydration state in Figure 4.5c differs from that at the stable hydration state in Figure 4.5d. At the high hydration state, two layers of water form in the interlayer region and these two layers of water are adjacent to the MOH layers. The two layers of water merge into one layer at the stable hydration state in Figure 4.5d. However, there is no discernible difference in the distribution of Cl⁻ at both high and stable hydration states, as shown in Figures 4.5c and d. Additionally, the distribution of O (from both CO₃²⁻ and H₂O) is similar for the two hydration states in the LDH-CO₃²⁻ system (Figure 4.5e and f). For LDH-C₂O₄²⁻, there are three O (C₂O₄²⁻) peaks at the high hydration state (Figure 4.5g), which convert into one O

4.2 The impact of different hydration states on $\text{Mg}_4\text{Al}_2\text{-LDH}$ systems

($\text{C}_2\text{O}_4^{2-}$) peak at the stable hydration state (Figure 4.5h), indicating that the $\text{C}_2\text{O}_4^{2-}$ plane is first tilted to the MOH layers before shifting to be parallel, respectively (see Figure 4.6).

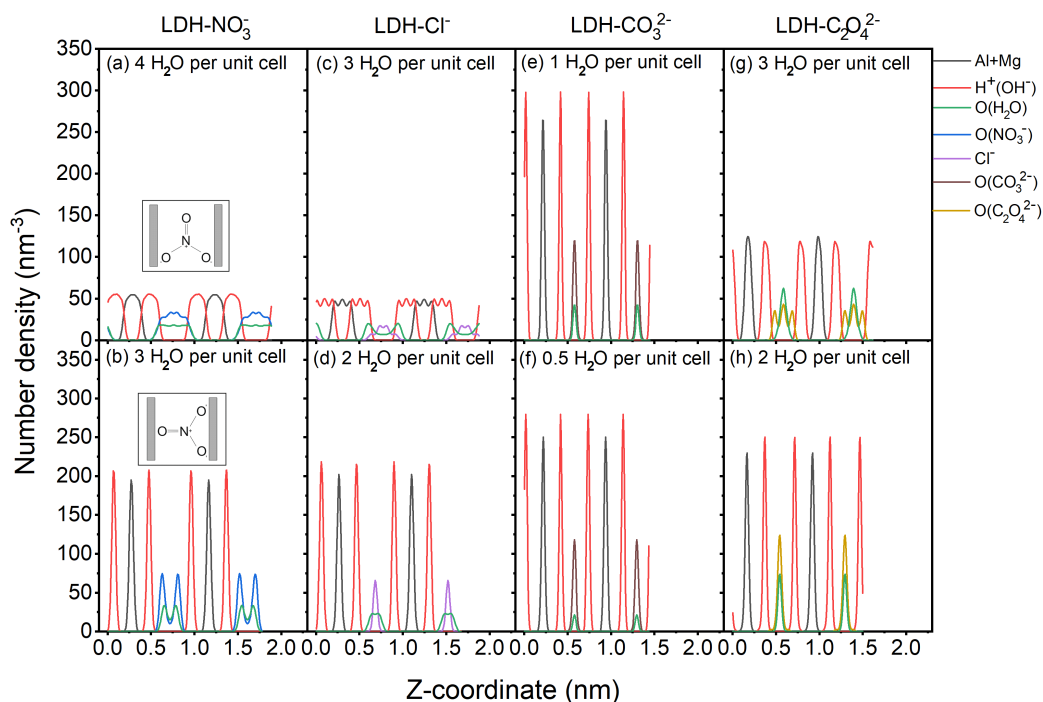


Figure 4.5: Number density of different types of atoms along z -direction for LDH- NO_3^- (a) high hydration state: four H_2O per unit cell and (b) stable hydration state: three H_2O per unit cell. The inserted illustrations are two schematic diagrams of how one NO_3^- can be intercalated between two MOH layers, represented by grey bars. For LDH- Cl^- (c) high hydration state: three H_2O per unit cell and (d) stable hydration state: two H_2O per unit cell. For LDH- CO_3^{2-} (e) high hydration state: one H_2O per unit cell and (f) stable hydration state: half H_2O per unit cell. For LDH- $\text{C}_2\text{O}_4^{2-}$ (g) high hydration state: three H_2O per unit cell and (h) stable hydration state: two H_2O per unit cell. Black depicts aluminium and magnesium, red the hydrogen atoms of the hydroxide group in the MOH layer, green the oxygen of the intercalated water molecules, blue for the oxygen atoms of nitrate, purple for chloride, brown for oxygen atoms in carbonate and orange for oxygen atoms in oxalate.

The angle distributions between the NO_3^- plane and the MOH layer are similar in the high and stable hydration states, as shown in Figure 4.6a, indicating that the majority of the NO_3^- anions are tilted by $\approx 70^\circ$ relative to the MOH layer in both states. This observation is consistent with the findings from Marappa et al. [371] and Perez-Sanchez et al. [76]. At both hydration states of the LDH- CO_3^{2-} system, all CO_3^{2-} anions are parallel to the MOH layer (see Figure 4.6b). In contrast, the orientation of the anion in the LDH- $\text{C}_2\text{O}_4^{2-}$ system is strongly affected by the change in hydration state, as shown in Figure 4.6c. Over 50% of the intercalated $\text{C}_2\text{O}_4^{2-}$ anions are tilted with respect to the MOH layer at high hydration state, while the anion plane of

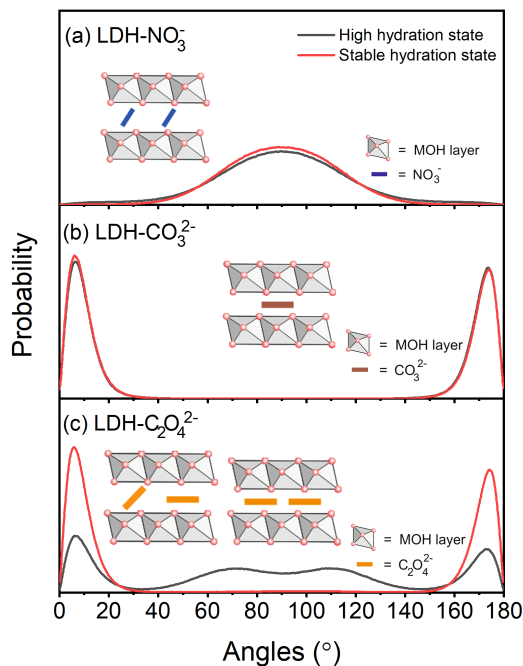


Figure 4.6: Angle distribution between the (a) NO_3^- , (b) CO_3^{2-} and (c) $\text{C}_2\text{O}_4^{2-}$ anion planes and the MOH layer at high and stable hydration states. The inserted illustrations in (a), (b) and (c) show the schematic diagrams at the two hydration states for the respective system, where the left section depicts the high and the right section the stable hydration state in (c). Black lines depict the high and red lines the stable hydration state. The definition of the anion plane for NO_3^- , CO_3^{2-} and $\text{C}_2\text{O}_4^{2-}$ is provided in Figure A.1 in Appendix A.1.

some $\text{C}_2\text{O}_4^{2-}$ is in parallel to the MOH layer. At the stable hydration state, all $\text{C}_2\text{O}_4^{2-}$ anions adopt a parallel orientation relative to the MOH layer, indicating that the orientation of the intercalated $\text{C}_2\text{O}_4^{2-}$ is strongly dependent on the number of present water molecules in the interlayer. This result is corroborated by the behavior of intercalated CO_3^{2-} which is also observed to be oriented parallel to the MOH layer.

To determine the relative position differences between water molecules and anions at high and stable hydration states, the RDFs $g(r)$ of the water molecules around the anions were calculated. From the cumulative number of $g(r)$, one can infer the coordination number of water molecules around the intercalated anions. For LDH- NO_3^- , LDH- Cl^- and LDH- CO_3^{2-} in Figures 4.7a, c and e, respectively, the peak of $g(r)$ appears at the same distance for both states. The cumulative number calculations in Figures 4.7b, d and f show that for these three types of LDHs, the coordination number of water around NO_3^- , Cl^- and CO_3^{2-} is proportional to the amount of intercalated water molecules. When the amount of intercalated water per unit cell for LDH- NO_3^- is reduced from four to three, the cumulative number decreases proportionally

4.2 The impact of different hydration states on $\text{Mg}_4\text{Al}_2\text{-LDH}$ systems

from 4.9 to 3.7. For LDH-Cl^- , the cumulative number changes from 4 to 2.7 when the amount of intercalated water is reduced from three H_2O to two H_2O per unit cell. The cumulative number of water around the CO_3^{2-} anions is halved from 2 to 1 after removing half of the total intercalated water, transitioning from the high hydration state to the stable state. The change in the number of coordinated water molecules around the anions for the different hydration states of LDH-NO_3^- , LDH-Cl^- and LDH-CO_3^{2-} can be explained by the ionic radii of the involved species. The tilted NO_3^- (179 pm) [372], Cl^- (172 pm) [372] and CO_3^{2-} (178 pm) [372] all exhibit a larger ionic radius than the water molecules (138 pm) [373] and determine the LDH interlayer spacing. As a result, as shown in the respective inlays in Figures 4.7b, d and f, the intercalated water molecules primarily fill the empty space, explaining the correlation between the coordination number and the amount of intercalated water.

For $\text{LDH-C}_2\text{O}_4^{2-}$ (the ionic radius of $\text{C}_2\text{O}_4^{2-}$ is 200 pm [374]), the $g(r)$ peaks of the two hydration states display slightly different shapes (see Figure 4.7g). There is a noticeable peak shift of the $g(r)$ for the stable hydration state (two water molecules) compared to the high hydration state (three water molecules), as illustrated in Figure 4.7g. Furthermore, the cumulative numbers of $g(r)$ (Figure 4.7h) at these two states are close to each other, which differs from the proportional behavior observed in the other three systems. These are related to the change in angle distribution, as illustrated in Figure 4.6c. After the water content was reduced from three to two H_2O per unit cell, all $\text{C}_2\text{O}_4^{2-}$ anions were orientated parallel to the MOH layer. Naturally, this implies a reorganization of the intercalated water molecules that surround the oxalate (see Figure 4.5h) and a smaller interlayer distance. Since the $g(r)$ calculation in the z -dimension is more constrained at the stable hydration state, the corresponding $g(r)$ peak is shifted to a higher distance value. This also leads to a larger cutoff distance for calculating the cumulative numbers, resulting in a similar number of water molecules coordinated to the $\text{C}_2\text{O}_4^{2-}$ anion in both hydration states. Therefore, the number of coordinated water molecules around $\text{C}_2\text{O}_4^{2-}$ is not proportional to the amount of intercalated water molecules, although the ionic radius of $\text{C}_2\text{O}_4^{2-}$ is also larger than that of the water molecules.

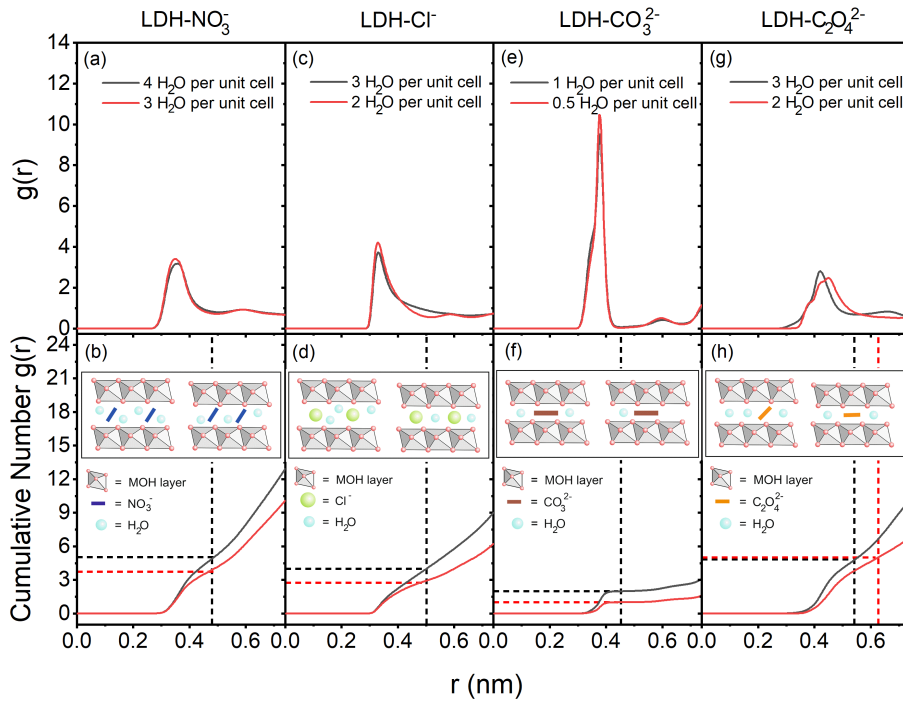


Figure 4.7: RDFs $g(r)$ of water molecules around the intercalated anions and their respective cumulative numbers for the four types of LDHs at high hydration state (in black continuous lines) and stable hydration state (in red continuous lines). For LDH- NO_3^- (a) $g(r)$ and (b) cumulative number at four and three H_2O per unit cell. For LDH- Cl^- (c) $g(r)$ and (d) cumulative number at three and two H_2O per unit cell. For LDH- CO_3^{2-} (e) $g(r)$ and (f) cumulative number at one and a half H_2O per unit cell. For LDH- $\text{C}_2\text{O}_4^{2-}$ (g) $g(r)$ and (h) cumulative number at three and two H_2O per unit cell. The inserted illustrations in Figures (b), (d), (f) and (h) show the schematic diagrams for the four types of LDHs at the two hydration states, the left section showing the high and the right section the stable hydration state. The vertical dashed lines mark the cutoff distances for the first peak of $g(r)$ and the cutoffs for calculating the cumulative number of $g(r)$. The horizontal dashed lines were included to determine the number of coordinated water molecules around the anions at different states.

The self-diffusion coefficient of the intercalated anions was calculated based on their MSD values (see Appendix A.1 Figure A.2) to determine the stability of the LDH. The simulations indicate that when there are fewer water molecules in the interlayer region, the self-diffusion coefficient of an intercalated anion decreases. The LDH- NO_3^- system is an exception to this observation, as the self-diffusion coefficient of NO_3^- ($D_{\text{NO}_3^-}$) increases when the amount of intercalated water is reduced from three H_2O (stable state) to two H_2O per unit cell (Figure 4.8a). This can be attributed to the implied subtle change in interlayer distance (see Figure 4.4a). As a result, water removal creates empty space in the interlayer for the lower hydration state, allowing for less impaired relocation of the intercalated NO_3^- across the xy -plane. The self-

4.2 The impact of different hydration states on Mg₄Al₂-LDH systems

diffusion coefficient of the intercalated Cl⁻ ion (D_{Cl^-}) does not increase when the amount of water is lower than that of the stable hydration state containing two water molecules (Figure 4.8b). A D_{Cl^-} plateau appears, similar to the hydration energy plateau shown in Figure 4.4b. Overall, the change in D_{Cl^-} with respect to the hydration state is more significant compared to the change in $D_{\text{NO}_3^-}$, which can be attributed to the size difference of these two monovalent anions (the size of Cl⁻ (172 pm) is smaller than NO₃⁻ (179 pm)). The self-diffusion coefficient of the intercalated CO₃²⁻ ($D_{\text{CO}_3^{2-}}$) approaches zero when the amount of intercalated water is reduced to three water molecules per unit cell, as shown in the inlay of Figure 4.8c. There is no discernible decrease of $D_{\text{CO}_3^{2-}}$ after the amount of intercalated water is reduced to less than three water molecules per unit cell. Similar to $D_{\text{NO}_3^-}$, the self-diffusion coefficient of C₂O₄²⁻ ($D_{\text{C}_2\text{O}_4^{2-}}$) increases slightly when the number of intercalated water molecules per unit cell is one lower than for the stable hydration state, as shown in the inlay of Figure 4.8d. However, it is noteworthy that $D_{\text{C}_2\text{O}_4^{2-}}$ is generally low (similar to $D_{\text{CO}_3^{2-}}$). The relatively low self-diffusion coefficients $D_{\text{CO}_3^{2-}}$ and $D_{\text{C}_2\text{O}_4^{2-}}$ are presumably due to the ability of both the CO₃²⁻ and C₂O₄²⁻ ions to form multiple ionic bonds, resulting in a stronger interaction with the MOH layer than the other two investigated ions. In conclusion, it was discovered that the self-diffusion coefficient is highly dependent on the hydration state and anion type. Lower D values for the intercalated anions indicate a highly stable LDH host-guest system. At the stable hydration state, the self-diffusion coefficients of the four anions decrease in the following order: NO₃⁻ > Cl⁻ > CO₃²⁻ and C₂O₄²⁻. Concomitantly, the stability order obtained from the calculations agrees well with affinity orders from literature: NO₃⁻ < Cl⁻ < CO₃²⁻. For CO₃²⁻ and C₂O₄²⁻, the self-diffusion coefficients at the stable hydration state are equivalent. The intercalated CO₃²⁻ and C₂O₄²⁻ ions possess a lower D value compared to the intercalated NO₃⁻ and Cl⁻ anions studied in this work, indicating that both the LDH-CO₃²⁻ and LDH-C₂O₄²⁻ systems have more stable structures.

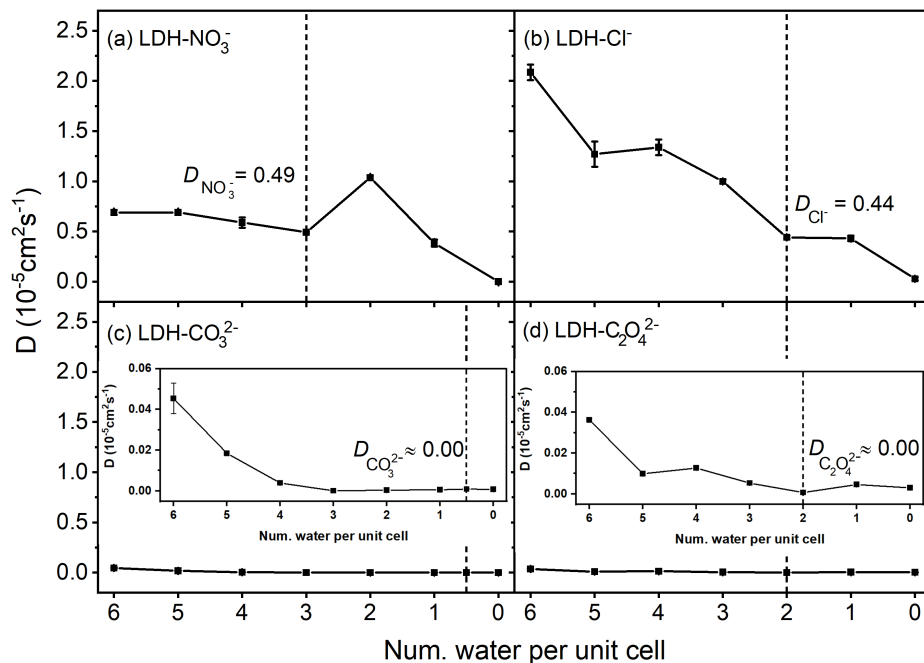


Figure 4.8: Self-diffusion coefficient at different hydration states in x - and y -direction, calculated for (a) NO_3^- , (b) Cl^- , (c) CO_3^{2-} and (d) $\text{C}_2\text{O}_4^{2-}$. The inlays in (c) and (d) show the enlarged view in y -direction of (c) and (d). The dotted vertical lines highlight the stable hydration states for each of the investigated LDH systems.

The hydrogen-bonded network of each LDH system was investigated at different hydration states, since Sasai et al. [368] concluded that this network is important for the stability of the LDH structure. The average number of hydrogen bonds per intercalated water was calculated by dividing the sum of the three types of hydrogen bonds (as shown in Figure 4.9a) by the total number of intercalated water molecules. Hydrogen bonds formed between intercalated anions and -OH groups of the MOH layer were excluded from the calculation as these were not directly connected to the intercalated water molecules. As shown in Figure 4.9b, the average number of hydrogen bonds per water molecule typically increases at the start of the dehydration process and then reaches a plateau for all four systems. A possible reason for the formation of this plateau could be the reorganization of the intercalated water molecules to maintain the hydrogen-bonded network for the stability. The plateau commences at two water per unit cell for LDH- $\text{C}_2\text{O}_4^{2-}$ and LDH- Cl^- , and at three water per unit cell for LDH- CO_3^{2-} and LDH- NO_3^- . For LDH- NO_3^- , LDH- Cl^- and LDH- $\text{C}_2\text{O}_4^{2-}$, the beginning of the plateau correlates with the stable hydration state described in Figure 4.4a, b and d. The beginning of the plateau for the LDH- CO_3^{2-} is at a higher hydration state than the stable hydration state. The interlayer distance of LDH- CO_3^{2-} does not significantly change from the beginning state of the plateau to its stable hydration state (see Figure 4.4c) and all intercalated CO_3^{2-} are parallel to the

MOH layer as shown in Appendix A.1 Figure A.3. This implies that LDH- CO_3^{2-} is stable in all of these hydration states, since the removal of a water molecule does not destabilize the existing hydrogen-bonded network. The LDH- Cl^- system formed the lowest average number of hydrogen bonds per water molecule as shown in Figure 4.9b. This could be attributed to the fact that the Cl^- provides a lower number of acceptors in the calculation of the hydrogen bonds than the other three intercalated anions. At the plateaus of the four systems, the LDH- CO_3^{2-} and LDH- $\text{C}_2\text{O}_4^{2-}$ formed more hydrogen bonds than the LDH- NO_3^- and LDH- Cl^- , supporting the observation that these two systems have more stable structures. As a result, the intercalated $\text{C}_2\text{O}_4^{2-}$ ion is likely to behave similarly to the intercalated CO_3^{2-} , which can not be easily replaced by NO_3^- nor Cl^- via an ion exchange mechanism.

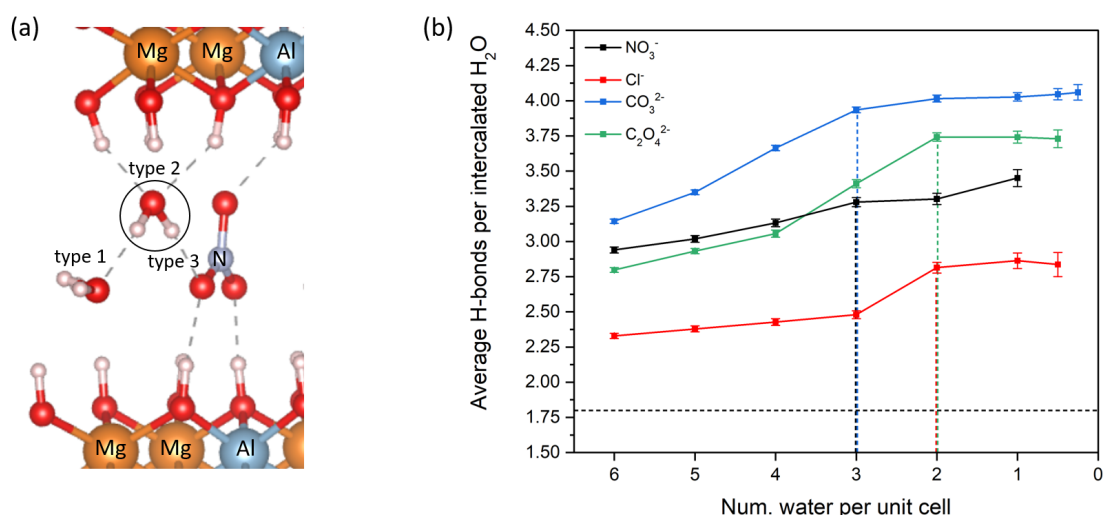


Figure 4.9: (a) Schematic representation of a hydrogen-bonded network formed around an intercalated water molecule in an LDH- NO_3^- system. Three types of hydrogen bonds are considered: hydrogen bonds formed between water molecules (type 1), hydrogen bonds formed between water and -OH groups of the MOH layer (type 2), and hydrogen bonds formed between water and intercalated anions (type 3). Average hydrogen bonds (H-bonds) per intercalated water was obtained by dividing the sum of the three types of hydrogen bonds by the total number of intercalated water. (b) Average H-bonds per intercalated water at different hydration states for LDH- NO_3^- (black), LDH- Cl^- (red), LDH- CO_3^{2-} (blue) and LDH- $\text{C}_2\text{O}_4^{2-}$ (green). The dotted vertical lines highlight the beginning state of the plateau for each of the investigated LDH systems. The dotted horizontal line stands for the number of hydrogen bonds per water of bulk water obtained from MD calculations.

4.3 Changing the composition of LDH

Previous simulation results on $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ show that the intercalated $\text{C}_2\text{O}_4^{2-}$ anions are unlikely to be easily replaced by Cl^- through anion exchange. As a result, the intercalated $\text{C}_2\text{O}_4^{2-}$, which is a strong corrosion inhibitor for several Mg alloys [30], may not be released to inhibit corrosion without an additional trigger. This can be attributed to the strong interaction

The impact of intercalated water on LDH properties

between the intercalated $\text{C}_2\text{O}_4^{2-}$ and the MOH layer of LDHs. As a strategy to weaken the interaction and support the release of intercalated $\text{C}_2\text{O}_4^{2-}$, the charge density of the MOH layer was adjusted by changing the $\text{Mg}^{2+}/\text{Al}^{3+}$ ratio from 2 to 3 as shown in Figure 4.10. For the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system, the influence of the hydration on the DDEC6 partial charges was studied and, as seen from Figure 4.11, the influence can be neglected as in the previously mentioned four $\text{Mg}_4\text{Al}_2\text{-LDH}$ systems. The averaged partial charges for the same atoms over different hydration states are comparable between the $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system and the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system. In the MD simulations for both systems, the same partial charge values are applied to the intercalated $\text{C}_2\text{O}_4^{2-}$ and H_2O . To keep the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system neutral, the partial charges of the atoms in the MOH layer are slightly adjusted. An overview of the values used in the MD simulations is reported in Table 4.2.

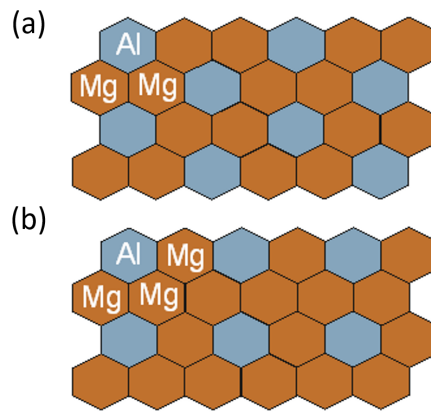


Figure 4.10: Schematic representation of the distribution of Al and Mg atoms in the MOH layer of (a) $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ and (b) $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ systems.

Table 4.2: DDEC6 derived partial charges (q , in a.u.) of different types of atoms used in MD simulations for $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ and $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ systems

	Element	$\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$	$\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$
MOH layer	Al	1.7700	1.7400
	Mg	1.4400	1.4500
	O(LDH)	-1.0800	-1.0900
	H(LDH)	0.4000	0.4000
Intergallery	O($\text{C}_2\text{O}_4^{2-}$)	-0.5340	
	C($\text{C}_2\text{O}_4^{2-}$)	0.4980	
	O(H_2O)	-0.8476	
	H(H_2O)	0.4238	

According to the previous definition, the minimum hydration energy of the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system at four H_2O molecules per unit cell is the most stable hydration state, as shown in Figure 4.12. The corresponding interlayer distance is 0.79 nm, which is larger than that of the

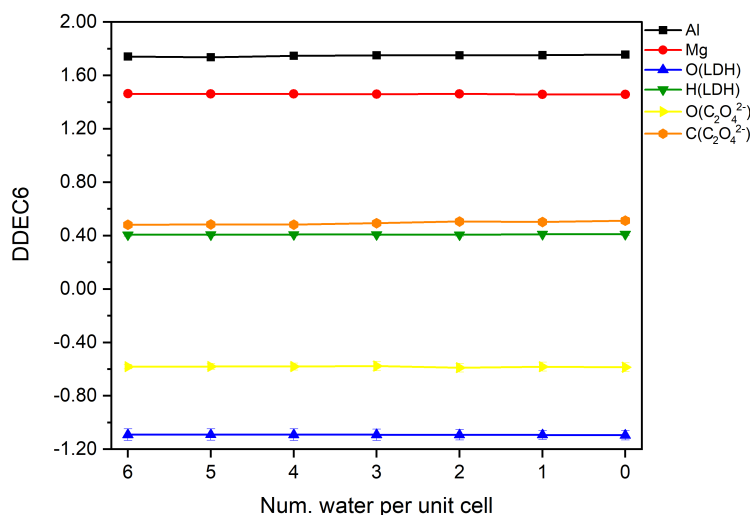


Figure 4.11: DDEC6 charges of different types of atoms for the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system at different hydration states. The color code is shown in the legend and as follows: black for aluminium, red for magnesium, blue for hydroxide oxygen, dark green for hydroxide hydrogen, yellow for oxalate oxygen and orange for oxalate carbon.

$\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system (0.75 nm at two H_2O per unit cell). The $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system seems to provide a better framework for the release of intercalated $\text{C}_2\text{O}_4^{2-}$ than the corresponding $\text{Mg}_4\text{Al}_2\text{-LDH}$ due to its higher water content and larger interlayer distance at the stable hydration state. However, the self-diffusion coefficient of $\text{C}_2\text{O}_4^{2-}$ ($D_{\text{C}_2\text{O}_4^{2-}}$) remains close to zero at the stable hydration state for the $\text{Mg}_6\text{Al}_2\text{-LDH}$ system, as shown in Figure 4.13. This indicates that the intercalated $\text{C}_2\text{O}_4^{2-}$ in the $\text{Mg}_6\text{Al}_2\text{-LDH}$ is as stable as that in the $\text{Mg}_4\text{Al}_2\text{-LDH}$. There is an obvious increase of the $D_{\text{C}_2\text{O}_4^{2-}}$ when the number of intercalated water molecules per unit cell is lower than the stable hydration state for the $\text{Mg}_6\text{Al}_2\text{-LDH}$ system (black line in Figure 4.13). Similar as the increase of $D_{\text{NO}_3^-}$ from three (stable state) to two H_2O molecules per unit cell (Figure 4.8a), the increase of $D_{\text{C}_2\text{O}_4^{2-}}$ can be attributed to the less impaired relocation due to the larger empty space generated in the interlayer after the removal of water.

Therefore, the higher water content of the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system at its stable hydration state will not accelerate the release of $\text{C}_2\text{O}_4^{2-}$ compared to the $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system because they both have low $D_{\text{C}_2\text{O}_4^{2-}}$ at the stable hydration states. The higher water content is most likely due to the larger surface area of the MOH layer compared to that in the $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system when there are the same number of Al atoms in the MOH layer present to balance the same amount of intercalated $\text{C}_2\text{O}_4^{2-}$ anions. The surface area ratio of the MOH layer between the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system and the $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system is around 4:3 based on the number of metal atoms (Al and Mg) in the MOH layers. The water content ratio at the stable hydration states for these two system is 2:1, which is slightly higher than

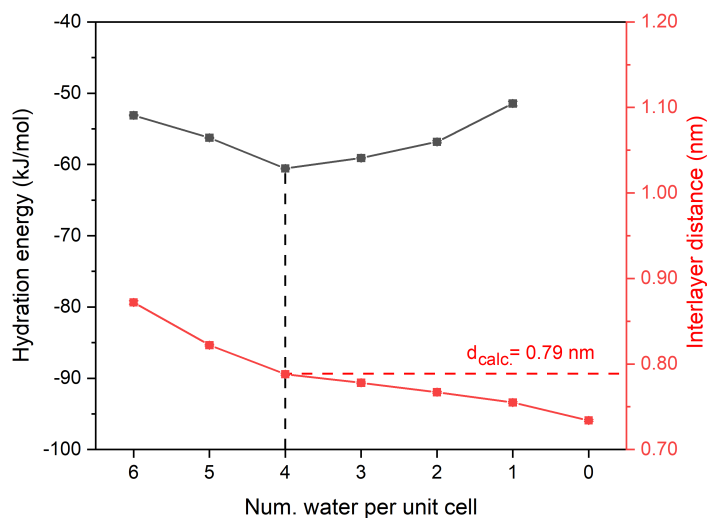


Figure 4.12: Hydration energy (black) and interlayer distance (red) as function of the water content for the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system. The dotted black lines are used as guide for the eye to highlight the most stable hydration state and the dotted red lines were included to mark the corresponding interlayer distance.

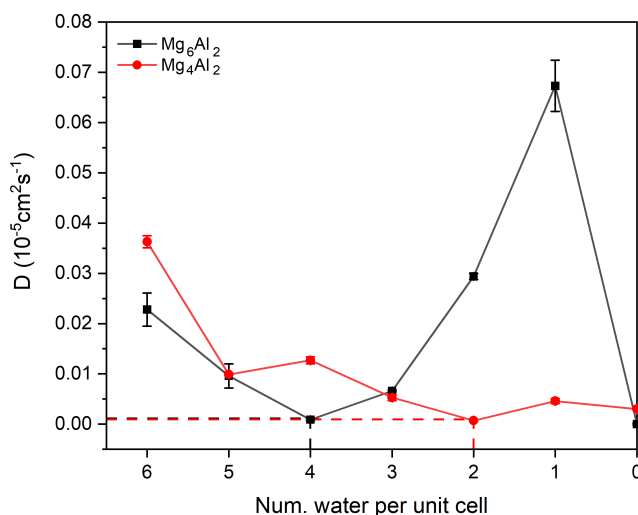


Figure 4.13: Self-diffusion coefficient of $\text{C}_2\text{O}_4^{2-}$ at different hydration states in x - and y -direction for the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ (black) and $\text{Mg}_4\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ (red) systems.

the theoretical surface area ratio (4:3). This can be attributed to the wave structure of the MOH layer in the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system as illustrated in Figure 4.14. Moreover, this wave

structure enlarged the contact area between the intercalated $\text{C}_2\text{O}_4^{2-}$ and MOH layer, resulting in a strong interaction and the low $D_{\text{C}_2\text{O}_4^{2-}}$, indicating a stable system.

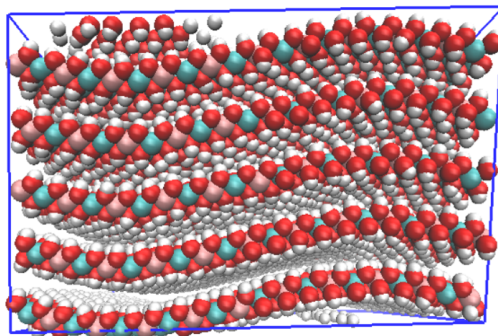


Figure 4.14: Illustration of the wave structure of the metal hydroxide layers for a Mg_6Al_2 -LDH- $\text{C}_2\text{O}_4^{2-}$ system. Pink depicts aluminium, cyan magnesium, red oxygen atoms of the hydroxide group, white hydrogen atoms of the hydroxide group.

4.4 Conclusion

In this work, the influence of the water content on the structure of LDH systems has been investigated for four intercalated anions (NO_3^- , Cl^- , CO_3^{2-} and $\text{C}_2\text{O}_4^{2-}$). The stable hydration state was identified for each host-guest system by employing a combination of density functional theory and molecular dynamic simulations. The favorable number of water molecules in the intergallery of the respective Mg_4Al_2 -LDH was determined as three for the NO_3^- , as two for the Cl^- and the $\text{C}_2\text{O}_4^{2-}$ systems, and as half for the CO_3^{2-} in each unit cell. It is noteworthy that the interlayer distances of LDH- NO_3^- , LDH- Cl^- and LDH- CO_3^{2-} at the stable hydration state are close to their respective experimental reference values. Hence, the theoretically determined value for the oxalate ion is a good estimate for the future experimental investigation.

The atom distribution, radial distribution, orientation, self-diffusion of the intercalated anions and the hydrogen-bonded network connected to the intercalated water molecules have been discussed to provide deeper insights into the interlayer region and to unravel the influence of different hydration states on the properties of the respective LDH systems. The simulation results showed that the behavior of the intercalated $\text{C}_2\text{O}_4^{2-}$ anion resembles intercalated CO_3^{2-} in the Mg_4Al_2 -LDH systems. The two systems exhibited similar properties like interlayer distances, anion orientations, and low self-diffusion coefficients of the intercalated anions at their stable hydration states. The theoretical investigation indicated that the Mg_4Al_2 -LDH- $\text{C}_2\text{O}_4^{2-}$ is a highly stable host-guest system analogous to Mg_4Al_2 -LDH- CO_3^{2-} . The Mg_6Al_2 -LDH- $\text{C}_2\text{O}_4^{2-}$ system was investigated to verify whether the lower charge density of the MOH layer can weaken the high stability of the intercalated $\text{C}_2\text{O}_4^{2-}$. The results showed that the intercalated $\text{C}_2\text{O}_4^{2-}$ is still highly stable in the Mg_6Al_2 -LDH- $\text{C}_2\text{O}_4^{2-}$ system, and the MOH layer is observed to be wavy rather than flat as was observed for the four Mg_4Al_2 -LDH systems.

The impact of intercalated water on LDH properties

In conclusion, in addition to the type of the intercalated anions, the amount of intercalated water also has an influence on the structure and stability of the LDHs by affecting the configuration of the intercalated molecules in the interlayer region and in turn the anion exchange capacity of LDHs. For the intercalated anions, a high self-diffusion coefficient can assist them to be able to be released from the LDH and thereby further protect Al/Mg alloys from corrosion. Therefore, the rational design of LDH-based systems used in corrosion protection, does not only require the selection of potent inhibitors but also needs to take the hydration state into account.

5 Machine-learned force field

To validate the wave-structure for Mg_6Al_2 -LDH intercalated with oxalate ($\text{C}_2\text{O}_4^{2-}$) reported in Section 4.3, a machine-learned force field (ML-FF) was developed in this work that can combine the accuracy of DFT methods with the efficiency of classical MD simulation. To construct ML-FFs, suitable datasets derived from DFT calculations must first be generated in order to factor in the relationship between the material structure and the forces acting on the atoms or/and the energy of the structure in the training process. [231,235] Unlike classical force fields, an ML-FF does not require the assumption of bonding patterns, instead it constructs important connections solely on the basis of the atomic coordinates of the structure. [231] ML-FFs have been applied to study electronic effects [263,375], thermodynamic properties [376], reactions [377–379], nuclear quantum effects [261,380], excited states [381] and spectroscopy [382] in materials science.

Machine learning models based on graph neural networks (GNNs) are among the fastest growing approaches and have been widely used in chemistry because of their ability to directly work on the chemical graphs of molecules and materials. [266,267] Wieder et al. [383] gathered 80 GNNs that were used to predict molecular properties using different datasets in their review article. Furthermore, GNNs have also been used to investigate crystal materials. Xie et al. [384] created a crystal graph convolutional neural network framework based on atomic connections in crystals that is directly capable of learning material properties. SchNet [380] enables high accuracy estimations of atomic interactions by combining continuous-filter convolutional (cfconv) layers with filter-generating networks [264,385] and incorporating periodic boundary conditions of crystal materials. [264] The cfconv layers are able to model the local correlation of a molecule with unevenly spaced atoms. [264] Figure 5.1 shows an overview of the SchNet architecture, which consists of multiple layers, including an embedding layer, three interaction layers and a pooling layer. At each layer, the material structure is represented atom-wise similar as pixels in an image. [264] The initial inputs for the architecture are atomic nuclear charges $Z = (Z_1, \dots, Z_n)$ and atomic positions $R = (\mathbf{r}_1, \dots, \mathbf{r}_n)$ of the material structure. Using the SchNet architecture, Schütt et al. [380] predicted the potential-energy surfaces and energy-conserving force fields of small organic molecules. Furthermore, Westermayr et

al. [381] used the SchNet architecture and its automatically determined representation of molecular structures to explore multiple electronic states.

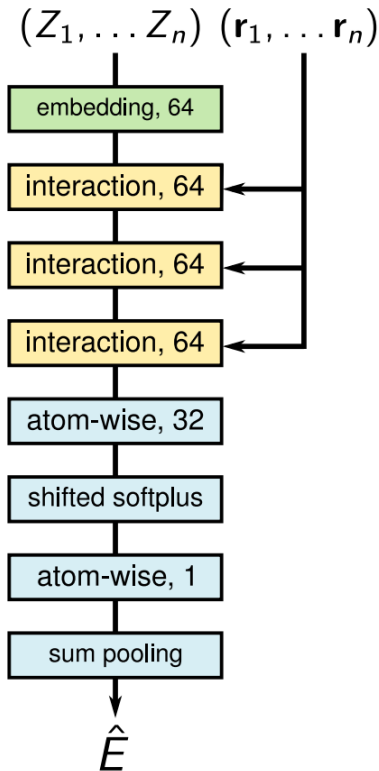


Figure 5.1: Illustration of SchNet architectural overview (Source: [264]).

5.1 Methodology

To benefit from these advantages, the SchNet architecture was used in this work to develop an ML-FF using the Keras graph convolution neural networks (KGCCNN) package developed by Reiser et al. [386]. KGCCNN is a TensorFlow-Keras-based neural network package that focuses on transparent tensor structures between layers. [386] Aside from the ability to train a SchNet-based ML-FF, KGCCNN also provides an interface for integrating it with the functionality of the atomic simulation environment (ASE) [387] package, such as molecular dynamics simulations (e.g. GROMACS package [354]). The workflow used in the construction of ML-FF is depicted in Figure 5.2. The input LDH dataset was generated using the plane-wave (PW) DFT code Vienna Ab Initio Simulation Package (VASP) [341, 342], where the process is similar to the aforementioned DFT calculations but without geometry optimization. Because the goal was to find a link between geometry and energy (and force), a diverse set of geometries was advantageous for increasing data diversity. DFT calculations were performed for three different size LDH systems to increase the variance in the training data. These three sizes are defined as small, which is the same as one unit cell, medium, which is a supercell twice

the size of the unit cell, and large, which is four times the size of the unit cell. The large, medium and small size LDH systems are shown from left to right in Figure 5.2. Following that, a SchNet-based ML-FF was built within KGCNN [386]. The constructed ML-FF can be used to run MD simulations using the interface functionality of ASE [387]. The MD simulation trajectories were subsequently used as input file for VASP [339] to validate the MD simulation results and thus the quality of the used ML-FF. The DFT calculations for the validations will be added to the initial LDH dataset to retrain the model and to increase its domain of applicability, particularly if structural artifacts are detected in the ML-FF MD simulations, resulting in an active learning approach.

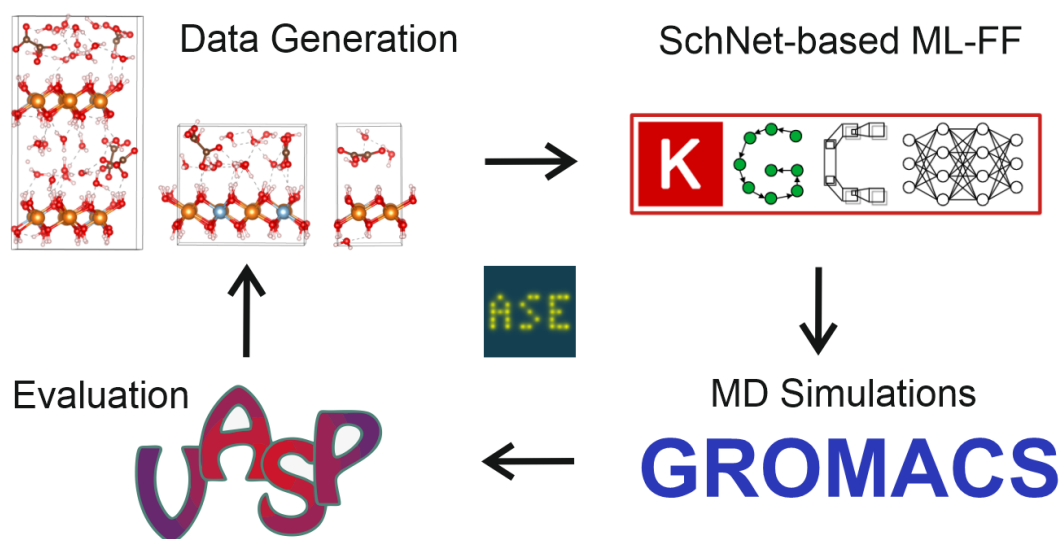


Figure 5.2: The workflow used in the construction of ML-FF is depicted schematically. First, a dataset of LDHs is generated by DFT calculations (top left). A SchNet-based ML-FF is built in the KGCNN package using the LDH dataset. The SchNet-based ML-FF can then be used to run MD simulations, with the corresponding evaluation done using DFT calculations in VASP. Finally, the DFT calculations from the evaluation stage can be added to the initial LDH dataset to retrain the model. The ASE software package is used for setting up and linking the atomic simulations described in the feedback loop.

Besides changing the LDH size, two other strategies were used to increase data diversity: manually changing the relative positions of the oxalate and metal hydroxide (MOH) layers, and running MD simulations at high temperatures using the classical CLAYFF. A total of 1370 different structures were created based on the small size LDH, and the difference between the structures was obtained by changing the relative position of the oxalate and MOH layers, followed by filling water molecules randomly in the empty space. As a result, the water content in these structures may slightly differ. In addition, 1000 LDH structures based on the medium and large size LDHs (500 each) obtained by running classical MD simulations at 1150 K and 2150 K, respectively were added to the pool of available training data. Hence, the initial dataset for the ML-FF construction consisted of 2370 samples.

A molecular structure can be represented as a graph, with nodes representing atoms and edges representing covalent bonds. [273] However, unlike molecules, LDH crystal systems lack comprehensive covalent bonding information that can be directly used as edges in their graphs, so edge selection in a crystal system necessitates more care. In this study, the edges of the LDHs were chosen using a combination of the k-nearest neighbors [388] and radius-based [389] algorithms. To calculate the edges of each atom in the LDHs, the intersection of its 25 nearest neighbors and neighboring atoms within 5 Å was chosen with the consideration of image graphs (periodic boundary conditions of crystals as illustrated in Figure 5.3). Figure 5.4 depicts the LDH SchNet architecture, where the LDH structure is represented as a graph to be used as input for model construction. The atomic number of each atom (Z_i) and the distance between two atoms for the selected edges (L_i) were used to represent node and edge information, respectively. In the embedding stage, nodes and edges were embedded with latent vectors to obtain the initial node and edge embedding. During the interaction stage, the edge embedding was initially updated on the basis of the initial edge embedding (\mathbf{X}_E) and the initial node embedding (\mathbf{X}_V). Afterwards, the local edge aggregation function collected the updated edge embedding from the previous step for all incoming edges at each node. Based on the aggregated result for each node and the initial node embedding, the node embedding was updated which completed a round of communication between the edge and the node. A model can have multiple rounds of communication between edges and nodes, with the SchNet architecture (Figure 5.1) displaying three rounds of communication. The final updated node embedding was used for energy prediction and the force was obtained by differentiating the predicted energy with respect to the atomic positions. The following hyperparameters were used in the SchNet model:

- **Node embedding dimensions:** 128
- **Edge embedding:** 25-dimensional Gaussian basis expansion [390] on a [0,5] Å interval
- **Depth:** 6
- **Interaction layer dimensions:** 128
- **Node pooling:** Sum
- **Epochs:** 1000
- **Optimizer:** Adam [391]

Finally, the obtained model (SchNet-based ML-FF) can be used as a force field in the MD simulations.

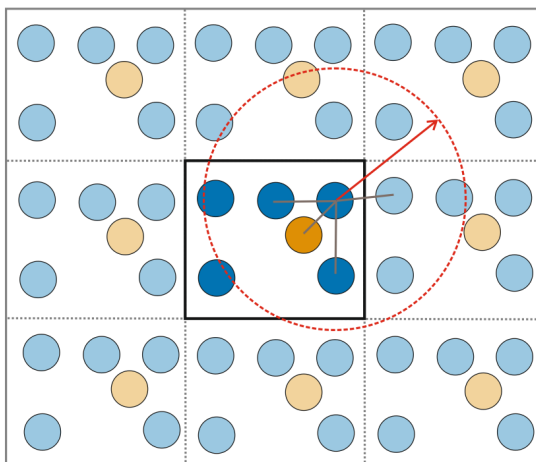


Figure 5.3: Edge generation in two dimensions using the k-nearest neighbors algorithm (four nearest neighbors around the selected centroid, edges marked in dark brown lines) and the radius-based algorithm (the radius is represented by the red arrow).

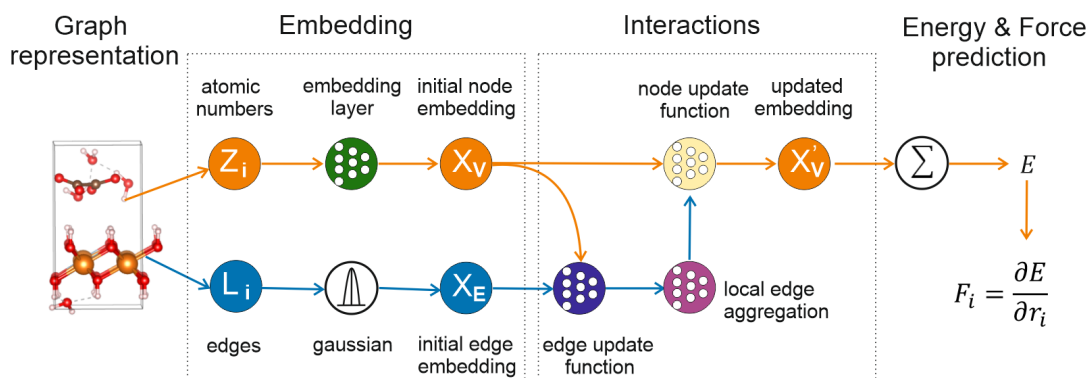


Figure 5.4: The SchNet architecture for the LDH system is depicted. To construct the model, each LDH structure is represented as a graph. The atomic number of each atom represents the node information, while the distance between two atoms represents the edge information. To obtain the initial node and edge embeddings (X_V and X_E), nodes and edges are embedded with latent vectors during the embedding stage. During the interaction stage, the node embedding is iteratively updated. For energy prediction, the final updated node embedding is used, whereas force prediction is obtained by differentiating the predicted energy with respect to the atomic positions.

In order to develop the initial ML-FF, the available training data was divided into five folds and a cross validation approach was adopted to validate the accuracy and robustness of the trained model. After the initial ML-FF was generated, it was used to run MD simulations using the velocity Verlet algorithm [392] via the KGCNN-ASE interface. The Maxwell-Boltzmann distribution [393] was used to set the momentum for the atoms of LDHs at room temperature

(300 K). The time step was set to 1 fs and the total number of time steps was 500. The 500 time step trajectories were saved for visualization and distance analysis to further investigate the ML-FF behavior. Only the second half of the total trajectory was sampled to evaluate the ML-FF due to its higher diversity, and 50 structures were extracted from the second half of the MD simulation, with each 5th trajectory extracted beginning with the 250th time step. On these sampled LDH structures, DFT calculations were performed, and the comparison of DFT calculated values and ML-FF predicted values can be used to evaluate the behavior of the ML-FF. Finally, the initial dataset was updated by incorporating the sampled structures with DFT calculated results to retrain the ML-FF model. The prediction accuracy of the model can be increased using an active learning strategy by performing iterative updates to the dataset and the associated ML-FF model which is discussed in detail in Section 5.3 and 5.4. The retrained ML-FF was subsequently used to perform MD simulations of an expanded LDH system created by replicating the large size LDH in the initial dataset $2 \times 2 \times 1$ times. Using the Berendsen thermostat [394] and the same time step (1 fs) for 5000 time steps, the canonical ensemble (NVT) was used to reach the equilibrium state for the expanded LDH system.

5.2 Performance evaluation of the initial ML-FF model

From the learning curves obtained from the model training, displayed in Figure 5.5, it is apparent that the energy and force mean absolute errors (MAEs) decrease for the training dataset in the cross validations after 100 epochs. As the number of epochs increases, the energy MAEs of the training and validation sets continue to decrease until they reach a plateau, while the force MAEs remain at lower values. The model obtained at the end of training (at the 1000th epoch) in the 5th fold of the cross validations was saved as the SchNet-based ML-FF model for the subsequent MD simulations.

The predicted values in the validation set are compared to the DFT calculated results for both energies and forces in Figure 5.6 to assess the behavior of the saved model more directly. Since energy is an extensive property, the scaled energy of the three considered LDH sizes was determined and results in an MAE of 0.033 (Figure 5.6a). The comparison between the predicted energy and DFT calculated energy for each size of LDHs is shown in Figure 5.7, and the MAEs are 0.391 eV, 0.191 eV, and 0.058 eV for the large, medium, and small size of LDHs, respectively. Naturally, the large LDH has a higher MAE value, as the total energy of a system scales with the number of atoms. The force MAE of the LDH structures in the validation set is 0.100 eV/Å, as shown in Figure 5.6b. This value is in the same order as the force MAE of some organic molecules in the validation set from the work of Schütt et al. [264]. The saved ML-FF is thus able to accurately predict energies and forces for the structures in the validation set and was subsequently tested for its applicability in MD simulations to estimate its robustness.

5.2 Performance evaluation of the initial ML-FF model

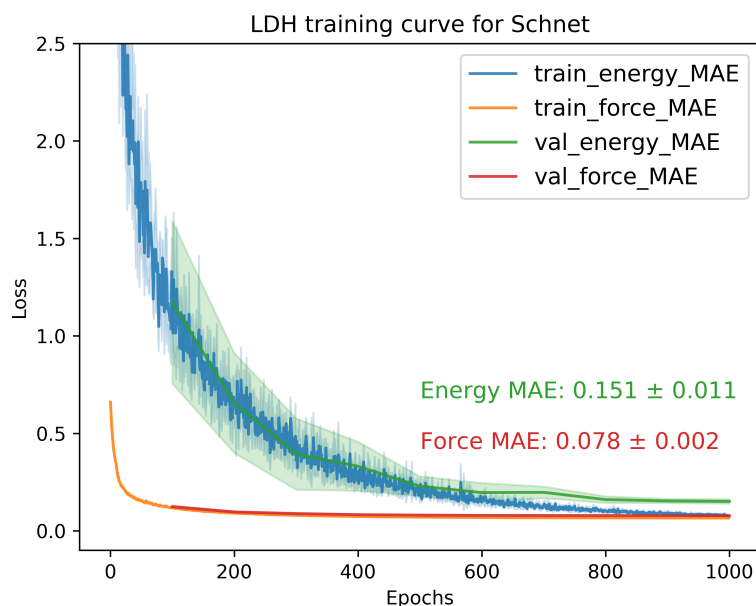


Figure 5.5: LDH training curve with respect to epoch variation. The blue and orange lines represent the energy and force MAEs for the training dataset. The green and red lines depict the energy and force MAEs for the validation dataset. The corresponding standard deviation interval of the 5-fold cross validations is marked in a slightly lighter color for each error. The energy and force MAEs of the validation dataset obtained from the last epoch are displayed in colors that correspond to the legend.

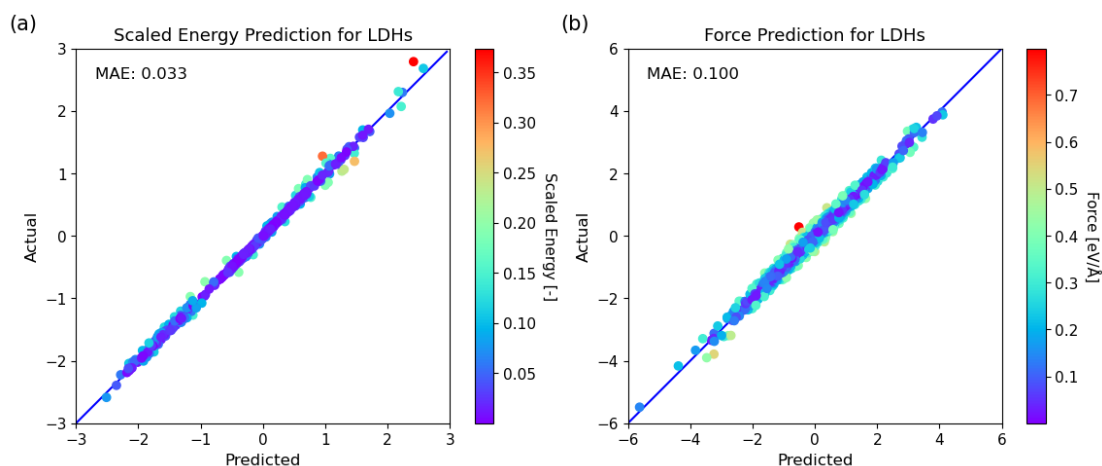


Figure 5.6: Comparison of ML-FF model predicted values and actual values obtained from DFT calculations for (a) scaled energy and (b) force of the LDH structures in the validation date set. The color bar depicts the difference between the predicted and actual values. The blue lines represent perfect correlation.

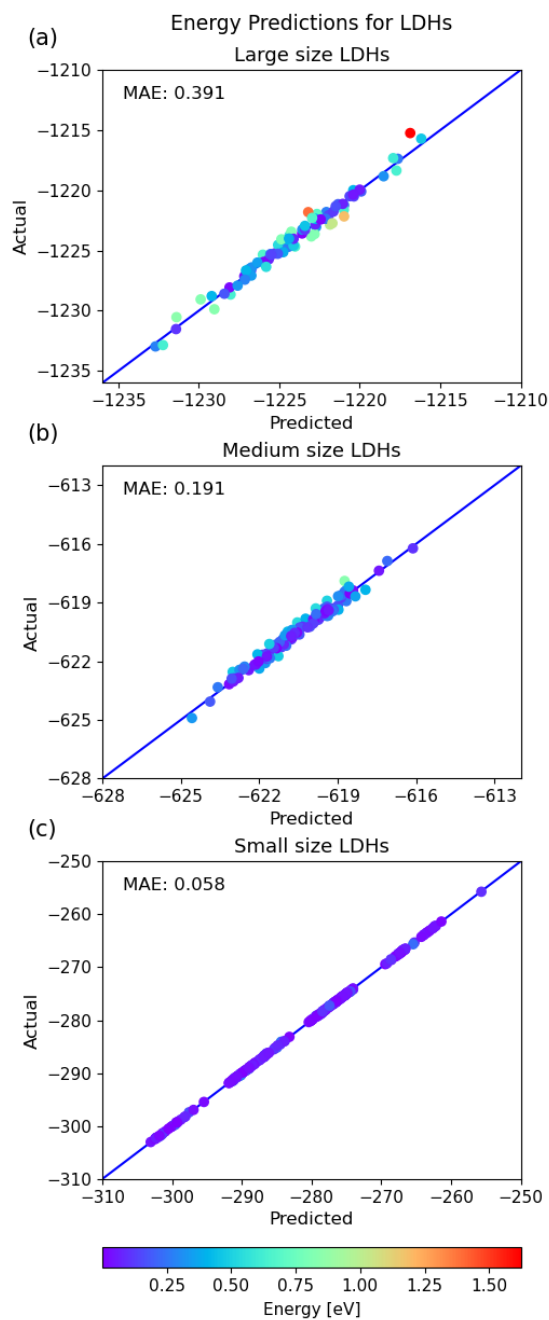


Figure 5.7: Comparison of ML-FF model predicted values and actual values obtained from DFT calculations for the energy of three different sizes of LDHs. The corresponding sizes of the LDHs are **(a)** large, **(b)** medium and **(c)** small. The color bar depicts the difference between the predicted and actual values. The blue lines represent perfect correlation.

5.2 Performance evaluation of the initial ML-FF model

The unit cell LDH was used as the starting geometry for the MD simulation, with the force field being the saved ML-FF model. The potential energy obtained from the MD simulations (referred to as ML-FF energy) shown in Figure 5.8a increases with the simulation time, with a noticeable increase around 0.4 ps. This increase can be attributed to three observed factors: the stretching of the carbon-carbon (referred to as C-C) atom distance within the oxalate molecule (shown in orange dots in Figure 5.8a), and the stretching of the oxygen-hydrogen (referred to as O-H) atom distances in water molecules and MOH layers (shown in green and purple dots in Figure 5.8a, respectively). The O-H distances in water were averaged over all water molecules in the system, and the corresponding standard deviation interval is denoted by the green area in Figure 5.8a. The standard deviation interval of O-H distance in MOH layers is shown in Figure A.4 in Appendix A.2 after performing the same averaged calculation. The stretching of the averaged O-H distances in water molecules and MOH layers is not as pronounced as the increase in C-C distance, but each averaged distance has a large standard deviation as shown in Figure 5.8a and Figure A.4. The large standard deviation is due to the fact that only a few hydroxide ions in water molecules and MOH layers exhibited remarkable stretching, for example, as shown Figure 5.8b, the O-H distance of the prominent hydroxide ions in the inset image stretched significantly at the end of the simulation. The stretching of the C-C distance occurs prior to the sharp increase in ML-FF energy, whereas the stretching of the O-H distances in both water molecules and MOH layers occurs almost simultaneously with the sharp increase in ML-FF energy, as shown in Figure 5.8a. As a result, the sharp increase in ML-FF energy is caused primarily by O-H stretching in both water molecules and MOH layers.

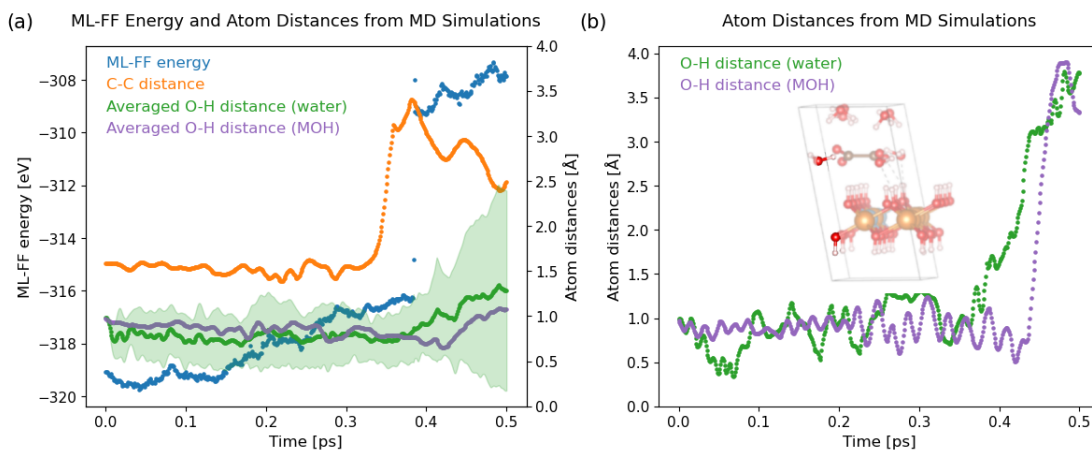


Figure 5.8: (a) ML-FF energy (in blue), C-C atom distances in oxalate (in orange), and averaged O-H atom distances in water molecules (in green) and MOH layers (in purple) varied with simulation time for the unit cell LDH. (b) Two highlighted O-H distances marked in the inlet image varied with simulation time.

Figure 5.9 depicts a comparison of the ML-FF model predicted and DFT calculated energies and forces for the 50 extracted structures from the second half of the MD simulation. The ML-FF model predicted much lower energies and forces than the DFT calculated values, resulting

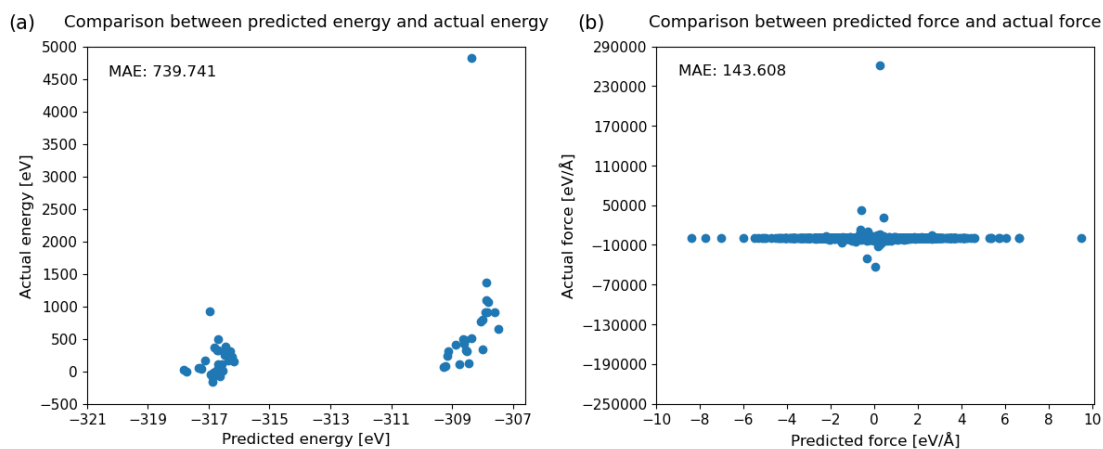


Figure 5.9: Comparison of (a) predicted energy from the initially saved ML-FF model and the actual energy obtained from DFT calculation and (b) predicted force from the initially saved ML-FF model and actual force obtained from DFT calculation for the unit cell LDH.

in high MAE values (739.741 eV for the energy and 143.608 eV/Å for the force). The high MAE values indicate that the ML-FF model used in the MD simulation underestimated the effect of the three observed atom distance stretches on the energy and force of the simulated system. This can be attributed to a lack of stretched compounds inside LDH structures in the initial dataset, causing the model to incorrectly predict their impact on the prediction of energy and force of the system. The active learning approach will be used in the following section to solve this problem.

5.3 Active learning with unit cell LDH

Following the performance evaluation of the initially saved model, these 50 extracted structures were added to the initial dataset, along with their DFT calculated energies and forces, and the ML-FF model was retrained. The first iteration is defined as validating the behavior of the one time retrained ML-FF model in MD by comparing it to DFT results. After implementing nine iterations of the described active learning procedure, the model accuracy is significantly improved. Figure 5.10 depicts the behavior change of the retrained model used in each iteration, including ML-FF energy, C-C distances and O-H distances. As illustrated in Figure 5.10a, there is no significant increase in ML-FF energy over the nine iterations, which indicates that the retrained models performed better. However, the behavior of the retrained model in the first iteration is significantly different compared to the other eight iterations, with lower ML-FF energies and shorter O-H atom distances in both water molecules and MOH layers, while the C-C atom distances remain unexpectedly higher, as shown in Figure 5.10. The lower ML-FF energy and shorter O-H atom distance indicate that the retrained model learned from the added stretched compounds that O-H stretch can result in a high energy. Thus, the model avoided O-H stretching as much as possible in the first iteration, resulting in

a compressed O-H atom distance, and the model predicted that the compressed structure had a low energy.

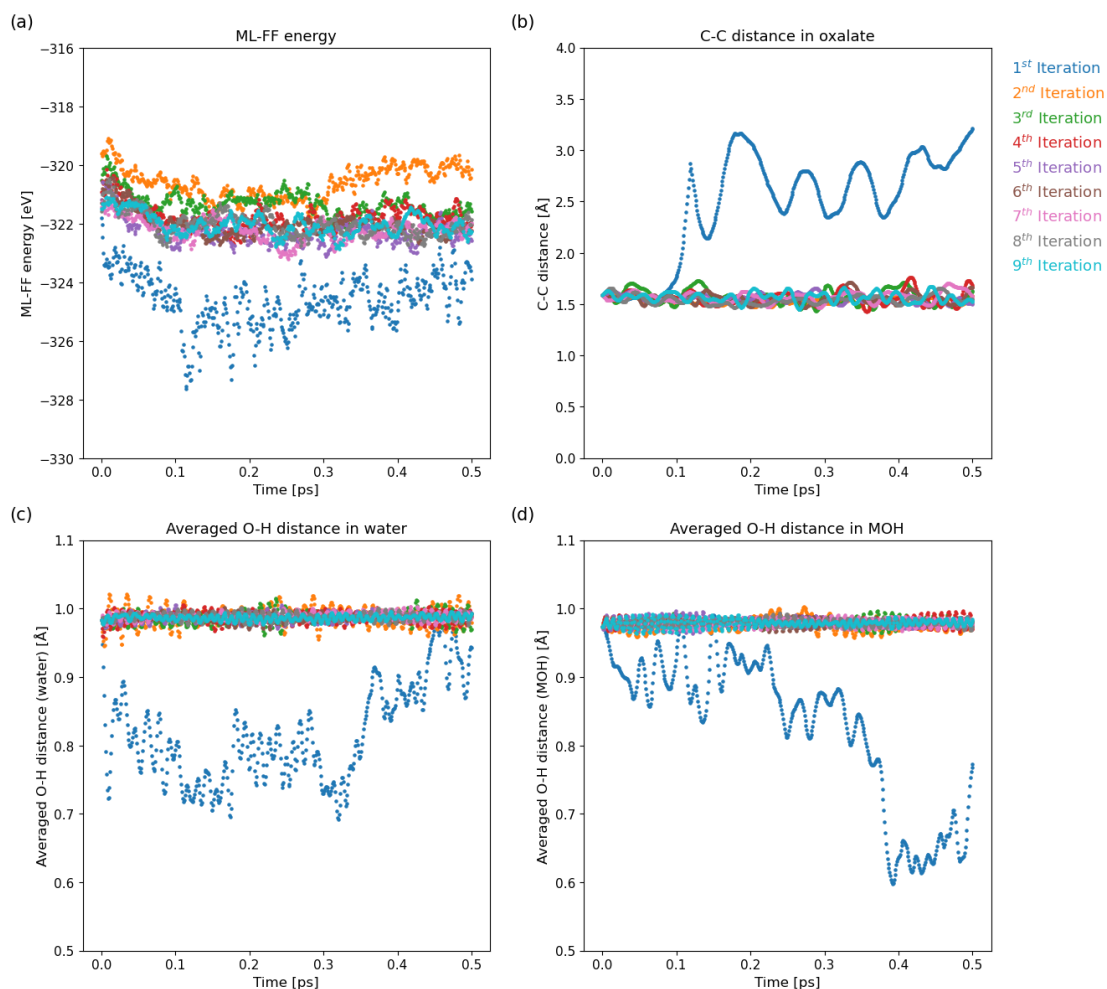


Figure 5.10: (a) ML-FF energy, (b) C-C atom distances in oxalate, (c) averaged O-H atom distance in water molecules and (d) averaged O-H atom distance in MOH layers varied with simulation time from the 1st to the 9th iteration for the unit cell LDH. From the 1st to the 9th iteration, the results are shown in the following color order: blue, orange, green, red, purple, brown, pink, gray and cyan.

The DFT validation of the first iteration, shown in blue in Figure 5.11a and b, reveals that the retrained model underestimated the effect of the overly compressed structures on the energy and force predictions. This limitation was overcome in the second iteration after retraining the ML-FF with an updated training dataset containing compressed structures with high energies and large forces. Furthermore, no C-C distance stretching was observed in the second iteration (Figure 5.10b). From the third to the ninth iteration, the previously observed stretching or compression of the O-H atom distance vanished, and the atom distances of C-C and O-H remain around the values corresponding to their bond lengths of 1.6 Å for a C-C bond in the

oxalate compound [395] and 1 Å for an O-H bond in water [357] and the MOH layer [185] (as seen in Figure 5.10**b**, **c** and **d**). For the energy and force, the predictions improved significantly through the first and the second iteration, where the MAE of the energy decreased about 15% (from 739.741 eV to 625.851 eV) and 99% (from over 600 eV to a value less than 1 eV), respectively. The MAE of the force decreased about 47% (from 143.608 eV/Å to 76.281 eV/Å) and 99% (from around 76 eV/Å to less than 0.5 eV/Å), respectively. From the third to the ninth iteration, the retrained model improved slightly for the energy and force predictions at first until the predicted values correlated well with the actual values obtained from DFT after the fourth iteration (Figure 5.11**c** and **d**). Figure 5.12 shows a clear trend in the error minimization concerning the energy and force MAEs as the number of iterations increased from two to four. From the sixth to the ninth iteration, the force MAEs stabilized at a value below 0.09 eV/Å, whereas the energy MAEs fluctuated.

The energy and force MAEs obtained from the sixth to ninth iterations as well as their averaged values are listed in Table 5.1 and compared to the corresponding MAE values of the initial model. Counterintuitively, neither iteration resulted in a lower energy MAE value compared to the initial model. This can be partially attributed to the fact that the starting geometry unit cell LDH contains approximately three additional water molecules compared to the unit cell LDH structures used in the initial dataset. Furthermore, it indicates that more iterations may be required to further improve the energy prediction of the model. All of the force MAEs, on the other hand, are lower than the force MAE value of the initial model. This suggests that the retrained models used in these four iterations outperform the initial model in terms of force predictions.

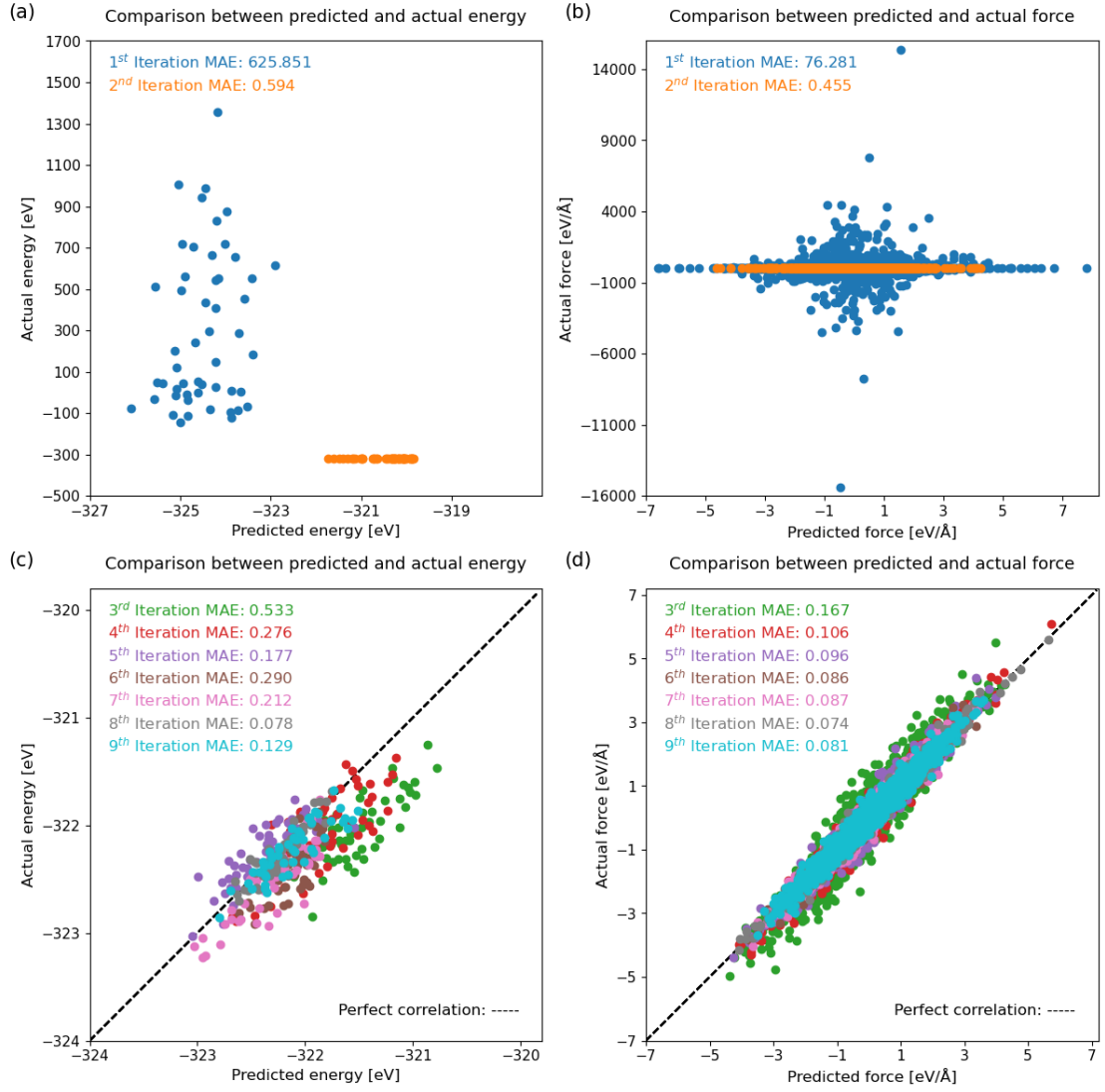


Figure 5.11: Comparison of predicted energy from the retrained ML-FFs and actual energy obtained from DFT calculations (**a** and **c**) and predicted force from the retrained ML-FFs and actual force obtained from DFT calculations (**b** and **d**) for the unit cell LDH. (**a**) and (**b**) show the results of the 1st and the 2nd iteration. (**c**) and (**d**) show the results of the 3rd to the 9th iterations. From the 1st to the 9th iteration, the results are shown in the following color order: blue, orange, green, red, purple, brown, pink, gray and cyan. The corresponding MAEs are colored in the same way. The black dashed lines represent the perfect correlation, with no difference between predicted and actual values.

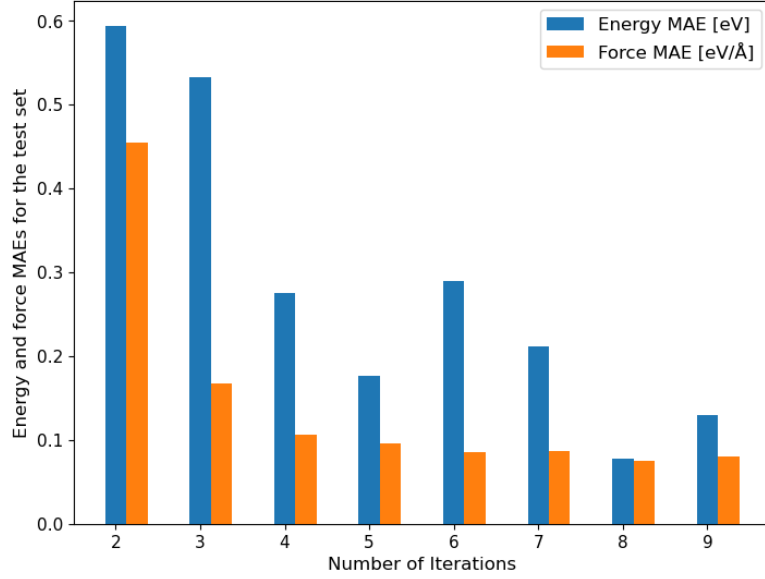


Figure 5.12: The energy and force MAE values varied with the number of iterations for the unit cell LDH.

Table 5.1: Comparison of the energy and force MAEs between the retrained models with the unit cell LDH and the initial model. The results from the 6th to the 9th iteration and the average values are listed and compared to the values of the initial model.

Num. Iteration	Energy MAE [eV]	Force MAE [eV/Å]
6	0.290	0.086
7	0.212	0.087
8	0.078	0.074
9	0.129	0.081
Averaged	0.177	0.082
Initial model	0.058	0.100

5.4 Active learning with supercell LDHs

Medium and large size supercells were used in the active learning process to further improve the behavior of the ML-FF extracted from the fifth iteration of Section 5.3. The interaction of the oxalate molecules in supercells was taken into account in the active learning process when compared to the retraining process using unit cell LDH with only one oxalate per structure. Figure 5.13 shows the results of a comprehensive investigation into the variation of the C-C atom distances over nine iterations for the medium size LDH. The C-C atom distances of the two oxalate molecules (Figure 5.13a C1-C2 and Figure 5.13b C3-C4) are close to the bond length mentioned in the previous paragraphs for the majority of the iteration process. In the third iteration, a sharp increase in the C-C atom distance for C3-C4 distance appeared at the early stage of the simulation and followed by a large variation in the remaining simulation time as shown in Figure 5.13b. The sharp increase of the C1-C2 distance in Figure 5.13a occurred at the end of the simulation time. Aside from the C-C atom distance within an oxalate molecule, the distance between two neighboring carbon atoms (C2 and C3) from two distinct oxalates was calculated and was found to strongly decrease as the C1-C2 and C3-C4 distances increase, as shown in Figure 5.13c. This observation implies that the model used in the third iteration did not recognize the two oxalate compounds as separated units. Instead, it tried to bring these two carbon atoms closer together to form a new oxalate molecule.

In the third iteration, the ML-FF model underestimated the influence on both the energy and force predictions of two neighboring carbon atoms belonging to two distinct oxalates moving closer to each other. The energy predicted by this ML-FF model is significantly lower than the actual value calculated using DFT, as shown in green in Figure 5.14a. The predicted forces moreover do not correlate well with the forces calculated by DFT (Figure 5.14b). Furthermore, Figure 5.14a and b show that for some of the structures, the ML-FF used in the fifth iteration provides a lower energy prediction and a larger force prediction compared to their respective DFT values, while the C1-C2 and C3-C4 distances are close to the expected bond length (Figure 5.13a and b). The highest predicted force, denoted by a red circle in Figure 5.14b, is caused by a hydrogen atom of a water molecule. Therefore, the higher MAE values in the fifth iteration compared to the iterations in Figure 5.14c and d could be attributed to the changes in the O-H distances in the water molecules. The averaged O-H distance in water molecules was calculated and, as shown in Figure 5.15, only a small variation was observed towards the end of the simulation with a large standard deviation interval. To eliminate the averaging effect in LDH systems, the O-H distance associated with the largest predicted force value was investigated. The distance of this O-H remains within the confidence interval of the averaged O-H distance for the first 0.4 ps, followed by a sharp increase at the end of the simulation time. It should be noted however that the sharp increase mentioned here is significantly smaller than the O-H stretch observed for the initially saved ML-FF model (the green plot in Figure 5.8b). The averaged O-H atom distances in the water molecules and MOH layers for the remaining eight iterations are provided in Figure A.5 in Appendix A.2. From the sixth iteration on, the energy and force MAEs stabilized at values less than 0.19 eV and 0.07 eV/Å, respectively.

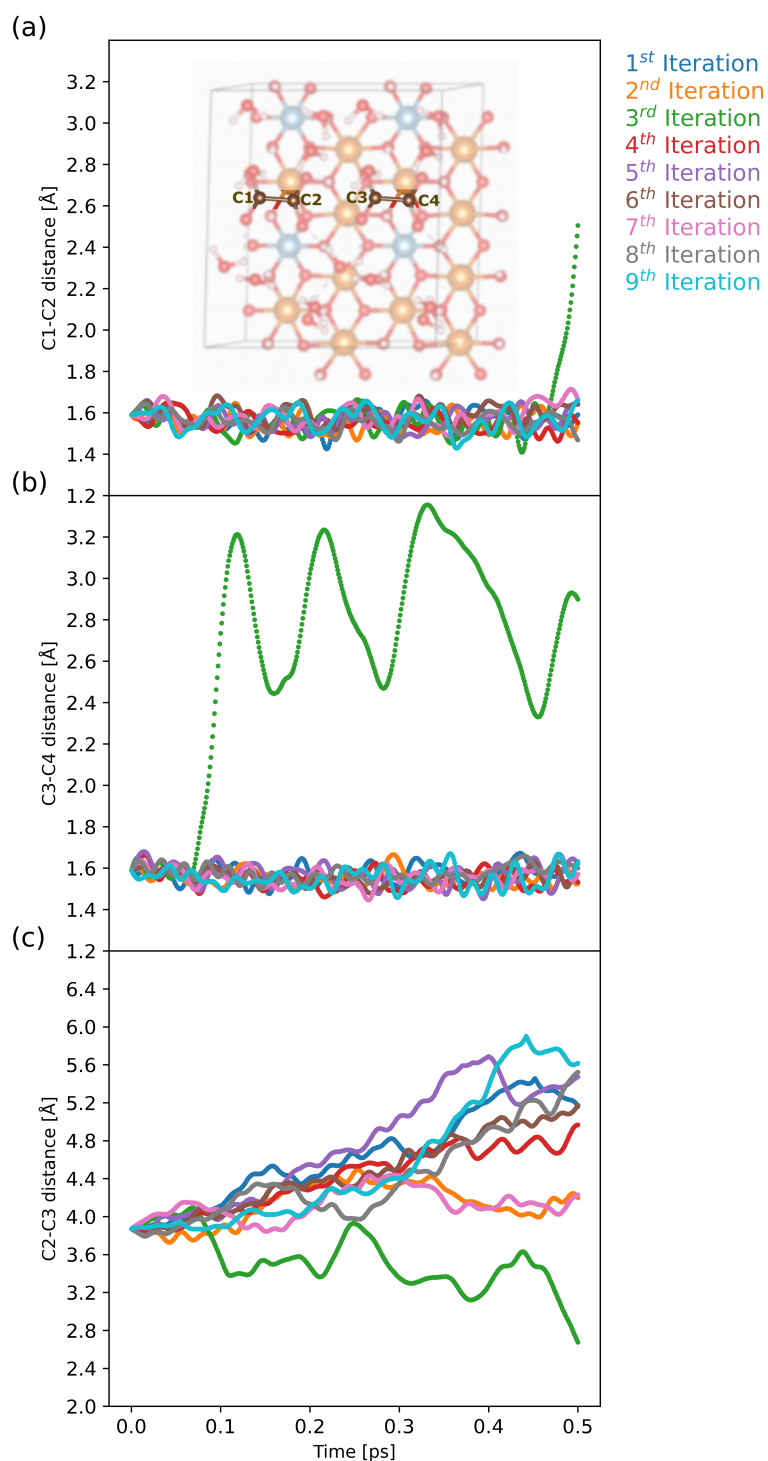


Figure 5.13: Three C-C atom distances in the two oxalates varied with simulation time in different iterations for the medium size LDH. As shown in the inserted image, the four carbon atoms in the two oxalates are designated as C1, C2, C3 and C4. (a) C1-C2, (b) C3-C4 and (c) C2-C3 distances varied with simulation time from the 1st to the 9th iteration. From the 1st to the 9th iteration, the results are shown in the following color order: blue, orange, green, red, purple, brown, pink, gray and cyan.

Meanwhile, the C-C atom distances of the two oxalates and the averaged O-H atom distances in the water molecules and MOH layers are close to their bond lengths.

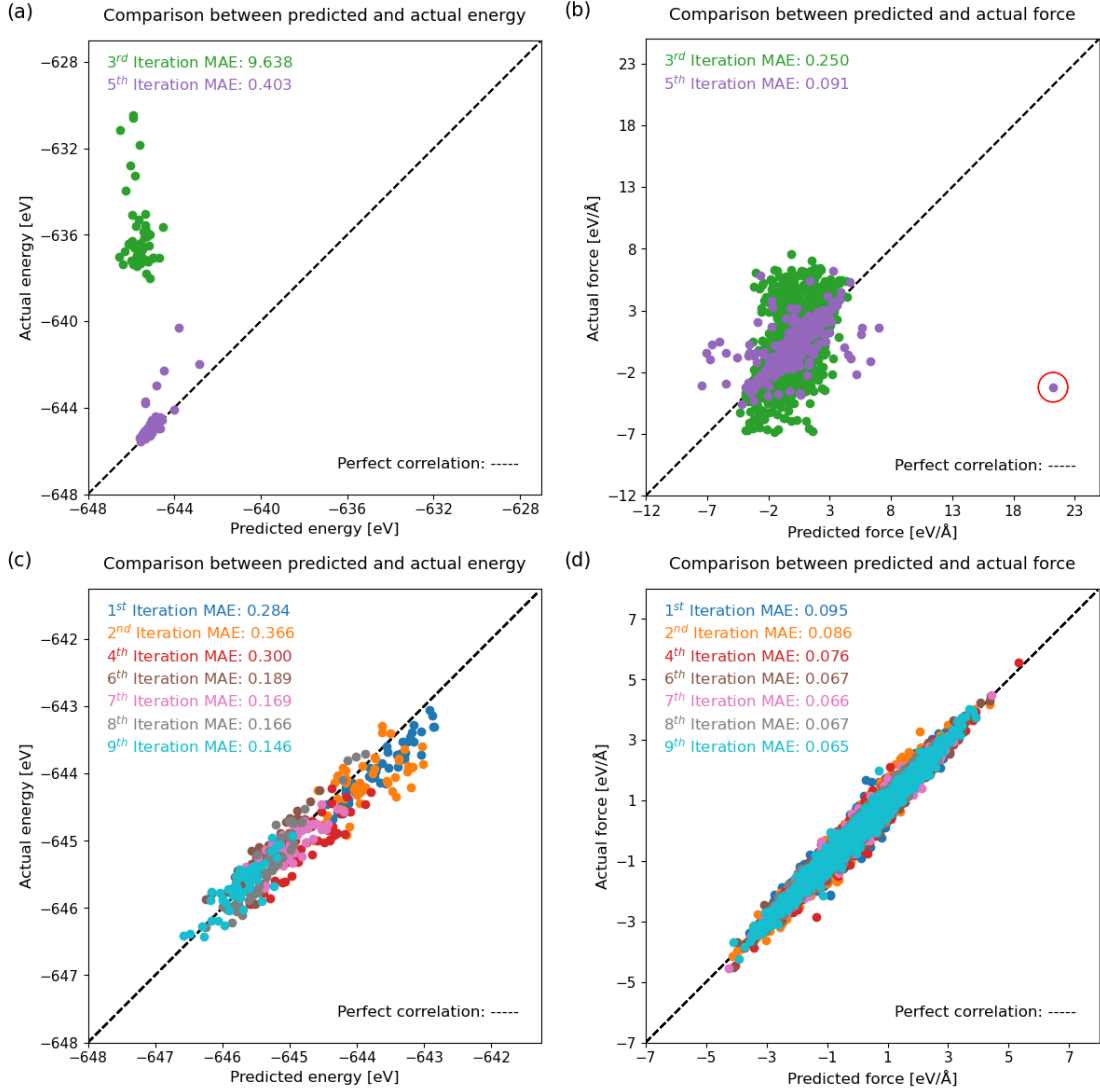


Figure 5.14: Comparison of predicted energy from the retrained ML-FFs and actual energy obtained from DFT calculations (**a** and **c**) and predicted force from the retrained ML-FFs and actual force obtained from DFT calculations (**b** and **d**) using the medium size LDH. (**a**) and (**b**) show the results of the 3rd and 5th iterations. (**c**) and (**d**) show the results of the remaining iterations. From the 1st to the 9th iteration, the results are shown in the following color order: blue, orange, green, red, purple, brown, pink, gray and cyan. The corresponding MAEs are colored respectively. The purple point with the largest predicted force is highlighted by a red circle. The black dashed lines represent perfect correlation between predicted and actual values.

In addition to active learning with medium size LDH, the model was retrained with large size

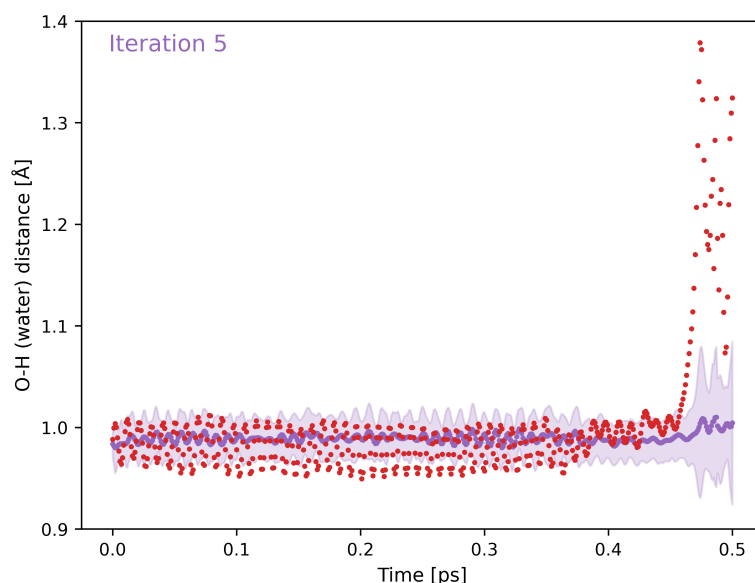


Figure 5.15: Averaged O-H atom distance in water molecules varied with simulation time for the medium size LDH in the 5th iteration, marked in purple. The standard deviation interval is in a slightly lighter purple. The O-H atom distance corresponding to the hydrogen atom showing the largest predictive force is plotted in red to show its variation with time.

LDH to investigate potential further improvement of the model performance. No stretching of the C-C atom distance within the same oxalate is present in the large size LDH, nor does the approaching of two carbon atoms from separate oxalates appear. This can be attributed to the fact that each LDH structure contains four oxalates, which can provide enough information for the model to learn that breaking an oxalate molecule into two C_2O_2^- fragments and pulling two carbon atoms from two different oxalates closer together is not energetically favorable. Figure 5.16 shows the calculated average C-C atom distance in oxalates and O-H atom distances in water molecules and MOH layers. Over the nine iterations, the averaged distances fluctuated around the expected chemical bond lengths.

Figure 5.17 shows that in the first iteration, there was already a decent correlation between the predicted and DFT calculated energy and force values, and there was a slight improvement from the first to the second iteration, after which the model behavior was maintained. This indicates that if the initial geometry is of large size LDH, only two iterations are required to improve the model behavior whereas at least six iterations are required to achieve a similar model behavior using a medium size LDH. The energy and force MAEs of the retrained models obtained from the sixth to the ninth iteration using medium and large supercells as initial geometries are listed in Table 5.2. For all four iterations of both sizes LDHs, their energy and force MAEs are lower than the values obtained from the initially saved model, indicating a significant improvement in the ML-FF through active learning.

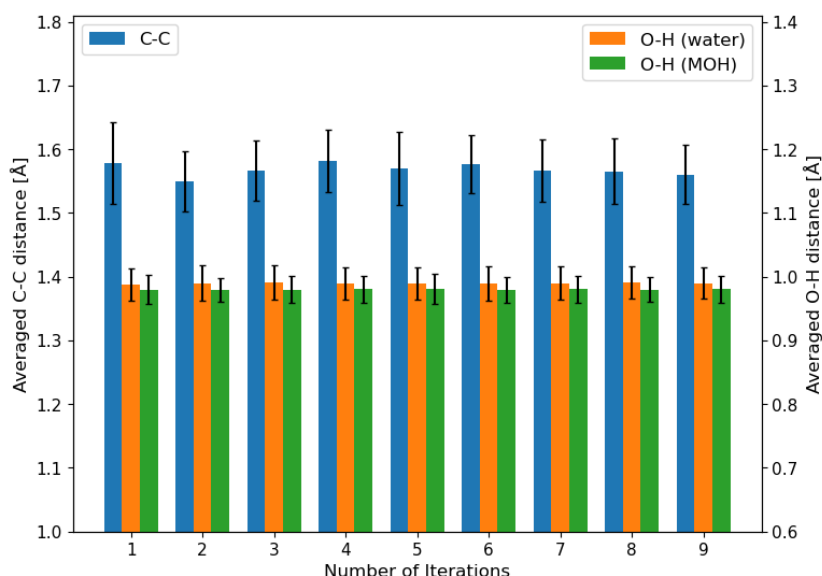


Figure 5.16: Averaged C-C atom distance in oxalates, averaged O-H atom distance in water molecules and averaged O-H atom distance in MOH layers varied with the number of iterations ranging from 1 to 9 for the large size LDH.

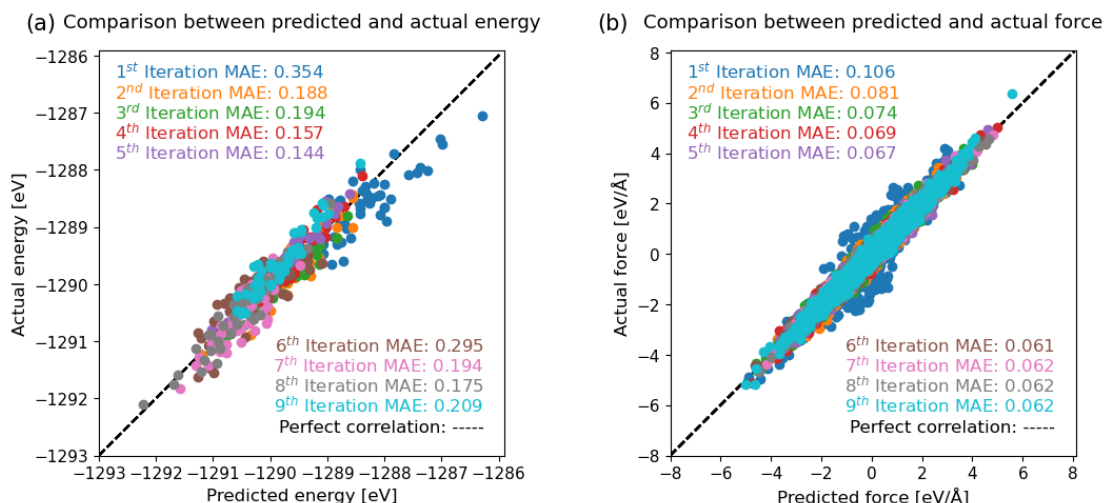


Figure 5.17: Comparison of (a) predicted energy from the retrained ML-FFs and actual energy obtained from DFT calculations and (b) predicted force from the retrained ML-FFs and actual force obtained from DFT calculations from the 1st to the 9th iteration using the large size LDH. The corresponding MAEs are colored in the same way. The black dashed lines represent the perfect correlation, with no difference between predicted and actual values.

Table 5.2: The comparison of the energy and force MAEs between the retrained models based on two supercells (the medium and large size LDHs) and the initial model. The results from the 6th to the 9th iterations and the corresponding average values are listed and compared to the values of the initial model.

Num. Iteration	Energy MAE [eV]		Force MAE [eV/Å]	
	Medium	Large	Medium	Large
6	0.189	0.295	0.067	0.061
7	0.169	0.194	0.066	0.062
8	0.166	0.175	0.067	0.062
9	0.146	0.200	0.065	0.062
Averaged	0.168	0.216	0.066	0.062
Initial model	0.191	0.391	0.100	

5.5 Testing the retrained model on an expanded LDH system

Finally, the retrained model obtained from the ninth iteration in the active learning process with the large size supercell LDH was used to run a longer MD simulation to reach the equilibrium state, using an expanded LDH system (four times the large size supercell LDH) as the starting geometry. The results of the energy and atom distances for the expanded LDH are shown in Figure 5.18. The ML-FF energy decreased at the start of the simulation and then fluctuated around the low energy, as shown in Figure 5.18a. There was no significant stretch or compression of the C-C and O-H atom distances in the expanded LDH system, and the distances fluctuated around their bond lengths, as shown in Figure 5.18b, c, d. This implies that the model retrained with smaller size LDHs can be scaled up to simulate expanded LDH systems.

Although no wave-like structure was observed in the MD simulation using the retrained ML-FF for this expanded LDH, it cannot yet be concluded that the wave-like structure observed in the classical MD simulation is physically incorrect due to the significant size and time scale differences. In the future investigations, it is recommended to use the retrained ML-FF to run MD simulations using the same size LDH as in the classical MD simulation, with a simulation time of ns.

5.6 Conclusion

A graph neural network was used to create an ML-FF for an LDH system in this study. The development of this ML-FF focused on the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system, with the goal of validating the previously observed wave structure in the traditional MD simulations. The ML-FF development process was divided into four stages: initial ML-FF generation, ML-FF behavior evaluation, active learning to improve ML-FF behavior and final MD simulation test on an expanded LDH system. The initial ML-FF was created using a dataset of 2370 LDH structures and a 5-fold cross validation, but underestimated the energy and force predictions

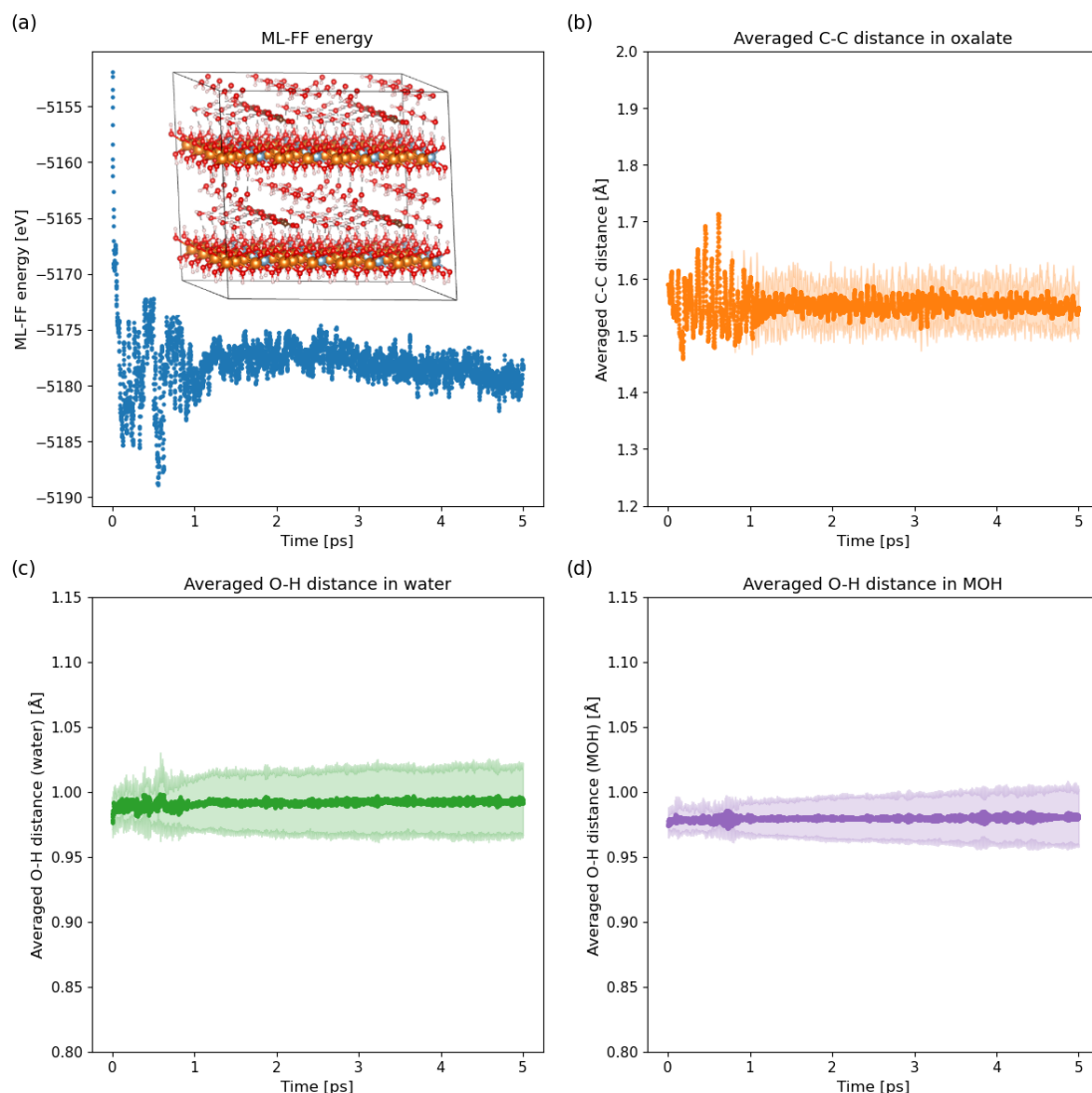


Figure 5.18: **(a)** ML-FF energy, **(b)** C-C atom distances in oxalate, **(c)** O-H atom distances in water molecules and **(d)** O-H atom distances in MOH layers varied with simulation time. The corresponding standard deviation interval is marked in a slightly lighter color for each property. The applied ML-FF model was obtained from the 9th iteration in the active learning process with the large size LDH. The used expanded LDH is displayed in the inlet in the top left panel.

of stretched structures during its performance evaluation. This underestimation was most likely caused by the absence of a stretched structure in the initial dataset used to train the model. Active learning was introduced to improve model behavior by retraining the model with updated datasets. For the active learning using the unit cell LDH, the model behavior was significantly improved by first recognizing that stretched O-H distances in the water molecules and MOH layers result in large energy and force values, and then recognizing that compressed O-H and stretched C-C distances can also result in high energies and large

Machine-learned force field

forces. Medium (two times of unit cell) and large (four times of unit cell) supercell LDHs were used to further improve the ML-FF prediction accuracy. During the active learning with the medium supercell LDH, the ML-FF was able to learn that it is not energetically favorable to pull two neighboring carbon atoms from two distinct oxalate compounds closer together by breaking the central C-C bond in one of the compounds to form two C_2O_2^- . Compared to using medium supercell LDH, using large supercell LDH requires fewer iterations to achieve the same predictive behavior by active learning since there are more atoms in the large supercell that can provide resourceful information for the model development. Furthermore, energy and force prediction of the retrained model improved significantly through active learning, indicating that the retrained ML-FF is a promising force field for accurately simulating LDHs at significantly lower computational costs. Finally, in the MD simulation with an expanded LDH structure as starting geometry, the retained model was able to keep the O-H and C-C distances to values around their expected chemical bond lengths.

The ML-FF developed in this work serves as the foundation for investigating large scale LDH systems with thousands of atoms and represents a DFT level accuracy tool for verifying the wave-structure observed in the classical MD simulation mentioned in the previous chapter, with comparable efficiency. Furthermore, this work provides a solid framework for training a new and general applicable force field for different LDH systems intercalated with other corrosion inhibitors.

6 Screening for benign corrosion inhibitors

This chapter is extracted from the article 'Predicting corrosion inhibition efficiencies of small organic molecules using data-driven techniques' submitted to the journal npj Materials Degradation by X. Li, B. Vaghefinazari, T. Würger, S.V. Lamaka, M.L. Zheludkevich, C. Feiler.

The selection of potent corrosion inhibitors is another critical step in facilitating a long-term and environmentally friendly corrosion protection solution based on layered double hydroxides (LDHs). Oxalate ($\text{C}_2\text{O}_4^{2-}$) has been demonstrated to be an effective inhibitor for several magnesium (Mg) alloys. [30] However, the simulation results in Section 4 indicate that LDHs intercalated with $\text{C}_2\text{O}_4^{2-}$ will be a highly stable system, and thus the intercalated $\text{C}_2\text{O}_4^{2-}$ may not be voluntarily released, limiting its corrosion protection capability for Mg alloys. To achieve highly effective active corrosion protection, new promising organic inhibitors for the design of novel LDHs are required. However, due to the large number of organic compounds with potentially useful properties [295], identifying an effective organic corrosion inhibitor to be intercalated into LDHs (see Figure 6.1) to protect a specific type of Mg alloy can be very challenging. Furthermore, purely experimental studies on the intercalation of new organic molecules into LDHs can be time-consuming due to the large number of candidate molecules to choose from. [9] Fortunately, machine learning-based approaches promise to generate a shortlist of candidate inhibitors prior to large-scale experimental testing and computer simulations to model their properties.

As described in Section 1.3.3.2, a number of quantitative structure-property relationship (QSPR) models have been successfully developed to predict the corrosion inhibiting effect of small organic compounds for various types of metal alloys. In this section, support vector regression (SVR) and kernel ridge regression (KRR)-based QSPR models were developed to assist in the identification of effective organic corrosion inhibitors for AZ91.

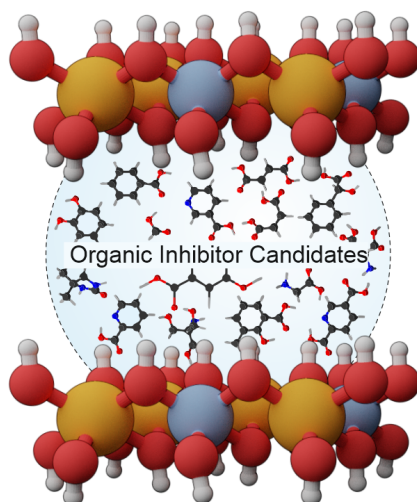


Figure 6.1: Schematic representation of a layered double hydroxide system with a large number of organic inhibitor candidates.

6.1 Methodology

Following the process of the QSPR model generation, the corrosion inhibition responses of 58 small organic molecules on Mg alloy AZ91 from a previous study [30] were used as the training dataset in this study. AZ91 was chosen as the substrate because previous experimental work [396] demonstrated that LDHs can be synthesized directly on the surface of this alloy as a conversion layer. The corrosion inhibition efficiencies (IEs) of the samples in the used dataset are more variable than those used in other Mg alloy prediction models [293, 295, 296] so far, necessitating the use of a machine learning algorithm with good generalization capabilities. SVR is a generalization-capable algorithm that has been used to develop a predictive model to investigate the effect of the outdoor environment on the corrosion rates of metallic materials [316, 317] and to identify effective corrosion inhibitors for Q235 steel using a limited dataset [300]. Therefore, SVR was chosen for the QSPR model construction in this work to investigate its applicability for Mg-based datasets and the quality of the predictions was evaluated using experimental blind testing. Meanwhile, the KRR approach was chosen as a benchmark for comparing the performance of the SVR model because it has been shown to be capable of developing a reliable predictive model for the inhibition efficiency of small organic compounds for CPMg220 [295].

Figure 6.2 depicts an overview of the workflow used in this study. Following the feature generation and feature selection steps, a model was generated based on the AZ91 experimental dataset, and the model was validated using experimental blind tests. Prior to the model construction, the molecular structures of the 58 organic compounds extracted from the experimental dataset were optimized using the quantum chemical software package Gaussian 16 [397] at the TPSSh/def2SVP level of theory. After combining the DFT calculated features (highest occupied (HOMO) and lowest unoccupied molecular orbital (LUMO)) with the molec-

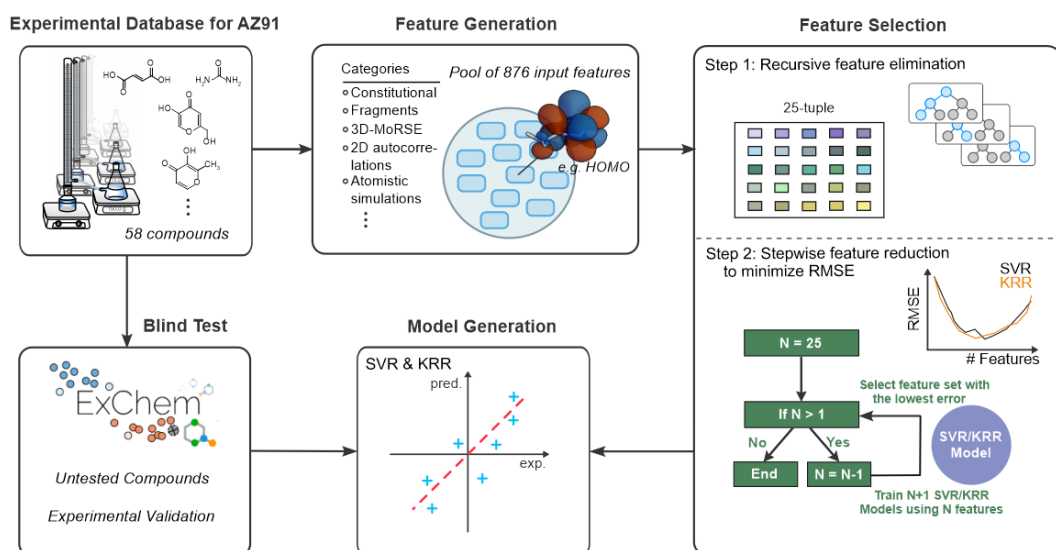


Figure 6.2: Schematic representation of the ML workflow used in this study. A database of 58 small organic molecules and their corrosion responses to AZ91 are employed as a training database. To encode their molecular structures, a pool of molecular descriptors is generated first, and a two-step sparse feature selection approach is used. Following that, the most relevant descriptors are subsequently used to train supervised machine learning models to predict the behavior of untested chemicals. The small organic molecules used in the blind test are selected based on a previously published ExChem [295] approach.

ular descriptors obtained from the chemoinformatics software package alvaDes [322], the initial feature set in this work consisted of more than 800 features, which significantly exceeded the number of compounds in the initial dataset. Using too many features in a model construction can lead to high model complexity, long training times, and overfitting. [398] Hua et al. [399] demonstrated that the optimal number of features is always less than the sample size, regardless of whether the features are uncorrelated, slightly correlated, or highly correlated. Therefore, feature selection was required in this work to select relevant features.

Initially, random forests-based recursive feature elimination (RFE) was used to select a 25-tuple of input features from the pool of available descriptors. This step can initially reduce the feature space, resulting in a shorter computational time. The selection procedure was repeated 50 times to account for the randomness introduced by random forest method. These 50 distinct groups of features obtained in step 1 were fed into the 5-fold cross validation of the SVR model. The 25-tuple with the lowest averaged test root mean squared error (RMSE) of the cross validation was then chosen as the foundation for searching for the most relevant features in the SVR and KRR models, respectively, because 25 features is still a very large number given the size of the data compared to the Refs [293, 296, 318]. These 25 features were thus reduced stepwise (one feature per step) in the following feature selection to remove insignificant features in the model training. Instead of applying RFE, the SVR and KRR models

were used directly to select features for each step, along with hyperparameter optimization and cross validations.

Structure similarity was calculated using the distance metric, with the selected input features being the coordinates of each compound in the corresponding high dimensional feature space. To compute the similarity, the radial basis function (RBF) kernel used in the SVR and KRR models was used, which is defined as:

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \quad (6.1)$$

where x and y are the vectors of the selected input features for two compounds, respectively.

Blind tests were performed using the same experimental setup and conditions as in the work of Lamaka et al. [30], from which the initial dataset was extracted, to validate the two (SVR and KRR) models developed in this work. For the blind tests, Trimesic acid and Pyromellitic acid were suggested by experimentalists based on chemical intuition, while the remaining candidates were chosen using the ExChem method described in a previous work [295] and a database of 7094 commercially available compounds provided by Thermo Fisher Scientific. The IE of compounds was calculated using a hydrogen evolution test, in which the amount of hydrogen produced by magnesium corrosion was measured while immersed in a NaCl solution. 0.5 g magnesium AZ91 chips with a surface area of $430 \pm 29 \text{ cm}^2/\text{g}$ were immersed in 0.5 wt.% NaCl solution without (reference solution) and with the untested compounds. The chemical composition of the AZ91 chip is identical to the work of Lamaka et al., which is provided in Appendix A.3 Table A.1. The compound concentration was 0.05 M, and the pH of the solution was adjusted to 7 ± 0.1 by NaOH/HCl. For each solution, the hydrogen evolution measurements were repeated three times, and the average of the calculated IEs was used for the corresponding blind test data point. The IE was defined by the following equation:

$$\text{IE} = \frac{V_{\text{H}_2}^0 - V_{\text{H}_2}^{\text{Inh}}}{V_{\text{H}_2}^0} 100\% \quad (6.2)$$

where $V_{\text{H}_2}^0$ and $V_{\text{H}_2}^{\text{Inh}}$ are the volumes of H_2 evolved after 20 h of immersion in the reference NaCl solution and the NaCl solution containing the investigated chemical compound, respectively. More details on the hydrogen evolution tests are available in the original publication [30].

6.2 Selecting relevant features for each model

In this work, feature selection was investigated as a key factor in the development of ML models for predicting corrosion IEs of small organic molecules on Mg alloy AZ91. At each step in the stepwise feature selection, there are multiple ways to remove a feature from the previous step, for example, there are twenty-five ways to remove one feature from the initially selected 25 features. The $n-1$ feature set with the lowest averaged test RMSE was selected at each step (plotted in Figure 6.3) and the preserved features were used as the basis for the next step until

6.2 Selecting relevant features for each model

only one input feature left to build the model. For the selected feature set, the removed feature was defined as the least important feature in the previous step. The averaged RMSEs for the corresponding train sets in Figure 6.3 were listed in Appendix A.3 Table A.2. Applying this approach, the most relevant features with the lowest averaged test RMSE were selected for the SVR and KRR models, respectively. The continued stepwise procedure yielded an order of importance for the selected features based on their removed order after selecting the optimal features for each model.

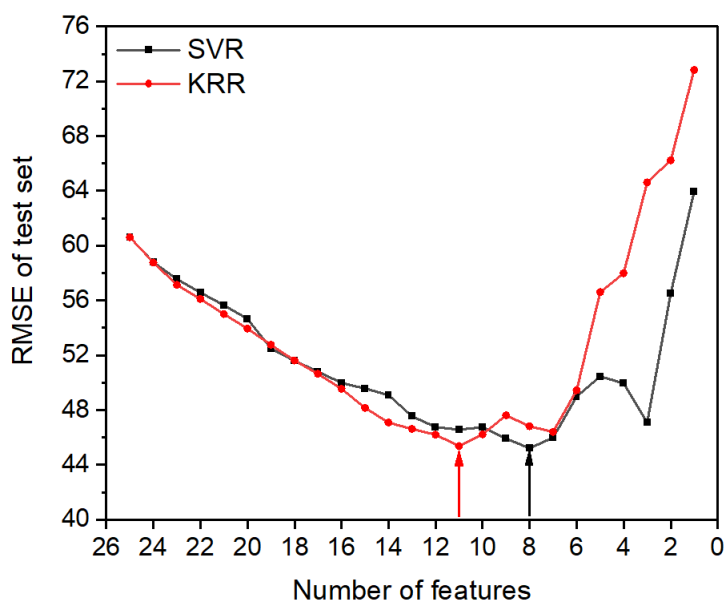


Figure 6.3: 25-tuple features selected after applying the random forest-based RFE in Step 1 were removed one-by-one and the minimum averaged RMSE of the test sets in the cross-validations varied with the number of features in the SVR (black line) and KRR (red line) models.

According to the trend of the black line in Figure 6.3, the optimal number of features for the SVR model is equal to eight, as the resulting model has the lowest RMSE. The selected molecular descriptors are P_VSA_LogP_2, Mor28e, HOMO, MATS4v, Mor06s, GATS4p, MATS8m, and Mor15v, in order of feature importance. Except for the HOMO obtained from DFT calculations, the other seven features are from three descriptor categories obtained from the chemoinformatics software package alvaDesc [322] (P_VSA-like descriptors, 3D-MoRSE descriptors [400] and 2D autocorrelations). P_VSA-like descriptors [401] are based on the van der Waals surface area of the compounds, which is calculated by adding up all atomic contributions. Information about the entire molecule structure is incorporated into 3D-MoRSE descriptors [400] by summarizing the atomic pairwise information related to the scattering parameter based on electron diffraction and then weighted by either of the properties, e.g. mass, Sanderson electronegativity, van der Waals volume and atomic polarizability. The 2D autocorrelations descriptors [402] are calculated to provide interdependencies between atomic properties

(analogous to the 3D-MoRSE descriptors) that are linked by a log function. [403] All three descriptor categories calculate the spatial distribution of a generic molecular property rather than focusing solely on atomic configurations.

The optimal number of features in the KRR model is eleven, as shown by the red line in Figure 6.3. The eleven features were identified as Mor15v, HOMO, MATS8m, Mor30e, nRNH2, C-018, GATS4p, MATS2i, Mor11e, Mor06s, Mor28e, in order of their feature importance. It is worth noting that six of the eleven features are identical to those selected for the SVR model. Mor15v, HOMO, MATS8m, GATS4p, Mor06s, and Mor28e are the overlapping features. This finding implies that DFT-derived HOMO energies, 3D-MoRSE descriptors and 2D autocorrelations descriptors appear to encode critical structural information for predicting the corrosion inhibition efficiency of small organic molecules for AZ91. Furthermore, this observation is consistent with the findings of Schiessler et al. [293], in which DFT calculated features and 3D-MoRSE descriptors were identified as important input features for an artificial neural network using IEs of small organic molecules as a target property for the Mg-based alloy ZE41.

Aside from these three feature groups, a number of features encoding functional group counts and atom-centered fragments were identified for the top eleven features in the KRR model, such as nRNH2, which directly encodes the number of aliphatic primary amines. The five compounds in the dataset with nRNH2 moieties are all amino acids with negative inhibition efficiencies (Cysteine, Glutamic acid, Glycine, DL-norleucine and DL-phenylalanine). The corrosion-accelerating behavior of amino acids can be attributed to the solubility of their corresponding magnesium complex in water. [404, 405] The atom-centered fragment feature C-018 represents =CHX, where "=" represents a double bond and X can be any of the following heteroatoms: O, N, S, P, Se, or any halogen. In the =CHX fragment, a sp^2 -hybridized carbon atom is directly connected to a hydrogen and one of the heteroatoms denoted as X. This specific functional group is found in the training compounds Kojic acid, Maltol and Uracil (X represents either O or N), and all three organic molecules have negative IE values, as shown in Appendix A.3 Figure A.6. The complexes formed by these three compounds with magnesium have been shown to be water soluble. [405–407] In comparison to the ability of organic inhibitors to form complexes with metal ions, the solubility of these complexes in water appears to be a more important factor in determining their efficiency. This observation is consistent with the findings of Lamaka et al. [30] and Anjum et al. [408] that organic compounds with low water solubility in their complexes have a high inhibiting effect because they delay corrosion by forming a protective barrier layer.

6.2 Selecting relevant features for each model

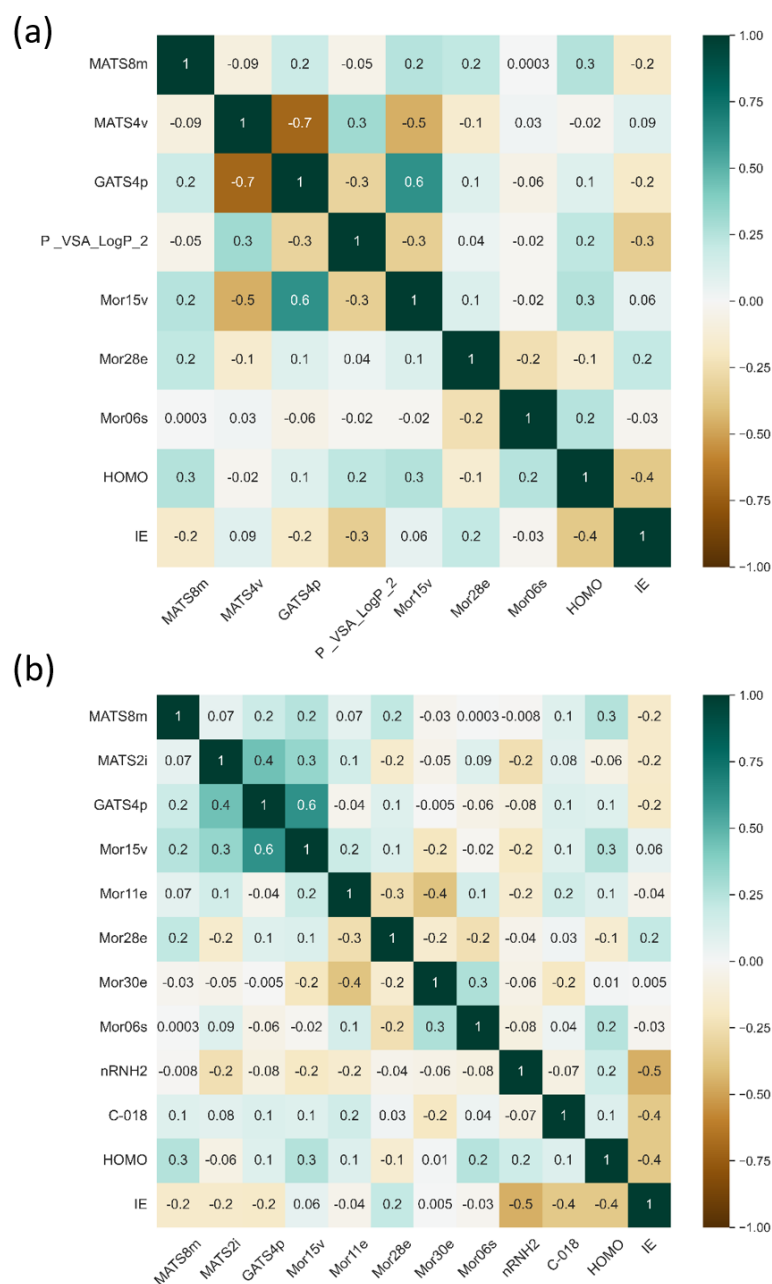


Figure 6.4: Pearson correlation coefficients of the two models. (a) Pearson correlation between the 8-tuple features selected for the SVR model and IEs. (b) Pearson correlation between the 11-tuple features selected for the KRR model and IEs.

Some molecular descriptors obtained from chemoinformatics tools are esoteric and difficult to associate with physical properties because they are derived from extensive mathematical manipulations of the chemical graph. Pearson tests provided a better understanding of the correlation between the used input features and IEs, as well as a statistical significance measure. The Pearson correlation coefficient, which ranges from -1 to 1, quantifies the

linear relationship between two sets of data, with 0 indicating no correlation and -1 and 1 indicating exact negative and positive correlations, respectively. [282] The correlation between the individual features and the IEs is not strong for either model because the values of the determined correlation coefficients in Figure 6.4a and b are not higher/lower than ± 0.5 , where the most pronounced negative and positive correlations are -0.5 and 0.2 , respectively. This observation agrees well with the findings of Guyon et al. [409] that the selected features are not always the most relevant in terms of the target property on their own. None of the correlations between the selected features are considered strong (> 0.9), and the majority of the correlations (over 90%) are interpreted as weak relationships ($0.1 - 0.39$) or negligible (< 0.1) as defined by Schober et al. [282]. Furthermore, p-values between the used input features and IEs were calculated and are illustrated in Appendix A.3 Figure A.7, where the p-value indicates whether the correlation is statistically significant. The weak correlations between the majority of the selected features largely ensure that there is no redundant feature selected as input for the models. Despite the fact that the majority of the selected features are only weakly correlated with the target property, the results indicate that they can still be used to build a predictive model as a group due to underlying synergistic effects, which is consistent with previous works. [293, 296]

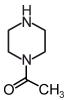

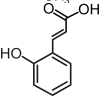




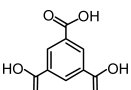
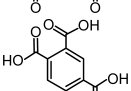

6.3 Performance evaluation of each model

In conjunction with the feature importance investigation, hyperparameters for the SVR and KRR models were optimized using a grid search with 5-fold cross validations. As a result, SVR ($\text{random_state}=10$, $C=17$, $\gamma=0.1$) and KRR ($\text{random_state}=10$, $\alpha=0.1$, $\gamma=0.1$) hyperparameter sets were selected respectively. The random state parameter (random_state) was chosen to be the same for both models, indicating the same split of the train and test sets during the cross-validation. Following the selection of hyperparameters, the entire initial dataset was used to fit the SVR and KRR models, which were then used to predict the behavior of blindly tested compounds to assess their robustness. The experimental and predicted values for the ten compounds in the blind test are listed in Table 6.1. For both models, the predicted values for the piperazine derivatives **1** and **2** are highlighted in brown because they are significantly less negative than the corresponding experimental values and fall outside the range of inhibition efficiencies of the chemicals used as the initial dataset in this work. It is worth noting, however, that both compounds were correctly predicted to accelerate AZ91 dissolution. These two compounds were excluded from the subsequent analysis because they are far outside the range of IEs known in the initial dataset.

For the entire initial data set, the SVR and KRR models performed similarly well, as shown by the blue points in Figure 6.5a and b, where the predicted and experimental values correlated well with an RMSE of around 10%. Both models provide accurate predictions for half of the tested compounds in the blind test, but the appearance of underestimated or overestimated values, depicted in Figure 6.5 by blue and red dashed circles or ellipses, results in relatively high RMSE values for both employed models. As shown in Figure 6.5, both the SVR and KRR

6.3 Performance evaluation of each model

Table 6.1: Experimental and predicted values (IEs in %) for the blindly tested compounds. Red values indicate overestimated, while blue values indicate underestimated predictions. The brown values represent predictions that are qualitatively correct but where the actual value is far outside of the range of the IEs used to train the model.

Index	Compound	IE (exp.)	SVR (pred.)	KRR (pred.)
1	 1-Acetylpiperazine	-563	-172	-108
2	 1-Amino-4-methylpiperazine	-517	-195	-109
3	 2-Hydroxycinnamic acid	-24	51	2
4	 2-Hydroxyphenylacetic acid	-14	-21	-53
5	 3-Methylpyrazole	16	9	-17
6	 5-Nitouracil	79	-68	-82
7	 Tartronic acid	30	37	60
8	 Trimesic acid	67	-89	23
9	 Pyromellitic acid	52	34	23
10	 Trimethylolpropane	20	-42	-49

models underestimate 5-Nitouracil (**6**, $IE_{pred,SVR} = -68\%$, $IE_{pred,KRR} = -82\%$) and Trimethylolpropane (**10**, $IE_{pred,SVR} = -42\%$, $IE_{pred,KRR} = -49\%$) in a similar way. The SVR model contains two more outliers (2-Hydroxycinnamic acid (**3**) and Trimesic acid (**8**)), as shown in Figure 6.5a. Although there are two more outliers in the SVR model, it is important to note that the predicted values of the other four compounds in the blind test set correlate well with the corresponding experimental values, for acetic acid **4** ($IE_{pred,SVR} = -21\%$, $IE_{exp.} = -14\%$), the pyrazole **5** ($IE_{pred,SVR} = 9\%$, $IE_{exp.} = 16\%$) as well as the aliphatic (**7**, $IE_{pred,SVR} = 37\%$, $IE_{exp.} = 30\%$) and aromatic (**9**, $IE_{pred,SVR} = 34\%$, $IE_{exp.} = 52\%$) carboxylic acids, with an RMSE of 11% and an R^2 of 0.782 in the SVR model. For the same non-outlier compounds (**4**, **5**, **7**, **9**), the RMSE and R^2 calculated in the KRR model are 33% and 0.385, respectively. These findings suggest that the SVR and KRR models are both able to provide good estimates for the four

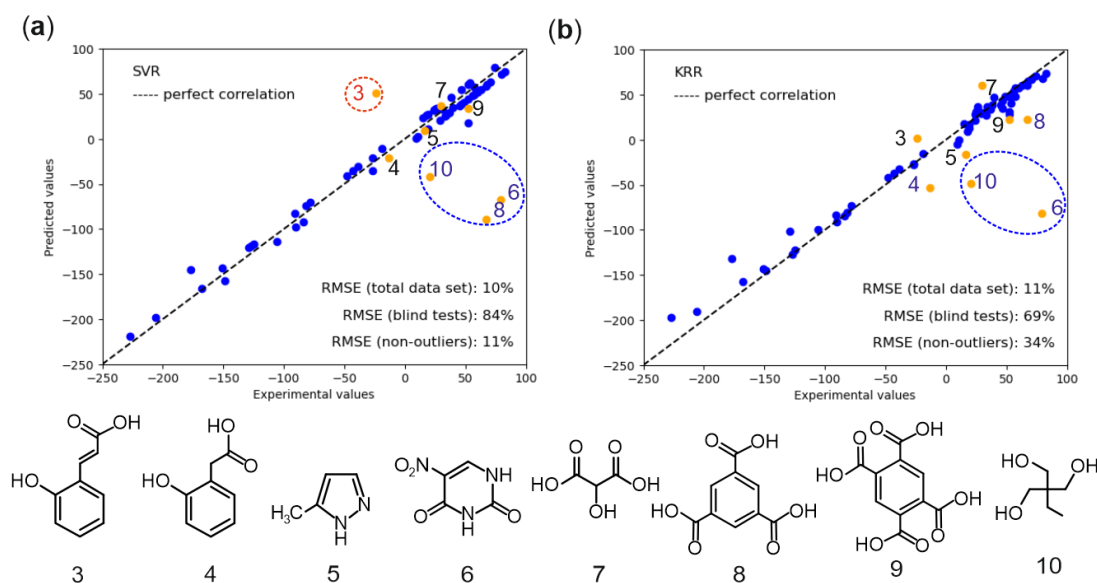


Figure 6.5: Performance evaluation of the (a) SVR and (b) KRR models. The correlation between predicted and experimentally measured values (IEs in %) is shown in both plots. The blue dots represent the initial dataset in its entirety (58 compounds, names and IEs are listed in Appendix A.3 Table A.3). The blindly tested compounds are represented by orange dots. Please note that the plots do not include 1-Acetylpiiperazine (1) or 1-Amino-4-methylpiperazine (2). Although their estimates are qualitatively correct (1: $IE_{pred,SVR} = -172\%$, $IE_{pred,KRR} = -108\%$, $IE_{exp} = -563\%$; 2: $IE_{pred,SVR} = -195\%$, $IE_{pred,KRR} = -109\%$, $IE_{exp} = -517\%$), their measured values are outside the models range. The corresponding structures of the plotted blind test compounds are shown at the bottom of the figure for clarity. The over- and underestimated compounds are denoted by red and blue dashed circles or ellipses, respectively.

blind test compounds (4, 5, 7, 9). For the compounds where the predictions yielded reliable estimates based on the SVR (4, 5, 7, 9) and KRR (3, 4, 5, 7, 8, 9) models, the Pearson correlation coefficient and p-value were calculated between their predicted and experimental values. The predicted values of the SVR model (coefficient=0.93, p-value=0.071) have a higher correlation with the experimental results than the KRR model (coefficient=0.60, p-value=0.214), while the p-value of the SVR model indicates statistical relevance of the prediction. In terms of prediction accuracy for non-outlier compounds for the given dataset, the SVR model outperforms the KRR model, whereas the KRR model has fewer outliers.

Moreover, modulators containing an aliphatic primary amine ($nRNH_2$), such as an amino acid, or fragments with the general formula $R=CHX$ cause elevated corrosion rates in experimental studies. [30] The results suggest that small organic molecules with either of the above mentioned functional moiety can most likely be excluded from the screening for effective corrosion inhibitors. However, they might be useful in other applications, such as battery electrolyte additives, where controlled dissolution of the Mg-based anode material is required. [410] One out of the ten compounds in the blind test set (5-Nitouracil (6)) contains a $=CHX$ fragment,

indicating a negative IE value. Despite the predicted negative inhibition efficiency, the experimental result demonstrated that 5-Nitrouracil provided adequate inhibition performance. This could be attributed to the nitro compound of 5-Nitrouracil, which have been shown to aid in the corrosion protection of a variety of alloys. [411–413] While Uracil has a negative IE value (-151%), its substitution with a nitro moiety, 5-Nitrouracil, results in a highly potent corrosion inhibitor (78%), indicating that the nitro moiety is important in corrosion protection. Furthermore, in experimental tests, some organic compounds with nitro groups inhibited commercially pure Mg and Mg alloy ZE41 with high efficiency. [30, 295] Due to the scarcity of information on the effect of the nitro function in the initial dataset, only two compounds (5-Nitrobarbituric acid and 3-Methyl-2-nitrobenzoic acid) exhibited this functional group, neither model used in the study could accurately predict the effect of this functional moiety. This strongly suggests that future experimental datasets need to include more compounds containing a nitro moiety to enable the model to recuperate the impact of this group on the corrosion inhibiting effect.

To gain more insights of the compounds which are outliers, the pairwise distances based on input features were calculated between the compounds in the blind test and those in the initial dataset to evaluate the highly similar structures for each blind test compound. A value of 1 in the similarity matrix suggests high similarity while a value of 0 indicates no similarity. The similarity matrices of the SVR and KRR models for the eight blind test compounds and the initial dataset are shown in Figure 6.6a and b, respectively. Furthermore, Figure 6.6 depicts the top 5 similar structures (with names and inhibition efficiencies) of 5-Nitrouracil (**6**) for both models as an example. An order of similarity from the highest to the lowest can be extracted for these five structures in SVR (Uracil, Glycine, 5-Nitrobarbituric acid, DL-Phenylalanine, Glutamic acid) and KRR (Uracil, Maltol, Kojic acid, Fumaric acid, Urea). Some of the top 5 similar structures have significant similarity differences, such as those shown in Figure 6.6 b between Uracil and Urea. This indicates the limitation of the dataset used in this work because there are only 58 data points in total. As a result, there are insufficient structures in the dataset that have similarities higher than or comparable to the similarity between Uracil and the blind test compound 5-Nitrouracil (**6**).

For each outlier compound, the IEs of its top 5 similar structures are ordered by similarity, as shown in Table 6.2. The average of the IEs of these 5 similar structures is listed in the last column of Table 6.2. The average IE values are in fairly good agreement with the predicted values. This demonstrates that the models can recognize the similarity linkage in the dataset and make predictions accordingly. However, due to the small size of the dataset, the similarity linkage is limited, resulting in the appearance of these outliers. The learning curves for the SVR and KRR models (as illustrated in Appendix A.3 Figure A.8) show that in the cross validation, the averaged RMSEs of the test sets decrease as the size of the training set increases, despite the fact that the RMSEs of the test sets for both models are higher than the RMSEs of the train sets. One possible solution is to increase the dataset size so that the averaged RMSEs of the test sets can be consistently reduced by adding more training data.

Screening for benign corrosion inhibitors

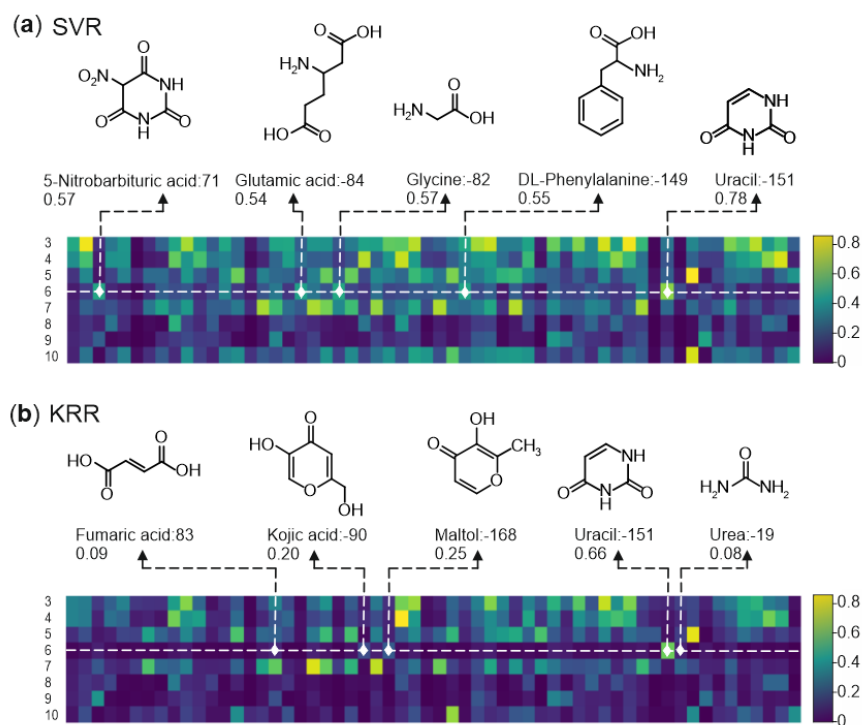


Figure 6.6: Similarity matrix of the 8 blind tested compounds and the 58 compounds in the dataset for the (a) SVR model and (b) KRR model. As an example, the top 5 similar structures containing names and inhibition efficiencies are plotted in the figure for 5-Nitrouacil (6). The similarity values are shown below the names. The color scale corresponds to the matrix values, with dark blue indicating low / no similarity, green moderate and yellow high similarity values.

Table 6.2: The IEs in % of the extracted top 5 similar structures from Ref. [30] are listed in order of similarity from high to low (from 1st to 5th, please note that 1st, 2nd, 3rd, 4th, 5th do not indicate the same structures but refer to those that are most similar to the outlier compounds.) and the average of these 5 IEs for the SVR and KRR outliers. The red value indicates overestimated, while blue values indicate underestimated predictions.

		IE _{exp}	IE _{pred}	1 st	2 nd	3 rd	4 th	5 th	Average
SVR	3	-24	51	26	56	18	45	24	34
	6	79	-68	-151	-82	71	-149	-84	-79
	8	67	-89	38	11	-90	-125	52	-23
	10	20	-42	-27	-27	24	34	-106	-20
KRR	6	79	-82	-151	-168	-90	83	-19	-69
	10	20	-49	-27	-91	54	-27	-129	-44

6.4 Conclusion

In this study, QSPR models based on SVR and KRR were developed to select an efficient organic corrosion inhibitor from a large number of compounds. These two models were developed using experimental inhibition efficiencies of 58 organic compounds for an AZ91 alloy that were retrieved from literature. A two-step feature selection method was proposed in this work to select relevant features while improving the prediction accuracy of the model. After first reduction of the pool of available features to a 25-tuple using RFE, this feature set is subsequently systemically screened for the best n -tuple to train the predictive model, rather than relying on human intuition to choose the number and composition of input features. To assess the performance of the two models, ten untested chemicals were subjected to experimental blind tests. Both models correctly predicted 2-Hydroxyphenylacetic acid, 3-Methylpyrazole, Tartronic acid, and Pyromellitic acid, with the SVR model predicted values being closer to the actual values than the KRR model. 1-Acetylpiperazine and 1-Amino-4-methylpiperazine were correctly predicted as strong accelerators, but the predicted values were not quantitatively correct. This could be because their experimental values are outside of the IE range in the initial dataset. Outliers were identified by both models as 5-Nitrouracil and Trimethylolpropane, with the SVR model identifying two more (2-Hydroxycinnamic acid and Trimesic acid). To gain a better understanding of the appearance of these outliers, the top 5 highly similar structures for each outlier were extracted from the initial experimental dataset. The similarity values of a portion of the extracted top similar structures are relatively low and vary significantly for each outlier, indicating that there are insufficient structures with high similarities to each outlier in the initial dataset. More training samples could be fed into the model training to overcome this initial dataset limitation and improve the accuracy of predictions for untested structures.

7 Conclusion & Outlook

7.1 Conclusion

Atomistic methods based on density functional theory (DFT) and molecular dynamics (MD) simulations, data-driven inhibitor selection workflows and the development of machine-learned force fields depict three key elements of the simulation-assisted design of layered double hydroxides (LDHs) that provided the motivation for this work. The goal of this thesis is to gain a better understanding of how anion exchange process and hydration state influence on the structure and stability of LDHs intercalated with inhibitors through an atomistic investigation combined with machine learning techniques to support the design of LDH-based corrosion protection strategies.

Studies on the anion exchange process in combination with the influence of hydration states on LDHs intercalated with NO_3^- and Cl^- were carried out towards a more complete understanding of anion exchange mechanisms by combining the output of DFT and MD simulations. It was shown that, at the start of the anion exchange, the intercalated NO_3^- was released first, followed by the new intercalation of the Cl^- . Following the new Cl^- intercalation, more water molecules were released from the LDH system, resulting in a decrease of the interlayer distance from 0.89 nm of pure LDH- NO_3^- system. Furthermore, it was found that the hydration state has an influence the stability of an LDH system by changing the configuration of the interlayer region. However, this influence was found to be limited for highly stable systems such as LDH- CO_3^{2-} . The stable hydration state of each LDH system was identified and the stability order was in good agreement with the affinity order described in literature. The hydration state should thus be considered in the design of novel LDH, and the designed LDH should avoid the formation of highly stable systems.

A machine-learned force field (ML-FF) was developed for the $\text{Mg}_6\text{Al}_2\text{-LDH-C}_2\text{O}_4^{2-}$ system to verify the observed wave structure in classical force field MD simulations and to provide a solid framework for training a new and general applicable force field. It was demonstrated that the obtained ML-FF predicts the force and energy of the investigated system with DFT accuracy while avoiding the formation of unrealistically long or short bond lengths. Furthermore, it

Conclusion & Outlook

has been demonstrated that this force field can be scaled up to study large LDH systems, indicating that the developed ML-FF is a promising approach for simulating LDHs on a large scale with considerably lower computational cost. Efficient strategies for identifying effective corrosion inhibitors for intercalation were developed after the establishment of a framework for simulating the properties of the LDH system. Following a two-step input feature selection workflow, quantitative structure-property relationship (QSPR) models based on support vector regression (SVR) and kernel ridge regression (KRR) were developed to generate short lists of promising small organic molecule corrosion inhibitors. The performance of the two models was assessed by experimental blind testing of a ten-membered set of untested chemicals. The SVR model had 6/10 qualitatively correct predictions (4/8 quantitatively correct) and the KRR model had 8/10 qualitatively correct predictions (6/8 quantitatively correct) although the inhibition efficiencies of two compounds were well-beyond the range of the training dataset. However, more training data describing a wider area of chemical space is required to make the developed models more robust as well as to improve the accuracy of the predicted values. Furthermore, the results suggested that more compounds containing a nitro moiety should be included in future experimental datasets to allow the model to capture the potential positive impact of this group on corrosion inhibiting effect.

In summary, the results presented in this work lay the foundation for a comprehensive investigation of LDH-based systems using multi-scale computer simulations ranging from ML-FF development to the selection of potent corrosion inhibitors prior to testing to atomistic simulations that can be employed to gain a deeper understanding of key properties of LDH for corrosion control.

7.2 Outlook

This thesis included a comprehensive study of the atomistic simulation, data-driven inhibitor selection and machine-learned force field development, which is a good starting point for connecting them into a feedback loop, as shown in Figure 7.1. In this loop, the effective inhibitors selected from the QSPR-based models can be intercalated into the LDH to train the corresponding machine-learned force field, which can be employed in atomistic simulations to study the anion exchange process and estimate the structure and stability of the resulting LDH.

Further adaption of the ML-FF is needed to facilitate simulations of LDHs intercalated with other inhibitors, e.g. 5-Nitouracil, which has been identified as an effective inhibiting agent. In the long term, ML-FF development can become more generic by combining all of the different LDH datasets. In this case, a new ML-FF does not need to be retrained for each novel proposed LDH system. This generic and transferable ML-FF could be used to investigate various LDH systems in corrosion protection or another application field such as photocatalytic applications to investigate their chemical reactions. Further improvement of the QSPR based models can be done by including more experimentally tested compounds, or changing the tar-

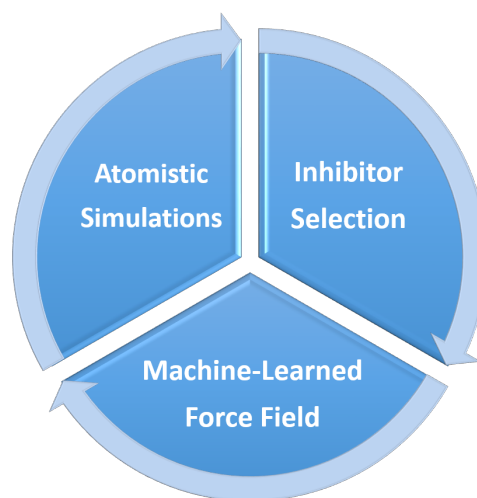


Figure 7.1: Schematic representation of a feedback loop between atomistic simulations, inhibitor selection and machine-learned force field development.

get metric format (e.g. from inhibition efficiency to inhibition power). Compounds containing nitro moieties should be further tested to verify their potential positive inhibition efficiency and to ensure that their chemical composition is well represented in the training dataset. The increased chemical versatility will expand the domain of applicability of QSPR-based models.

Figure 7.2 depicts a possible workflow for using the multi-scale simulation techniques developed in this thesis to guide the experimental design of such corrosion protection systems. An inhibitor with a predicted IE higher than 90% can be selected from a large pool of candidates and will be intercalated into the LDH. The corresponding ML-FF will be trained and used to investigate the anion exchange process and the stability of this type of LDH to estimate if the intercalated inhibitor will be voluntarily released from the LDH system via anion exchange. On the experimental side, LDH with a higher stability than LDH- NO_3^- , the most commonly used precursor for novel LDH design, but a lower stability than LDH- CO_3^{2-} , which is known to be a highly stable system, will be suggested for synthesis. This requirement ensures that the novel LDH can be synthesized via anion exchange with NO_3^- while avoiding the formation of a highly stable LDH system. The combination of inhibitor pre-selection and LDH stability research can significantly guide experimental work while saving money and time and reducing negative environmental impacts, resulting in the most efficient design of novel LDH used in corrosion protection.

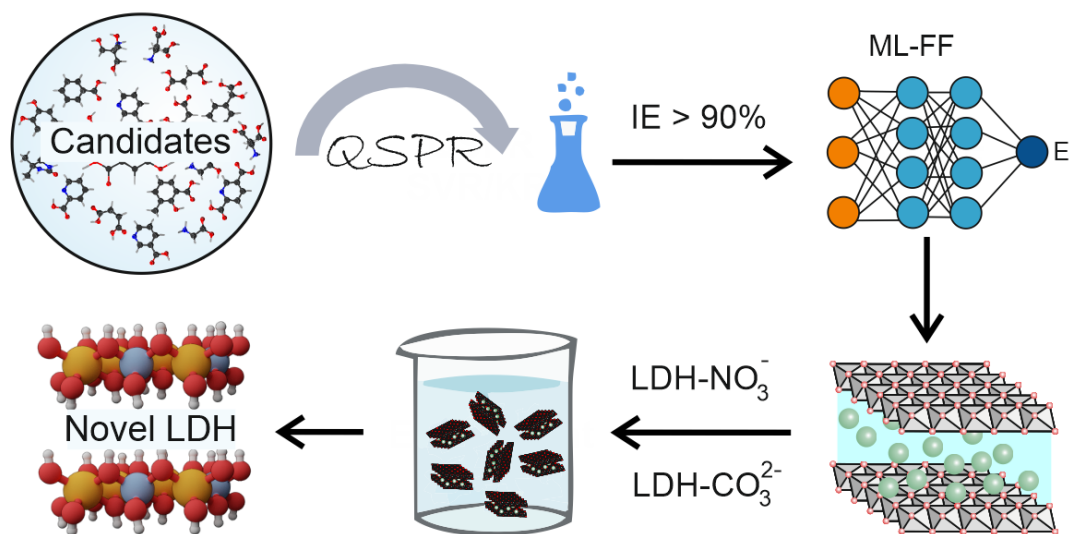


Figure 7.2: Schematic workflow depicting the use of multi-scale simulation techniques to aid in the experimental design of layered double hydroxides. The SVR/KRR-based QSPR model is used to pre-select organic inhibitor candidates. Only candidates with an IE value higher than 90% will be transferred to the development of the corresponding LDH ML-FF. The stability of this novel LDH will be investigated using atomistic simulations, and only the LDH with a stability between LDH-NO_3^- and LDH-CO_3^{2-} will be suggested for experimental synthesis.

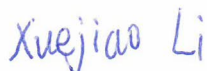
Own contribution to the manuscripts included in the dissertation

1. **X. Li**, T. Würger, C. Feiler, RH. Meißner, M. Serdechnova, C. Blawert, ML. Zheludkevich. 'Atomistic Insight into the Hydration States of Layered Double Hydroxides', in ACS omega 7.14 (2022): 12412-12423. (incorporated in **Section 4**)

X. Li conceived and planned this work in discussion with C. Feiler and ML. Zheludkevich. T. Würger provided valuable advice on the DFT calculations. X. Li carried out and analyzed the atomistic simulations included in this work. X. Li wrote and revised the paper in response to all comments and corrections made by all authors.

2. **X. Li**, B. Vaghefinazari, T. Würger, S.V. Lamaka, M.L. Zheludkevich, C. Feiler. 'Predicting corrosion inhibition efficiencies of small organic molecules using data-driven techniques' submitted to the journal npj Materials Degradation. (incorporated in **Section 6**)

X. Li conceived and planned this work in discussion with C. Feiler, S.V. Lamaka and ML. Zheludkevich. B. Vaghefinazari carried out the experimental work included in this work. X. Li built the machine learning models and interpreted the results. C. Feiler gave advices on the interpretation of the results. X. Li wrote and revised the paper in response to all comments and corrections made by all authors.



Doctoral researcher's signature



Supervisor's signature and stamp

A Appendix

A.1 The impact of intercalated water on LDH properties

A.1.1 The definition of the planes for the intercalated anions

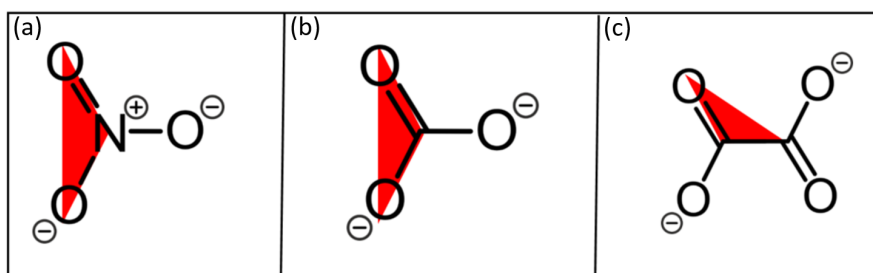


Figure A.1: The definition of the planes for the intercalated anions. The anion plane (a) for NO_3^- was defined by one nitrogen and two oxygens forming a triangle as marked in red, (b) for CO_3^{2-} by one carbon and two oxygens forming a triangle as marked in red and (c) for $\text{C}_2\text{O}_4^{2-}$ by two carbons and one oxygen atom marked as a triangle in red.

A.1.2 MSD at different hydration states for the intercalated anions

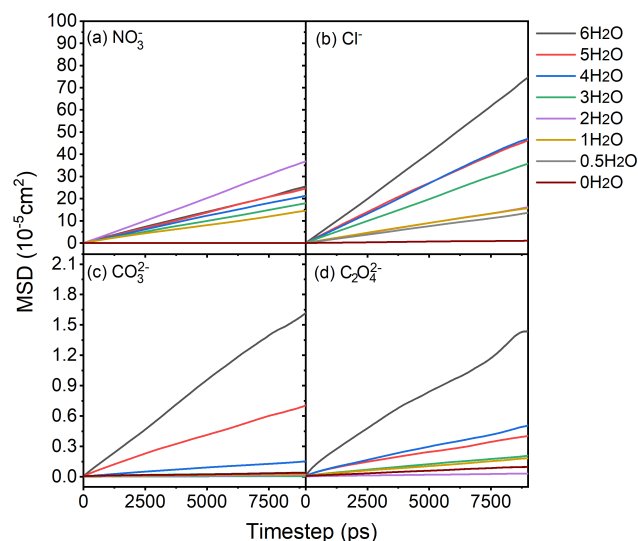


Figure A.2: MSD at different hydration states for (a) NO_3^- , (b) Cl^- , (c) CO_3^{2-} and (d) $\text{C}_2\text{O}_4^{2-}$.

A.1.3 Angle distribution of the intercalated CO_3^{2-} at different hydration states

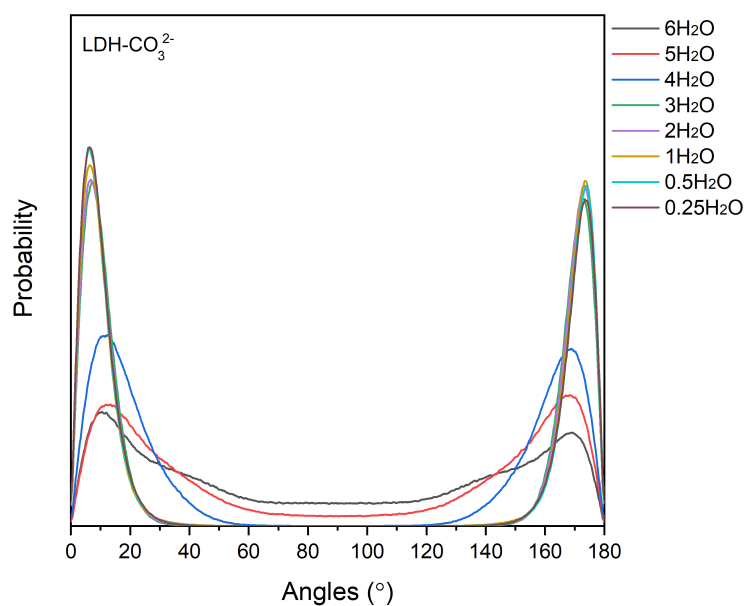


Figure A.3: Angle distribution between CO_3^{2-} plane and the MOH layer at different hydration states.

A.2 Machine-learned force field

A.2.1 Averaged O-H atom distance in the MOH layers

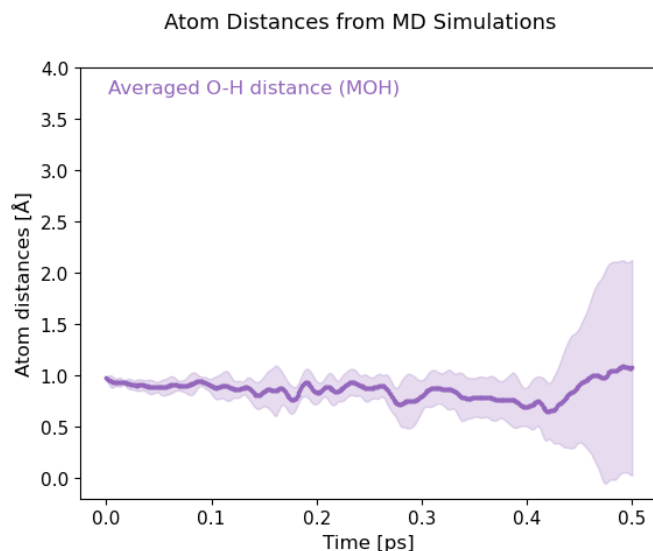


Figure A.4: The O-H atom distance in the MOH layer of LDH (in purple) changes over simulation time for the unit cell LDH. The O-H distance represents the averaged value of all hydroxides ions in the MOH layers, and the standard deviation interval is indicated in light purple.

A.2.2 Averaged O-H atom distances in water molecules and MOH layers

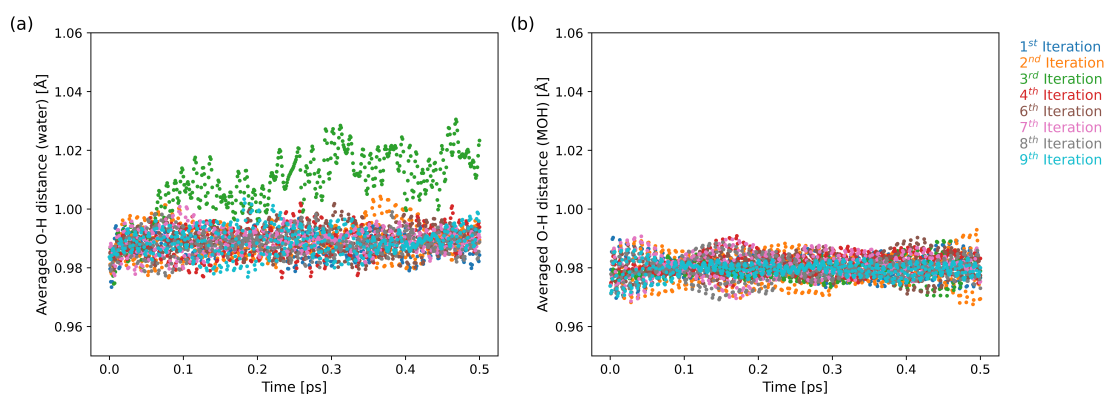


Figure A.5: (a) averaged O-H atom distance in water molecules, (b) averaged O-H atom distance in MOH layers varied with simulation time from the 1st to the 9th iteration excluding the 5th iteration for the medium size LDH. From the 1st to the 9th iteration, the results are shown in the following color order: blue, orange, green, red, purple, brown, pink, grey and cyan.

A.3 Screening for benign corrosion inhibitors

A.3.1 The chemical composition of AZ91

Table A.1: The chemical composition of AZ91 alloy as analysed by spark (optical) emission spectroscopy. The values are in ppm or in at.% when indicated.

Element	
Ag	< 1
Al	$7.95 \pm 0.29\%$
Ca	10 ± 2.2
Ce	< 9
Cu	48 ± 8.6
Fe	22 ± 6.7
La	< 2
Mn	0.23 ± 0.01
Ni	< 2
Si	561 ± 87.1
Sn	24 ± 7.8
Zn	$0.54 \pm 0.04\%$
Zr	< 6
Mg	$91.20 \pm 0.32\%$

A.3.2 Averaged RMSEs and number of features

Table A.2: At each step, the averaged RMSEs of the test and train sets corresponding to the probability with the lowest RMSE of the test set in the 5-fold cross validations are listed. The red values represent the averaged RMSEs of the optimized SVR and KRR models.

Num. features	SVR_test	SVR_train	KRR_test	KRR_train
25	60.61	7.75	60.61	7.72
24	58.79	7.73	58.76	7.71
23	57.56	7.61	57.13	7.72
22	56.57	7.60	56.11	7.98
21	55.64	7.62	55.00	7.96
20	54.66	7.63	53.94	8.01
19	52.49	7.74	52.75	8.38
18	51.58	7.77	51.61	8.42
17	50.81	7.68	50.63	8.58
16	49.98	7.71	49.54	9.01
15	49.57	7.74	48.15	9.71
14	49.08	7.74	47.08	9.86
13	47.54	7.84	46.62	10.57
12	46.76	7.88	46.19	10.75
11	46.56	7.84	45.36	10.89
10	46.74	7.89	46.21	12.54
9	45.90	8.47	47.61	14.22
8	45.20	9.35	46.80	18.55
7	45.98	10.02	46.40	22.15
6	48.97	21.01	49.44	30.72
5	50.45	23.31	56.61	23.34
4	49.94	26.53	57.98	30.23
3	47.08	29.59	64.62	38.09
2	56.50	50.11	66.24	44.06
1	63.94	63.12	72.83	70.32

A.3.3 Compounds containing =CHX fragment

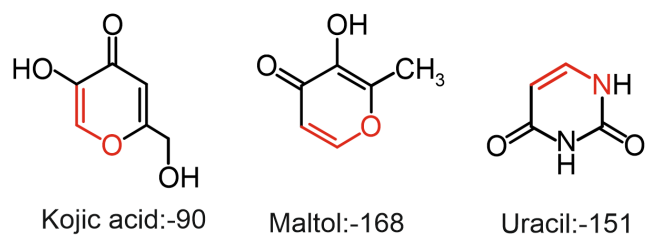
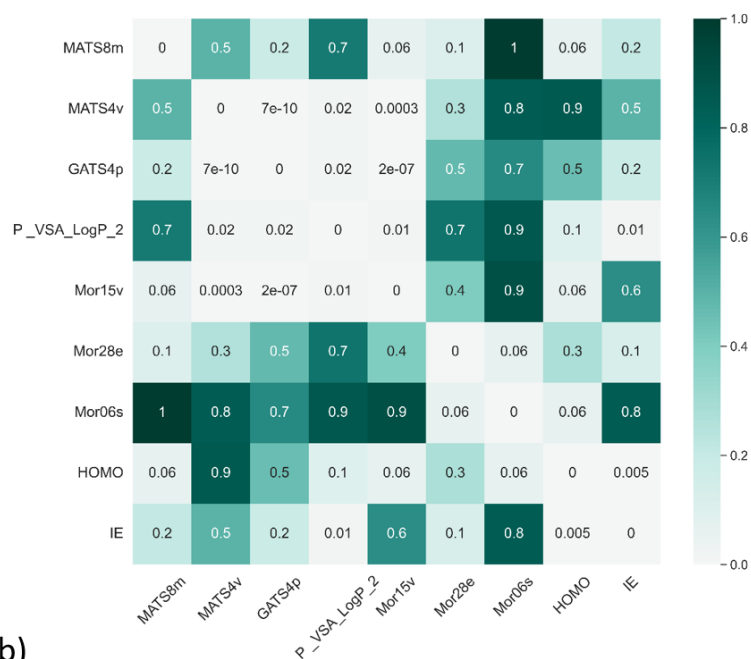


Figure A.6: Compounds in the used dataset containing the atom-centred fragment =CHX (highlighted in red) and corresponding IE efficiencies: Kojic acid, Maltol and Uracil from left to right.

A.3.4 p-value between features and target

(a)



(b)

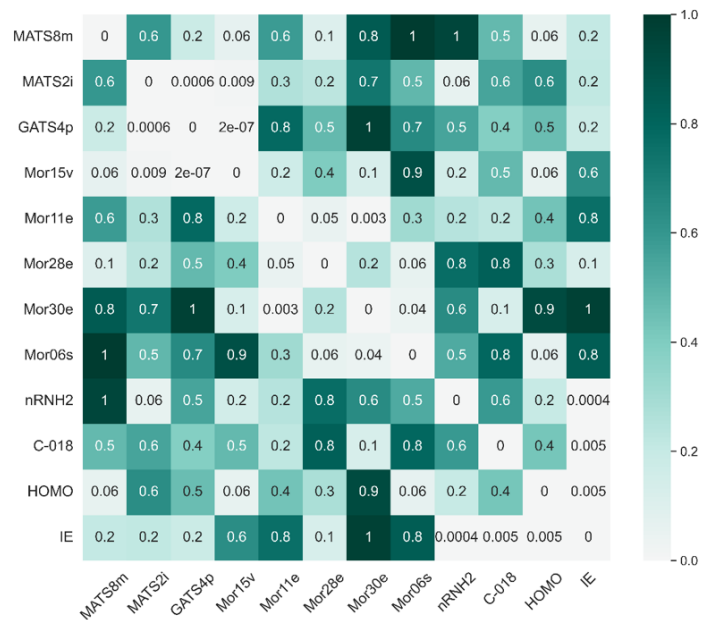


Figure A.7: p-value calculation for the two models. (a) p-value among the selected 8-tuple features for the SVR model and IEs. (b) p-value among the selected 11-tuple features for the KRR model and IEs.

A.3.5 Experimental database

Table A.3: Names and IEs of the 58 chemicals in the dataset extracted from Ref. [30]

Name	IE(%)
5-Amino-salicylic acid	80
5-Methyl-salicylic acid	56
5-Nitrobarbituric acid	71
5-Sulfosalicylic acid	-125
2,2'Bipyridyl	47
1,2,4-Triazole	-43
Acetic acid	60
Adipic acid	-48
Ascorbic acid	38
Benzoic acid	40
Bicine	-227
Citric acid	-78
L-Cysteine	-206
Diglycolic acid	54
Dodecylbenzene sulfonic acid	67
Formic acid	19
Fumaric acid	83
Glucose	29
L-Glutamic acid	-84
Glyceric acid	57
Glycerol	24
Glycine	-82
Glycolic acid	49
Kojic acid	-90
Maleic acid	26
Maltol	-168
Mandelic acid	11
Nicotinic acid	24
DL-Norleucine	-127
Oxalic acid	63
D-Panthenol	-27
DL-Phenylalanine	-149
Phthalic acid	33
Picolinic acid	45
Piperazine	-106
Propionic acid	34
Para-Toluic acid	-39

A.3 Screening for benign corrosion inhibitors

Pyrazole	31
2,3-Dipicolinic acid	47
2,5-Dipicolinic acid	68
Quinaldic acid	74
Quinic acid	37
Salicylaldehyde	18
Salicylhydroxamic acid	-177
Salicylic acid	26
Tartaric acid	52
Thiourea	57
Uracil	-151
Urea	-19
3,5-Dimethylpyrazole	-27
3,5-Dinitrosalicylic acid	15
3-Amino-1,2,4-triazole	19
3-Methyl-2-nitrobenzoic acid	25
3-Methylcatechol	-129
3-Methylsalicylic acid	52
4-Amino-salicylic acid	52
4-Methylsalicylic acid	9
4-Tert-butylbenzoic acid	-91

A.3.6 Learning curves

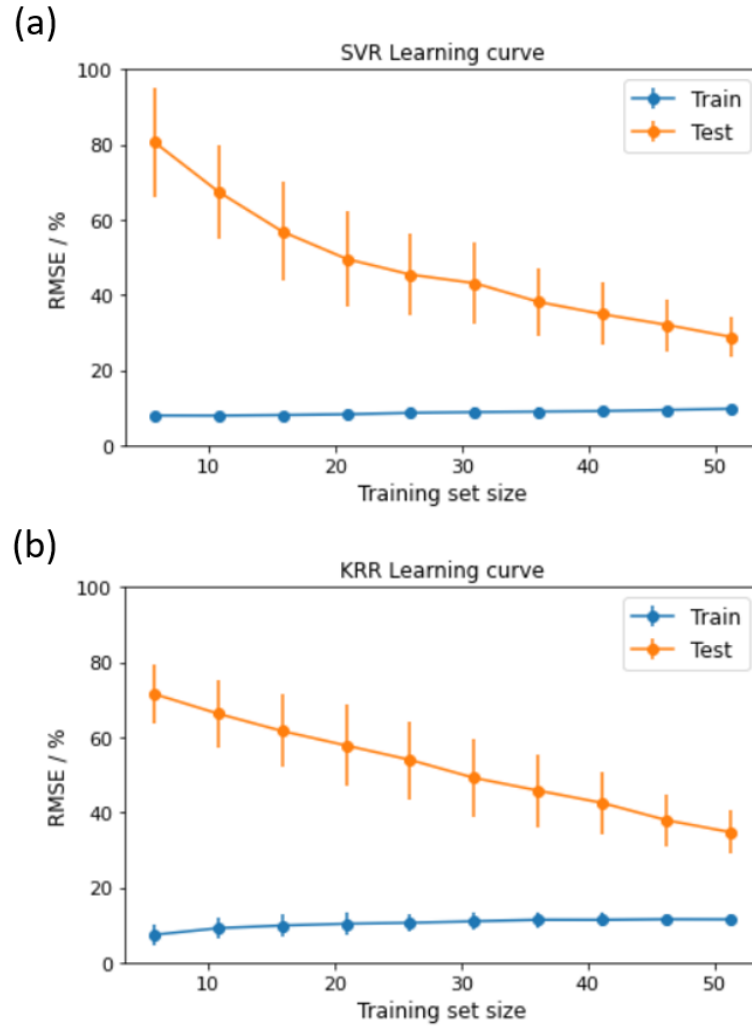


Figure A.8: Learning curves for the two models displaying the averaged RMSEs of the train and test sets as the training set size varies. **(a)** SVR model. **(b)** KRR model. Blue depicts the averaged RMSEs of the train sets and orange for the averaged RMSEs of the test sets in both figures.

Bibliography

- [1] Feister, T., Kim, H., Gwinn, A., Schiller, T. & Austin, M. Failure predictions in warm forming of 7075-t6 aluminum structural parts. In *IOP Conference Series: Materials Science and Engineering*, vol. 418, 012024 (IOP Publishing, 2018).
- [2] Scallan, P. *Process planning: the design/manufacture interface* (Elsevier, 2003).
- [3] Jing, C., Dong, B., Raza, A., Zhang, T. & Zhang, Y. Corrosion inhibition of layered double hydroxides for metal-based systems. *Nano Materials Science* **3**, 47–67 (2021).
- [4] Varshney, D. & Kumar, K. Application and use of different aluminium alloys with respect to workability, strength and welding parameter optimization. *Ain Shams Engineering Journal* **12**, 1143–1152 (2021).
- [5] Tan, J. & Ramakrishna, S. Applications of magnesium and its alloys: a review. *Applied Sciences* **11**, 6861 (2021).
- [6] Pater, Z. & Tomczak, J. Experimental tests for cross wedge rolling of forgings made from non-ferrous metal alloys/próby doświadczalne walcowania poprzeczno-klinowego odkuwek ze stopów metali nieżelaznych. *Archives of Metallurgy and Materials* (2012).
- [7] Mouritz, A. P. *Introduction to aerospace materials* (Elsevier, 2012).
- [8] Musfirah, A. & Jaharah, A. Magnesium and aluminum alloys in automotive industry. *Journal of Applied Sciences Research* **8**, 4865–4875 (2012).
- [9] Tabish, M. *et al.* Reviewing the current status of layered double hydroxide-based smart nanocontainers for corrosion inhibiting applications. *Journal of Materials Research and Technology* **10**, 390–421 (2021).
- [10] Hamilton, B. Aluminum cost profiles. *Unpublished Report, March* 26–30 (2013).
- [11] Wulandari, W., Brooks, G., Rhamdhani, M. & Monaghan, B. Magnesium: current and alternative production routes (2010).
- [12] Predko, P. *et al.* Promising methods for corrosion protection of magnesium alloys in the case of mg-al, mg-mn-ce and mg-zn-zr: A recent progress review. *Metals* **11**, 1133 (2021).

Bibliography

- [13] Snow, D. A. *Plant engineer's reference book* (Elsevier, 2001).
- [14] Tedim, J. *et al.* Layered double hydroxides for corrosion-related applications—main developments from 20 years of research at ciceco. *Frontiers in Chemistry* **14**, 12 (2022).
- [15] Kordas, G. Corrosion barrier coatings: Progress and perspectives of the chemical route. *Corrosion and Materials Degradation* **3**, 376–413 (2022).
- [16] Koch, G. *et al.* International measures of prevention, application, and economics of corrosion technologies study. *NACE international* **216**, 2–3 (2016).
- [17] Dong, Y., Wang, F. & Zhou, Q. Protective behaviors of 2-mercaptobenzothiazole intercalated zn–al-layered double hydroxide coating. *Journal of Coatings Technology and Research* **11**, 793–803 (2014).
- [18] Scharnagl, N., Blawert, C. & Dietzel, W. Corrosion protection of magnesium alloy az31 by coating with poly (ether imides)(pei). *Surface and Coatings Technology* **203**, 1423–1428 (2009).
- [19] Zheludkevich, M. *et al.* Active protection coatings with layered double hydroxide nanocontainers of corrosion inhibitor. *Corrosion Science* **52**, 602–611 (2010).
- [20] Zheludkevich, M. L., Shchukin, D. G., Yasakau, K. A., Möhwald, H. & Ferreira, M. G. Anticorrosion coatings with self-healing effect based on nanocontainers impregnated with corrosion inhibitor. *Chemistry of Materials* **19**, 402–411 (2007).
- [21] Buchheit, R. G., Guan, H., Mahajanam, S. & Wong, F. Active corrosion protection and corrosion sensing in chromate-free organic coatings. *Progress in Organic Coatings* **47**, 174–182 (2003).
- [22] Mahajanam, S. & Buchheit, R. Characterization of inhibitor release from zn–al-[v10o28] 6- hydrotalcite pigments and corrosion protection from hydrotalcite-pigmented epoxy coatings. *Corrosion* **64**, 230–240 (2008).
- [23] Pommiers, S., Frayret, J., Castetbon, A. & Potin-Gautier, M. Alternative conversion coatings to chromate for the protection of magnesium alloys. *Corrosion Science* **84**, 135–146 (2014).
- [24] Bouali, A. *et al.* Layered double hydroxides (ldhs) as functional materials for the corrosion protection of aluminum alloys: A review. *Applied Materials Today* **21**, 100857 (2020).
- [25] Gray, J. & Luan, B. Protective coatings on magnesium and its alloys—a critical review. *Journal of alloys and compounds* **336**, 88–113 (2002).
- [26] Xie, Y., Holmgren, S., Andrews, D. M. & Wolfe, M. S. Evaluating the impact of the us national toxicology program: A case study on hexavalent chromium. *Environmental health perspectives* **125**, 181–188 (2017).

-
- [27] Regulation, E. No 1907/2006 of the european parliament and of the council of 18 december 2006 concerning the registration. *Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 45*, 1–849 (1999).
- [28] Winkler, D. A. *et al.* Towards chromate-free corrosion inhibitors: structure–property models for organic alternatives. *Green Chemistry* **16**, 3349–3357 (2014).
- [29] Zhang, G. *et al.* Corrosion protection properties of different inhibitors containing peo/ldhs composite coating on magnesium alloy az31. *Scientific Reports* **11**, 1–14 (2021).
- [30] Lamaka, S. *et al.* Comprehensive screening of mg corrosion inhibitors. *Corrosion Science* **128**, 224–240 (2017).
- [31] Hu, H., Nie, X. & Ma, Y. Corrosion and surface treatment of magnesium alloys. *Magnesium Alloys-Properties in Solid and Liquid States* 67–108 (2014).
- [32] Latnikova, A. *Polymeric capsules for self-healing anticorrosion coatings*. Ph.D. thesis, Universität Potsdam (2012).
- [33] Wei, H. *et al.* Advanced micro/nanocapsules for self-healing smart anticorrosion coatings. *Journal of Materials Chemistry A* **3**, 469–480 (2015).
- [34] Javidparvar, A. A., Naderi, R., Ramezanzadeh, B. & Bahlakeh, G. Graphene oxide as a ph-sensitive carrier for targeted delivery of eco-friendly corrosion inhibitors in chloride solution: Experimental and theroretical investigations. *Journal of industrial and engineering chemistry* **72**, 196–213 (2019).
- [35] Javidparvar, A. A., Naderi, R. & Ramezanzadeh, B. Manipulating graphene oxide nanocontainer with benzimidazole and cerium ions: Application in epoxy-based nanocomposite for active corrosion protection. *Corrosion Science* **165**, 108379 (2020).
- [36] Fan, F., Zhou, C., Wang, X. & Szpunar, J. Layer-by-layer assembly of a self-healing anticorrosion coating on magnesium alloys. *ACS Applied Materials & Interfaces* **7**, 27271–27278 (2015).
- [37] Li, G. L., Zheng, Z., Moehwald, H. & Shchukin, D. G. Silica/polymer double-walled hybrid nanotubes: synthesis and application as stimuli-responsive nanocontainers in self-healing coatings. *ACS nano* **7**, 2470–2478 (2013).
- [38] Gao, L., He, J., Hu, J. & Wang, C. Photoresponsive self-healing polymer composite with photoabsorbing hybrid microcapsules. *ACS applied materials & interfaces* **7**, 25546–25552 (2015).
- [39] Yeole, K., Agarwal, I. & Mhaske, S. The effect of carbon nanotubes loaded with 2-mercaptobenzothiazole in epoxy-based coatings. *Journal of Coatings Technology and Research* **13**, 31–40 (2016).

Bibliography

- [40] Katariya, M. N., Jana, A. K. & Parikh, P. A. Corrosion inhibition effectiveness of zeolite zsm-5 coating on mild steel against various organic acids and its antimicrobial activity. *Journal of industrial and Engineering Chemistry* **19**, 286–291 (2013).
- [41] Rassouli, L., Naderi, R. & Mahdavian, M. Study of the impact of sequence of corrosion inhibitor doping in zeolite on the self-healing properties of silane sol–gel film. *Journal of Industrial and Engineering Chemistry* **66**, 221–230 (2018).
- [42] Ghazi, A., Ghasemi, E., Mahdavian, M., Ramezanzadeh, B. & Rostami, M. The application of benzimidazole and zinc cations intercalated sodium montmorillonite as smart ion exchange inhibiting pigments in the epoxy ester coating. *Corrosion Science* **94**, 207–217 (2015).
- [43] Hosseini, M., Jafari, M. & Najjar, R. Effect of polyaniline–montmorillonite nanocomposite powders addition on corrosion performance of epoxy coatings on al 5000. *Surface and Coatings Technology* **206**, 280–286 (2011).
- [44] Hosseini, M., Raghbi-Boroujeni, M., Ahadzadeh, I., Najjar, R. & Dorraji, M. S. Effect of polypyrrole–montmorillonite nanocomposites powder addition on corrosion performance of epoxy coatings on al 5000. *Progress in organic coatings* **66**, 321–327 (2009).
- [45] Anjum, M. J. *et al.* In-situ intercalation of 8-hydroxyquinoline in mg-al ldh coating to improve the corrosion resistance of az31. *Corrosion Science* **157**, 1–10 (2019).
- [46] Hayatdavoudi, H. & Rahsepar, M. Smart inhibition action of layered double hydroxide nanocontainers in zinc-rich epoxy coating for active corrosion protection of carbon steel substrate. *Journal of Alloys and Compounds* **711**, 560–567 (2017).
- [47] Asl, V. Z. *et al.* The effect of cerium cation on the microstructure and anti-corrosion performance of ldh conversion coatings on az31 magnesium alloy. *Journal of Alloys and Compounds* **821**, 153248 (2020).
- [48] Tedim, J. *et al.* Enhancement of active corrosion protection via combination of inhibitor-loaded nanocontainers. *ACS applied materials & interfaces* **2**, 1528–1535 (2010).
- [49] Li, D. *et al.* Anticorrosion organic coating with layered double hydroxide loaded with corrosion inhibitor of tungstate. *Progress in organic coatings* **71**, 302–309 (2011).
- [50] Yu, X. *et al.* One-step synthesis of lamellar molybdate pillared hydrotalcite and its application for az31 mg alloy protection. *Solid State Sciences* **11**, 376–381 (2009).
- [51] Poznyak, S. *et al.* Novel inorganic host layered double hydroxides intercalated with guest organic inhibitors for anticorrosion applications. *ACS Applied Materials & Interfaces* **1**, 2353–2362 (2009).
- [52] Zhang, F. *et al.* Corrosion resistance of superhydrophobic layered double hydroxide films on aluminum. *Angewandte Chemie* **120**, 2500–2503 (2008).

- [53] Wang, X. & Wang, D.-Y. Fire-retardant polylactic acid-based materials: Preparation, properties, and mechanism. In *Novel fire retardant polymers and composite materials*, 93–116 (Elsevier, 2017).
- [54] Yan, L. & Chen, X. Nanomaterials for drug delivery. In *Nanocrystalline materials*, 221–268 (Elsevier, 2014).
- [55] Daniel, S. & Thomas, S. Layered double hydroxides: fundamentals to applications. In *Layered double hydroxide polymer nanocomposites*, 1–76 (Elsevier, 2020).
- [56] Gaines, R. V., Dana, J. D. & Dana, E. S. *Dana's new mineralogy: The system of mineralogy of James Dwight Dana and Edward Salisbury Dana* (Wiley, 1997).
- [57] Mishra, G., Dash, B. & Pandey, S. Layered double hydroxides: A brief review from fundamentals to application as evolving biomaterials. *Applied Clay Science* **153**, 172–186 (2018).
- [58] Parida, K. & Mohapatra, L. Carbonate intercalated zn/fe layered double hydroxide: a novel photocatalyst for the enhanced photo degradation of azo dyes. *Chemical Engineering Journal* **179**, 131–139 (2012).
- [59] Barahuie, F., Hussein, M. Z., Abd Gani, S., Fakurazi, S. & Zainal, Z. Synthesis of protocatechuic acid–zinc/aluminium–layered double hydroxide nanocomposite as an anticancer nanodelivery system. *Journal of Solid State Chemistry* **221**, 21–31 (2015).
- [60] Wijitwongwan, R., Intasa-ard, S. & Ogawa, M. Preparation of layered double hydroxides toward precisely designed hierarchical organization. *ChemEngineering* **3**, 68 (2019).
- [61] Nestroinaia, O. V., Ryltsova, I. G. & Lebedeva, O. E. Effect of synthesis method on properties of layered double hydroxides containing ni (iii). *Crystals* **11**, 1429 (2021).
- [62] Das, J. & Parida, K. Heteropoly acid intercalated zn/al htlc as efficient catalyst for esterification of acetic acid using n-butanol. *Journal of Molecular Catalysis A: Chemical* **264**, 248–254 (2007).
- [63] Ryu, S.-J., Jung, H., Oh, J.-M., Lee, J.-K. & Choy, J.-H. Layered double hydroxide as novel antibacterial drug delivery system. *Journal of Physics and Chemistry of Solids* **71**, 685–688 (2010).
- [64] Aisawa, S. *et al.* Intercalation behavior of amino acids into zn–al-layered double hydroxide by calcination–rehydration reaction. *Journal of Solid State Chemistry* **177**, 3987–3994 (2004).
- [65] Erickson, K. L., Bostrom, T. E. & Frost, R. L. A study of structural memory effects in synthetic hydrotalcites using environmental sem. *Materials Letters* **59**, 226–229 (2005).

Bibliography

- [66] Haraketi, M., Hosni, K. & Srasra, E. Intercalation of salicylic acid into znal and mgal layered double hydroxides for a controlled release formulation. *Colloid Journal* **78**, 533–541 (2016).
- [67] Prevot, V., Forano, C., Besse, J. & Abraham, F. Syntheses and thermal and chemical behaviors of tartrate and succinate intercalated zn₃al and zn₂cr layered double hydroxides. *Inorganic Chemistry* **37**, 4293–4301 (1998).
- [68] Rives, V., del Arco, M. & Martín, C. Intercalation of drugs in layered double hydroxides and their controlled release: A review. *Applied clay science* **88**, 239–269 (2014).
- [69] Mishra, G., Dash, B., Pandey, S., Sethi, D. & Kumar, C. G. Comparative evaluation of synthetic routes and antibacterial/antifungal properties of zn–al layered double hydroxides containing benzoate anion. *Environmental Engineering Science* **35**, 247–260 (2018).
- [70] George, G. & Saravanakumar, M. Synthesising methods of layered double hydroxides and its use in the fabrication of dye sensitised solar cell (dssc): A short review. In *IOP Conference Series: Materials Science and Engineering*, vol. 263, 032020 (IOP Publishing, 2017).
- [71] Kameliya, J. *et al.* Layered double hydroxide materials: A review on their preparation, characterization, and applications. *Inorganics* **11**, 121 (2023).
- [72] Radha, A., Kamath, P. V. & Shivakumara, C. Mechanism of the anion exchange reactions of the layered double hydroxides (ldhs) of ca and mg with al. *Solid State Sciences* **7**, 1180–1187 (2005).
- [73] Kooli, E., Holgado, M., Rives, V., Sanroman, S. & Ulibarri, M. A simple conductivity study of decavanadate intercalation in hydrotalcite. *Materials research bulletin* **32**, 977–982 (1997).
- [74] Mikami, N., Sasaki, M., Horibe, S. & Yasunaga, T. Kinetic study of ion exchange of phosphate ions for chloride in a hydrotalcite-like compound, ni_{0.75}al_{0.25}(oh)₂cl_{0.25}.mh₂o, using the pressure-jump technique. *The Journal of Physical Chemistry* **88**, 1716–1719 (1984).
- [75] Badreddine, M., Legrouri, A., Barroug, A., De Roy, A. & Besse, J.-P. Influence of ph on phosphate intercalation in zinc-aluminum layered double hydroxide. *Collection of Czechoslovak chemical communications* **63**, 741–748 (1998).
- [76] Perez-Sanchez, G., Galvão, T., Tedim, J. & Gomes, J. A molecular dynamics framework to explore the structure and dynamics of layered double hydroxides. *Applied Clay Science* **163**, 164–177 (2018).
- [77] Poznyak, S. *et al.* Novel inorganic host layered double hydroxides intercalated with guest organic inhibitors for anticorrosion applications. *ACS applied materials & interfaces* **1**, 2353–62 (2009).

- [78] Newman, S., Williams, S., Coveney, P. & Jones, W. Interlayer arrangement of hydrated mg/al layered double hydroxides containing guest terephthalate anions: Comparison of simulation and measurement. *Journal of Physical Chemistry B - J PHYS CHEM B* **102** (1998).
- [79] Sideris, P. J., Nielsen, U. G., Gan, Z. & Grey, C. P. Mg/al ordering in layered double hydroxides revealed by multinuclear nmr spectroscopy. *Science* **321**, 113–117 (2008).
- [80] Zhang, Y., Xu, H. & Lu, S. Preparation and application of layered double hydroxide nanosheets. *RSC advances* **11**, 24254–24281 (2021).
- [81] Silva, C. G., Bouizi, Y., Fornés, V. & García, H. Layered double hydroxides as highly efficient photocatalysts for visible light oxygen generation from water. *Journal of the American Chemical Society* **131**, 13833–13839 (2009).
- [82] Gunjakar, J. L., Kim, T. W., Kim, H. N., Kim, I. Y. & Hwang, S.-J. Mesoporous layer-by-layer ordered nanohybrids of layered double hydroxide and layered metal oxide: highly active visible light photocatalysts with improved chemical stability. *Journal of the American Chemical Society* **133**, 14998–15007 (2011).
- [83] Parida, K., Satpathy, M. & Mohapatra, L. Incorporation of fe 3+ into mg/al layered double hydroxide framework: effects on textural properties and photocatalytic activity for h 2 generation. *Journal of Materials Chemistry* **22**, 7350–7357 (2012).
- [84] Kim, S. J., Lee, Y., Lee, D. K., Lee, J. W. & Kang, J. K. Efficient co-fe layered double hydroxide photocatalysts for water oxidation under visible light. *Journal of Materials Chemistry A* **2**, 4136–4139 (2014).
- [85] Chaara, D. *et al.* Removal of nitrophenol pesticides from aqueous solutions by layered double hydroxides and their calcined products. *Applied Clay Science* **50**, 292–298 (2010).
- [86] Wang, Q. *et al.* Morphology and composition controllable synthesis of mg-al-co3 hydrotalcites by tuning the synthesis ph and the co2 capture capacity. *Applied Clay Science* **55**, 18–26 (2012).
- [87] Pshinko, G. Layered double hydroxides as effective adsorbents for u (vi) and toxic heavy metals removal from aqueous media. *Journal of Chemistry* **2013** (2013).
- [88] Wong, Y. *et al.* Efficient delivery of sirna to cortical neurons using layered double hydroxide nanoparticles. *Biomaterials* **31**, 8770–8779 (2010).
- [89] Li, S. *et al.* Cellular uptake and gene delivery using layered double hydroxide nanoparticles. *Journal of Materials Chemistry B* **1**, 61–68 (2013).
- [90] Thyveetil, M.-A., Coveney, P. V., Greenwell, H. C. & Suter, J. L. Computer simulation study of the structural stability and materials properties of dna-intercalated layered double hydroxides. *Journal of the American Chemical Society* **130**, 4742–4756 (2008).

Bibliography

- [91] Erastova, V., Degiacomi, M. T., Fraser, D. G. & Greenwell, H. C. Mineral surface chemistry control for origin of prebiotic peptides. *Nature communications* **8**, 1–9 (2017).
- [92] Grégoire, B. *et al.* Insights into the behaviour of biomolecules on the early earth: The concentration of aspartate by layered double hydroxide minerals. *Geochimica et Cosmochimica Acta* **176**, 239–258 (2016).
- [93] Liu, H.-M., Zhao, X.-J., Zhu, Y.-Q. & Yan, H. Dft study on mgal-layered double hydroxides with different interlayer anions: structure, anion exchange, host–guest interaction and basic sites. *Physical Chemistry Chemical Physics* **22**, 2521–2529 (2020).
- [94] Rives, V., Del Arco, M. & Martín, C. Layered double hydroxides as drug carriers and for controlled release of non-steroidal antiinflammatory drugs (nsaids): a review. *Journal of Controlled Release* **169**, 28–39 (2013).
- [95] Hou, X., Bish, D. L., Wang, S.-L., Johnston, C. T. & Kirkpatrick, R. J. Hydration, expansion, structure, and dynamics of layered double hydroxides. *American Mineralogist* **88**, 167–179 (2003).
- [96] Hibino, T. Anion selectivity of layered double hydroxides: effects of crystallinity and charge density. *European Journal of Inorganic Chemistry* **2018**, 722–730 (2018).
- [97] Plank, J., Dai, Z. & Andres, P. Preparation and characterization of new ca–al–polycarboxylate layered double hydroxides. *Materials Letters* **60**, 3614–3617 (2006).
- [98] Monteiro, B. *et al.* Investigation of layered double hydroxides intercalated by oxomolybdenum catecholate complexes. *Inorganic chemistry* **47**, 8674–8686 (2008).
- [99] Richetta, M. *et al.* Preparation, intercalation, and characterization of nanostructured (zn, al) layered double hydroxides (ldhs). *Surface and Interface Analysis* **50**, 1094–1098 (2018).
- [100] Miyata, S. Anion-exchange properties of hydrotalcite-like compounds. *Clays and Clay minerals* **31**, 305–311 (1983).
- [101] Allada, R. k., Navrotsky, A., Berbeco, H. T. & Casey, W. H. Thermochemistry and aqueous solubilities of hydrotalcite-like solids. *Science* **296**, 721–723 (2002).
- [102] Peltier, E. *et al.* Calorimetric determination of the enthalpies of formation of hydrotalcite-like solids and their use in the geochemical modeling of metals in natural waters. *Clays and Clay Minerals* **54**, 409–417 (2006).
- [103] Prasad, B. E., Kamath, P. V. & Vijayamohanan, K. Anion exchange reaction potentials as approximate estimates of the relative thermodynamic stabilities of mg/al layered double hydroxides containing different anions. *Langmuir* **27**, 13539–13543 (2011).

- [104] Bouali, A. C. *et al.* Zn-al ldh growth on aa2024 and zinc and their intercalation with chloride: Comparison of crystal structure and kinetics. *Applied Surface Science* **501**, 144027 (2020).
- [105] Palin, L., Milanesio, M., van Beek, W. & Conterosito, E. Understanding the ion exchange process in ldh nanomaterials by fast in situ xrpd and pca-assisted kinetic analysis. *Journal of Nanomaterials* **2019** (2019).
- [106] Tedim, J. *et al.* Zn–al layered double hydroxides as chloride nanotraps in active protective coatings. *Corrosion Science* **55**, 1–4 (2012).
- [107] Dwiasi, D. W., Mudasir, M. & Roto, R. Ion exchange of benzoate in ni-al-benzoate layered double hydroxide by amoxicillin. *Open Chemistry* **17**, 1043–1049 (2019).
- [108] Frost, R. L. *et al.* Thermal decomposition of hydrotalcite with chromate, molybdate or sulphate in the interlayer. *Thermochimica Acta* **429**, 179–187 (2005).
- [109] Theiss, F. L., Ayoko, G. A. & Frost, R. L. Thermogravimetric analysis of selected layered double hydroxides. *Journal of thermal analysis and calorimetry* **112**, 649–657 (2013).
- [110] Sotiles, A. R., Gomez, N. A. G. & Wypych, F. Thermogravimetric analysis of layered double hydroxides intercalated with sulfate and alkaline cations $[m62+ al3 (oh) 18][a+ (so4) 2] 12h2o$ ($m2+= mn, mg, zn$; $a+= li, na, k$). *Journal of Thermal Analysis and Calorimetry* **140**, 1715–1723 (2020).
- [111] Dobrea, I. D. *et al.* Raman spectroscopy—useful tool for studying the catalysts derived from mo and v-oxyanion-intercalated layered double hydroxides. *Applied Clay Science* **104**, 205–210 (2015).
- [112] Chagas, L. H., Chiaro, S. S. X., Leitão, A. A. & Diniz, R. Raman spectroscopy for characterization of hydrotalcite-like materials used in catalytic reactions. *Recent Developments in Atomic Force Microscopy and Raman Spectroscopy for Materials Characterization* 133 (2021).
- [113] Winsberg, E. Computer Simulations in Science. In Zalta, E. N. & Nodelman, U. (eds.) *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2022), Winter 2022 edn.
- [114] Dove, M. T. An introduction to atomistic simulation methods. *Seminarios de la SEM* **4**, 7–37 (2008).
- [115] Wimmer, E. The growing importance of computations in materials science. current capabilities and perspectives. *MATERIALS SCIENCE-WROCLAW-* **23**, 325 (2005).
- [116] Ceriotti, M. Unsupervised machine learning in atomistic simulations, between predictions and understanding. *The Journal of chemical physics* **150**, 150901 (2019).

Bibliography

- [117] Kondov, I. & Sutmann, G. *Multiscale modelling methods for applications in materials science: CECAM tutorial, 16-20 September 2013, Forschungszentrum Jülich; lecture notes*, vol. 19 (Forschungszentrum Jülich, 2013).
- [118] Schönemann, M., Bockholt, H., Thiede, S., Kwade, A. & Herrmann, C. Multiscale simulation approach for production systems. *The International Journal of Advanced Manufacturing Technology* **102**, 1373–1390 (2019).
- [119] Wieser, C., Prill, T. & Schladitz, K. Multiscale simulation process and application to additives in porous composite battery electrodes. *Journal of Power Sources* **277**, 64–75 (2015).
- [120] Eidel, B., Hartmaier, A. & Gumbsch, P. Atomistic simulation methods and their application on fracture. In *Multiscale Modelling of Plasticity and Fracture by Means of Dislocation Mechanics*, 1–57 (Springer, 2010).
- [121] Talirz, L., Ghiringhelli, L. M. & Smit, B. Trends in atomistic simulation software usage. *arXiv preprint arXiv:2108.12350* (2021).
- [122] Liew, K. M., Jianwei, Y. & Zhang, L.-W. *Mechanical behaviors of carbon nanotubes: theoretical and numerical approaches* (William Andrew, 2016).
- [123] Nguyen, V. P., Stroeve, M. & Sluys, L. J. Multiscale continuous and discontinuous modeling of heterogeneous materials: a review on recent developments. *Journal of Multiscale Modelling* **3**, 229–270 (2011).
- [124] Lu, G. & Kaxiras, E. An overview of multiscale simulations of materials. *arXiv preprint cond-mat/0401073* (2004).
- [125] Peng, G. C. *et al.* Multiscale modeling meets machine learning: What can we learn? *Archives of Computational Methods in Engineering* **28**, 1017–1037 (2021).
- [126] Alber, M. *et al.* Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ digital medicine* **2**, 115 (2019).
- [127] Gormley, A. J. & Webb, M. A. Machine learning in combinatorial polymer chemistry. *Nature Reviews Materials* **6**, 642–644 (2021).
- [128] Hart, G. L., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. *Nature Reviews Materials* **6**, 730–755 (2021).
- [129] Friederich, P., Häse, F., Proppe, J. & Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nature Materials* **20**, 750–761 (2021).
- [130] Zaverkin, V., Holzmüller, D., Steinwart, I. & Kaestner, J. Fast and sample-efficient interatomic neural network potentials for molecules and materials based on gaussian moments. *Journal of Chemical Theory and Computation* **17**, 6658–6670 (2021).

-
- [131] Magana, A. J. Modeling and simulation in engineering education: A learning progression. *Journal of Professional Issues in Engineering Education and Practice* **143**, 04017008 (2017).
- [132] Yin, C. & McKay, A. Introduction to modeling and simulation techniques. In *Proceedings of ISCIIA 2018 and ITCA 2018* (Leeds, 2018).
- [133] Kim, N., Kim, Y., Tsotsis, T. T. & Sahimi, M. Atomistic simulation of nanoporous layered double hydroxide materials and their properties. i. structural modeling. *The Journal of chemical physics* **122**, 214713 (2005).
- [134] Ozaki, H., Tada, K. & Kiyobayashi, T. Monte-carlo simulation combined with density functional theory to investigate the equilibrium thermodynamics of electrode materials: lithium titanates as model compounds. *Physical Chemistry Chemical Physics* **21**, 15551–15559 (2019).
- [135] Irigoyen, M., Matxain, J. M. & Ruipérez, F. Combined dft and md simulation protocol to characterize self-healing properties in disulfide-containing materials: Polyurethanes and polymethacrylates as case studies. *Front. Mater.* **9**: 859482. doi: 10.3389/fmats (2022).
- [136] Altundal, O. F., Haslak, Z. P. & Keskin, S. Combined gcmc, md, and dft approach for unlocking the performances of cofs for methane purification. *Industrial & engineering chemistry research* **60**, 12999–13012 (2021).
- [137] Zhao, X.-J. *et al.* Anion exchange behavior of m ii al layered double hydroxides: a molecular dynamics and dft study. *Physical Chemistry Chemical Physics* **22**, 19758–19768 (2020).
- [138] Zhang, S.-T. *et al.* Dft-based simulation and experimental validation of the topotactic transformation of mgal layered double hydroxides. *ChemPhysChem* **17**, 2754–2766 (2016).
- [139] Shandilya, A., Schadler, L. S. & Sundararaman, R. First-principles identification of localized trap states in polymer nanocomposite interfaces. *Journal of Materials Research* **35**, 931–939 (2020).
- [140] Costa, D. G. *et al.* Structural model proposition and thermodynamic and vibrational analysis of hydrotalcite-like compounds by dft calculations. *The Journal of Physical Chemistry C* **114**, 14133–14140 (2010).
- [141] Tavares, S. R., Wypych, F. & Leitão, A. A. Dft study of the intercalation of layered double hydroxides and layered hydroxide salts with dodecylsulfate and dodecylbenzene sulfonate: Exfoliation and hydration properties. *Applied Clay Science* **143**, 107–114 (2017).

Bibliography

- [142] Moraes, P. I. R., Tavares, S. R., Vaiss, V. S. & Leitão, A. A. Ab initio study of layered double hydroxides containing iron and its potential use as fertilizer. *The Journal of Physical Chemistry C* **120**, 9965–9974 (2016).
- [143] Zhang, Y., Cheng, X., Wu, C., Köhler, J. & Deng, S. Electronic structure and lithium diffusion in $\text{LiAl}_2(\text{OH})_6\text{Cl}$ studied by first principle calculations. *Molecules* **24**, 2667 (2019).
- [144] Singh, P. & Harbola, M. K. Density-functional theory of material design: fundamentals and applications-i. *Oxford Open Materials Science* **1**, itab018 (2021).
- [145] Iftimie, R., Minar, P. & Tuckerman, M. E. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proceedings of the National Academy of Sciences* **102**, 6654–6659 (2005).
- [146] Tran, N. V., Tieu, A. K. & Zhu, H. First-principles molecular dynamics study on the surface chemistry and nanotribological properties of MgAl layered double hydroxides. *Nanoscale* **13**, 5014–5025 (2021).
- [147] Kohn, W., Becke, A. D. & Parr, R. G. Density functional theory of electronic structure. *The Journal of Physical Chemistry* **100**, 12974–12980 (1996).
- [148] Baseden, K. A. & Tye, J. W. Introduction to density functional theory: calculations by hand on the helium atom. *Journal of Chemical Education* **91**, 2116–2123 (2014).
- [149] Van Mourik, T., Bühl, M. & Gageot, M.-P. Density functional theory across chemistry, physics and biology (2014).
- [150] Sham, L. J. & Kohn, W. One-particle properties of an inhomogeneous interacting electron gas. *Physical Review* **145**, 561 (1966).
- [151] Giustino, F. *Materials modelling using density functional theory: properties and predictions* (Oxford University Press, 2014).
- [152] Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **140**, A1133 (1965).
- [153] Girvin, S. M. & Yang, K. *Modern condensed matter physics* (Cambridge University Press, 2019).
- [154] Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical review* **136**, B864 (1964).
- [155] Sahni, V., Bohnen, K.-P. & Harbola, M. K. Analysis of the local-density approximation of density-functional theory. *Physical Review A* **37**, 1895 (1988).
- [156] Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Physical review letters* **45**, 566 (1980).

-
- [157] Burke, K. Perspective on density functional theory. *The Journal of chemical physics* **136**, 150901 (2012).
- [158] Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **77**, 3865 (1996).
- [159] Perdew, J. P. *et al.* Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B* **48**, 4978 (1993).
- [160] Johnson, B. G., Gill, P. M. & Pople, J. A. The performance of a family of density functional methods. *The Journal of chemical physics* **98**, 5612–5626 (1993).
- [161] García, A., Elsässer, C., Zhu, J., Louie, S. G. & Cohen, M. L. Use of gradient-corrected functionals in total-energy calculations for solids. *Physical Review B* **46**, 9829 (1992).
- [162] Bloch, F. Quantum mechanics of electrons in crystal lattices. *Z. Phys* **52**, 555–600 (1928).
- [163] Meyer, B. The pseudopotential plane wave approach. *NIC Series* **31**, 71 (2006).
- [164] Khosravi, M., Murthy, V. & Mackinnon, I. D. Evaluation of dft methods to calculate structure and partial atomic charges for zeolite n. *Computational Materials Science* **171**, 109225 (2020).
- [165] Xu, S.-M. *et al.* Theoretical and experimental study on miimiii-layered double hydroxides as efficient photocatalysts toward oxygen evolution from water. *The Journal of Physical Chemistry C* **119**, 18823–18834 (2015).
- [166] Tavares, S. R., Nangoi, I. M. & Leitão, A. A. Computational investigation of two-dimensional ldhs and the modification of their electronic structure induced by defects. *Applied Surface Science* **532**, 147159 (2020).
- [167] Thomas, G. S. & Kamath, P. V. Line broadening in the pxrd patterns of layered hydroxides: The relative effects of crystallite size and structural disorder. *Journal of Chemical Sciences* **118**, 127–133 (2006).
- [168] Galvao, T. L. *et al.* How density functional theory surface energies may explain the morphology of particles, nanosheets, and conversion films based on layered double hydroxides. *The Journal of Physical Chemistry C* **121**, 2211–2220 (2017).
- [169] Costa, D. G., Rocha, A. B., Souza, W. F., Chiaro, S. S. X. & Leitão, A. A. Comparative structural, thermodynamic and electronic analyses of znalan- hydrotalcite-like compounds (an- cl-, f-, br-, oh-, co32- or no3-): An ab initio study. *Applied clay science* **56**, 16–22 (2012).
- [170] Tavares, S. R., Moraes, P. I. R. & Leitão, A. A. Dft and force-field based md simulations of formamide intercalation in ldh and its exfoliated form. *Applied Surface Science* **552**, 149450 (2021).

Bibliography

- [171] Thyveetil, M.-A. *Large-scale simulations of layered double hydroxide nanocomposite materials* (University of London, University College London (United Kingdom), 2008).
- [172] Qin, W., Li, X., Bian, W.-W., Fan, X.-J. & Qi, J.-Y. Density functional theory calculations and molecular dynamics simulations of the adsorption of biomolecules on graphene surfaces. *Biomaterials* **31**, 1007–1016 (2010).
- [173] Kumar, P. P., Kalinichev, A. G. & Kirkpatrick, R. J. Molecular dynamics simulation of the energetics and structure of layered double hydroxides intercalated with carboxylic acids. *The Journal of Physical Chemistry C* **111**, 13517–13523 (2007).
- [174] Ruengkajorn, K., Erastova, V., Buffet, J.-C., Greenwell, H. C. & O'Hare, D. Aqueous immiscible layered double hydroxides: synthesis, characterisation and molecular dynamics simulation. *Chemical Communications* **54**, 4394–4397 (2018).
- [175] Erastova, V., Degiacomi, M. T., O'Hare, D. & Greenwell, H. C. Understanding surface interactions in aqueous miscible organic solvent treated layered double hydroxides. *RSC advances* **7**, 5076–5083 (2017).
- [176] Wang, J., Kalinichev, A. G., Amonette, J. E. & Kirkpatrick, R. J. Interlayer structure and dynamics of cl-bearing hydrotalcite: far infrared spectroscopy and molecular dynamics modeling. *American Mineralogist* **88**, 398–409 (2003).
- [177] Kirkpatrick, R., Kalinichev, A. & Wang, J. Molecular dynamics modelling of hydrated mineral interlayers and surfaces: structure and dynamics. *Mineralogical Magazine* **69**, 289–308 (2005).
- [178] Chen, M., Zhu, R., Lu, X., Zhu, J. & He, H. Influences of cation ratio, anion type, and water content on polytypism of layered double hydroxides. *Inorganic chemistry* **57**, 7299–7313 (2018).
- [179] Murthy, V., Smith, H. D., Zhang, H. & Smith, S. C. Molecular modeling of hydrotalcite structure intercalated with transition metal oxide anions: Cro42–and vo43–. *The Journal of Physical Chemistry A* **115**, 13673–13683 (2011).
- [180] Tsukanov, A. & Psakhie, S. G. Energy and structure of bonds in the interaction of organic anions with layered double hydroxide nanosheets: A molecular dynamics study. *Scientific reports* **6**, 1–8 (2016).
- [181] Cygan, R. T., Greathouse, J. A., Heinz, H. & Kalinichev, A. G. Molecular models and simulations of layered materials. *Journal of Materials Chemistry* **19**, 2470–2481 (2009).
- [182] Paquet, E. & Viktor, H. L. Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review. *BioMed research international* **2015** (2015).
- [183] Chen, J. The development and comparison of molecular dynamics simulation and monte carlo simulation. In *IOP Conference Series: Earth and Environmental Science*, vol. 128, 012110 (IOP Publishing, 2018).

- [184] González, M. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **12**, 169–200 (2011).
- [185] Cygan, R. T., Liang, J.-J. & Kalinichev, A. G. Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *The Journal of Physical Chemistry B* **108**, 1255–1266 (2004).
- [186] Filippova, V., Kunavin, S. & Pugachev, M. Calculation of the parameters of the lennard-jones potential for pairs of identical atoms based on the properties of solid substances. *Inorganic Materials: Applied Research* **6**, 1–4 (2015).
- [187] Limas, N. G. & Manz, T. A. Introducing ddec6 atomic population analysis: part 4. efficient parallel computation of net atomic charges, atomic spin moments, bond orders, and more. *RSC advances* **8**, 2678–2707 (2018).
- [188] Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *Journal of computational chemistry* **25**, 1656–1676 (2004).
- [189] Vega, C., Abascal, J. L., Conde, M. & Aragones, J. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday discussions* **141**, 251–276 (2009).
- [190] Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).
- [191] Allen, M. P. *et al.* Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins* **23**, 1–28 (2004).
- [192] Sri Harish, M. & Patra, P. K. Temperature and its control in molecular dynamics simulations. *Molecular Simulation* **47**, 701–729 (2021).
- [193] Faller, R. & De Pablo, J. J. Constant pressure hybrid molecular dynamics–monte carlo simulations. *The Journal of Chemical Physics* **116**, 55–59 (2002).
- [194] Kalinichev, A. G., Padma Kumar, P. & James Kirkpatrick, R. Molecular dynamics computer simulations of the effects of hydrogen bonding on the properties of layered double hydroxides intercalated with organic acids. *Philosophical Magazine* **90**, 2475–2488 (2010).
- [195] Smith, D. E. Molecular computer simulations of the swelling properties and interlayer structure of cesium montmorillonite. *Langmuir* **14**, 5959–5967 (1998).
- [196] Zhang, L., Lu, X., Liu, X., Zhou, J. & Zhou, H. Hydration and mobility of interlayer ions of (na x, ca y)-montmorillonite: a molecular dynamics study. *The Journal of Physical Chemistry C* **118**, 29811–29821 (2014).

Bibliography

- [197] Wang, J., Kalinichev, A. G. & Kirkpatrick, R. J. Effects of substrate structure and composition on the structure, dynamics, and energetics of water at mineral surfaces: A molecular dynamics modeling study. *Geochimica et cosmochimica acta* **70**, 562–582 (2006).
- [198] Peng, C. *et al.* Molecular dynamics simulation of nh₄-montmorillonite interlayer hydration: Structure, energetics, and dynamics. *Applied Clay Science* **195**, 105657 (2020).
- [199] Jorgensen, W. L. & Tirado-Rives, J. Monte carlo vs molecular dynamics for conformational sampling. *The Journal of Physical Chemistry* **100**, 14508–14513 (1996).
- [200] Ulmschneider, J. P., Ulmschneider, M. B. & Di Nola, A. Monte carlo vs molecular dynamics for all-atom polypeptide folding simulations. *The Journal of Physical Chemistry B* **110**, 16733–16742 (2006).
- [201] Reimer, U., Wahab, M., Schiller, P. & Mögel, H.-J. Monte carlo simulation of the adsorption equilibrium of a model surfactant solution on hydrophilic solid surfaces. *Langmuir* **17**, 8444–8450 (2001).
- [202] Zehl, T., Wahab, M., Schiller, P. & Moegel, H.-J. Monte carlo simulation of surfactant adsorption on hydrophilic surfaces. *Langmuir* **25**, 2090–2100 (2009).
- [203] Liu, L., Yang, X. & Xu, Z. Gibbs ensemble monte carlo simulation of adsorption for model surfactant solution in confined slit pores. *The Journal of chemical physics* **128**, 184712 (2008).
- [204] Coudert, F.-X. & Fuchs, A. H. Computational characterization and prediction of metal–organic framework properties. *Coordination Chemistry Reviews* **307**, 211–236 (2016).
- [205] Al-Jadir, T. M. & Siperstein, F. R. Monte carlo simulation of adsorption of polar and nonpolar gases in (fp) yeu metal–organic framework. *Journal of Chemical & Engineering Data* **61**, 4209–4214 (2016).
- [206] McIntyre, S. M. *et al.* Monte carlo simulations to examine the role of pore structure on ambient air separation in metal–organic frameworks. *Industrial & Engineering Chemistry Research* **57**, 9240–9253 (2018).
- [207] Long, R. *et al.* Screening metal-organic frameworks for adsorption-driven osmotic heat engines via grand canonical monte carlo simulations and machine learning. *Iscience* **24**, 101914 (2021).
- [208] Zhang, H., Diao, R., Mostofi, M. & Evans, B. Monte carlo simulation of the adsorption and displacement of ch₄ by co₂ injection in shale organic carbon slit micropores for co₂ enhanced shale gas recovery. *Energy & Fuels* **34**, 150–163 (2019).
- [209] Amamra, S. *et al.* Synthesis and characterization of layered double hydroxides aimed at encapsulation of sodium diclofenac: Theoretical and experimental study. *Journal of Molecular Liquids* **338**, 116677 (2021).

- [210] Earl, D. J. & Deem, M. W. Monte carlo simulations. *Molecular modeling of proteins* 25–36 (2008).
- [211] Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
- [212] Schmidt, J., Marques, M. R., Botti, S. & Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials* **5**, 1–36 (2019).
- [213] Popel, M. *et al.* Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature communications* **11**, 1–15 (2020).
- [214] Sharma, S., Bhatt, M. & Sharma, P. Face recognition system using machine learning algorithm. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 1162–1168 (IEEE, 2020).
- [215] Sarker, I. H. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science* **2**, 1–21 (2021).
- [216] Mohammed, M., Khan, M. B. & Bashier, E. B. M. *Machine learning: algorithms and applications* (Crc Press, 2016).
- [217] Han, J., Kamber, M. & Pei, J. Data mining: concepts and. *Techniques (3rd ed)*, Morgan Kauffman (2011).
- [218] Sarker, I. H. *et al.* Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data* **7**, 1–29 (2020).
- [219] Faber, F. A., Lindmaa, A., Von Lilienfeld, O. A. & Armiento, R. Machine learning energies of 2 million elpasolite (a b c 2 d 6) crystals. *Physical review letters* **117**, 135502 (2016).
- [220] Schmidt, J., Chen, L., Botti, S. & Marques, M. A. Predicting the stability of ternary intermetallics with density functional theory and machine learning. *The Journal of chemical physics* **148**, 241728 (2018).
- [221] Kim, K. *et al.* Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary heusler compounds. *Physical Review Materials* **2**, 123801 (2018).
- [222] Graser, J., Kauwe, S. K. & Sparks, T. D. Machine learning and energy minimization approaches for crystal structure predictions: a review and new horizons. *Chemistry of Materials* **30**, 3601–3612 (2018).
- [223] Oliynyk, A. O., Adutwum, L. A., Harynuk, J. J. & Mar, A. Classifying crystal structures of binary compounds ab through cluster resolution feature selection and support vector machine analysis. *Chemistry of Materials* **28**, 6672–6681 (2016).

Bibliography

- [224] Zhuo, Y., Mansouri Tehrani, A. & Brgoch, J. Predicting the band gaps of inorganic solids by machine learning. *The journal of physical chemistry letters* **9**, 1668–1673 (2018).
- [225] Isayev, O. *et al.* Materials cartography: representing and mining materials space using structural and electronic fingerprints. *Chemistry of Materials* **27**, 735–743 (2015).
- [226] De Jong, M. *et al.* A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Scientific reports* **6**, 1–11 (2016).
- [227] Nguyen, A.-T., Reiter, S. & Rigo, P. A review on simulation-based optimization methods applied to building performance analysis. *Applied energy* **113**, 1043–1058 (2014).
- [228] Wang, X., Xu, Y., Zheng, H. & Yu, K. An extendible, graph-neural-network-based approach for accurate force field development of large flexible organic molecules. *arXiv preprint arXiv:2106.00927* (2021).
- [229] Gavezzotti, A. *Molecular aggregation: structure analysis and molecular simulation of crystals and liquids*, vol. 19 (OUP Oxford, 2006).
- [230] Vanduyfhuys, L. *et al.* Extension of the quickff force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal–organic frameworks (2018).
- [231] Unke, O. T. *et al.* Machine learning force fields. *Chemical Reviews* **121**, 10142–10186 (2021).
- [232] Mueller, J. E., Van Duin, A. C. & Goddard III, W. A. Development and validation of reaxff reactive force field for hydrocarbon chemistry catalyzed by nickel. *The Journal of Physical Chemistry C* **114**, 4939–4949 (2010).
- [233] Senftle, T. P. *et al.* The reaxff reactive force-field: development, applications and future directions. *npj Computational Materials* **2**, 1–14 (2016).
- [234] Van Duin, A. C., Dasgupta, S., Lorant, F. & Goddard, W. A. Reaxff: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A* **105**, 9396–9409 (2001).
- [235] Deringer, V. L., Caro, M. A. & Csányi, G. A general-purpose machine-learning force field for bulk and nanostructured phosphorus. *Nature communications* **11**, 1–11 (2020).
- [236] Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters* **98**, 146401 (2007).
- [237] Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M. & Schmidhuber, J. Flexible, high performance convolutional neural networks for image classification. In *Twenty-second international joint conference on artificial intelligence* (2011).

- [238] Wang, L., Wu, W., Xiao, J. & Yi, Y. Large scale artificial neural network training using multi-gpus. *arXiv preprint arXiv:1511.04348* (2015).
- [239] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444 (2015).
- [240] Yao, K., Herr, J. E., Toth, D. W., Mckintyre, R. & Parkhill, J. The tensormol-0.1 model chemistry: a neural network augmented with long-range physics. *Chemical science* **9**, 2261–2269 (2018).
- [241] Lubbers, N., Smith, J. S. & Barros, K. Hierarchical modeling of molecular energies using a deep neural network. *The Journal of chemical physics* **148**, 241715 (2018).
- [242] Smith, J. S., Isayev, O. & Roitberg, A. E. Ani-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific data* **4**, 1–8 (2017).
- [243] Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nature communications* **8**, 1–8 (2017).
- [244] Quaranta, V., Hellström, M. & Behler, J. Proton-transfer mechanisms at the water–zno interface: The role of presolvation. *The journal of physical chemistry letters* **8**, 1476–1483 (2017).
- [245] Li, J. *et al.* Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chemical science* **12**, 5302–5314 (2021).
- [246] Kolb, B., Luo, X., Zhou, X., Jiang, B. & Guo, H. High-dimensional atomistic neural network potentials for molecule–surface interactions: Hcl scattering from au (111). *The journal of physical chemistry letters* **8**, 666–672 (2017).
- [247] Boes, J. R. & Kitchen, J. R. Neural network predictions of oxygen interactions on a dynamic pd surface. *Molecular Simulation* **43**, 346–354 (2017).
- [248] Bag, S., Konrad, M., Schloder, T., Friederich, P. & Wenzel, W. Fast generation of machine learning-based force fields for adsorption energies. *Journal of Chemical Theory and Computation* **17**, 7195–7202 (2021).
- [249] Järvi, J., Rinke, P. & Todorović, M. Detecting stable adsorbates of (1s)-camphor on cu (111) with bayesian optimization. *Beilstein journal of nanotechnology* **11**, 1577–1589 (2020).
- [250] Deringer, V. L., Proserpio, D. M., Csányi, G. & Pickard, C. J. Data-driven learning and prediction of inorganic crystal structures. *Faraday discussions* **211**, 45–59 (2018).
- [251] Cubuk, E. D., Malone, B. D., Onat, B., Waterland, A. & Kaxiras, E. Representations in neural network based empirical potentials. *The Journal of chemical physics* **147**, 024104 (2017).

Bibliography

- [252] Artrith, N., Morawietz, T. & Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Physical Review B* **83**, 153101 (2011).
- [253] Eckhoff, M. & Behler, J. From molecular fragments to the bulk: development of a neural network potential for mof-5. *Journal of chemical theory and computation* **15**, 3793–3809 (2019).
- [254] Uhrig, R. E. Introduction to artificial neural networks. In *Proceedings of IECON'95-21st Annual Conference on IEEE Industrial Electronics*, vol. 1, 33–37 (IEEE, 1995).
- [255] Gurney, K. *An introduction to neural networks* (CRC press, 2018).
- [256] Zubatiuk, T. & Isayev, O. Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. *Accounts of Chemical Research* **54**, 1575–1585 (2021).
- [257] Pinheiro, M., Ge, F., Ferré, N., Dral, P. O. & Barbatti, M. Choosing the right molecular machine learning potential. *Chemical Science* **12**, 14396–14413 (2021).
- [258] Smith, J. S., Isayev, O. & Roitberg, A. E. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chemical science* **8**, 3192–3203 (2017).
- [259] Han, J., Zhang, L., Car, R. *et al.* Deep potential: A general representation of a many-body potential energy surface. *arXiv preprint arXiv:1707.01478* (2017).
- [260] Zhang, L., Han, J., Wang, H., Car, R. & Weinan, E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical review letters* **120**, 143001 (2018).
- [261] Chmiela, S. *et al.* Machine learning of accurate energy-conserving molecular force fields. *Science advances* **3**, e1603015 (2017).
- [262] Asnaashari, K. & Krems, R. Gradient domain machine learning with composite kernels: improving the accuracy of pes and force fields for large molecules. *Machine Learning: Science and Technology* **3**, 015005 (2021).
- [263] Sauceda, H. E., Chmiela, S., Poltavsky, I., Müller, K.-R. & Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *The Journal of chemical physics* **150**, 114102 (2019).
- [264] Schütt, K. *et al.* Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **30** (2017).
- [265] Wang, Z. *et al.* Symmetry-adapted graph neural networks for constructing molecular dynamics force fields. *Science China Physics, Mechanics & Astronomy* **64**, 1–9 (2021).

- [266] Reiser, P. *et al.* Graph neural networks for materials science and chemistry. *arXiv preprint arXiv:2208.09481* (2022).
- [267] Fung, V., Zhang, J., Juarez, E. & Sumpter, B. G. Benchmarking graph neural networks for materials chemistry. *npj Computational Materials* **7**, 1–8 (2021).
- [268] Ekambaram, V. N. *Graph-structured data viewed through a Fourier lens* (University of California, Berkeley, 2014).
- [269] Park, C. W. *et al.* Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *npj Computational Materials* **7**, 1–9 (2021).
- [270] Wang, F. *et al.* Graph attention convolutional neural network model for chemical poisoning of honey bees' prediction. *Science Bulletin* **65**, 1184–1191 (2020).
- [271] Yan, C. *et al.* Retroxpert: Decompose retrosynthesis prediction like a chemist. *Advances in Neural Information Processing Systems* **33**, 11248–11258 (2020).
- [272] Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature communications* **9**, 1–10 (2018).
- [273] Gong, W. & Yan, Q. Graph-based deep learning frameworks for molecules and solid-state materials. *Computational Materials Science* **195**, 110332 (2021).
- [274] Le, T., Epa, V. C., Burden, F. R. & Winkler, D. A. Quantitative structure–property relationship modeling of diverse materials properties. *Chemical reviews* **112**, 2889–2919 (2012).
- [275] Nolte, T. M. & Ragas, A. M. A review of quantitative structure–property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps. *Environmental Science: Processes & Impacts* **19**, 221–246 (2017).
- [276] dos Santos, V. H. J. M., Pontin, D., Rambo, R. S. & Seferin, M. The application of quantitative structure–property relationship modeling and exploratory analysis to screen catalysts for the synthesis of oleochemical carbonates from CO₂ and bio-based epoxides. *Journal of the American Oil Chemists' Society* **97**, 817–837 (2020).
- [277] Katritzky, A. R., Sild, S., Lobanov, V. & Karelson, M. Quantitative structure–property relationship (qspr) correlation of glass transition temperatures of high molecular weight polymers. *Journal of chemical information and computer sciences* **38**, 300–304 (1998).
- [278] Swamidass, S. J. *et al.* Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics* **21**, i359–i368 (2005).
- [279] Bouchlaghem, Y., Akhiat, Y. & Amjad, S. Feature selection: A review and comparative study. In *E3S Web of Conferences*, vol. 351, 01046 (EDP Sciences, 2022).

Bibliography

- [280] Wadhwa, P. & Mittal, A. Quantitative structure-property relationship (qspr) modeling applications in formulation development. In *Computer Aided Pharmaceutics and Drug Delivery*, 543–560 (Springer, 2022).
- [281] Khan, P. M. & Roy, K. Current approaches for choosing feature selection and learning algorithms in quantitative structure–activity relationships (qsar). *Expert opinion on drug discovery* **13**, 1075–1089 (2018).
- [282] Schober, P., Boer, C. & Schwarte, L. A. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia* **126**, 1763–1768 (2018).
- [283] Sánchez-Marono, N., Alonso-Betanzos, A. & Tombilla-Sanromán, M. Filter methods for feature selection—a comparative study. In *International Conference on Intelligent Data Engineering and Automated Learning*, 178–187 (Springer, 2007).
- [284] Zeng, X., Chen, Y.-W. & Tao, C. Feature selection using recursive feature elimination for handwritten digit recognition. In *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 1205–1208 (IEEE, 2009).
- [285] Kohavi, R. & John, G. H. Wrappers for feature subset selection. *Artificial intelligence* **97**, 273–324 (1997).
- [286] Kuhn, M., Johnson, K. *et al.* *Applied predictive modeling*, vol. 26 (Springer, 2013).
- [287] Breiman, L. Bagging predictors. *Machine learning* **24**, 123–140 (1996).
- [288] Lal, T. N., Chapelle, O., Weston, J. & Elisseeff, A. Embedded methods. In *Feature extraction*, 137–165 (Springer, 2006).
- [289] Chen, R., Dewi, C., Huang, S. & Caraka, R. Selecting critical features for data classification based on machine learning methods. *j. big data* **7** (1), 1–26 (2020).
- [290] Thomas, T., P. Vijayaraghavan, A. & Emmanuel, S. *Applications of Decision Trees*, 157–184 (Springer Singapore, Singapore, 2020). URL https://doi.org/10.1007/978-981-15-1706-8_9.
- [291] Darst, B. F., Malecki, K. C. & Engelman, C. D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC genetics* **19**, 1–6 (2018).
- [292] Biau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).
- [293] Schiessler, E. J. *et al.* Predicting the inhibition efficiencies of magnesium dissolution modulators using sparse machine learning models. *npj Computational Materials* **7**, 1–9 (2021).
- [294] Uyanık, T., Karatuğ, Ç. & Arslanoğlu, Y. Machine learning approach to ship fuel consumption: A case of container vessel. *Transportation Research Part D: Transport and Environment* **84**, 102389 (2020).

-
- [295] Würger, T. *et al.* Exploring structure-property relationships in magnesium dissolution modulators. *npj Materials degradation* **5**, 1–10 (2021).
- [296] Feiler, C. *et al.* In silico screening of modulators of magnesium dissolution. *Corrosion science* **163**, 108245 (2020).
- [297] Chen, F. F. *et al.* Correlation between molecular features and electrochemical properties using an artificial neural network. *Materials & Design* **112**, 410–418 (2016).
- [298] Hofmann, T., Schölkopf, B. & Smola, A. J. Kernel methods in machine learning. *The annals of statistics* **36**, 1171–1220 (2008).
- [299] Condon, D. Performance of artificial neural networks on small structured datasets .
- [300] Liu, Y. *et al.* A machine learning-based qsar model for benzimidazole derivatives as corrosion inhibitors by incorporating comprehensive feature selection. *Interdisciplinary Sciences: Computational Life Sciences* **11**, 738–747 (2019).
- [301] Saucedo, H. E., Chmiela, S., Poltavsky, I., Müller, K.-R. & Tkatchenko, A. Construction of machine learned force fields with quantum chemical accuracy: Applications and chemical insights. *Machine Learning Meets Quantum Physics* 277–307 (2020).
- [302] Thurnhofer-Hemsi, K., López-Rubio, E., Molina-Cabello, M. A. & Najarian, K. Radial basis function kernel optimization for support vector machine classifiers. *arXiv preprint arXiv:2007.08233* (2020).
- [303] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. & Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **408**, 189–215 (2020).
- [304] Kaur, P., Pannu, H. S. & Malhi, A. K. Plant disease recognition using fractional-order zernike moments and svm classifier. *Neural Computing and Applications* **31**, 8749–8768 (2019).
- [305] Bhowmik, T. K., Ghanty, P., Roy, A. & Parui, S. K. Svm-based hierarchical architectures for handwritten bangla character recognition. *International Journal on Document Analysis and Recognition (IJDAR)* **12**, 97–108 (2009).
- [306] Je, H.-M., Kim, D. & Bang, S. Y. Human face detection in digital video using svmensemble. *Neural processing letters* **17**, 239–252 (2003).
- [307] Awad, M. & Khanna, R. Support vector regression. In *Efficient learning machines*, 67–80 (Springer, 2015).
- [308] Schölkopf, B., Luo, Z. & Vovk, V. *Empirical inference: Festschrift in honor of Vladimir N. Vapnik* (Springer Science & Business Media, 2013).

Bibliography

- [309] Okujeni, A. *et al.* A comparison of advanced regression algorithms for quantifying urban land cover. *Remote Sensing* **6**, 6324–6346 (2014).
- [310] Wehbe, B., Hildebrandt, M. & Kirchner, F. Experimental evaluation of various machine learning regression methods for model identification of autonomous underwater vehicles. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 4885–4890 (IEEE, 2017).
- [311] Hofmann, M. Support vector machines-kernels and the kernel trick. *Notes* **26**, 1–16 (2006).
- [312] Chudzian, P. Radial basis function kernel optimization for pattern classification. In *Computer Recognition Systems 4*, 99–108 (Springer, 2011).
- [313] Shi, H., Xiao, H., Zhou, J., Li, N. & Zhou, H. Radial basis function kernel parameter optimization algorithm in support vector machine based on segmented dichotomy. In *2018 5th International Conference on Systems and Informatics (ICSAI)*, 383–388 (IEEE, 2018).
- [314] Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Statistics and computing* **14**, 199–222 (2004).
- [315] Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T. & Prachayasittikul, V. A practical overview of quantitative structure-activity relationship (2009).
- [316] Fang, S., Wang, M., Qi, W. & Zheng, F. Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials. *Computational Materials Science* **44**, 647–655 (2008).
- [317] Zhi, Y., Fu, D., Zhang, D., Yang, T. & Li, X. Prediction and knowledge mining of outdoor atmospheric corrosion rates of low alloy steels based on the random forests approach. *Metals* **9**, 383 (2019).
- [318] Winkler, D. A. *et al.* Using high throughput experimental data and in silico models to discover alternatives to toxic chromate corrosion inhibitors. *Corrosion Science* **106**, 229–235 (2016).
- [319] White, P. A. *et al.* Towards materials discovery: assays for screening and study of chemical interactions of novel corrosion inhibitors in solution and coatings. *New Journal of Chemistry* **44**, 7647–7658 (2020).
- [320] Kokalj, A. Molecular modeling of organic corrosion inhibitors: Calculations, pitfalls, and conceptualization of molecule–surface bonding. *Corrosion Science* **193**, 109650 (2021).
- [321] Galvão, T. L., Novell-Leruth, G., Kuznetsova, A., Tedim, J. & Gomes, J. R. Elucidating structure–property relationships in aluminum alloy corrosion inhibitors by machine learning. *The Journal of Physical Chemistry C* **124**, 5624–5635 (2020).

- [322] Mauri, A. alvades: A tool to calculate and analyze molecular descriptors and fingerprints. In *Ecotoxicological QSARs*, 801–820 (Springer, 2020).
- [323] Landrum, G. Rdkit: Open-source cheminformatics software (2016). URL https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4.
- [324] Alamri, A. H. & Alhazmi, N. Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors. *Journal of Saudi Chemical Society* **26**, 101536 (2022).
- [325] Khaled, K. Modeling corrosion inhibition of iron in acid medium by genetic function approximation method: A qsar model. *Corrosion science* **53**, 3457–3465 (2011).
- [326] El Sayed, H., El Nemr, A., Essawy, S. A. & Ragab, S. Corrosion inhibitors part v: Qsar of benzimidazole and 2-substituted derivatives as corrosion inhibitors by using the quantum chemical parameters. *Progress in Organic Coatings* **61**, 11–20 (2008).
- [327] Breedon, M., Per, M. C., Cole, I. S. & Barnard, A. S. Molecular ionization and deprotonation energies as indicators of functional coating performance. *Journal of Materials Chemistry A* **2**, 16660–16668 (2014).
- [328] Higuchi, T. Mechanism of sustained-action medication. theoretical analysis of rate of release of solid drugs dispersed in solid matrices. *Journal of pharmaceutical sciences* **52**, 1145–1149 (1963).
- [329] Costa, P. & Lobo, J. M. S. Modeling and comparison of dissolution profiles. *European journal of pharmaceutical sciences* **13**, 123–133 (2001).
- [330] Ambrogi, V., Fardella, G., Grandolini, G., Perioli, L. & Tiralti, M. C. Intercalation compounds of hydrotalcite-like anionic clays with anti-inflammatory agents, ii: uptake of diclofenac for a controlled release formulation. *Aaps Pharmscitech* **3**, 77–82 (2002).
- [331] Shkirskiy, V. *et al.* Factors affecting moo42–inhibitor release from zn2al based layered double hydroxide and their implication in protecting hot dip galvanized steel by means of organic coatings. *ACS applied materials & interfaces* **7**, 25180–25192 (2015).
- [332] Prasanna, S. V. & Kamath, P. V. Chromate uptake characteristics of the pristine layered double hydroxides of mg with al. *Solid State Sciences* **10**, 260–266 (2008).
- [333] Xu, Y. *et al.* Removal efficiency of arsenate and phosphate from aqueous solution using layered double hydroxide materials: intercalation vs. precipitation. *Journal of Materials Chemistry* **20**, 4684–4691 (2010).
- [334] Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry* **30**, 2157–2164 (2009).

Bibliography

- [335] Galvão, T. L. *et al.* Control of crystallite and particle size in the synthesis of layered double hydroxides: Macromolecular insights and a complementary modeling tool. *Journal of Colloid and Interface Science* **468**, 86–94 (2016).
- [336] Li, X. *et al.* Atomistic insight into the hydration states of layered double hydroxides. *ACS omega* **7**, 12412–12423 (2022).
- [337] Iqbal, M. A., Sun, L., Barrett, A. T. & Fedel, M. Layered double hydroxide protective films developed on aluminum and aluminum alloys: synthetic methods and anti-corrosion mechanisms. *Coatings* **10**, 428 (2020).
- [338] Serdechnova, M. *et al.* Interlayer intercalation and arrangement of 2-mercaptobenzothiazolate and 1, 2, 3-benzotriazolate anions in layered double hydroxides: In situ x-ray diffraction study. *Journal of Solid State Chemistry* **233**, 158–165 (2016).
- [339] Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **47**, 558 (1993).
- [340] Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B* **54**, 11169 (1996).
- [341] Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science* **6**, 15–50 (1996).
- [342] Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B* **49**, 14251 (1994).
- [343] Blöchl, P. E. Projector augmented-wave method. *Physical review B* **50**, 17953 (1994).
- [344] Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b* **59**, 1758 (1999).
- [345] Klimeš, J., Bowler, D. R. & Michaelides, A. Van der waals density functionals applied to solids. *Physical Review B* **83**, 195131 (2011).
- [346] Klimeš, J., Bowler, D. R. & Michaelides, A. Chemical accuracy for the van der waals density functional. *Journal of Physics: Condensed Matter* **22**, 022201 (2009).
- [347] Román-Pérez, G. & Soler, J. M. Efficient implementation of a van der waals density functional: application to double-wall carbon nanotubes. *Physical review letters* **103**, 096102 (2009).
- [348] Dion, M., Rydberg, H., Schröder, E., Langreth, D. C. & Lundqvist, B. I. Van der waals density functional for general geometries. *Physical review letters* **92**, 246401 (2004).

-
- [349] Lee, K., Murray, É. D., Kong, L., Lundqvist, B. I. & Langreth, D. C. Higher-accuracy van der waals density functional. *Physical Review B* **82**, 081101 (2010).
- [350] Thonhauser, T. *et al.* Van der waals density functional: Self-consistent potential and the nature of the van der waals bond. *Physical Review B* **76**, 125112 (2007).
- [351] Kratzer, P. & Neugebauer, J. The basics of electronic structure theory for periodic systems. *Frontiers in chemistry* **7**, 106 (2019).
- [352] Manz, T. A. & Limas, N. G. Introducing ddec6 atomic population analysis: part 1. charge partitioning theory and methodology. *RSC advances* **6**, 47771–47801 (2016).
- [353] Limas, N. G. & Manz, T. A. Introducing ddec6 atomic population analysis: part 2. computed results for a wide range of periodic and nonperiodic materials. *RSC advances* **6**, 45727–45747 (2016).
- [354] Abraham, M., Van der Spoel, D., Lindahl, E. & Hess, B. The gromacs development team, gromacs user manual version 2016.3 (2017).
- [355] Hockney, R. W., Goel, S. & Eastwood, J. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* **14**, 148–158 (1974).
- [356] Essmann, U. *et al.* A smooth particle mesh ewald method. *The Journal of chemical physics* **103**, 8577–8593 (1995).
- [357] Berendsen, H., Grigera, J. & Straatsma, T. The missing term in effective pair potentials. *Journal of Physical Chemistry* **91**, 6269–6271 (1987).
- [358] Cadena, C. & Maginn, E. J. Molecular simulation study of some thermophysical and transport properties of triazolium-based ionic liquids. *The Journal of Physical Chemistry B* **110**, 18026–18039 (2006).
- [359] Smith, D. E. & Dang, L. X. Computer simulations of nacl association in polarizable water. *The Journal of Chemical Physics* **100**, 3757–3766 (1994).
- [360] Schmid, N. *et al.* Definition and testing of the gromos force-field versions 54a7 and 54b7. *European biophysics journal* **40**, 843–856 (2011).
- [361] Malde, A. K. *et al.* An automated force field topology builder (atb) and repository: version 1.0. *Journal of chemical theory and computation* **7**, 4026–4037 (2011).
- [362] Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126**, 014101 (2007).
- [363] Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81**, 3684–3690 (1984).

Bibliography

- [364] Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A* **31**, 1695 (1985).
- [365] Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics* **52**, 255–268 (1984).
- [366] Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52**, 7182–7190 (1981).
- [367] Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids* (Oxford university press, 2017).
- [368] Sasai, R. *et al.* Why do carbonate anions have extremely high stability in the interlayer space of layered double hydroxides? case study of layered double hydroxide consisting of mg and al (mg/al= 2). *Inorganic chemistry* **58**, 10928–10935 (2019).
- [369] Sun, Z., Jin, L., Shi, W., Wei, M. & Duan, X. Preparation of an anion dye intercalated into layered double hydroxides and its controllable luminescence properties. *Chemical Engineering Journal* **161**, 293–300 (2010).
- [370] Bocclair, J. W., Brateman, P. S., Brister, B. D. & Yarberry, F. Layer- anion interactions in magnesium aluminum layered double hydroxides intercalated with cobalticyanide and nitroprusside. *Chemistry of materials* **11**, 2199–2204 (1999).
- [371] Marappa, S., Radha, S. & Kamath, P. V. Nitrate-intercalated layered double hydroxides–structure model, order, and disorder. *European Journal of Inorganic Chemistry* **2013**, 2122–2128 (2013).
- [372] Jenkins, H. & Thakur, K. Reappraisal of thermochemical radii for complex ions. *Journal of Chemical Education* **56**, 576 (1979).
- [373] Heyrovská, R. Dependence of ion–water distances on covalent radii, ionic radii in water and distances of oxygen and hydrogen of water from ion/water boundaries. *Chemical physics letters* **429**, 600–605 (2006).
- [374] Buchner, R., Samani, F., May, P. M., Sturm, P. & Hefter, G. Hydration and ion pairing in aqueous sodium oxalate solutions. *ChemPhysChem* **4**, 373–378 (2003).
- [375] Saucedo, H. E., Gastegger, M., Chmiela, S., Müller, K.-R. & Tkatchenko, A. Molecular force fields with gradient-domain machine learning (gdml): Comparison and synergies with classical force fields. *The Journal of Chemical Physics* **153**, 124109 (2020).
- [376] Morawietz, T., Singraber, A., Dellago, C. & Behler, J. How van der waals interactions determine the unique properties of water. *Proceedings of the National Academy of Sciences* **113**, 8368–8373 (2016).

- [377] Unke, O. T., Castro-Palacio, J. C., Bemish, R. J. & Meuwly, M. Collision-induced rotational excitation in $n\text{2}+(2\sigma\text{g}^+, v=0)\text{-ar}$: Comparison of computations and experiment. *The Journal of Chemical Physics* **144**, 224307 (2016).
- [378] Denis-Alpizar, O., Unke, O. T., Bemish, R. J. & Meuwly, M. Quantum and quasiclassical trajectory studies of rotational relaxation in ar-n2^+ collisions. *Physical Chemistry Chemical Physics* **19**, 27945–27951 (2017).
- [379] Andrade, M. F. C., Ko, H.-Y., Zhang, L., Car, R. & Selloni, A. Free energy of proton transfer at the water–tio₂ interface from ab initio deep potential molecular dynamics. *Chemical Science* **11**, 2335–2341 (2020).
- [380] Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet—a deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **148**, 241722 (2018).
- [381] Westermayr, J., Gastegger, M. & Marquetand, P. Combining schnet and sharc: The schnarc machine learning approach for excited-state dynamics. *The journal of physical chemistry letters* **11**, 3828–3834 (2020).
- [382] Raimbault, N., Grisafi, A., Ceriotti, M. & Rossi, M. Using gaussian process regression to simulate the vibrational raman spectra of molecular crystals. *New Journal of Physics* **21**, 105001 (2019).
- [383] Wieder, O. *et al.* A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies* **37**, 1–12 (2020).
- [384] Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters* **120**, 145301 (2018).
- [385] Jia, X., De Brabandere, B., Tuytelaars, T. & Gool, L. V. Dynamic filter networks. *Advances in neural information processing systems* **29** (2016).
- [386] Reiser, P., Eberhard, A. & Friederich, P. Graph neural networks in tensorflow-keras with raggedtensor representation (kgcnn). *Software Impacts* 100095 (2021). URL <https://www.sciencedirect.com/science/article/pii/S266596382100035X>.
- [387] Larsen, A. H. *et al.* The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter* **29**, 273002 (2017).
- [388] Zhang, Z. Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine* **4** (2016).
- [389] Behley, J., Steinhage, V. & Cremers, A. B. Efficient radius neighbor search in three-dimensional point clouds. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 3625–3630 (IEEE, 2015).

Bibliography

- [390] Cheng, G., Gong, X.-G. & Yin, W.-J. Crystal structure prediction by combining graph network and optimization algorithm. *Nature communications* **13**, 1–8 (2022).
- [391] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [392] Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of chemical physics* **76**, 637–649 (1982).
- [393] Rowlinson*, J. The maxwell–boltzmann distribution. *Molecular Physics* **103**, 2821–2828 (2005).
- [394] Rühle, V. Berendsen and nose-hoover thermostats. *Am. J. Phys* **575** (2007).
- [395] Zhou, G. & Li, W.-K. The abnormally long cc bond in the oxalate ion. *Journal of Chemical Education* **66**, 572 (1989).
- [396] Shulha, T. *et al.* In situ formation of ldh-based nanocontainers on the surface of az91 magnesium alloy and detailed investigation of their crystal structure. *Journal of Magnesium and Alloys* (2021).
- [397] Frisch, M. *et al.* Gaussian 16 revision c. 01, 2016. *Gaussian Inc. Wallingford CT* **1** (2016).
- [398] Ying, X. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, vol. 1168, 022022 (IOP Publishing, 2019).
- [399] Hua, J., Xiong, Z., Lowey, J., Suh, E. & Dougherty, E. R. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics* **21**, 1509–1515 (2005).
- [400] Devinyak, O., Havrylyuk, D. & Lesyk, R. 3d-morse descriptors explained. *Journal of Molecular Graphics and Modelling* **54**, 194–203 (2014).
- [401] Labute, P. A widely applicable set of descriptors. *Journal of Molecular Graphics and Modelling* **18**, 464–477 (2000).
- [402] Hollas, B. An analysis of the autocorrelation descriptor for molecules. *Journal of mathematical chemistry* **33**, 91–101 (2003).
- [403] Caballero, J. Computational modeling to explain why 5, 5-diarylpentadienamides are trpv1 antagonists. *Molecules* **26**, 1765 (2021).
- [404] Reid, B., Agri-Minerals, P. E. & Headquarters, C. Nop petition for inclusion of magnesium oxide to the national list of substances allowed. *Cell* **850**, 261–0807 (2013).
- [405] Case, D. R., Zubieta, J., Gonzalez, R. & Doyle, R. P. Synthesis and chemical and biological evaluation of a glycine tripeptide chelate of magnesium. *Molecules* **26**, 2419 (2021).

-
- [406] Murakami, Y. Complexing behavior of kojic acid with metal ions. i. mg (ii) and mn (ii) chelates. *Bulletin of the Chemical Society of Japan* **35**, 52–56 (1962).
- [407] Kufelnicki, A. Complexes of uracil (2, 4-dihydroxypyrimidine) derivatives. part i. cu (ii), ca (ii) and mg (ii) coordination with uracil and related compounds in aqueous solution. *Polish Journal of Chemistry* **76**, 1559–1570 (2002).
- [408] Anjum, M. J. *et al.* Green corrosion inhibitors intercalated mg: Al layered double hydroxide coatings to protect mg alloy. *Rare Metals* **40**, 2254–2265 (2021).
- [409] Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Machine learning* **46**, 389–422 (2002).
- [410] Würger, T. *et al.* Data-driven selection of electrolyte additives for aqueous magnesium batteries. *Journal of Materials Chemistry A* (2022).
- [411] Deyab, M. Corrosion inhibition of heat exchanger tubing material (titanium) in msf desalination plants in acid cleaning solution using aromatic nitro compounds. *Desalination* **439**, 73–79 (2018).
- [412] Aslam, J. *et al.* Inhibitory effect of 2-nitroacridone on corrosion of low carbon steel in 1 m hcl solution: An experimental and theoretical approach. *Journal of Materials Research and Technology* **9**, 4061–4075 (2020).
- [413] Eddy, N. O., Ameh, P. O. & Essien, N. B. Experimental and computational chemistry studies on the inhibition of aluminium and mild steel in 0.1 m hcl by 3-nitrobenzoic acid. *Journal of Taibah University for Science* **12**, 545–556 (2018).