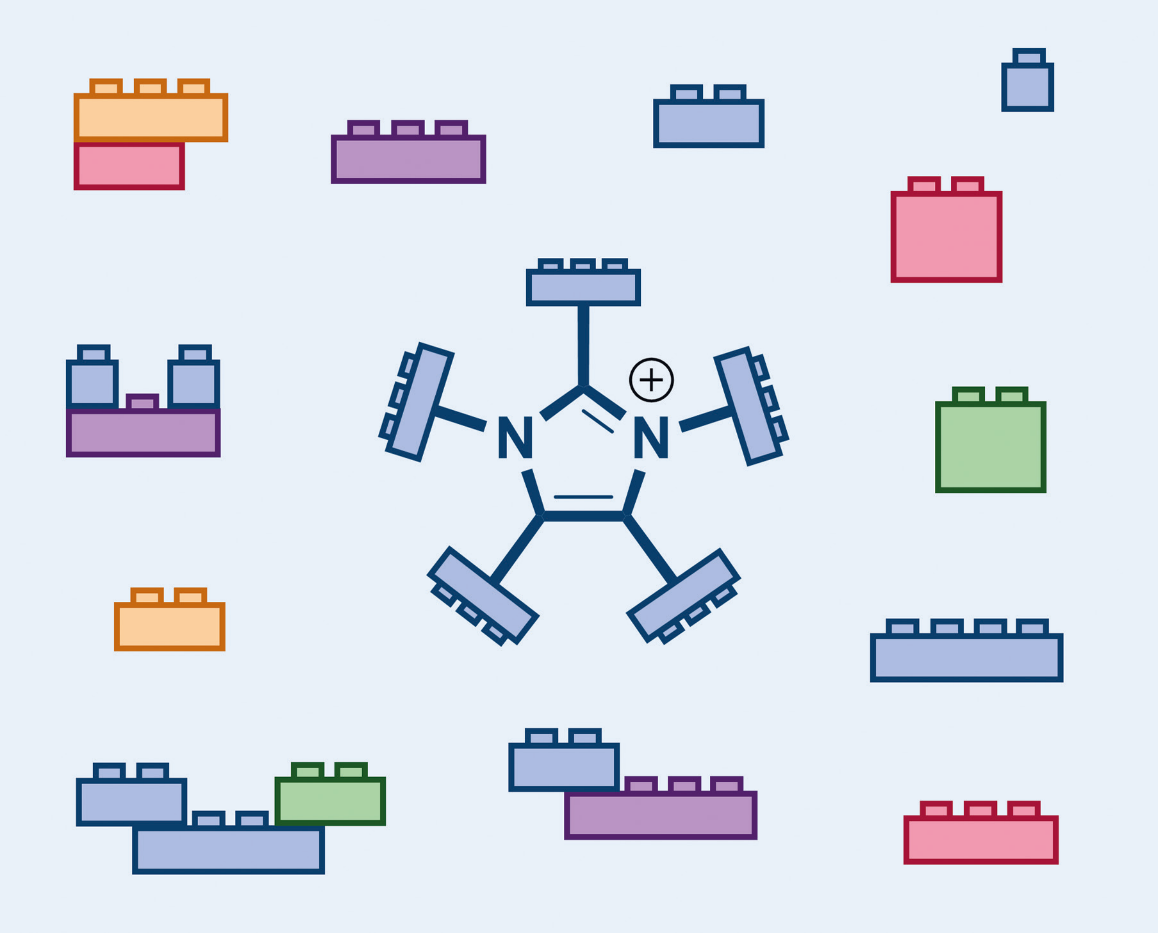


Accelerating the discovery of alkaline-stable anion exchange membrane materials via computational exploration

Fabian Paul Tipp

Energie & Umwelt / Energy & Environment
Band / Volume 701
ISBN 978-3-95806-900-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment Band / Volume 701

Forschungszentrum Jülich GmbH
Institute of Energy Technologies (IET)
Theorie und computergestützte Modellierung von Materialien
in der Energietechnik (IET-3)

Accelerating the discovery of alkaline-stable anion exchange membrane materials via computational exploration

Fabian Paul Tipp

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment Band / Volume 701

ISSN 1866-1793 ISBN 978-3-95806-900-8

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

Herausgeber und Vertrieb: Forschungszentrum Jülich GmbH
Zentralbibliothek, Verlag
52425 Jülich
Tel.: +49 2461 61-5368
Fax: +49 2461 61-6103
zb-publikation@fz-juelich.de
www.fz-juelich.de/zb

Umschlaggestaltung: Grafische Medien, Forschungszentrum Jülich GmbH

Druck: Grafische Medien, Forschungszentrum Jülich GmbH

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment, Band / Volume 701

D 82 (Diss. RWTH Aachen University, 2025)

ISSN 1866-1793
ISBN 978-3-95806-900-8

Contents

Abstract

Hydrogen fuel cells and water electrolyzers using anion exchange membranes (AEMs) host an alkaline chemical environment that mitigates the need for rare platinum group metals (PGMs) as electrocatalyst. Despite the extraordinary potential for widespread adoption of these devices, they are held back by challenges associated with the AEM itself. The current generation of AEMs is particularly limited in terms of durability in the caustic conditions they inhabit, as the cationic moieties of the AEMs are vulnerable to the high concentration of hydroxide anions. Designing cationic moieties that are highly stable in alkaline conditions is, therefore, of crucial importance.

Promising classes of cationic compounds for the implementation in AEMs are those based on the imidazolium group. Through adjustments of the chemical structure via different substitution schemes, the alkaline stability is highly adjustable. Experimental exploration of the vast chemical space of imidazolium-based compounds is slow due to the high resource demand and an insufficient understanding of the structure-stability relationship. The work described in this thesis has thus been predominantly focused on leveraging computational means to accelerate the discovery of alkaline stable imidazolium-based compounds.

As the first step in this undertaking, a reliable computational descriptor for the alkaline stability of a given imidazolium has been identified in the Gibbs free energy change of the C-2 hydroxide attack by comparing the degradation energetics computed through *ab initio* simulations with experimental stability values reported in literature. Thereafter, the calculation of this descriptor was streamlined and fully automated. Next, the descriptor was applied to an extensive molecular dataset containing 5832 imidazolium-based compounds, which was constructed by systematically adding 18 different substituents to a base imidazolium structure. The dataset contains numerous compounds for which the descriptor computationally predicts an exceptional alkaline stability. The computed values were verified as part of this PhD research program by synthesizing five of the molecules and performing two different degradation tests on each, that showed good agreement with the computational predictions. Among the tested compounds, an especially stable compound was identified that surpasses the stability of comparable penta-substituted imidazoliums by substitution of the C-4 and C-5 sites with a methyl group instead of the commonly employed phenyl substituent.

The dataset includes a number of additional compounds with even higher computed alkaline stability. These will be promising candidates for future experimental exploration. Additionally, the dissertation provides further in-depth insights for the design of stable compounds based on new understanding of the structure-stability relationship of imidazolium-based compounds, that was gained through analysis of the molecular dataset. The design of novel compounds will benefit, in particular, from a quantitative assessment of the influence that a selection of 18 different substituents exerts upon addition to each site of the imidazolium ring. A machine learning model was trained for the expanded computational exploration of imidazolium-based compounds. The model is able to approximate the alkaline stability in under one second on a typical laptop processor.

Computational insights were also gained regarding a recently reported azacyclic cationic group, 1,6-diazabicyclo[4.4.4]tetradecan-1,6-ium (in-DBD), which possesses an exceptionally high alkaline stability. Simulations of several degradation pathways were performed, indicating the hydroxide-induced elimination reaction as the most prominent pathway. The impact of small perturbations to the in-DBD structure was investigated to guide the introduction of the cation into a polymer structure without incurring an adverse effect on stability. It was found that the methylation at the β-carbons is especially favorable as it adds protection against the critical elimination degradation reaction.

This work has made significant progress in the pursuit of materials for highly durable anion exchange membranes. A particularly stable cationic group could be identified and design aids for further improvements have been developed.

Kurzfassung

Brennstoffzellen und Wasserelektrolyseure, die auf Anionenaustauschmembranen (AEMs) basieren, ermöglichen die Verwendung von Katalysatoren, die nicht auf Platinmetalle angewiesen sind. Trotz des großen Potentials für die flächendeckende Verwendung dieser Geräte, ist sie derzeit durch Limitationen der Anionenaustauschmembran begrenzt. Die aktuelle Generation von AEMs ist besonders durch die geringe Lebenszeit unter stark basischen Betriebsbedingungen limitiert. Die kationischen Gruppen, welche für den Anionentransport notwendig sind, werden von Hydroxidanionen chemisch angegriffen und irreversibel zerstört. Daher ist die Entwicklung kationischer Gruppen entscheidend, welche den stark basischen Bedingungen standhalten können.

Eine vielversprechende Gruppe von kationischen Verbindungen basiert auf der Imidazoliumgruppe. Anpassungen der chemischen Struktur dieser Verbindungen durch verschiedene Substitutionsmuster haben einen großen Einfluss auf die resultierende alkalische Stabilität. Die experimentelle Identifikation von stabilen Verbindungen ist allerdings langsam, da sie einen hohen Ressourceneinsatz für jede einzelne untersuchte Verbindung benötigt, und auf einem limitierten Verständnis der zugrundeliegenden Struktur-Stabilitäts-Beziehung basiert. In der hier beschriebenen Arbeit wurde daher die beschleunigte Erkundung neuer alkalistabiler Imidazolium-basierter Verbindungen durch computergestützte Methoden vorangetrieben.

Die Änderung der Gibbs-Energie während des Angriffs von Hydroxid auf das C-2 Atom des Imidazoliumringes wurde als zuverlässiger Deskriptor für die alkalische Stabilität von Imidazolium-basierten Verbindungen identifiziert. Die Berechnung des Deskriptors wurde daraufhin automatisiert, um die alkalische Stabilität eines molekularen Datensatz mit 5832 Imidazolium-basierten Verbindungen computergestützt ermitteln zu können. Die berechneten Stabilitätswerte wurden anschließend validiert, indem fünf der Verbindungen synthetisiert und jeweils in zwei unterschiedlichen experimentellen Stabilitätstests untersucht wurden. Die experimentellen Messungen haben hierbei eine gute Übereinstimmung mit den computergestützten Stabilitätsvorhersagen gezeigt. Außerdem konnte ein besonders stabiles Molekül experimentell charakterisiert werden, welches unter alkalischen Be-

dingungen deutlich langlebiger ist als vergleichbare pentasubstituierte Imidazoliumverbindungen.

Weitere Verbindungen mit höherer vorhergesagter alkalischer Stabilität sind im molekularen Datensatz vertreten und sind somit vielversprechende Ziele weiterer experimenteller Untersuchungen. Um den Designprozess weiterer Verbindungen zu unterstützen, wurde die Beziehung zwischen Struktur und Stabilität durch Analyse des molekularen Datensatzes untersucht. Besonders die quantitativen Einflüsse von 18 verschiedenen Substituenten an den jeweiligen Substitutionsstellen des Imidazoliumringes auf die resultierende Stabilität sind für die Weiterentwicklung von Bedeutung. Für die computergestützte Weiterentwicklung von Imidazolium-basierten Verbindungen wurde zudem ein künstliches neuronales Netzwerk entwickelt, welches die alkalische Stabilität einer neuen Verbindung in unter einer Sekunde näherungsweise bestimmen kann.

Neben der Arbeit an Imidazolium-basierten Verbindungen wurde auch das vielversprechende 1,6-diazabicyclo[4.4.4]tetradecan-1,6-ium (in-DBD) Kation mit *ab initio* Simulationen untersucht, da diese Verbindung in einer kürzlich erschienenen Veröffentlichung als besonders stabil in basischer Umgebung identifiziert wurde. Durch die Simulation verschiedener Degradationspfade wurde die Hydroxid-induzierte Eliminationsreaktion als besonders relevant identifiziert. Es wurde daraufhin der Einfluss kleiner Perturbationen auf die Stabilität der in-DBD Struktur untersucht, um Methoden zum Andocken der Verbindung an ein Polymerrückgrat zu identifizieren. Hierbei wurde die Methylierung der β-Kohlenstoffatome als besonders vielversprechender Ansatz aufgezeigt.

Die vorliegende Arbeit enthält signifikante Fortschritte für die Erkundung alkalistabiler Verbindungen zur Verwendung in Anionenaustauschmembranen. Eine besonders stabile Verbindung konnte experimentell charakterisiert werden und mehrere Ansätze zur beschleunigten Erkundung weiterer stabiler Verbindungen wurden entwickelt.

Acknowledgement

Many vital contributions towards the presented PhD thesis have been made by multiple individuals, whom I would like to acknowledge here.

First and foremost, I would like to express my deepest gratitude and thanks to the director of the IET-3, Prof. Dr. Michael Eikerling, for allowing me to perform my dissertation in his institute. He has provided me with invaluable scientific guidance throughout the years and has always ensured that any resources required for my research were available. Michael has always been available for valuable in-depth discussion on all parts of the dissertation.

I also sincerely thank my direct supervisor Dr. Kourosh Malek. He has enabled me many opportunities and has always given me valuable guidance and support. Kourosh's support was not just limited to this dissertation but he has always ensured that I develop the skills needed for my further career.

My deep gratitude also goes out to my day-to-day supervisor Dr. Mehrtoos Eslamibidgoli. Mehrtoos has always supported me in the work of this dissertation and was always at hand for fruitful discussions. His contributions were especially valuable through the many inspiring ideas he brought to me.

Special thanks also go out to André Colliard and Max Dreger who have supported me all the way from the start of our bachelor studies to this point. They have always been there for valuable discussions and have always motivated me.

Additionally, I would like to thank all members of the IET-3. The welcoming and supportive atmosphere in the institute have made my workdays highly enjoyable and made my long commute more than worthwhile. The many extracurricular activities have also added a highly valuable balance to the dissertation work.

The research visit at the Simon Fraser University has been highly rewarding and valuable to this work. I would like to extend my deepest appreciation and thanks to Prof. Dr.

Steven Holdcroft for providing me the possibility to perform research in his group and providing me with valuable guidance before, during and after the visit.

I am very grateful to Dr. Kate Fraser for providing her experimental expertise during the computational part and the enormous support during the experimental part of the project, patiently providing me the required knowledge in experimentally working with imidazolium-based compounds.

My deep appreciation also goes to Dr. Chase Radford for his extensive guidance in performing and analyzing alkaline degradation tests, performing measurements after I have returned to Germany and even advising me on the data analysis after he started a new job outside of the Holdcroft group.

Many thanks also go to Dr. Torben Saatkamp for his support on correctly utilizing the dynamic vapor sorption and performing further measurements of my samples after my return to Germany.

I am very thankful to all members of the Holdcroft Group for providing an outstandingly welcoming and supportive environment during my visit and always taking the time to help me with any problems I encountered while adjusting from computational to experimental chemistry.

This endeavor would not be possible without the generous funding from the Mitacs Globalink program (Application Ref. IT34669), the Center for Advanced Simulation and Analytics, and the Ministry of Education and Research (BMBF) under the German-Canadian Materials Acceleration Centre (GC-MAC) grant number 01DM21001A, for which I would like to express my deepest gratitude.

Lastly, I would like to express my highest gratitude to my friends and family who have always supported me along the way, especially during the particularly work-intensive parts of the thesis. My mother Bettina has especially contributed by her outstanding language proficiency and has helped me tremendously in identifying any unclear or erroneous formulations. The insights into the academic environment provided by my father, Andreas, were highly insightful and have always helped me pick the correct path in my scientific career. My sister Leonie has always provided me with balance during the most stressful phases of this work and ensured my plants remain in good health. My grandmother Marianne made sure that I took sufficient breaks and was well-fed. My friend and roommate Vladimir has always been supportive of my work and his deep knowledge of organic chemistry was highly valuable at many points. My friend Ben was there for me even after the longest workdays and his company during many of the home office days strongly contributed to making the work enjoyable.

Chapter 1

Introduction

The daunting challenges directly related to climate change must be countered with the transition away from fossil fuels and towards renewable energy technologies. Within a sustainable energy system, energy demands are met through renewable sources such as geothermal, hydro, biomass, solar, and wind power.[1, 2] Utilization of these technologies does, however, entail additional complications compared to fossil fuel energy systems. Some technologies, especially those based on solar and wind, deliver power dependent on external factors, exhibiting strong variance. There is a notorious temporal mismatch between energy supply and demand, also described as the intermittency challenge.[3] Additionally, the transportation of sustainably produced energy is challenging.[4] These obstacles must be addressed through a suitable energy storage regime.[5, 6]

1.1 Electrochemical Technologies for the Hydrogen Economy

A promising solution to the obstacles of intermittent power delivery and portability is based on hydrogen as an energy carrier.[7] Molecular hydrogen (H_2) possesses an especially high specific energy that is significantly greater than that of common fossil fuels.[8] Hydrogen can be produced from water in a water electrolyzer (WE) when electrical energy is supplied and then be stored and distributed for later use in stationary and portable applications. On demand, electricity can be produced from hydrogen and oxygen in a fuel cell (FC).

The literature contains numerous reports on electrochemical energy conversion devices, such as those based on a liquid alkaline electrolyte[9, 10], solid oxides[11, 12] and ion exchange membranes[13, 14, 15, 16]. Water electrolyzers and hydrogen fuel cells based on ion exchange membranes are especially prominent in the contemporary literature and can be further classified into those based on proton exchange membranes (PEMs)[13, 14] and on anion exchange membranes (AEMs)[15, 16]. Their prominence stems from their suitability to operate under a wide range of conditions and in numerous applications through properties that are further outlined in the following.

1.1.1 Proton Exchange Membrane Devices

Proton exchange membrane fuel cells (PEMFCs) and PEM water electrolyzers (PEMWEs) are a promising class of hydrogen energy conversion devices. PEMFCs operate below 120 °C and they can reach efficiencies of about 65 %.[17] These attributes make them highly suitable for mobile applications such as hydrogen powered vehicles but also for stationary applications such as grid stabilization.[18, 19] Proton exchange membrane water electrolyzers can reach efficiencies of about 75 - 80 %[17, 20] and they can be utilized flexibly for the production of green hydrogen, with commercial products being available in the kilowatt to megawatt range.[17]

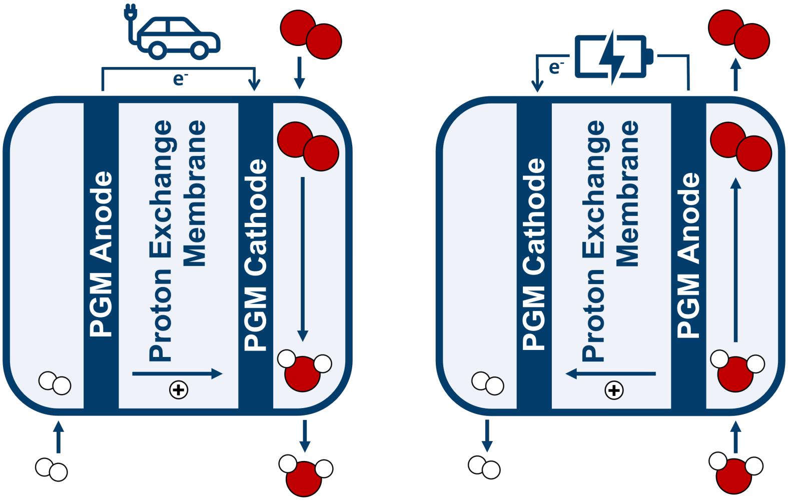


Figure 1.1: Simplified depiction of a proton exchange membrane fuel cell (left) and a proton exchange membrane water electrolyzer (right). White circles represent hydrogen atoms, red circles represent oxygen atoms, the car pictograms represents an electrical load and the battery pictogram represents a voltage source. The electrocatalysts employed in the electrodes are typically based on platinum group metals (PGMs).[17]

Both PEMFCs and PEMWEs operate in a similar fashion, with one performing the inverse process of the other. A simplified schematic of how they function is provided in Figure 1.1. For a PEMFC, molecular hydrogen is supplied to the anode and split into hydrogen cations, typically sitting on a carrier water molecule or embedded into a hydrogen bond network, that are transported through the PEM to the cathode,

$$H_2 \rightarrow 2H^+ + 2e^-. \tag{1.1}$$

At the cathode, the hydrogen cations react with supplied oxygen gas to form water, following the next reaction,

$$2H^+ + \frac{1}{2}O_2 + 2e^- \rightarrow H_2O. \tag{1.2}$$

The electrons that are extracted from H_2 at the anode side are conducted through an external circuit before being recombined with the protons that have migrated through the PEM at the cathode side, allowing electrical work to be done.

Analogously in a PEMWE, water is supplied at the anode and split into molecular oxygen, hydrogen cations and two electrons per water molecule,

$$H_2O \rightarrow 2H^+ + \frac{1}{2}O_2 + 2e^-. \tag{1.3}$$

The hydrogen cations are conducted through the PEM towards the cathode. At the cathode, electrons are provided by a voltage source to reduce the hydrogen cations to molecular hydrogen,

$$2H^+ + 2e^- \rightarrow H_2. \tag{1.4}$$

As the name already suggests, the proton exchange membrane is a crucial part for PEMFCs and PEMWEs to function effectively. The membrane must be conductive to protons with as little resistance as possible to optimize the energy efficiency. Additionally, the membrane should be electrically insulating, be impermeable to gases and have a high chemical, mechanical and thermal stability. Currently one of the most suitable and commonly employed membrane material in PEM devices is the sulfonated fluoropolymer Nafion. However, while Nafion results in comparatively effective devices, it does come with limitations, among them being the high cost and reliance on fluorocarbons. Therefore, research towards suitable alternatives to Nafion is ongoing.[21]

Another crucial component of any PEM-based device is the electrocatalyst. For the electrochemical reactions to take place effectively, a catalyst is required at both the anode and the cathode for PEMFCs and PEMWEs. In the chemical environment of PEM-based devices, platinum group metals are usually the only viable catalyst choice, with iridium commonly being employed for PEMWEs[22] and platinum for PEMFCs[23]. The scarcity and cost of these rare earth metals is a formidable challenge for the widespread adoption of PEM-based hydrogen energy conversion devices.[24, 25]

1.1.2 Anion Exchange Membrane Devices

A promising alternative to devices based on proton exchange membranes are those based on anion exchange membranes. AEMs are permeable to anions, usually hydroxide in practical applications, leading to a significant change in electrochemical mechanisms and conditions in the devices.[26] To illustrate the differences, a simplified schematic of the inner workings of AEM devices is provided in Figure 1.2.

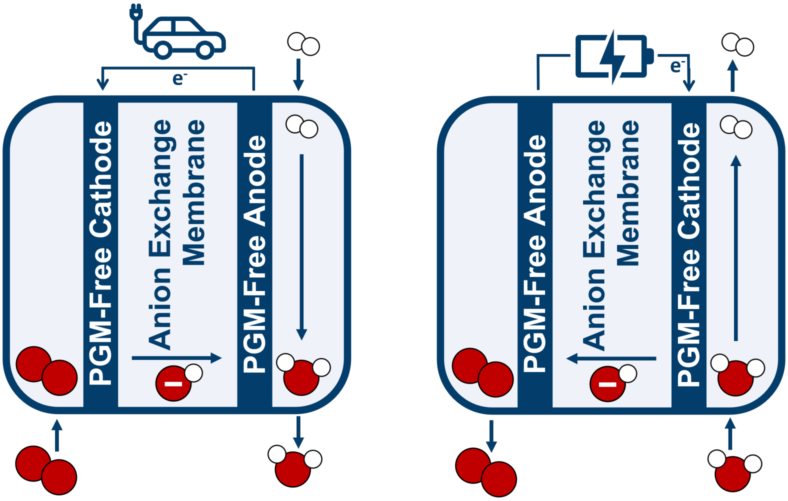


Figure 1.2: Simplified depiction of an anion exchange membrane fuel cell (left) and an anion exchange membrane water electrolyzer (right). White circles represent hydrogen atoms, red circles represent oxygen atoms, the car pictogram represents an electrical load and the battery pictogram represents a voltage source. The transfer of water through the AEM is not depicted for reasons of clarity.[26, 27]

In an anion exchange membrane fuel cell (AEMFC), molecular oxygen is introduced on the cathode side and is reduced to hydroxide anions upon combination with water, which is transported through the AEM from the anode side to the cathode side,[26]

$$\frac{1}{2}O_2 + H_2O + 2e^- \rightarrow 2OH^-. \tag{1.5}$$

The generated hydroxide is transported through the AEM to the anode side. At the anode site, molecular hydrogen is oxidized to form water upon reaction with the hydroxide anions,[26]

$$2OH^- + H_2 \rightarrow 2H_2O + 2e^-. \tag{1.6}$$

The electrons from the hydrogen oxidation at the anode are conducted through an external circuit before being used in the oxygen reduction at the cathode, performing electrical work in the process.

Processes in an anion exchange membrane water electrolyzer (AEMWE) are generally the reverse of those in an AEMFC. At the cathode, water molecules are reduced to molecular hydrogen and hydroxide anions,[27]

$$2H_2O + 2e^- \rightarrow H_2 + 2OH^-. \tag{1.7}$$

The hydroxide anions are transported through the AEM to the anode side. At the anode, hydroxide is reduced to molecular oxygen and water,

$$2OH^- \rightarrow H_2O + \frac{1}{2}O_2 + 2e^-. \tag{1.8}$$

AEM-based devices rely on the transport of hydroxide ions, compared to hydrogen cations in PEM-based devices. This leads to AEM-based devices operating in a strongly alkaline environment compared to the acidic environment in PEM-based devices. Combined with the differing electrochemical processes, this change in chemical environment allows for the utilization of electrocatalysts based on elements other than platinum group metals. While the design of optimal non-precious metal catalysts is still an ongoing area of research, significant progress has been made in recent years towards outperforming platinum group metals with oxides based on abundant elements such as nickel.[28]

1.2 Anion Exchange Membranes

A major challenge to the widespread adoption of AEM-based devices involves the AEM itself. The AEM must fulfill a number of roles that are highly influenced by its chemical composition. The optimal chemical architecture of AEMs is therefore an active area of research.[29]

1.2.1 Required Properties

The design of an optimal AEM is demanding, as many different properties must be imparted simultaneously for optimal device performance. The most important properties are the following:[30]

- High anion conductivity, especially for hydroxide anions in order to minimize losses due to resistance and achieve high performance.
- Low permeability to gases to prevent hydrogen and oxygen from mixing and ensure a safe and efficient operation.
- High electrical resistance to prevent a short circuit between anode and cathode.
- High degree of chemical, thermal and mechanical stability to prevent premature device failure.

Especially insufficient chemical stability is a major obstacle of state-of-the-art AEMs as the strongly alkaline environment can degrade many of the employed materials.[31]

1.2.2 Structure of the Membrane

An AEM is generally made from a polymer backbone, which is augmented by cationic moieties. The polymer backbone is mainly responsible for the mechanical stability of the membrane, while the cationic groups facilitate anion conduction. Figure 1.3 provides a simplified schematic of this structure.

1.2.2.1 Ion Exchange Capacity

An important characteristic of any AEM is the density of fixed positive charge carriers. Typically, this density is reported as the ion exchange capacity (IEC), which is defined as the ratio of the number of ion pairs in the membrane and its molecular weight, typically computed per repeat unit of the polymer.[31, 32]

$$\text{IEC} = \frac{\text{Number of Ionic pairs}}{\text{Molecular Weight}} \tag{1.9}$$

A higher IEC is generally regarded to have a positive impact on anionic conductivity.[33] However, adverse effects may arise from a high IEC in the form of excess water uptake. If the water uptake is too high, the membrane swells, possibly causing unwanted changes to the mechanical and structural properties and also reducing the density of fixed charge carriers.[34]

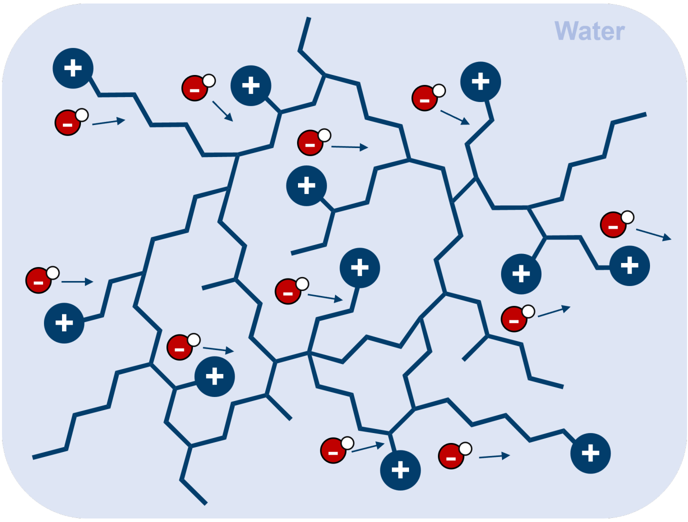


Figure 1.3: Schematic of the structure of an AEM. The blue lines represent the polymer backbone and the blue circles represent the cationic moieties. The light blue background symbolizes that an AEM is generally filled with water. The red and white circles depict hydroxide ions which are transported through the structure.

1.2.2.2 Backbone Materials

The specific makeup of the AEM backbone is a key determinant of the membrane's mechanical properties. A variety of different polymer backbone designs has been reported in the literature. The chemical makeup is often based on poly(arylene ether), polyolefins, polyphenylene, poly(phosphazene) or fluorinated polymers such as poly(vinylidene fluoride).[31, 35]

To conjoin the polymer backbone structure with the cationic moieties, two general strategies are employed. First, the polymer backbone can be a pre-established unit to which cationic chemical groups are attached. The schematic in Figure 1.3 depicts this approach. Alternatively, the cationic moieties can be embedded as structural parts of the backbone, being attached to each other, either directly or through flexible side chains.[31, 32]

1.2.2.3 Cationic Materials

The cationic groups are integral to the conductive properties of any given AEM. Naturally, the chemical structure must incorporate at least one positive formal charge. The positive charge is commonly introduced through nitrogen atoms, however different charge carriers such as phosphor or cobalt are also possible. Figure 1.4 displays a selection of reported cationic structures as an overview.[36]

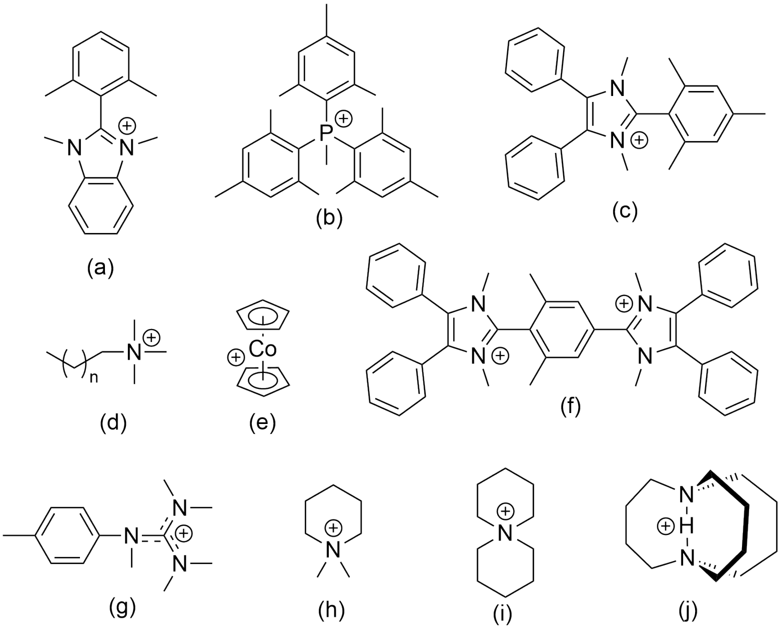


Figure 1.4: Overview of different chemical groups reported in literature for the use in anion exchange membranes. a) Example of a benzimidazolium, b) example of a phosphonium, c) example of a penta-substituted imidazolium, d) example of a quaternary ammonium, e) cobaltocenium, f) example of a bis-arylimidazolium, g) example of a guanidinium, h) example of a piperidinium, i) 6-azaaspiro[5,5]undecan-6-ium, j) 1,6-diazabicyclo[4.4.4]tetradecan-1,6-ium.[32, 36, 37]

The specific chemical structure of the cationic group has a profound impact on the overall properties of the resulting AEM, especially on its conductivity and chemical stability. The chemical stability of the AEM is highly dependent on the cationic functional group as it is a notoriously vulnerable site for hydroxide induced chemical attack. The conductivity of the AEM is affected directly by the choice of cationic group as it impacts the mechanism of ionic transport. Additionally, the cationic group determines the maximum ion exchange capacity of the resulting AEM, as the IEC of the AEM is limited by the charge density of integrated cationic groups. Therefore, cationic groups with a high charge to molecular weight ratio are typically desirable as they provide the highest degree of freedom to dictate the IEC of the AEM.[34]

Cationic moieties that are widely used in AEMs are based on quaternary ammonium (QA).[38, 39] Quaternary ammoniums are highly flexible in terms of their chemical structure and lead to membranes with a high IEC by having a low molecular weight. Typically, their synthesis is comparatively simple, making QAs a cost-effective option.[36, 31] Additionally, QAs generally result in AEMs with high mechanical strength, which is partially due to their ability to self-cross-link.[31] However the chemical stability is a major obstacle to the widespread adoption of QA-based membranes. The positive charge is localized on a single nitrogen atom, making it a vulnerable target of nucleophilic attack by hydroxide anions. Multiple different degradation mechanisms are possible, with the most prevalent being Hofmann elimination, nucleophilic substitution and an elimination reaction.[34]

Various chemical compounds apart from quaternary ammoniums have been explored to improve upon the cationic group's properties, especially chemical stability. Most commonly, these alternative cations still utilize nitrogen atoms to induce localization of the positive charges, such as in guanidinium, pyridinium or imidazolium. Additionally, other strategies have been developed that rely on other charge carriers, such as phosphor in the case of phosphonium or metal cations in the case of cobaltocenium. All of these come with their own advantages and drawbacks and none have emerged as the ideal material so far. However, those groups based on imidazolium have emerged as especially promising and are the main focus of this work.[31]

1.3 Imidazolium-Based Materials

A promising alternative to quaternary ammonium as cationic functional groups in AEMs are compounds based on imidazolium, which have emerged as a strong research focus.[34]

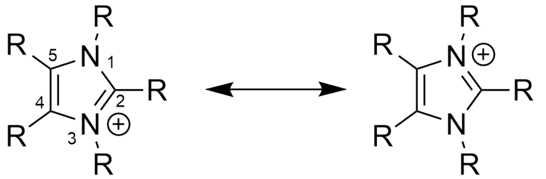


Figure 1.5: General structure of an imidazolium. A numbering scheme for the atoms of the heterocycle is provided, which is used throughout this work. Each of the heterocycle's five atoms can be substituted with a different chemical group. The two most relevant resonance structures are shown.

A notable difference between quaternary ammonium and imidazolium is the localization of the positive charge. In QAs the charge is localized on a single nitrogen atom and therefore especially vulnerable to nucleophilic attack by hydroxide anions. In an imidazolium cation, the positive charge is delocalized throughout the whole heterocycle, mainly on the two nitrogen atoms as can be seen in Figure 1.5. With the charge being delocalized, the susceptibility to nucleophilic attack is markedly reduced, which generally leads to a higher stability in a highly alkaline environment.[34, 40]

Other benefits of imidazolium-based cationic groups are a high anionic conductivity and a comparatively simple synthesis.[34] However, even with the delocalized cationic charge, alkaline stability remains a major challenge. Fortunately, the imidazolium group can be augmented in a straightforward fashion through addition of different substitutents, which significantly impacts the overall properties. Improving alkaline stability of imidazolium-based compounds through various substitution schemes is a vital area of ongoing research.[41]

1.3.1 Degradation Pathways

To combat the degradation of imidazolium-based compounds in strongly alkaline environments, understanding the exact pathways by which degradation takes place is crucial. Various mechanisms have been reported in literature, for example deprotonation of substituents, deprotonation of the heterocycle, S_N2 reaction of the nitrogen substituents and ring opening reaction after direct hydroxide attack of the heterocycle. The latter two are irreversible reactions and have been reported to play an especially important role, they are depicted in Figure 1.6.[32, 42]

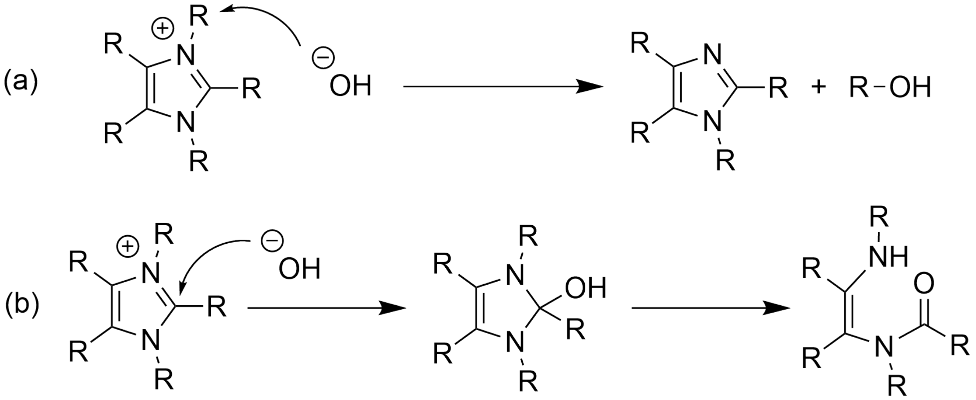


Figure 1.6: Two irreversible alkaline degradation pathways of imidazolium-based materials reported in literature.[42] An S_N2 reaction of hydroxide and one of the nitrogen substituents (a) and a nucleophilic addition-elimination resulting in a ring opening of the heterocycle (b).

The reported degradation reactions result in the positive charge being removed from the imidazolium, turning it into an imidazole. The loss of fixed cationic charges causes a reduction in anionic conductivity.

The rate of the degradation reaction is dependent on multiple parameters, including the hydroxide, the imidazolium and environmental aspects. The electrophilicity of the hydroxide depends on its solvation environment, which is typically made up of water in

AEMFC and AEMWE applications. The amount of water present is highly important, as the more water molecules are present in the hydroxide's proximity, the more its nucleophilicity of the hydroxide is compensated. Therefore, the lower the water content of an AEM, the faster the degradation is to be expected.[43, 44] Another important aspect is how exposed the cationic functional groups are to the hydroxide filled environment, which depends on the exact preparation of the AEM and the backbone material employed.[41] Additionally, other typical factors influencing the speed of a chemical reaction remain relevant, especially reactant concentrations and temperature.[41]

Many of the influential properties are hard to control, at least without significant impact on the performance of the entire device. Therefore, major research interest is aimed toward the decrease of degradation by adjusting the chemical structure of the imidazolium group itself. The chemical structure of an imidazolium can be modified with a high flexibility due to the five atoms of the heterocycle that can be independently substituted. Depending on the exact substitution scheme, great variance in the resulting alkaline stability has been observed.[40, 41]

1.3.2 Overview of Literature-Reported Structures

The flexibility to substitute imidazolium at five independent locations opens up a wide chemical space. With a selection of n different substituents, it is possible to construct n^5 distinct compounds when neglecting the small amount of duplicates due to symmetry. Even with just a moderate selection of 20 different substituents, millions of distinct compounds are possible, considering that $20^5 = 3.2 \times 10^6$. Naturally, only a very small subset of the chemical space has been explored so far, owed to the high cost of experimental investigations.[41]

Those imidazolium-based compounds that have been studied can be categorized into three major groups. Single substituted imidazoliums, substituted benzimidazoliums and two substituted imidazolium rings connected by an aromatic ring. These base groups are depicted in Figure 1.7.

The choice of substituents has generally been based on simple alkanes, ethers and substituted or simple phenyl groups. For the C-4 and C-5 locations, a simple phenyl group or hydrogen is commonly employed. The N-1 and N-3 substituent choice commonly falls on small alkane groups, oftentimes methyl. The C-2 location has been left unsubstituted in earlier approaches but is now commonly substituted by either a small alkane or a substituted phenyl like mesityl.[41, 32]

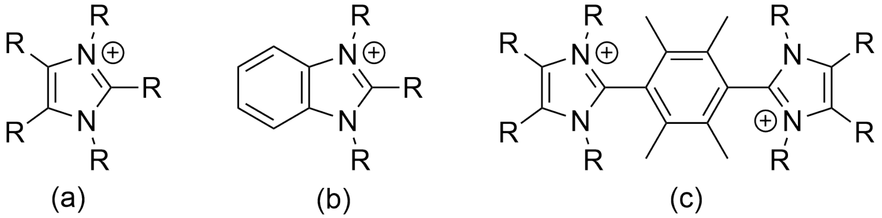


Figure 1.7: Three base compounds from which a majority of imidazolium-based compounds for AEM applications reported in literature are built,[41] (a) substituted single imidazolium, (b) a substituted benzimidazolium, (c) substituted bis-arylimidazolium.

1.3.3 Influence of Chemical Structure on Stability

Even small adjustments to the chemical structure of imidazolium have been shown to cause significant changes in alkaline stability.[32, 41, 40] Understanding the relationship between chemical structure and alkaline degradation is crucial in efforts to rapidly screen and find imidazolium-based compounds with exceptional alkaline stability. The aspects of this relationship reported in the literature so far can be categorized into steric and electronic effects.

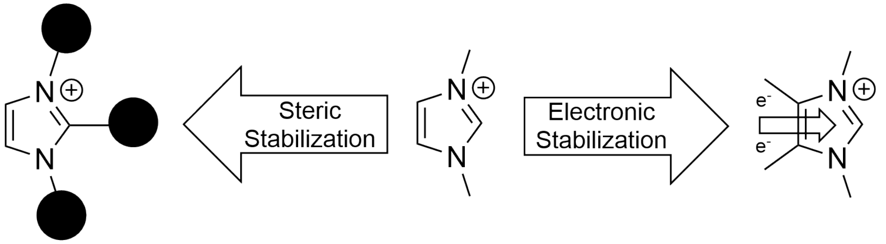


Figure 1.8: Schematic visualizing the two major contributing effects of the structure-stability relationship of imidazolium-based compounds. The solid black circles represent sterically demanding substituents that suppress hydroxide attack. The small hollow arrow on the right, with e^- represents substituents that donate electron density into the imidazolium ring, making it less electrophilic and therefore more resistant to hydroxide attack.

Steric effects rely on substituents that physically block hydroxide from attacking the imidazolium. As the ring opening degradation reaction shown in Figure 1.6(b) is initiated by a hydroxide anion attacking the C-2 atom, substitution at this location is especially impactful. It was found in multiple studies that imidazoliums with only a hydrogen at the C-2 site degrade exceptionally quickly in alkaline environments.[40, 45, 46] Even just the addition of a methyl group at the C-2 location is rewarded with significant increases in alkaline stability.[40] Incorporation of bulky substituents at C-2 results in an additional boost in alkaline stability with methylated phenyl rings showing especially strong effects, i.e., with a mesityl substituent.[32, 40]

The N-1 and N-3 sites are also close to the degradation location. They are right next to the C-2 site at which the ring opening degradation initiates and the substituents at these sites are the target of the S_N2 degradation pathway. Substitution at N-1 and N-3 sites with bulky substituents has been shown to increase alkaline stability, likely due to steric effects. Specifically switching from two methyl substituents to butyl has resulted in a significant stabilization.[32, 40]

Electronic effects contributing to the stabilization of imidazolium are typically based on donation of electron density. When electron density is donated into the imidazolium ring, the positive charge is partially compensated and thereby rendered less electrophilic. Through a decrease in electrophilicity, the hydroxide anion is partially hindered from nucleophilic attack, resulting in higher stability. With added electron density, the energy level of the lowest unoccupied molecular orbital (LUMO) is also lifted. As the highest occupied molecular orbital (HOMO) of the hydroxide typically lies below the LUMO of the imidazolium, this effect moves the two energy levels further apart, thereby inhibiting the reaction of the two molecules.[47, 41]

Electron donating substituents are reported to enable this stabilization. Alkyl substituents such as methyl have been reported as particularly suitable for this purpose through hyperconjugation between the alkyls C-H bond and the imidazolium ring.[48] Even substitution at the C-4 and C-5 sites with methyl has been reported to increase alkaline stability significantly despite its relatively large distance to the reaction sites of the degradation reaction.[40]

1.3.4 Synthesis of Imidazolium

When designing an alkaline stable imidazolium-based structure, it is important to keep the synthesizability of these compounds in focus. Even the most stable and conductive imidazolium is of no use if it cannot be practically synthesized. Fortunately, the synthesis of imidazolium-based compounds can be straightforward in many cases.

A common synthesis pathway for imidazolium is to first form the imidazole corresponding to the desired imidazolium. The imidazole can be obtained through the Debus-Radziszewski imidazole synthesis from a correspondingly substituted 1,2-dicarbonyl and aldehyde.[49]

After acquiring the imidazole, the imidazolium can be formed. While imidazole is a base and can be protonated to form imidazolium even in water with a neutral pH (pK_a of unsubstituted imidazolium: 7.0), this is not the case in the alkaline environment an AEM is employed in.[50] Therefore, in order to form the imidazolium, non-hydrogen substituents

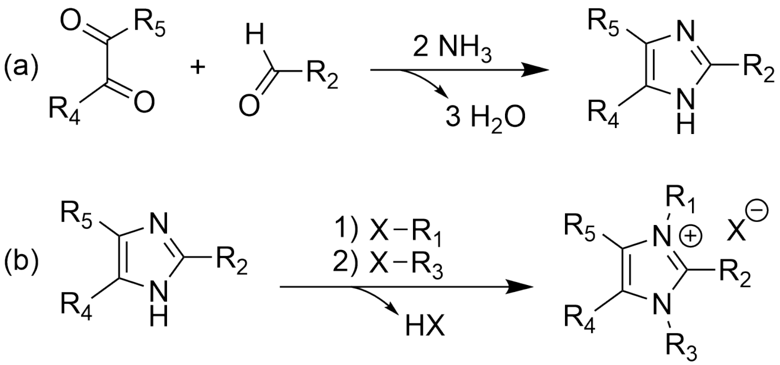


Figure 1.9: Common synthesis pathway of imidazolium. In the first step (a), the imidazole is formed through the Debus-Radziszewski reaction. In the subsequent step (b), the imidazolium is formed by reaction with the halide of the desired nitrogen substituents. The X represents either Br, Cl, or I.

are introduced to both nitrogen atoms. This is commonly done by reacting with the halide of the desired substituent, e.g. methyl iodide.[32]

1.4 Goal and Scope of this Work

While progress is being made in the experimental exploration of alkaline stable cationic moieties for the implementation in AEMs, it is held back by the high resource demand involved in the characterization of each sample. The small sample size of experimentally verified structures also limits the understanding of the structure-stability relationship that is needed for efficient molecular design. For the acceleration of the molecular discovery process, computational methods hold great potential.

In many cases, high throughput computational scanning can be challenging, as the different behavior of the involved compounds needs to be taken into account. However, for imidazolium-based compounds, an enormous chemical space of possible compounds is available that is based around the same core imidazolium group. As these compounds share the core functional group, they will all exhibit a certain consistency in their chemical behavior, making them especially favorable for accurate high throughput calculations to assess and compare their alkaline stability. Due to these especially favoring attributes of imidazolium-based compounds, this work centers around imidazolium-based materials in Chapter 3 to Chapter 5. Chapter 6 applies parts of the methodology developed on imidazolium-based compounds to a novel type of azacylic cation, 1,6-diazabicyclo[4.4.4]tetradecan-1,6-ium (in-DBD), that exhibits particularly promising properties in terms of its chemical stability.

A multi-step approach was chosen to develop a robust workflow for the accelerated materials discovery of alkaline stable AEM materials. First of all, in **Chapter 3**, a computational descriptor for the alkaline stability of imidazolium-based compounds was developed, based on simulations of imdidazolium degradation pathways using density functional theory and coupled cluster calculations. The energetics of prominent degradation reactions were compared to experimental stability values reported in literature to yield a descriptor that is reliable and requires only moderate computational resources.

The obtained descriptor is then utilized in **Chapter 4**, where the calculation pipeline is automated to perform stability computations of a large number of compounds. A dataset containing 5832 compounds was systematically generated from a base imidazolium and a set of 18 substituents. The stability descriptor was computed for each contained compound and the dataset was analyzed to identify trends and compounds of interest. Five of the compounds contained were synthesized to verify the computational predictions through experimental stability measurements.

The experimentally verified dataset of imidazolium-based compounds forms the basis of **Chapter 5**, in which further insights are described that were generated from these data. The dataset was analyzed to identify quantitative relationships between chemical structure and alkaline stability, focusing on the impact of individual substituents. Further, the size of the dataset allowed for the training of various machine learning models to approximate the descriptor while requiring a fraction of the resources that the initial *ab initio* simulations require.

Lastly, the workflow of the previous chapters is applied to the novel in-DBD cation, as described in **Chapter 6**. The degradation pathways of in-DBD were characterized at the density functional theory level to identify the main weak points. Additionally, the impact of perturbations to the base structure on alkaline stability were investigated to aid the design of attachment strategies to a polymer backbone. This was done by computing the degradation energetics of the modified structures.

Chapter 2

Theoretical Background

In this chapter, the theoretical background into the utilized computational methods is provided. This work was mainly performed through computational chemistry simulations, with density functional theory being the most prominently used. Additionally, machine learning models were trained on the computed data.

2.1 Computational Chemistry

Describing chemical systems mathematically has been a long standing goal, with many major contributions such as the Schrödinger equation [51], the Born-Oppenheimer approximation [52] and the Hartree-Fock method [53] having been described nearly a century ago. Many excellent textbooks, such as those by Frank Jensen [54] or by Attila Szabo and Neil Ostlund [55] exist describing computational chemistry methods in detail, thereofore, only a limited scope of essential concepts are covered in the following.

2.1.1 Fundamentals of Molecular Quantum Mechanics

To investigate materials theoretically at the atomistic level, classical mechanics do not provide suitable explanations and quantum mechanics must be employed. For the investigation of a quantum mechanical system, it is crucial to obtain the wave function Ψ of that given system, as Ψ fully describes system's state.[56]

2.1.1.1 The Schrödinger Equation

For most chemical applications, the non-relativistic time-independent Schrödinger equation gives access to Ψ of the system of interest,[56, 55]

$$\hat{H}\Psi = E_{\text{total}}\Psi. \tag{2.1}$$

The total energy of the system is denoted by E_{total} and $\hat{H}$ refers to the Hamiltonian operator, which includes all energy contributions. In the case of a chemical system of nuclei and electrons, $\hat{H}$ takes the following form:[55]

$$\hat{H} = -\sum_{i=1}^{N}\frac{\nabla_i^2}{2} - \sum_{A=1}^{M}\frac{\nabla_A^2}{2M_A} - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}. \tag{2.2}$$

In $\hat{H}$ the following five energy contributions are collected as a sum in the order of their appearance on the right hand side of Equation 2.2:[55]

1. The kinetic energy of the electrons, with i being the indices of the electrons.
2. The kinetic energy of the nuclei, with A being the indices of the nuclei and M_A representing the corresponding atomic mass.
3. The potential energy stemming from the Coulomb attraction between electrons and nuclei, with Z_A referring to the charge of nucleus A and r_{iA} being the distance between electron i and nucleus A.
4. The potential energy from the Coulomb repulsion between electrons, with r_{ij} representing the distance between electron i and electron j.
5. The potential energy due to the Coulomb repulsion between the nuclei, with R_{AB} being the distance between nucleus A and nucleus B.

2.1.1.2 Born-Oppenheimer Approximation

Solving the Schrödinger equation with the Hamiltonian of the form in Equation 2.2 is only possible for the simplest of systems such as a single hydrogen atom. In practical applications, a number of approximations have to be made. Typically, an initial approximation is applied to the Hamiltonian itself, the Born-Oppenheimer approximation. As a nucleus is orders of magnitude heavier than an electron, about 2000 times heavier in the case of hydrogen, the nuclei move much slower than the electrons and are assumed to be stationary in this approximation. Within this approximation, the kinetic energy of the nuclei is zero and the potential energy due to the repulsion of the nuclei is a constant that can be added after solving the Schrödinger equation. By following these steps, one is left with a simplified electronic Hamiltonian $\hat{H}_{\text{elec}}$,[55]

$$\hat{H}_{\text{elec}} = -\sum_{i=1}^{N} \frac{\nabla_i^2}{2} - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}}. \tag{2.3}$$

The electronic Hamiltonian can then be inserted into the Schrödinger equation to give an expression describing the electronic wave function Ψ_{elec} and electronic energy E_{elec},

$$\hat{H}_{\text{elec}}\Psi_{\text{elec}} = E_{\text{elec}}\Psi_{\text{elec}}. \tag{2.4}$$

Solving for the electronic wave function Ψ_{elec} is the goal in most of quantum chemistry and for simplicity the subscript "elec" is dropped in the following.

2.1.1.3 The Variational Principle

Even with the simplified Schrödinger equation (2.4) obtained within the Born-Oppenheimer approximation, finding the exact solution for Ψ is usually not possible and an approximation of Ψ is aimed for. When developing approximate solutions of the Schrödinger equation, a method for assessing the quality of the approximated wave function Ψ_{trial} is paramount. For this, the variational principle is employed, which can be expressed in the following way,[56]

$$E_0 \leq \frac{\int \Psi_{\text{trial}}^* \hat{H} \Psi_{\text{trial}} d\tau}{\int \Psi_{\text{trial}}^* \Psi_{\text{trial}} d\tau}. \tag{2.5}$$

In essence, the principle states that the energy expectation value of any trial wave function Ψ_{trial} is always higher than or equal to the exact ground state energy E_0. As it is not possible to undershoot the energy with an approximate wave function, the lower the energy expectation value of Ψ_{trial} is, the closer the approximation is usually to the exact form of Ψ. Using this relation, it is possible to compare two approximate solutions to the Schrödinger equation and iteratively improve Ψ_{trial}.[56]

2.1.2 The Hartree-Fock Method

While theoretically an infinite search over all possible functions would give the exact form of Ψ for any given Hamiltonian when using the variational principle, this is obviously not possible in any practical application. A major challenge is that the wave function is highly dimensional, with a N electron system having $3N$ spatial dimensions and N spin dimensions. This high dimensionality makes the search for a suitable approximation to

Ψ exceptionally demanding and simplifying the problem of dimensionality is one of the approaches taken within computational chemistry.[57]

2.1.2.1 The Mean-Field Approximation

One way of decreasing the dimensionality of the problem is to disentangle the $4N$ dimensional total wave function into N separate four dimensional functions, describing each electron separately. This single electron approach is, however, generally not conducive for obtaining sufficiently accurate solutions to the Schrödinger equation, as the electrons interact via Coulomb interactions and are further subdued to exchange and correlation effects. However, it is possible to apply the approximation that each electron only reacts to the mean field of all other electrons. With this approximation, each electron can be described through a separate single-particle wave function, which can then all be combined into a single many-particle function to yield Ψ.[57, 55]

2.1.2.2 Orbitals

A wave function describing a single electron is defined as an *orbital.* Orbitals can be further subdivided into spatial orbitals and spin orbitals. A spatial orbital $\psi_i(\mathbf{r})$ is a function of the three dimensional position vector $\mathbf{r}$ in such a way that $|\psi_i(\mathbf{r})|^2$ gives the probability distribution of finding electron i at location $\mathbf{r}$. A spin orbital $\chi_i(\mathbf{x})$ is a function of $\mathbf{x}$ which represents the spin coordinate in addition to the spatial coordinates contained in $\mathbf{r}$. With electrons being fermions, the spin coordinate can be either spin up or spin down and therefore two spin orbitals can be formed from a given spatial orbital.[55]

2.1.2.3 Antisymmetry Principle and Slater Determinants

In order to obtain a total wave function Ψ for a given electronic system, the N separate spin orbitals χ_i must be combined into a single function that depends on all coordinates, $\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x_N})$. The naive approach would be to simply form the product of all spin orbitals, known as the Hartree product $\Psi^{\mathbf{HP}}$,[55]

$$\Psi^{\mathbf{HP}}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x_N}) = \chi_i(\mathbf{x}_1)\chi_2(\mathbf{x}_2)...\chi_\mathbf{N}(\mathbf{x_N}). \tag{2.6}$$

While simple, the Hartree product neglects an important property of the exact wave function, the antisymmertry principle. The antisymmetry principle states that a given many-electron wave function has to be antisymmetric in regard to the interchange of two electrons, i.e. their spin and spatial coordinates $\mathbf{x}$,[55]

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x_i}, \ldots, \mathbf{x_j}, \ldots, \mathbf{x_N}) = -\Psi(\mathbf{x}_1, \ldots, \mathbf{x_j}, \ldots, \mathbf{x_i}, \ldots, \mathbf{x_N}). \tag{2.7}$$

The Hartree product does not fulfill the antisymmetry principle as it generally shows no symmetry in regard to interchange of electrons as can be seen on the following example with two electrons:[55]

$$\Psi^{\mathbf{HP}}(\mathbf{x}_1, \mathbf{x}_2) = \chi_1(\mathbf{x}_1)\chi_2(\mathbf{x}_2) \neq \Psi^{\mathbf{HP}}(\mathbf{x}_2, \mathbf{x}_1) = \chi_1(\mathbf{x}_2)\chi_2(\mathbf{x}_1) \tag{2.8}$$

However the properties of a matrix' determinant can be exploited to take any Hartree product and antisymmetrize it. The resulting function is called a Slater determinant and has the following form for an N electron system,[55]

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x_N}) = \sqrt{\frac{1}{N!}} \begin{vmatrix} \chi_i(\mathbf{x}_1) & \chi_j(\mathbf{x}_1) & \cdots & \chi_k(\mathbf{x}_1) \\ \chi_i(\mathbf{x}_2) & \chi_j(\mathbf{x}_2) & \cdots & \chi_k(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(\mathbf{x}_N) & \chi_j(\mathbf{x}_N) & \cdots & \chi_k(\mathbf{x}_N) \end{vmatrix} . \tag{2.9}$$

A matrix determinant is antisymmetric in regard to the interchange of two rows. Through the way that the matrix is constructed for the Slater determinant, after the exchange of two electrons, the original matrix can be reconstructed by simply interchanging two rows of the matrix. Therefore, the entire determinant is antisymmetric in regard to electron exchange, regardless of the choice of spin orbitals. The term $\sqrt{1/N!}$ then ensures that the resulting wave function remains normalized as long as all spin orbitals are orthonormal.[55]

2.1.2.4 Applying the Mean-Field Approximation

With the definition of spin orbitals and how to combine them into a antisymmetrized wave function being established, it is now possible to combine these concepts with the variational principle to obtain practical approximations to Ψ. When constructing a wave function from a single Slater determinant, the mean-field approximation is implicitly applied. By applying the variational principle to the spin orbitals forming a single Slater determinant, it is possible to find the Slater determinant with the lowest energy expectation value. This, in turn, is the exact solution for an electronic system within the mean-field approximation. The variational freedom lies in the choice of the spin orbitals, which must be adjusted in such a way that the resulting energy expectation value is as low as possible.[55]

2.1.2.5 The Hartree-Fock Equation

In practice, the spin orbitals are solved for through the Hartree-Fock equation, which is an equation corresponding to the Schrödinger equation but only applied to a single spin orbital,[55]

$$f(i)\chi(\mathbf{x_i}) = \epsilon\chi(\mathbf{x_i}). \tag{2.10}$$

Instead of the total energy E and the Hamiltonian $\hat{H}$ in the Schrödinger equation, the orbital energy ϵ and the Fock operator $f(i)$ is used respectively. The Fock operator is an effective one-electron operator that takes the following form,[55]

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + v^{\mathrm{HF}}(i). \tag{2.11}$$

The Fock operator is similar to the Hamiltonian within the Born-Oppenheimer approximation, being the sum of the kinetic energy of the electrons, the electron-nucleus attraction and the electron-electron repulsion. The main difference is that the electron-electron repulsion is approximated as the mean-field potential v^{HF}, acting on electron i by all other electrons.[55]

While the Fock equation is directly dependent on only one of the electrons, specifically the ith electron, the dependence on the average potential of all other electrons still requires information on their state. The equations are therefore solved iteratively through the self-consistent field method. Herein, an initial guess of the electron density is used in the first computation of the spin orbitals. In the next iterations, the electron density from the results of the previous computation is used to compute the new spin orbitals and again receive a new electron density. After some number of iterations, the procedure has converged when the electron density changes less than a set threshold from one iteration to the next. Then the spin orbitals from the final calculation can be combined into a single Slater determinant, which in turn is an approximation of the exact Schrödinger equation.[55]

2.1.2.6 Basis Sets

Computing the Hartree-Fock wave function approximation by hand is only possible for the simplest of systems and for any practical application, the calculation must be performed on a computer system. In this, the challenge arises that solving the Hartree-Fock equations requires solving a differential equation, which is not a straightforward task to encode in a computer program, at least in an efficient manner. Computers are however proficient in linear algebra and many efficient algorithms already exist for it. Therefore, the quantum chemical problem is converted into one of linear algebra by the use of a basis set.[54]

When performing a quantum chemical calculation using a basis set, each spatial orbital ψ_i is represented as a linear combination of M basis functions ϕ_a and M coefficients c_{ai},[54]

$$\psi_i = \sum_a^M c_{ai}\phi_a. \tag{2.12}$$

In essence, any set of basis functions could be used. If the functions form a complete basis, any ψ_i could be reproduced as linear combination and no loss in accuracy would occur. However, a complete basis would have to include an infinite number of functions for most systems, which is not feasible for practical computations. Therefore, a certain degree of basis-set superposition error is accepted in exchange for having a finite number of basis functions. Herein, it is of great importance that the limited number of basis functions can closely reproduce the true form of ψ_i to limit the introduced error.[56]

To form a set of basis functions that is well suited at reproducing the true form of ψ_i, they should model its physical behavior. One such class are the Slater type functions ϕ^{SF} that are centered on the nuclei of the studied molecular system inherently capture a large degree of the underlying physical behavior,[55]

$$\phi^{SF} = e^{-\xi|\mathbf{r}-\mathbf{R}_A|}. \tag{2.13}$$

The function has its maximum at the center of a given nucleus $\mathbf{R}_A$ and decays towards zero when the distance to the nucleus goes towards infinity. Also, the slope of the function aligns well with the underlying physical behavior and can be adjusted through changing ξ to tune the function to a specific system and to form multiple functions for each nucleus. Although well suited physically, Slater-type functions are limited computationally as performing algorithmic computations with them is resource-intensive due to a lack of efficient algorithms for their mathematical treatment. Therefore, often Gaussian-type functions ϕ^{GF} are used as computationally efficient algorithms for solving integrals involving them exist,[55]

$$\phi^{GF} = e^{-\alpha|\mathbf{r}-\mathbf{R}_A|^2}. \tag{2.14}$$

As Gaussian-type functions reproduce the underlying physics to a lesser degree than Slater-type functions, a greater number of the former is required to achieve the accuracy of the latter. One popular compromise is to combine multiple Gaussian-type functions into a single contracted Gaussian function that more closely mimics the behaviors of Slater-type functions while still exhibiting greater computational efficiency.[55]

Many pre-established basis sets are available that have been tuned to the commonly used chemical elements and are available with a differing number of functions per nucleus. The choice of basis set is then dependent on the accuracy requirements and the specifics of the studied system.[58]

2.1.2.7 Post Hartree-Fock Methods

The Hartree-Fock method is already a great leap in the field of computational chemistry. However, the neglect of instantaneous electron-electron correlation produces errors that are not insignificant for the quantitative description of many chemical systems, even if an infinite basis set was to be used.[55]

For the many cases in which additional accuracy is required, more advanced methods have been developed. Many of these methods build upon the single Slater determinant wave function obtained from a Hartree-Fock calculation, such as Møller-Plesset perturbation theory and the Configuration Interaction method. In Møller-Plesset perturbation theory, the initial Hartree-Fock solution is augmented by adding contributions from the electron correlation as a perturbation to the system. In the Configuration Interaction method, the Slater determinant from a Hartree-Fock calculation is augmented with additional Slater determinants into a total wave function as a linear combination of Slater determinants. The additional Slater determinants are formed from virtual orbitals not occupied by electrons, produced within the Hartree-Fock method. The coefficients of the linear combination of Slater determinants are then computed with the variational principle.[56]

These so called post Hartree-Fock methods can greatly increase the accuracy and can even reach the exact non-relativistic ground state within the used basis set. This increased accuracy does come at the cost of computational demand, which can make the computation of even medium sized systems impossible depending on the method used.[54]

2.1.3 Density Functional Theory

An alternative approach to derive approximations of the exact wave function is found in density functional theory (DFT). The idea at the heart of density functional theory is closely related to that of the Hartree-Fock approach, namely, a reduction in dimensionality. The exact wave function of a system with N electrons is dependent on $3N$ spatial coordinates. The total electron density on the other hand is only dependent on the three spatial dimensions, independent of the electron count. Describing a quantum chemical system through its electron density is the main concern of density functional theory in the aim of achieving a high accuracy at a low computational cost.[56, 57]

2.1.3.1 The Electron Density

The wave function $\Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ of an N-electron system fully describes that system, naturally including the distribution of all N electrons. Specifically, the square absolute value of the wave function $|\Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)|^2$ gives the probability of simultaneously finding electron 1 at location $\mathbf{r}_1$, electron 2 at location $\mathbf{r}_2$ and so on. However, this does not achieve any reduction in dimensionality and, therefore, density functional theory is based on a more condensed version of this probability distribution, the total electron density.[57]

The total electron density $\rho(\mathbf{r})$ is the probability of finding any electron at location $\mathbf{r}$. It can be obtained with ease from the normalized wave function $\Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ of any N-electron system through the following relation,[57]

$$\rho(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)|^2 d\mathbf{r}_2 \ldots d\mathbf{r}_N. \tag{2.15}$$

The electron density can be used to compute the number of electrons in a specific area of space by integrating over it. Naturally, when integrating over the whole space the number of electrons N in the system is obtained,[57]

$$\int \rho(\mathbf{r}) d\mathbf{r} = N. \tag{2.16}$$

2.1.3.2 Hohenberg-Kohn Theorem

As shown in Equation 2.15, the electron density can be obtained straightforwardly from a given wave function. Nevertheless, this does not imply that the inverse operation is possible, deriving the wave function from the electron density. If this inverse operation is possible, it would be implied that the electron density is able to fully describe the underlying system, giving a solid groundwork to density functional theory.

In 1964, Hohenberg and Kohn published a formal proof that all ground-state electronic properties can be uniquely determined by the electron density.[59] Apart from the formal proof, an intuitive explanation can be formulated by showing that the electron density contains all information needed to form the Hamiltonian of a given system, from which in turn the wave function could be derived. The information needed to form the Hamiltonian for a chemical system within the Born-Oppenheimer approximation is the number of electrons, the position of the nuclei and their nuclear charge. All of these are contained in the electron density in the following way:[57]

- The number of electrons can be obtained by integrating the electron density over all space as shown in Equation 2.16.

- The position of the nuclei can be identified as the electron density exhibits a cusp at these locations.
- The nuclear charge of each nucleus can be derived by the height of these cusps.

This proof is significant, as it legitimizes the use of the electron density to understand chemical systems as all ground-state properties can be obtained from it. Nevertheless, to gain the benefits of solely utilizing the electron density to describe a system, one needs a method to obtain it without first acquiring the wave function of that system. For this, Hohenberg and Kohn have also made advancements in the form of the Hohenberg-Kohn variational theorem. The theorem states that a given trial electron density function $\rho'(\mathbf{r})$ gives an energy that is greater than or equal to the true ground state energy E_0,[56]

$$E[\rho'(\mathbf{r})] \geq E_0. \tag{2.17}$$

Analogous to deriving approximations to the wave function, the electron density can be developed by starting from a guess and then improving upon it by minimizing the energy. One additional challenge herein is the calculation of the energy. While for a trial wave function computing the energy is trivial, the same cannot be said for a trial electron density. One needs to find the energy functional of the electron density $E[\rho]$ that has been proven by Hohenberg and Kohn to exist, however its explicit form is still elusive.[56, 57]

2.1.3.3 Orbital Free DFT

Developing an energy functional for the electron density $E[\rho]$ is a main concern of density functional theory. Analogously to the electron Hamiltonian described in Equation 2.4, the energy functional should include contributions of the electrons kinetic energy $T_\mathrm{e}[\rho]$, the electron-nucleus attraction $V_\mathrm{ne}[\rho]$ and the electron-electron repulsion $V_\mathrm{ne}[\rho]$,[54]

$$E[\rho] = T_\mathrm{e}[\rho] + V_\mathrm{ne}[\rho] + V_\mathrm{ne}[\rho]. \tag{2.18}$$

Deriving the functional of the electron-nucleus attraction is straightforward by calculating the integral of the classical Coulomb attraction between each of the M nuclei and the electron density over the entire system space,[54]

$$V_\mathrm{ne}[\rho] = -\sum_{A}^{M} \int \frac{Z_A(\mathbf{R}_A)\rho(\mathbf{r})}{|\mathbf{R}_A - \mathbf{r}|} d\mathbf{r}. \tag{2.19}$$

The functional of the electron-electron repulsion is more complex to derive. As a first approximation, a double integral can be calculated over the total system that computes the Coulomb interaction of the electron density with itself,[54]

$$V_{\mathrm{ee}}[\rho] \approx \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \tag{2.20}$$

However, this classical description of the electron-electron repulsion is not sufficient. The total electron density does not give insights from how many electrons and at which proportions the density at any given point originates. Therefore, the functional in Equation 2.20 accounts for electron-electron repulsion of an electron with itself, adding nonphysical contributions to the energy. Forming a functional of the kinetic energy is also challenging due to the missing information about individual electrons in the electron density. Moreover, other quantum mechanical effects such as electron exchange are not covered.[54]

An early attempt to form an energy functional despite the stated challenges is that of the Thomas-Fermi-Dirac (TFD) model.[60] Within the TFD model a uniform electron gas is assumed, giving the following approximate functionals of the kinetic energy $T_{\mathrm{e}}[\rho]$ and the electron-electron repulsion $V_{\mathrm{ee}}[\rho]$,[54]

$$T_{\mathrm{e}}[\rho] \approx \frac{3}{10}(3\pi^2)^{\frac{2}{3}} \int \rho^{\frac{5}{3}}(\mathbf{r}) d\mathbf{r}, \tag{2.21}$$

$$V_{\mathrm{ee}}[\rho] \approx \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - \frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho^{\frac{4}{3}}(\mathbf{r}) d\mathbf{r}. \tag{2.22}$$

While the assumption of the electron density being comprised of a uniform electron gas can be suitable for the valence electrons of some metallic systems, it is problematic for molecular systems or metallic systems with localized electrons. This is most notably illustrated by the TFD model not being able to predict bonding between atoms. Attempts have been made at improving the energy functionals with some success. For example by including corrections accounting for the gradient in the electron density, it is possible to improve the model sufficiently for it to predict atomic bonding. However, all these approaches developed thus far lack accuracy for practical applications, especially compared to wave function based methods.[54]

2.1.3.4 Kohn-Sham Theory

As the construction of a sufficiently accurate energy functional of the pure electron density has proved elusive so far, an alternative proposed by Kohn and Sham has emerged.[61] The core idea of this approach is to initially assume the electrons to be independent and introduce orbitals to preserve the fine grained information on the behavior of individual electrons. Within Kohn-Sham theory, the especially problematic kinetic energy functional can now be constructed with a much higher accuracy through the additional information contained in the orbitals ψ_i,[54]

$$T_{\mathrm{KS}} = \frac{1}{2}\sum_i \int \psi_i^*(\mathbf{r})\nabla_i^2\psi_i(\mathbf{r}). \tag{2.23}$$

The term for the kinetic energy would be exact if the electrons were non-interacting, as assumed in this approach. For typical chemical systems, a small error is introduced, that is however much smaller than that of orbital-free approaches, such as the expression presented in Equation 2.21. To obtain the Coulomb contributions to the total energy, the electron density can be approximately obtained as the sum of the individual electron density of each orbital,[54]

$$\rho(\mathbf{r}) \approx \sum_i |\psi_i(\mathbf{r})|^2. \tag{2.24}$$

The electron density can now be used in the functional of the electron-nucleus attraction and electron-electron repulsion described in equations 2.19 and 2.20, respectively. The small errors inherent in this description are combined into a fourth term, called the exchange-correlation functional $E_{\mathrm{XC}}[\rho]$, to give an exact energy functional of the electron density $E[\rho]$,[54]

$$E[\rho] = T_{\mathrm{KS}} + V_{\mathrm{ne}} + V_{\mathrm{ee}} + E_{\mathrm{XC}}. \tag{2.25}$$

While an exact energy functional is now derived in theory, this is only under the condition that the exchange-correlation functional is known. For most non-trivial systems this is not the case. However, the contribution of E_{XC} is small compared to the other contributions, roughly one order of magnitude smaller. Therefore, the errors in approximations to the true form of E_{XC} have a much lower impact than the error in the kinetic energy of orbital-free DFT. Finding accurate and efficient implementations of the exchange-correlation functional is the core aim of modern density functional theory.[54]

2.1.3.5 Exchange-Correlation Functionals

As the accuracy of Kohn-Sham DFT is directly tied to that of the employed exchange-correlation functional $E_{\mathrm{XC}}[\rho(\mathbf{r})]$, substantial effort is invested in their development and a variety of different approaches have been developed so far. Perhaps the simplest approach is that of the local-density approximation (LDA). Within the LDA, the local electron density is treated as a uniform electron gas, resulting in a significant simplification. The exchange-correlation functional is then defined as the integral over the entire system of the electron density multiplied by the exchange-correlation energy that a single electron would have in a uniform electron gas of the given density $\epsilon_{\mathrm{XC}}[\rho(\mathbf{r})]$,[56, 62]

$$E_{\mathrm{XC}}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\epsilon_{\mathrm{XC}}[\rho(\mathbf{r})]d\mathbf{r}. \tag{2.26}$$

The local-density approximation allows for computationally efficient simulations and can often reach satisfactory accuracy, especially for the prediction of structural properties. However, the achievable accuracy decreases the further the actual electron density deviates from the assumed uniform electron gas, such as in the atomic bonds of molecules.[62]

To improve the description of systems with significant inhomogeneity in the electron density, a non-local correction can be included. The correction is often included through the gradient of the electron density at any given point, giving rise to the class of functionals within the generalized gradient approximation (GGA). These functional can often describe molecular systems much better, especially the bond lengths and binding energies.[62, 56]

A further class of popular functionals explicitly include the results of a Hartree-Fock calulation to obtain the exchange correlation energy. These functionals are called hybrid functionals and typically include a fraction of the Hartree-Fock exchange energy. Through adding the "exact" Hartree-Fock exchange energy, highly accurate calculations can be performed in many cases.[56, 58] One popular choice of hybrid functional is the B3LYP functional used in this work.[63, 64]

2.2 Machine Learning

Computational resources are being employed with great success in real world application across nearly all scientific domains. A major challenge herein is the development of efficient and accurate algorithms that correctly treat the research questions. Taking computational quantum chemistry as an example, great strides have been made on improving the algorithms but at the cost of many decades of intense research. Naturally, the aspiration arises for computers to autonomously develop the algorithms to solve a given problem, in a sense on-the-fly learning how to solve it. This aim is commonly denoted as the field of machine learning. Many excellent resources [65, 66, 67, 68, 69] are available for a detailed study of the involved concepts, therefore only the basic concepts are provided in the following.

2.2.1 Training Regimes

When developing a system that can autonomously "learn" how to solve a given problem, a core question is how to implement the problem into this system. The system must be given enough information to understand what the desired output corresponding to the input data is and how the two relate to each other. To fulfill this task, different strategies are being pursued. These can be categorized into supervised learning, unsupervised learning and reinforcement learning.[65]

2.2.1.1 Supervised Learning

In supervised learning, the problem statement and relevant information is given to the machine learning system in the form of examples. A dataset is prepared with a set of inputs that are each given a label. The label is the set of attributes that one wants the machine learning model to predict, when given the corresponding data point.[65]

For instance, the data points could be molecules and the label is the corresponding toxicity level of that molecule. A well-designed machine learning algorithm then processes the example molecules from the dataset during training to "understand" what makes a molecule toxic and what does not. When a model is sufficiently trained, it is then subsequently possible to get a prediction from the machine learning model on yet unseen data, e.g., whether a novel molecule is toxic or not as illustrated in Figure 2.1. Feeding a model with labeled data to improve it is referred to as *training*, while using a trained model to perform predictions is referred to as *inference*.[66, 65]

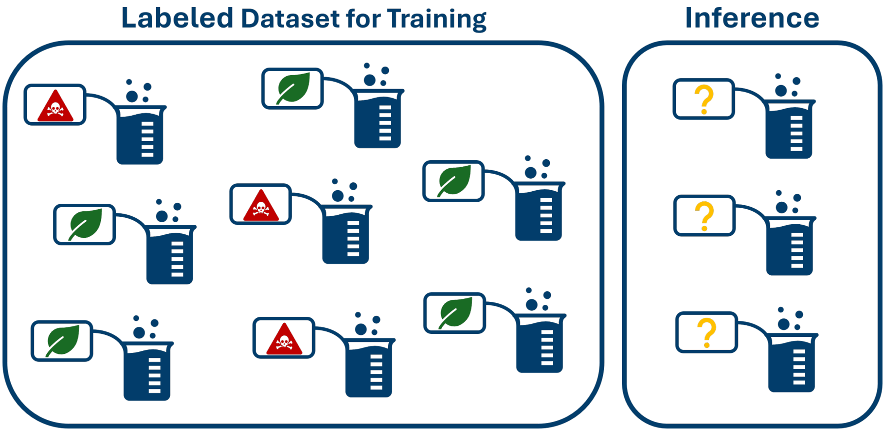


Figure 2.1: Simplified illustration of a supervised learning regime. A dataset of labeled data is given in the form of chemicals and information on their toxicity (left). The machine learning model learns the relationship between the chemical makeup and toxicity using the examples. When sufficiently trained, the machine learning model can be used for inference to predict the toxicity of previously unseen chemicals (right).

This training regime can be used for a wide range of problems. The label can be categorical or numerical. Labeling a chemical as toxic or non-toxic is an example of a categorical label, while the boiling point in °C of that chemical would be an example of a numerical label.[70]

2.2.1.2 Unsupervised Learning

In some applications, a sufficiently large labeled dataset is not practically obtainable. For instance, if a model is sought after that should predict if a molecule is a suitable catalyst for a novel reaction. While data points in the form of molecular structures are easily obtained, labeling them with their catalytic performance is expensive due to each label requiring extensive experimental testing. In other cases it is not even known what the label should be for a given data point. For example, a particle accelerator produces a tremendous amount of raw data, within which data of novel events could be contained. If a machine model is to be used to predict novel events through supervised learning, it would be required to obtain a dataset that is labeled on whether a data point is novel or not. However, if there was a dataset that contains the novel event, it would not be novel anymore and an alternative to supervised learning is needed.[66, 65]

For these cases involving unlabeled datasets, unsupervised learning is an ideal methodology. In an unsupervised learning approach, the algorithm attempts to find patterns and structures in the provided data. Still, a goal needs to be defined for the algorithm to work towards, such as predicting the label in a supervised learning approach. Prominent forms of unsupervised learning can have the following goals:[66]

- Dimensionality Reduction: Reduce the number of features of each data point while preserving as much information as possible, e.g., by combining multiple features into a new, more complex feature.
- Clustering: Divide the data into a certain number of clusters in such a way that the members of each cluster are as similar as possible, a schematic of which is provided in Figure 2.2.
- Anomaly Detection: Understand patterns in the data to identify data points which differ significantly from the majority.

Unsupervised learning methods can be used in conjunction with supervised learning, such as in the case of semi-supervised learning. When a large unlabeled dataset is given, it may be too costly to label enough data points to perform a robust supervised learning approach. To circumvent this problem, the unlabeled dataset can first be structured into clusters through an unsupervised learning approach. Then, under the assumption that all data points in a cluster share the same label, only a single instance from each cluster can be labeled and that label applied to all other data points in the current cluster. Hereby,

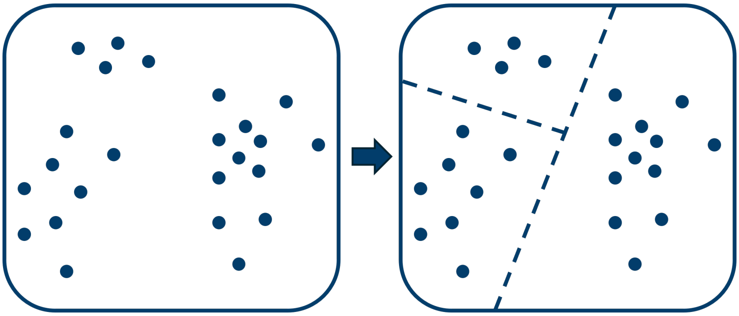


Figure 2.2: Simplified illustration of the goal for an unsupervised learning system performing clustering. An unlabeled dataset is given on the left and the model has the goal of finding a method that divides that data into clusters so that the data points in each cluster are as similar as possible.

a large dataset can be equipped with labels while only having to perform the expensive labeling process for a small fraction of the contained data.[65]

2.2.1.3 Reinforcement Learning

Supervised and unsupervised learning are often applied to large static datasets that are provided to the models. However, it is not always practical or possible to collect a sufficiently large dataset and it might not yet be known what form the data should have to solve a certain problem. In these cases, algorithms based on reinforcement learning are ideal.[66, 71]

To train a model using reinforcement learning, an agent is created that can interact with its environment. This agent may interact with the physical environment, e.g., by controlling a robot arm, or it can act fully in a digital environment. After the agent performs an action in its environment, the outcome is given a score, as illustrated in Figure 2.3. For instance, if an agent is controlling a robotic arm in a robotic laboratory, it may be scored based on how much of a target molecule is synthesized. In response to the score given, the underlying model controlling the agent is updated, e.g., reinforcing the current behavior upon receiving a high score or making changes upon receiving a low score.[71]

Within this training regime, no initial data is needed. As long as a way of scoring the outcome of a given set of actions is provided, it is possible to train an agent to maximize this score. Designing the environment may however be challenging especially for actions to be taken in physical space. To reach sufficient training times, it may be required to run the training in a simulation of the physical world, in which case great care must be taken that all relevant aspects of the environment are adequately simulated.[72]

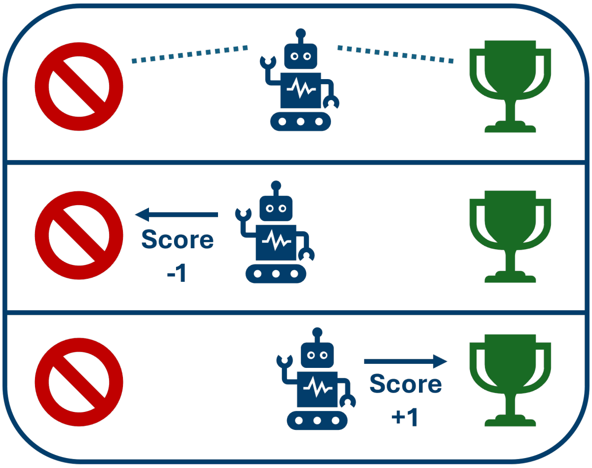


Figure 2.3: Simplified methodology of a reinforcement training. An agent represented as a robot interacts with the environment. Depending on the desirability of the outcome, a low or high score is given and the policy by which the agent acts is updated accordingly.

2.2.2 Model Architectures

The training methods, such as supervised or reinforcement learning, specify a goal for a machine learning model to achieve. This goal must then be used to train an underlying model to perform the desired task. For the architecture of this underlying model, a vast range of possibilities are available. The architectures range from a simple linear regression to complex deep neural networks with billions of parameters. In the following, some of the most important model architectures are described.[65]

2.2.2.1 Linear Regression

One of the simplest machine learning models is that of a linear regression model. As the name already suggests, a linear regression model constructs a linear relationship between the input data and predicted output data. The construction of a linear regression model requires a set of N real-valued labeled examples $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ with the D-dimensional feature vectors $\mathbf{x}_i$ and their associated target values y_i. The model then calculates a linear combination of the features in example $\mathbf{x}$,[65]

$$f_{\mathbf{w},b}(\mathbf{x}) = \mathbf{w}\mathbf{x} + b. \tag{2.27}$$

The model is parameterized by both the real number b and the D-dimension vector of real numbers $\mathbf{w}$. For a given choice of parameters, the model can be used to predict the target value y for a given $\mathbf{x}$ through $y \leftarrow f_{\mathbf{w},b}(\mathbf{x})$. Naturally, the optimal choice of b and $\mathbf{w}$

should be made in such a way as to maximize the accuracy of the prediction. The quality of the predictions must be quantified in order to maximize it. Typically, the performance of a linear regression is expressed as a loss function that sums the squares of the errors in the predictions, the mean squared error (MSE), given by,[66, 65]

$$\text{MSE} = \frac{1}{N}\sum_i (f_{\mathbf{w},b}(\mathbf{x}_i) - y_i)^2. \tag{2.28}$$

For minimizing the MSE of a linear regression model, a closed form solution is known for obtaining the optimal values for the parameters in the form of the *Normal Equation*. For performance measures other than the MSE, a generic optimization algorithm is often used to obtain the optimal values for b and $\mathbf{w}$, with *Gradient Descent* being a popular choice. In *Gradient Descent*, the parameters are initialized with random values and the local gradient of the performance measure function is determined with regard to the parameters. The parameters are then adjusted in the direction of the descending gradient in an iterative manner until a satisfactory performance is reached or the gradient diminishes.[66]

2.2.2.2 Decision Trees and Random Forests

While linear regression models can be powerful and computationally efficient tools, they reach their limit when the underlying structure of the data becomes too complex. Machine learning models that are based on a decision tree architecture can contain an arbitrary number of parameters to be adjusted and can therefore reproduce very complex relationships within the data.[66, 65]

In principle, a decision tree is a multi-layer workflow, through which input data are sent and sorted by a true or false question at each layer. After travelling through all layers of the decision tree, an output is given to each data point, depending on the location of the decision tree in which it ends. In Figure 2.4, a simple example is given of a decision tree that predicts the yield of a chemical reaction based on reaction temperature and time. First, the input data is sorted by whether the reaction temperature is above or below 100 °C. For a temperature above 100 °C, the yield is immediately predicted to be 0 %. In the case of a temperature below or equal to 100 °C, the data is further sorted by the reaction time, predicting a yield of 85 % or 75 % for a reaction time above and below 24 hours, respectively.[65]

The number of layers and sorting steps of a decision tree can be infinite in principle, allowing the model to learn complex relationships between input and output data. Methods such as the classification and regression tree (CART) algorithm[73] are available to "grow" a decision tree when given a labeled dataset. The algorithm determines which fea-

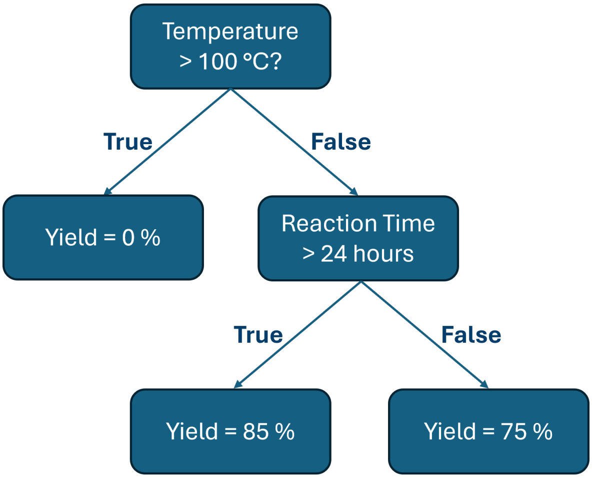


Figure 2.4: Example of a simple decision tree that predicts the yield of a given reaction based on reaction time and temperature. The data points are first sorted depending on if their temperature is above or below 100 °C. In the case of a temperature above 100 °C a yield of 0 % is predicted. In the case of a temperature below 10 °C, the data is further sorted by the reaction time, resulting in predicted yields of 75 % or 85 %.

ture of the input data best splits the total data according to their label at the first layer, repeating the process for the remaining subset of the data set at each deeper layer. Additional performance can be gained by combining multiple decision trees into a single model and combining their individual predictions into a single prediction. This combination of decision trees into a single model is known as a random forest model.[66]

2.2.2.3 Artificial Neural Networks

Artificial neural networks are a popular class of machine learning models due to their versatility and ability to perform highly complex tasks. As the name suggests, artificial neural networks are inspired by the neural networks found in nature, i.e. in the human brain. An artificial neural network is a model $f_{\mathrm{NN}}(\mathbf{x})$ that processes a D-dimensional feature vector $\mathbf{x}$ to give an output y,[65]

$$y = f_{\mathrm{NN}}(\mathbf{x}). \tag{2.29}$$

The input data $\mathbf{x}$ are inserted at the input layer and is processed by sending them through highly connected network of neurons before eventually reaching the output layer where the output data y are extracted. This process of passing data from one layer to the next

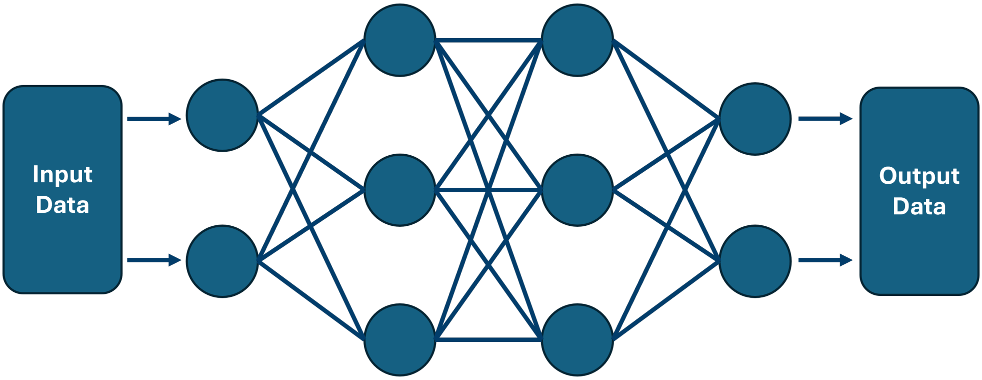


Figure 2.5: Schematic of an artificial neural network with two hidden layers. The input data is passed through the network from left to right along the path of the arrows.

can be described as a nested function,[65]

$$y = f_{\mathrm{NN}}(\mathbf{x}) = f_3(\boldsymbol{f}_2(\boldsymbol{f}_1(\mathbf{x}))). \tag{2.30}$$

The inner functions of the nested function, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ in the case of Equation 2.30, return a vector as follows,[65]

$$\boldsymbol{f}_l(\mathbf{z}) = g_l(\mathbf{W}_l z + \mathbf{b}_l). \tag{2.31}$$

The index l denotes which layer is currently processed. The D-dimensional vector $\mathbf{z}$ is equal to $\mathbf{x}$ in the innermost function and the output of the previous function otherwise. For each layer, the function is parameterized by the weight matrix $\mathbf{W}_l$ and the bias vector $\mathbf{b}_l$. The parameters $\mathbf{W}_l$ and $\mathbf{b}_l$ are learned in order to perform the desired prediction. To ensure that also non-linear relationships can be learned, a non-linear activation function g_l is included.[65]

A schematic of a simple artificial neural network is shown in Figure 2.5. The exact shape of a neural network can be constructed flexibly. A *Perceptron* [74] is among the simplest of cases, being made up of only an input layer that is directly connected to an output layer. Introducing one or more hidden layer in between input and output layers greatly increases the complexity of data processing that is possible within the network. The connectivity between the neurons is also flexible. In the simplest case of a feed-forward network the neurons of one layer are only connected to those of the following layer. Additional complexity can be added by connecting neurons to previous layers or even within one layer in the case of recurrent neural networks. After constructing an artificial neural network, a vast range of data-processing tasks can be performed by the network depending on the choice of $\mathbf{W}_l$ and $\mathbf{b}_l$. These parameters are usually determined by training the network to reproduce the behavior within a training dataset. In a process called back propagation, the gradient of the performance measure function is computed for each parameter of the model and then the parameters are iteratively adjusted in order to minimize the error of the model's prediction.[66]

Chapter 3

Identification of a Computational Stability Descriptor

In order to accelerate the discovery of alkaline stable compounds through computational means, the relevant degradation reactions must be thoroughly understood. The impact of the specific molecular structure on the degradation rate must be uncovered and efficient computational descriptors of the alkaline stability identified.

First computational insights on the degradation mechanistics of imidazolium-based compounds have already been reported in literature. Through calculation of the energy barriers of the degradation reaction of selected compounds, it was found that the ring opening reaction is especially relevant due to its comparatively low energy barrier.[42, 75] It has also been reported that the computed LUMO energy can be correlated to alkaline stability, with a higher LUMO energy being associated with greater stability.[47, 75] In another study, the spatial distribution of the LUMO was additionally taken into account, suggesting a higher degradation rate when concentrated at the C-2 site.[76] Still, a comprehensive study of both the energetic and electronic effects on alkaline stability and a comparison to a varied sample of experimentally tested imidazolium-based compounds is necessary.

A challenge for comparing computational findings to the experimental stability for a wide range of compounds is the testing regime. Experimental stability data reported in literature were generally not generated following a standardized testing protocol. Differences exist in chemical conditions under which degradation tests have been performed, with variations in temperature, concentration, solvent, and hydration level.[41] Additionally, the manner in which the degree of degradation is monitored during the testing procedure is not consistent. The different methods include monitoring the remaining cationic material through ^{1}H NMR spectroscopy, ion exchange capacity, conductivity and gravimetry.[41]

One set of structurally diverse imidazolium-based compounds that had been tested in an identical testing regime were reported by Fan et al.[32] These 11 compounds had been all tested for their alkaline stability in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether. Their chemical structures are diverse and are based on simple imidazolium, benzimidazolium and bis-arylimidazolium. They are depicted in Figure 3.1 along with the half life in the KOH degradation test reported by Fan et al.

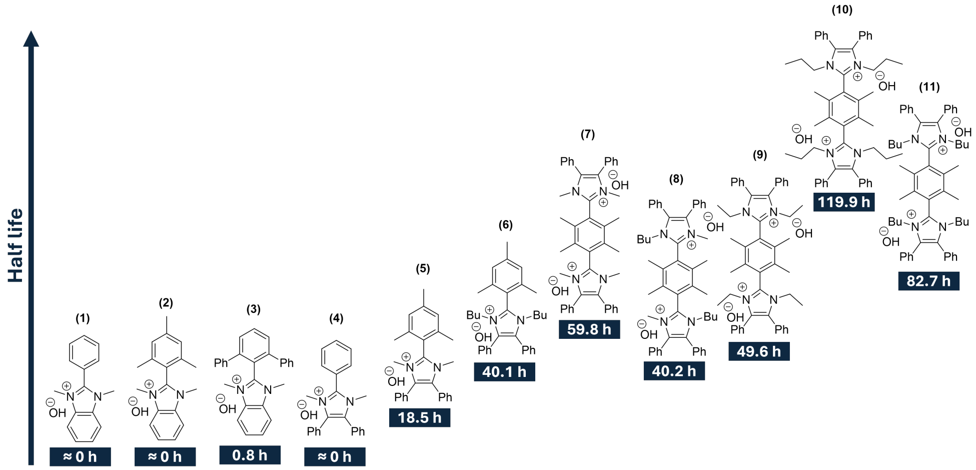


Figure 3.1: Chemical structure of imidazolium-based compounds and their half life in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether at room temperature as reported by Fan et al.[32, 77]

In the work described in this chapter, the imidazolium-based molecules depicted in Figure 3.1 were used for an in-depth computational investigation of the factors contributing to their alkaline stability. Computational results were compared to experimental reports and a computationally efficient descriptor of alkaline stability was thereby identified. The work described in this chapter has been published in *Macromolecules*[77], from which the Figures in this chapter have been reproduced.

3.1 Computational Details

Version 5.0.3 of the ORCA quantum chemistry package[78] was used for all *ab initio* calculations in conjunction with the 6-311G(d,p) basis set.[79, 80] For the DFT calculations, the B3LYP functional[63, 64] was employed with an implicit water solvation environment in the form of the conductor-like polarizable continuum model.[81]

The 3D coordinates of the non-degraded imidazolium molecules studied in this chapter were generated with version 2022.9.4 of RDKIT[82] from their corresponding SMILES representation. To ensure that a meaningful starting conformer was used, a conformer scan was performed with the ”–quick” parameter preset of version 2.12 of CREST[83, 84] in an implicit water solvation environment based on the analytical linearized Poisson-Boltzmann (ALPB) model.

The following methodology was used to compute the energetics of the degradation reactions for each molecule:

1. A DFT level geometry optimization was performed on the lowest energy conformation found for each non-degraded compound.
2. A single hydroxide anion was manually added about 5 Å above the imdiazolium plane and an additional DFT level geometry optimization was performed. The resulting structures served as the reference system for each compound.
3. Starting from the reference structures, the degradation products and intermediates were manually created and a geometry optimization was performed for each.
4. The optimized structures were used in nudged elastic band calculations[85] at the DFT level to generate the transition state of the corresponding degradation reactions.
5. The Gibbs free energy of each structure generated in the previous steps was computed through a frequency analysis. It was ensured that all reaction educts, intermediates and products show no imaginary frequencies while the transition states show exactly one imaginary frequency and this frequency aligns with the reaction coordinate of the studied degradation pathway.
6. The Gibbs free energy is refined for each optimized structure by performing a DLPNO-CCSD(T) single-point calculation[86] on each, to replace only the electronic energy component of the free energy.

3.2 Reaction Energetics

The reaction energetics of the two most relevant and competing degradation pathways, depicted in Figure 1.6, were computed for the 11 compounds. The dealkylation of one of the nitrogen substituents through a $\mathrm{S_N2}$ reaction (Figure 1.6a) and the ring opening reaction via a hydroxide attack on C-2 (Figure 1.6b). For the dealkylation, the Gibbs energy of activation $\Delta G^{\ddagger}_{\mathrm{DA}}$ and Gibbs energy change ΔG_{DA} were computed. For the ring opening reaction, the Gibbs energy of activation $\Delta G^{\ddagger}_{\mathrm{HA}}$ and Gibbs energy change ΔG_{HA} was computed for the initial hydroxide attack along with the total Gibbs energy change of the full ring opening reaction ΔG_{RO}. Computed energetics of degradation reactions are reported together with experimental half lifes reported by Fan et al.[32] in Table 3.1.

Table 3.1: Degradation reaction energetics for the dealkylation and ring opening reaction of the compounds depicted in Figure 3.1. The half life is that reported by Fan et al.[32] for a degradation test in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether at room temperature. The highest and lowest value of each data column is highlighted in bold for added clarity.[77]

Mol	ΔG_{DA} **[kcal/mol]**	$\Delta G^{\ddagger}_{DA}$ **[kcal/mol]**	ΔG_{HA} **[kcal/mol]**	$\Delta G^{\ddagger}_{HA}$ **[kcal/mol]**	ΔG_{RO} **[kcal/mol]**	**Half Life [h]**
1	-33.38	25.01	**-26.30**	**1.34**	**-31.93**	**0**
2	-31.58	25.20	-14.96	9.72	-30.53	**0**
3	-32.39	25.65	-15.89	15.04	-29.12	0.8
4	-30.96	26.62	-11.62	5.56	-22.24	**0**
5	-29.12	27.51	-2.10	15.28	-19.46	18.5
6	**-33.86**	**23.61**	-1.98	12.74	-21.45	40.1
7	-29.16	29.08	0.47	16.09	-20.17	59.8
8	-30.53	24.12	0.65	15.80	-22.11	40.2
9	-32.09	28.56	2.36	18.00	-20.52	49.6
10	**-28.47**	**30.42**	**7.02**	**20.44**	-15.08	**119.9**
11	-29.15	29.77	6.26	19.02	**-14.88**	82.7

The dealkylation reaction exhibits Gibbs energies of activation ranging from 23.6 to 30.4 kcal/mol. The values for the hydroxide attack of C-2 range from 1.3 to 20.4 kcal/mol. The hydroxide attack on C-2 requiring a lower Gibbs energy of activation compared to the dealkylation indicates that it is the relevant degradation reaction for all studied compounds. The dependence of the energies on the chemical structure is also much stronger for the C-2 hydroxide attack than for the dealkylation. The strong impact of the chemical structure on the relevant degradation reaction is a promising property, as it suggests a high potential gain in alkaline stability if a suitable chemical structure is designed.

The reaction energetics of the dealkylation reaction do not show a significant correlation to the experimental stability measurements, further strengthening the assumption that it is not the relevant degradation pathway for these compounds. Contrarily, the energetics of the C-2 hydroxide attack appear to be strongly linked to the experimental half life measurements. Both ΔG_{HA} and $\Delta G^{\ddagger}_{HA}$ are correlated to the experimental stability, which can be explained by the two having a Bell-Evans-Polanyi[87, 88] type correlation to each other. This correlation can be seen best when plotting ΔG_{HA} and $\Delta G^{\ddagger}_{HA}$ in a scatter plot in Figure 3.2.

As both ΔG_{HA} and $\Delta G^{\ddagger}_{HA}$ are correlated to the alkaline stability of the tested imidazolium-based compounds, both could be used as a computational stability descriptor. For the

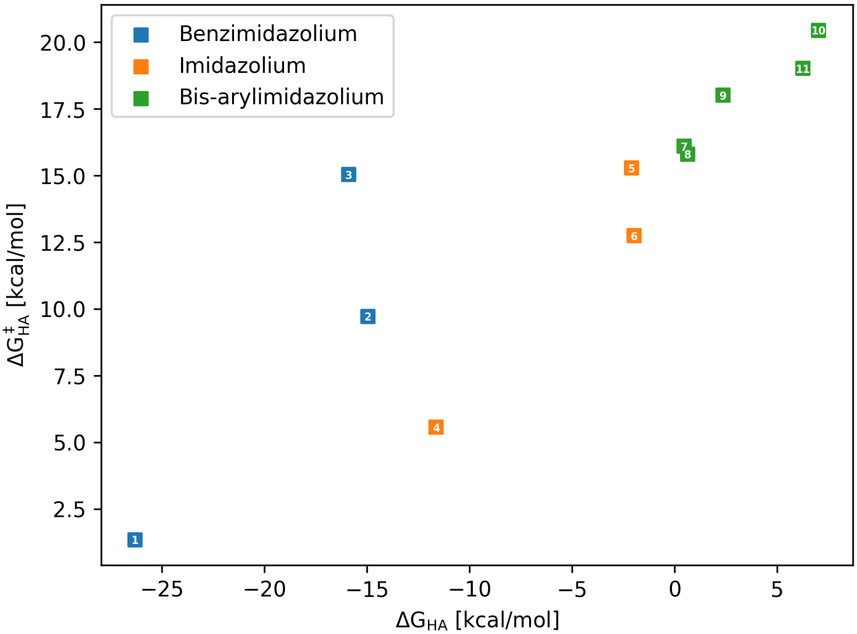


Figure 3.2: Scatter plot of the Gibbs energy of activation of the C-2 hydroxide attack $\Delta G^{\ddagger}_{HA}$ versus the Gibbs free energy change of the C-2 hydroxide attack ΔG_{HA}. The molecule number is given for each data point as provided in Figure 3.1.[77]

exploration of the vast chemical space of imidazolium-based compounds, ΔG_{HA} is more suitable as its calculation requires significantly less computational resources and is easier to automate. The calculation of $\Delta G^{\ddagger}_{HA}$ requires finding the transition state of the reaction, which is computationally costly and often requires manual intervention and verification. For the calculation of ΔG_{HA}, only the optimized structure of the reaction product and starting material is needed.

To quantitatively asses the performance of ΔG_{HA} as a computational stability descriptor, it is plotted in a scatter plot versus the experimental half life reported by Fan et al. in Figure 3.3.

The scatter plot in Figure 3.3a of ΔG_{HA} versus the reported experimental half life shows a strong exponential correlation. When taking the natural logarithm of the experimental half life, a linear regression can be performed. The linear regression is shown in Figure 3.3b, it has a coefficient of correlation R^2 of 0.943. ΔG_{HA} is hereby shown to be well equipped to quantitatively compute the alkaline stability of a diverse set of imidazolium-based compounds while imposing only a moderate computational cost compared to computation of the Gibbs energy of activation.

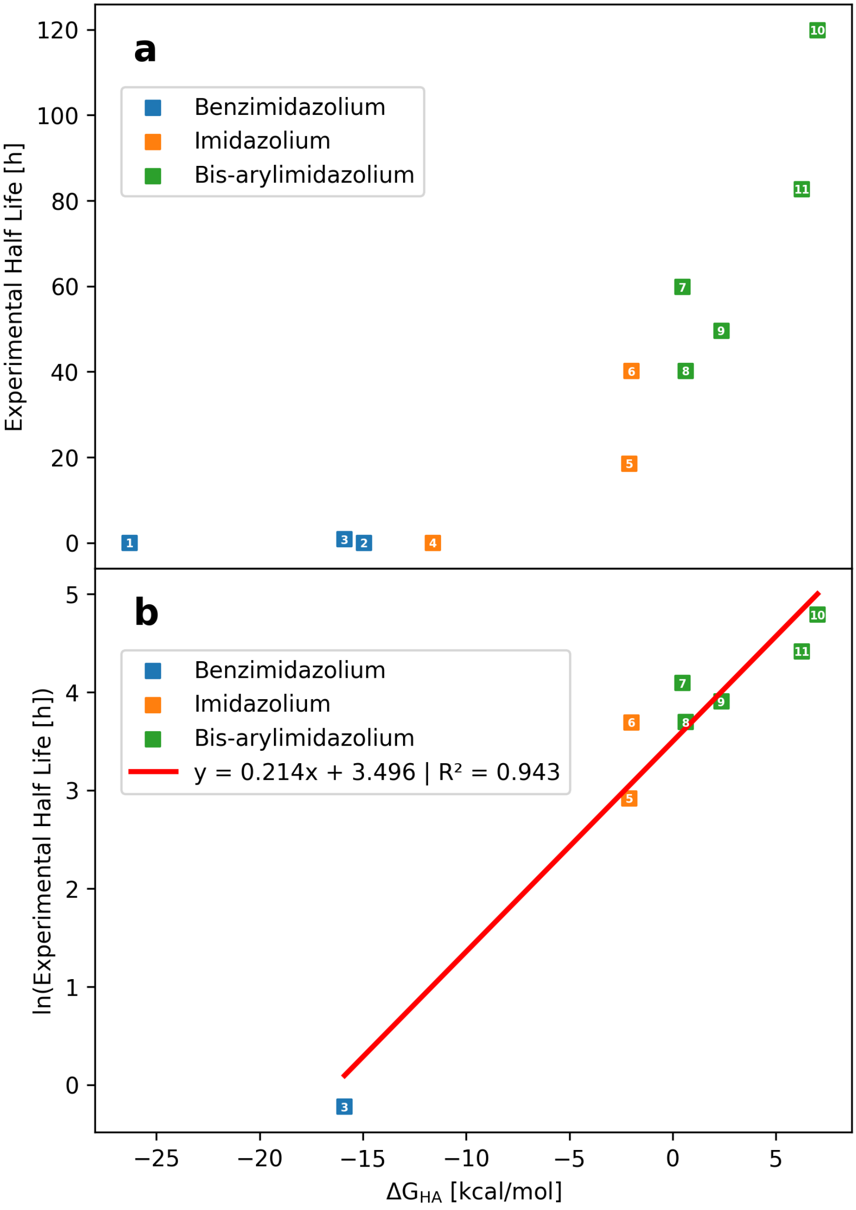


Figure 3.3: Scatter plot of the Gibbs free energy change of the C-2 hydroxide attack ΔG_{HA} versus the half life (a: unmodified, b: its natural logarithm) in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether at room temperature as reported by Fan et al.[32] The molecule number is given for each data point as provided in Figure 3.1. For compounds **1**, **2** and **4** the natural logarithm could not be computed as it is undefined for the reported half life of 0 hours.[77]

3.3 Electronic Properties

As shown in the previous section 3.2, degradation reaction energetics can have good agreement with experimental stability measurements. For the Gibbs energy change, the computational cost is also only moderate, as only the degradation product and starting material need to be computed. However, significantly lower computational demand would be required for the calculation of most electronic properties as they typically only require a single point calculation of the imidazolium structure.

As the LUMO energy of imidazolium-based compound has been proposed numerous times as being correlated to alkaline stability[45, 47, 48, 89], it is evaluated for the compounds studied in this work. In Figure 3.4, the LUMO energy is plotted against the experimental half life measurements reported by Fan et al.

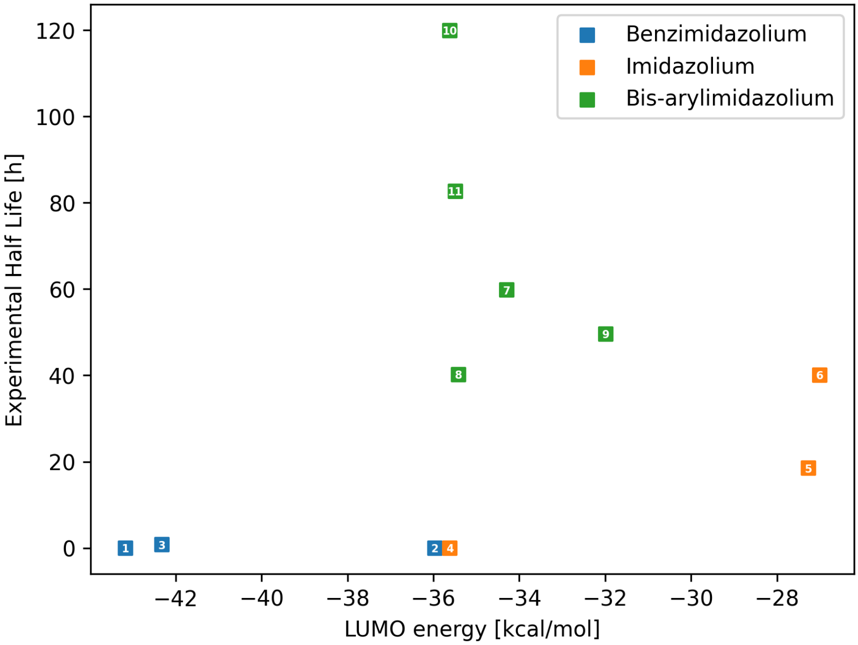


Figure 3.4: Scatter plot of the LUMO energy versus the half life in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether at room temperature as reported by Fan et al.[32] The molecule number is given for each data point as provided in Figure 3.1.[77]

As can be seen in Figure 3.4, the LUMO energy is not correlated to the experimental stability of the compounds studied in this work. This is likely due to the large variation in the chemical structure for the sample studied in this work compared to those reported in literature.

It has also been reported by Lee et al.[76], that the LUMO energy alone is not a sufficient descriptor of alkaline stability and that an analysis of the LUMO distribution is necessary. Especially the distribution of the LUMO around the reactive C-2 site is relevant.

Analyzing the LUMO distribution through an isosurface analysis, like reported by Lee et al., is however unpractical for a computational stability descriptor. Assigning a quantitative value to the LUMO isosurface, as needed for a descriptor, would typically require manual analysis. For an automated analysis of the isosurface, advanced and potentially error prone techniques, such as machine learning, would be required. A more straightforward methodology is utilized in this work in the form of Mulliken frontier molecular orbital population analysis (FMOPA).

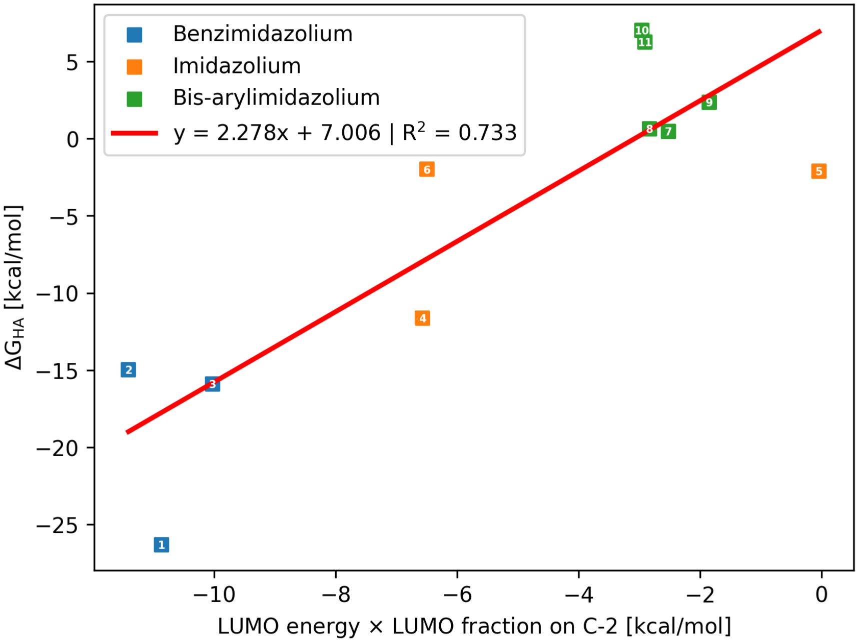


Figure 3.5: Scatter plot of the LUMO energy multiplied with the LUMO fraction on C-2 versus the half life in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether at room temperature as reported by Fan et al.[32] The molecule number is given for each data point as provided in Figure 3.1.[77]

If a FMOPA is applied to the LUMO, the LUMO distribution is given in a machine readable format. The LUMO is distributed to all atoms of the given structure, assigning each atom a fraction of the LUMO localized on it. From this, the LUMO fraction on the C-2 atom can be automatically extracted for a given compound as it is the most relevant atom to the C-2 hydroxide attack. When combining the LUMO fraction on C-2 with the LUMO energy, the correlation with the experimental half life can be improved over the LUMO energy alone and is moderately correlated with ΔG_{HA}. To illustrate this, the product of the LUMO fraction on C-2 and the LUMO energy is plotted versus the experimental half life in Figure 3.5 to give a R^2 of 0.733.

The correlation of the LUMO-based descriptor is only moderate, as it still neglects important aspects affecting the degradation reactions. Most importantly, the steric effects included in the energetic descriptors are fully neglected. However, the LUMO-based descriptor comes with a significantly lower computational cost and can be used as a first assessment before utilizing the more costly energetic-based descriptor.

3.4 Structure-Stability Relationship

With both an electronic and energetic descriptor being established, an in-depth exploration of the chemical space of imidazolium-based compounds can be undertaken. However, even with just the knowledge of the factors contributing to the alkaline stability, at least qualitative insights into the structure-stability relationship can be made. The factors contributing to a high ΔG_{HA} are of interest as they are correlated with an increase in alkaline stability for imidazolium-based compounds.

A major factor impacting ΔG_{HA} is steric strain in the C-2 hydroxide attack product, with increased steric strain leading to a higher ΔG_{HA} and, therefore, higher alkaline stability. This effect can be seen best when comparing compound **1** and **2** as they differ only in their C-2 substituent. Compound **1** is substituted with a phenyl group while compound **2** is substituted by the significantly more sterically demanding mesityl group. The different substituents lead to molecule **2** having a ΔG_{HA} value that is about 11.3 kcal/mol higher. The impact on the alkaline stability is obscured due to the stability being too low to measure in the harsh conditions of an ultra-dry KOH solution, resulting in a half life of zero hours reported by Fan et al. for both. However, Fan et al. also reported stability measurements in a much milder 3M NaOD/D_2O/CD_3OD solution at 80 °C that shows the superior alkaline stability of compound **2** over **1** with half lifes of 436 hours and < 0.1 hours, respectively.

The steric effects are due to direct interference of the mesityl's methyl groups with the attacking hydroxide and an indirect destabilization of the hydroxide attack product. The plane of the mesityl group is rotated about 90° in reference to the plane of the imidazolium ring. Thereby, the two ortho methyl groups of the mesityl are directly above and below the C-2 atom, interfering with the attacking hydroxide group, as can be seen in Figure 3.6.

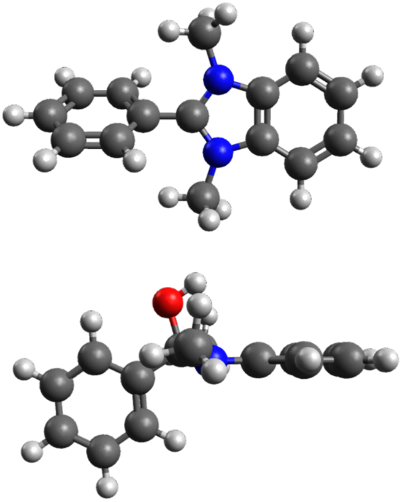
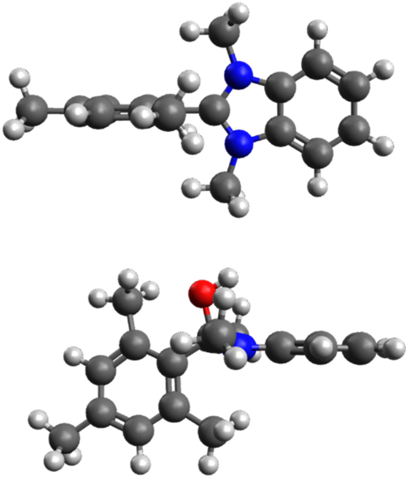


Figure 3.6: Molecular structure of compounds **1** (left) and compound **2** (right) before hydroxide attack as an unperturbed cation (top) and after C-2 hydroxide attack (bottom). The mesityl group of compound **2** exerts a higher degree of steric interference with the hydroxide than the phenyl group of compound **1**.[77]



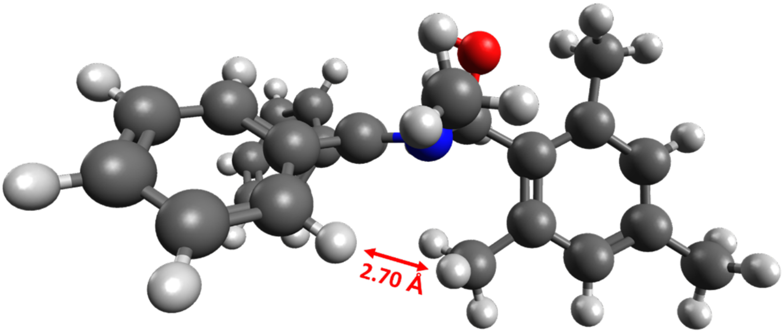


Figure 3.7: Molecular structure of the C-2 hydroxide attack product of compound **5**. As the hybridisation of the C-2 atom is changed from the planar sp^2 to the tetrahedral sp^3, the C-2 substituent is pushed towards the imidazolium ring and its substituents. The distance between the two closest atoms of the C-2 and C-4 substituent of 2.70 Å is depicted as a red double arrow.[77]

Apart from direct steric interference with the attacking hydroxide, a steric destabilization of the hydroxide attack product also leads to an increase in ΔG_{HA}. As an unperturbed cation, the three bonds of the C-2 atom lie in one plane due to sp^2 hybridisation. Upon hydroxide attack, the C-2 atom transitions into a sp^3 hybridisation and the four bonds form an approximate tetrahedron. Through the sp^3 hybridisation, the C-2 substituent is tilted towards the imidazolium ring and its substituents, leading to steric interference.

This can be clearly seen in the example of the C-2 hydroxide attack product of compound **5** depicted in Figure 3.7. This effect has significant implications on the impact of the C-4 and C-5 substituents on the alkaline stability. Even though they are located too far away from the C-2 site to have a direct steric protective effect versus hydroxide attacks, they still protect the imidazolium by sterically destabilizing the attack product.

The compounds **7** to **11** belonging to the group of bis-arylimidazolium show an exceptionally high alkaline stability. To a large degree, the high stability can be attributed to the comprehensive protection by the sterically demanding substituents, especially the central methylated aromatic ring. Additionally, the structure leads to stabilizing electronic effects. The LUMO distribution is shifted away from the imidazolium rings in favor of the central aromatic ring. In conjunction with the LUMO being diluted across two imidazolium rings, this leads to an average LUMO fraction on each C-2 atom of 7.52 % for the compounds based on bis-arylimidazolium (compounds **7** to **11**) versus an average of 20.5 % across the other compounds **1** to **6**.

3.5 Summary and Conclusions

Through the manual computational study of a sample of (benz)imidazolium-based compounds with a range of different chemical structures, a computationally obtainable stability descriptor was identified in the Gibbs free energy change of the C-2 hydroxide attack reaction ΔG_{HA}. The product of the LUMO energy and the fraction of the LUMO localized on the C-2 atom was identified as a less accurate but computationally significantly cheaper stability descriptor. The identified descriptors can be applied to evaluate a novel imidazolium-based compound before performing costly experimental testing. Additionally, the descriptors give comparatively specific guidelines on how to improve the stability of a given compound, e.g., by augmenting the chemical structure to destabilize the C-2 hydroxide attack product or moving the LUMO distribution away from the C-2 site. Lastly, the identified descriptors allow the computational assessment of alkaline stability for a molecular dataset of a vast range of imidazolium-based compounds to identify additional structure-stability relationships and promising materials.

The computation of the energetic descriptor was performed manually, and is not yet suitable for high throughput scanning of compounds. To this end, the computational method needs to be streamlined and automated, which is the subject of the following Chapter 4.

Chapter 4

Molecular Dataset and Experimental Validation

In the undertaking to unravel the intricate relationship between the chemical structure of an imidazolium-based compound and its alkaline stability, access to a diverse dataset of structures and their alkaline stability is of great value. While experimentally constructing such a dataset is only viable for a small number of molecules, computational methods allow for the stability assessment of a large number of compounds. Using the previously identified descriptor of the Gibbs free energy change of the C-2 hydroxide attack ΔG_{HA}, it is possible to compute the stability for thousands of compounds with high accuracy. This chapter describes the automated computation of ΔG_{HA}. To ensure the accuracy of the dataset, five compounds were selected to be tested experimentally for their alkaline stability to evaluate the computed stability of ΔG_{HA}.

4.1 Computational Methodology

The alkaline stability descriptor ΔG_{HA} covered in the previous chapter was initially computed by manually creating the chemical structure of both the starting materials and the reaction products. The electronic energies were also computed through costly DLPNO-CCSD(T) calculations. To compute ΔG_{HA} for thousands of compounds, the process should be automated and the computational cost for each compound should be reduced as much as possible while still retaining a high degree of accuracy. To achieve both of these aims, a number of alterations have been made to the computational methodology compared to the one used in the previous Chapter 3. Costly DLPNO-CCSD(T) calculations were eliminated in favor of the electronic energy given by the DFT calculations. Also, the imidazolium cation is computed in a separate calculation from the hydroxide anion for the educt state, thereby eliminating a lengthy geometry optimization when hydroxide is placed close to the imidazolium and to simplify the automatization.

A significant challenge in automatizing the ΔG_{HA} calculation is the consistency of the conformers. If the educt and product states are simply generated independently, it often occurs that the conformational states of the substituents are significantly different to each other, in a way that would not occur during the reaction. The changes in conformation then contribute a significant and partially random energetic contribution to ΔG_{HA}, which leads to a loss in accuracy.

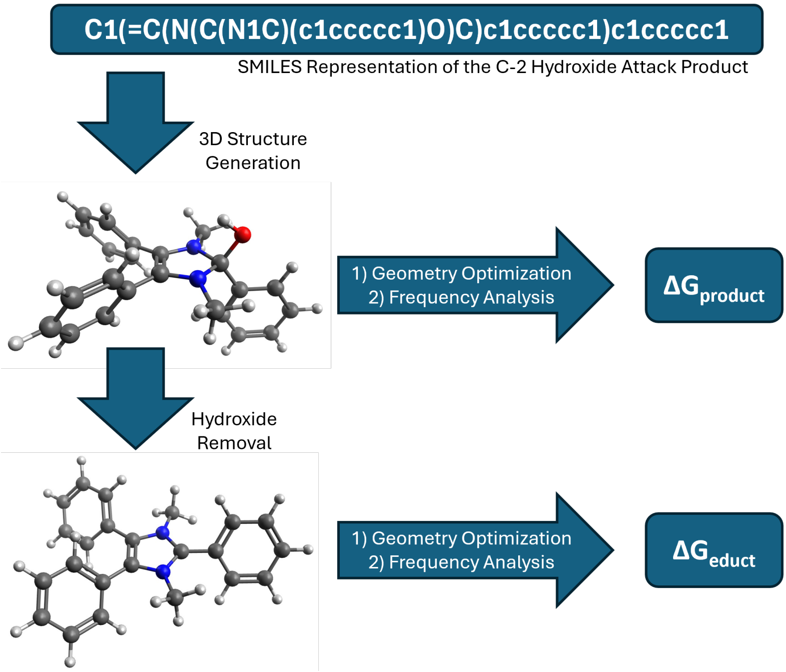


Figure 4.1: Simplified overview of the computational methodology used to automatically generate the unperturbed imidazolium and its C-2 hydroxide attack product while ensuring the conformational states of the substituents are not unphysically altered between educt and product structure.

To ensure both the educt and product structure of the C-2 hydroxide attack are in a comparable conformational state while keeping the process fully automated, the computational methodology illustrated in Figure 4.1 is used. Initially, the SMILES representation of the product structure is obtained and a 3D structure is generated from it using a conformer search. The hydroxide is then removed from the product structure to obtain the educt structure after a geometry optimization, leaving the conformational states almost fully intact.

By then computing the Gibbs free energy of the product structure G_{product}, the educt structure G_{educt} and a single hydroxide anion G_{OH}, the descriptor ΔG_{HA} can be computed through

$$\Delta G_{\text{HA}} = G_{\text{product}} - (G_{\text{educt}} + G_{\text{OH}}). \tag{4.1}$$

4.1.1 Computational Details

The SMILES representation of the product structure was automatically generated with a custom Python code. From the SMILES representation, the initial 3D structure was generated with the ETKDGv3 algorithm implemented in version 2023.09.4 of RDKIT using the random seed ”1” and the ”useRandomCoords” option.[82] An initial force field geometry optimization was performed in RDKIT with the MMFF94 force field.[90]. From that structure, the conformer with the lowest energy was selected from a conformer search using version 2.12 of CREST and the ”squick” preset.[83, 84] A geometry optimization was then performed from the product conformer using the GFN2 method in an implicit water solvation environment through the linearized Poisson-Boltzmann model contained in version 6.6.1 of xTB.[91] Subsequently, a frequency analysis was performed at the GFN2 level and in case of imaginary frequencies, a new geometry that is slightly distorted along the imaginary frequency was used as the starting point for the geometry optimization; the process was repeated until the optimized structure exhibited no imaginary frequencies anymore. From the optimized product structure, the attached hydroxide was removed and the structure was optimized in the same manner as the product structure. Afterwards, the educt structure, the product structure and a lone hydroxide anion were optimized at the DFT level using version 5.0.4 of ORCA[78] with the B3LYP functional[63, 64], the 6-311G(d,p) basis set[79, 80] and an implicit water solvation environment based on the conductor-like polarizable continuum model.[81] Finally, the Gibbs energy of all three optimized structures was obtained through a frequency analysis at the DFT level with the same parameters as the geometry optimization.

4.1.2 Validation of the Computational Pipeline

To ensure that the changes made to streamline the automatic calculation of ΔG_{HA}, the descriptor was calculated using the changed computational pipeline for the same eleven structures **1** to **11** studied in the previous chapter. The ΔG_{HA} values were then analogously compared to the half life times in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether reported by Fan et al.[32] to ensure the computed values are still correlated to the experimental stability.

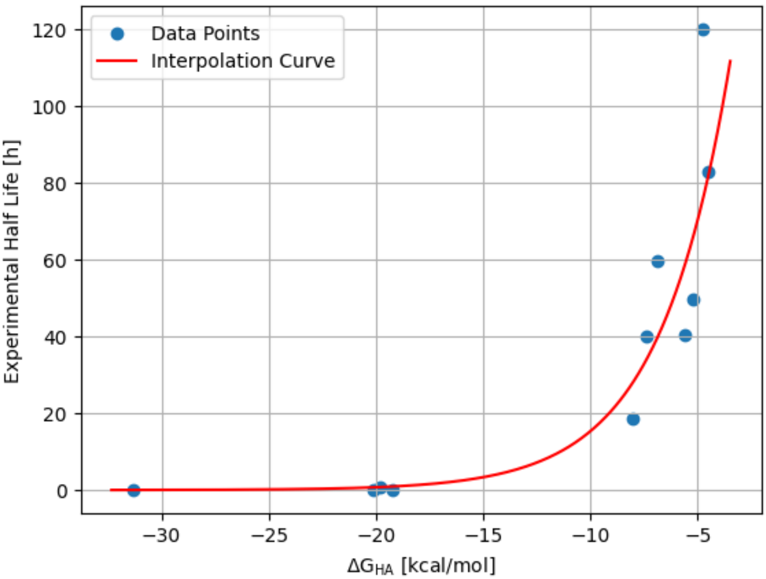


Figure 4.2: Scatter plot of the ΔG_{HA} values computed with the streamlined methodology described in this chapter for compounds **1** to **11** depicted in Figure 3.1 versus their half life in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether at room temperature as reported by Fan et al.[32] An exponential function of the form $a \cdot e^{bx}$ was fitted to the data and is shown in red.

As can be seen in Figure 4.2, the adapted computational methodology results in ΔG_{HA} values that correlate well with experimental stability data. A quantitative relationship between computed and experimental values can be established through an exponential fit to the data of the form $a \cdot e^{bx}$ as shown in Figure 4.2. The exponential fit can be used to interpolate the experimental half life in a 0.5 M ultra-dry solution of KOH in DMSO/crown ether at room temperature through the following relation.

$$\text{Computationally Predicted Half Life [h]} = 322.79\text{h} \cdot e^{0.30465 \cdot \frac{\Delta G_{HA}}{\text{kcal/mol}}} \tag{4.2}$$

In the current form, Equation 4.2 computationally predicts the half life in the specific conditions used by Fan et al.[32] To find compounds with exceptional alkaline stability, it is sufficient to know the relative stability of one compound to another. Therefore, the unit of hours is dropped in the subsequent text and the values computed by 4.2 are simply referred to as computed relative stability values, making them more universally applicable.

4.2 Molecular Dataset Creation

With the computational methodology established, a molecular dataset is needed to apply it to. This dataset was created automatically from an unsubstituted imidazolium ring, to which a set of substituents is systematically added. A set of 18 substituents was chosen that should not impose excessive difficulty regarding their synthesis and covers a sizeable portion of the chemical space. The substituents are a mixture of aromatic and aliphatic hydrocarbons as to not introduce heteroatoms that could open up novel degradation pathways, e.g., would be the case with a vulnerable ether group. The set of substituents also covers different sizes and complexities, it is depicted in Figure 4.3.

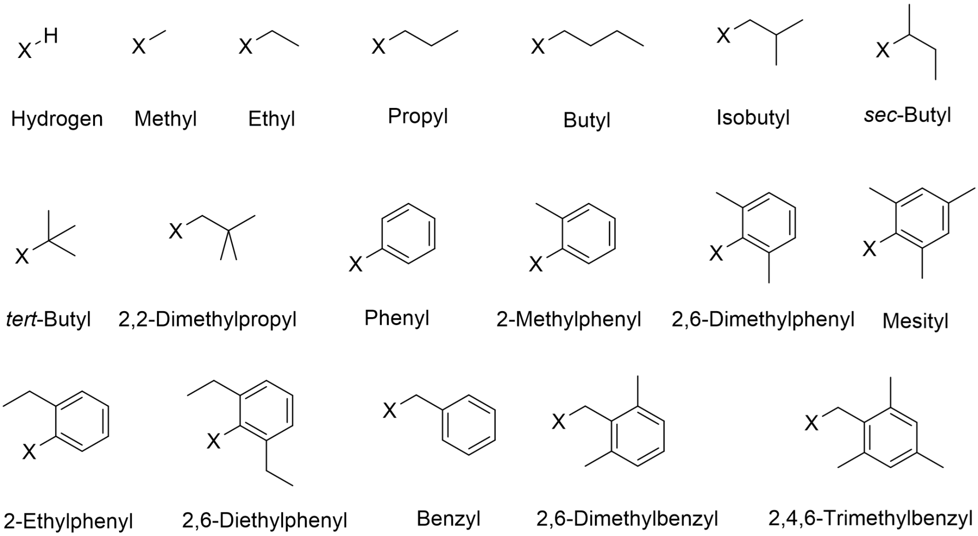


Figure 4.3: Chemical structures of the set of substituents used in the automatic generation of a molecular dataset of imidazolium-based compounds. The X in each structure indicates the position at which the substituent connects to the imidazolium.

Generating all fivefold combinations of the 18 substituents would result in $18^5 \approx 1.9$ million penta-substituted imidazolium compounds, which would require a prohibitively large amount of computational resources even with the streamlined descriptor calculation methodology. While some compounds would be duplicates due to symmetry, removing these would still only eliminate an insignificant number of compounds. In order to limit the number of structures, two limitations are imposed. The two nitrogen substituents are always identical and the C-4 and C-5 substituents are always identical.

By limiting the substituents combinations, only all threefold combinations of the 18 substituents must be generated. Thereby, the total number of structures is reduced to a manageable $18^3 = 5832$ variations, while still covering a large subset of the chemical

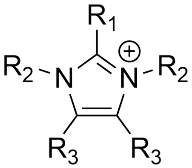


Figure 4.4: Chemical structure of the base imidazolium group used as the starting point of the molecular dataset creation. The substituents at the two nitrogen atoms are always identical, as are the two substituents at the C-4 and C-5 site.

space. Additionally, those compounds remaining should, on average, be easier to synthesize than those eliminated, as the difficulty of having different N-1/N3 substituents and C-4/C-5 substituents is eliminated. The three distinct substituents of each compound are referred to as R_1 for the C-2 substituent, R_2 for the N-1/N-3 substituent and R_3 for the C-4/C-5 substituent, as also depicted in Figure 4.4.

4.2.1 Distribution of the Computed Stability

Applying the described computational methodology to the set of 5832 imidazolium-based compounds results in a broad range of computed stability values. For initial visualization of the data, the computed relative stability values are plotted versus the molecular weight of the corresponding compound in Figure 4.5.

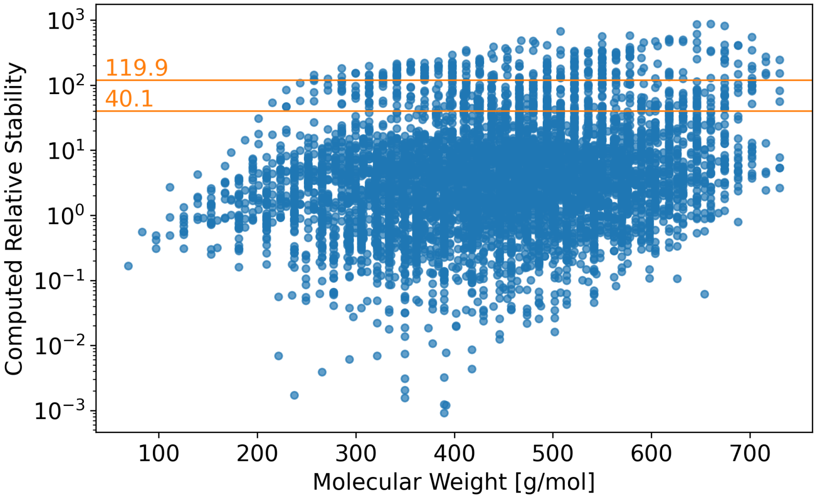
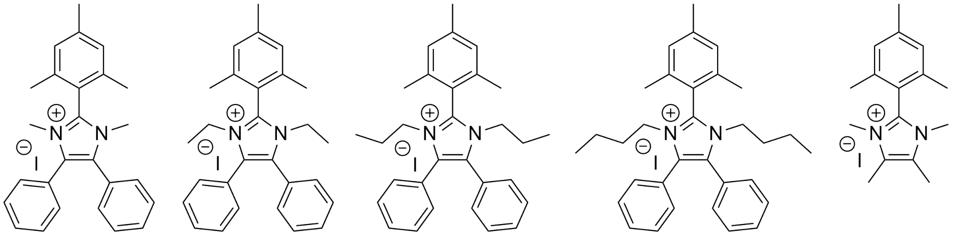


Figure 4.5: Scatter plot of the computed relative stabilities versus the molecular weight. Two orange horizontal lines are shown at y values of 40.1 and 119.9, representing the highest stability reported by Fan et al.[32] for a penta-substituted simple imidazolium and over all imidazolium-based structures, respectively.

A majority of the compounds possesses a computed stability that is significantly below the state-of-the-art. Some compounds however show a promisingly high computed stability. About 11 % of the investigated compounds posses a computed relative stability above 40.1. They are therefore computationally predicted to be more stable than the most stable simple penta-substituted imidazolium (molecule **6**, depicted in Figure 3.1) among the eleven compounds studied in Chapter 3. Furthermore, about 5 % of the compounds even possess a computed relative stability above 119.9, indicating a higher stability than the most stable of the compounds in Figure 3.1, the bis-arylimidazolium-based molecule **10**.

4.3 Experimental Validation

While the molecular dataset shows promise as it contains imidazolium-based compounds for which an exceptional alkaline stability is computed, it must still be verified that the computational predictions hold up for compounds not used to derive the descriptor. Concisely, it must be verified that the computed stability can predict the stability of novel compounds. To verify the accuracy of the computed stabilities, five compounds were selected from the dataset to be synthesized and have their alkaline stability experimentally measured.



(5) Rel. Stab.: 37.1
(12) Rel. Stab.: 60.4
(13) Rel. Stab.: 57.3
(6) Rel. Stab.: 43.2
(14) Rel. Stab.: 108.2

Figure 4.6: Chemical structure of the molecules selected from the dataset for experimental validation. The computed relative stability is shown below each molecule number.

The five compounds selected for experimental validation are depicted in Figure 4.6. The compounds **5** and **6** were chosen as they were tested by Fan et al. already and give a point of reference to examine the reproducibility of different degradation tests. Specifically, molecule **6** is included, as it is the penta-substituted imidazolium with the highest stability reported by Fan et al. Compounds **12** and **13** are included as they are a slight variation of the previously studied compounds **5** and **6**, while having a higher computed stability and lower molecular weight than molecule **6**. Including these four molecules, which differ only slightly in terms of the length of the N-1/N-3 substituent, also allows evaluating how accurately the descriptor can compute the relative stability of similar compounds. Lastly, molecule **14** is included as it possess a significantly higher computed stability than the other molecules while also having a significantly lower molecular weight, making it a promising candidate for implementation into an anion exchange membrane.

The five compounds were all synthesized in the Holdcroft lab of the Simon Fraser University and tested for their experimental alkaline stability. Two different alkaline stability tests were performed to increase the confidence in the results. One degradation test was performed in an alkaline solution with limited hydration and another was performed in a dynamic vapor sorption chamber. Testing the molecules in two vastly different conditions also gives insights towards the impact of these conditions on the relative stability between the molecules. It can be tested if one compound possesses superior stability in most conditions, or if each set of conditions corresponds to its own ideal chemical structure for maximum stability. This is an important aspect, as it determines if one AEM material is superior in terms of stability or if each application field has a different optimum that needs to be found, including its own descriptor.

4.3.1 Synthesis of the Compounds

Molecules **5**, **6**, **12** and **13** are identical in terms of to their R_1 and R_3 substituents and only the R_2 substituent varies. Therefore, the four imidazoliums can all be synthesized from the same imidazole precursor, i.e., molecule **15** depicted in Figure 4.7. The precursor was already available in large quantities from previous experiments in the Holdcroft lab and did not have to be synthesized again; a procedure for the synthesis of the precursor is available in literature.[92] Starting from molecule **15**, the R_2 substituent had to be added to the two nitrogen atoms of the imidazole. This was achieved in a two-step synthesis through the iodide of the respective R_2 group, i.e., methyl iodide for molecule **5**, ethyl iodide for molecule **12** and so on. The imidazole precursor was dissolved is DMSO and deprotonated through addition of base. The monosubstituted imidazolium was synthesized by adding a slight excess of alkyl iodide to the dissolved imidazole precursor. The monoalkylated imidazole was then extracted and washed before dissolving it again and adding an excess of alkyl iodide to substitute the second nitrogen atom to receive the desired imidazolium. The imidazolium was then extracted, washed, and its structure and purity were confirmed with ^{1}H NMR. An overview of the synthesis for the four compounds is given in Figure 4.7.

Molecule **14** differs in its R_3 substituent from the other four compounds and, therefore, cannot be synthesized from the available precursor imidazole **15**. The precursor imidazole **16** had to be synthesized from the corresponding mesitaldehyde and butane-2,3-dione

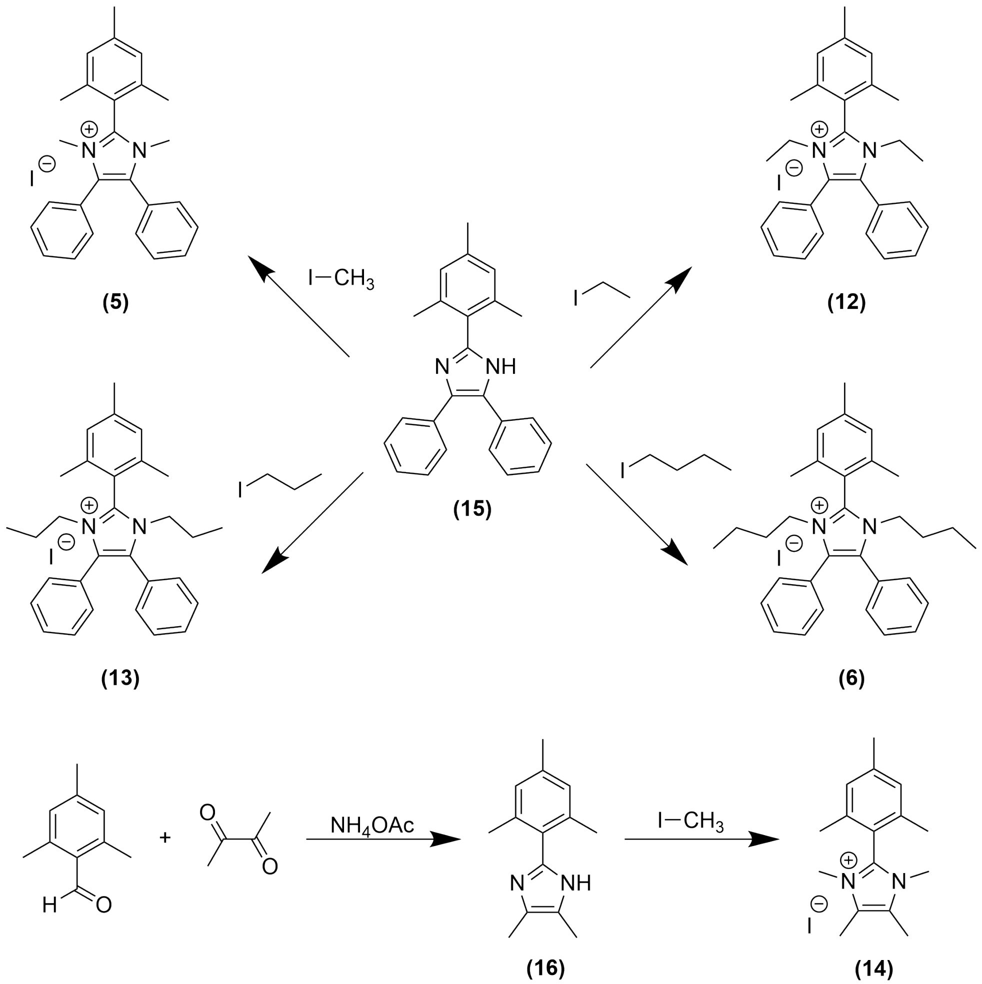


Figure 4.7: Overview schematic showing the synthesis pathways for target molecules **5**, **6**, **12**, **13** and **14**. All molecules are synthesized through alkylation with the corresponding alkyl iodid from an imidazole precursor. The molecules **5**, **6**, **12** and **13** were synthesized from the imidazole **15**, which was already available as is. The precursor **16** for the synthesis of imidazolium **14** had to be synthesized first from mesitaldehyde and butane-2,3-dione.

along with NH_4OAc in acetic acid. Analogously to the molecule **5**, the imidazolium **14** could then be synthesized by methylating the imidazole precursor through methyl iodide. A schematic of the synthesis pathway of imidazolium **14** is included in Figure 4.7.

4.3.2 Alkaline Solution Degradation Test

All five compounds were tested for their stability in an alkaline solution degradation test, using the protocol reported by Hugar et al. as reference.[93] Each imidazolium in its Iodine form (0.03 mol/L) was added to a 2 molar KOH solution in deuterated methanol. The sodium salt of 3-(Trimethylsilyl)-1-propanesulfonic acid (NaDSS) was added as an internal standard that is stable in alkaline conditions. All components were dissolved in deuterated methanol and flame sealed into a NMR tube and are depicted in Figure 4.8. The samples were stored in an 80 °C oil bath and taken out periodically to measure ^{1}H NMR spectra. A detailed description of the procedure is available in the appendix.

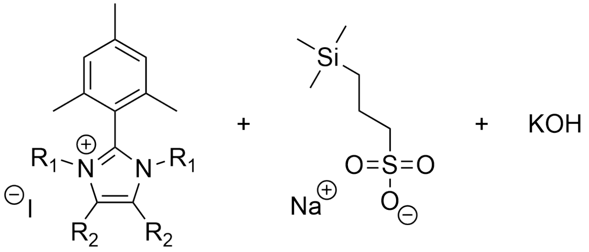


Figure 4.8: Structure of the compounds dissolved in deuterated methanol for the alkaline solution-based degradation test. The sodium salt of 3-(Trimethylsilyl)-1-propanesulfonic acid (NaDSS) is stable in alkaline solution and is used as an internal standard to monitor the imidaozlium degradation in ^{1}H NMR spectroscopy.

As the NaDSS is stable in alkaline solution, its amount in the solution is constant and it can be used as a reference point. By comparing the area of the imidazolium signals in a ^{1}H NMR to that of the NaDSS, the fraction of remaining imidazolium can be calculated. The proportion of the NMR signal areas at day 0 (the day that the solution was set up) were set to represent 100 % of the imidazolium being intact.

The remaining fraction of all five compounds is plotted versus the time that has passed since setting up the solution in Figure 4.9. From Figure 4.9 it is already possible to qualitatively assess the relative stability of the compounds. In line with the computational stability predictions, molecule **5** is significantly less stable than the other compounds, molecule **14** is significantly more stable than the other compounds and molecules **6**, **12** and **13** lie between **5** and **14** in terms of alkaline stability.

To quantify the alkaline stability of the studied molecules further, an exponential curve of the form $e^{a \cdot t}$ was fitted to the data shown in Figure 4.9. By setting the exponential function equal to 0.5 and solving for the time t, the amount of time can be calculated until only half of the imidazolium is remaining, i.e. the half life $t_{\frac{1}{2}}$.

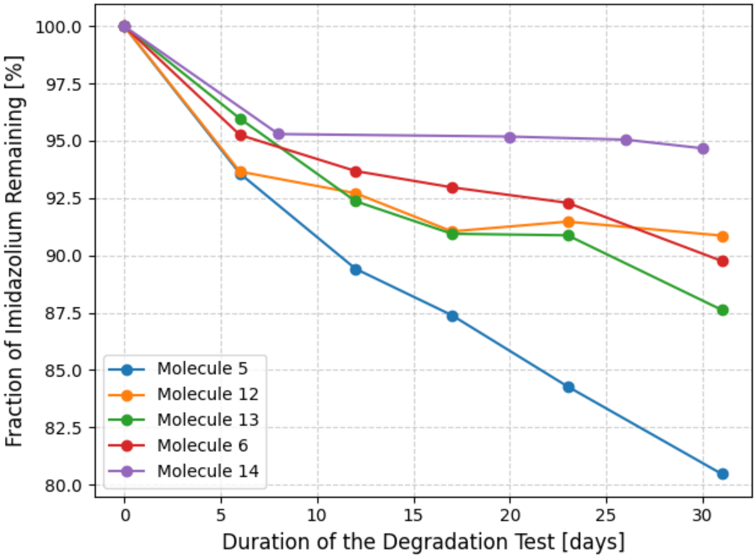


Figure 4.9: Evolution of the amount of each imidazolium remaining throughout the duration of the alkaline solution degradation test.

$$0.5 = \exp(a \cdot t_{\frac{1}{2}}) \tag{4.3}$$

$$t_{\frac{1}{2}} = \frac{\ln(0.5)}{a} \tag{4.4}$$

The hereby derived half life times for the five molecules are reported in Table 4.1.

Table 4.1: Half life times calculated from the experimental measurements for the five molecules chosen for experimental validation of the computationally predicted alkaline stability in a 2 M KOH solution of deuterated methanol at 80 °C.

Molecule	**5**	**12**	**13**	**6**	**14**
Half Life [days]	91.4	169.0	147.8	179.3	324.6

While precautions were taken to keep the solution as dry as possible, some water is inherently present in deuterated methanol. Therefore, the hydroxide is hydrated partially and less reactive than in the ultra-dry KOH solution of DMSO/crown ether reported by Fan et al.[32] On an absolute scale, the resulting half life times are much larger than those measured by Fan et al. The relative half life times are more consistent between the two different solution based testing regimes of the work by Fan et al. and this work. In this work, molecule **6** has a half life that is about twice that of molecule **5**, in the testing regime of Fan et al., molecule **6** has an about 50 % longer half life. While only a sample of two, the qualitative relative stability of the compounds appears to stay mostly intact upon differences in the testing conditions. To investigate further, an additional degradation test of all five compounds was performed under significantly different conditions through dynamic vapor sorption.

4.3.3 Dynamic Vapor Sorption Degradation Test

The alkaline stability of the compounds was also tested through a dynamic vapor sorption (DVS) degradation test, derived from a procedure originally reported by Kreuer and Jannasch [94] and refined for the study of small molecules by Radford et al.[37] This degradation test method allows for the creation of exceptionally harsh conditions by systematically removing the stabilizing hydration layer around the hydroxide.

For the DVS degradation test, the synthesized iodide form of the imidazolium was exchanged for the hydroxide form through an ion exchanger. The hydroxide salt was then progressively dried through 60 °C hot air with a defined relative humidity that is decreased step wise. As the salt dries, the hydroxide has a lower hydration number and therefore becomes more nucleophilic. At a certain drying stage, the hydroxide begins to degrade the imidazolium. The lower the relative humidity is before degradation occurs, the higher the alkaline stability of the imidazolium. A detailed description of the testing procedure is available in the appendix along with the protocol by which the humidity is decreased.

The degradation of the imidazolium can be measured indirectly through the mass of the sample. The degradation products of imidazolium are neutral species and, therefore, possess a lower hydrophilicity than the intact imidazolium. Through the lower hydrophilicity, water is less tightly bound to the sample and evaporates off; this can be observed through a weight loss of the sample. Additional weight loss during degradation may also occur when parts of the degradation products are volatile and leave the sample. As the mass also changes naturally along with a change in the relative humidity, only the last 30 minutes of each step of the DVS procedure are used to allow for the sample mass to stabilize first. The hourly mass losses, relative to the original stabilized sample weight, of the last 30 minutes of the steps of the DVS procedure are plotted against the relative humidity at that step in Figure 4.10. For cases in which multiple steps have the same targeted relative humidity values, only the first occurrence is used.

The most noteworthy points in Figure 4.10 for each compound are the global minima of the mass change, as maximum degradation occurs at these points. The data points directly leading up to the point of fastest degradation are also of other interest. Other data points may be misleading as they are not directly representative of rate at which

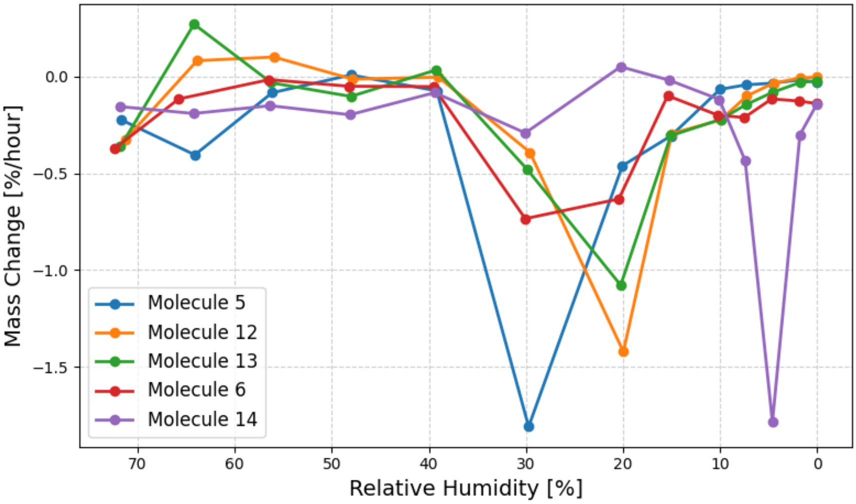


Figure 4.10: Line plot of the mass change, relative to the original stabilized sample mass, the five tested compounds encountered upon contact with 60 °C hot air at different relative humidity levels. The lower the relative humidity is at which the compound starts to undergo significant mass loss, the higher the alkaline stability.

degradation is occurring. First of all, all points to the right of the global minimum can be misleading as they suggest a lower degradation rate even though the relative humidity is even further decreased. This is however caused by the compounds being depleted from the preceding degradation. Secondly, the data points with high relative humidity values may show a mass loss that is not due to degradation but due to the sample not being in equilibrium with the relative humidity yet, as there is a large excess of solvent at the beginning of the degradation test. To only include the data points most representative of the hydroxide-induced degradation, in Figure 4.11 all data points are removed that are to the right of the global minimum and that are to the left of the local maximum that is directly to the left of the global minimum.

From Figure 4.11 the relative stability can be assessed qualitatively. Molecules **5** and **6** have their most pronounced mass loss at about 30 % relative humidity, with the mass loss of molecule **5** being over double that of molecule **6**. Molecule **5** is, therefore, the least stable of the tested compounds, followed by molecules **6**. Molecules **12** and **13** show a higher stability as their maximum hourly mass loss does not occur until 20 % relative humidity is reached. By a large margin, molecule **14** shows the highest stability, with a maximum mass loss not occurring until the relative humidity has reached about 4.5 %.

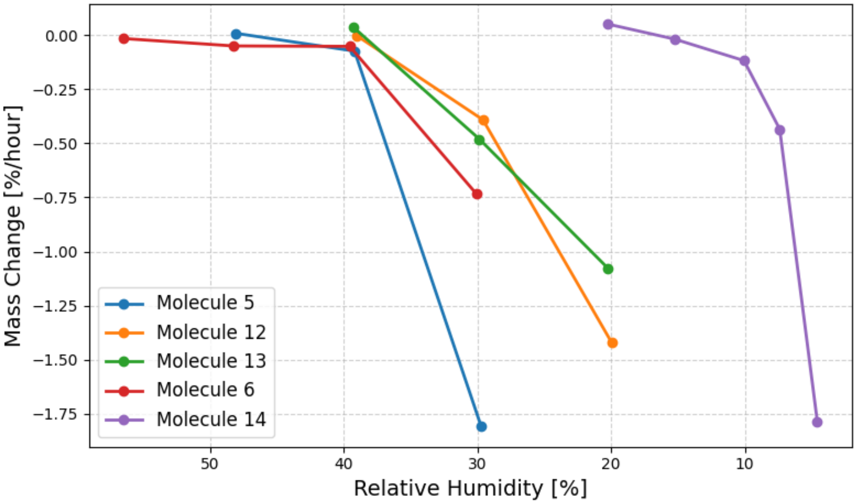


Figure 4.11: Line plot of the mass change, relative to the original stabilized sample mass, the five tested compounds encountered upon contact with 60 °C hot air at different relative humidity levels. The data points after the global minimum in mass change have been removed as a significant portion of the imidazolium is already degraded at that point. Additionally, the data points at high humidity values have been removed as the sample may not have fully stabilized to the relative humidity yet.

To further quantify the stability values, the total mass loss is observed over a longer time frame. The time frame picked is that from the end of step 4 to the end of step 9 of the DVS procedure provided in the appendix. This part of the testing procedure starts off at 70 % relative humidity, followed by a step-wise decrease to 30 % relative humidity over 480 minutes and a final 240 minutes at 70 % relative humidity to rehydrate any potentially remaining imidazolium. The total mass loss occurring in this time frame is provided in Table 4.2.

Table 4.2: The total mass loss that each studied molecule encountered from the end of step 4 to the end of step 9 of the DVS procedure detailed in the appendix.

Molecule	**5**	**12**	**13**	**6**	**14**
Total Mass Loss [%]	10.58	1.75	1.97	2.70	0.02

4.3.4 Comparison of Computations and Experiment

To validate the computational predictions made with the stability descriptor ΔG_{HA} applied to the molecular dataset, the computed relative stability values are plotted versus the measured stability values of the five experimentally molecules **5**, **6**, **12**, **13** and **14**.

In Figure 4.12, the computed relative stability is plotted against the measured half life times in the alkaline solution degradation test and versus the mass loss measured in the DVS degradation test. The values are also reported in Table 4.3.

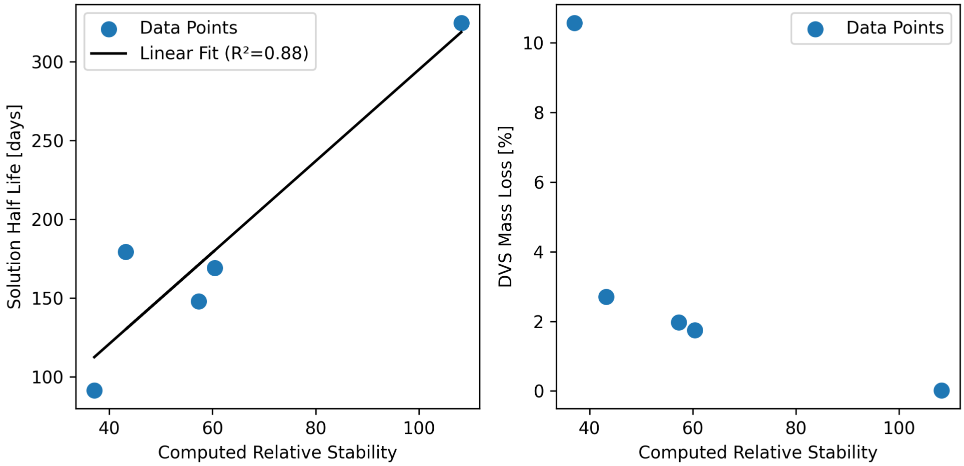


Figure 4.12: Scatter plots comparing the computational predictions of the relative alkaline stability of molecule **5**, **6**, **12**, **13** and **14** versus both the half life times measured in an alkaline solution degradation test (left) and versus the mass loss occurred in the dynamic vapor sorption degradation test (right).

Table 4.3: Overview of the computational predictions of the relative alkaline stability of molecule **5**, **6**, **12**, **13** and **14** along with the half life times measured in an alkaline solution degradation test and the mass loss occurred in the dynamic vapor sorption degradation test.

Molecule	**5**	**12**	**13**	**6**	**14**
Computed Relative Stability	37.1	60.4	57.3	43.2	108.2
Alkaline Solution Half Life [days]	91.4	169.0	147.8	179.3	324.6
DVS Total Mass Loss [%]	10.58	1.75	1.97	2.70	0.02

For the alkaline solution degradation test, the computational predictions of the relative stability were mostly correct. Molecule **6** is a minor outlier as it was computed to be slightly less stable than molecules **12** and **13**, while possessing a slightly longer half life in experimental testing. Regardless of the outlier, the computed stability values and alkaline solution half life times show a solid correlation, with an R^2 value of 0.88.

Within the dynamic vapor sorption degradation test, molecule **6** inhibits a lower alkaline stability than molecules **12** and **13**, as also concluded by the computational descriptor. The stability calculations were also able to qualitatively predict the relative stability between the five compounds in the DVS degradation test. There is no linear correlation between the computed relative stability values and the mass loss in the DVS test. This is however not to be expected as the computed stability values were fitted to half life measurements in an alkaline solution, which does not scale in the same manner as a mass loss during a specific procedure with alkaline stability.

4.4 Summary and Conclusions

In the work described in this chapter, a molecular dataset of 5832 penta-substituted imidazolium compounds was generated. The computational descriptor of alkaline stability, developed in Chapter 3, was applied to assign each member of the dataset a computed relative stability value. Within the dataset, multiple hundred compounds are computationally predicted to posses an alkaline stability that is higher than that of the most stable penta-substituted imidazolium reported in the publication by Fan et al.[32]

To verify that the computational descriptor can be utilized to computationally assess the alkaline stability of unseen compounds, five molecules were selected from the dataset for experimental validation. The compounds were synthesized and their alkaline stability was experimentally measured in two different stability tests to investigate the influence of different chemical conditions. A degradation test in a 2 molar alkaline solution and through dynamic vapor sorption was performed.

The experimental tests confirmed that computationally obtained stability metrics can be used to assess the approximate stability of novel imidazolium-based compounds. For the dynamic vapor sorption degradation test, the stability order of the studied compounds was correctly computed. For the degradation test in alkaline solution, the relative stability was computed with a high accuracy apart from one slight outlier for molecule **6**. However, even when including the outlier, an R^2 value of 0.88 was reached.

The generated dataset is a valuable resource for the development of novel alkaline stable materials for anion exchange membranes. From the dataset, molecules can be selected with exceptional alkaline stability that fulfill other desired properties, such as a specific ion exchange capacity. Among the molecules chosen for experimental validation, the exceptionally stable imidazolium **14** was identified that has has shown a significantly higher stability in experimental tests than the most stable penta-substituted imidazolium **6** among those reported in the work by Fan et al.[32], which was used as reference.

The molecular dataset can also be the starting point to gain additional insights into the structure-stability relationship of imidazolium-based compounds to guide the design of molecules not already contained in it. As the dataset contains thousands of compounds, the impact of specific substituents can be assessed statistically. Additionally, the dataset is sufficiently large to train machine learning models to approximate alkaline stability at a low computational cost. This is described in the following Chapter 5.

Chapter 5

Analysis of the Molecular Dataset

The molecular dataset of 5832 imidazolium-based compounds and their experimentally verified computed relative stability under alkaline conditions described in Chapter 4, allows for a quantitative analysis into the structure-stability relationship. The number of samples allows for a statistical evaluation of the impact each individual substituent choice has at each position on the imidazolium core structure. Additionally, correlations with properties of the imidazolium cation can be analyzed. Lastly the dataset is large enough to train machine learning models on the data to vastly accelerate the stability assessment of novel compounds.

5.1 Independent Influence of the Substituent Choice

To initiate the investigation of the structure-stability relationship within the computed molecular dataset, the independent influence of each substituent is evaluated. This is done by first calculating the average computed relative stability for all compounds sharing the same R_1 substituent and subsequently doing the same for the R_2 and R_3 substituent.

5.1.1 Influence of the R_1 substituent

In Figure 5.1, the average computed relative stability is reported for all molecules that share the same substituent choice for R_1.

The choice of R_1 has the strongest impact on the computed stability compared to R_2 and R_3. Three substituents lead to compounds with significantly higher stability than the others: 2,6-dimethylphenyl, mesityl and 2,6-diethylphenyl. These substituents share the characteristic of being a phenyl group substituted in both ortho positions, providing comprehensive steric protection to the vulnerable C-2 atom. The plane of the ortho-

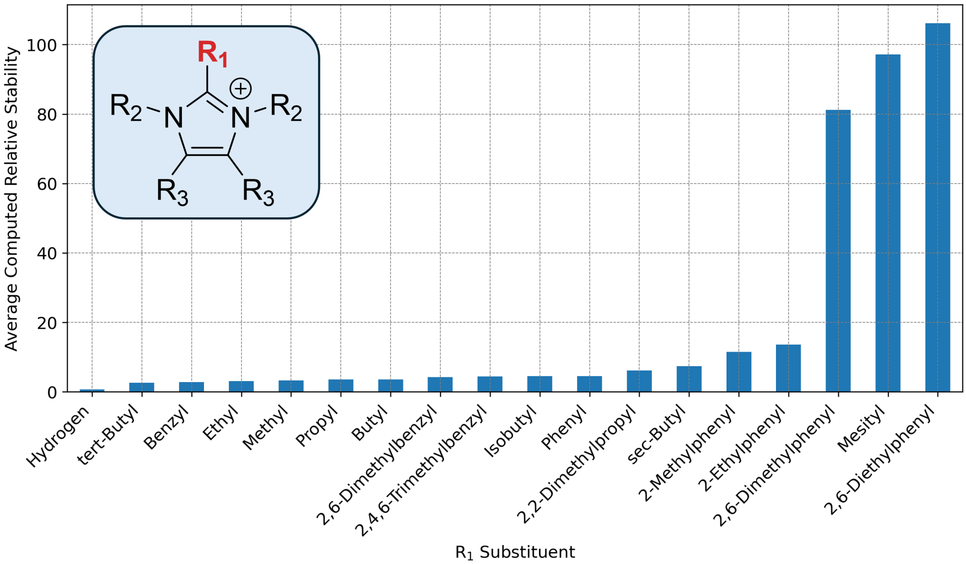


Figure 5.1: Average computed relative stability of the studied compounds grouped by their R_1 substituent.

substituted phenyl ring is rotated by a 90° angle towards the plane of the imidazolium ring and the ortho positions of the phenyl ring then protrude on top and below the C-2 atom. An additional effect of both ortho positions of the phenyl being substituted is due to one of them being pushed towards the imidazolium ring in the product state of the C-2 hydroxide attack, leading to an energetically unfavorable structure. These two effects can be seen well for the examples shown in Figure 5.2, with R_1 = 2,6-dimethylphenyl, R_2 = methyl and R_3 = hydrogen.

The structurally similar 2-methylphenyl and 2-ethylphenyl lead to significantly lower computed stability values. This is likely due to only one of the two ortho positions of the phenyl being substituted, leaving one side of the imidazolium ring to be significantly less protected by steric effects. The structurally similar 2,6-dimethylbenzyl and 2,4,6-trimethylbenzyl are constituted of a phenyl ring that is substituted in both ortho positions. However, the single CH_2 group of the benzyl substituents that connects the phenyl ring to the imidazolium ring introduces a significant degree of flexibility to the structure that leads to the whole substituent turning towards one side of the imidazolium ring and leaving the other side poorly protected. Example structures of both cases are shown in Figure 5.3.

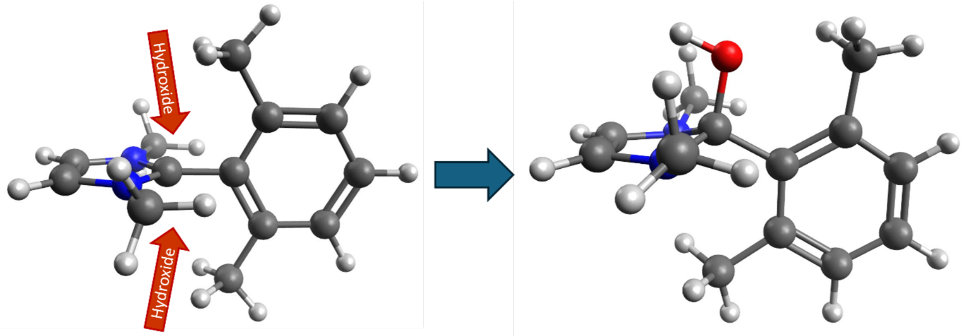


Figure 5.2: The structure of an imidazolium with R_1 = 2,6-dimethylphenyl, R_2 = methyl and R_3 = Hydrogen, before (left) and after a C-2 hydroxide attack (right). The two methyl groups of the 2,6-dimethylphenyl sterically protect both the top and bottom site of the C-2 atom and lead to a strained product of the attack reaction as the bottom methyl group is pushed towards the imidazolium ring.

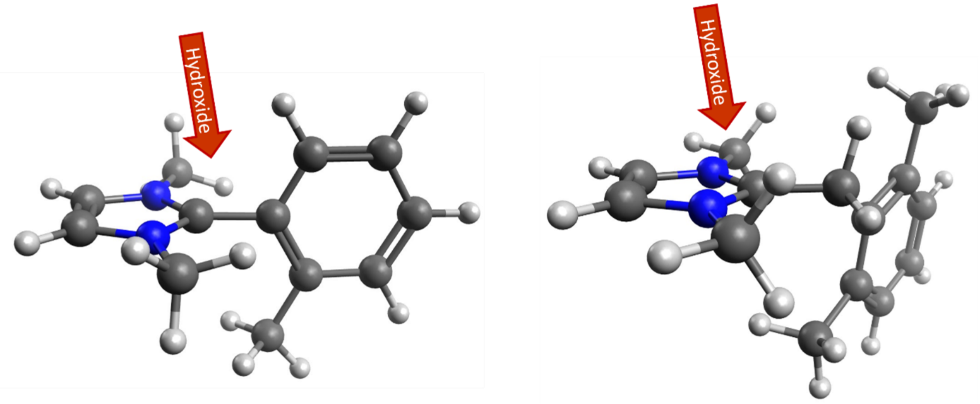


Figure 5.3: Structures of two compounds with reduced steric protection from C-2 hydroxide attack from one side, with R_1 = 2-Methylphenyl (left) and R_1 = 2,6-dimethylbenzyl (right). For both molecules R_2 = methyl and R_3 = Hydrogen.

5.1.2 Influence of the R_2 substituent

The impact of R_2 is visualized in Figure 5.4 in the same manner as before, showing the average computed relative stability of all compounds grouped by their R_2 substituent.

The impact of R_2 on the computed stability is less prominent than that of R_1, as the distribution of stability across the different substituents is more uniform for R_2. The highest computed stability is found for compounds with R_2 = mesityl, due to the high degree of steric protection provided on both sides of the imidazolium ring, as can be seen in Figure 5.5.

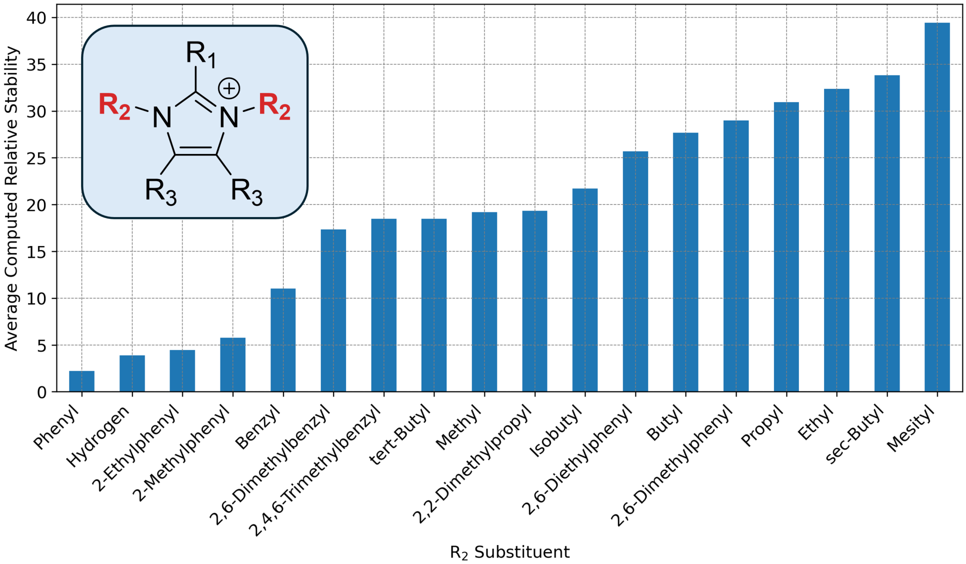


Figure 5.4: Average computed relative stability of the studied compounds grouped by their R_2 substituent.

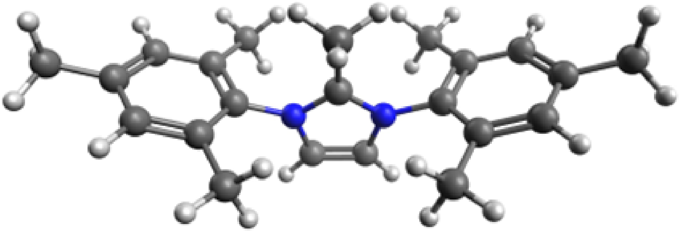


Figure 5.5: Example of a structure with R_2 = mesityl, R_1 = methyl and R_3 = Hydrogen, showing the steric protection of the imidazolium ring from both sides offered by the two mesityl groups.

Analogous to the choice of R_1, the benzyl-based and mono-substituted phenyl substituents lead to significantly decreased computed stability compared to mesityl and 2,6-dimethylphenyl, as they each only protect one side of the imidazolium ring sterically. Alkane groups, while less sterically demanding, still promote higher stability compared to most other substituents. This can likely be attributed to their effect of donating electron density into the imidazolium ring and, thereby, disfavoring the nucleophilic hydroxide attack.

An intriguing effect occurs with the choice of phenyl for R_2, as it leads to the lowest computed stability of all substituents. This effect can be attributed to a stabilizing effect of the hydroxide attack product in comparison to the educt state. A parallel orientation of the phenyl ring with the imidazolium ring is favorable electronically, as it allows for the optimal overlap of p-orbitals to form a conjugated system, lowering the total energy of the molecule. Counteracting the drive towards parallel ring rotation are steric effects, which become more pronounced the more sterically demanding the involved substituents are. Figure 5.6 illustrates this effect, where the introduction of a methyl group as R_1 leads to a significant increase in the torsion angle between the phenyl and imidazolium ring, disturbing the stabilizing effect of a conjugated system.

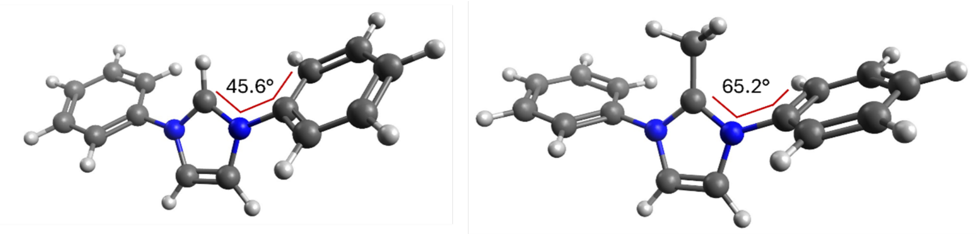


Figure 5.6: Illustration of the torsion angle between the ring of a phenyl substituent as R_2 and the imidazolium ring with a hydrogen (left) and a methyl group (right) as the R_1 substituent. With more steric demand next to the phenyl ring, the torsion angle increases significantly. The red lines indicate the four atoms used to derive the torsion angles. The R_3 substituent is hydrogen in both cases.

Through the described trade-off between electronic and steric effects, the educt side of the imidazolium C-2 hydroxide attack is energetically disfavored, especially for sterically demanding neighboring substituents. The product side of the attack reaction differs in its stereo chemistry in a significant detail. Before the C-2 hydroxide attack, the C-2 atom is sp^2 hybridized, and the imidazolium ring and the bond to the R_1 substituent lie in one plane. As the hydroxide forms a bond with the C-2 atom, its hybridization changes to sp^3 and the R_1 substituent significantly protrudes out of the plane of the imidazolium ring.

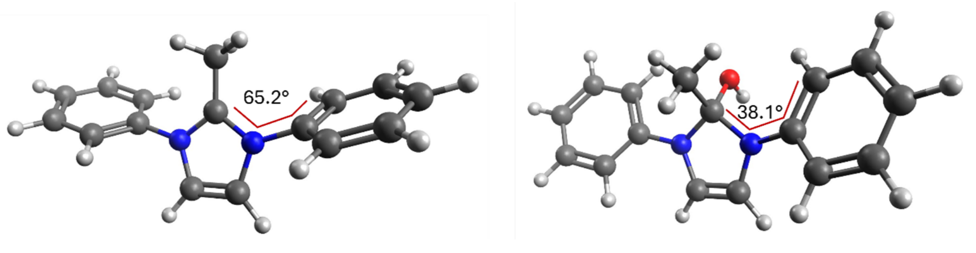


Figure 5.7: Comparison of the torsion angle between the plane of the imidazolium ring and a phenyl ring as the R_2 substituent with a methyl group for R_1 and hydrogen for R_3. After a C-2 hydroxide attack, the R_1 substituent is pushed out of the imidazolium ring plane and allows for the phenyl rings to reduce their torsion angle relative to the imidazolium ring plane. The red lines indicate the four atoms used to derive the torsion angles.

When the R_1 group is pushed out of the plane of the imidazolium ring, the steric strain is reduced, allowing the phenyl rings to rotate into a parallel position relative to the imidazolium ring as illustrated in Figure 5.7. This increases the stabilizing effects from the higher degree of conjugation, thereby decreasing the energy of the attack product structure compared to the educt structure.

5.1.3 Influence of the R_3 substituent

The influence of the R_3 substituent is analyzed in Figure 5.8 in an analogous manner to R_1 and R_2 by showing the average computed relative stability for each studied substituent.

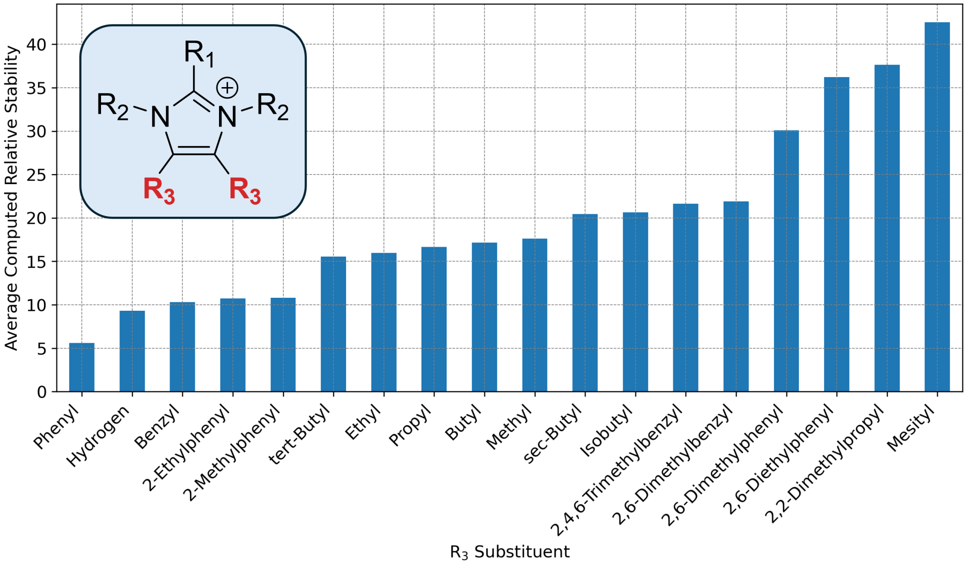


Figure 5.8: Average computed relative stability of the studied compounds grouped by their R_3 substituent.

The computed relative stability is highest for the mesityl group as R_3, closely followed by other sterically demanding groups, such as 2,2-dimethylpropyl or 2,6-diethylphenyl. As the R_3 substituent is located at a significant distance to the vulnerable C-2 atom, the stabilization through these bulky groups is not caused predominantly by direct steric repulsion but by second order steric effect in the product state of the C-2 hydroxide attack. As the C-2 atom changes its hybridization from sp^2 to sp^3 during the hydroxide attack, the R_1 substituent is moved out of the plane of the imidazolium ring and towards the R_3 substituent, as can be seen in the example in Figure 5.9. This can lead to steric strain in the structure of hydroxide attack product, thereby stabilizing the compound in its educt state. The amount of steric strain introduced depends on the steric demands of both the

R_1 and R_3 substituent, explaining why R_3 can have a profound impact on the computed stability despite the large distance to the attack site.

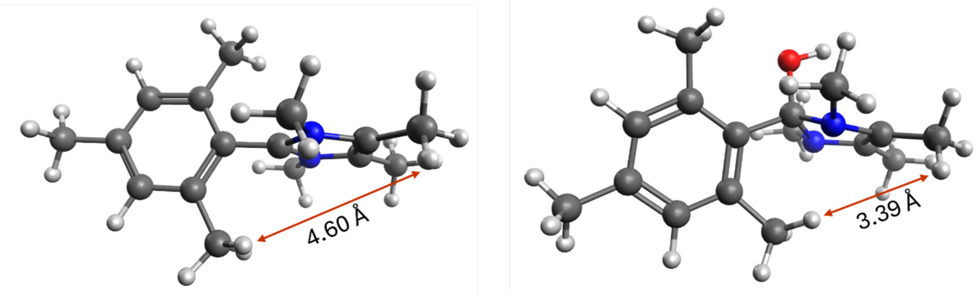


Figure 5.9: Illustration of the secondary steric effects on the C-2 hydroxide attack reaction. The R_1 substituent is pushed towards the R_3 substituents during the reaction and introduces steric strain in the product structure, thereby destabilizing it and suppressing the reaction. The distance between the two closest atoms of the R_1 and the R_3 substituents is provided by a two-sided red arrow.

The choice of a phenyl ring as the R_3 substituent leads to the lowest average computed stability among the studied substituents. This is despite the steric demand of phenyl being much greater than that of, e.g., hydrogen or methyl. The explanation of this finding may be similar to that of the low stability computed for phenyl as the R_2 substituent. In the educt state, the phenyl rings are sterically hindered by the R_2 substituent from turning in parallel to the plane of the imidazolium ring, which stabilizes the molecule by forming a conjugated system.

When the C-2 atom is attacked by hydroxide, the geometry changes and the R_1 substituent is pushed out of the plane of the imidazolium ring, allowing for the phenyl rings to align with the plane of the imidazolium, thereby stabilizing the product structure energetically. This effect can be seen in Figure 5.10, where the dihedral angle of the phenyl ring plane and the imidazolium ring plane is reduced from about 60° to about 40° during the C-2 hydroxide attack.

5.2 Noteworthy Molecular Structures

The compound in the generated molecular dataset with the highest computed relative stability of 879.6 is molecule **17**, which is substituted by the mesityl group at all possible locations ($R_1 = R_2 = R_3 =$ mesityl). The large mesityl groups provide direct steric protection from hydroxide and the reaction product is further destabilized by the bulky R_1 substituent being pushed into the bulky R_3 substituent upon C-2 hydroxide attack. However, the proximity of these bulky groups may pose great challenges in terms of the

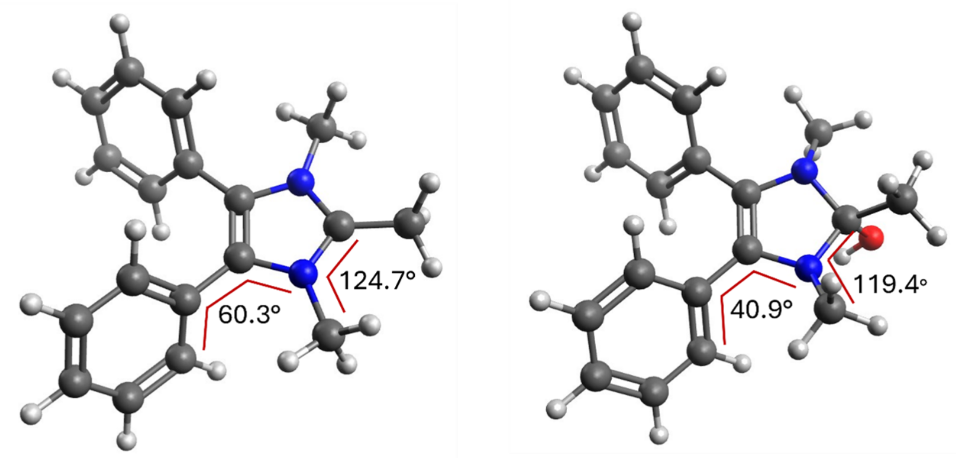
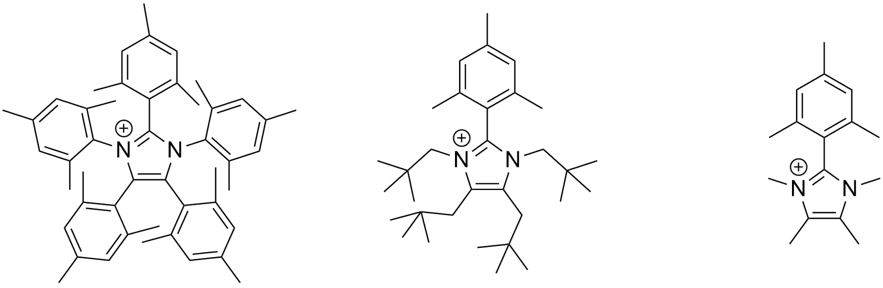


Figure 5.10: Illustration of the destabilizing effects introduced by phenyl as the R_3 substituent. In the educt structure, the phenyl ring is sterically hindered by the R_2 substituent from turning to a position parallel to the plane of the imidazolium ring, which would offer a stabilizing effect through conjugation. In the product state, the R_2 substituent is more flexible through the moved R_1 substituent and allows a less inhibited rotation of the phenyl groups. Relevant angles and dihedral angles are shown, and the atoms involved in the measurement are indicated by red lines.

synthesizability. Additionally, the large substituents lead to a high molecular weight of about 660.0 g/mol that strongly limits the possible ion exchange capacity, therefore potentially inhibiting anionic conductivity. To reduce these two problems, candidates are presented with high stability but with a lower molecular weight.



(17)
Molecular Weight
660.0 g/mol

Computed Relative Stability
879.6

(18)
Molecular Weight
467.8 g/mol

Computed Relative Stability
487.8

(14)
Molecular Weight
243.37 g/mol

Computed Relative Stability
108.2

Figure 5.11: Exemplary chemical structures from the molecular dataset with exceptional computed relative stability at different molecular weights.

The most stable compound with a molecular weight below 500 g/mol is molecule **18**, which has a computed relative stability of 467.8. Molecule **18** possesses mesityl for R_2, while the other two substituents are exchanged for 2,2-dimethylpropyl (R_1 = mesityl, R_2 = R_3 = 2,2-dimethylpropyl). The mesityl group offers direct steric protection against hydroxide attack and destabilizes the hydroxide attack product as it is pushed towards the R_3 groups upon C-2 hydroxide attack. The 2,2-dimethylpropyl group as R_2 and R_3 has a reduced weight compared to mesityl, but similar to the mesityl group protrudes significantly out of the plane of the imidazolium ring. Thereby 2,2-dimethylpropyl still offers direct steric protection in the R_2 position and leads to a destabilization of the product structure in the R_3 position.

For a molecular weight below 250 g/mol, molecule **14** has the highest computed relative stability of 108.2. This is achieved through a mesityl group as R_1 and methyl groups at the other locations (R_1 = mesityl, R_2 = R_3 = methyl). Even with this low limit for molecular weight, the very strong stabilizing effect of mesityl in the R_1 position is indispensable for the construction of a stable structure. The methyl groups add additional protection from hydroxide attack while adding only slightly to the total molecular weight of the compound. As molecule **14** was synthesized for the experimental validation of the molecular dataset, it could be verified that the compound can be synthesized straightforwardly. Its especially high alkaline stability could also be verified in the two degradation tests employed as described in Chapter 4.

5.3 Relationship of Molecular Features and ΔG_{HA}

To further understand the relationship between the molecules and the stability descriptor ΔG_{HA}, molecular features were generated and their relationship to ΔG_{HA} was investigated. A number of properties were extracted from the performed DFT calculations and then used to probe their relationship to ΔG_{HA} using methods with varying sophistication, ranging from a correlation analysis to an artificial neural network. Additionally, a molecular fingerprint was generated for each molecule and then used as a feature to infer ΔG_{HA} using another artificial neural network.

5.3.1 Linear Correlation Analysis

Molecular features were generated for each molecule that have the potential to be related to the hydroxide-induced degradation. These include electronic properties like the charge and orbital distribution, but also structural features such as the molecular weight.

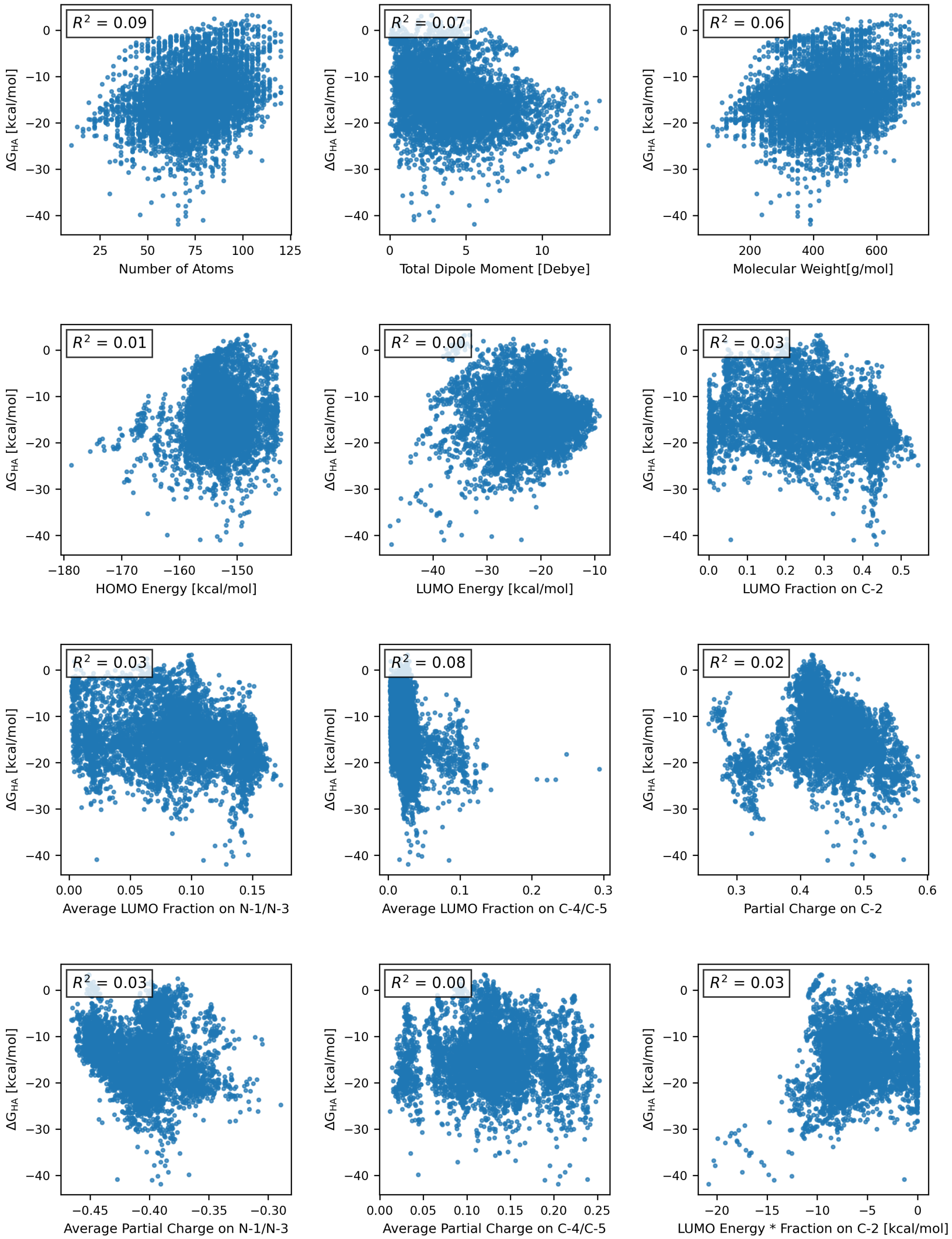


Figure 5.12: Scatter plots of various molecular properties that can be obtained from a DFT single point calculation versus the computational descriptor of alkaline stability ΔG_{HA}. The coefficient of correlations R^2 was calculated for each property and is provided in the top left of each scatter plot.

The following molecular properties were extracted from the final DFT calculation of each imidazolium contained in the molecular dataset:

- The total dipole moment.
- The molecular weight and the number of atoms in the molecule.
- The energy of the HOMO and LUMO orbitals.
- The average fraction of the LUMO localized on the C-2 atom, on the N-1/N-3 atom and on the C-4/C-5 atom according to a Mulliken frontier orbital population analysis.
- The average partial charge of the C-2 atom, of the N-1/N-3 atom and of the C-4/C-5 atom according to a Mulliken population analysis.

The listed properties can be obtained from a single-point DFT calculation of the optimized imidazolium structure and, therefore, pose a significantly lower computational cost than ΔG_{HA}. The number of atoms and the molecular weight can even be directly derived from the chemical structure alone. To investigate if these properties are related to ΔG_{HA} individually, their correlation to ΔG_{HA} is assessed by calculating the coefficient of correlation R^2 for each. The R^2 value and the scatter plots of all correlations analyses are provided in Figure 5.12.

The correlations of ΔG_{HA} with the extracted properties are generally weak, with the coefficient of correlation R^2 being below 0.1 in all cases. The considered properties are, therefore, not suitable for predicting ΔG_{HA} and thus the alkaline stability through simple linear correlations.

One trend that is however noticeable from the scatter plots relates to the number of atoms and the molecular weight. The higher the number of atoms in an imidazolium or the higher the molecular weight, the higher the value of ΔG_{HA} that can be reached. This trend is due to bulky substituents such as the mesityl generally leading to high ΔG_{HA} values and these substituents add substantially to the total molecular weight of the compound. However, the correlation is weakened as some large substituents such as 2-ethylphenyl add significant weight but do not offer much enhanced steric protection due to their asymmetric structure. Additionally, outliers like the phenyl substituents counteract the correlation. The phenyl substituent adds significantly more weight to the molecule than simpler groups like alkanes but leads to a destabilization of the compound through more complex effects than simple steric shielding.

5.3.2 Machine Learning on Molecular Properties

As simple linear correlations between the extracted molecular properties are not sufficient to reliably estimate ΔG_{HA} of a given compound, more sophisticated methods are tested to identify if there may be an underlying relationship between molecular properties and ΔG_{HA}. For these standard machine learning models, version 1.6.1 of scikit-learn was used.[95]

While each individual molecular property is not correlated with ΔG_{HA}, it is tested if a linear combination of extracted properties does show a correlation. For this, the linear regression model of scikit learn is trained on all molecules and all features shown in Figure 5.12 with default settings. The parameters of the linear regression model are tuned in such a way that they minimize the square of the error between inferred ΔG_{HA} values and the ground truth.

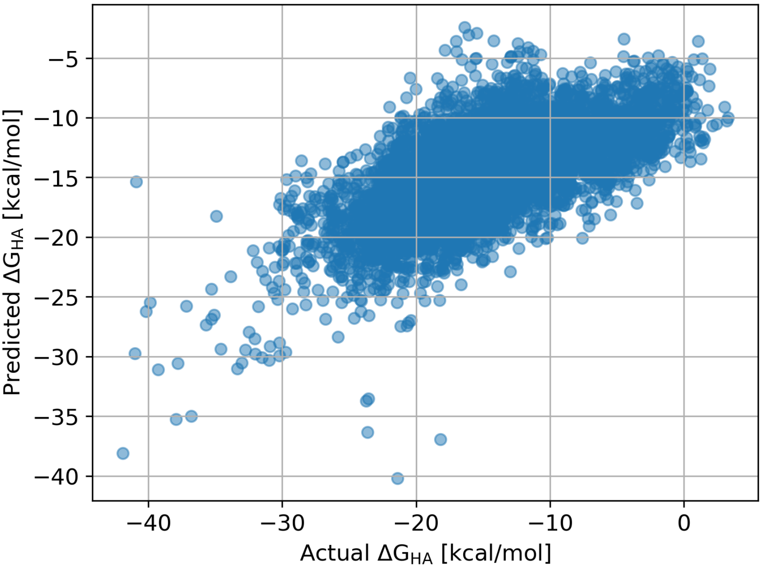


Figure 5.13: Values for ΔG_{HA} computed from DFT versus the inferred ΔG_{HA} values derived from a linear regression model utilizing the extracted molecular properties shown in Figure 5.12. The mean absolute error of the prediction is 3.59 kcal/mol.

Combining all the extracted properties into a single linear regression model improves the predictive ability towards the computed ΔG_{HA} values significantly as can be seen in Figure 5.13. However, the mean absolute error of the predictions is still prohibitively high at 3.59 kcal/mol, making it too unreliable to infer ΔG_{HA}. The error is also a best case scenario, as the model was trained on the entire dataset and also tested on the dataset it was trained on, most likely resulting in significantly higher errors when performing predictions for novel data points.

The molecular properties can be scaled by subtracting the mean of each sample and dividing it by the standard deviation of that feature, resulting in each feature having a mean of 0 and a standard deviation of exactly 1. Scaling the data in such a way has no impact on the performance of the linear regression model, however, it makes the coefficients within the data comparable. The higher the absolute value of the coefficient is, the higher its impact within the model, giving an estimate of how important the feature is for predicting the ΔG_{HA} values. The highest absolute value of 7.99 is hereby found for the coefficient of the average LUMO fraction on N-1/N-3, followed by 6.35 and -5.32 for the coefficients of the number of atoms and of the LUMO fraction on C-2. This indicates that especially the size of the molecules and their LUMO distribution on and around the reactive C-2 site are important parameters when predicting ΔG_{HA} for the analyzed dataset. The coefficient of the LUMO energy is -1.61, indicating that it is significantly less important in the prediction of ΔG_{HA}, at least when utilizing a linear regression model.

The linear regression and the linear correlation analysis may show weak correlations as the underlying relationship is not a linear one. To allow more complex relationships to be grasped by the model, a polynomial regression model with up to third-order interactions was trained. For the third-order polynomial regression model, all possible squares and cubes of the molecule properties were computed and were added as features. On this extended set of features another linear regression model was trained.

The results of the third-order linear regression model are shown in Figure 5.14. The mean absolute error is reduced to 1.96 kcal/mol, compared to the 3.59 kcal/mol for the linear regression model. The error is still relatively large, despite the linear regression model being trained on the full dataset and being evaluated on the data it has been trained on. Therefore, the predictive ability on novel data will be significantly lower.

To allow for still more complex relationships to be identified between the extracted molecular properties and ΔG_{HA}, a simple artificial neural network was constructed. The neural network takes all extracted molecular properties as input and predicts ΔG_{HA} by passing the data through three dense layers with 1024 neurons each. The input data is scaled by subtracting the mean and dividing by the standard distribution of each corresponding property. The model was constructed with the *MLPRegressor* function of scikit-learn version 1.6.1.[95] using the random state 1 and the default parameters otherwise.

When training the artificial neural network on the full dataset, it is able to reproduce the ΔG_{HA} values with a mean absolute error of 0.65 kcal/mol, performing much better than linear and logistic regression. However, especially due to the large number of parameters

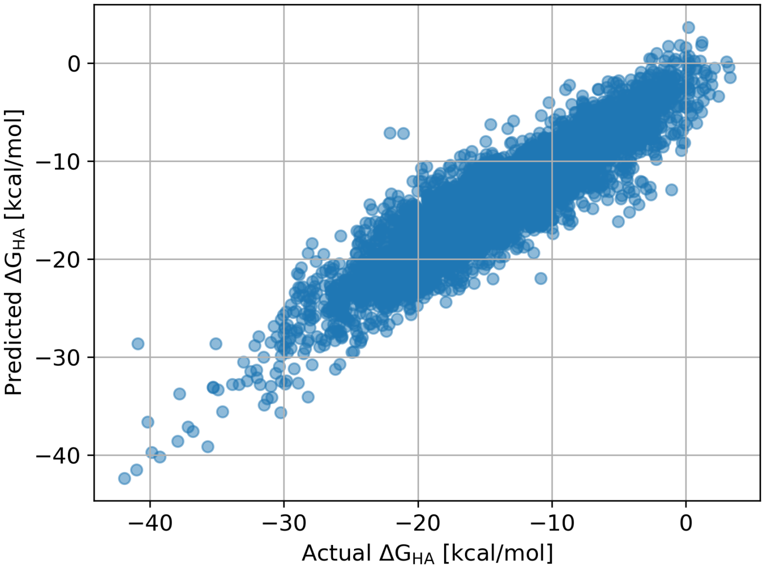


Figure 5.14: Values for ΔG_{HA} computed from DFT versus the predicted ΔG_{HA} values derived from a third-order polynomial regression model utilizing the extracted molecular properties shown in Figure 5.12. The mean absolute error of the prediction is 1.96 kcal/mol.

in neural networks, the performance on the dataset on which the training was performed does not reflect the performance on novel data. To test the performance on unseen data, the model was only trained on a randomly selected sample containing 80 % of the dataset and evaluated on the other 20 % of the dataset that the model has not seen during training. This process was repeated five times in such a way that each data point is part of the validation set once. The neural network was then able to predict ΔG_{HA} with a mean absolute error of 1.50 kcal/mol for molecules not seen during training, depicted in Figure 5.15.

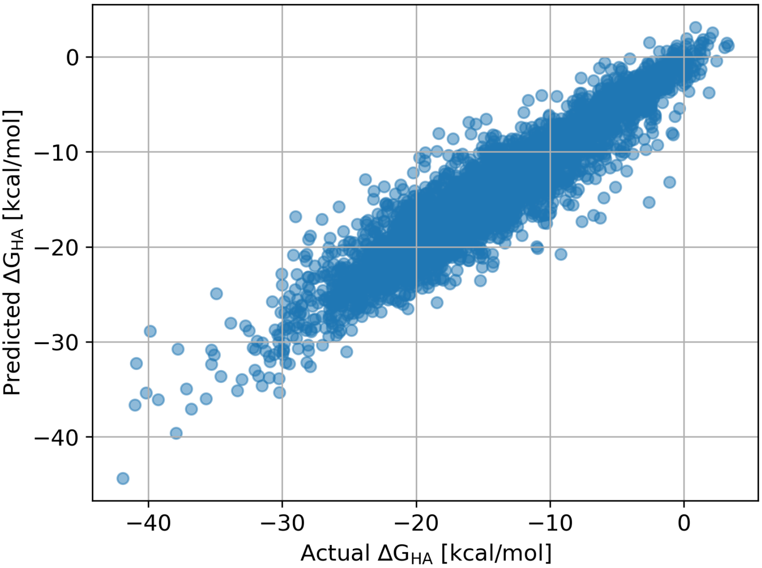


Figure 5.15: Values for ΔG_{HA} computed from DFT versus the predicted ΔG_{HA} values of the validation sets derived from an artificial neural network trained on the extracted molecular properties shown in Figure 5.12. The mean absolute error of the prediction is 1.50 kcal/mol.

5.3.3 Machine Learning on the Molecular Structure

Training machine learning models on properties extracted from the imidazolium-based compounds leads to a loss of information, especially regarding the very significant steric effects. Additionally, most of the properties require performing a DFT calculation which incurs a high computational costs. Both of these disadvantages can be alleviated by training a machine learning model to directly predict ΔG_{HA} from the molecular structure. The steric information would be available to the model and no costly *ab initio* calculations are required, allowing for efficient scanning of molecules.

Typical machine learning models require the input to be a vector of a predetermined size. This is straightforward when using a predetermined number of molecular properties as the input but is more challenging for molecular structures as they do not natively take the form of a vector and vary in size. In the following, the problem is circumvented by converting the chemical structure into a vectorized molecular fingerprint to then input into a regular artificial neural network.

To convert the molecular structures into vector form, the SMILES representation of each compound is converted into the Morgan fingerprint using version 2024.09.4 of rdkit[82] and the default parameters. The fingerprint is then used as the input into a simple artificial network with three hidden layers consisting of 1024 neurons each, illustrated in Figure 5.16. The model is created with the *MLPRegressor* function of scikit-learn version 1.6.1.[95] using the random state 1 and the default parameters otherwise. When training the model on the entire dataset, it is able to predict ΔG_{HA} with a mean absolute error of 0.29 kcal/mol. However, the model should not just be able to reproduce the results it was trained on but also predict ΔG_{HA} for novel compounds. To test the performance on unseen data, the molecular dataset is randomly split into a training set containing 80 % of the data and a validation set containing the remaining 20 %. The model was then trained on the training set and the error was computed for the validation set. The process was repeated five times in such a manner that each molecule is in exactly one of the validation sets. The average performance on data points not in the dataset is an average absolute error of 1.31 kcal/mol, shown in Figure 5.17.

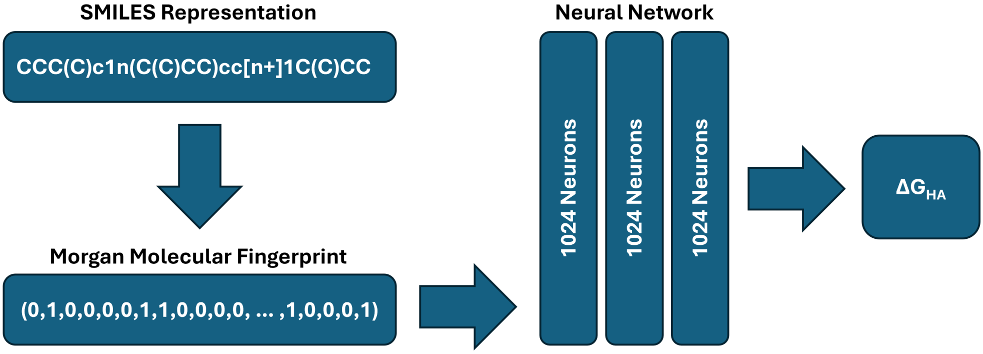


Figure 5.16: Schematic of the machine learning pipeline. The SMILES representation of a given compound is converted into a vector using the Morgan fingerprint. The generated vector is then fed into a regular artificial neural network to predict ΔG_{HA}.

While predicting ΔG_{HA} through the described artificial neural network introduces an error of slightly above 1 kcal/mol, it is highly computationally efficient. Inferring ΔG_{HA} of all 5832 compounds in the molecular dataset can be done in under one second on an average laptop CPU. Therefore, the model can be used to efficiently obtain a first assessment of the alkaline stability of a given compound and only performing the more costly DFT calculations for the most promising candidates.

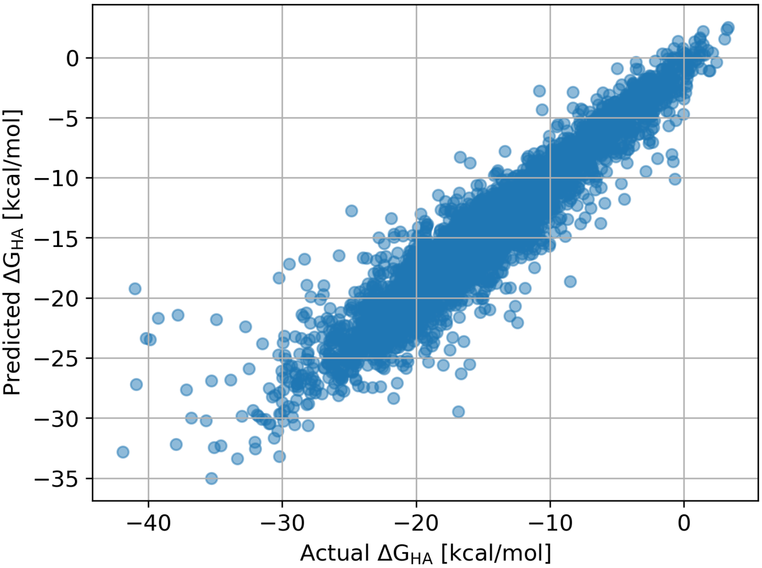


Figure 5.17: Values for ΔG_{HA} computed from DFT versus the predicted ΔG_{HA} values of the validation sets derived from an artificial neural network trained on the Morgan fingerprint representation of the molecules. The mean absolute error of the prediction is 1.31 kcal/mol.

5.4 Summary and Conclusions

The molecular dataset that was generated and experimentally validated, described in Chapter 4, implicitly contains in-depth information about the intricate structure-stability relationship of imidazolium-based compounds. An analysis of the dataset was performed and described in this chapter. The effect that the individual substituents R_1, R_2 and R_3 have on the alkaline stability were visualized in Figures 5.1, 5.4 and 5.8, respectively. Using these findings, the design of alkaline stable imidazolium-based compounds can be guided without performing additional computations or even accessing the molecular dataset. Additionally, especially pronounced substituent effects could be identified and investigated in-depth.

The mesityl group and ortho bi-substituted phenyls result in a significant increase in computed stability in any location, but especially when attached to the C-2 atom. The cause of the stabilizing effect is the direct steric protection of the reactant from hydroxide attacks and the destabilization of the products of a hydroxide attack. A particularly low stability was computed when phenyl, and to a lesser extent its mono-substituted derivatives, were chosen as R_2 or R_3 substituents, which is caused by stabilization of the product structure of the C-2 hydroxide attack due to the drive of the phenyl group to create a conjugated system. This finding is especially intriguing as phenyl is a common substituent group for imidazolium-based compounds.

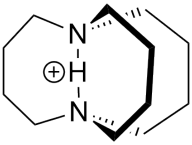
A number of molecular properties were extracted from each compound and their relationship with the computed stability was investigated. Each individual property exhibited a weak correlation with the computed stability or none at all. The lack of strong correlations show the complexity of the structure-stability relationship that cannot be captured by individual properties of the imidazolium cation. However, combining the properties together in linear regression or polynomial regression models is able to roughly reproduce the energetic stability descriptor ΔG_{HA}. Only processing data through a much more sophisticated artificial neural network yields a meaningful prediction of ΔG_{HA}, using the properties extracted from the DFT calculations.

Lastly, an artificial neural network was employed to compute ΔG_{HA} from the imidazolium cation's structure alone. The molecular structure was directly converted from the SMILES format to a molecular fingerprint and used as input for an artificial neural network. With the molecular fingerprint alone, the model could predict ΔG_{HA} of a molecule not contained in the dataset with a mean absolute error of 1.31 kcal/mol. As only a SMILES string is required for the prediction, no costly quantum chemical computations have to be performed, allowing for incredibly fast stability approximations within a fraction of a second on a regular laptop CPU. The trained neural network is therefore a valuable tool to quickly assess a given compound to focus the costly quantum chemical calculations and experimental validations on only the most promising of candidates.

Chapter 6

Computational Exploration of the in-DBD Cation

Recently an alternative to imidazolium-based compounds as the cationic moiety in anion exchange membranes has been reported by Radford et al.[37] in the form of 1,6-diazabicyclo[4.4.4]tetradecan-1,6-ium (in-DBD, molecule **19**), depicted in Figure 6.1. Even in extremely harsh alkaline conditions, in-DBD has been found by Radford et al. to exhibit extraordinary durability, outcompeting state-of-the-art cationic materials like those based on imidazolium, in the tested conditions.



(19)

Figure 6.1: Chemical structure of 1,6-diazabicyclo[4.4.4]tetradecan-1,6-ium, which is referred to as in-DBD or molecule **19** in this work.

The in-DBD molecule has an intricate structure that stabilizes the central hydrogen atom through both endohedral protection and an intra-bridgehead hydrogen bonding. Through the robust fixation of the central hydrogen atom, a Hofmann elimination is assumed to be strongly inhibited.[37]

One challenge arising from the intricate structural stabilization of in-DBD is its implementation into a polymeric membrane. The symmetric in-DBD structure must be augmented in some manner to include an attachment point to the polymer backbone. Any changes to the structure may however interfere with the stabilizing effects present in the undisturbed structure. Designing a strategy for polymer attachment therefore requires an extensive understanding of the degradation mechanics and the impact that structural changes have on them.

The strategies employed for guiding the design of alkaline stable imidazolium-based compounds are well-suited to the challenges of designing in-DBD polymer attachment sites. The structure of in-DBD and the relevant degradation reactions were studied extensively through density functional theory calculations. The insights gained were then applied to assess the alkaline stability of in-DBD derivatives that were equipped with various attachment sites to guide the design of a strategy without negative influence on stability.

6.1 Computational Methodology

The cations were constructed manually in version 1.2.0 of the molecular modeling software Avogadro.[96, 97] The lowest energy conformers were generated with CREST version 2.12 [83, 84] in an implicit water environment (analytical linearized Poisson-Boltzmann model) using the GFN2 method of xTB version 6.6.1 [91] as the quantum chemical backend. The generated conformer structures are the starting point for the density functional theory investigations that are performed with version 5.0.4 of ORCA [78], the B3LYP functional [63, 64] with the D4 dispersion correction [98], the def2-TZVP basis set [99, 100] and an implicit water solvation through the conductor-like polarizable continuum model.[81] For each cation, a geometry optimization was performed, a single hydroxide ion was added close to the cation and another geometry optimization was performed. Starting from the optimized geometry with a hydroxide nearby, the degradation products were constructed in Avogadro and another geometry optimization was done. Lastly, the optimized geometries of degradation product and the educt with a nearby hydroxide are given as the input for an automated nudged elastic band transition state search [85] implemented in ORCA, to give the transition state of each respective degradation reaction. To ensure the correct structures were obtained, a frequency analysis was performed for all ground states and transition states, and it was confirmed that they possess zero and one imaginary frequency mode, respectively.

6.2 Geometry of the in-DBD Structure

To effectively study the degradation characteristics of in-DBD, it is paramount to understand its structure and especially its symmetry. The structure of in-DBD after a DFT level geometry optimization in an implicit water environment is depicted in Figure 6.2.

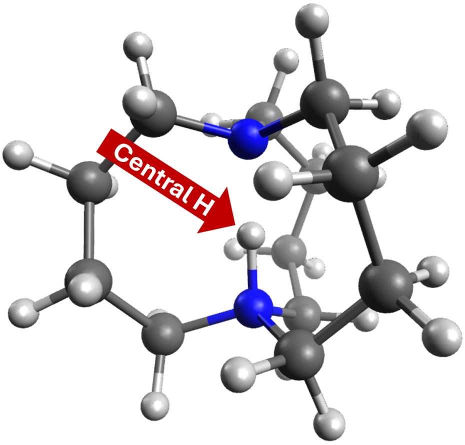


Figure 6.2: Ground state structure of the in-DBD cation **19** in an implicit water solvation environment as computed at the DFT level. The central hydrogen atom is highlighted with a red arrow for clarity and due to its importance to the chemical properties of in-DBD.

The structure possesses a threefold rotational axis of symmetry that passes through both nitrogen atoms. A rotation of 120° along that axis leaves the molecule unchanged and, therefore, the carbon atoms can be classified into four groups that each contain three chemically equivalent members. A higher degree of symmetry would be present if the central hydrogen atom was located exactly in the center of the two nitrogen atoms, as three additional twofold symmetry axes would be added that are orthogonal to the threefold main symmetry axis and pass through the central hydrogen. However, after a DFT level geometry optimization, the hydrogen is located slightly closer to one of the nitrogen atoms than to the other. The distances between the central hydrogen and the two nitrogen atoms are 1.12 and 1.47 Å. To validate the structure and understand the potential energy surface of the central hydrogen, the minimum energy path (MEP) was computed through the nudged elastic band method for the shift of the hydrogen from one nitrogen to the other. For 10 steps of the MEP, the electronic energy and the Gibbs free energy were computed and the results are plotted in Figure 6.3.

The MEP provides mechanistic insight into why the hydrogen is not located centrally between the nitrogen atoms after a DFT geometry optimization. The geometry optimization adjusts the atomic positions to minimize the electronic energy, which is at a maximum when the hydrogen is centered and at a minimum at 1.12 Å from either of the nitrogen atoms. However, when computing the Gibbs free energy, the molecule is at an energetic minimum for the hydrogen being roughly centered between the nitrogen atoms. The free energy curve forms a plateau with slight variations, which likely stem from noise in the computation and is below the typical error of a DFT calculation.

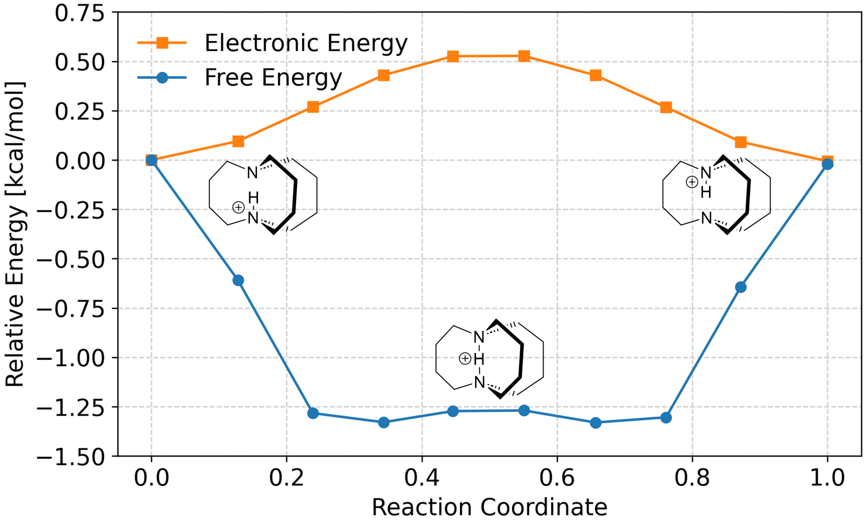


Figure 6.3: Minimum energy path of the hydrogen shift from one nitrogen to the other as computed with the nudged elastic band method at the DFT level. For 10 steps, the electronic energy and the Gibbs free energy were computed and plotted.

As the free energy is generally reflective of the chemical behavior in standard experiments, the hydrogen should be centered between the nitrogen atoms. Assuming the hydrogen at the center of the molecule increases the symmetry by introducing three twofold axes of symmetry that are orthogonal to the main symmetry axis. The molecule then belongs to the D3 point group. This has the consequence that both nitrogen atoms are chemically equivalent and that all 12 carbon atoms can be grouped into two groups of six chemically equivalent atoms.

As depicted in Figure 6.4, the six atoms that are bound directly to nitrogen are all chemically equivalent and will be referred to as α-carbons. The other 6 carbon atoms are also chemically equivalent and will be referred to as β-carbons. In the Lewis structures shown in the subsequent work, the hydrogen atom will also be shown to be located in the center of the two nitrogen atoms.

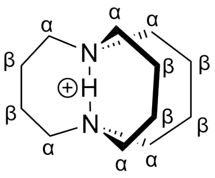


Figure 6.4: Structure of the in-DBD cation **19** with the nomenclature for the carbon atoms used in this work.

The chemical equivalency of the 6 carbon atoms in both the α and β group is also supported by the ^{1}H and ^{13}C NMRs of in-DBD reported by Radford et al., as both NMRs only show two signals that stem from the cation.

6.3 Degradation Pathways of in-DBD

The three degradation reactions of in-DBD suggested in the work by Radford et al. were investigated in detail through DFT calculations. The three reactions are a deprotonation of the central hydrogen atom through a hydroxide attack, the addition of hydroxide to an α-carbon and an elimination reaction through a hydroxide attack.

For all three degradation reactions, the Gibbs energy of activation $\Delta G^{\ddagger}$ and Gibbs free energy change ΔG were computed. The calculated energetics are collected in Figure 6.5.

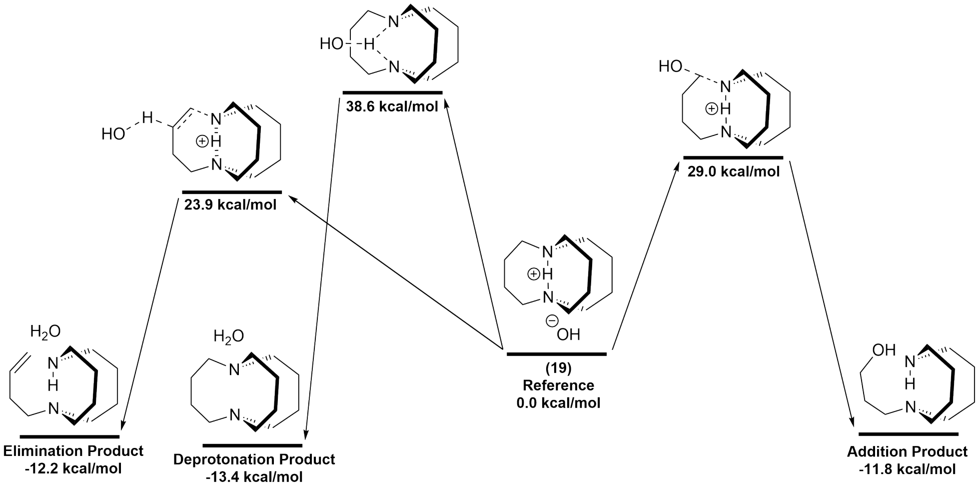


Figure 6.5: Three postulated degradation pathways of in-DBD **19** in an aqueous alkaline environment along with the corresponding Gibbs free energy changes and free energy barriers computed at the DFT level.

The Gibbs free energies of the three degradation products lie within a range of less than 2 kcal/mol of each other. The Gibbs free energy change of the degradation pathways are all negative at -11.8 kcal/mol to -13.4 kcal/mol. The products are, therefore, all thermodynamically stable with no large differences in their stability. The Gibbs energy of activation of the degradation pathways lie between 23.9 kcal/mol and 38.6 kcal/mol. The largest barrier with 38.6 kcal/mol is observed for the deprotonation, a barrier of 29.0 kcal/mol is obtained for the addition reaction and the lowest value of 23.9 kcal/mol belongs to the elimination pathway. Radford et al. only observed the elimination product in their degradation experiments, which can be explained by this degradation pathway possessing the lowest activation barrier as computed in this work. In the following subsections, the three degradation reactions are investigated in more detail.

6.3.1 Deprotonation of in-DBD

As reported by Radford et al., the hydroxide-induced deprotonation of the central hydrogen atom is energetically unfavorable. The Gibbs energy of activation was determined to be 38.6 kcal/mol, making it the highest barrier found among the three studied degradation reactions, which also explains the lack of its experimental observation in the degradation tests.

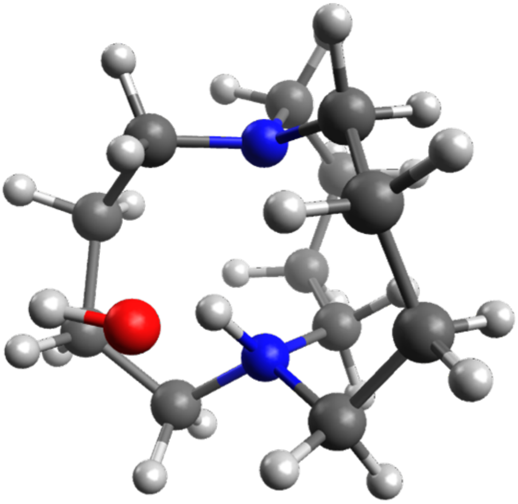


Figure 6.6: Transition state of the hydroxide-induced degradation of in-DBD, as computed at the DFT level.

In Figure 6.6, the transition state of the deprotonation reaction is shown. The hydroxide approaches the cation until it reaches a distance of about 2.5 Å to the central hydrogen atom that is to be removed. In the transition state, the hydrogen has also shifted from its original position towards the hydroxide. This movement of the hydrogen atom can be quantified by the angle between the top nitrogen atom, the bottom nitrogen atom and the deprotonated hydrogen (N-N-H angle). The angle is 0.2° for the undisturbed in-DBD, as the hydrogen is directly on the line between the two Nitrogen atoms. In the transition state, the angle is expanded to 58.6°. Additionally, the mean distance between the hydrogen and the closest nitrogen atom is 1.11 Å in the undisturbed case and is stretched to 1.17 Å in the transition state. The distance to the other nitrogen atom is affected more significantly, increasing from 1.48 Å to 2.35 Å.

6.3.2 Hydroxide Addition to in-DBD

The simulation of the addition of hydroxide at one of the α-carbons concluded a Gibbs free energy of activation of 29.0 kcal/mol. This degradation reaction is therefore more favorable than the deprotonation. The barrier is however still 4.7 kcal/mol higher than the elimination reaction, which is a large enough difference to explain why it was not observed in the experimental degradation tests performed by Radford et al.

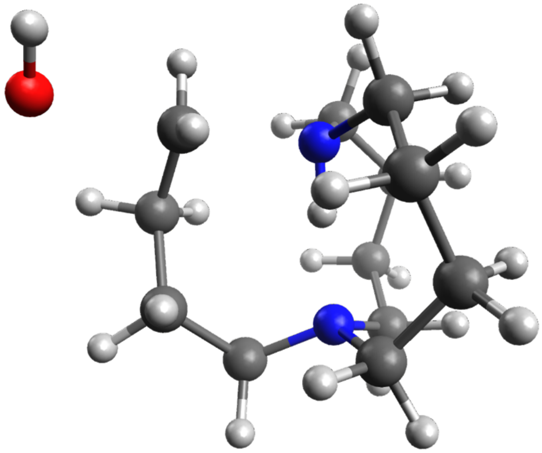


Figure 6.7: Transition state of the hydroxide addition at an α-carbon of in-DBD, computed at the DFT level.

In the transition state, shown in Figure 6.7, the α-carbon is penta-coordinated and the involved oxygen, carbon and nitrogen atoms nearly form a straight line. The transition state of this degradation pathway resembles the typical geometry of a S_N2 reaction. In the undisturbed in-DBD, the bond distance between the α-carbon and the nitrogen is 1.50 Å and increases to 2.02 Å in the transition state. The distance of the hydroxide to the α-carbon in the transition state is 2.07 Å. This reaction mechanism could be investigated experimentally as it leads to a stereo-inversion (Walden inversion), which would be observable if one of the α-carbons hydrogens were replaced by a different substituent.

6.3.3 Elimination Reaction of in-DBD

From the studied hydroxide-induced degradation pathways of in-DBD, the elimination reaction exhibits the lowest Gibbs energy of activation of 23.9 kcal/mol. This is also consistent with Radford et al. reporting the elimination as the only observed degradation product in the performed degradation tests.

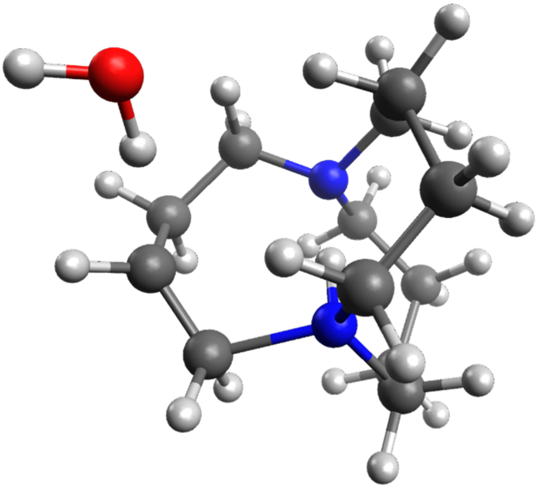


Figure 6.8: Transition state of the hydroxide-induced elimination reaction of in-DBD, computed at the DFT level.

In the transition state of the elimination reaction, depicted in Figure 6.8, the reacting hydrogen atom is almost fully transferred to the hydroxide already, with its distance to oxygen being 1.00 Å. The distance to the β-carbon, that the hydrogen was previously bound to, has reached 2.02 Å from the 1.10 Å in the initial in-DBD structure. The bond length between the α-carbon and the nitrogen is stretched slightly to 1.78 Å from 1.49 Å in its unperturbed state. The bond length of the involved α-carbon and β-carbon is slightly shortened to 1.42 Å, in comparison to the 1.53 Å in the unperturbed case.

6.3.4 Comparison to Other Azacyclic Cations

To set a reference point for the degradation energetics of in-DBD, the energetics for two similar azacyclic compounds were computed that do not possess the unique stabilizing properties of in-DBD. The two compounds are 6-azaspiro[5.5]undecan-6-ium (ASU, molecule **20**) and *N*,*N*-dimethylpiperidin-1-ium (DMP, molecule **21**).

Analogous to in-DBD, ASU is susceptible to a hydroxide addition to an α-carbon and to an elimination reaction through hydroxide attack at the hydrogen atom attached to a β-carbon. As no central hydrogen is present in ASU, a deprotonation reaction analogous to that of in-DBD is not possible.

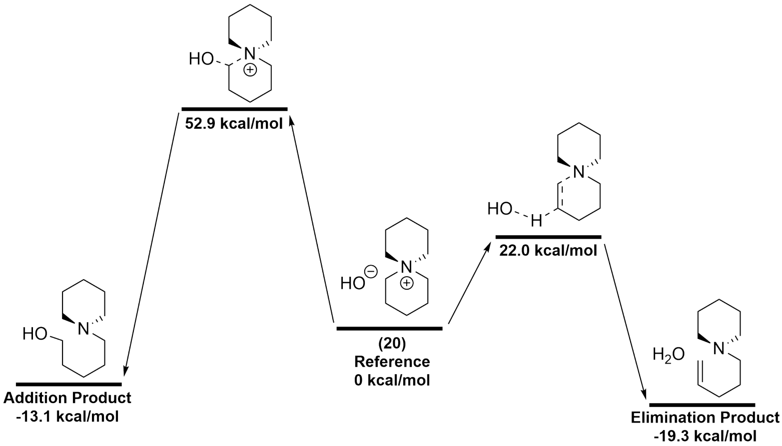


Figure 6.9: Two postulated degradation pathways of ASU **20** in an aqueous alkaline environment along with the corresponding Gibbs free energy changes and free energy barriers computed at the DFT level.

The energetics of the two studied degradation pathways of ASU are reported in Figure 6.9. The hydroxide addition reaction at an α-carbon is energetically unfavorable, having a Gibbs energy of activation of 52.9 kcal/mol. This is significantly higher than the 29.0 kcal/mol for the comparable addition reaction of in-DBD. The high energy barrier is likely caused by one ring of the ASU structure protecting the vulnerable α-carbons of the other ring.

The elimination reaction of ASU is the least inhibited energetically, having a $\Delta G^{\ddagger}$ of 22.0 kcal/mol. The energy barrier of degradation is therefore significantly lower than the 29.0 kcal/mol for the addition reaction of in-DBD and also of the 23.9 kcal/mol of the in-DBD elimination reaction. The computations therefore suggest a lower relative alkaline stability of ASU over in-DBD, as also found experimentally by Radford et al.

DMP possesses two methyl groups that may be the target of a hydroxide-induced demethylation, which was studied on top of the addition and elimination reactions studied for ASU. The energetics of these three reactions are given in Figure 6.10.

Similar to ASU, the addition reaction exhibits the highest energy barrier, requiring a Gibbs energy of activation of 51.6 kcal/mol. The methyl groups may offer steric protection to the α-carbon attacked by the hydroxide anion. The β-carbons are, however, located further away from the methyl groups and therefore potential sterical protection. This may in part explain the significantly lower $\Delta G^{\ddagger}$ of 22.1 kcal/mol required for the elimination reaction initiated through a hydroxide attack near a β-carbon. The demethylation requires the lowest observed $\Delta G^{\ddagger}$ for DMP at 20.6 kcal/mol. This is significantly below the 23.9

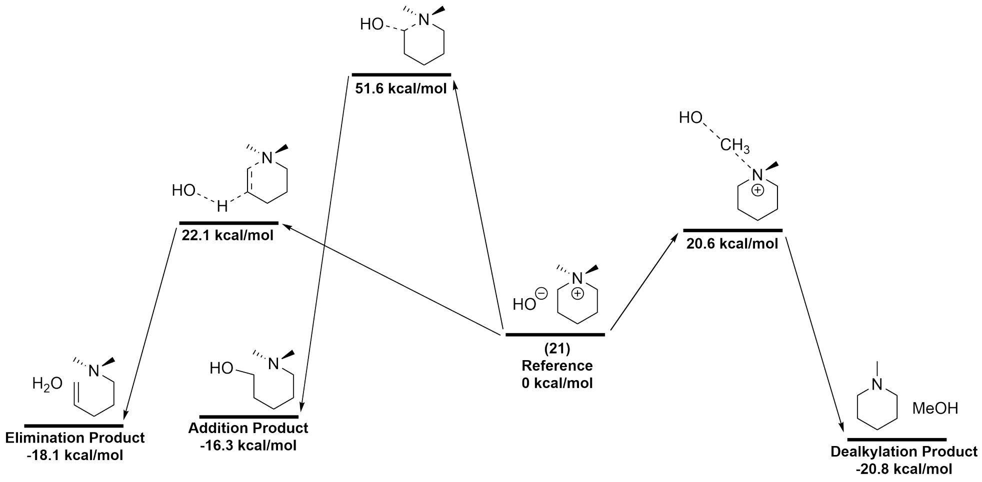


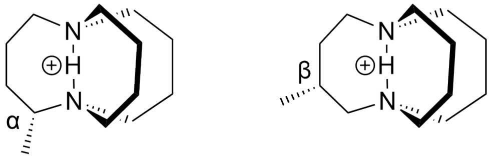
Figure 6.10: Two postulated degradation pathways of DMP (**21**) in an aqueous alkaline environment along with the corresponding Gibbs free energy changes and barriers computed at the DFT level.

kcal/mol of the lowest $\Delta G^{\ddagger}$ found for in-DBD, indicating that DMP should be considerably less stable in alkaline conditions.

6.4 Mono-Methylated in-DBD

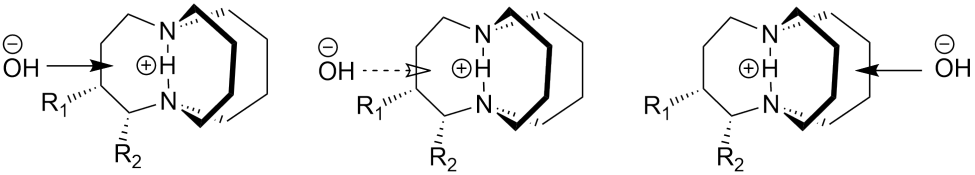
To incorporate in-DBD into an anion exchange membrane, it must be incorporated into a polymer. The cation can either be part of the polymer chain itself or be attached to a polymer backbone. In either case, the bare in-DBD needs to be augmented at some point, possibly altering the degradation pathways of the highly symmetric unsubstituted in-DBD. Analogous to the previous work on imidazolium-based compounds, the impact of structural changes to the degradation energetics are investigated. As the starting point of this investigation, a very simple augmentation of the in-DBD cation is modelled, the attachment of a single methyl group. As depicted in Figure 6.4, there are only two chemically distinct carbon atoms, therefore, only two different locations for the monomethylation must be investigated; the methylation at one α-carbon and the methylation at one β-carbon. The resulting structures were generated, and their geometry optimized; they are depicted in Figure 6.11.

As the elimination reaction was shown to be by far the most relevant degradation reaction of the unsubstituted in-DBD cation, this reaction will be studied for the two compounds. Additionally, the deprotonation reaction is also studied as this pathway might be espe-



(22) **(23)**

Figure 6.11: Two studied structures to understand the impact of mono-methylation of the in-DBD cation and its alkaline stability. A single methyl group was added to an α-carbon (**22**) and to a β-carbon (**23**).



[1] **[2]** **[3]**

Figure 6.12: The three studied approach directions of the hydroxide for the α-methylated (**22**; R_1=H, R_2=CH_3) and β-methylated (**23**; R_1=CH_3, R_2=H) in-DBD. A solid arrow indicates a path above the respective bond, while a dashed arrow indicates a path below the respective bond. The three approach directions are labeled in square brackets for future reference.

cially susceptible to changes in the in-DBD structure by exposing the central hydrogen atom.

The in-DBD molecule can be described as a structure made up of three rings, each containing four carbon atoms and the same two nitrogen atom. Consequently, there are also three openings between the rings from which the hydroxide might attack. As the added methyl groups disturb the C_3 symmetry axis of in-DBD, all three sides are chemically distinct, possibly resulting in different degradation energetics. To find the attack route with the lowest activation energy, the hydroxide attacks are modelled separately for each of the three different routes.

In the following, the approach directions are labeled as specified in Figure 6.12 as [**1**], [**2**] and [**3**]. The resulting energetics of the corresponding deprotonation reaction pathways are reported in Table 6.1.

For each structure, the lowest found activation energy is relevant for its alkaline stability. It was found that the lowest activation energy for both monomethylated structures is slightly above the activation energy of the unsubstituted in-DBD (38.6 kcal/mol). Therefore, the calculations suggest that a mono-methylation of in-DBD does not pose an adverse effect on the stability towards hydroxide-induced deprotonation and might even have a slight stabilizing effect.

Table 6.1: Computed Gibbs free energy change ΔG and Gibbs energy of activation $\Delta G^\ddagger$ of the two mono-methylated variations of in-DBD **22** and **23** for different hydroxide attack pathways. The lowest $\Delta G^\ddagger$ value for each structure is highlighted in bold.

Deprotonation Reaction Energetics [kcal/mol]	α-Methylated (**22**)		β-Methylated (**23**)	
	ΔG	$\Delta G^\ddagger$	ΔG	$\Delta G^\ddagger$
Approach Direction [**1**]	-12.6	39.7	-12.3	40.8
Approach Direction [**2**]	-12.7	**39.1**	-12.7	39.3
Approach Direction [**3**]	-12.0	44.8	-12.7	**38.8**

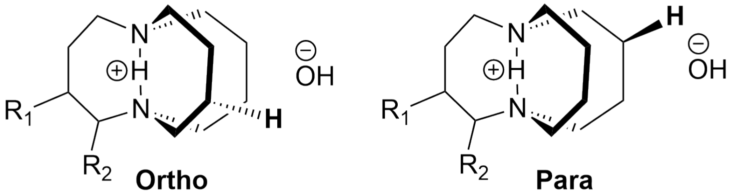


Figure 6.13: The two hydrogen atoms for each approach direction that are investigated as potential attack sites for the hydroxide-induced elimination reaction of the α-methylated (**22**; R_1=H, R_2=CH_3) and β-methylated (**23**; R_1=CH_3, R_2=H) in-DBD. The hydrogen that shares the same nitrogen as its closest neighbor with the methyl group is referred to as ortho, the other is referred to as para.

For the study of the elimination degradation pathway, the number of possible hydroxide attack sites increases and need to be considered in individual simulations. After the hydroxide has approached from the specified orientation, two β hydrogen atoms are readily accessible to the hydroxide due to their proximity. To reference these two possibilities, the hydrogen that has that same nitrogen as its closest neighbor as the methyl group is referred to as ortho, the other is referred to as para, as shown in Figure 6.13. These distinctions lead to six studied cases for each of the two molecules. The Gibbs free energy change ΔG and the Gibbs energy of activation $\Delta G^\ddagger$ of the elimination reaction are reported in Table 6.2.

The lowest computed Gibbs free energy of activation for the β-methylated in-DBD (**23**) is 24.0 kcal/mol, showing only a small deviation from the 23.9 kcal/mol for the original in-DBD. For α-methylated in-DBD **22**, the lowest found $\Delta G^\ddagger$ value is 21.9 kcal/mol for the ortho attack from the approach direction [**2**], which results in an elimination of the β-hydrogen that is located right next to the added methyl group. The eliminated hydrogen may be more susceptible to attack through steric interaction with the methyl group, lowering the degradation barrier by about 2 kcal/mol in comparison to the unsubstituted in-DBD. One notable case is the ortho attack from approach direction [**1**] for the β-

Table 6.2: Computed Gibbs free energy change ΔG and Gibbs energy of activation $\Delta G^{\ddagger}$ of the two mono-methylated variation of in-DBD **22** and **23** for different hydroxide attack pathways. The lowest $\Delta G^{\ddagger}$ value for each structure is highlighted in bold.

Elimination Reaction Energetics [kcal/mol]		α-Methylated (**22**)		β-Methylated (**23**)	
		ΔG	$\Delta G^{\ddagger}$	ΔG	$\Delta G^{\ddagger}$
Approach Direction [**1**]	Ortho	-12.6	23.6	-13.9	30.3
	Para	-10.8	24.0	-10.4	24.8
Approach Direction [**2**]	Ortho	-14.2	**21.9**	-11.7	24.3
	Para	-4.4	24.6	-5.3	**24.0**
Approach Direction [**3**]	Ortho	-12.9	24.2	-10.6	24.6
	Para	-10.5	25.0	-11.5	24.9

methylated compound **23** as it results in an exceptionally high $\Delta G^{\ddagger}$ value of 30.3 kcal/mol. To understand this deviation the transition state is shown in Figure 6.14.

The especially high Gibbs energy of activation for the ortho elimination reaction of the β-methylated in-DBD **23** from approach direction [**1**] is likely caused through steric effects. The attached methyl group is in close proximity with the attacking hydroxide, leading to additional steric strain in the transition state. This is the only one of the six studied cases in which the methyl group is directly bound to the same carbon from which the hydrogen is eliminated by the hydroxide. As the β-methylated in-DBD **23** also has additional β-hydrogens that do not show a significantly higher $\Delta G^{\ddagger}$ value of the elimination reaction, the stability in comparison to the original in-DBD is likely not strongly affected. The hydroxide will generally attack using one of the attack vectors with the lower barrier and only one out of six β-hydrogens is protected by the methyl-induced increase in $\Delta G^{\ddagger}$.

6.5 Poly-Methylated in-DBD

To further investigate the impact of minor additons to the in-DBD structure, three compounds were crafted by adding multiple methyl groups to in-DBD.

6.5.1 Vicinal Dimethylated in-DBD

Two different dimethylated variations of in-DBD were studied, with neighboring carbon atoms being substituted in both cases. In the case of molecule **24**, the neighboring α and β position were substituted while both β positions are substituted for molecule **25**. The different substitution schemes lead to different possible attack pathways for the hydroxide. For the α,α-dimethylated compound **24**, an approach from all three directions is chemically differentiable while for compound **25** the two approach directions [**1**] and [**2**] are chemically equivalent, as shown in Figure 6.15.

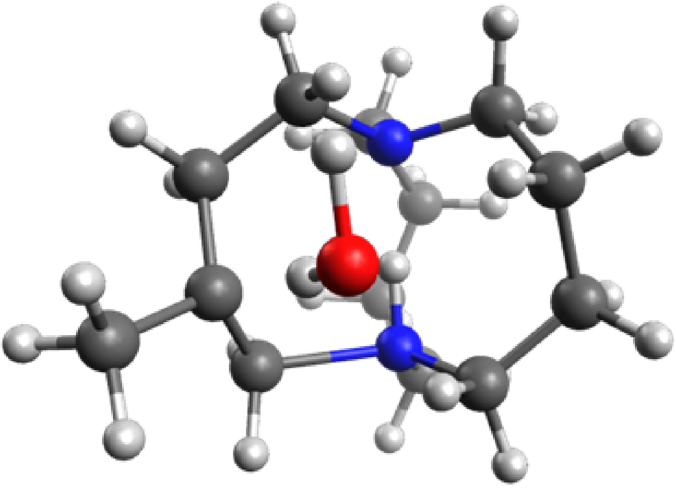


Figure 6.14: Transition state of the hydroxide-induced ortho elimination reaction of the β-methylated in-DBD (**23**) from approach direction [**1**], computed at the DFT level.

The energetics of the deprotonation from the different attack pathways are reported for compound **24** and **25** in Table 6.3. The resulting Gibbs energies of activation are slightly higher than that of the unsubstituted in-DBD for the hydroxide approach directions [**1**] and [**2**], which are both in close proximity to the introduced methyl group that likely provides a slight steric protection. A slight decrease in the Gibbs energies of activation are found for the approach direction [**3**] for both compounds.

Table 6.3: Computed Gibbs free energy change ΔG and Gibbs energy of activation $\Delta G^\ddagger$ of the two dimethylated variation of in-DBD **24** and **25** for different hydroxide attack pathways resulting in a deprotonation. The lowest $\Delta G^\ddagger$ value for each structure is highlighted in bold.

Deprotonation Reaction Energetics [kcal/mol]	α,β-Methylated (**24**)		β,β-Methylated (**25**)	
	ΔG	$\Delta G^\ddagger$	ΔG	$\Delta G^\ddagger$
Approach Direction [**1**]	-12.0	42.0	-12.2	39.2
Approach Direction [**2**]	-12.4	42.2		
Approach Direction [**3**]	-12.1	**37.5**	-12.7	**36.9**

For the elimination reaction of compound **24**, two chemically distinct β-hydrogens are possible targets of the elimination reaction and the resulting attack pathways are computed separately. They are labeled as ortho and para as shown in Figure 6.16. For approach direction [**3**] of compound **25**, both vulnerable β-hydrogens are chemically equivalent and

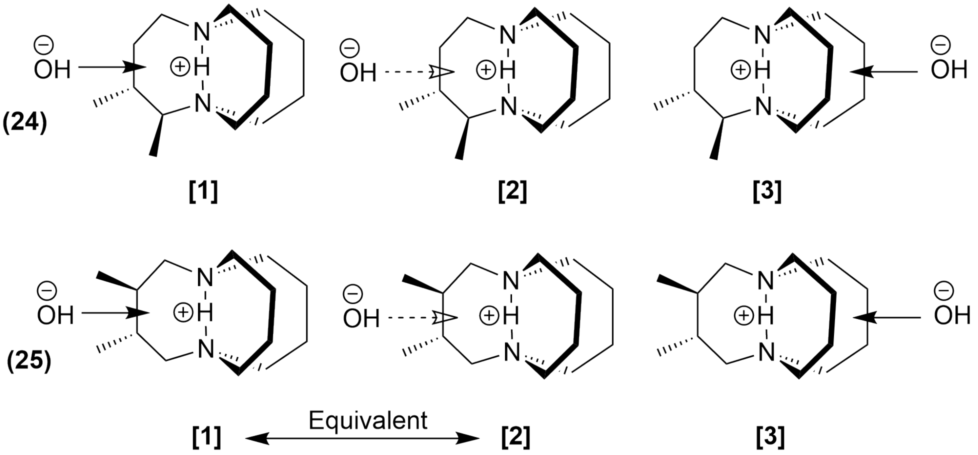


Figure 6.15: The hydroxide approach directions studied for the two dimethylated compounds **24** and **25**. A solid arrow indicates a path above the respective bond, while a dashed arrow indicates a path below the respective bond. The approach directions are labeled in square brackets for future reference. For molecule **25** the directions [**1**] and [**2**] are chemically equivalent.

for approach direction [**1**]/[**2**], the attack of the β-hydrogen at the methylated carbon atom is modelled.

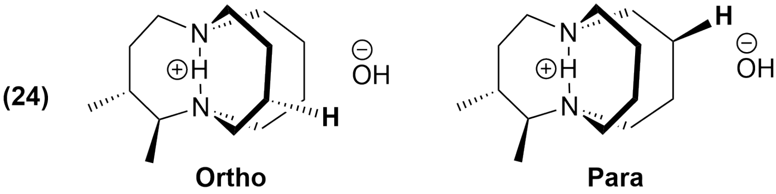


Figure 6.16: The two hydrogen atoms explicitly shown are those investigated as potential attack sites for each approach direction for the hydroxide-induced elimination reaction of the α,β-dimethylated compound **24**. The hydrogen that shares the same nitrogen as its closest neighbor with the methyl groups is referred to as ortho, the other is referred to as para.

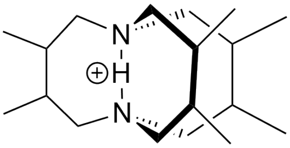
The resulting energetics of the elimination reaction for the different reaction pathways of compounds **24** and **25** are reported in Table 6.4. For both compounds, the lowest activation barrier is approximately as high as that of unsubstituted in-DBD and was found for approach direction [**3**], which is furthest away from the methylated positions. For the other approach directions, a slightly higher activation barrier was found, suggesting that the methyl groups lead to a local steric protection against hydroxide-induced elimination.

Table 6.4: Computed Gibbs free energy change ΔG and Gibbs energy of activation $\Delta G^\ddagger$ of the two dimethylated variation of in-DBD **24** and **25** for different hydroxide attack pathways resulting in an elimination. The lowest $\Delta G^\ddagger$ value for each structure is highlighted in bold.

Elimination Reaction Energetics [kcal/mol]		α,β-Methylated (**24**)		β,β-Methylated (**25**)	
		ΔG	$\Delta G^\ddagger$	ΔG	$\Delta G^\ddagger$
Approach Direction [**1**]	Ortho	-15.3	26.9	-11.5	28.8
	Para	-3.7	25.6		
Approach Direction [**2**]	Ortho	-6.3	23.9		
	Para	-10.5	24.7		
Approach Direction [**3**]	Ortho	-12.7	**23.7**	-10.7	**24.4**
	Para	-4.6	25.1		

6.5.2 Hexa-Methylated in-DBD

As the methylation of β-carbons has resulted in a local steric protection around the substituted atom, fully substituting all six β-carbon atoms may result in a comprehensive steric protection and is investigated through the β-hexamethylated compound **26** shown in Figure 6.17.



(26)

Figure 6.17: The structure of β-hexamethylated in-DBD **26**. As all six β-carbons are methylated, the molecule belongs to the point group D_3, therefore having the same degree of symmetry as unsubstituted in-DBD.

As all six β-carbons are substituted, the structure thus possesses the same degree of symmetry as unsubstituted in-DBD. Therefore, only a single reaction pathway must be studied for both the deprotonation and the elimination degradation reactions. The resulting energetics are reported in Table 6.5.

Table 6.5: Computed Gibbs free energy change ΔG and Gibbs energy of activation $\Delta G^\ddagger$ of β-hexamethylated in-DBD **26** for the deprotonation and elimination degradation pathways.

β-hexamethylated in-DBD (**26**)	ΔG	$\Delta G^\ddagger$
Deprotonation Energetics [kcal/mol]	-11.0	39.2
Elimination Energetics [kcal/mol]	-15.5	30.7

The reaction barrier of the deprotonation is only minorly affected compared to unsubstituted in-DBD, increasing slightly from 38.6 kcal/mol to 39.2 kcal/mol. The effect on the deprotonation is likely small as two counteracting effects are present. On the one hand, the methyl groups add steric protection around the central hydrogen atom. However, the methyl groups also cause the three rings of in-DBD to repel each other more strongly, thereby leading to the carbon chains to adopt a straighter and less jagged chain, opening up a bigger gap between the rings, as shown in Figure 6.18.

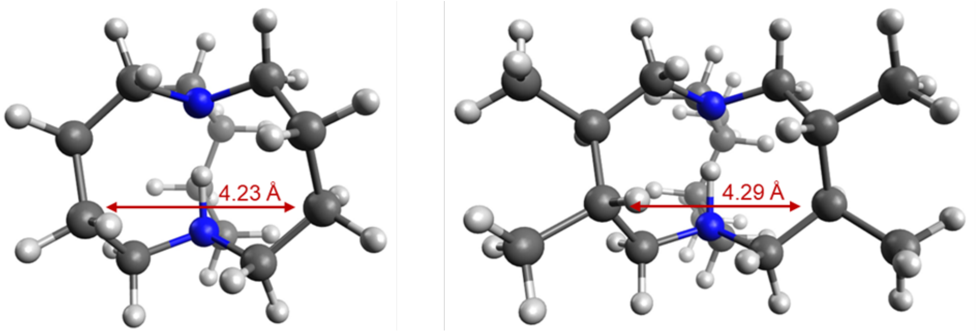


Figure 6.18: Optimized structure of unsubstituted in-DBD **19** (left) and β-hexamethylated in-DBD **26** (right). The distance between the β-carbons of neighboring rings is increased slightly by the introduction of the methyl groups, as indicated by the red arrows.

For the elimination reaction, the steric protection of the methyl groups strongly dominates, and the reaction barrier is increased to 30.7 kcal/mol compared to the 23.9 kcal/mol in the unsubstituted in-DBD. The β-hexamethylated in-DBD **26** is therefore suggested to be significantly more resistant to hydroxide-induced degradation than unsubstituted in-DBD **19**.

6.6 Benzofused in-DBD

The addition of a benzene ring that is fused to one of the in-DBD rings at the β-carbons is an interesting derivative of in-DBD for incorporation into a polymer structure. The aromatic ring is attached to the core in-DBD cation as a rigid structure that could be used as an anchor for polymer attachment while also keeping the polymer and the cation separated spatially.

As the aromatic ring points towards only one of the neighboring in-DBD subrings, the symmetry of in-DBD is broken and different approach pathways of the hydroxide have to be studied, as shown in Figure 6.19.

The energetics resulting from the different approach directions are summarized in Table 6.6. The lowest found reaction barrier for the deprotonation is significantly lower at 29.8 kcal/mol in comparison to the 38.6 kcal/mol of the unsubstituted in-DBD. The barrier of

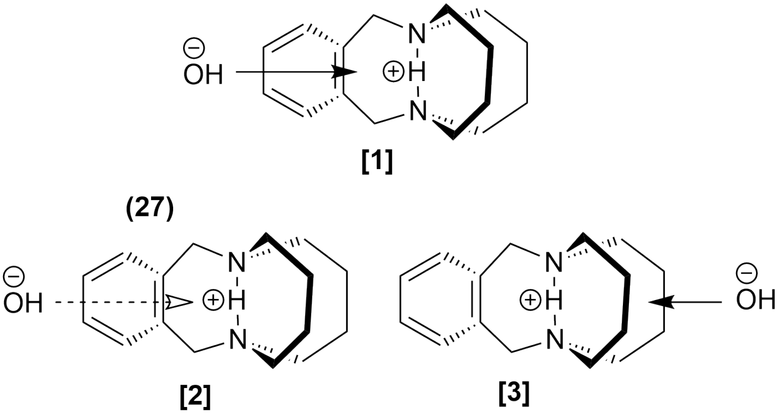


Figure 6.19: The hydroxide approach directions studied for the two benzofused in-DBD **27**. A solid arrow indicates a path above the respective bond, while a dashed arrow indicates a path below the respective bond. The approach directions are labeled in square brackets for future reference.

Table 6.6: Computed Gibbs free energy change ΔG and Gibbs energy of activation $\Delta G^\ddagger$ of the benzofused in-DBD **27** for the deprotonation and elimination degradation pathways. The lowest $\Delta G^\ddagger$ value for each degradation type is highlighted in bold.

Reaction Energetics [kcal/mol]	Deprotonation		Elimination	
	ΔG	$\Delta G^\ddagger$	ΔG	$\Delta G^\ddagger$
Approach Direction [**1**]	-14.9	35.2	-19.2	24.7
Approach Direction [**2**]	-17.2	32.0	-15.8	**22.0**
Approach Direction [**3**]	-15.9	**29.8**	-16.5	25.7

the elimination reaction is also lowered in comparison to the unsubstituted in-DBD, from 23.9 kcal/mol to 22.0 kcal/mol. The molecule is therefore suggested to be significantly more susceptible to hydroxide attacks and is expected to possess a lower alkaline stability. The higher susceptibility to degradation is likely caused by the strong effects of the rigid aromatic ring, which strongly perturbs the intricate symmetric structure of unsubstituted in-DBD as shown in Figure 6.20. The phenyl ring is tilted towards one of the rings of the core in-DBD structure, possibly adding additional strain to the structure that favors the ring opening of the elimination reaction.

6.7 Summary and Conclusions

In this computational study, in-DBD **19** was investigated with special attention to the hydroxide-induced degradation. Three possible degradation pathways were investigates, namely an addition, an elimination and a deprotonation reaction, corresponding to those

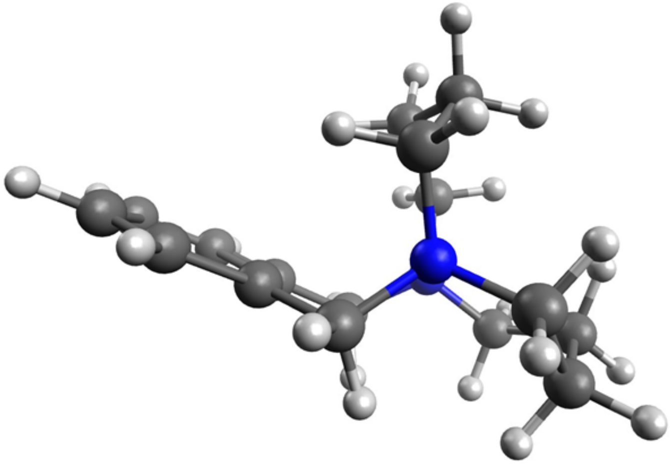
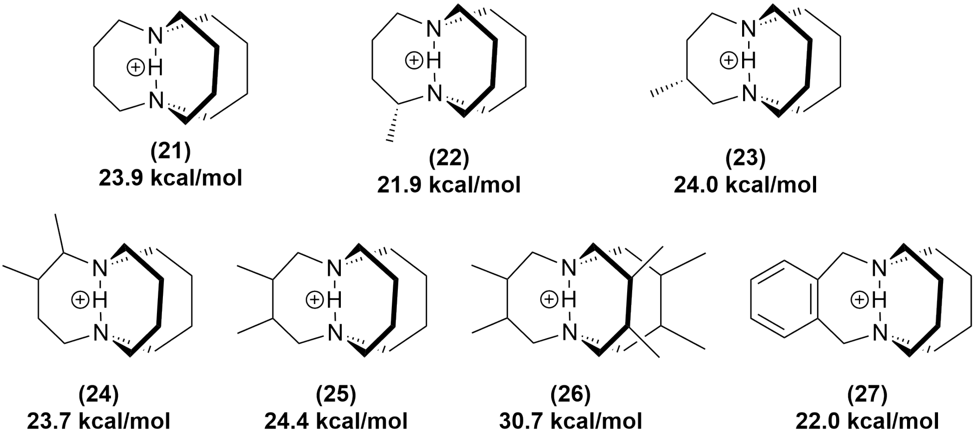


Figure 6.20: Optimized 3D structure of the benzofused in-DBD **27**. The addition of the rigid aromatic ring strongly perturbs the structure of the unsubstituted in-DBD promoting hydroxide-induced degradation.

reported by Radford et al. The deprotonation and addition reactions possess significantly higher energy barriers than the elimination reaction. The elimination reaction is therefore expected to be the main alkaline degradation cause of in-DBD, which is consistent with the experimental findings by Radford et al.[37]

To investigate the suitability of in-DBD to be incorporated into a polymer, the effect of a minimal addition to the base structure was studied through model systems. The addition of a single methyl group at the β position was found to have a small impact on the elimination reaction barrier (molecule **23**), while a reduction of about 2 kcal/mol was found for methylation in the α position (molecule **22**). This suggests that in-DBD may be attached to a polymer through an aliphatic substitution at the β position without adverse effects to the alkaline stability. The attachment through the α position may however lead to a higher susceptibility to alkaline degradation through the elimination pathway.

More complex derivatives were also studied. It was found that a dimethylation of two neighboring carbon atoms has only a minor effect on the susceptibility to hydroxide attacks, with a slightly higher resistance when two β-carbons are methylated (**25**) instead of an α- and a β-carbon (**24**). A methylation of all β-carbons of in-DBD, resulting in molecule **26**, has a strong effect on the elimination reaction barrier by sterically protecting all vulnerable hydrogen atoms, raising the $\Delta G^‡$ value by 6.8 kcal/mol, suggesting the compound to be much more durable in an alkaline environment. The fusion of a benzene ring to two neighboring β-carbons (**27**) lowers the elimination reaction barrier by 1.9 kcal/mol, as the rigid aromatic ring distorts the symmetric in-DBD structure.



(21) **23.9 kcal/mol**

(22) **21.9 kcal/mol**

(23) **24.0 kcal/mol**

(24) **23.7 kcal/mol**

(25) **24.4 kcal/mol**

(26) **30.7 kcal/mol**

(27) **22.0 kcal/mol**

Figure 6.21: The structure of in-DBD **19** and the six derivatives **22** to **27** that were studied in this work for their alkaline stability. The lowest Gibbs free energy of activation $\Delta G^{\ddagger}$ that was found for each structure is reported; it stems from the elimination reaction in each case.

In summary, this work suggests that in-DBD may be attached to a polymer backbone through an aliphatic connection without significant reduction in alkaline stability, especially when the substitution site is a β-carbon and if more than one aliphatic group is added. If synthetically feasible, a substitution at all β-carbon sites would enable a significant increase in alkaline stability while also offering attachment sites for a polymer backbone.

Significant advances in the design of stable AEM materials are expected when further work is focused on experimental measurements to evaluate the computational findings described in this chapter. In particular, the synthesis and characterization of alkaline stability of the hexamethylated in-DBD **26** is promising. If the computational findings are reproduced, molecule **26** is a prospective material for exceptionally stable AEMs, significantly surpassing the durability of even unsubstituted in-DBD. If experimental investigations of molecule **26** do not reproduce the exceptional stability suggested by the calculations, further refinement of the computational modeling is enabled, possibly uncovering further degradation pathways.

In addition to further experimental testing, the expansion of the computational work is also expected to produce valuable insights. Incorporating different numbers of explicit water molecules will enable a deeper understanding of the influence that hydration of the hydroxide has on the degradation mechanics. Higher hydration of hydroxide generally leads to a decreased degradation rate, and increased understanding of the underlying mechanisms could aid the design of materials that take advantage of this effect locally, i.e. by incorporating hydrophilic groups near sites that are vulnerable to hydroxide attack.

Finally, a combination of experimental and computational efforts should be aimed at incorporating in-DBD into a polymer structure while retaining or improving its alkaline stability. Larger substituents should be added to in-DBD to investigate their impact on alkaline stability before then constructing model polymer structures containing the in-DBD core group. For these model polymer structures, stability computations should be performed that take reaction barriers, hydration levels and motion of the hydroxide anions into account. The stability trends suggested by the stability computations can then be compared to experimental measurements to refine the computational modelling. With a refined computational model, the design of an in-DBD containing polymer with high resistance to hydroxide attack can be derived.

Conclusions

In this work, different routes towards the design of anion exchange membranes with high stability in an alkaline chemical environment have been explored, using *ab initio* simulations, dataset analysis and machine learning. A major focus was set on cations based on the imidazolium group and a further excursion was made into the in-DBD group. A method has been developed to computationally assess the alkaline stability of imidazolium-based compounds and was then applied to an expansive molecular dataset of 5832 compounds. The computational stability calculations were verified by experimentally synthesizing and testing five of the compounds in the dataset, thereby also identifying an exceptionally stable candidate compound. By analyzing the dataset and training machine learning models on the data, an in-depth understanding of the structure-stability relationship has been gained and a method has been devised for making computational stability assessments within the fraction of a second using the trained machine learning models. Lastly, *ab initio* simulations gave insights into the exceptionally stable in-DBD cation, recently reported by Radford et al., to understand the relevant degradation pathways and guide the strategy of attaching it to a polymer backbone for use in an AEM.

The identification of a computationally derived stability descriptor was reported in Chapter 3. The degradation pathways of a selection of imidazolium-based compounds were studied using highly accurate *ab initio* methods. It was identified that the hydroxide attack of the C-2 atom is the most important degradation step, due to its comparatively low energy barrier. Comparisons with experimental stability data reported in literature confirmed the importance of the C-2 hydroxide attack, as the reported half life values are correlated with both the computed Gibbs energy of activation $\Delta G^{\ddagger}_{\mathrm{HA}}$ and the Gibbs energy change ΔG_{HA}. As the Gibbs energy change requires significantly lower computational resources and manual intervention, it was identified as an ideal computational descriptor of alkaline stability for imidazolium-based compounds.

With the viable computational stability descriptor identified, it was applied to a large number of compounds, as described in Chapter 4. A selection of 18 synthetically feasible and varied substituents were systematically added to the imidazolium base structure while making precautions to eliminate compounds that are expected to pose difficulties during synthesis due to a broken symmetry. The computational workflow of deriving the stability descriptor was streamlined and fully automated through a Python script to efficiently evaluate each member of the generated molecular dataset. The resulting dataset contains a significant number of compounds that are computationally predicted to possess an exceptional alkaline stability. To verify the computed stabilities, five compounds were selected and their computed stability was verified experimentally in two different stability tests. The computed stability values showed good agreement with the experimental stability measurements and a stable penta-substituted imidazolium **14** was identified with exceptional alkaline stability in conjunction with a high ion exchange capacity.

The generated dataset not only allows the selection of promising compounds contained, but the number of contained compounds also allowed for a quantitative assessment of the structure-stability relationship that was not previously possible. The dataset was thoroughly analyzed, the results of which were described in Chapter 5. The effect of substituent choice on the computed alkaline stability could be identified for each of the three distinct substitution sites. For the C-2 substituent R_1, the substituents based on ortho substituted phenyl showed a substantially higher stability than the alternatives, due to direct and indirect steric effects. The N-1/N-3 substituent choice was shown to be less binary compared to that of C-2, generally showing a higher stability with increased steric demand. The insights into the effects of the C-4/C-5 substituent are especially illuminating, as they show that a phenyl substituent in this position leads to the lowest computed stability among the 18 studied substituents. This is especially relevant since a large fraction of the compounds reported in literature possess a phenyl group in this position. The calculations predicted that even a simple methyl group could possess a significantly higher stability, which was also experimentally verified as molecule **14** and **5** differ only in their C-4/C-5 substituent, with the choice falling on methyl and phenyl, respectively. In the solution based degradation test, the switch from a phenyl to a methyl group tripled the half life in alkaline solution and nearly eliminated the mass loss in the dynamic vapor sorption test cycle. Lastly, the stability calculation for novel compounds was drastically accelerated through a machine learning model that only requires the chemical structure as input and returns the computational stability descriptor within a fraction of a second with a mean absolute error of 1.31 kcal/mol.

In addition to the significant advances in the design of alkaline stable imidazolium-based cationic groups, an alternative cation, in-DBD **19**, was investigated through *ab initio* simulations described in Chapter 6. The in-DBD cation has been shown to possess exceptional alkaline stability by Radford et al., making it an intriguing candidate for the implementation in durable anion exchange membranes. However, the stability is related to intricate effects of the highly symmetric in-DBD structure. Adding chemical groups to the core in-DBD structure may profoundly affect the hydroxide-induced degradation reaction, but is needed to incorporate the cation into a polymer for the use in AEMs. To guide the design of a suitable polymer anchor, the degradation pathways were thoroughly investigated computationally. In agreement with the experimental data published by Radford et al., the elimination reaction was found to be the dominating factor in the alkaline in-DBD degradation. To understand the impact of changes to the in-DBD structure, small additions were made to it and their effects on the degradation pathways were computed. It was found that methylation is expected to have little to no adverse effect on stability, with a methylation on the β-carbons even showing a protective effect. By adding a methyl group to each of the six β-carbons, the core in-DBD is protected, leading to a considerably higher resistance to hydroxide attack. Fusing a phenyl ring to the structure as an anchor leads to a distortion of the symmetric in-DBD structure that comes with an increased vulnerability to hydroxide attack.

Outlook

An apparent expansion of the work presented in this dissertation involves additional experimental exploration of the many promising but experimentally still unstudied compounds available in the computed molecular dataset. The experimentally studied imidazolium **14** is already much more stable than the comparable penta-substituted imidazoliums used as reference in Chapter 3. However, a number of compounds are computationally predicted to exhibit an alkaline stability that is multiple times higher but could not yet be studied. Additional degradation tests at varying conditions should also be performed to allow comprehensive study of the effect different conditions have on stability and allow for further comparisons to compounds already reported in literature.

Replacing the C-4 and C-5 substituent from the commonly used phenyl groups for methyl groups has been showcased in this work as a highly effective method for simultaneously increasing alkaline stability and decreasing molecular weight. This strategy promotes the design of compounds that can exhibit high durability and ion exchange capacity, making them ideal for the use in effective anion exchange membranes. Computational and experimental work should be aimed towards investigating if this strategy holds for similar classes of compounds, such as bis-arylimidazolium. A notable candidate for this strategy is the bis-arylimidazolium-based compound **10**, which already exhibits very high alkaline stability despite being substituted by phenyl at C-4 and C-5 and has the potential to reach truly exceptional stability.

In addition to the investigation of individual compounds, an expansion of the molecular dataset to multivalent cations is promising, such as those based on bis-arylimidazolium. This expanded exploration of the chemical space is also expected to uncover compounds with exceptional alkaline stability. With the molecular dataset expanded from penta-substituted imidazoliums, it would contain a much larger variety of novel electronic and structural interactions. Investigating these will further deepen the understanding of the structure-stability relationship of imidazolium-based compounds and guide additional structural design improvements.

Further refinements of the computational stability calculations are also of high value. The reactivity of hydroxide is determined to a high degree by the water content around it, as seen, for example, in the dynamic vapor sorption test procedure. By incorporating varying amounts of explicit water molecules into the simulation of the degradation reactions, further insights can be obtained. With thus obtained insights, the computational stability descriptor can be tuned to a specific chemical environment, better modelling the conditions within different degradation tests and application fields.

Continued refinement of the calculations is possible by taking reaction dynamics into account in the form of molecular dynamics calculation. Within *ab initio* molecular dynamics simulations, the hydroxide attack can be examined in increased detail, especially providing an understanding of the events prior to the attack. With this understanding, additional strategies for suppressing hydroxide attacks might be foreseeable, e.g., by inhibiting hydroxide anions from approaching the attack site in the first place. Performing *ab initio* molecular dynamics simulations may also uncover yet unknown degradation pathways that were not explicitly studied in static simulations, e.g. uncovering a weak point of an added substituent.

An important aspect to be explored is the impact on alkaline stability incurred by the cations upon incorporation into a polymer structure. Model polymer structures should be constructed computationally, noting changes to the cation properties and evaluating their effect on degradation pathways. For this step, the utilization of molecular dynamics is especially worthwhile, as it allows for the investigation of the hydroxide motion and its approach towards the vulnerable sites of the cationic groups. With the comprehension gained from these calculations, strategies can be developed for incorporating the cations into a polymer structure while retaining or even increasing their stability. Possible strategies might include limiting encounters of hydroxide and cations or increasing hydration near vulnerable sites to locally reduce the hydroxide's reactivity.

The methodologies developed in this PhD project and the future work suggested should also be further applied to the highly promising in-DBD cation. An expansive investigation of the structure-stability relationship should be gained through the automated stability computation of a large number of in-DBD derivatives. Additional refinements to the computational methodology should be applied including explicit water molecules and performing molecular dynamics simulations. The impact of incorporating the cation into a model polymer structure should also be studied further. From an experimental perspective, in particular the hexa-methylated in-DBD **26** is of special interest, as it introduces an anchor site for polymer attachment while simultaneously adding additional protection to the core in-DBD group from hydroxide attack. Synthesis and stability measurements of molecule **26** are expected to be rewarding as they will deepen the understanding of the degradation reactions and possibly uncover a highly stable alternative to unsubstituted in-DBD. The work on in-DBD is of great interest as unsubstituted in-DBD already exhibits very high alkaline stability and design improvements through computational exploration may promote it to be an exceptionally durable material for AEMs.

Lastly, other aspects, apart from alkaline stability, should be investigated that are crucial for highly effective AEMs, including their conductivity, water uptake and mechanical strength. The proposed molecular dynamics simulations may be a valuable strategy to assess the ionic conductivity and mechanical properties in addition to the alkaline stability. The proposed study of explicit water molecules can additionally aid in the understanding of the water uptake and swelling behavior.

The combination of the methods and results obtained in this PhD project with the further work suggested, results in a powerful methodology to accelerate the design of effective and durable AEM materials. With these materials uncovered, it is possible to construct long-lived and performant AEM-based fuel cells and water electrolyzers, suitable for widespread adoption due to their non-reliance on rare platinum group metals. These technologies will play a pivotal role in transitioning to a sustainable energy economy and mitigating the adverse effects of climate change as much as possible.

Bibliography

[1] Keshani Attanayake, Isuru Wickramage, Udul Samarasinghe, Yasangi Ranmini, Sandali Ehalapitiya, Ruwan Jayathilaka, and Shanta Yapa. Renewable energy as a solution to climate change: Insights from a comprehensive study across nations. *Plos one*, 19(6):e0299807, 2024.

[2] Phebe Asantewaa Owusu and Samuel Asumadu-Sarkodie. A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Engineering*, 3(1):1167990, 2016.

[3] Erik Delarue and Jennifer Morris. Renewables intermittency: operational limits and implications for long-term energy system models. Technical report, MIT Joint Program on the Science and Policy of Global Change, 2015.

[4] Les Duckers. Energy storage. *Central European Review of Economics and Management (CEREM)*, 6(3):33–45, 2022.

[5] Hans Joshua C Conde, Christian M Demition, and Jaime Honra. Storage is the new black: A review of energy storage system applications to resolve intermittency in renewable energy systems. *Energies*, 18(2):354, 2025.

[6] Balázs M Fekete, Mihály Bacskó, Jiaqi Zhang, and Mengye Chen. Storage requirements to mitigate intermittent renewable energy sources: analysis for the us northeast. *Frontiers in Environmental Science*, 11:1076830, 2023.

[7] Stephen Okiemute Akpasi, Ifeanyi Michael Smarte Anekwe, Emmanuel Kweinor Tetteh, Ubani Oluwaseun Amune, Sherif Ishola Mustapha, and Sammy Lewis Kiambi. Hydrogen as a clean energy carrier: advancements, challenges, and its role in a sustainable energy future. *Clean Energy*, 9(1):52–88, 2025.

[8] Marika Wieliczko and Ned Stetson. Hydrogen technologies for energy storage: A perspective. *MRS Energy & Sustainability*, 7:E41, 2020.

[9] Martín David, Carlos Ocampo-Martínez, and Ricardo Sánchez-Peña. Advances in alkaline water electrolyzers: A review. *Journal of Energy Storage*, 23:392–403, 2019.

[10] Thomas Benjamin Ferriday and Peter Hugh Middleton. Alkaline fuel cell technology-a review. *International journal of hydrogen energy*, 46(35):18489–18510, 2021.

[11] Mandeep Singh, Dario Zappa, and Elisabetta Comini. Solid oxide fuel cell: Decade of progress, future perspectives and challenges. *International Journal of Hydrogen Energy*, 46(54):27643–27674, 2021.

[12] Yuhao Xu, Shanshan Cai, Bo Chi, and Zhengkai Tu. Technological limitations and recent developments in a solid oxide electrolyzer cell: A review. *International journal of hydrogen energy*, 50:548–591, 2024.

[13] Yun Wang, Daniela Fernanda Ruiz Diaz, Ken S Chen, Zhe Wang, and Xavier Cordobes Adroher. Materials, technological status, and fundamentals of pem fuel cells–a review. *Materials today*, 32:178–203, 2020.

[14] P Millet, R Ngameni, SA Grigoriev, N Mbemba, F Brisset, A Ranjbari, and C Etiévant. Pem water electrolyzers: From electrocatalysis to stack development. *International Journal of hydrogen energy*, 35(10):5043–5052, 2010.

[15] Maša Hren, Mojca Božič, Darinka Fakin, Karin Stana Kleinschek, and Selestina Gorgieva. Alkaline membrane fuel cells: anion exchange membranes and fuels. *Sustainable Energy & Fuels*, 5(3):604–637, 2021.

[16] Naiying Du, Claudie Roy, Retha Peach, Matthew Turnbull, Simon Thiele, and Christina Bock. Anion-exchange membrane water electrolyzers. *Chemical reviews*, 122(13):11830–11895, 2022.

[17] Yun Wang, Yiheng Pang, Hui Xu, Andrew Martinez, and Ken S Chen. Pem fuel cell and electrolysis cell technologies and hydrogen infrastructure development–a review. *Energy & environmental science*, 15(6):2288–2328, 2022.

[18] Yun Wang, Ken S Chen, Jeffrey Mishler, Sung Chan Cho, and Xavier Cordobes Adroher. A review of polymer electrolyte membrane fuel cells: Technology, applications, and needs on fundamental research. *Applied energy*, 88(4):981–1007, 2011.

[19] H Sayed-Ahmed, Árpad Istvan Toldy, and A Santasalo-Aarnio. Dynamic operation of proton exchange membrane electrolyzers—critical review. *Renewable and Sustainable Energy Reviews*, 189:113883, 2024.

[20] Kewei Hu, Jiakun Fang, Xiaomeng Ai, Danji Huang, Zhiyao Zhong, Xiaobo Yang, and Lei Wang. Comparative study of alkaline water electrolysis, proton exchange membrane water electrolysis and solid oxide electrolysis through multiphysics modeling. *Applied Energy*, 312:118788, 2022.

[21] Shahbaz Ahmad, Tahir Nawaz, Asghar Ali, Mehmet Fatih Orhan, Ayesha Samreen, and Arunachala M Kannan. An overview of proton exchange membranes for fuel cells: Materials and manufacturing. *International Journal of Hydrogen Energy*, 47(44):19086–19131, 2022.

[22] Maximilian Bernt, Alexandra Hartig-Weiß, Mohammad Fathi Tovini, Hany A El-Sayed, Carina Schramm, Jonas Schröter, Christian Gebauer, and Hubert A Gasteiger. Current challenges in catalyst development for pem water electrolyzers. *Chemie Ingenieur Technik*, 92(1-2):31–39, 2020.

[23] Yuqing Guo, Fengwen Pan, Wenmiao Chen, Zhiqiang Ding, Daijun Yang, Bing Li, Pingwen Ming, and Cunman Zhang. The controllable design of catalyst inks to enhance pemfc performance: a review. *Electrochemical Energy Reviews*, 4:67–100, 2021.

[24] Ifan Erfyl Lester Stephens, Jan Rossmeisl, and Ib Chorkendorff. Toward sustainable fuel cells. *Science*, 354(6318):1378–1379, 2016.

[25] Steffen Kiemel, Tom Smolinka, Franz Lehner, Johannes Full, Alexander Sauer, and Robert Miehe. Critical materials for water electrolysers at the example of the energy transition in germany. *International Journal of Energy Research*, 45(7):9914–9935, 2021.

[26] Dario R Dekel. Review of cell performance in anion exchange membrane fuel cells. *Journal of Power Sources*, 375:158–169, 2018.

[27] Hamish Andrew Miller, Karel Bouzek, Jaromir Hnat, Stefan Loos, Christian Immanuel Bernäcker, Thomas Weißgärber, Lars Röntzsch, and Jochen Meier-Haack. Green hydrogen from anion exchange membrane water electrolysis: a review of recent developments in critical materials and operating conditions. *Sustainable Energy & Fuels*, 4(5):2114–2133, 2020.

[28] Jian Wang, Yang Gao, Hui Kong, Juwon Kim, Subin Choi, Francesco Ciucci, Yong Hao, Shihe Yang, Zongping Shao, and Jongwoo Lim. Non-precious-metal catalysts for alkaline water electrolysis: operando characterizations, theoretical calculations, and recent advances. *Chemical Society Reviews*, 49(24):9154–9196, 2020.

[29] Jiaqiang Huang, Zongxue Yu, Junlei Tang, Pingquan Wang, Qiuyue Tan, Juan Wang, and Xianzhang Lei. A review on anion exchange membranes for fuel cells: Anion-exchange polyelectrolytes and synthesis strategies. *International Journal of Hydrogen Energy*, 47(65):27800–27820, 2022.

[30] Gautam Das, Ji-Hyeok Choi, Phan Khanh Thinh Nguyen, Dong-Joo Kim, and Young Soo Yoon. Anion exchange membranes for fuel cell application: a review. *Polymers*, 14(6):1197, 2022.

[31] Kimberly FL Hagesteijn, Shanxue Jiang, and Bradley P Ladewig. A review of the synthesis and characterization of anion exchange membranes. *Journal of materials science*, 53(16):11131–11150, 2018.

[32] Jiantao Fan, Sapir Willdorf-Cohen, Eric M Schibli, Zoe Paula, Wei Li, Thomas JG Skalski, Ania Tersakian Sergeenko, Amelia Hohenadel, Barbara J Frisken, Emanuele Magliocca, et al. Poly (bis-arylimidazoliums) possessing high hydroxide ion exchange capacity and high alkaline stability. *Nature communications*, 10(1):2306, 2019.

[33] Philip Overton, Wei Li, Xinzhi Cao, and Steven Holdcroft. Tuning ion exchange capacity in hydroxide-stable poly (arylimidazolium) ionenes: increasing the ionic content decreases the dependence of conductivity and hydration on temperature and humidity. *Macromolecules*, 53(23):10548–10560, 2020.

[34] Jince Thomas, Alex Schechter, Flavio Grynszpan, Bejoy Francis, and Sabu Thomas. *Alkaline Anion Exchange Membranes for Fuel Cells: From Tailored Materials to Novel Applications*. John Wiley & Sons, 2024.

[35] Wei You, Kevin JT Noonan, and Geoffrey W Coates. Alkaline-stable anion exchange membranes: A review of synthetic approaches. *Progress in Polymer Science*, 100:101177, 2020.

[36] Jiandang Xue, Junfeng Zhang, Xin Liu, Tong Huang, Haifei Jiang, Yan Yin, Yanzhou Qin, and Michael D Guiver. Toward alkaline-stable anion exchange membranes in fuel cells: cycloaliphatic quaternary ammonium-based anion conductors. *Electrochemical Energy Reviews*, pages 1–53, 2022.

[37] Chase L Radford, Torben Saatkamp, Andrew J Bennet, and Steven Holdcroft. An organic proton cage that is ultra-resistant to hydroxide-promoted degradation. *Nature Communications*, 15(1):3395, 2024.

[38] Guillaume Couture, Ali Alaaeddine, Frédéric Boschet, and Bruno Ameduri. Polymeric materials as anion-exchange membranes for alkaline fuel cells. *Progress in Polymer Science*, 36(11):1521–1557, 2011.

[39] Ruijie Yu, Haitao Yang, Xiaohua Yu, Jiaxin Cheng, Yuhua Tan, and Xin Wang. Preparation and research progress of anion exchange membranes. *International Journal of Hydrogen Energy*, 50:582–604, 2024.

[40] Kristina M Hugar, Henry A Kostalik IV, and Geoffrey W Coates. Imidazolium cations with exceptional alkaline stability: a systematic study of structure–stability relationships. *Journal of the American Chemical Society*, 137(27):8730–8737, 2015.

[41] Umme Salma and Nazmus Shalahin. A mini-review on alkaline stability of imidazolium cations and imidazolium-based anion exchange membranes. *Results in Materials*, 17:100366, 2023.

[42] Hai Long and Bryan Pivovar. Hydroxide degradation pathways for imidazolium cations: A dft study. *The Journal of Physical Chemistry C*, 118(19):9880–9888, 2014.

[43] Dario R Dekel, Michal Amar, Sapir Willdorf, Monica Kosa, Shubhendu Dhara, and Charles E Diesendruck. Effect of water on the stability of quaternary ammonium groups for anion exchange membrane fuel cell applications. *Chemistry of Materials*, 29(10):4425–4431, 2017.

[44] Ning Zhang, Jun Huo, Boyun Yang, Xuehua Ruan, Xiaopeng Zhang, Junjiang Bao, Wenxu Qi, and Gaohong He. Understanding of imidazolium group hydration and polymer structure for hydroxide anion conduction in hydrated imidazolium-g-ppo membrane by molecular dynamics simulations. *Chemical Engineering Science*, 192:1167–1176, 2018.

[45] Bencai Lin, Huilong Dong, Youyong Li, Zhihong Si, Fenglou Gu, and Feng Yan. Alkaline stable c2-substituted imidazolium-based anion-exchange membranes. *Chemistry of Materials*, 25(9):1858–1867, 2013.

[46] Dongyang Chen and Michael A Hickner. Degradation of imidazolium-and quaternary ammonium-functionalized poly (fluorenyl ether ketone sulfone) anion exchange membranes. *ACS applied materials & interfaces*, 4(11):5775–5781, 2012.

[47] Fenglou Gu, Huilong Dong, Youyong Li, Zhihong Si, and Feng Yan. Highly stable n3-substituted imidazolium-based alkaline anion exchange membranes: experimental studies and theoretical calculations. *Macromolecules*, 47(1):208–216, 2014.

[48] Huilong Dong, Fenglou Gu, Min Li, Bencai Lin, Zhihong Si, Tingjun Hou, Feng Yan, Shuit-Tong Lee, and Youyong Li. Improving the alkaline stability of imidazolium cations by substitution. *ChemPhysChem*, 15(14):3006–3014, 2014.

[49] Klaus Ebel, Hermann Koehler, Armin O Gamer, and Rudolf Jäckh. Imidazole and derivatives. *Ullmann's Encyclopedia of Industrial Chemistry*, 2000.

[50] Jonathan Clayden, Nick Greeves, and Stuart Warren. *Organic chemistry.* Oxford University Press, USA, 2012.

[51] Erwin Schrödinger. Quantisierung als eigenwertproblem. *Annalen der physik*, 385(13):437–490, 1926.

[52] Max Born and Werner Heisenberg. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[53] Vladimir Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61:126–148, 1930.

[54] Frank Jensen. *Introduction to computational chemistry.* John wiley & sons, 2017.

[55] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory.* Courier Corporation, 1996.

[56] Peter W Atkins and Ronald S Friedman. *Molecular quantum mechanics, Fifth Edition.* Oxford University Press, USA, 2011.

[57] Feliciano Giustino. *Materials modelling using density functional theory: properties and predictions.* Oxford University Press, 2014.

[58] Markus Bursch, Jan-Michael Mewes, Andreas Hansen, and Stefan Grimme. Best-practice dft protocols for basic molecular computational chemistry. *Angewandte Chemie International Edition*, 61(42):e202205735, 2022.

[59] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[60] Llewellyn H Thomas. The calculation of atomic fields. In *Mathematical proceedings of the Cambridge philosophical society*, volume 23, pages 542–548. Cambridge University Press, 1927.

[61] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[62] Peter W Atkins and Ronald S Friedman. *Molecular quantum mechanics, Fourth Edition.* Oxford University Press, USA, 2005.

[63] Axel D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988.

[64] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2):785, 1988.

[65] Andriy Burkov. *The hundred-page machine learning book*, volume 1. Andriy Burkov Quebec City, QC, Canada, 2019.

[66] A Géron. Hands-on machine learning with scikit-learn, keras & tensorflow. *O'Reilly*, 2019.

[67] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[68] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[69] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1):381–386, 2020.

[70] Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018.

[71] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.

[72] Gabriel Dulac-Arnold, Nir Levine, Daniel J Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, 110(9):2419–2468, 2021.

[73] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA, 1984.

[74] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[75] Wenpin Wang, Shubo Wang, Xiaofeng Xie, Vijay Ramani, et al. Density functional theory study of hydroxide-ion induced degradation of imidazolium cations. *International journal of hydrogen energy*, 39(26):14355–14361, 2014.

[76] Boryeon Lee, Dayoung Yun, Ji-Su Lee, Chi Hoon Park, and Tae-Hyun Kim. Development of highly alkaline stable oh—conductors based on imidazolium cations with various substituents for anion exchange membrane-based alkaline fuel cells. *The Journal of Physical Chemistry C*, 123(22):13508–13518, 2019.

[77] Fabian P Tipp, Kate Fraser, Mohammad J Eslamibidgoli, Kourosh Malek, Steven Holdcroft, and Michael H Eikerling. Stability descriptors for (benz) imidazolium-based anion exchange membranes. *Macromolecules*, 57(4):1734–1743, 2024.

[78] Frank Neese, Frank Wennmohs, Ute Becker, and Christoph Riplinger. The orca quantum chemistry program package. *The Journal of chemical physics*, 152(22), 2020.

[79] RBJS Krishnan, J Stephen Binkley, Rolf Seeger, and John A Pople. Self-consistent molecular orbital methods. xx. a basis set for correlated wave functions. *The Journal of chemical physics*, 72(1):650–654, 1980.

[80] Michael J Frisch, John A Pople, and J Stephen Binkley. Self-consistent molecular orbital methods 25. supplementary functions for gaussian basis sets. *The Journal of chemical physics*, 80(7):3265–3269, 1984.

[81] Vincenzo Barone and Maurizio Cossi. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *The Journal of Physical Chemistry A*, 102(11):1995–2001, 1998.

[82] Rdkit: Open-source cheminformatics. https://www.rdkit.org, doi: 10.5281/zenodo.7541264.

[83] Philipp Pracht, Fabian Bohle, and Stefan Grimme. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics*, 22(14):7169–7192, 2020.

[84] Stefan Grimme. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *Journal of chemical theory and computation*, 15(5):2847–2862, 2019.

[85] Vilhjálmur Ásgeirsson, Benedikt Orri Birgisson, Ragnar Bjornsson, Ute Becker, Frank Neese, Christoph Riplinger, and Hannes Jónsson. Nudged elastic band method for molecular reactions using energy-weighted springs combined with eigenvector following. *Journal of chemical theory and computation*, 17(8):4929–4945, 2021.

[86] Yang Guo, Christoph Riplinger, Ute Becker, Dimitrios G Liakos, Yury Minenkov, Luigi Cavallo, and Frank Neese. Communication: An improved linear scaling per-

turbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [dlpno-ccsd (t)]. *The Journal of chemical physics*, 148(1), 2018.

[87] Ronald Percy Bell. The theory of reactions involving proton transfers. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 154(882):414–429, 1936.

[88] MG Evans and M Polanyi. Further considerations on the thermodynamics of chemical equilibria and reaction rates. *Transactions of the Faraday Society*, 32:1333–1360, 1936.

[89] Anthony J Arduengo III, HV Rasika Dias, Richard L Harlow, and Michael Kline. Electronic stabilization of nucleophilic carbenes. *Journal of the American Chemical Society*, 114(14):5530–5534, 1992.

[90] Thomas A Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.

[91] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation*, 15(3):1652–1671, 2019.

[92] Jiantao Fan, Andrew G Wright, Benjamin Britton, Thomas Weissbach, Thomas JG Skalski, Jonathan Ward, Timothy J Peckham, and Steven Holdcroft. Cationic polyelectrolytes, stable in 10 m kohaq at 100° c. *ACS Macro Letters*, 6(10):1089–1093, 2017.

[93] Kristina M Hugar, Wei You, and Geoffrey W Coates. Protocol for the quantitative assessment of organic cation stability for polymer electrolytes. *ACS Energy Letters*, 4(7):1681–1686, 2019.

[94] Klaus-Dieter Kreuer and Patric Jannasch. A practical method for measuring the ion exchange capacity decrease of hydroxide exchange membranes during intrinsic degradation. *Journal of Power Sources*, 375:361–366, 2018.

[95] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[96] Avogadro: An open-source molecular builder and visualization tool. http://avogadro.cc/, Version 1.2.0.

[97] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4:1–17, 2012.

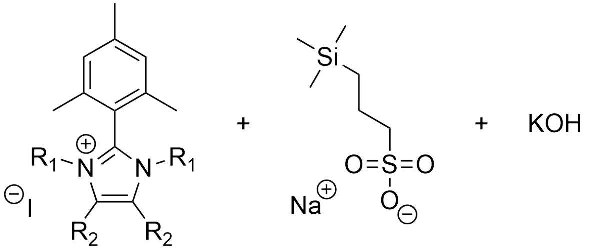
[98] Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Hagen Neugebauer, Sebastian Spicher, Christoph Bannwarth, and Stefan Grimme. A generally applicable atomic-charge dependent london dispersion correction. *The Journal of chemical physics*, 150(15), 2019.

[99] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18):3297–3305, 2005.

[100] Florian Weigend. Accurate coulomb-fitting basis sets for h to rn. *Physical chemistry chemical physics*, 8(9):1057–1065, 2006.

Appendix

Procedure of the Solution Degradation Test



In a glove box, the imidazolium in its iodide form (imidazolium **5**: 7 mg, imidazolium **12**: 8 mg, imidazolium **13**: 8 mg, imidazolium **6**: 9 mg, imidazolium **14**: 6 mg), 56 mg of grinded KOH and 3 mg of the sodium salt of 3-(Trimethylsilyl)-1-propanesulfonic acid (NaDSS) were added to a vial. Outside of the glovebox under argon flow, 0.50 mL deuterated methanol was added to the solids, the mixture was added to a NMR tube and flame sealed. After measuring an initial ^{1}H-NMR, the tubes were placed in an oil bath at 80 °C and further ^{1}H NMR measurements were performed at intervals of approximately five days. A final ^{1}H NMR was performed after at least 30 days.

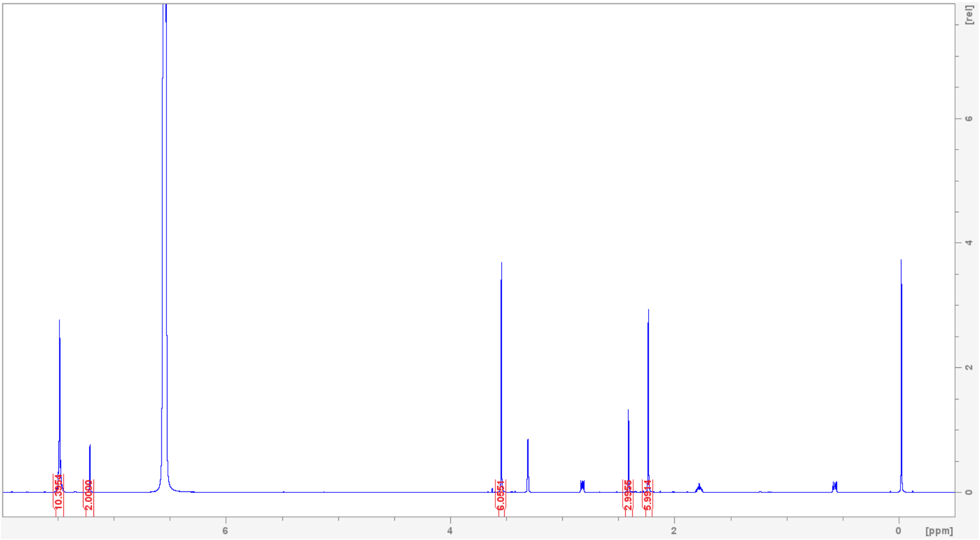
The fraction of imidazolium remaining was quantified by determining the ratio of the integral of the 2H signal of the mesityl group at around 7.25 pm and the integrals of all non-obstructed signals of the NaDSS and dividing those values by the respective value at day 0 of the degradation test.

Procedure of the Dynamic Vapor Sorption Degradation Test

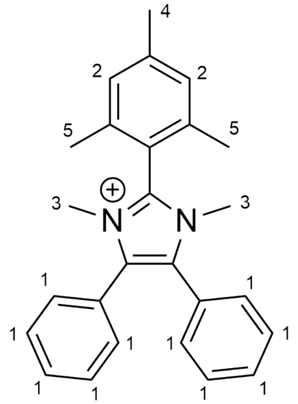
About 30 mg of the iodide form of the imidazolium was dissolved in approximately 1 mL of a degassed mixture of methanol and deionized water at varying ratios (imidazolium **5**: 1:1, imidazolium **12**: 1:1, imidazolium **13**: 2:1, imidazolium **6**: 2:1, imidazolium **14**: 1:1). The solution was passed twice through approximately 1 g of Amberlyst A-26(OH) ion exchange resin and rinsed once with an additional 0.5 mL of the solvent mixture. The solution was then given into the Dynamic Vapor Sorption device and the following sequence of target relative humidity (RH) stages was started at 60 °C:

1. 120 min, 40 % RH
2. 600 min, 70 % RH
3. 120 min, 80 % RH
4. 120 min, 70 % RH
5. 120 min, 60 % RH
6. 120 min, 50 % RH
7. 120 min, 40 % RH
8. 120 min, 30 % RH
9. 240 min, 70 % RH
10. 120 min, 30 % RH
11. 180 min, 20 % RH
12. 180 min, 15 % RH
13. 240 min, 10 % RH
14. 240 min, 70 % RH
15. 120 min, 30 % RH
16. 120 min, 10 % RH
17. 180 min, 7.5 % RH
18. 180 min, 5 % RH
19. 180 min, 2.5 % RH
20. 180 min, 1 % RH
21. 240 min, 70 % RH

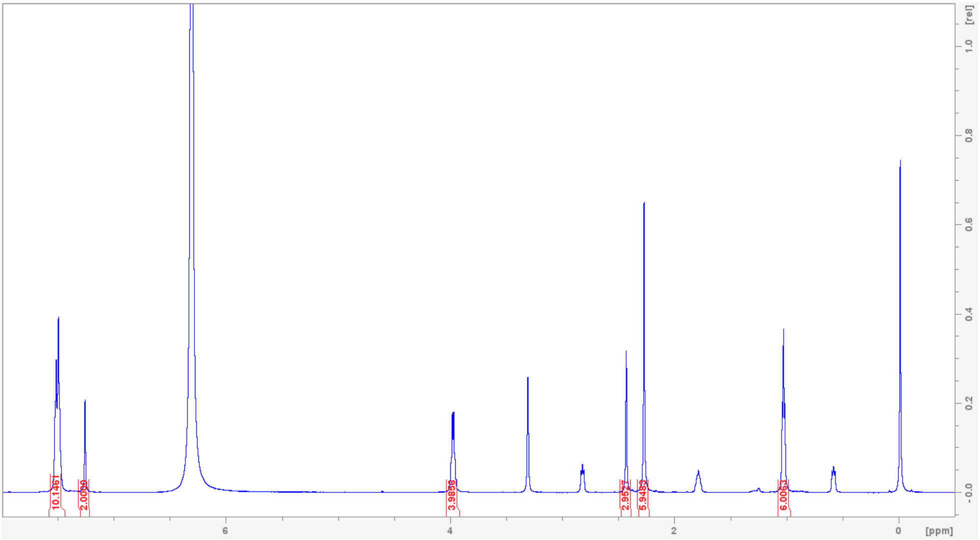
^{1}H NMR of Imidazolium 5



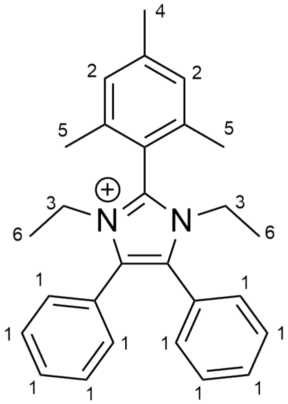
^{1}H-NMR: (600 MHz, MeOD-D$_4$) δ [ppm] = 7.48 (m, 10H, H-1), 7.22 (s, 2H, H-2), 3.55 (s, 6H, H-3), 2.41 (s, 3H, H-4), 2.24 (s, 6H, H-5).



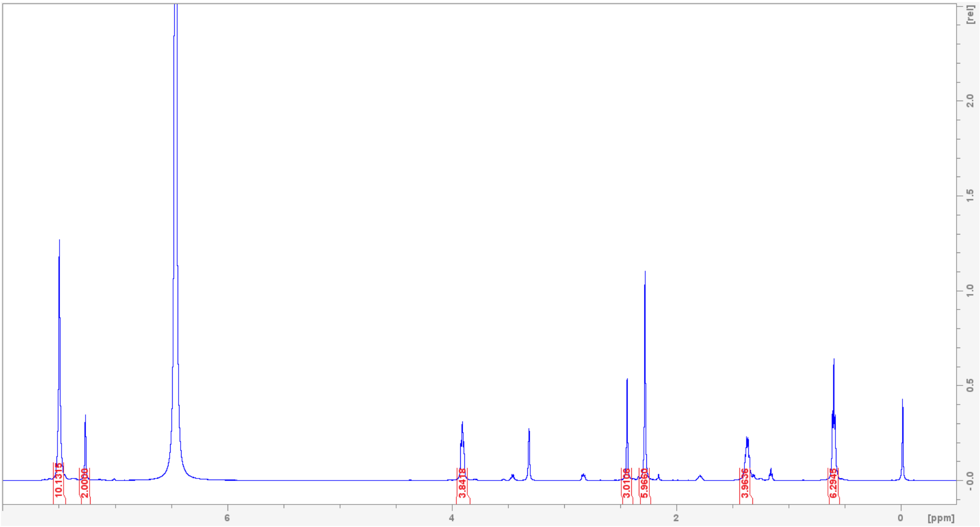
^{1}H NMR of Imidazolium 12



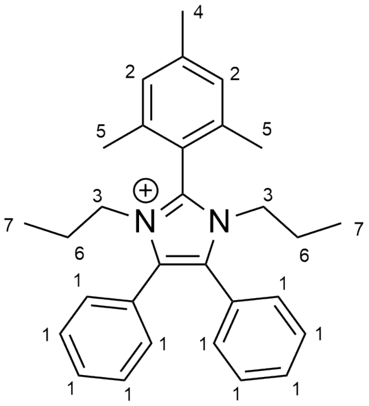
^{1}H-NMR: (600 MHz, MeOD-D$_4$) δ [ppm] = 7.50 (m, 10H, H-1), 7.26 (s, 2H, H-2), 3.97 (q, 4H, H-3), 2.43 (s, 3H, H-4), 2.27 (s, 6H, H-5), 1.03 (t, 6H, H-6).



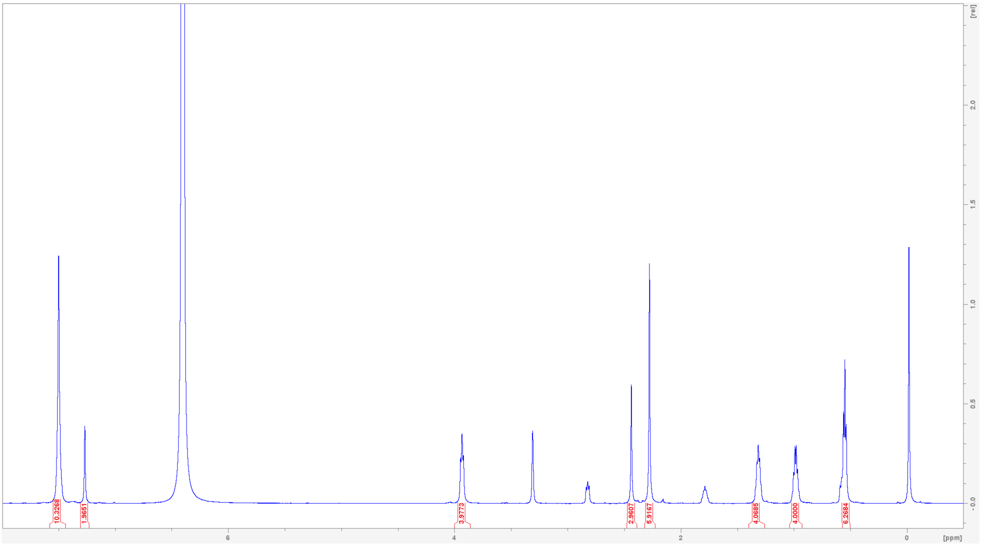
^{1}H NMR of Imidazolium 13



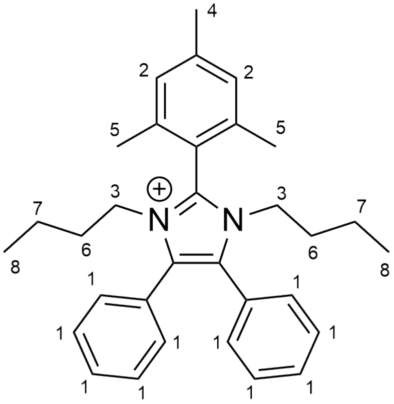
^{1}H-NMR: (600 MHz, MeOD-D$_4$) δ [ppm] = 7.49 (m, 10H, H-1), 7.26 (s, 2H, H-2), 3.90 (t, 4H, H-3), 2.44 (s, 3H, H-4), 2.28 (s, 6H, H-5), 1.37 (m, 4H, H-6), 0.59 (t, 6H, H-7).



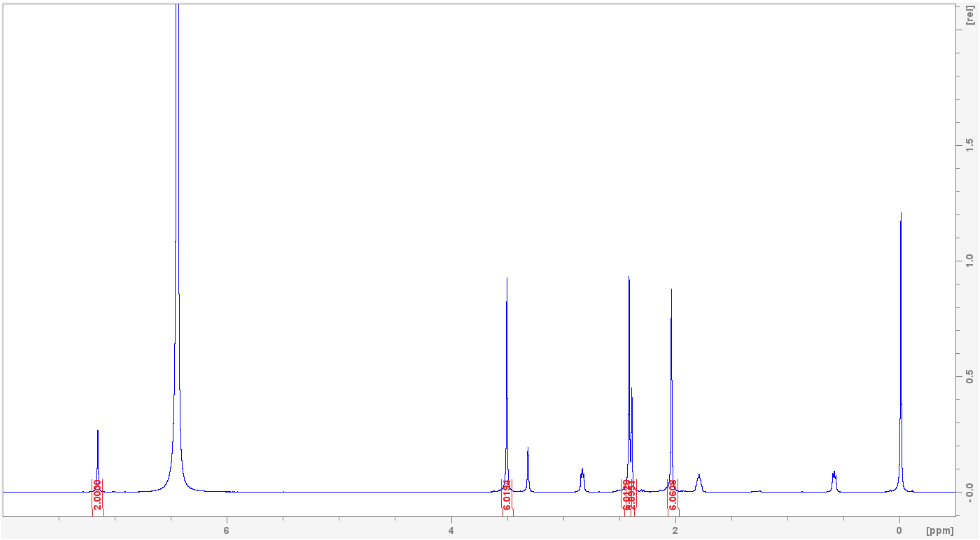
^{1}H NMR of Imidazolium 6



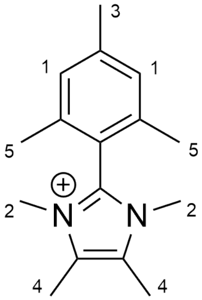
^{1}H-NMR: (600 MHz, MeOD-D_4) δ [ppm] = 7.50 (m, 10H, H-1), 7.27 (s, 2H, H-2), 3.93 (t, 4H, H-3), 2.44 (s, 3H, H-4), 2.28 (s, 6H, H-5), 1.32 (m, 4H, H-6), 0.99 (m, 4H, H-7), 0.55 (t, 6H, H-8).



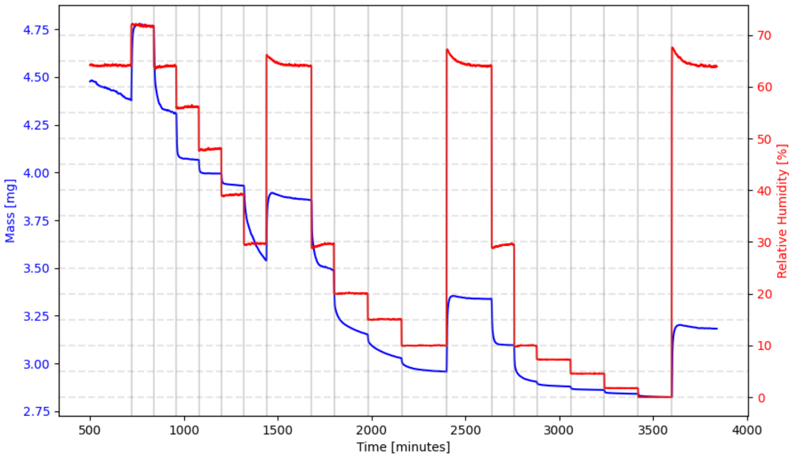
^{1}H NMR of Imidazolium 14



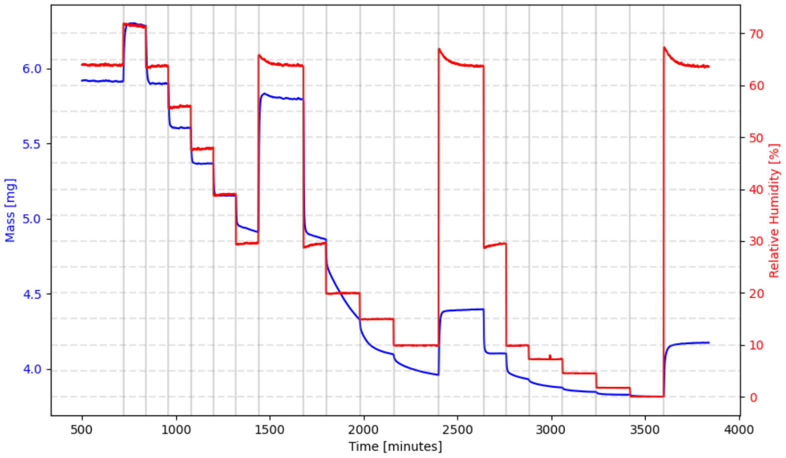
^{1}H-NMR: (600 MHz, MeOD-D$_4$) δ [ppm] = 7.15 (s, 2H, H-1), 3.50 (s, 6H, H-2), 2.41 (s, 6H, H-4/H-5), 2.38 (s, 3H, H-3), 2.03 (s, 6H, H-4/H-5).



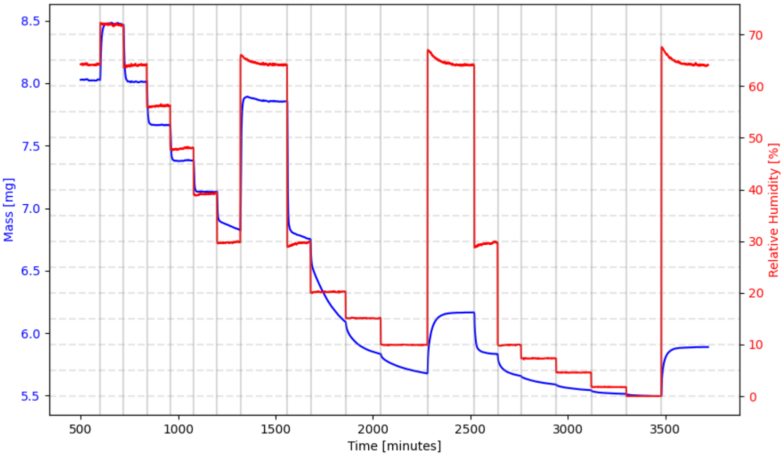
DVS Measurements of Imidazolium 5



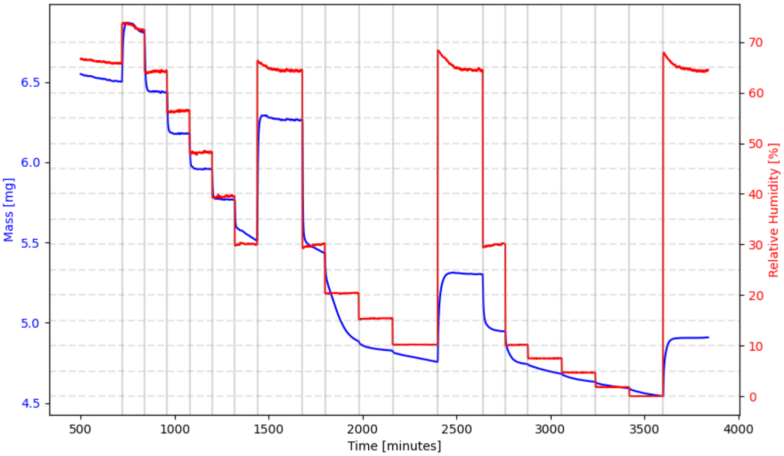
DVS Measurements of Imidazolium 12



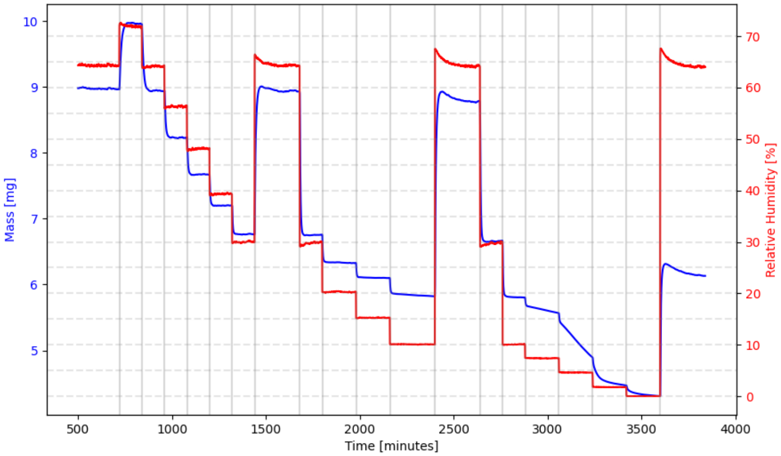
DVS Measurements of Imidazolium 13



DVS Measurements of Imidazolium 6



DVS Measurements of Imidazolium 14



Statement of Contributions

The selection of the substituents used in Chapter 4 was performed with the help of Dr. Mohammad Javad Eslamibidgoli, Sai Govind Hari Kumar and Dr. Kate Fraser. All five compounds **5**, **6**, **12**, **13** and **14** used for experimental verification were synthesized by myself. Dr. Kate Fraser synthesized compounds **6** and **13** in parallel and her batch was used for the degradation test as they were available at an earlier time. For the solution-based degradation test, the samples were prepared together with Dr. Chase Radford. Setup of the DVS procedure for the five compounds was performed together with either Dr. Chase Radford or Dr. Torben Saatkamp. The last ^{1}H-NMR spectrum of compounds **5**, **6**, **12** and **13** and the last three ^{1}H-NMR spectra of compound **14** were performed by Dr. Chase Radford and Dr. Torben Saatkamp, as they had to be taken after the research stay was over. For non-substantive tasks including spelling and grammar checks, the generation of simple Python and bash scripts, and literature research, generative artificial intelligence was utilized.

List of Publications

Contributions to the following publications published so far were made in the time frame of the PhD Project.

- Tipp, F. P., Fraser, K., Eslamibidgoli, M. J., Malek, K., Holdcroft, S., & Eikerling, M. H. (2024). Stability Descriptors for (Benz) imidazolium-Based Anion Exchange Membranes. *Macromolecules*, 57(4), 1734-1743.
- Cheong, O., Tipp, F. P., Eikerling, M. H., & Kowalski, P. M. (2024). Entropy Effects on Reactive Processes at Metal-Solvent Interfaces. *The Journal of Physical Chemistry C*, 128(19), 7892-7902.
- Chen, Y., Rodenbücher, C., Morice, A., Tipp, F., Kowalski, P. M., & Korte, C. (2024). Understanding the Mid-Infrared Spectra of Protic Ionic Liquids by Density Functional Theory. *The Journal of Physical Chemistry B*, 128(47), 11723-11729.

Band / Volume 689
Development of an oxygen ion conducting solid oxide electrolysis cell based on gadolinium-doped cerium oxide as fuel electrode and electrolyte material
D. Ramler (2026), ix, 162 pp
ISBN: 978-3-95806-879-7

Band / Volume 690
Design of Local Multi-Energy Systems: Impact of Coupled Energy Vector Integration and Grid Service Participation
P. S. Glücker (2026), xxviii, 145 pp
ISBN: 978-3-95806-880-3

Band / Volume 691
A Parallel-in-Space Simulator for Accelerating Power System Simulation on Graphics Processing Units
J. Zhang (2026), 112 pp
ISBN: 978-3-95806-882-7

Band / Volume 692
Entwicklung von Reparaturmethoden für einkristalline Bauteile mittels thermischer Spritzverfahren
M. L. Létang (2026), X, 211 pp
ISBN: 978-3-95806-883-4

Band / Volume 693
Assessing the Environmental Implications of Offshore Wind Energy Advancements on the Future German Electricity Sector
A. Benitez (2026), xi, 176 pp
ISBN: 978-3-95806-885-8

Band / Volume 694
Entwicklung von Korrosionsschutzschichten für Protonen-Austausch-Membran-Wasserelektrolyseure
T. Sievert (2026), 201 pp
ISBN: 978-3-95806-888-9

Band / Volume 695
Hierarchical Modeling of Electrocatalytic Reactions from a Local Perspective
X. Zhu (2026), ix, 121 pp
ISBN: 978-3-95806-889-6

Schriften des Forschungszentrums Jülich
Reihe Energie & Umwelt / Energy & Environment

Band / Volume 696
Nanocrystalline Silicon Carbide in Transparent Passivating Contact Solar Cells
A. Eberst (2026), xiii, 225 pp
ISBN: 978-3-95806-891-9

Band / Volume 697
Theory of Electronic and Ionic Perturbations at Supported Electrocatalyst Nanoparticles
Y. Zhang (2026), XI, 131 pp
ISBN: 978-3-95806-896-4

Band / Volume 698
Aufbau und Einsatz eines on-board Messsystems zur Untersuchung der Abgaszusammensetzung von Fahrzeugen betrieben mit konventionellen und alternativen Kraftstoffen
V. Polinowski (2026), VIII, 269 pp
ISBN: 978-3-95806-897-1

Band / Volume 699
Ab Initio-based large-scale Atomistic Simulations of Cathode Materials for Secondary Batteries: From Computational Methodologies to Applications towards improved Structural and Chemical Stability
K. Köster (2026), x, 204, x pp
ISBN: 978-3-95806-898-8

Band / Volume 700
Materials Design, Processing and Application of Proton-Conducting Oxides for Electrochemical Energy Conversion
Y. Zeng (2026), vii, 117 pp
ISBN: 978-3-95806-899-5

Band / Volume 701
Accelerating the discovery of alkaline-stable anion exchange membrane materials via computational exploration
F. P. Tipp (2026), xii, 135 pp
ISBN: 978-3-95806-900-8

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp

Energie & Umwelt / Energy & Environment
Band / Volume 701
ISBN 978-3-95806-900-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum