Dynamic Solvent Models and Exploring the Parameter Space of Hydrogen Fluoride, Hafnium, and Zirconium

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Marco Manni

October 2022

The herein presented research was conducted from January 2019 until October 2022 under the guidance of Prof. Dr. Bettina Keller at the institute of chemistry and biochemistry of the Freie Universität Berlin.

1st Reviewer:
Prof. Dr. Bettina G. Keller
Department of Biology, Chemistry and Pharmacy
Physical and Theoretical Chemistry
Freie Universität Berlin
Arnimallee 22, 14195 Berlin

2nd Reviewer:
Prof. Dr. Beate Paulus
Department of Biology, Chemistry and Pharmacy
Physical and Theoretical Chemistry
Freie Universität Berlin
Arnimallee 22, 14195 Berlin

Date of Disputation: 26.01.2023

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. Bettina Keller for supporting me throughout the project. Many discussions were held, with critical impact on the course of this work. Despite two years of pandemic in the middle of this work, Bettina managed to keep the group together as well as ensure continuous communication throughout the years of home-office.
She gave me many helpful tips and pointers throughout the work and had a big impact on the final outcome. Furthermore, I want to express my gratitude to Prof. Dr. Beate Paulus for taking the role of the second supervisor.
The following names should not remain unmentioned and I am grateful for their respective contributions:
Jennifer Anders for close cooperation, for supplying calculated partial charges, helping with simulating NMR-spectra as well as scientific and general discussions.
Dr. Anselm Loges for the close cooperation regarding the geochemical twins. His expertise helped advance the project a lot and his experimental results were the foundation of one of the projects.
Dr. Patrick Pröhm for measuring the requested NMR-spectra as well as being very helpful in providing related information afterwards. Stefanie Kieninger for the many helpful scientific and non-scientific discussions.
Marius Wenz for being a cheerful colleague to share an office with. The rest of the AG Keller in the time of 2019-2022 for helpful inputs and being great.
The SFB 1349 for funding this thesis throughout the years and providing a useful environment for networking as well as opportunities for growth.

Declaration of Authorship

I, Marco Manni, declare that this thesis titled, "Dynamic Solvent Models and Exploring the Parameter Space of Hydrogen Fluoride, Hafnium, and Zirconium" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Contents

Abstract

Solvation is the interaction of solute and solvent. Every biological interaction happens in a solvent. Most technical procedures occur within solvents and geological processes are too mediated by their solvents. Understanding these effects of solvation is therefore critical for the understanding of biology, technology, and geology.
While static properties of solvation shells, like coordination numbers and radial distribution functions, are well understood, the dynamic properties of these open systems are rarely studied.
Furthermore, an interesting solvation-based phenomenon is the separation of the geochemical twins Zirconium and Hafnium in fluoride-bearing media. In this work, I present a method for evaluating Markov models of solvation shells and investigate ways of combining the solvent models with solute models.
Moreover, I attempt to find a suitable two-site and three-site model of hydrogen fluoride, specialized for its interaction with metal ions. By simulating aqueous and pure HF for several combinations of q, σ_F, and ϵ_F and evaluating density, peaks of radial distribution functions as well the solvation free energy of NaF in HF, I hoped to find a suitable model.
A three-site model for HF is parameterized by recreating the electrostatic potential of HF with a classical force field, focusing on the location of maximum potential which takes a conical shape around the tip of the ellipsoid and is not located at the poles.
A new method is presented which allows the automatic detection of coordination polyhedra based on reference structures and Steinhardt-order parameters.
The Lennard-Jones parameter space for tetravalent cations is explored and analyzed in terms of static solvation shell quantities.
Finally, the thermal contraction of the solvation shells of Zr^{4+} and Hf^{4+} in 1 M HF was investigated using classical MD simulations. The Markov models of solvation shells indicated that solvent dynamics couple close to the solute. Additive combined models yielded slightly higher timescales compared to their individual components. The opposite is true for the multiplicative models which performed just as well or even worse than their components.
The parameterization of HF, for the two-site model, yielded two parameter combinations that could reproduce three of the five target quantities, the relevant peaks of the F-H and H-H radial distribution functions, as well as the solvation free energy of NaF in HF. After choosing a topology for the three-site model, the Lennard-Jones parameter scans were unable to yield stable simulations of aqueous HF. The project was therefore discontinued and I settled for a recently published HF model.
The parameterization of metal cations yielded a very robust result. Static solvation shell properties exist on continuous regions of similar value in the parameter space. These regions appear as a diagonal lines in the $\log(\epsilon_M) - \sigma_M$ parameter space.
This behavior is also observed in the coordination polyhedra found by the novel method.
The thermal contractions of solvation shells of tetravalent cations could be observed for the four ionic ligands. The contractions are a result of water molecules increasing their distance to the central cation. Their missing repulsive Coulomb interaction allows the ionic ligands to move in closer to the central cation, thus causing the thermal contraction. Furthermore, we observed a two-state system for the solvation shells at high temperatures which consists of octahedral and tetrahedral solvation shells interchanging each other. The herein presented results offer new methods for analyzing solvation shells. Firstly by constructing Markov models of these open

systems to study their dynamics. Secondly, by automatically determining the coordination polyhedron, which is essentially an analysis of the angular distribution of solvent molecules in the solvation shell.
The parameterization attempts of HF depict the difficulty of finding parameter combinations that match all fitting targets, albeit the searched parameter space was rather small.
The parameter space for tetravalent cations shows an extremely robust result which can yield the basis for future parameterization attempts.
Finally, the peculiar thermal contractions could be explained through classical MD simulations. This shows the power of this method for studying hard ionic systems.

Zusammenfassung

Solvatation ist die Wechselwirkung zwischen gelöstem Stoff und Lösungsmittel. Jede biologische Wechselwirkung findet in einem Lösungsmittel statt. Die meisten technischen Verfahren finden in Lösungsmitteln statt, und auch geologische Prozesse werden durch ihre Lösungsmittel vermittelt. Das Verständnis der Auswirkungen der Solvatation ist daher entscheidend für das Verständnis von Biologie, Technik und Geologie.
Während die statischen Eigenschaften von Solvathüllen wie Koordinationszahlen und radiale Verteilungsfunktionen, gut bekannt sind, werden die dynamischen Eigenschaften dieser offenen Systeme nur selten untersucht.
Ein interessantes Phänomen, das auf Solvatation beruht, ist die Trennung der geochemischen Zwillinge Zr und Hf in fluoridhaltigen Medien. In dieser Arbeit stelle ich eine Methode zur Auswertung von Markov-Modellen von Solvathüllen vor und untersuche Möglichkeiten zur Kombination der Lösungsmittelmodelle mit den Modellen für gelöste Stoffe.
Außerdem versuche ich, ein geeignetes Zwei- und Drei-Teilchen-Modell für Fluorwasserstoff zu finden, das auf seine Wechselwirkung mit Metallionen spezialisiert ist. Indem ich wässrige und reine HF für verschiedene Kombinationen von q, σ_F und ϵ_F simuliere und die Dichte, die Maxima der Radialverteilungsfunktionen sowie die freie Solvatationsenergie von NaF in HF auswerte, hoffe ich, ein geeignetes Modell zu finden.
Das Drei-Teilchen-Modell für HF wird parametrisiert, indem das elektrostatische Potenzial von HF mit einem klassischen Kraftfeld nachgebildet wird, wobei der Schwerpunkt auf dem Ort des maximalen Potenzials liegt, das eine konische Form um die Spitze des Ellipsoids annimmt und nicht an den Polen liegt.
Es wird eine neue Methode vorgestellt, die die automatische Erkennung von Koordinationspolyedern auf der Grundlage von Referenzstrukturen und Steinhardt Parametern ermöglicht.
Der Lennard-Jones-Parameterraum für vierwertige Kationen wurde erforscht und im Hinblick auf statische Solvatationsschalengrößen analysiert.
Schließlich wurde die thermische Kontraktion der Solvathüllen von Zr^{4+} und Hf^{4+} in 1 M HF mit klassischen MD-Simulationen untersucht. Die Markov-Modelle der Solvathüllen deuten darauf hin, dass die Dynamik des Lösungsmittels eng mit der des gelösten Stoffes verbunden ist. Additiv kombinierte Modelle ergaben im Vergleich zu ihren Einzelkomponenten geringfügig bis deutlich höhere Zeitskalen. Das Gegenteil gilt für die multiplikativen Modelle, die genauso gut oder sogar schlechter abschnitten als ihre Komponenten.
Die Parametrisierung von HF für das Zwei-Teilchen-Modell ergab zwei Parameterkombinationen, die drei der fünf Zielgrößen, die Maxima der radialen Verteilungsfunktionen F-H und H-H sowie die freie Solvatationsenergie von NaF in HF, reproduzieren konnten. Nach der Wahl einer Topologie für das Drei-Teilchen-Modell waren die Lennard-Jones-Parameter-Scans nicht in der Lage, stabile Simulationen von wässrigem HF zu liefern. Das Projekt wurde daher eingestellt, und ich entschied mich für ein kürzlich veröffentlichtes HF-Modell.
Die Parametrisierung der Metallkationen führte zu einem sehr robusten Ergebnis. Statische Solvathülleneigenschaften existieren in kontinuierlichen Regionen mit ähnlichen Werten im Parameterraum. Diese Regionen erscheinen als diagonale Linie im Parameterraum $\log(\epsilon_M) - \sigma_M$.
Dieses Verhalten ist auch bei den Koordinationspolyedern zu beobachten, die mit der neuen Methode gefunden wurden.

Die thermischen Kontraktionen der Solvathüllen von vierwertigen Kationen konnten für die vier ionischen Liganden beobachtet werden. Die Kontraktionen sind das Ergebnis von Wassermolekülen, die ihren Abstand zum zentralen Kation vergrößern. Ihre fehlende abstoßende Coulomb-Wechselwirkung ermöglicht es den ionischen Liganden, sich näher an das zentrale Kation heranzubewegen, was die thermische Kontraktion verursacht. Darüber hinaus haben wir bei hohen Temperaturen ein Zwei-Zustands-System für die Solvatationsschalen beobachtet, das aus oktaedrischen und tetraedrischen Solvatationsschalen besteht, die sich gegenseitig austauschen. Die hier vorgestellten Ergebnisse bieten neue Methoden zur Analyse von Solvathüllen. Erstens durch die Konstruktion von Markov-Modellen dieser offenen Systeme zur Untersuchung ihrer Dynamik. Zweitens durch die automatische Bestimmung des Koordinationspolyeders, das im Wesentlichen eine Analyse der Winkelverteilung der Lösungsmittelmoleküle in der Solvathülle ist.
Die Parametrisierungsversuche von HF zeigen, wie schwierig es ist, Parameterkombinationen zu finden, die alle Anpassungsziele erfüllen, auch wenn der gesuchte Parameterraum recht klein war.
Der Parameterraum für vierwertige Kationen zeigt ein äußerst robustes Ergebnis, das die Grundlage für zukünftige Parametrisierungsversuche liefern kann.
Schließlich konnten die eigenartigen thermischen Kontraktionen durch klassische MD-Simulationen erklärt werden. Dies zeigt die Leistungsfähigkeit dieser Methode für die Untersuchung harter ionischer Systeme.

Chapter 1

Introduction

1.1 Motivation

Fluorine stands out among all elements as the most electronegative. Its unique ability to draw electrons towards it results in the unique properties of elemental fluorine and fluorinated compounds. In nature, fluorine occurs mostly as ionic fluoride in minerals. This means that fluorinated molecules can show properties unseen in nature. Fluorinated organic compounds find use as pharmaceuticals [1], water-repelling and non-stick coatings [2], surfactants [3], and agrochemicals [4]. In the laboratory, metal fluoride complexes are used as catalysts in fluorination reactions [5]. In nature, metal fluoride complexes occur in the aqueous fluid phases of the earth's crust. Aqueous fluids play a role in redistribution processes of elements in the earth. It is critical to understand these processes in order to understand the formation of economically interesting deposits. Fluorine plays a special role here by mobilizing certain high field strength elements (HFSEs) and even separating pairs of ions which are known as geochemical twins.
These pairs are Zr and Hf as well as Y and Ho. They are called geochemical twins because their similar charges and atomic radii cause them to occur together and in ratios close to those found in bulk silicate earth. Deviations from this trend occur as Hf-enriched minerals and are caused by mobilizations of the HFSEs in the late hydrothermal stage of crystallization. This is attributed to the presence of depolymerizing elements such as F, Li, B, Na, and K [6–8].
The fractionation of the geochemical twins Zr and Hf is therefore a fluorine-specific effect which will play a central role in this thesis.
Understanding the fractionation of Zr and Hf is also of technical importance, because they vary strongly in their ability to absorb neutrons. This makes Hf-free Zr a key compound in alloys which are used for the cladding of nuclear fuel rods [9]. Furthermore, by replacing Hf with Zr in the crystal structure, zirconolite ($CaZrTi_2O_7$) might be used as storage containers for actinides and enriched Pu [10, 11].
I used molecular dynamics (MD) simulations to investigate systems of highly charged metal ions in water under the presence of fluoride. Being hard Lewis-acids and bases respectively the ions tend for form strong complexes with ionic character. These hard systems are well suited for classical MD simulation where bond breaking and bond forming are not possible. In turn, classical MD-simulations are highly computationally efficient, compared to *ab-initio* MD. In general, MD simulations offer an excellent resolution in space and time simultaneously, making them powerful tools to gain chemical insights.
Experimentally, solvation shell geometries, in terms of ion-ligand distances, coordination numbers and coordination polyhedra can be extracted from X-Ray absorption, X-ray scattering and neutron scattering experiments [12]. Computationally, distances and coordination numbers are straight-forward to extract from simulations.

The coordination polyhedron is difficult to determine automatically. Therefore, I introduce a new method for determining the coordination polyhedron based on a library of reference structures.
Dynamic solvation shell properties can give insights into chemical systems. Including the solvation shell dynamics into Markov models for the solute can also lead to increased convergence of the models. In this work, a method for constructing and combining Markov models of solvation shells is investigated.
While classical MD is a computationally efficient way of gaining insight into chemical processes with high spatial and temporal resolution, it suffers from a few drawbacks. The simulation results are heavily dependant on the parameterization of the system. An extensive parameter search is almost always required when investigating a new system. Thus, parameterization makes up a large portion of this work. Chemical reactions, i.e. bond breakage and formation, cannot occur in a classical MD-simulations. Hard ionic systems are well suited for classical MD simulations due to their comparatively low covalent contribution to bond formation.

1.2 Markov Models of Solvation Shells

Solvation is an omnipresent phenomenon that influences the thermodynamic and dynamic properties of chemical species. Examples include cooperative solvation effects that drive protein folding [13], conformationally dependent membrane permeabilities of cyclic peptides [14, 15], the free energy surface of alanine dipeptide [16] or of the RNA hairpin-loop motif [17], the association of globular protein-protein complexes [18, 19], the freezing point of aqueous solutions of HF at various concentrations [20, 21], the nucleation and growth of minerals in solution [22], and the performance of Li ion batteries [23]. To understand these effects, one needs to understand the interplay between solvent and solute, i.e. the coupling of conformational solute dynamics to the dynamics of its solvation shells. These dynamics can couple so tightly that it becomes adequate to think of solute and first solvation shell as a single functional entity [24].
From a theoretical point of view, solvation shells are a special case of open diffusive systems of equivalent particles. Open and diffusive systems pose a problem because only a handful of the thousands or ten-thousands or more particles constitute the solvation shell. Recording the position of every particle is computationally demanding. Numerating every particle is also nonsensical in real life, therefore the equivalence of solvent particles facilitates the problem. From the viewpoint of the solute, the solvation shell can be represented by the solvent particle density. This density does not differentiate between specific water molecules (e.g. water molecule #718 vs water molecule #21) but only accounts for positions occupied by solvent particles.
Previous works attempted to tackle the problem using shape descriptors like Steinhardt parameters [25]. Lechner and Dellago[26] used a modification of these descriptors to differentiate between different crystal structures, Shevchuk [27] used them to identify phase transitions in the crystallization of NaCl and Bandyopadhyay [28] applied them to assess the two-state behavior of water. Gu et al. [29] introduced a "solvent fingerprint", which captures the total degree of solvation for each atom of the solute. This approach was further developed by Harrigan et al. [30], who computed solvent fingerprints for spheres of different radii, thus preserving some spatial resolution. The solvent fingerprint is based on solvent-solute distance and can report on wetting/de-wetting processes in large bio-molecular rearrangements.

However, as most angular information is discarded, there is too little information to resolve the detailed coupling of solvent molecules to a solvated complex or similar. Markov models for water dynamics were previously applied by Shevchuck [31] to study the proton transfer via the Zundel complex. Hamm [32] constructed MSMs based on the root mean squared deviation (RMSD) of $(H_2O)_6$ clusters and used the resulting first right eigenvector as an order parameter for the two-state behavior of water. Schulz [33] built models for water trimers based on relative coordinates to yield insight into the dynamics of the hydrogen-bond network.
In this work, we introduce a model that resolves the solute dynamics in parallel with the dynamics of the solvation shell. This would allow studying the precise coupling between solute and solvation shell. To include angular information, we use the spherical harmonics as ansatz functions for a Markov model. To do this we will use the framework of variational Markov models [34], which allows for arbitrary ansatz functions. The eigenvectors of the Markov model will be linear combinations of the ansatz functions.
We study the dynamics of the Hexa-aqua-aluminium-complex $[Al(H_2O)_6]^{3+}$ and its first solvation shell. The complex with its six ligands acts as a model solute, we describe its conformational dynamics based on the rotation of individual water ligands around the Al-O-axis. The conformational space of this motion exhibits four maxima for a single ligand and two maxima for the whole complex. The solvation shell density is described as a function of the angular parts of spherical coordinates, i.e. on the θ-ϕ-plane. The solute dynamics are described using a variational model, a model constructed on the torsion angles of the ligands, and a model based on individual torsion angles, as well as the torsion angles of neighboring ligands. Descriptions of the solvent are solely variational and are later included in the solute models. Two methods of joining Markov models are tested, additive and multiplicative models. These joint models give further insight into the interplay between solvent and solute.

1.3 Parameterizing Hydrofluoric Acid

Hydrogen fluoride is a colorless gas at room temperature with a stinging odor. It is highly corrosive toward metals and has the unique ability to dissolve glass. This substance is miscible with water at any ratio and forms hydrofluoric acid as a result. Hydrofluoric acid is mainly used for the production of fluorocarbons. These substances find use as coolants and in the production of polymers. Hydrofluoric acid is also used for the etching of silicium-based silicon wafers and hence, plays a crucial role in the electronics industry. Further uses include the separation of metals from ores, the production of pharmaceuticals, insecticides, and herbicides, prevention of caries as well as the production of crystal glass and ceramics [35]. In 1984, Cournoyer and Jorgensen fitted the first three-site model of HF based on the physical properties of the liquid instead of gas phase data [36]. More than a decade later, in 1997, Jedlovsky published an improved three-site model for HF in[37], where length, fluorine-virtual site distance, and charges have been set to reproduce the experimental values of the bond length, dipole moment, and quadrupole moment. The Lennard-Jones parameters were determined by fitting the internal energy and density of the system. Eight years later, Kreitmeir slightly modified the model of Jedlovsky in [38] to better reproduce pair correlation functions measured by neutron diffraction for the liquid phase. In their model, only the F-H distance was changed, while all other parameters were kept as in [37]. The F-X distance was determined from the ratio of F-X distance to the

F-H distance reported in the publication from the 1990s. Orabi and Faraldo-Gomez published the latest model for hydrogen fluoride in 2020 in [39]. The F-H distance and Lennard-Jones parameters, ϵ_F, and σ_F were inherited from Cournoyer and Jorgensen. The charges and the position of the virtual site were set to reproduce the *ab initio* geometry of the $(HF)_2$ dimer, the density and the enthalpy of vaporization of liquid HF at 296 K and 1.184 atm.
The previously mentioned publications all parameterized their models based on physical data of the liquid, as well data for the $(HF)_2$ dimer. However, no attempts have been made to incorporate the interaction of HF with metal ions into the parameterization procedure. An HF-model specialized for metal cation would prove useful to describe the fluorine-specific separation of the geochemical twins Zr and Hf, as well as Y and Ho.
In this work, I will try to parameterize a two-site model of HF by simulating pure and aqueous HF for a variety of Lennard-Jones parameter combinations and charges. The charges were provided in cooperation with J. Anders who computed them using quantum mechanical methods. From the simulations, densities, maxima of the radial distribution functions, and the solvation free energy of NaF in HF were extracted and compared to reference values. In the next step, I looked for parameter combinations that recreated several of the target quantities simultaneously. Unfortunately, the two-site model could recreate three quantities simultaneously at best.
In a second attempt, I tried to parameterize a three-site model of HF. Charges and location of the virtual site were determined by fitting the electrostatic potential of the molecule to *ab inito* calculations. With this result, another set of Lennard-Jones parameter combinations was simulated. Unfortunately, this attempt proved unfruitful and the project was discontinued from hereon.

1.4 Tetravalent Metal Cations in Fluorous Environments

Metal cations are a superb system for investigating solvation shell properties. Their descriptions in classical MD simulations essentially consist of four numbers: their charge, their mass, and the two Lennard-Jones parameters σ_M and ϵ_M. The former can be understood as an ionic radius as it determines the location of the minimum of the Lennard-Jones potential, whereas the latter can be understood as the ions polarizability due to the fact that it determines the depth of the minimum of the Lennard-Jones potential. Since the charge and mass of the ions are known *a priori* only the Lennard-Jones Parameters need to be determined. The group of Kenneth Merz and his co-workers provided a plethora of Lennard-Jones parameters for monovalent [40, 41], divalent [42], trivalent and tetravalent metal cations [43, 44] in combination with a multitude of water models. Some of these parameterizations were performed for a modified Lennard-Jones potential to better recreate experimental quantities. The target quantities for metal ions typically include coordination number (CN), hydration free energy (HFE) as well as the ion oxygen distance (IOD). The previously mentioned publications contain parameter combinations that could recreate either IODs or HFEs for the metal cations but not both. Zhang et al in [45] reported parameter combinations that were able to recreate both quantities, however, these parameter combinations lie in rather extreme parts of the parameter spaces with $\epsilon_M > 10^3$ kcal/mol and $\sigma_M < 1$ Å. The works by Zhang et al and the group of K. Merz both refer to experimental data of Hf or Zr in the absence of fluorine [46, 47]. However, the complexation of Zr and Hf in fluoride-bearing solutions differs from the respective complexation in fluoride-free solutions [48, 49]. That is, one observes an 8-fold

square antiprismatic coordination in fluoride-free solutions and a 6-fold octahedral coordination in fluoride-bearing solutions. This work aims to find suitable metal parameters for Zr and Hf in fluorous environments and furthermore investigate the dependence of Lennard-Jones Parameters on static solvation shell properties. Furthermore, a novel method is introduced which allows for the automatic detection of coordination polyhedra (CP) which can be used as a fitting target as well. Metal fluorides of the form MF_4 are simulated at room temperature in water for a large section of the parameter space. Static solvation shell properties are evaluated, including the novel coordination polyhedron.

1.5 Thermal Contractions of the Solvation Shells of Geochemical Twins

The geochemical twins are the two pairs of Zirconium and Hafnium as well as Yttrium and Holmium. These pairs are called twins because of their similar ionic radii and charge which causes the ion pairs to occur in similar ratios in different rocks. The geochemical twins are difficult to separate, however, fluorous environments are able to perform this fractionation [50]. Separating Zr from Hf is relevant in the production of the outer cladding of fission rods. The Zr used for this process must be free of Hf, due to the large neutron cross section of Hf [51]. That is, impurities of Hf in the Zr-alloy would make it difficult for neutrons to permeate the cladding of the fission rods. Several technical processes for this fractionation have been developed as well, using molten copper halides [52] or re-crystallization of K_2ZrF_6 [53] among others. A. Loges studied Zr and Hf in 1 M HF at temperatures ranging from 100 °C-400 °C using X-ray Absorption Near Edge Spectroscopy (XANES) as well as Extended X-ray Absorption Fine Structure spectroscopy (EXAFS). The spectroscopic methods yield insight into the solvation shells of the ions. The results are pre-published in [54]. The XANES data reveal that the solvation shell must be in an octahedral configuration. The fits to the EXAFS-Data indicate that the mean distance of ligands to the central cation decreases with increasing temperature. This peculiarity will be the focus of this project.
From solubility experiments [49], one can deduce that the stoichiometry of the solvation shell must be Zr:F:O = 1:2:2. This method cannot detect neutral ligands. Therefore, the complexes must be of the constitution $[MF_2(OH)_2(H_2O)_2]^0$, with M = Zr, Hf. I wanted to identify the reason behind the thermal contractions observed in the experimental results. The known constitution of the complexes allows for 5 different configurations of the ligands. These complexes were simulated in classical MD simulations in the NVT ensemble under experimental conditions, i.e. 1 M HF and Temperatures 25 °C-400 °C. The boxes were set up such that the density corresponds to the isobaric temperature of water at 400 bar, corresponding to the experimental conditions. The mean distances of the solvation shells were analyzed as well as distance distributions for the different ligands. The stabilities of the complexes were assessed and a thermal contraction could be observed. An explanation for the contraction is provided as well.

Chapter 2

Research Questions

One of the questions addressed in this work is: How to describe the dynamic properties of solvation shells, which are essentially open systems, in terms of Markov models? Furthermore, how can the Markov models of the solvation shells be combined with the Markov models of the solute? Also, does the quality of the Markov models improve in terms of absolute timescales, resolved processes, and convergence of timescales?
The geochemical twins Zr and Hf play a major role in this work. In order to adequately describe these cations in fluorous environments, the following question was tackled: What are the Lennard-Jones parameters and topology for the hydrogen fluoride molecule which is adequate to describe its interactions with metal cations as anhydrous HF and aqueous HF?
The static solvation shell properties of metal cations were investigated too. I asked, how do the solvation shell properties like coordination number, coordination polyhedron, ion-oxygen distances, and fluoride content of the solvation shells depend on the Lennard-Jones parameters of the metal cation? Furthermore, will any of the parameter combinations be suitable candidates for the description of Hf or Zr in fluorous environments?
Finally, A. Loges measured thermal contractions of the first solvation shells of Zr and Hf in 1 M HF at temperatures up to 400 °C. Can this peculiarity be explained with classical MD? If yes, what is the cause of this phenomenon?

Chapter 3

Theoretical Foundation

3.1 Solvation

Liquids take on the shape of their container without changing their volume. This ability to adapt and rearrange also emerges when foreign objects are introduced into the liquid. These object can be macroscopic, like electrodes or microscopic, like ions. In both cases the liquid forms a layer around the introduced object. This process is called solvation. Through solvation, the liquid becomes the solvent and the object becomes the solute.
The solvent molecules in close vicinity to the solute are different from the bulk solvent molecules. The solvent forms one or more layers of molecules around the solute. These layers are called solvation shells. The immediate surroundings of the solute is the first solvation shell.
We can study the constitutions, geometries and energetics of solvation shells to understand the process of solvation. Why do some substances readily dissolve in water while others remain underwater undissolved for millenia? Why do some substances heat up the solvent while dissolving when others cool it down? To answer these questions and others one needs to consider the phenomenon of solvation.

3.1.1 Solvent Models in Computational Chemistry

Solvent models are methods in computational chemistry to take the liquid environment of a chemical system into consideration. One distinguishes between implicit solvent models and explicit solvent models. Implicit solvent models are computationally cheaper but they only model an averaged behaviour of the solvent as a whole. They are widely used for estimating solvation free energies but typically do not account for hydrophobic effects, the viscosity of the solvent or the hydrogen bond network with the solvent. Explicit models include atomistic representations of the solvent molecules. This allows the formation of many layers of solvation shells and the exchange of solvent molecules between the shells as well as with the bulk. In order to simulate bulks of water, large boxes need to simulated in conjunction with periodic boundary conditions.

Implicit Solvent Models

This section is based on [55, 56]. Implicit solvent models aim to approximate the electrostatic contribution to the free energy of solvation by modelling the solvent as a continuum dielectric. Consider a sphere of charge q and radius r, e.g. a hard ion. The charge distribution ρ on the sphere is given by

$$\rho(\mathbf{s}) = \frac{q}{4\pi r^2} \tag{3.1}$$

where **s** in equation 3.1 refers to a surface element of the sphere. The electrostatic potential ϕ at position **r** outside of the sphere is then

$$\phi(\mathbf{r}) = -\frac{q}{4\pi\epsilon_0\epsilon r} \tag{3.2}$$

where r is the distance to the center of the sphere, ϵ_0 is the vacuum permittivity and ϵ refers to the dielectric constant of the environment. The work of charging the sphere G can be computed by integrating over the entire surface S.

$$G = -\frac{1}{2}\int_S \rho(\mathbf{s})\phi(\mathbf{s})d\mathbf{s} \tag{3.3}$$

$$= -\frac{1}{2}\int_S \frac{q}{4\pi r^2}\frac{q}{4\pi\epsilon_0\epsilon r}d\mathbf{s} \tag{3.4}$$

$$= -\frac{q^2}{8\pi\epsilon_0\epsilon r} \tag{3.5}$$

Here, the factor $4\pi r^2$ in the first term cancels out with the same term stemming from the integral over the surface. For the gas phase, $\epsilon = 1$, computing the difference between gas phase and solution (i.e. $\epsilon \neq 1$), for a mole N_A of spheres leads to the Born equation, eq 3.6.

$$G_P = -\frac{N_A q^2}{8\pi\epsilon_0 r}\left(1 - \frac{1}{\epsilon}\right) \tag{3.6}$$

Deriving the Born-equation for non-spherical solutes leads to a multitude of approaches. A popular method is the generalized Born equation (GBE), which becomes the Born equation in the limit of vanishing interatomic distances and the subsequently becomes Coulomb's law for large distances with a gaussian function mediating between the two extremes. Consider a solute with N atoms, the GBE can be written as

$$G_P = -\frac{N_A q^2}{8\pi\epsilon_0 r}\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{\gamma_{ij}} \tag{3.7}$$

with

$$\gamma_{ij} = \left(d_{ij}^2 + r_i r_j \exp\left(-\frac{-d_{ij}^2}{r_i r_j}\right)\right)^{1/2}$$

where d_{ij} is the distance between atom i and j.
Treating the solute as a continuous conductor leads to the conductor-like-screening-model (COSMO). The method assumes that the electric potential of the solvent must cancel out the electric potential of the solute. One can therefore compute the charge distribution of the solvent with a charge distribution of the molecule gained from high level theoretical calculations. In COSMO, the charges of the solvent q are scaled by $f(\epsilon)$, compared to the charges of solute $q*$. That is

$$q = f(\epsilon)q* \tag{3.8}$$

$$f(\epsilon) = \frac{\epsilon + 1}{\epsilon + x} \tag{3.9}$$

where the value of x is set to 0.5 for neutral molecules and 0 for ions. An improvement of this method is COSMO-RS [57] which aims to reproduce solvation shells too.

Explicit Solvent Models

Explicit solvent models are atomistic representations of solvation shells and bulk water in an MD simulation. That is, the solute of interest is simulated together with thousands to ten-thousands of solvent molecules. Bond lengths, bond angles, partial charges and, Lennard-Jones parameters of the solvent molecule all need to be parameterized in order to produce meaningful results. Many such parameterizations have been performed for the most common solvent: water. These water models are categorized by the amount of sites they employ. In this context a site is either a real atom or a so called dummy atom. The dummy atoms carry only charge but no mass and are used to mimic the true charge distribution of the solvent molecules and improve performance.
3-site models include the simple point charge (SPC) [58] which stands out by its use of a tetrahedral H-O-H angle, and improved version this model called SPC/E [59] which includes a correction term for the average polarization. The transferable intermolecular potential functions (TIPS) [60] was chosen to yield reasonable structural and energetic results for monomers, dimers and pure liquids. An improvement of the TIPS model with better density and energy for liquid water is published as TIP3P [61].
4-site models for water contain one virtual site (dummy atom) which is located close to the center of mass of the water molecule. One such model called TIP4P was published together with TIP3P [61]. Many specialized versions of TIP4P were released, like TIP4P-Ew [62] which is made for Ewald simulations and TIP4P/2005 in [63] which aims to describe the entire phase diagram of condensed water.
5-site models, like TIP5P [64] exist too. These models contain two virtual sites which mimic the free electrons of the oxygen atom in water. Due to their high computational cost and poor performance [63], these models are rarely used.

3.1.2 Accessing Solvation Shells Experimentally

This section is based on the review published in [65].

The main tools for studying solvation shells are spectroscopic methods like X-Ray absorption. X-Ray absorption near edge structure (XANES) can be used to determine coordination numbers and coordination polyhedra by comparing the measured spectra to reference compounds of known geometry. Extended X-ray absorption fine structure (EXAFS) data together with a guess for the first solvation shell can be fitted to obtain interatomic distances in the solvation shell. Neutron- and X-Ray diffraction experiments yield radial distribution functions, sometimes called pair correlation functions. These functions are extremely valuable for understanding the solvation shells. The function contains a peak for every solvation shell, integrating over a single peak between two local minima yields the coordination number of that solvation shell. The locations of the peaks are indicative for the size of the solvation. NMR-Experiments can be used to determine the exchange rate of water molecules in the solvation shells.
The free energy of solvation is the central quantity when studying the energetics of solvation. It is experimentally accessible via calorimetry. The mobility of ions in solution is related to dynamic properties like conductivity, electrical mobility and diffusion. Quasi elastic neutron scattering (QENS) can be used to determine the self-diffusion coefficient of hydrated ions. NMR measurements can be used to determine rotational correlation times of waters in the first solvation shell. QENS and NMR

measurements are also able to determine the residence times of water molecules in the first solvation shells.
Another way of accessing the above mentioned quantities are MD simulations. They offer the largest simultaneous temporal and spatial resolution of all methods, however, they are dependent on their parametrization for classical MD or limited in the system size accessible for *ab-initio* methods, due to the computational cost.

3.2 MD Simulations

This section is grossly based on [66]. Classical molecular dynamics (MD) simulations propagate a system of N particles $\mathbf{x}(t) = \{\mathbf{x}_i(t)\}_{i=1}^{N}$, with positions $\mathbf{x}_i$ and velocities $\mathbf{v}_i$ in time according to Newton's laws of motion. That is, the acceleration of each particle is derived from the forces acting upon it.

$$\mathbf{a}_i(t) = \frac{\mathbf{F}_i(t)}{m_i} \tag{3.10}$$

In equation 3.10 $\mathbf{a}_i(t)$ is the acceleration of particle i at time t, $\mathbf{F}_i(t)$ is the force acting upon the particle at that time and m_i is its mass. The computation of the term $\mathbf{F}_i(t)$ plays a central role in the field of MD. The force is computed from the analytic derivative of the potential energy function V of the system via equation 3.11.

$$\mathbf{F}_i(t) = -\nabla_i V(\mathbf{x}(t)) \tag{3.11}$$

Here $\nabla_i = \begin{bmatrix} \frac{\partial}{\partial x_i} \\ \frac{\partial}{\partial y_i} \\ \frac{\partial}{\partial z_i} \end{bmatrix}$ with x_i,y_i and z_i being the cartesian coordinates of particle i. The gradient of the potential energy function is called the force fied.
The terms in the potential energy function are typically split into bonded and non-bonded terms, as is shown in equation 3.12.

$$V(\mathbf{x}(t)) = V_{bonded}(\mathbf{x}(t)) + V_{non-bonded}(\mathbf{x}(t)) \tag{3.12}$$

$$\begin{aligned} V_{bonded}(\mathbf{r}(t)) = &\sum_{\text{bonds}} \frac{1}{2} k_d (d(t) - d_{eq})^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta(t) - \theta_{eq})^2 \\ &+ \sum_{\substack{\text{proper} \\ \text{dihedrals}}} k_\phi (1 + \cos(n\phi(t) - \gamma)) + \sum_{\substack{\text{improper} \\ \text{dihedrals}}} \frac{1}{2} k_\zeta (\zeta(t) - \zeta_{eq})^2 \end{aligned} \tag{3.13}$$

Bonded terms are computed for sets of atoms connected through covalent bonds. The bonded terms include harmonic potentials for bonds, angles, and improper dihedrals and periodic potentials for proper dihedrals, compare 3.13. Here k_d, k_θ, and k_ζ refer to the force constants of the harmonic potentials for bonds, angles, and improper dihedrals respectively, and tune the strength of the potentials. The terms d_{eq}, θ_{eq}, and ζ_{eq} refer to equilibrium bond-distance, angle, and improper dihedral angle respectively. These parameters tune the locations of the minimums of their respective potentials. Finally, $d(t)$, $\theta(t)$ and $\zeta(t)$ refer to the instantaneous bond distance, angle and improper dihedral angle respectively. The periodic potential for the proper dihedrals contains four terms, k_ϕ, which tunes the strength of the potential, n, which governs the periodicity, i.e. the number of minima, of the potential, γ, the

phase shift and finally, $\phi(t)$, the instantaneous dihedral angle.
Illustrations for the bonded terms are shown in figure 3.1. The bond potential governs stretch vibrations, the angle potential governs bending vibrations. Improper dihedrals apply to out-of-plane vibrations and proper dihedrals to rotations around a bond axis.

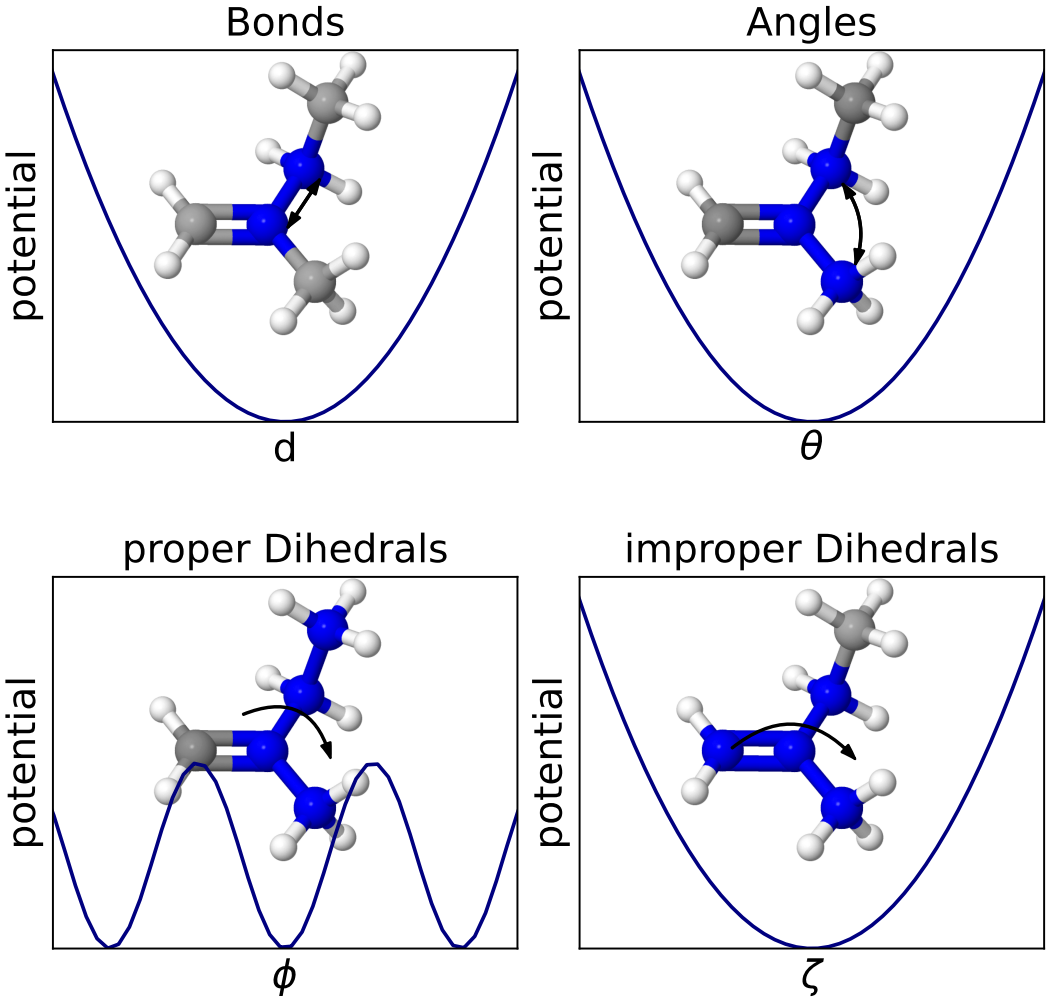

FIGURE 3.1: Illustrations of the bonded terms in the potential energy function shown on 2-Methyl-1-butene.

$$V_{non-bonded}(\mathbf{r}(t)) = \sum_{i=1}^{N}\sum_{j=i+1}^{N} \underbrace{\frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{d_{ij}}}_{\text{Coulomb}} + \sum_{i=1}^{N}\sum_{j=i+1}^{N} \underbrace{4\epsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{d_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{d_{ij}^{6}}\right)}_{\text{Lennard-Jones}} \tag{3.14}$$

Non-bonded terms are computed for all pairs of atoms that are not connected via bonds, angled or dihedrals. In practice, a cutoff radius is chosen for the short-range non-bonded terms like the Lennard-Jones potential. One can also split the electrostatic terms into long-range and short-range terms for the Ewald summation method [67, 68]. The purpose of the cutoff is to mitigate the computational load. The non-bonded terms include the Coulomb potential and the Lennard-Jones potential and are depicted in equation 3.14. The terms are computed for each unique pair of atoms i, j that are not connected via any bonded-term potenial. The distances between

these atoms are denoted as d_{ij}. The Coulomb potential governs the electrostatic interactions between the particles, their respective charges are q_i and q_j and ϵ_0 refers to the vacuum permittivity. The Lennard-Jones potential includes two effects. The repulsive term proportional to d_{ij}^{-12} models the Pauli-repulsion and effectively prevents particles from overlapping. The attractive term proportional to d_{ij}^{-6} models the London dispersion forces, which stem from fluctuating dipoles of the atoms and are a type of van-der-Waals force. The remaining terms in this potential are ϵ_{ij}, which affects the depth of the minimum of the potential, and σ_{ij} which determines the location of the minimum. These parameters carry the index ij because they are computed for each interaction pair, typically using the Lorentz-Berthelot rules, equations 3.15 and 3.16 [69]. The non-bonded potentials are illustrated in figure 3.2.

$$\sigma_{ij} = \left(\sigma_{ii} + \sigma_{jj}\right)/2 \tag{3.15}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{3.16}$$

In equations 3.15 and 3.16 σ_{ii} and ϵ_{ii} refer to the Lennard-Jones interaction parameters of atom i with itself. These values are part of the input of an MD-Simulation, like an atom's charge.

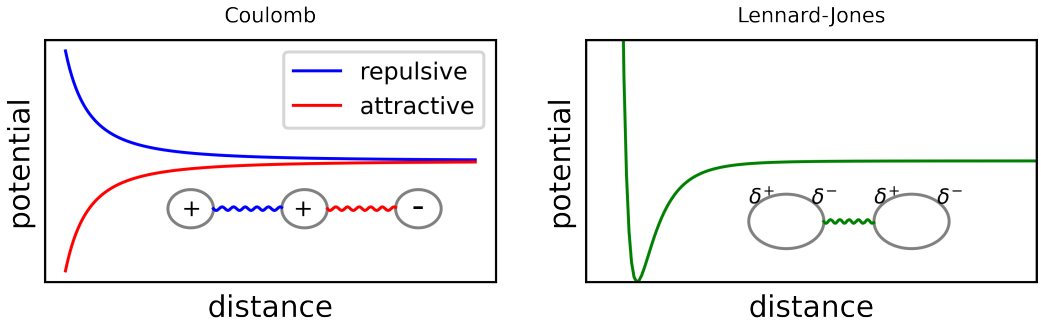

FIGURE 3.2: Illustrations of the non-bonded terms in a force field. The Coulomb interaction on the left is attractive for particles of opposite charge and repulsive for particles of the same charge. The Lennard-Jones interaction models the van-der-Waals forces resulting from fluctuating dipoles and the Pauli repulsion.

Once the forces are computed, one can go on to solve the equations of motion, these are equations 3.17 and 3.18.

$$\dot{\mathbf{x}}(t) = \frac{\partial}{\partial t}\mathbf{x}(t) = \mathbf{v}(t) \tag{3.17}$$

$$\dot{\mathbf{v}}(t) = \frac{\partial}{\partial t}\mathbf{v}(t) = \mathbf{a}(t) \tag{3.18}$$

The solution to this essentially second-order differential equation can be found by writing the positions at time $t + \Delta t$ as a Taylor expansion.

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{3.19}$$

In equation 3.19 Δt refers to the increment by which the system is propagated in time, the timestep. $\mathcal{O}(\Delta t^3)$ refers to the error of the expansion being in the order of

Δt^3. The Verlet-algorithm is derived by adding the positions at time $t - \Delta t$ to the positions at time $t + \Delta t$.

$$\mathbf{x}(t - \Delta t) = \mathbf{x}(t) - \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 - \mathcal{O}(\Delta t^3) \tag{3.20}$$

That is, equation 3.19 + 3.20 and isolating $\mathbf{x}(t + \Delta t)$ yields:

$$\mathbf{x}(t + \Delta t) = 2\mathbf{x}(t) - \mathbf{x}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{3.21}$$

Equation 3.21 implies that the velocities become obsolete to update the position. Should one require the velocities, e.g. to compute instantaneous temperatures, they can be derived from the finite difference of positions, as is shown in equation 3.22.

$$\mathbf{v}(t) = \frac{\mathbf{x}(t + \Delta t) - \mathbf{x}(t - \Delta t)}{2\Delta t} \tag{3.22}$$

Another popular integrator is the Leap-Frog algorithm [70]. This can be seen as computing four half steps in the order velocity, position, position, velocity.

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}(t) + \mathbf{a}\left(t + \frac{\Delta t}{2}\right)\frac{\Delta t}{2} \tag{3.23}$$

$$\mathbf{x}\left(t + \frac{\Delta t}{2}\right) = \mathbf{x}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\frac{\Delta t}{2} \tag{3.24}$$

$$\mathbf{x}\left(t + \Delta t\right) = \mathbf{x}\left(t + \frac{\Delta t}{2}\right) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\frac{\Delta t}{2} \tag{3.25}$$

$$\mathbf{v}\left(t + \Delta t\right) = \mathbf{v}\left(t + \frac{\Delta t}{2}\right) + \mathbf{a}\left(t + \Delta t\right)\frac{\Delta t}{2} \tag{3.26}$$

Equations 3.24 and 3.25 can be combined into single update of the position, similarly equations 3.26 and 3.23 can be combined into a single update of the velocities. This results in velocities being updated at time intervals $t + n\frac{\Delta t}{2}$, where n is an odd integer. The position are updated at time intervals $t + m\frac{\Delta t}{2}$, where m is an even integer. Just like even and odd numbers appearing in an alternating fashion on the number line, the positions and velocity are seemingly updated at alternating points on the timeline, leaping over one another as the title of the algorithm suggests. The advantage of this algorithm is the fact that it is time-reversible and energy-conserving.
Typical simulation boxes contain 10^2-10^5 molecules, one of the largest box simulation box reported contained 1.6 billion atoms and was simulated with a performance of 8.3 ns/day on the Fugaku supercomputer [71]. This is extremely impressive because the computational load required for an MD simulation grows like N^2.
While 10^5 molecules seems like a lot, it is less than a droplet in macroscopic terms. Even a billion, 10^{12}, molecules are nothing compared to macroscopic scales. To overcome this problem one can employ periodic boundary conditions (PBCs). PBCs emulate copies of the simulation box at each of its faces. This simulates an infinite system and prevents interactions with the vacuum. They can be implemented computationally efficiently by making particles that leave the box on one side, reappear at the other end. This results in just a single periodic image that needs to be simulated. The distance in a periodic system can easily be computed from a single box.

$$\Delta w_{ij}^{PBC} = \begin{cases} \Delta w_{ij} - \frac{L_w}{2}, & \text{if } \Delta w_{ij} > \frac{L_w}{2} \\ \Delta w_{ij} + \frac{L_w}{2}, & \text{if } \Delta w_{ij} < \frac{L_w}{2} \text{ , for } w \in \{x, y, z\} \\ \Delta w_{ij}, & \text{otherwise} \end{cases} \tag{3.27}$$

In equation 3.27, w is placeholder for a cartesian coordinate x, y or z and Δw_{ij} is the difference of two cartesian coordinates of particles i and j. L_w refers to the box length along the corresponding coordinate.
The components of an MD simulation that have been discussed so far allow for the simulation at constant particle number, constant volume, and constant energy, the NVE ensmeble. Under laboratory conditions, it is not volume and energy that constant, but rather pressure and temperature, therefore, it is critical to take these quantities into account. Several algorithms have been devised to tackle these problems, they are called barostats and thermostats respectively. The Berendsen barostat and Berendsen thermostat [72] shift the instantaneous pressure and temperature towards their respective target values in a first-order process. This exponential decay towards the target value is very effective in reaching the desired pressure or temperature respectively, however, the fluctuations are not physical and hence it is not the correct ensemble that is being simulated. The Andersen barostat and Andersen thermostat [73] have a similar name but different methods of operating. The Andersen thermostat picks particles with a certain probability and reassigns their velocity to one drawn from the Maxwell-Boltzmann distribution of the target temperature. The Andersen barostat extends the Hamiltonian of the system which affects the resulting equations of motions. Another famous barostat is the Parrinello-Rahman barostat [74, 75] which is similar to the Andersen-barostat while additionally being able to handle a box that changes shape in the process.

3.3 Parametrizing MD Force Fields

This section is based on the review published in [76].
Imagine if you will, a system containing a molecular solvent with monoatomic cations and anions dissolved therein. What parameters need to be determined to simulate this system with classical molecular dynamics?
For the solvent molecules one needs to determine values for the bond lengths, bond angles if it is a non-linear molecule, force constants for the harmonic potentials governing bond lengths and bond angles, Lennard-Jones parameters for all atom types as well as partial charges. The ions are a simpler chemical species and only require the determination of suitable Lennard-Jones parameters.
How can one determine suitable values for all these parameters? Bond-lengths and bond-angles are typically taken from crystal structures, microwave spectra or high level quantum mechanical (QM) calculations in the ground state. For bond-angles, QM calculations or crystal structure data can be used. The vibrations of bonds and angles are modeled as harmonic oscillators. The force constants for bonds and angles can be derived from fitting to QM potential energy surfaces (PES) or from infrared and Raman spectroscopy.
Determining partial charges in molecules is more challenging as they cannot be directly measured. For simple diatomic molecules, like HF, one can determine the partial charges of both atoms from the ratio of dipole moment and bond length. Larger molecules require more sophisticated computational methods. Many methods are

based on fitting the electrostatic potential on a set of grid points. These methods include the MK (Merz-Kollman) method, ChelpG (CHarges from ELectrostatic Potentials using a Grid-based method and RESP (Restrained ElectroStatic Potential). The methods mainly differ in the ways the grid are set up. One drawback of methods based on electrostatic potentials is the fact that deeply embedded atoms are poorly represented. Another method to tackle this problem is the Mulliken charge analysis. This method is based on the density matrix of electronic structure calculations. These charges require little computational effort to compute but depend heavily on the used basis set. Bader charges are obtained via the atoms-in-molecules (AIM) method. Here, the molecule is divided into volumes, the boundaries are drawn at the local minima of the electronic density. All charge within a given volume is then assigned to the corresponding atom. A comparison of methods to determine partial charges was published in [77].
The last set of parameters that need to be determined are the Lennard-Jones parameters σ and ϵ. The first parameter, σ has the dimension of length and determines the location of the minimum of the LJ-potential. One can therefore think of σ like a Van-der-Waals radius. The second parameter, ϵ has the dimension of energy and determines the depth of the minimum of the LJ potential, i.e. the strength of the Lennard-Jones potential. One can think of ϵ as the polarizability of the atoms.
Finding suitable parameters for this potential is challenging because many parameter combinations need to be tested to determine a result. The parameter combinations are judged by their ability to recreate target quantities which are relevant for the system. For systems of metal ions in aqueous solution these target quantities are typically ion-oxygen distances, hydration free energies and coordination numbers. Coordination polyhedra and fluorine content of the solvation shells are also possible target quantities and are discussed in the results section of this work. For pure liquids such as hydrofluoric acid, one can also compute densities, dielectric constant, heat capacity and self diffusion constant to judge the tested parameter combinations.
Naturally, one could simply run an MD-simulation for every parameter combination of interest to determine a suitable one. Alternatively, the target quantity can be fitted to a polynomial as a function of the Lennard-Jones parameters to determine the parameter combination which yields the exact target quantity. The more target quantities are involved the more challenging the task of parameterizing becomes. Typically there exist several parameter combinations to reproduce a given target quantity. The overlap of the sets of suitable parameter combinations shrinks which each new target quantity included. Therefore one might have to prioritize certain quantities to be included into the model.

3.4 Evaluation

3.4.1 Markov Model Theory

Let $\Omega \subset \mathbb{R}^M$ denote the M-dimensional state space of a molecular system. Ω is a finite-dimensional real vector space. $\mathbf{x}(t) \in \Omega$ denotes the state of the system at time t. We assume that $\mathbf{x}(t)$ follows a dynamic that is ergodic and reversible. Note that Ω is not necessarily the $(3N_{atom} - 6)$-dimensional configurational space of the molecular system, where N_{atom} is the number of atoms in the system. It is more often a suitable low-dimensional subspace of the full configurational space.

The slow dynamic modes of this system can be extracted by analyzing the dominant eigenfunctions and eigenvalues of the associated propagator $\mathcal{P}(\tau)$ [78–80]

$$p_{t+\tau}(\mathbf{y}) \quad = \quad \mathcal{P}(\tau)p_t(\mathbf{x}) := \int_\Omega \mathrm{p}(\mathbf{x},\mathbf{y},\tau)p_t(\mathbf{x})d\mathbf{x} \tag{3.28}$$

where $p_t(\mathbf{x})$ denotes a probability density in Ω that evolves in time, and $p(\mathbf{x},\mathbf{y};\tau)$ denotes the transition probability density, i.e. the conditional probability to find the system in $\mathbf{y}\,\mathrm{d}\mathbf{y}$ at time $t+\tau$, given that is has been in state $\mathbf{x}$ a time t. The eigenfunctions $l_i(\mathbf{x})$ and eigenvalues $\lambda_i(\tau)$ are defined by:

$$\mathcal{P}(\tau)l_i(\mathbf{x}) = \lambda_i(\tau)l_i(\mathbf{x}) \tag{3.29}$$

The eigenvalue $\lambda_0(\tau) = 1$ always exists and is the largest by absolute value. The associated eigenfunction $\mathcal{P}(\tau)l_0(\mathbf{x}) = l_0(\mathbf{x}) = \mu(\mathbf{x})$ is the stationary density. Eigenfunctions with eigenvalues close to $\lambda_0(\tau) = 1$ present the slow dynamic modes of the system. For reversible dynamics, the detailed balance condition holds:

$$\mu(\mathbf{x})\mathrm{p}(\mathbf{x},\mathbf{y},\tau) = \mu(\mathbf{y})\mathrm{p}(\mathbf{y},\mathbf{x},\tau) \tag{3.30}$$

In this case, the propagator is self-adjoint with respect to the weighted scalar product.

$$\langle g|f\rangle_{\mu^{-1}} \quad = \quad \int_\Omega g(\mathbf{x})\mu^{-1}(\mathbf{x})f(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{3.31}$$

If equation 3.30 holds, then the matrix representation of the propagator is self-adjoint w.r.t. the weighted scalar product and thus has a real-valued eigenspectrum. Using the self-adjointness one can derive a linear variational principle [34, 81] which allows one to approximate the first m eigenfunctions by linearly expanding them in a set of basis functions $B = (\phi_1, \phi_2, \ldots \phi_m)$

$$\hat{l}_j(\mathbf{x}) = \sum_{i=1}^{m} a_{ij}\phi_i(\mathbf{x}) \tag{3.32}$$

and varying the expansion coefficients a_{ij} to maximize $\langle \hat{l}_j|\mathcal{P}(\tau)\hat{l}_j\rangle_{\mu^{-1}}$ under the constraint that $l_j(\mathbf{x})$ is orthogonal to the other $m-1$ eigenfunctions. This leads to a generalized eigenvalue problem

$$\mathbf{C}(\tau)\mathbf{A} = \mathbf{C}(0)\mathbf{\Lambda}(\tau)\mathbf{A} \tag{3.33}$$

in which $\mathbf{A}$ is a matrix that contains the expansion coefficients of the first m eigenfunctions, and $\mathbf{\Lambda}(\tau)$ is a diagonal matrix that contains the approximation of the first m eigenvalues. Eq. 3.33 is also called a discretization of $\mathcal{P}(\tau)$.
The matrix elements of $\mathbf{C}(\tau)$ are given as $C_{ij}(\tau) = \langle \phi_j|\mathcal{P}(\tau)\phi_i\rangle_{\mu^{-1}}$. Because the state space Ω on which $\mathcal{P}(\tau)$ and the basis functions $\phi_i(\mathbf{x})$ are defined is extremely high-dimensional, the matrix elements cannot be evaluated by standard numerical integration techniques. Instead one uses the fact that the matrix element can be rewritten as a time-lagged correlation of functions $\chi_i(\mathbf{x}) = \phi_i(\mathbf{x})\mu^{-1}(\mathbf{x})$ and $\chi_j(\mathbf{x}) = \phi_j(\mathbf{x})\mu^{-1}(\mathbf{x})$

$$\begin{aligned} C_{ij}(\tau) = \langle \phi_j|\mathcal{P}(\tau)\phi_i\rangle_{\mu^{-1}} \quad &= \quad \int_\Omega \phi_j(\mathbf{y})\mu^{-1}(\mathbf{y}) \int_\Omega p(\mathbf{x},\mathbf{y};\tau)\phi_i(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \\ &= \quad \int_\Omega \int_\Omega \chi_j(\mathbf{y})p(\mathbf{x},\mathbf{y};\tau)\mu(\mathbf{x})\chi_i(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \end{aligned}$$

$$\begin{aligned} &= \mathbb{E}[\chi_i(t), \chi_j(t+\tau)] \\ &= \mathrm{cor}(\chi_i, \chi_j, \tau) \end{aligned} \tag{3.34}$$

and estimates them from time series data $\mathbf{x}(t)$ of the dynamic process

$$C_{ij}(\tau) = \mathrm{cor}(\chi_i, \chi_j, \tau) = \lim_{T\to\infty} \frac{1}{T-\tau} \int_{t=0}^{T-\tau} \chi_i(\mathbf{x}(t))\chi_j(\mathbf{x}(t+\tau))\,\mathrm{d}t \tag{3.35}$$

where T is the length of the time series. The time series $\mathbf{x}(t)$ are usually molecular-dynamics trajectories generated by classical molecular-dynamics simulations.
The matrix elements of $\mathbf{C}(0)$ are given as $C_{ij}(0) = \langle \phi_j \,|\, \phi_i \rangle_{\mu^{-1}}$, and are estimated as correlation functions time lag $\tau = 0$.

$$C_{ij}(0) = \mathrm{cor}(\chi_i, \chi_j, \tau = 0) = \lim_{T\to\infty} \frac{1}{T} \int_{t=0}^{T} \chi_i(\mathbf{x}(t))\chi_j(\mathbf{x}(t))\,\mathrm{d}t \tag{3.36}$$

This yields the following strategy for estimating variational Markov models [34]: (*i*) choose a basis set $\bar{B} = (\chi_1, \chi_2, \ldots \chi_m)$; (*ii*) carry out a simulation of the dynamic process and estimate $\mathbf{C}(\tau)$ and $\mathbf{C}(0)$ via 3.35 and 3.36; (*iii*) solve eq. 3.33 to obtain the expansion coefficients for the eigenfunctions defined in eq. 3.32 and the associated eigenvalues. Matrix representations of the propagator that are estimated using this strategy are called Markov models.
The various Markov models techniques can be classified according to the type of basis set they use for the discretization of $P(\tau)$. In Markov state models (MSMs) [78–80, 82] the state space Ω is discretized into non-overlapping sets $(S_1, S_2, \ldots S_m)$, such that the states cover the entire state space. The basis functions are then the characteristic functions associated with each set.

$$\chi_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S_i \\ 0 & \text{otherwise} \end{cases} \tag{3.37}$$

In MSMs, the elements of the matrix $\mathbf{T}(\tau) = \mathbf{C}^{-1}(0)\mathbf{C}(\tau)$ represent transition probabilities, and $\mathbf{T}(\tau)$ is called a transition matrix.
In core-set Markov models, [83–86] the state space Ω is discretized into non-overlapping sets $(S_1, S_2, \ldots S_m)$, such that the states do not cover the entire state space. The sets are called core sets and need to be located at the cores of the long-lived conformations of the system. The basis functions of a core-set model are related to the committor functions $q_i(\mathbf{x})$ of each core set, where $q_i(\mathbf{x}(t))$ denotes the probability that $\mathbf{x}(t)$ reaches the core set S_i before it enters any of the other core sets.
Markov models obtained by time-lagged independent component analysis (tICA) [87, 88] can also be understood as variational Markov models. Here the state space Ω is the entire configurational space, and the basis functions are the observables $o_i(\mathbf{x})$ that one chooses to include in the tICA model, usually shifted such that the expectation value of the observable is zero.
If the state space can be decomposed into two subspaces $\Omega = \Omega_1 \times \Omega_2$, one can use different discretization strategies for each of the two subspaces, e.g. a MSM on Ω_1 and a tICA Markov model on Ω_2, and combine them into a model for the dynamics in the full state space Ω. Strictly speaking, the Markov models will be constructed on the space that is the direct vector sum of all subspaces, i.e. $\Omega = \bigoplus_i \Omega_i$. Alternatively, one can think of our procedure as using features from solvent and solute before dimensionality reduction with tICA. The resulting tICs can then be decomposed into contributions from solvent or solute for graphical representation. Each

of the two subspaces needs to be a mathematically well-defined finite-dimensional vector space. This requirement is difficult to meet when modeling the dynamics of solvation shells.

3.4.2 Markov Models of Solvation Shells

Consider a simulation box with one solute consisting of N_{solute} atoms and N solvent molecules each of which have $N_{solvent}$ atoms. The full configurational space of this system is $\mathbb{R}^{3N_{solute}} \times \mathbb{R}^{3N_{solvent} \cdot N}$. In current simulations, the number of solvent molecules, N, ranges from 10^2 (*ab-initio* MD simulations) to 10^{12}, which creates an extremely high-dimensional configurational space. Constructing a Markov model in this space is not practical. A chemical intuitive way to reduce the full configurational space is to consider the solute plus its solvation shell and to disregard the surrounding bulk solvent. The solvation shell consists of the $N_{shell}(t)$ solvent molecules that are not further than the first minimum of the solvation shell away from the solute. This number changes over time because solvent molecules diffuse in and out of the solvation shell. That is, the solvation shell is an open system. Because of this exchange "$\mathbb{R}^{3N_{solute}} \times \mathbb{R}^{3N_{shell} \cdot N}$" does not constitute a mathematically well-defined vector space, and any Markov model constructed on these coordinates would have to take the exchange process into account.
Therefore, a different approach is chosen. In a pure solvent, the solvent molecules are chemically equivalent. The dynamics of the solute is influenced by the distribution and possibly the orientation of the solvent molecules in the solvation shell. Hence, the solvent molecule density in the solvation shell is considered a reaction coordinate. At time t, each solvent molecule k in the solvation shell is represented by a single point in the three-dimensional space, $\mathbf{x}_k(t) \in \mathbb{R}^3$. For water, $\mathbf{x}_k(t)$ would be, for example, the position of the oxygen atom. The solvent molecule density $\rho(\mathbf{x}, t)$ parametrically depends on the positions of the $N_{shell}(t)$ solvent molecules in the solvation shell at time t

$$\rho(\mathbf{x}, t) = \rho(\mathbf{x} \,|\, \mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_{N_{shell}(t)}(t)) \quad = \quad \frac{1}{N_{shell}(t)} \sum_{k=1}^{N_{shell}(t)} \delta(\mathbf{x} - \mathbf{x}_k(t)) \quad (3.38)$$

where $\delta(\mathbf{x} - \mathbf{x}_k(t))$ is the Dirac delta-function. By introducing eq. 3.38, we replaced the open system of the solvation shell with a time-dependent function $\rho : \mathbb{R}^3 \times \mathbb{R}_{\geq 0} \to \mathbb{R}^3_{\geq 0}$, which captures changes in the solvation shell due to rearrangements of the current solvent molecules within the shell, as well as changes due to exchange processes with the bulk solvent.
The solvent molecule density is a (time-dependent) function and thus a vector in an infinite-dimensional vector space that evolves in time. To incorporate it into a Markov model, we need to represent the function $\rho(\mathbf{x}, t)$ by a finite-dimensional vector $\mathbf{c}(t) \in \mathbb{R}^C$ that evolves in time. Thus, we expand the solvent molecule density in a set of basis functions $b_i : \mathbb{R}^3 \to \mathbb{R}$ and truncate this expansion at C basis functions.

$$\rho(\mathbf{x}, t) = \sum_i^{\infty} c_i(t) b_i(\mathbf{x}) \quad \approx \quad \sum_i^{C} c_i(t) b_i(\mathbf{x}) \quad (3.39)$$

A Latin letter is used for these basis functions to emphasize the fact that these basis functions are only an auxiliary tool to map $\rho(\mathbf{x}, t)$ on a finite-dimensional vector

space, and that they are not directly linked to the basis functions in the variational principle.
If an orthonormal basis set is used, i.e. if $\langle b_i|b_j\rangle = \delta_{ij}$, the i^{th} expansion coefficient is given as $c_i(t) = \langle b_i|\rho(t)\rangle$. The expansion coefficient can be calculated from the positions of the solvent molecules in the solvation shell as:

$$\begin{aligned} c_i(t) = \langle \rho | b_i \rangle &= \int_{\mathbb{R}^3} \rho(\mathbf{x},t) b_i(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^3} \frac{1}{N_{shell}(t)} \sum_{k=1}^{N_{shell}(t)} \delta(\mathbf{x} - \mathbf{x}_k(t)) b_i(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{N_{shell}(t)} \sum_{k=1}^{N_{shell}(t)} \int_{\mathbb{R}^3} \delta(\mathbf{x} - \mathbf{x}_k(t)) b_i(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{N_{shell}(t)} \sum_{k=1}^{N_{shell}(t)} b_i(\mathbf{x}_k(t)) \end{aligned} \tag{3.40}$$

Thus, the expansion coefficient $c_i(t)$ is obtained by evaluating $b_i(\mathbf{x})$ at the positions of each solvent molecule in the solvation shell and averaging over the number of molecules in the solvation shell.
Together with the basis set, the vector of expansion coefficients
$\mathbf{c}^\top(t) = (c_1(t), c_2(t) \dots c_C(t))$ approximates the state of the solvent molecule density. The state space of the solute and the solvation shell, as approximated by eq. 3.39, is $\Omega = \Omega_{solute} \times \mathbb{R}^C$, where Ω_{solute} denotes the state space of the solute. We can now use standard discretization techniques to construct Markov models in this space.

3.4.3 Basis Sets for the Solvation Shell

The quality of the solvation shell Markov models will depend on the choice of basis functions $b_i(x)$ in eq. 3.39. Each solvent molecule in the solvation shell is described by a set of spherical coordinates (r, θ, ϕ) with the solute at the origin. However, the radius r is not used in either of the two basis sets, which is justified if there is little variation of r within the solvation shell.
Spherical harmonics $Y_l^m(\theta,\phi)$ are functions of degree l and order m defined by

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} e^{im\phi} P_l^m(\cos(\theta)) \tag{3.41}$$

where $P_l^m(\cos(\theta))$ are the associated Legendre functions.
$Y_l^m(\theta,\phi)$ are complex-valued functions. Real linear combinations of the spherical harmonics are given via $Y_{l,\text{real}}^m$

$$Y_{l,\text{real}}^m(\theta,\phi) = \frac{1}{\sqrt{2}} \left(Y_l^{|m|}(\theta,\phi) + \frac{m}{|m|} Y_l^{-|m|}(\theta,\phi) \right) \cdot \sqrt{\frac{m}{|m|}}, \qquad m \neq 0 \tag{3.42}$$

where $\sqrt{\frac{m}{|m|}} = 1$ if $m > 0$, or $\sqrt{\frac{m}{|m|}} = i$ if $m < 0$. For $m = 0$, $Y_l^m(\theta,\phi)$ is already a real-valued function.

3.4.4 Radial Distribution Functions

An extremely useful tool for studying solvation phenomena are radial distribution functions (RDFs). RDFs give insight into the radial structure of the solvation shells

and are also experimentally available through X-ray or neutron diffraction. Not only can solute-solvent distances be extracted from the extremal points of the RDFs, but coordination numbers are also available as well via integration over the respective peaks. The RDF, sometimes denoted g(r), can be extracted from a simulation by computing equation 3.43.

$$g(r) = \frac{H(r)}{dV(r)N_{frames}\rho} \tag{3.43}$$

Where $H(r)$ is the histogram entry at radius r, $dV(r)$ is the bin volume, N_{frames} is the number of frames and ρ denotes the particle density of the entire simulation box. We can thus understand $g(r)$ as the ratio of mean bin particle density and average particle density.

3.4.5 Steinhardt Order Parameters

Steinhardt parameters are shape descriptors. They assign a value to any configuration of particles in space and thus allow one to group structures in a meaningful way [89]. Steinhardt Parameters are based on spherical harmonics $Y_l^m(\theta, \phi)$ and have been used previously to quantify structures of solvation shells [26, 28]. Consider N neighboring particles with positions $\mathbf{x}_k = (r_k, \theta_k, \phi_k)$ around a central particle, which is located in the origin.

$$q_l^m := \frac{1}{N} \sum_{k=1}^{N} Y_l^m(\theta_k, \phi_k) \tag{3.44}$$

$$Q_l := \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |q_l^m|^2} \tag{3.45}$$

Where Q_l is the Steinhardt parameter. The q_l^m are summed over m to generate rotationally invariant parameters. Note that equation 3.44 is equivalent to equation 3.40 for $Y_l^m = b_i$ and $N = N_{shell}$. This implies that Steinhardt parameters can be understood as functions of the expansion coefficients of the solvent particle density, for a special basis set.

3.5 Thermodynamic Integration

The free enthalpy G is related to the partition function Z of the system. They are related via

$$G = -k_B T \ln Z \tag{3.46}$$

$$Z = \int_\Gamma \exp(-\frac{U(\vec{q})}{k_B T}) d\vec{q} \tag{3.47}$$

where k_B is the Boltzmann constant, T the temperature and U the potential. Γ is the phase space and $\vec{q} = (p, q)$ are positions and momenta. The difference of free

enthalpies ΔG_{AB} between states A and B is given by:

$$\Delta G_{BA} = G_B - G_A = -k_B T \ln \frac{Z_B}{Z_A} \tag{3.48}$$

$$\frac{Z_B}{Z_A} = \frac{\int_\Gamma \exp(-\frac{U_B(\vec{q})}{k_B T}) d\vec{q}}{\int_\Gamma \exp(-\frac{U_A(\vec{q})}{k_B T}) d\vec{q}} \tag{3.49}$$

$$= \int_\Gamma \exp\left(-\frac{U_B(\vec{q}) - U_A(\vec{q})}{k_B T}\right) \frac{\exp(-\frac{U_A(\vec{q})}{k_B T})}{Z_A(\vec{q})} d\vec{q} \tag{3.50}$$

$$= \int_\Gamma \exp\left(-\frac{U_B(\vec{q}) - U_A(\vec{q})}{k_B T}\right) p_A(\vec{q}) d\vec{q} \tag{3.51}$$

$$\Delta G_{AB} = -k_B T \ln \left\langle \exp\left(-\frac{U_B(\vec{q}) - U_A(\vec{q})}{k_B T}\right) \right\rangle_A \tag{3.52}$$

Where equation 3.52 implies that the energy difference between two states is given by computing the average of $\exp\left(-\frac{U_B(\vec{q}) - U_A(\vec{q})}{k_B T}\right)$ in the ensemble of A. In practice, one replaces the phase space integrals and time averages of MD-Simulations. This method requires sufficient phase space overlap between the sampled states A and B. Luckily, these states need not correspond to physical entities. Consequently one can construct intermediate states to increase the phase space overlap. A typical choice is a linear interpolation between A and B.

$$U_\lambda = \lambda U_B + (1 - \lambda) U_A \tag{3.53}$$

and the free enthalpy difference becomes

$$\Delta G_{BA} = G(\lambda = 1) - G(\lambda = 0) = \int_0^1 \frac{d}{d\lambda} G(\lambda) d\lambda \tag{3.54}$$

where the term of $\frac{d}{d\lambda} G(\lambda)$ is given by:

$$\frac{d}{d\lambda} G(\lambda) = -k_B T \int_\Gamma p_\lambda(\vec{q}) \left(U_B(\vec{q}) - U_A(\vec{q})\right) d\vec{q} = -k_B T \langle U_B - U_A \rangle_\lambda \tag{3.55}$$

With these preparations, one can determine the free enthalpy difference of two states by choosing a set of values of λ, running an MD simulation for each λ-point, and computing 3.55. Finally 3.54 is computed numerically to obtain ΔG_{BA}.
In this work, the states A and B correspond to a complete absence or presence of Van-der-Waals (VdW) and Coulomb forces respectively. Typically the Coulomb forces are switched off first and the VdW-forces are switched off second. When gradually switching the forces on, the order is reversed. However, it is also possible to switch both forces on or off simultaneously.

3.6 Other Methods

3.6.1 Restricted Electrostatic Potentials

Restricted electrostatic potentials (RESPs) are fits of the electrostatic (Hartree) potential to a set of grid points $\{\mathbf{r}_k\}_{k=1}^N$. The Hartree potential V_H is obtained by solving

the Poisson equation 3.56, for a given electronic density n(**r**).

$$\nabla^2 V_H[n](\mathbf{r}) = -4\pi n\mathbf{r} \tag{3.56}$$

The electronic density can be obtained from density functional theory (DFT) methods, which are not discussed herein. The curious reader is referred to [90].
Given a Hartree Potential and a set of grid points, the RESP is computed by minimizing the left-hand-side of equation 3.57.

$$R_{ESP} = \frac{1}{N}\sum_{k=1}^{N}\left(V_H(\mathbf{r}_k) - V_{RESP}(\mathbf{r}_k)\right)^2 \tag{3.57}$$

Where V_{RESP} is computed from point charges q_a around atom a, using equation 3.58

$$V_{RESP}(\mathbf{r}_k) = \sum_a \frac{q_a}{|\mathbf{R}_a - \mathbf{r}_k|} \tag{3.58}$$

with $|\mathbf{R}_a - \mathbf{r}_k|$ denoting the distance between atomic nucleus a and grid point k.

3.7 Controlling Pressure in MD Simulations

3.7.1 Computing the Instantaneous Pressure

This section is based on information found in [91–93].
Deriving an equation for the instantaneous pressure starts with the Hamiltonian

$$H(p,x) = T(p) + U(x) = \sum_i^N \frac{p_i^2}{2m_i} + \sum_i^N U(x_i) \tag{3.59}$$

where p_i, m_i, and x_i refer to the momentum, mass, and position of particle i respectively. $T(p)$ denotes the kinetic energy and U is the potential.
With the Hamiltonian defined, one can write down the equipartition theorem (EPT), equation 3.60.

$$\frac{1}{2}\left\langle x_i \frac{\partial H}{\partial x_i}\right\rangle = \frac{1}{2}\left\langle p_i \frac{\partial H}{\partial p_i}\right\rangle = \frac{1}{2}k_B T \tag{3.60}$$

Here, $\langle ... \rangle$ denotes the ensemble-average, k_B is Boltzmann's constant and T is the temperature. The theorem states that every degree of freedom contains the same average energy.
Using the EPT, we can connect the average kintetic energy $\langle T(p)\rangle$, with the virial $\sum_i^N x_i F$ via:

$$\langle T(p)\rangle = \left\langle \sum_i^N \frac{p_i^2}{2m_i} \right\rangle \tag{3.61}$$

$$= \left\langle \frac{1}{2}\sum_i^N p_i \frac{\partial H}{\partial p_i} \right\rangle \tag{3.62}$$

$$\overset{\text{EPT}}{=} \left\langle \frac{1}{2}\sum_i^N x_i \frac{\partial H}{\partial x_i} \right\rangle \tag{3.63}$$

$$= \left\langle \frac{1}{2}\sum_i^N x_i \frac{\partial U}{\partial x_i} \right\rangle \tag{3.64}$$

$$= \left\langle \frac{1}{2}\sum_i^N x_i F \right\rangle \tag{3.65}$$

$$\tag{3.66}$$

With this preparation complete, one can take a closer look at the virial. Typically it is split into inner and outer virial.

$$\sum_i^N x_i F = \sum_i^N x_i F_{inner} + \sum_i^N x_i F_{outer} = \Xi + W \tag{3.67}$$

The outer force is exerted by wall elements dA along their normal **n** onto nearby particles.

$$F_{outer} = P\mathbf{n}dA \tag{3.68}$$

Here, P is the pressure. If one assumes that this force is exerted onto particles close to the wall, we can replace the position of particle i, x_i with the position of the wall element **x**.

$$\sum_i^N x_i F_{outer} = \sum_i^N x_i P\mathbf{n}dA = \mathbf{x}P\mathbf{n}dA \tag{3.69}$$

Finally, one can integrate over the entire surface ∂V and replace the surface integral by a volume integral using the divergence theorem to solve for the final expression.

$$\left\langle \sum_i^N x_i F_{outer} \right\rangle = P\int_{\partial V} \mathbf{x}\mathbf{n}dA = P\int_V \mathrm{x}dV = 3PV \tag{3.70}$$

Using the previous preparation, this result can be used to derive the Clausius virial theorem:

$$3PV = 2\langle T(p)\rangle - \sum_i^N \langle x_i F_{inner}\rangle \tag{3.71}$$

This allows one compute the instantaneous pressure P from the volume of the system, its average kinetic energy $\langle T(p)\rangle$, and inner virial $\sum_i^N x_i F_{inner}$. The inner virial depends on the particle positions as well as the forces between them, making it a dependent on the forcefield that is used.

3.7.2 Barostat Algorithms

The Berendsen Barostat

This section describes the Beredensen barostat originally published in [72]. The idea is similar to the thermostat algorithm of the same name, i.e. rescale the box as needed to reach the desired pressure via a first order process.

$$\frac{dP}{dt} = \frac{P_{Target} - P}{\tau} \tag{3.72}$$

Where P_{Target} is the target pressure, P is the instantaneous pressure and τ is the time constant. The equations of motion are extended.

$$\frac{dx}{dt} = v + \alpha x \tag{3.73}$$

$$\frac{dV}{dt} = 3\alpha V \tag{3.74}$$

The value of α is determined by combining equation 3.72 and the definition of the isothermal compressibility β, equation 3.75.

$$\beta = -\frac{1}{V}\frac{\partial V}{\partial P} \tag{3.75}$$

Transforming the derivative, rearranging and inserting equation 3.74, one arrives at:

$$\frac{dP}{dt} = -\frac{1}{\beta V}\frac{dV}{dt} = -\frac{3\alpha}{\beta} \tag{3.76}$$

One can combine this result with equation 3.72 to arrive at the final expression.

$$\alpha = -\frac{\beta}{3\tau}\left(P_{Target} - P\right) \tag{3.77}$$

This result indicates, that the coupling strength of the Berendsen barostat depends in the ration of isothermal compressibility β and time constant τ, both of which are input parameters for an MD-simulation.
The Berendsen barostat yields the correct pressure but not provenly the correct ensemble. It is therefore useful to equilibrate the system towards target pressure before using a different method for the production runs.

The Andersen Barostat

The Andersen thermostat, first mentioned in [73], works by mimicking a piston acting on the system and subsequently deriving new equations of motion from the extended Hamiltonian.
First, the positions x and momenta p get scaled by $\sqrt[3]{V}$.

$$\rho := \frac{x}{\sqrt[3]{V}} \tag{3.78}$$

$$\pi := \frac{p}{\sqrt[3]{V}} \tag{3.79}$$

With the scaled coordinates one can write down the extended Hamiltonian. The new terms correspond to the dynamics of the volume.

$$H(\rho, \pi, V, \dot{V}) = \frac{m_i}{2} V^{\frac{2}{3}} \sum_i^N \pi_i^2 + \sum_i^N U\left(\sqrt[3]{V}\rho_i\right) + \frac{M}{2}\dot{V}^2 + P_{Target}V \tag{3.80}$$

Where M refers to the mass of the piston and $\dot{V}$ is the piston velocity.
Deriving the equations of motion from the extended Hamiltonian is done by evaluating the following expressions:

$$\frac{d\rho}{dt} = \frac{dH}{d\pi} \tag{3.81}$$

$$\frac{d\pi}{dt} = -\frac{dH}{d\rho} \tag{3.82}$$

$$\frac{dV}{dt} = \frac{dH}{d\dot{V}} \tag{3.83}$$

$$\frac{d\dot{V}}{dt} = -\frac{dH}{dV} \tag{3.84}$$

The Andersen Barostat on its own yields the isobaric-isoenthalpic ensemble (NPH). In order to simulate the NPT ensemble, it needs to be used in conjunction with a thermostat algorithm. The Andersen Barostat also serves as the basis for the Parrinello-Rahman and Martyna-Tuckerman-Klein barostats.

The Parrinello-Rahman Barostat

The Parrinello-Rahman barostat was first mentioned in [74, 75]. It can be understood as an improvement to the Andersen barostat which allows the simulation box to change shape. Let $\mathbf{a}, \mathbf{b}, \mathbf{c}$ be the box vectors, then

$$V = \det(\mathbf{H}) \tag{3.85}$$

with H = {$\mathbf{a}, \mathbf{b}, \mathbf{c}$ }. The scaled coordinates are

$$r = \chi\mathbf{a} + \eta\mathbf{b} + \zeta\mathbf{c} = \mathbf{H}s \tag{3.86}$$

where $s = (\chi, \eta, \zeta) \in [0, 1]$. The extended Hamiltonian now becomes

$$H(s, \dot{s}, \mathbf{H}, \dot{\mathbf{H}}) = \frac{m}{2}\sum_i^N \dot{s}_i^T \mathbf{H}^T \mathbf{H}\dot{s}_i + \sum_i^N U(\mathbf{H}s_i) + \frac{M}{2}\dot{\mathbf{H}}^T\dot{\mathbf{H}} + P_{Target}V \tag{3.87}$$

where M is again the mass of the piston. In GROMACS, this parameter is not given as an input but is determined from other quantities.

$$M^{-1} = \frac{4\pi^2\beta}{3\tau^2 L} \tag{3.88}$$

Here, L is the largest box matrix element. Since the piston mass determines the coupling strength, it is important to note that for the Parrinello-Rahman barostat, the coupling strength depends on β/τ^2. This stands in contrast to the Berendsen barostat where the coupling strength is determined by β/τ.

3.8 Controlling Temperature in MD Simulations

This section is based on references [94–96]. The instantaneous temperature $T(t)$ is directly related to the instantaneous kinetic energy $K(t)$ via equation 3.89

$$\langle K(t) \rangle = \frac{1}{2} N_{DOF} k_B T(t) \tag{3.89}$$

where $\langle K(t) \rangle$ denotes the ensemble average, N_{DOF} is the number of degrees of freedom and k_B is the Boltzmann constant. Therefore, controlling the kinetic energy of the system is equivalent to controlling the temperature.
Many algorithms exist to keep the temperature in an MD simulation almost constant, such that the correct stationary distribution is sampled. Here, I will only present the stochastic velocity-rescaling thermostat by Bussi, because it was used in all simulations.
Velocity-rescaling thermostat revolve around a scaling factor α which is applied to velocties $\mathbf{v}_i$. This rescaling occurs at a given frequency of timesteps, e.g. every time step. The scaling factor is computed via

$$\alpha = \sqrt{\frac{K_{Target}}{K(t)}} \tag{3.90}$$

where K_{Target} is the target kinetic energy. This target value is determined by propagating the total kinetic energy alongside the MD-simulation with an auxiliary continuous stochastic dynamics, that is

$$dK_{Target} = \left(K_{Target} - K(t)\right) \frac{dt}{\tau_T} + 2\sqrt{\frac{K_{Target} K(t)}{N_{DOF} \tau_T}} dW(t) \tag{3.91}$$

where τ_T is the time constant of the thermostat and $dW(t)$ is a Wiener process. In practice, a Wiener process is a random number drawn from a gaussian distribution. If this thermostat is used in conjunction with an Andersen barostat or Parrinello-Rahman Barostat, one simply needs to extend the kinetic energy term to contain the velocity of the piston η. That is

$$K^* = K + \frac{M}{2} \eta^2 \tag{3.92}$$

with M being the mass of the piston as described in the previous section.

Chapter 4

Methods

4.1 Markov Models of Solvation Shells

4.1.1 Simulations of $[Al(H_2O)_6]^{3+}$

The system was set up by creating a cubic box with side lengths of 3.14 nm. After filling the box with 1024 TIP4P/2005 water molecules[63], four water molecules were replaced by one Al^{3+}-ion and three Cl^{-}-ion. The Lennard-Jones parameters for the aluminium-ion were taken from [97] and the parameters for chloride were taken from the Amber force field [98]. The energy was minimized using the steepest descent method up to a maximum force of 100.0 kJ/mol/nm. Following the minimization, an NPT-run was performed at T = 300 K (velocity-rescale thermostat [95], $\tau_T = 0.1$ ps) and p = 1.0125 bar (Berendsen barostat [72], $\tau_p = 2.0$ ps), using the leap-frog integrator [99]. The simulation ran for 50 ns with an integration timestep of 2 fs, outputs were saved every 10 fs, yielding $5 \cdot 10^6$ frames per trajectory. Bonds were constrained using the LINCS algorithm [100].
All simulations (minimization and NPT-run) employed the PME-scheme [101], periodic boundary conditions in all directions, and a cutoff of 1 nm for short-range Coulomb and Van-der-Waals interactions. Lorentz-Berthelot [69] rules were employed to compute interaction Lennard-Jones parameters. Simulations were performed in GROMACS 2019.4 [102], using the AMBER03 [98] force field. All further calculations were performed in Python 2.7 with the help of NumPy [103], SciPy [104], PyEMMA [105], and Matplotlib [106].

4.1.2 Determination of Torsion Angles in $[Al(H_2O)_6]^{3+}$

Figure 4.1 illustrates how the torsion angle of each ligand was determined. The nuclei of the hexa-aqua-aluminium-complex are represented as points of different colors, the thick blue line connects the four points that were used to calculate the dihedral angle. The black lines visualize the octahedral complex, while grey lines are used to represent O-H bonds.
For each torsion angle, it was ensured that the two oxygen atoms did not lie on the same axis.
This procedure relies on inter-atomic distances only and is thus free from possible errors that could stem from the fitting procedure.

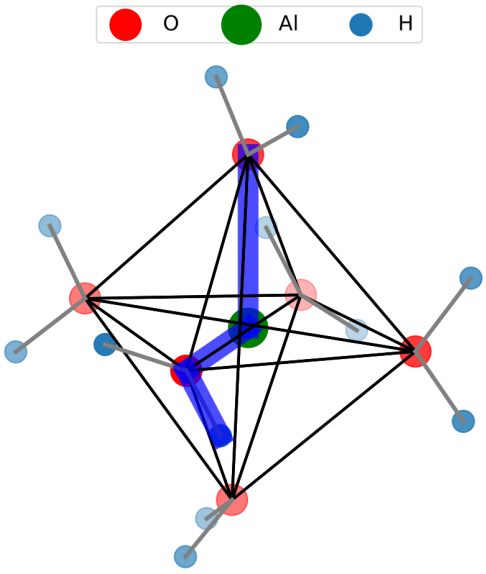

FIGURE 4.1: Ball and stick representation of the hexa-aqua-aluminium-complex with 14 additional black lines to emphasize the octahedral geometry. The thick blue line is representative of the quartet of atoms used to compute the torsion angle of a given hydrogen atom.

4.1.3 Combining Markov Models

Combined Markov models, denoted by $\bigoplus$ or $\bigotimes$ are generated by combining their basis functions into a joint basis. Consider two vector spaces U with basis set $(u_1, u_2, ..., u_N)$ and V with basis set $(v_1, v_2, ..., v_M)$, with $M, N \in \mathbb{N}$. Adding them to construct $S := U \bigoplus V$ yields the following basis $(u_1, u_2, ..., u_N, v_1, v_2, ..., v_M)$. Alternatively they can be multiplied yielding $P := U \bigotimes V$ with the basis set $(p_1, p_2, ..., p_{N\times M})$ where $p_{i+(j-1)N} = u_i \times v_j, i = 1...N, j = 1...M$.
The additive models were constructed by adding the basis functions as described above. For the multiplicative models, projections onto their slowest processes were used as basis functions for practical reasons.

4.1.4 State Space Discretizations of $[Al(H_2O)_6]^{3+}$

To study the effect of the surrounding medium on the quality of the resulting Markov models, several discretization strategies were employed: (i) 36 bins for the interval $[-\pi, \pi]$ for the torsion angles of individual ligands. (ii) A combination of 4 bins for one ligand and 2 bins for its four neighbours, yielding a total of $4 \times 2^4 = 64$ crisp states. (iii) 121 spherical harmonics to approximate the angular distribution of all 12 hydrogen atoms. (iv) 121 spherical harmonics to approximate the distribution of oxygen atoms in the first solvation shell. (v) Combinations of strategies (ii) and (iii) with (iv) in forms of vector space products or vector space sums.

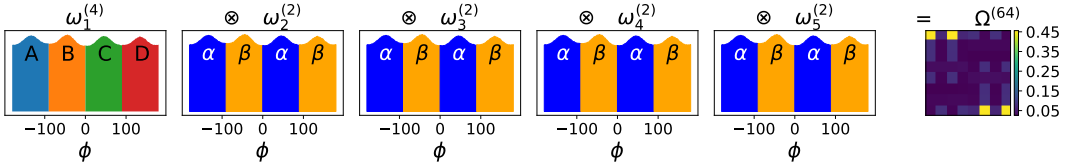

FIGURE 4.2: Discretization (ii), using four states for one ligand and two states for its four neighbouring ligands. The first five panels depict the distributions of torsion angles discretized into either four or two states. The last panel is the first left eigenvector of the resulting MSM, reshaped into a matrix for illustration purposes.

Figure 4.2 further illustrates strategy (ii). It yields one trajectory per ligand and includes all four configurations of the ligand, while at the same time accounting for its four neighbours.

4.2 The Parameterization of HF

4.2.1 Simulations of Hydrofluoric Acid and Pure HF

Simulations were prepared in a cubic box with edge lengths of 3.1415 nm. The necessary amount of HF-molecules to reach target concentrations was added and afterwards, the box was filled with TIP4P/2005 water [63]. All simulations were carried out using the GROMACS [102] simulation package with the AMBER03 [98] force field, optimized for the TIP4P/2005 water model [63]. Before all production runs, the system was relaxed with a steepest descent method using a step size of 0.0001 until a maximum force of 100 kJ/mol/nm was reached or 100000 steps were performed. This is followed by an NVT-equilibration for 100 ps, using 2 fs timestep at 300 K using the velocity-rescale thermostat [95] (τ_T=0.1 ps) and the leap-frog integrator. Initial velocities were drawn from a Maxwell-Boltzmann distribution. Periodic boundary conditions were applied in all three dimensions and electrostatic interactions were treated using the particle mesh Ewald method (PME) [101]. The PME-order was set to 4 and the grid spacing in the Fourier space was set to 0.16. Short-range interactions, i.e. the short-range part of the Coulomb interaction and the Lennard-Jones interaction were dealt with using a distance cutoff of 1 nm. The Verlet cutoff scheme was employed along with the grid neighbor search type. The Lorentz-Berthelot [69] rules were applied for Lennard-Jones interactions. Hydrogen-bonds were constrained using the LINCS algorithm with an order of 4 and 1 iteration.
Production runs were carried out in the NPT ensemble for 1.1 ns, where the first 100 ps are regarded as equilibration and not considered in the evaluation. These runs used a timestep of 2 fs, velocity-rescale thermostat (τ_T=0.1 ps) at 300 K, Berendsen barostat [72] (τ_P=2 ps) at 1.0125 bar. All other parameters were inherited from the NVT-run.
All following simulations for the parameterization of HF were run with the above protocol unless stated otherwise.
An overview of the simulation boxes is shown in table 4.1.
Several HF-models were tested. Simulations were performed for each combination of the following three parameters

$\epsilon \in \{0.2, 0.5, 0.525, 0.55, 0.575, 0.6, 0.625, 0.65, 0.675, 0.7, 1.0, 1.4\}$ in kcal/mol,

TABLE 4.1: Simulation boxes of hydrofluoric acid.

Box Number	HF	H_2O	c(HF)/mol/l	% HF
1	37	984	1.99	3.62
2	118	881	6.35	11.81
3	373	637	20.08	36.93
4	500	541	26.91	48.03
5	800	322	43.06	71.30
6	1180	144	63.52	89.12
7	2611	0	140.56	100

$\sigma \in \{0.25, 0.26, 0.2625, 0.265, 0.2675, 0.27,$
$0.2725, 0.275, 0.2725, 0.28, 0.2825, 0.285, 0.2875, 0.29, 0.30\}$ in nm,

q $\in \{0.441078, 0.441974, 0.4223, 0.742186\}$ in e.

This adds up to 720 parameter combinations. The rationale behind the chosen charges is shown in table 4.2. The densities taken from Eurofluor [107] were given for dif-

TABLE 4.2: Comments on the partial charges.

value / e	comment
0.441078	CHELPG, CCSD(T) aug-cc-pVQZ for geometry optimzation
0.441974	ESP Kollmann-Mehthod, PBE0, def2-TZVP
0.4223	dipole moment/bond length
0.742186	AIM, PBE0 def2-QZVP, Pop.grid: 0.04 Bohr

ferent concentrations than the ones simulated. Therefore, the literature values were used to fit a polynomial of degree five to interpolate the densities at the simulated concentrations. The sum of mean squared errors towards the literature values, Δ_ρ was computed via:

$$\Delta_\rho = \sum_{i=1}^{N} \frac{1}{N} \left(\rho_{lit}(\omega_i) - \rho_{sim}(\omega_i)\right)^2$$
$$\omega_i \in \{3.62, 11.81, 36.93, 48.03, 71.30, 89.12\}(\%HF)$$

Where ρ_{sim} is the simulated density and ω describes the fraction of HF. Parameter sets with $\Delta_\rho < 1000\,\mathrm{kg}^2/\mathrm{m}^6$ were considered successful. The peaks of the radial distribution functions were extracted for the box of pure HF, using the SciPy function find_peak().

4.2.2 Determination of the Solvation Free Energy of NaF in HF

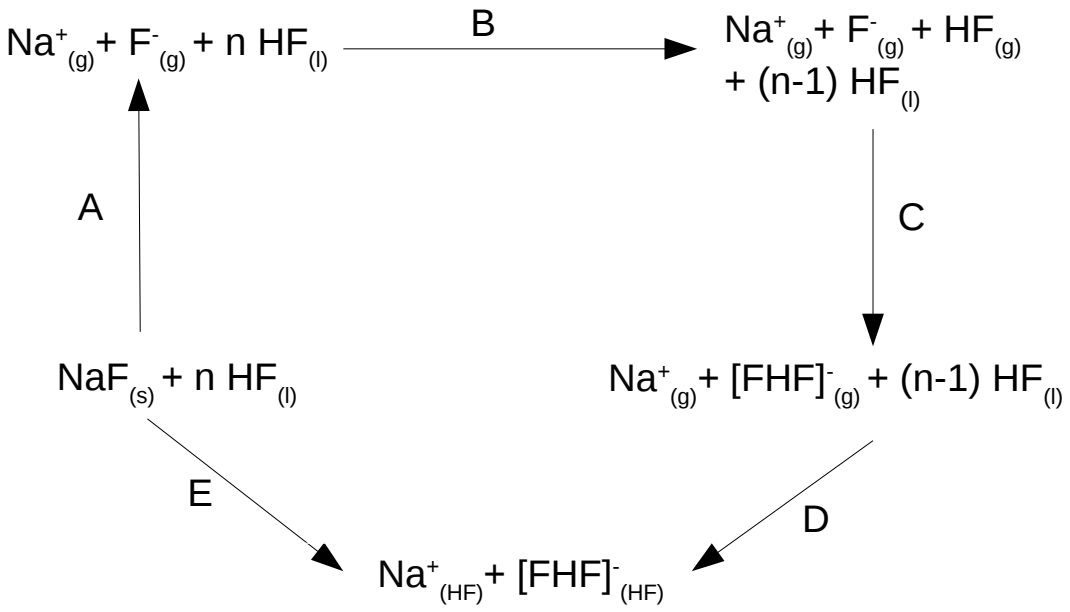

FIGURE 4.3: Thermodynamic cycle set up to evaluate the enthalpy of solvation of NaF in HF. A: Lattice enthalpy of NaF. B: Evaporation enthalpy of HF. C: Enthalpy of formation of bifluoride. D: Solvation enthalpy of Na^+ and $[FHF]^-$ in HF. E: Calorimetrically determined enthalpy of solvation of NaF in HF.

Figure 4.3 illustrates the thermodynamic cycle that was set up to determine the heat of solvation of NaF in HF. The values of A, B, C, and E were taken from literature, and D was determined in this work. The corresponding equation of the thermodynamic cycle is:

$$A + B + C + D = E \tag{4.1}$$

An overview of all important enthalpies at standard conditions is given below in table 4.3. The ΔG referring to step D in figure 4.3 was determined by computing ΔH

TABLE 4.3: Meaning and values of symbols occuring in figure 4.3.

Symbol	Meaning	Value/[kcal$\times mol^{-1}$]	Source
A	$\Delta H^0_{Lattice}(NaF)$	222.19	[108]
B	$\Delta H^0_{g\longrightarrow l}(HF)$	-5.24	[109]
C	$\Delta H^0([FHF^-])$	-45.4	[110]
D	$\Delta H^0_{solv}(Na^+ + [FHF^-])$	determined from the other quantities	This work
E	$\Delta H^0_{solv}(NaF)$	-15.15	[109]

for the corresponding reaction and converting the thermodynamic quantities using the Gibbs-Helmholtz-relation. The change in entropy ΔS was estimated by J. Anders using the quantum mechanical rigid-rotor and harmonic-oscillator approximation

(RRHO) [111]. The combined solvation entropy of Na^+ and $[HFH]^-$ was estimated to be 0.0818 kJ/mol/K.
From equation (1) I estimated the target value for $D = -186.7$ kcal/mol $= -781.2$ kJ/mol. We, therefore, arrive at a solvation free enthalpy of $\Delta G = -757.9$ kJ/mol at 284 K.

4.2.3 Modeling Bifluoride for MD Simulations

The model of the bifluoride ion $[FHF]^-$ requires charges, F-H bond distances, F-H bond potential, and F-H-F bond angle potential. These quantities were determined using the ORCA software with the def2-TZVP basis set, DFT-grid:m4 method, and PBE0 functional with D3BJ dispersion correction. Partial charges were computed by J. Anders using the ESP-Kolman method and AIM of systems in vacuum. The molecule was designed to have a net negative charge, i.e. $q_H + 2q_F = -1$ e. F-H bond distances, d_{F-H}, were taken from structure optimizations. Bonde length force-constants, k_{F-H}, and bond angle force-constants, k_{F-H-F}, were computed by fitting a quadratic function around the minimum of a potential energy surface for the molecule. The results are summarized in table 4.4.

TABLE 4.4: Parameters of the bifluoride ion and the methods used to obtain them.

Quantity	Value	Method
d_{F-H}	0.1142526 nm	Geometry optimization
q_F	-0.738084 e	Kollmann method
k_{F-H}	25264.53 kcal/mol/nm^2	quadratic fit to PES
k_{F-H-F}	30.4373 kcal/mol/rad^2	quadratic fit to PES

4.2.4 Thermodynamic Integration of NaF in HF

For these simulations, one Na^+ and one biflouride FHF^- were placed in a cubic box with an edge length of 3.145 nm and solvated with 2610 HF molecules. All other box- and simulation parameters for minimization and NVT-equilibration are identical to the previous section, except temperature which was adjusted to 284 K to sample the liquid state. Production runs were performed in the NPT ensemble using a leap-frog integrator. Simulations ran for 1.1 ns in total, where the first 100 ps were treated as equilibration time and discarded before the analysis. Van der Waals (VdW) and Coulomb interactions for the ions were varied simultaneously. That is

$$V^{ions}_{couple}(\mathbf{x}(t);\lambda_{couple}) = V^{ions}_{bonded}(\mathbf{x}(t)) + \lambda_{couple} V^{ions}_{non-bonded}(\mathbf{x}(t)) \tag{4.2}$$

$$V^{ions}_{decouple}(\mathbf{x}(t);\lambda_{decouple}) = V^{ions}_{bonded}(\mathbf{x}(t)) + \lambda_{decouple} V^{ions}_{non-bonded}(\mathbf{x}(t)) \tag{4.3}$$

where $V^{ions}(\mathbf{x}(t);\lambda)$ refers to the parts of the potential acting on the ions only and $\mathbf{x}(t)$ are the atomic positions at time t.
Hysteresis was monitored and λ-points were chosen such that both results were in agreement with each other. Both results refer to one run where the interactions are slowly switched on (coupling) and another where they are slowly switched off (de-coupling). The λ_{couple}-points for the coupling runs were : {0.0, 0.0005, 0.001, 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}. Equally many points where chosen for the decoupling runs with

$\lambda_{decouple} = 1 - \lambda_{couple}$
Softcore potentials were employed to improve convergence, softcore potential parameters were sc-powr=1, sc-sigma=0.3 and sc-alpha=1.0, as suggested by [112].

4.2.5 Simulating NMR Spectra of NaF in HF

The simulations were run for all concentrations mentioned in table 4.1, the HF-model was an older version with $\sigma_F = 0.332840$ nm, $\epsilon_F = 1.082496$ kcal/mol and $q_F = -q_H = -0.64942$ e. Additionally, one sodium ion and one fluoride ion were added. The Lennard-Jones parameters of the sodium ion were taken from the AMBER force field. The setup was otherwise identical to the setup used to determine the radial distribution functions and densities.
From the resulting NPT-runs, 100 structures of the first solvation shell of the sodium ion were extraced. That is, for each coordination number CN$\in \{4,5,6,7,8\}$ and each fluoride coordination number $CN_F \in \{0,1,...,CN\}$ three randomly chosen examples. Based on these structures, J. Anders computed the chemical shifts for each atomic species. The chemical shifts were then used to simulate spectra, according to the occurrences of the differently constituted solvation shells.

4.2.6 Measuring Spectra of NaF in HF

NMR spectra were recorded by Patrick Pröhm on a JEOL 400 MHz ECS spectrometer. All reported chemical shifts are referenced to the Ξ values given in IUPAC recommendations of 2008 using the ^{2}H signal of the deuterated solvent as internal reference [113]. Samples were sealed in 3.2 mm PFA tubes. For the measurements, the PFA tubes were put into standard NMR tubes containing acetone-d6 for external locking. The measuring frequencies are : ^{1}H 400MHz, ^{19}F 376MHz, ^{23}Na 106MHz. The samples were measured at 21.7 °C and were constituted as follows: 28 mg NaF in 400 mg aHF, 29 mg NaF in 400 mg H2O, and NaF saturated in 10% HF.

4.2.7 Computing the Electrostatic Potential for HF

The RESP of HF was computed using CP2K. The charge is set to 0, multiplicity to 1. The Goedecker-Teter-Hutter pseudo-potentials were used for core electrons and the MOLOPT basis set was employed for valence electrons. H- and F-atoms used the DZVP-MOLOPT-SR-GTH basis set. H-atoms used the GTH-PBE-q1 pseudopotential, whereas the GTH-PBE-q7 pseudopotential was employed for the F-atoms. Four multigrids were used, with a cutoff of 1000 Ry and a relative cutoff of 60 Ry. Electronic structure calculations were performed using the quickstep code, with the gaussians and plane waves scheme. The default threshold was set to 1^{-10} Hartree. The maximum number of iterations for calculating the LUMO energies with the orbital transform eigensolver was set to 10000, The target accuracy for the SCF convergence was set to 10^{-8} Hartree. The atomic guess was set as the initial guess for the SCF densities, with a maximum of 60 inner SCF cycles. The threshold for the outer SCF cycle was also 10^{-8} Hartree, with a maximum of six outer SCF cycles. The PBE functional was employed together with Grimme's D3 correction. The unit cell was cubic with an edge length of 1 nm.

4.2.8 Fitting a 3-Site Model to the Electrostatic Potential of HF

Let the 3-site model of HF consist of two atomic sites for the H- and F-atoms with charge q as well as a virtual site X with charge $-2q$, located along the H-F bond-axis at distance R_{FX} away from the F-atom. The values of q and R_{FX} were determined by fitting the Coulomb-potential V_C for a test charge of -1 e to the Hartree-potential V_H obtained from quantum mechanical calculations at the PBE level. The parameters were chosen to minimize the function

$$\Delta = \delta_{ESP} + \delta_Z \tag{4.4}$$

with

$$\delta_{ESP} = \sum_k \left(V_H(k) - V_C(k)\right)^2 \tag{4.5}$$

and

$$\delta_Z = \sum_k \left(\frac{d}{dZ}V_H(k) - \frac{d}{dZ}V_C(k)\right)^2 \tag{4.6}$$

where the sum iterates over all points of the 3-dimensional grid that fulfill

$$1.4R_{VdW} \leq r(k) \leq 2R_{VdW} \tag{4.7}$$

where $r(k)$ is the distance of the kth gird-point towards an atom. The Van-der-Waals radii were set to 147 pm and 120 pm for F and H respectively.
The parameter R_{FX} was varied linearly from 0 Å to 0.5 Å in 100 steps, excluding the last point. The charge q was varied from 0.4 e to 1 e in 100 steps, excluding the last point.
After this initial fit to determine R_{FX} and q, MD simulations were run for several parameter combinations of σ_F and ϵ_F.
With
$\sigma_F \in \{0.125, 0.15, 0.175, 0.2, 0.25, 0.25, 0.26, 0.2625, 0.265, 0.2675, 0.27,$
$0.2725, 0.275, 0.2725, 0.28, 0.2825, 0.285, 0.2875, 0.29, 0.30\}$ in nm

$\epsilon_F \in \{0.1, 0.125, 0.15, 0.175, 0.2, 0.3, 0.4, 0.5, 0.525, 0.55,$
$0.575, 0.6, 0.625, 0.65, 0.675, 0.7, 1.0, 1.1, 1.2, 1.3, 1.4\}$ in kcal/mol.

4.2.9 Validating the 3-Site model by Orabi and Faraldo-Gomez [39]

The validation was carried out with 500 HF molecules using the topology described in [39]. The system was generated by placing 500 HF molecules into an empty, cubic box with an edge length of 2.56 nm corresponding to a density of 989.43 g/cm^3 for pure HF.
In all simulations, the bond length was constrained to the gas phase value of HF, i.e. 0.917 Å. Furthermore, the timestep was set to 1 fs for all runs. Electrostatics were treated using the particle mesh Ewald (PME) method, all real-space cutoffs were set to 1.0 nm. The PME-order was set to 6 with a Fourier-spacing of 0.1 nm. Constraints were enforced using the LINCS algorithm with LINCS-order set to 4 and LINCS-iter set to 2. The LINCS-warnangle was set to 30°.
After setting up the system, a steepest descent minimization was performed using an energy step size of 0.001 nm, a tolerance of 0.1 kJ/mol/nm, and a maximum of 100000 steps.
Periodic boundary conditions were employed for all dimensions in the following

runs. The minimized system was subjected to a 100 ps NVT-equilibration using the velocity-rescale thermostat ($\tau_T = 0.1$ ps) and a temperature of 296 K. This combination of thermostat, target temperature, and τ_T was employed for all runs hereafter. Initial velocities were drawn from the Maxwell-Boltzmann distribution for 296 K with the random seed set to 666420. This tempering was followed by a 0.3 ns NPT-equilibration using the Berendsen barostat ($\tau_P = 0.1$ ps) set to a target pressure of 1.199688 bar. The compressibility was set set to 4.5×10^{-5} bar^{-1}.
The final production run employed the Parrinello-Rahman barostat ($\tau_P = 0.1$ ps) set to 1.199688 bar for 10 ns. The other simulation parameters were inherited from the NPT equilibration.
Density, isothermal compressibility, and dielectric constant were computed directly from the trajectory using the python package MDtraj. Radial distribution functions were computed using the GROMACS function rdf. The diffusion constant was found by fitting the mean squared displacement, computed using the GROMACS functions msd, to a linear function. The first 10 ps and the final 20 ps were omitted from the fit, to ensure fitting the diffusive regime and to avoid the ballistic and confined regimes respectively.
The boxes of aqueous HF were computed using a similar protocol as mentioned above. The total number of solvent particles was kept fixed at 500. The water-model used was TIP4P/2005 [63]. Densities and diffusion constants were extracted as mentioned above. Heat capacities were computed using the GROMACS function energy.

4.3 Simulations of Metal Ions in Fluorous Environments

4.3.1 Parameter Scans of M^{4+}

The systems are simulated using classical MD by the GROMACS package, using the AMBER force field. The water model is TIP4P/2005 [63] and the HF-Model is a 3-site model published by Orabi et al. in 2020 [39]. The system was generated by placing 505 water molecules into an empty, cubic box with an edge length of 2.56 nm corresponding to a density of 998.2 g/cm^3 for pure HF. In the next step, five molecules were replaced by one tetravalent cation and four F-anions. The mass of the cation was set to 91.224 amu, the value of zirconium. The F-anions inherited the same LJ parameters as the bound F-atom in HF.
In all simulations, the bond length of HF was constrained to the gas phase value of 0.917 Å. Furthermore, the timestep was set to 1 fs for all runs. Electrostatics were treated using the particle mesh Ewald (PME) method, all real-space cutoffs were set to 1.0 nm. The PME order was set to 6 with a Fourier spacing of 0.1 nm. Constraints were enforced using the LINCS algorithm with LINCS-order set to 4 and LINCS-iter set to 2. The LINCS-warnangle was set to 30°.
After setting up the system, a steepest descent minimization was performed using an energy step size of 0.001 nm, a tolerance of 0.1 kJ/mol/nm, and a maximum of 100000 steps.
The minimized system was subjected to a 100 ps NVT-equilibration using the velocity-rescale thermostat ($\tau_T = 0.1$ ps) and a temperature of 296 K. This combination of thermostat, target temperature, and τ_T was employed for all runs hereafter. This tempering was followed by a 5 ns NPT-equilibration using the Berendsen barostat ($\tau_P = 0.1$ ps) set to a target pressure of 1.199688 bar.
The final production run employed the Parrinello-Rahman barostat ($\tau_P = 0.1$ ps) set to 1.199688 bar for 3 ns.

This protocol was then used to scan 208 combinations of Lennard-Jones parameters for the metal ion. The values for σ_M were linearly distributed from 0.2 nm to 0.5 nm with a stepsize of 0.02 nm, resulting in 16 values for σ_M. The other parameter, ϵ_M was chosen from logarithmically distributed values ranging from 10^{-6} kJ/mol to 1 kJ/mol with 2 values per order of magnitude, resulting in 13 values for ϵ_M in total.

4.3.2 Simulations of M^{4+} in Aqueous HF Under Extreme Conditions

The systems are simulated using classical MD by the GROMACS package, version 2019.4 [114] using the AMBER force field. The water model is TIP4P/2005 [63] and the HF-Model is a 3-site model published by Orabi et al. in 2020 [39]. The boxes were set up to recreate a concentration of 1.0 M HF, as well as the density of water at 400 bar and the given temperature. The densities were taken from [115]. An overview is shown in table 4.5. The systems were generated by placing manually generated

TABLE 4.5: Overview of the simulation boxes. Each box is assigned one number for reference. The density ρ at temperature T were taken from [115]. L refers to the length of one side of the cubic simulation box. The last five columns contain the number of molecules of that respective species inserted into the simulation box.

Number	ρ / kg/m^3	T / K	L / nm	H_2O	OH^-	F^-	HF	M^{4+}
1	1011.95	298.15	2.4578	500	2	2	7	1
2	976.09	373.15	2.4876	500	2	2	7	1
3	890.93	473.15	2.5671	500	2	2	8	1
4	764.35	573.15	2.7049	500	2	2	10	1
5	671.84	623.15	2.8285	500	2	2	12	1
6	523.30	673.15	3.0821	500	2	2	16	1

$[MF_2(OH)_2(H_2O)_2]$ octahedral complexes into an empty, cubic box. Subsequently, 498 water molecules and the corresponding amount of HF molecules were added to the box. The mass of the cation was set to 178.49 amu, the value of hafnium. The F-anions inherited the same LJ parameters as the bound F-atom in HF. The hydroxide ions were given a partial charge of 0.32 e and -1.32 e for the hydrogen and oxygen-atom respectively, following the approach of [116]. The bond distance in the hydroxide was fixed to 0.09572 nm, the O-H bond length of TIP4P-water. The oxygen inherited its Lennard-Jones parameters from the TIP4P/2005 water model. For the metal ion, Fourierused the Lennard-Jones parameters reported by Zhang et al. [45] for zirconium in TIP4P/2005 water, i.e. σ = 0.046835 nm and ϵ = 4388.3373 kcal/mol.
In all simulations, Fourierconstrained the bond length of HF to the gas phase value of 0.917 Å. Furthermore, the timestep was set to 1 fs for all runs. Electrostatics were treated using the particle mesh Ewald (PME) method [101], and all real-space cutoffs were set to 1.0 nm. The PME-order was set to 6 with a Fourier-spacing of 0.1 nm. Constraints were enforced using the LINCS-algorithm [100] with LINCS-order set to 4 and LINCS-iter set to 2. The LINCS-warnangle was set to 30°. We employed periodic boundary conditions in all directions. After setting up the system, a steepest descent minimization was performed using an energy step size of 0.001 nm, a tolerance of 0.1 kJ/mol/nm, and a maximum of 100000 steps. The minimized system was simulated for 3 ns in the NVT ensemble using the velocity-rescale thermostat

($\tau_T = 0.1$ ps) and the leap-frog integrator. The first nanosecond was discarded as equilibration time. At the beginning of all NVT runs, the velocities were drawn from the Maxwell-Boltzmann distributions of the target temperatures.
We used 5 different starting structures, varying all possible relative placements of the ligands, the structures are shown in figure 5.45.

4.3.3 Identifying Coordination Polyhedra

Steinhardt parameters [25] were computed using the angular coordinates of fluorine- and oxygen-atoms in the first solvation shell. Steinhardt parameters are shape descriptors and are used to represent structures numerically. They assign a value to any configuration of particles in space and thus allow one to group structures in a meaningful way. They are useful for characterising solvation shells, because they are rotationally invariant. Plenty of shape descriptors have been used in literature, the interested reader is referred to [89].
Steinhardt parameters are based on spherical harmonics $Y_l^m(\theta, \phi)$ and take the configuration of the entire system into account.
Consider N neighboring particles with positions $\mathbf{q}_k = (r_k, \theta_k, \phi_k)$ around a central particle, which is located in the origin.

$$q_l^m := \frac{1}{N} \sum_{k=1}^{N} Y_l^m(\theta_k, \phi_k) \tag{4.8}$$

$$Q_l := \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |q_l^m|^2} \tag{4.9}$$

Where Q_l is the Steinhardt parameter. The q_l^m are summed over m to generate rotationally invariant parameters.
To make sense of the Steinhardt parameters, 125 reference polyhedra coordinates were taken from [117]. The first 10 Steinhardt parameters were computed for these structures. To deliver a guess for the structure at hand, the euclidean distance in this 10-dimensional space, between the time average of the computed Steinhardt parameters for the first solvation shell and the reference values, was used to find the closest reference structures.

4.3.4 Distributions

To ensure comparability of the distance distribution plots for the ligands of the $[MF_2(OH)_2(H_2O)_2]$ complex, I chose 200 equally spaced bins in the range of 0.15 nm up to and excluding 0.5 nm. The boundaries cover the observed ranges of distances. The histograms are not normalized and show bin counts on the y-axes.

4.3.5 Pearson Correlation Coefficient

The Pearson correlation coefficient $p(x, y)$ is a measure of the linear correlation between two observables x and y.

$$p(x, y) = \frac{\mathbb{E}\left[(x - \mathbb{E}(x))(y - \mathbb{E}(y))\right]}{\sigma_x \sigma_y} \tag{4.10}$$

Here $\mathbb{E}(x)$ denotes the mean value and σ_x the standard deviation of x respectively. The Pearson correlation coefficient is bound by -1 and +1, where +1 indicates that the

two observables are identical. If p(x,y)=0 then there is no linear correlation between the two observables.

4.3.6 Potential Energy Scans

We computed the Lennard-Jones and Coulomb potential acting on one fluoride atom in the all-trans configuration for several cation-oxygen distances. The positions of the hydroxide-oxygens and one fluoride ion were kept fixed at 0.19 nm and 0.185 nm respectively. The water-oxygens were represented by a point carrying a charge of -1.128 e, which is the charge of the virtual site in the TIP4P/2005 water model. This point also inhered the Lennard-Jones parameters of the TIP4P/2005 water oxygen. The two water-oxygens were moved simultaneously. The hydrogen atoms were omitted for these calculations. This simplification is justified by the fact that the Coulomb potential dominates the Lennard-Jones potential by several orders of magnitude in the relevant distance regime.

4.3.7 Weighted Averages

The weighted averages were computed using

$$\langle r \rangle = \sum_{t=1}^{T} \frac{1}{T} \frac{\sum_{i=1}^{N} w_i(t) r_i(t)}{\sum_{i=1}^{N} w_i(t)}$$

where $w_i(t) = r_i^{-2}(t)$ are the weights that represent the sensitivity of the XANES data. Here, $r_i(t)$ denotes the distance of the i^{th} closest ligand at time t. For the six closest ligands $N = 6$.

4.3.8 Equilibrium Constants

The equilibrium constants K for the Van-'t-Hoff analysis, figure 5.53, were computed by counting the number of frames in which the sixth closest water molecule was at distances $r > 3$ Å and those where it was at distances $r \leq 3$ Å. The ratio of these numbers of frames was then taken as the equilibrium constant. This procedure is expressed in the equation below.

$$K = \frac{\text{\#frames where } r_6 > 3 \text{ Å}}{\text{\#frames where } r_6 \leq 3 \text{ Å}}$$

Chapter 5

Results & Discussions

5.1 Dynamics of $[Al(H_2O)_6]^{3+}$ and its Solvation Shell

5.1.1 The Solute: Hexa-Aqua-Aluminium Complex

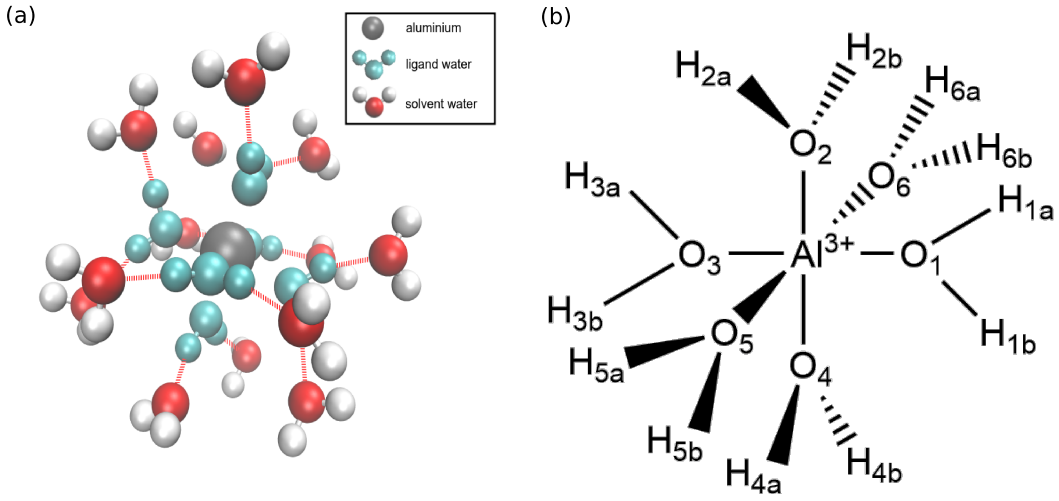

FIGURE 5.1: (a): Hexa-aqua-aluminium-complex (green and grey) and its first solvation shell (red and white). (b): Alternative representation of the Hexa-aqua-aluminium-complex.

During the equilibration of the simulation box, six water molecules (numbered 1 to 6 in figure 5.1) spontaneously bound to the Al^{3+} ion and formed the octahedral hexa-aqua-aluminium-complex $[Al(H_2O)_6]^{3+}$ (figure 5.1.a). The complex remained stable throughout the 50 ns simulation, and none of the six water ligands exchanged with the surrounding solvent water molecules. Because of this stability, the hexa-aqua-aluminium complex can be treated as the solute of our system, whose conformational state space is denoted Ω_{solute}.
Once aligned into an octahedral complex, the relative positions of the water ligands did not change, such that the only flexible degrees of freedom are the rotations of the water ligands around Al^{3+}-oxygen axis (figure 5.1.b). Therefore these six torsion angles $(\theta_1, \theta_2 \dots \theta_6)$ are chosen as the state space of the solute: $\Omega_{solute} \subset \mathbb{R}^6$. figure 5.2.a shows the equilibrium distribution of the torsion angles of all ligands. The distribution shows four maxima: at $\theta_3 = 90°$ and at $\theta_3 = 270°$ when the water ligand lies in the xy-plane, as well as at $\theta_3 = 0°$ and at $\theta_3 = 180°$ when the water ligand lies in the yz-plane. The maxima are not very pronounced, which means that these

states are only separated by small barriers. Correspondingly, the observed transitions between maxima are on the timescale of picoseconds. A rotational transition between the horizontal orientation of the water ligand ($\theta = 0°$ or $\theta = 180°$) and the vertical orientation of the water ligand ($\theta = 90°$ or $\theta = 270°$), or vice versa, will be called a flip.
The orientations of neighboring water molecules are correlated (figure 5.2.d). The product of the marginal distributions in figure 5.2.c predicts the joint distribution in the uncorrelated case. The simulated joint distribution in figure 5.2.d consists of several local maxima that are separated by patches of low probability density. The distribution in figure 5.2.c is rather homogeneous and lacks regions with zero probability density. The difference between both distributions implies a strong correlation in the motion of the ligands.
The shown pair of angles, namely θ_1 and θ_3 serve as examples. All other pairs of torsion angles yield qualitatively identical results.

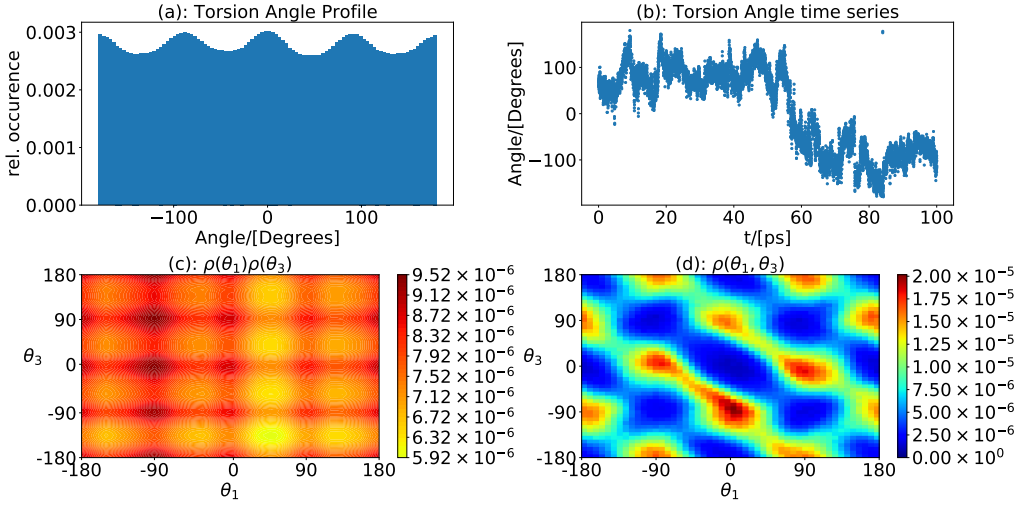

FIGURE 5.2: (a): Distribution of torsion angles for all six ligands. (b): Exemplary trajectory of θ_3. (c): Product of marginal distributions of θ_1 and θ_5. (d): Joint Distribution of θ_1 and θ_5.

In conclusion, the interactions between hydrogen atoms cause more or less favorable configurations to arise. Therefore, the dynamics of this configurational space can be studied using Markov models.

5.1.2 Markov State Model for Individual Ligands

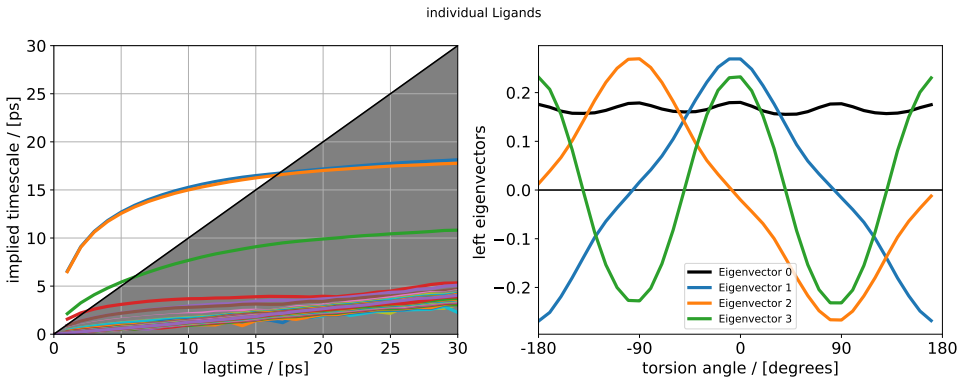

FIGURE 5.3: Left: Implied timescales vs. lagtime for discretization strategy (i). Right: first four left eigenvectors.

The first MSM is constructed on the torsion angle of an individual ligand. The implied timescales and slowest processes are shown in figure 5.3. There are three slow processes, two of which are degenerate and occur at ca. 16 ps while a third, much faster process lies at ca. 6 ps. This faster process is only poorly resolved and does not converge. The stationary distribution shows four slight peaks, similar to the underlying distribution. The two degenerate processes correspond to changes in the torsion angle of 180°. The next fastest process is a 90° rotation.
There are four stable states for each ligand, indicating the entire complex could be described in an MSM using $4^6 = 4096$ states. However, this model converges poorly as many configurations of this large state space are energetically degenerate. Thus an intermediate strategy is chosen to describe the complex, as is depicted in figure 4.2.

5.1.3 Markov State Model for a Ligand Surrounded by its Neighbours

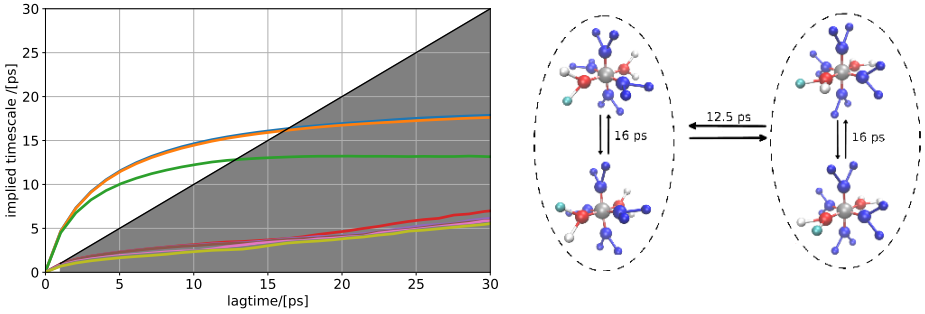

FIGURE 5.4: Left: Implied timescales vs. lagtime for discretization strategy (ii). Right: Representations of the three slowest processes. The central ion is depicted in gray. The ligand that was observed is plotted in the foreground and one of its hydrogen atoms is marked turquoise, its four neighbours are shown in dark blue.

The definition of states for this model was explained in the Methods section. The timescales and a representation of the eigenvectors of the resulting MSM can be seen in figure 5.4. Again one finds three slow processes, two of which are degenerate

and lie at about 16 ps and the next fastest process lies at about 12.5 ps. There are four dominant states in this MSM. Representative structures are shown next to the timescales.
The slowest two transitions are once again semicircular rotations of the ligand observed. The timescale remains similar to the one of the previous model. However, the timescale of the third process increases dramatically, despite being a 90° rotation as well. The difference stems from the fact that this rotation is coupled to four other 90° rotations of the neighboring ligands to ensure an optimal configuration of the complex.

5.1.4 Variational Models for the Complex

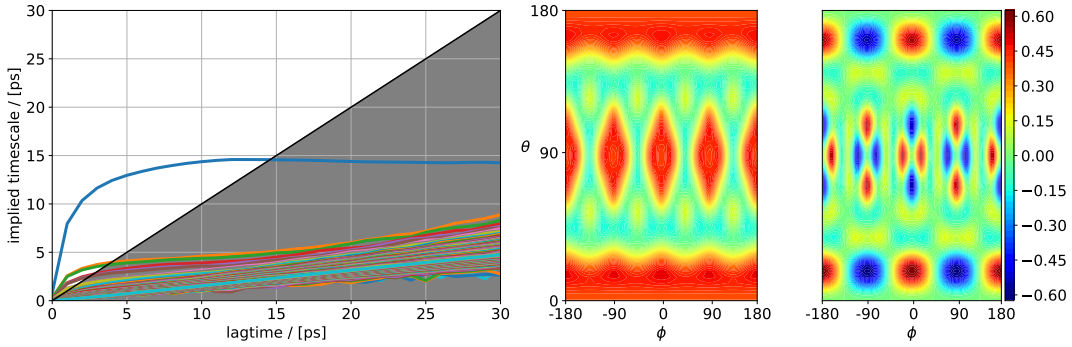

FIGURE 5.5: Left: implied timescales vs lagtime for discretization strategy (iii). Center: stationary distribution c_0, Right: first process c_1. The eigenvectors are depicted as functions of spherical coordinates. The representation is akin to a world-map.

Variational models to describe the dynamics of the hexa-aqua-aluminium complex were employed as well. The model shown in figure 5.5 used 121 spherical harmonics ($Y_l^m(\theta, \phi)$ up to $l = 10$) to approximate the distribution of all 12 H-atoms of the complex. Surprisingly, only one dominant timescale at ca. 15 ps emerges. The corresponding eigenvector implies a transition akin to the fastest process of the previous model, that is, a simultaneous change of all torsion angles by 90°. The discrepancy in the timescale for the same process between the two models may arise from a projection error in the variational model.
In this model, changes of a torsion angle by 180° are projected out. That is because the distribution of H-atoms does not change when a water-molecule is rotated by 180°.
NMR-studies by Hertz et al [118] measured the rotational correlation time τ_C of the cation-proton vector. This correlation time is the time of rotation of a vector from a proton in a hydrated water molecule to the center of an ion, which is exactly the motion of water molecules investigated herein. The NMR-experiments determined $\tau_C = 53 \pm 13$ ps. Since the above model shows an extremely robust result of 15 ps for a quarter rotation, one can assume that an entire rotation would require 60 ps. The results of the Markov model agree with the experiment.

5.1.5 Solvation Shell

Figure 5.6 shows the Al^{3+}-O radial distribution function (RDF). The three peaks correspond to the six water ligands (blue), the first solvation shell (orange), and the

second solvation shell (green). Between the first and the second peak, the distribution function drops to zero, which is in line with our observation that the ligand water molecules do not exchange with the surrounding bulk solvent. By contrast, one observes a frequent exchange of water molecules between the first and the second solvation shell.

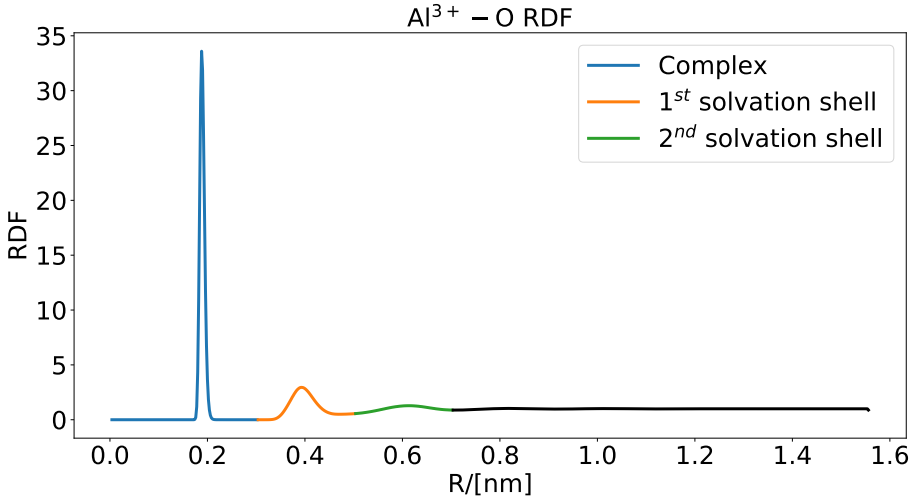

FIGURE 5.6: $Al^{3+}-O$ radial distribution function.

The solvation shells are defined by a relatively small interval of radii, hence one assumes that their dynamics can be sufficiently described using solely the angular components of spherical coordinates θ and ϕ.
Each peak of the RDF is highly structured in the $\theta - \phi$-plane and anti-correlated with its neighbours. This is illustrated in the distributions of oxygen-atoms in figure 5.7.

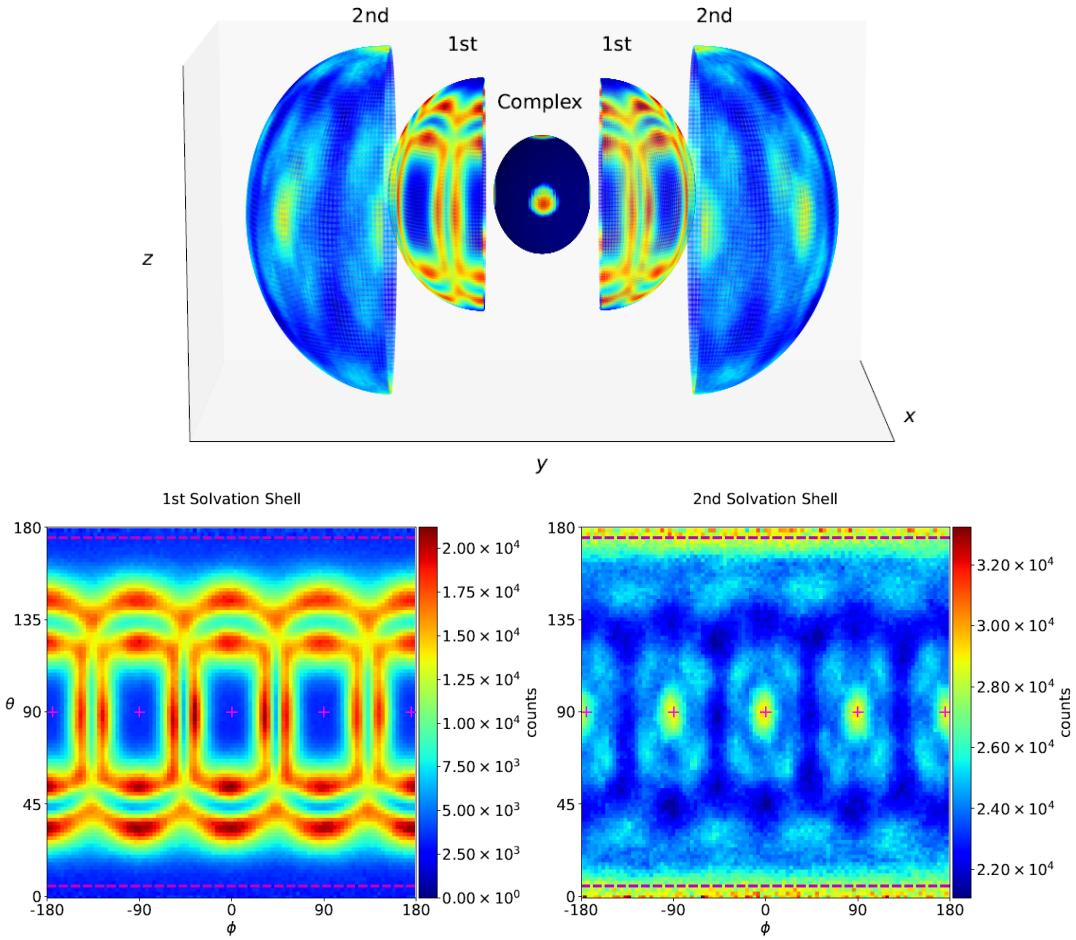

FIGURE 5.7: Top: Distribution of oxygen atoms for each peak of the RDF as colored spheres. The logarithm of the distribution of the innermost sphere is represented for visualization purposes. Bottom: Distribution of oxygen atoms for the first and second solvation shell as contour diagrams. The angular positions of the six water ligands are shown as purple crosses or dashed lines in both plots. The color bars are identical for the top and bottom parts of the figure.

The central sphere depicts two of the six spots that represent the octahedrally structured complex. The central hemispheres correspond to the first solvation shell. Six regions remain practically unvisited. The boundaries of those regions are where most particles are located. The outermost hemispheres depict the distribution within the second solvation shell. It qualitatively resembles the complex, with six highly populated regions. All distributions have in common that favored regions in one shell correspond to unfavored regions in adjacent shells. This anticorrelation hints at a coupling between the shells.

To illustrate this coupling a snapshot of the complex with its first solvation shell is shown in figure 5.1(a).

The oxygen atoms of the complex remain mostly static. Two oxygen atoms had their location fixed in the fitting procedure. However, the ligands can still rotate around their respective Al^{3+}-O-axis. The structure of the first solvation shell becomes clear if one imagines that the solvent particles of the first solvation shell follow the ligand-hydrogen-atoms in their rotation.

The dynamics of the first solvation shell are similar to the solute dynamics in terms of implied timescales and first eigenvectors. Figure 5.8 shows this result.

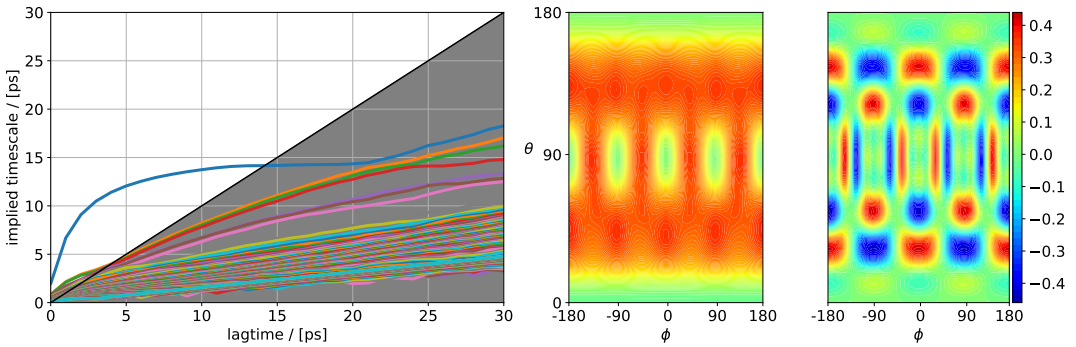

FIGURE 5.8: Left: implied timescales vs lagtime for discretization strategy (v). Center: stationary distribution s_0, Right: first process s_1.

Again one dominant timescale, which converges to roughly 15 ps, is found. The stationary distribution roughly resembles what is shown in figure 5.7, the difference stems from the different color bars in the two figures. The first process has a similar symmetry as the slowest process of strategy (iii).
There are striking similarities between the structure and dynamics of the complex and its first solvation shell. Radial and angular dynamics occur on a timescale of roughly 15 ps, for both the solute and the solvent. This is a further hint for coupling between these two subsystems, the precise details of this coupling will be explored next.

5.1.6 Joint Models

Combined models were constructed by either adding or multiplying the respective vector spaces. The resulting Markov models are evaluated in terms of implied timescale values and convergence, as well as their eigenvector sign structure.

Additive Models

The timescale comparison of the additive models is shown below in figure 5.9. The joint models perform slightly better than the individual models in terms of absolute values. The implied timescales of the joint models all lie above the implied timescales of the individual models. Convergence could not be reached faster. However, the sum of the complex-MSM and solvent model yield a significantly better timescale for their third process. The sum of both variational models shows a slight improvement to the slowest process in terms of absolute timescales.

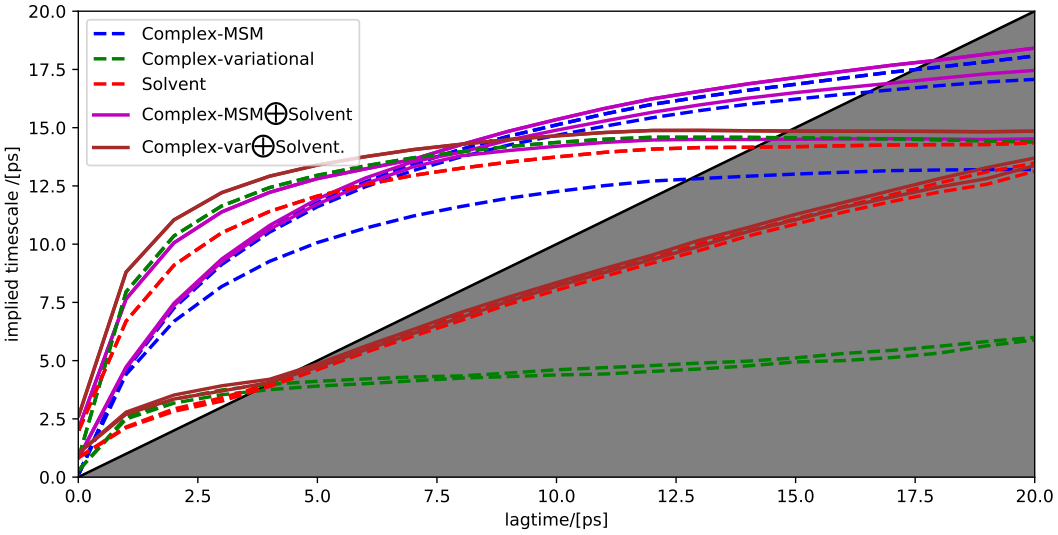

FIGURE 5.9: Implied timescale comparison of different additive joint models as well their constituents.

In figure 5.10 the first two eigenvectors of the first additive model are depicted. The stationary distributions as well as the first processes for both subsystems, i.e. complex and solvation shell, are equivalent to those already shown for the individual systems. Here, the fourth eigenvector is depicted which corresponds to the 12 ps transition in figure 5.4.
The same observation can be made in figure 5.11. The two eigenvectors of both subsystems are equivalent to those already discussed for the individual systems.

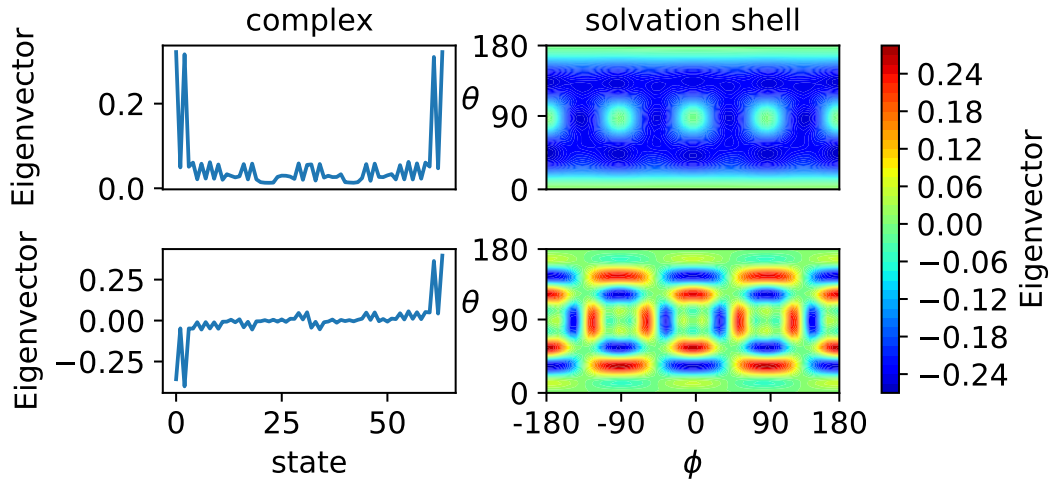

FIGURE 5.10: First and fourth left eigenvectors of the sum of the Markov state model for the complex and the variational model of the solvent.

Interestingly, the sign structure for the complex and the solvation shell is equivalent. That is, positive regions in the first process of the complex have corresponding regions in the first process of the solvation shell.

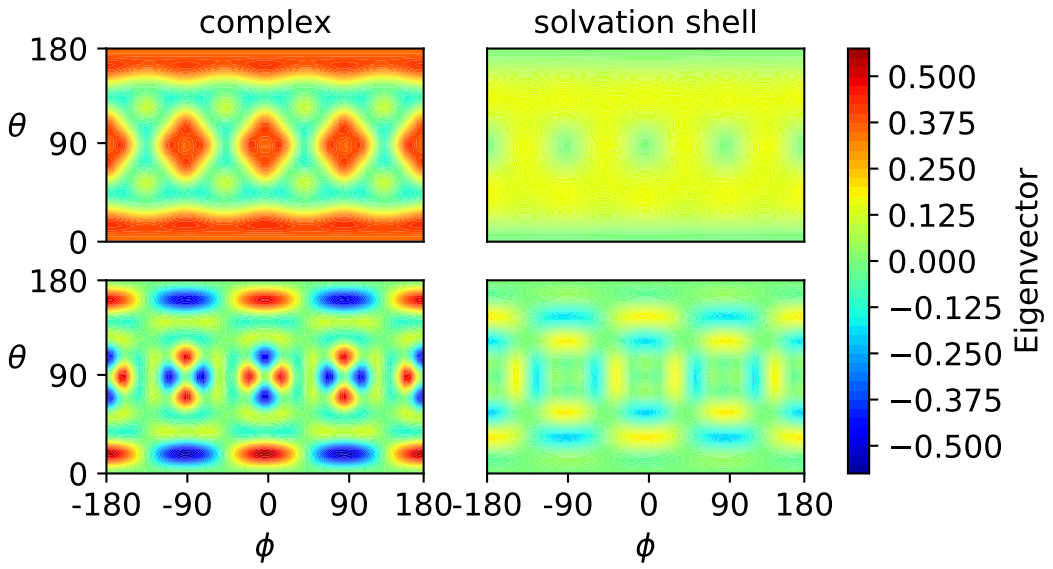

FIGURE 5.11: First and second left eigenvectors of the sum of the variational model for the complex and the variational model of the solvent.

Multiplicative Models

The multiplicative models show mixed results. While combining the two variational models in this way yields no improvement whatsoever, the variational model combined with the solvation shell performs worse. Depicted in figure 5.12 are the first processes of both combined models because higher processes for the MSM ⊗ solvent model perform even worse.

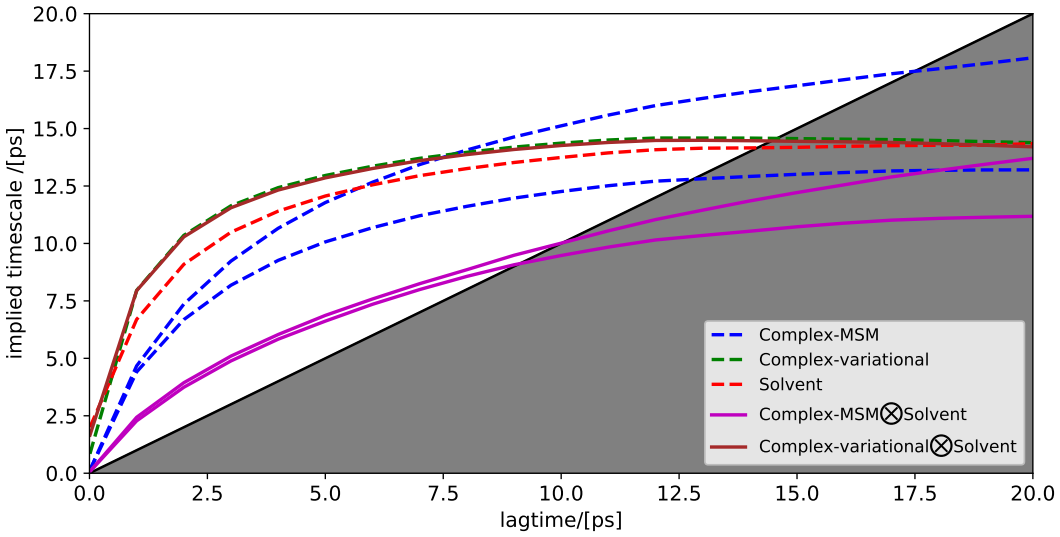

FIGURE 5.12: Implied timescale comparison of different multiplicative joint models as well their constituents.

In the following figures, let c_i denote the i^{th} eigenvector of the complex and let s_i denote the i^{th} eigenvector of the solvation shell.
In figure 5.13 the first two eigenvectors of the product of both variational models

are depicted. The stationary distribution has two contributions $c_0 \otimes s_0$ and $c_1 \otimes s_1$. The first contribution is simply the product of both stationary distributions and is expected to show up in this eigenvector. The second contribution is the product of both first processes. This observation can only be explained if both processes occur simultaneously. That is, if the projection of the solvation shell dynamics onto s_1 is positive or negative, the projection of the complex dynamics onto c_1 is positive or negative too. Therefore, the product of both will always have the same sign and appear like a stationary distribution.
Eigenvector 1 has two contributions as well, these are $c_0 \otimes s_1$ and $c_1 \otimes s_0$. Both of these contributions are products of projections of individual processes multiplied with a constant process. This does not change the underlying sign structure. Both contributions occurring with the same sign indicate that both individual processes occur with equivalent sign structures, given that the stationary distributions are positive.

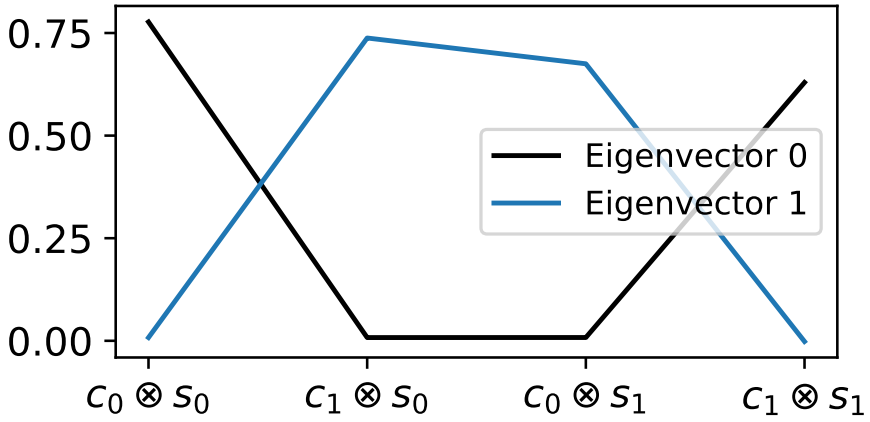

FIGURE 5.13: First and second left eigenvectors of the product of the variational model for the complex and the variational model of the solvent.

In figure 5.14, the first four eigenvectors of the product of the Markov state model for the complex and the variational model for the solvation shell are shown.
In this model, eigenvector 3 is equivalent to eigenvector 1 in figure 5.11. The stationary distribution contains two contributions, the product of both stationary distributions and the product of the coupled processes c_3 and s_1. These observations are equivalent to the previously discussed model.
The interpretation of eigenvectors 1 and 2 remains unclear. However, their corresponding timescales are also poorly converged.

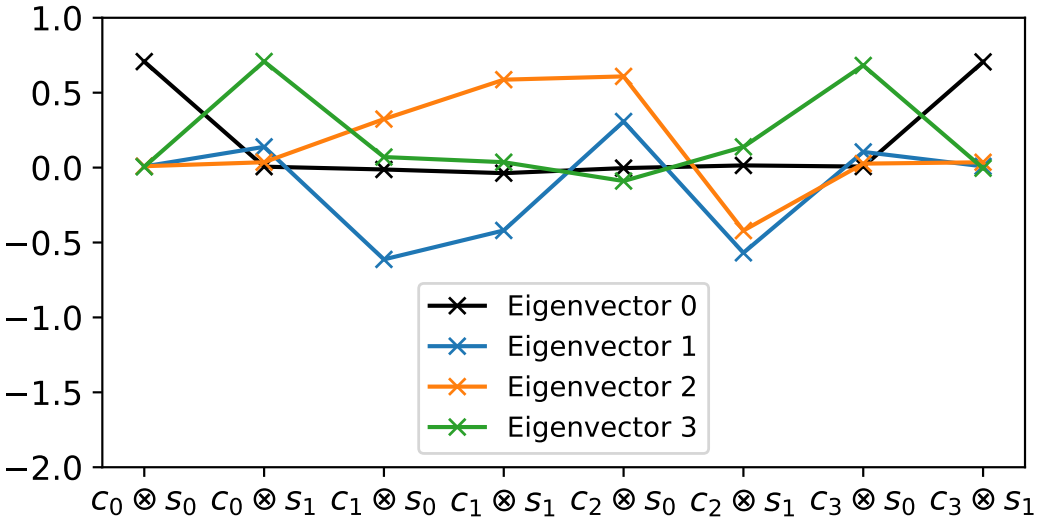

FIGURE 5.14: First and second left eigenvectors of the product of the Markov state model for the complex and the variational model of the solvent.

Combined models perform differently. While additive models always resulted in at least slightly improved timescales, the multiplicative models performed just as well as their components or even worse.
Interpreting the eigenvectors for the additive models is straightforward, one simply has to look at the contributions of each subsystem to the combined eigenvector. The eigenvectors of the multiplied models require a deeper understanding of the projections of the processes onto the individual eigenvectors and are thus more challenging to interpret.
This analysis showed no advantage of the multiplicative models and leads to the conclusion that additive models are preferred.

5.2 The Parameterization of HF with Two Sites

5.2.1 Approach

The goal of this project is to establish a model of HF, suitable for describing its interaction with metal ions in classical MD simulations. A system of metal ions in hydrofluoric acid consists of three components, water, HF, and the metal ion. One can therefore identify six interaction pairs: (i) water-water, (ii) water-HF, (iii) water-cation, (iv) HF-HF, (v) HF-cation and (vi) cation-cation. Not all of these interactions need to be accounted for in this project. It is assumed that (i) is sufficiently parametrized by the water model and that (iii) is dealt with by the model of the cation. Interaction (vi) is negligible for dilute systems.
The remaining interaction pairs are the focus of this project. Suitable parameters are determined for (ii) by computing density curves for several temperatures. Densities as well as radial distribution functions of pure HF are used for (iv). Finally, (v) is included by estimating the solvation free energy of NaF in pure HF and computing this quantity using the free energy perturbation method.

5.2.2 Densities

The first quantity to be tackled was the density as a function of the HF concentration. The experimental values were taken from Eurofluor [107]. Because the simulations were conducted at different concentrations than reported experimentally, the literature curve was fitted to a polynomial of order 5. The fit is shown below in figure 5.15.

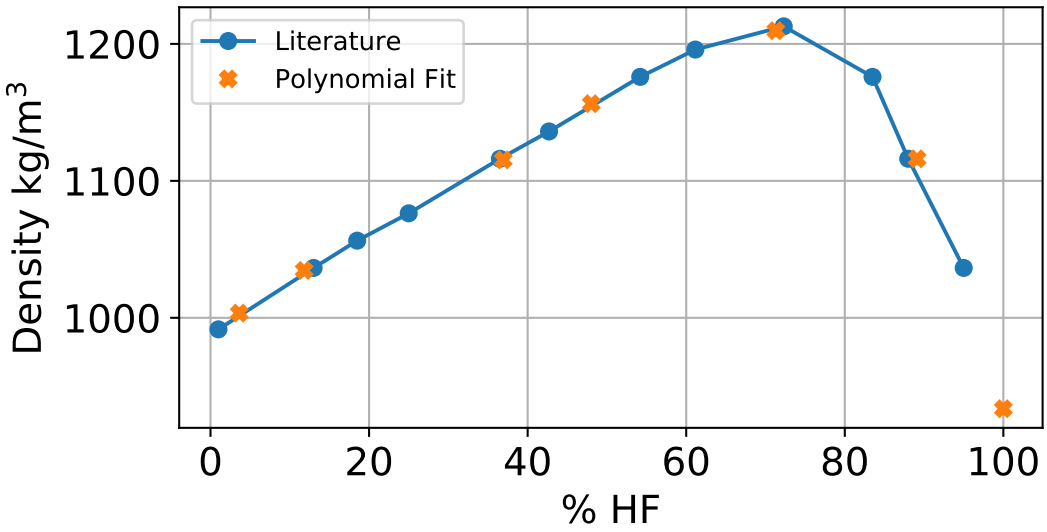

FIGURE 5.15: Literature densities [107] of HF and polynomial fit. The crosses correspond to the concentrations of the simulation boxes.

The quality of the fit suffices for comparing the results of the MD simulations to it. The literature densities were measured at 300 K, however, hydrofluoric acid is a gas at this temperature and standard pressure. Therefore, the density curves were fitted only up to the second last concentration. The five best parameter combinations and their respective density profiles are shown in figure 5.16.

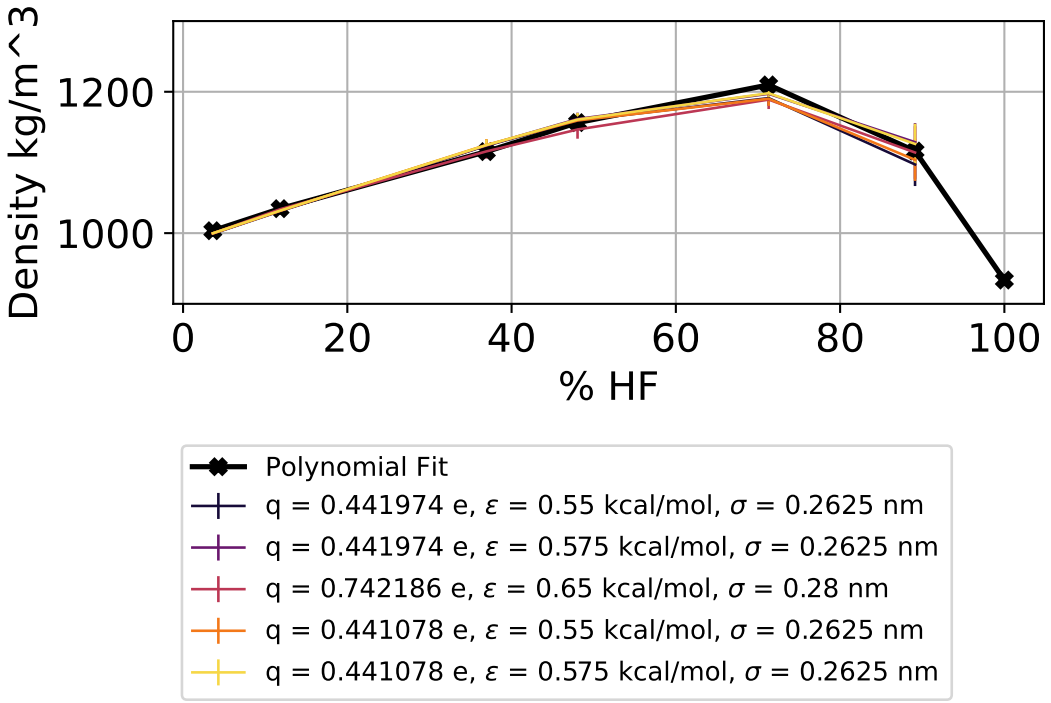

FIGURE 5.16: Five best parameter combinations to reproduce the densities of hydrogen fluoride, excluding pure HF.

Initially, the simulations were evaluated by including the results for pure HF as well. However, the error bars increase drastically for this temperature, as can be seen in figure 5.17.

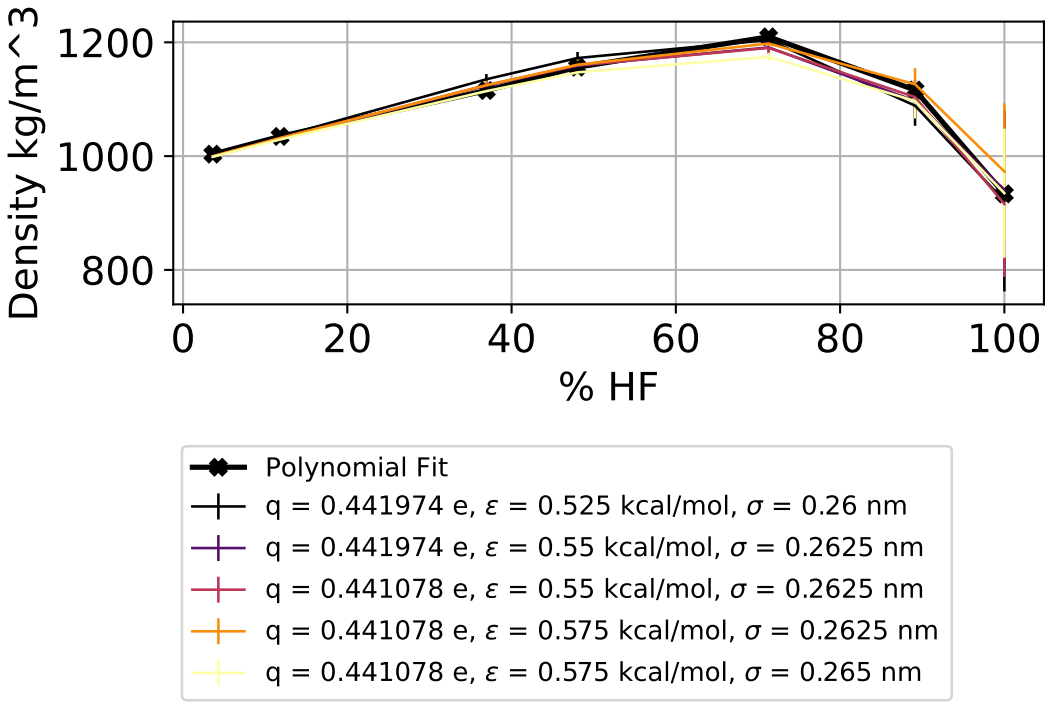

FIGURE 5.17: Five best parameter combinations to reproduce the densities of hydrogen fluoride, including pure HF.

Nevertheless, one can see that for both fitting schemes, the best parameter combinations lie closely together in the parameter space.

5.2.3 Radial Distribution Functions

The second fitting target was the locations of maxima of the radial distribution functions (RDFs). That is, the first maximum of the F-F and H-H RDF as well as the second maximum of the F-H RDF. The first maximum of the F-H RDF corresponds to the intramolecular F-H bond length, which is a topology parameter and constrained during the simulations.
The experimental curves are shown below in figure 5.18. The left part of the H-H RDF is convex, this property cannot be reproduced by the 2-site model. The curves were taken from figure 2 of [119] and digitized manually. More features that were not reproducible by the MD simulations are the many maxima occurring after the first two peaks.

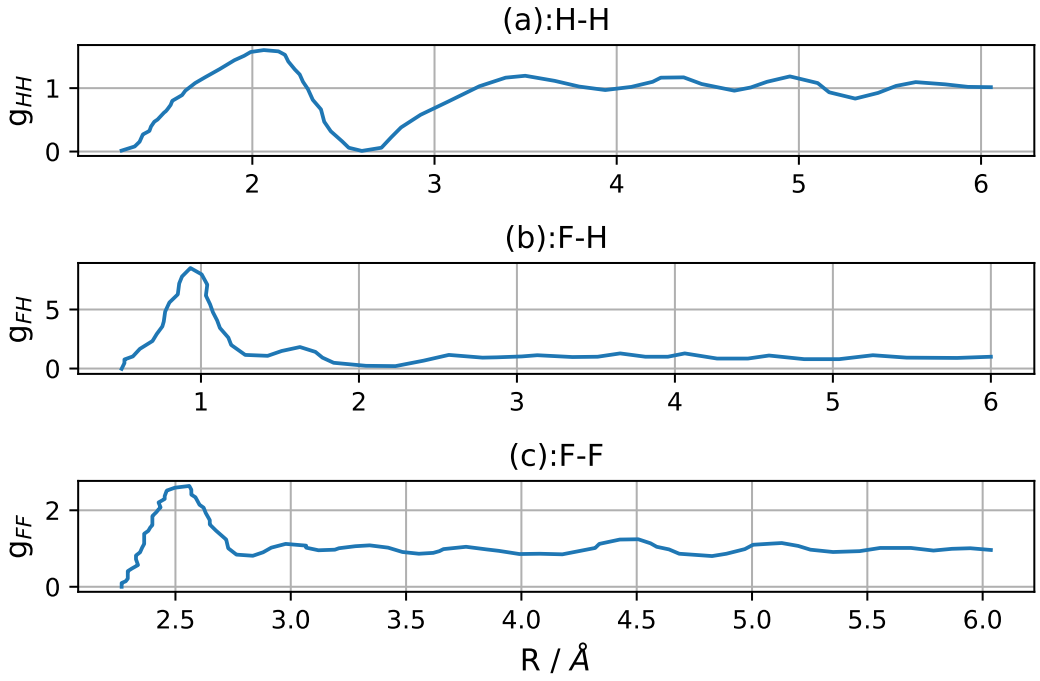

FIGURE 5.18: Radial distribution functions of pure HF taken from [119].

The location of the first peak of the H-H-RDF could be reproduced sufficiently. However, the shape of the RDF does not match, and the second maximum strays from the literature curve as well. This indicates that smaller H-H distances are undersampled. The wrong location of the second peak could indicate that the H-F···H angle might be too large. The five best results are shown in figure 5.19.

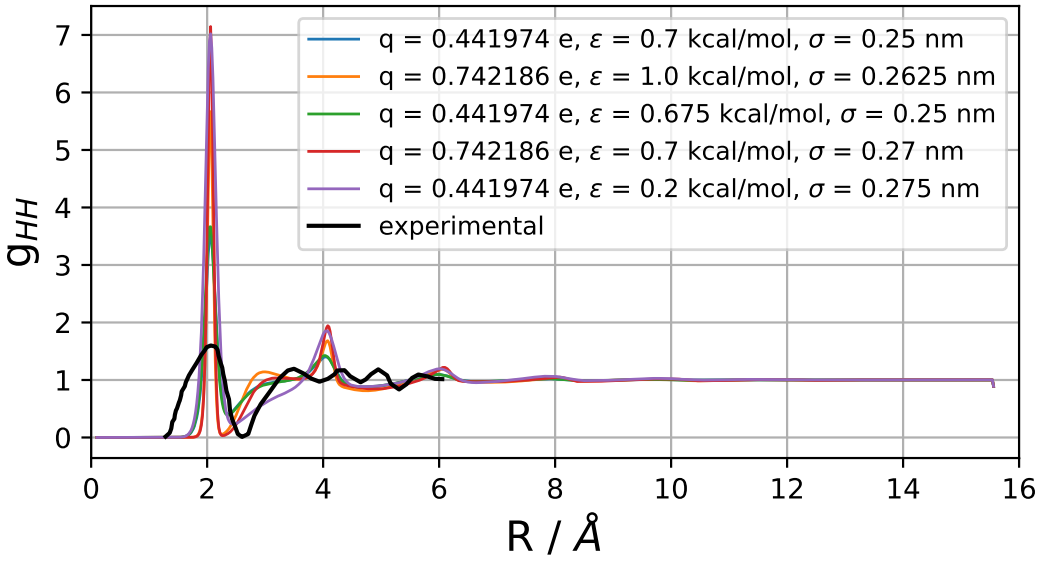

FIGURE 5.19: Experimental H-H RDF and simulated RDF of the five best parameter combinations.

The F-H RDF performs slightly better than the H-H RDF. The first peak is much narrower than the literature curve because this bond is constrained during the simulations. The second peak location could be reproduced perfectly. The intensity of the peak does not match, however. Again the following peak strays significantly from the literature curve. This could be a result of wrong angles in the H-F chain or wrong distances or wrong relative orientations between the H-F chains. The best five combinations are shown in figure 5.20.

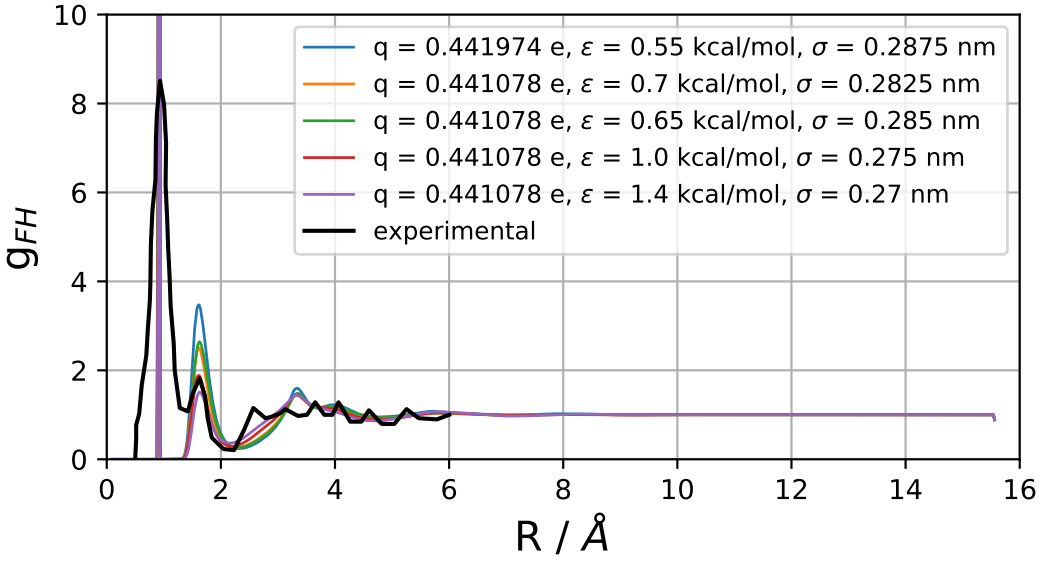

FIGURE 5.20: Experimental F-H RDF and simulated RDF of the five best parameter combinations.

Finally, the F-F RDF can reproduce the peak location but neither its height nor the shape of the right side of the peak. The second maximum is located very far from the second maximum in the literature curve. This indicates that the F-H···F angle might

be too large as well. The best five parameter combinations are shown in figure 5.21.

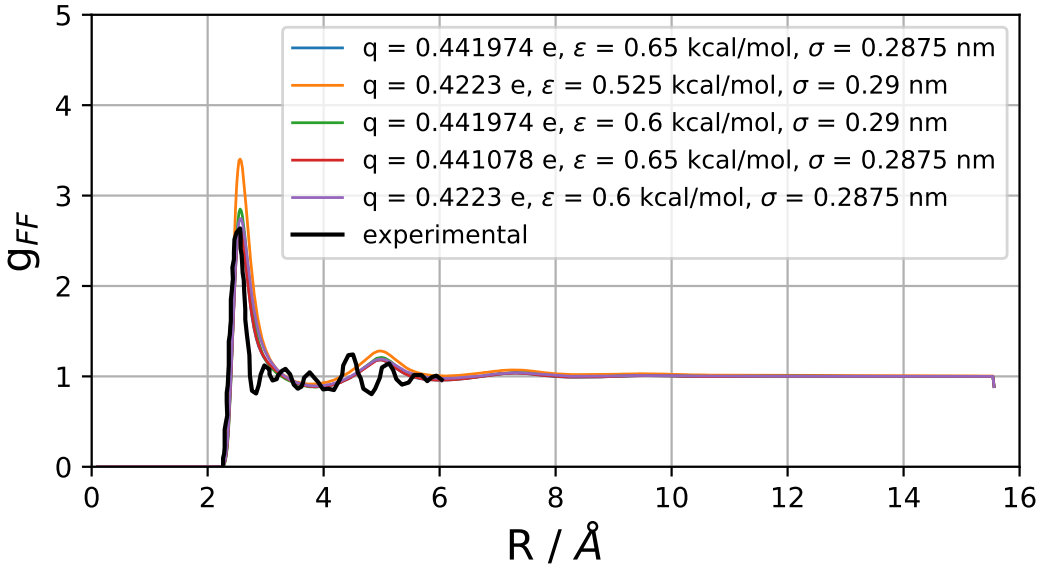

FIGURE 5.21: Experimental F-F RDF and simulated RDF of the five best parameter combinations.

This overall trend indicated that the classical simulations are only able to reproduce a single H-F molecule and its immediate surroundings to some extent. The angles within the H-F chain and the distances and orientations between the H-F chains must differ from the simulations.

5.2.4 Hydration Enthalpies of NaF in HF

Thermodynamic integration runs require ample amounts of fine-tuning. A cachet of a good simulation setup is matching results for forward and backward thermodynamic integration simulations. That is, switching the interactions on or off respectively yields the same result, with the opposite sign. Visually this can be seen in hysteresis plots. If the curves for forward and backward integration meet at $\lambda = 1$, the runs are converged. However, this graph is rather a guideline and not as accurate as a numerical analysis. Such a graphical evaluation is shown figure 5.22 for some parameter combinations. The plots show the alignment of both curves, indicating a reliable result. Many lambda points were required towards the end(beginning) of the forward(backward) integration.
The quality of the thermodynamic integration was ensured by fine-tuning the lambda-points until the results of the forward and backward integrations yielded agreeable results.

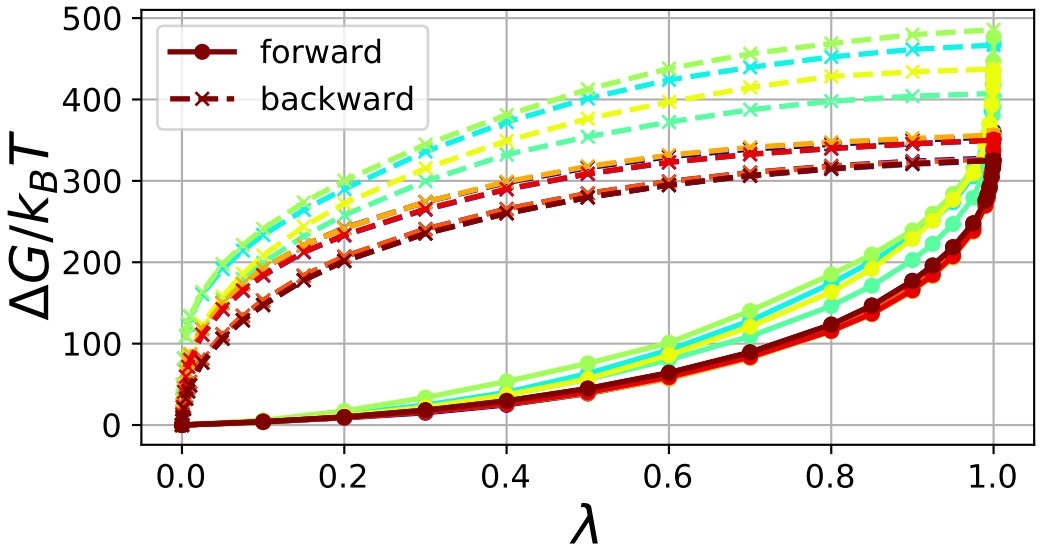

FIGURE 5.22: Hysteresis plots for several parameter combinations. The dashed lines indicate thermodynamic integration runs where the interactions were slowly switched on, while the solid lines indicate those runs where the interactions were slowly switched off.

The scan of the parameter space, shwown in figure 5.23, indicates that the solvation free energy is mostly dependent on the charge of the ions. This becomes evident by the fact that for $q \approx 0.74$ e all solvation-free energies lie below -900 kJ/mol. The second most impactful parameter is σ, the Van-der-Waals radius. Smaller values lead to larger solvation enthalpies. Likely due to electrostatic interactions growing inversely with distance. That is, a smaller σ, allows for solvent molecules to move in closer to the solute, which leads to larger electrostatic interactions between solvent and solute. The depth of the Lennard-Jones potential is governed by ϵ this parameter had the smallest influence on the solvation free energy, with differences only becoming noticeable over large jumps in the parameter space.

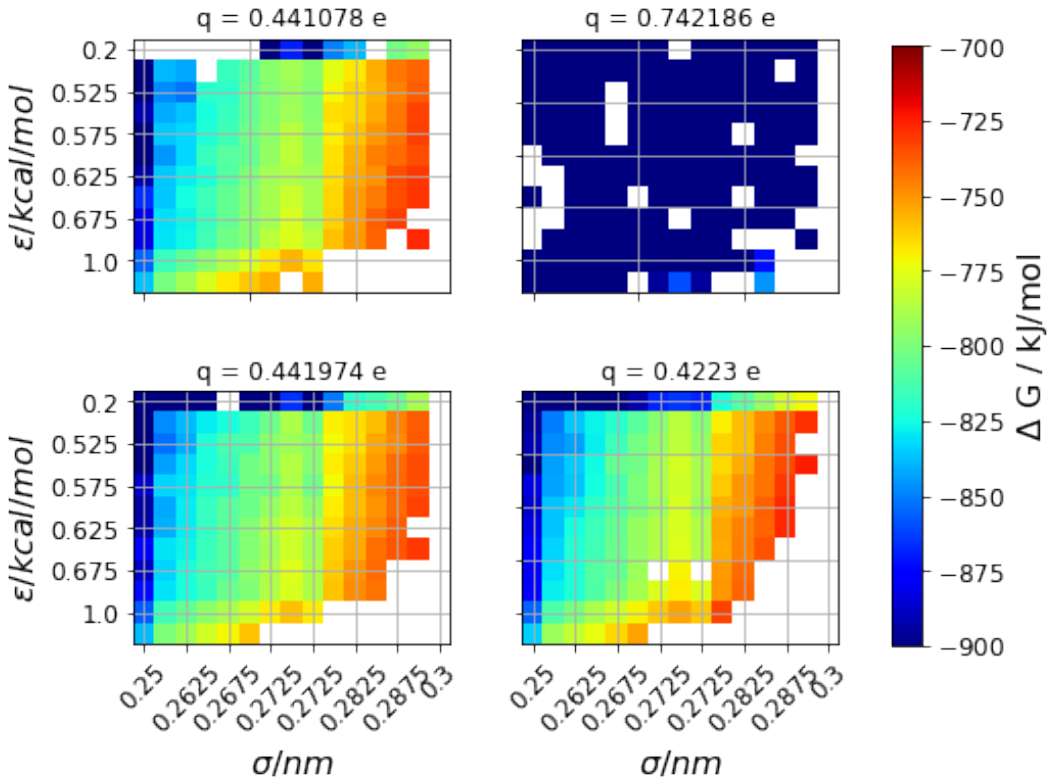

FIGURE 5.23: Resulting solvation free energies for all parameter combinations. The color bar ranges from -700 to -900 kJ/mol only for better visibility. The white dots indicate faulty runs which were not completed.

5.2.5 Overall Performance of the Sampled Parameter Space

The parameter combinations were evaluated by their ability to recreate the five physical quantities. The quantities were discussed in the previous sections and consist of the first maxima of the H-H and F-F radial distribution functions, titled g_{HH} and g_{FF} respectively. The second maximum of the F-H radial distribution function, titled g_{FH}. The overall density profile for HF-contents up to 90%, called Density, and the free energy of solvation of Na^+ and $[FHF]^-$ in pure HF.
The maxima of the radial distribution functions were considered recreated if the simulated maximum was within 0.1 nm of the experimental value.
The density profile was considered recreated if the overall mean squared error satisfied:

$$\Delta_\rho < 1000 \left(\frac{kg}{m^3}\right)^2 \tag{5.1}$$

The free energy of solvation ΔG_{solv} was considered recreated if the following expression was satisfied:

$$abs\left(\Delta G_{solv} - (-757.9\ \text{kJ/mol})\right) \leq \delta\Delta G_{solv} \tag{5.2}$$

where $\delta\Delta G_{solv}$ is the uncertainty associated with the estimated free energy.
An overview of the performance of the parameter combinations is shown in figure 5.24.

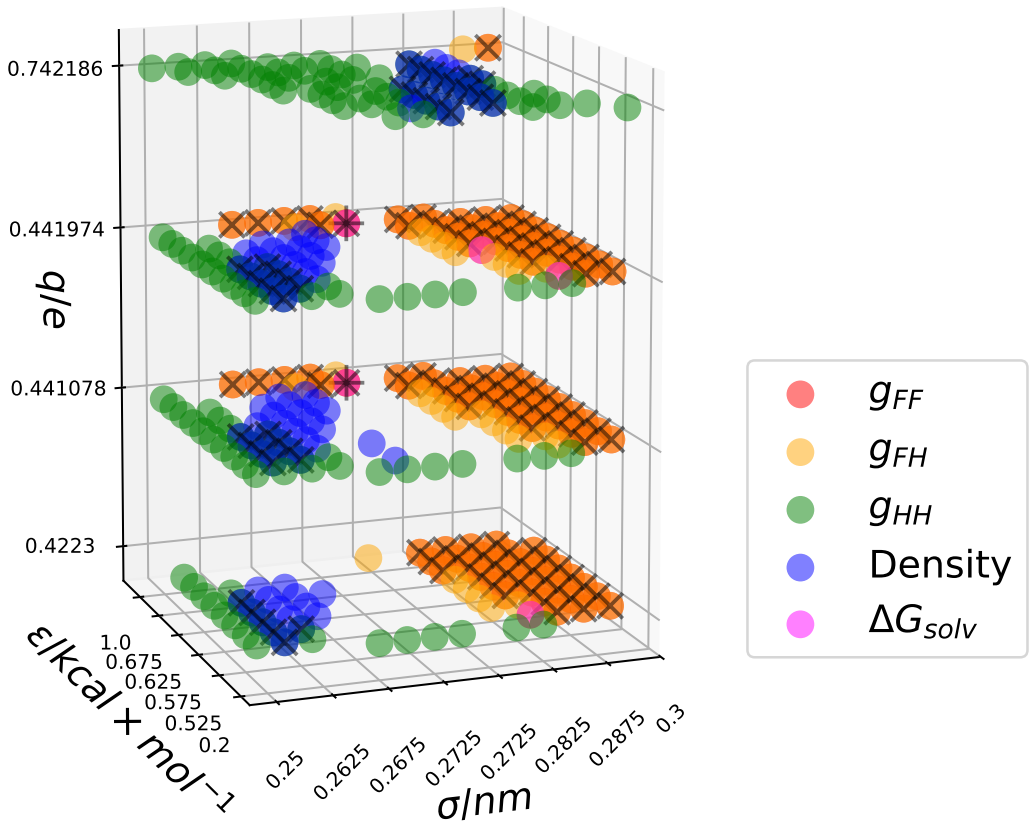

FIGURE 5.24: Parameter combinations which successfully recreated various experimental properties. Different colors indicate which property was recreated. The points in parameter space marked by an "x" indicate points where two properties were recreated simultaneously. Two points were able to recreate three properties, these are marked with an eight-spoked asterisk.

No parameter combinations were able to recreate all five quantities simultaneously. Some combinations were able to recreate two quantities. These doubly successful combinations could either recreate g_{FF} and g_{FH} or g_{HH} and the density. These are marked with crosses in the plot. Two parameter combinations were able to recreate three quantities. These were able to reproduce g_{FF}, g_{FH} and ΔG_{solv}.
The triply successful parameter combinations combined properties which are governed by the HF-HF interaction and the HF-M^+ interaction. That is, the HF-H_2O interaction is missing. The doubly successful combinations included the HF-H_2O interaction via the density as well as the HF-HF interaction in parts. Therefore, no parameter combinations yielded a satisfactory result.
From the locations of the points clouds in parameters space, it seems as if no overlap of all five quantities could be created by choosing the parameters accurately within the sampled paramter space. Extending the range of the Lennard-Jones parameters, especially σ could prove fruitful.

5.2.6 Simulating NMR-Spectra of Na^+ in Aqueous HF

The simulated NMR-spectrum of NaF in aqueous HF at various concentrations displays some interesting features, which could potentially be used to identify suitable parameter combinations. The spectrum is displayed in figure 5.25. One such feature is the shifting of the ^{23}Na signal towards smaller chemical shifts with increasing HF concentrations. Another interesting feature is the left shoulder of the double peak of the ^{1}H signal which grows with increasing HF concentrations. The fine structure of all peaks is unfortunately no suitable fitting target, due to the solvation shell dynamics occurring on faster timescales than the NMR-spectra can measure.

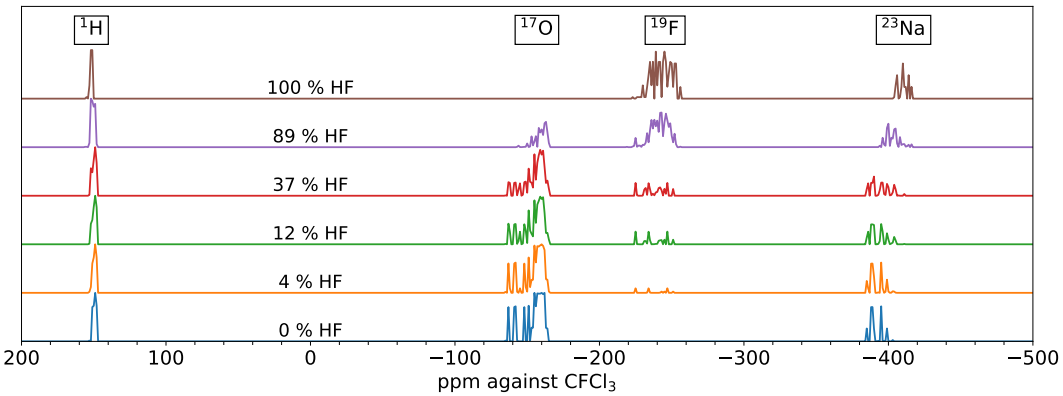

FIGURE 5.25: Calc. NMR-shifts against ^{19}F-$CFCl_3$ of 100 $[Na(H_2O)_x(HF)_y]^+$ with $x, y \in [0, 8]$, CN = {4;5;6;7;8} and 3 conformers per species.

Dr. Patrick Pröhm measured NMR-spectra for NaF in water, aqueous HF, and aHF. All spectra for different nuclei showed only one singlet. The location of the singlets is shown in table 5.1.
The measured and calculated chemical shifts are referenced with respect to different substances. Because the referencing is only a linear shift, one can still evaluate the shifts of peaks with increasing HF concentrations. By adding 148.15 ppm to all reported shifts, one can even compare with the calculated shifts.
Unfortunately, the fluorine signals are hard to compare to the calculations. The chemical shifts were computed only for different solvation shells of the sodium ion. The solvation of the fluorine ion was neglected at this point. Therefore the corresponding measured signal cannot be compared to the calculated one.
The calculated signals of ^{23}Na and ^{19}F differ strongly from the measured ones. The error in the fluorine signal could stem from the method focusing on sodium solvation shells as mentioned above. The cause of the error for the sodium signals is unclear. The ^{1}H-signals however, agree nicely with the experiments, deviating only by 1-2 ppm.
The other two nuclei can be compared with the simulated spectrum, specifically the peak shift with increasing HF concentrations. The measured signals show a shift of 17.54 ppm towards the downfield for ^{23}Na and 4.55 ppm towards the upfield for ^{1}H. The sodium signal shows a multiplet in the calculations. I determined the centers of the multiplets and a corresponding shift of 18 ppm towards the downfield for sodium and 3 ppm for hydrogen. The locations of the peaks and a zoomed-in spectrum is shown in figure 5.26. These are comparable with the experimentally measured shifts of 17.54 ppm towards the downfield for sodium and 4.54 ppm towards the upfield for hydrogen.

TABLE 5.1: Chemical shifts of experimental NMR-spectra of NaF in water, aqueous HF and aHF against ^{1}H-TMS as reported by the experimentalists and against ^{19}F-$CFCl_3$ to better compare with figure 5.25.

% HF	^{23}Na vs ^{1}H-TMS	^{1}H vs ^{1}H-TMS	^{19}F vs ^{1}H-TMS
0	-0.36	3.76	-120.44
10	-2.80	4.35	-161.89
100	-17.90	8.31	-195.51
% HF	^{23}Na vs ^{19}F-$CFCl_3$	^{1}H vs ^{19}F-$CFCl_3$	^{19}F vs ^{19}F-$CFCl_3$
0	147.79	151.91	27.71
10	145.35	152.5	-13.74
100	130.25	156.46	-47.36

It remains unclear whether the solvation of sodium was correctly determined or not. The shifts of the calculated signals with increasing HF-concentrations as well as the locations of the ^{1}H-signals speak for it. The fact that the sodium signal errs by 500 ppm speaks strongly against it. Judging solvation shells in MD simulations with this method has the advantage that the dynamics are inherently included, since all observed coordination numbers with all observed constitutions of solvation shells are include. This is very laboursome as the chemical shifts needs to be computed for every solvation shell. However, if only the distribution of coordination numbers and their constitutions changes with changing MD-parameters, then many sets of parameters can be evaluated with the same chemical shifts.

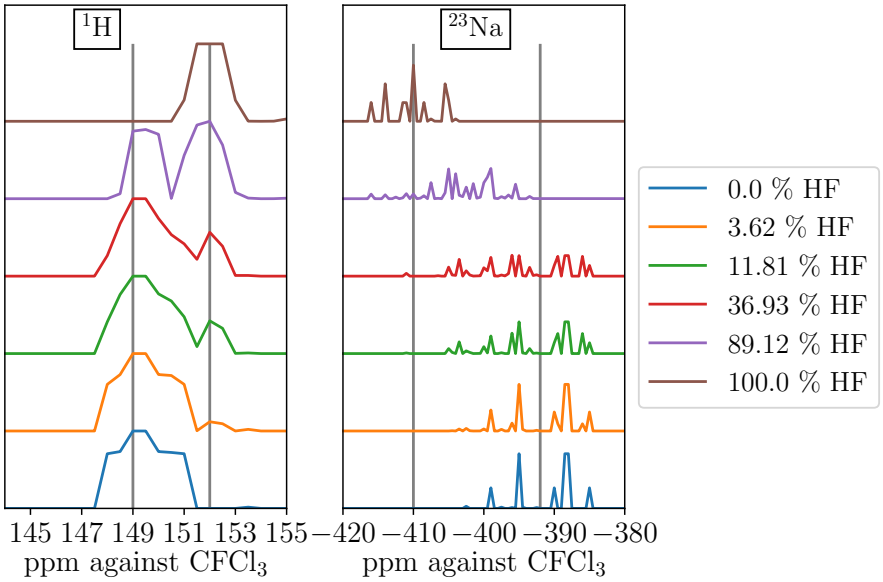

FIGURE 5.26: Calc. NMR-shifts against ^{19}F-$CFCl_3$ of 100 $[Na(H_2O)_x(HF)_y]^+$ with $x, y \in [0, 8]$, CN = {4;5;6;7;8} and 3 conformers per species. Zoomed in on hydrogen and sodium signals.

In this section, the quantum mechanical calculations of partial charges as well as the calculations of chemical shifts for differently constituted solvation shells were performed by Jennifer Anders. Furthermore, Jennifers Anders computed the entropy of sodium bifluoride to determine a reference solvation free energy of the substance in

HF. Dr. Patrick Pröhm measured the NMR spectra mentioned above.
I performed all mentioned MD simulations, constructed the thermodynamic cycle to compute the enthalpy of solvation, extracted the relevant quantities from the simulations, and compared the resulting parameter combinations. Furthermore, I provided the snapshots of differently constituted solvation shells of Na^+ in water, hydrofluoric acid, and HF and computed the spectra using the provided chemical shifts.

5.3 The Parameterization of HF with Three Sites

5.3.1 Approach

Many HF-models reported in literature make use of a virtual site X to improve the performance of the model [36–39]. The models are all constructed as F-X-H with $q_F = q_H = -2q_X$, i.e. all the negative charge is placed on the virtual site X. This allows the maximum of the negative charge to vary in position. The Hartree potential was computed using the PBE-functional and used as a reference. A classically computed potential was compared to the reference and the error Δ, as described in the methods section, is computed to find values for the charge q_F and the fluorine-virtual site distance R_{FX}. The term δ_Z was designed to find parameter combinations that yield the correct circle of maximum electrostatic potential around the pole of the H-F bond axis.
After the initial two parameters were set, the Lennard-Jones parameters were attempted to be determined like in the previous section. That is computing physical quantities for a variety of parameter combinations in hope of finding a suitable candidate.

5.3.2 Choice of Topology

The *ab initio* ESP is depicted in figure 5.27. The right panel of the figure depicts the same potential with a more detailed scale around the maximum of the potential. It reveals that the maximum of the potential lies not at the tip of the ellipsoid but rather forms a ring around it.

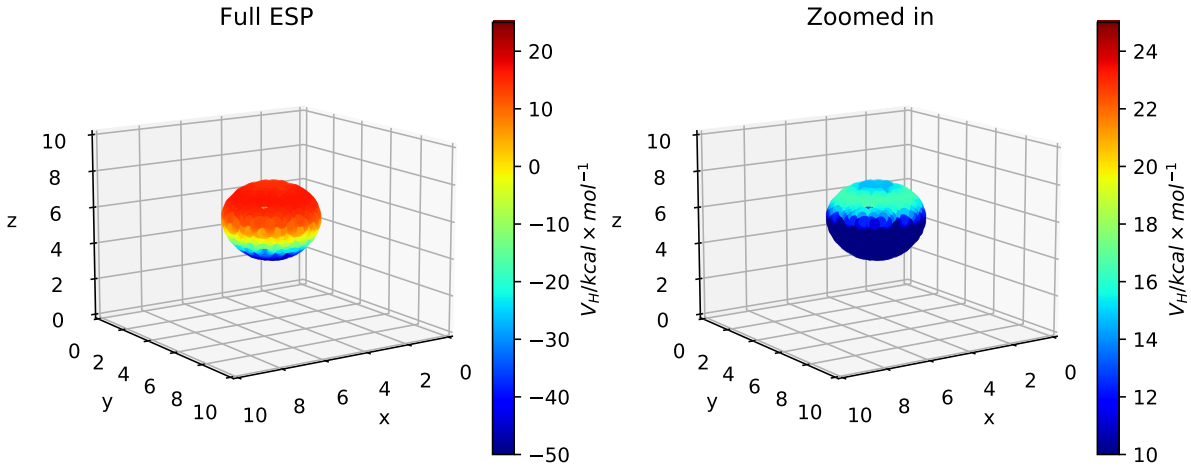

FIGURE 5.27: Electrostatic potential of HF, computed with the PBE functional.

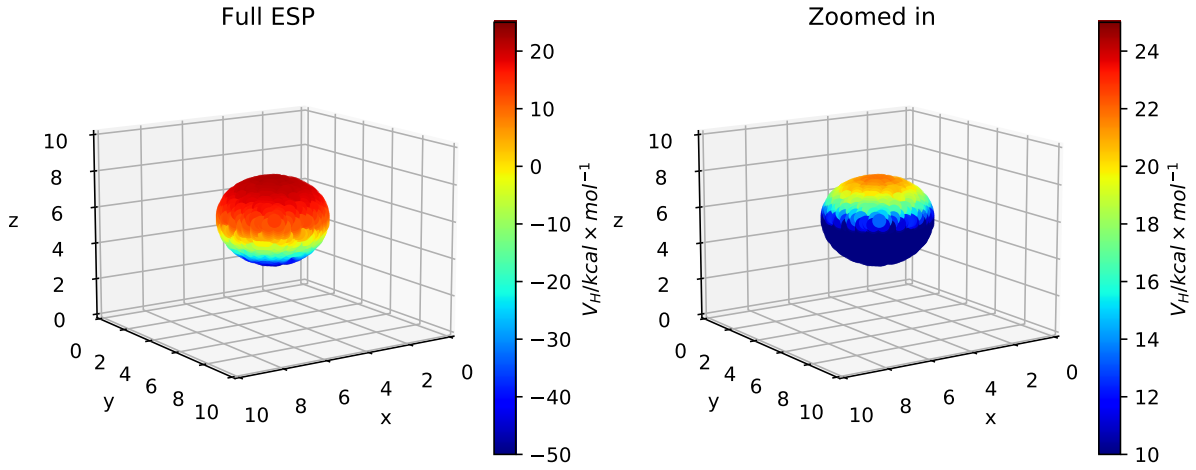

FIGURE 5.28: Classically computed electrostatic potential of HF, computed with the model of Orabi and Faraldo-Gomez [39].

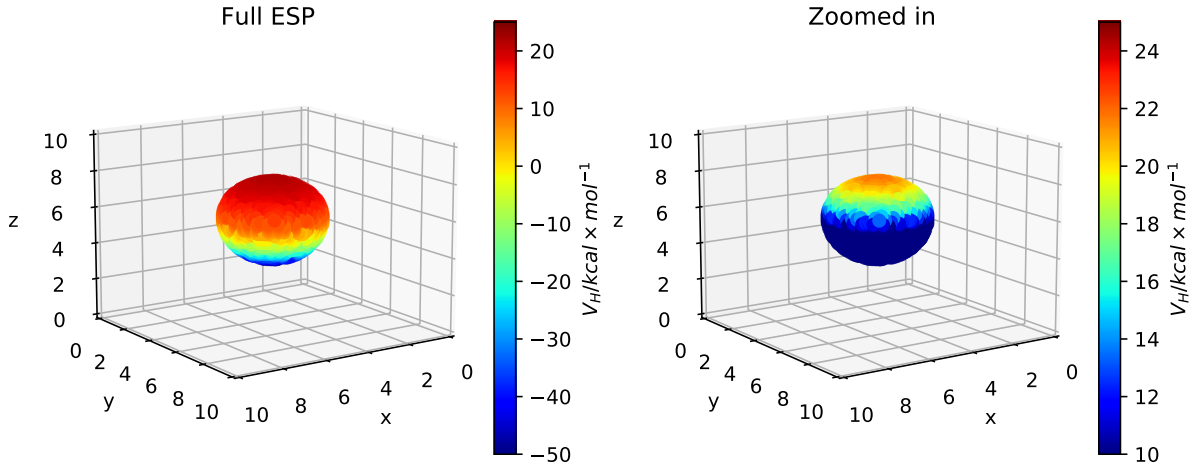

FIGURE 5.29: Classically computed electrostatic potential of HF, computed with the model of Cournoyer and Jorgensen [36].

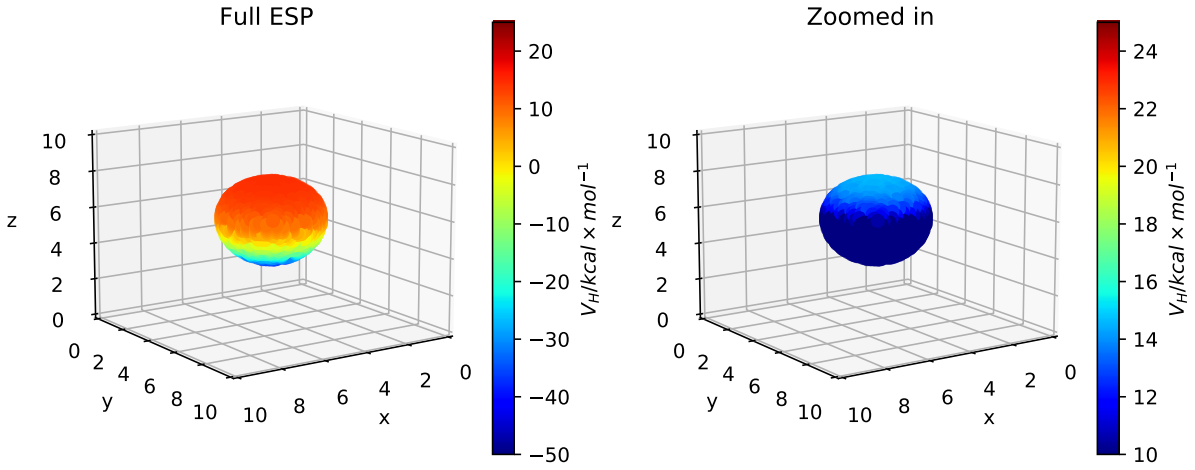

FIGURE 5.30: Classically computed electrostatic potential of HF, computed with the model of Jedlovszky and Vallauri [37].

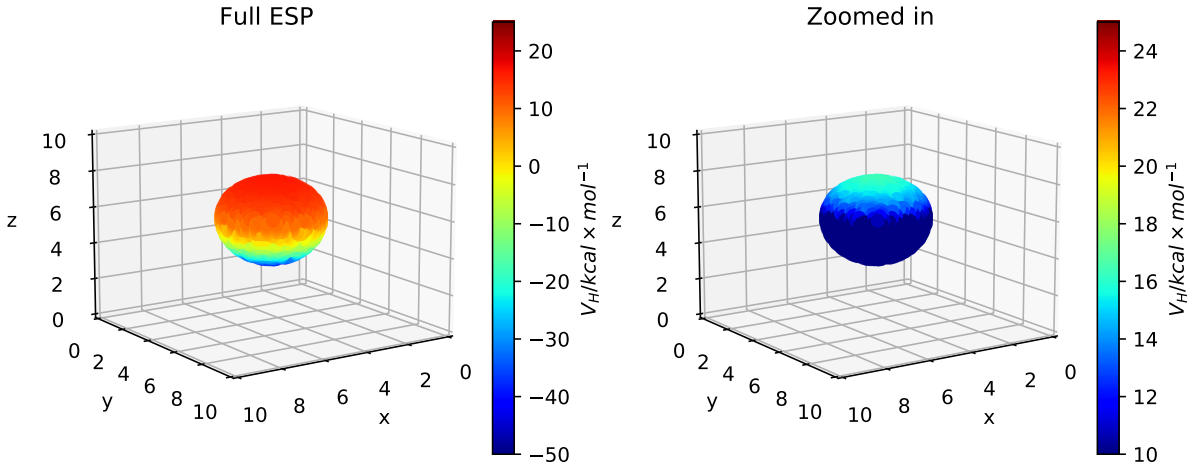

FIGURE 5.31: Classically computed electrostatic potential of HF, computed with the model of Kreitmeir [38].

The ESP-plots in figures 5.28, 5.31, 5.29 and 5.30 all fail to show the ring of maximum potential, instead the maximum lies on the pole of the ellipsoid.

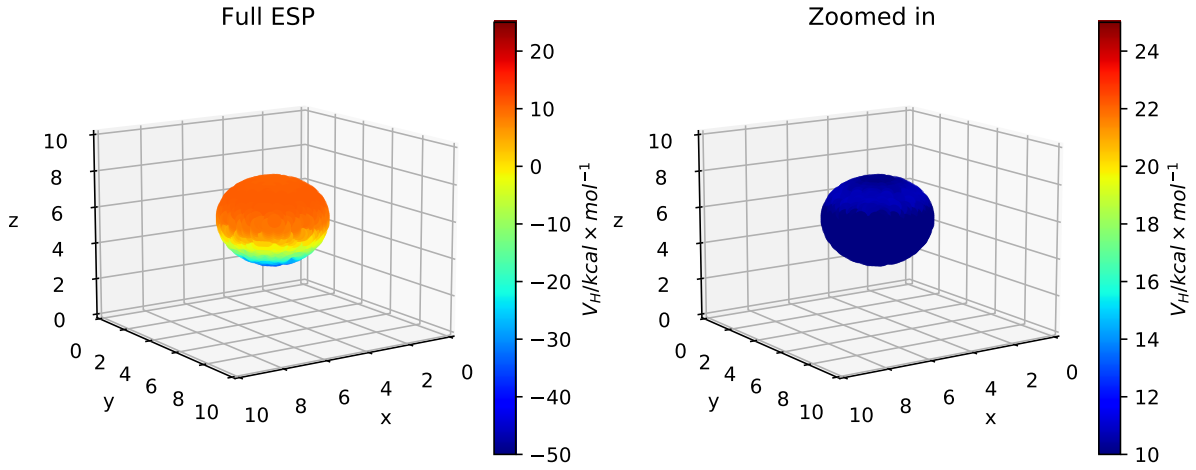

FIGURE 5.32: Classically computed electrostatic potential of HF, computed with the model determined in this work.

The model determined in this work, depicted in figure 5.32 shows a similar feature as the PBE surface. The difference between the ring of maximum value and the value at the pole is not as large as the reference. The range of values is also smaller than the reference. The fact that our work model is the only one depicting the ring of maximum value around implies that the term δ_Z works as intended. The parameters for the different models are shown in table 5.2.

TABLE 5.2: Overview of topologies of 3-site HF-models.

parameter	this work	OFG model [39]	CJ model [36]	JV model [37]	K model [38]
r_{FH} / Å	0.917	0.917	0.917	0.973	0.950
r_{FX} / Å	0.26262626	0.141218	0.166	0.1647	0.1608
q_F / e	0.67272727	0.705	0.725	0.592	0.592
q_H / e	0.67272727	0.705	0.725	0.592	0.592
q_X / e	-1.34545454	-1.410	-1.450	-1.184	-1.184

5.3.3 Radial Distribution Functions

Radial distribution functions were computed for pure HF at 273 K. The best five parameter combinations to reproduce the g_{HH} are shown below in figure 5.33.

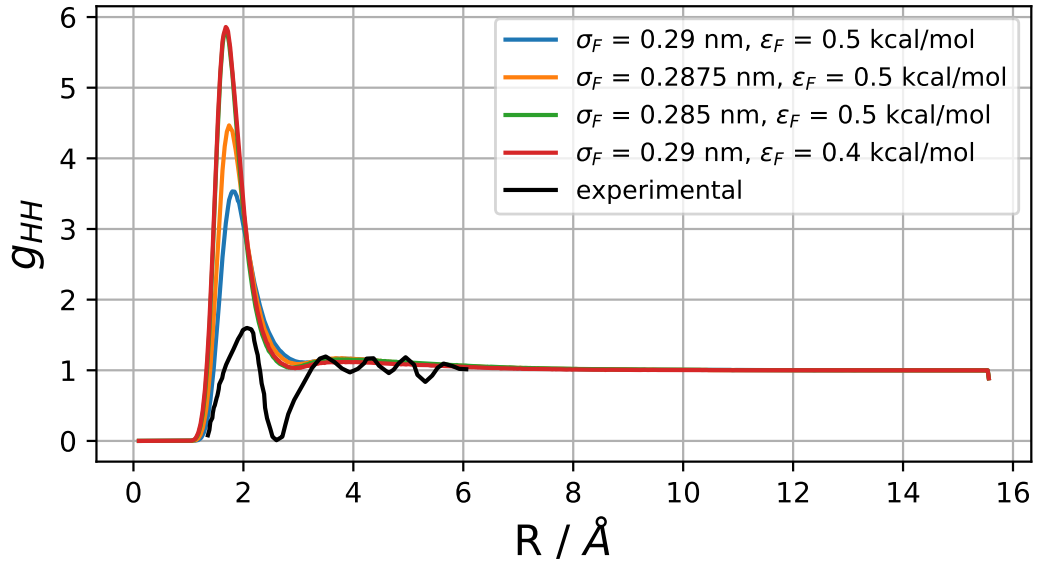

FIGURE 5.33: Experimental H-H RDF and simulated RDF of the five best parameter combinations of the three-site model.

The first peak does not have the same peculiar shape as the experimental curve, however, the width of the peak matches the experiment. This finding indicates that the range of H-H distances has been improved, compared to the two-site model. The following peaks have not been resolved, which suggests that shorter chains occur in the simulations compared to the experiment and the 2-site model.
The five best reproduce F-H distance distributions are shown in figure 5.20. Just like the 2-site model, there are three peaks in the radial distribution function. The first and second peak locations are comparable to the 2-site model. The third peak however, lies at shorter distances and closer to the third peak of the experimental spectrum. Therefore, the performance is slightly improved compared to the 2-site model.

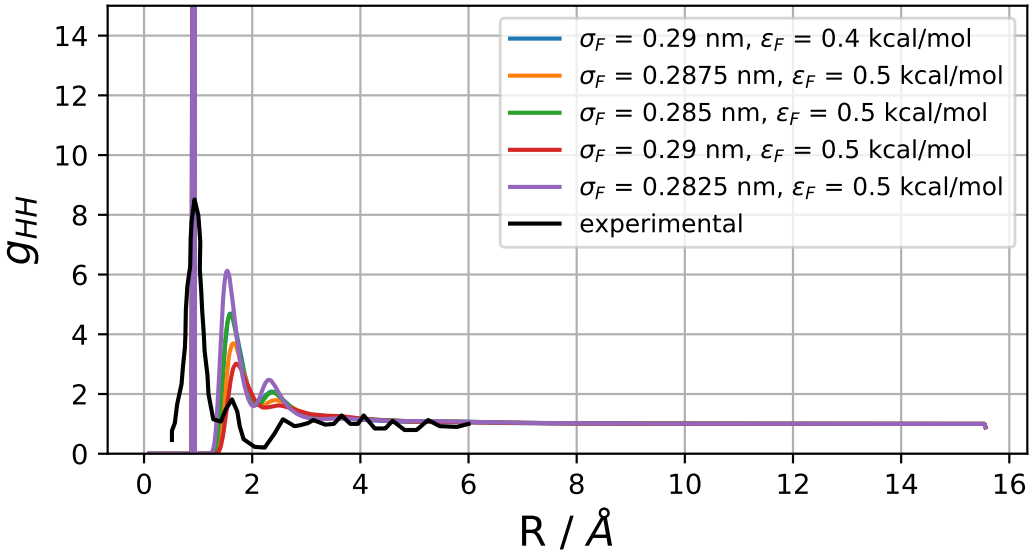

FIGURE 5.34: Experimental F-H RDF and simulated RDF of the five best parameter combinations of the three-site model.

The F-F radial distribution functions for the five best parameter combinations are shown in figure 5.21. Similar to the H-H RDF, one broad peak can be observed. The second peak is very shallow and barely visible. In comparison to the 2-site model, the performance is worse. The peak heights do not match as well, the same is true for the peak locations. Qualitatively the RDFs look similar to the ones observed for the 2-site model.

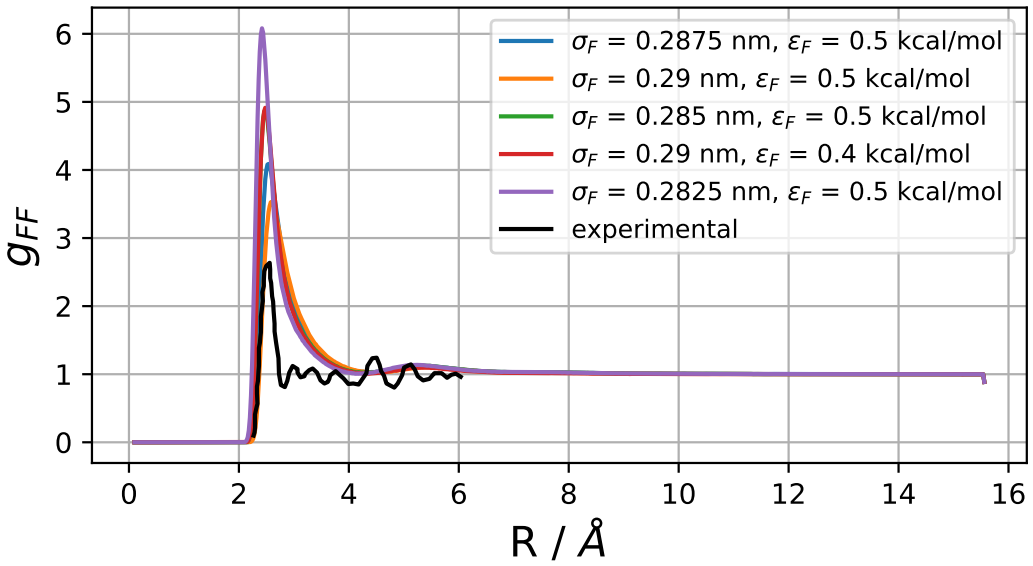

FIGURE 5.35: Experimental F-F RDF and simulated RDF of the five best parameter combinations of the three-site model.

5.3.4 Concluding Remarks

Radial distribution functions were the only quantity that could be extracted for the three-site model. The density curves could not be computed because for no parameter combinations could all concentrations be simulated successfully. The project was discontinued at this point and it was decided to use the model of Orabi and Faraldo-Gomez [39] from here on.

5.4 Validating the HF-model

The HF model reported by Orabi and Faraldo-Gomez (OFG) was implemented in GROMACS and simulated to check whether the results reported in [39] could be reproduced. An overview of this comparison is shown below in table 5.3. Most values

TABLE 5.3: Comparison of the HF-model performance reported in the literature with my implementation in GROMACS.

Property	experimental	OFG model (lit.) [39]	OFG model (own simulation)
Diffusion constant $D \times 10^{-9}$ / ms^{-1}	10.0 ± 1.5 [120]	11.7	12.03
Isothermal compressibility β / MPa^{-1}	5.87 [121]	5.0	0.0041
Dielectric constant ϵ / 1	83.6 [122]	34.2	33.62
Density ρ / gcm^{-3}	0.952 [123], 0.962 [124], 0.964 [125], 0.991 [126]	0.987	0.982
r_{max}(gFF) / Å	2.52 [119]	2.63	2.64
r_{max}(gFH) / Å	1.60 [119]	1.73	1.72
r_{max}(gHH) / Å	2.03 [119]	2.40	2.38

of my simulation lie close to the values reported in the literature. Only one discrepancy arises, the value of the isothermal compressibility. In [39] the authors reported the literature value and their computed value for the isothermal compressibility β to be 5.87 and 5.0 MPa^{-1} respectively. This is likely an error in the reported units.
In [121], Lagemann reports an isothermal compressibility of β =587.5 $\times 10^{-12}$ $\frac{cm^2}{dyne}$. Since 1 $\frac{cm^2}{dyne}$ = 10 $\frac{1}{Pa}$, I obtain $\beta = 4.1 \times 10^{-9}$ $\frac{1}{Pa}$ = 4.1 $\frac{1}{GPa}$. This value is also much closer to the values of water and ethanol which are $\beta_{H_2O} = 0.46$ $\frac{1}{GPa}$ [127] and $\beta_{EtOH} = 1.1$ $\frac{1}{GPa}$ [128] respectively.
So far only the performance for pure HF can be judged. However, aqueous HF is a system of interest too. Three physical properties of aqueous HF at various concentrations were computed and compared to available experimental data, the results are displayed in table 5.4. The densities lie in the same order of magnitude and are close to the experimental values. Furthermore, the qualitative trend of finding a local maximum at around 80% HF could also be recreated. The computed diffusion constants are in the same order of magnitude as the available experimental data. Due to a lack of data points, the trend of monotonously increasing diffusion constants with increasing concentration of HF is difficult to judge. For the heat capacities at constant pressure, I find that the computed values are in the same order of magnitude as the experimental ones. However, the parabolic trend found in the experimental data, could not be recreated.

TABLE 5.4: The performance of mixed models. The experimental density values were taken from a polynomial fit to experimental data obtained at 299.8 K. The heat capacities were measured at 300 K. The simulations were carried out at 296 K and 1.199688 bar.

% HF	ρ/ g$\times$ cm^{-3} (Expt.)	ρ/ g$\times$ cm^{-3} (Comp.)	$D_{\mathrm{HF}} \times 10^{-9}$ / m^2/s (Expt.)	$D_{\mathrm{HF}} \times 10^{-9}$ / m^2/s (Comp.)	C_P / J/K/mol (Expt.)	C_P / J/K/mol (Comp.)
00	0.989 [107]	0.998 $\pm$0.011	2.299 [129]	1.91 $\pm$0.10	83.79 [107]	88.48
20	1.064 [107]	1.026 $\pm$0.012	1.835 [130]	2.47 $\pm$0.45	72.55 [107]	80.018
40	1.131 [107]	1.039 $\pm$0.015	n.A.	3.71 $\pm$0.49	64.68 [107]	81.70
60	1.120 [107]	1.054 $\pm$0.017	n.A.	4.74 $\pm$0.16	62.14 [107]	74.86
80	1.192 [107]	1.055 $\pm$0.021	n.A.	7.37 $\pm$0.52	64.49 [107]	68.60
100	0.952 [123], 0.962 [124] 0.964 [125], 0.991 [126]	0.980 $\pm$0.031	10.0 $\pm$ 1.5 [120]	11.61 $\pm$0.78	71.13 [131], 50.6 [132]	60.92

The properties of aqueous HF could be recreated sufficiently well. Densities are most satisfying. Diffusion constants are somewhat hard to judge due to a lack of experimental data. The heat capacities follow a different trend, compared to what is found experimentally. Overall, all investigated quantities lie in the correct order of magnitude.

5.5 Exploring the Parameter Space of Tetravalent Cations in Aqueous HF

5.5.1 Testing the New Method

The novel method of determining coordination polyhedrons will be applied in this section. It comes with one caveat: The number of Steinhardt-parameters used for computing the Euclidean norm in the space of Steinhardt-parameters called the polyhedron distance from hereon, affects the reference structure which minimizes the polyhedron distance. This is shown in figure 5.36.
The three reference structures all contain six ligands. Increasing the number of Steinhardt parameters in the norm, increases the polyhedron distance for the octahedron so much, that the polyhedron distance for pentagonal pyramid becomes smaller. This means that the structure which minimizes the polyhedron distance is not the actual structure of the complex if too many Steinhardt parameters are used. Luckily, the tolerance for this effect to happen is rather large, if the number of Steinhardt parameters is around ten, one should be able to obtain correct predictions. The reason

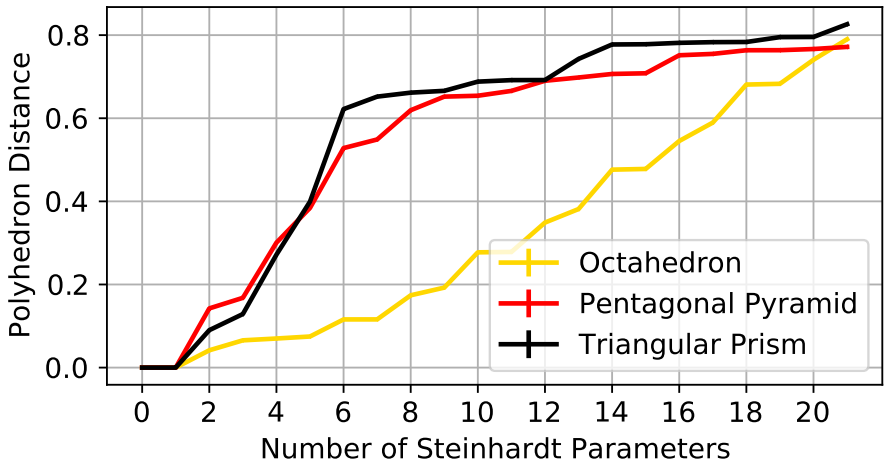

FIGURE 5.36: Polyhedron distance towards three reference structures for a truly octahedral structure vs the number of Steinhardt parameters included in the polyhedron distance.

behind this phenomenon is likely a cause of the fact that norms increase with an increasing amount of dimensions. That is

$$D = \left(\sum_{i=1}^{L} (Q_i - Q_i^{ref})^2 \right)^{(1/2)} \tag{5.3}$$

the polyhedron distance D, gains terms for each of new Steinhardt-parameter included, therefore D increases with increasing L.

5.5.2 The Parameter Space

Previous attempts for parametrizing hafnium and zirconium ions mainly focused on reproducing the hydration free energy (HFE) or the ion-oxygen distance (IOD). Two authors published Lennard-Jones parameters for the relevant ions, Zhang et al in [45] and Merz and his coworkers in [43, 44]. The resulting parameter combinations

of both publications are shown in figure 5.37. While the results of the group of Merz reproduced either HFEs or IODs with different parameter combinations, the results by Zhang et al recreated both simultaneously. However, the parameter combinations which could recreate both quantities lie at extreme locations in the parameter space. That is $\epsilon_F > 10^3$ kcal/mol and $\sigma_F < 0.5$ Å. This implies an extremely small and extremely polarizable fluoride ion, which stands in contrast to the fact that fluoride is the hardest anion of all elements. Furthermore, both publications span a range of 9 orders of magnitude for ϵ_F. The models form three clusters based on the quantities which were reproduced. That is, the difference between parameterization targets is larger than the difference between zirconium and hafnium.

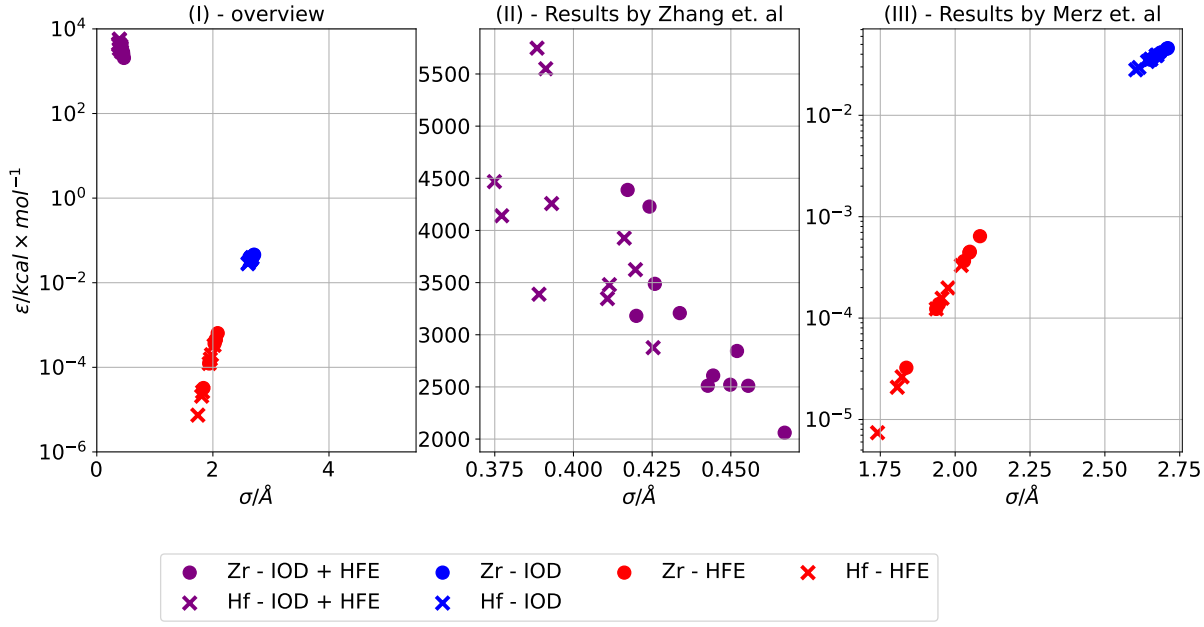

FIGURE 5.37: (I): Overview of Lennard-Jones parameters for zirconium and hafnium found in the literature. Several markers correspond to parameter combinations for several water models. Points marked with HFE or IOD refer to parameter combinations fit to reproduce the hydration free energy or ion-oxygen distance respectively. (II): A more detailed representation of the results published in [45]. (III): More detailed representation of the results published in [43, 44].

The entire parameter space depicted in figure 5.37 was sampled and the static solvation shell properties were evaluated. The static solvation shell properties include coordination numbers (CN), fluorine coordination numbers (CNF), coordination polyhedra (CP), and ion-oxygen distances (IODs).

5.5.3 Coordination Numbers

The mean coordination numbers in the first solvation shell and corresponding standard deviations are depicted in figure 5.38. First and second shell coordination numbers were computed as the number of oxygen and fluorine atoms within thin the first and second minimum of the radial distribution function respectively. Most coordination numbers are four or smaller, this indicates that the metal ions could be coordinated by fluoride ions. Furthermore, one can observe that similar values of coordination numbers lie on coherent lines in the parameter space. These lines appear as diagonals in the semi-logarithmic plots.

The first shell coordination numbers are relatively stable since most standard deviations are zero. The fluorine coordination numbers were computed and the results

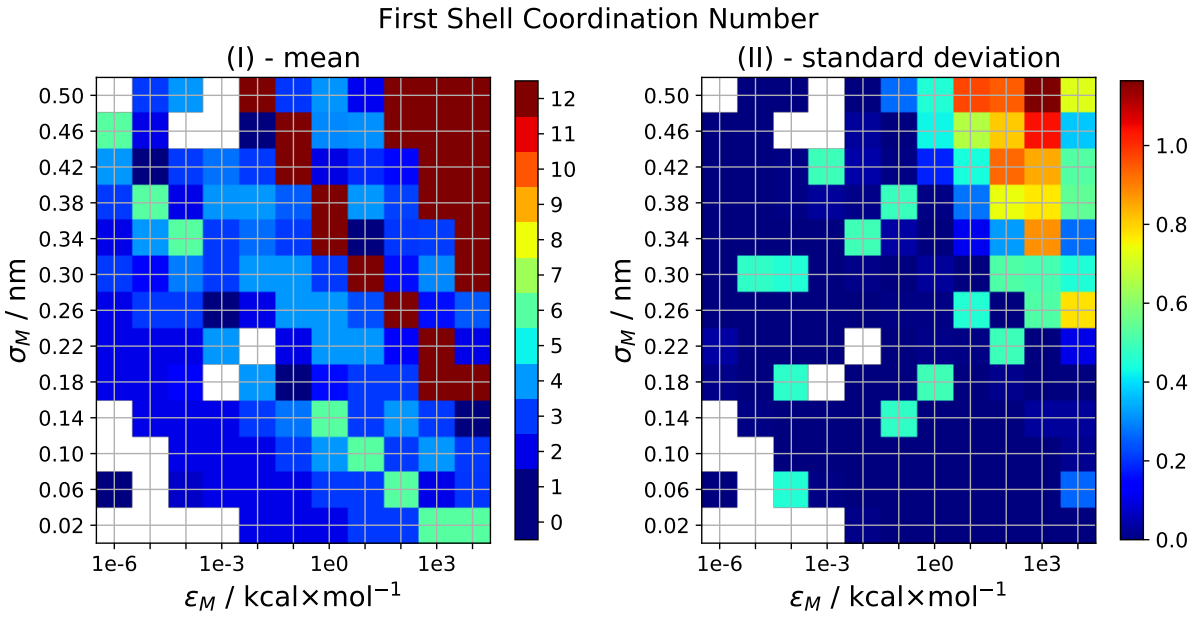

FIGURE 5.38: Overview of mean coordination numbers within the first coordination shell (left) and corresponding standard deviations (right). CNs larger than 12 were colored red for clarity.

depicted in figure 5.39 confirm the suspicion that the low coordination numbers of the first solvation are due to fluoride ion coordinating to the metal cation.
This also becomes evident in the similarity of the corresponding standard deviation. Interestingly, the parameter combinations which yielded CN > 4 are free of fluoride ions. The second solvation shell CNs were evaluated as well. The results are depicted in figure 5.40. In this plot, the trend of diagonal lines of similar coordination numbers becomes more evident. Furthermore, the coordination numbers increase with increasing ϵ_F and increasing σ_F.

5.5.4 Coordination Polyhedra

The coordination polyhedra (CPs) were determined based on the first or second solvation shell. The second shell was considered if the first shell contained less than four particles. The results are shown in figure 5.41. Just like the CNs, the CPs lie on diagonal lines in the parameter space. The most prominent polyhedra are octahedrons, tetrahedrons, and icosahedrons.

5.5.5 Ion Oxygen Distances

The radial distribution functions were evaluated to determine the ion-oxygen distances. Precisely it was the distances between the metal cation and fluoride ion or water oxygen atoms. The results, in terms of first and second shell peak and minimum locations, are shown in figure 5.42. The trend of lines of similar values continues for this property as well.
One can observe a qualitatively similar trend for all the discussed properties, that is diagonal lines in the semi-logarithmic parameter space. This trend is extremely

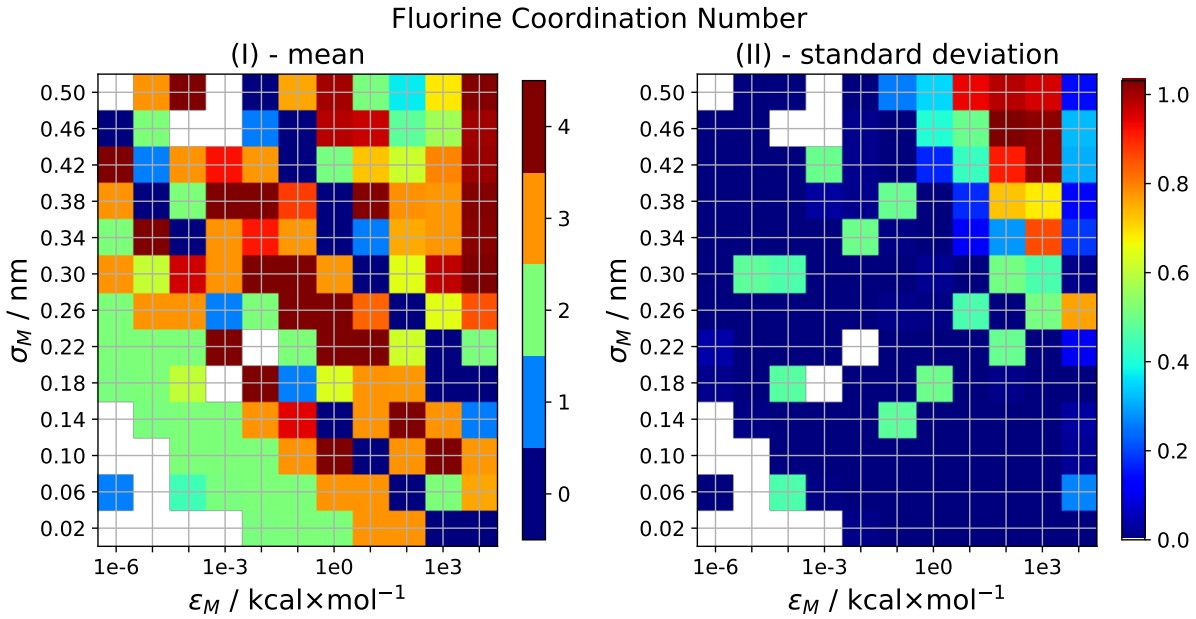

FIGURE 5.39: Overview of mean coordination numbers of fluorine atoms within the first coordination shell (left) and corresponding standard deviations (right).

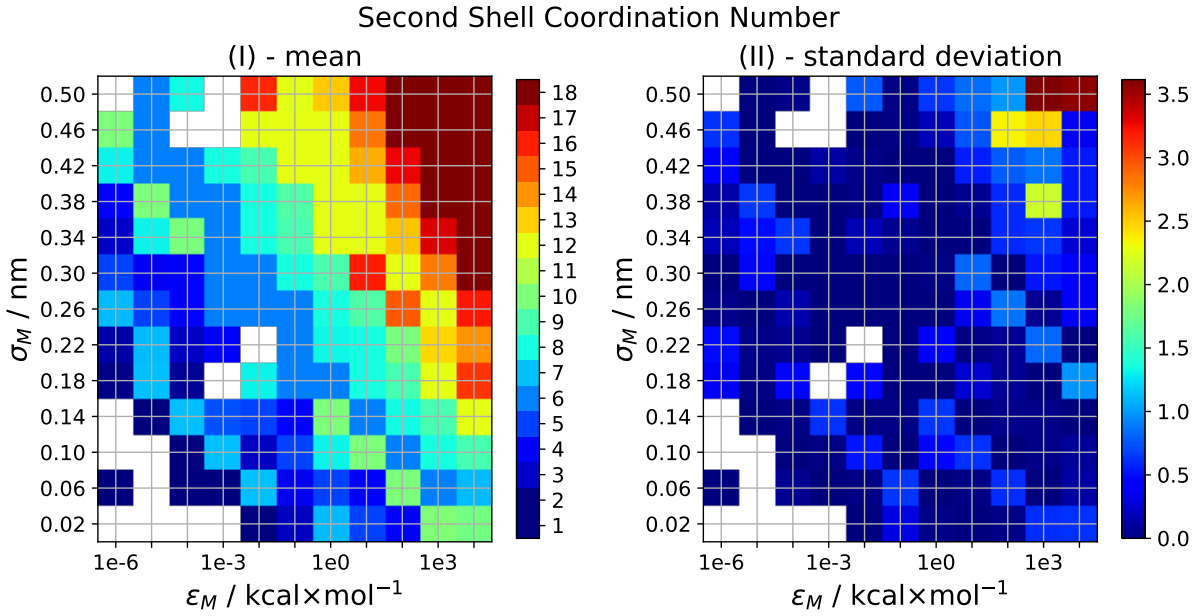

FIGURE 5.40: Overview of mean coordination numbers within the second coordination shell (left) and corresponding standard deviations (right). CNs larger than 18 were colored red for clarity.

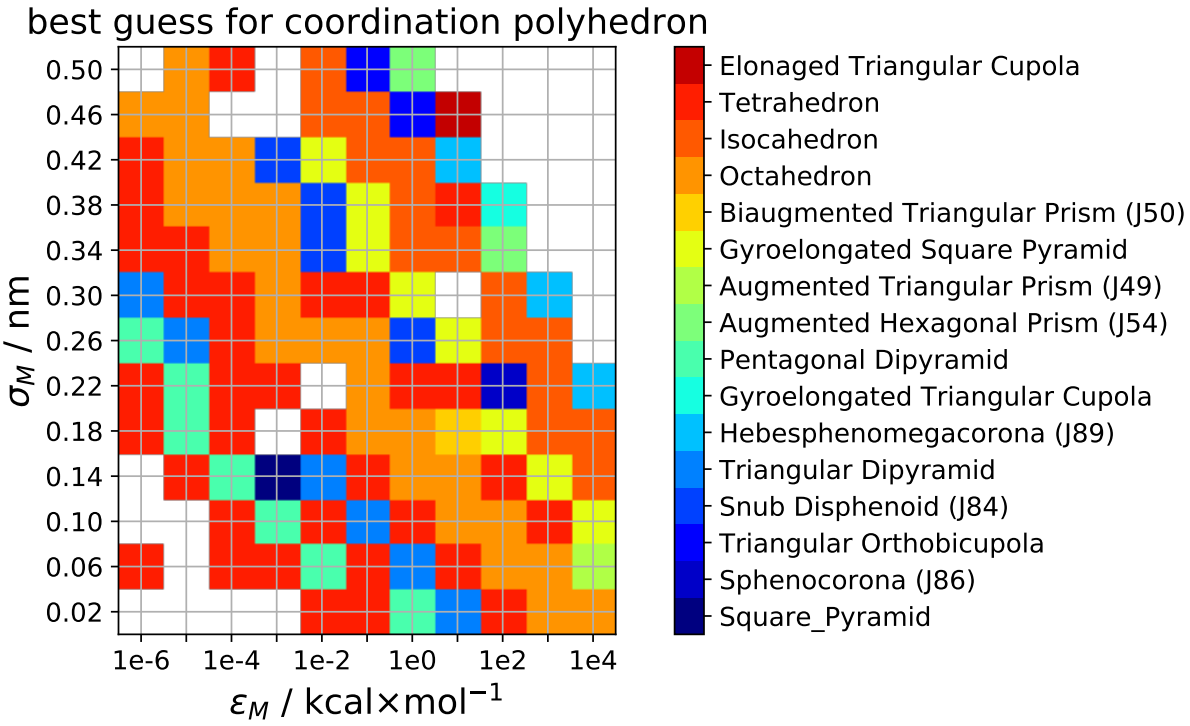

FIGURE 5.41: Best guess for coordination polyhedra of either the first or the second solvation shell. The second solvation shell was chosen if the first solvation shell contained less than four particles. Coordination shells with more than 16 particles were excluded to keep the color bar clear.

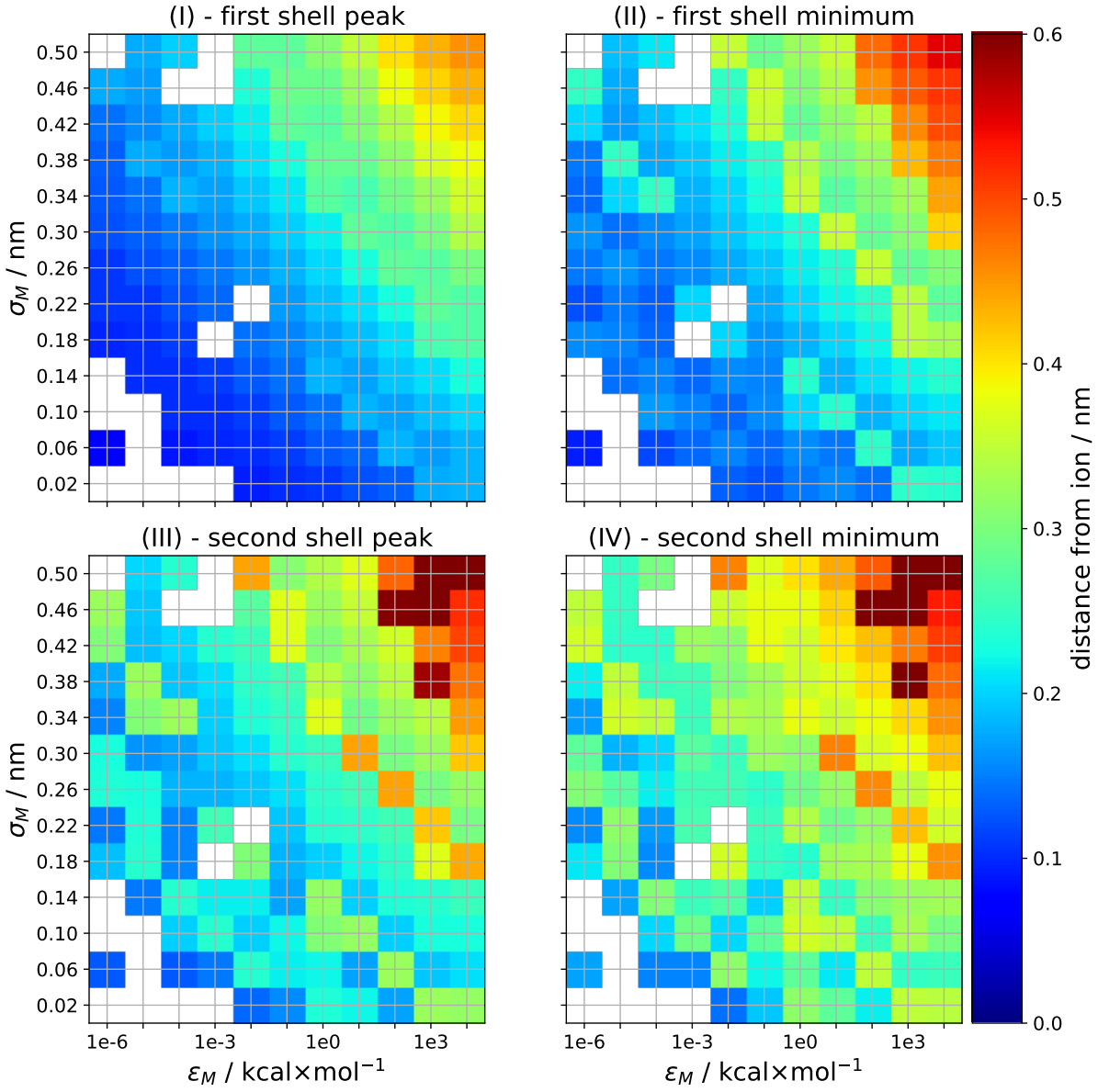

FIGURE 5.42: Overview of locations of the first two maxima and minima of the cation-solvent radial distribution function.

robust and does not break over 9 orders of magnitude for ϵ_F and σ_F up to 5 Å. However, not all quantities follow these trends. Zhang et al provided similar plots for the hydration free energy in [45]. There it can be seen that the HFE lies on curved lines in the semi-logarithmic parameter space. The results presented herein provide a basis for future parameterization efforts of tetravalent metal cations in general.
In order to investigate the thermal contractions of the solvation shells of Zr^{4+} and Hf^{4+}, one needs to find Lennard-Jones parameters that yield an octahedral structure with two fluoride ions, two hydroxide ions and two water molecules. Since hydroxide ions are not present in these simulations, one can look for two fluoride ions in an octahedral structure.
The results in this section show that parameter combinations which fulfill the criteria of fluoride-content and coordination polyhedron exist on diagonal lines in the semi-logarithmic parameter space. Therefore, a whole set of suitable parameter combinations exists which fulfill the two criteria.

5.6 Simulating Geochemical Twins Under Extreme Conditions

5.6.1 Parameter Scan

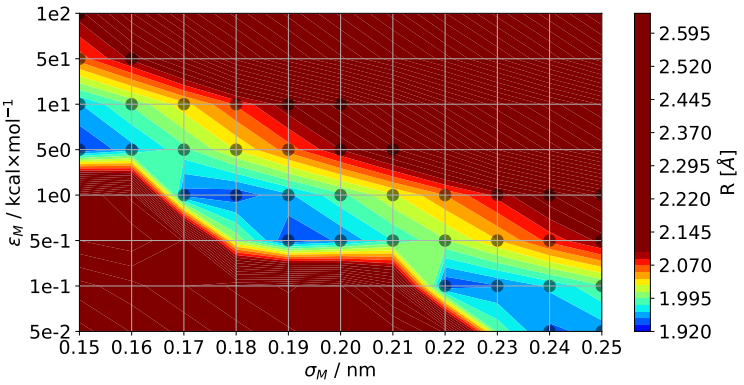

FIGURE 5.43: Mean distances of the six F- or O-atoms which are closest to the metal ion. The black dots indicate octahedral structures. The averaging is performed over all atoms and all timeframes. The colorbar was chosen to differentiate only in the relevant space, i.e. between 1.9 Åand 2.1 Å.

To extend the previous section, I simulated a part of the parameter space with a finer resolution and under conditions similar to the experimental conditions described in section 4.3.2. The results are shown in figure 5.43. Highlighted are all parameter combinations that yield octahedral structures with mean cation-ligand distances matching the experimentally measured range. The suitable values lie on a diagonal in the parameter space, as discussed in the previous chapter. I chose two parameter combinations from this diagonal to investigate the thermal contractions of the solvation shells. Additionally, one set of parameters which has been published recently [45] was included. The Lennard-Jones potentials and overall interaction potentials for the cation-fluoride and cation-hydroxide oxygen interactions for the three parameter sets are shown in figure 5.44.

5.6.2 Stabilities of the Complex Configurations

There are five possible configurations for an octahedral complex with the composition $[M(OH)_2F_2(H_2O)_2]^0$. Figure 5.45 shows these five complex structures. I simulated each of the five complexes in 1 M aqueous HF and temperatures ranging from 25 °C to 400 °C. To reproduce the pressure of $p = 400$ bar I adjusted the density of the simulation to the corresponding density of water at these conditions (see Method section). This temperature scan was repeated for all three Lennard-Jones parameters for the M^{4+} cation.
The stability of the five complex structures is summarized in Tab. 5.5. The parameter combination $\sigma_M = 0.047$ nm and $\epsilon_M = 4388.337$ kcal/mol was taken from [45]. Overall, the anions are less likely to leave the complex, whereas one observes that water molecules exchange with the surrounding solvent, especially at high temperatures.

TABLE 5.5: Stability of $[M(OH)_2F_2(H_2O)_2]^0$ in the MD simulations.

No.	name	coordination sphere		
		$\sigma_M = 0.36$ nm $\epsilon_M = 0.001$ kcal/mol	$\sigma_M = 0.18$ nm $\epsilon_M = 5$ kcal/mol	$\sigma_M = 0.047$ nm $\epsilon_M = 4388.337$ kcal/mol
1	all trans	Completely stable up to and including T=200 °C, at higher temperatures the water molecules exchange. The ions remain bound for all temperatures.	The ions remain stable at all temperatures, the water molecules for T=350 °C and above.	Stable at all temperatures.
2	F^--trans	Completely stable up to and including T=100 °C, at higher temperatures the water molecules exchange. The ions remain bound for all temperatures.	The ions remain stable at all temperatures, the water molecules for T=300 °C and above.	Stable at all temperatures, except T=400 °C where one water molecule exchanges.
3	OH^--trans	Completely stable up to and including T=100 °C, at T=200 °C one water molecule begins exchanging, and at higher temperatures both water molecules exchange. The ions remain bound for all temperatures.	The ions remain stable at all temperatures, the water molecules for T=300 °C and above.	Stable until T=350 °C, where one water molecule exchanges.
4	H_2O-trans	Completely stable up to and including T=200 °C, at higher temperatures both water molecules exchange. The ions remain bound for all temperatures.	The ions remain stable at all temperatures, the water molecules for T=350 °C and above.	Stable at all temperatures.
5	all cis	The fluoride ions remain stable. At every temperature below T=300 °C, one hydroxide ion leaves. The water molecules exchange at all temperatures.	The fluoride ions remain stable. One hydroxide ion leaves at each temperature, except for T=350 °C and T=400 °C where both hydroxide ions remain. The water molecules exchange at all temperatures.	Fluoride ions remain stable. Hydorixde leaves for all temperatures. Water exchanges for all temperatures.

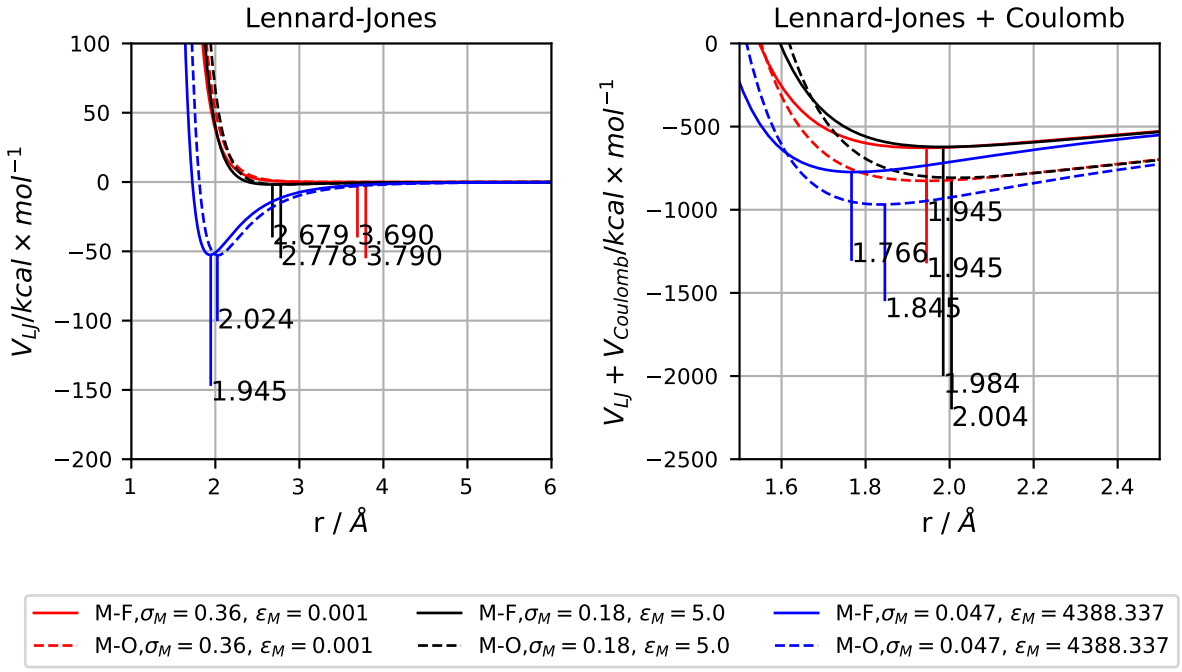

FIGURE 5.44: Interaction potentials between the tetravalent cation and the fluoride-ion (solid lines) and hydroxide-oxygen (dashed lines) respectively. The r-values of the minima are shown explicitly.

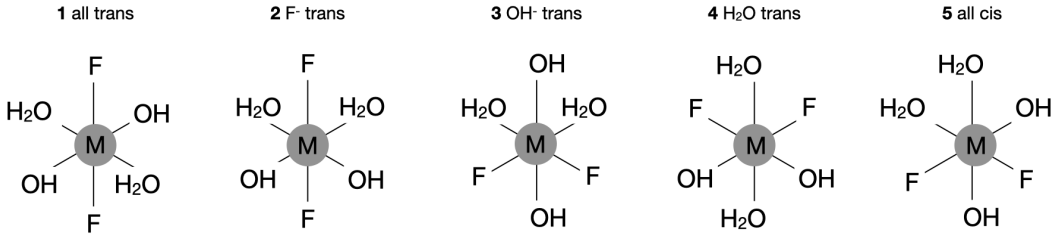

FIGURE 5.45: Possible coordination spheres of the complex $[M(OH)_2F_2(H_2O)_2]^0$, figure made by B. Keller.

5.6.3 Cation-Ligand-Distances

The mean distances of the six closest ligands are shown in figure 5.46. Because the water molecules exchange, the six closest ligands, and not the six initially bound ligands, are used to determine the mean distances to compare to the experimental values. The parameter set from [45] (blue line) underestimates the average ligand distance by almost 10%. The other two parameter sets (red and black lines) yield average ligand distances that are in the same range as the experimental values. One observes a dramatic thermal expansion for $\sigma_M = 0.36$ nm and $\epsilon_M = 0.001$ kcal/mol, contrary to the experimental findings. The other two parameter combinations show thermal expansion too, albeit of lesser magnitude.
To better understand the unusual thermal expansion of $\sigma_M = 0.36$ nm and $\epsilon_M = 0.001$ kcal/mol, I analyzed the distance distributions of the individual ligands (columns 1 to 6 in figure 5.47) for each of the five complex structures (rows 1 to 5 in figure 5.47). There exists a thermal contraction for the anions in the complex (r_1 to r_4). The two water molecules (r_5 to r_6) oscillate between a tightly bound conformation with $r \approx 2.1$ Å and a loosely bound conformation with $r \approx 3.8$ Å. That is, at high

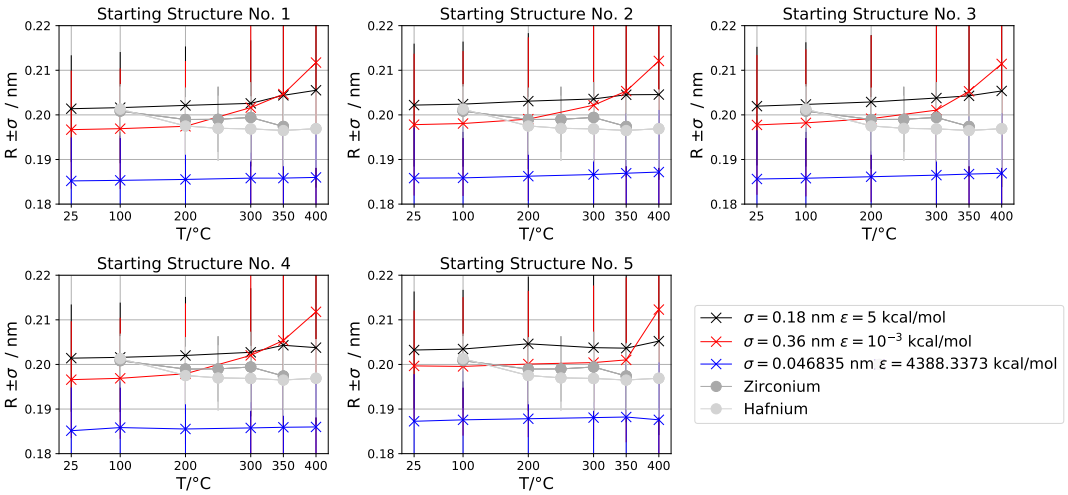

FIGURE 5.46: Mean distances of the six F- or O-atoms that are closest to the metal ion. The error bars represent the standard deviation. The averaging is performed over all atoms and all timeframes. The grey curves correspond to the experimental values measured by A. Loges. Some error bars are truncated for better visibility.

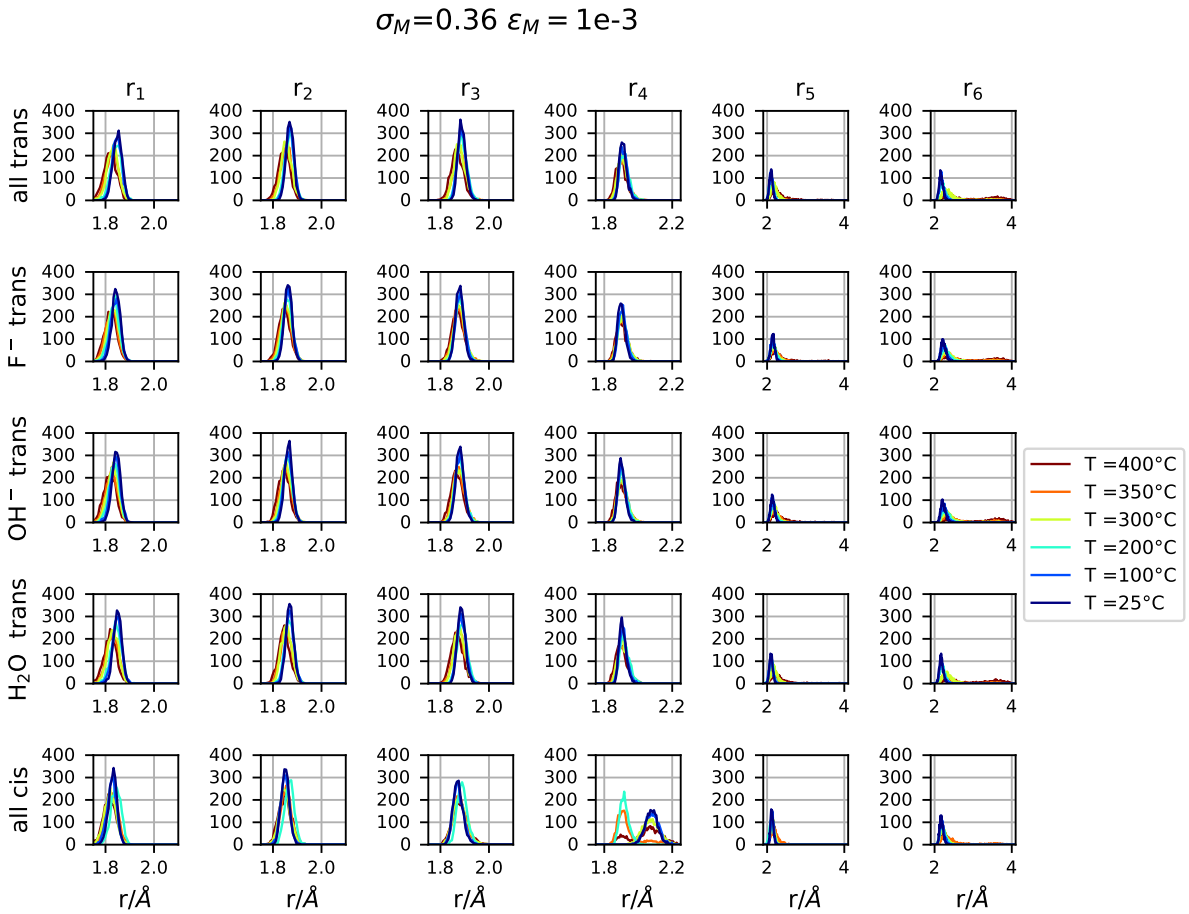

FIGURE 5.47: Probability densities of ion-ligand distances, for the six closest ligands, for each starting structure and temperature. r_i denotes the distance of the i^{th} ligand closest to the ion.

temperatures, the complex oscillates between a compact and an elongated/distorted octahedral structure. The all-cis complex structure 5 shows a slightly different behavior with an abrupt contraction for ligand 4. However, since the hydroxide ions leave the complex, one sees the distances for two different kinds of ligands in these subplots.

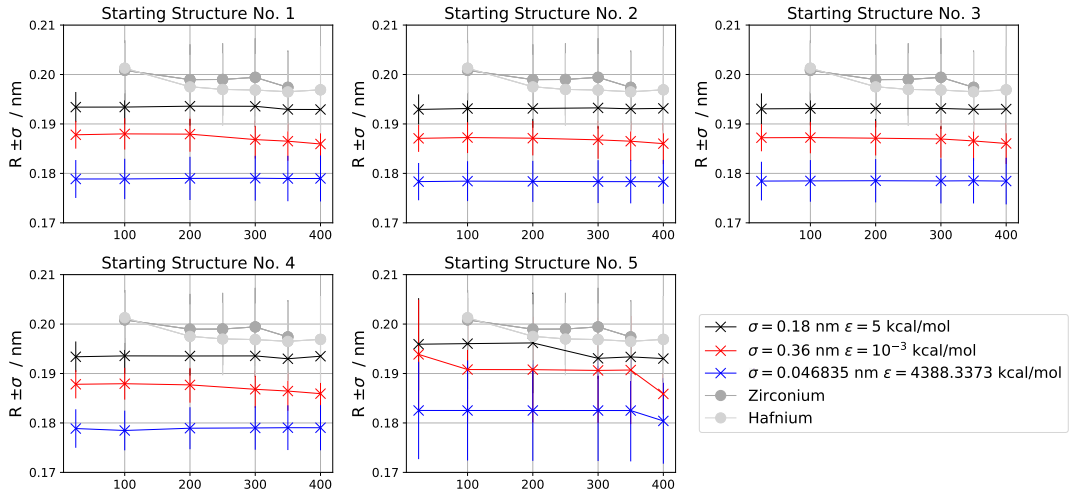

FIGURE 5.48: Mean distances of the four F- or O-atoms which are closest to the metal ion. The error bars represent the standard deviation. The grey curves correspond to the experimental values measured by A. Loges. Some errorbars are truncated for better visibility.

Figure 5.48 shows the mean diistances of the four closest ligands as a function of temperature. The parameters $\sigma = 0.36$ nm and $\epsilon = 0.001$ kcal/mol (red line) show a thermal contraction that is in the same range as the experimentally observed thermal contraction.

5.6.4 Cation-Ligand Distance Correlations

The results so far suggest the following mechanism: The water molecules are less tightly bound than the anions. At higher temperatures, they can fluctuate between a short distance to the M^{4+} cation (octahedral complex) and a larger distance to the M^{4+} cation (distorted octahedral complex). In the distorted octahedral complex, the anions can shift closer to M^{4+} cation leading to a thermal contraction of these four ligands.
To test whether this mechanism indeed takes place, I computed the Pearson correlation coefficients for several combinations of ion-ligand distance. The pairs were chosen such that we include correlations between two ligands of the same type e.g r_1 and r_2 which are likely to be the fluoride ions, as well as one interaction for each combination of different ligands, e.g. r_1 and r_3 which correspond to fluoride ion and hydroxide ion distances respectively and r_1 and r_5 which correspond to fluoride ion and water molecule distances respectively. The results are shown in figure 5.49. For

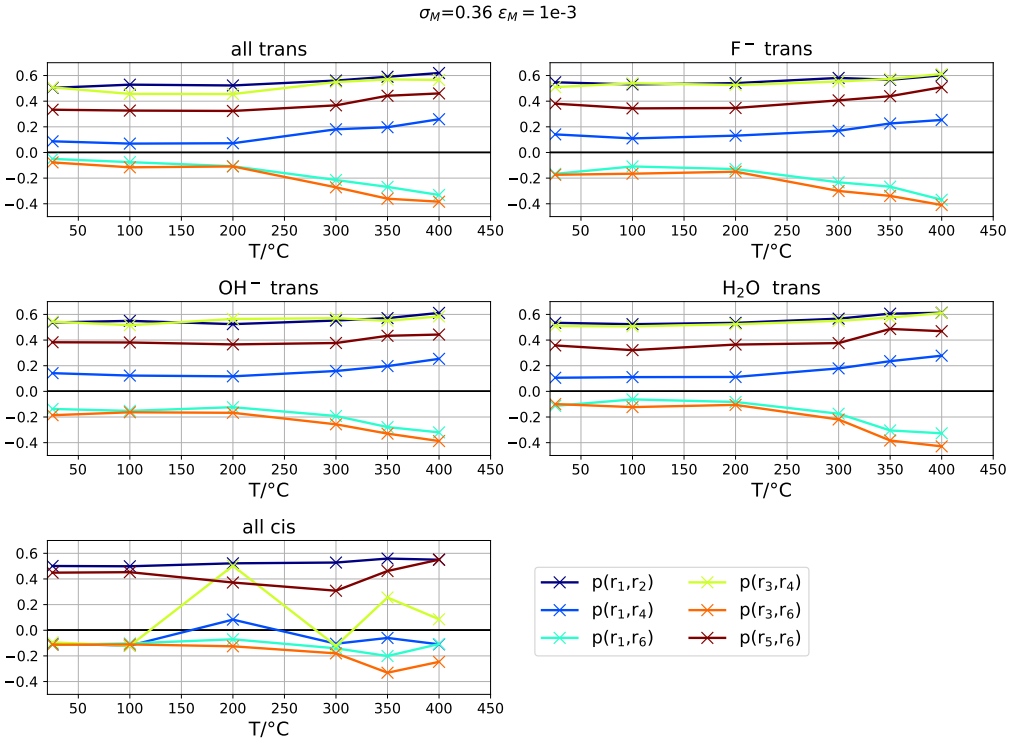

FIGURE 5.49: Pearson correlation coefficients for relevant pairs of ligands. r_i denotes the distance of the i^{th} ligand closest to the ion. The line at p=0 is thickened for better visibility.

all starting structures, except all cis, one observes positive correlation coefficients for combinations of r_1 to r_4, i.e. between the ions. This means that as the fluoride ions get closer to the cation, the other fluoride ion and hydroxide ions get closer, too. At the same time, there are negative correlation coefficients for correlations between r_1-r_4 and r_5-r_6, i.e. between the ions and the water molecules. The negative correlations indicate that as the water molecules move away from the cation, the ionic ligands move towards the central cation and vice versa. Furthermore, there is also a positive correlation between r_5 and r_6, which means that if one water molecule moves away from the cation, the other does too.
Note that the correlations are very high since the Pearson correlation coefficient is

bound by -1 and +1. The all cis configuration is again the odd one out likely due to its instability. The negative correlation between r_3 and r_4 can be fully understood by one hydroxide ion being exchanged by one water molecule as was observed.
Interestingly, the magnitude of all correlations increases with increasing temperature. This means that the overall motion of the six ligands becomes more concerted at higher temperatures compared to lower ones.

5.6.5 Potential Energy Scan

I computed the potential acting on one of the fluoride ions in the all-trans configuration for several cation-water distances. The result is shown in figure 5.50. This simulates the effect of two water molecules leaving the solvation shell. I find that as a consequence of increasing water-cation distances, the minimum of cation-fluoride potential shifts towards smaller values. This confirms the suspicion that the contraction of the four ligands is a consequence of the vanishing Coulomb repulsion originating from the two water molecules.
This potential scan has been conducted in vacuum. The absence of solvent interactions likely explains why the minima of the potential energy functions are located at larger distances than the cation-ligand distances found in the simulations.

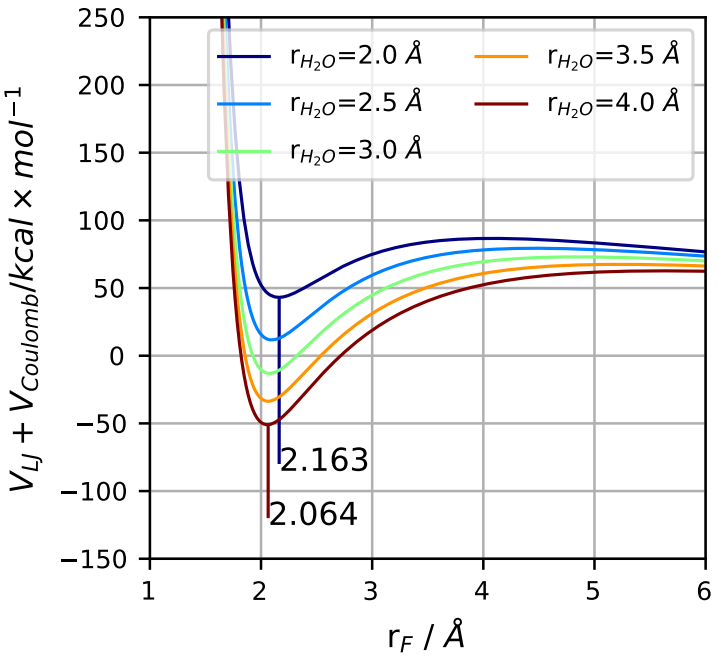

FIGURE 5.50: Lennard-Jones and Coulomb potential vs the cation-fluoride distance for several cation water distances for the all-trans configuration. The vertical lines indicate the minima of the functions. Here, r_F and r_{H_2O} denote the cation-fluoride and cation-water distances respectively. The Lennard-Jones parameters were $\sigma_M = 0.36$ nm and $\epsilon_M = 10^{-3}$kcal/mol.

5.6.6 The Two States of the Solvation Shell at Higher Temperatures

A second peak appears in the distributions of water distances at higher temperatures. This interesting feature is depicted in detail in figure 5.51. The existence of the

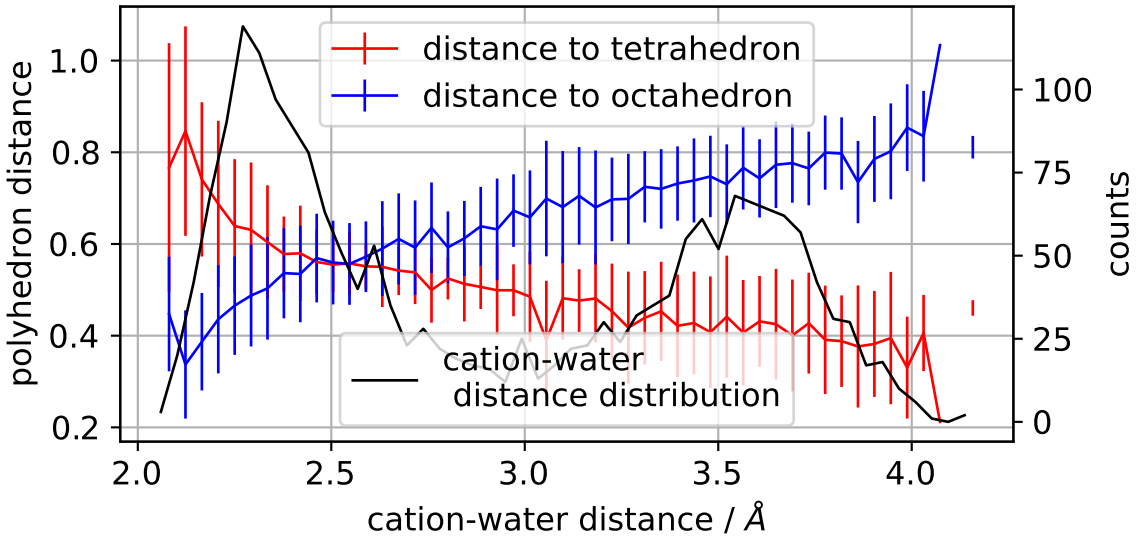

FIGURE 5.51: Red and blue curves: Mean distance towards tetrahedral and octahedral reference structures respectively including standard deviations. Averages are computed over all frames within a bin. The black curve shows the distribution of cation-ligand distances for the sixth furthest ligand. All data refer to the all-trans configuration at 400 °C.

second maximum implies the existence of a corresponding structure of the solvation shell. Visual inspections hinted at this second structure to be a tetrahedron formed with the four closest ions. To test this hypothesis, I computed the Steinhardt parameters for the closest four ligands and those for the closest six ligands. Next, I used the Steinhardt parameters for the closest four ligands to compute the distance towards the tetrahedral reference structure. I used the Steinhardt parameters for the closest six ligands to determine the distance towards the octahedral reference structure.
Finally, I discretized the cation-water distance into 50 bins and computed the mean polyhedron distance to reference structures for each bin. These are the results shown as red and blue curves in figure 5.51.
One can see that the polyhedron distance towards the tetrahedral structure shrinks for smaller cation-water distances and grows for larger cation-water distances. The opposite is true for polyhedron distance towards the octahedral structure. That is, it shrinks for smaller cation-water distances and grows for larger cation-water distances. The first peak must belong to octahedral structures because this was the starting structure of the complexes. The minimized distance towards octahedral structures further proves this statement and gives this analysis confidence. Therefore, the second peak can be assigned to tetrahedral structures as the corresponding polyhedron distance shrinks towards similar values seen for the first peak and the octahedral structure.
Snapshots of both structures are compared in figure 5.52.
The point $R = 3$ Å was chosen as a boundary between the two states. The transitions over the boundary were studied to determine the equilibrium constant for this shift of the solvation shell. The temperature dependence of the equilibrium constants was used for a Van-'t-Hoff analysis, revealing the difference in enthalpy and the difference in entropy of both solvation shells. From the slope and y-intercept of

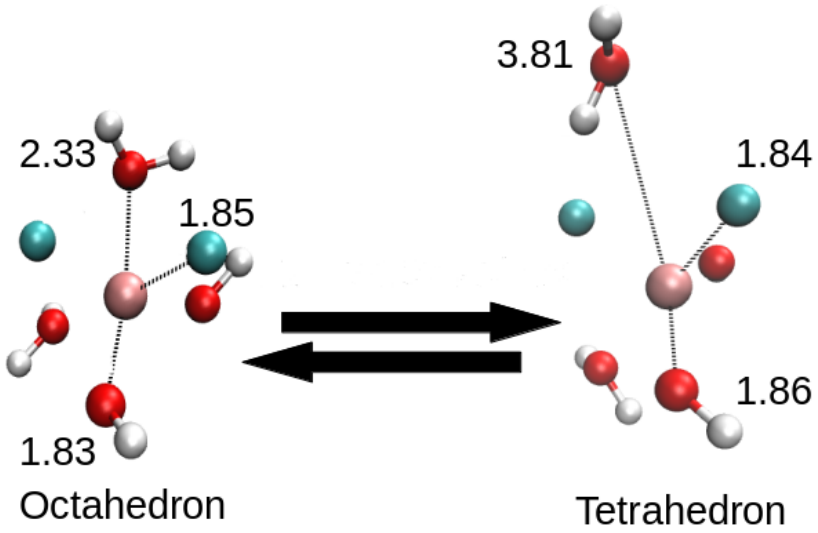

FIGURE 5.52: Both solvation shells of the tetravalent metal ion at high temperatures. The numbers next to an atom are the distances from the center in Å.

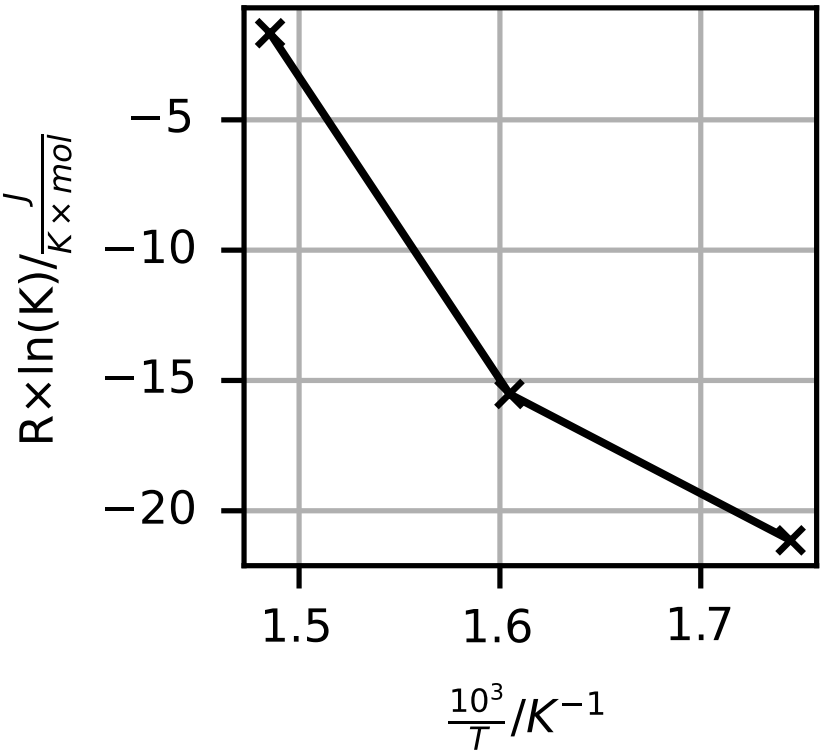

FIGURE 5.53: Temperature dependence of the equilibrium constant for the transition of a water molecule over $R = 3$ Å in the solvation shell of a tetravalent metal cation.

the graph in figure 5.53 the enthalpy difference and entropy difference were determined respectively. The enthalpy difference was found to be $\Delta H = -74.07$ kJ/mol. The entropy difference was found to be $\Delta S = 106.60$ J/K/mol.

In this project, the MD simulation as well as their analysis was performed by me. Dr. A. Loges provided the mean cation-ligand distances based on the EXAFS spectra.

Chapter 6

Conclusions

The Markov models for open systems, specifically solvation shells, were based on the time series of expansion coefficients in terms of spherical harmonics. Using a basis set which relies solely on angular coordinates becomes justified if all particles in consideration appear at a common distance towards the central point. The resulting implied timescales and eigenvectors can be analyzed like any other Markov model.
Multiplying models for solute and solvent yielded no increase in absolute timescales and no increased convergence of the models. In contrast, additive models yielded a slight increase in absolute timescales. The sign structure of the combined eigenvectors yielded information about the co-occurrence of processes in the solute and solvation shell.
Judging by the value and convergence of implied timescales, the additive models performed better than the multiplicative models. While the added models had slightly to much higher timescales, the multiplicative model performed just as well as the individual models or worse. Practically, the multiplied models require a projection onto slow processes before being analyzed whereas the additive models are computationally more demanding. That is because the correlation matrix grows with the square of the number of basis functions used.
Unfortunately, the search for a suitable HF-model was unfruitful. The parameter combinations investigated did not recreate a satisfactory amount of physical quantities. Even a three-site model, parametrized to recreate the electrostatic potential, was unable to yield stable simulations of aqueous HF, within the investigated parameter space.
Fortunately, a publication in 2020 provided an excellent model for HF which was suitable for simulation of aqueous HF as well.
The parameter scans of tetravalent metal cations in fluorous environments yielded extremely robust results. The investigated quantities all form coherent regions in the parameter space which look like diagonal lanes in the $\log(\epsilon_M) - \sigma_M$-plane. This trend was observed for 9 orders of magnitude in ϵ_M and $\sigma_M < 5$ Å.
Most metal ions in the parameter space are coordinated by some of the fluoride ions. Uncoordinated metal ions were the exception in this broad search. As a results, the most common coordination polyhedra were either tetrahedrons or octahedrons as is found by the newly developed method. From analyzing the mentioned static quantities of the solvation shells, it becomes hard to pinpoint a specific parameter combination which is suitable to be used for Zr and Hf, but instead, an entire lane in parameter space. It should be mentioned that other quantities described in literature, i.e. the hydration free energy in [45], occur in coherent regions which are not lanes in the semi-logarithmic parameter space.
Differentiating Zr from Hf cations in classical MD is rather difficult because the two chemical species exhibit similar charges as well as similar ionic radii. Recently, Dr. A. Loges found that the solvation shells of both metal ions contract with increasing

temperature, in 1 M HF.
Classical MD simulations were able to recreate the phenomenon over an entire lane in parameter space and for all conformations of the $[MF_2(OH)_2(H_2O)_2]^0$ complex. The thermal contractions were a result of the correlated motion of the ligands in the complex. The cation-ligand distance of the neutral ligands is anticorrelated with cation-ligand distance of the ionic ligands. Therefore, when the water molecules move further away from the cation, their missing coulomb repulsion allows the ligands to move in closer to the central cation.
Interestingly, at extremely high temperatures of 300 °C and above, the cation-water distance distribution shows a second peak. This second configuration corresponds to a tetrahedral configuration of the complex which coexists with the octahedral one. Difference in enthalpy and entropy of the two solvation shell modifications could be determined via Van-'t-Hoff analysis.

Chapter 7

Outlook

Future projects could aim to include radial dynamics into the Markov models of solvation shells and look for improved accuracy or faster model convergence. The basis set used for the radial dynamics should be able to recreate the radial distribution functions as well as possess knots that are varied enough to depict the dynamics accurately. Furthermore, it would be useful if the basis function would not vary too strongly in their values, such that the coefficients, which are essentially the eigenvector of the correlation matrix, are able to reproduce the dynamic processes within computational accuracy.
The dynamic properties reported in [65] can be used as reference to validate dynamic solvation shell models. One could even think of using these dynamic properties as fitting targets for parametrizations.
Another route would be investigating the solvation shells of systems with higher complexity and less symmetry, e.g. small organic molecules like fluorinated amino acids with this method.
The parameter search of HF would benefit by expanding the explored parameter space, especially σ_F. Crystallographic studies of coordination compounds of metal ions with HF [133–137] could be used to fine-tune Lennard-Jones parameters for a metal-ion specific HF-model.
The polyhedron detection method could be expanded by increasing the database of reference polyhedra, especially for smaller coordination numbers.
The parameter scan for metal ions would benefit from scanning the hydration-free energy of the metal ions for several Lennard-Jones parameters. It would also be interesting to see if this quantity changes with increasing HF concentrations. Beyond these additions, the scans of the parameter space can serve as a basis for future parameterization attempts of tetravalent metal cations.
I noticed that the $[MF_2(OH)_2(H_2O)_2]$ octahedron did not stay in its starting configuration for the entire simulation. Therefore, it would be interesting to investigate the exchange rates between these configurations. Furthermore, analyzing the thermal contraction of several parameter combinations could be fruitful to find parameter combinations that distinguish between zirconium and hafnium.

Bibliography

1. Muller, K., Faeh, C. & Diederich, F. Fluorine in pharmaceuticals: looking beyond intuition. *science* **317,** 1881–1886 (2007).
2. Carlson, S. *et al.* Hydrophobicity of self-assembled monolayers of alkanes: fluorination, density, roughness, and Lennard-Jones cutoffs. *Langmuir* **37,** 13846–13858 (2021).
3. Chowdhury, M. S. *et al.* Functional surfactants for molecular fishing, capsule creation, and single-cell gene expression. *Nano-micro letters* **13,** 1–9 (2021).
4. Jeschke, P. The unique role of fluorine in the design of active ingredients for modern crop protection. *ChemBioChem* **5,** 570–589 (2004).
5. Campbell, M. G. & Ritter, T. Modern carbon–fluorine bond forming reactions for aryl fluoride synthesis. *Chemical Reviews* **115,** 612–633 (2015).
6. Neves, J., Nunes, J. & Sahama, T. G. High hafnium members of the zircon-hafnon series from the granite pegmatites of Zambézia, Mozambique. *Contributions to Mineralogy and Petrology* **48,** 73–80 (1974).
7. Yin, R. *et al.* Extreme fractionation from zircon to hafnon in the Koktokay No. 1 granitic pegmatite, Altai, northwestern China. *American Mineralogist* **98,** 1714–1724 (2013).
8. Kudryashov, N., Skublov, S., Galankina, O., Udoratina, O. & Voloshin, A. Abnormally high-hafnium zircon from rare-metal pegmatites of the Vasin-Mylk deposit (the northeastern part of the Kola Peninsula). *Geochemistry* **80,** 125489 (2020).
9. Motta, A. T. *et al.* Hydrogen in zirconium alloys: A review. *Journal of Nuclear Materials* **518,** 440–460 (2019).
10. Cerefice, G, Draye, M, Noyes, K & Czerwinski, K. Hafnium hydroxide complexation and solubility: The impact of hydrolysis reactions on the disposition of weapons-grade plutonium. *MRS Online Proceedings Library (OPL)* **556** (1999).
11. Zhang, Y., Gregg, D. J., Kong, L., Jovanovich, M. & Triani, G. Zirconolite glass-ceramics for plutonium immobilization: The effects of processing redox conditions on charge compensation and durability. *Journal of Nuclear Materials* **490,** 238–241 (2017).
12. Persson, I. Hydrated metal ions in aqueous solution: How regular are their structures? *Pure and Applied Chemistry* **82,** 1901–1917 (2010).
13. Huggins, D. J. Studying the role of cooperative hydration in stabilizing folded protein states. *Journal of Structural Biology* **196,** 394–406 (Dec. 2016).
14. Witek, J. *et al.* Kinetic Models of Cyclosporin A in Polar and Apolar Environments Reveal Multiple Congruent Conformational States. *J. Chem. Inf. Model.* **56,** 1547–1562 (July 2016).

15. Witek, J. *et al.* Interconversion Rates between Conformational States as Rationale for the Membrane Permeability of Cyclosporines. *ChemPhysChem* **7,** 608 (Sept. 2017).
16. Bolhuis, P. G., Dellago, C. & Chandler, D. Reaction coordinates of biomolecular isomerization. *Proceedings of the National Academy of Sciences* **97,** 5877–5882 (2000).
17. Sorin, E. J., Rhee, Y. M. & Pande, V. S. Does water play a structural role in the folding of small nucleic acids? *Biophysical Journal* **88,** 2516–2524 (2005).
18. Brotzakis, Z. F. & Bolhuis, P. G. Unbiased Atomistic Insight into the Mechanisms and Solvent Role for Globular Protein Dimer Dissociation. *J. Phys. Chem. B* **123,** 1883–1895 (Feb. 2019).
19. Hua, L., Huang, X., Zhou, R. & Berne, B. J. Dynamics of water confined in the interdomain region of a multidomain protein. *The Journal of Physical Chemistry B* **110,** 3704–3711 (2006).
20. Cotton, F. A., Wilkinson, G., Murillo, C. A., Bochmann, M. & Grimes, R. *Advanced inorganic chemistry* (Wiley New York, 1988).
21. Greenwood, N. N. & Earnshaw, A. *Chemistry of the Elements* (Elsevier, 1998).
22. Di Tommaso, D. & de Leeuw, N. H. First principles simulations of the structural and dynamical properties of hydrated metal ions Me2+ and solvated metal carbonates (Me= Ca, Mg, and Sr). *Crystal growth & design* **10,** 4292–4302 (2010).
23. Piao, Z., Gao, R., Liu, Y., Zhou, G. & Cheng, H.-M. A review on regulating Li+ solvation structures in carbonate electrolytes for lithium metal batteries. *Advanced Materials,* 2206009 (2022).
24. Schirò, G. *et al.* Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins. *Nat. Commun.* **6,** 6490 (2015).
25. Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Physical Review B* **28,** 784 (1983).
26. Lechner, W. & Dellago, C. Accurate determination of crystal structures based on averaged local bond order parameters. *The Journal of Chemical Physics* **129,** 114707 (2008).
27. Shevchuk, R. Markov State Model of Ion Assembling Process. *The Journal of Physical Chemistry A* **120,** 2783–2788 (2016).
28. Dibyendu Bandyopadhyay, S. M. S. K. G. & Choudhury, N. Correlation of Structural Order, Anomalous Density, and Hydrogen Bonding Network of Liquid Water, 1–13 (July 2013).
29. Gu, C. *et al.* Building Markov state models with solvent dynamics. *BMC Bioinformatics* **14,** S8. `https://doi.org/10.1186/1471-2105-14-s2-s8` (2013).
30. Harrigan, M. P., Shukla, D. & Pande, V. S. Conserve water: A method for the analysis of solvent in molecular dynamics. *Journal of Chemical Theory and Computation* **11,** 1094–1101 (2015).
31. Shevchuk, R., Agmon, N. & Rao, F. Network analysis of proton transfer in liquid water. *J. Chem. Phys.* **140,** 244502–9 (2014).
32. Hamm, P. Markov state model of the two-state behaviour of water. *J. Chem. Phys.* **145,** 134501–12 (Oct. 2016).

33. Schulz, R *et al.* Collective hydrogen-bond rearrangement dynamics in liquid water. *J. Chem. Phys.* **149,** 244504 (Dec. 2018).
34. Nüske, F., Keller, B. G., Pérez-Hernández, G., Mey, A. S. & Noe, F. Variational approach to molecular kinetics. *Journal of Chemical Theory and Computation* **10,** 1739–1752 (2014).
35. Eurofluor. *Ein Überblick über die Fluorchemie 4. Ausgabe* `https://www.eurofluor.org/download/german/`. Accessed: 2022-08-10. 2019.
36. Cournoyer, M. E. & Jorgensen, W. L. An improved intermolecular potential function for simulations of liquid hydrogen fluoride. *Molecular Physics* **51,** 119–132 (1984).
37. JEDLOVSZKY, B. P. & Vallauri, R. Computer simulation study of liquid HF with a new effective pair potential model. *Molecular Physics* **92,** 331–336 (1997).
38. Kreitmeir, M. *et al.* Structure of dense hydrogen fluoride gas from neutron diffraction and molecular dynamics simulations. *The Journal of chemical physics* **122,** 154511 (2005).
39. Orabi, E. A. & Faraldo-Gómez, J. D. New molecular-mechanics model for simulations of hydrogen fluoride in chemistry and biology. *Journal of chemical theory and computation* **16,** 5105–5126 (2020).
40. Li, P., Song, L. F. & Merz Jr, K. M. Systematic parameterization of monovalent ions employing the nonbonded model. *Journal of chemical theory and computation* **11,** 1645–1657 (2015).
41. Sengupta, A., Li, Z., Song, L. F., Li, P. & Merz Jr, K. M. Parameterization of monovalent ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *Journal of chemical information and modeling* **61,** 869–880 (2021).
42. Li, Z., Song, L. F., Li, P. & Merz Jr, K. M. Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *Journal of chemical theory and computation* **16,** 4429–4442 (2020).
43. Li, Z., Song, L. F., Li, P. & Merz Jr, K. M. Parametrization of Trivalent and Tetravalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *Journal of chemical theory and computation* **17,** 2342–2354 (2021).
44. Li, P., Song, L. F. & Merz Jr, K. M. Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water. *The Journal of Physical Chemistry B* **119,** 883–895 (2015).
45. Zhang, Y., Jiang, Y., Qiu, Y. & Zhang, H. Rational Design of Nonbonded Point Charge Models for Highly Charged Metal Cations with Lennard-Jones 12-6 Potential. *Journal of Chemical Information and Modeling* (2021).
46. Marcus, Y. Ionic radii in aqueous solutions. *Chemical Reviews* **88,** 1475–1498 (1988).
47. Marcus, Y. Thermodynamics of solvation of ions. Part 5.—Gibbs free energy of hydration at 298.15 K. *Journal of the Chemical Society, Faraday Transactions* **87,** 2995–2999 (1991).
48. Louvel, M., Sanchez-Valle, C., Malfait, W. J., Testemale, D. & Hazemann, J.-L. Zr complexation in high pressure fluids and silicate melts and implications for the mobilization of HFSE in subduction zones. *Geochimica et Cosmochimica Acta* **104,** 281–299 (2013).

49. Migdisov, A. A., Williams-Jones, A., Van Hinsberg, V & Salvi, S. An experimental study of the solubility of baddeleyite (ZrO2) in fluoride-bearing solutions at elevated temperature. *Geochimica et Cosmochimica Acta* **75,** 7426–7434 (2011).
50. Bau, M. Controls on the fractionation of isovalent trace elements in magmatic and aqueous systems: evidence from Y/Ho, Zr/Hf, and lanthanide tetrad effect. *Contributions to Mineralogy and Petrology* **123,** 323–333 (1996).
51. Eagleson, M. *et al. Concise encyclopedia chemistry* (Walter de Gruyter, 1994).
52. Xu, L *et al.* Zirconium and hafnium separation with molten salt extraction in Sn–Cu–Zr–Hf and Cu–Zr–Hf alloy systems. *Separation Science and Technology* **51,** 1664–1674 (2016).
53. Poriel, L, Favre-Réguillon, A, Pellet-Rostaing, S & Lemaire, M. Zirconium and hafnium separation, part 1. Liquid/liquid extraction in hydrochloric acid aqueous solution with Aliquat 336. *Separation science and technology* **41,** 1927–1940 (2006).
54. Loges, A. *et al.* Fluoride complexation of Zr and Hf in hydrothermal aqueous fluids.
55. Onufriev, A. The generalized Born model: its foundation, applications, and limitations. *Departments of Computer Science and Physics, Blacksburg, Virginia, USA* (2010).
56. Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1,** 699–709 (2011).
57. Klamt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *The Journal of Physical Chemistry* **99,** 2224–2235 (1995).
58. Berendsen, H., Postma, J., Van Gunsteren, W. & Hermans, J. *Intermolecular Forces, Edited by: Pullman, B. 331–342* 1981.
59. Berendsen, H., Grigera, J. & Straatsma, T. The missing term in effective pair potentials. *Journal of Physical Chemistry* **91,** 6269–6271 (1987).
60. Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society* **103,** 335–340 (1981).
61. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79,** 926–935 (1983).
62. Horn, H. W. *et al.* Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of chemical physics* **120,** 9665–9678 (2004).
63. Abascal, J. L. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *The Journal of Chemical Physics* **123,** 234505 (2005).
64. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of chemical physics* **112,** 8910–8922 (2000).
65. Ohtaki, H. & Radnai, T. Structure and dynamics of hydrated ions. *Chemical reviews* **93,** 1157–1204 (1993).

66. Sutmann, G. in *Quantum Simulations of Complex Many-Body Systems:From Theory to Algorithms,* (eds Grotendorst, J., Marx, D. & Muramatsu, A.) 211–254 (John von Neumann Institute for Computing, Jülich, 2002).

67. Ewald, P. P. Ewald summation. *Ann. Phys* **369,** 1–2 (1921).

68. Wells, B. A. & Chaffee, A. L. Ewald summation for molecular simulations. *Journal of chemical theory and computation* **11,** 3684–3695 (2015).

69. Lorentz, H. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Annalen der Physik* **248,** 127–136 (1881).

70. Fincham, D. Leapfrog rotational algorithms. *Molecular Simulation* **8,** 165–178 (1992).

71. Jung, J. *et al.* New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems. *Journal of computational chemistry* **42,** 231–241 (2021).

72. Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A. & Haak, J. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81,** 3684–3690 (1984).

73. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **72,** 2384–2393 (1980).

74. Parrinello, M. & Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Physical review letters* **45,** 1196 (1980).

75. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52,** 7182–7190 (1981).

76. Li, P. & Merz Jr, K. M. Metal ion modeling using classical mechanics. *Chemical reviews* **117,** 1564–1686 (2017).

77. Sigfridsson, E. & Ryde, U. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *Journal of Computational Chemistry* **19,** 377–395 (1998).

78. Schütte, C., Fischer, A., Huisinga, W. & Deuflhard, P. "A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo". *J. Comput. Phys.* **151,** 146. https://doi.org/10.1006/jcph.1999.6231 (1999).

79. Swope, W. C., Pitera, J. W. & Suits, F. "Describing Protein Folding Kinetics by Molecular Dynamics Simulations". *J. Phys. Chem. B* **108,** 6571–6581. https://doi.org/10.1021/jp037421y (2004).

80. Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics* **134,** 174105 (2011).

81. Noé, F. & Nuske, F. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Modeling & Simulation* **11,** 635–655 (2013).

82. Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A. & Swope, W. C. "Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics". *J. Chem. Phys.* **126,** 155101. https://doi.org/10.1063/1.2714538 (2007).

83. Buchete, N.-V. & Hummer, G. "Coarse Master Equations for Peptide Folding Dynamics". *J. Phys. Chem. B* **112,** 6057–6069. https://doi.org/10.1021/jp0761665 (2008).

84. Vanden-Eijnden, E. & Venturoli, M. "Markovian milestoning with Voronoi tessellations". *J. Chem. Phys.* **130,** 194101. https://doi.org/10.1063/1.3129843 (2009).

85. Schütte, C., Noé, F., Lu, J., Sarich, M. & Vanden-Eijnden, E. "Markov state models based on milestoning". *J. Chem. Phys.* **134,** 204105. https://doi.org/10.1063/1.3590108 (2011).

86. Lemke, O. & Keller, B. G. "Density-based cluster algorithms for the identification of core sets". *J. Chem. Phys.* **145,** 164104. https://doi.org/10.1063/1.4965440 (2016).

87. Pérez-Hernández, G., Paul, F., Giorgino, T., Fabritiis, G. D. & Noé, F. "Identification of slow molecular order parameters for Markov model construction". *J. Chem. Phys.* **139,** 015102. https://doi.org/10.1063/1.4811489 (2013).

88. Schwantes, C. R. & Pande, V. S. "Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9". *J. Comp. Theory Comput.* **9,** 2000–2009. https://doi.org/10.1021/ct300878a (2013).

89. Keys, A. S., Iacovella, C. R. & Glotzer, S. C. Characterizing structure through shape matching and applications to self-assembly. *Annual Review of Condensed Matter Physics* **2,** 263–285 (2011).

90. Koch, W. & Holthausen, M. C. *A chemist's guide to density functional theory* (John Wiley & Sons, 2015).

91. R¨uhle, V. *Pressure coupling / barostats* hhttps://www2.mpip-mainz.mpg.de/~andrienk/journal_club/barostats.pdf (2021).

92. *MD Thermostats handout* http://www.strodel.info/index_files/lecture/MDthermostats_handout.pdf (2021).

93. $Der_V irialsatz_d er_s tatistischen_M echanik$ https://de.wikipedia.org/wiki/Virialsatz#Der_Virialsatz_der_statistischen_Mechanik (2021).

94. Hünenberger, P. H. Thermostat algorithms for molecular dynamics simulations. *Advanced computer simulation,* 105–149 (2005).

95. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126,** 014101 (2007).

96. Bussi, G., Zykova-Timan, T. & Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of chemical physics* **130,** 074101 (2009).

97. Faro, T. M., Thim, G. P. & Skaf, M. S. A Lennard-Jones plus Coulomb potential for Al 3+ ions in aqueous solutions. *The Journal of chemical physics* **132,** 114509 (2010).

98. Best, R. B. & Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *The Journal of Physical Chemistry B* **114,** 14916–14923 (2010).

99. Van Gunsteren, W. F. & Berendsen, H. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation* **1,** 173–185 (1988).

100. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18,** 1463–1472 (1997).

101. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98,** 10089–10092 (1993).

102. Berendsen, H. J., van der Spoel, D. & van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics communications* **91,** 43–56 (1995).

103. Oliphant, T. A Guide to NumPy, vol. 1. *Spanish Fork: Trelgol Publishing* (2006).

104. Oliphant, T. E. Python for Scientific Computing. *Computing in Science & Engineering* **9,** 10–20. `https://doi.org/10.1109/mcse.2007.58` (2007).

105. Scherer, M. K. *et al.* PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *Journal of Chemical Theory and Computation* **11,** 5525–5542 (2015).

106. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9,** 90–95 (2007).

107. Working Group on Storage, T. & (STS), S. *GENERAL PROPERTIES OF ANHYDROUS HYDROGEN FLUORIDE (AHF) AND HYDROFLUORIC ACID SOLUTIONS (HF)* tech. rep. 2018.04.09 (EUROFLUOR, the European Technical Committee for Fluorine, Avenue E. Van Nieuwenhuyse 4, B-1160 Brussels, Belgium, 2018). `https://www.eurofluor.org/publications-and-recommendations/`.

108. Baker, D. *General Chemistry, (Ebbing, Darrell D.)* 1997.

109. Higgins, T. L. & Westrum Jr, E. F. THERMOCHEMICAL STUDY OF THE SODIUM AND AMMONIUM HYDROGEN FLUORIDES IN ANHYDROUS HYDROGEN FLUORIDE1. *The Journal of Physical Chemistry* **65,** 830–836 (1961).

110. Wenthold, P. G. & Squires, R. R. Bond dissociation energies of F2-and HF2-. a gas-phase experimental and G2 theoretical study. *The Journal of Physical Chemistry* **99,** 2002–2005 (1995).

111. McQuarrie, D. A. *Statistical mechanics* (Sterling Publishing Company, 2000).

112. GROMACS. *Free Energy of Solvation* `https://tutorials.gromacs.org/free-energy-of-solvation.html#`. Accessed: 2020-08-10.

113. Harris, R. K. *et al.* Further conventions for NMR shielding and chemical shifts (IUPAC Recommendations 2008). *Pure and Applied Chemistry* **80,** 59–84 (2008).

114. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1,** 19–25 (2015).

115. Wagner, W. & Pruß, A. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *Journal of physical and chemical reference data* **31,** 387–535 (2002).

116. Chen, W., Wallace, J. A., Yue, Z. & Shen, J. K. Introducing titratable water to all-atom molecular dynamics at constant pH. *Biophysical journal* **105,** L15–L17 (2013).

117. Marsh, D. A. J. *Polyhedron Generator (https://drajmarsh.bitbucket.io/poly3d.html)* `https://drajmarsh.bitbucket.io/poly3d.html` (2021).

118. Hertz, H., Tutsch, R & Versmold, H. Molecular motion and structure around the hydrated ions Li+ and Al3+. *Berichte der Bunsengesellschaft für physikalische Chemie* **75,** 1177–1191 (1971).

119. McLain, S. E., Benmore, C. J., Siewenie, J. E., Urquidi, J. & Turner, J. F. On the structure of liquid hydrogen fluoride. *Angewandte Chemie* **116,** 1986–1989 (2004).

120. O'Reilly, D. Comment on "Self-Diffusion in Liquid Hydrogen Fluoride". *The Journal of Chemical Physics* **52,** 5974–5975 (1970).

121. Lagemann, R. T. & Knowles, C. H. Velocity of compressional waves in liquid hydrogen fluoride and some thermodynamic properties derived therefrom. *The Journal of Chemical Physics* **32,** 561–564 (1960).

122. Fredenhagen, K. & Dahmlos, J. Die dielektrizitätskonstante des flüssigen fluorwasserstoffs. *Zeitschrift für anorganische und allgemeine Chemie* **178,** 272–274 (1929).

123. Simons, J. & Bouknight, J. The density and surface tension of liquid hydrogen fluoride. *Journal of the American Chemical Society* **54,** 129–135 (1932).

124. Franck, E. & Spalthoff, W. Fluorwasserstoff I. Spezifische Wärme, Dampfdruck und Dichte bis zu 300° C und 300 at. *Zeitschrift für Elektrochemie, Berichte der Bunsengesellschaft für physikalische Chemie* **61,** 348–357 (1957).

125. Sheft, I., Perkins, A. J. & Hyman, H. H. Anhydrous hydrogen fluoride: Vapor pressure and liquid density. *Journal of Inorganic and Nuclear Chemistry* **35,** 3677–3680 (1973).

126. Streng, A. G. Miscibility and compatibility of some liquefied and solidified gases at low temperatures. *Journal of Chemical & Engineering Data* **16,** 357–359 (1971).

127. Fine, R. A. & Millero, F. J. Compressibility of water as a function of temperature and pressure. *The Journal of Chemical Physics* **59,** 5529–5536 (1973).

128. Domenico, P. & Mifflin, M. Water from low-permeability sediments and land subsidence. *Water Resources Research* **1,** 563–576 (1965).

129. Holz, M., Heil, S. R. & Sacco, A. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. *Physical Chemistry Chemical Physics* **2,** 4740–4742 (2000).

130. Noulty, R. A. & Leaist, D. G. Diffusion of aqueous hydrofluoric acid and aqueous potassium fluoride. *Electrochimica acta* **30,** 1095–1099 (1985).

131. Horvath, A. Heat capacity of liquid hydrogen fluoride-A discrepancy (1972).

132. Hu, J.-H., White, D. & Johnston, H. The Heat Capacity, Heat of Fusion and Heat of Vaporization of Hydrogen Fluoride1. *Journal of the American Chemical Society* **75,** 1232–1236 (1953).

133. Tramšek, M, Benkič, P, Turičnik, A, Tavčar, G & Žemva, B. Coordination compounds with XeF2, AsF3 and HF as ligands to metal ions: a review of reaction systematics, Raman spectra and metal, fluoro-ligand polyhedra. *Journal of fluorine chemistry* **114,** 143–148 (2002).

134. Turičnik, A., Benkič, P. & Žemva, B. Syntheses and crystal structures of [Mg (HF) 2](SbF6) 2 and [Ca (HF) 2](SbF6) 2: new examples of HF acting as a ligand to metal centres. *Journal of fluorine chemistry* **121,** 245–251 (2003).

135. Tramšek, M. & Žemva, B. Synthesis of novel salts with HF, AsF3 and XeF2 as ligands to metal cations. *Journal of fluorine chemistry* **127,** 1275–1284 (2006).

136. Tramšek, M., Goreshnik, E., Lozinšek, M. & Žemva, B. HF molecules and poly (hydrogen fluoride) anions as ligands to metal centers. *Journal of Fluorine Chemistry* **130,** 1093–1098 (2009).

137. Bunič, T., Tramšek, M., Goreshnik, E. & Žemva, B. Synthesis and structural investigation of the compounds containing HF2- anions: Ca (HF2) 2, Ba4F4 (HF2)(PF6) 3 and Pb2F2 (HF2)(PF6). *Journal of Solid State Chemistry* **181,** 2318–2324 (2008).