Development of Models and Methods to Simulate Peptide-Assisted Nucleation and Growth of Calcium-Minerals

Dissertation

zur Erlangung des Grades eines
'Doktor rerum naturalium (Dr. rer. nat.)'

der Fachbereiche
08 - Physik, Mathematik und Informatik,
09 - Chemie, Pharmazie und Geowissenschaften,
10 - Biologie,
Universitätsmedizin
der Johannes Gutenberg-Universität

Jens Kahlen
geb. in Bonn

Max Planck Graduate Center
mit der Johannes Gutenberg-Universität Mainz

Angefertigt am MPI für Polymerforschung

Mainz, 12. Mai 2014

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 03. Juni 2014

Dissertation der Universität Mainz - D77

Statutory Declaration

I hereby declare that I wrote the dissertation submitted without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages which are appropriated verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of Johannes Gutenberg-University Mainz to insure standard scientific practice.

Jens Kahlen

Abstract

In the past years, several so far unknown phenomena - such as the existence of different pre-nucleation structures - have been observed experimentally, and have generated a new understanding of the processes that occur at the molecular scale during nucleation and growth of crystals. The implications that pre-nucleation structures might have in the process of biomineralisation are not yet well-understood. The mechanisms, by which biomolecular modifiers such as peptides might interact with pre-nucleation structures and thus influence the nucleation process of minerals, are numerous. Molecular simulations are well suited for the analysis of the formation of pre-nucleation structures in the presence of modifiers. This thesis presents an approach to study the interactions of peptides with the solvated constituents of the forming crystals by means of Molecular Dynamics simulations.

As a first step towards informative simulations, the quality of existing forcefields with respect to the description of oligoglutamates interacting with calcium ions in aqueous solution was tested. It was found that large discrepancies between well-established forcefields exist, and that non of the tested forcefields gave a realistic description of the ion-pairing of these complex ions. Therefore, a strategy to optimise existing biomolecular forcefields in this respect was developed and it was found that relatively small changes in the parameters of the ion–peptide van-der-Waals interactions were sufficient to obtain a reliable model for the system of interest.

The comprehensive sampling of the phase space of the systems is particularly challenging due to the large number of degrees of freedom and the strong interactions between calcium ions and glutamate in solution. The Biasing Potential Replica Exchange Molecular Dynamics method was therefore adapted to the enhanced sampling of oligoglutamates, and peptides of different chain lengths were simulated in the presence of calcium ions. With the help of the sketch-map analysis method numerous stable ion–peptide clusters, which might affect the formation of pre-nucleation structures, were identified from these simulations. Depending on the chain length of the peptide, these clusters exhibit characteristic distances between the calcium ions. These distances resemble some of the distances between calcium ions in those phases of calcium oxalate crystals, which are grown in the presence of oligoglutamates. The analogy of distances between calcium ions in solvated ion–peptide clusters and in calcium oxalate crystals might be interpreted as an indication for the importance of ion–peptide clusters in the nucleation and growth of biominerals and it is a possible explanation for the ability of oligoglutamates to influence the phase of the forming crystals, as it has been observed in experiments.

Zusammenfassung

In den vergangenen Jahren wurden einige bislang unbekannte Phänomene experimentell beobachtet, wie etwa die Existenz unterschiedlicher Prä-Nukleations-Strukturen. Diese haben zu einem neuen Verständnis von Prozessen, die auf molekularer Ebene während der Nukleation und dem Wachstum von Kristallen auftreten, beigetragen. Die Auswirkungen solcher Prä-Nukleations-Strukturen auf den Prozess der Biomineralisation sind noch nicht hinreichend verstanden. Die Mechanismen, mittels derer biomolekulare Modifikatoren, wie Peptide, mit Prä-Nukleations-Strukturen interagieren und somit den Nukleationsprozess von Mineralen beeinflussen könnten, sind vielfältig. Molekulare Simulationen sind zur Analyse der Formation von Prä-Nukleations-Strukturen in Anwesenheit von Modifikatoren gut geeignet. Die vorliegende Arbeit beschreibt einen Ansatz zur Analyse der Interaktion von Peptiden mit den in Lösung befindlichen Bestandteilen der entstehenden Kristalle mit Hilfe von Molekular-Dynamik Simulationen.

Um informative Simulationen zu ermöglichen, wurde in einem ersten Schritt die Qualität bestehender Kraftfelder im Hinblick auf die Beschreibung von mit Calciumionen interagierenden Oligoglutamaten in wässrigen Lösungen untersucht. Es zeigte sich, dass große Unstimmigkeiten zwischen etablierten Kraftfeldern bestehen, und dass keines der untersuchten Kraftfelder eine realistische Beschreibung der Ionen-Paarung dieser komplexen Ionen widerspiegelte. Daher wurde eine Strategie zur Optimierung bestehender biomolekularer Kraftfelder in dieser Hinsicht entwickelt. Relativ geringe Veränderungen der auf die Ionen–Peptid van-der-Waals-Wechselwirkungen bezogenen Parameter reichten aus, um ein verlässliches Modell für das untersuchte System zu erzielen.

Das umfassende Sampling des Phasenraumes der Systeme stellt aufgrund der zahlreichen Freiheitsgrade und der starken Interaktionen zwischen Calciumionen und Glutamat in Lösung eine besondere Herausforderung dar. Daher wurde die Methode der Biasing Potential Replica Exchange Molekular-Dynamik Simulationen im Hinblick auf das Sampling von Oligoglutamaten justiert und es erfolgte die Simulation von Peptiden verschiedener Kettenlängen in Anwesenheit von Calciumionen. Mit Hilfe der Sketch-Map Analyse konnten im Rahmen der Simulationen zahlreiche stabile Ionen-Peptid-Komplexe identifiziert werden, welche die Formation von Prä-Nukleations-Strukturen beeinflussen könnten. Abhängig von der Kettenlänge des Peptids weisen diese Komplexe charakteristische Abstände zwischen den Calciumionen auf. Diese ähneln einigen Abständen zwischen den Calciumionen in jenen Phasen von Calcium-Oxalat Kristallen, die in Anwesenheit von Oligoglutamaten gewachsen sind. Die Analogie der Abstände zwischen

Calciumionen in gelösten Ionen-Peptid-Komplexen und in Calcium-Oxalat Kristallen könnte auf die Bedeutung von Ionen-Peptid-Komplexen im Prozess der Nukleation und des Wachstums von Biomineralen hindeuten und stellt einen möglichen Erklärungsansatz für die Fähigkeit von Oligoglutamaten zur Beeinflussung der Phase des sich formierenden Kristalls dar, die experimentell beobachtet wurde.

Contents

List of Figures

List of Tables

5.2 Statistics of successful exchange attempts between neighbouring replicas of the two systems containing a decamer of glutamate ("decaGLU") during the BP-REMD simulations. Exchange attempts are either made between even replica pairs (i.e. 0-1, 2-3, 4-5, and 6-7) or between odd replica pairs (i.e. 1-2, 3-4, and 5-6) at once. Replica 0: $\lambda = 0$, replica 1: $\lambda = 0.21$, replica 2: $\lambda = 0.40$, replica 3: $\lambda = 0.54$, replica 4: $\lambda = 0.68$, replica 5: $\lambda = 0.81$, replica 6: $\lambda = 0.92$, replica 7: $\lambda = 1.0$.100

Related Publications

[8] J. Kahlen, L. Salimi, M. Sulpizi, C. Peter, and D. Donadio. Interaction of charged amino-acid side chains with ions: An optimization strategy for classical forcefields. *The Journal of Physical Chemistry B*, 2014.doi: 10.1021/jp412490c.

Chapter 1

Introduction

1.1 General Aspects of Biomineralisation

The term *biomineralisation* describes a multitude of processes by which crystalline or amorphous inorganic solids are built in a biological environment. The degree to which the biological environment actively controls this mineralisation process varies. A coarse classification of mineralisation distinguishes biologically *induced* and biologically *controlled* mineralisation. In biologically induced mineralisation, inorganic solids are formed as byproducts of other important metabolic processes. On the other hand, biologically controlled mineralisation is a targeted, highly regulated method by which solid inorganic structures are formed that fulfil specific biological functions [9]. The extraordinary attributes of these biominerals (e.g. the mechanical stability of bones), that result from sophisticated fabrication strategies, have generated a large interest in the underlying mechanisms of these complex processes. In the following, we will therefore focus on biologically controlled mineralisation.

Prominent examples of natural materials that are produced *via* biomineralisation are the shells of crustaceans, the skeleton of vertebrates or teeth. Biominerals fulfil a great variety of functions in different organisms. Corresponding to this variety in functionalities of biominerals, there is a broad diversity in their fabrication processes, their composition and their structure. The most abundant constituents of biominerals are calcium carbonate, calcium phosphate, zinc oxide, calcium oxalate and silicon dioxide. They can consist of single crystals or can appear in a polycrystalline or amorphous form. In contrast to products of synthetic crystallisation processes, biominerals often exhibit properties such as well-defined structures and compositions, complex morphologies with preferential crystallographic orientations and high levels of spatial organisation [9]. The possibility to assemble into

hierarchical structures can be seen in the formation of mesocrystals (small crystals embedded in an organic matrix, which exhibit features of a single crystal due to their uniform crystallographic alignment) [10] or even biogenic composite materials in which inorganic crystals grow in organic matrices from an amorphous precursor.

The influence of the biological environment on the processes occurring during biomineralisation is manifold and can be described at various levels of "resolution". Starting at environmental and genetic parameters at a macroscopic scale, biomineralisation is regulated by a cascade of various factors [9, 3]. From the numerous processes occurring at different scales, we want to focus on the "physiochemical" perspective at a molecular scale here. At this scale, biomineralisation can be viewed as a combination of physical and chemical processes that occur during the crystallisation from complex solutions. The biological control over these processes is promoted through biogenic macromolecular structures that might consist of peptides, proteins, polysaccharides or lipids.

As we will see in the following, substantial research efforts have been devoted to understanding the many different ways in which biomolecules might influence the crystallisation process. A nice overview that illustrates the numerous possible mechanisms, by which additives in general may exert control on crystallisation, is given in figure 1.1. Not all of these mechanisms are observed in biomineral systems, but in almost all of the diagram's branches examples of biomolecular modifiers can be found. The possible regulatory functions of biomolecules in the different stages of biomineral formations are manifold: they might manipulate crystal nucleation (presumably by interacting with the solvated constituents of the mineral prior to or during nucleation or by serving as nucleation seeds) [11], crystal growth (by modulating the transport of ions to the surface or by selectively adsorbing on certain crystal faces where they either promote or suppress the growth of this face) [12, 13, 14], and they might stabilise amorphous intermediate phases [15, 16]. In some systems, biomacromolecules such as peptides or proteins get incorporated into the inorganic crystals, yielding organic–inorganic hybrid materials with fascinating mechanical properties [17, 18]. They may also act as structural templates by forming larger aggregates, which serve as matrices or compartments for the subsequent crystal growth [19]. By means of these different mechanisms, the biological additives may provide control over various aspects of the morphology of the forming crystals, such as size, shape and even composition.

It is important to note, that large differences in the effectiveness of biological modifiers exist. Properties such as the size and flexibility of the modifier molecules and the number and chemical sequence of their functional groups

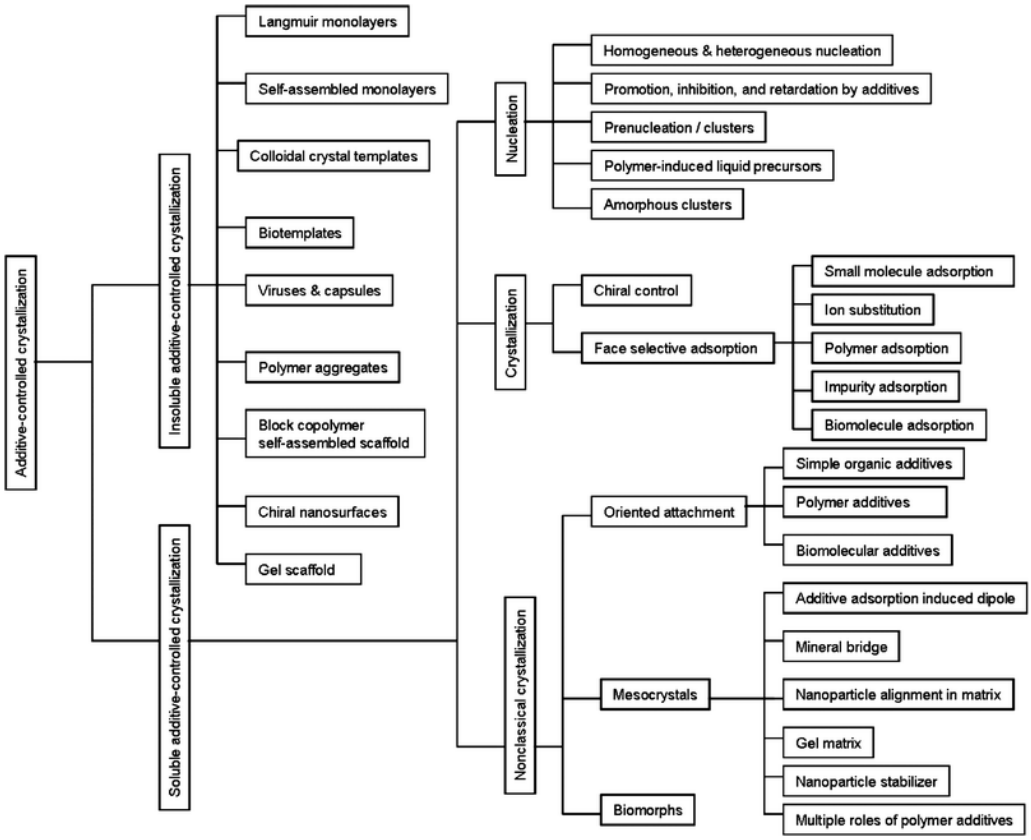


Figure 1.1: Overview over the many possible mechanisms by which additives such as biomolecules might manipulate the crystallisation process. Reproduced from Song and Cölfen [1] with permission of The Royal Society of Chemistry.

play important roles [20, 21, 22, 23, 24]. As a consequence, the additives might for example affect the crystallisation process only if the chain-length exceeds a certain threshold or in case of a specific sequence of functional groups. Even though a lot of effort has been devoted to understanding these correlations, it has not yet been possible to completely unravel the basic principles underlying the specificity of modifier–mineral interactions at a molecular scale.

Besides the general scientific interest to decode the underlying principles of the effects of biomolecular modifiers on crystallisation processes, the study of biomineralisation is directed by the aim to imitate the principal biological strategies to control the morphologies of amorphous and crystalline inorganic materials. The application of biomimetic pathways to the formation of minerals with controlled morphologies has become a very important research area in materials science. If it was possible to understand the molecular mechanisms that give rise to the formation of complex shapes, hierarchical structures and unique properties of biominerals, they could be applied as fabrication strategies in materials design. Possible applications of such products with tailored properties are seen in the fields of microelectronics, optics, catalysis, pigments, pharmaceuticals, cosmetics and biomedical implants [25].

1.2 Biomineralisation of Calcium-Containing Minerals

As described above, biomineralisation is an enormous topic with a large variety in components, phenomena and theories regarding its physical principles. In this vastly complex field, we have here focussed on phenomena occurring in a specific system. Calcium-containing minerals were selected as our subject of study due to their high abundance in nature - they belong to the most significant naturally occurring minerals - and rich polymorphism, which have created considerable interest in research. The number of experimental and theoretical studies on the formation of these minerals is steadily increasing, and a large variety of mechanisms, by which biogenic modifiers might manipulate this process, has been proposed. The most prominent examples of this class of minerals are calcium carbonate, calcium phosphate and calcium oxalate.

Calcium carbonate ($CaCO_3$) is a major constituent of rocks like marble, chalk, limestone or coral reefs. As a product of biomineralisation, it is the main component of shells of marine organisms, snails, and eggshells, where it has a protective role. Furthermore, it fulfils supportive functions in skele-

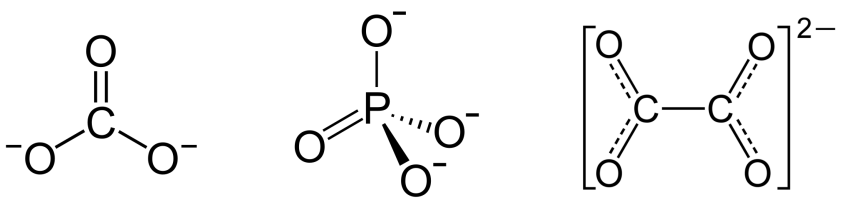


Figure 1.2: Schematic structure of the anions that form calcium carbonates, calcium phosphates and calcium oxalates together with the cation Ca^{2+}; left: carbonate, middle: phosphate, right: oxalate.

tons or might even serve as an optical material in primitive vision systems [26, 27, 28, 6]. Six different forms of calcium carbonate exist: three crystalline polymorphs (calcite, aragonite, and vaterite), two hydrate phases, and amorphous calcium carbonate (ACC). The diversity in material properties of the different forms has raised a lot of attention and has made calcium carbonate one of the most extensively studied biominerals [19]. Numerous modifiers are known which affect the growth of calcium carbonate. Most of these modifiers have in common, that they contain negatively charged functional groups. Examples of such additives are acidic amino acids (e.g. aspartate or glutamate) and corresponding peptides or proteins [29, 30, 31], small organic molecules such as acetate or citrate [6], and synthetic polyelectrolytes such as polyacrylates [32] or polystyrene sulfonates [33, 34]. Compared to other minerals, many details of the biogenic control over mineral formation of calcium carbonate have been unravelled, as will be described below.

Calcium phosphate is considered as a key component in human health [35]. Hydroxyapatite (HAP), which is one of the numerous phases of calcium phosphate, is the most common mineral in vertebrate tissues, where it confers mechanical stability to bones and teeth. An overview over the different crystalline and amorphous forms of calcium phosphate can be found in table 1 of Bleek and Taubert [35]. Inspired by the various natural occurrences, calcium phosphate powders, coatings, composites and ceramics have been studied with respect to possible industrial applications [36, 37] and further applications in the fields of drug delivery and bone or dental repair are expected [38]. Biomineralisation of calcium phosphate has therefore received a lot of attention in research [39, 40]. The modifiers that have been studied are similar to those that have been examined for calcium carbonate, and they consist mostly of charged species such as polyelectrolytes (oligopeptides or synthetic polymers) or low-molecular-weight additives [41]. For example, charged peptides consisting of aspartate (negatively charged), or arginine

(positively charged) have been found to affect both the nucleation rate and the morphology of HAP [42, 43].

Calcium oxalate can be found in more than 1000 genera of plants, where it participates in the regulation of the concentration of free calcium ions in the cytosol [44, 45]. The hydrated forms of calcium oxalate occur in the rare minerals whewellite (monohydrate), weddellite (dihydrate) and caoxite (trihydrate). In industry, it is mostly recognised for the formation of scale, which blocks pipes, pumps or heat exchangers and thus causes energy losses. Most interesting with respect to biomineralisation is the formation and suppression of kidney stones, which mainly consist of calcium oxalate monohydrate ("COM": $CaC_2O_4 \cdot H_2O$). Several natural compounds are known for their potencies to inhibit the growth of kidney stones, e.g. citrate [46], amino acids [47, 48], acid-rich urine-proteins [49, 50], and osteopontin (an aspartic and glutamic acid-rich protein) [51, 52, 53]. This shows that - similar to calcium carbonate and calcium phosphate - oligomers of negatively charged amino acids strongly affect the crystallisation process of calcium oxalate.

An Example: Biomineralisation of Calcium Oxalate in the Presence of Oligoglutamates In order to illustrate the approach, by which the effects of potential modifiers on the crystallisation process might be studied and to exemplify the phenomena which are observed in such experiments, we will shortly summarise a recent study on the influence of oligomers of glutamate (the deprotonated form of the natural amino acid L-glutamic acid) on the crystallisation of calcium oxalate. Fischer et al. have conducted a systematic analysis of the effect of the chain length and the number of complexing functional groups of oligomers of glutamate on the crystallisation of calcium oxalate [2]. For this purpose, they have studied the kinetics of the precipitation process and the characteristics (morphology, phase and stability) of the resulting calcium oxalate crystals in the presence of oligoglutamates ("oligoGLU": cf. figure 1.3) with well-defined chain lengths. Three homooligopeptides of glutamate with chain lengths of five ("pentaGlu"), ten ("decaGlu"), and twenty units ("icosaGlu") were used during the experiments. Due to a pH value of 8.5, the carboxylate groups of the side chains and the C-terminus were completely deprotonated, while the N-terminus was protonated, thus yielding a zwitterionic structure.

During the experiments, Fischer et al. monitored the desupersaturation rates *via* light-transmission and conductivity measurements in order to measure the length of the nucleation period, the maximum growth rates and the equilibrium solubility of calcium ions. Furthermore, they analysed the size,

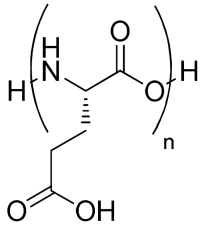


Figure 1.3: Molecular structure of oligomers of glutamic acid. In the experiments of Fischer et al., pentamers (n = 5), decamers (n = 10), and icosamers (n = 20) of glutamic acid were used in a deprotonated form ("glutamate") [2].

shape, water content and stability of the resulting crystals. They found that the influence of glutamate on both nucleation and crystal growth strongly depends on the chain length. While pentaGLU shows only very limited effects, the presence of decaGLU or icosaGLU significantly modifies the process and the product.

Figure 1.4 shows the changes in the concentration of free calcium ions in solution in the presence of different oligoglutamates during the precipitation. The induction time as one descriptor of the nucleation process is increased by a concentration of 0.1 mmol·L^{-1} of decaGLU (factor of 2.6) or icosaGLU (factor of 4.6) with respect to the precipitation in the absence of peptides, while pentaGLU has an almost negligible effect on the induction time. A further indication of the manipulation of the nucleation process by oligoglutamates can be seen in the phase of the resulting crystals. Under the given reaction conditions, calcium oxalate trihydrate ("COT": $CaC_2O_4 \cdot 3H_2O$) is formed in the absence of modifiers and in the presence of pentaGLU. An increase in the concentration of longer peptides (decaGLU or icosaGLU) however leads to a gradual change towards the formation of calcium oxalate dihydrate ("COD": $CaC_2O_4 \cdot 2H_2O$). Fischer et al. assume that complexes of peptides and calcium ions might serve as nucleation clusters and thus determine the phase of the resulting crystal [2].

The maximum growth rate of calcium oxalate crystals is also affected by the chain length of the oligoglutamates (cf. gradients of curves in figure 1.4). This might have several reasons. The equilibrium calcium concentrations in the end of the experiments (cf. figure 1.4) indicate that the peptides influence the solubility of calcium depending on their chain length. The increased solubility of ions, which is observed for longer chains of oligoglutamate, reduces the level of supersaturation and thus the driving force for precipitation. Furthermore, it is likely that the peptides adsorb to the crystal surface where

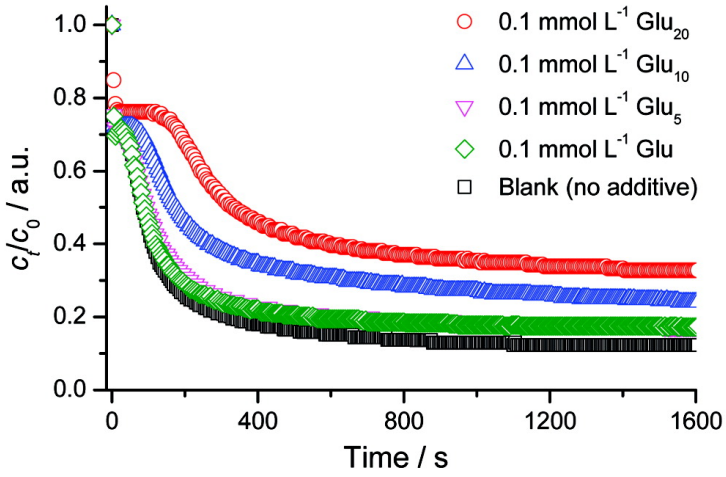


Figure 1.4: Relative concentration of free Ca^{2+} ions in solution during the experiments of Fischer et al.. The ion concentration c_t was measured *via* Ca^{2+}-selective electrode and normalised by the initial concentration c_0. It was measured in the absence and in the presence of oligomers of glutamate (GLU). Reprinted with permission from Fischer et al. [2]. Copyright 2011 American Chemical Society.

they slow down the crystal growth.

The classical model of crystal growth assumes that crystal surfaces consist of terraces, steps and islands (cf. figure 1.5). Single solvated ions adsorb to the different regions of the crystal surface with different probabilities, the highest probability being located at kinks. This attachment at kinks leads to a step advancement perpendicular to the step edge. At equilibrium, the rates of attachment and thermally activated detachment of ions from the crystal are equal. In the presence of modifiers, kinetic and thermodynamic

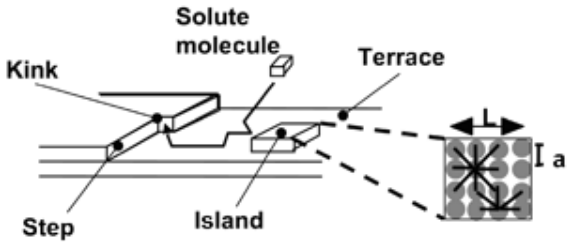


Figure 1.5: Simple model of crystal surfaces with terraces, steps, kinks and islands. Reprinted with permission from De Yoreo and Vekilov [3]. Copyright 2003 Mineralogical Society of America.

phenomena might affect the crystal growth [54, 55]. Modifiers might change the step speed or the step-edge energy *via* four different mechanisms: 1) step pinning, 2) incorporation, 3) kink blocking, and 4) step-edge adsorption [3]. Each of these mechanisms exhibits a different characteristic dependence of the step speed on supersaturation and impurity concentration, as illustrated in figure 26 of De Yoreo and Vekilov [3].

The *step pinning* model was first introduced by Cabrera and Vermilyea [56]. It assumes that impurities such as modifiers block the attachment of ions to the step edges. The steps may continue to advance between these pinning sites, but a step curvature is induced which raises the step edge energy. Depending on the modifier concentration and the corresponding density of pinning sites along the step, this curvature can result in a reduced step speed or in a complete inhibition [3]. If molecules of the modifier adsorb to the crystal surface, they might become captured by advancing steps and become incorporated into the growing crystal. The *incorporation* of impurities usually distorts the crystal structure and thus raises the internal energy and the entropy of the crystal. While an increase in the entropy of the crystal stabilises the crystal, an increase in the internal energy has the opposite effect. Depending on which of the two effects predominates, the solubility of the constituents of the crystal changes, which alters the crystal growth rate as well [57, 58, 59]. In contrast to step pinning, *kink blocking* is only a temporary effect. For very short residence times, modifier molecules adsorb to kink sites and thus reduce the effective kink site density. This temporary retardation of kink propagation reduces the rate of step growth [3]. The model of *step-edge adsorption* assumes that molecules of the modifier adsorb to step edges, where they modify the step-edge energy. At a sufficiently high concentration of adsorbed modifier molecules, the minimum energy step shape might be altered, resulting in a new crystal shape [3]. An example of this effect is described in the work of Orme et al., who observed that the adsorption of individual enantiomers of amino acids to step edges induced chirality in the step and facet morphologies of calcite [21, 24].

In the experiments of Fischer et al., it was observed that the dependence of crystal growth inhibition on peptide concentration resembles a Freundlich-type adsorption isotherm, where the surface coverage is proportional to the concentration in solution. A simple model to explain this observation assumes that the inhibition of crystal growth is directly proportional to the coverage of modifier molecules on top of the surface. If the surface coverage is proportional to the overall peptide concentration, the fraction of the surface covered by peptide might be described by a Freundlich-like isotherm. However, this model does not contain any information on the exact mechanism, by which the adsorbed peptides inhibit the crystal growth.

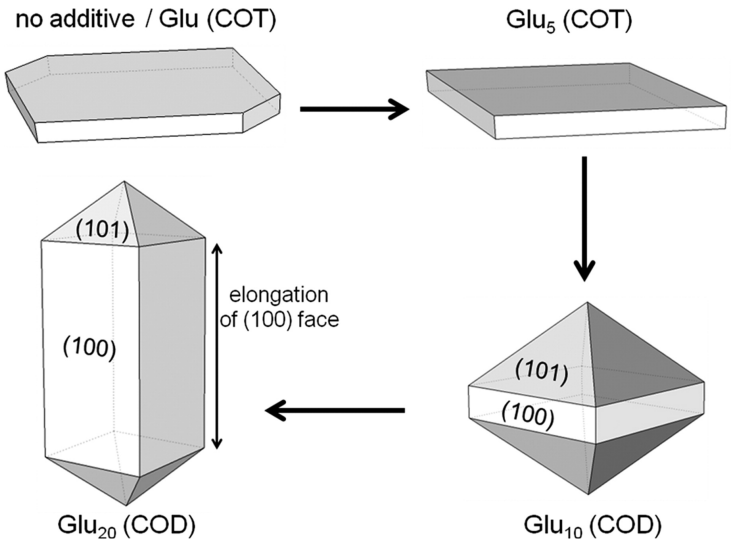


Figure 1.6: Schematic diagram of the morphological evolution of calcium oxalate crystals in dependence of the chain length of the oligoglutamates present during precipitation. Reprinted with permission from Fischer et al. [2]. Copyright 2011 American Chemical Society.

An interesting effect of the interactions between oligoglutamates and calcium oxalate crystals is the dependence of the shape and size of the final crystals on the concentration and chain length of the peptide additives. Fischer et al. have summarised the morphological evolution of the calcium oxalate crystals in dependence of the chain length of the oligoglutamates in the scheme shown in figure 1.6. In general, the equilibrium shape of a crystal is determined by the minimal surface free energy of the crystal. Therefore, faces with small surface free energies are preferentially expressed [3]. As described above, additives might interact in several ways with crystal surfaces (for example *via* step-edge adsorption) and thereby modify the surface free energy of the respective crystal face, which can result in a change of the overall crystal shape. Depending on the crystallographic direction, the components of the crystal are exposed to the surrounding solution to different degrees at the various faces of the crystal. For example, calcium ions are more exposed to the solution at the 100 faces than at 101 faces of calcium oxalate crystals. It can thus be assumed, that oligoglutamates are more likely to interact with 100 faces of calcium oxalate *via* their negatively charged carboxylate groups, than with the 101 faces. Fischer et al. therefore assume, that the specific adsorption of the longer chains of oligoglutamates to the 100 faces promotes the growth in the [001] direction and results in smaller 101 faces (cf. figure 1.6) [2].

Apart from the adsorption of oligoglutamates to the surface of calcium oxalate crystals, it is likely that the peptides also get incorporated into the crystal. The incorporation of peptides (which resemble sequences of the protein osteopontin) into COM has been described before by O'Young et al. [60]. Long-term observations by Fischer et al. showed that the initial phase of calcium oxalate crystals, which precipitate in the presence of the longer chains of oligoglutamates (decaGLU and icosaGLU), is preserved even if it is metastable under normal conditions. In the absence of peptides or in the presence of pentaGLU, COT is initially formed and then slowly converts into the thermodynamically stable phase of COM. But in the presence of longer peptide chains, the mixture of COD and COT, which precipitates from solution, is preserved for several months [2]. A possible explanation for this observation is, that the longer peptides get incorporated into the crystal and thereby stabilise the initial phase. The incorporation of peptides into the crystal would also explain, why the final crystals exhibit large sizes of a few micrometers in spite of crystal growth hindrance through peptide adsorption.

Based on their experimental results, Fischer et al. assume that oligomers of glutamate manipulate the formation of calcium oxalate crystals at different stages of the crystallization process. In solution, the peptides and calcium ions might form a coordination complex that alters the nucleation process, while the adsorption of peptides to the crystal surface changes the crystal growth kinetics and the morphology of the final crystals [2]. These explanations are plausible, but they lack any detailed information on the processes occurring at a molecular scale. Several questions remain unanswered: Do peptides and calcium ions form aggregates in solution? Do these aggregates really influence the phase of the forming crystals? How do the peptides adsorb to the various faces of calcium oxalate? How do they inhibit the growth of the crystal faces? How do the peptides get incorporated into the crystal? Why is the effect of pentaGLU limited, while decaGLU significantly alters the crystallisation process? As will be shown in the next section, molecular simulations are well suited to study many of these processes in detail.

1.3 Biomineralisation at a Molecular Scale

The numerous phenomena that occur on different time and length scales during biomineralisation can be studied by various methods of computer simulations that range from *ab initio* calculations over all-atom simulations to coarse-grained simulations. Chemical aspects such as hydrolysis reactions at the aqueous interface need to be treated on a quantum mechanical

level. Classical molecular dynamics simulations are applied to study molecular processes at the organic/inorganic interface such as solvation properties, ordering phenomena regarding the structure of the aqueous phase, or the adsorption of molecules to crystal surfaces. Advanced sampling algorithms and rare-event methods are used to study the nucleation of minerals, while coarse-grained models are developed to model the self-assembling of nanoparticles into crystals for example [61]. In the last years, molecular simulations have mostly been focussed on two aspects of biomineralisation: the adsorption of organic modifiers on crystal surfaces and the details of the nucleation process, including its manipulation by modifiers.

Face-selective Adsorption The adsorption of organic molecules to crystal surfaces and step edges is that aspect of modifier–crystal interactions, which has been studied most intensively *via* molecular simulations in the past years [62, 63]. From experiments it is known that several factors determine the strength of interactions between modifiers (such as peptides) and crystal surfaces and thus the effectiveness of peptides to inhibit crystal growth: the modifier's functional groups and their sequence, the chain length and flexibility of the modifier, its concentration and the degree of supersaturation in solution [24].

The peptide sequence is believed to influence the crystal growth through the stereochemical relationship between step and modifier. Early studies focussed on modifications of growing calcite crystals through proteins and peptides containing aspartic acid residues. They compared the molecular structures of newly expressed crystal faces (i.e. faces *un*expressed in the absence of the modifier) to the functional groups of the modifiers and to the intramolecular distances between these functional groups and postulated that stereochemical effects were responsible for the changes in crystal morphologies [20]. In the case of calcite crystals grown in the presence of aspartic-rich peptides, a stereochemical match of negatively charged carboxylate groups of the peptides to calcium ions at the surface of the calcite crystal lattice was found [24]. Similar effects during the adsorption of peptides to calcium oxalates and calcium phosphates were investigated intensively by the groups of Graeme Hunter and Mikko Karttunen. Interested in the effects of the phosphorylation of peptides and proteins such as osteopontin on the crystal growth, they used a combination of experimental techniques and classical and *ab initio* molecular simulations to study the exact mechanisms, by which crystal growth is inhibited [52, 60, 64, 65, 66]. According to their findings, the adsorption of peptides to crystal faces is governed by electrostatics and facilitated by the conformational flexibility of the polypeptide chain. That

means that the potency to interact with crystal surfaces is higher for those modifiers, which are able to fit the arrangement of their functional groups (which are oppositely charged with respect to the crystal surface) to the crystal lattice, compared those modifiers with rigid structures (e.g. imposed by secondary structure) in which a good stereochemical match of the interacting sites is necessary. In the case of the adsorption of acidic peptides to calcium-containing biominerals they observed strong interactions between the negatively charged carboxylate groups and the calcium ions of calcium-rich crystal faces. Similar observations were made by Thomas et al., who studied the formation of calcium oxalate dihydrate in the presence of polyacrylic acid. It was found that the flexibility of the polyacrylic acid chains, which consist of 25 monomeric units, enables the oligomers to adopt conformations in which the oxygen atoms of the carboxylate groups interact with the calcium ions at the surfaces of COD [67]. The dominant role of anionic side chains in the inhibition of calcium-containing minerals was analysed systematically by the groups of Michael Ward and Jeffrey Wesson. By means of Atomic Force Microscopy (AFM) they measured the adhesion forces between functional groups (e.g. carboxylate-, hydroxyl-, amine- or methyl-terminated tips) and different faces of COM crystals in aqueous solution and found that the adhesion forces for the carboxylate group are significantly larger than those of most other functional groups [68, 69]. Furthermore, they were interested in the importance of the length of the side chains and analysed polyacrylate, polyaspartate and polyglutamate, which mostly differ in the length of the side chains, as modifiers. The surface morphology and step growth of COM in the presence of the modifiers were monitored by means of AFM. It was discovered that the steps and faces of the crystal, to which the three polymers preferentially adsorb, differ depending on the length of the side chain, even though all three polymers interact *via* carboxylate groups with the COM crystals [70].

The size of a modifier, i.e. its chain length, can also have a tremendous influence on its potency to inhibit crystal growth. The reasons for this observation seem to be numerous. For relatively small modifiers such as citrate, classical models suffice to describe the inhibition effects on COM: electrostatic interactions lead to the adsorption of the negatively charged citrate to positively charged faces of COM where it hinders growth via step pinning, while the negatively charged faces (exposing oxalate) are not affected by the presence of citrate [71, 46]. For larger modifiers, such simple electrostatic considerations are not sufficient to understand the interactions with crystal surfaces. When Elhadj et al. measured the inhibition of the growth of calcite by oligomers of aspartate, they found that the peptide concentration, which is necessary to completely stop the step growth, decreases by five orders of

magnitude, when the chain length is increased from one to six monomeric units [22]. However, with increasing chain length kinetic effects also gain in relevance. Friddle et al. studied the growth of COM under the influence of aspartic acid-rich polypeptides with a chain length of 27 residues. As observed in previous experiments with similar components, they observed high densities of pinning sites on calcium-rich surfaces by means of AFM. The inhibition of crystal growth through the peptides however significantly depends on the peptide chain length and the degree of supersaturation in the solution. At high degrees of supersaturation, a high step speed is observed over a wide range of peptide concentrations. But below a certain degree of supersaturation, the step speed suddenly drops to zero. A possible explanation for this observation is the slow process of peptides binding to surfaces, to which they adsorb through site-specific interactions. Here, a number of kinetic barriers need to be overcome in order to reach a lowest-energy well-bound state: the displacement of shielding counter-ions, the coordination with oppositely charged sites and the configurational relaxation of the polymer. The potential to inhibit the step advancement depends on the degree, to which such well-bound states have been adopted by the polymers. Weakly bound peptides might be displaced by the advancing step, while well-bound peptides can block the access of dissolved ions to kinks. Depending on the degree of supersaturation and on the peptide concentration, two time scales compete with each other: the terrace lifetime (terrace width divided by step speed), which depends on the degree of supersaturation, and the characteristic time that the peptide needs to relax into a well-bound state, which significantly depends on the peptide chain length.

Another important factor in the adsorption process is the structuring of solvent at the interface between crystal and solution. In aqueous solutions, crystals induce a strong structuring of water on top of their surfaces [62]. This modification in the local water structure affects the barriers which the solvated ions and the modifier molecules face during their adsorption to the crystal surfaces [72, 73]. Modifier molecules, which are adsorbed to the crystal, may corrupt the structure of these water layers and thereby facilitate the desolvation of the ions and lead to *increased* crystal growth rates [74, 22, 24]. This effect depends on the concentration of the modifier. At low peptide concentrations, Elhadj et al. observed enhanced step growth rates, whereas higher concentrations of the same peptides led to a complete inhibition of crystal growth [22].

Nucleation from Solution In classical nucleation theory the crystal formation process starts with stochastic fluctuations in the local ion concentra-

tion in solution, which leads to the aggregation of ions and to a formation of an interface between nucleus and surrounding solution. As this interface generates a positive surface free energy and the structure formation is accompanied by an entropic penalty, the emerging clusters are unstable. The driving forces of the aggregation are the supersaturation of the solution and the favourable bulk lattice energy of the forming crystal. As the size of the clusters (which are assumed to be of spherical shape) increases, the ratio of surface to volume decreases. At some point during growth, the cost of the surface energy is outweighed by the bulk lattice energy, the system passes through a free energy maximum and the unstable aggregates gain in stability [75, 76].

However, in the past years more and more evidence has been found that - even in the absence of any modifier - the nucleation of mineral crystals may not follow the route described in classical nucleation theory. Several types of non-crystalline ionic structures have been observed experimentally, which serve as precursors to the forming crystals: amorphous precursors to the different polymorphs (such as amorphous calcium carbonate: "ACC"), droplets of a dense ion-rich liquid phase as products of liquid–liquid phase separation (sometimes referred to as a liquid condensed phase: "LCP"), and pre-nucleation clusters ("PNC"). Based on these observations, it has been proposed that the formation of several minerals follows a multi-step crystallization process, which starts with a liquid precursor phase (LCP or PNC), that is transformed into a solid amorphous precursor, and eventually becomes crystalline [77, 76, 78]. The exact mechanisms of transformation between these different types of species are still unclear and it is still discussed intensively, under which conditions they occur (e.g. pH values, temperature range, presence of additives). While most of the studies on non-classical nucleation of minerals have been focussed on calcium carbonate in the past, relatively little is known about other systems.

The smallest stable structures that have been observed beyond simple ion pairs in solution prior to nucleation are pre-nucleation clusters. A precise definition of PNCs has lately been formulated by Gebauer et al. [76]. These supramolecular ionic polymers have been studied experimentally as well as in computer simulations and were found to be composed of alternating ions (e.g. calcium and carbonate) with dynamic topologies consisting of chains, rings and branches [79, 80]. The disordered, flexible and strongly hydrated structure of PNCs ensures that they remain in equilibrium with the solution and avoid a phase boundary [80, 6]. Gebauer et al. have shown that PNCs of calcium carbonate form in undersaturated, saturated and supersaturated conditions [79]. The cluster size distribution varies considerably with the degree of supersaturation and the pH value, varying from small (< 20 ions)

clusters at a pH of 8.5 to larger clusters (up to 200 ions) at a pH of 11.5, which corresponds to hydrodynamic diameters of 2-3 nm in the case of calcium carbonate [80, 79]. Due to their constantly changing structure, the PNCs of calcium carbonate have been named as a "dynamically ordered liquid-like oxyanion polymer" (DOLLOP). In contrast to amorphous calcium carbonate (ACC), the dominant coordination number of calcium and carbonate ions in DOLLOP is 2 [80, 75]. An interesting observation was made by Wallace et al., who found that the free energy of calcium-carbonate clusters essentially decreases monotonically with increasing size, which is a clear disagreement with classical nucleation theory.

Based on other experimental findings, an alternative route to the mineralisation of calcium carbonate, which evolves from liquid–liquid phase separation, has been proposed [82, 83, 77, 81]. Stable droplets of a dense ion-rich liquid phase of $CaCO_3{\cdot}nH_2O$ were found to develop in supersaturated calcium carbonate solution through a liquid–liquid phase separation in the absence of any additives. In view of the liquid nature of the phase, it has been referred to as a "liquid condensed phase" (LCP). In the experiments of Bewernitz et al., the droplets were approximately 60 nm in diameter and displayed an emulsion-like behaviour with a resistance to coalescence [77], which has also been observed by Wolf et al. [83].

For a few years, the correlation between these two precursor structures has remained unclear, but recently, it has been found that LCP may evolve from PNCs. Under certain conditions (e.g. increasing supersaturation), a breakdown of the equilibrium between PNCs and solution is observed. At this point, thermodynamics favours liquid–liquid separation, PNCs change their speciation and become less dynamic, they develop interfaces, and become nanodroplets [81, 76]. Wallace et al. analysed this process in detail and found that the evolution of coordination numbers of calcium and carbonate ions and of the ion diffusivity show that the structural and dynamical properties of the clusters smoothly adopt those values found in LCP [81]. PNCs can therefore be regarded as the direct precursors of liquid nanodroplets [76]. An open question in this context that has not yet been answered is, at what size large PNCs with slow dynamics can be viewed as nanophases, and can be associated with an interfacial surface rather than a hydration layer [76].

The next step of the nucleation process is the transformation of LCP to an amorphous precursor, such as ACC in the case of calcium carbonate. The nanoscopic droplets aggregate and form larger entities that undergo progressive dehydration to give solid nanoparticles [76]. Indirect evidence of this process was found by Wallace et al., who observed in computer simulations, that the structure of these agglomerated and dehydrated clusters is consistent with the structure of ACC [81].

Of the different precursor structures of crystalline calcium carbonate, ACC is the one that has been studied most extensively [16, 84, 85, 86]. ACC occurs in a variety of types with varying degrees of water content [87], incorporation of impurities (such as organic molecules) [88] and polyamorphism [89]. Compared to the crystalline polymorphs of calcium carbonate, ACC features a highly disordered structure with a higher solubility. It is the most unstable form of solid calcium carbonate at ambient conditions and is subsequently transformed into more stable bulk crystalline polymorphs *via* dehydration and subsequent crystallization according to Ostwalds rule of stages [90, 87].

In summary, Gebauer et al. describe the crystallisation of calcium carbonate along the *PNC pathway* as a "progressive, step-wise loss of hydration water, according to the sequence $Ca^{2+}_{(aq)}/HCO^{-}_{3\,(aq)}/CO^{2-}_{3\,(aq)} \rightarrow$ PNCs $\rightarrow$ dense liquid nanodroplets $\rightarrow$ liquid ACC $\rightarrow$ solid ACC $\rightarrow$ anhydrous crystalline polymorphs" [76]. However, it may be that not all of these steps always occur during mineralisation. For instance, the direct creation of a nanoscopic solid phase through aggregation of PNCs is still discussed as a possible pathway which omits LCP [76]. The pH value might also influence the nucleation route, as LCP - in contrast to PNCs - has been observed at near-neutral pH in supersaturated calcium carbonate solutions [91, 77]. The hydrogen carbonate ions ("bicarbonate") that occur at these pH values might therefore trigger the direct formation of LCP without prior formation of PNCs [77].

Additives, such as polyelectrolytes, may also significantly affect the nucleation process. The stabilising effect of certain poly(carboxylic acids) on LCP of calcium carbonate was first described by Gower and Odom, who named this phenomenon "polymer-induced liquid precursors" (PILP) [92, 78]. Several polyelectrolytes were shown to stabilise LCP of calcium carbonate probably by effectively inhibiting the release of hydration water and by stabilising the nanodroplets colloidally [93, 94, 77, 76]. The polyelectrolytes may not only stabilise LCP, but they may also direct the transformation of LCP towards the formation of non-equilibrium crystalline morphologies [92, 95, 78]. But the manipulation of the morphology of the forming crystals by modifiers might start even earlier during precipitation. It has been observed experimentally, that calcite (the thermodynamically stable crystalline phase of calcium carbonate at ambient temperature and pressure) probably results from a stable form of ACC, while vaterite (a metastable crystalline polymorph) shows structural similarities to a less stable form of ACC. These different forms of ACC in turn apparently originate from different PNCs that were formed at different pH values and exhibit different stabilities themselves [79, 76]. These findings suggest that the structure of PNCs correlates with the morphology

of the evolving crystalline phase. If additives such as peptides or other polyelectrolytes were able to manipulate the structure of PNCs in solution, they might influence the morphology of the resulting crystal through a cascade of structural motif encoding. This would be a plausible explanation of the ability of many organic macromolecules to steer the selection of crystalline polymorphs, as it is often observed during biomineralisation.

Pre-nucleation clusters have not only been reported for calcium carbonate, but also for calcium phosphate, iron oxide, silica, and even some amino acids, but the structure and stability of the precursors and the details of the transition pathways during crystallisation are less well-known than in the case of calcium carbonate [76, 35]. Even for calcium carbonate, which has been studied extensively in the past, numerous open questions regarding the exact mechanisms of precipitation from supersaturated solutions remain. For example, the dynamics of the transformations between the different precursor structures and the average lifetimes of PNCs and nanodroplets are still unknown. The molecular structural explanation for polyamorphism (the polymorphism of ACC) and its connection to PNCs and LCP are still unclear. It is furthermore of interest, whether the formation of solid amorphous phases from liquid nanodroplets proceeds *via* nucleation (possibly induced by modifiers) or simply dehydration. Another open question is, whether crystals may nucleate directly from PNCs, or phase-separated nanodroplets are required for the nucleation of solids [76].

1.4 Motivation and Outline of the Thesis

A lot of progress in the understanding of the details of crystal nucleation and growth and the effectiveness of additives at a molecular scale has been made in the past years. The adsorption of modifiers to crystal surfaces, which may significantly affect the crystal growth during biomineralisation, has been studied for several combinations of minerals and modifiers already. The simulation of such hetero-interfaces by means of classical molecular simulations is a great challenge, as hard and soft matter are usually described by different types of models (i.e. forcefields) which are often not compatible. Simulations of these systems therefore either need to be based on extensive development and testing of classical forcefields or should employ first-principles DFT-based molecular dynamics in order to circumvent these problems. The latter approach is currently taken by my colleague Leila Salimi (MPI-P Mainz and University of Mainz) for studying the interactions of oligoglutamates and calcium oxalate crystals with the goal to rationalise some of the experimental

findings of Fischer et al. on this system [2].

In contrast to crystal growth, the earliest stages of the crystallisation process, i.e. phenomena occurring prior to and during nucleation, are not yet well-understood. Even for calcium carbonate, one of the best-studied systems in this respect, important details of the processes at a molecular scale are still unknown. The knowledge regarding phenomena such as pre-nucleation structures is even more limited for other mineral systems (e.g. calcium oxalate). Several observations of Fischer et al., such as the retardation of nucleation (cf. figure 1.4) and changes in the degree of hydration of calcium oxalate (cf. figure 1.6) in the presence of oligoglutamates, are strongly depended on the concentration and especially the chain length of the modifier. These phenomena might originate from the early stages of nucleation, where single solvated ions (or small clusters of them) interact with the biomolecular modifiers. Relatively little is known about the interactions of modifiers (such as peptides) with pre-nucleation structures. Several scenarios regarding the manipulation of the nucleation process by additives are possible: (i) complexation of single solvated ions by the modifier, (ii) stabilisation of pre-nucleation structures (PNCs or LCPs) by the modifier, (iii) manipulation of the structures of PNCs and LCPs by the modifier, (iv) inhibition of the aggregation of LCPs by the modifier, (v) stabilisation of amorphous polymorphs (e.g. *via* incorporation) by the modifier.

The major challenge with respect to simulations of nucleation processes are the time scales on which nucleation occurs. Due to the discrepancy between the necessary time scales (order of seconds) and the currently feasible simulation times of classical simulations (order of microseconds), it is not yet possible to simulate the complete nucleation process. Nevertheless, it is possible to obtain important information on the interactions of crystal constituents and modifiers and on the emerging structures (such as PNCs) *via* classical molecular simulations. Of the five scenarios mentioned above, the first three can be studied by classical MD simulations, while the latter ones require coarse-grained or mesoscale simulations.

The complexation of single solvated ions by the modifier might have several implications for the nucleation process. Depending on the relative concentrations of solvated ions and modifier molecules, the modifier might "withdraw" considerable amounts of free ions from solution. Frey-Wyssling studied the precipitation of calcium oxalate experimentally, and found that the ratio of calcium to oxalate ions affects the hydration of the forming crystals. An excess of calcium ions leads to the formation of calcium oxalate dihydrate, while an excess of oxalate ions results in the monohydrate form [96, 45]. If a modifier preferentially interacted with only one type of ions (e.g. calcium) and avoided contact to the other type of ions (e.g. oxalate),

an imbalance in the free ion concentrations of the two ionic species would be created. Even if the respective total ion concentrations in the system were equivalent, this might result in phenomena comparable to those observed by Frey-Wyssling.

The modifier might not only interact with single ions in solution, but also with pre-nucleation structures such as PNCs and LCPs. Bewernitz et al. observed that polyaspartate stabilises LCPs of calcium carbonate (PILP process) [77]. A similar effect might occur during the formation of calcium oxalate in the presence of oligoglutamates. The stabilisation of LCPs by the peptides might be responsible for the significant retardation in nucleation, which has been observed by Fischer et al. [2]. However, the modifier might not only stabilise pre-nucleation structures, but it could even function as a "nucleation-seed" and initialise their formation. Depending on the size of the modifier molecules, they may each interact with multiple ions simultaneously. The pattern of functional groups of the modifier may impose a certain structural pattern on the relative positions of the ions with respect to each other. This was for example observed by Bulo et al., who analysed the distances between calcium ions which were bound to polyacrylates in solution and found "a local arrangement reminiscent of the crystal structure of $CaCO_3$ minerals" [97]. If the modifier served as a nucleation centre which favours distinct structural patterns, the complete crystallisation process might be influenced through a cascade of structural motif encoding, and modifiers might thereby steer important crystal properties such as the selection of polymorphs. According to this scenario, the differences in the phase of calcium oxalate crystals, which Fischer et al. observed in dependence of the chain length of the oligoglutamates [2], might therefore originate from the complexation of single calcium (and carbonate) ions by the peptides.

Based on these considerations, we decided to focus our studies on the interactions of oligoglutamates and solvated calcium ions, which probably dominate the earliest stages of nucleation of calcium-containing minerals in the presence of these peptides. Certainly, oxalate anions need to be included in such studies in order to gain a comprehensive understanding of all details of these nucleation processes. However, as no compatible, reliable model for oxalate is available at present and as important aspects of the nucleation phenomena are probably governed by the interactions of calcium cations with the negatively charged side chains of glutamate, we decided to focus on oligoglutamate–calcium interactions here. In view of the different scenarios discussed above, several questions might be posed: How strong are the interaction between oligoglutamates and calcium ions and how do they depend on the chain length of the peptide? How flexible are the ion–peptide aggre-

gates? Does the number of calcium ions interacting with one oligoglutamate molecule depend on the chain length? How strongly are the "bound" calcium ions exposed to the solvent (and thus accessible to oxalate ions)? Does the binding of calcium ions to oligoglutamates impose any structural pattern on the calcium ions? As the carboxylate group is the most abundant anionic group in proteins and calcium ions occur in numerous electrolyte systems, the investigation of these interactions may not only help to better understand the phenomena observed in the experiments of Fischer et al. [2], but is of a broader general interest.

Peptides such as oligoglutamates exhibit a complex conformational phase space. Therefore, it is necessary to resort to established biomolecular forcefields, which have been parametrised with a special focus on peptides and proteins, in any classical Molecular Dynamics simulation of such systems. After a detailed description of the methods applied in this work in chapter 2, we will present the results of an analysis of existing, well-established forcefields from literature with respect to their description of the ion-pairing of charged peptides with calcium ions in solution (chapter 3). In chapter 4 we present an optimisation strategy that we developed in order to fine-tune the interactions of charged peptides with monoatomic ions in solution with respect to thermodynamic and structural properties. Finally, a strategy to enhance the conformational sampling of ion–peptide complexes and to identify relevant structures from the simulation data is presented in chapter 5.

Chapter 2

Theory and Methods

2.1 Molecular Dynamics Simulations

The study of atomic-level phenomena in condensed and soft matter systems is normally aggravated by the vast numbers of particles and degrees of freedom, which render an analytical description impossible. In these cases, molecular simulations can be used to gather information on phenomena occurring at the molecular scale. If chemical reactions and other quantum effects are irrelevant in the system under consideration, a quantum mechanical description is not necessary and classical Molecular Dynamics (MD) simulations can be used to study the system. Classical MD simulations, as they have been performed in the course of the present study, consider atoms as point particles, whose motion can be described by classical mechanics (omitting any quantum effects). In order to describe the interactions of atoms on a classical level, empirical interaction potentials (forcefields) are used, which need to be fitted to experimental data or to the results of accurate quantum mechanical calculations [98]. Further details on forcefields and the respective methods of parametrisation can be found in the section on forcefields (cf. below). Starting from an initial set of positions and momenta of the point particles and from their interaction potentials, it is possible to calculate the time-evolution ("trajectory") of the system by means of MD simulations. With the help of statistical mechanics, it is then possible to extract macroscopic thermodynamic properties from the trajectory. These observations can then be compared to experimental results or can be used to predict phenomena that have not yet been observed experimentally.

In a system of N particles, starting from an initial set of positions ($3N$-dimensional vector) and momenta ($3N$-dimensional vector), the trajectory of all particles through the $6N$-dimensional phase space can be calculated

based on the Hamiltonian of the system, which is the sum of the kinetic energy $K(\mathbf{p}^N)$ and the potential energy $V(\mathbf{q}^N)$:

$$H(\mathbf{p}^N, \mathbf{q}^N) \;=\; K(\mathbf{p}^N) + V(\mathbf{q}^N) \;=\; \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{q}^N) \tag{2.1}$$

Here, $\mathbf{p}^N \equiv \{\mathbf{p}_1, \ldots, \mathbf{p}_N\}$ is the $3N$-dimensional momentum vector, $\mathbf{q}^N \equiv \{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$ is the $3N$-dimensional position vector, and m_i is the mass of particle i. The time evolution of the system can be described *via* equations of motions associated with each coordinate:

$$\begin{aligned} \frac{d\mathbf{q}_i}{dt} \;&=\; \frac{\partial H}{\partial \mathbf{p}_i} \;=\; \frac{\mathbf{p}_i}{m}, \\ \frac{d\mathbf{p}_i}{dt} \;&=\; -\frac{\partial H}{\partial \mathbf{q}_i} \;=\; -\nabla_{\mathbf{q}_i} V(\mathbf{q}^N) \end{aligned} \tag{2.2}$$

Several numerical algorithms have been developed to solve this set of first order differential equations. A prominent example is the velocity-verlet algorithm, which is time reversible, conserves the phase space volume and is efficient as the expensive force calculations are only performed once per time step in contrast to other algorithms [99, 100]. It thus fulfils the three most important criteria of numerical integration algorithms [101]. The Verlet algorithm can be derived from a Taylor expansion of the coordinates at times $t + \Delta t$ and $t - \Delta t$:

$$\begin{aligned} \mathbf{q}_i(t + \Delta t) \;&=\; \mathbf{q}_i(t) + \frac{\partial \mathbf{q}_i}{\partial t}\Delta t + \frac{\mathbf{F}_i(t)}{2m}\Delta t^2 + \mathcal{O}(\Delta t^3) \\ \mathbf{q}_i(t - \Delta t) \;&=\; \mathbf{q}_i(t) - \frac{\partial \mathbf{q}_i}{\partial t}\Delta t + \frac{\mathbf{F}_i(t)}{2m}\Delta t^2 + \mathcal{O}(\Delta t^3) \end{aligned} \tag{2.3}$$

$\mathbf{F}_i(t)$ denotes the force acting on particle i at time t. Combining these two equations yields:

$$\mathbf{q}_i(t + \Delta t) \;=\; 2\mathbf{q}_i(t) - \mathbf{q}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m}\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{2.4}$$

The particle velocity can be obtained from:

$$\frac{\partial \mathbf{q}_i}{\partial t} \;=\; \frac{\mathbf{q}_i(t + \Delta t) - \mathbf{q}_i(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \tag{2.5}$$

The goal of MD simulations is to determine thermodynamic averages and fluctuations of the properties of interest. In this respect, the "ergodic hypothesis" is of great importance, which states that all accessible microstates of phase space are equiprobable over a long period of time. That means

that all important regions of the phase space are in principle sampled in an infinitely long trajectory, and the time averages of the properties of a system are thus equal to the statistical ensemble averages.

By solving Newton's equations of motion (eq. 2.2) a microcanonical (NVE) ensemble is generated, i.e. the energy is kept constant while the temperature can not be directly controlled. It is however often desirable to perform MD simulations at a specific temperature, which for example resembles the temperature in a comparable experiment. Therefore, the temperature of the simulated system needs to be regulated. For this purpose numerous algorithms are available, but not all of them generate well-defined ensembles. In the present study the stochastic velocity rescaling thermostat was used. As the name indicates, the general idea of velocity rescaling is to rescale the velocities of the particles in the system such that a target total kinetic energy is met. The most straightforward way to do this, is to calculate the average kinetic energy at the target temperature $\overline{K} = \frac{1}{2} N_f k_\mathrm{B} T$, where N_f is the number of degrees of freedom, k_B is the Boltzmann constant and T is the temperature. In order to ensure that the total kinetic energy of the system K is equal to this target average kinetic energy, all particle velocities are multiplied by the factor alpha:

$$\alpha \;=\; \sqrt{\frac{\overline{K}}{K}} \tag{2.6}$$

This approach has the drawback, that the average kinetic energy is fixed at one value and the canonical equilibrium distribution of the kinetic energy is not sampled. In view of this deficiency, the idea of stochastic velocity rescaling, as it was proposed by Bussi et al., is to replace the target kinetic energy $\overline{K}$ by K_t, which is drawn from the canonical equilibrium distribution of the kinetic energy in a stochastic procedure [102]:

$$\overline{P}(K_t)dK_t \propto K_t^{(N_f/2-1)} e^{-\beta K_t} dK_t \tag{2.7}$$

The rescaling factor α then reads:

$$\alpha \;=\; \sqrt{\frac{K_t}{K}} \tag{2.8}$$

The normal equations of motion (eq. 2.2) are used to propagate the system between the rescalings. In this manner, the correct sampling of the canonical NVT ensembles is possible.

2.2 Forcefields

Structure and Parameters of Forcefields Solving the equations of motion (eq. 2.2) requires knowledge on the interaction potential $V(\mathbf{q}^N)$, which is a multi-body potential, i.e. the potential energy of each atom depends on the complete local environment. As the corresponding calculations of the interactions usually become very demanding, classical simulations are based on the assumption that the multi-body interactions are equal to the sum of pairwise (radial symmetric) interactions, i.e. $V(\mathbf{q}^N) = \sum_{ij} V_{ij}(|\mathbf{q}_i - \mathbf{q}_j|)$. Accordingly, typical biomolecular forcefields (FF) are based on classical pairwise additive non-polarisable potentials. This simplification massively reduces the computational costs compared to multi-body potentials, polarisable FF and First-Principles Molecular Dynamics (FPMD) simulations. In the following, we will briefly introduce the underlying model structure and the model parameters that need to be adapted to the specific system of interest.

In the framework of classical potentials, the interactions V_{ij} comprise an electrostatic contribution according to Coulomb's law (depending on the atomic partial charges q_i, which are not to be confused with a position vector $\mathbf{q}_i$) and van-der-Waals interactions. The latter ones are represented by a Lennard-Jones potential in many popular biomolecular forcefields. The non-bonded interactions between two atoms i and j, separated by a distance $r_{ij} = |\mathbf{q}_i - \mathbf{q}_j|$, thus reads:

$$\begin{aligned} V^{\text{non-bonded}}(r_{ij}) &= V_{\text{Coulomb}}(r_{ij}) + V_{\text{LJ}}(r_{ij}) \\ &= \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \\ &= \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \end{aligned} \tag{2.9}$$

FF such as GROMOS [103] or OPLS-AA [104] use heuristic combination rules to define the heteroatomic Lennard-Jones interaction parameters $C_{ij}^{(6)}$ and $C_{ij}^{(12)}$ or σ_{ij} and ε_{ij} based on the homoatomic parameters. These two FF apply the geometric-mean combination rule ($\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}}$ and $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$), while other FF use the Lorentz-Berthelot combination rule ($\sigma_{ij} = \frac{\sigma_{ii}+\sigma_{jj}}{2}$ and $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$). Chemical peculiarities of the non-bonded interactions such as the specific mutual polarisation of atoms i and j (which also depends on the environment of the two atoms) are therefore not fully taken into account [8].

Besides the Lennard-Jones potential V_{LJ} several other approaches to model van-der-Waals interactions classically exist. The Buckingham potential $V_{\text{Buckingham}}$ (eq. 4.1) relies on the same approach to model the attractive

part of the van-der-Waals interactions ($V^{\text{attr}}(r_{ij}) = -\frac{C_{ij}^{(6)}}{r_{ij}^6}$) as the Lennard-Jones potential, but uses a different description for the short range repulsion of the van-der-Waals interactions. This approach offers advantages for properly modelling interactions involving minerals as will be described in more detail in chapter 4.

The nature of these long-range interactions has some implications on the practical execution of MD simulations. The evaluation of the pairwise interactions of a particle becomes very expensive when the local environment (r_{ij}), which is considered, is large. Therefore, the evaluation of pairwise interactions $V_{ij}(r_{ij})$ is normally limited to a local environment of radius r_{cut}. In order to circumvent the occurrence of artefacts that may arise from this truncation of interactions, several strategies have been developed. For example, the van-der-Waals interactions can be shifted to zero at the cutoff distance r_{cut} in order to avoid discontinuous forces [105]. The electrostatic interactions pose a special challenge to MD simulations due to their slow decay with $\frac{1}{r}$. The most frequently applied method to prevent artefacts from the truncation of electrostatic interactions at r_{cut} is the Ewald-summation and variants thereof [106, 105]. It splits the electrostatic interactions into two contributions, a short-range and a long-range part. While the former part is evaluated in real space, the latter one is solved in Fourier space and includes the interactions of the respective charge with all its period images. The use of periodic boundary conditions is usually necessary due to the limited system sizes of MD simulations. In order to prevent artefacts that would results from "surfaces" at the borders of the simulation box, the simulated system ("unit cell") is virtually replicated infinitely in all directions [105]. Practically, this means that particles moving out of the unit cell in the course of a simulation are placed back into the cell at the opposite side of the cell. By means of the "minimum image convention" it is assured that each particle only interacts with the closest replica.

Intramolecular interactions are mostly described by "bonded interactions", which are combinations of linear (two-body), angular (three-body) and dihedral (four-body) potentials. In principle, any functional form that captures the character of the respective interaction can be used for these potentials. The chemical bond between two atoms is often modelled as a harmonic potential with the adjustable parameters bond length b_{ij} and force constant k_{ij}^b:

$$V_b(r_{ij}) \;=\; \frac{1}{2} k_{ij}^b (r_{ij} - b_{ij})^2 \tag{2.10}$$

The angular potential can be modelled by a similar harmonic potential which offers the adjustment of the equilibrium angle θ^0_{ijk} and the force constant k^θ_{ijk}:

$$V_a(\theta_{ijk}) \;=\; \frac{1}{2}k^\theta_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2 \tag{2.11}$$

A dihedral angle describes the relative positions of four atoms (i, j, k, l) that are connected by three consecutive bonds. It is defined as the angle between the normal vectors of the two planes which are spanned by the atoms i, j, k and j, k, l. The dihedral potential can be modified by means of three parameters, the multiplicity n which is restricted to positive integers in order to maintain periodicity, the factor k and the phase shift δ:

$$V_d(\phi_{ijkl}) \;=\; k^\phi_{ijkl}(1 + \cos(n_{ijkl}\phi - \delta)) \tag{2.12}$$

Often, a second type of dihedral potential is employed with the goal to ensure the planarity of certain groups or in order to distinguish between enantiomers. In order to maintain a certain structure by means of such "improper dihedrals" ξ, harmonic potentials are often applied:

$$V_{id}(\xi_{ijkl}) \;=\; k^\xi_{ijkl}(\xi_{ijkl} - \xi_0)^2 \tag{2.13}$$

with the two adjustable parameters k^ξ_{ijkl} and ξ_0.

Existing Approaches of Parametrisation Due to the empirical nature of forcefields, the validity of forcefield parameters is always restricted to a specific chemical environment and to a certain thermodynamic state. For example, the partial charges, bonded and non-bonded interaction parameters of a carbon atom differ significantly between carbon atoms in an aromatic ring and those in alkanes [107, 108]. Furthermore, it can be a challenging task to find one set of parameters that suits liquid as well as solid environments. The polarisation of an ion pair (which is reflected by the values of forcefield parameters such as the nonbonded heteroatomic Lennard-Jones parameters) is often very different in aqueous solution than in a mineral solid phase. Any simulation that attempts to model the transition of ions from a solvated state to an incorporation into a crystal will encounter this difficulty. Therefore, the details of the parametrisation process of forcefields depend directly on the targeted applications and many different classes of forcefields have been created for different types of systems.

The biomolecular forcefields (FF) employed in this thesis (GROMOS and OPLS-AA) have been designed specifically for solutions containing peptides,

proteins, sugars, nucleotides or lipids. The target data used in the parametrisation process of biomolecular forcefields are experimental (spectroscopic, thermodynamic, and crystallographic) data as well as data computed by means of quantum-mechanical (QM) methods. Experimental target data include densities, heats of vaporisation, solvation free energies, vibrational spectra and electron or X-ray diffraction structures, while QM target data may consist of minimum energy geometries, conformational energies and barrier heights, vibrational spectra, dipole moments or electrostatic potentials [109]. Each FF resorts to a different subset of target data.

However, all forcefields rely on the same type of target data with respect to the bonded interactions: the parameters of bond stretching and angle bending are optimised to reproduce experimental or computed infrared spectra and geometries, while dihedral parameters are supposed to mimic experimental and computed QM conformational energies [109].

The non-bonded Lennard-Jones (LJ) parameters are usually based on experimental thermodynamic properties of pure liquids, such as alkanes, alcohols, amides, and aromatics, which are supposed to resemble important functional groups found in peptides and proteins [109, 107]. In iterative procedures, the LJ parameters are optimised to reproduce the densities and heats of vaporisation in the respective condensed-phase simulations. Furthermore, some FF include the free enthalpies of solvation in different solvents (such as water and cyclohexane) in the optimisation procedure [107] and Reif et al. have introduced absolute intrinsic single-ion hydration free energies to the parametrisation of charged amino-acid side chains [108]. These measures are supposed to increase the accuracy of the description of solvated biomolecules such as peptides and proteins.

Two different strategies to identify suitable values for the atomic partial charges exist. While some biomolecular FF rely entirely on QM calculations (e.g. in the gas-phase) in order to determine the partial charges, other FF adjust the partial charges together with the LJ parameters during the fitting to experimental thermodynamic data. In this case, the possible values of the partial charges are constrained by the derivation of LJ parameters [109]. Furthermore, it is necessary to consider that the parametrisations of the different parts of a FF (bonded and non-bonded interactions) mutually influence each other. Therefore, the FF parametrisation is always an iterative self-consistent process and the effects of any changes to a FF on the remaining parts of the FF need to be checked. An overview over some forcefield optimisation strategies with respect to the interactions between ions is presented in chapter 4.1.

2.3 Potential of Mean Force

The potential of mean force (PMF) is the projection of the free energy of a system on one reaction coordinate [110]. In the case of two interacting ions, the distance between the two ions often is a good reaction coordinate. If the reaction coordinate of interest has been sampled extensively during the simulation, the PMF can be simply obtained by an inversion of the probability distribution along the reaction coordinate. As will be shown in the next section, the PMF between two ions (i.e. the free energy as a function of the distance between the ions) can then be used to determine the association constant of the ions in solution. However, due to sampling problems, this is not always possible. For example, the strong interactions between calcium and acetate ions prevent the accurate sampling of less favourable parts of the phase space within reasonable time scales. In this case, the comprehensive sampling of phase space can be supported with different methods, such as umbrella sampling [111], metadynamics [112] or thermodynamic integration [113]. Here, the PMF between two ions has been determined with the "blue-moon ensemble" method [114], which is a variant of thermodynamic integration. The PMF is calculated from a series of independent simulations in which the two ions are constrained at different distances using the linear constraint solver (LINCS) algorithm [115]. The distance between the calcium ion and the carbon atom of the carboxylate group of the acetate anion was used as reaction coordinate. The PMF is then obtained from integrating the constraint force f_c (including an entropic correction term) [116, 117]:

$$V^{\mathrm{PMF}}(r) \;=\; \int_{r_m}^{r} \left[\langle f_c \rangle_s + \frac{2k_{\mathrm{B}}T}{s} \right] ds + C \tag{2.14}$$

The value of the integration constant is chosen such that the PMF is equal to zero at the farthest distance considered during the constrained simulations, where the interactions between the two ions are negligible.

2.4 Calculation of the Association Constant

In order to compare the simulation results to experimental data, the association constant of calcium acetate is computed from the calculated PMF. The association constant of a 1:1 salt is generally defined as the ratio of the activity a_{AC} of the ion complex and the activities a_{A} and a_{C} of the free anions and cations in solution:

$$K_a \;=\; \frac{a_{\mathrm{AC}}}{a_{\mathrm{A}} \cdot a_{\mathrm{C}}} \tag{2.15}$$

The ion complex AC comprises all sorts of different ion pairs that might occur for the respective ion pair (cf. figure 4.3): contact ion pairs (CIP), solvent-shared ion pairs (SIP) and possibly (depending on the ionic species) solvent-separated ion pairs (2SIP). Therefore, this equation holds regardless whether a simple one-step ion pairing model or the so-called "multi-step Eigen mechanism" of ion-pairing [118, 119] is considered. Equation 2.15 holds only for dilute conditions where polyionic clusters can be neglected.

According to Chialvo et al. [120] the association constant (eq. 2.15) can be computed from molecular simulations by integrating over the radial distribution function $g(r)$ of the two ionic species. A prerequisite for the validity of this approach is a low ion concentration, as the $g(r)$ must not be affected by multi-body effects and the importance of like-ion pairs must be negligible. In the limit of infinite dilution (the salt concentration ρ_0 goes to zero), the association constant can then be expressed as [120]:

$$\lim_{\rho_0 \to 0} K_a \;=\; 4\pi \int_0^{R_{\mathrm{cut}}} g_{\mathrm{AC}}^{\mathrm{id}}(r)\; r^2 \mathrm{d}r \tag{2.16}$$

The radial distribution function of the two ions at infinite dilution $g_{\mathrm{AC}}^{\mathrm{id}}(r)$ can be easily obtained through the PMF that was determined in the NVT ensemble:

$$g^{\mathrm{id}}(r) \;=\; \exp\left(-\frac{V^{\mathrm{PMF}}(r)}{k_{\mathrm{B}}T}\right) \tag{2.17}$$

The restriction to the infinite dilute regime is no hindrance for our analysis, as most experimentalists also report the extrapolation of their results to infinite dilution in order to ease comparison [121, 7, 122]. An important quantity in equation 2.16 is the upper integration limit R_{cut}, which defines the border between the two states of *paired* (comprising CIPs and SIPs) and *unpaired* ions. The choice of the value of R_{cut} is to a certain extent arbitrary as it marks the distance from which the deviations of $g^{\mathrm{id}}(r)$ from one are negligible. However, the results obtained for K_a *via* equation 2.16 critically depend on the choice of R_{cut}. Therefore, it should be assured that the chosen value fits to the sensitivity, by which the experimental method is able to detect the different types of ion pairs.

2.5 Biasing-Potential Replica Exchange Molecular Dynamics

The adequate sampling of the phase space of oligopeptide–ion complexes is a demanding challenge. Brute-force molecular dynamics simulations are inappropriate for this task especially when the number of interesting degrees

of freedom in the system is large. The various Replica Exchange Molecular Dynamics (REMD) techniques that have been invented in the past years are well-suited to tackle this problem [123, 124]. By modifying the Hamiltonian of the system in an appropriate manner, these methods increase the frequency of transitions between the different minima of the energy landscape. In the case of oligopeptides interacting with ions, major structural changes are linked to changes in the peptide backbone dihedral angles. Therefore, the barriers to backbone rearrangements need to be manipulated for an enhanced sampling of the conformational phase space. In the following, we will describe how the Biasing-Potential Replica Exchange Molecular Dynamics (BP-REMD) method - which belongs to the group of Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) methods - can be applied to fulfil this task.

2.5.1 Theory of H-REMD Simulations

A problem frequently encountered in canonical simulations of complex (such as biomolecular) systems is, that at a fixed temperature, these systems display locally rugged energy landscapes with a large number of local minima. The heights of the barriers, which separate these minima, are often considerably larger than the thermal energy available to the system. In this state, the barriers are sampled infrequently on the timescales of typical MD simulations and the systems are often referred to as being "kinetically trapped". Escaping from these local minima is often not trivial. It is therefore difficult to obtain accurate canonical distributions by standard Molecular Dynamics or Monte Carlo methods [124]. In order to explore the huge configuration space of such systems and thus to find the global minimum state and to obtain canonical ensemble averages of the physical quantities of interest, it is necessary to find a method to overcome the large energy barriers [125].

The H-REMD approach is a special variant of the multidimensional replica exchange methods by Sugita, Kitao, and Okamoto [123, 124]. It was first formulated by Fukunishi et al. and is based on the idea that targeted changes in the Hamiltonian of the system may result in an energy landscape which is smoother than the original one, thus facilitating the exploration of the phase space. In Replica Exchange simulations multiple non-interacting copies ("replicas") of the system are run in parallel simulations. In all but one replicas, the Hamiltonian is modified such that the probabilities to overcome the barriers of the energy landscape are increased and sampling of phase space is enhanced. Only one of the replicas (often the "lowest" replica) coincides with the real system of interest. At a given frequency, pairs of (adjacent) replicas are allowed to exchange their configurations (atom positions and mo-

menta). By exchanging configurations between replicas, it is assumed that configurations from all regions of phase space visit all replicas, if the energy of the respective configuration fits to the replica. The probability of a successful exchange attempt depends on the states of the respective replicas. In a system composed of N atoms, the coordinate and momentum vectors are denoted by $\mathbf{q} = \{q_1, \ldots, q_N\}$ and $\mathbf{p} = \{p_1, \ldots, p_N\}$, and the Hamiltonian is the sum of the kinetic energy $K(\mathbf{p})$ and the potential energy $V(\mathbf{q})$ [126]:

$$H(\mathbf{p}, \mathbf{q}) \;=\; K(\mathbf{p}) + V(\mathbf{q}) \tag{2.18}$$

It is normally assumed that the probability $P_i(\mathbf{p}_i, \mathbf{q}_i)$ to find the system in state $\mathbf{p}_i, \mathbf{q}_i$ in replica i obeys a Boltzmann distribution [125]:

$$P_i(\mathbf{p}_i, \mathbf{q}_i) \;=\; \frac{1}{Z_i} \exp[-\beta H_i(\mathbf{p}_i, \mathbf{q}_i)] \tag{2.19}$$

Here, H_i is the Hamiltonian of replica i, Z_i is the partition function and β is the inverse temperature defined as $(k_\mathrm{B}T)^{-1}$. The joint probabilities of all M replicas, i.e. the probability of the "macroscopic" state in which replica i is in the respective state $\mathbf{p}_i, \mathbf{q}_i$ is then given by:

$$P_\mathrm{all} \;=\; \prod_i^M P_i(\mathbf{p}_i, \mathbf{q}_i) \tag{2.20}$$

The transition probability for an exchange of configuration $\mathbf{p}_i, \mathbf{q}_i$ of the m^th replica with configuration $\mathbf{p}_j, \mathbf{q}_j$ of replica n shall be denoted by $W(\mathbf{p}_i, \mathbf{q}_i, H_m; \mathbf{p}_j, \mathbf{q}_j, H_n)$. If the complete system is to be in Boltzmann equilibrium, the detailed balance has to hold:

$$\begin{aligned} P_\mathrm{all}[(\ldots; \mathbf{p}_i, \mathbf{q}_i, H_m; \mathbf{p}_j, \mathbf{q}_j, H_n; \ldots)] W(\mathbf{p}_i, \mathbf{q}_i, H_m; \mathbf{p}_j, \mathbf{q}_j, H_n) \;=\; \\ P_\mathrm{all}[(\ldots; \mathbf{p}_j, \mathbf{q}_j, H_m; \mathbf{p}_i, \mathbf{q}_i, H_n; \ldots)] W(\mathbf{p}_j, \mathbf{q}_j, H_m; \mathbf{p}_i, \mathbf{q}_i, H_n) \end{aligned} \tag{2.21}$$

Combining the previous equations yields:

$$\frac{W(\mathbf{p}_i, \mathbf{q}_i, H_m; \mathbf{p}_j, \mathbf{q}_j, H_n)}{W(\mathbf{p}_j, \mathbf{q}_j, H_m; \mathbf{p}_i, \mathbf{q}_i, H_n)} \;=\; \exp(-\beta[H_m(\mathbf{q}_j) + H_n(\mathbf{q}_i)] - [H_m(\mathbf{q}_i) + H_n(\mathbf{q}_j)]) \tag{2.22}$$

The Hamiltonians of two replicas of an H-REMD simulation at constant temperature only differ in the form of their potential energy function, but not in the function of the kinetic energy. Therefore, the right-hand side of this equation is independent of the individual momenta of the two replicas. In principle, the function W can be chosen freely under the constraint that it

satisfies equation 2.22. Usually, the Metropolis MC-like acceptance criteria are chosen for the transition probability:

$$\begin{aligned} W(\mathbf{p}_i, \mathbf{q}_i, H_m; \mathbf{p}_j, \mathbf{q}_j, H_n) &= 1 \qquad \text{for } \Delta \leq 0 \\ &= \exp(-\Delta) \qquad \text{for} \Delta > 0 \\ \text{with } \Delta &= \beta[H_m(\mathbf{q}_j) + H_n(\mathbf{q}_i)] - [H_m(\mathbf{q}_i) + H_n(\mathbf{q}_j)] \end{aligned} \tag{2.23}$$

An optimal choice regarding the modifications of the Hamiltonian would yield a random walk of the configurations through replica space. That means that the probability to find a configuration (a copy of the molecular system) in a specific replica is equally distributed over all replicas. The "higher" replicas, which feature significant modifications in the original Hamiltonian, exhibit higher frequencies of conformational changes. If the generated configurations are compatible with the respective Hamiltonian (i.e. the exchange probability according to eq. 2.23 is not too low), they are passed on to the "lower" replicas and finally to the target replica with the "correct" Hamiltonian, in which sampling of phase space is slow.

This concept was first introduced in the form of temperature replica exchange simulations (T-REMD, also referred to as *parallel tempering*) [127, 128, 129]. In these simulations all replicas have the same interaction potential but differ in temperature. As a change in the temperature is equivalent to scaling the whole Hamiltonian by one constant factor in the equilibrium thermodynamics (cf. eq. 2.19), T-REMD is sometimes referred to as a special case of H-REMD [125]. A problem, which is frequently encountered during T-REMD simulations, results from the number of necessary replicas, which increases as $O(f^{0.5})$ with the number of degrees of freedom f of the system [130]. For large systems, T-REMD simulations are therefore computationally inefficient [125]. Furthermore, an increase in temperature does not always help to overcome certain barriers of the energy landscape. For example, entropy-driven phenomena that hinder conformational changes in the system will even increase in strength upon an increase in temperature. In these cases, it is worthwhile to resort to H-REMD simulations. By scaling only specific parts of the Hamiltonian with the H-REMD approach, those interactions that hinder conformational changes can be manipulated selectively.

The dihedral angle BP-REMD method is a special variant of the H-REMD method, which has been designed specifically for the enhancement of the conformational sampling of peptides and proteins [131, 132]. The conformations of peptide backbones are generally characterised by one pair of dihedral angles φ and ψ per amino acid monomer. These dihedral angles exhibit

distinct patterns when the peptide or protein forms secondary structures (alpha-helices or beta-sheets: cf. figure 3.1). Due to numerous factors such as intramolecular hydrogen bonds or steric hindrances, the conformations are often relatively rigid. In order to enhance the conformational sampling in these systems, a biasing potential is introduced in BP-REMD simulations, which enforces changes in the dihedral angles φ and ψ along the backbone. The level of biasing - and with it the frequency of conformational transitions - is gradually increased along the replicas. This ensured that the system can escape from getting trapped in local energy minima. A significant advantage of this method over T-REMD simulations is that the number of necessary replicas is reduced considerably [125, 131]. As a first step of the BP-REMD method, a biasing potential needs to be constructed that forces the peptide backbone into configurations that are scarcely accessible in an unbiased situation. The construction of this biasing-potential is presented in the next section, while a detailed description of the execution of the BP-REMD simulations is given in chapter 5.

2.5.2 The Biasing-Potential

Following the idea of Straatsma and McCammon [133], we introduce a biasing-potential that is supposed to decrease the effective barriers to backbone rotation. The potential energy of rotation around the dihedral angles φ and ψ in vacuum is usually employed to construct the standard backbone dihedral potential of the FF. However, in solution several effects might impose additional barriers to torsions of the backbone. In order to set up the biasing-potential, the height and position of these barriers need to be determined. Therefore, we determined the PMFs of the two backbone dihedral angles φ and ψ.

PMFs of φ and ψ For this purpose, several different techniques such as umbrella sampling or metadynamics can be employed. Here, the PMFs of φ and ψ of the mid-monomer of a pentamer of glutamate were determined with well-tempered metadynamics [134] using the PLUMED2 [135] plug-in for GROMACS. The settings during these simulations were chosen in accordance with those of Barducci et al., who have studied the free energy of rotation around the dihedral angles of an alanine dipeptide: The height of the Gaussians in the beginning was set to 1.2 kJ/mol, the deposition interval was 120 fs, a bias factor of 5 was used and the width of the Gaussian hills was set to 0.1745 rad corresponding to 10 °. For each of the two dihedral angles, the peptide solvated in 3563 water molecules was simulated in a dodecahedron box in the NVT ensemble. In order to distinguish between intramolecular effects that hinder the torsion around these dihedral angles

(a) dihedral angle φ (b) dihedral angle ψ

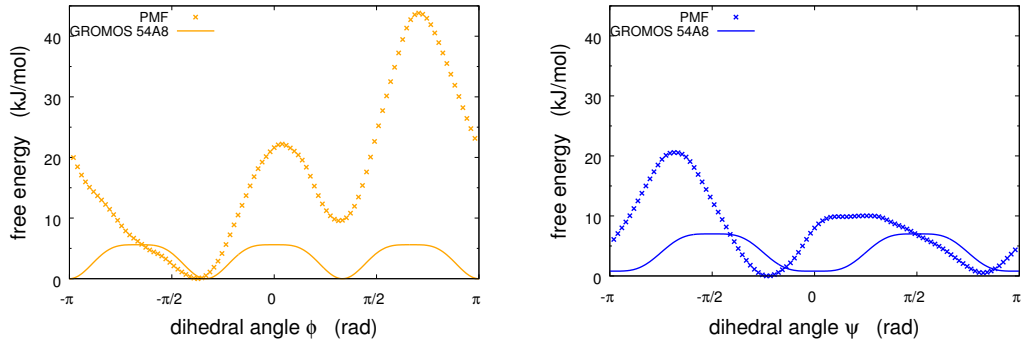


Figure 2.1: Comparison of the PMFs of the peptide backbone dihedral angles φ and ψ to the standard backbone dihedral potential of the GROMOS 54A8 FF.

and other effects such as saltbridges induced by the presence of calcium ions, the PMFs were determined in the absence of any calcium ions. 4 ns of simulation are sufficient to reach convergence of the PMFs. Figure 2.1 shows the PMFs of both backbone dihedral angles in comparison to the dihedral potentials as implemented in the GROMOS 54A8 FF. The direct comparison illustrates that the actual barriers to backbone torsion in solution can be significantly larger than one might expect from the FF potentials. Therefore, it is not sufficient to only remove the barriers implemented in the FF in an attempt to enhance the conformational sampling of a peptide in solution. Instead, conformational changes can be sampled efficiently only by introducing a bias potential constructed from the PMF.

Constructing the Biasing-Potential For practical reasons, the biasing-potential needs to be fitted to the functional form that is employed for peptide backbone dihedral angles in the respective MD software. GROMACS uses the following general form for dihedral angle potentials [136, 137]:

$$V(\varphi) \;=\; k \cdot [1 + \cos(n\varphi - \delta)] \tag{2.24}$$

While the multiplicity n is restricted to positive integers in order to maintain periodicity, the parameter k and the phase shift δ can be adjusted without limitations. Straatsma and McCammon have used the same functional form to construct the biasing-potential from the PMFs of the dihedral angles. Owing to the irregular shape of the PMFs, it is however necessary to use the following expansion of the above formula [133]:

Table 2.1: Parameter values of the curves that has been fitted to the PMFs of the dihedral angles φ and ψ. The functional form of the fitted curves (eq. 2.25) corresponds to form of the peptide backbone dihedral potentials as implemented in GROMACS.

i	$k_i^{(\varphi)}$ [kJ · mol^{-1}]	$\delta_i^{(\varphi)}$ [rad]	$k_i^{(\psi)}$ [kJ · mol^{-1}]	$\delta_i^{(\psi)}$ [rad]
1	13.97	2.021	-2.812	1.168
2	7.865	-0.978	7.288	1.852
3	-6.49	3.141	2.629	0.091
4	1.229	1.455	0.446	-0.416
5	0.366	-1.894	0.410	1.134
6	0.189	2.629	-0.188	2.96

(a) dihedral angle φ (b) dihedral angle ψ

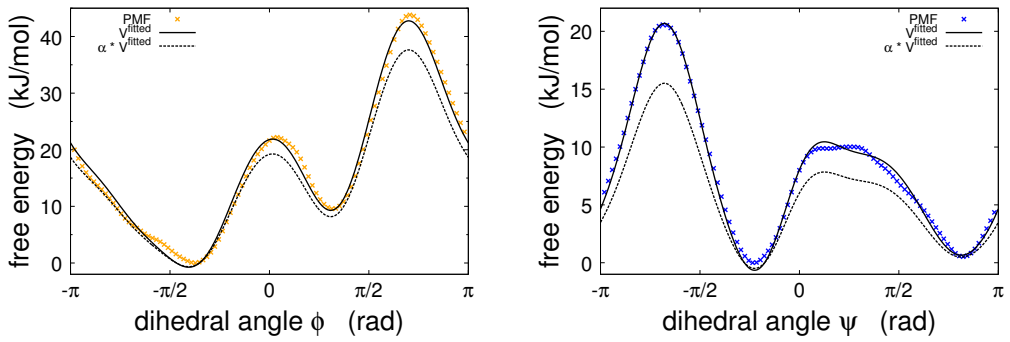


Figure 2.2: Comparison of the fitted potentials (V^{fitted}) to the PMFs of the peptide backbone dihedral angles φ and ψ.

$$V(\varphi) \;=\; \sum_{i=1}^{6} k_i \cdot [1 + \cos(n_i \varphi - \delta_i)] \tag{2.25}$$

Here, the multiplicities n_i are set to $n_1 = 1$, $n_2 = 2$, etc. The same functional form is used for the dihedral angle ψ. Applying this approach by Straatsma and McCammon, we have fitted the PMFs of φ and ψ to the formular given in eq. 2.25 using the non-linear least squares method as implemented in Matlab [138]. The results of this optimisation procedure, during which the values of δ_i are confined to the interval $[-\pi, \pi]$, can be found in table 2.1. The fitted potentials are shown in figure 2.2 together with the original PMFs.

(a) dihedral angle φ (b) dihedral angle ψ

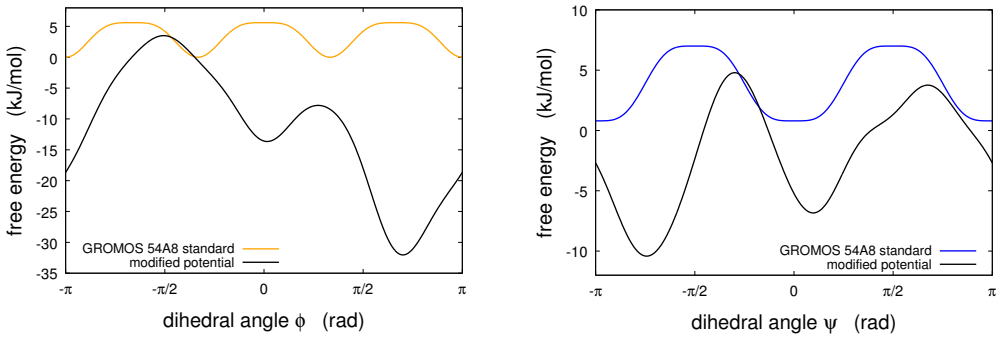


Figure 2.3: Comparison of the standard dihedral potentials V^{FF} of the GROMOS 54A8 FF and the modified potentials V^{modified} of the two peptide backbone dihedral angles φ and ψ.

In order to remove the barriers to backbone torsion and thus to enhance the conformational sampling during simulations, the biasing-potential needs to be subtracted from the standard FF backbone dihedral potential:

$$V^{\mathrm{modified}}(\varphi) \;=\; V^{\mathrm{FF}}(\varphi) - V^{\mathrm{bias}}(\varphi) \tag{2.26}$$

However, in order to ensure that the low energy regions of the PMF remain favourable during conformational sampling, the fitted functions are not used directly as biasing potentials. Instead, they are rescaled by a factor smaller than one [133]:

$$V^{\mathrm{bias}}(\varphi) \;=\; \alpha \cdot V^{\mathrm{fitted}}(\varphi) \tag{2.27}$$

The fitted potentials of both dihedral angles φ and ψ are rescaled according to equation 2.27, with a factor α of 0.88 for φ and 0.75 for ψ in the most strongly biased system in the replica-exchange simulation runs. The rescaled fitted functions $\alpha \cdot V^{\mathrm{fitted}}(\varphi)$ are shown in figure 2.2. This rescaling ensures that the highest energy barrier of the resulting dihedral potential $V^{\mathrm{modified}}(\varphi)$ (eq. 2.26) is reduced to approximately 5 kJ/mol. The remaining maximum barrier height of 5 kJ/mol corresponds to nearly 2 $k_{\mathrm{B}}T$, which is a barrier that can be easily overcome during simulations at a temperature of 300 K.

Figure 2.3 shows the standard dihedral potentials V^{FF} of the GROMOS FF together with the modified potentials V^{modified} that were obtained *via* equation 2.26. The modified potentials were then used in the BP-REMD simulations as in chapter 5.

2.6 Dimensionality Reduction in the Analysis of Simulation Data: The Sketch-Map Approach

The dynamics of systems with a large number of degrees of freedom (such as polymers or proteins in solution) can be very complex. It is often insufficient to rely only on chemical intuition when attempting to analyse the conformations and the dynamics of such complex high-dimensional systems. Pattern recognition is a classical problem known in many different sciences (physics, chemistry, biology, social sciences, psychology, marketing), and the analysis of numerous sets of high-dimensional data has shown, that often only a small number of dimensions is sufficient to rationalise them [139]. The energetically accessible regions of the phase space of large biomolecules in solution seems to have a highly complex, non-linear and possibly fractal structure. Nevertheless, there is evidence that these regions lie on a sub-structure of phase space that has a low dimensionality [140, 141]. Therefore, it is worth attempting to find low-dimensional representations of the free-energy surfaces in order to obtain meaningful insights into the phenomena occurring in these systems [142]. For this purpose, the high-dimensional vectors ($\overline{X}_i \in \mathbb{R}^D$), that represent one point in the full-dimensional phase space, need to be projected into low-dimensional vectors ($\overline{x}_i \in \mathbb{R}^d$, with $d < D$). In the optimal case, this transformation maintains the characteristics of the high-dimensional vector in the lower dimensions. This problem has been studied for many decades and numerous algorithms (e.g. linear and non-linear, local and global dimensionality reduction algorithms) have been designed [143, 139]. The most prominent example of these algorithms is Multidimensional Scaling (MDS), which is sometimes referred to as the "grandfather" of global dimensionality reduction algorithms. The approach taken by MDS to identify the low-dimensional representations of phase space $\overline{x}_i$ is to attempt to preserve the "distance" (or dissimilarity) between any two high-dimensional vectors in the low-dimensional space:

$$R_{ij} \;=\; \|\overline{X}_i - \overline{X}_j\| \;\approx\; \|\overline{x}_i - \overline{x}_j\| \;=\; r_{ij} \tag{2.28}$$

The determination of all vectors $\overline{x}_i$ is an optimisation problem, which is solved by minimising the following cost function:

$$\chi^2 \;=\; \sum_{i \neq j} (R_{ij} - r_{ij})^2 \tag{2.29}$$

The application of classical dimensionality reduction algorithms, such as

MDS, in the analysis of molecular simulations is often problematic, as these algorithms are very sensitive to noise [144] (molecular trajectories are typically very noisy [145]) and often rely on assumptions on the structure of the low-energy regions of phase space (e.g. a linear subspace of the full dimensionality space, a Euclidean manifold or a convex subset) [142]. These assumptions are often not correct. According to the theory of energy landscapes, energetically accessible regions of phase space are local basins which are connected by a spiders web of transition pathways [146]. Ceriotti et al. analysed the free energy landscape of a polypeptide (dodecamer of alanine) [142]. Plotting a histogram of the distances between pairs of high-dimensional data (R_{ij}), they found that a part of this histogram resembles a uniform distribution of points in the full high-dimensional space. That means that some features of the free-energy surface are *inherently* high-dimensional, which renders a completely faithful embedding in lower dimensions impossible, i.e. equation 2.28 cannot be satisfied. The simulation data simply does not satisfy the assumptions made in conventional manifold learning algorithms [142]. The sketch-map approach has been designed based on this insight: any projection from high to low dimension is always accompanied by a loss in information. Nevertheless, a meaningful projection can be obtained, if a sensible decision is made about what information is to be retained and what information is irrelevant [142].

Theory of Sketch-Map Under the premiss that a reproduction of all high-dimensional details is impossible, Ceriotti et al. assumed that a rough sketch of the energy landscape (a "sketch-map") is sufficient to gain insight into the collective behaviour of biomolecules in solution [142]. Therefore, they formulated the goal to "reproduce the *proximity information* from the high-dimensionality description in a space of lower dimensionality even when a faithful embedding is not possible" [142]. As an example, they studied the 24-dimensional phase space of the dihedral angles of a decamer of polyalanine. The pairwise "distance" R_{ij} between the two structures i and j in this phase space is the difference in all 24 dihedral angles: $R_{ij} = \|\overline{X}_i - \overline{X}_j\|$ with $\overline{X} \in \mathbb{R}^{24}$. In this case, the minimum image convention is applied when calculating the norm in order to take the periodicity of the dihedral angles into account. Analysing the distribution of pairwise "distances" R_{ij} of this 24-dimensional phase space (similar to fig. 2.4), they found that large values of R_{ij} represent large structural dissimilarities, which are inherently high-dimensional and thus cannot be reproduced accurately in lower dimensions. On the other hand, small values of R_{ij} are likely to represent only small structural variations similar to what one would expect for the fluctuations

within a high-dimensional harmonic basin. In view of the analysis of conformational changes of the peptide, both of these two regions are irrelevant. However, somewhere in between there is a characteristic length scale at which the most valuable topological information about the free-energy landscape is encoded [142]. The goal of sketch-map is to focus on reproducing the relative positions of nearby basins, which are likely to represent different stable conformations that are directly connected *via* transition pathways. For this purpose, all the high-dimensional, unfittable data on the internal structure of basins (small values of R_{ij}) and on the relative positions of distant basins (large values of R_{ij}) can be discarded. This approach corresponds to a non-linear dimensionality reduction. Non-linearity may be introduced by transforming the distances between data points, before distance matching according to equation 2.29 is carried out. Accordingly, sketch-map corresponds to MDS, in which the distances in *both* the high- and low-dimensional spaces are transformed by a sigmoid function [142]:

$$s_{\sigma,a,b}(r) \;=\; 1 - [1 + (2^{a/b} - 1)(r/\sigma)^a]^{-b/a} \tag{2.30}$$

These sigmoid functions (cf. figure 2.5) map the respective distances r_{ij} and R_{ij} from $\mathbb{R}^+$ monotonically to $[0, 1)$. The low-dimensional representations $\overline{x}_i$ of the high-dimensional data is then determined by a modified cost function:

$$\chi^2 \;=\; (\sum_{i \neq j} w_i w_j)^{-1} \sum_{i \neq j} w_i w_j [S(R_{ij}) - s(r_{ij})]^2 \tag{2.31}$$

$S(R)$ and $s(r)$ are sigmoid functions according to equation 2.30, w_i and w_j are weights that can be assigned to the individual data points, and R_{ij} and r_{ij} are the "distances" (i.e. the vector norm) between two data points ($R_{ij} \;=\; \|\overline{X}_i - \overline{X}_j\|$ and $r_{ij} \;=\; \|\overline{x}_i - \overline{x}_j\|$ with $\overline{X}_i \in \mathbb{R}^D$, $\overline{x}_i \in \mathbb{R}^d$, $d < D$, and $d = 2$ for sketch-map). It is important to point out, that the vector norm does not need to be the Euclidean distance.

In summary, the difference between normal MDS and sketch-map might be described by their goals to achieve $R_{ij} \;\approx\; r_{ij}$ (MDS) *versus* $S(R_{ij}) \;\approx\; s(r_{ij})$ (sketch-map). Transforming the distances through the sigmoid functions has the effect, that three different types of distances are recognised by the sketch-map approach: if points are close together in high-dimensional space they will also be so in the low dimensional space, even though the relative positions may differ significantly in high and low dimensions. The same is valid for points that are far away from each other. Only those distances R_{ij}, which are close to the σ-value of the sigmoid function, will be projected faithfully into low-dimensional space. Therefore, the resulting projection may be de-

scribed as a sketch of the conformational space that is locally accurate but globally distorted.

As the minimisation of the cost function (eq. 2.31) scales quadratically with the number of data points, it is necessary to select a small number of landmark points (e.g. 2-10 % of all data points from simulation). Different methods can be applied in order to select the landmark points from the relevant regions of phase space. Furthermore, weights can be assigned to the landmark frames (cf. equation 2.31) in order to focus the optimisation on the low-energy parts of the landscape. As the cost function is difficult to optimise, a combination of strategies is required to minimize χ^2 effectively (cf. below) [142].

Once the minimisation of the cost function is completed for the landmark points, the projection $\overline{x}$ of any high-dimensional point $\overline{X}$ can be calculated by minimising the following equation:

$$\chi^2(\overline{x}) \;=\; (\sum_{i=1}^{N} w_i)^{-1} \sum_{i=1} w_i [S(\|\overline{X} - \overline{X}_i\|_{(D)}) - s(\|\overline{x} - \overline{x}_i\|_{(d)})]^2 \qquad (2.32)$$

$\overline{X}_i$ denotes one of the landmark points and $\overline{x}_i$ is its low-dimensional projection. This equation is solved for every data point $\overline{X}$ separately and only distances between $\overline{X}$ and all landmark points are considered during the optimisation [142].

In summary, the sketch-map optimisation procedure consists of the following steps:

- analysis of the pairwise distances R_{ij} of the high-dimensional data
- adjustment of the sigmoid functions with respect to the characteristics of the histogram of pairwise distances (eq. 2.30)
- selection of landmark points
- iterative sketch-map optimisation of the projection of the landmark points (eq. 2.31)
- "out-of-sample embedding" of the remaining data points (eq. 2.32)

The application of the sketch-map analysis to the systems under consideration is described in detail in the following section.

Application of Sketch-Map In the course of this study, four different enhanced sampling simulations of oligopeptides in solution were performed (cf. chapter 5). These simulations differed in the chain length of the peptides (pentamer *vs.* decamer) and in the presence of counter-ions. In order to analyse structural differences of the forming ion–peptide complexes, the sketch-map method was applied to the results of each of the four simulations. The trajectories of the 400 ns simulations each contain approximately 800.000 data points. As some parts of the sketch-map analysis are very time-consuming and scale with the number of data points, 40,000 data points from each simulation were selected randomly. These data points are snapshots from the trajectories and thus contain one configuration of the peptide (in the absence of counter-ions) or of the ion–peptide complexes. The very first step of a sketch-map analysis is to choose collective variables (CV) which characterise the high-dimensional data points. As we are interested in the different possible structures of peptide-containing complexes, the peptide backbone dihedral angles have been chosen as descriptors. These CVs (8 dihedral angles in the case of the pentamer and 18 dihedral angles in the case of the decamer) span the high-dimensional phase space which is to be projected on two dimensions with the help of sketch-map.

The sketch-map analyses were performed with the code provided on the sketch-map webpage according to the detailed instructions of the tutorial on the webpage ([147]). The simulations of the pentamer and the decamer were analysed separately. In order to illustrate the procedure of sketch-map, the analysis of the two decamer simulations (with and without counter-ions) are presented in the following. The data of the pentamer simulations have been analysed accordingly.

The first step of the sketch-map procedure is the *analysis of pairwise distances* R_{ij} of the high-dimensional data. The corresponding histograms of the two decamer simulations are shown in figure 2.4. The two distributions exhibit the same features as the histogram of pairwise distances of a dodecamer of alanine presented by Ceriotti et al. [142]: a peak at small values of R_{ij} (here: $R_{ij} = 0 \ldots 2.5$) which corresponds to Gaussian fluctuations around one stable type of conformation, a broad peak at large values of R_{ij} (here: $R_{ij} > 4$) which is similar to a uniform distribution of data points in the high-dimensional space (i.e. the conformations are simply very different from each other), and an intermediate range of values ($R_{ij} \approx 3.5$), which carry information on the relative positions of different nearby stable conformations.

The next step is to *adjust the sigmoid functions* to the characteristics of the histogram of pairwise distances. A key-feature of the sketch-map approach is that the distance (or similarity) of points in low dimensions is also

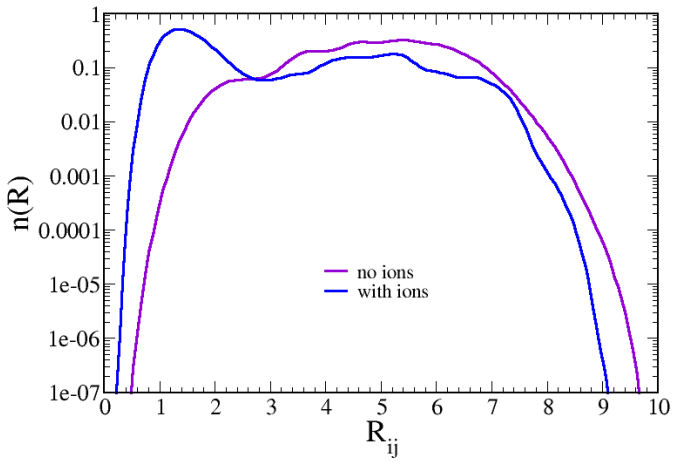


Figure 2.4: Histogram of pairwise "distances" R_{ij} between data points in the 18-dimensional space of peptide backbone dihedral angles of a decamer of glutamate (with and without counter-ions).

not evaluated using Euclidean distances, but with another sigmoid function [142]. In order to ensure that the distances of interest (here: $R_{ij} \approx 3.5$) are faithfully transformed from high- into low-dimensional space, the value of the σ-parameter of the sigmoid-functions (cf. eq. 2.30: $s_{\sigma,a,b}(r = \sigma) = 0.5$) needs to be chosen accordingly: $\sigma = 3.5$. All other distances ($R_{ij} < 3.5$, $R_{ij} > 3.5$) will be distorted during the projection. The value of σ should be the same for both sigmoid-functions. The parameters a and b of the sigmoid function (cf. eq. 2.30), which should be different for the low- and high-dimensional functions, can used to further adjust the sigmoid function to the features of the histogram of pairwise distances [142, 147]. The two sigmoid functions employed in the transformation of data of the decamer simulations are shown in figure 2.5.

As it is not feasible to optimise the projection of all 40,000 data points at once, the *selection of landmark points* is necessary. Accordingly, a subset of 1,000 representative data points need to be selected. As we are interested in a comparison of the conformations of the decamer of glutamate in solution with and without counter-ions, it is advisable to use the same projection from high- into low-dimensional space for both data sets. Therefore, 500 landmark-points were selected from each of the two simulations by means of the "farthest point strategy", which guarantees that all the sampled regions of phase space are represented in this selection [147]. Alternatively, a random selection of landmark-points is possible. Furthermore, it is possible to assign weights to the landmark-points in order to ensure that the reproduction of

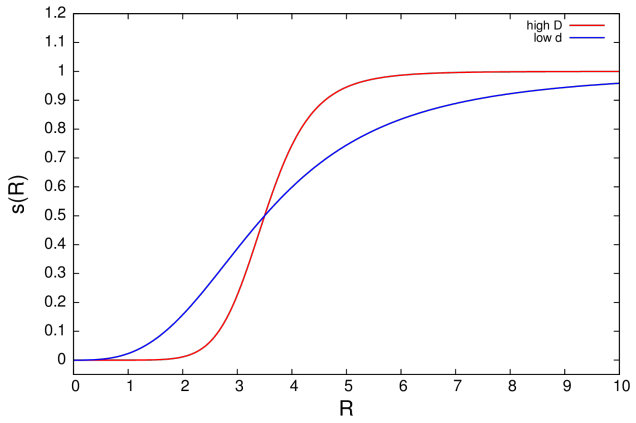


Figure 2.5: Sigmoid functions used to transform the pairwise "distances" R_{ij} and r_{ij} of the high- and low-dimensional space according to the sketch-map approach.

the structure of the low-energy parts of the energy-landscape has a higher priority during the sketch-map optimisation. The weights can be assigned based either on an estimate of the free energy, or by computing the number of trajectory frames within each landmarks Voronoi polyhedron in the high-dimensional space [142]. The latter method was applied here.

As a next step, the *iterative sketch-map optimisation* of the projection of the landmark points can be started. In order to generate a starting point for the sketch-map optimisation, linear Multidimensional Scaling (l-MDS) is applied first. That means that low dimensional representations $\overline{x}_i$ of the landmark points are determined from minimising the stress function given in equation 2.29. Figure 2.6 shows a qualitative and a quantitative analysis of the l-MDS projection of the landmark-points. The two-dimensional representation of the 18-dimensional landmark-points (fig. 2.6b) shows that most data points are grouped together into large clusters by the l-MDS projection and that no clear pattern can be identified. The size of the data points in this figure corresponds to the weight of the respective landmark-point and the colour denotes the value of one of the backbone dihedral angles . Since similar structures are expected to be grouped together in the two-dimensional representation, a structural pattern in the colour of the data points can be viewed as a qualitative indicator for a good projection. However, no such pattern can be observed for the l-MDS embedding. These qualitative impressions are confirmed by the "D-d plot" (fig. 2.6a). This plot shows a two-dimensional histogram of the probability that, for any high-dimensional pairwise distance $D = R_{ij}$, the distance between the corresponding projections (r_{ij}) will be d. If an optimal linear projection was possible, this plot

(a) "D-d plot" (b) 2-D representation

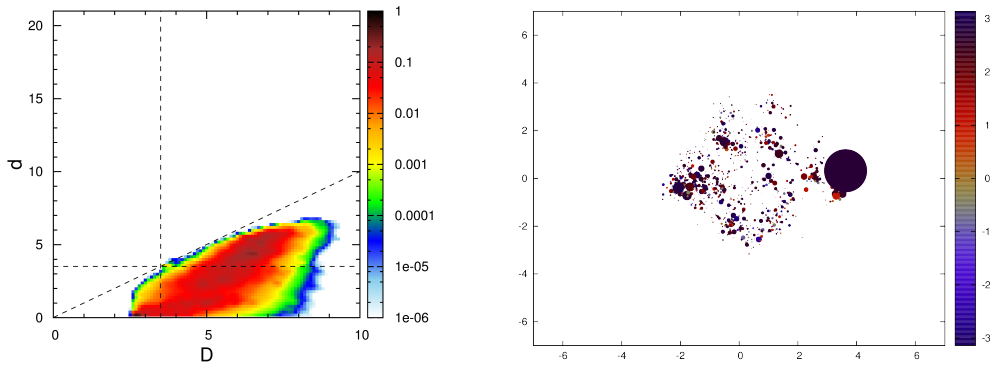


Figure 2.6: Results of the first step of the iterative sketch-map optimisation: the projection of the landmark points *via* linear Multidimensional Scaling. a) "D-d plot": Histogram which shows the probability that, for any high-dimensional pairwise distance $D = R_{ij}$, the distance between the corresponding projections (r_{ij}) will be d. The dashed line along the $D = d$ diagonal indicates the optimal case of a completely faithful embedding. The other two dashed lines denote the value of σ. A high population of the areas ($D < \sigma$, $d > \sigma$) and ($D > \sigma$, $d < \sigma$) is undesirable. b) Two-dimensional representation of the 18-dimensional landmark points of a decamer of glutamate. The data points are coloured according to the value of one of the backbone dihedral angles and the size of the points indicates the respective weight.

would consist of a series of delta function along the $D = d$ diagonal (dashed line in fig. 2.6a) [142]. The other two (horizontal and vertical) dashed lines denote the value of σ. High populations in the areas ($D > \sigma$, $d < \sigma$) and ($D < \sigma$, $d > \sigma$) indicate a poor quality of the projection, as the original similarity (or dissimilarity) of the respective high-dimensional data points has not been conserved during the projection, i.e. distant points (large D value) are projected to be close in low-dimensional space (small d value) and *vice versa*. The large number of data points (fig. 2.6a) in the area ($D > \sigma$, $d < \sigma$) therefore underlines the necessity to apply other methods than l-MDS to the system under consideration.

The projection resulting from l-MDS is a relatively poor starting point for the sketch-map optimisation. Applying the sketch-map optimisation directly to the outcome of l-MDS may result in a projection of inferior quality. Instead, the cost functions of l-MDS (eq. 2.29) and sketch-map (eq. 2.31) can be combined in a procedure, in which the contribution of l-MDS is iteratively reduced while that of sketch-map is increased accordingly [147]. The quality

of the successive optimizations can be monitored by means of D-d plots (cf. figures 2.6a and 2.7a), while the convergence can be nicely illustrated by the changes in the two-dimensional representations of the data points (cf. figures 2.6b and 2.7b). Figure 2.7 shows the converged results of the sketch-map projection of landmark-point of the two decamer simulations after seven steps of iteration. Compared to the results of l-MDS, the two-dimensional representation of the data points (fig. 2.7b) features a higher degree of structuring. This representation exhibits many more details of the relative distances between data points and shows that sketch-map was successful in assigning the different 18-dimensional conformations to different regions of the two-dimensional plot. The D-d plot (fig. 2.7a) shows only a small number of data points in the unfavourable areas ($D > \sigma$, $d < \sigma$) and ($D < \sigma$, $d > \sigma$). The area of the highest probability (red colour in fig. 2.7a) illustrates the concept of sketch-map and the underlying sigmoid-function nicely. A completely faithful projection is only guaranteed for distances similar to the value of σ (with $D \approx d$ for $D = \sigma$). Dissimilar structures (large D-values) are also far apart in the low-dimensional representation (large d-values), but the relative positions have been distorted (mostly $d > D$). Extremely similar structures (small D values) are also close together in the low-dimensional representation (small d values), but the relative distances are not projected faithfully (mostly $D > d$). In view of the general impossibility to project *all* distances from high- to low-dimensional space faithfully, the goal of sketch-map to embed only a limited range of distances with high accuracy, while allowing for a distortion of all other distances, has been achieved.

Finally, the projections of the landmark-points can be used for the "*out-of-sample embedding*" of the remaining data points of each simulation. From equation 2.32, the position $\overline{x}$ of each data point with respect to the projections of the landmark-points can be calculated. Figure 2.8 shows the projections of all 40,000 data points from each of the two decamer simulations. Sketch-map assigns the projected data points of both simulations to several distinct regions. By colouring the data points according to properties of the high-dimensional data (e.g. dihedral angle values, radius of gyration, coordination numbers, etc.), it is possible to evaluate the quality of the projection and to analyse the characteristics of the various clusters of data points in the two-dimensional representation. The corresponding analyses of all four enhanced sampling simulations are presented in chapter 5.

(a) "D-d plot" (b) 2-D representation

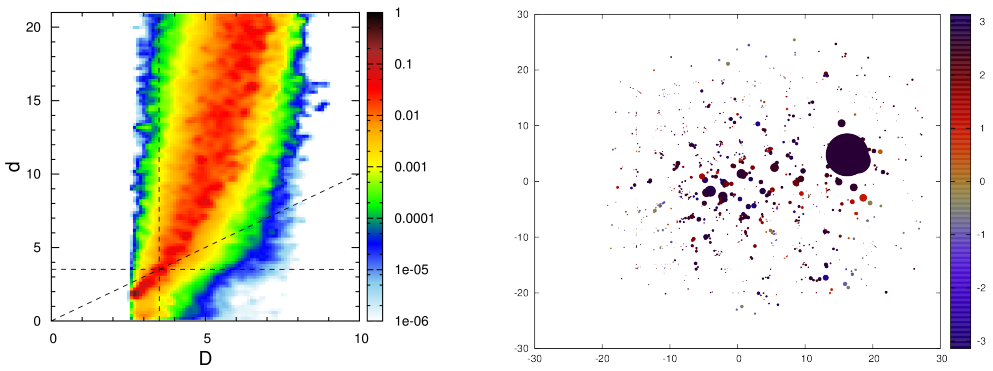


Figure 2.7: Results of the last step of the iterative sketch-map optimisation: the projection of the landmark-points based on the sketch-map cost function (eq. 2.31). a) "D-d plot": Histogram which shows the probability that, for any high-dimensional pairwise distance $D = R_{ij}$, the distance between the corresponding projections (r_{ij}) will be d. b) Two-dimensional representation of the 18-dimensional landmark points of a decamer of glutamate. The data points are coloured according to the value of one of the backbone dihedral angles and the size of the points indicates the respective weight.

(a) no counter-ions (b) with counter-ions

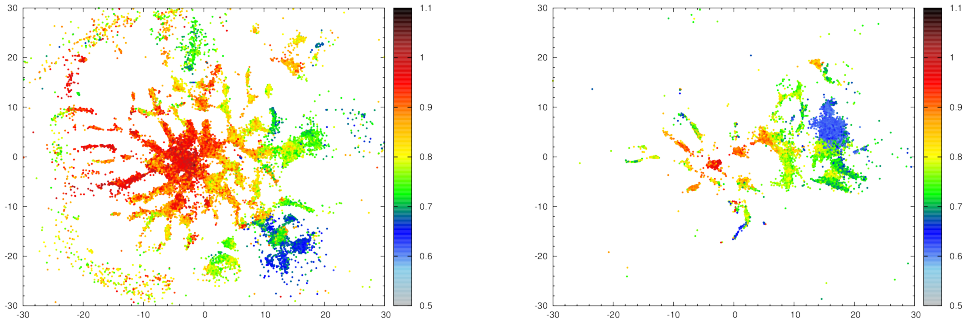


Figure 2.8: Sketch-map projections of 40,000 data points (18-dimensional) from the two simulations of a decamer of glutamate in aqueous solution with and without counter-ions. The colour-code indicates the radius of gyration of the peptide (in nm) in the respective configuration.

Chapter 3

Oligoglutamate in Contact with Calcium Ions: Test of Existing Forcefields

Prior to the application of any forcefield (FF) in molecular simulations, its validity for the system of interest needs to be tested and confirmed. Considering the goal to simulate the influence of peptides on the nucleation process of minerals in solution, it is important to ensure that the interactions of the peptides and the constituents of the mineral are accurately reproduced by the FF. In this chapter, we will show that some of the well-established biomolecular FF exhibit significant differences in the description of calcium ions interacting with the charged side chains of glutamate, which result from deficiencies in the parametrisation of interactions between calcium ions and oxygen atoms of the peptide.

3.1 Motivation

The necessity to review standard biomolecular forcefields Well-established classical biomolecular forcefields (FF) such as GROMOS, OPLS, CHARMM or AMBER have proven to give accurate descriptions of many different processes involving biomolecules [109]. However, the validity of a particular FF parametrisation needs to be confirmed for every new type of system which is modelled. Systems, in which biomineralisation occurs, contain numerous species that exhibit different characteristics in their interactions with other species: solvated ions, crystals, biomolecules and solvent molecules. In spite of their reputation as reliable descriptors of biomolecular systems, the FF mentioned above are often not explicitly parametrised for

specific ion–protein or mineral–protein interactions. The processes occurring during biomineralisation, such as the formation of clusters of ions and biomolecules or the adsorption of these species to crystal surfaces, may critically depend on subtle processes such as the structuring of surrounding water molecules [75]. Therefore, it needs to be assured that the interactions between the different species of the system are well-parametrised and balanced with respect to each other.

The non-bonded interaction parameters of ions are mostly parametrised based on single-ion hydration properties such as the free energy of solvation [108]. This approach ensures the proper modelling of interactions between these ions and water, but concentrated electrolyte solutions and the interactions with other ionic species are not necessarily well represented by such models. Depending on the pH of the solution, biomolecules such as acidic peptides may as well occur in a charged, deprotonated state. The interactions of such polyatomic ions with other ions in solution might result in a complex ion-pairing behaviour. Besides capturing the thermodynamics, the parametrisation of these interactions should also include structural details such as the coordination mode of the ions. Several studies in the literature point out the necessity to re-evaluate the parametrisation of the interactions of charged amino-acid side chains with water and mono-atomic ions [108, 4]. The simulated dynamics of protein–ion interactions of the three commonly used forcefields GROMOS, OPLS-AA and CHARMM were observed to show large deviations. Furthermore, none of the standard parameter sets of these three FF was able to reproduce experimental thermodynamic data of the respective electrolyte solution [4]. These findings illustrate that the validity of the FF needs to be tested if the interactions of different ionic species are to be simulated.

System under consideration As described in the first chapter, we want to study the interactions of oligoglutamates with calcium ions in order to gain insights into the mechanisms that dominate the earliest stages of the nucleation of calcium-containing minerals such as calcium oxalate. The interactions of oligoglutamates with other monovalent, mono-atomic ions have been studied both experimentally [148] and by computer simulations [149, 150] before, and were shown to be of a complex nature. No such reliable information is available for calcium ions in contact with oligoglutamates, but it is to be expected that the interactions of the negatively charged peptide side chains with the divalent calcium ions might be even more complex. The simulations presented in this chapter are supposed to show, how oligoglutamates respond to the presence of calcium ions, how strong these interactions

are, and whether any differences between standard biomolecular FF are observable with respect to the description of this system. In order to limit the complexity of the simulated system, the chain length of the peptide was limited to a trimer and only one calcium ion was added to the solution. The forcefields GROMOS and OPLS-AA have been selected as well-established representatives of both united-atom and all-atom forcefields for this study.

3.2 Methods

In accordance with a pH value of 8.5, which was applied in the experiments of Fischer et al. [2], the glutamic acid molecules were fully deprotonated in our simulations. Accordingly, the C-terminus was deprotonated and negatively charged as well, while the N-terminus was protonated and thus positively charged. Classical Molecular Dynamics (MD) simulations of two different systems were carried out: a trimer of glutamate (triGLU) dissolved in explicit water was simulated in the presence and in the absence of one calcium ion. Besides the standard versions of the GROMOS 54A7 [103, 151] and OPLS-AA FF [152, 153], we have also considered the modified versions of these FF as they were proposed by Project et al. [4, 5]. Both of these modified versions feature revised non-bonded interactions between calcium ions and the oxygen atoms of the carboxylate group of glutamate (table 4.2), which lead to improved ion–ion interactions. Taking into account which water models were used during the original FF parametrisation processes and following the results of Hess and van der Vegt [154], we used the SPC/E water model[155] in conjunction with the GROMOS FF and the TIP4P-Ew water model [156] together with the OPLS-AA FF. The models for the calcium ion were also chosen in accordance with the standards of the respective FF. In conjunction with the OPLS-AA FF the ion model according to Åqvist [157] is usually applied, while the GROMOS FF comes with its own set of parameters for the calcium ion [158].

We used the GROMACS simulation package [136, 137] (version 4.5.5) to execute the simulations. The systems were simulated in cubic boxes of approximately 74 nm^3 with periodic boundary conditions. They contained 2462 (GROMOS) or 2399 (OPLS-AA) water molecules during the simulations without calcium ion. For the simulations with calcium ions, one of the water molecules was replaced by a calcium ion. All simulations were conducted in the NVT ensemble at a temperature of 300 K with the temperature being controlled *via* stochastic velocity rescaling [102] with a coupling time of 0.1 ps. The equations of motion were integrated with the leap-frog algorithm with a time step of 2 fs. The center of mass translation of the simulation box was

removed every 20 fs and the neighbour list was updated every 20 fs. In all simulations, the PME method [159] was used to treat long-range electrostatic interactions with a grid spacing of 0.12 nm and an interpolation of 4. For the OPLS-AA FF, a switching function was used to switch the Lennard-Jones interactions to zero between 1.05 and 1.1 nm, while the real space cut-off for electrostatic interactions was set to 1.4 nm. On the other hand, the Lennard-Jones interactions of the GROMOS FF were truncated at 1.4 nm and real-space Coulomb interactions were cut off at 1.0 nm. In accordance with the respective FF standards, no long-range dispersion correction was applied to simulations using the GROMOS FF, while the respective corrections for pressure and energy were applied in simulations of the OPLS-AA FF.

In order to evaluate the impact of the calcium ion on the conformations of the peptide, three independent simulations with different starting positions of the calcium ion with respect to the peptide were carried out, applying the standard parametrisations of both FF. All simulations were run for 200 ns, with the exception that one of the simulations with calcium ion (standard OPLS-AA FF) was extended to 400 ns in order to study the conformational changes in a longer interval. In order to ensure a proper equilibration, the first 40 ns of the simulations were not considered for data analysis.

3.3 Results

The simulations of triGLU in the absence of calcium ions reveal small differences in the dynamics and conformations of the peptide between the standard parametrisations of the GROMOS and the OPLS-AA FF. The conformations of a peptide can be characterised by pairs of the two dihedral angles φ and ψ of the peptide's backbone (fig. 3.1). Each monomer-unit - except for those at the ends of the peptide - contributes one pair of dihedral angles to the overall conformation. Therefore, the pair of dihedral angles of the mid-monomer unit is a good indicator for the peptide's conformation. The two-dimensional histogram of these two dihedral angles ("Ramachandran plot") of triGLU in the absence of calcium ions is shown in figure 3.1 for the two standard parametrisations of the two FF. In both cases, the conformations most frequently sampled correspond to a backbone structure that is found in β-sheets of longer peptides ($-130° \leq \varphi \leq -50°$ and $90° \leq \psi \leq -170°$) and α-helices are also observed ($-150° \leq \varphi \leq -50°$ and $-60° \leq \psi \leq 0°$). Furthermore, left-handed helices ($\varphi \approx 60°$) occur with the GROMOS FF occasionally. Despite the simplicity of the simulated system, long simulation times of several hundreds of nanoseconds are necessary for sufficient conformational sampling.

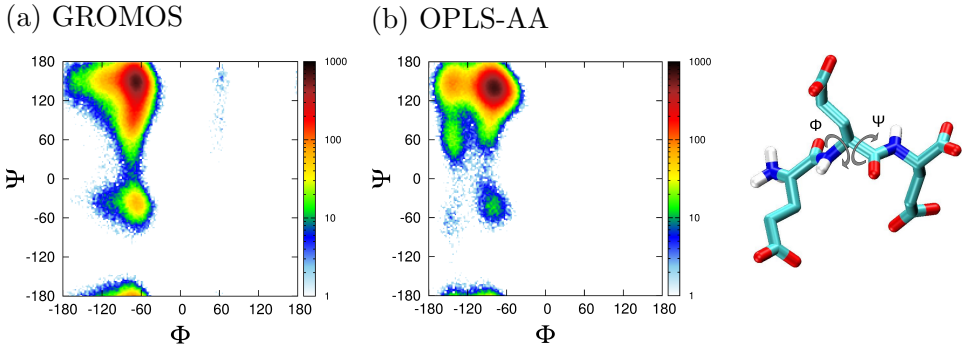


Figure 3.1: Ramachandran plots of the backbone dihedral angles φ and ψ of a trimer of glutamate in the absence of calcium ions during 200 ns simulations at 300 K. (a) GROMOS 54A7 - SPC/E, (b) OPLS-AA - TIP4P-Ew, right: triGLU in the united atom representation of the GROMOS FF.

The addition of a calcium ion to the system reveals further differences between the two forcefields. In general, the divalent ion may induce salt-bridges between the negatively charged carboxylate groups of the side chains of the peptide (figure 3.2), which are deprotonated in accordance with the pH value of 8.5 that is applied during experiments. While the attraction between carboxylate groups and the calcium ion is so strong for the OPLS-AA forcefield, that a stable close contact is formed, the same interaction is less attractive with the GROMOS forcefield. In the latter case, the attractive interactions between water and carboxylate groups outweigh the interaction between the peptide and the calcium ion, which results in a solvent-shared contact (cf. molecular structures in fig. 3.2). The radial distribution function (RDF) of the calcium ion and the oxygen atoms of the carboxylate groups of triGLU show that the contact ion-pair is by far the most frequently sampled conformation during the 200 ns simulation with the standard OPLS-AA FF. Once this ion–peptide complex has formed, it persists for time scales of the order of hundreds of nanoseconds. The same complex structure is hardly ever sampled during the simulation with the standard GROMOS 54A7 FF. Instead, mostly one shell of water molecules mediates the interaction between the calcium ion and the carboxylate groups. The significant peak of the RDF is therefore shifted towards larger distances (fig. 3.2).

The nature of the ion-pairing of the calcium ion and glutamate strongly affects the sampled conformational phase space and the dynamics of the system. The solvent-shared contact observed with the GROMOS FF is not strong enough to impose persistent restrictions on the conformations of the

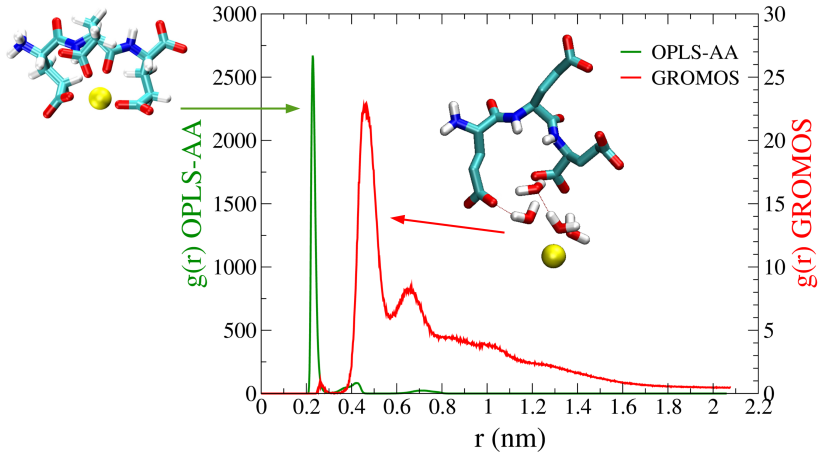


Figure 3.2: Radial distribution function of the calcium ion and the oxygen atoms of the carboxylate groups of triGLU during a 200 ns simulation at 300 K for the standard parametrisations of the GROMOS 54A7 and the OPLS-AA FF.

peptide. Therefore, the Ramachandran plot (fig. 3.3a) looks almost identical to the one of the simulation without ions (figure 3.3a), with the exception that conformations corresponding to left-handed helices ($\varphi \approx 60°$) are sampled slightly more often. For the OPLS-AA FF the close contact between calcium ion and carboxylate groups is so stable, that only minor conformational changes of the peptide are observed during several hundreds of nanoseconds. Figure 3.3 shows, that only four major conformational changes occur during a 400 ns simulation after a calcium-mediated salt-bridge has formed. Depending on the initial position of the calcium ion with respect to the carboxylate groups, the resulting peptide conformations can deviate strongly from those conformations that are observed in simulations without calcium ion (figure 3.3b compared to figure 3.1b). Furthermore, the regions of conformational phase space, which are sampled during the three independent simulations, significantly differ from each other depending on the initial calcium position. This is not the case with the standard GROMOS FF, where all three simulations with different starting positions of the calcium ion yield very similar results with respect to the conformations sampled.

Similar observations of deviating FF predictions regarding the interactions between calcium ions and molecules with carboxylate functional groups were made by Project et al. [4]. In an attempt to optimise these interactions, Project et al. suggested to modify the Lennard-Jones parameters (2.9) of the interactions between the calcium ion and the oxygen atoms of the carboxylate

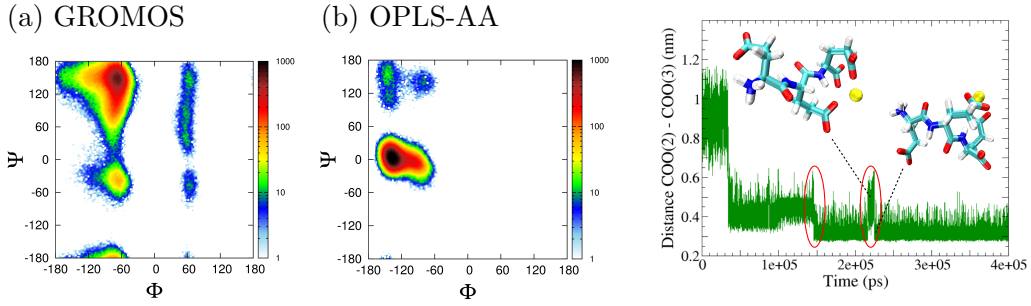


Figure 3.3: Ramachandran plots of the backbone dihedral angles φ and ψ of a trimer of glutamate in the presence of one calcium ion during 200 ns simulations at 300 K. (a) GROMOS 54A7 - SPC/E, (b) OPLS-AA - TIP4P-Ew. Furthermore, the distance between the carboxylate groups of two of the side chains of triGLU is plotted as a function of simulation time for a 400 ns simulation with the standard OPLS-AA forcefield (right). The scarce conformational changes are marked (red circles) and representative snapshots of two main conformations sampled are shown.

groups for the GROMOS and OPLS-AA FF [4, 5]. Their reparametrisation was based on a comparison of simulated and experimental data of the dissociation constant of calcium and formate ions, with the aim of an improved residence time of calcium ions in the vicinity of carboxylate groups. As a result, the repulsive term of the Lennard-Jones potential of the GROMOS FF between calcium and oxygen was weakened and the depth of the attractive well was increased (fig. 3.4a) in order to strengthen the attraction between these two atoms. The OPLS-AA FF in its standard parametrisation overestimates the attraction between calcium and oxygen. Therefore, the proposed modification increases the repulsion, while leaving the attractive well of the potential unaltered (fig. 3.4b). A comparison of the modified Lennard-Jones parameters with the respective standard FF parameters can be found in table 4.2.

The forcefield-modifications proposed by Project et al. were used to simulate the system of triGLU in the presence of one calcium ion again. These simulations reveal the crucial role of the interactions between calcium ion and oxygen atom of the carboxylate group for the whole system. A slight variation in these interaction parameter values has a pronounced effect on the nature and stability of the contact between the calcium ion and the peptide. The data in figure 3.5 show how the probability of a contact varies in response to the FF-modifications. Based on the analysis of the simulation data, a contact between the calcium ion and a carboxylate or carbonyl group

(a) GROMOS 54A7 (b) OPLS-AA

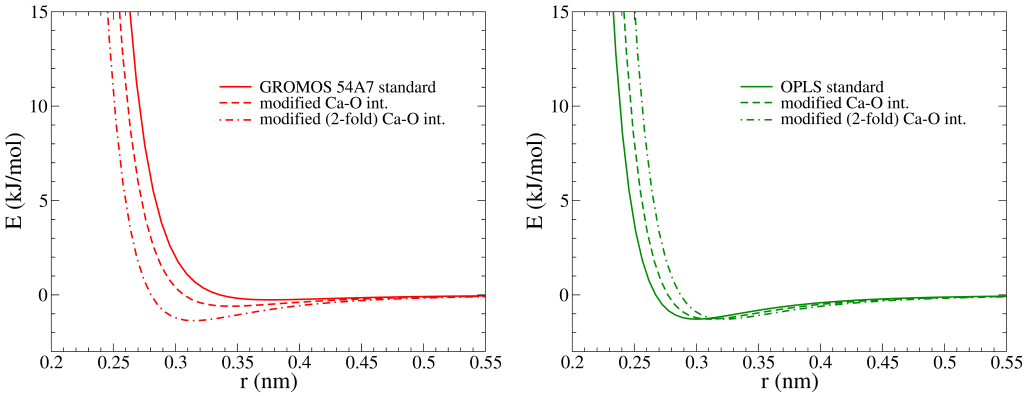


Figure 3.4: Modifications of the Lennard-Jones potential between a calcium ion and an oxygen atom of the carboxylate group of the side chain of glutamate. Solid lines: standard forcefields, dashed lines: modification as proposed by Project et al. (GROMOS: [4], OPLS-AA: [5]), dash-dotted line: 2-fold amplification of the modification as proposed by Project et al..

was defined as a distance of less than 0.3 nm between the calcium ion and the centre of mass of the functional group. In figure 3.5 the simulation data have been binned into segments of 1 ns, for which the probability of a contact was determined as the fraction of time for which a contact existed. In order to analyse, if the whole spectrum of calcium–peptide interactions - ranging from weak interactions (as with the standard GROMOS FF) to very strong interactions (as with the standard OPLS-AA FF) - can be achieved with both FF by modifications of the calcium–oxygen interactions, we introduced further modifications to both FF. These modifications were oriented towards the modifications of Project et al., but exceeded those by a factor of two (cf. figure 3.4: "modified (2-fold) Ca O int."). As can be seen from figure 3.5, it is indeed possible to obtain the complete range of calcium–peptide interactions with both FF *via* modifications of the calcium–oxygen interactions.

Even though the same set of experimental data has been used by Project et al. for the reparametrisation of both forcefields, the resulting binding affinities between calcium ions and the functional groups of triGLU still differ for the two FF (cf. fig. 3.5: middle column "modifed Ca O int."). Furthermore, Project et al. used formate as a representative for the carboxylate groups of the side chains of glutamate [4], which may not be the optimal choice due to its lack of an aliphatic carbon and the resulting differences in the electronic structure (i.e. partial charges) of the different carboxylate groups of

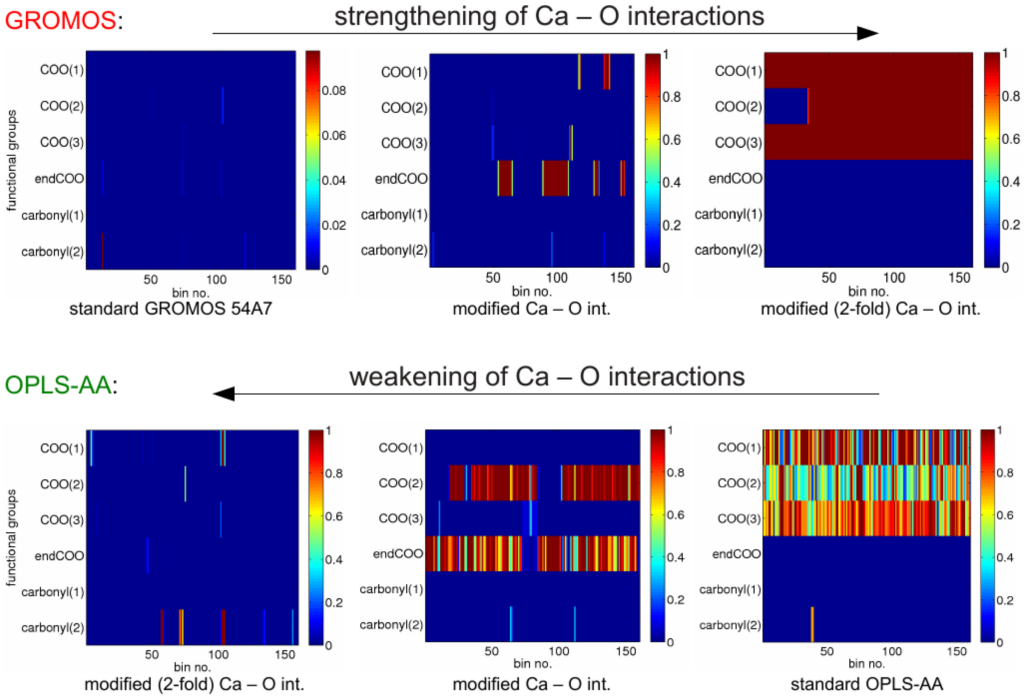


Figure 3.5: Probabilities of a contact ion-pair of the calcium ion and the different oxygen-containing functional groups of the trimer of glutamate during the last 160 ns of a 200 ns simulation at 300 K. The numbering of the functional groups starts at the N-terminus of the peptide. The origin of the x-axis (bin number 1) corresponds to the 41st nanosecond of the respective simulation. Colour-code: red: high contact probability, blue: low contact probability

formate and glutamate. Besides these deficiencies, there are other ways of determining the association constant from molecular simulations (cf. chapter 2), which are more precise than the method applied by Project et al.. Therefore, an improved optimisation of the calcium–glutamate interactions is necessary, which yields a reliable set of interaction parameters that guarantee for a realistic ion-pairing behaviour.

3.4 Conclusions

Even though substantial efforts have been devoted to the parametrisation of such established, well-tested biomolecular FF like GROMOS or OPLS-AA, these FF do not necessarily describe all the details of ion-pairing accurately, especially for complex polyatomic ions. In the case of calcium ions interacting with a peptide of glutamate, we observed significant differences in the parametrisations of the two FF that lead to large discrepancies in the observed structures and dynamics of ion-pairing in solution. Before these biomolecular FF can be applied to study phenomena occurring prior to or during the nucleation in biomineral systems, it is necessary to refine the FF with respect to details of calcium–peptide interactions. In the case of calcium ions interacting with oligoglutamates, it is necessary to identify a realistic description of the interactions of calcium with the carboxylate group of the peptide's side chains. In the next chapter, we present a general strategy that is suitable to optimise established biomolecular FF with respect to the pairing of complex ions in solution. The findings presented in this chapter illustrate furthermore, that a comprehensive conformational sampling of the relatively small system of a trimer of glutamate dissolved in an electrolyte solution can be computationally demanding. Therefore, the challenge of enhanced conformational sampling in polyelectrolyte solutions will be addressed in chapter 5 after a reliable FF parametrisation has been identified in chapter 4.

Chapter 4

Ion-pairing of Complex Ions

Our findings presented in the previous chapter revealed that standard biomolecular forcefields need to be refined in order to ensure a reliable description of the interactions between charged amino-acid side chains and monatomic ions in solution. Therefore, a general procedure to systematically improve existing classical forcefields, so to match both thermodynamic and structural properties of ion-pairing, is needed. In the following we describe the development of such a procedure and we will demonstrate, that - in the case of calcium ions interacting with oligoglutamates in solution - small modifications of the van-der-Waals ion–ion interaction parameters suffice for a systematic improvement of the essential thermodynamic and structural properties of ion-pairing.

The findings presented in this chapter have been accepted for publication in the following article: J. Kahlen, L. Salimi, M. Sulpizi, C. Peter, and D. Donadio. Interaction of charged amino-acid side chains with ions: An optimisation strategy for classical forcefields. *The Journal of Physical Chemistry B*. Reprinted with permission from [8]. Copyright 2014 American Chemical Society.

4.1 Existing approaches to optimise ion interactions in classical forcefields

Our findings regarding the necessity to optimise existing biomolecular forcefields with respect to ion-pairing properties is not surprising, as they confirm the observations described in the literature for similar systems. In the past years, several studies have pointed out the importance of a reevaluation of the parametrisation of the interactions of charged amino-acid side chains with water and ions [108, 4]. It was observed for example, that the sim-

ulated dynamics of protein–ion interactions shows large deviations among three commonly used forcefields GROMOS, OPLS-AA and CHARMM, and that none of the standard parameter sets of these three FF was able to reproduce experimental thermodynamic data of the respective electrolyte solution [4]. Several groups have tried to assess and systematically optimise the description of solvated ions and of ion association properties of different FF, following different approaches [4, 160, 161, 108, 6], which will be discussed hereafter.

Possible optimisations of model parameters Several different approaches to parametrise classical interaction potentials exist. A realistic description of the ion-pairing of complex ions, including structural details such as the binding-mode, delicately depends on a correct balance of ion–ion, ion–water and water–water interactions. If the relative magnitude of these interactions is incorrect (leading to wrong solvation properties of the simulated ion-pair), several changes to the non-bonded interactions can be made. Two major decisions have to be made during such reparametrisation process. The first question concerns which interaction between the different types of species needs to be revised (ion–ion, ion–water, or water–water interactions). As it is often advisable to work with one of the established water models, the water–water interactions are not subject to modifications. If the ion–water interactions have been set-up reliably, based on single-ion solvation properties, it is then sufficient to focus on the modification of the ion–ion interactions. The other decision that needs to be made applies to the question on how the non-bonded interaction potential (equation 2.9) between different ions or possibly between ions and water is modified. This is a controversial issue.

As described in chapter 2, the model parameters that describe the non-bonded interactions between two atoms i and j in a classical FF are the atomic partial charges q_i and q_j and the van-der-Waals interaction parameters σ_{ij} and ε_{ij}, which are ususally determined from the homoatomic parameters σ_{ii}, ε_{ii}, σ_{jj} and ε_{jj} *via* combination rules. Consequently, the van-der-Waals interactions between two atoms i and j can be modified either by a change in the homoatomic parameter values (with a subsequent application of a combination rule) or by a direct manipulation of σ_{ij} and ε_{ij} (leaving the homoatomic parameter values unaltered).

The validity of an application of heuristic combination rules depends on the two atom types i and j. Their usage is well tested and confirmed for atom types frequently occuring in biomolecular FF (such as carbon, nitrogen or

oxygen) and for the interactions of some alkali-halides in aqueous solution. In the case of ionic species, the self-interaction parameters σ_{ii} and ε_{ii} are often fitted to single-ion properties, such as ionic solvation free energies, and do not necessarily yield a realistic description of ion-pair formation and of the resulting thermodynamic properties of electrolyte solutions at higher ion concentrations.

If the pair-specific heteroatomic Lennard-Jones parameters σ_{ij} and ε_{ij} are changed, the modification only affects the specific interactions between atom types i and j, while the interaction with any other atom type k remains unaltered. This approach overrules the commonly used heuristic combination rules, but is the one that has the smallest impact on the transferability of the FF. In fact, if the homoatomic Lennard-Jones parameters σ_{ii} and ε_{ii} or the atomic partial charges were modified, this would have an impact on the interactions with any other atom type in the system. The transferability of this parametrisation to other systems would therefore be hampered. Furthermore, Fyta and Netz have shown that for some ion pairs, such as KF, the application of combination rules fails to reproduce experimental thermodynamic properties of ion-pairs, irrespective of the choice of the homoatomic parameters [161]. Only if the heteroatomic interaction parameters σ_{ij} and ε_{ij} are optimised independently from the self-interaction parameters (and thereby ignoring the combination rule), both single-ion and ion-pair properties can be modelled in a satisfactory manner.

Optimisation strategies in the literature Reif et al. have recently published a modified version 54A8 of the GROMOS FF that was largely motivated by the need to recalibrate the solvation properties of charged amino-acid side chains and their interactions with ions. During reparametrisation, they did not focus on the interactions of amino acids with specific ions (ion–ion interactions) but rather aimed at a general parametrisation of solvated charged amino-acid side chains. They chose ionic side chain analogues as representatives for the functional groups of the peptide side chains. For the optimisation of the solvation properties of these monovalent polyatomic ions (ion–water interacions), absolute intrinsic single-ion hydration free energies served as experimental target data. The changes made to the FF included nearly exclusively the atomic partial charges. Homoatomic Lennard-Jones parameters were only modified if absolutely necessary. As these changes made to the FF strongly affect the interactions of charged amino-acid side chains with other solvated ions, it is necessary to investigate the consequences with respect to ion-pairing properties. This process has been initiated [162],

and it is likely that the parametrisation of other ionic species needs to be revised in order to be compatible with the newly developed description of charged amino-acid side chains [108].

Alternatively one can leave the interactions between charged amino-acid side chains and water unaltered and focus on the ion–ion interacions [4]. The underlying assumption of this approach is that biomolecular FF have been designed to yield a thorough description of solvated proteins. Furthermore, they assume that the monoatomic ions that interact with the ionic amino-acid side chains have also been parametrised carefully, based on single-ion solvation properties. The only remaining degrees of freedom that need to be optimised are then the direct ion–ion interactions. Following this idea, Project et al. used formate as a model for the carboxylate group of an aspartate side chain and reparametrised the calcium–carboxylate interactions based on a comparison between simulated and experimental data on the association constant of calcium-formate [4]. In contrast to Reif et al., they did not change the partial charges in order to modify the non-bonded interactions, but concentrated on the heteroatomic Lennard-Jones parameters.

Fyta and Netz have optimised both ion–water and ion–ion interactions simultaneously [161]. They did not study charged amino acid side chain–ion interactions but only force field parameters for alkali and halide ions. Following a similar approach as Reif et al., they parametrised ion–water interactions by using single-ion solvation free energies as experimental target data. In this first step, the model parameters they adjusted were the homoatomic Lennard-Jones parameters σ_{ii} and ε_{ii} of the ion. Adjusting these two parameters to only one experimental entity leaves a certain degree of uncertainty. The parameter combinations that satisfy the experimental data turn out to correspond to continuous lines in the two-dimensional space spanned by the Lennard-Jones diameter σ_{ii} and the Lennard-Jones energy ε_{ii}. This uncertainty in the two-dimensional Lennard-Jones parameter space was used to take more experimental data into account, especially at non-vanishing concentrations of solvated ions. Experimental data at finite ion concentrations such as osmotic pressures or solute activities can be compared to simulation results *via* Kirkwood-Buff theory [163, 164, 165, 160]. For this purpose, Fyta and Netz applied the classical Lorentz-Berthelot mixing rule to calculate the heteroatomic Lennard-Jones interaction parameters for ion pairs and then used the Kirkwood-Buff theory to compare to experimental data. When this second step was not successful (e.g. for ion-pairs containing fluoride or iodide) and no combination of homoatomic Lennard-Jones parameters could be found that matches experimental data of both single ions and finite ion concentrations, then the heteroatomic Lennard-Jones interaction parameters (mostly ε_{ij}) were directly optimised to match experimental data. For this op-

timisation of the heteroatomic Lennard-Jones interaction parameters, scaling factors were introduced that quantify deviations from the standard mixing rules. Optimal values of this scaling factor of ε_{ij} were identified to be between 0.8 and 1.65 depending on the ion pair.

Another interesting approach has been followed by Raiteri et al., who studied the interactions of charged organic species (similar to amino-acid side chains) with different forms of calcium carbonate [6]. Hetero-interfaces between hard and soft matter pose a great challenge to simulations, as different types of forcefields have evolved for minerals and biomolecules over the past decades. Having been developed by different scientific communities, the forcefields for biological aqueous soft matter systems and for hard matter minerals have been created with different approaches to omit electronic polarisability and to cast its effects into effective classical pairwise additive potentials. These differences affect the transferability of forcefields between hard and soft systems and make them largely incompatible [61]. With the goal of reproducing experimental thermodynamic properties, Raiteri et al. have created a new FF for the nonbonded interactions among the different types of ions and between ions and water. In contrast to all other approaches mentioned so far, Raiteri et al. used the Buckingham potential as the basis for some nonbonded van-der-Waals interactions:

$$V_{\text{Buckingham}}(r_{ij}) \;=\; A_{ij} \cdot \exp\left(-\frac{r_{ij}}{\rho_{ij}}\right) \;-\; \frac{C_{ij}}{r_{ij}^6} \tag{4.1}$$

This potential provides a more realistic description of short range repulsion, essential to properly model minerals. They have optimised the ion–water interactions based on the free energy of solvation of the ions, with experimental data and quantum mechanical calculations serving as references. Regarding the ion–(organic) ion interactions, they rescaled the repulsive part of the Buckingham potential to match experimental free energies of ion pairing.

4.2 Forcefields under investigation

We have selected two well-established forcefields from the increasing number of available biomolecular FF: GROMOS and OPLS-AA. From the family of GROMOS FF [103], we have selected the latest version (54A8)[108], which incorporates changes in the description of charged amino-acid side chains. In view of the aim of our work, to analyse the interactions of such charged amino-acid side chains with other solvated ions, a comparison of this version to its predecessors is of special interest. For calcium acetate in water, there is no difference between the former versions 53A6 and 54A7 [151]. Besides

these "standard" versions, we have also considered the modified version of GROMOS 54A7 as it was proposed by Project et al. [4]. The other FF of interest is OPLS-AA in its standard version [152, 153] and its modification according to Project [5]. Both modified versions of GROMOS and OPLS-AA that were proposed by Project et al. feature revised heteroatomic Lennard Jones parameters $\sigma_{\mathrm{Ca-OM}}$ and $\varepsilon_{\mathrm{Ca-OM}}$ (table 4.2), which lead to optimised ion–ion interactions. Taking into account which water models were used during the original FF parametrisation processes and following the results of Hess and van der Vegt [154], we used the SPC/E water model[155] in conjunction with the GROMOS FF and the TIP4P-Ew model [156] together with the OPLS-AA FF.

We also consider, for the sake of comparison, the FF by Raiteri et al. [6], as it has been constructed with a special focus on the description of the interactions of charged organic species (such as acetate) with calcium ions. In contrast to the other FF, it is largely based on Buckingham potentials and applies a flexible water model. However, in our implementation we used the rigid SPC/E water model instead of the SPC/Fw model [166] for the ease of use. As these two water models differ only slightly in the partial charges, this deviation from the original FF is not expected to drastically change the solvation properties of the ions. Indeed, our implementation of the forcefield (applying the SPC/E model) does preserve the important characteristics of the Potential of Mean Force of calcium acetate shown in Fig. 3 of Ref. 6. A comparison of the PMFs is shown in figure 4.1.

The models for the calcium ion were also chosen in accordance with the standards of the respective FF. In conjunction with the OPLS-AA FF the ion model according to Åqvist [157] is usually applied, while the GROMOS FF comes with its own set of parameters for the calcium ion [158]. The calcium parameters of Raiteri's FF can be found in the Supporting Information of Ref. 6 and in Ref. 75. The models for the acetate anion were constructed based on the parameters of the side chain of glutamic acid of the respective FF (GROMOS and OPLS-AA). The acetate model of Raiteri et al. is described in detail in the supporting information of Ref. 6. Table 4.1 gives an overview of the atomic partial charges of the different acetate models. The charge of the calcium ion is not rescaled.

Table 4.2 contains an overview of the nonbonded interaction parameters between the calcium ion and an oxygen atom of acetate for all the different forcefields considered here. Note that table 4.2 and the following figures already contain values and data of FF modifications (dashed lines) that will be discussed only in the last part of the manuscript.

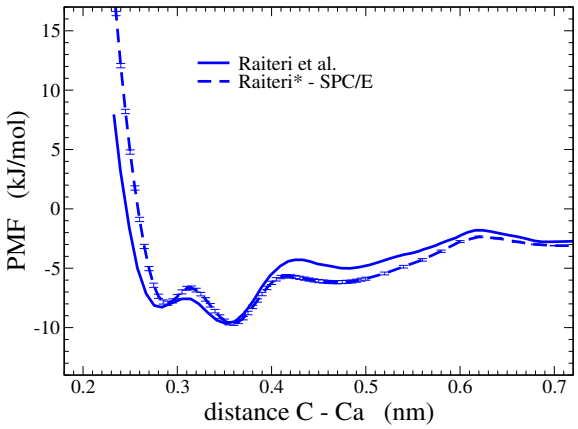


Figure 4.1: Comparison of the Potential of Mean Force of calcium acetate according to our implementation of Raiteri's forcefield with the original PMF taken from Fig. 3 in Raiteri et al. [6] Abbreviations: Raiteri *et al.* - PMF taken from Fig. 3 in Raiteri et al. [6]; Raiteri* - our implementation of the forcefield of Raiteri et al. [6] (cf. Methods section)

Table 4.1: Atomic partial charges of acetate. Abbreviations: OM - oxygen atom of the carboxylate group; C_{COO} - carbon atom of the carboxylate group; UA - united atoms; C_{CH_3} - carbon atom of the methyl group

Atom	GROMOS 54A7 [e]	GROMOS 54A8 [e]	OPLS-AA [e]	Raiteri *et al.*[6] [e]
OM	-0.635	-0.715	-0.80	-0.63
C_{COO}	0.270	0.270	0.70	0.23
CH_3 (UA)	0.0	0.16	-	-
C_{CH_3}	-	-	-0.28	-0.27
H	-	-	0.06	0.1

Table 4.2: Nonbonded interaction parameters between the calcium ion (Ca) and an oxygen atom of the carboxylate group of the acetate anion (OM). Abbreviations: GROMOS 54A7 EP - GROMOS 54A7 forcefield with modified Ca-OM interactions according to Project et al. [4]; GROMOS 54A8 Ca-OM mod. - GROMOS 54A8 forcefield with modified Ca-OM interactions that correspond to 60 % of the modifications as suggested by Project et al. [4]; OPLS-AA EP - OPLS-AA forcefield with modified Ca-OM interactions according to Project [5]; OPLS-AA Ca-OM mod. - OPLS-AA forcefield with an increase in σ_{Ca-OM} by 5.8 %; Raiteri *et al.* - forcefield as presented in Raiteri et al. [6]

LJ-type FF	σ [nm]	ε [kJ·mol^{-1}]	$C^{(6)}$ [kJ $\cdot$ mol^{-1} $\cdot$ nm^6]	$C^{(12)}$ [kJ $\cdot$ mol^{-1} $\cdot$ nm^{12}]
GROMOS				
54A7	0.3359	0.2625	$1.508 \cdot 10^{-3}$	$2.165 \cdot 10^{-6}$
54A7 EP [4]	0.3069	0.5988	$2.0 \cdot 10^{-3}$	$1.67 \cdot 10^{-6}$
54A8	0.3359	0.2625	$1.508 \cdot 10^{-3}$	$2.165 \cdot 10^{-6}$
54A8 Ca-OM mod.	0.3181	0.4351	$1.803 \cdot 10^{-3}$	$1.868 \cdot 10^{-6}$
OPLS-AA				
standard	0.2672	1.2857	$1.872 \cdot 10^{-3}$	$0.6811 \cdot 10^{-6}$
EP [5]	0.277	1.2857	$2.323 \cdot 10^{-3}$	$1.049 \cdot 10^{-6}$
Ca-OM mod.	0.2827	1.2857	$2.625 \cdot 10^{-3}$	$1.340 \cdot 10^{-6}$

Buckingham-type FF	A [eV]	ρ [Å]	C [eV Å^6]
Raiteri *et al.*[6]	2564	0.2715	0.0

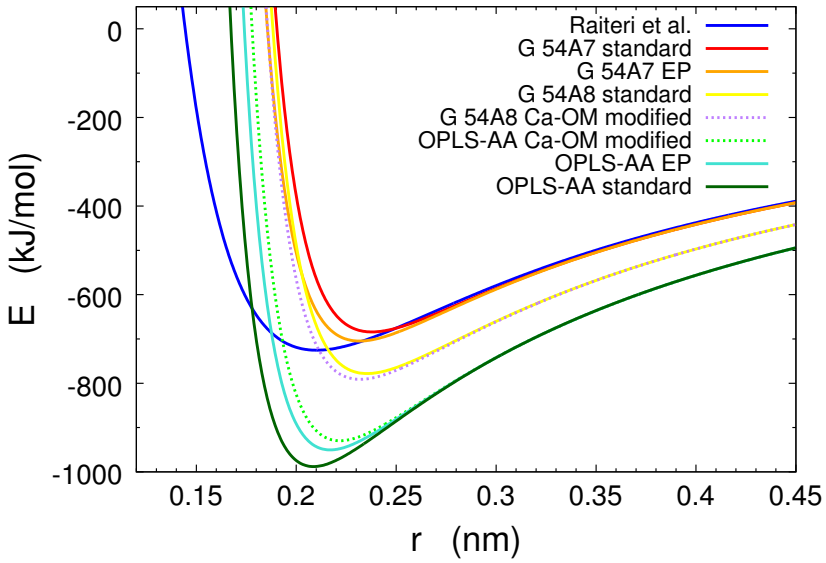


Figure 4.2: Non-bonded (van-der-Waals and Coulomb) interactions between the calcium cation and an oxygen atom of the carboxylate group of the acetate anion. Abbreviations: cf. caption of table 4.2.

Figure 4.2 shows a plot of the total nonbonded (Lennard-Jones and Coulomb) interactions between a calcium ion and an oxygen atom of the carboxylate group, according to the functional forms and parameter sets considered in this work. This plot reveals large difference between the standard parametrisations of the GROMOS and the OPLS-AA FF. The depths of the attractive minima of these interactions differ by approximately 300 kJ/mol for the two FF. This observation already hints at a possible explanation for the differences between these FF that are described in chapter 3.

4.3 Computational Details

4.3.1 Computational details of the classical MD simulations

We used the GROMACS simulation package [136, 137] (version 4.5.5) to calculate the Potential of Mean Force (PMF) between one calcium and one acetate ion from classical Molecular Dynamics simulations as described in chapter 2. Two system types, both in the shape of a rhombic dodecahedron box with periodic boundary conditions but with different sizes, were used. A

"large" simulation box was used in order to allow large distances r_m between the two ions to probe the unbonded state, where the PMF goes asymptotically to zero. This box contained 1031 water molecules besides the ion pair. Furthermore, a "small" simulation box containing 122 water molecules and the ion pair was used to directly compare the results of classical simulations to first-principles MD simulations, in which the electronic problem is solved at the level of density functional theory (DFT). The PMF was determined based on a total number of 91 (large box) or 72 (small box) distances ranging from 0.2 to 1.2 nm (large system) or 0.2 to 0.72 nm (small system). Intervals of 0.005 nm were used between 0.2 and 0.5 nm, while intervals of 0.02 nm were used for distances between 0.5 nm and 1.0 nm and intervals of 0.04 nm for distances larger than 1.0 nm. All simulations were conducted in the NVT ensemble at a temperature of 300 K with the temperature being controlled *via* stochastic velocity rescaling [102] with a coupling time of 0.1 ps. The equations of motion were integrated with the leap-frog algorithm with a time step of 2 fs. The center of mass translation of the simulation box was removed every 20 fs and the neighbour list was updated every 10 fs. In all simulations, the PME method [159] was used to treat long-range electrostatic interactions with a grid spacing of 0.116 nm and an interpolation of 4. Owing to the small size of the smaller systems, a real space cut-off of 0.72 nm was used and the Lennard-Jones interacions were truncated at the same distance. In the larger system, the settings for the calculation of the non-bonded interactions differed for the two FF. For the OPLS-AA FF, a switching function was used to switch the Lennard-Jones interactions to zero between 1.05 and 1.1 nm while the real space cut-off for electrostatic interactions was set to 1.4 nm. On the other hand, the Lennard-Jones interactions of the GROMOS FF were truncated at 1.4 nm and real-space Coulomb interactions were cut off at 1.0 nm. In accordance with the respective FF standards, no long-range dispersion correction was applied to simulations using the GROMOS FF in the large simulation box, while the respective corrections for pressure and energy were applied in all other simulations. During the simulations that used the small simulation box in conjunction with the GROMOS FF, long-range dispersion corrections were applied in order to partially compensate for the short cut-off distances. Each set-up was equilibrated for 2 ns with additional 8 ns of simulation afterwards for data analysis.

In order to compare the simulation results to experimental data, the association constant of calcium acetate is computed from the calculated PMF

(cf. chapter 2):

$$\lim_{\rho_0 \to 0} K_a \;=\; 4\pi \int_0^{R_{\mathrm{cut}}} \exp\left(-\frac{V_{\mathrm{AC}}^{\mathrm{PMF}}(r)}{k_{\mathrm{B}}T}\right)\, r^2 \mathrm{d}r \tag{4.2}$$

The restriction of this formula to the infinitely dilute regime is no hindrance for our analysis and the comparison to experimental data, as most experimentalists also report the extrapolation of their results to infinite dilution in order to ease comparison [121, 7, 122]. As described in chapter 2, the choice of the value of the important quantity R_{cut} should correspond to the distance at which two ions are considered non-bonded by experimental means. For most FF considered here, $g^{\mathrm{id}}(r)$ of calcium acetate obtained *via* equation 2.17 shows only small deviations from one for ion–ion distances larger than those of SIPs (data not shown). Therefore, the value of R_{cut} was set to the respective (FF-dependent) distance of the minimum, which separates the SIP and 2SIP maxima of $g^{\mathrm{id}}(r)$. This choice also considers that the potentiometric method, used to collect the reference experimental data, identifies both CIP and SIP complexes as associated species [167].

4.3.2 DFT calculations

First-principles DFT-based Born-Oppenheimer molecular dynamics (BOMD) simulations were performed to calculate the binding free energy of calcium and acetate in solution. These simulations were carried out by Leila Salimi (Max Planck Institute for Polymer Research and Johannes Gutenberg University Mainz) under the supervision of Davide Donadio (Max Planck Institute for Polymer Research, Mainz) and Marialore Sulpizi (Johannes Gutenberg University Mainz). The generalized gradient approximation exchange-correlation functional by Perdew-Burke-Ernzerhof (PBE) [168] including Grimme (D3) corrections for dispersion [169] were used. The pseudopotential method was adopted so that only valence electrons have been treated explicitly. Pseudopotentials were built according to the scheme by Goedecker, Teter and Hutter (GTH) [170]. The calculations were carried out with the CP2K/Quickstep package [171], which employs a mixed basis approach to represent the electronic Kohn-Sham wavefunctions [172]: a double zeta plus polarzation function (DZVP) localized basis set was chosen for the real space representation of the valence electrons, while plane waves up to a cutoff energy of 280 Ry were used to expand the density in reciprocal space.

The simulation system consisted of cubic cell of size 1.56 nm containing 123 water molecules and one ion pair. The BOMD equations of motion were

integrated using a time step of 0.5 fs and the canonical NVT ensemble was simulated by keeping the temperature at 300 K using a Nose-Hoover thermostat [173]. The PMF was obtained by thermodynamic integration using the Blue-moon ensemble approach [114], and performing constrained dynamics fixing the distance between the Ca ion and the carbon atom of the carboxylate group of the acetate molecule at 14 distances ranging from 0.286 to 0.565 nm. In the close contact region (0.279-0.325 nm) the distance intervals are separated by 0.007 nm (except 0.01 nm for 0.3 to 0.31 nm and 0.015 nm for 0.31 to 0.325 nm). Then 0.02 nm was used for distances between 0.325 to 0.565 nm. All bonds were constrained using the SHAKE algorithm [174]. After 2 ps of equilibration in the NVT ensemble, the constraint reaction force was averaged over 8 to 13 ps trajectories. The errors in the PMF are estimated from block data analysis.

In spite of reported shortcomings in reproducing the charge distribution around solvated ions [175, 176], DFT with semilocal functionals, such as PBE, is nevertheless expected to provide a correct picture of the structure and thermodynamics of monovalent ion pairs in water [177]. First-principle molecular dynamics with semilocal density functionals was also shown to render the hydration free energy of a set of monovalent ions within 4% with respect to experimental values [178], and even to provide structural and vibrational properties of more complex anions, such as microhydrated sulfate, in good agreement with experiments [179].

4.4 Results and Discussion

In order to focus on the essential aspects of the interactions between charged amino-acid side chains and divalent monoatomic ions, a model system was studied. The acetate anion served as an analogue for the negatively charged carboxylate group of the side chains of glutamate, which interact with calcium ions in solution.

Any procedure that has the target to create or refine forcefield parametrisations is guided by two main questions: (*i*) Which are the target properties that need to be properly described by the forcefield? This question is closely conditioned by the availability of reference data. (*ii*) Which part of the forcefield should be modified? The answers to these questions strongly influence the outcome of the parametrisation process.

Target Properties The goal of our reparametrisation process was to describe *structural* and *thermodynamic* properties of the ion-pairing of calcium

acetate properly, as both properties have a significant impact on nucleation and crystal growth of biominerals.

Metal cations can bind to the negatively charged carboxylate group in several different ways: if the two ions are in direct contact to each other, they are referred to as a contact ion pair (CIP), whereas the so-called solvent-shared ion pair (SIP) is characterised by a stable layer of solvent in between the two ions (figure 4.3). Furthermore, the CIP state can be subdivided into one state, in which both carboxylate oxygen atoms are coordinated to the cation (bidentate CIP: from now on referred to as *biCIP*), and another state, in which only one of the oxygen atoms is binding to the cation (monodentate CIP: from now on referred to as *monoCIP*).

In the previous chapter we already observed that the two FF GROMOS and OPLS-AA predict very different types of coordination modes for the interaction between glutamate side chains and calcium ions. While the ions form a stable CIP when simulated with the OPLS-AA FF, they form a less stable SIP with the GROMOS FF. At first sight, this differentiation between the different types of ion pairs might appear like a minor structural detail. However, the binding-mode (biCIP vs. monoCIP) of carboxylate groups in metalloproteins is believed to have a major influence on important phenomena in living organisms, such as the signal transduction pathway, and in catalytic processes [180, 181, 182, 183, 184, 185]. Furthermore, within the context of biomineralisation it is not unlikely that the binding-mode of the carboxylate group of peptide side chains and metal cations affects the nucleation and crystal growth of the respective mineral [14]. In order to obtain meaningful insights into such crystallisation processes at a molecular scale, it is therefore essential to verify that the applied forcefield yields a realistic description of the association of the ions involved.

Experimental evidence on structural properties such as the most stable coordination mode of the acetate ion to calcium is sparse. Kondoh and Oi used a combination of ^{13}C NMR spectroscopy and IR spectroscopy to analyse the interaction of alkaline earth metal ions with acetate and found that the acetate ion forms a (presumably monodentate) CIP with a calcium ion [186].

In molecular simulations, information on the probabilities of the different binding modes of calcium acetate can be obtained from the Potential of Mean Force (PMF). As described in chapter 2, the PMF quantifies the difference in free energy between system configurations along some degree of freedom. Here, the PMF was determined as a function of the distance between one calcium and one acetate ion (figure 4.3d). The minima of the PMF curves correspond to the different types of ion pairs that are observed for calcium

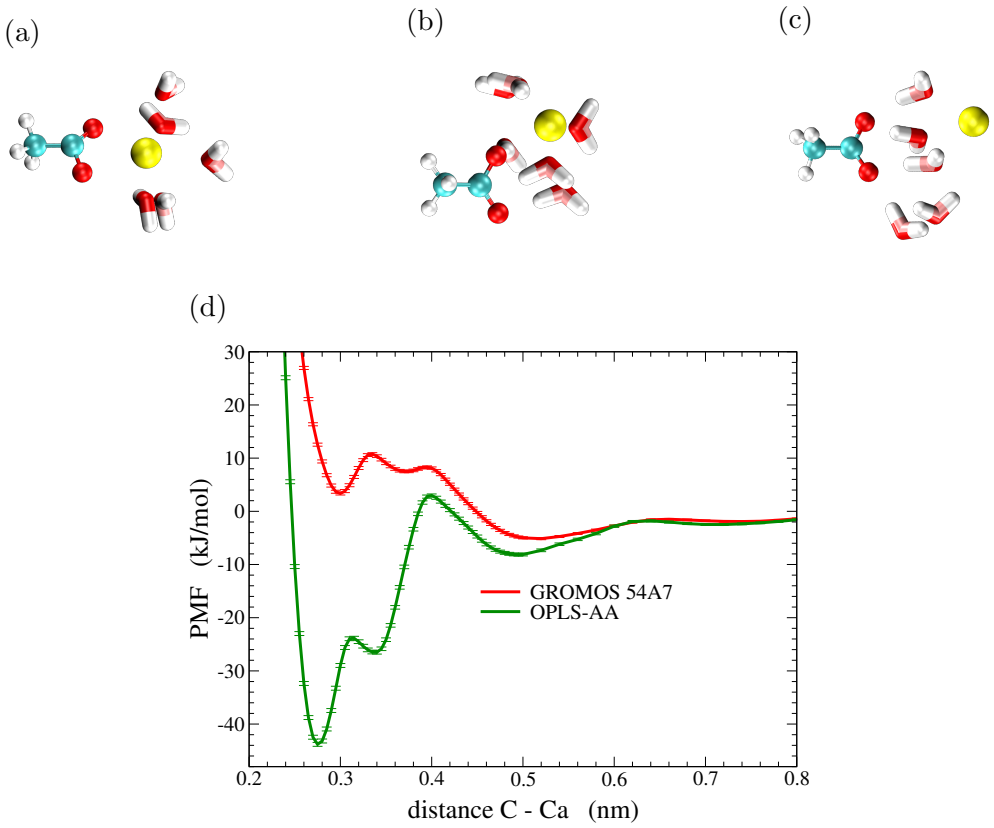


Figure 4.3: Potential of Mean Force of calcium acetate and snapshots of the different types of ion pairs which correspond to the different minima in the PMF: (a) bidentate CIP (distance between the carbon atom of the carboxylate group of the acetate anion ("C") and the calcium cation ("Ca"): 0.28 - 0.3 nm), (b) monodentate CIP (C-Ca distance 0.34 - 0.36 nm), (c) SIP (C-Ca distance approx. 0.5 nm), (d) Potential of Mean Force of calcium acetate as a function of the distance C-Ca. By construction, the PMF converges to zero at a distance of 1.20 nm.

acetate: solvent-shared ion pairs (SIP) at C - Ca distances of around 0.5 nm, monodentate contact ion pairs (monoCIP) at distances of approximately 0.34 - 0.36 nm and bidentate contact ion pairs (biCIP) at distances of 0.28 - 0.3 nm.

The difference between the depths of CIP and SIP minima of the PMF is very important as it gives an estimate of the free energy that is necessary to separate a contact ion pair. Figure 4.3d shows that there is a marked difference between the standard FF versions of GROMOS and OPLS-AA regarding the relative weight of the two main minima of the PMF. For the standard OPLS-AA forcefield biCIP is the stable configuration, while using GROMOS CIP is only metastable, in favour of SIP pairing. These two PMF curves thus give a quantitative explanation for the differences in the interactions of a calcium ion with a trimer of glutamate modelled by the two different FF (cf. previous chapter). In accordance with the depth of the CIP minimum of the PMF of calcium acetate, the OPLS-AA FF predicts CIPs between the carboxylate groups of the peptide side chains and the calcium ion. On the other hand, calcium acetate preferentially forms SIPs according to the standard GROMOS FF, which is also observed for the ion-pairing of glutamate side chains and calcium ions when modelled with this FF.

In order to be able to evaluate which of the FF gives a more realistic description of ion-pairing of calcium acetate, quantitative reference data are needed. As there is no reliable, fully quantitative experimental data available that allows detailed conclusions on the structural properties of the pairing of calcium acetate, we have collaborated with J. Prof. Marialore Sulpizi (University of Mainz) and Leila Salimi (MPI-P Mainz), who have studied the ion-pairing of calcium acetate by means of first-principles DFT-based Born-Oppenheimer molecular dynamics (BOMD) simulations as described in section 4.3.2. From the results of these BOMD simulations (black curve in figure 4.5), it is possible to draw conclusions on the probabilities of the different possible coordination modes and on the water structure around these ion complexes.

It is necessary to perform both types of calculations (classical MD simulations and BOMD calculations) in a system of similar size in order to take size effects into account and thus to ensure a fair comparison. For this reason, we performed the classical PMF calculations in a system of the same size as the FMPD calculations. The system contained only 122 water molecules and one ion pair. The classical simulations show that going from the large system (1031 water molecules and one ion pair) that can be considered uneffected by finite size effects, to the smaller system, which is more prone to finite size errors, the main characteristics of the PMF curves are preserved. Figure 4.4 shows for one of the GROMOS and one of the OPLS-AA FF, that the PMF

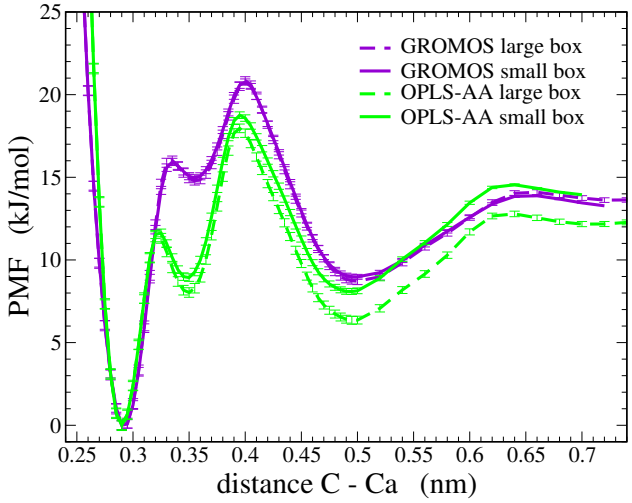


Figure 4.4: Potential of Mean Force of calcium acetate as a function of the distance between the calcium cation and the carbon atom of the carboxylate group of the acetate anion for different system sizes. The curves have been shifted vertically such that the minima of biCIP are aligned at 0 kJ/mol. "Small box": 122 water molecules and one ion pair. "Large box": 1031 water molecules and one ion pair.

curves of the small and the large system differ by no more than 2 kJ/mol at any ion–ion distance. Given the rather large uncertainties in the reference data ("DFT"-curve in figure 4.5) and the overall large free energy differences of the ion-pairing of calcium acetate, this error can be neglected.

Figure 4.5 shows a comparison of the classical PMF calculations (small system) to the results of the DFT-based calculations. In order to ease the comparison of the free energy difference between CIP and SIP states, the PMF curves have been shifted vertically, so that the CIP minimum (biCIP or monoCIP) with the lowest energy of each curve (where the error in the DFT-based reference data is smaller than for longer ion–ion distances) is aligned at 0 kJ/mol. Regarding the balance between the population of the CIP and SIP states, all FF variations studied here lie between the two extreme cases of the standard OPLS-AA and the standard GROMOS FF. The standard OPLS-AA FF strongly *underestimates* the probability of SIPs (the free energy of the SIP state is approx. 35 kJ/mol higher than the one of the CIP state), while the standard GROMOS FF strongly *overestimates* the probability of SIPs (the free energy of the SIP state is approx. 8 kJ/mol lower than the one of the CIP state). All the proposed modifications of the GROMOS and the OPLS-AA FF improve the balance of CIP and SIP minima with respect to the original FF. But only two of the FF from the literature do not only

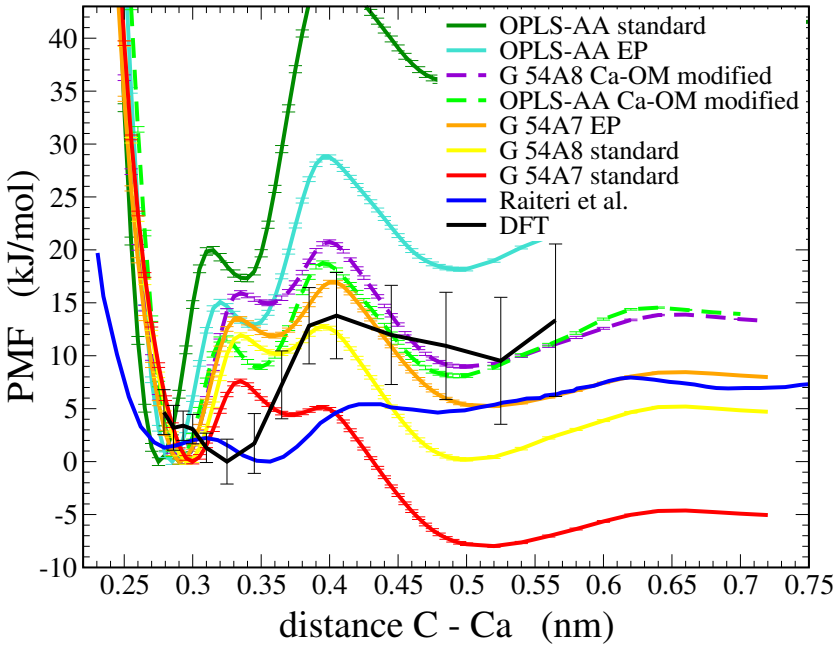


Figure 4.5: Potential of Mean Force of calcium acetate as a function of the distance between the calcium cation and the carbon atom of the carboxylate group of the acetate anion. The curves have been shifted vertically such that their lowest CIP-minimum (either bidentate or monodentate) is aligned at 0 kJ/mol. Abbreviations: cf. caption of table 4.2; Raiteri *et al.* - PMF taken from Fig. 3 in Raiteri et al. [6]; DFT - First-Principles Molecular Dynamics simulations based on DFT.

show the correct trend regarding the relative probabilities of CIPs and SIPs, but also agree quantitatively with the DFT-based results (within the error bars of the SIP minimum). The first one is the modification of GROMOS 54A7 as it was proposed by Project et al. ("G 54A7 EP"), which differs from the original version only in an increased value of $\sigma_{\mathrm{Ca-OM}}$ [4]. The other FF is the one proposed by Raiteri et al. which is parametrised specifically for mineralisation from solution [6]. Note that figure 4.5 already contains data of FF modifications (dashed lines) that will be discussed below.

Comparing the PMF curves of the classical FF to the PMF of the DFT-based simulations draws our attention to another detail of the PMF: the relative probabilities of bidentate and monodentate CIPs. According to the PMF of the DFT-based simulations, the minima of biCIP and monoCIP are comparable in depth (considering the error bars) with a slight tendency towards the monoCIP state. However, considering the limitations in sampling in these BOMD simulations, it is likely that the uncertainty in the PMF is

larger than it appears from the error bars in figure 4.5. Of the classical FF, only the one suggested by Raiteri et al. shows a similar trend towards the monodentate structure while all other classical FF predict that the bidentate state is more favourable.

The different GROMOS and OPLS-AA FF underestimate the probability of monoCIPs (with respect to biCIPs) in contrast to the FF by Raiteri et al. and to BOMD. Therefore, it is worth analysing the differences between the classical FF based on the Lennard-Jones interaction potential (GROMOS and OPLS-AA) and based on the Buckingham potential (FF of Raiteri et al.) more closely. The two FF from the literature, where the free energy difference between CIPs and SIPs best matches the DFT results (FF according to Raiteri et al. and the modification of the GROMOS 54A7 FF proposed by Project et al.) are in principle very similar. One difference is in the water models applied in conjunction with the two FF: while the SPC/E water model [155] was used for the GROMOS FF, Raiteri et al. utilised the flexible SPC/Fw water model [166] with their FF. But apart from its flexibility, the SPC/Fw model is quite similar to the SPC/E model. Therefore, it is unlikely that this difference in the water model leads to the observed differences between the two FF. If the water model had such a strong influence on the PMF curves, the OPLS-AA - TIP4P-Ew combination would differ more from the GROMOS - SPC/E results than it actually does. All the other main features of the FF (van-der-Waals and Coulomb interactions between the ions and between each ion and water) are very similar, with two exceptions: (*i*) The repulsion between oxygen atoms of the acetate anion and of the water molecules is softer for the FF of Raiteri et al. (cf. figure 4.6). (*ii*) Raiteri et al. add a repulsive Buckingham potential to the interaction between the hydrogen atoms of water molecules and the oxygen atoms of the acetate anion.

Both these differences affect the water structure of the first solvation shell around acetate, which substantially differs for the two FF. Figure 4.7 shows details of the water structure around the acetate anion for biCIPs. Raiteri's FF (as it was implemented by us) predicts a structure of the first solvation shell that deviates remarkably from the DFT reference: The OM-OW radial distribution function (figure 4.7a) shows that both oxygen atoms (OM: oxygen atom of the acetate anion; OW: oxygen atom of a water molecule) are able to approach each other very closely. A comparison of these results to original data of Raiteri et al. generated with LAMMPS revealed that our implementation of this FF captures these important details correctly (data not shown).

This artifact in OM-OW distances influences the hydrogen-bonding angle between acetate and water when calcium is in direct contact: The minimum

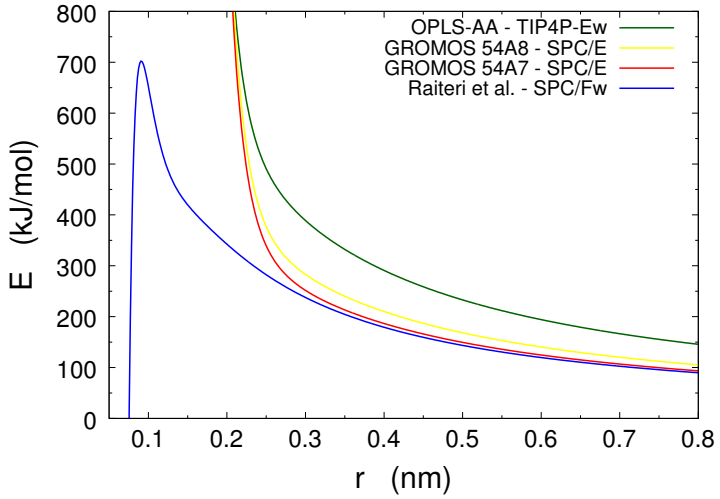
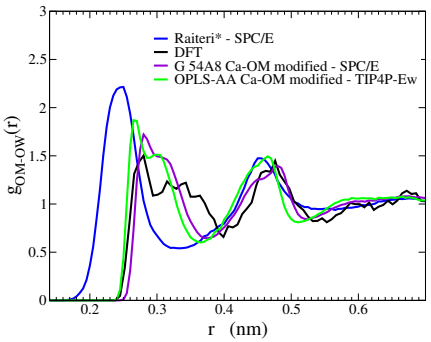


Figure 4.6: Non-bonded (van-der-Waals and Coulomb) interactions between the oxygen atom of water and an oxygen atom of the acetate anion.

(a) RDF OM-OW



(b) hydrogen-bond angle distribution

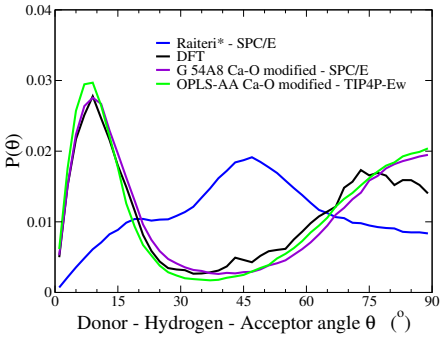
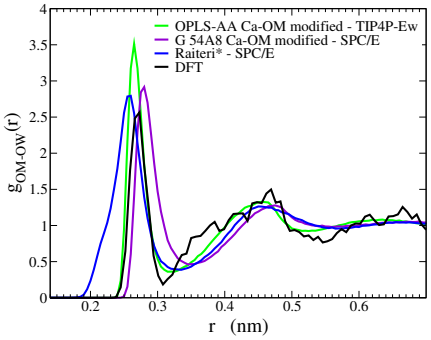


Figure 4.7: Structure of water around the acetate anion for a biCIP configuration. Abbreviations: OM: oxygen atom of the acetate anion; OW: oxygen atom of a water molecule; Raiteri* - our implementation of the forcefield of Raiteri et al. [6] (cf. Methods section); DFT - First-Principles Molecular Dynamics simulations based on DFT; G 54A8 Ca-OM modified - GROMOS 54A8 forcefield with modified interactions between the calcium ion and the oxygen atoms of the carboxylate group of the acetate anion that correspond to 60 % of the modifications as suggested by Project et al. [4]; OPLS-AA Ca-OM modified - OPLS-AA forcefield with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 %.

(a) RDF OM-OW



(b) hydrogen-bond angle distribution

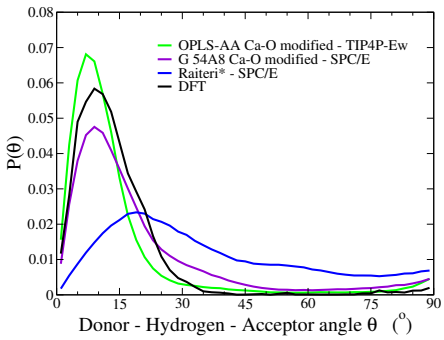


Figure 4.8: Structure of water around the acetate anion for an unpaired ion configuration. Abbreviations: cf. caption of 4.7.

distance between the two oxygen atoms can be so small that the hydrogen atom of the hydrogen-bond-donating water molecule is pushed away from the axis between the two oxygen atoms (figure 4.7b: maximum for $\Theta \approx 45\,^{\circ}$). This behaviour is not observed as strongly when the two ions are farther apart and the solvation shell of water around the acetate anion is thus not affected by the presence of the calcium ion. In this unpaired situation, both the RDF (OM-OW) and the hydrogen-bond angle distribution of the FF of Raiteri et al. are a lot more similar to the respective DFT-based results (figure 4.8). In spite of the success of Raiteri's FF in describing the single solvated ions and the thermodynamic properties of ion-pairing (cf. below), it is thus worth considering an alternative FF, if structural details of ion-pairing are of special interest.

Comparing the PMF calculations from classical and BOMD simulations provides information on the local short range interactions between calcium and acetate, especially the relative probabilities of the different types of ion-pairs (CIP vs. SIP), and the water structure around the ion complexes. However, due to the size limitations of BOMD this comparison gives information on the *relative* depth of the minima, but not on the *absolute* depth of the minima which governs the total attraction between the two ions. In order to evaluate the classical FF with respect to the total attraction strength between the two ions, further reference data are needed.

For calcium acetate, reliable experimental data on the thermodynamics of ion-pairing are mostly limited to the association constant [7], which is proportional to the ratio of concentrations of paired (ρ_{AC}) and unpaired ions

(ρ_A, ρ_C) in solution:

$$K_a \propto \frac{\rho_{AC}}{\rho_A \cdot \rho_C} \tag{4.3}$$

In MD simulations, the association constant of calcium acetate can be computed from the PMF according to equation 4.2. For this purpose, the PMF needs to be determined over a wide range of distances between the two ions, including large distances where the PMF converges to zero (cf. chapter 2). Therefore, we performed the same classical PMF calculations that are shown in figure 4.5 in a larger system, in which large distances between the two ions can be probed. The PMFs of calcium acetate, as calculated in the "large" cell with the different FF analysed here, are shown in figure 4.9. Note that figure 4.9 already contains data of FF modifications (dashed lines) that will be discussed later. The PMFs obtained from the different GROMOS FF versions illustrate how the probabilities of contact ion pairs and solvent-mediated ion pairs are altered, depending on the non-bonded interaction parameters that are changed. If only the heteroatomic Ca-OM Lennard-Jones interaction parameters are changed (54A7 modified according to Project et al.: attraction is increased, repulsion is decreased [4]) compared to the standard 54A7 version, the CIP minima of the PMF become much deeper, while the SIP minimum remains unaltered. On the other hand, if only partial charges are changed (54A8) compared to the standard 54A7 version, not only the CIP minima of the PMF become deeper, but also the SIP minimum is affected, since non-contact ion–ion interactions (mediated by water) are dominated by electrostatic interactions.

Based on the PMFs shown in figure 4.9 the association constants of the respective FF have been calculated applying equation 4.2. Figure 4.10 shows the logarithm of the association constant for the different FF of interest. The standard GROMOS and OPLS-AA FF and their modifications that have been published so far, clearly fail to reproduce the attraction between the two ions as observed from experiments [7] (grey-shaded area). For the standard GROMOS 54A7 FF, the attraction between the calcium cation and the acetate anion is by far too weak. The modifications of this FF, that have been published, go in the correct direction by enhancing the ion-ion attraction (leading to an increased association constant), but do not reach the range of the experimental data. On the other hand, the standard OPLS-AA FF stongly overestimates the attraction between the ions. The modification proposed by Project [5] improves this shortcoming, yet the association constant of the modified FF does not reach the range of the experimental data either. Of all the FF from the literature that we analysed so far, the one proposed by Raiteri et al. [6] approximates the experimental association

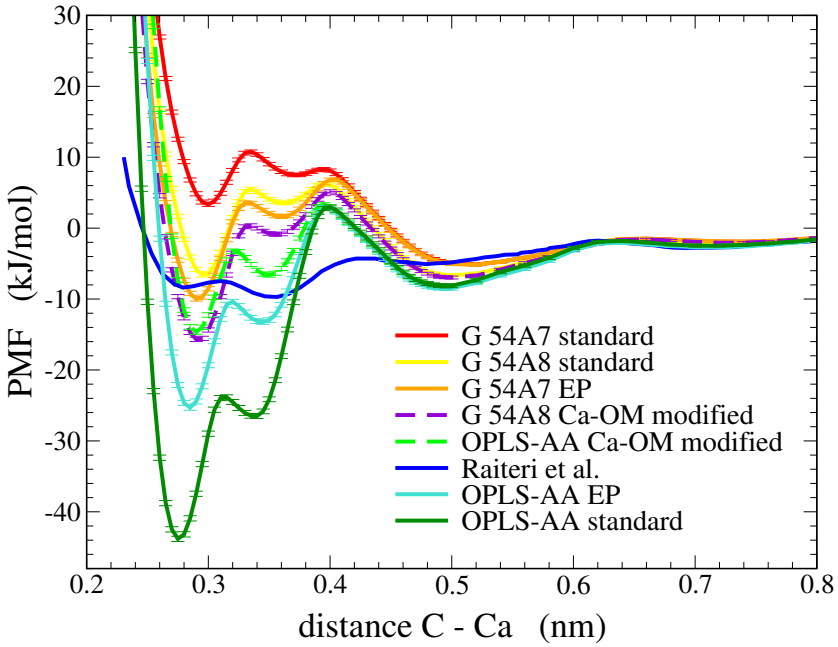


Figure 4.9: Potential of Mean Force of calcium acetate as a function of the distance C-Ca. By construction, the PMF converges to zero at a distance of 1.20 nm. Abbreviations: cf. caption of table 4.2; Raiteri *et al.* - PMF taken from Fig. 3 in Raiteri et al. [6]

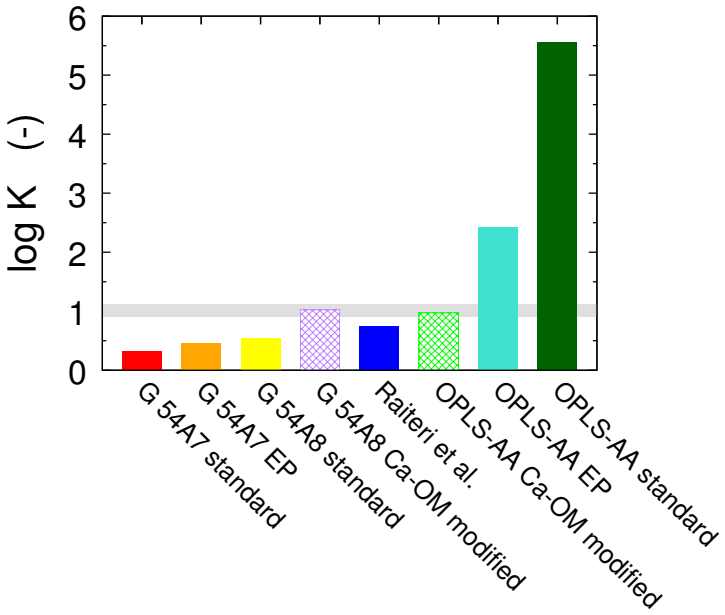


Figure 4.10: Association constant of calcium acetate. The grey-shaded area indicates the range of experimental data according to Daniele et al. [7]. Abbreviations: cf. caption of table 4.2; Raiteri *et al.* - common logarithm of the association constant calculated from the PMF of calcium acetate as presented in Fig. 3 of Raiteri et al. [6]

constant best, even though it still underestimates the attraction of the ions. Therefore, none of the selected (unmodified) forcefields from literature reproduces the experimental association constant correctly. Note that figure 4.10 already contains data of FF modifications (chequered bars) that will be discussed later.

The data on the experimental association constant of calcium acetate and the PMF calculated from BOMD simulations suffice as reference data in order to evaluate existing forcefields and to test possible forcefield modifications. These reference data allow us to consider both the *local* ion-pair structure (comparison to BOMD simulations) and an *integral* (thermodynamic) quantity (comparison to association constant) during the evaluation of classical FF. Our results show that only two FF from literature reproduce the relative probabilities of CIPs and SIPs correctly and that none of the FF from literature reproduces the experimental association constant. Therefore, a refinement of the existing biomolecular FF with respect to these properties is necessary for a proper description of ion-pairing.

Forcefield Modifications In spite of the success of Raiteri's FF in describing the single solvated ions and the thermodynamic properties of ion-pairing [6], the artifacts that this model exhibits in the structure of the first solvation shell around ion complexes (figure 4.7) lead us to decide to omit this FF in the further refinement procedure. Furthermore, an advantage of using standard biomolecular FF such as GROMOS or OPLS-AA as starting points for the FF optimisation is that no further reparametrisation is needed to simulate ion–peptide interactions.

As the next step, it needs to be decided which of the forcefield parameters are to be modified. Again, two decisions need to be made. Firstly, it is necessary to clarify which of the ion–ion or ion–water interactions (which all have a significant impact on ion-pairing) should be modified. Secondly, the non-bonded interaction parameters (e.g. van-der-Waals parameters or partial charges) that are to be optimised, need to be chosen. Guided by the aim to use well-established biomolecular forcefields, it is important to maintain compatibility with the standard parametrisations of these forcefields. Therefore, we assume that the ion–water interactions were parametrised carefully based on single-ion hydration properties. This enables us to focus the optimisation efforts on the ion–ion interactions. Furthermore, we will restrict any changes of the FF to the pair-specific heteroatomic Lennard-Jones parameters in order to ensure compatibility with the modelling of other compounds.

The modifications of the GROMOS and OPLS-AA FF that were proposed by Project et al. [4, 5] change the short-range interactions between calcium and acetate with the correct tendency. Yet, neither of these two FF reproduces the experimental association constant of calcium acetate ("G 54A7 EP" and "OPLS-AA EP" in figure 4.10). This modified version of the OPLS-AA FF furthermore does not reproduce the relative depths of the CIP and SIP minima as predicted by the DFT-based simulations (figure 4.5). Therefore, it is necessary to optimise the van-der-Waals ion–ion interactions even further.

The modification of the OPLS-AA FF as proposed by Project consists of an increase in the Lennard-Jones parameter $\sigma_{\mathrm{Ca-OM}}$ by 3.7 % while the other Lennard-Jones parameter $\varepsilon_{\mathrm{Ca-OM}}$ remains unchanged (OM represents the atom type of the carboxylate oxygen of the acetate anion: cf. table 4.1) [5]. For the GROMOS FF, Project et al. increased $C^{(6)}_{\mathrm{Ca-OM}}$ by 32.7 % and decreased $C^{(12)}_{\mathrm{Ca-OM}}$ by 22.9 % [4]. We further optimised these Lennard-Jones

parameters with two major targets in mind: the new parameters should modify the ion–ion interactions such that (*i*) the association constant matches the experimental results, and (*ii*) the relative depths of the CIP and SIP minima of the PMF agree with those from the BOMD simulations (within errorbars). For the OPLS-AA FF, these targets are met if $\sigma_{\mathrm{Ca-OM}}$ is increased by 5.8 % with respect to the original FF parameter, while the energy parameter $\varepsilon_{\mathrm{Ca-OM}}$ remains unaltered. In contrast to the work of Project et al., our optimisation of the GROMOS FF is based on its latest version (54A8), which comprises an increased partial charge of the OM atom type with respect to its predecessors (table 4.1). Starting from the standard parameter values of version 54A8, we increased $C^{(6)}_{\mathrm{Ca-OM}}$ by 19.6 % and decreased $C^{(12)}_{\mathrm{Ca-OM}}$ by 13.7 % in order to meet our two goals. Table 4.2 contains the new model parameters and figure 4.2 shows the resulting optimised ion–ion interactions for both FF ("G 54A8 Ca-OM modified" and "OPLS-AA Ca-OM modified").

The results in figure 4.10 and figure 4.5 show that the modified versions of the GROMOS and OPLS-AA FF proposed here both satisfy the two conditions formulated above: (*i*) The calculated association constants fall into the range of experimental data and (*ii*) the relative depths of the CIP and SIP minima of the PMF curves agree well with the results of the BOMD simulations. From figure 4.9 we see that the PMF profiles of these optimised FF differ from those of the respective original FF ("G 54A8 standard" and "OPLS-AA standard") only at small C-Ca distances (range of contact ion pairs) while the minima of the SIP states remain unchanged. This was to be expected as only the van-der-Waals interactions were modified and the far-reaching Coulomb interactions - which have the potential to considerably manipulate the minimum of the SIP state - were kept constant.

In spite of the success of these modified FF in modelling important properties of the ion-pairing of calcium acetate, the PMF curves of the two FF possess two features which diverge from the DFT-based reference (figure 4.5): The free energy of the monoCIP state differs significantly from the one of the biCIP state, which leads to a strong underestimation of the probability of monoCIPs. Furthermore, the free energy barrier between SIPs and CIPs is more pronounced for both FF than for the DFT-based simulations. These two inaccuracies affect the sampling and the dynamics of structural changes of the ion-pairing of calcium acetate. However, a further improvement of the PMF characteristics appears to be not possible, if the FF modifications are limited to the direct ion–ion interactions. This raises the question whether a further optimisation of these modified FF is possible, if the search domain is not limited to ion–ion interactions but also includes ion–water interactions.

A modification of the ion–water interactions might seem to be imprudent, as these interactions have already been optimised on single ion hydration properties before. However, it is worth analysing how small changes in ion–water interactions affect the characteristics of the ion–ion PMF in order to assess the effect of future FF reparametrisations.

For the FF models, the most important ion–water interactions are those between the calcium ion and the oxygen atoms of water ("Ca-OW") and those between the oxygen atoms of acetate and water ("OM-OW"). As mentioned above, the only significant differences between the GROMOS FF (based on version 54A7) and the one by Raiteri et al. are the repulsions of the OM-OW and of the OM-HW interactions. Raiteri's FF performs better with respect to both the depth of the monoCIP minimum and the barrier between CIPs and SIPs. Therefore, it is worth studying whether a softer OM-OW repulsion can be introduced into the Lennard-Jones type FF, leading to an improved depth of the monoCIP minimum and to a decreased barrier between CIPs and SIPs without corrupting the water structure as strongly as Raiteri's FF does. Furthermore, we studied whether a manipulation of the Ca-OW interactions can lead to similar effects. For this analysis, we have chosen the OPLS-AA FF with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 % ("OPLS-AA Ca-OM modified"). Based on this FF, we have introduced small modifications in the four FF parameters that dominate the (non-electrostatic) ion-water interactions: $\sigma_{\mathrm{Ca-OW}}$, $\varepsilon_{\mathrm{Ca-OW}}$, $\sigma_{\mathrm{OM-OW}}$, and $\varepsilon_{\mathrm{OM-OW}}$.

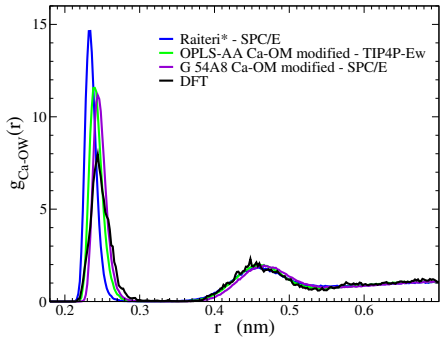
The Ca-OW radial distribution function (figure 4.11) of the optimal OPLS-AA FF shows that this FF nicely reproduces the positions of the RDF-peaks of the DFT-based simulations. From this perspective, no modifications in $\sigma_{\mathrm{Ca-OW}}$ are needed.

However, the solvation structure of the calcium ion as it is modelled by the OPLS-AA FF is still far from being optimal. Figure 4.12 shows that the coordination number of water around the calcium ion is overestimated for ion pair structures (figure 4.12a) as well as for the unpaired calcium ion (figure 4.12b). Therefore, it seems worthwhile to modify both $\sigma_{\mathrm{Ca-OW}}$ and $\varepsilon_{\mathrm{Ca-OW}}$ with the goal to improve the characteristics of the PMF curve and the coordination of water around the calcium ion.

Figure 4.13 illustrates the scope of the changes in the van-der-Waals interactions that are introduced by the modifications of $\sigma_{\mathrm{Ca-OW}}$ (decreasing $\sigma_{\mathrm{Ca-OW}}$ by 0.002 nm) and $\varepsilon_{\mathrm{Ca-OW}}$ (decreasing $\varepsilon_{\mathrm{Ca-OW}}$ by 0.132 kJ/mol).

The OM-OW RDF and the respective hydrogen-bond angle distribution (figure 4.7) of the optimal OPLS-AA FF nicely match BOMD simulations as

(a) bidentate CIP



(b) unpaired ions

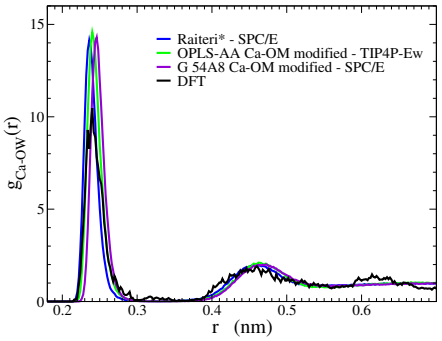
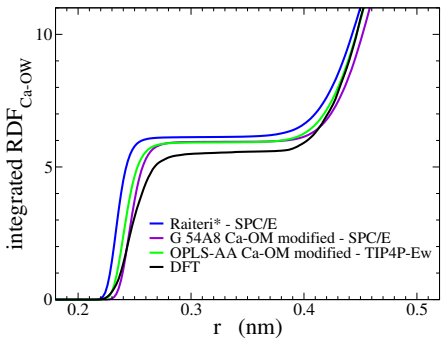


Figure 4.11: Structure of water around the calcium cation for a biCIP structure and an unpaired ion structure. Abbreviations: cf. caption of 4.7.

(a) bidentate CIP state



(b) unpaired ions

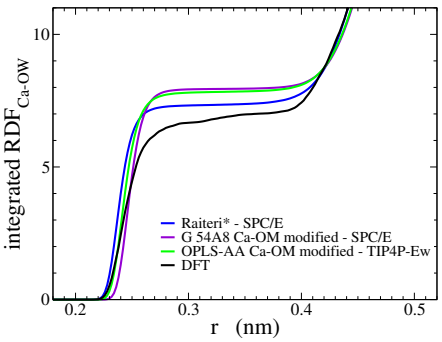


Figure 4.12: Number of water molecules around the calcium cation as a function of the Ca-OW distance for a biCIP structure and an unpaired ion structure. Abbreviations: cf. caption of figure 4.7.

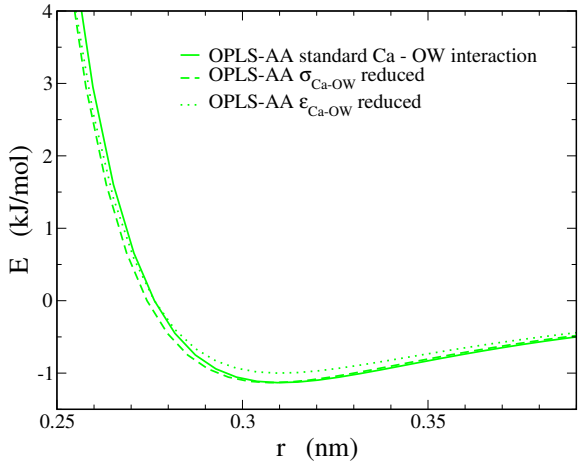


Figure 4.13: Lennard-Jones interaction potentials between the calcium cation ("Ca") and the oxygen atom of water ("OW") of the standard OPLS-AA forcefield and of two modified versions: $\sigma_{\mathrm{Ca-OW}}$ reduced - decreasing the standard OPLS-AA parameter $\sigma_{\mathrm{Ca-OW}}$ by 0.002 nm; $\varepsilon_{\mathrm{Ca-OW}}$ reduced - decreasing the standard OPLS-AA parameter $\varepsilon_{\mathrm{Ca-OW}}$ by 0.132 kJ/mol.

for the water structure around the acetate anion. The remaining small dissimilarities between the results of classical and BOMD simulations have been observed for the solvatation of glutamate before [187], and can be attributed to the different approaches of the two simulation methods to account for the omitted electronic polarisation. Therefore, the ion–water interactions should only be moderately changed if at all. Figure 4.14 illustrates the scope of the changes in the van-der-Waals interactions that we introduced into the optimal OPLS-AA FF by the modifications of $\sigma_{\mathrm{OM-OW}}$ (decreasing $\sigma_{\mathrm{OM-OW}}$ by 0.002 nm) and $\varepsilon_{\mathrm{OM-OW}}$ (decreasing $\varepsilon_{\mathrm{OM-OW}}$ by 0.074 kJ/mol).

The modifications of the four ion–water FF parameters $\sigma_{\mathrm{Ca-OW}}$, $\varepsilon_{\mathrm{Ca-OW}}$, $\sigma_{\mathrm{OM-OW}}$, and $\varepsilon_{\mathrm{OM-OW}}$ all have a very similar effect on the PMF of calcium acetate: figure 4.15 shows that all of these modifications affect all three minima of the PMF curve. They all improve both the depth of the monoCIP minimum with respect to the biCIP minimum and they improve the barrier between CIP and SIP states. But they also strongly influence the relative depths of the CIP and SIP states (affecting the relative probabilities of CIPs and SIPs) and very likely have a strong effect on the total attraction of the two ions which will become manifest in a significant alteration of the calculated association constant. The optimisation of the relative probability of monoCIPs through changes in the ion–water interactions thus has significant side effects on the original goals to meet the experimental association con-

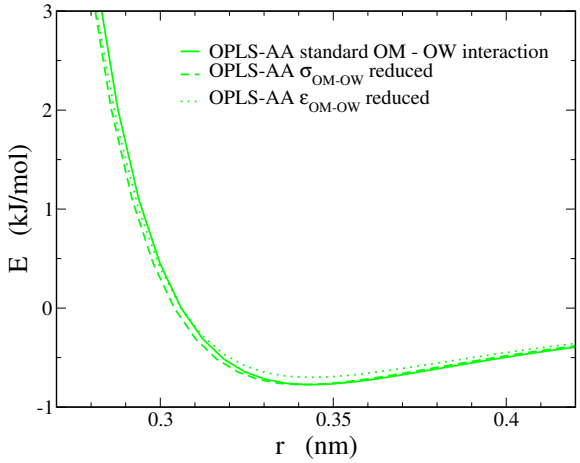


Figure 4.14: Lennard-Jones interaction potentials between the oxygen atom of water ("OW") and an oxygen atom of acetate ("OM") of the standard OPLS-AA forcefield and of two modified versions: $\sigma_{\mathrm{OM-OW}}$ reduced - decreasing the standard OPLS-AA parameter $\sigma_{\mathrm{OM-OW}}$ by 0.002 nm; $\varepsilon_{\mathrm{OM-OW}}$ reduced - decreasing the standard OPLS-AA parameter $\varepsilon_{\mathrm{OM-OW}}$ by 0.074 kJ/mol.

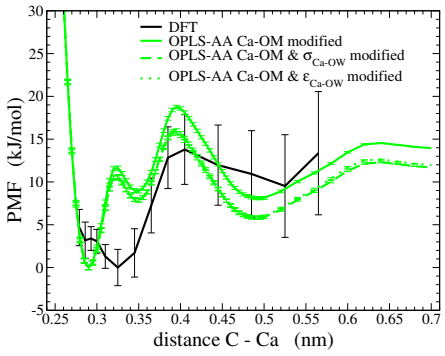
stant and the correct balance between CIPs and SIPs. It therefore appears not to be a viable strategy to improve FF with respect to the relative probabilities of biCIPs and monoCIPs, unless one undertakes a global optimisation approach involving all the Lennard-Jones pairing parameters.

4.5 Conclusions

If parametrised with care, classical interaction potentials are in principle capable of describing the peculiarities of ion-pairing in solution. Existing well-established biomolecular forcefields such as GROMOS or OPLS-AA can describe important aspects of ion pairing even if complex ions such as charged peptide side chains are involved. However, for the system of calcium acetate dissolved in water, none of the unmodified standard forcefields analysed here gave a realistic description of thermodynamic *and* structural details of the ion-pair formation. Therefore, small modifications of the forcefield parameters were necessary.

The strategy that we have developed for the optimisation of classical forcefields with respect to the interactions of complex ions has been adapted to the basic conditions that experimental reference data is limited and that

(a) modified Ca-OW interactions



(b) modified OM-OW interactions

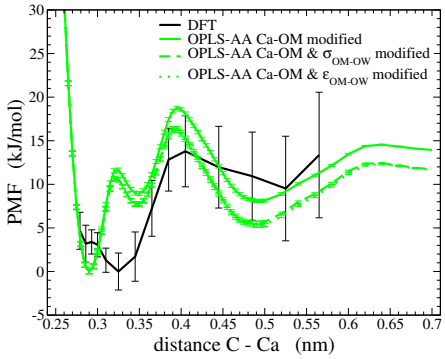


Figure 4.15: Influence of modifications in the Ca-OW and OM-OW interaction potentials on the Potential of Mean Force of calcium acetate. The curves have been shifted vertically such that their lowest CIP-minimum (either bidentate or monodentate) is aligned at 0 kJ/mol. Abbreviations: OPLS-AA Ca-OM modified - OPLS-AA forcefield with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 %; OPLS-AA Ca-OM & $\sigma_{\mathrm{Ca-OW}}$ modified - OPLS-AA forcefield with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 % and a decrease in $\sigma_{\mathrm{Ca-OW}}$ by 0.002 nm; OPLS-AA Ca-OM & $\varepsilon_{\mathrm{Ca-OW}}$ modified - OPLS-AA forcefield with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 % and a decrease in $\varepsilon_{\mathrm{Ca-OW}}$ by 0.132 kJ/mol; OPLS-AA Ca-OM & $\sigma_{\mathrm{OM-OW}}$ modified - OPLS-AA forcefield with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 % and a decrease in $\sigma_{\mathrm{OM-OW}}$ by 0.002 nm; OPLS-AA Ca-OM & $\varepsilon_{\mathrm{OM-OW}}$ modified - OPLS-AA forcefield with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 % and a decrease in $\varepsilon_{\mathrm{OM-OW}}$ by 0.074 kJ/mol; DFT - First-Principles Molecular Dynamics simulations based on DFT.

the ion–water interactions are already well optimised in the original FF.

Due to the influence of both the ion–water and the ion–ion interactions on important characteristics of the pairing of ions, it might seem worth *re*considering the common practice, to first optimise ion–water interactions based on single ion hydration properties and to adjust the ion–ion interactions to properties of ion-pairs afterwards. However, if one chooses to optimise the ion–ion and ion–water interactions simultaneously, the final set of parameters that results from such a high-dimensional optimisation procedure would very likely represent a completely new FF.

Our study shows that such a major effort is actually not necessary. In order to describe the pairing of calcium acetate in solution properly, it is sufficient to modify the van-der-Waals ion–ion interaction parameters of one of the existing biomolecular forcefields. Thereby, a further modification of the standard ion–water interactions, which were usually optimised on single ion hydration properties, can be omitted. A prerequisite for the optimisation of ion–ion interaction parameters is the availability of thermodynamic as well as structural reference data. For calcium acetate experimental data are sparse. Even though computationally costly, it is sufficient to complement these experimental data with BOMD simulations in order to fine-tune the forcefield, so that it gives a realistic description of the free energy as well as structural details of ion-pairing.

This approach to forcefield-optimisation has been tested for two different popular forcefields (GROMOS and OPLS-AA). In both cases, the aforementioned modification of van-der-Waals ion–ion interactions allows us to reproduce successfully the essential properties of the reference data. Following this approach, we found that both the modified OPLS-AA FF with an increase in $\sigma_{\mathrm{Ca-OM}}$ by 5.8 % and the modified GROMOS 54A8 FF with an increase in $C^{(6)}_{\mathrm{Ca-OM}}$ by 19.6 % and a decrease in $C^{(12)}_{\mathrm{Ca-OM}}$ by 13.7 % describe the ion-pairing of calcium acetate well. The underlying criteria of this choice are the agreement with the experimental association constant (figure 4.10), and with the most important characteristics of the PMF of the two ions as calculated from BOMD simulations (figure 4.5: relative probabilities of CIPs and SIPs and height of the free energy barrier between CIPs and SIPs). Based on these optimised biomolecular forcefields, the direct interactions of charged glutamate side chains with solvated calcium ions can be described properly.

Applying this realistic description of ion–peptide interactions, it is now possible to look into the accurate sampling of the conformational phase space of complexes of oligoglutamates with calcium ions in solution. The study of these complexes will give us the opportunity to analyse the interactions between oligopeptides of different chain lengths and solvated ions and their

influence on nucleation and crystal growth of biominerals. The results of this study are presented in chapter 5.

Chapter 5

Conformational Sampling of Oligoglutamates in Contact with Calcium Ions

The identification of a suitable forcefield for modelling the interactions of calcium ions with the side chains of glutamates, that was presented in the previous chapter, enables us to study the formation of complexes of calcium ions and oligoglutamates that might play important roles during the nucleation and crystal growth of calcium-containing minerals. As the conformational phase space of such complexes increases considerably with the peptide's chain length and - as observed earlier - the presence of ions significantly obstructs the conformational sampling of the peptides, enhanced-sampling methods are needed to gain a sufficient sampling of structures within reasonable times and computational efforts. In the following, we demonstrate how the Hamiltonian Replica Exchange method can be applied successfully in combination with the sketch-map analysis method to obtain an overview over the most important structures of these complexes *via* Molecular Dynamics simulations. The various peptide–ion complexes exhibit characteristic patterns of distances between the calcium ions that resemble some of the calcium distances found in calcium oxalate crystals that form in the presence of oligoglutamates.

5.1 Motivation

Complexes of oligoglutamates and calcium ions in solution As described in chapter 1, biomolecules might influence the formation of minerals in solution in various ways. Besides interacting directly with existing crystal surfaces, peptides may as well affect the crystallisation in numerous ways

by interacting with the solvated constituents of the mineral. Fischer et al. studied the effect of oligoglutamates on the formation of calcium oxalate and found that the chain length of these peptides influences the nucleation rate, the solubility of calcium in solution and the phase of the forming crystal [2]. While peptides consisting of five or less monomers have only a limited effect, peptides of ten or more monomer-units considerably alter the characteristics of the final crystal. When decamers of glutamate (from now on referred to as decaGLU) were present in solution during these experiments (concentration of 0.1 mmol·L^{-1}), the induction time of crystal nucleation increased by a factor of two to three compared to those experiments with pentamers of glutamate (pentaGLU). A possible explanation for this observation might be that the peptides withdraw (depending on their chain length) different amounts of calcium ions from solution. Another remarkable observation during these experiments concerns the composition of the crystals. In the absence of any peptides, as well as in the presence of pentaGLU, calcium oxalate trihydrate (COT) was formed, while decaGLU stabilised calcium oxalate dihydrate (COD) and led to the simultaneous crystallisation of COT and COD. Based on these findings, Fischer et al. concluded that "Calcium ions precomplexed by the peptides may serve as nucleation centers reducing the free energy of nuclei formation, and the structure of the formed calcium-ion complex may be determining for the structure of the resulting crystal." [2] A similar hypothesis was described by Bulo et al. regarding the precipitation of calcium carbonate in the presence of polyacrylates [97]. In the deprotonated state, polyacrylates, which possess side chains similar to the peptides studied here, form complexes with calcium ions. According to Bulo et al., the distances between the calcium ions of these complexes agree in some cases remarkably well with those distances found in the crystal structure of calcite.

Regarding the models for nucleation and growth of mineral phases, there has been a slight shift of paradigm in the last years. Recent studies show that in order to understand the mechanisms by which molecular modifiers are able to confer a specific structural pattern to the forming crystal, it is necessary to consider non-classical nucleation pathways. These might involve pre-nucleation clusters (PNCs) or liquid droplets of the mineral constituents, which are formed by liquid–liquid phase separation. PNCs are dynamic (meta-)stable aggregates of the constituents of the forming solid. They are molecular precursors to the phase nucleating from solution [76]. The existence of PNCs has been shown for several minerals such as calcium carbonate [79, 80] and calcium phosphate [188, 189, 190]. The importance of PNCs can be seen from the fact that they can have encoded structural motifs resembling one of the amorphous or crystalline polymorphs [76]. On

the other hand, liquid droplets of calcium carbonate, that form through liquid–liquid phase separation from the bulk solution, were observed in several experimental set-ups [82, 32, 91, 77, 81]. These droplets are considered to be liquid precursor phases to amorphous calcium carbonate (ACC). PNCs and liquid–liquid phase separation have been observed during the crystallisation in pure systems, i.e. without any additives. However, biomolecular modifiers have been shown to stabilise both PNCs and liquid mineral droplets [6, 191, 77, 192]. For example, Picker et al. found that peptides of glutamic acid stabilise PNCs of calcium carbonate [191]. If biomolecular modifiers were able to not only stabilise but also to manipulate the structure of PNCs or liquid mineral droplets, ion–peptide complexes might be viewed as the starting point of a cascade of structural motif encoding.

Here, we want to study peptide–ion complexes formed by oligoglutamates and calcium ions by means of Molecular Dynamics simulations. The interaction of other (monovalent) cations with oligoglutamates in solution has been studied both experimentally [148] and by computer simulations [149, 150] before. However, no such information is available for calcium ions in contact with oligoglutamates. Fischer et al. observed remarkable differences in the impact of pentaGLU and decaGLU on crystal nucleation and growth of calcium oxalate [2]. Therefore, our study is focused on the interactions of those two oligopeptides with calcium ions in solution, with the goal to unravel some of the processes that lead to the differences in crystallisation observed in the presence of pentaGLU and decaGLU. The complexity of the structure of such ion–peptide aggregates underlines that the accurate modelling of such structures delicately depends on a proper FF parametrisation. The optimised FF presented in the previous chapter, which have been fine-tuned with respect to structural and thermodynamic properties of ion-pairing, are well-suited for these simulations.

Enhanced-sampling *via* BP-REMD simulations A fundamental difficulty, which is frequently encountered in molecular simulations, is the adequate sampling of phase space [133]. This is especially the case, if the minima of the potential energy surface are separated by regions of high potential energy, which are less likely to be sampled. As shown in chapter 3, the strong interactions between the charged side chains of oligoglutamates and the calcium ions may lead to the formation of rigid aggregates that strongly limit the frequency of conformational changes and thus hinder the exploration of the conformational phase space. One approach that is often applied to circumvent this problem is to increase the temperature during the simulation in order to facilitate the escape from local potential energy min-

(a) GROMOS FF (b) OPLS-AA FF

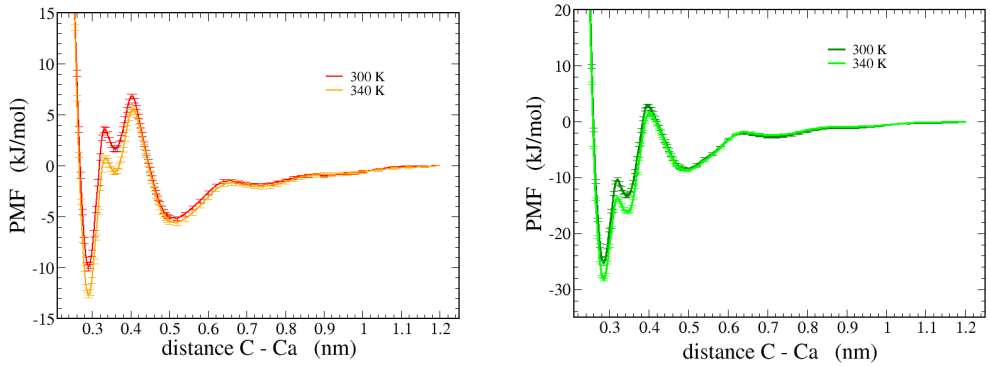


Figure 5.1: Influence of an increase in temperature on the Potential of Mean Force (PMF) between a calcium and an acetate ion. The PMF is plotted as a function of the distance between the carbon atom of the carboxylate group and the calcium ion.

ima [127, 128, 129]. In the system under investigation, the PMF calculations of calcium acetate at different temperatures reveal that an increase in temperature strengthens the attraction between calcium and carboxylate groups even further for both force fields (figure 5.1).

This shows that the free energy of ion-pairing of calcium ions and the carboxylate groups of glutamate side chains in aqueous solution is dominated by a gain in entropy. Any attempt to enhance the conformational sampling in this system through an increase in temperature (as in temperature Replica Exchange MD) will therefore not be successful. Instead, another method to enhance the conformational sampling is needed.

In general, a number of different advanced sampling techniques is available for molecular simulations [124, 193], such as metadynamics [112, 134, 194], implicit solvent simulations [195], or Hamiltonian replica exchange methods [123, 125, 132]. As each of these methods has its strengths and weaknesses, the optimal choice among these methods depends on the system of interest and the available resources. Metadynamics has the drawback, that a successful application critically depends on the choice of suitable collective variables (CVs). The CVs have to be good descriptors of the phenomenon of interest and for efficiency reasons, they need to be limited in their number. Therefore, a sensible choice is often difficult *a priori*. Due to these difficulties, metadynamics is rather well-suited as a second enhanced sampling method that is applied after the characteristics of the system under investigation

have been studied by means of another enhanced sampling technique. Simulations, in which an explicit treatment of the solvent molecules is omitted by the application of the implicit solvent concept, have the advantage that the computational costs of the simulation are significantly reduced. Therefore, the size of the system studied and the simulated time can be extended to values which are not accessible to explicit solvent simulations. However, for these simulations it is necessary to develop a system-specific implicit-solvent model, which adequately captures the influence of the solvent on the processes of interest. This task is very time-consuming, if possible at all.

The basic idea of Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) simulations is to modify the Hamiltonian of the system in a way that facilitates the sampling of rare events, e.g. conformational changes. As there are many different ways to manipulate the Hamiltonian of a system, many different variants of the H-REMD method exist. Some of these methods have been specifically designed for the conformational sampling of biomolecules. Ostermeir and Zacharias compared different H-REMD methods to sample peptide and protein conformations and found that one of the most promising techniques was the dihedral angle biasing-potential (BP-REMD) method [132]. The BP-REMD method - as it is applied here - is based on the observations of Straatsma and McCammon that due to non-bonded interactions and steric effects, the free energy barriers to dihedral rotation in biomolecules are a lot larger than the dihedral angle potential energy function of the FF [133]. Therefore, it is not sufficient to scale-down the dihedral potential function of the FF in order to enhance the sampling of rotational isomeric states in molecular simulations of polypeptides in water. Instead, a biasing-potential of the magnitude of the *actual* free energy barriers to dihedral rotation is needed, which effectively flattens the free energy barriers that hinder conformational transitions. As in any other Replica Exchange method, several non-interacting copies of the system ("replicas") are simulated in parallel with the possibility to exchange configurations between the different replicas. REMD methods therefore have the advantage that they are suitable for parallel computing due to the small amount of necessary communication between the different replicas. Even more important is the fact, that the BP-REMD method has been specifically designed for systems such as the one under investigation. Therefore, this method was selected here to increase the sampling of structures of oligoglutamates and calcium ions. A detailed description of this method, its theoretical foundations and the determination of a suitable biasing-potential is given in chapter 2.

5.2 Methods

Systems under investigation In their experiments, Fischer et al. used a calcium concentration of 1 mmol·L^{-1} irrespective of the chain length and the concentration of the peptides. The peptide concentration was varied between 0.001 mmol·L^{-1} and 0.2 mmol·L^{-1} [2]. Regarding the peptide concentration, Fischer et al. did not consider the concentrations of the peptide's monomers (and thus the number of functional groups), but only the number of peptide molecules in solution. Therefore, the ratio of calcium ions per peptide is always 10 at a peptide concentration of 0.1 mmol·L^{-1}, irrespective of the chain length of the peptide. Accordingly, the ratio of calcium ions per side chain of the oligopeptides differs depending on the peptide chain lenght at a given peptide concentration. Our goal is to study a system which qualitatively resembles the situation in solution during nucleation and crystal growth of calcium-containing minerals. However, a calcium concentration of 1 mM corresponds to 1660.54 nm^3/ion. The peptide concentration (e.g. 0.1 mM) is typically even one order of magnitude lower in the experiments. This means that very large systems would be necessary in order to model realistic ion and peptide concentrations. Considering the computational costs of large systems and the necessary time scales that need to be simulated in the given system, it is clear that such system sizes cannot be achieved here. But as we are interested in processes of nucleation and crystal growth, during which the local concentrations of ions are significantly higher than the average concentrations in the system, it is acceptable (and necessary) to simulate systems of higher ion concentrations. In accordance with those experiments of Fischer et al. with peptide concentrations of 0.1 mM, we have chosen a ratio of calcium ions per peptide molecule of 10, independent of the chain length of the peptide. In order to neutralise the simulated systems, 15 chloride ions were added to the system of one (fully deprotonated) pentaGLU ion with 10 calcium ions, while 10 chloride ions were necessary to neutralise the system of one (fully deprotonated) decaGLU ion with 10 calcium ions. In order to reveal the influence of the calcium ions on the conformations of the oligoglutamates, the same systems were also simulated without any calcium and chloride ions.

Computational details The BP-REMD method (cf. chapter 2) was applied to the four different systems (pentaGLU with and without counter-ions and decaGLU with and without counter-ions). In these simulations, the "higher" replicas differ from the reference replica by modified peptide backbone dihedral potentials for the dihedral angles φ and ψ that will allow for an enhanced rate of conformational changes. The details of the two biasing-

potentials, which were used to modify the dihedral potentials accordingly, are described in chapter 2. In order to construct the biasing-potentials, the free energy barriers to backbone dihedral rotation (i.e. the PMF) were determined for the central monomer-unit of pentaGLU *via* well-tempered metadynamics [134]. It is important to note, that these PMFs were measured in the absence of calcium. As we will see later on, calcium ions contribute additionally to the torsional free energy barriers of the peptide. But in an attempt to consider the influence of calcium ions and other non-bonded interactions and steric effects on the PMF separately, the PMFs were determined in the absence of calcium ions. The various replicas of one simulation differed in the extent to which the dihedral biasing-potentials were applied: while the "lowest" replica possessed unmodified peptide backbone dihedral potentials, the biasing-potentials were fully exerted only in the "highest" replica. The BP-REMD simulations were executed employing the H-REMD utilities implemented in GROMACS 4.6. In GROMACS, the potentials of bonded interactions can be interpolated smoothly from state A ($\lambda = 0$) to state B ($\lambda = 1$) using the λ-dependence of potentials. The resulting potential of dihedral angle φ of any interpolated state reads [136, 137]:

$$V(\varphi) \;=\; [(1-\lambda)k^A + \lambda k^B] \cdot (1 + \cos[n\varphi - (1-\lambda)\delta^A - \lambda\delta^B]) \qquad (5.1)$$

The parameter λ can be assigned to any value between 0 and 1, while k^A, k^B, δ^A and δ^B represent the respective constants of the dihedral potentials of state A and state B (general form of dihedral angle potentials: cf. eq. 2.24). The same function holds for the dihedral angle ψ.

In order to weaken the interactions between calcium ions and oligopeptides and thus to remove further obstacles to conformational changes, it would have been advantageous to simultaneously modify the van-der-Waals interactions between calcium ions and oxygen atoms of the carboxylate groups of the peptides in these simulations. Based on the analyses of different FF parametrisations presented in the previous chapter, the choice of suitable modified Lennard-Jones parameters would have been straightforward. However, the simultaneous modification of these interaction parameters during BP-REMD simulations was unfortunately not yet possible with the latest available GROMACS version. Therefore, the replicas of one simulation differed only in the peptide backbone dihedral potentials.

The number of replicas, the λ-values of each replica and the frequency of exchange attempts were chosen based on several BP-REMD test runs of 10 or 20 ns. The final values of these parameters were selected based on the goal to achieve rates of successful exchange attempts between neighbouring replicas of 20 to 30 %. This goal was chosen in accordance with the findings of

Rathore et al. and Kone and Kofke, who found out that an acceptance ratio of 20 to 30 % yields the optimal performance in temperature REMD simulations [196, 197]. The two systems containing pentaGLU were simultated with six replicas each. The λ-values of these replicas were chosen to be 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0. Eight different replicas were used in the BP-REMD simulations of the two systems containing decaGLU, with λ-values of 0.0, 0.21, 0.40, 0.54, 0.68, 0.81, 0.92, and 1.0. In all simulations, exchange attempts were made every 2 ps. The systems were simulated in boxes of a rhombic dodecahedron shape with periodic boundary conditions. The pentaGLU systems contained 3563 (pentaGLU without counter-ions) or 3538 (pentaGLU with counter-ions) water molecules in a box of approx. 108 nm^3. The decaGLU systems contained 8203 (decaGLU without counter-ions) or 8183 (decaGLU with counter-ions) water molecules in a box of approx. 249 nm^3. We used the GROMACS simulation package [136, 137] (version 4.6.3) for our simulations. The GROMOS 54A8 FF [108] with our optimised van-der-Waals interactions between calcium ions and oxygen atoms of the carboxylate group of the side chains of the oligoglutamate (LJ parameters of the Ca-OM interaction) as presented in chapter 4 (cf. table 4.2) was used in combination with the standard GROMOS parameters for calcium [158] and with the SPC/E water model [155]. All simulations were conducted in the NVT ensemble at a temperature of 300 K with the temperature being controlled *via* stochastic velocity rescaling [102] with a coupling time of 0.1 ps. The equations of motion were integrated with the leap-frog algorithm with a time step of 2 fs. The center of mass translation of the simulation box was removed every 200 fs and the neighbour list was updated every 20 fs. In all simulations, the PME method [159] was used to treat long-range electrostatic interactions with a grid spacing of 0.1 nm and an interpolation of 4. The Lennard-Jones interactions and the real-space Coulomb interactions were truncated at 1.4 nm. In accordance with the FF standards, no long-range dispersion correction was applied. Each replica was equilibrated separately for 20 ns (pentaGLU) or 40 ns (decaGLU), before the 400 ns BP-REMD simulations were started. The simulations were executed on the Hydra-Cluster of the Rechenzentrum Garching computing centre of the Max Planck Society on 12 nodes (pentaGLU) or 24 nodes (decaGLU), with 16 cores per node.

In order to be able to evaluate the success of the BP-REMD simulations, the systems of interest (replica 0 of the BP-REMD simulations) were also simulated as standard MD ("brute-force") simulations. All settings and parameter values were exactly as in the BP-REMD simulations, except that these simulations were performed in single precision in order to reach longer simulation times. The simulated time was 2 μs for the system of pentaGLU without calcium or chloride ions and 1 μs for the pentaGLU system contain-

ing counter-ions.

5.3 Results

Before we give a detailed description of the results of the BP-REMD simulations here, it is necessary to specify the meaning of the two terms "system" and "replica", which are frequently used in the context of replica exchange methods. Here, we refer to a *system* as a fixed set of molecules with a continuous trajectory. On its way through replica space, the Hamiltonian of the system changes according to the replica. In contrast, a *replica* is specified by a certain Hamiltonian. During the simulation, the systems are exchanged between the replicas. Therefore, the trajectories of molecular coordinates and velocities are discontinuous within one replica. The number of replicas and systems is always equal, and the systems are numbered according to the replica, in which they started the BP-REMD simulation.

5.3.1 Evaluation of the applied method

Some first conclusions on the success of the application of replica exchange methods can be drawn from the statistics of successful exchange attempts. These statistics are given in table 5.1 for the two BP-REMD simulations containing pentaGLU and in table 5.2 for the two BP-REMD simulations containing decaGLU.

The optimal acceptance ratio is between 20 and 30 % [196, 197]. With respect to this optimum, the fractions of successful exchange attempts are reasonable in all four simulations. The BP-REMD simulations could be further optimised by changing the "spacing" between the different replicas, i.e. the extent to which the biasing-potential is added to the standard dihedral potential in the respective replica (λ-values). This measure helps to obtain a similar acceptance ratio of exchange attempts in all replicas. Furthermore, the number of replicas could be increased for the decaGLU simulations, as a smaller spacing between the replicas increases the rate of successful exchange attempts.

However, the rate of successful exchange attempts between neighbouring replicas contains no information on the question, whether a random walk through replica space was achieved, i.e. whether all systems visited all replicas. In order to obtain a qualitative answer to this question, it is possible to analyse the "trajectory" of one system through replica space. Such trajectories are shown in figure 5.2 for all four BP-REMD simulations.

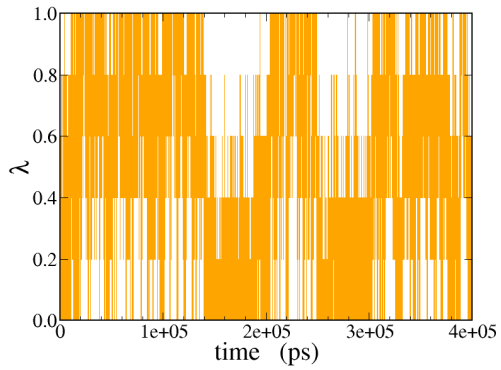
Table 5.1: Statistics of successful exchange attempts between neighbouring replicas of the two systems containing a pentamer of glutamate ("pentaGLU") during the BP-REMD simulations. Exchange attempts are either made between even replica pairs (i.e. 0-1, 2-3, and 4-5) or between odd replica pairs (i.e. 1-2, 3-4) at once. Replica 0: $\lambda = 0$, replica 1: $\lambda = 0.2$, replica 2: $\lambda = 0.4$, replica 3: $\lambda = 0.6$, replica 4: $\lambda = 0.8$, replica 5: $\lambda = 1.0$.

	pentaGLU no ions		pentaGLU with ions	
	total number	%	total number	%
total exchange attempts	200094		200052	
exchange attempts even/odd	100047		100026	
successful 0-1 exchanges	29009	29.0	41040	41.0
successful 1-2 exchanges	26746	26.7	32662	32.7
successful 2-3 exchanges	27237	27.2	24178	24.2
successful 3-4 exchanges	26729	26.7	29823	29.8
successful 4-5 exchanges	9326	9.3	14560	14.6

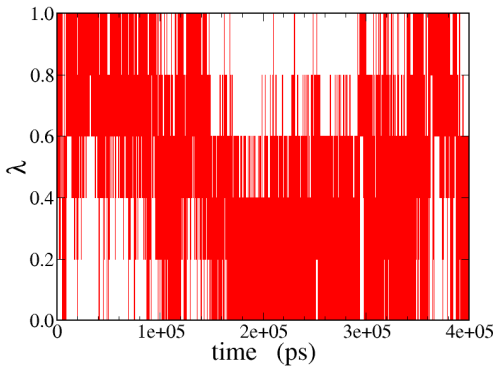
Table 5.2: Statistics of successful exchange attempts between neighbouring replicas of the two systems containing a decamer of glutamate ("decaGLU") during the BP-REMD simulations. Exchange attempts are either made between even replica pairs (i.e. 0-1, 2-3, 4-5, and 6-7) or between odd replica pairs (i.e. 1-2, 3-4, and 5-6) at once. Replica 0: $\lambda = 0$, replica 1: $\lambda = 0.21$, replica 2: $\lambda = 0.40$, replica 3: $\lambda = 0.54$, replica 4: $\lambda = 0.68$, replica 5: $\lambda = 0.81$, replica 6: $\lambda = 0.92$, replica 7: $\lambda = 1.0$.

	decaGLU no ions		decaGLU with ions	
	total number	%	total number	%
total exchange attempts	200039		200024	
exchange attempts even/odd	100020/100019		100012	
successful 0-1 exchanges	11552	11.5	17858	17.9
successful 1-2 exchanges	12656	12.7	15124	15.1
successful 2-3 exchanges	14621	14.6	12008	12.0
successful 3-4 exchanges	19753	19.7	15934	15.9
successful 4-5 exchanges	21993	22.0	23517	23.5
successful 5-6 exchanges	18700	18.7	24702	24.7
successful 6-7 exchanges	21415	21.4	10431	10.4

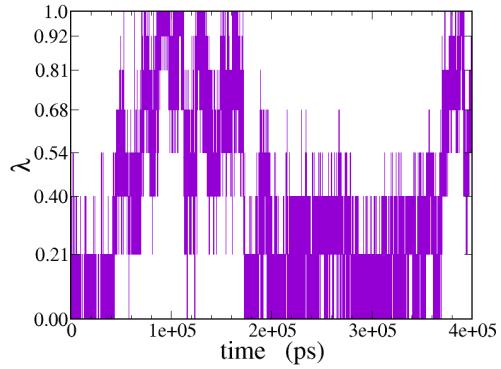
(a) pentaGLU without calcium ions



(b) pentaGLU with calcium ions



(c) decaGLU without calcium ions



(d) decaGLU with calcium ions

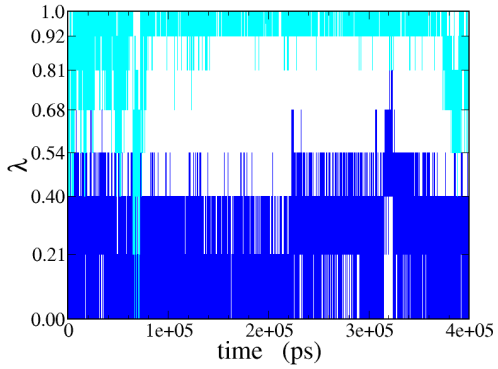


Figure 5.2: "Trajectory" of one system through replica space during each BP-REMD simulations. For each simulation, the system that started from replica 0 ($\lambda = 0$) in the beginning of the simulations is shown. For the system containing decaGLU with counter-ions, the trajectory of system 4 (starting from replica 4 with $\lambda = 0.68$) through replica space is also shown in order to illustrate, that the whole range of replicas was in principle accessible to all systems during the simulation.

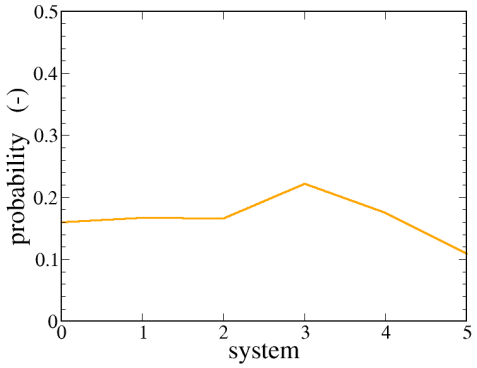
The sampling of replica space is sufficient during the simulation of pentaGLU without calcium ions and acceptable for the simulations of pentaGLU with ions and decaGLU without ions, even though further sampling in the latter two simulations would be advantageous. However, in the case of decaGLU with counter-ions, a complete random walk in replica space is clearly not achieved. System 0 (dark blue curve: starting from replica 0 with $\lambda = 0$) never reaches the highest replica ($\lambda = 1$) and mostly samples a limited region of replica space, even though the whole range of replicas is in principle accessible to all systems during the simulation, as the trajectory of system 4 (light blue curve: starting from replica 4 with $\lambda = 0.68$) illustrates (fig. 5.2d).

A quantitative analysis of the trajectories through replica space is given in figure 5.3, which shows histograms of the occurrence of each system of the simulation in replica 0 ($\lambda = 0$). A random walk of systems through replica space would yield a uniform distribution of equal probability for each system to visit one replica (1/6 for pentaGLU due to 6 replicas and 1/8 for decaGLU corresponding to 8 replicas). This is nearly achieved during the BP-REMD simulation of pentaGLU without calcium ions. The two simulations of pentaGLU with counter-ions and decaGLU without ions exhibit the necessity for a further improvement of the simulation set-up and the simulation of decaGLU in the presence of counter-ions needs some major revision of the set-up in order to achieve optimal BP-REMD simulations.

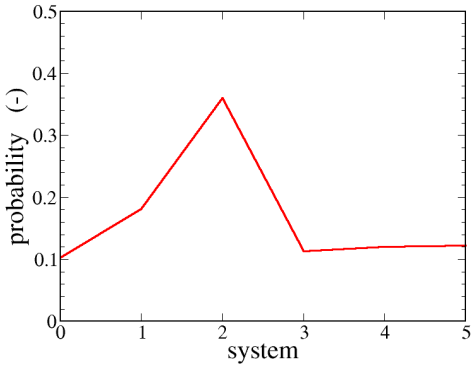
The non-uniform sampling of replica space during the simulation of decaGLU in the presence of counter-ions reveals that the Hamiltonians of neighbouring replicas are too dissimilar. One would indeed expect that a replica that locks into a very specific conformation (e.g. an ion-bridged loop) would be particularly reluctant towards an exchange to a Hamiltonian where the corresponding backbone torsion is more strongly unfavoured. In some replicas, the systems might also not have enough time to undergo major conformational changes and thus to adopt a state which is more compatible with the next replica. Obvious measures to correct for these shortcomings would be to increase the number of replicas and to decrease the frequency of exchange attempts. A more elaborate approach to support the random walk of systems in replica space would be to manipulate further parts of the Hamiltonian in order to smooth the energy landscape even further. For example, the van-der-Waals interactions between calcium ions and carboxylate groups of the peptides could be weakened in order to further alleviate the escape from local energy minima. This measure would enhance the conformational sampling in one replica even further and decrease the possible dissimilarities of conformations of neighbouring replicas.

Despite the above limitations, the BP-REMD simulations can neverthe-

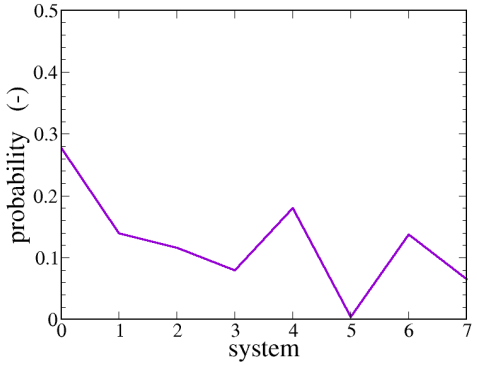
(a) pentaGLU without calcium ions



(b) pentaGLU with calcium ions



(c) decaGLU without calcium ions



(d) decaGLU with calcium ions

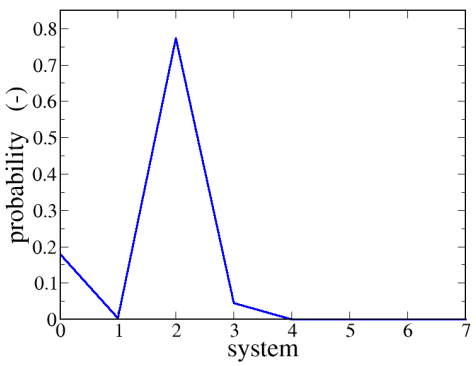


Figure 5.3: Histograms of the probability for each system to visit replica 0 ($\lambda = 0$) during the respective BP-REMD simulation. A random walk of systems through replica space would yield a uniform distribution of equal probabilities for each system to occur in one replica (1/6 for pentaGLU due to 6 replicas and 1/8 for decaGLU corresponding to 8 replicas).

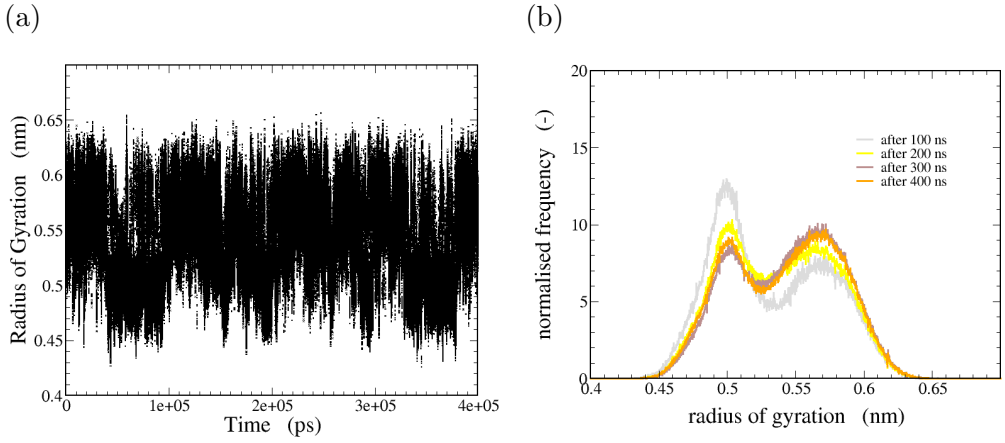


Figure 5.4: Evolution of the radius of gyration of pentaGLU during the BP-REMD simulation of pentaGLU without counter-ions. The radius of gyration is shown as a function of time (left) and in form of a histogram at several points of the trajectory (right).

less be viewed as a success. Frequent exchanges between replicas with different peptide backbone dihedral angle potentials were achieved and thus enhanced conformational sampling can be expected. However, especially the presence of calcium ions necessitates a further optimisation of the simulation set-up in order to obtain a fully comprehensive sampling of conformational phase space with reasonable computational effort. Besides improving simulation set-ups (such as an increased number of replicas), it might also be necessary to simulate the systems containing decaGLU longer than 400 ns. Furthermore, it is in principle possible to use the data of higher replicas during the analysis as well. For this purpose, the weighted histogram analysis method (WHAM) [198] that has been adapted for the analysis of results of H-REMD simulations [126] needs to be implemented.

Apart from these possibilities to further improve the replica exchange simulations, it remains to be shown that the BP-REMD simulations are more efficient than standard brute-force simulations. Figure 5.4a shows that conformational changes occur frequently during the BP-REMD simulation of pentaGLU without ions. The impression of an extensive sampling of the conformational phase space during this simulation is supported by figures 5.2a and 5.3a. The evolution of the histogram of the radius of gyration of pentaGLU (figure 5.4) shows that the sampling is close to convergence after 400 ns.

These findings can be compared directly with the results of the respective

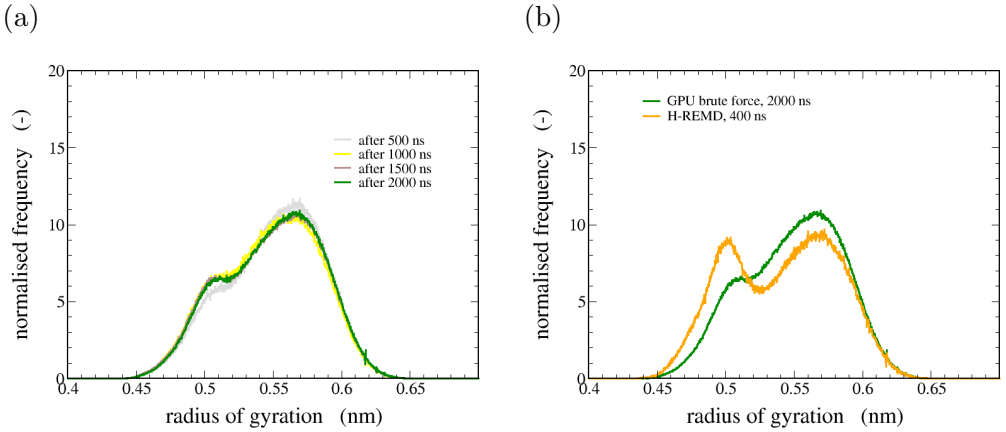


Figure 5.5: Radius of gyration of pentaGLU in the simulations without counter-ions. The histogram of the radius of gyration is shown at several points during the brute-force simulation (left) and the final histograms in the end of the BP-REMD and the brute-force simulation are compared (right).

brute-force simulation (figure 5.5). After 1000 ns of brute-force simulation, the histogram of the radius of gyration shows no significant changes any more (fig. 5.5a). However, this does not mean that this simulations has converged in the sense that the phase space has been sampled comprehensively. A comparison to the results of the BP-REMD simulation (fig. 5.5b) reveals that there is a region in conformational phase space, with a radius of gyration of pentaGLU of approximately 0.5 nm, which has not yet been sampled during the 2000 ns brute-force simulation. This is remarkable as out of the four systems considered here, this is the system in which comprehensive conformational sampling is most easily achieved.

In the presence of ions, the differences between brute-force and BP-REMD simulations in the efficiency of phase space sampling are even more striking. Figure 5.6 depicts a comparison of the histograms of the radius of gyration of pentaGLU in the end of the BP-REMD and the brute-force simulation. Areas of phase space with a radius of gyration of pentaGLU of less than 0.45 nm were not sampled at all during the brute-force simulation. It is unclear how long the brute-force simulation would need to be extended in order to ensure a comprehensive sampling of the conformational phase space. For pentaGLU, many microseconds of simulation time are probably necessary. As the phase space of the decamer of glutamate is a lot more complex, a sufficient sampling would probably take even a lot longer. This shows that simulations applying the BP-REMD method are definitively more efficient than brute-

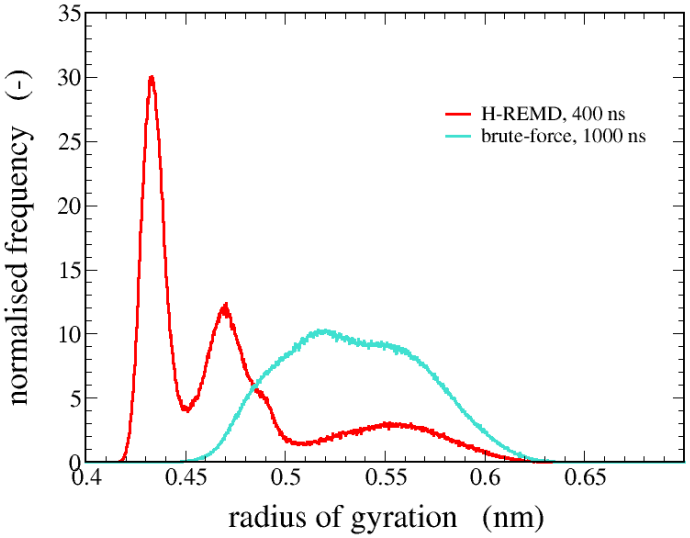


Figure 5.6: Comparison of the final histograms of the radius of gyration of pentaGLU in the end of the BP-REMD and the brute-force simulations of the system that contained counter-ions.

force simulations. Especially if calcium ions are present in solution, the brute-force approach is futile.

5.3.2 Influence of the peptide chain length on the formation of oligoGLU–calcium ion complexes

The total number of calcium ions that interact simultaneously with one oligopeptide molecule strongly depends on the chain length of the peptide. Figure 5.7 shows the histograms of the number of calcium ions in contact with pentaGLU and decaGLU during the BP-REMD simulations. From the PMF calculations presented in the previous chapter (cf. figure 4.9), the distances between calcium ions and the carbon atom of a carboxylate group in the states of a contact ion pair (CIP) and a solvent-shared ion pair (SIP) are well known. Therefore, a "contact" between a calcium ion and the side chains of the oligopeptides is here defined by a maximum distance of 0.66 nm, such that both CIPs and SIPs are considered. As it was to be expected, the calcium "uptake" of decaGLU is higher than that of pentaGLU. According to the positions of the peaks of the histograms, one decaGLU molecule interacts with approximately two more calcium ions than pentaGLU on average. This difference might be one explanation for the retardation of nucleation observed during the experiments [2], which is a lot stronger in the presence

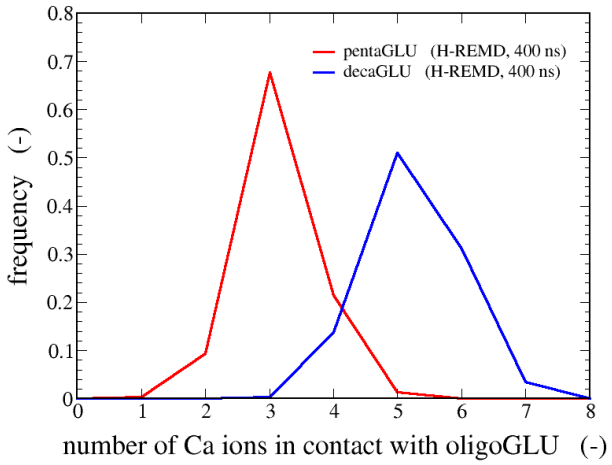


Figure 5.7: Histograms of the number of calcium ions that directly interact with the oligoglutamate ions during the BP-REMD simulations of pentaGLU and decaGLU. A "contact" between calcium ions and oligoGLU is defined by a maximum distance of 0.66 nm between calcium ion and carbon atom of the carboxylate groups of the side chains of the peptides (Ca-C). According to the findings of the previous chapter, a Ca-C distance of 0.66 nm includes both contact ion-pairs (CIP) and solvent-shared ion-pairs (SIP) in simulations with the optimised GROMOS 54A8 FF (cf. fig. 4.9).

of decaGLU than in the presence of pentaGLU. As a consequence of this extraction of free calcium ions from solution by oligoglutamates, the apparent concentration of calcium in solution is decreased, thus lowering the supersaturation as a driving force for precipitation.

The calcium ions induce salt-bridges between the side chains of oligoGLUs. These salt-bridges may form between any pair of side chains, leading to extended structures (figure 5.8a: salt-bridges between neighbouring side chains) or globular ones (figure 5.8b: salt-bridges between "distant" side chains).

The interactions with calcium ions modify the low-energy conformations of the oligopeptides which are sampled frequently. As figure 5.9 shows, the presence of ions leads to more compact structures. For both peptide chain lengths analysed here, the formation of ion–peptide complexes with small radii of gyration and end-to-end distances of the peptide is very probable. Naturally, the complexes that decaGLU forms in the presence of ions are significantly larger than those of pentaGLU. It can be expected that the diffusion of these larger complexes is slower than the one of the pentaGLU complexes. The different mobilities of these structures might contribute to the differences in the rates of crystal nucleation and growth that were ob-

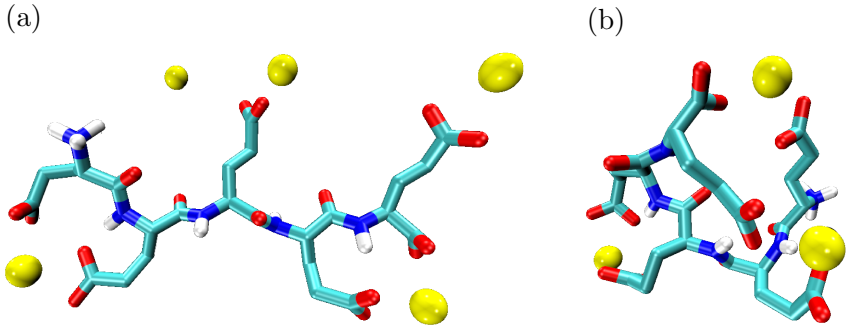


Figure 5.8: The calcium ions induce salt-bridges between the carboxylate groups of the side chains of oligoglutamates. These salt-bridges may form between any pair of side chains, leading to extended structures (left: salt-bridges between neighbouring side chains of pentaGLU) or globular ones (right: salt-bridges between "distant" side chains).

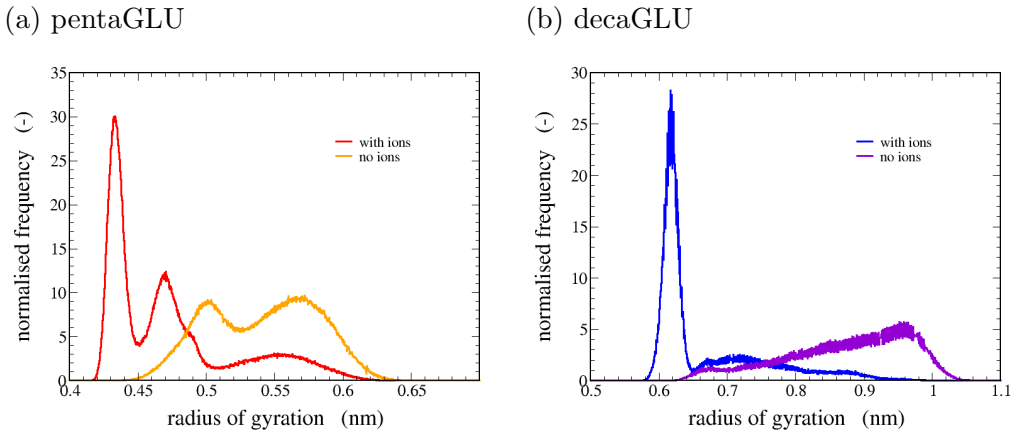


Figure 5.9: Comparison of the histograms of the radius of gyration of pentaGLU (left) and decaGLU (right) during the BP-REMD simulations with and without counter-ions.

served in experiments between pentaGLU and decaGLU [2].

Although the conformational sampling during the BP-REMD simulations is probably not comprehensive for those systems that contain calcium ions, the results still contain enough information to draw qualitative conclusions on the probabilities of the different conformations. In order to systematically analyse, how many different types of ion–peptide complexes exist and to determine their respective probabilities, some clustering of the sampled structures needs to be carried out. Here, the sketch-map method was used for this task.

At this point, it is necessary to introduce some terms and definitions which will be used throughout the following discussion. It is especially necessary to characterise the different phrases used to discriminate the various possible "spatial arrangements" of the peptides and ions in the system. A single point in the conformational phase space (i.e. a "snapshot" of a trajectory) will be referred to as a *configuration*. Configurations, which exhibit only minor differences in structural properties such as the peptide backbone dihedral angles, are summarised as one type of *conformation*. As will be shown later, some conformations, which are sampled during the simulations with counterions, feature the same specific pattern of salt-bridges (number and position of salt-bridges within the peptide) in spite of their differences in the backbone dihedral arrangement. These different conformations are said to belong to one type of *complex structure*.

From each of the four BP-REMD simulations, 40,000 configurations were analysed and clustered based on the similarity of the peptide's backbone conformation according to the sketch-map method as described in chapter 2. Figure 5.10 shows the projections of the configurations on a two-dimensional space for each of the four systems. The different densely populated regions in the two-dimensional space are referred to as *clusters*. Each cluster usually consists of a very limited number of conformations and they are coloured according to the absolute probability of data points per area in figure 5.10.

From these projections it is possible to identify some of the most probable structures of the ion–peptide complexes. As the same landmark points are used for the projection of the systems with and without ion (landmarks only differ between pentaGLU and decaGLU), the positions of the clusters in the two-dimensional projections can be compared between the simulations with and without ions. From these sketch-map projections it can therefore be inferred, that the presence of ions confines the conformational phase space of both types of oligopeptides (pentaGLU and decaGLU) in the sense that the number of low-energy conformations is reduced. As the phase space of pentaGLU is less complex than that of decaGLU, we will first focus on

pentaGLU here in order to illustrate the possible conlusions that can be drawn from the sketch-map analysis.

The sketch-map projection presented in figure 5.10 shows that some clusters have the same position in the projections of configurations of the pentaGLU simulation with and without calcium ions, but they are sampled with different frequencies. Other clusters exist only in the one case but not in the other. The reason for these differences is probably the presence of calcium ions, which attract the carboxylate groups of the side chains into positions, which are unfavourable in the absence of calcium ions due to electrostatic repulsion of the negatively charged side chains.

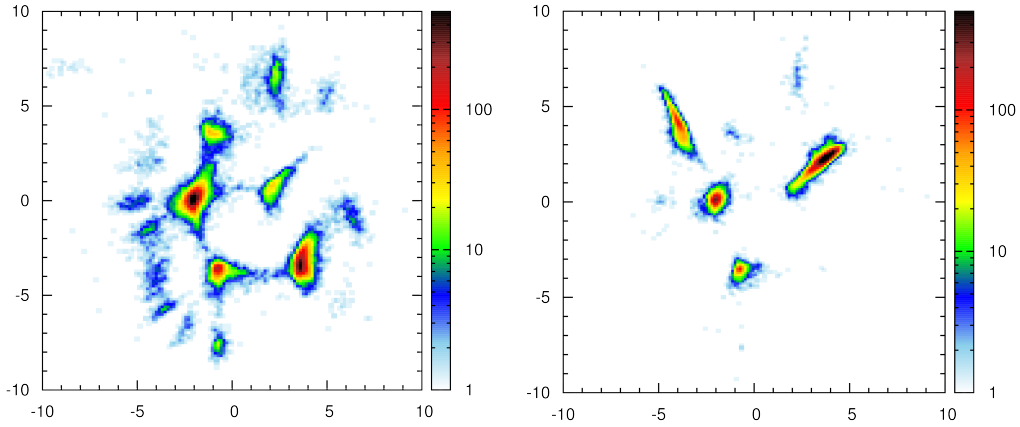
In order to analyse how the different clusters of the sketch-map projection differ from each other, the two-dimensional data points can be coloured according to different properties of the respective high-dimensional configuration, as for example the radius of gyration. Figure 5.11 shows the sketch-map projection of configurations of pentaGLU in the presence of calcium ions with a colour-code according to the radius of gyration of the peptide.

This colour-code shows that the sketch-map clustering (which is here based on backbone dihedral angles) is able to organise the ion–peptide configurations into meaningful conformations: configurations with similar radii of gyration are grouped into a collective cluster. Based on the radius of gyration and on the probabilities (cf. figure 5.10b) of the different clusters, 12 regions of the two-dimensional space were chosen for further analysis. Representative snapshots from these regions are shown in figure 5.11.

The most "central" cluster consists of elongated conformations (red and yellow colouring according to the radius of gyration), while the clusters surrounding this central cluster mostly contain more compact conformations (blue colouring according to the radius of gyration). It is interesting to see that there are several different clusters that contain compact conformations in the 2D space. As skecth-map assigns the respective configurations to different regions in the 2D space, they must differ considerably in their backbone conformations. However, in view of our main interest to study the interference of oligopeptides with nucleation or crystal growth of calcium-containing minerals, properties such as the number and positions of salt-bridges of the ion–peptide configurations are probably more important than the backbone conformations of the peptide alone. Omitting any information on backbone conformations and focusing solely on properties such as salt-bridge formation, four different types of complex structures can be identified among the twelve different regions of the 2D space. The colour of the frames surrounding the snapshots in figure 5.11 indicates to which type of complex structure the respective snapshot belongs. The structures that are indicated by a blue frame possess two salt-bridges between the neighbouring side chains 1 and 2

(a) pentaGLU without calcium ions

(b) pentaGLU with calcium ions



(c) decaGLU without calcium ions

(d) decaGLU with calcium ions

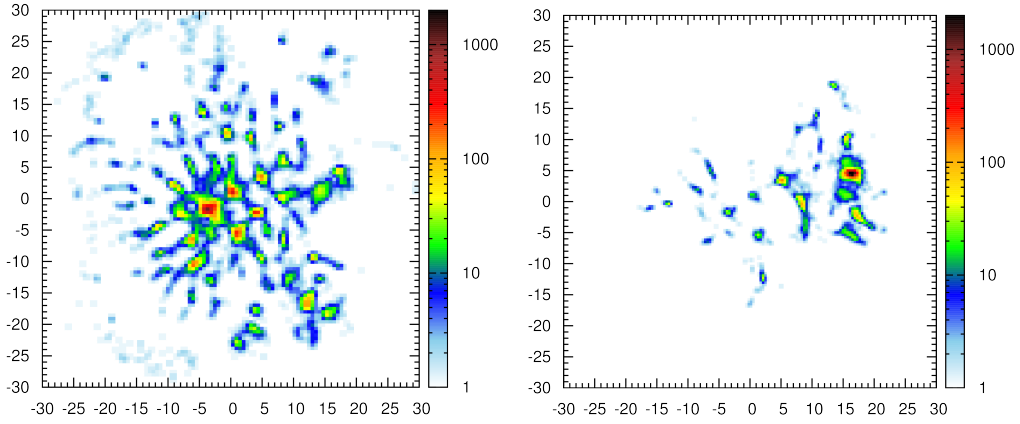


Figure 5.10: Two-dimensional sketch-map projections of 40,000 configurations from each of the four BP-REMD simulations. The clustering is based on the similarity of the peptide's backbone conformation and the various regions of the projection represent different types of structures of the ion–peptide complexes. The colour scheme reflects the absolute probability of data points per area.

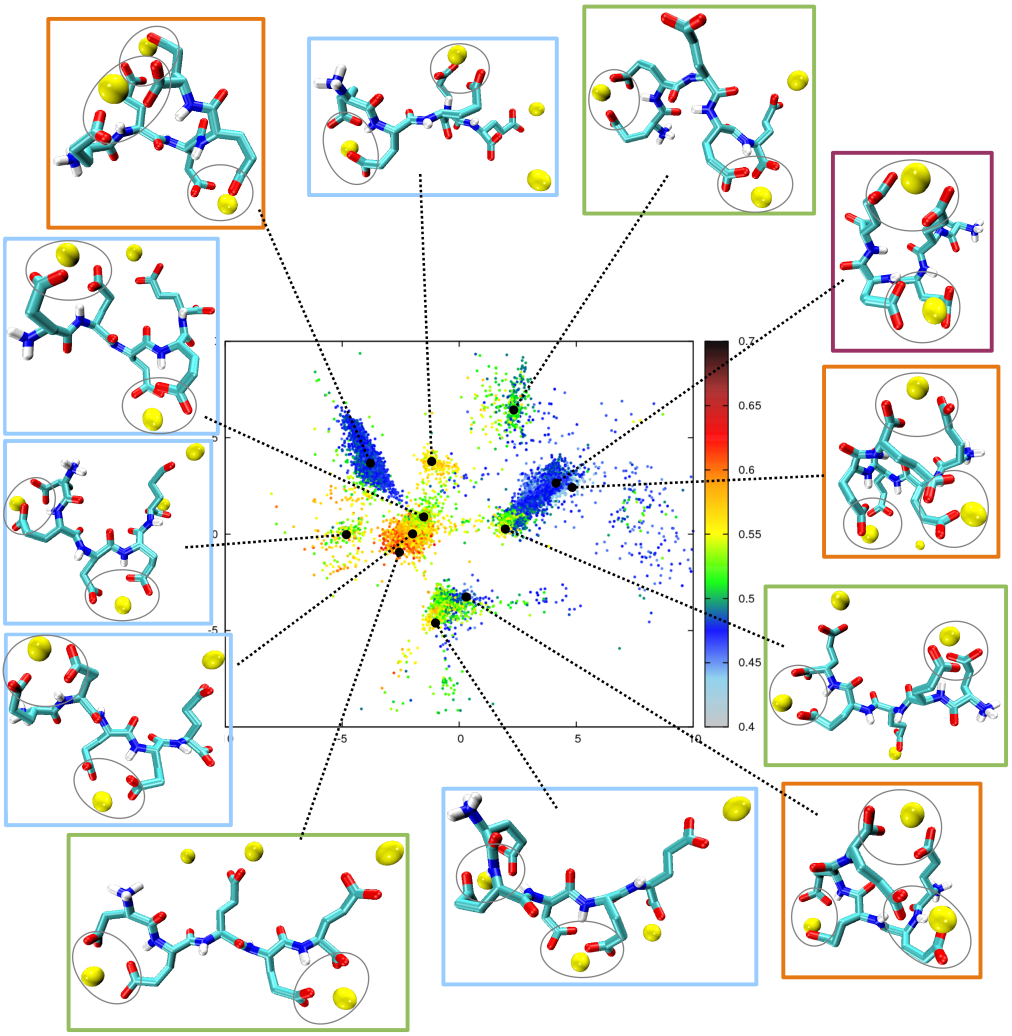


Figure 5.11: Two-dimensional sketch-map projections of 40,000 configurations from the BP-REMD simulation of pentaGLU with counter-ions. The colour scheme of the data points reflects the radius of gyration of pentaGLU (the scale on the right side shows the values of the radius of gyration in nanometers). Configurations of 12 different regions are shown. Based on the number and positions of salt-bridges (grey circles), these 12 configurations have been categorised into four different structures of ion–peptide complexes, which are labeled by the colour of their frames.

and between side chains 3 and 4 (the numbering of side chains starts at the N-terminus of the peptides). Further calcium ions frequently interact with one of the carboxylate groups of the other side chains or of the deprotonated C-terminus. The radius of gyration of this structure varies in the range of 0.52 to 0.57 nm. The green frames indicate elongated structures with a large radius of gyration (varying between 0.52 and 0.65 nm) with two salt-bridges. The first salt-bridge is formed between side chains 1 and 2, while the second salt-bridge connects the 4th side chain with the C-terminus. The third type of structure, which is highlighted by orange frames, is the most compact structure observed for pentaGLU. The globular structure can be ascribed to two salt-bridges which connect the two ends of the peptide. These salt-bridges are formed between the 1st side chain and the C-terminus and between the 2nd and 5th side chain. A third salt-bridge connects side chains 3 and 4. The radius of gyration of this complex structure varies between 0.43 and 0.47 nm. As can be seen from figure 5.10b, the fourth type of structure (violet frame colour) is the most frequently sampled structure of the respective BP-REMD simulation. It is very similar to the previously described structure (orange frames), but features only two salt-bridges between the 2nd and 5th and between the 3rd and 4th side chains. As the salt-bridge between the 1st side chain and the C-terminus is missing compared to the previous type of structure, it is slightly less compact with a radius of gyration of approximately 0.47 nm.

Each of these types of complex structures can be formed by different peptide backbone conformations. This is nicely illustrated by the sketch-map projection (fig. 5.11), which assigns configurations of one type of complex structure to different regions in the two-dimensional space according to their backbone conformation. Furthermore, this is indicated by the range of values of the radius of gyration of each type of structure.

The two-dimensional sketch-map projection of configurations (cf. figures 5.10b and 5.11) contains further valuable information on possible transitions between the various clusters. It is important to consider that the exchanges of configurations between the replicas during the BP-REMD simulations lead to discontinuous trajectories within one replica. Therefore, only a limited number of conformational changes is observed within the replica of interest. It is thus unclear, whether those data points, that lie in between two clusters in the 2D space, really correspond to configurations that were sampled during a conformational change in the replica of interest. Nevertheless, it is plausible that these configurations indicate the position of low-energy pathways between different conformations and that conformational changes really proceed along these pathways. Therefore, we will refer to them as transition pathways in the following. It is interesting to see that many of the

non-central clusters are not connected through direct transition pathways. Major conformational changes from one non-central cluster to another one (e.g. from one compact structure to another compact structure) apparently almost always have to pass through the extended conformation of the central cluster as a transition state.

The two-dimensional sketch-map projection of ion–peptide configurations consisting of decaGLU and calcium ions exhibits similar features as the one of pentaGLU–calcium configurations. As for the pentaGLU system, the sketch-map method projects the configurations onto the two-dimensional space such that meaningful, well-distinguishable clusters are created. This can be seen from figure 5.12, in which the data points are coloured according to the radius of gyration of the respective configuration. Naturally, the conformational phase space of decaGLU–calcium ion configurations is a lot more complex than the one of pentaGLU with calcium ions. Therefore, the number of clusters is larger and their distribution in the two-dimensional projection is a lot more complex than for pentaGLU. In figure 5.12 representative configurations of the most important clusters, that have been identified during the BP-REMD simulation, are shown. The clusters of compact conformations are again arranged around a central cluster of elongated conformations. Many clusters are connected *via* well-defined transition pathways. And clusters with similar features in terms of the number of saltbridges and the radius of gyration are positioned in different areas of the two-dimensional projection, which shows that they differ with respect to their peptide backbone conformations. As the conformational sampling during this simulation is probably not comprehensive, these conformations are most likely only a subset of the many stable structures that decaGLU forms in the presence of calcium ions. In contrast to pentaGLU, the conformations identified here can not be easily grouped into a smaller number of characteristic complex structures, as the number and positions of saltbridges are too diverse in all of the above conformations.

Considering the different types of structures of the ion–peptide complexes observed for pentaGLU and decaGLU, it is interesting to analyse the distances between the calcium ions of the complexes. As described above (section 5.1), it has been speculated in the literature that molecular modifiers in solution might collocate the constituents of a forming solid in a way that determines the amorphous or crystalline polymorph which is formed [97, 2]. Furthermore, recent results on the formation of "dynamically ordered liquid like oxyanion polymers" (DOLLOP) of calcium carbonate suggest that the ordering of ions in DOLLOP (a precursor of amorphous calcium carbonate)

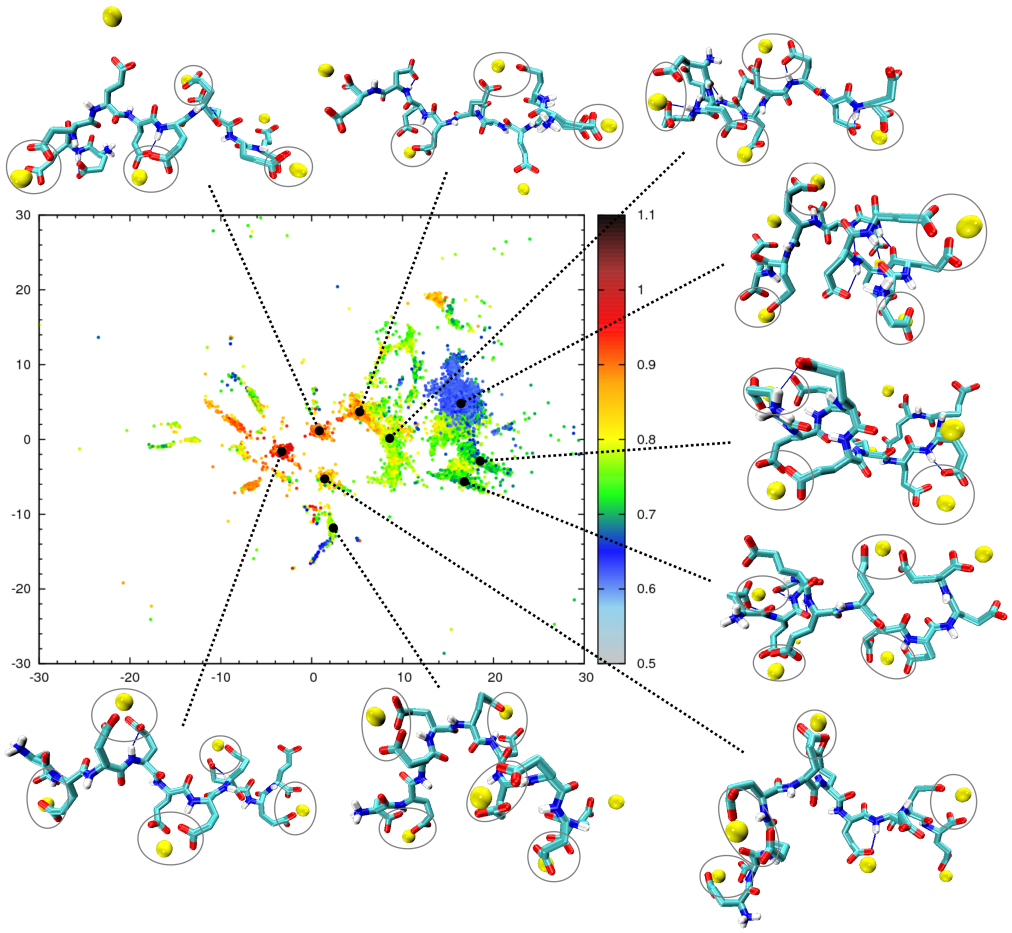


Figure 5.12: Two-dimensional sketch-map projections of 40,000 snapshots from the BP-REMD simulation of decaGLU with counter-ions. The colour scheme of the data points reflects the radius of gyration of decaGLU (the scale on the right side shows the values of the radius of gyration in nanometers). Snapshots of 9 different regions are shown, in which the positions of salt-bridges are indicated by grey circles.

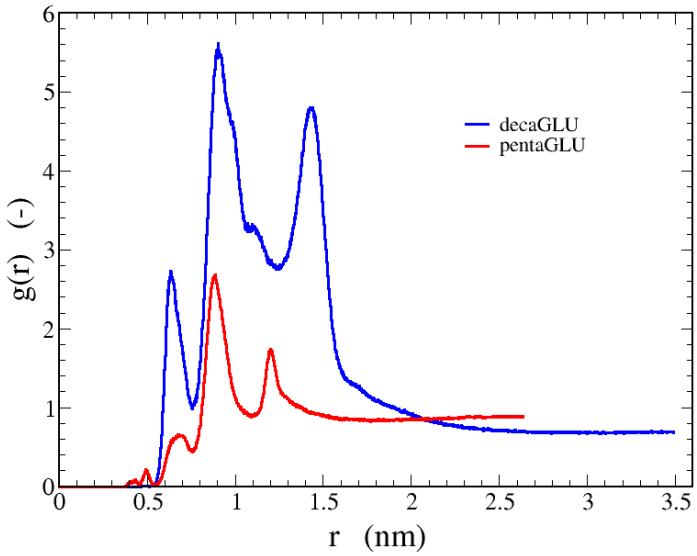


Figure 5.13: Radial distribution functions of calcium ions during the BP-REMD simulations of pentaGLU and decaGLU.

is characteristic for the type of polymorph into which DOLLOP later on transforms [76]. These findings suggest that polymorphism is modulated by highly dynamic nanoscale structures in solution and might thus be affected by ion–peptide complexes [191, 77, 192]. As the exact mechanisms, by which oligoglutamates manipulate the nucleation process of calcium oxalate, are still unknown, it is worth studying whether any characterstic distances between calcium ions can be identified for the different types of oligoGLU–calcium complexes. Figure 5.13 shows the radial distribution functions (RDF) of calcium ions during the BP-REMD simulations of pentaGLU and decaGLU.

The curves possess distinct peaks which can be assigned to the calcium ions "bound" to the oligopeptides and kept at very characteristic distances by the peptide conformations. The significant peaks are located at distances of approximately 0.68 nm, 0.88 nm, and 1.2 nm (pentaGLU) and at 0.63 nm, 0.9 nm, and 1.43 nm (decaGLU). While those peaks at shorter distances have similar positions for pentaGLU and decaGLU, the respective peaks at larger distances are very different in their position (1.2 nm *vs.* 1.43 nm). This shows that the range of "ordering" of calcium ions is larger for decaGLU due to the longer chain length, which might have implications for processes occuring prior to or during nucleation.

Figure 5.14 shows two pentaGLU structures that feature calcium distances corresponding to the peaks of the respective RDF. The calcium distance of approximately 0.68 nm is observed for example in representative

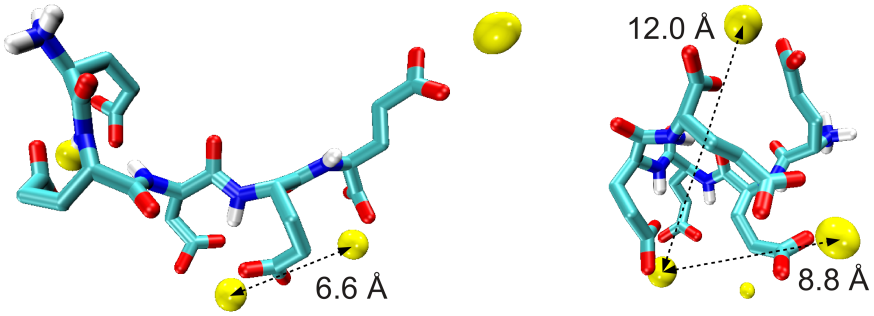


Figure 5.14: Two exemplary snapshots of pentaGLU structures from the BP-REMD simulation that feature calcium distances corresponding to the peaks of the respective RDF (cf. fig. 5.13).

snapshots of structure type 1 (blue frames in figure 5.11): it is the distance between one calcium ion in a saltbridge and another ion which is only loosely bound to one carboxylate group of a neighbouring side chain. This structure might be too unstable to direct the aggregation of any crystallising species. The calcium distances of 0.88 nm and 1.2 nm are both observed in snapshots of ion–peptide complexes of type 3 (orange frames in figure 5.11): both distances are found between calcium ions of two salt-bridges in a globular complex structure (fig. 5.14). The information depicted in figure 5.10b reveals that such structures were sampled frequently during the BP-REMD simulations, which indicates that these pentaGLU–calcium-ion complexes are relatively stable.

In order to compare these distances between the calcium ions bound to pentaGLU to those calcium distances in calcium oxalate crystals, two of the RDFs are plotted in figure 5.15. The RDF of the COT crystals shows several peaks of similar magnitude, of which the peak at 0.9 nm (grey dash-dotted line in fig. 5.15) coincides with the largest peak of the RDF of calcium ions bound to pentaGLU at a calcium distance of 0.88 nm. In figure 5.15, the positions of some of the most significant peaks of the RDF of the COD crystal (fig. 5.17) are indicated by the grey dashed lines. The positions of the peaks of the RDF of calcium ions in the vicinity of pentaGLU exhibit no similarity to the positions of peaks in the COD crystal (fig. 5.15).

The characteristic distances between calcium ions which are observed in the presence of decaGLU (fig. 5.13) can be found in many of the different structures that have been identified above. The complex structure depicted in figure 5.16 features three different distances between calcium ions which

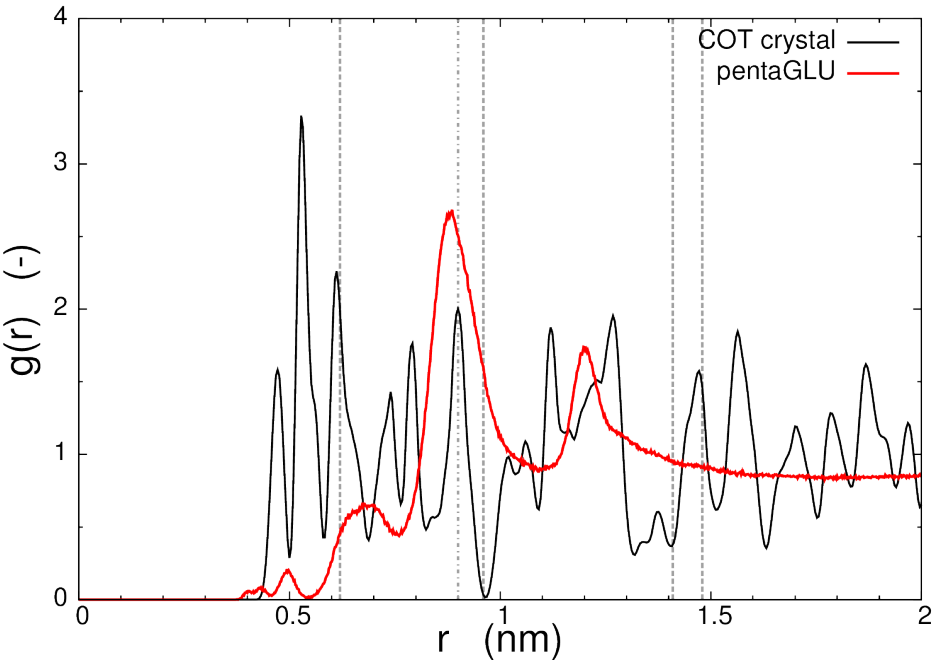


Figure 5.15: Radial distribution functions of calcium ions in calcium oxalate trihydrate (COT) crystals and of the BP-REMD simulation containing pentaGLU. The grey dash-dotted line (r = 0.9 nm) indicates the distance between calcium ions which can be found both in COT and in the complexes formed by pentaGLU and calcium ions in solution. The grey dashed lines denote the most significant peaks in the RDF of the calcium oxalate dihydrate (COD) cyrstal.

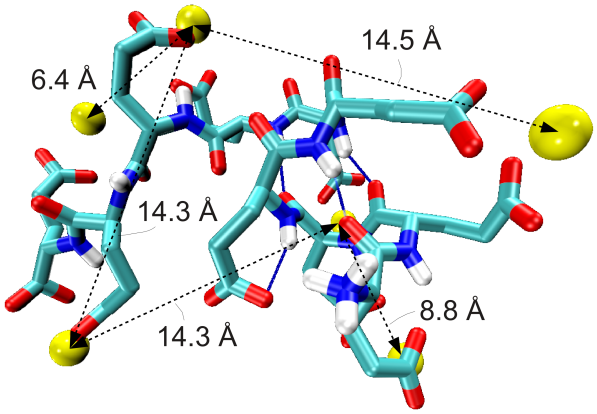


Figure 5.16: Exemplary snapshot of a decaGLU structure from the BP-REMD simulation that features calcium distances corresponding to the peaks of the respective RDF (cf. fig. 5.13).

correspond to the three largest peaks observed in figure 5.13.

The shortest distance (0.64 nm) is observed between a calcium ion that forms a salt-bridge and one ion which interacts with only one carboxylate group. All other distances occur between calcium ions that promote salt-bridges between carboxylate groups of various side chains. This structure features numerous salt-bridges and intramolecular hydrogen-bonds, and from the information of figure 5.10b it is known that this structure is the most frequently sampled structure during the BP-REMD simulation. Therefore, this structure is probably very stable. The snapshot of a decaGLU–calcium-ion complex (fig. 5.16) illustrates that even those calcium ions, which promote a salt-bridge between side chains of the peptide, are well-exposed to the solvent, which makes further interactions with anions such as oxalate or carbonate possible.

Comparing the positions of peaks of the RDF of calcium ions in the vicinity of decaGLU with the RDF of calcium ions in COD crystals (figure 5.17), several similarities become apparent. The pronounced peaks at 0.62, 0.96, 1.41, and 1.48 nm of the crystal RDF have some counterparts in the RDF of the decaGLU system (grey dashed lines in fig. 5.17), but not in the RDF of the pentaGLU system (grey dashed lines in fig. 5.15). It is interesting however, that the most pronounced peak of the RDF of the decaGLU system at 0.9 nm coincides with one of the peaks observed in COT (grey dash-dotted line in figures 5.15 and 5.17). At this point, it is important to recall the experimental observations, that COT forms in the presence of pentaGLU, while both COT and COD form when decaGLU is present in solution [2].

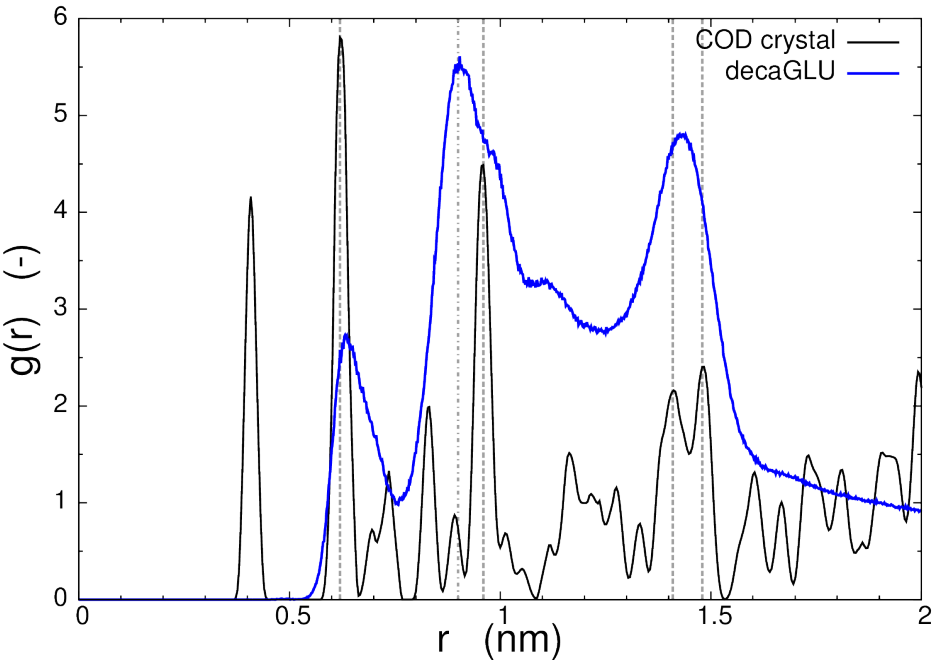


Figure 5.17: Radial distribution functions of calcium ions in calcium oxalate dihydrate (COD) crystals and of the BP-REMD simulation containing decaGLU. The grey dashed lines indicates the distances between calcium ions which can be found both in COD and in the complexes formed by decaGLU and calcium ions in solution. The grey dash-dotted line (r = 0.9 nm) denotes one of the significant peaks in the RDF of the calcium oxalate trihydrate (COT) cyrstal.

The observed similarities of the respective distances between calcium ions in calcium oxalate crystals and in oligoglutamate–calcium-ion complexes suggest that the accumulation of calcium ions around oligoglutamates might trigger the formation of ion clusters, which serve as precursors for the formation of COT or COD. The long-range order of calcium ions observed in the presence of decaGLU might be necessary for the formation of COD. As it is missing in pentaGLU–calcium-ion clusters due to the smaller size of the peptide, only precursors to COT might be formed.

For the system of calcium carbonate, it is well known that the formation of crystals results from several intermediate phases such as prenucleation clusters and amorphous precursors. Therefore, any structural motif encoding that happens at the stage of ion–peptide clusters would need to be transferred through the cascade of intermediate stages in order to affect the polymorphism of the resulting crystalline end product. For calcium carbonate, several studies suggest that this might be the case [6, 192]. Even though the exact mechanisms of the nucleation of calcium oxalate are still unknown (prenucleation structures such as PNCs or droplets of liquid precursors forming from liquid–liquid phase separation have not yet been detected for calcium oxalate), our results suggest that a similar structural motif encoding - introduced by oligopeptides - might also affect the formation of calcium oxalate crystals. This might explain some of the observed differences (COT *vs.* COD) in the crystallisation of calcium oxalate in the presence of oligoglutamates of various chain lengths [2].

5.4 Conclusions and Outlook

The experiments of Fischer et al. evidenced the preferences of oligoglutamates to adsorb to specific faces of calcium oxalate crystals [2]. However, these experiments were not able to provide detailed explanations for the observed changes in the phases of calcium oxalate crystals (COT *vs.* COD) precipitated from peptide solutions. The methods and results presented in this chapter outline a promising approach to study complexes of ions and oligopeptides in solution with the goal to identify processes prior to or during nucleation of biominerals which might affect the phase of the forming crystal.

The BP-REMD method was applied successfully in order to overcome the problems in conformational sampling, and some potentials to further optimise these simulations have been identified. Apart from those possibilities to further optimise the simulation set-up, the BP-REMD method has proven to be a powerful tool that significantly enhances the conformational sampling even

in systems with strongly interacting components. An interesting aspect of the BP-REMD method is the fact that it facilitates the detection of several different structures (which differ significantly in their backbone conformation) that exhibit similar properties with respect to the radius of gyration or the number and positions of salt-bridges. This is probably an advantage of the BP-REMD method compared to the sampling *via* metadynamics. The usage of the radius of gyration or the number of salt-bridges as collective variables during metadynamics simulations does not guarantee to find all of these different structures with similar radii of gyration and salt-bridges, which possess different backbone conformations. However, the BP-REMD method enforces changes in the backbone conformations such that similar structures with different backbone conformations can be detected. These differences in the backbone conformations might be important with respect to the interference of oligopeptides with processes of crystal nucleation and growth of calcium-containing minerals.

The presence of calcium ions poses considerable obstacles to the sampling of oligoglutamate conformations. However, the simulations presented in this chapter reveal that a completely comprehensive sampling of the conformational phase space might not always be necessary. Applying the sketch-map method to analyse the data, numerous significant structures of ion–peptide complexes that feature characteristic distances between calcium ions were found. These distances between calcium ions resemble some of those distances of calcium ions in calcium oxalate crystals. Especially the coincidence of some calcium distances in complexes of decaGLU and calcium ions with those in COD and COT crystals (fig. 5.17) might be an indication for a structural motif encoding, which is induced by the oligopeptides and which directs the crystallisation process of calcium oxalate towards the simultaneous formation of COT and COD, as it is observed during experiments. According to this hypothesis, the chain length of peptides plays an important role, as a certain chain length might be necessary in order to encode some specific structural patterns found in the different phases of calcium oxalate. This might explain the observed shift from COT to COD with an increase in the chain length of the oligoglutamates present during precipitation [2].

In order to gain further insight into the formation processes of prenucleation structures, it might be worth to analyse the two-dimensional results of the sketch-map method in more detail. The connections between the basins of the two-dimensional representation probably contain information on possible transition states between the various stable conformations of ion–peptide clusters. An interesting approach to intensify the sampling of the energy landscape even further, would be to use the position of a configuration in the two-dimensional sketch-map projection as collective variables

in a metadynamics simulation. This would probably significantly enhance the sampling of all relevant parts of the phase space and thus help to find further minima of the energy landscape.

CHAPTER 5. CONFORMATIONAL SAMPLING OF OLIGOGLUTAMATES IN CONTACT WITH CALCIUM IONS

Chapter 6

Conclusions and Outlook

Conclusions The earliest stages of biomineralisation are governed by the interactions of biomolecular modifiers with solvated ions and pre-nucleation structures. In order to gain insight into these processes at a molecular scale, the goal of this thesis was to study the complexation of single solvated ions (calcium) by oligopeptides (glutamate) in aqueous solution by means of Molecular Dynamics simulations and to analyse the resulting ion–peptide complexes.

The analysis of existing models (forcefields) regarding a proper description of the interactions of calcium and glutamate revealed the necessity to optimise these forcefields. Therefore, we developed a general strategy that is suitable to optimise established biomolecular FF with respect to the pairing of complex ions in solution. We found that small modifications in the van-der-Waals ion–ion interaction parameters are sufficient to obtain a realistic description of important aspects of ion pairing (even if complex ions such as charged peptide side chains are involved) based on existing well-established biomolecular forcefields such as GROMOS or OPLS-AA.

The comprehensive conformational sampling of oligoglutamates dissolved in an electrolyte solution can be computationally demanding due to the large number of degrees of freedom and the strong ion–peptide interactions, which - due to the formation of salt-bridges between peptide sidechains - further impede the conformational sampling of the peptide. The application of a biasing dihedral potential, that levels the torsional free energy barriers of the backbone dihedral angles of the free peptide (without ions), allowed us to successfully explore the vast phase space of calcium–oligoglutamate complexes. Another challenge was to analyse the complex conformational free-energy landscape, to identify suitable order parameters to compare the conformational sampling of different simulations, and to characterise representative structures of the ion–peptide complexes that may play a role during

nucleation processes. Here, the sketch map analysis helped to identify various stable conformations and it was found that the ion–peptide complexes exhibit characteristic distances between the different calcium ions. These distances depend on the chain length of the peptides and they resemble some of the distances between calcium ions of the different phases of calcium oxalate crystals, which have been formed in the presence of oligoglutamates. This analogy might be an indication for structural motif encoding and might thus explain the peptides' ability to manipulate the structure and phase of forming calcium oxalate crystals (COT *vs.* COD) as it was observed in the experiments of Fischer et al. [2].

Outlook It is important to point out that our findings and the speculations derived from these results do not deliver any information on the exact mechanisms of nucleation of calcium-containing minerals such as calcium oxalate. However, the methods and results presented here disclose a promising approach to investigate the phenomena occurring prior to and during nucleation. Numerous interesting systems might be studied in the future with similar methods: Given that reliable models for anions such as oxalate or carbonate - which fit to the optimised models for oligoglutamate–calcium-ion interactions of the present work - are identified, solutions of ions and peptides can be simulated in order to study the aggregation of ion-clusters in the vicinity of the oligoglutamates. These simulations might help to study non-classical prenucleation species such as pre-nucleation clusters (PNCs), and might furthermore clarify if modifiers such as oligopeptides are capable of inducing structural motifs in precursors of amorphous or crystalline polymorphs of minerals.

In order to further characterise the influence of oligopeptides on the nucleation process, it is worth including an icosamer of glutamate (GLU20), which was also considered in the experiments of Fischer et al. [2], in future studies. Furthermore, the structures of ion–peptide complexes are likely to depend significantly on the flexibility of the side chains. For example, the complexes formed by oligoaspartate peptides and calcium ions might exhibit different patterns of calcium distances (compared to those of oligoglutamates) due to the different length of the side chains, even though the functional groups interacting with calcium are identical. Therefore, it would be interesting to repeat the simulations and analyses presented in chapter 5 of this thesis with oligoaspartates and to compare these results with those of the simulations containing oligoglutamates.

In the literature it has been speculated, that supramolecular assemblies of modifiers rather than single molecules of the modifier might be respon-

sible for some of the manipulative effects that these molecules exert on the processes occurring prior to or during nucleation [6, 191]. Therefore, future investigations might be complemented by simulations of varying concentrations of oligopeptides.

Since these possibilities to extend the explicit-solvent MD simulations will be subject to even more severe sampling problems than the systems presented here, it is also possible to develop a solvent-free model based on the structures identified in our simulations. This would allow for studies of larger systems and longer time scales, which is probably necessary in order to fully understand the processes occurring prior to or during nucleation. The results presented in this chapter might also help to generate coarse-grained models of the constituents of the system, which could be used to the study the interaction of oligopeptides with crystal surfaces. A lot of research has already been devoted to study these interactions [199, 68, 71, 200, 53, 65, 66, 201] and experiments have shown that peptides of the size of pentaGLU are able to stabilise certain crystal faces which are not stable in the absence of peptides [2]. Simulations might help to further analyse, whether chemical recognition at specific binding sites is the origin of such preferences in the binding to individual crystal faces. It would be interesting to study, whether any correlation can be found between the distances of calcium ions exposed at the different crystal surfaces and the conformations that oligoglutamates can adopt on top of these surfaces.

This thesis can be viewed as a basis for such future studies, as several important forcefield and methodological challenges have been solved. This will help to conduct future simulations which may provide valuable mechanistic insight into the molecular processes that steer peptide-assisted nucleation and growth of calcium-based minerals.

Bibliography

[1] R.-Q. Song and H. Cölfen. Additive controlled crystallization. *CrystEngComm*, 13(5):1249, 2011. doi: 10.1039/c0ce00419g.

[2] V. Fischer, K. Landfester, and R. Muñoz Espí. Stabilization of Calcium Oxalate Metastable Phases by Oligo (L-glutamic acid): Effect of Peptide Chain Length. *Crystal Growth & Design*, 11:1880–1890, 2011.

[3] J. J. De Yoreo and P. G. Vekilov. Principles of crystal nucleation and growth. *Reviews in Mineralogy and Geochemisty*, 54:57–93, 2003. doi: 10.2113/0540057.

[4] E. Project, E. Nachliel, and M. Gutman. Parameterization of Ca+2 protein interactions for molecular dynamics simulations. *Journal of Computational Chemistry*, 29:1163–1169, 2008. doi: 10.1002/jcc.

[5] E. Project. *Elucidating the Ca^{2+} Binding to Calmodulin Family of Proteins by Molecular Dynamics Simulations*. PhD thesis, Tel Aviv University, 2010.

[6] P. Raiteri, R. Demichelis, J. D. Gale, M. Kellermeier, D. Gebauer, D. Quigley, L. B. Wright, and T. R. Walsh. Exploring the influence of organic species on pre- and post-nucleation calcium carbonate. *Faraday Discussions*, 159:61, 2012. doi: 10.1039/c2fd20052j.

[7] P. G. Daniele, C. Foti, A. Gianguzza, E. Prenesti, and S. Sammartano. Weak alkali and alkaline earth metal complexes of low molecular weight ligands in aqueous solution. *Coordination Chemistry Reviews*, 252(10-11):1093–1107, May 2008. doi: 10.1016/j.ccr.2007.08.005.

[8] J. Kahlen, L. Salimi, M. Sulpizi, C. Peter, and D. Donadio. Interaction of charged amino-acid side chains with ions: An optimization strategy for classical forcefields. *The Journal of Physical Chemistry B*, 2014. doi: 10.1021/jp412490c.

[9] S. Mann. *Biomineralization*. Oxford University Press, New York, 2001. ISBN 0-19-850882-4.

[10] H. Cölfen and S. Mann. Higher-order organization by mesoscale self-assembly and transformation of hybrid nanostructures. *Angewandte Chemie International Edition*, 42(21):2350–2365, 2003. doi: 10.1002/anie.200200562.

[11] Y. Levi, S. Albeck, A. Brack, S. Weiner, and L. Addadi. Control over aragonite crystal nucleation and growth: An in vitro study of biomineralization. *Chemistry A European Journal*, 4(3):389–396, 1998.

[12] L. Addadi and S. Weiner. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. *Proceedings of the National Academy of Sciences*, 82(12):4110–4114, 1985.

[13] S.-H. Yu and H. Cölfen. Bio-inspired crystal morphogenesis by hydrophilic polymers. *Journal of Materials Chemistry*, 14(14):2124, 2004. doi: 10.1039/b401420k.

[14] L. M. Hamm, A. F. Wallace, and P. M. Dove. Molecular dynamics of ion hydration in the presence of small carboxylated molecules and implications for calcification. *The Journal of Physical Chemistry B*, 114(32):10488–95, August 2010.

[15] J. Aizenberg, L. Addadi, S. Weiner, and G. Lambert. Stabilization of amorphous calcium carbonate by specialized macromolecules in biological and synthetic precipitates. *Advanced Materials*, 8(3):222–226, 1996. doi: 10.1002/adma.19960080307.

[16] L. Addadi, S. Raz, and S. Weiner. Taking advantage of disorder: Amorphous calcium carbonate and its roles in biomineralization. *Advanced Materials*, 15(12):959–970, 2003. doi: 10.1002/adma.200300381.

[17] A. Berman, J. Hanson, L. Leiserowitz, T. F. Koetzle, S. Weiner, and L. Addadi. Biological control of crystal texture: a widespread strategy for adapting crystal properties to function. *Science*, 259(5096):776–779, 1993.

[18] J. Aizenberg, J. Hanson, M. Ilan, L. Leiserowitz, T. Koetzle, L. Addadi, and S. Weiner. Morphogenesis of calcitic sponge spicules: a role for specialized proteins interacting with growing crystals. *The FASEB Journal*, 9(2):262–268, 1995.

[19] F. C. Meldrum and H. Cölfen. Controlling mineral morphologies and structures in biological and synthetic systems. *Chemical reviews*, 108 (11):4332–432, November 2008. doi: 10.1021/cr8002856.

[20] S. Mann, J. M. Didymus, N. P. Sanderson, B. R. Heywood, and E. J. A. Samper. Morphological influence of functionalized and non-functionalized α,ω-dicarboxylates on calcite crystallization. *J. Chem. Soc., Faraday Trans.*, 86:1873–1880, 1990. doi: 10.1039/FT9908601873.

[21] C. Orme, A. Noy, A. Wierzbicki, M. T. McBride, M. Grantham, H. Teng, P. Dove, and J. DeYoreo. Formation of chiral morphologies through selective binding of amino acids to calcite surface steps. *Nature*, 411(6839):775–779, 2001.

[22] S. Elhadj, E. A. Salter, A. Wierzbicki, J. J. De Yoreo, N. Han, and P. M. Dove. Peptide controls on calcite mineralization: polyaspartate chain length affects growth kinetics and acts as a stereochemical switch on morphology. *Crystal Growth Design*, 6(1):197–201, 2006. doi: 10.1021/cg050288+.

[23] L. Wang, S. R. Qiu, W. Zachowicz, X. Guan, J. J. DeYoreo, G. H. Nancollas, and J. R. Hoyer. Modulation of calcium oxalate crystallization by linear aspartic acid-rich peptides. *Langmuir*, 22(17):7279–7285, 2006. doi: 10.1021/la060897z.

[24] J. J. De Yoreo, A. Wierzbicki, and P. M. Dove. New insights into mechanisms of biomolecular control on growth of inorganic crystals. *CrystEngComm*, 9(12):1144–1152, 2007. doi: 10.1039/b713006f.

[25] K. Gorna, R. Muñoz Espí, F. Gröhn, and G. Wegner. Bioinspired mineralization of inorganics from aqueous media controlled by synthetic polymers. *Macromolecular bioscience*, 7(2):163–73, February 2007. doi: 10.1002/mabi.200600209.

[26] J. Aizenberg, A. Tkachenko, S. Weiner, L. Addadi, and G. Hendler. Calcitic microlenses as part of the photoreceptor system in brittlestars. *Nature*, 412(6849):819–822, 2001.

[27] J. Aizenberg, G. Lambert, S. Weiner, and L. Addadi. Factors involved in the formation of amorphous and crystalline calcium carbonate: a study of an ascidian skeleton. *Journal of the American Chemical Society*, 124(1):32–39, 2002. doi: 10.1021/ja016990l.

[28] L. Addadi, D. Joester, F. Nudelman, and S. Weiner. Mollusk shell formation: A source of new concepts for understanding biomineralization processes. *Chemistry A European Journal*, 12(4):980–987, 2006. doi: 10.1002/chem.200500980.

[29] K. Delak, S. Collino, and J. S. Evans. Expected and unexpected effects of amino acid substitutions on polypeptide-directed crystal growth. *Langmuir*, 23(24):11951–11955, 2007. doi: 10.1021/la702113x.

[30] R. A. Metzler, I. W. Kim, K. Delak, J. S. Evans, D. Zhou, E. Beniash, F. Wilt, M. Abrecht, J.-W. Chiou, J. Guo, S. N. Coppersmith, and P. U. P. A. Gilbert. Probing the organicmineral interface at the molecular level in model biominerals. *Langmuir*, 24(6):2680–2687, 2008. doi: 10.1021/la7031237.

[31] R. A. Metzler, G. A. Tribello, M. Parrinello, and P. U. P. A. Gilbert. Asprich peptides are occluded in calcite and permanently disorder biomineral crystals. *Journal of the American Chemical Society*, 132 (33):11585–91, August 2010. doi: 10.1021/ja103089r.

[32] J. Rieger, T. Frechen, G. Cox, W. Heckmann, C. Schmidt, and J. Thieme. Precursor structures in the crystallization/precipitation processes of $CaCO_3$ and control of particle formation by polyelectrolytes. *Faraday Discussions*, 136:265–277, 2007. doi: 10.1039/B701450C.

[33] T. Wang, H. Cölfen, and M. Antonietti. Nonclassical Crystallization: Mesocrystals and Morphology Change of $CaCO_3$ Crystals in the Presence of a Polyelectrolyte Additive. *Journal of the American Chemical Society*, 127(10):3246–3247, 2005. doi: 10.1021/ja045331g.

[34] J.-W. Shen, C. Li, N. F. A. van der Vegt, and C. Peter. Understanding the Control of Mineralization by Polyelectrolyte Additives : Simulation of Preferential Binding to Calcite Surfaces. *The Journal of Physical Chemistry C*, 117:6904–6913, 2013. doi: 10.1021/jp402341w.

[35] K. Bleek and A. Taubert. New developments in polymer-controlled, bioinspired calcium phosphate mineralization from aqueous solution. *Acta biomaterialia*, 9(5):6283–321, May 2013. doi: 10.1016/j.actbio.2012.12.027.

[36] S. Dorozhkin. Bioceramics based on calcium orthophosphates (review). *Glass and Ceramics*, 64(11-12):442–447, 2007. doi: 10.1007/s10717-007-0109-1.

[37] S. V. Dorozhkin. Nanodimensional and nanocrystalline apatites and other calcium orthophosphates in biomedical engineering, biology and medicine. *Materials*, 2(4):1975–2045, 2009. doi: 10.3390/ma2041975.

[38] M. Epple, K. Ganesan, R. Heumann, J. Klesing, A. Kovtun, S. Neumann, and V. Sokolova. Application of calcium phosphate nanoparticles in biomedicine. *J. Mater. Chem.*, 20:18–23, 2010. doi: 10.1039/B910885H.

[39] L. Wang and G. H. Nancollas. Calcium orthophosphates: Crystallization and dissolution. *Chemical Reviews*, 108(11):4628–4669, 2008. doi: 10.1021/cr0782574.

[40] L. Wang and G. H. Nancollas. Pathways to biomineralization and biodemineralization of calcium phosphates: the thermodynamic and kinetic controls. *Dalton Transactions*, pages 2665–2672, 2009. doi: 10.1039/B815887H.

[41] M. D. Sikiric and H. Furedi-Milhofer. The influence of surface active molecules on the crystallization of biominerals in solution. *Advances in Colloid and Interface Science*, 128-130:135–158, 2006.

[42] B. Palazzo, D. Walsh, M. Iafisco, E. Foresti, L. Bertinetti, G. Martra, C. L. Bianchi, G. Cappelletti, and N. Roveri. Amino acid synergetic effect on structure, morphology and surface properties of biomimetic apatite nanocrystals. *Acta Biomaterialia*, 5(4):1241 – 1252, 2009. doi: http://dx.doi.org/10.1016/j.actbio.2008.10.024.

[43] X. Yang, B. Xie, L. Wang, Y. Qin, Z. J. Henneman, and G. H. Nancollas. Influence of magnesium ions and amino acids on the nucleation and growth of hydroxyapatite. *CrystEngComm*, 13:1153–1158, 2011. doi: 10.1039/C0CE00470G.

[44] V. R. Franceschi and P. A. Nakata. Calcium oxalate in plants: Formation and function. *Annual Review of Plant Biology*, 56(1):41–71, 2005. doi: 10.1146/annurev.arplant.56.032604.144106.

[45] M. A. Webb. Cell-mediated crystallization of calcium oxalate in plants. *The Plant cell*, 11(4):751–61, 1999.

[46] S. R. Qiu, A. Wierzbicki, E. A. Salter, S. Zepeda, C. a. Orme, J. R. Hoyer, G. H. Nancollas, A. M. Cody, and J. J. De Yoreo. Modulation of calcium oxalate monohydrate crystallization by citrate through

selective binding to atomic steps. *Journal of the American Chemical Society*, 127(25):9036–44, June 2005. doi: 10.1021/ja043591s.

[47] L. Brecevic and D. Kralj. The influence of some amino acids on calcium oxalate dihydrate transformation. *Journal of Crystal Growth*, 79(13): 178 – 184, 1986. doi: http://dx.doi.org/10.1016/0022-0248(86)90433-1.

[48] R. W. Friddle, M. L. Weaver, S. R. Qiu, A. Wierzbicki, W. H. Casey, and J. J. De Yoreo. Subnanometer atomic force microscopy of peptide-mineral interactions links clustering and competition to acceleration and catastrophe. *Proceedings of the National Academy of Sciences of the United States of America*, 107(1):11–15, January 2010. doi: 10.1073/pnas.0908205107.

[49] N. Garti, F. Tibika, S. Sarig, and S. Perlberg. The inhibitory effect of polymeric carboxylic amino-acids and urine on calcium oxalate crystallization. *Biochemical and Biophysical Research Communications*, 97 (3):1154 – 1162, 1980. doi: http://dx.doi.org/10.1016/0006-291X(80)91496-5.

[50] B. Fellstrom, B. G. Danielson, S. Ljunghall, and B. Wikstrom. Crystal inhibition: the effects of polyanions on calcium oxalate crystal growth. *Clinica Chimica Acta*, 158(3):229 – 235, 1986. doi: http://dx.doi.org/10.1016/0009-8981(86)90286-X.

[51] A. Taller, B. Grohe, K. A. Rogers, H. A. Goldberg, and G. K. Hunter. Specific adsorption of osteopontin and synthetic polypeptides to calcium oxalate monohydrate crystals. *Biophysical journal*, 93(5):1768–77, September 2007. doi: 10.1529/biophysj.106.101881.

[52] B. Grohe, J. O'Young, D. A. Ionescu, G. Lajoie, K. A. Rogers, M. Karttunen, H. A. Goldberg, and G. K. Hunter. Control of calcium oxalate crystal growth by face-specific adsorption of an osteopontin phosphopeptide. *Journal of the American Chemical Society*, 129(48):14946–51, December 2007. doi: 10.1021/ja0745613.

[53] L. Wang, X. Guan, R. Tang, J. R. Hoyer, A. Wierzbicki, J. J. De Yoreo, and G. H. Nancollas. Phosphorylation of osteopontin is required for inhibition of calcium oxalate crystallization. *The Journal of Physical Chemistry B*, 112(30):9151–7, July 2008. doi: 10.1021/jp804282u.

[54] W.-K. Burton, N. Cabrera, and F. Frank. The growth of crystals and the equilibrium structure of their surfaces. *Philosophical Transactions*

of the Royal Society of London. Series A, Mathematical and Physical Sciences, pages 299–358, 1951.

[55] A. Chernov. The spiral growth of crystals. *Soviet Physics-Uspekhi*, 4 (1):116–148, 1961.

[56] N. Cabrera and D. Vermilyea. The growth of crystals from solution. In T. D. Doremus RH, Roberts BW, editor, *Growth and perfection of crystals*, pages 393–410. Wiley, New York, 1958.

[57] W. van Enckevort and A. van den Berg. Impurity blocking of crystal growth: a monte carlo study. *Journal of Crystal Growth*, 183(3):441 – 455, 1998. doi: http://dx.doi.org/10.1016/S0022-0248(97)00432-6.

[58] K. J. Davis, P. M. Dove, and J. J. De Yoreo. The Role of Mg^{2+} as an Impurity in Calcite Growth. *Science*, 290(5494):1134–1137, 2000. doi: 10.1126/science.290.5494.1134.

[59] J. Astilleros, C. Pina, L. Fernndez-daz, and A. Putnis. Molecular-scale surface processes during the growth of calcite in the presence of manganese. *Geochimica et Cosmochimica Acta*, 66(18):3177 – 3189, 2002. doi: http://dx.doi.org/10.1016/S0016-7037(02)00892-X.

[60] J. O'Young, S. Chirico, N. Al Tarhuni, B. Grohe, M. Karttunen, H. A. Goldberg, and G. K. Hunter. Phosphorylation of Osteopontin Peptides Mediates Adsorption to and Incorporation into Calcium Oxalate Crystals. *Cells, tissues, organs*, 189(1-4):51–5, January 2009. doi: 10.1159/000151724.

[61] J. H. Harding, D. M. Duffy, M. L. Sushko, P. M. Rodger, D. Quigley, and J. A. Elliott. Computational techniques at the organic-inorganic interface in biomineralization. *Chemical reviews*, 108(11):4823–54, November 2008. doi: 10.1021/cr078278y.

[62] J. H. Harding and D. M. Duffy. The challenge of biominerals to simulations. *Journal of Materials Chemistry*, 16(12):1105–1112, 2006. doi: 10.1039/b511650c.

[63] R. A. Latour. Molecular simulation of protein-surface interactions: Benefits, problems, solutions, and future directions (Review). *Biointerphases*, 3(3):FC2, 2008. doi: 10.1116/1.2965132.

[64] A. Langdon, G. R. Wignall, K. Rogers, E. S. Sø rensen, J. Denstedt, B. Grohe, H. A. Goldberg, and G. K. Hunter. Kinetics of calcium oxalate crystal growth in the presence of osteopontin isoforms: an analysis by scanning confocal interference microcopy. *Calcified Tissue International*, 84(3):240–8, March 2009. doi: 10.1007/s00223-008-9215-5.

[65] P. V. Azzopardi, J. O'Young, G. Lajoie, M. Karttunen, H. A. Goldberg, and G. K. Hunter. Roles of electrostatics and conformation in protein-crystal interactions. *PloS one*, 5(2):e9330, January 2010. doi: 10.1371/journal.pone.0009330.

[66] S. Hug, G. K. Hunter, H. Goldberg, and M. Karttunen. Ab initio simulations of peptide-mineral interactions. *Physics Procedia*, 4(2009): 51–60, January 2010. doi: 10.1016/j.phpro.2010.08.008.

[67] A. Thomas, E. Rosseeva, O. Hochrein, W. Carrillo-Cabrera, P. Simon, P. Duchstein, D. Zahn, and R. Kniep. Mimicking the growth of a pathologic biomineral: shape development and structures of calcium oxalate dihydrate in the presence of polyacrylic acid. *Chemistry - A European Journal*, 18(13):4000–9, March 2012. doi: 10.1002/chem.201102228.

[68] X. Sheng, M. D. Ward, and J. A. Wesson. Adhesion between molecules and calcium oxalate crystals: critical interactions in kidney stone formation. *Journal of the American Chemical Society*, 125(10):2854–5, March 2003. doi: 10.1021/ja029575h.

[69] X. Sheng, T. Jung, J. A. Wesson, and M. D. Ward. Adhesion at calcium oxalate crystal surfaces and the effect of urinary constituents. *Proceedings of the National Academy of Sciences of the United States of America*, 102(2):267–72, January 2005. doi: 10.1073/pnas.0406835101.

[70] T. Jung, X. Sheng, C. K. Choi, W.-S. Kim, J. a. Wesson, and M. D. Ward. Probing crystallization of calcium oxalate monohydrate and the role of macromolecule additives with in situ atomic force microscopy. *Langmuir*, 20(20):8587–96, September 2004. doi: 10.1021/la0488755.

[71] S. R. Qiu, A. Wierzbicki, C. A. Orme, A. M. Cody, J. R. Hoyer, G. H. Nancollas, S. Zepeda, and J. J. De Yoreo. Molecular modulation of calcium oxalate crystallization by osteopontin and citrate. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7):1811–5, February 2004. doi: 10.1073/pnas.0307900100.

[72] S. Kerisit and S. C. Parker. Free energy of adsorption of water and metal ions on the 1014 calcite surface. *Journal of the American Chemical Society*, 126(32):10152–10161, 2004. doi: 10.1021/ja0487776.

[73] S. Kerisit, D. J. Cooke, D. Spagnoli, and S. C. Parker. Molecular dynamics simulations of the interactions between water and inorganic solids. *J. Mater. Chem.*, 15:1454–1462, 2005. doi: 10.1039/B415633C.

[74] S. Piana, F. Jones, and J. D. Gale. Aspartic acid as a crystal growth catalyst. *CrystEngComm*, 9:1187–1191, 2007. doi: 10.1039/B708260F.

[75] P. Raiteri and J. D. Gale. Water Is the Key to Nonclassical Nucleation of Amorphous Calcium Carbonate. *Journal of the American Chemical Society*, 132:17623–17634, November 2010. doi: 10.1021/ja108508k.

[76] D. Gebauer, M. Kellermeier, J. D. Gale, L. Bergstrom, and H. Cölfen. Pre-nucleation clusters as solute precursors in crystallisation. *Chemical Society Reviews*, 2014. doi: 10.1039/C3CS60451A.

[77] M. A. Bewernitz, D. Gebauer, J. Long, H. Cölfen, and L. B. Gower. A metastable liquid precursor phase of calcium carbonate and its interactions with polyaspartate. *Faraday Discussions*, 159:291–312, 2012. doi: 10.1039/C2FD20080E.

[78] L. B. Gower. Biomimetic model systems for investigating the amorphous precursor pathway and its role in biomineralization. *Chemical Reviews*, 108(11):4551–4627, 2008. doi: 10.1021/cr800443h.

[79] D. Gebauer, A. Völkel, and H. Cölfen. Stable prenucleation calcium carbonate clusters. *Science*, 322(5909):1819–1822, 2008. doi: 10.1126/science.1164271.

[80] R. Demichelis, P. Raiteri, J. D. Gale, D. Quigley, and D. Gebauer. Stable prenucleation mineral clusters are liquid-like ionic polymers. *Nature communications*, 2:590, 2011. doi: 10.1038/ncomms1604.

[81] A. F. Wallace, L. O. Hedges, A. Fernandez-Martinez, P. Raiteri, J. D. Gale, G. A. Waychunas, S. Whitelam, J. F. Banfield, and J. J. De Yoreo. Microscopic Evidence for Liquid-Liquid Separation in Supersaturated $CaCO_3$ Solutions. *Science*, 341(6148):885–889, 2013. doi: 10.1126/science.1230915.

[82] M. Faatz, F. Gröhn, and G. Wegner. Amorphous calcium carbonate: Synthesis and potential intermediate in biomineralization. *Advanced Materials*, 16(12):996–1000, 2004. doi: 10.1002/adma.200306565.

[83] S. E. Wolf, J. Leiterer, M. Kappl, F. Emmerling, and W. Tremel. Early homogenous amorphous precursor stages of calcium carbonate and subsequent crystal growth in levitated droplets. *Journal of the American Chemical Society*, 130(37):12342–12347, 2008. doi: 10.1021/ja800984y.

[84] E. Beniash, L. Addadi, and S. Weiner. Cellular control over spicule formation in sea urchin embryos: A structural approach. *Journal of Structural Biology*, 125(1):50 – 62, 1999. doi: http://dx.doi.org/10.1006/jsbi.1998.4081.

[85] I. M. Weiss, N. Tuross, L. Addadi, and S. Weiner. Mollusc larval shell formation: amorphous calcium carbonate is a precursor phase for aragonite. *Journal of Experimental Zoology*, 293(5):478–491, 2002. doi: 10.1002/jez.90004.

[86] Y. Politi, T. Arad, E. Klein, S. Weiner, and L. Addadi. Sea urchin spine calcite forms via a transient amorphous calcium carbonate phase. *Science*, 306(5699):1161–1164, 2004. doi: 10.1126/science.1102289.

[87] A. V. Radha, T. Z. Forbes, C. E. Killian, P. U. P. A. Gilbert, and A. Navrotsky. Transformation and crystallization energetics of synthetic and biogenic amorphous calcium carbonate. *Proceedings of the National Academy of Sciences*, 107(38):16438–16443, 2010. doi: 10.1073/pnas.1009959107.

[88] R. S. K. Lam, J. M. Charnock, A. Lennie, and F. C. Meldrum. Synthesis-dependant structural variations in amorphous calcium carbonate. *CrystEngComm*, 9:1226–1236, 2007. doi: 10.1039/B710895H.

[89] D. Gebauer, P. N. Gunawidjaja, J. Y. P. Ko, Z. Bacsik, B. Aziz, L. Liu, Y. Hu, L. Bergström, C.-W. Tai, T.-K. Sham, M. Eden, and N. Hedin. Proto-calcite and proto-vaterite in amorphous calcium carbonates. *Angewandte Chemie International Edition*, 49(47):8889–8891, 2010. doi: 10.1002/anie.201003220.

[90] W. Ostwald. Studies upon the forming and changing solid bodies. *Zeitschrift für Physikalische Chemie*, 22:289–330, 1897.

[91] S. E. Wolf, L. Müller, R. Barrea, C. J. Kampf, J. Leiterer, U. Panne, T. Hoffmann, F. Emmerling, and W. Tremel. Carbonate-coordinated metal complexes precede the formation of liquid amorphous mineral emulsions of divalent metal carbonates. *Nanoscale*, 3:1158–1165, 2011. doi: 10.1039/C0NR00761G.

[92] L. B. Gower and D. J. Odom. Deposition of calcium carbonate films by a polymer-induced liquid-precursor (PILP) process. *Journal of Crystal Growth*, 210(4):719–734, 2000. doi: http://dx.doi.org/10.1016/S0022-0248(99)00749-6.

[93] M. Kellermeier, D. Gebauer, E. Melero-Garca, M. Drechsler, Y. Talmon, L. Kienle, H. Cölfen, J. M. Garcia-Ruiz, and W. Kunz. Colloidal stabilization of calcium carbonate prenucleation clusters with silica. *Advanced Functional Materials*, 22(20):4301–4311, 2012. doi: 10.1002/adfm.201200953.

[94] D. Gebauer, H. Cölfen, A. Verch, and M. Antonietti. The Multiple Roles of Additives in $CaCO_3$ Crystallization: A Quantitative Case Study. *Advanced Materials*, 21(4):435–439, 2009. doi: 10.1002/adma.200801614.

[95] B. Guillemet, M. Faatz, F. Gröhn, G. Wegner, and Y. Gnanou. Nanosized amorphous calcium carbonate stabilized by poly(ethylene oxide)-b-poly(acrylic acid) block copolymers. *Langmuir*, 22(4):1875–1879, 2006. doi: 10.1021/la052419e.

[96] A. Frey-Wyssling. Crystallography of the two hydrates of crystalline calcium oxalate in plants. *American Journal of Botany*, pages 130–141, 1981.

[97] R. E. Bulo, D. Donadio, A. Laio, F. Molnar, J. Rieger, and M. Parrinello. "Site Binding" of Ca^{2+} Ions to Polyacrylates in Water: A Molecular Dynamics Study of Coiling and Aggregation. *Macromolecules*, 40:3437–3442, 2007.

[98] W. F. van Gunsteren and H. J. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English*, 29 (9):992–1023, 1990.

[99] L. Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159: 98–103, Jul 1967. doi: 10.1103/PhysRev.159.98.

[100] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, 1982. doi: http://dx.doi.org/10.1063/1.442716.

[101] M. Tuckerman. *Statistical Mechanics and Molecular Simulations*. Oxford University Press, 2008.

[102] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, January 2007. doi: 10.1063/1.2408420.

[103] J. Hermans, H. J. C. Berendsen, W. F. Van Gunsteren, and J. P. M. Postma. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23(8):1513–1518, 1984. doi: 10.1002/bip.360230807.

[104] G. Kaminski, E. M. Duffy, T. Matsui, and W. L. Jorgensen. Free energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. *The Journal of Physical Chemistry*, 98(49): 13077–13082, 1994.

[105] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. Oxford University Press, 1989.

[106] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

[107] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry*, 25(13):1656–76, October 2004. doi: 10.1002/jcc.20090.

[108] M. M. Reif, P. H. Hünenberger, and C. Oostenbrink. New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *Journal of chemical theory and computation*, 8(10):3705–3723, 2012. doi: 10.1021/ct300156h.

[109] A. D. Mackerell. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry*, 25(13):1584–604, October 2004. doi: 10.1002/jcc.20082.

[110] J. G. Kirkwood. Statistical mechanics of liquid solutions. *Chemical Reviews*, 19(3):275–307, 1936. doi: 10.1021/cr60064a007.

[111] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[112] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12562–6, October 2002. doi: 10.1073/pnas.202427399.

[113] J. G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 1935.

[114] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters*, 156(5):472–477, 1989.

[115] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[116] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *The Journal of Chemical Physics*, 109(18):7737, 1998. doi: 10.1063/1.477419.

[117] B. Hess, C. Holm, and N. van der Vegt. Osmotic coefficients of atomistic NaCl (aq) force fields. *The Journal of chemical physics*, 124(16):164509, April 2006. doi: 10.1063/1.2185105.

[118] M. Eigen and K. Tamm. Schallabsorption in Elektrolytlösungen als Folge chemischer Relaxation. *Zeitschrift für Elektrochemie*, 66(2):93–107, 1962.

[119] Y. Marcus and G. Hefter. Ion pairing. *Chemical reviews*, 106(11): 4585–621, November 2006. doi: 10.1021/cr040087x.

[120] A. A. Chialvo, P. T. Cummings, H. D. Cochran, J. M. Simonson, and R. E. Mesmer. Na+Cl- ion pair association in supercritical water. *The Journal of Chemical Physics*, 103(21):9379–9387, 1995.

[121] P. G. Daniele, A. De Robertis, C. De Stefano, S. Sammartano, and C. Rigano. On the Possibility of determining the Thermodynamic Parameters for the Formation of Weak Complexes using a Simple Model for the Dependence on Ionic Strength of Activity Coefficients: Na+, K+, and Ca^{2+} Complexes of Low Molecular Weight Ligands in Aqueous S. *J. Chem. Soc. Dalton Trans.*, (11):2353–2361, 1985. doi: 10.1039/DT9850002353.

[122] A. De Robertis, P. Di Giacomo, and C. Foti. Ion-selective electrode measurements for the determination of formation constants of alkali

and alkaline earth metals with low-molecular-weight ligands. *Analytica Chimica Acta*, 300(1-3):45–51, January 1995. doi: 10.1016/0003-2670(94)00421-H.

[123] Y. Sugita, A. Kitao, and Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *The Journal of Chemical Physics*, 113(15):6042, 2000. doi: 10.1063/1.1308516.

[124] Y. Okamoto. Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *Journal of Molecular Graphics and Modelling*, 22(5):425–39, May 2004. doi: 10.1016/j.jmgm.2003.12.009.

[125] H. Fukunishi, O. Watanabe, and S. Takada. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*, 116(20):9058, 2002. doi: 10.1063/1.1472510.

[126] R. Affentranger, I. Tavernelli, and E. E. Di Iorio. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *Journal of Chemical Theory and Computation*, 2(2):217–228, March 2006. doi: 10.1021/ct050250b.

[127] R. H. Swendsen and J.-S. Wang. Replica Monte Carlo Simulation of Spin-Glasses. *Physical Review Letters*, 57(21):2607–2609, 1986.

[128] Y. Sugita and Y. Okamoto. Replica exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314:141–151, 1999.

[129] D. J. Earl and M. W. Deem. Parallel tempering: theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–6, December 2005.

[130] K. Hukushima and K. Nemoto. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996. doi: 10.1143/JPSJ.65.1604.

[131] S. Kannan and M. Zacharias. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. *Proteins: Structure, Function, and Bioinformatics*, 66(3):697–706, 2007. doi: 10.1002/prot.21258.

[132] K. Ostermeir and M. Zacharias. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochimica et Biophysica Acta*, 1834(5):847–53, May 2013. doi: 10.1016/j.bbapap.2012.12.016.

[133] T. P. Straatsma and J. A. McCammon. Treatment of rotational isomeric states. III. The use of biasing potentials. *The Journal of Chemical Physics*, 101(6):5032, 1994. doi: 10.1063/1.468409.

[134] A. Barducci, G. Bussi, and M. Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters*, 100(2):020603, January 2008. doi: 10.1103/PhysRevLett.100.020603.

[135] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, February 2014. doi: 10.1016/j.cpc.2013.09.018.

[136] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–18, December 2005. doi: 10.1002/jcc.20291.

[137] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4:435–447, 2008. doi: 10.1021/ct700301q.

[138] MATLAB. *version 7.8.0 (R2009a)*. The MathWorks Inc., Natick, Massachusetts, 2009.

[139] G. Rosman, M. M. Bronstein, A. M. Bronstein, and R. Kimmel. Nonlinear Dimensionality Reduction by Topologically Constrained Isometric Embedding. *International Journal of Computer Vision*, 89(1):56–68, February 2010. doi: 10.1007/s11263-010-0322-1.

[140] A. E. Garcia. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 68:2696–2699, Apr 1992. doi: 10.1103/PhysRevLett.68.2696.

[141] R. Hegger, A. Altis, P. H. Nguyen, and G. Stock. How complex is the dynamics of peptide folding? *Phys. Rev. Lett.*, 98:028102, Jan 2007. doi: 10.1103/PhysRevLett.98.028102.

[142] M. Ceriotti, G. A. Tribello, and M. Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32):13023–8, August 2011. doi: 10.1073/pnas.1108486108.

[143] J. D. Chodera and V. S. Pande. The social network (of protein conformations). *Proceedings of the National Academy of Sciences of the United States of America*, 108(32):12969–70, August 2011. doi: 10.1073/pnas.1109571108.

[144] C. Hou, J. Wang, Y. Wu, and D. Yi. Local linear transformation embedding. *Neurocomputing*, 72(1012):2368 – 2378, 2009. doi: http://dx.doi.org/10.1016/j.neucom.2008.12.002.

[145] M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi. Determination of reaction coordinates via locally scaled diffusion map. *The Journal of Chemical Physics*, 134(12):124116, 2011. doi: http://dx.doi.org/10.1063/1.3569857.

[146] D. Wales. *Energy landscapes: Applications to clusters, biomolecules and glasses*. Cambridge University Press, Cambridge, UK, 2003.

[147] `http://sketchmap.berlios.de/compose.php?page=tuts`.

[148] M. Satoh, Y. Fujii, F. Kato, and J. Komiyama. Solvent- and salt-induced coilhelix transition of alkali metal salts of poly(l-glutamic acid) in aqueous organic solvents. *Biopolymers*, 31(1):1–10, 1991. doi: 10.1002/bip.360310102.

[149] M. V. Fedorov, J. M. Goodman, and S. Schumm. To switch or not to switch: the effects of potassium and sodium ions on alpha-poly-L-glutamate conformations in aqueous solutions. *Journal of the American Chemical Society*, 131(31):10854–6, August 2009. doi: 10.1021/ja9030374.

[150] J. Dzubiella. Molecular insights into the ion-specific kinetics of anionic peptides. *The Journal of Physical Chemistry B*, 114(20):7098–103, May 2010. doi: 10.1021/jp1010814.

[151] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal : EBJ*, 40(7):843–56, July 2011. doi: 10.1007/s00249-011-0700-9.

[152] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[153] G. A. Kaminski and R. A. Friesner. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[154] B. Hess and N. F. A. van der Vegt. Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *The Journal of Physical Chemistry B*, 110(35):17616–26, September 2006. doi: 10.1021/jp0641029.

[155] H. Berendsen, J. Grigera, and T. Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91:6269–6271, 1987.

[156] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics*, 120(20):9665–78, May 2004. doi: 10.1063/1.1683075.

[157] J. Aqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *Journal of Physical Chemistry*, 94:8021–8024, 1990.

[158] W. F. van Gunsteren and H. J. C. Berendsen. *Groningen Molecular Simulation (GROMOS) Library Manual*. 1987.

[159] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577, 1995. doi: 10.1063/1.470117.

[160] B. Hess and N. F. A. van der Vegt. Cation specific binding with protein surface charges. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32):13296–300, August 2009. doi: 10.1073/pnas.0902904106.

[161] M. Fyta and R. R. Netz. Ionic force field optimization based on single-ion and ion-pair solvation properties: going beyond standard mixing rules. *The Journal of Chemical Physics*, 136(12):124103, March 2012. doi: 10.1063/1.3693330.

[162] M. M. Reif, M. Winger, and C. Oostenbrink. Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins. *Journal of Chemical Theory and Computation*, 9(2):1247–1264, January 2013. doi: 10.1021/ct300874c.

[163] J. G. Kirkwood and F. P. Buff. The Statistical Mechanical Theory of Solutions. I. *The Journal of Chemical Physics*, 19(6):774, 1951. doi: 10.1063/1.1748352.

[164] A. Ben-Naim. *Molecular Theory of Solutions*. Oxford University Press, Oxford, 2006. ISBN 978-0-19-929970-6.

[165] R. Chitra and P. E. Smith. Molecular Association in Solution: A Kirkwood-Buff Analysis of Sodium Chloride, Ammonium Sulfate, Guanidinium Chloride, Urea, and 2,2,2-Trifluoroethanol in Water. *The Journal of Physical Chemistry B*, 106(6):1491–1500, February 2002. doi: 10.1021/jp011462h.

[166] Y. Wu, H. L. Tepper, and G. A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of Chemical Physics*, 124(2):024503, January 2006. doi: 10.1063/1.2136877.

[167] P. Fournier, E. H. Oelkers, R. Gout, and G. Pokrovski. Experimental determination of aqueous sodium-acetate dissociation constants at temperatures from 20 to 240 °C. *Chemical Geology*, 151:69–84, 1998.

[168] J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77:3865, 1996.

[169] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. *The Journal of Chemical Physics*, 132:154104, 2010.

[170] S. Goedecker, M. Teter, and J. Hutter. *Phys. Rev. B*, 54:1703, 1996.

[171] The cp2k developers group. http://cp2k.berlios.de/.

[172] G. Lippert, J. Hutter, and M. Parrinello. *Mol. Phys.*, 92:477, 1997.

[173] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511, 1984.

[174] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23(3):327–341, 1977.

[175] A. J. Cohen, P. Mori-Sanchez, and W. Yang. Insights into Current Limitations of Density Functional Theory. *Science*, 321(5890):792–794, August 2008.

[176] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Challenges for Density Functional Theory. *Chem Rev*, 112(1):289–320, January 2012.

[177] C. Krekeler and L. Delle Site. Solvation of positive ions in water: The dominant role of water-water interaction. *Journal of Physics: Condensed Matter*, 19(19):192101, 2007. doi: 10.1088/0953-8984/19/19/192101.

[178] K. Leung, S. B. Rempe, and O. A. von Lilienfeld. Ab initio molecular dynamics calculations of ion hydration free energies. *J Chem Phys*, 130 (20):204507, 2009.

[179] Q. Wan, L. Spanu, and G. Galli. Solvation Properties of Microhydrated Sulfate Anion Clusters: Insights from *abInitio* Calculations. *J Phys Chem B*, 116(31):9460–9466, August 2012.

[180] T. Dudev and C. Lim. Monodentate versus Bidentate Carboxylate Binding in Magnesium and Calcium Proteins: What Are the Basic Principles? *The Journal of Physical Chemistry B*, 108(14):4546–4557, April 2004. doi: 10.1021/jp0310347.

[181] R. L. Rardin, W. B. Tolman, and S. J. Lippard. Monodentate carboxylate complexes and the carboxylate shift: implications for polymetalloprotein structure and function. *New Journal of Chemistry*, 15 (6):417–430, 1991.

[182] J. J. Falke, S. K. Drake, A. L. Hazard, and O. B. Peersen. Molecular tuning of ion binding to calcium signaling proteins. *Quarterly Reviews of Biophysics*, 27(3):219–90, August 1994.

[183] S. M. Gagné, M. X. Li, and B. D. Sykes. Mechanism of direct coupling between binding and induced structural change in regulatory calcium binding proteins. *Biochemistry*, 36(15):4386–92, April 1997. doi: 10.1021/bi963076+.

[184] T. Ozawa, M. Fukuda, M. Nara, A. Nakamura, Y. Komine, K. Kohama, and Y. Umezawa. How can Ca^{2+} selectively activate recoverin in the presence of Mg^{2+}? Surface plasmon resonance and FT-IR spectroscopic studies. *Biochemistry*, 39(47):14495–503, November 2000.

[185] C. S. Babu, T. Dudev, R. Casareno, J. A. Cowan, and C. Lim. A combined experimental and theoretical study of divalent metal ion selectivity and function in proteins: application to *E. coli* ribonuclease H1. *Journal of the American Chemical Society*, 125(31):9318–28, August 2003. doi: 10.1021/ja034956w.

[186] A. Kondoh and T. Oi. Interaction of Alkaline Earth Metal Ions with Acetic and Lactic Acid in Aqueous Solutions Studied by ^{13}C NMR Spectroscopy. *Zeitschrift für Naturforschung A, A Journal of Physical Sciences*, 52a:351–357, 1997.

[187] E. J. M. Leenders, P. G. Bolhuis, and E. J. Meijer. Microscopic Picture of the Aqueous Solvation of Glutamic Acid. *Journal of Chemical Theory and Computation*, 4(6):898–907, June 2008. doi: 10.1021/ct700344f.

[188] F. Betts and A. Posner. An x-ray radial distribution study of amorphous calcium phosphate. *Materials Research Bulletin*, 9(3):353 – 360, 1974. doi: http://dx.doi.org/10.1016/0025-5408(74)90087-7.

[189] A. Dey, P. H. H. Bomans, F. A. Müller, J. Will, P. M. Frederik, G. de With, and N. A. J. M. Sommerdijk. The role of prenucleation clusters in surface-induced calcium phosphate crystallization. *Nature Materials*, 9:1010–1014, 2010. doi: 10.1038/nmat2900.

[190] W. J. E. M. Habraken, J. Tao, L. J. Brylka, H. Friedrich, L. Bertinetti, A. S. Schenk, A. Verch, V. Dmitrovic, P. H. H. Bomans, P. M. Frederik, J. Laven, P. van der Schoot, B. Aichmayer, G. de With, J. J. DeYoreo, and N. A. J. M. Sommerdijk. Ion-association complexes unite classical and non-classical theories for the biomimetic nucleation of calcium phosphate. *Nature communications*, 4:1507, January 2013. doi: 10.1038/ncomms2490.

[191] A. Picker, M. Kellermeier, J. Seto, D. Gebauer, and H. Cölfen. The multiple effects of amino acids on the early stages of calcium carbonate crystallization. *Zeitschrift für Kristallographie - Crystalline Materials*, 227(11):744–757, 2012. doi: 10.1524/zkri.2012.1569.

[192] A. R. Finney and P. M. Rodger. Probing the structure and stability of calcium carbonate pre-nucleation clusters. *Faraday Discuss.*, 159: 47–60, 2012. doi: 10.1039/C2FD20054F.

[193] K. Tai. Conformational sampling for the impatient. *Biophysical Chemistry*, 107(3):213–20, March 2004. doi: 10.1016/j.bpc.2003.09.010.

[194] A. Barducci, M. Bonomi, and M. Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):826–843, 2011. doi: 10.1002/wcms.31.

[195] B. Roux and T. Simonson. Implicit solvent models. *Biophysical Chemistry*, 78(12):1 – 20, 1999. doi: http://dx.doi.org/10.1016/S0301-4622(98)00226-9.

[196] N. Rathore, M. Chopra, and J. J. de Pablo. Optimal allocation of replicas in parallel tempering simulations. *The Journal of Chemical Physics*, 122(2):024111, 2005. doi: http://dx.doi.org/10.1063/1.1831273.

[197] A. Kone and D. A. Kofke. Selection of temperature intervals for parallel-tempering simulations. *The Journal of Chemical Physics*, 122 (20):206101, 2005. doi: http://dx.doi.org/10.1063/1.1917749.

[198] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992. doi: 10.1002/jcc.540130812.

[199] A. Wierzbicki, C. S. Sikes, J. D. Sallis, J. D. Madura, E. D. Stevens, and K. L. Martin. Scanning electron microscopy and molecular modeling of inhibition of calcium oxalate monohydrate crystal growth by citrate and phosphocitrate. *Calcified Tissue International*, 56(4):297–304, April 1995.

[200] A. Gul and P. Rez. Models for protein binding to calcium oxalate surfaces. *Urological research*, 35(2):63–71, April 2007. doi: 10.1007/s00240-007-0087-3.

[201] G. K. Hunter, J. O'Young, B. Grohe, M. Karttunen, and H. A. Goldberg. The flexible polyelectrolyte hypothesis of protein-biomineral interaction. *Langmuir*, 26(24):18639–46, December 2010. doi: 10.1021/la100401r.

Acknowledgement

Danksagung

RESUME