Simulating the Role of Molecular Binding to Mineral Surfaces: from Biomineral Growth to Cell Attachment

by

DAVID JONATHAN SPARKS

Thesis

for the award of

DOCTOR OF PHILOSOPHY

Department of Materials Science and Engineering

The University of Sheffield

April 2015

"An expert is a person who has found out by his own painful experience

all the mistakes that one can make in a very narrow field."

Niels Bohr (1885 - 1962)

"David wordt later een hele goeie uitvinder, want hij kan goed rekenen

en heel goed technisch lego-en."

Liesbeth Sparks (ca. 1996)

Abstract

The interactions between organic molecules and minerals is fundamental to the under-standing of processes such as biomineralisation, the attachment of bacteria to surfaces and the design of synthetic materials within biomimetics. This thesis shows how molec-ular dynamics simulations can be employed to study the organic - inorganic interactions and give new insights into the molecular binding at mineral surfaces that play a role in these processes.

The incorporation of amino acids within calcium carbonate crystals was simulated and show a high energy associated with the incorporation of these molecules. The amino acids get incorporated in-between the lattice planes of the crystal, causing small anisotropic distortions to the crystal. The inclusion of these molecules occurs via a goodness-of-fit principle, where disruptions to the crystal lattice should be kept to a minimum. These simulations show good agreement with experimental X-ray data.

Simulations of multiple tripeptides show a different conformational behaviour of the peptides in solution than on the surface of calcium carbonate. Whereas the pep-tides exhibit a flexible behaviour in solution, binding to the mineral surface induces a disorder-to-order transition and the peptides become rigid. These changes in con-formational behaviour offer insight into the structure and behaviour of intrinsically disordered proteins.

The polymer poly acrylic acid was simulated to analyse its conformational behaviour. In the presence of counter ions the polymer exhibits a flexible, extended conformation, whereas a coiled conformation is found in the absence of counter ions. The simulations in this work agree well with experimental spectroscopy studies. The binding of the polymer to a mineral surface is not only governed by the number of functional groups, but also the flexibility of the polymer. These results give an insight in how such molecules can aid the attachment of bacteria to surfaces.

Acknowledgements

Declaration

I declare that the work presented here is original research carried out at the Department of Materials Science and Engineering, The University of Sheffield, between October 2011 and April 2015 in accordance with the Regulations of The University of Sheffield. Where other studies are referenced it is explicitly stated in the text. No part of this thesis has been submitted for any other academic award.

David Jonathan Sparks

Contents

List of Publications

The work presented here has led to the following papers which are in course of publication.

'Incorporation, Structure and Properties of Calcite Crystals Occluding Amino Acids', Yi-Yeoun Kim, Beatrice Demarchi, Miki Kunitake, David J. Sparks, Joe Carloni, Chiu C. Tang, Colin L. Freeman, Shefford Baker, Boaz Pokroy, Kirsty Pinkman, John H. Harding, Lara A. Estroff and Fiona, C. Meldrum. In preparation.

'Conformational Behaviour and Adsorption of Multiple Tripeptides onto the Surface of Calcite: a Simulation Study', David J. Sparks, Colin L. Freeman and John, H. Harding. In progress.

'Adsorption of Poly Acrylic Acid onto the Surface of Calcite: an Experimental and Simulation Study', David J. Sparks, Maria E. Romero-Gonzalez, Elfateh El-Taboni, Colin L. Freeman, Shaun A. Hall, Gabriella Kakonyi, Linda Swanson, Steve A. Banwart and John H. Harding, *Physical Chemistry Chemical Physics*, (2015). Submitted.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Biomineralisation

During the existence of the earth, both the relative amounts and the diversity of minerals on the surface of the planet have drastically changed. A wide range of chemical, biological and physical processes such as changes in temperature, pressure, and the abundance of volatile gases such as H_2O, CO_2 and O_2 have caused an increase in the number and type of minerals. It is estimated that over time the number of minerals that can be found on the earth's surface has increased from a few hundred to over 4400 [1–3]. The work described here is focused on the mineral calcium carbonate ($CaCO_3$), which has been present throughout the history of the earth, although its abundance has changed over time. $CaCO_3$ appears from ca. 2.7 Ga onwards, but in terms of volume the early Cambrian period (before 0.54 Ga) saw a significant increase in the amount of $CaCO_3$ deposited in the forms of calcite and aragonite [1]. This large change in mineral formation is attributed to the development of biological life on earth and the rise of *biomineralisation*, a process of combining hard and soft matter to create functional materials for organisms [4]. The ability of biological organisms to include hard, inorganic minerals completely changed their development and

increased their chances of survival. Integration of both organic and inorganic materials via these biomineralisation processes allowed organisms to form materials to fulfill a range of functions, such as protection, mechanical strength, buoyancy, and gravity sensing [5].

1.1.1 Calcium carbonate

Calcium carbonate ($CaCO_3$) exists in several different polymorphs. Calcite, aragonite and vaterite are the main three anhydrous crystalline polymorphs, and ikaite and monohydrocalcite are two hydrated polymorphs. Amorphous calcium carbonate is the amorphous, non-crystalline phase of $CaCO_3$.

Calcite

At standard temperature and pressure calcite is the thermodynamically most stable polymorph of calcium carbonate and is found abundantly in nature. Calcite appears as a trigonal crystal system but by combining several unit cells can be described as a hexagonal lattice (Figure 1.1) in space group $R\bar{3}c$ with lattice parameters $a = b = 4.991$ Å, $c = 17.062$ Å, $\alpha = \beta = 90°$ and $\gamma = 120°$ [6]. The unit cell of calcite has alternating layers of $CO_3{}^{2-}$ and Ca^{2+} ions perpendicular to the c-axis, with the C$-$O bonds of the $CO_3{}^{2-}$ ions in consecutive layers pointing in opposite directions. Calcite is a brittle material and has a hardness of 3 on the Mohs scale of mineral hardness[1]. The density of calcite is 2.7102 g cm^{-3}. The crystal of calcite can be cut along several different cleavage planes to expose a surface. Cleaving the crystal along the hexagonal ($10\bar{1}4$) plane yields the thermodynamically most stable surface, where both Ca^{2+} and $CO_3{}^{2-}$ ions are exposed on the surface (Figure 1.2) [8]. The model ($10\bar{1}5$) and ($10\bar{1}3$) surfaces, which are stepped surfaces based on the ($10\bar{1}4$) surface, are the next thermodynamically favourable surfaces, followed by the planar (0001) and ($10\bar{1}0$) surfaces (Table 1.1) [9]. As

---

[1]The Mohs scale is an ordinal scale of hardness of materials to describe the scratch resistance of a material [7].
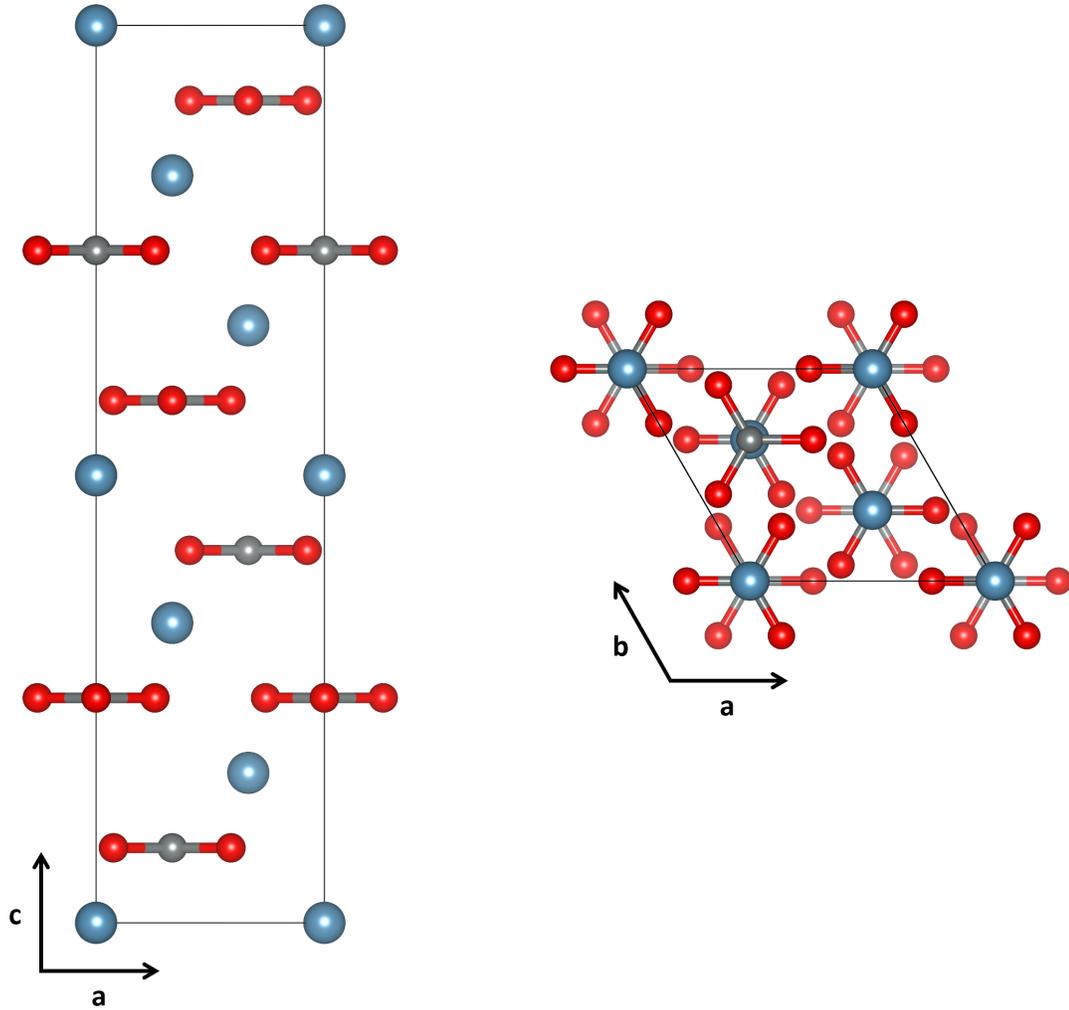
Figure 1.1: Schematic representation of the unit cell of calcite in space group $R\bar{3}c$ with cell parameters $a = b = 4.991$ Å, $c = 17.062$ Å, $\alpha = \beta = 90°$ and $\gamma = 120°$.

the $(10\bar{1}4)$ surface is most stable surface of calcite, it will therefore be the predominantly expressed face of growing calcite.

Aragonite

Aragonite is the second thermodynamically stable polymorph of calcium carbonate and, like calcite, also has a widespread availability in nature. The lattice system of aragonite is orthorhombic with lattice parameters $a = 4.962$ Å, $b = 7.969$ Å, $c = 5.743$ Å and $\alpha = \beta = \gamma = 90°$ and has the space group $Pmcn$ (Figure 1.3) [10, 11].

3
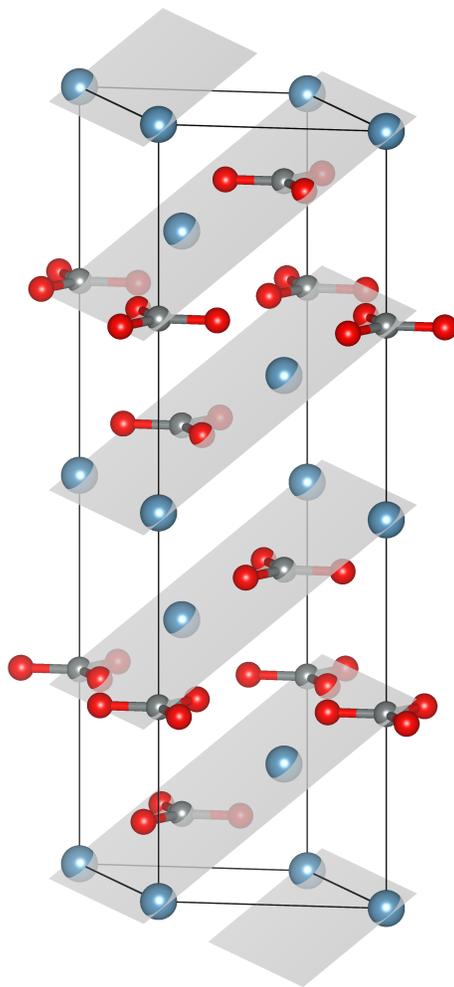
Figure 1.2: Schematic representation of the unit cell of calcite in with the hexagonal $(10\bar{1}4)$ lattice planes.

Surface	Surface energy (J m^{-2})
$(10\bar{1}4)$	0.60
$(10\bar{1}5)$	0.69
$(10\bar{1}3)$	0.87
$(0001)^a$	0.97
$(0001)^b$	0.99
$(10\bar{1}0)$	0.97

Table 1.1: The calculated surface energies of various calcite surfaces in solution. [a] layer terminated by Ca^{2+} ions; [b] layer terminated by CO_3^{2-} ions. Adapted from [9] with permission of The Royal Society of Chemistry.
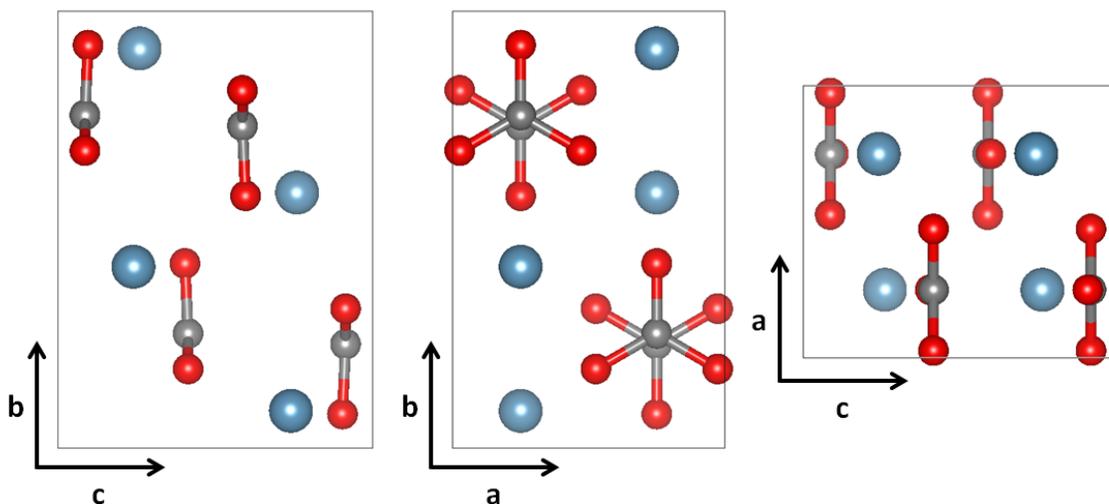
Figure 1.3: Schematic representation of the unit cell of aragonite in space group *Pmcn* with cell parameters $a = 4.962$ Å, $b = 7.969$ Å and $c = 5.743$ Å.

Vaterite

Whereas the crystal structures of calcite and aragonite are well defined, the structure of the metastable polymorph vaterite is not. Several crystal structures have been proposed over the last half century [12–14] and even with advances in experimental diffraction techniques and computational simulations the discussion about its structure is still ongoing. Recent *ab initio* calculations and experiments [15, 16] have hypothesised that vaterite exists as multiple structures in several hexagonal and monoclinic space groups ($P3_221$, $C2/c$ and $C\bar{1}$, Figure 1.4) in which the $CO_3{}^{2-}$ ions are dynamic and can be displaced and rotated. Demichelis *et al.* [15] hypothesise that this dynamic behaviour is the cause of the uncertainty in the structure of vaterite.

Hydrated polymorphs

Ikaite, or hexahydrate calcium carbonate ($CaCO_3 \cdot 6\,H_2O$), is a metastable polymorph of calcium carbonate that has been observed to exist in rare cases (low temperature and high pressure) in nature [17]. Due to its low stability ikaite will rapidly convert to the more stable polymorph calcite at standard conditions [18]. Ikaite exists in a monoclinic
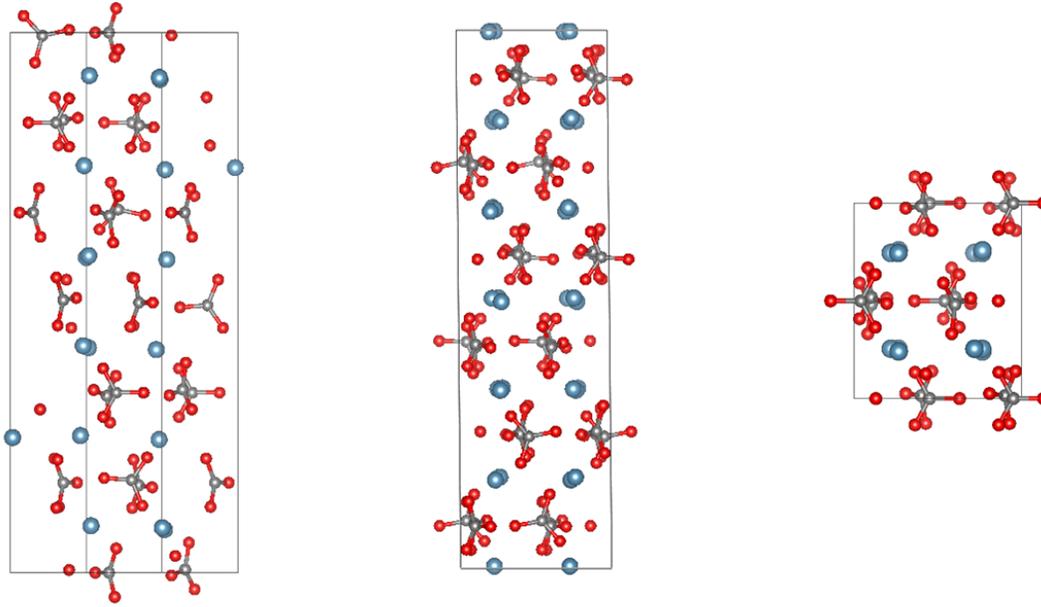
(a) Space group $P3_221$.　　　　(b) Space group $C2/c$.　　　　(c) Space group $C\bar{1}$.

Figure 1.4: Schematic representation of the unit cells of vaterite proposed by Demichelis *et al.* [15].

structure in space group $C2/c$ with $a = 8.792$ Å, $b = 8.310$ Å, $c = 11.021$ Å, $\alpha = \gamma = 90°$ and $\beta = 110.53°$ [19]. Monohydrocalcite ($CaCO_3 \cdot H_2O$) can also be observed at low temperatures and high pressures, although near ambient conditions, both dry and in the presence of water, the crystal will transform to either calcite or aragonite and is thus another metastable polymorph of calcium carbonate [20]. The space group of monohydrocalcite is the hexagonal $P3_21$ with $a = 6.0931$ Å and $c = 7.5446$ Å [21].

Amorphous calcium carbonate

Amorphous calcium carbonate (ACC) is isotropic, does not diffract X-rays and is the only polymorph of calcium carbonate that does not have a crystalline phase [5]. Both anhydrous and hydrated ACC exist, although the former is less stable. Hydrated ACC is a precurser for other crystalline polymorphs such as calcite within many organisms [22]. The main biological relevant form of ACC is approximately $CaCO_3 \cdot H_2O$. Hydrated

ACC structures can be stable for days to months, depending on the conditions and the presence of foreign molecules, and organisms have been seen to stabilise these ACC structures for extended periods of time [5]. X-ray experiments and computational studies have shown that although ACC is an amorphous phase, short range order does exist within the structure [23–26] and the overall structure of ACC is still not fully understood.

1.1.2 Nucleation and growth

The formation of the crystalline polymorphs of calcium carbonate starts via a nucleation event when concentrations of the calcium and carbonate ions in solution are high enough to form crystals. This nucleation event is still poorly understood, but several theories exist that describe the event. In the 1930s the concept of classical nucleation theory was developed [27], describing the nucleation of crystals from supersaturated solutions under the assumptions that growing clusters are spherical, grow one species at a time and that there is a clear distinction between the bulk and surface energies, with the bulk energy of a cluster equal to the relevant bulk energy of a perfect crystal and the surface energy of a cluster equal to the surface energy of the bulk crystal surface. The bulk energy of these crystals is lower than the individual ions in solutions, and therefore the formation of these crystals is thermodynamically favourable. However, the interfacial energy between the surface of these crystals and solution is positive, preventing small crystals from forming if the interfacial energy is larger than the bulk energy.

The bulk energy scales with respect to the volume of the crystals, and thus with the cube of the crystal radius (r^3). The interfacial energy of the forming crystals scales with respect to the surface area, which is scaled with the square of the radius (r^2). Although the unfavourable interfacial energy rises with an increasing particle radius, the favourable bulk energy increases quicker. This results in a critical radius at which the favourable bulk energy will dominate (Figure 1.5), leading to the formation of

small crystals that will continue to grow. Particles smaller than the critical radius will be thermodynamically unstable and will dissolve in the solution. The free energy change of the system (Δg) is given by the sum of the bulk and surface terms: $\Delta g = \Delta g_{bulk} + \Delta g_{surface}$ [28]. The free energy is minimised (vanishing first derivative), yielding a critical radius for nucleation:

$$r_{critical} = 2\frac{\Omega\alpha}{\Delta\mu} \tag{1.1}$$

where Ω is the volume per molecule, α is the interfacial free energy and $\Delta\mu$ is the change in chemical potential.
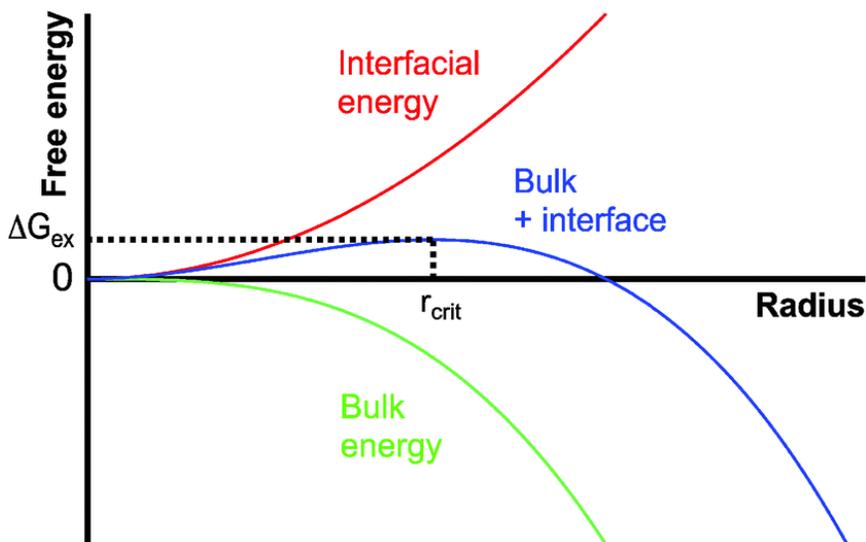
Figure 1.5: Schematic representation of the classical nucleation theory, where the sum of the bulk and surface energies will result in a critical particle radius after which the favourable bulk energy will dominate and crystals will continue to grow. From [29]. Published by The Royal Society of Chemistry.

However, the well established concept of the classical nucleation theory has been questioned with recent claims of thermodynamically stable clusters which are smaller than the assumed critical radius and stable larger pre-nucleation clusters of amorphous phase which exist for long time scales [29–31]. These pre-nucleation clusters are relatively large (up to 4 nm) agglomerations of calcium and carbonate ions that have not crystallised. These stable pre-nucleation clusters form via the aggregation of ions in supersaturated

solutions, and the formation of crystals from this phase has been hypothesised to be a stepwise pathway: "pre-nucleation clusters $\rightarrow$ dense liquid nanodroplets $\rightarrow$ liquid ACC $\rightarrow$ solid ACC $\rightarrow$ anhydrous crystalline polymorphs" [29]. Where the classical nucleation theory predicts the nucleation of $CaCO_3$ via the addition of ions to a single cluster, this new theory describes the nucleation as an aggregation of amorphous pre-nucleation clusters which subsequently crystallise (Figure 1.6). The debate between the two theories is still ongoing, with both experimental and computational simulation studies being utilised to shed light on this problem [26, 29, 32–35].
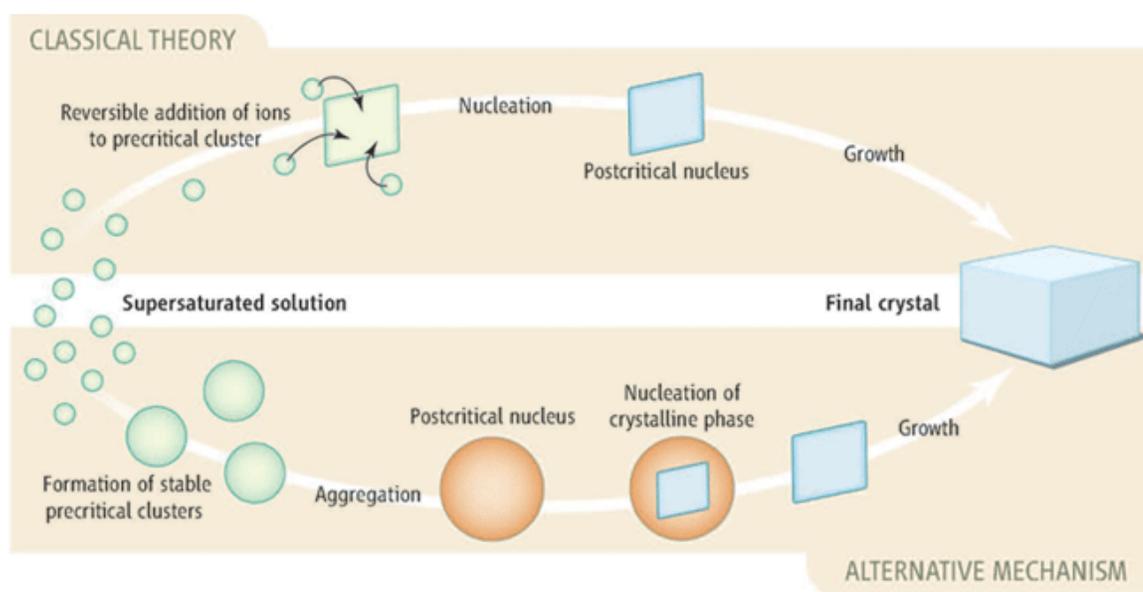
Figure 1.6: Schematic representation of the classical nucleation theory and the formation of pre-nucleation clusters. From [32]. Reprinted with permission from AAAS.

1.1.3 Biomineralisation in nature

Growth control

In nature organisms utilise the interactions between organic macromolecules such as proteins, peptides and polysaccharides, to control the nucleation and growth of the mineral phase into (single) crystals. By using both organic and inorganic matter they are able to create biocomposite materials that exhibit intricate hierarchical architec-

tures and increased mechanical properties compared to the organic or inorganic phases alone.

Perhaps the best known example of this biomineralisation process is that of coccolithophores, eukaryotic unicellular micro algae with a mineralised calcium carbonate exoskeleton. The coccolithophores concentrate HCO_3^- and Ca^{2+} ions from their environment which are subsequently crystallised into intricate single crystal $CaCO_3$ platelets [36, 37] to form this exoskeleton. During the crystallisation process an organic phase composed of coccolith associated polysaccharides (CAPs) interacts with the acute steps[2] on the surface of growing calcite and by blocking the acute growth sites the CAPs promote the growth of surfaces parallel to the c-axis [37]. The highly intricate structure of the coccolith platelets formed via this process is shown in Figure 1.8. The interactions between the CAPs and the mineral phase is an excellent example of the control these organisms can exert on the mineral phase to produce the shapes and structures required.
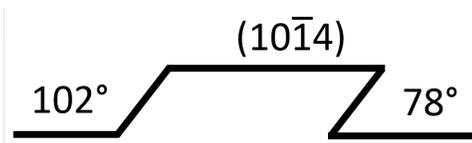
Figure 1.7: The edges of a growing ($10\bar{1}4$) calcite surface have an obtuse ($102°$) and acute ($78°$) step on either side of the surface.

Polymorph selection

In many organisms the control over the growth and nucleation of $CaCO_3$ crystals extends further to the control over polymorph selection [39, 40]. One of these examples is the mollusc species *Atina rigida* where an aragonite structure, nacre (Figure 1.9a), is formed on the inside of the mollusc shell. Remarkably, this aragonite structure is complemented with a prismatic layer of calcite crystals on the outside of the shell (Figure 1.9b) [4]. The small aragonite platelets on the inside of the shell are surrounded

---

[2]The edges of a growing ($10\bar{1}4$) calcite surface will show an obtuse and acute step on either side of the surface, shown in Figure 1.7.
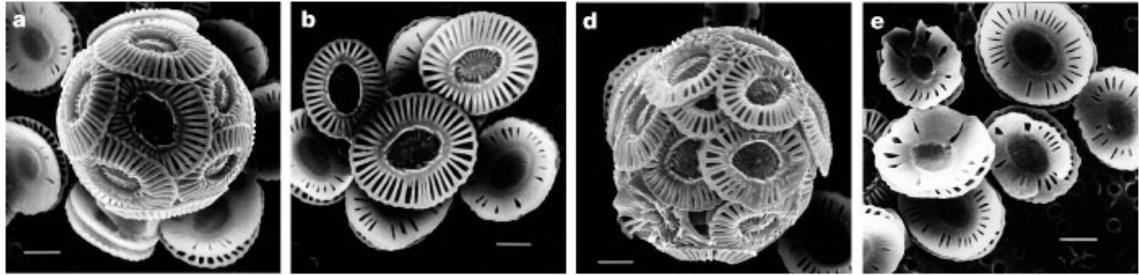
Figure 1.8: Scanning electron microscope (SEM) images of the exoskeleton of *Emiliania huxleyi* coccolithophores. The exoskeleton consists of small $CaCO_3$ platelets produced by the organism and the complex structures of these platelets are formed due to the interactions between polysaccharides and the mineral phase during their formation. Scale bars represents 1 μm. Adapted by permission from Macmillan Publishers Ltd: Nature, [38] copyright 2000.

by a thin organic matrix when crystallised [4,41] and are packed tightly within lamella sheets. Neighbouring sheets are rotated by 90 degrees relative to one another, with separate shell layers tilted by another 45 degrees (Figure 1.10) to form a crossed-lamellar structure that minimises the anisotropy of the structure and transforms the normally brittle aragonite into an elastic material [41].

The fracture resistance of these crossed-lamellar structures is five orders of magnitude higher than that of geological aragonite [42]. The outer layer of the shell (Figure 1.9b), consisting of calcite crystals, also display mechanical properties that exceed their geological counterparts, due to the organic molecules included within the crystal structure of these materials [43]. The example of these mollusc shells shows that the organisms not only control the polymorph selection of the mineral phases, but in doing so are able to construct a highly hierarchical structures with mechanical properties that far exceed their geological counterparts.

1.1.4 Biomineralisation *in vitro*

Within the biomineralisation examples set out above there is an important interplay between the crystallising mineral and the organic phase present. A range of biomolecules within the organic phase supports the mineral phase to take its form by promoting

11

(a) Nacreous part of the shell. (b) Prismatic part of the mollusc shell.

Figure 1.9: The inner shell of a mollusc is comprised of the composite material nacre, or mother of pear, an iridescent material composed of aragonite platelets and an organic matrix that has mechanical properties that exceed its geological counterparts. The prismatic structure on the outside shell consists of calcite crystals. Reprinted (adapted) with permission from [44]. Copyright 2008 American Chemical Society.
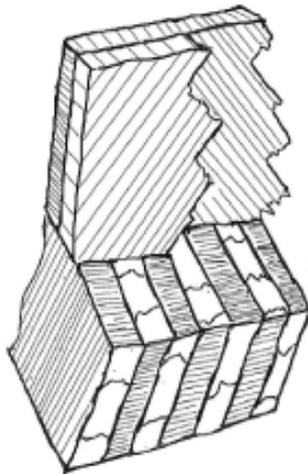
Figure 1.10: The crossed-lamellar structure of nacre uses an hierarchical structure to transform brittle aragonite into an elastic material. Reprinted from [41] with permission from Elsevier.

nucleation events, managing the growth of certain crystal polymorphs and controlling the growth of the mineral in certain crystallographic directions. A vast number of *in vitro* experiments have been performed to mimic the control that biomolecules have over the crystallisation of the mineral phase as seen in nature [45]. Both macromolecules found in nature such as proteins, peptides and polysaccharides, and novel biomolecules such as peptides, peptoids and polymers have been used to assess their control over the nucleation, polymorph selection and growth of calcium carbonate.

In vitro growth control

In vitro recrystallisation experiments in the presence of a range of isolated intra-crystalline protein fractions from both sea urchin spines and mollusc samples were performed by Albeck *et al.* [46]. Protein fractions extracted from sea urchin spines were observed to interact only with crystal planes parallel to the crystallographic c-axis, promoting an elongated growth of the crystal along this axis. The major protein fraction extracted from the mollusc showed the opposite behaviour and was found to bind primarily to planes perpendicular to the c-axis, inducing the formation of smooth (001) surfaces. The authors suggest that these two organisms are clearly able to direct a high level of control over the crystallisation of the mineral phase by their ability to secrete these intra-crystalline proteins in "a perfectly orchestrated sequence in time and space" [46]. Two other nacre proteins that were isolated from an abalone shell, AP8-α and AP8-β, also displayed control over the morphology of growing calcite crystals *in vitro* [47]. With either of the two proteins present in solution the normally rhombo-hedral calcite crystals developed rounded, acute edges and were elongated along the crystallographic c-axis when analysed with scanning electron microscopy (SEM) (Figure 1.11).

Synthetic organic molecules were able to induce a similar control over the morphology of growing $CaCO_3$ crystals. Chen *et al.* [48] synthesised a range of peptoids (Figure 1.12a),
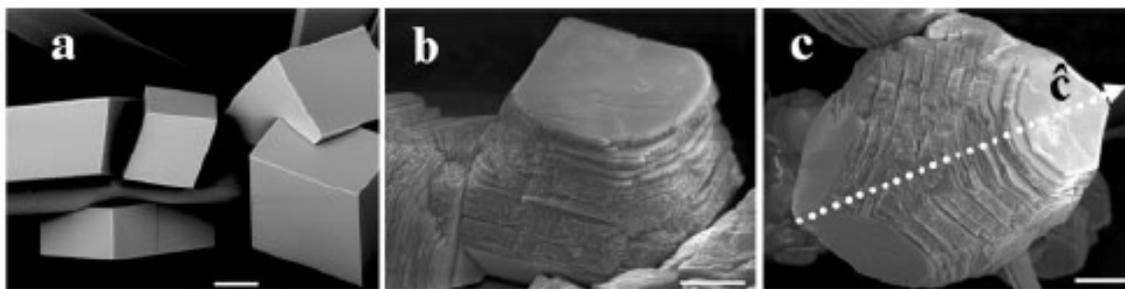
Figure 1.11: SEM images of: a) Calcite growth in the absence of organic additives. The rhombohedral structure is clearly visible. b,c) Calcite crystal grown in the presence of 10 μg mL^{-1} AP8-α. The acute edges of the crystal appear rounded, whereas the obtuse edges remain untouched. Scale bar represents a) 10 μm, b,c) 5 μm. From [47]. Reprinted with permission from the Wiley Company.

synthetic poly-N-substituted glycines in which the side chains of the peptide are attached to the nitrogen of the backbone, rather than the α-carbon, to study their effect on the morphology of calcite. In high concentrations the peptoids were able to control the morphology of the crystallising calcite (Figure 1.12b), with the formation of elongated spindles, truncated rhombohedral, twinned spherical shaped, cross-shaped and twisted paddle shaped structures being obtained in the presence of different peptoids. With the changes in morphology it can be seen that both the number of functional groups and the position of these functional groups throughout the peptoids are important factors in the control over the final morphology of the calcite crystals.

In addition to peptoids, synthetic polymers have also been found to alter the growth of calcite particles. In the presence of rigid synthetic polymers [49, 50] (Figure 1.14) the morphology of growing calcite changed as the polymers were able to block the growth of calcite in all but the crystallographic c-axis (Figure 1.13), promoting the crystal growth on the (101) surface, along the c-axis. Other rigid polymers have been supported on metal substrates to produce self assembled monolayers (SAMs), and the nucleation of calcite crystals can be induced on top of these SAMs. Several research groups such as Aizenberg et $al.$ [51–53], Travaille et $al.$ [54, 55] and Hu et $al.$ [56] were able to control the CaCO$_3$ nucleation on a range of SAMs. The nucleating crystals showed distinct homogeneous surfaces, but different SAMs induced different nucleating crystallographic
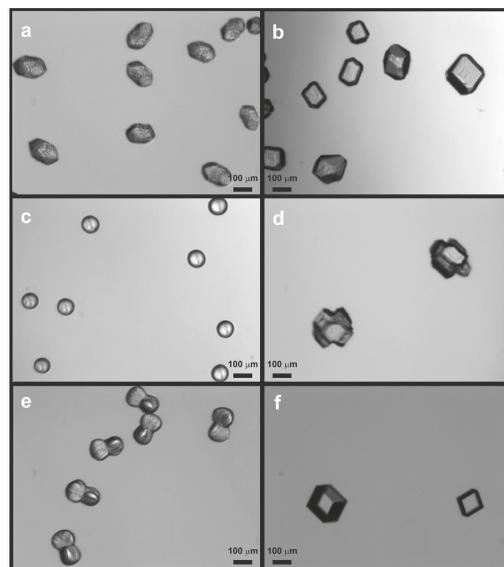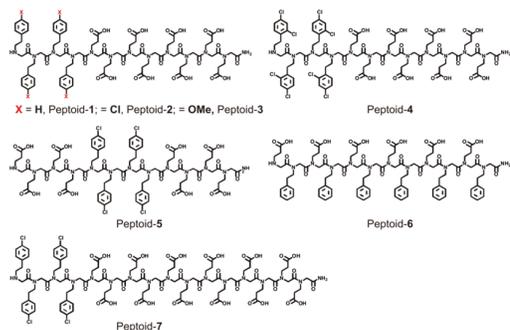
(a) Molecular structures of the peptoids.　　(b) SEM images of calcite crystals.

Figure 1.12: The effect of peptoids on the morphology of growing calcite crystals. SEM images a-e show elongated spindles, truncated rhombohedral, twinned spherical shaped, cross-shaped and twisted paddle shaped structures in the presence of peptoids 2, 3, 4, 5 and 7 respectively. SEM image f is a control with no additives. Reprinted (adapted) with permission from [48]. Copyright 2011 American Chemical Society.

planes. These experiments show that the flexibility of the underlying organic matrix and the interface that can be formed between the nucleating mineral phase and the organic phase are important features of the nucleation events.
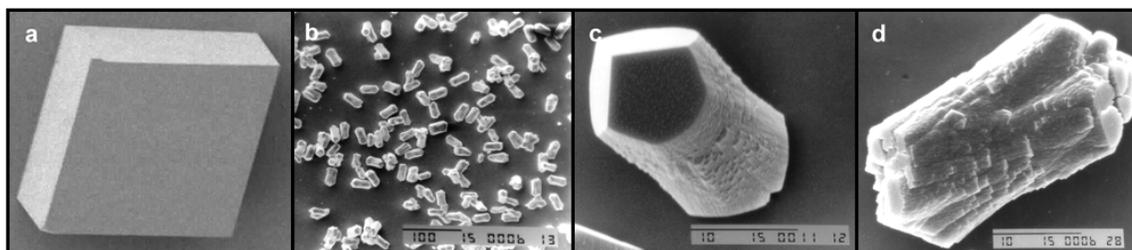
Figure 1.13: SEM images of calcite crystals formed in the presence of rigid synthetic polymers during calcite precipitation. Reprinted (adapted) with permission from [49]. Copyright 2002 American Chemical Society.

Polymorph selection

As mentioned in Chapter 1.1.3, the shell of certain species of mollusc consists of both aragonite and calcite polymorphs. Experimental studies [40,57] have been performed to
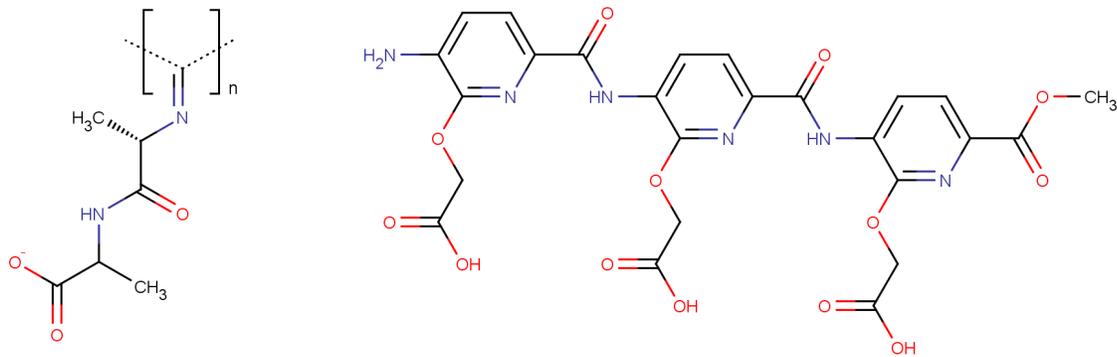
Figure 1.14: Molecular structures of the rigid alanyl-alanine-derived poly(isocyanide) polymer used within the precipitation experiments of Donners *et al.* and the [49] and the tricarboxylic acid derivative used in the work of Estroff *et al.* [50].

induce calcium carbonate precipitation in the presence of organic matter isolated and purified from the aragonite and the calcite partitions of the mollusc shells. These studies show that, under the same conditions, aragonite crystals formed in the presence of the organic phase isolated from the aragonite partition and calcite crystals were precipitated in the presence of biomolecules from the calcite partition. Another biomolecule thought to be involved with specific polymorph selection is the protein n16, one of many [58] proteins found in the Japanese pearl oyster *Pinctada fucata*, and several studies [59–62] have used the n16 sequence in crystallisation experiments to show that the peptide can induce aragonite nucleation *in vitro*. It is clear from these *in vitro* experiments that the organic phase is responsible for the polymorph selection of these crystals *in vivo*, although the exact control of the organic phase on the nucleation and growth is still poorly understood.

Kinetic control

Henriksen *et al.* [37] performed *in vitro* calcite precipitation experiments in the presence of CAPs isolated from coccolithophores and found a preferential binding of the CAPs to the acute stepped surface of calcite, and blocking the acute growth sites, slowing down,

16

or inhibiting, the growth of the calcite. Similar results of the inhibition of calcite growth were observed by Lakshtanov *et al.* [63,64] who analysed the inhibition of calcite *in vitro* by (branched) polysaccharides and found an increasing inhibition of calcite precipitation with increasing concentration of polysaccharide in solution. The authors speculated that the growth of calcite is inhibited by the polysaccharides attaching to and blocking the active sites on the growing surface, and that both the structure and position of the functional groups on the polysaccharides were responsible for their control of calcite inhibition.

Surprisingly, when analysing the growth kinetics of calcite crystals using *in situ* atomic force microscopy (AFM) in *in vitro* experiments with AP8-α and AP-β, the opposite effect is observed. Both AP8-α and AP-β have been seen to alter the calcite growth kinetics by increasing the step growth speed at both the obtuse and acute steps on the calcite ($10\bar{1}4$) surface. With low concentrations of the proteins (0.15 μM), growth accelerations of 3.3-fold and 2.3-fold for AP8-α, and 2.2-fold and 1.5-fold for AP8-β were observed at the obtuse and acute steps respectively. The mechanism for this increase in growth kinetics proposed by the authors is that the proteins act as surfactants, with large hydrophilic regions of the proteins that are rich in aspartic acid[3]. These hydrophilic regions interact with water molecules and ions in solution, modifying the thermodynamics at the step edges by lowering the energy barrier for the attachment of calcium and carbonate ions to the growing surfaces. Elhadj *et al.* [65,66] measured the increase of step growth at the ($10\bar{1}4$) calcite surface in the presence of various aspartic acid (di)peptides and analysed the correlation between the step growth acceleration and the molecular charge and hydrophilicity of the peptides. Their results showed that both the net charge and hydrophilicity of the peptides are good descriptors for the acceleration in step growth. *In situ* AFM measurements were also performed to measure the growth kinetics in the presence of peptoids (Figure 1.12a), and extraordinary increases of up to 23-fold in step growth were observed for both the obtuse and acute steps on

---

[3]A list of all amino acids is included in Appendix A.

the $(10\bar{1}4)$ surface. The authors suggest a few possible mechanisms for this dramatic increase, with the disruption of the water layer on the mineral surface by the peptoids being the most likely source of this increased step growth.

Macromolecular attachment

The attachment of the organic phase to the surface of minerals also plays a significant role in the formation of bacterial biofilms, for example. Microbial cells tend to agglomerate and attach to surfaces via the excretion of extracellular polymer substance (EPS). This EPS consists of organic material that is made up of macromolecules such as polysaccharides, proteins and lipids [67]. Functional groups such as carboxyl groups play an important role in the attachment of macromolecules such as lipopolysaccharides, mycolic acid and alginate that are found in the EPS. These surface - macromolecular interactions are of great interest in the study of the attachment of bacterial cells to the surface of minerals.

1.2   Computational studies

Computational methods can be used to build models of biomineral systems to provide atomistic details of the processes involved with biomineralisation. The range of length scales, from μm to cm, involved in these biomineralisation processes make the simulation of these processes difficult, and we must therefore consider different computational approaches for each problem [68, 69]. There are a number of computational techniques available which are described in Chapter 2. A large range of computational studies has been employed to study biomineralisation processes, ranging from the simulation of the aqueous interface to the nucleation of $CaCO_3$ and the binding of biomolecules to the surfaces of these crystals. A small selection of these studies is discussed below.

1.2.1 Nucleation

Simulation studies on the stability of small calcite particles in water have shown that the water structure on the surface of these particles stabilises the particles [70, 71]. Simulation studies of particle formation from an amorphous precursor have also been performed, and to overcome the long time scales it takes for these events to take place, Quigley *et al.* [25, 34] used accelerated molecular dynamics (MD) simulations to map the relative stabilities of different polymorphs as a function of particle size. Small nano particles formed of 75 or 192 $CaCO_3$ formula units were seen to be stable as both vaterite-like and calcite structures, although the vaterite-like structures collapsed to amorphous particles with continuous simulations. In the case of larger particles (300 $CaCO_3$ formula units), only the original metastable amorphous phase and calcite particles were observed. The same research group performed simulations of the ACC nucleation in the presence of biomolecules [72] and showed that in the presence of the eggshell protein OC-17 the energy barrier between the amorphous and calcite phases disappears, causing the amorphous particles to rapidly nucleate to calcite particles.

Simulations performed by Raiteri and Gale [26] with ACC particles showed that both the water interface and the inclusion of water molecules throughout the amorphous phase lowered the free energy of the ACC and stabilised the particles in an aqueous environment, supporting recent claims by Gebauer *et al.* [29] that crystallisation of $CaCO_3$ occurs via these stable ACC particles. Simulations to analyse the $CaCO_3$ nucleation behaviour on the surface of SAMs have also been performed by various research groups [56, 73–75] to mimic experiments [51–55]. These simulations indicate that the ionisation of the functional groups on the SAMs is important for the nucleation of the mineral phase. Furthermore, the flexibility of the SAMs, its ability to form a close match between the epitaxy of the nucleating calcite and the charge density on the SAMs are important for the nucleation to occur.

1.2.2 The effect of the aqueous interface

From the simulations described above it is clear that water plays an important role in the nucleation events of $CaCO_3$, and de Leeuw and Parker [8, 9] have simulated the adsorption of water on the surfaces of calcite, aragonite and vaterite to calculate the surface and hydration energies for these crystals. Other studies [70, 76–78] have also simulated the adsorption of water molecules on calcite surfaces and found a highly ordered water structure, with two water layers at 2.3 and 3.2 Å from the surface, stabilised by hydrogen bonding between the molecules. Experimentally these water layers can also be seen with X-ray scattering experiments [79]. From these simulations it is clear that this water double layer is an important feature of the mineral surfaces and should not be ignored.

1.2.3 Organic - mineral interactions and the control of growth

Within biomineralisation processes the organic-mineral interactions are of particular interest and a large number of simulations have been performed to study the binding of biomolecules to the surfaces of minerals. Simulations with a range of mono-, di- and poly-saccharides on various calcite surfaces have been performed by various research groups [80–82] and show that the adsorption of these molecules disrupts the ordered water layers on the surface of calcite. In the simulations of a mono-saccharide an energy barrier of ca. 50 kJ mol^{-1} was calculated between 5.1 and 6.6 Å from surface due to the water layers [81]. Yang *et al.* [82] observed that the binding energy for poly-saccharides depends on the site of adsorption, with a higher energy of adsorption seen at acute stepped surface sites than flat or obtuse sites. Simulations of methylamine and methanoic acid [83] showed a higher energy of adsorption for the latter, due to stronger interactions between the calcite surface ions and the carboxyl functional group. The importance of the functional groups in the binding process were also observed by Perry *et al.* [80] who suggested that the placement of the functional groups on the

di-saccharides used in their simulations was important for the interaction with the underlying surface, and therefore the strength of the binding. Simulations involving larger biomolecules have also been performed, with the adsorption of the protein OC-17 to flat calcite surfaces being performed by Freeman *et al.* [84]. Simulating the adsorption of the protein in several conformations on a solvated calcite surface they observed that the minimisation of the surface waters during the binding process resulted in the most favourable binding energies.

1.3   Research objectives and outline

I have shown above, using several examples, that the interactions between the organic and inorganic phases play an important role in the biomineralisation process. The role of the binding of biomolecules to the mineral surfaces can be considered from the point of view of the biomolecule or the surface. The presence of the biomolecules can control the mineral growth of the surfaces that it binds to, and if the molecules become incorporated within the mineral itself, it can profoundly influence the properties of the resulting composite material. On the other hand, the interplay between the arrangement of the functional groups on the biomolecule and the mineral surface can determine the strength of the attachment and the conformation of a biomolecule.

These biomineralisation processes can be applied in a range of research disciplines. The strong interactions between biomolecules and mineral surfaces for example plays an important factor in the bacterial binding to surfaces to form biofilms [85]. In the research area of carbon sequestration the trapping of CO_2 from the atmosphere into $CaCO_3$ storage becomes feasible with the recent developments of Chen *et al.*, who observed calcite growth accelerations of 23-fold for their synthetic peptoids [48, 86]. Within the area of composite materials, the highly hierarchical architectures and inclusion of organic macromolecules seen in biological systems are just some of the solutions employed to improve the properties of a material [41, 43, 87–89] and similar methods can be em-

ployed within synthetic materials design to produce composite materials with improved mechanical and physical properties [90].

The research described in this thesis uses a range of simulation methods to investigate the interactions between biomolecules and the mineral calcium carbonate. The work has been performed to gain a better understanding of the interplay between hard, inorganic matter and soft, organic matter, with the objective to answer the questions of how these biomolecules interact with the mineral phase; what are the key interactions between the two phases; and what effect does the mineral phase have on the conformation of these biomolecules. The aim of this research is to elucidate the specific interactions between the organic molecules and the mineral phase. The methodology behind the computational simulations performed in this work is set out in Chapter 2. Next, the research performed is separated into four chapters, each of which investigates the organic-inorganic interactions from a different point of view.

In Chapter 3 the incorporation of small biomolecules within a calcite crystal is investigated to analyse the effect of the inclusion of biomolecules on the crystal lattice and to suggest possible modes of incorporation of these biomolecules. This research was performed with the collaboration of the experimental research group of Fiona Meldrum at the University of Leeds.

Chapter 4 reports research into the behaviour of disordered tripeptides on the surface of calcite and discusses to what extent the peptide-peptide and mineral-peptide interactions control the conformational behaviour of the binding biomolecules. As part of this research, I have carried out calcite precipitation experiments at the NanoGeoScience laboratory at Copenhagen University. The aim of this study is to gain an understanding of the interactions that govern the binding of flexible molecules to the surface of calcite.

In Chapter 5 the polymer poly acrylic acid is used as a model to study the attachment of bacterial cells to the surface of calcite. The conformational behaviour of the polymer

is analysed in solution and on the surface of the mineral, and the adsorption energy of the biomolecules to the mineral surface is calculated. This research has been performed in collaboration with the experimental research group of Maria Romero-Gonzalez at the University of Sheffield.

Chapter 6 is a preliminary study, linked to the work on the inclusions of amino acids into calcite set out in Chapter 3. Within this chapter the binding of amino acids to the surface of amorphous calcium carbonate has been simulated, to get an understanding of the strength of the mineral-amino acid interactions.

Chapter 2

Molecular modelling

2.1   Introduction

Computational simulations of molecular systems have become a widely used aid in a range of scientific applications. Examples of these applications include the calculation of electronic structures via quantum mechanical methods, the prediction of physical properties in, for example, polymers, the elucidation of chemical processes such as the reaction pathway of enzymes, and the prediction of activity and toxicology of drugs in the pharmaceutical industry. There are a range of computational methods available for the simulations of these systems, and the choice of method depends strongly on the length and time scales of the problem (Figure 2.1) [91].

Quantum mechanical methods provide an accurate description of the forces, enabling the prediction of accurate energies and structures, but at great computational costs [92]. Molecular mechanics use a simplified model of the inter-atomic forces, providing a method to reach longer time and length scales. Coarse graining covers a range of methods designed to cover even longer time and length scales by grouping atoms together into beads, whilst continuum methods neglect the atomistic structure of the system altogether. In this chapter an overview of these methods is given, and an in depth
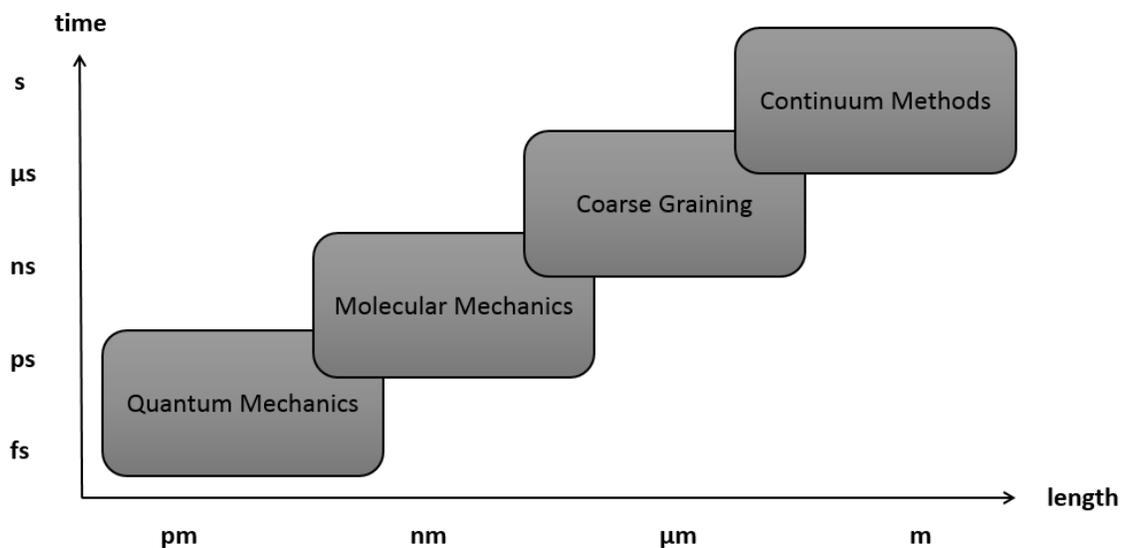
Figure 2.1: The hierarchy of the computational methods utilised at approximate time and length scales, ranging from quantum mechanical methods at small time and length scales to continuum methods for larger systems.

discussion regarding the methods used within this work (molecular dynamics) is provided.

Quantum mechanics

Within theoretical physics and chemistry, quantum mechanics (QM) is used to explicitly model electrons and derive structures and properties from the electron distribution. As the electron distribution is involved in chemical reactions such as the forming and breaking of bonds, these reactions can also be studied using QM methods. The electronic state of a system for a given arrangement of nuclei can be described by the electronic Schrödinger equation $\mathcal{H}\Psi = E\Psi$, where $\mathcal{H}$ is the Hamiltonian operator, Ψ the wavefunction describing the electrons, and E is the corresponding energy. The Hamiltonian operator $\mathcal{H}$ can be written as Equation 2.1:

$$\mathcal{H} = \frac{-\hbar^2}{2m}\nabla^2 + U(r) \tag{2.1}$$

where $\hbar$ is the reduced Planck constant, m is the mass of an electron, ∇^2 is the sum of the second derivatives with respect to the electron positions and $U(r)$ is the potential energy. The Schrödinger equation is a second order partial differential equation for which the wave function Ψ can only be solved analytically for a few simple systems (particle in a box, harmonic oscillator, hydrogen atom etc.). For poly electronic systems the Schrödinger equation cannot be solved exactly, but can be approximated. Within these systems one must also include inter-electron interactions. However, the positions of the electrons depend on the wave function, for which the Hamiltonian is needed, and the Hamiltonian depends on the positions of the electrons. To solve this problem, one can use an iterative self consistent field approach such as the well-known Hartree-Fock (HF) method. Within the HF approximation, the Schrödinger equation for an electron is solved by assuming that a given electron is moving in a field created by the nuclei and other electrons. An alternative method that can be used to investigate the electronic structure is Density Functional Theory (DFT). DFT is based on the Hohenberg-Kohn theorem that states that the ground state properties of a system are determined by the electron density of the system. By using a functional of the electron density the electronic structure and properties can be approximated [93]. Many semi-empirical methods exist that use simplifications and parameters to gain computational speed-ups with these calculations, such as the AM1 method used within this work to calculate the atomic charges of organic molecules (see Appendix B for further details). As QM methods are generally not used within this work, they will not be discussed further.

Classical methods

Due to the large number of particles involved in QM calculations, these methods are extremely computationally expensive and not suitable for the simulation of larger systems. Using the Born-Oppenheimer approximation, the motion of the nuclei and electrons can be separated and the total energy of the system can be written in a functional form

of an inter-atomic potential $U(r^N)$. Such an inter-atomic potential can be used to predict atomic structures by minimising the energy of the system. Both Monte Carlo (MC) and MD methods can be utilised to obtain configurational properties of the system of interest. MC methods can be used when only static configurational properties are required (such as equilibrium thermodynamic properties). With MD methods, the inter-atomic potentials can be used in combination with Newton's laws of motion to create time dependent trajectories of a system to obtain dynamic properties of a system, such as transport coefficients and time correlation functions. The MD methods are used throughout this work and are described in more detail in Sections 2.3 and 2.4.

Coarse graining

Coarse grained models have been developed to extend the system size and simulation time of classical models. Many coarse graining methods, in particular those involved with biomolecular simulations, group atoms together as pseudo-atoms or beads. This minimises both the number of particles in a system and the number of interactions that have to be calculated, and increases the size of time steps that can be used during simulations. The parameters for these coarse graining methods can be fitted to experimental data or to atomistic scale simulation data, an approach that is increasingly taken. Using coarse graining methods, the simulation of large systems and long time scales is made possible. However, this achievement of longer time scales and larger simulations has the drawback of the inevitable loss of detailed information due to the lower resolution of the system.

2.2 Force fields

2.2.1 Potentials

The forces within a system, and therefore all the interactions between all the particles, can be calculated via quantum or classical mechanical methods. In general, if the de Broglie wavelength of a particle in the system is larger than the size of the system, quantum mechanical methods should be used to calculate the forces within the system. If the de Broglie wavelength is smaller than the size of the system however, the wave nature of the particles in question is negligible, and classical mechanics can be utilised. Due to the large number of particles in MD simulations, the computationally expensive QM methods to calculate the forces are often replaced by the less computationally expensive classical methods. The Born-Oppenheimer approximation can be used to separate the total wave function into a nuclear and electronic part (Equation 2.2):

$$\Psi_{(nuclei,electrons)} = \Psi_{(nuclei)}\Psi_{(electrons)} \tag{2.2}$$

This separation is possible due to the mass disparity between the electrons and nuclei: as the nuclei have a large mass compared to the electrons, they can be considered as stationary particles. As the fast electrons can rapidly adjust to changes in nuclei positions, the electron wave function and energy are functions of the nuclei positions, and thus the total energy of the system can be described in terms of the nuclei positions only. In order to speed up calculations, the positions of and the interactions between the nuclei can be described by a inter-atomic potential $U(r^N)$, or force field: a set of simplistic mathematical models that have been parametrised using QM calculations and/or experimental data. The interactions described through such a force field are the inter-molecular potentials, encompassing van der Waals forces and Coulomb interactions, and intra-molecular potentials, which include two-, three- and four-body terms.

2.2.2 Inter-molecular potentials

van der Waals

The van der Waals pair-wise interactions between particles can be described by a number of mathematical models. The Lennard-Jones 12-6 potential is a commonly used potential which describes both the repulsive and attractive interactions. The short range repulsive term arises from the fact that two electrons are prohibited to occupy the same quantum state (Pauli exclusion principle). The long range attractive term is due to the attraction between (induced) dipoles in the system. Within the Lennard-Jones potential the attractive and repulsive behaviours vary with the separation of the particles and are described as r_{ij}^{-6} and r_{ij}^{-12} respectively, where r_{ij} is the separation between particles i and j. The physical basis for the attractive term comes from the dipole-dipole interaction, which goes as $1/r^6$. The repulsive term (r_{ij}^{-12}) was chosen arbitrarily for computational convenience as it is a square of the already calculated attractive term, and an exponential form would be more appropriate than the current term. The combination of these components describe the pair-wise energy between the particles, and the Lennard-Jones potential contains two parameters that can be adjusted: the depth of the energy well (ϵ_{ij}) and the equilibrium separation distance at which the force is zero (r_{ij}^{min}). The potential is most often written in the form of Equation 2.3:

$$U(r_{ij}) = \epsilon_{ij} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{6} \right] \tag{2.3}$$

This expression can be re-written in a simplistic AB-form (Equation 2.4), where $A = 4\epsilon\sigma^{12}$ and $B = 4\epsilon\sigma^6$:

$$U(r_{ij}) = \left(\frac{A}{r_{ij}^{12}} \right) - \left(\frac{B}{r_{ij}^6} \right) \tag{2.4}$$

Alternatively, a Buckingham potential can be used which takes the form of Equation 2.5 and has three constants that can be adjusted: A, ρ and C:

$$U(r_{ij}) = A\exp(-\frac{r_{ij}}{\rho}) - \frac{C}{r_{ij}^6} \tag{2.5}$$

Note that at small r_{ij} values the Lennard-Jones potential rises to infinity and becomes highly repulsive. The Buckingham potential becomes highly attractive at small r_{ij} values as the attraction term goes to negative infinity. As such, care must be taken that the inter-atomic distances do not become too small, as this can result in particles clumping together. In both the Lennard-Jones and Buckingham potentials the inter-action energy approaches zero at high separation values. In order to avoid calculating these interactions at high separations, these potentials can be cut-off at a specific distance. To avoid any discontinuity, a smooth taper function can be used to reduce the potentials to zero, such as the use of a scaling factor applied to the region around the cut-off.

Coulomb electrostatic

The charge distribution of a molecule can be modelled in a number of ways and a common approach is the placement of partial point charges throughout the molecule. The total electrostatic energy of a set of charges can be described via a Coulomb sum with the functional form of Equation 2.6:

$$U(r_{ij}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{r_{ij}} \tag{2.6}$$

where N is the number of point charges in the system, q_i and q_j are the particle charges and r_{ij} is the separation between the particles. Note that these electrostatic interactions are given in terms of reduced electrostatic units, which correspond to being divided by a factor of $1/4\pi\epsilon_0$, where ϵ_0 is the vacuum permittivity constant. The problem with

calculating this Coulomb sum is that it is conditionally convergent: the interaction energy decays with an increasing separation, but the number of interactions increases with the separation.
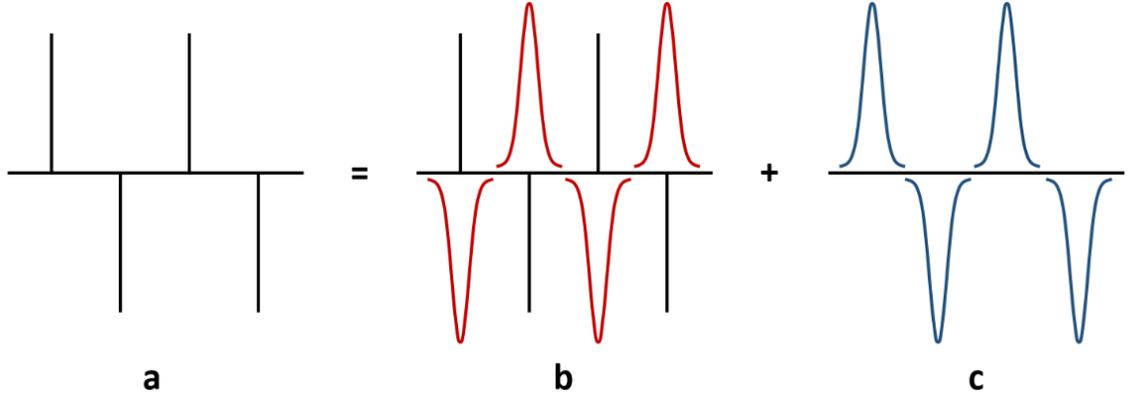
Figure 2.2: Representation of the Ewald summation in which the point charges (a) are screened by a Gaussian distribution summed in real space (b), which are then cancelled out by a Gaussian distribution which is summed in reciprocal space (c). It is also necessary to remove the interactions between Gaussians and point charges on the same site.

The most common method to overcome this problem is by using the Ewald summation [94]. With the Ewald method, the point charges of the system are screened by a Gaussian charge distribution with the opposite sign (Figure 2.2b). A second Gaussians distribution is added (Figure 2.2c) to cancel out the first set of Gaussians. The summation corresponding to Figure 2.2b can easily be calculated from Equation 2.7:

$$U_{(\text{real})} = \frac{1}{2} \sum_{i \neq j}^{N} \frac{q_i q_j}{r_{ij}} \text{erfc}(\eta^{\frac{1}{2}} r_{ij}) \tag{2.7}$$

where N is the number of point charges in the system, $\text{erfc}(x)$ is the complementary error function given as Equation 2.8:

$$\text{erfc}(x) = \left(\frac{2}{\sqrt{\pi}} \right) \int_x^\infty \exp(-u^2) du \tag{2.8}$$

and the width of the Gaussians is given by $\sqrt{2/\eta}$. This summation is calculated in

real space and converges rapidly. The summation of the second set of Gaussians (Figure 2.2c) converges rapidly in reciprocal space via Equation 2.9:

$$U_{(\text{reciprocal})} = \frac{1}{2L^3} \sum_{k \neq 0} \frac{4\pi}{k^2} q_i q_j \exp(ik \cdot r_{ij}) \exp\left(\frac{-k^2}{4\eta}\right) \tag{2.9}$$

where L is the side of the cubic system and k is a reciprocal lattice vector, given as $k = (2\pi/L)n$, with n the lattice repeat vector, describing the periodic boundary conditions of the system. A self energy correction term between a charge and a Gaussian on the same site (Equation 2.10) then needs to be subtracted from the system. The total electrostatic energy of the system can then be calculated from these three terms (Equation 2.11). The width of the added Gaussians is defined by the variable η and the convergence of the real and reciprocal terms of Equation 2.11 depend on this parameter, chosen by the user. A small value of η ensures a quick convergence of the reciprocal term, whereas the real term will converge rapidly with a large value of η. A balance between the two can be found experimentally by simulating a system using a range of Gaussian widths and plotting the coulombic energy. A correct convergence of both the reciprocal and real terms will result in a plateau of the coulombic energy.

$$U_{(\text{self})} = \left(\frac{\eta}{\pi}\right)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2 \tag{2.10}$$

$$U_{(\text{electrostatic})} = U_{(\text{real})} + U_{(\text{reciprocal})} - U_{(\text{self})} \tag{2.11}$$

2.2.3   Intra-molecular potentials

Two body terms - bonds

As described in Section 2.2.1, the Born-Oppenheimer approximation can be used to separate the nuclei and electronic contributions, and thus the total energy can be decoupled

into electronic, vibrational and rotational contributions. The latter two terms can be described by a intra-molecular potential: a simplistic mathematical model parametrised using QM calculations and/or experimental data. The intra-molecular interactions between two particles, or bond stretching, is a two-body term that can be described in a number of ways, most commonly described as a harmonic spring. The functional form of a simple harmonic spring used in this work is shown in Equation 2.12:

$$U(r_{ij}) = \frac{1}{2}k(r_{ij} - r_0)^2 \tag{2.12}$$

where k is the bond stretching force constant of the spring and r_0 is the reference or natural bond length. This potential gives a fair representation of the bond behaviour close to the equilibrium distances, but as $r_{ij} - r_0$ increases, $U(r_{ij})$ tends towards infinity. As such, the harmonic potential is unbreakable and gives a poor representation far from the equilibrium distances. More sophisticated potentials such as the Morse potential and reactive potentials, in which the breaking and forming of bonds is possible, can be used. However, as in many biomolecular systems the bonds do not deviate far from their equilibrium values, a simple harmonic spring can be used to give a good approximation of the bond length. The stiffness of the spring also has an impact on the timestep used in MD simulations (see Section 2.3), where a stiffer spring requires a lower timestep in order to maintain a stable simulation.

Three body terms - angles

The angles between three particles can be described by a three-body potential and, as with bond stretching, this angle bending (Figure 2.3a) is often described using a simple harmonic potential (Equation 2.13):

$$U(\theta_{ijk}) = \frac{k_\theta}{2}(\theta_{ijk} - \theta_0)^2 \tag{2.13}$$

where θ_{ijk} is the angle between the three particles and θ_0 is the reference angle.
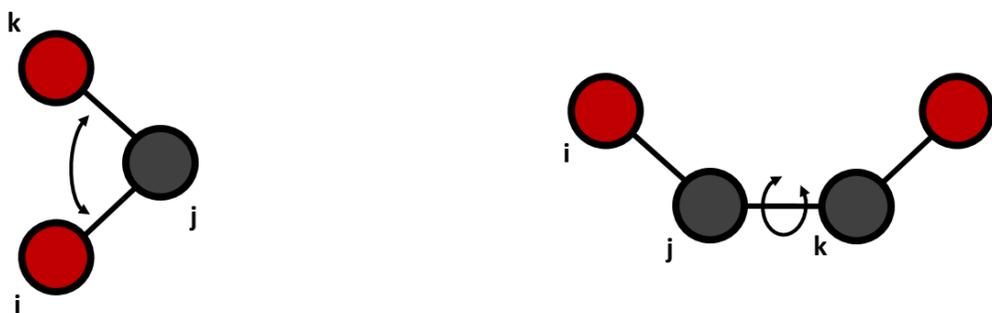
(a) Three-body angle term.

(b) Four-body dihedrals term.

Figure 2.3: Diagrams of the three- and four-body terms of the intra-molecular forces described by force fields.

Four body terms - dihedrals and out of plane

The dihedral or torsion angle between four bonded particles describes the angle between two planes (Figure 2.3b). This dihedral can be described by a potential with a cosine functional form (Equation 2.14):

$$U(\phi_{ijkl}) = A[1 + \cos(m\phi_{ijkl} - \delta)] \qquad (2.14)$$

where A is a constant responsible for the height of the energy barriers, m determines the multiplicity of the dihedral (number of minima) and δ determines the energy minimum position of the rotation. In addition to dihedral angles, four-body terms can describe the out-of-plane bending motions of particles and this motion can be described by r_{i_jkl}, where r_{i_jkl} is the distance of particle i from the plane described by particles j, k and l. For example, the four atoms in a carbonate ion (CO_3) are bonded together in a planar orientation. In order to force the carbonate ion to remain planar, the out-of-plane movement of the carbon atom needs to be restricted, which can be accomplished with a potential such as Equation 2.15:

$$U(r_{i\_jkl}) = Ar_{i\_jkl}^2 + Br_{i\_jkl}^4 \qquad (2.15)$$

2.2.4   Force fields used in this work

As described above, the interactions between atoms can be described by the inter-atomic potential $U(r^N)$, or force field. These force fields are simplistic models that are empirically parametrised using experimental data and/or QM calculations. It must be noted that the quality of the force fields used within molecular modelling dictates the quality of the results, and as such, these force fields must be properly constructed and tested. For the modelling in this work, force fields for calcium carbonate, water and organic molecules have been taken from the literature and are described below.

Calcium carbonate

There has been a considerable effort in the generation of suitable potentials for calcium carbonate in the last few decades, and Raiteri *et al.* [95] give a substantial overview of these efforts, of which a few are described below.

Several $CaCO_3$ potentials were developed in the early '90s, [96–98], of which the Pavese potential [98] has seen the most use. This potential describes the carbonate ion as a molecular group with harmonic angles and improper torsion terms and uses a shell model for the oxygen on the carbonate group. The improper torsion term is mainly used to maintain a planar configuration of the carbonate ion. In a revised version [99] a Morse potential was added to the carbon to oxygen interaction. Fisler *et al.* [100] subsequently made revisions to the Pavese potential by changing the carbonate group, and De Leeuw and Parker [101] combined a shell model of water with the Pavese potential to model the hydrated surfaces of $CaCO_3$. Bruneval *et al.* [102] later combined the potentials by Fisler *et al.* and De Leeuw and Parker to create a potential with polarisable descriptors for the water. Freeman *et al.* [103] combined parts of the Pavese potential and the potential by De Leeuw and Parker with a TIP3P water model [104] to create a potential to use within biomineralisation simulations. In 2010 Raiteri *et al.* [95, 105] fitted a new potential to experimental dissolution enthalpies of the ions in water, with the purpose

of developing a potential that could accurately describe the thermodynamic energies of $CaCO_3$ and the energy difference in the calcite-aragonite phase transition.

The Raiteri potential is extensively used throughout the biomineralisation community and is used for the MD simulations within this work. The parameters for this potential are shown in Table 2.1.

Water

There are a range of simple water models available to use within MD simulations that have different methods of describing the geometry of the the water molecule and the inter- and intra-molecular interactions. The SPC (Simple Point Charge) [106] and TIP3P (Transferable Interaction Potential 3P) [104] are the most common water models that contain three sites for the electrostatic interactions (Figure 2.4). Refined four and five site models (TIP4P and TIP5P) are also available, but are computationally more expensive. Within this work the TIP3P water model is used, as it gives structural and thermodynamic properties in close agreement with experimental values at room temperature [104] and is computationally inexpensive to simulate. Additionally, most organic force fields have been parametrised using TIP3P. The potential parameters for TIP3P are described in Table 2.2.
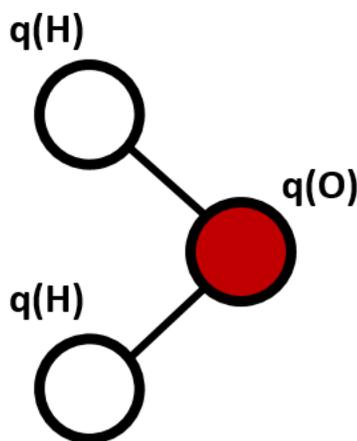
Figure 2.4: Model of a three site water molecule, such as TIP3P.

Charges			
Atom	Charge (au)	Atom	Charge (au)
Ca	+2.000	C(carbonate)	+1.123
O(carbonate)	-1.041	Ow(water)	-0.834
Hw(water)	+0.417		

Bonds (Equation 2.12)		
Interaction	k (kJ mol^{-1} Å^{-2})	r_0 (Å)
C-O (carbonate)	3463.8	1.313

Angles (Equation 2.13)		
Interaction	k (kJ mol^{-1} rad^{-2})	θ_0 (deg)
O-C-O (carbonate)	1160	120.0

Out-of-plane term (Equation 2.15)		
Interaction	A (kJ mol^{-1} rad^{-2})	B (deg)
C-O/O/O (carbonate)	2006.508	360.0

Lennard-Jones potential (Equation 2.3)		
Interaction	ϵ (kJ mol^{-1})	σ (Å)
Ca-Ow(water)	0.09166	3.35

Buckingham potential (Equation 2.5)			
Interaction	A (kJ mol^{-1})	ρ (Å)	C (kJ mol^{-1} Å^{-6})
Ca-O(carbonate)	305050.84	0.27151	0.0
Ca-C(carbonate)	1157823690	0.12	0.0
O(carbonate)-O(carbonate)	6159641.3	0.19891	2691.844
O(carbonate)-Ow(water)	1209391.23	0.2020	1166.507
O(carbonate)-Hw(water)	38237.1	0.2170	0.0

Table 2.1: CaCO$_3$ inter- and intra-molecular terms from the Raiteri potential [105].

Charges

Atom	Charge (au)	Atom	Charge (au)
Ow	-0.834	Hw	+0.417

Bonds (Equation 2.12)

Interaction	k (kJ mol^{-1} Å^{-2})	r_0 (Å)
Ow-Hw	3768.12	0.9572

Angles (Equation 2.13)

Interaction	k (kJ mol^{-1} rad^{-2})	θ_0 (deg)
Hw-Ow-Hw	104.52	460.52

12-6 potential (Equation 2.4)

Interaction	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
Ow-Ow	2436397.374231	2491.2067290

Table 2.2: Inter- and intra-molecular terms for the TIP3P water model [104].

Biomolecules

Biomolecular MD simulations have been performed for decades and a large number of force fields are available to model these systems, such as the CHARMM (Chemistry at Harvard Molecular Mechanics) [107], GROMOS (Groningen Molecular Simulation) [108], OPLS (Optimized Potentials for Liquid Simulations) [109], COMPASS (Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies) [110] and AMBER (Assisted Model Building with Energy Refinement) [111] force fields. All these force fields were developed in the 1980s or 1990s and have had a substantial number of revisions over the years, and comparison studies [112, 113] have suggested that the differences between these force fields is minimal. The AMBER force field is one of the most used force fields for organic systems, in particular for system that include proteins, peptides and nucleic acids. Within this work the AMBER *ff12SB* force field is used, which is parametrized using the TIP3P water model. The calculation of the atomic charge distribution was performed using the AM1 method, as described in Appendix B.

Biomineral cross-terms

The potentials used within biomineralisation simulations must include terms that describe the interactions between the water/organic phase and the mineral phase, but existing potentials that describe these two phases are quite dissimilar. Mineral potentials often describe the crystalline phase as ionic solids with large charges, whereas potentials for organic molecules emphasise the intra-molecular bonding terms and use smaller atomic charges. In order to describe the interactions between these two phases appropriately, the two sets of potentials need to be combined to create the cross-terms between the mineral and organic phase. One of the methods that could be used to combine two sets of potentials is the Lorentz-Berthelot mixing rules to obtain Lennard-Jones cross potentials [114], as given in Equations 2.16 and 2.17:

$$\sigma_{AB} = \frac{\sigma_A + \sigma_B}{2} \tag{2.16}$$

$$\epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B} \tag{2.17}$$

The Lorentz-Berthelot mixing rules describe the Lennard-Jones interactions between the mineral and organic phases, but due to the inconsistencies in the atomic charges between the two phases, the binding energies of organic molecules at the surface of the mineral can be largely overestimated. Another technique to combine two sets of potentials is by using the Schröder method [115], and Freeman *et al.* [103] described a generalised methodology to generate a robust and reliable potential set for biomineral systems based on the Schröder method without the need to fit a whole new potential to new systems.

The methodology described by Freeman *et al.* is a stepwise procedure to find a short range potential that is consistent with the reduced electrostatic interactions between the mineral ions and the organic molecules. Firstly, the scaled charges in a chosen

mineral structure are determined. The scaled charges need to fulfil the conditions set out in Equations 2.18 and 2.19:

$$q_{Anew}q_{Bnew} = q_{Amineral}q_{Borganic} \tag{2.18}$$

$$\sum_I q_{Inew} = 0 \tag{2.19}$$

where A and B are atoms in the original mineral and organic phase respectively, q_{Anew} and q_{Bnew} are the new scaled charges for the new chosen mineral structure used to fit the short range potential and I is any atom within the new system. The new charges q_{Inew} can be scaled with respect to the original mineral, as shown in Equation 2.20:

$$q_I = \alpha_I q_{Bnew} \tag{2.20}$$

where α_I is the scaling factor derived from the original mineral to be used to obtain the new short range potential (Equation 2.21):

$$\alpha_I = \frac{q_{Imineral}}{q_{Bmineral}} \tag{2.21}$$

Combining these equations with the original condition that the combined charges of the new system must equal the combined charges of the old system (Equation 2.18), the charges q_{Bnew} can be deduced from Equation 2.22, and by using Equation 2.18 all other charges q_I for the new system can be determined.

$$\alpha_A q_{Bnew}^2 = q_{Amineral}q_{Borganic} \tag{2.22}$$

The second step in the methodology is the refitting of the inter-atomic mineral potentials

with the new, scaled, charges. As the charges of the mineral are scaled down, the repulsive term of the potential has to be reduced such that the structure of the mineral kept intact. These newly fitted inter-atomic potentials are the cross-terms between the organic and mineral phases, and for each atom in an organic potential that is also present in the mineral phase, these cross-terms must be included. As the cross-terms used depend on the organic potentials used, they must be refitted when changing the organic force field. Note that using this method the mineral potential remains unchanged and is only used to generate the cross-terms. The results obtained from simulations that use these cross-terms have been compared with DFT calculations, and in the case of the interactions of small organic molecules with calcite surfaces they produce adsorption energies that agree well with the DFT calculations [83, 103]. All cross-terms used throughout this work are described in their respective chapters.

2.3   Molecular dynamics

As mentioned in 2.2.1 using the Born-Oppenheimer approach, an inter-atomic potential $U(r^N)$ can be used to describe the structure and energy of a system. Using this in combination with Newtons laws of motion, the time dependent interactions between particles can be determined. These molecular dynamics simulations involve the calculation of the movement of atoms over a period of time. By solving Newtons motion equations of a set of particles, the interactions between these particles can be calculated and both thermodynamic and dynamic properties of the system can be determined.

2.3.1   Newton's laws of motion

The positions and momenta of the particles in the system are governed by the interactions between the particles. The movement of the particles in a system are determined by Newton's laws of motion:

1. Particles continue to move at constant velocity unless a force acts upon them.

2. The force on a particle is equal to the rate of change of momentum.

3. For each action there is an equal and opposite reaction.

From Newton's second law, the interactions between particles can be described as Equation 2.23:

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \tag{2.23}$$

where F_i is the the force exerted on particle i and m_i, a_i and r_i are the mass, acceleration and position of the particle respectively at time t. The force exerted on the particles is taken from the gradient of the potential energy of the system (Equation 2.24), where the potential energy U is differentiated with respect to r_i, and is the result of the interactions of all particles in the system. As the motions of the particles in question are coupled to the interactions between all particles in the system, it is impossible to analytically solve these equations of motion for all but the simplest examples. The equations of motion within MD simulations can be solved numerically using a finite difference method.

$$F_i = -\nabla_{r_i} U \tag{2.24}$$

2.3.2 Finite difference method

The finite difference method is a method to integrate Newton's laws of motion to obtain an evolving trajectory of the particles in the system. There are many integration schemes available, of which the time-reversible and computationally inexpensive *Verlet* algorithm [116] is the most popular. This algorithm assumes that the positions r of the particles at time $t+\delta t$ can be approximated by a Taylor expansion (Equation 2.25):

$$r(t + \delta t) = r(t) + \delta t v(t) + \tfrac{1}{2}\delta t^2 a(t) + \tfrac{1}{6}\delta t^3 b(t) + O(\delta t^4) \tag{2.25}$$

where $O(\delta t^4)$ is the error associated to the truncation of the Taylor expansion. Similarly, the position of the particles at time $t - \delta t$ can be approximated by Equation 2.26:

$$r(t - \delta t) = r(t) - \delta t v(t) + \tfrac{1}{2}\delta t^2 a(t) - \tfrac{1}{6}\delta t^3 b(t) + O(\delta t^4) \tag{2.26}$$

A summation of Equations 2.25 and 2.26 gives

$$r(t + \delta t) + r(t - \delta t) = 2r(t) + a(t)\delta t^2 + O(\delta t^4) \tag{2.27}$$

and rearrangement of this equations gives

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \tag{2.28}$$

where the positions of the particles at time $t + \delta t$ can be calculated without the velocities. The velocities of the particles can subsequently be determined from the difference in positions (Equation 2.29):

$$v(t) = [r(t + \delta t) - r(t - \delta t)]/2\delta t \tag{2.29}$$

Note that the positions of the particles contain a truncation error in the order of δt^4, and the error associated with the velocities is in the order of δt^2. Several variations of the *Verlet* algorithm have been developed such as the *Velocity Verlet* and the *leap-frog* algorithm [117]. The leap-frog algorithm calculates the velocities at time $t + \tfrac{1}{2}\delta t$ (Equation 2.30) and subsequently the positions at $t + \delta t$ (Equation 2.31), such that the two calculations "leap" over each other, and both the positions and velocities are

43

calculated correct to the second order.

$$v(t + \tfrac{1}{2}\delta t) = v(t - \tfrac{1}{2}\delta t) + a(t)\delta t \tag{2.30}$$

$$r(t + \delta t) = r(t) + v(t + \tfrac{1}{2}\delta t)\delta t \tag{2.31}$$

The velocities at time t can be calculated as the average of the velocities at $t + \tfrac{1}{2}\delta t$ and $t - \tfrac{1}{2}\delta t$ (Equation 2.32):

$$v(t) = \tfrac{1}{2}[v(t + \tfrac{1}{2}\delta t) + v(t - \tfrac{1}{2}\delta t)] \tag{2.32}$$

Other, higher order integration methods such as the Runge-Kutta and Gear predictor-corrector methods are also available and details of these can be found elsewhere [114, 118]. The length of the time step δt is the bottleneck of simulations and is responsible for the computational cost of the simulations. Using a short time step will not allow the simulation to cover much of the phase space throughout the simulations. A large time step will cover more of the phase space within simulations but are more prone to develop instabilities within the integration algorithm due to the increased truncation error associated with the larger time step.

2.3.3 Running molecular dynamics simulations

Within MD simulations the initial state of a system contains the positions and momenta of the particles. The initial momenta of particles in MD simulations are typically assigned randomly from a Maxwell-Boltzmann distribution, whilst ensuring that the total momenta of the system is equal to zero. With the starting positions and momenta an integration algorithm can be implemented to propagate the system. The starting positions of the particles are unlikely to be in a low energy region since, for computational

convenience, they are usually setup in a highly special configuration. Therefore an equilibrium period is needed in order to relax the system. To check whether a simulation has reached an equilibrium minimum, the thermodynamic variables of the system can be tracked. These variables should fluctuate around the mean but not show any systematic drift.

Periodic boundary conditions

In MD simulations, a large number of particles within the simulation box are in contact with the boundaries of the system. If the purpose of the simulation is to calculate the properties of a bulk system, the behaviour of those particles interacting with the boundaries must be excluded. As interactions between particles and the boundaries can extend up to 10 Å into the simulation box, the number of particles that would have to be excluded is extensive. Using periodic boundary conditions (PBC) the particles can be simulated as though no boundaries are present, effectively calculating bulk properties.

The basis of PBC is to place replicas of the simulation box around the original simulation to give a periodic array, as indicated by the two-dimensional example in Figure 2.5. The movements of all particles within the original simulation box are replicated throughout the neighbouring boxes, and a particle i interacts not only with the other particles within its own simulation box, but with all particles in neighbouring replicas, thus providing a bulk simulation. Should particle i move through the boundary into a neighbouring box, an image particle i' will enter the box from the opposite side, keeping the number of particles constant. This method enables the whole system to be treated as a pseudo-infinite system.

Minimum image convention and potential cut off

Within MD simulations the most time consuming calculations are the non-bonded inter-actions between particles. Whereas the number of bonded interactions is proportional to the number of particles N in the system, the number of non-bonded (pair) inter-actions increases as N^2. As the value of the non-bonded potentials used for these interactions drastically falls at longer distances, the interactions at larger distances can be discarded in order to speed up the simulations. The minimum image convention can be applied, in which a particle interacts only with at most one replica of every other particle in the system. A potential cut-off can also be applied, in which any interactions outside the cut-off distance are discarded. In simulations using PBC this cut-off must be chosen such that there are no interactions between a particle and its replica, and as such the cut-off distance should be less than half the simulation cell when simulating a cubic box. To avoid any discontinuity of these potentials at the cut-off distance, a smooth taper function can be used to reduce the potentials to zero.

2.3.4 Analysis

The trajectory of the MD simulations can be analysed in a number of ways. Firstly, the trajectory can be visually analysed to explore the dynamic behaviour of the particles in a time dependent manner. Moreover, a range of thermodynamic and structural properties can be calculated from the trajectories.

Energies

Thermodynamic properties such as the energy, heat capacity and temperature can easily be obtained from MD simulations. The total potential energy of the system is evaluated at each time step in the simulation as the sum of all values evaluated from the inter- and intra-molecular potentials. Additionally, the kinetic energy of the system at time
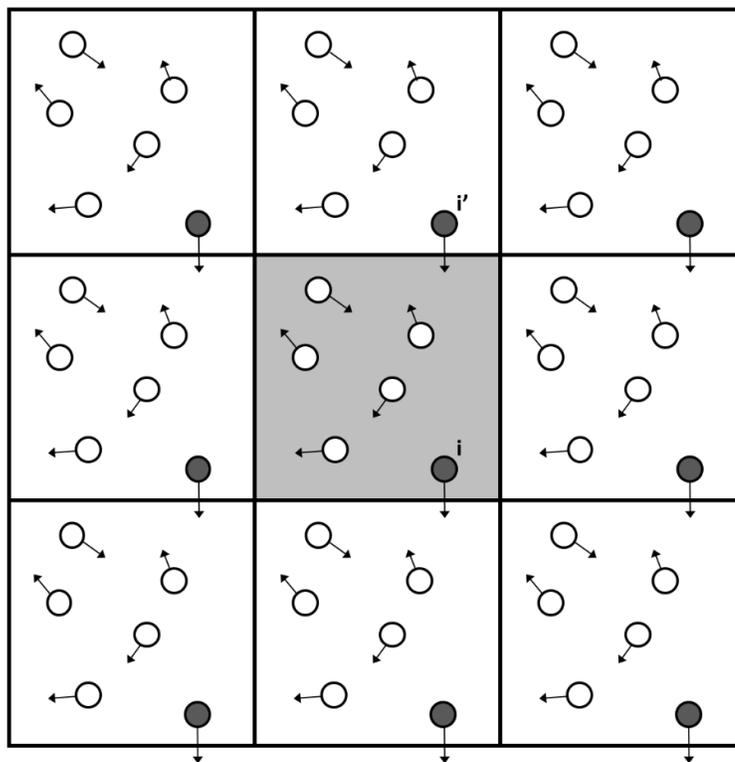
Figure 2.5: Two dimensional example of the use of periodic boundary conditions, where replicas are placed around the original simulation cell.

t can be directly calculated from the velocities of the particles via Equation 2.33:

$$U_{kinetic}(t) = \frac{1}{2}\sum_i^N m_i v_i^2(t) \qquad (2.33)$$

The changes in the energy of the system can be analysed over the length of the simulation and fluctuations in these energies can give indications of changes throughout in the system.

The energies of the system can also be used to calculate the thermodynamic differences between two configurations (e.g. binding energies, conformational relaxation etc.) by subtracting the two mean energies from one another. The error of the difference between these two configurations can be calculated using the standard error (SE) of the difference between the means of two samples, as described in Equation 2.34:

$$SE = \sqrt{[(\frac{sd_a^2}{n_a}) + (\frac{sd_b^2}{n_b})]} \qquad (2.34)$$

where sd_a and sd_b are the standard deviations for the data of configuration a and b respectively, and n_a and n_b are the population size of a and b respectively.

Radial distribution function

Another tool that can be utilised to analyse structural features of a system is the radial distribution function (RDF), or pair correlation function. The RDF, or $g(r)$, describes the probability of finding a particle at distance r from another reference particle relative to the probability for a random distribution. The RDF calculates the number of particles found between a distance r and $r + \delta r$ from the reference particle. As this is described in a spherical region around the reference particle, the form of the RDF is (Equation 2.35):

$$g(r) = \frac{V}{4\pi r^2 \delta r N} \langle \sum_{i=1}^{N} \sum_{j \neq i}^{N} \delta(r - r_{ij}) \rangle \qquad (2.35)$$

where V is the volume of the simulation, $4\pi r^2 \delta r$ is the volume of a spherical shell of width δr, and N is the number of particles. The $g(r)$ is dimensionless and the values are averaged over the entire simulation. In a perfect crystal, the RDF has a number of sharp, intense peaks corresponding to the lattice equilibrium positions of the particles. The movement of the particles away from these equilibrium positions will result in a broadening of the peaks in the RDF.

Diffusion coefficient

The translational movement of particles throughout a simulation can be described by the mean-square displacement of the particles. This displacement can be calculated

as an ensemble average over all the particles, relative to the starting positions and is related to the particle diffusion coefficient, D, in Equation 2.36:

$$D = \frac{1}{6Nt} \lim_{t \to \infty} \sum_{i=1}^{N} \langle |r_i(t_0 + t) - r_i(t_0)|^2 \rangle \tag{2.36}$$

where t_0 is the initial time origin and $r_i(t_0 + t)$ is the position of particle r_i at time t. The diffusion coefficient is not used within this work and is only included for completeness. Details of the practicalities of calculating D can be found in Allen and Tildesley [114].

2.4 Statistical mechanics

The exact state of a system (the microstate) can be described by the three positional and three momentum values of each particle of that system, and for a system containing N particles, the microstate of the system can thus be described with $6N$ values. The $6N$-dimensional space that can be described by a combination of these $6N$ values make up the *phase space.*

All possible microstates for a system consistent with a given thermodynamic state defined by macroscopic variables such as the pressure, temperature and volume can be described as an ensemble. MD simulations are calculations that generate a series of time dependent microstates within this phase space, a trajectory that describes the dynamic changes in the $6N$ variables. MD simulations are a deterministic method, such that future microstates (future $6N$ variables) are determined by the current microstate (current $6N$ variables).

Properties such as the energy, heat capacity and pressure of a system can be calculated from the positions and momentum of all particles in the system. Instantaneous values for these properties can be calculated from a microstate, and the time average ($A_{average}$)

values can be calculated from the system trajectory (Equation 2.37):

$$A_{average} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(r^N(t), p^N(t)) dt \tag{2.37}$$

where A is the property of interest and r^N and p^N are the positions and momentum values for particle N respectively. The property values can also be calculated over all possible microstates for the system (ensemble), given Equation 2.38:

$$\langle A \rangle = \iint A(r^N, p^N) \rho(r^N, p^N) dr^N dp^N \tag{2.38}$$

where $\langle A \rangle$ is the ensemble average and $\rho(r^N, p^N)$ is the probability density, i.e. the probability of finding a specific microstate with p^N and r^N within the phase space. An essential assumption in using MD simulations to construct these averages is the *ergodic hypothesis* which states that averages over time obtained from MD simulations can be used to calculate the ensemble average - or that Equation 2.37 can be used to calculate Equation 2.38.

2.4.1 Ensembles

The probability density for these ensemble averages depends on the ensemble that is simulated. MD simulations can sample from a range of different ensembles, which are described below.

Microcanonical (NVE) ensemble

Molecular dynamics simulations are most naturally performed within a subset of a microcanonical ensemble, an isolated system where Newton's laws of motion conserve linear momentum of particles. In this ensemble the number of particles (N), volume (V) and energy (E) are held constant, and it is described as an isolated system: no

exchange of energy or matter with the outside is possible. As the energy is constant, the probability of finding a microstate depends on the energy of the system and can be written as Equation 2.39 over the energy range $\delta[E - \mathcal{H}(r^N, p^N)]$:

$$\rho(r^N, p^N) = \begin{cases} \dfrac{1}{\Omega_{NVE}} & \text{for } E < \mathcal{H}(r^N, p^N) < E + \delta E \\ 0 & \text{for other cases} \end{cases} \tag{2.39}$$

where Ω_{NVE} is the partition function for the NVE ensemble, shown in Equation 2.40 and $\mathcal{H}$ is the Hamiltonian function, the sum of the kinetic and potential energy of the microstate. The factor $N!$ arises to remove indistinguishable particle count from the partition function. The h^{3N} factor represents the volume of phase space occupied by the microstate and, from Heisenberg uncertainty principle, this is the smallest volume in phase space that can be specified.

$$\Omega_{NVE} = \frac{1}{N! h^{3N}} \iint dr^N dp^N \delta[E - \mathcal{H}(r^N, p^N)] \tag{2.40}$$

Canonical (NVT) ensemble

In the canonical ensemble (NVT) the number of particles (N), volume (V) and temperature (T) are kept constant. This ensemble is regarded as a closed system: it can exchange energy (heat) with its environment, but not matter. Therefore the energy of the system is allowed to fluctuate, and can exchange heat via a heat bath (thermostat) connected to the system. The probability density within an NVT ensemble is given by Equation 2.41:

$$\rho(r^N, p^N) = \frac{1}{Q_{NVT}} \exp[-\beta \mathcal{H}(r^N, p^N)] \tag{2.41}$$

where Q_{NVT} is the partition function, a normalisation of the accessible microstates at a given temperature, which is given by Equation 2.42. The variable β is defined as

Equation 2.43 in which k_B is the Boltzmann constant and T the temperature.

$$Q_{NVT} = \frac{1}{N!h^{3N}} \iint dr^N dp^N \exp[-\beta \mathcal{H}(r^N, p^N)] \tag{2.42}$$

$$\beta = \frac{1}{k_B T} \tag{2.43}$$

Isothermal-isobaric NPT ensemble

With a constant number of particles (N), constant pressure (P) and constant temperature (T), an isothermal-isobaric (NPT) ensemble is classed as a closed system. As with an NVT ensemble, there is an energy exchange with the outside but no exchange of matter. The energy and volume of the system are allowed to fluctuate, regulated via the connection to a thermostat and barostat. The probability density of the system can be written as Equation 2.44:

$$\rho(r^N, p^N, V) = \frac{1}{Q_{NPT}} \exp[-\beta\{\mathcal{H}(r^N, p^N) + PV\}] \tag{2.44}$$

where Q_{NPT} is the partition function of the NPT ensemble which is a transform of the NVT partition function with respect to the volume (Equation 2.45):

$$Q_{NPT} = \int_0^\infty dV Q_{NVT} \exp(-\beta PV) \tag{2.45}$$

2.4.2 Thermostats and barostats

Microcanonical ensembles are commonly used within MD simulations, where the number of particles, volume and energy are conserved. The temperature within such a system can thus fluctuate. In order to perform simulations whereby the temperature is fixed, i.e. a canonical NVT ensemble, a thermostat must be implemented to ensure

the temperature is kept constant. Within a system the temperature is related to the kinetic energy of the system following Equation 2.46:

$$\langle k \rangle_{NVT} = \tfrac{3}{2} N k_B T \tag{2.46}$$

where $\langle k \rangle_{NVT}$ is the time average kinetic energy of the NVT ensemble. As the temperature of this system is set by the kinetic energy of the particles, the temperature of these simulations can be controlled by scaling the velocities of the particles. The direct scaling of the velocities, however, does not generate a proper canonical ensemble. Instead, a heat-bath can be linked to the system as an external source of thermal energy. The Nosé-Hoover thermostat [119] considers the heat-bath to be an integral part of the system, following Equation 2.47:

$$\frac{\delta v_i(t)}{\delta t} = \frac{F_i(t)}{m_i} - \xi(t) v_i(t) \tag{2.47}$$

where v_i is the velocity, F_i the total force on the particles and ξ is the friction coefficient of the heat-bath. Integrating the equations of motion gives Equation 2.48:

$$v(t + \delta t) = v(t) + [\frac{F(t)}{m} - \xi(t)v(t)]\delta t \tag{2.48}$$

The Hamiltonian of the extended system given as Equation 2.49:

$$\mathcal{H}_{\text{(Nosé-Hoover)}} = \sum_{i=1}^{N} \frac{p_i^2}{2m_i s^2} + U(r^N) + \frac{\xi^2 Q}{2} + g k_B T \ln s \tag{2.49}$$

with N particles of mass m_i and positions p_i, Q the fictitious mass, s the number of degrees of freedom and ξ the friction coefficient of the heat-bath. Simulations of the system using the extended Hamiltonian and a properly chosen friction coefficient ξ return a proper canonical distribution of the particle coordinates and momentum.

A similar approach can be taken to keep the pressure of the system constant in isothermal-isobaric NPT ensembles. Here, a pressure-bath is added to extend the system, defined by a piston with fictitious mass W and friction coefficient η. The Hamiltonian as described in Equation 2.49 is extended by adding a term for the pressure $P_{ext}V$ and the pressure-bath $\frac{1}{2}(\eta^2 W)$. The fictitious mass of the of the heat-bath (Q) and pressure-bath (W) determine the coupling between the heat or pressure-bath and the simulation, and can be expressed in terms of a relaxation time. These relaxation times must be chosen to efficiently generate the proper ensemble. Within the simulations in this work, the relaxation times were chosen such that the system was equilibrating quickly, whilst ensuring that a constant temperature and pressure was maintained.

2.5   Biased potential methods

A common (and serious) problem that occurs in MD simulations is the difficulty of sampling the whole of the phase space during a simulation, and the presence of (multiple) local energy minima with large free energy barriers can prevent simulations from visiting all configurations. Additionally, rare events, for instance phase transitions and conformational changes such as protein folding, take place on time scales beyond the reach of MD simulations. Brute force methods can be employed for the latter problem, although these have the disadvantages of a lack of statistical averages, and are very computationally expensive. Methods such as coarse graining can also be employed, although these require extensive parametrization and offer no atomistic detail.

Several accelerated molecular dynamics methods have been developed to overcome the problems of local energy minima and appropriate phase space sampling. These methods include (but are not limited to) potential of mean force (PMF), umbrella sampling and metadynamics.

2.5.1 Potential of mean force and umbrella sampling

With PMF calculations, the simulations are confined to a reaction coordinate $R(x)$, and the free energy surface along this reaction coordinate is calculated. The free energy $F_R(R')$ for a particular point R' along the reaction coordinate is given by Equation 2.50:

$$F_R(R') = -k_B T \ln P_R(R') + constant \tag{2.50}$$

where $P_R(R')$ is the probability of finding the system at this reaction coordinate. The PMF method relies on a properly sampled system in which all of the phase space is visited. If the variation of the free energy along a reaction coordinate is more than a few $k_B T$, the simulation will not sample enough of the phase space due to high energy barriers, and will lead to large inaccuracies in the obtained PMF values. The umbrella sampling method [120] can be employed to overcome the sampling problem. A reaction coordinate between two states is defined and the simulation samples along this reaction coordinate. A bias is added to the system to force the simulation along this reaction pathway, therefore overcoming any local minimum that the simulation might be stuck. From these biased simulations a potential of mean force can be generated that describes the free energy of the system along the chosen reaction pathway. Using umbrella sampling the system can sample a larger area of the phase space and enables the visualisation of the local energy minima along the reaction coordinates.

2.5.2 Metadynamics

Another method to improve the sampling of a system and to overcome energy barriers within MD simulations is the metadynamics method developed by Liao and Parinello [121]. Within metadynamic simulations the system is coaxed to explore new configurations by adding a biased potential to drive the system forwards. This biased

potential is added as a function of Collective Variables (CVs), which are numerical descriptors of the the system as a function of atomic positions. Throughout the simulations, in addition to the standard potential $U(r^N)$, a biased potential $V_G(r^N, t)$ in the form of Equation 2.51 is added to the system:

$$V_G(r^N, t) = \omega \sum_{t'=\tau_G, 2\tau_G, \ldots (t' < t)} \exp \left(-\sum_{\alpha=1}^{d} \frac{[S_\alpha(r^N) - s_\alpha(t')]^2}{2\delta s_\alpha^2} \right) \qquad (2.51)$$

where d is the number of CVs used, $S_\alpha(r^N)$ is the function of the coordinates in the system that defines the CV, s_α is the value of the CV, ω and δs_α are the height and width of the added Gaussians respectively and τ_G is the frequency in which these Gaussians are added to the system. By depositing Gaussians along the trajectory at small intervals (τ_G), the system is driven forward in a random walk. Local energy minima that exist within the system will get filled up with these Gaussians, driving the system away from configurations it has visited previously towards new configurations. Figure 2.6 shows the addition of Gaussians (top graph) during the simulation. The simulation has started in the central local minima, and after the addition of 20 Gaussians (blue line) the system is driven to explore a different region of the phase space.

As the biased potential $V_G(r^N, t)$ is a history dependent potential, an estimate of the underlying free energy surface of the system can be recreated from the Gaussians added during the simulation (Figure 2.6). The bottom graph shows this recovered free energy surface. This free energy is a sum of the added Gaussians and as such is a function of the collective variables used to drive the system. Equation 2.52 shows that the slow deposition of the biased potential tends towards the negative of the free energy F.

$$\lim_{t \to \infty} V_G(s, t) = -F(s) \qquad (2.52)$$
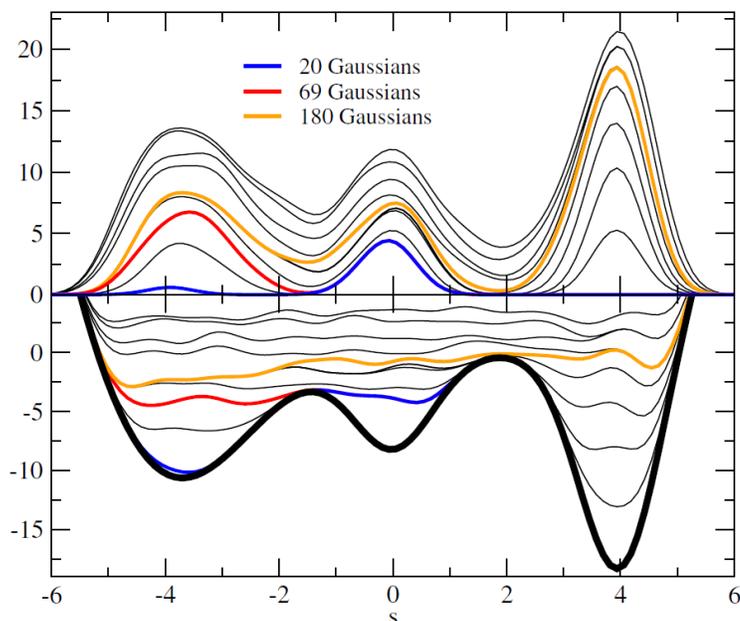
Figure 2.6: Top: Time evolution of the added bias potential along a trajectory with metadynamics simulations. Blue line: after the addition of 20 Gaussians the system escapes the energy minimum. Red line: the biased potential after the addition of 69 Gaussians. Yellow line: biased potential after 180 Gaussians. Bottom: The free energy surface approximation generated with the sum of the added Gaussians. From [122]. © IOP Publishing. Reproduced with permission. All rights reserved.

The system in metadynamics is biased along a trajectory described by the CVs, and thus the choice of CVs is an important factor in order to create reliable trajectories and free energy surfaces. The properties of the CVs must adhere to the following principles:

- The CVs must be able to clearly distinguish between different states (configurations) of the system. This includes not only the start and final configurations but also any intermediate configurations.

- The CVs must be able to describe all the slow variables that occur during the process of interest. If a CV descriptor is omitted, this process will not be observed.

- The number of CVs must be limited to avoid long simulation times to fill the free energy surfaces.

The CVs can be any function of the coordinates in the system, such as geometry

variables, coordination number or potential energy. A correct set of variables will normally return a smooth free energy [122]. Metadynamics have been used in a range of simulation studies. Quigley *et al.* [25, 34] used a combination of Steinhardt order parameters [123] and the potential energy of the system as CVs to construct a free energy landscape of the stability of $CaCO_3$ nano particles in water. Bulo *et al.* [124] used the distance between atoms, the coordination number, contact number and radius of gyration as CVs to describe the binding of Ca^{2+} ions to poly acrylate chains in water.

Metadynamics simulations provide a method to simulate larger systems whilst ensuring that all of the phase space is sampled. Unlike methods such as umbrella sampling, the history dependent bias potential that is added to the system allows for a estimation of the free energy of the system to be be obtained. Instead of driving the system along a set pathway, the metadynamics method enables the system to explore more of the phase space. The main drawback of metadynamics is the difficulty of choosing the appropriate CVs to describe the reaction, which greatly affect the results of the experiment [125].

Chapter 3

Inclusion of amino acids into calcite crystals

3.1 Introduction

3.1.1 Inclusions within minerals

The inclusion of biomolecules within mineral crystals is a well known and well documented phenomenon. In nature, many biomolecules are responsible for the specific nucleation and growth of minerals (Chapter 1.1.3) and these biomolecules often get incorporated within the crystal. In the case of calcium carbonate, one example of this phenomenon is with the structure of nacre (Figure 3.1) where, in addition to the organic matrix in between the aragonite tablets, intra-crystalline proteins have been observed within the aragonite single crystals [126]. The incorporation of these intra-crystalline biomolecules in calcium carbonate is a widespread phenomenon, and recently Pokroy *et al.* [87] studied the behaviour of biogenic calcite found in five different organisms to study the effects of the inclusion of these biomolecules on the calcite crystal. They found that these intra-crystalline proteins adhere to the calcite basal planes which lie perpendicular to the *c*-axis of the crystal, an observation that was shared by Li

et al. [127]. Using experimental techniques (annular dark-field scanning transmission electron microscopy, ADF-STEM, in combination with focussed ion beam, FIB) they were able to slice the surface of a sample of biogenic calcite from a mollusc shell and create a three dimensional tomogram of the included biomolecules (Figure 3.2). The biomolecules that were observed to be included within the single calcite crystals were not incorporated in an isotropic manner. Perpendicular to the *c*-axis of the crystal the particles were elongated, whereas along the *c*-axis they appeared more uniform and circular, supporting the suggestion by Pokroy *et al.* that these inclusions adhere to the planes perpendicular to the *c*-axis. The inclusion of these biomolecules within the framework of the minerals leads to nanocomposite materials which exhibit enhanced physical properties (such as hardness, fracture toughness and modulus) compared to their geological mineral counterparts [41]. Thus, the inclusion of these biomolecules within calcite crystals is an excellent method to (synthetically) alter the physical properties of these minerals.
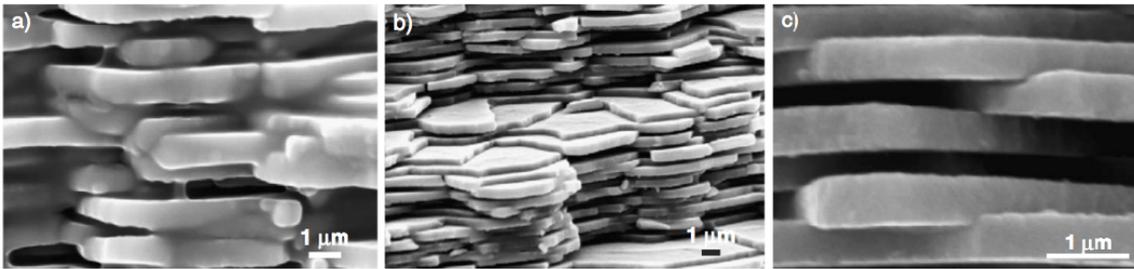
Figure 3.1: SEM images of the lamellar growth of aragonite platelets in nacre. The organic phase incorporated in between the platelets and within the single aragonite crystals are the cause of the high hierarchical structure of the material. Reprinted from [128] with permission from Elsevier.

The synthetic alteration of (calcium carbonate) crystals has been well documented, with examples of the inclusion of organic dye molecules such as Congo Red within aragonite and calcite to add colour to the normally transparent crystals going back centuries [129]. More recently, Li *et al.* [130] used an ammonium diffusion method to grow single calcite crystals from solution in the presence of an agarose (a linear polysaccharide) hydrogel with the aim to mimic the fibrous organic matrix found in biogenic calcite.
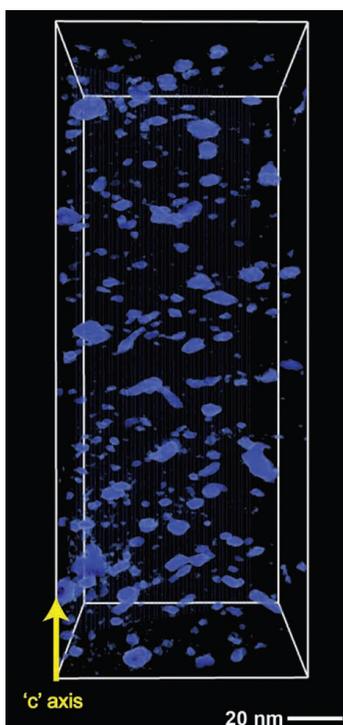
Figure 3.2: Three dimensional tomograph for biogenic calcite from *Atrina rigida*. The organic phase is seen to be distributed throughout the whole crystal. From [127]. Reprinted with permission from the Wiley Company.

Using the same techniques as described above (ADF-STEM and FIB) they were able to show that the fibrous biomolecules get incorporated within calcite, forming a three dimensional network throughout the crystal, whilst maintaining a perfect rhombohedral single calcite crystal. Calcite crystals were also grown by Pokroy *et al.* [87] in the presence of the well known intra-crystalline protein caspartin [131] to analyse the effect of the incorporation of this protein on the calcite lattice. Their experiments resulted in the growth of single calcite crystals with different morphologies from those crystals grown in the absence of caspartin. The lattice of the crystals with inclusions was expanded anisotropically, with an increased lattice distortion in the *c*-direction of the crystal.

The changes in physical properties of these nanocomposite materials in nature have also been observed by Kim *et al.* [88] who artificially created a nanocomposite material using calcium carbonate precipitation experiments in the presence of functionalised polystyrene polymer micelles with diameters of 220 to 250 nm. This research group

managed to synthesise single crystal calcite with organic inclusions of up to 20 vol %. The inclusion of these macromolecules affected the physical properties of the mineral, with a decrease in the modulus and hardness of the calcite being observed. However, upon nano indentation, cracks appeared in the crystal faces and the polymer micelles were found to bridge the gaps over these cracks, enhancing the fracture toughness of the material when compared to pure calcite (Figure 3.3).

The same research group has managed to refine the process of the incorporation of a range of biomolecules within calcite whilst still obtaining single crystals. They have included large co-polymer micelles [132], reaching up to 13 wt % or 30 vol % of inclusions that were uniformly distributed throughout the crystal without changing the morphology of the crystal (Figure 3.4). The incorporation of such large macromolecules is surprising but can be explained by a strong affinity between specific calcite surfaces and the biomolecules, and the balance between the fast growth mechanism of the calcite and the adsorption of the biomolecules. Presumably the affinity between the biomolecules and the surface is strong enough to allow a molecule to reside on the surface long enough for the fast growing calcite to precipitate around the molecule, enclosing it within the crystal. Thus the specific interactions between the host crystal and the biomolecule to be included plays an important factor in these inclusions.
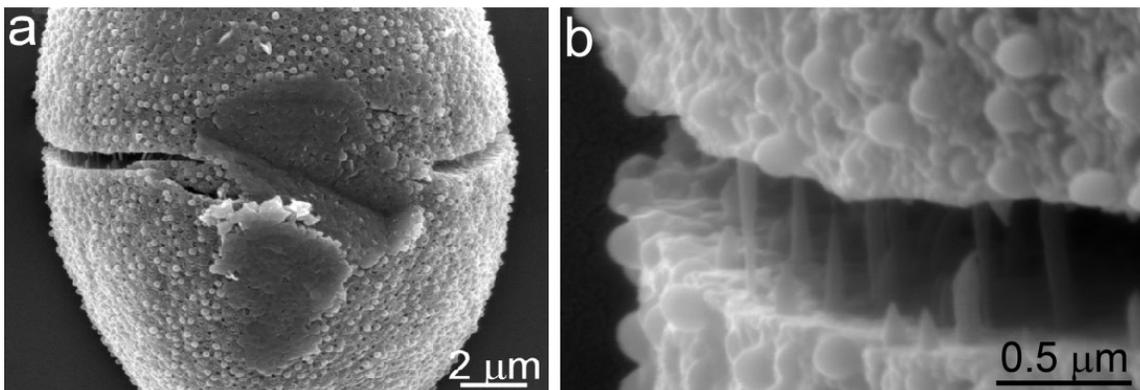
Figure 3.3: Upon nano indentation the polymer micelles included within the calcite crystal elongate and bridge the gap of the formed crack, inhibiting the crack propagation and enhancing the fracture toughness of the nano composite material. From [88]. Reprinted with permission from the Wiley Company.
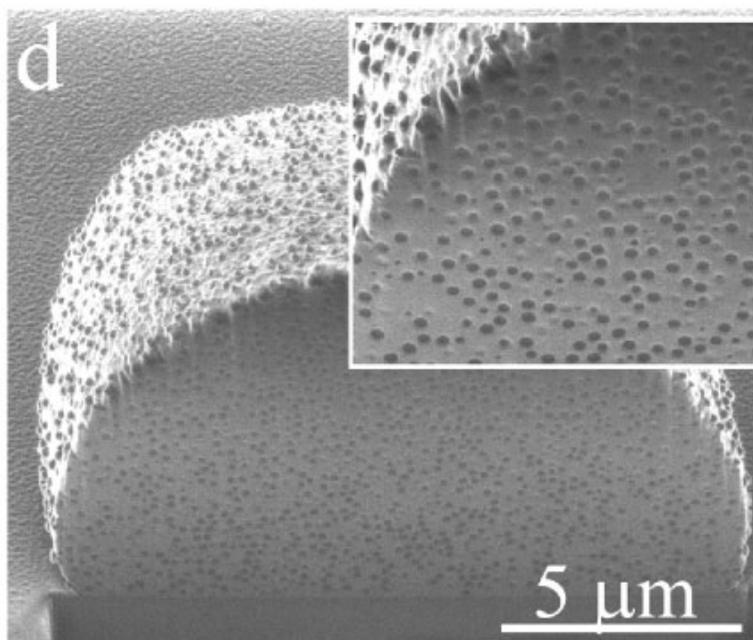
Figure 3.4: The polymer micelles included within the crystal are uniformly distributed throughout the crystal, and not just present near the surface of calcite. From [88]. Reprinted with permission from the Wiley Company.

3.1.2 Amino acid inclusions

The importance of specific interactions between included biomolecules and the host crystal has recently been investigated by Borukhin *et al.* [133], who published research on precipitation experiments of calcium carbonate in the presence of small amino acids. They adopted a systematic approach to map the incorporation of amino acids and the effects of the incorporation of these amino acids on the lattice parameters of calcite. The results from this study of the 20 common amino acids[1] (with the exception of the insoluble tyrosine) showed that most, but not all, amino acids can be incorporated within a calcite crystal using an ammonium diffusion method (Figure 3.5). Of all amino acids, aspartic acid (Asp) is the most readily incorporated within the calcite crystal and a molar incorporation of 0.82 mol % was observed. It should be noted that these experiments were performed with low amino acid concentrations in the solution (10 to 40 mM). Using X-ray diffraction the effects of the inclusion of the amino acids on the c-lattice parameters of the calcite crystal were analysed (Figure 3.6). Upon inclusion of

---

[1]A list of all amino acids is included in Appendix A

Asp a shift in the $(10\bar{1}4)$ peaks to larger d-spacings can be observed, indicating that the included amino acids have a significant effect on the lattice parameters of calcite. The data in Figure 3.7 shows the distortion of the calcite crystal lattice upon the inclusion of the amino acids. When normalising the lattice distortion to the molecular weight of the included amino acids, the degree of packing of these species within the crystal can be analysed, and the relative packing of two amino acids, histidine (His) and glycine (Gly), can be seen to be enhanced. The effect of His, with a large imidazole ring, can be explained by its bulk, but the inability of the small amino acid Gly to pack effectively into the calcite lattice is not well understood.
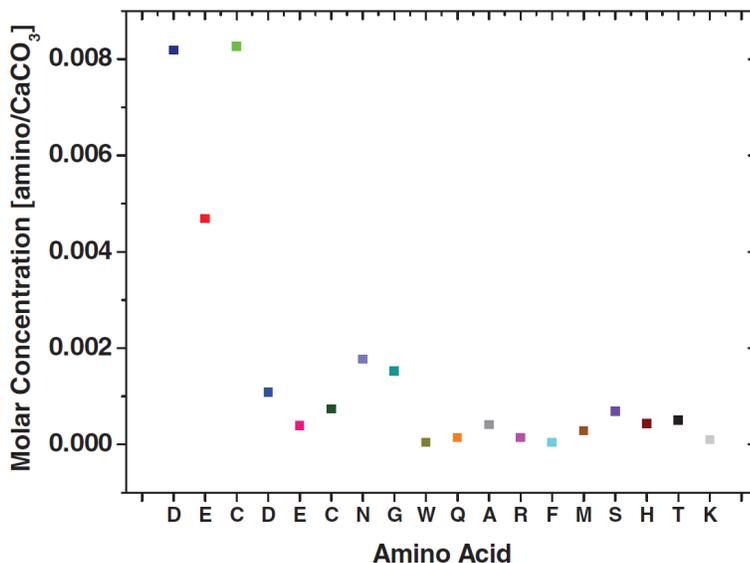
Figure 3.5: Molar concentrations of amino acids included within the calcite crystal using precipitation experiments. From [133]. Reprinted with permission from the Wiley Company.

Based on the high levels of incorporation shown by Asp and the unusual packing behaviour of the small amino acid Gly, both amino acids were selected by Kim *et al.* [134] for further in-depth analysis. Their analysis consisted of both experimental work and computational simulations, which are described in this work. Using an ammonium diffusion method [135] the research group studied the inclusion of Asp and Gly for a range of amino acid concentrations in the initial precipitation solution. The use of higher amino acid concentrations within the initial solutions resulted in much higher levels
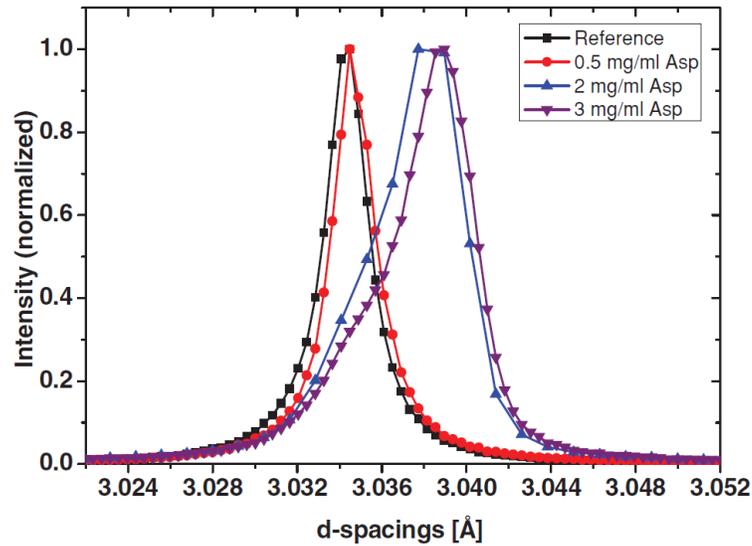
Figure 3.6: X-ray diffraction data for the d-spacing in the $(10\bar{1}4)$ diffraction peaks upon inclusion of aspartic acid within the calcite crystal. An increase in the d-spacing can be observed upon inclusion. From [133]. Reprinted with permission from the Wiley Company.
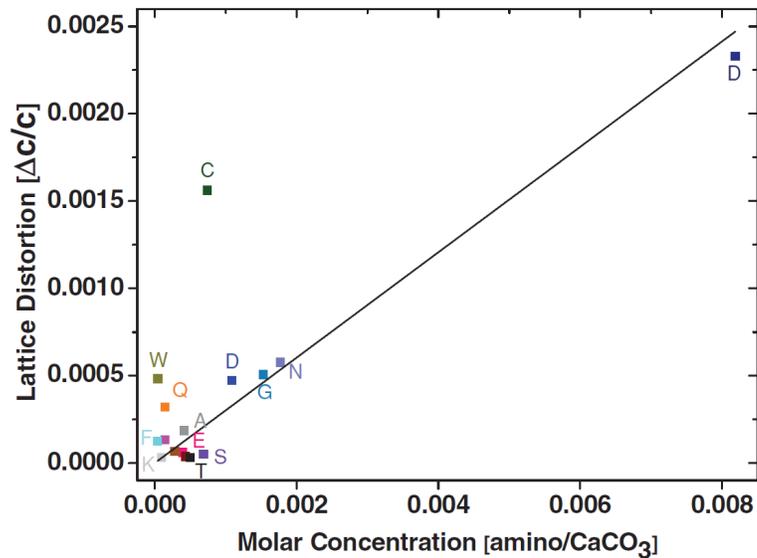
Figure 3.7: The distortion of crystal lattice in the *c*-direction upon inclusion of amino acids. From [133]. Reprinted with permission from the Wiley Company.

of amino acid inclusions than reported by Borukhin *et al.* [133]. The amino acid Asp showed a maximum incorporation of 4.0 mol % with an initial amino acid concentration of 50 mM (Figure 3.8). The inclusion of Gly resulted in similar incorporations, although higher amino acid concentrations in solution were needed. At similar concentrations of initial solution Gly was incorporated at lower levels than Asp, with a maximum inclusion of 1.7 mol % at 100 mM. However, higher inclusions of Gly of 4.4 mol % and 6.9 mol % were obtained by increasing the concentration of amino acid concentrations to 200 mM and 400 mM respectively. This data suggests that Asp is more readily incorporated within the calcite crystal than Gly, supporting the findings by Borukhin *et al.* [133].

The calcite crystals with incorporated amino acids were subjected to high resolution synchrotron powder X-ray diffraction experiments to analyse the effect of the inclusions, with both the peak shifts and peak broadening in the diffraction data being analysed. The peak shifts are a measure of the lattice distortions within the crystals and the data shows that in both the *c*- and *a*-directions the crystal lattice becomes distorted upon inclusion of the amino acids (Figure 3.9). Inclusion of Asp and Gly gave a maximum distortion in the $\Delta c/c$ direction of 0.003 and 0.005 respectively, relative to the original lattice parameters. The lattice distortion in the $\Delta a/a$ direction was an order of magnitude lower with a maximum distortion of 0.00045 and 0.00065 for Asp and Gly respectively. The discrepancy between the *c*- and *a*-axis lattice distortions can be attributed to the anisotropic behaviour of the calcite crystal, with layers of calcium and carbonate ions stacked horizontally in the (001) direction (Figure 1.1).

The peak broadening, a measure of the degree of local strain (microstrain) within the crystal, of the inclusions is shown in Figure 3.10. Not all diffraction peaks are broadened equally and this asymmetrical broadening is another indication that the crystal is distorted anisotropically. Maximum peak broadening (measured as the full width half maximum, FWHM) for the (006) plane were found at molar inclusions of 1.01 mol % and 1.70 mol % for Asp and Gly respectively. These inclusions caused a

peak broadening of 0.05 for Asp and 0.07 for Gly. At higher inclusions of Asp the peaks broadening levels out, whereas with Gly a contraction of the peaks is observed, an indication that at these levels the amino acids could get incorporated within the lattice via a different mode, such as aggregation of the biomolecules.
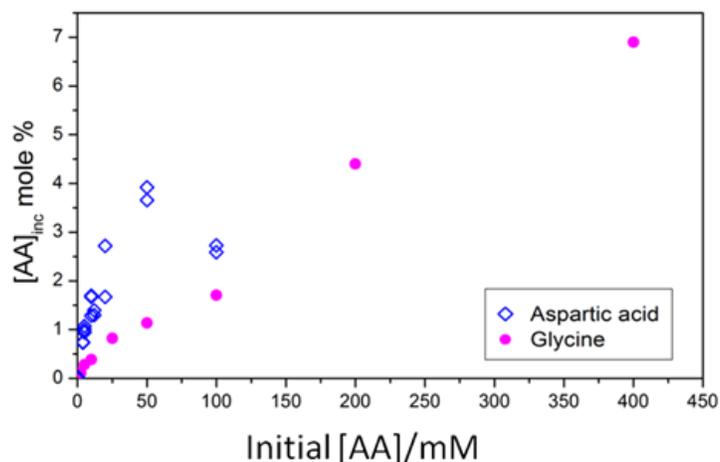
Figure 3.8: The effect of the amino acid concentration on the iInclusion of Asp and Gly within the calcite crystal. The data shows a linear relationship between inclusion and amino acid concentration of initial solution. From Kim *et al.* [134].
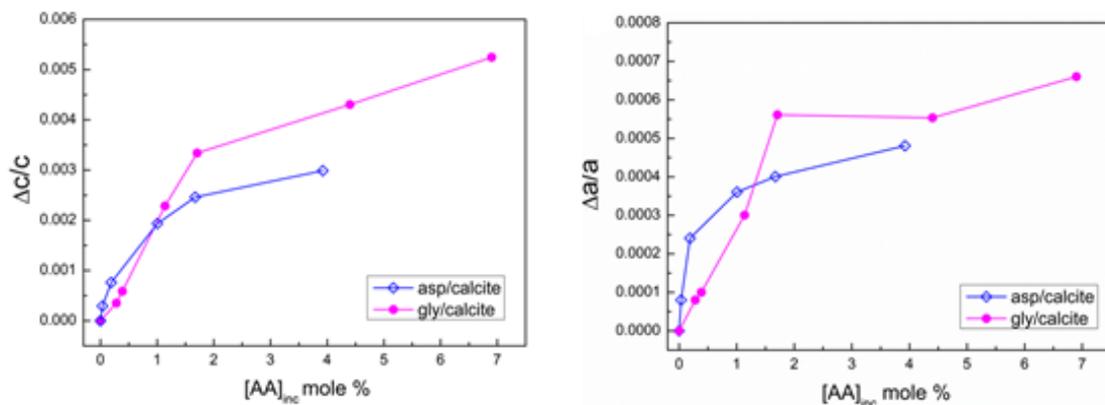
Figure 3.9: High resolution synchotron powder X-ray diffraction data for calcite crystals with both Asp and Gly included for the *a*-direction and *c*-direction. The crystal gets distorted in both directions, although the amount of distortion is higher within the *c*-direction. From Kim *et al.* [134].

Raman spectroscopy of the samples was performed to analyse the crystal lattice by investigating the configuration and molecular environment of the carbonate groups. The vibrational frequencies at $\sim$283 cm^{-1}, corresponding to the optical phonons of
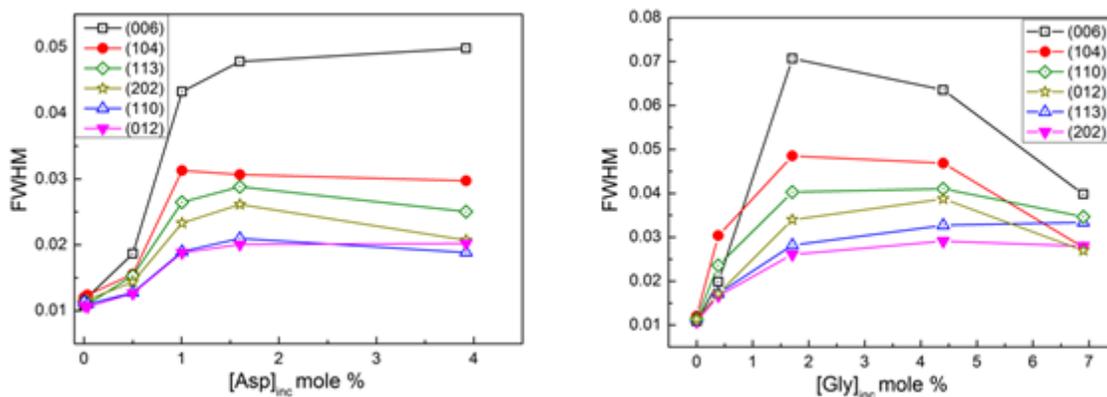
Figure 3.10: High resolution synchotron powder X-ray diffraction data shows peak broadening of the calcite peaks within the crystals. This indicates a degree of microstrain within the crystal upon inclusion. From Kim *et al.* [134].

carbonate lattice vibrations (Figure 3.11a), and $\sim$1085 cm^{-1}, corresponding to the carbonate $v1$ symmetric stretch (Figure 3.11b) were investigated [136]. The data shows that upon inclusion of amino acids the two vibrational peaks shifted by approximately 4 cm^{-1} and 1.5 cm^{-1} respectively. The same shifts are observed for both Asp/Gly inclusion, and these shifts are another indication of the disruption to the crystal lattice upon the inclusion of biomolecules.

(a) Carbonate lattice mode vibrations.

(b) Carbonate $v1$ symmetrical stretch.

Figure 3.11: Schematic representation of the Raman modes of vibration for carbonates.

Visual analysis of the morphology of the crystals was performed using scanning electron microscopy (Figure 3.12). The morphologies of the crystals remain perfect rhombohedral until initial concentrations of 5 mM Asp and 100 mM of Gly. The discrepancy between the onset of rounded edges and elongation is consistent with the lower level

of inclusions of Gly as seen in Figure 3.8. Similar morphology changes of calcite crystals have been observed in the presence of many other biomolecules, as discussed in Chapter 1.1.4.

The similarities between the calcite crystals with amino acid inclusions and biogenic calcite can further be seen when analysing the mechanical properties of these crystals. The hardness of these crystals (the resistance of the material to localised plastic deformation) was measured using nano indentation and the data shows an increase in hardness of the material with an increasing amount of included amino acid. The hardness increased from 2.3±0.2 GPa for pure calcite to 4.1±0.3 GPa and 3.4±0.4 GPa for the inclusion of Asp and Gly respectively. These values are similar to those found in the literature of 2.30±0.14 to 2.54±0.07 GPa for Icelandic Spar calcite (geological calcite) and 3.47±0.21 to 4.19±0.27 GPa for the biogenic calcite of a mollusc [43].
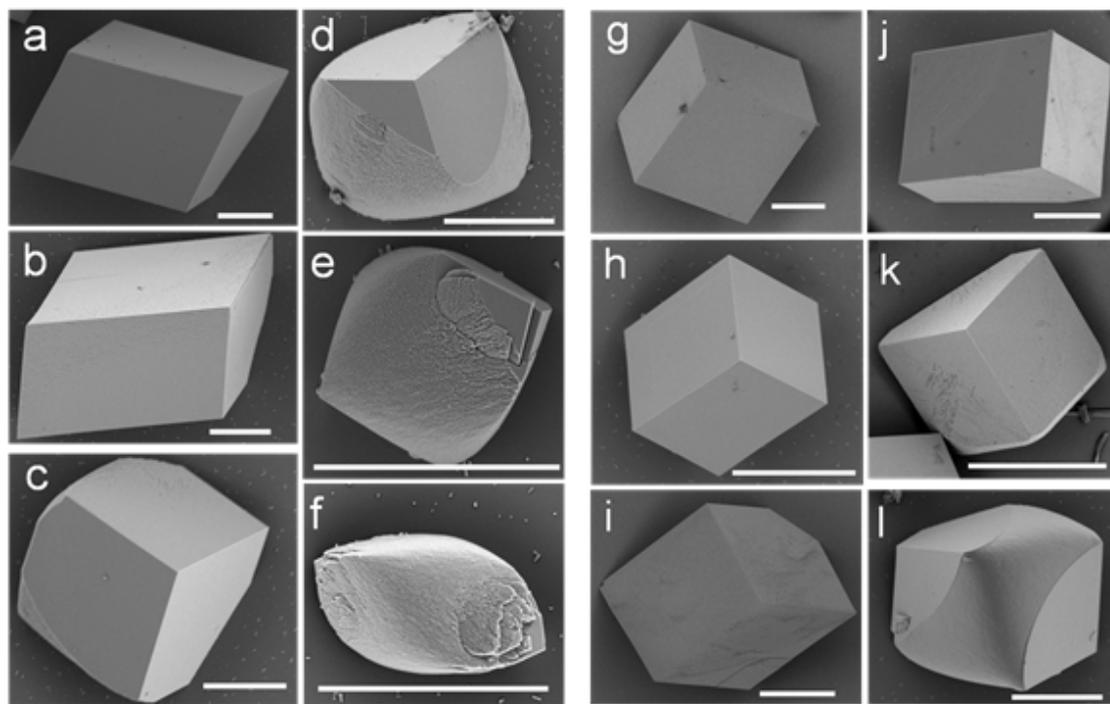
Figure 3.12: SEM images of calcite with Asp and Gly inclusions shows changes in morphology as a function of Asp (a-f) and Gly (g-l) concentration in solution. a. 1 mM, b. 5 mM, c. 10 mM, d. 20 mM, e. 50 mM and f. 100 mM. (scale bar = 10 μm). g. 5 mM, h. 10 mM, i. 20 mM, j. 50 mM, k. 100 mM and l. 200 mM. (scale bar = 20 μm). From Kim *et al.* [134].

3.1.3 Aims

The inclusion of these biomolecules within the calcite crystal can be observed via many experimental techniques, as described above. However, the modes of inclusion on an atomistic length scale will not necessarily be observable using these techniques. Using computational simulations, the interactions between the biomolecules and the ions within the crystal can be observed in great detail, and possible modes of the inclusion of these species can be suggested. Within this work molecular dynamics simulations were performed to analyse the interactions between the amino acids Asp and Gly and the calcite crystal. Using this technique the behaviour of the crystal lattice could be observed when these biomolecules were inserted within small vacancies in the crystal. Close collaboration with the experimental research groups of Fiona Meldrum[2] and Lara Estroff[3] enabled the conditions of the simulations to be in line with experimental protocols.

3.2   Methods

3.2.1   Preparation of amino acids

The charged state of amino acids changes depending on the pKa of the amino acid functional groups and the environment the amino acid is exposed to. The pH range of the experimental set-up for the precipitation experiments typically exceeds pH 9 [135] and as the pKa values of the amino acids used in this work are close to the experimental pH values, a range of charged states must be considered. Four differently charged amino acids were used within this work, to include the zwitterionic states of the amino acids (Figure 3.13).

The amine functional groups of the amino acids were considered to be either neutral

---

[2]School of Chemistry, University of Leeds, Leeds.
[3]Department of Materials Science and Engineering, Cornell University, Ithaca, NY.

(NH_2) or positively charged (NH_3^+), whereas the carboxyl groups were always considered to be negatively charged (COO^-) due to the pKa of these groups. The pKa values for the amine and carboxyl functional groups and the overall charges of these amino acids are detailed in Table 3.1.

(a) Asp^{1-}.   (b) Asp^{2-}.   (c) Gly^0.   (d) Gly^{1-}.
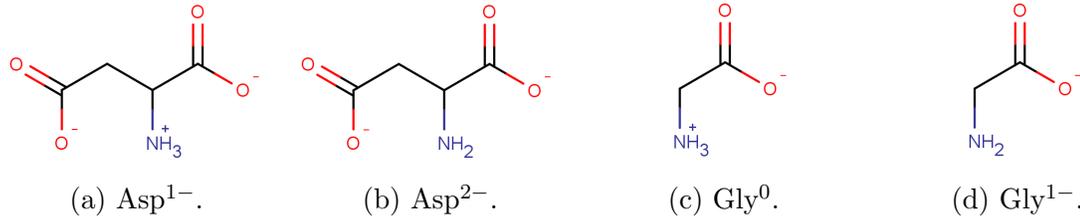
Figure 3.13: Chemical structures of the four amino acids used within this work.

Amino acid	pKa(NH_2)	pKa(COOH)	pKa(COOH sidechain)	Total charge
Asp^{1-}	9.82	2.10	3.86	-1
Asp^{2-}	9.82	2.10	3.86	-2
Gly^0	9.60	2.34	-	0
Gly^{1-}	9.60	2.34	-	-1

Table 3.1: Details of the pKa values of Asp and Gly and the total charge of the species used in this work.

3.2.2   Preparation of mineral

A calcite crystal of 144 unit cells measuring 34 Å x 44 Å x 34 Å and containing 864 $CaCO_3$ formula units was created such that the symmetric periodicity of the crystal was kept intact, thus creating a bulk crystal system when simulated with periodic boundary conditions. The orientation of the crystal was chosen such that the (001) planes of calcite were positioned perpendicular to the z-axis. The reasoning for this orientation was that anisotropic expansion of the crystal lattice along the crystal a/b- and c-directions was also along the Cartasian x/y- and z-directions. Defects were created within the crystal by removing Ca^{2+} and CO_3^{2-} ions situated next to each other - either in the same or adjacent calcite layers - to produce a cavity and these cavities were filled with small amino acids. The amino acids were simulated within an

aqueous environment to obtain an initial configuration, before being inserted within the cavities in a random orientation using the program `Packmol` [137]. Figure 3.14 details the method used to create the cavities within the crystal and the inclusion of the amino acid Asp. To avoid an overall net charge of the system, the number of ions deleted from the crystal varied, as set out in the solution chemistry as shown in equations 3.1 to 3.4:

$$CaCO_3(s) + 2\,nAsp^{1-}(aq) + nCa^{2+}(aq) \rightleftharpoons CaCO_3 \cdot 2\,nAsp^{1-} \cdot nCa(s) \quad (3.1)$$
$$+ 4\,nCaCO_3(s)$$

$$CaCO_3(s) + nAsp^{2-}(aq) + nCa^{2+}(aq) \rightleftharpoons CaCO_3 \cdot nAsp^{2-} \cdot nCa(s) \quad (3.2)$$
$$+ 2\,nCaCO_3(s)$$

$$CaCO_3(s) + nGly^0(aq) \rightleftharpoons CaCO_3 \cdot nGly^0(s) \quad (3.3)$$
$$+ 2\,nCaCO_3(s)$$

$$CaCO_3(s) + 2\,nGly^{1-}(aq) + nCa^{2+}(aq) \rightleftharpoons CaCO_3 \cdot 2\,nGly^{1-} \cdot nCa(s) \quad (3.4)$$
$$+ 4\,nCaCO_3(s)$$
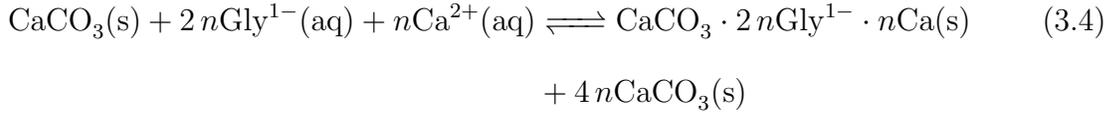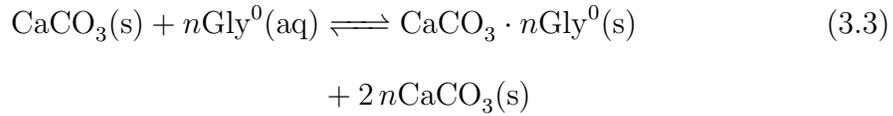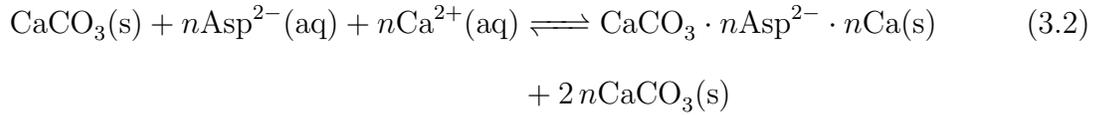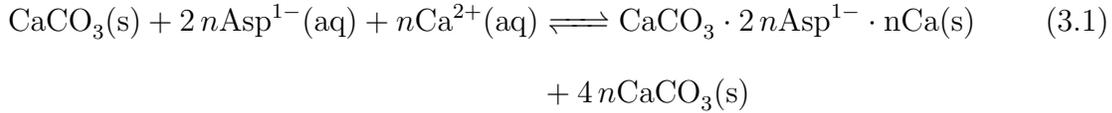
Within these equations, the first term denotes a solid block of calcite ($CaCO_3(s)$) that is the whole simulation cell. Next, the amino acid and a calcium counter ion are simulated in an aqueous environment (aq). These amino acids with their counter ion are then placed within the inclusion site of the calcite block, described by the first term on the right hand side of the equation. In the case of Asp^{1-} the inclusion site is generated by the removal of three Ca^{2+} and three CO_3^{2-} ions from the calcite block. These ions are removed from the inclusion site to become part of the bulk lattice in a way analogous to the definition of a Schottky defect. Thus, for n inclusions, $3nCa^{2+}$ and $3nCO_3^{2-}$ ions make place for $2nAsp^{1-}$ and nCa^{2+}.

The energetics of the system have to be compared to the reference state of all separate ions of the system. This comparison can be made with either the separate ions in solution or as part of the bulk calcite crystal. In this work the energetics of the excess Ca^{2+} and CO_3^{2-} ions were calculated with respect to bulk calcite. With this method the results did not depend on the reference state of the system, and any effect of the solvation of the mineral ions and the effects of structured water layers surrounding these ions in solution did not affect the energetics. An increasing number of amino acids was incorporated within the crystal in the range of 0 up to 5.2 wt %, corresponding to 0 to 2.3 mol %, or 0 to 25 molecules.

The detection of small biomolecules in calcite crystals and their distribution throughout the crystal is not possible using experimental techniques, although the detection of larger biomolecules is possible [88, 127, 130, 132]. The results of these studies show that the large biomolecules are distributed evenly throughout the crystal and are not agglomerated on the surface. Additionally, they observed that although the biomolecules agglomerated in solution, these clusters did not interact with the growing crystal surface [132]. Based on these results the amino acids within this work were modelled as individual inclusions distributed throughout the crystal instead of agglomerated clusters. As the amino acids are thought to be incorporated within the crystal by a rapid crystal growth front, the movement of the biomolecule whilst adsorbed on the surface will be minimal, resulting in a small, confined vacancy in which the amino acid is situated. Based on this presumption the vacancies within the crystal were modelled such that the amino acids were confined within a small space.

3.2.3 Modelling protocols

The models of the amino acids in this work were prepared using the `Amber 12 Antechamber` package [138]. The atomic charges of the molecules were calculated using the AM1-BCC method [139], further explained in Appendix B. The inter- and intra-molecular inter-

73
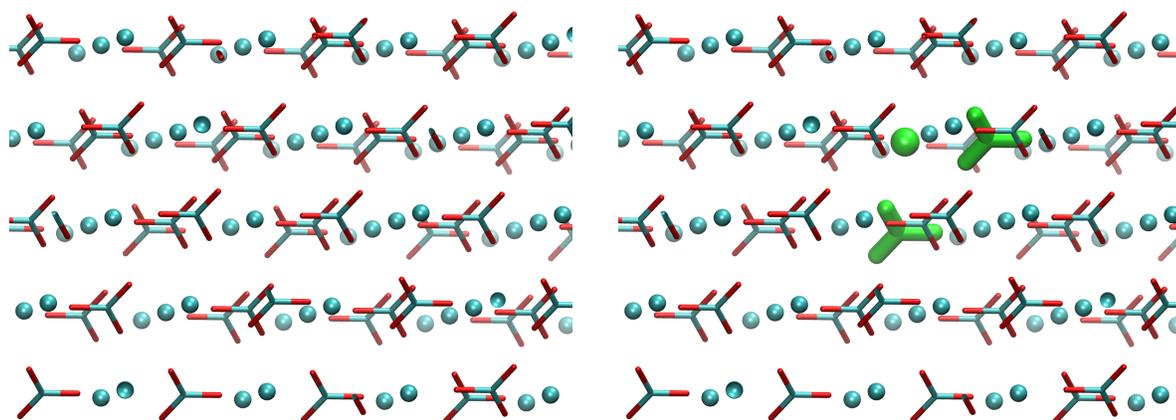
(a) The calcite crystal without any defects or inclusions within its structure.

(b) Two carbonate ions and one calcium ion are selected for removal (coloured green).
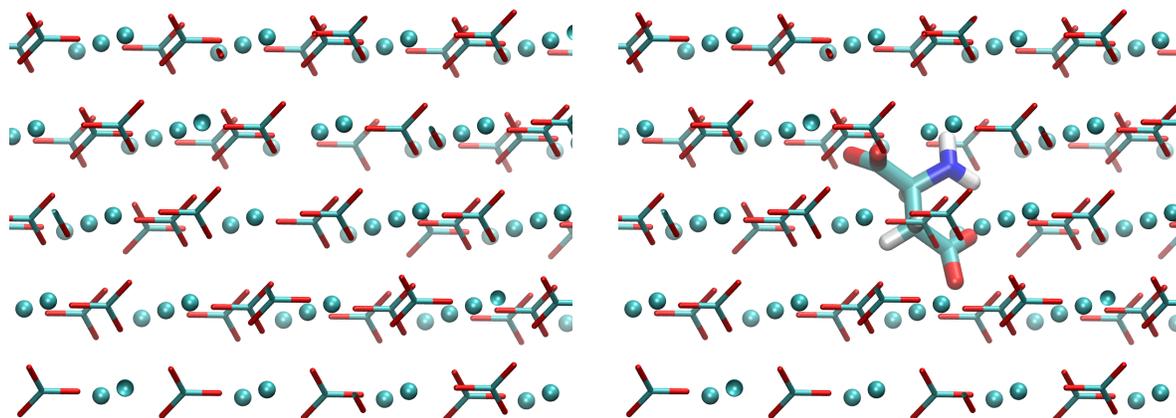
(c) The carbonate and calcium ions are removed from the crystal to form a cavity.

(d) Using the program `Packmol`, the amino acid Asp^{2-} is inserted within the cavity.

Figure 3.14: The mechanism of inclusion of amino acids within the calcite crystal, in which several ions are selected for removal before an amino acid molecule is inserted within the created vacancy. Oxygen, carbon, nitrogen and hydrogen atoms are depicted in red, cyan, blue and white respectively, with calcium ions represented as cyan spheres.

actions of the calcite were described by the potentials developed by Raiteri *et al.* [26]. The well used potentials from the `AMBER` *ff12SB* force field [111] were used for the amino acids and water molecules. The cross-terms between the organic (amino acids and water) and inorganic ($CaCO_3$) phases were calculated using the method described by Freeman *et al.* [103]. For further details on the potentials see Chapter 2.2.4. The potential details for these simulations are shown in Appendix E.

Inclusion simulations

All simulations of the calcite with inclusions (Figure 3.15) were carried out using the molecular dynamics package `DL POLY Classic` [140]. The systems were equilibrated at a temperature of 300 K for 100 ps with a timestep of 0.1 fs in an NVT ensemble with a Nosé-Hoover thermostat [119] with a relaxation time of 0.5 ps. The configurational energy of the system during the equilibration simulations was analysed to ensure a fully equilibrated system. A second equilibration simulation for 100 ps with a timestep of 1.0 fs in an NPT ensemble at 300 K and and at standard pressure (10^5 Pa) using a Nosé-Hoover thermostat and a Hoover barostat [141], both with a relaxation time of 0.5 ps, was performed. Subsequently, a 2.0 ns simulation using a timestep of 1.0 fs in an NPT ensemble with the same thermostat and barostat variables was carried out for data analysis.

Anisotropic expansion

In order to analyse the anisotropic behaviour of the calcite crystal, simulations were carried out in which the crystal was independently expanded in the a/b- and c-directions. The initial setup of the crystal was chosen such that the (001) planes of calcite were positioned perpendicular to the Cartasian z-axis, to enable this expansion to correspond to the Cartasian x/y- and z-directions. Three separate systems were simulated with 2.3 mol % Asp^{2-}, 2.8 mol % Gly^0 and no inclusions. These amounts of inclusions were
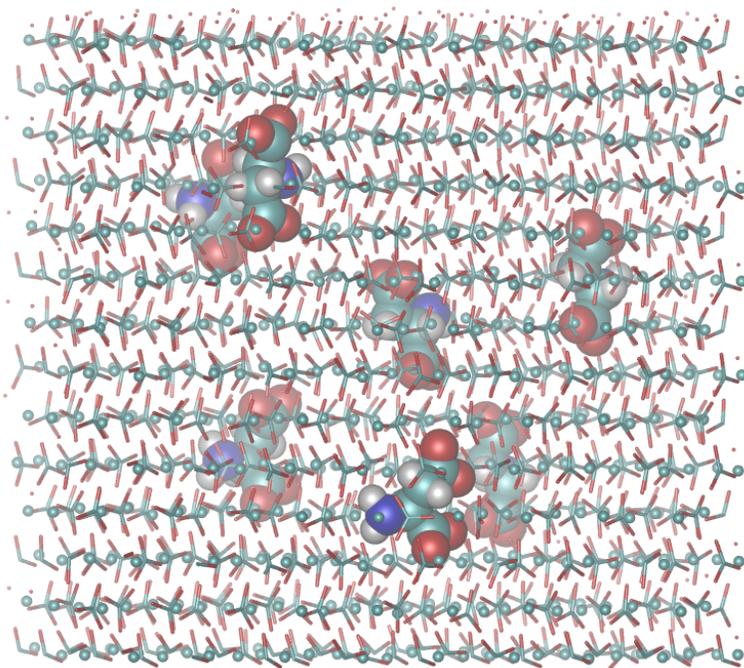
Figure 3.15: Model of the calcite crystal with seven amino acids included throughout the crystal.

chosen to reflect the maximum level of inclusion that were observed with experiments by the research group of Meldrum. The Asp^{2-} and Gly^0 systems were chosen as they represented the closest pH match to experiments. The inclusions were inserted into the crystal at random positions and in random orientations, as described above. Each simulation box was expanded manually along either the crystal a/b-axes or the c-axis independently in increments of 0.05 % of the initial lattice parameter. After each expansion the systems were subjected to an equilibration simulation at a temperature of 300 K in an NVT ensemble with a Nosé-Hoover thermostat with a relaxation time of 0.5 ps for 10 ps with a timestep of 1.0 fs, followed by a 0.5 ns data production run.

For these anisotropic expansion simulations an NσT ensemble (constant number of particles, constant stress and constant temperature) could be implemented. By applying a constant stress along all cell parameters, this ensemble allows those cell parameters to vary anisotropically. However, for this work we opted to use an NVT ensemble in which the cell size was manually maintained and expanded. By altering the cell dimensions independently, the anisotropic behaviour of the crystal could be mapped in detail. This

method not only provided the energy minimum of the lattice parameters but also the energy gradients surrounding this minima.

Additional simulations

Additional simulations were carried out to calculate the formation of vacancies within the crystal. In order to keep the overall charge of the system neutral, the energy of pair defects was considered, where both a cation and an anion are removed from the crystal. The defects were created within the crystal by removing a Ca^{2+} and $CO_3{}^{2-}$ ion situated next to each other - either in the same or adjacent calcite layers. An increasing number of defects were created throughout the crystal in the range of 0 to 10. The systems were equilibrated at a temperature of 300 K and at standard pressure (10^5 Pa) for 100 ps with a timestep of 1.0 fs in an NPT ensemble using a Nosé-Hoover thermostat and a Hoover barostat [141], both with a relaxation time of 0.5 ps. Data was collected during subsequent simulations of 1.0 ns with a timestep of 1.0 fs using the same parameters.

The inclusion energy of water molecules within the calcite crystal was modelled by removing a Ca^{2+} and $CO_3{}^{2-}$ ion situated next to each other - either in the same or adjacent calcite layers - from the crystal. Subsequently two TIP3P [104] water molecules were inserted within the created vacancy in a random orientation using the program `Packmol`. An increasing amount of vacancies with included water molecules was created throughout the crystal in the range of 0 to 10 vacancies. The systems were equilibrated at a temperature of 300 K and at standard pressure (10^5 Pa) for 100 ps with a timestep of 1.0 fs in an NPT ensemble using a Nosé-Hoover thermostat and a Hoover barostat [141], both with a relaxation time of 0.5 ps. Data was collected during subsequent simulations of 1.0 ns with a timestep of 1.0 fs using the same parameters.

To calculate the energies of inclusion for the systems according to the solution chemistry shown in Equations 3.1 to 3.4, additional simulations were carried out for the calculation

of the system energy of the amino acids, calcium and carbonate ions in solution. The necessary simulations performed are detailed in Table 3.2. All of these simulations were carried out using `DL POLY Classic`. Using the program `Packmol` all molecules were randomly inserted in a box of 30 Å x 30 Å x 30 Å, containing 1100 TIP3P water molecules. The simulation of pure water was of the same dimensions and had no other species included. All simulations were equilibrated at a temperature of 300 K for 1.0 ns with a timestep of 1.0 fs in a NVT ensemble using a Nosé-Hoover thermostat with a relaxation time of 0.5 ps, and subsequently equilibrated for 1.0 ns with a timestep of 1.0 fs in a NPT ensemble using a Nosé-Hoover thermostat and a Hoover barostat, both with a relaxation time of 0.5 ps. The configurational energy of the systems was obtained from subsequent data runs of 1.0 ns in an NPT ensemble with the same variables.

Equation	Species to simulate
3.1	2 Asp^{1-}, 1 Ca^{2+}
3.2	1 Asp^{2-}, 1 Ca^{2+}
3.3	1 Gly^{0}
3.4	2 Gly^{1-}, 1 Ca^{2+}
3.1 to 3.4	2 $\text{CO}_3{}^{2-}$, 2 Ca^{2+}
3.1 to 3.4	3 $\text{CO}_3{}^{2-}$, 3 Ca^{2+}
3.1 to 3.4	H_2O^a

Table 3.2: Details of species to be simulated in an aqueous environment. ^{a} simulation of pure water, without any added ions.

3.3 Results and discussion

3.3.1 Energies of inclusion

In all systems the calculations of the energetics of the excess Ca^{2+} and $\text{CO}_3{}^{2-}$ ions were performed with respect to bulk calcite so that the results did not depend on the reference state of the system. The energy of inclusion of the amino acids cannot be directly calculated, but can be determined using the solution chemistry as shown in

Equations 3.1 to 3.4. The reaction energy of the inclusion of n amino acids within the solid crystal: amino acid(aq) $\rightleftharpoons$ amino acid(s) can therefore be calculated as Equations 3.5 to 3.8:

$$E_{Asp^{1-}inclusion} = \left(E_{CaCO_3 \cdot 2\,nAsp^{1-} \cdot nCa(s)} + E_{4\,nCa^{2+}(s)} + E_{4\,nCO_3^{2-}(s)}\right) \tag{3.5}$$
$$- \left(E_{CaCO_3(s)} + E_{2\,nAsp^{1-}(aq)} + E_{nCa^{2+}(aq)}\right)$$

$$E_{Asp^{2-}inclusion} = \left(E_{CaCO_3 \cdot nAsp^{2-} \cdot nCa(s)} + E_{2\,nCa^{2+}(s)} + E_{2\,nCO_3^{2-}(s)}\right) \tag{3.6}$$
$$- \left(E_{CaCO_3(s)} + E_{nAsp^{2-}(aq)} + E_{nCa^{2+}(aq)}\right)$$

$$E_{Gly^0 inclusion} = \left(E_{CaCO_3 \cdot nGly^0(s)} + E_{2\,nCa^{2+}(s)} + E_{2\,nCO_3^{2-}(s)}\right) \tag{3.7}$$
$$- \left(E_{CaCO_3(s)} + E_{nGly^0(aq)}\right)$$

$$E_{Gly^{1-}inclusion} = \left(E_{CaCO_3 \cdot 2\,nGly^{1-} \cdot nCa(s)} + E_{4\,nCa^{2+}(s)} + E_{4\,nCO_3^{2-}(s)}\right) \tag{3.8}$$
$$- \left(E_{CaCO_3(s)} + E_{2\,nGly^{1-}(aq)} + E_{nCa^{2+}(aq)}\right)$$

The two terms in brackets represent the final and initial states of the system. The final state also includes water without ions in it. Instead, the ions removed from the inclusion site are added as extra bulk crystal. The inclusion energies calculated for all four amino acid inclusions via the method in Equations 3.5 to 3.8 are shown in Figure 3.16. This plot shows the amount of energy (in kJ mol^{-1}, representing the total energy of the inclusion per simulation cell, not per mol amino acid included) that is required to include a certain mol % amino acid within the crystal. The energy for each inclusion rises with an increasing amount of included amino acid. It is obvious from this graph that the inclusion energies are particularly high, with the energy of including 5 mol % amino acid being between 10,000 and 18,000 kJ mol^{-1}. The fact that these energies of inclusions are so high and that a rapid inclusion of these species are observed throughout experiments strongly suggests that a kinetic pathway of inclusion for these small organic molecules is being used, rather than a thermodynamic mechanism being

found. The difference between individual charged amino acids is hard to discern from these simulations. Out of all four species, Gly^{1-} is seen to have the highest energy of inclusion, whereas Gly^{0} exhibits the lowest energy of inclusion, similar to Asp^{2-}. The inclusion of Asp^{1-} lies in between these values, and a crossover with Asp^{2-} at low levels of incorporation and with Gly^{1-} at higher levels of incorporation is observed.
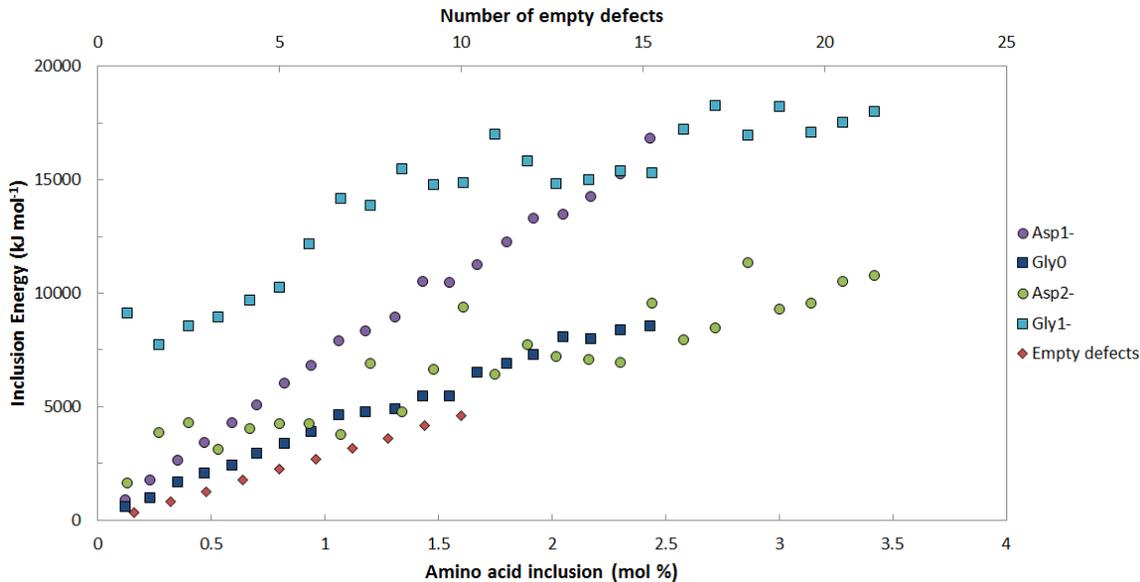
Figure 3.16: Inclusion energies calculated using molecular dynamics simulations for Asp^{1-}, Asp^{2-}, Gly^{0} and Gly^{1-}. The graph shows the amount of energy that is required to include a certain mol % amino acid within the crystal.

In order to analyse the high energy of inclusion observed in the above simulations, the inclusion energies of the amino acids in calcite were compared to vacancies in the crystal and the inclusion of water molecules within the cavities created. Fisler *et al.* [100] previously calculated the energy of a Schottky pair defect - a defect where both a Ca^{2+} and a $CO_3{}^{2-}$ were removed from the crystal - as 46.378 eV (4474.796 kJ mol^{-1}). Note that this value represents the removal of a defect to an infinite distance and not - as is the norm for Schottky defect calculations - to the surface of the calcite. This value should therefore be interpreted as a maximum defect energy. The energy for n pair defects can be calculated via Equation 3.9:

$$E_{pairdefect} = (E_{\mathrm{CaCO_3}defects} + nE_{\mathrm{Ca}^{2+}(s)} + nE_{\mathrm{CO_3}^{2-}(s)}) - (E_{\mathrm{CaCO_3}(s)}) \qquad (3.9)$$

As with the inclusion of amino acids, these energies are calculated with respect to bulk calcite. Up to ten defects were simulated, randomly placed throughout the crystal, and the average energy per defect in these simulations (Table 3.3) ranges from 338.8 to 461.9 kJ mol^{-1} for a single defect to ten defects in the crystal respectively. These simulations show that even without larger biomolecules included these calcite crystals are subjected to a high degree of strain, evident from the relative high energies for the crystals with empty defects, and that the individual defect energy rises with an increasing amount of defects. The inclusion of two water molecules within these empty defects led to a slight increase of the defect energies of between 25 to 50 kJ mol^{-1}, as shown in Table 3.3.

Number of defects	Total vacancy energy (kJ mol^{-1})	Average vacancy energy (kJ mol^{-1})	Total water energy (kJ mol^{-1})	Average water energy (kJ mol^{-1})
1	338.5	338.8	384.9	384.9
2	805.0	402.5	891.7	445.9
3	1250.3	416.8	1399.7	466.6
4	1763.0	440.8	1865.8	466.5
5	2226.8	445.4	2352.3	470.5
6	2680.3	446.7	2912.0	485.3
7	3146.5	449.5	3405.7	486.5
8	3596.8	449.6	3987.5	498.4
9	4156.8	461.9	4414.1	490.5
10	4616.0	461.6	5071.1	507.1

Table 3.3: Defect energies for vacancy defects and included water molecules within these defects calculated with the MD simulations within this work. The energy in kJ mol^{-1} represents the total energy of the defects or inclusion per simulation cell, not per mol of excluded $CaCO_3$/included water.

3.3.2 Lattice distortion

The influence of the amino acid inclusions on the crystal lattice during the simulations was also analysed and it was found that there is indeed a certain degree of disruption to the lattice. The global disruption to the lattice was analysed using the radial distri-

bution function (RDF), or pair correlation function. In an isotropic system, the RDF, $g(r)$, describes the probability of finding an atom at distance r from another reference atom with respect to the probability for a random system, as described in Chapter 2.3.4 and is a useful tool to describe the structure of a system. The distribution of the calcium ions throughout the system during the molecular dynamics simulations was analysed by calculating the RDF of the calcium to calcium distances, shown in Figure 3.17. The data for calcite without any inclusions shows a series of well defined, regular peaks, indications of the regular crystal without distortions. The data for the inclusions of Asp^{2-} and Asp^{1-} and the data for Gly^0 and Gly^{1-} showed a similar behaviour and for clarity only the data for Asp^{2-} and Gly^0 is shown. The inclusion of 2.3 mol % Asp^{2-} within the crystal shows no shift in the peaks with respect to the crystal without inclusions. The peaks are however broadened, with the first peak showing a FWHM broadening of 0.045 Å. Upon the inclusion of 2.8 mol % Gly^0, the first peak is shifted to the right by 0.07 Å and shows a broadening of 0.07 Å. The regularity seen for the lattice without inclusions has made way for irregular single peaks, showing that throughout the crystal the calcium to calcium distances are disordered. The same behaviour can be observed with the radial distribution functions of the calcium to carbonate-oxygen distances within the crystal (Figure 3.18). In this instance the inclusion of both Asp and Gly causes the regularity of the peaks in the graph to be distorted and shifted away from the pure calcite peaks. The first peak broadening for the inclusion of Asp^{2-} and Gly^0 is 0.09 and 0.1 Å respectively.

A full width at half maximum (FWHM) analysis of the first peak in the calcium to calcium and calcium to carbonate-oxygen RDF plots has also been carried out. The FWHM is a description of the width of a peak at half the maximum height of the peak. The RDF data for both Asp^{2-} and Gly^0 were analysed and are shown in Figure 3.19. This data shows that upon inclusion of Asp and Gly the disruption to the calcium to carbonate disruption is larger than the calcium to calcium distances. The onset of the disruption to the crystal can be seen to start early on in the inclusion process for Gly^0
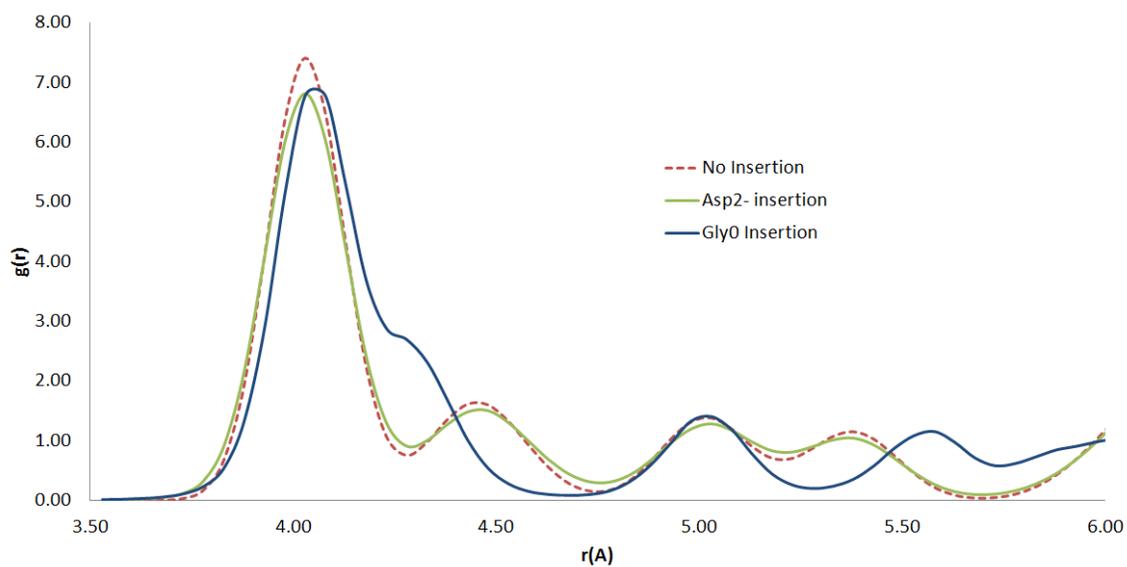
Figure 3.17: Radial distribution function of the calcium to calcium distance within the calcite crystal with no inclusion (red dotted line), 2.3 mol % Asp^{2-} inclusion (green solid line) and 2.8 mol % Gly^0 inclusion (blue solid line).
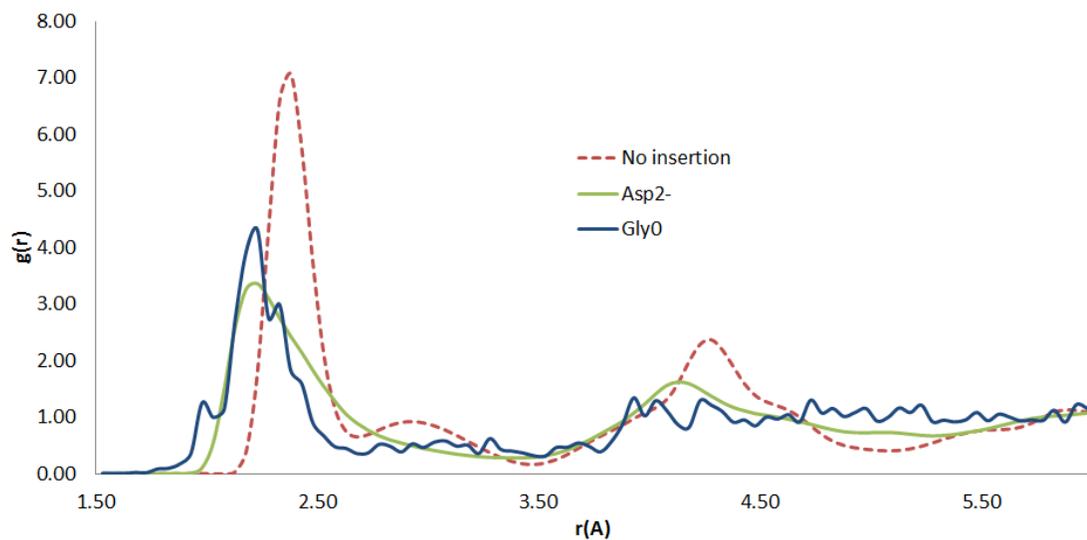
Figure 3.18: Radial distribution function of the calcium to carbonate-oxygen distance within the calcite crystal with no inclusion (red dotted line), 2.3 mol % Asp^{2-} inclusion (green solid line) and 2.8 mol % Gly^0 inclusion (blue solid line).

(0.25 mol %), whereas in the case of Asp^{2-} the onset of this disruption starts later (0.75 mol %), indicating that the calcite crystal can accommodate the inclusion of Asp better than the inclusion of Gly.
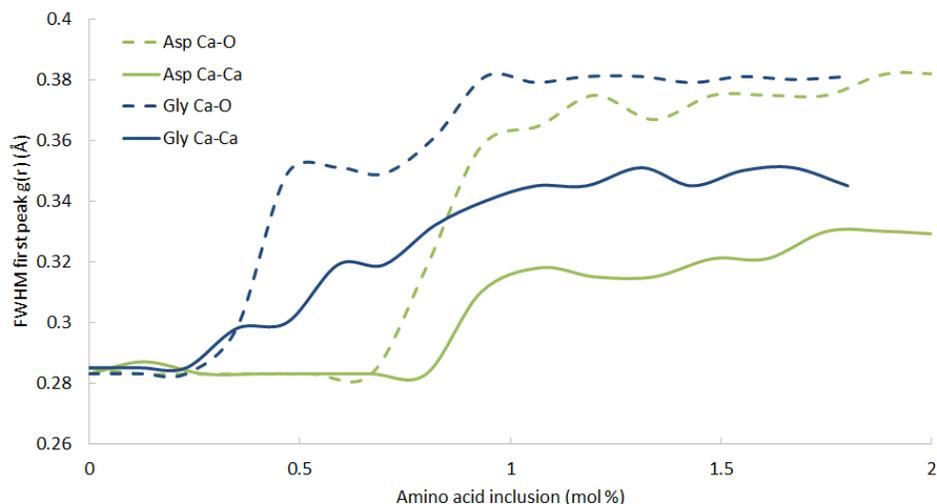
Figure 3.19: Full width at half maximum analysis of the calcium and calcium to carbonate-oxygen RDF data for Asp^{2-} and Gly^{0} inclusions.

The RDF between the calcium to carbonate-oxygen can also be compared to the RDF of the calcium to carboxyl-oxygen of the included amino acids. These data (Figure 3.20) show that the peaks of the calcium to carboxyl-oxygen for the inclusion of Asp and Gly are comparable to the calcium to carbonate-oxygen distances in the original calcite lattice, suggesting that the carboxyl functional groups of the included amino acids occupy the positions of the carbonate ions in calcite. This effect is more pronounced with the inclusion of Asp, showing well defined peaks, whereas Gly shows a more disrupted positioning. Asp is seen to be incorporated within the calcite crystal such that both carboxyl functional groups mimic the distance and positions of the carbonate ions. With the inclusion of Gly, only one carboxyl functional group is present, resulting in a less favourable positioning.

Visual analysis of the inclusions of these amino acids confirm this behaviour, where the Asp carboxyl groups mimic the positions of carbonate groups. The amine functional group is situated in the place of the vacant calcium (Figure 3.21). Although during the
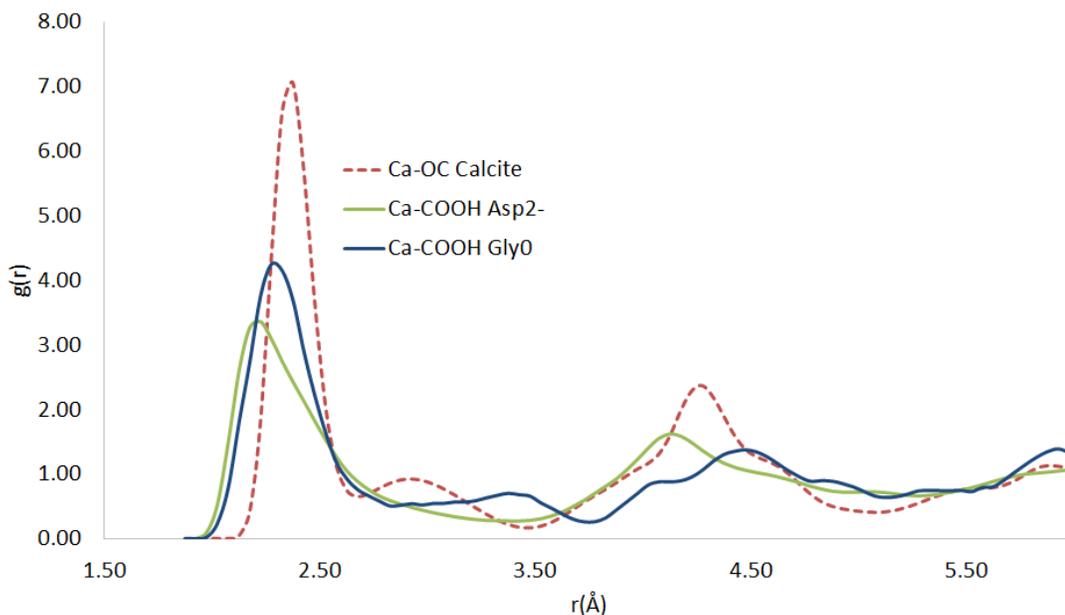
Figure 3.20: Radial distribution function of the calcium to carbonate-oxygen within calcite without inclusions (red dotted line) and the calcium to carboxyl-oxygen distances for 2.3 mol % Asp^{2-} inclusion (green solid line) and 2.8 mol % Gly^0 inclusion (blue solid line).

setup of the simulations the amino acids are placed within the vacancies of the crystal in a random orientation, throughout the simulations the majority of the amino acids rotate and position themselves such that the carboxyl groups mimic the carbonate vacancies in the crystal. There are a few exceptions and, as mentioned before, the inclusion energy (Figure 3.16) for some of the amino acids show an irregular behaviour, particularly pronounced with the inclusion energy for Asp^{2-}. The discrepancy of these data points can be attributed to local distortions caused by the misalignments of the included amino acids in the crystals. When analysing the simulations in which the energy of inclusion was particularly high, the amino acids were found not to occupy the ion sites directly, and most likely the amino acid has been positioned such that a rotation or translation of the molecule would result in a large energy barrier, causing the configuration to get stuck in a local rather than global minimum.
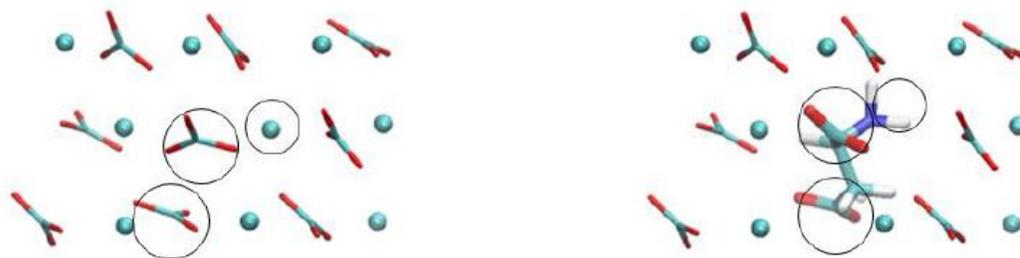
Figure 3.21: Visual analysis of the inclusion of Asp^{2-} within the calcite crystal. The amine functional group is situated in the place of the vacant calcium ion and the carboxyl groups mimic the positions of carbonate groups.

3.3.3 Anisotropy of calcite

In addition to analysing the local distortions in the lattice due to the interaction between the amino acids and the calcite crystal, simulations were performed to analyse the effect that the inclusion of amino acids have on the global lattice parameters of calcite. Data obtained from X-ray diffraction experiments by Borukhin *et al.* [133] and Kim *et al.* [134] and Raman spectroscopy experiments by the latter research group all show that the calcite crystal responds anisotropically to the inclusion of amino acids, with the *c*-direction of the crystal experiencing a significantly higher disruption than the *a*- and *b*-directions. With the molecular dynamics simulations of calcite without inclusions and with the inclusions of Asp^{2-} and Gly^{0}, the total configurational energy of the system was monitored whilst the simulation box was expanded manually along the crystal a/b-axes or the *c*-axis independently. This method allowed the generation of a three dimensional plot of the configurational energy of the crystal under the strain of expansion. These three dimensional plots are shown in Figures 3.22 to 3.27, which show the configurational energy of the systems as a function of the change of lattice parameters ($\Delta ab/ab$) and ($\Delta c/c$) with respect to the original calcite lattice parameters. In these plots the configurational energy is normalised between 0 and 1. For clarity, these plots have also been converted to contour plots, shown in Figure 3.28.
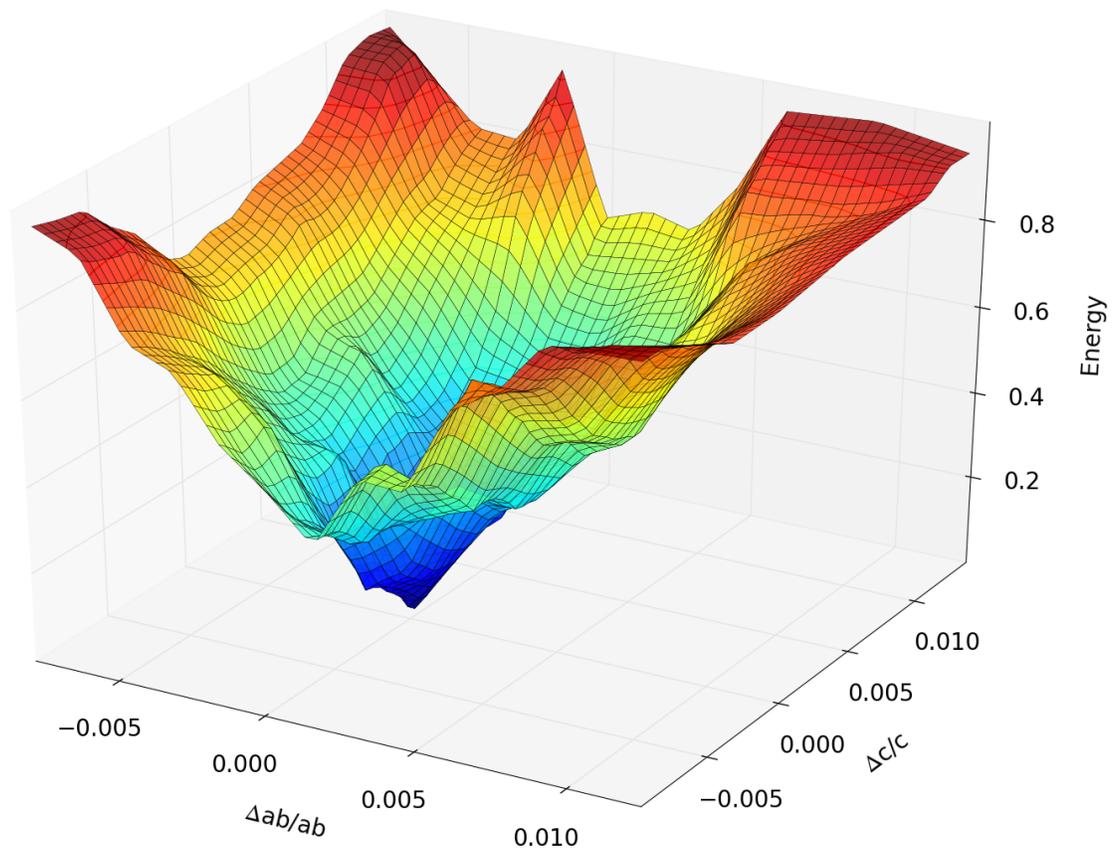
Figure 3.22: Three dimensional surface of anisotropy of calcite with no inclusions.

(a) a/b-direction.

(b) c-direction.
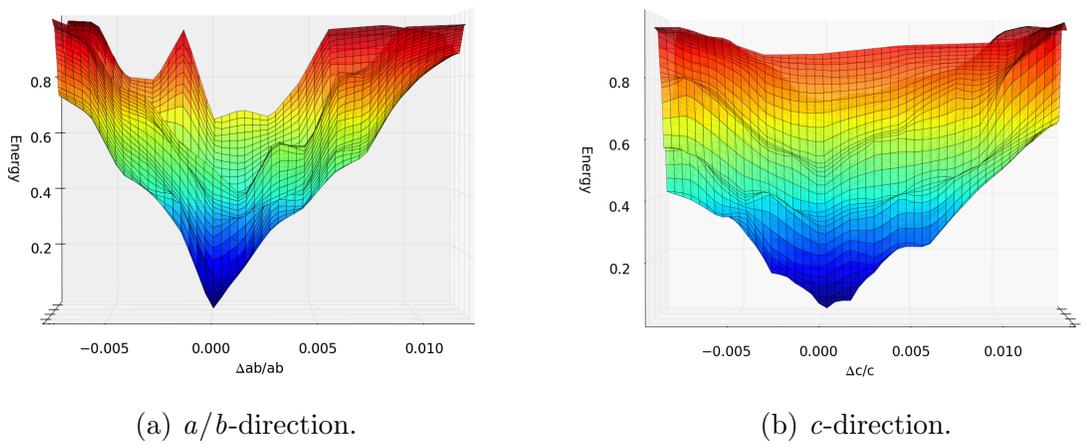
Figure 3.23: Three dimensional surface of anisotropy of calcite with no inclusions.
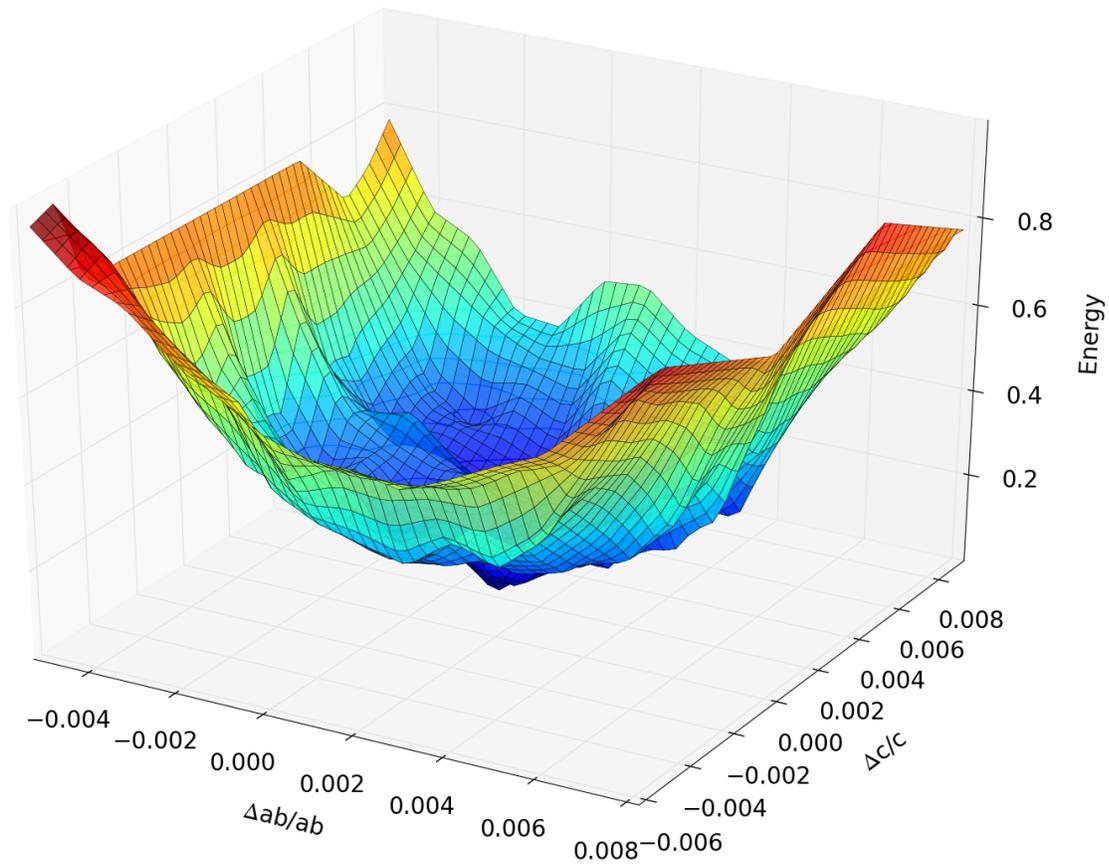
Figure 3.24: Three dimensional surface of anisotropy of calcite with included Asp^{2-}.

(a) a/b-axis

(b) c-axis
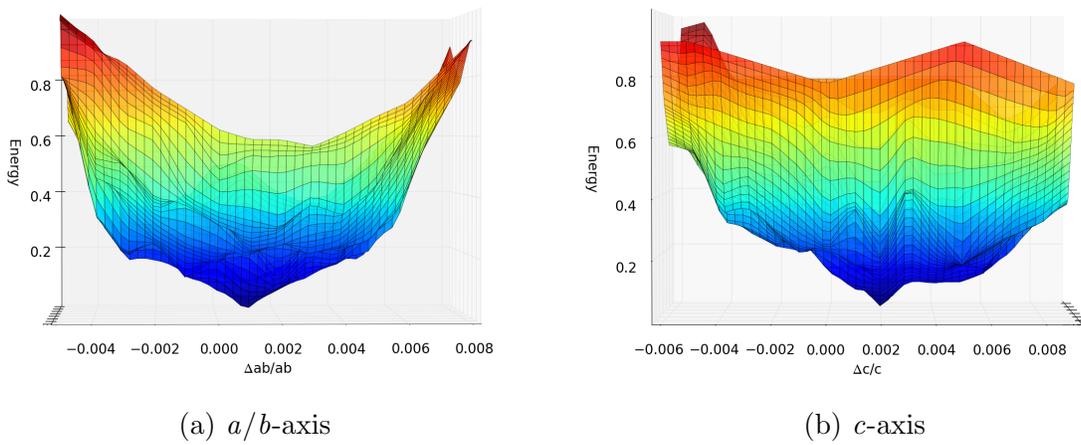
Figure 3.25: Three dimensional surface of anisotropy of calcite with included Asp^{2-}.
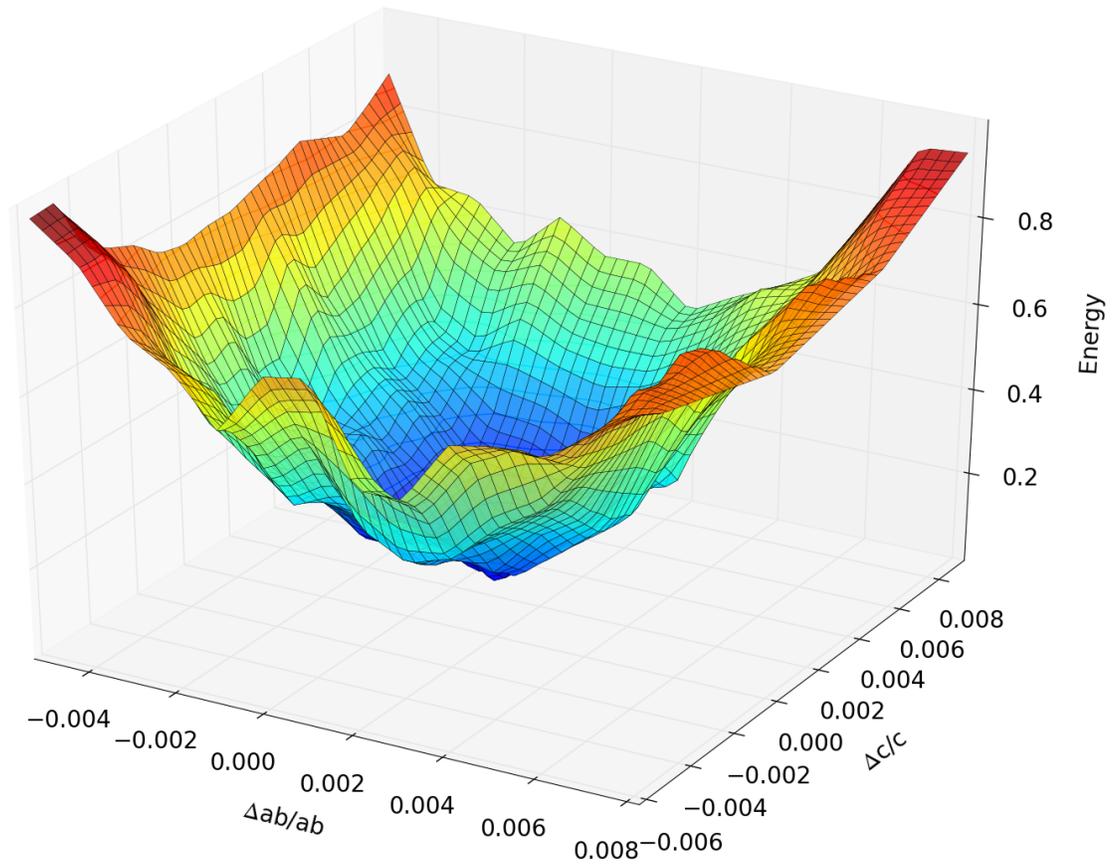
Figure 3.26: Three dimensional surface of anisotropy of calcite with included Gly^0.

(a) a/b-axis

(b) c-axis
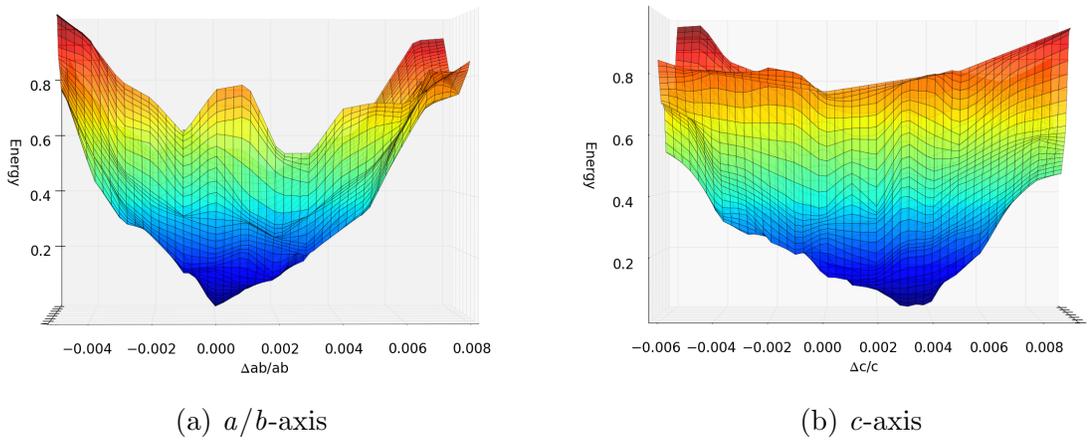
Figure 3.27: Three dimensional surface of anisotropy of calcite with included Gly^0.

When analysing the data for the expansion of calcite without any inclusions, the energy minimum is found when no expansion has occurred, which can be expected from a calcite structure in its natural, low energy conformation. Expansion of this crystal in both the a/b- and c-directions results in unfavourable configurations as the configurational energy increases. The expansion in the a/b-direction is less favourable by a factor of two than the c-direction, as can be seen from the steeper increase in energy along this direction. The expansion of the crystal with 2.3 mol % included Asp^{2-} shows a favourable expansion along the c-direction, and the energy minimum for the crystal is found to be shifted such that the lattice parameters are distorted by $\Delta ab/ab = 0.0008$ and $\Delta c/c = 0.002$ with respect to the original calcite structure. This preference of expansion along the c-axis can further be observed by the differences in gradient of the configurational energy between the a/b- and c-directions (Figures 3.25a and 3.25b respectively).

A similar effect on the calcite crystal is seen for the inclusion of 2.8 mol % Gly^0 (Figures 3.26 and 3.27), where the minimum energy configuration is found with an expansion of 0.0002 and 0.003 for $\Delta ab/ab$ and $\Delta c/c$ respectively. Similarly, the configurational energy of the system show that the steeper gradients are observed when moving along the ab-axis, revealing the anisotropic behaviour of the calcite crystal upon inclusion of Gly. Data of these simulations are in good agreement with the X-ray diffraction data obtained by Kim *et al.* [134] who found values of $\Delta ab/ab = 0.0004$ and $\Delta c/c = 0.0025$ for similar Asp inclusions and $\Delta ab/ab = 0.0005$ and $\Delta c/c = 0.0038$ for similar Gly inclusions.

When analysing biogenic calcite samples that have intra-crystalline biomolecules included within the crystal, the same anisotropic effects can be observed, as shown by Pokroy *et al.* [87]. This study analysed the disruption to the lattice parameters of biogenic calcite and found maximum $\Delta a/a$ distortions of 0.0008 and maximum $\Delta c/c$ distortions of 0.002 compared to geological calcite samples. In all cases the distortion in the c-direction was significantly higher than the a-direction. They explained this
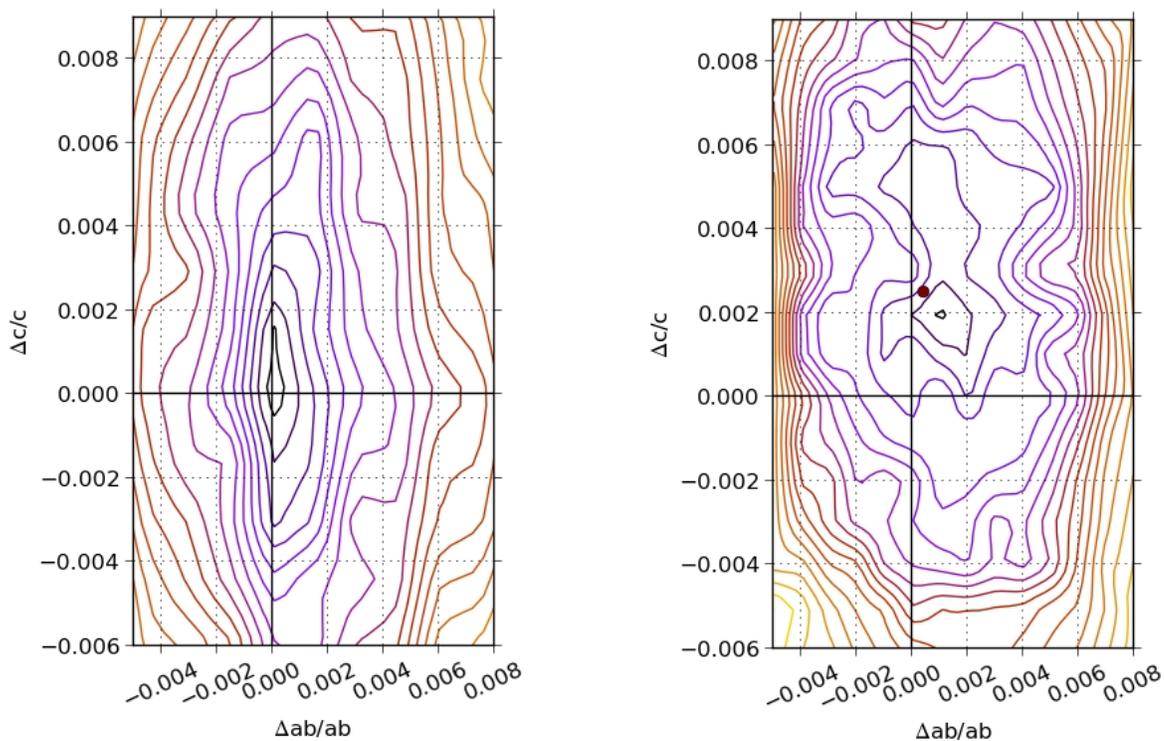
(a) No inclusion.

(b) 2.3 mol Asp^{2-} inclusion.
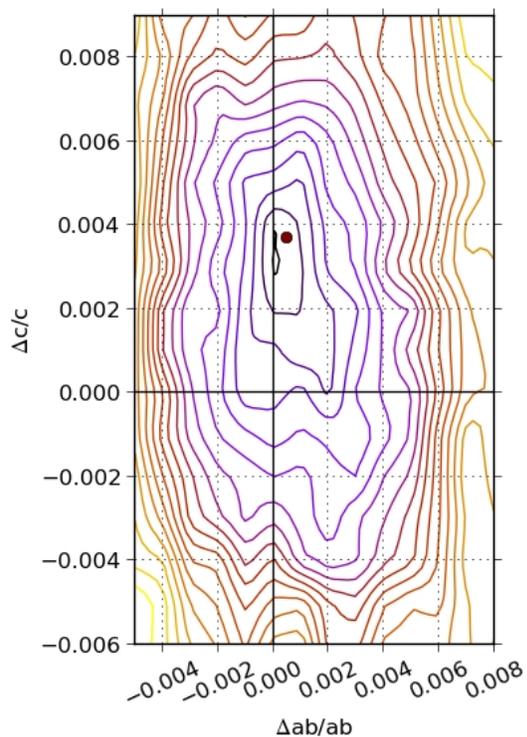
(c) 2.8 mol Gly^0 inclusion.

Figure 3.28: Contour plots of the anisotropy of calcite crystal. The 0,0 point indicated by the crossed lines represents the energy minimum of calcite without inclusions. The red dot on the plots for Asp and Gly represents the expansion seen in experimental data by Kim *et al.* [134].

anisotropic behaviour of calcite with the fact that the intra-crystalline biomolecules are situated on the basal plane of calcite, perpendicular to the c-axis of the crystal. This causes the carbonate and calcium layers of the crystal to be pushed away from each other by the incorporation of these biomolecules, resulting in a greater disruption in the c-direction than in the a- or b-directions of the crystal. The same can be observed when visually analysing the molecular dynamics simulations in this work (Figure 3.29). The included amino acids position themselves in between the planes perpendicular to the c-axis of the crystal. This causes the crystal planes to be pushed further apart, resulting in an anisotropic distortion.
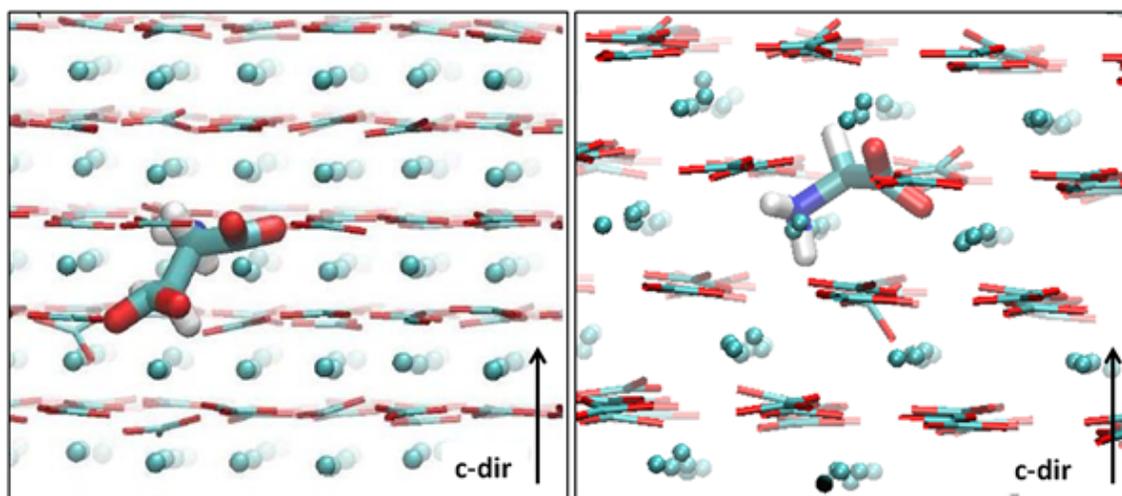
Figure 3.29: The amino acids Asp and Gly included within the calcite crystal are observed to position themselves between the planes perpendicular to the c-axis of the crystal, causes the crystal planes to be pushed further apart, resulting in an anisotropic distortion.

3.4 Conclusion

The molecular dynamics simulations of the inclusion of amino acids performed here show that the inclusion of Asp and Gly within a calcite crystal results in a high energy of inclusion. Together with the high inclusion levels found in experiments this supports

the argument that the inclusion of these species is controlled via a kinetic rather than via a thermodynamic pathway. The high energies obtained from the vacancy defect simulations and the inclusion of water reaffirm this point of view.

Structural data obtained from these simulations suggest that Asp is more readily incorporated within the crystal whilst causing a minimal disruption to the crystal lattice. The inclusion of Gly on the other hand causes more pronounced lattice distortions, evident from the RDF and FWHM data. This can be explained by the geometry of Asp where the distances between both carboxyl functional groups are such that upon inclusion the amino acid can mimic both the distances and positions of vacant carbonate ions, with the amine functional group occupying a vacant calcium position. With no side chain and only one carboxyl group, the positioning of Gly within the crystal is less favourable. These results are in good agreement with recent experimental results [134], that suggest an easier incorporation of Asp compared to Gly. The defect energy and the structural data from these simulations suggest that a 'goodness of fit' principle is maintained, whereby molecules will get incorporated within the calcite crystal if the disruption to the lattice can be minimised.

Simulations on the expansion of the calcite crystal in both the a/b- and c-directions show the same anisotropic of the calcite behaviour upon inclusion as seen in various experiments. The expansion of the crystal in the c-direction is an order of magnitude larger than the expansion in the a/b-directions. The energy minima at expansions of $\Delta ab/ab = 0.0008$ and $\Delta c/c = 0.002$ for Asp and $\Delta ab/ab = 0.0002$ and $\Delta c/c = 0.003$ for Gly are similar to those from X-ray diffraction data. These values also show a remarkable similarity to biogenic calcite with intra-crystalline proteins, suggesting that the inclusion of both small and large biomolecules result in similar structures. The current simulations show that the included biomolecules are positioned in between the planes perpendicular to the c-axis of the crystal, as previously proposed by Borukhin *et al.* [133].

Chapter 4

Conformational behaviour of tripeptides on calcite surfaces

4.1 Introduction

As described in Chapter 1.1, biomolecules such as proteins, peptides, polysaccharides and polymers can alter the polymorph selection, crystal face selection and nucleation and growth rates of calcium carbonate crystals. Some of the proteins that are involved with the biomineralisation processes in natural organisms have been found to be intrinsically disordered proteins (IDPs) - proteins that lack a fixed three-dimensional ordered structure. In this chapter the results of computational simulations and experimental studies of a short tripeptide (Glu-Asn-Gly) are discussed. This tripeptide is used as model system for IDPs and its conformational behaviour and adsorption onto calcite are studied.

4.1.1 Intrinsically disordered proteins

The paradigm in structural biology that the function of a protein is linked to its three-dimensional structure is widely used and accepted. However, many proteins and protein

domains exhibit little or no globular structure whatsoever, suggesting that this view should be reassessed. Over the years the appearance of intrinsically disordered (or unstructured) proteins has been discussed in a range of reviews [142–145]. IDPs are of two kinds: intrinsic molten globular structures, containing partial folded regions, or random coiled structures [144]. The structures of proteins and the disorder within them can be assessed using various techniques, such as X-ray crystallography, nuclear magnetic resonance (NMR), ultraviolet-visible (UV-vis) spectroscopy, hydrodynamic techniques (gel filtration, dynamic light scattering) or circular dichromism (CD) spectroscopy. CD spectroscopy can provide structural information of proteins by detecting the presence of secondary structures such as α-helices, β-sheets, β-turns and random coils. This technique is widely used to assess the secondary structures of IDPs.

The amino acid sequence (primary protein structure) can also provide information on the structure of the protein. Amino acids[1] that induce flexibility, low hydrophobicity and a high net charge are indicators of the absence of structure in proteins and the presence of amino acids such as Ala, Glu, Gly, Lys, Pro, Gln, Arg, Ser, Asp and Asn are found to promote the disordered structures in proteins [144, 145]. *Vice versa*, the inclusion of amino acids such as Cys, Phe, Ile, Leu, Val, Trp and Tyr are depleted in disordered protein structures.

Robinson and Robinson [146] performed a statistical study of unstable and flexible residue triplets and found that Glu and Asn are often found within these triplets. With triplets involving Asn, the nearest neighbours most frequently found on either side of this residue were Gly and Glu, with the tendency for Gly to be on the right of Asn, stabilising the normally flexible amino acid Asn. Within other studies of flexible domains of proteins and peptides, such as hairpin turns and reverse turns, the residues Asn and Gly are often seen to be involved [147–149]. This Asn-Gly repeat is the same sequence that is found within many biomineralisation proteins, such as AP7 and n16, as described below.

---

[1]A list of amino acids is included in Appendix A.

There are several advantages of proteins or protein domains to have a disordered structure. The flexibility involved with the disorder creates a structure that is adaptable and can bind to a target (e.g. another protein or a surface) via multiple points of contact. The flexibility of the peptide chain allows the spacing between these points of contact to vary, depending on the target. Whereas globular proteins with a highly ordered three-dimensional structure will have limited binding targets, unstructured proteins have versatile interactions and can bind to multiple targets, a so-called one-to-many signalling mechanism. Some IDPs have also shown the opposite behaviour, a many-to-one mechanism, where recognition of the target does not depend on a precise sequence but rather the presence of select residues in a chain [150]. Furthermore, IDPs have been shown to display disorder-to-order transitions, an induced folding of their structure when binding to a target. This disorder-to-order transition is coupled to a decrease in the conformational entropy of the protein, and thought experiments by Schulz [151] proposed that the binding of IDPs could lead to a gain in free energy (Figure 4.1) and lead to a high specificity coupled to a low affinity of the protein.
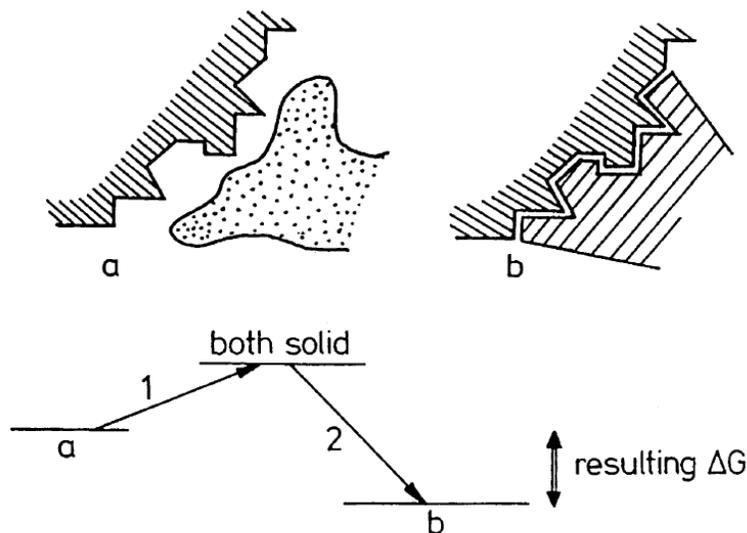
Figure 4.1: Thought experiment of disorder-to-order transition within IDPs. The binding of an IDP to a target is coupled to a gain in free energy. The flexibility and disorder-to-order transition of the IDP when binding to a target leads to a combination of a high specificity and low affinity of the protein. Reprinted from [151] with permission from Elsevier.

4.1.2 Intrinsically disordered proteins and peptides in biomineralisation

Studies in the biomineralisation research area have uncovered a vast range of IDPs within the organic matrices of organisms that have various effects on the crystal growth of calcium carbonate. An overview of developments of IDPs in biomineralisation can be found in a review by Evans [152]. Biomineralisation associated proteins are often found to exhibit open, extended, conformations and are found to have a large acidic residue content. The presence of acidic residues (Asp and Glu) is not surprising, because these negatively charged residues are found to be able to coordinate to Ca^{2+} ions. Three examples of IDPs in the area of biomineralisation are given below: The AP7 & AP24 proteins from a mollusc shell, the Starmaker protein and the small n16 protein. Many more IDPs are involved in the biomineralisation of $CaCO_3$, such as the Asprich protein family [153, 154], SM50 [155, 156], Lustrin A [157] and PM27 [158], but will not be discussed here.

Growth control

The proteins AP7 and AP24 (Aragonite Protein with molecular weight of 7 kDa (1 Da = 1 g mol^{-1}) and 24 kDa respectively) are found in abalone molluscs [159–162] (Table 4.1). Although UV-vis and CD spectroscopy studies of both proteins show small regions of α-helices and β-sheets in the secondary structure, the majority of the proteins exhibit extended and random coil conformations. With *in vitro* experiments both proteins are seen to affect the morphology of calcite crystals, where rounding of the calcite steps is observed (Figure 4.2). Additionally, at high concentrations of AP7 in solution, the protein can induce the formation of amorphous and aragonite aggregates [162]. The 30 amino acid N-terminus of these proteins (AP7-N and AP24-N) contain a high amount of the acidic Asp and Glu residues and are Ca^{2+} binding domains. When analysing the effect of the N-terminus binding domains alone on calcite, a similar but less pronounced

effect can be seen compared to the full proteins. Interestingly, Kim *et al.* [160] used the peptide sequence of AP7-N and scrambled the primary amino acid sequence to form a random version of the peptide, AP7-Ns. The effect of AP7-Ns on growing calcite crystals was similar to that of the native AP7-N, suggesting that the amino acid composition of these proteins is more important than their amino acid sequences.
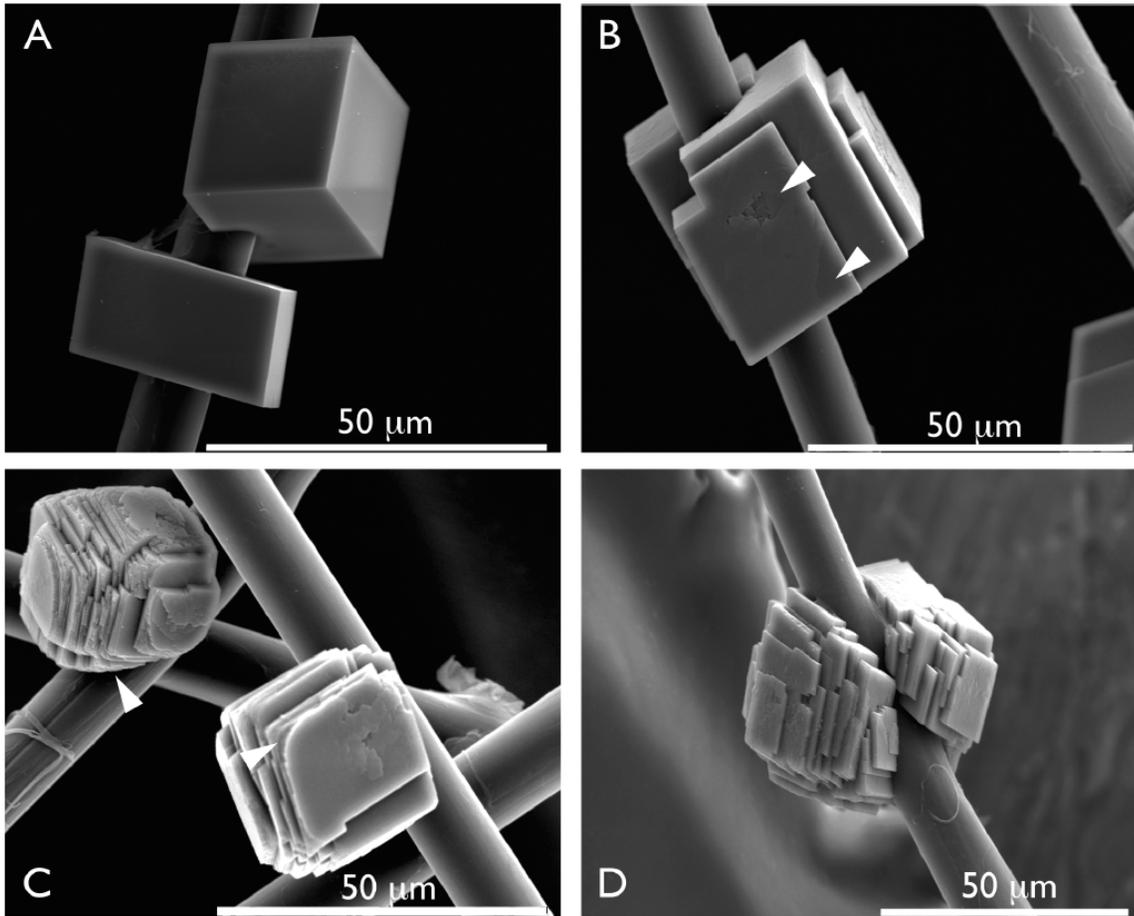
Figure 4.2: SEM images of the morphology of calcite crystals in the presence of a) no additives, b-d) 5, 10, 15 μM AP7. The calcite crystals show rounded edges when nucleated in the presence of the AP7 protein. Reprinted (adapted) with permission from [162]. Copyright 2009 American Chemical Society.

Polymorph selection

Otoliths are small mineral crystals present in the inner ear of fish that are part of their gravity and acceleration detection system. These otoliths are made up of small aragonite

Protein	Amino acid sequence
AP7-N	DD<u>NG</u>NYG<u>NG</u>MASVRTQGNTYDDLASLISYL
AP7-Ns	AYVQGNTDNADMD<u>NG</u>IRYSLLYNDTSLGSG
AP24-N	ADDDEDASSGLCNQYNQNVTTRPNNKPKMF
Stm[a]	NN<u>NG</u>TDNDESAADQRHIFTVQFNVGTPAPA
	DGDSVTTDGKDSAEKNEAPGDSSDTTEKPG
	TTDGKDSAEQHGVTTDGKDEAEQHGVTTDG
	QDSAEKRGEADGAPDKPDTQ<u>NG</u>TDDTDSDQ
	ETDASHHKTGDSD<u>EN</u>KDKPSAEDHTDGNHA
n16-N	AYHKKCGRYSYCWIPYDIERDRYD<u>NG</u>DKKC
n16-Ns	EPRYCKWCDNKHGDRAGCKYSIDYYKIRDY
n16-NN	AYHKKCGRYSYCWIPYNIQRNRYN<u>NG</u>NKKC
n16.3	AYHKKCGRYSYCWIPYDIERDRYD<u>NG</u>DKKC
	CFCRYAWSPWQCNEEERYEWLRCGMRFYSL
	CCYTDDD<u>NGNGNGNGNGNG</u>LNYLKSLYGG
	YG<u>NG</u>NGEFWEEYIDERYDN

Table 4.1: Primary amino acid sequences for peptides seen to affect the nucleation and growth of $CaCO_3$. The Glu-Asn and Asn-Gly repeat units are underlined in each sequence. [a] Only residues 55 to 205 shown. Sequences from [59, 159, 163, 164].

crystals enveloped in an organic matrix that is mainly made up of (glyco)proteins[2]. One of the proteins in this organic matrix that has been found to be involved in the aragonite polymorph selection of the otoliths is the protein Starmaker (Stm, Table 4.1). *In vivo* experiments [164] performed on wild-type and Stm knock-out zebra fish showed that in the absence of the Stm protein the normal aragonite composition of the otoliths disappeared in favour of the calcite polymorph (Figure 4.3). CD spectroscopy and *in silico* predictions [165] show that the secondary structure of Stm is highly disordered and the presence of large hydrophilic regions and a high number of acidic residues (24.6 % Asp, 10.6 % Glu) cause the protein to adopt an open, extended conformation.

Another protein involved with polymorph selection is the n16 protein from the Japanese pearl oyster which is involved in the formation of nacre platelets. Experimental studies have shown that the protein can stabilise aragonite formation *in vitro* [59]. Using the N-terminus domain (n16-N) of the protein, thought to be involved in the binding

---

[2]Glycoproteins are proteins that have oligosaccharides bound to the residue side chains.
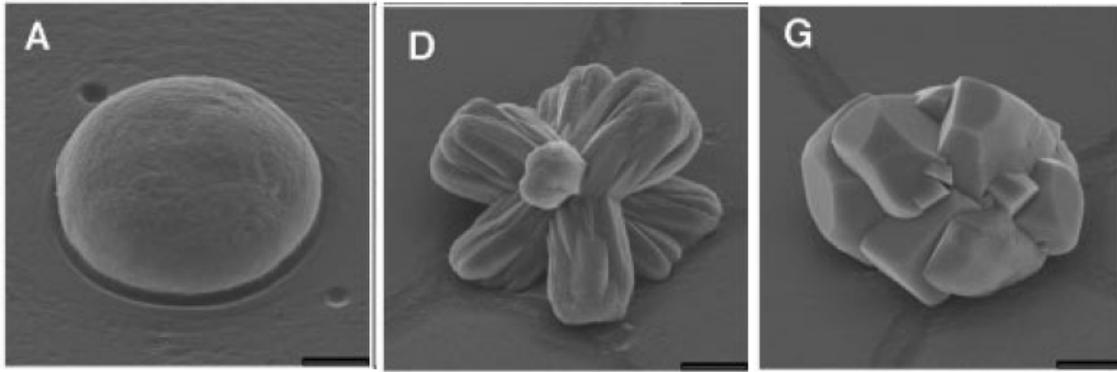
Figure 4.3: Effect of the starmaker (Stm) protein on $CaCO_3$ crystals in zebra fish. In wild type (a), aragonite crystals are formed. With a decreased amount of Stm (b), the formed aragonite crystals are misshapen. With the Stm gene completely knocked out, calcite crystals are formed (c). From [164]. Reprinted with permission from AAAS.

to $CaCO_3$, researchers were able to control the nucleation of $CaCO_3$, with vaterite, aragonite, calcite and ACC particles being formed [61, 62, 166–168]. Two synthetic modifications were made to the primary sequence of the n16-N peptide, one where the sequence was scrambled to produce a random peptide (n16-Ns) and one where the acidic residues were substituted (Asp $\rightarrow$ Asn, Glu $\rightarrow$ Gln) to produce an uncharged peptide (n16-NN, Table 4.1). The *in vitro* experiments [61, 62, 169] using these two modified peptides showed that scrambling the primary sequence did not alter the effect of the polypeptide much, but the substitution of the Ca^{2+} binding residues diminished the effects, indicating that the overall amino acid sequence is less important to the function of the peptide than the inclusion of specific residues.

Experiments performed [170] with a 108 amino acid N-terminus peptide chain (n16.3) showed that the IDPs oligomerize in solution to form protein films. The formation of these supra-molecular clusters is promoted by the many Asn-Gly repeat units in the peptide which, apart from inducing disorder in the protein, provide a strong agglomeration effect.

Computational simulations on the n16-N peptide have been performed by Brown *et al.* [171] using accelerated molecular dynamics simulations to ensure a sufficient sampling of the conformational space. Using replica exchange solute tempering, a simulation

method in which the phase space of a system is explored at different temperatures, the authors were able to confirm the findings by Collino and Evans [166] that the peptide conformation is mainly a random coil. The agglomeration of these peptides in solution was also analysed, and although the simulations were too short for statistical analysis, evidence of a large number of hydrogen bonds between two peptides indicated agglomeration is likely, as has been observed in experiments [170].

4.1.3   Aims

In the three examples described above the structural conformations of the peptides are disordered and the function of these peptides is thought to be carried out with a disorder-to-order transition upon the binding of these molecules to $CaCO_3$, a transition that is seen as thermodynamically favourable for these disordered peptides. Within this work we have employed the conformational analysis of a short tripeptide, Glu-Asn-Gly, to analyse the conformational behaviour of peptides in solution and at the surface of $CaCO_3$. This short tripeptide was selected for various reasons. Firstly, the Asn-Gly repeating sequence can be observed in proteins involved with biomineralisation processes such as AP7, Stm and n16, and primarily within their $CaCO_3$ binding domains. This sequence also induces the agglomeration of the peptides and is prone to induce disorder within peptides due to their flexibility. The addition of the Glu residue was based on the statistical preference of this residue next to the Asn residue found in peptides and proteins. In addition, the Glu-Asn-Gly peptide is readily available to purchase from chemical companies and as such provides an easy route for experimental analysis of the tripeptide. Using this small peptide as a model system of IDPs, the aim of this project is to gain an understanding of the peptide - peptide interactions of flexible molecules and the peptide - calcite interactions.

4.2 Conformational behaviour

4.2.1 Methods

The conformational behaviour of the Glu-Asn-Gly tripeptides in solution and on the $(10\bar{1}4)$ surface of calcite was analysed using molecular dynamics simulations. Details of the simulation setup, protocols and analysis techniques are described below.

System setup

In order to analyse the conformational behaviour and preferred secondary structure of the Glu-Asn-Gly peptide in solution, simulations of one, two, three, six and ten peptides in an aqueous environment were performed. The tripeptide Glu-Asn-Gly (Figure 4.5 and Appendix C) was prepared using the `Amber` 12 `Antechamber` package [138] and the atomic charges of the molecules were calculated using the AM1-BCC method [139] (see Appendix B).

For the simulations of the peptides in solution, one, two, three, six or ten peptides and 20,000 TIP3P water molecules [104] were inserted in a simulation box in random positions and orientations using the program `Packmol` [137], ensuring that the density of the system was 1.0 g ml^{-1}. For the simulations involving the adsorption of the peptides to the surface of calcite, a surface of calcite with a $(10\bar{1}4)$ orientation, with a depth of 26 Å (10 CaCO$_3$ layers) and surface dimensions of 81 x 83 Å, containing 3380 CaCO$_3$ formula units was generated using the program `buildcalcite.jar`[3]. The surface was solvated with 20,000 TIP3P water molecules, and the Glu-Asn-Gly peptides were placed near the surface ($<$5 Å) in a random orientation. The simulation was setup as seen in Figure 4.4, to allow periodic boundary conditions to be used in all three directions.

---

[3]Program developed by Robert Darkins at The London Centre for Nanotechnology, University College London.

102
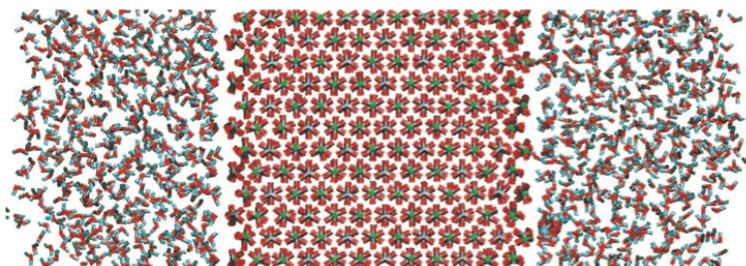
Figure 4.4: Schematic representation of the simulation setup used in this work. Using periodic boundaries the system is enclosed within a three dimensional box.
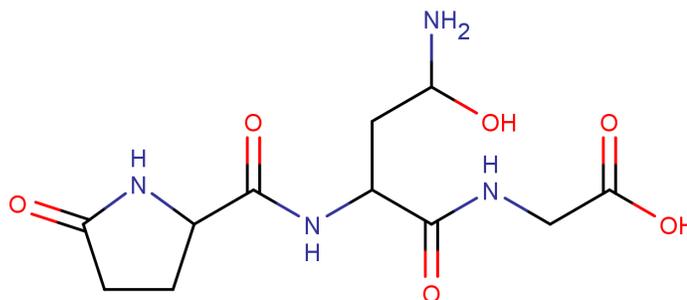
Figure 4.5: Chemical structure of the Glu-Asn-Gly peptide used throughout this work.

Modelling protocols

For the peptide simulations the `AMBER` *ff12SB* potentials [111] were used to describe the intra- and inter-molecular interactions between the peptides and water molecules. In the simulations involving the mineral phase, the potentials developed by Raiteri *et al.* [26] were used to describe the inter- and intra-molecular interactions of the calcite. The cross-terms between the organic and inorganic phases were calculated using the method described by Freeman *et al.* [103], see Chapter 2.2.4. The details of the potentials used are described in Appendix F.

All simulations of the tripeptides, water and calcite were carried out using the molecular dynamics package `DL POLY Classic` [140]. The systems were equilibrated at a temperature of 300 K for 100 ps with a timestep of 0.1 fs in an NVT ensemble with a Nosé-Hoover thermostat [119] with a relaxation time of 0.5 ps, and subsequently for 500 ps with a timestep of 1.0 fs using the same parameters. The density of the water-

box was monitored during the equilibration period to ensure the correct density was maintained. The configurational energy of the system during the equilibration simulations was analysed to ensure a fully equilibrated system. Production runs of 5.0 ns using a timestep of 1.0 fs in an NVT ensemble were performed, from which both the conformations and energetics of the Glu-Asn-Gly peptides were analysed.

Conformational analysis

The conformational behaviour of proteins and peptides has been studied since the first determination of protein crystal structures in the 1950s. In 1963 Ramachandran *et al.* [172] studied the dihedral angles of short polypeptide chains and proposed that the three-dimensional conformations of these chains can be described by the ϕ (phi, C'-N-C$_\alpha$-C') and ψ (psi, N-C$_\alpha$-C'-N) backbone dihedral angles (Figure 4.6) and can be plotted on a two dimensional map. By systematically rotating the ϕ and ψ dihedral angles in increments through 360°, the conformational space of the molecule could be explored, and the resultant plot of ϕ and ψ backbone dihedral angles is called a Ramachandran map [92]. Ramachandran *et al.* used a hard-sphere model to represent the atoms of the short peptides in their study, and molecular conformations that led to steric clashes between two atoms (atoms that occupy the same space) were classed as disallowed. Two atomic radii were used for the hard spheres in order to define an allowed region and an outer limit region, as shown in Figure 4.7a.

Lovell *et al.* [173] refined the map generated by Ramachandran *et al.* by analysing 97,368 residues from 500 protein structures. In their study they used a soft sphere atomistic model and added a hydrogen bonding term for the repulsive and attractive forces between the atoms. By using a large data set they were able to obtain a detailed map of the allowed regions of the ϕ and ψ angles (Figure 4.7b). Lovell *et al.* separately analysed the conformational behaviour of the Gly residue, as its behaviour within proteins is seen to be different than other residues. The Ramachandran map of 7705 Gly
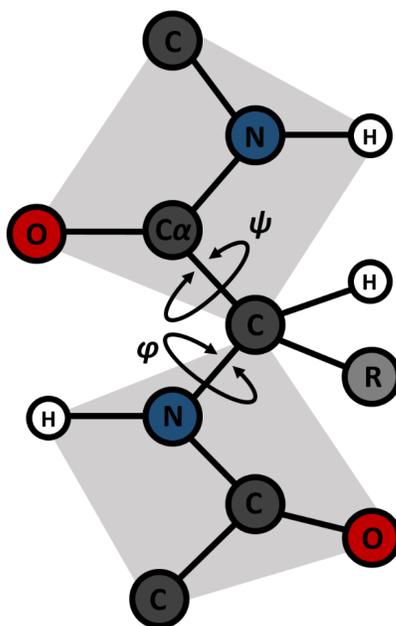
Figure 4.6: The three-dimensional conformations of peptides can be described by the ϕ (C'-N-C$_\alpha$-C') and ψ (N-C$_\alpha$-C'-N) backbone dihedral angles, which describe the orientations of the two amide planes.

residues in their data set is shown in Figure 4.7c. Due to the lack of a side chain there is less steric hindrance for this residue, enabling the ϕ and ψ angles to rotate freely and explore the right quadrant of the Ramachandran map, although the inaccessible conformations around $\phi = 0°$ are still observed due to the large steric clashes of a planar conformation. The ϕ and ψ angles of the peptide bonds also give an indication of the secondary structure of the peptide. As indicated in Figure 4.7a, peptides or proteins with ϕ and ψ angles in the upper left quadrant of the Ramachandran map will exhibit β-sheet secondary structures. The secondary structures of a right handed α-helix (Rα), turns (γ-turn and type II-turn) and the left handed α-helix (Lα) can also be deduced from the dihedral angles.

This technique of analysing the dihedral angles of a peptide has been utilised to study the conformational behaviour of the Glu-Asn-Gly peptide in the simulation work presented here. The distribution of the ϕ and ψ angles throughout the simulations gives an indication of the preferred secondary structure of the peptide. Additionally, using this technique the changes of the peptide conformation when interacting with other

105
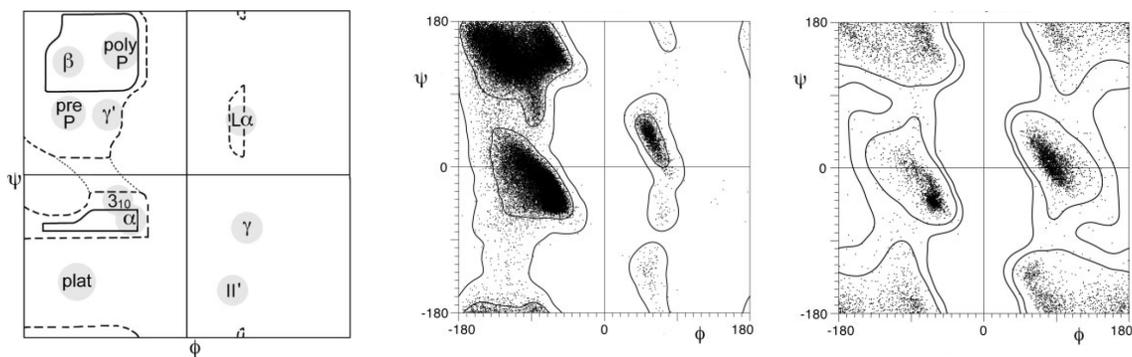
(a) Ramachandran map denotes the secondary structures associated with dihedral angles.

(b) Ramachandran map of all residues based on 97,368 residues from 500 protein structures.

(c) Ramachandran map of Gly residue only, based on 7705 residues within the data set.

Figure 4.7: The ϕ and ψ backbone dihedral angles of peptides and proteins can be plotted in a Ramachandran map, giving an indication of the conformation of the molecule. From [173]. Reprinted with permission from the Wiley Company.

peptides and when binding to the surface of calcite can be observed. Throughout the simulations the ϕ, ψ backbone dihedral angles were measured and saved every 20 fs, to produce 50,000 conformations per ns of simulation time. The dihedral data for all peptides was collected and normalised, before being plotted as a Ramachandran map with 3° x 3° bins, a bin width chosen to retain enough detail of the individual dihedral angles whilst ensuring a clear statistical representation of the data.

4.2.2   Conformation in solution

The conformational data for a single Glu-Asn-Gly peptide and multiple (three, six and ten) peptides in an aqueous environment were analysed and are described below. Note that as the conformational data was obtained in a similar fashion for the single peptides as the multiple peptides, more data points are available for analysis of the multiple peptides in solution.

106

Single peptide

The conformational data for a single Glu-Asn-Gly peptide in solution is shown in Figure 4.8a. Throughout the simulations the peptide is seen to explore the whole conformational space of the ϕ and ψ dihedral angles, an effect that with larger peptides is often only seen when a bias is applied to the dihedral angles (systematic rotation around the angles). As the tripeptides in these simulations are relatively small, the ns time scale simulations were enough to sample most of the conformations. Although the entirety of the Ramachandran map was sampled, the vast majority of the peptide conformations were observed in the upper left quadrant of the map ($\phi = $ -160° and $\psi = 70$°) appearing to be the lowest energy conformation of the tripeptide free in solution. As discussed previously, residues in peptides and proteins exhibiting similar ϕ and ψ dihedral angles to these are primarily associated with a β-sheet secondary structure.

Two peptides

The data for two peptides in solution sample the same extent of the conformational space as a single peptide in solution (Figure 4.8b). However, during the simulations the peptides are seen to interact with each other, causing the lowest energy conformation of the peptides to shift away from the upper left quadrant of the Ramachandran map. Three major conformations are observed, around $\phi = 130$°, $\psi = $ -160°; $\phi = 130$°, $\psi = $ -70° and $\phi = 50$°, $\psi = 50$°. Analysis of the structures observed throughout these simulations reveals strong interactions between the two peptides, causing an aggregation of the peptides. The interactions between the two peptides are caused by a number of hydrogen bonds between the two molecules (Figure 4.9). These hydrogen bonds between the two peptides are long lasting throughout the entire simulation, and the total life times for these bonds are shown in Table 4.2. A hydrogen bond is defined between a donor atom (D) with a hydrogen (H) on it and an acceptor atom (A) as shown in Figure 4.11, where the distance between D and A is less than 3.5 Å and
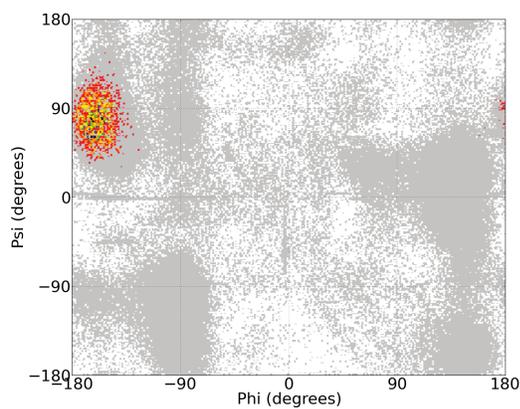
(a) One peptide.
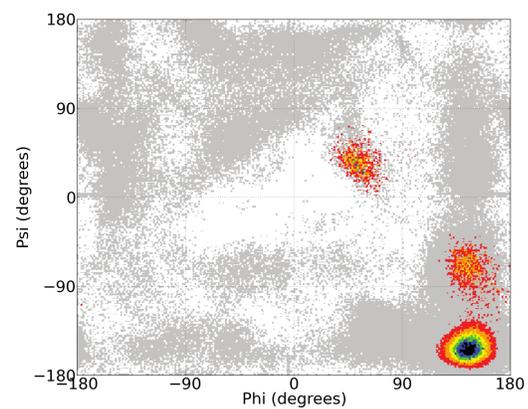
(b) Two peptides.
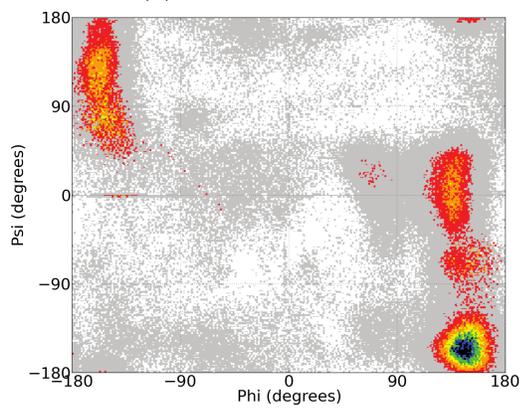
(c) Three peptides.
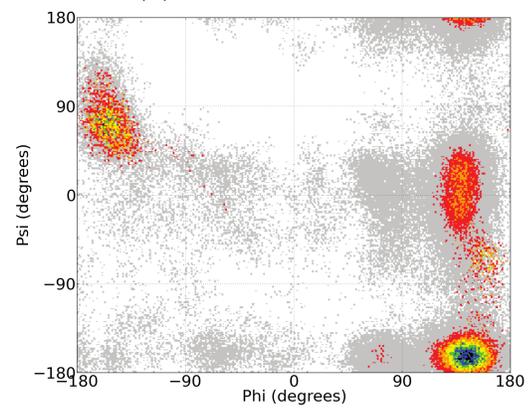
(d) Six peptides.
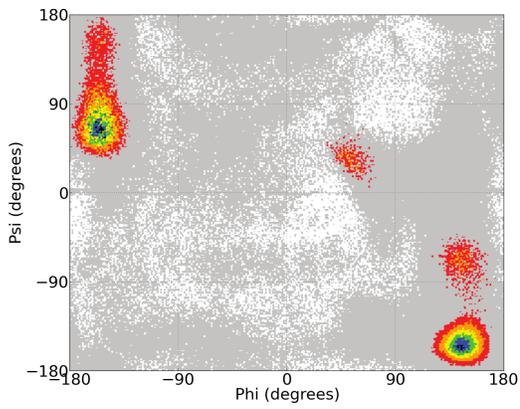
(e) Ten peptides.

Figure 4.8: Ramachandran maps obtained with MD simulations for the Glu-Asn-Gly peptide(s) in solution. The colour coding represent a linear scale for the number of counts in each bin, with increasing count from grey, red, orange. yellow, green, blue, black.

the D-H-A angle deviates no more than 20° from a linear bond. The hydrogen bonds between the Glu-Asn-Gly peptides are formed between the amide-hydrogens as donors and the carboxyl functional groups as acceptors (Figure 4.10).
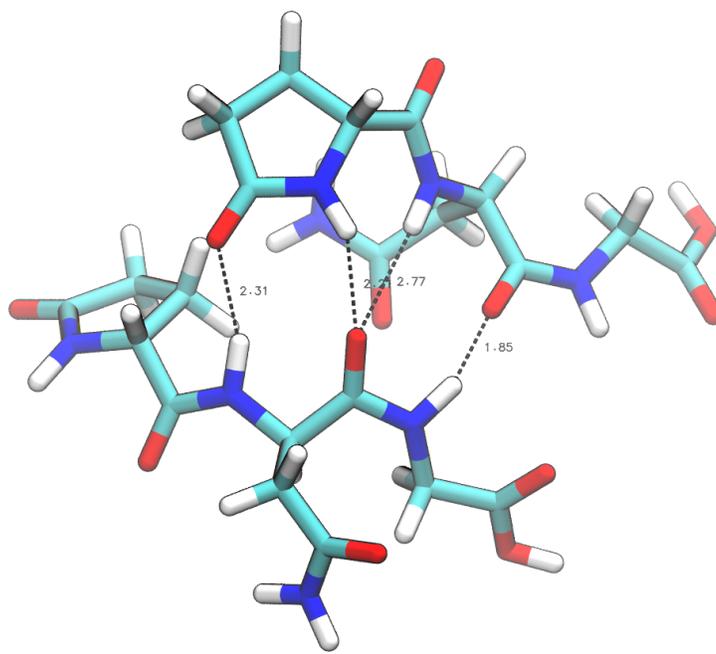
Figure 4.9: Hydrogen bond interactions between two peptides in solution observed throughout the MD simulations.

The conformation of the peptides seen at $\phi = 50°$, $\psi = 50°$ is normally associated with a left handed α-helix secondary structure, whereas the other two conformations ($\phi = 130°$, $\psi = $ -160° and $\phi = 130°$) are normally only associated with Gly residues, due to the steric strain between the side chains and the rest of the molecule. The agglomeration of these peptides and the strong interaction between them must overcome the energy barrier associated with this conformation and stabilises these conformations to make it them energetically viable. Interestingly, the peptide conformation at $\phi = 130°$ and $\psi = $ -160° shows a concentrated and detailed outline on the Ramachandran map, suggesting that the peptides dihedral angle is in a fixed position for an extended period of time. The well defined edges of the data observed in this data are not seen with single peptides in solution, which exhibit data that is more spread out, most likely caused by the more flexible behaviour of a single peptide. The rigid behaviour of the peptides here are most

likely caused by the strong hydrogen bonding between the two peptides, causing the flexible molecule to become rigid upon the peptide - peptide binding.
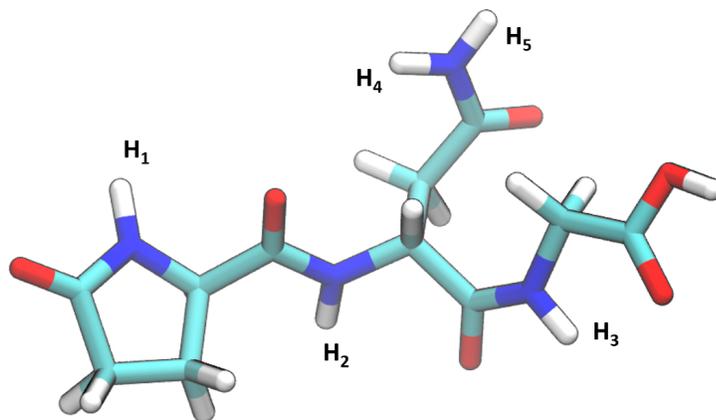
Figure 4.10: The amide functional groups on the Glu-Asn-Gly peptide. These amides (H_1 to H_5) are primarily responsible as hydrogen bond donors.
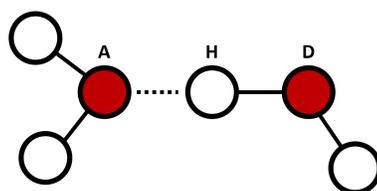
Figure 4.11: Hydrogen bonds (represented as a dotted line) in this work are defined between an electronegative donor atom (D) with a hydrogen (H) on it and an electronegative acceptor atom (A), where the distance between D and A is less than 3.5 Å and the D-H-A angle deviates no more than 20° from a linear bond.

Three peptides

In the case of three peptides in solution, a mixture of the conformational behaviour for a single peptide and two peptides is seen in the Ramachandran map (Figure 4.8c). Interactions between two of the peptides are seen to shift the ϕ and ψ angles for these peptides towards the lower right quadrant of the Ramachandran map, whereas one of the peptides appears as a single free peptide in solution, exhibiting the same conformational behaviour as a single peptide (upper left quadrant). The dihedral angles for the single peptide and the dimer do not completely mimic the data in Figures 4.8a and 4.8b,

Donor Peptide-1	Total life time of H-bond (ns)	Donor Peptide-2	Total life time of H-bond (ns)
N-H$_1$	0.33	N-H$_1$	0.34
N-H$_2$	0.90	N-H$_2$	1.10
N-H$_3$	1.86	N-H$_3$	1.41
N-H$_4$	4.66	N-H$_4$	2.91
N-H$_5$	3.50	N-H$_5$	3.58
Average	2.25	Average	1.87

Table 4.2: Hydrogen bonding between peptides during a 5.0 ns MD simulations of two Glu-Asn-Gly peptides in solution. The total life time of the hydrogen bonds is measured from the total number of frames (saved every 20 fs) that the donor atom displayed a hydrogen bond interaction.

presumably due to the fact that there are still some remnant interactions between the single peptide and the dimer. The hydrogen bonding between the peptides is shown in Table 4.3. The detailed edges of the conformational data for the two interacting peptides and the wide spread of the data for the single peptide again points to the difference in flexibility between the single peptide and the dimer.

Donor Peptide-1	Total life time of H-bond (ns)	Donor Peptide-2	Total life time of H-bond (ns)	Donor Peptide-3	Total life time of H-bond (ns)
N-H$_1$	1.45	N-H$_1$	2.47	N-H$_1$	0.75
N-H$_2$	0.48	N-H$_2$	0.40	N-H$_2$	0.16
N-H$_3$	2.13	N-H$_3$	2.17	N-H$_3$	0.53
N-H$_4$	3.34	N-H$_4$	4.01	N-H$_4$	0.13
N-H$_5$	2.15	N-H$_5$	3.02	N-H$_5$	0.53
Average	1.91	Average	2.41	Average	0.42

Table 4.3: Hydrogen bonding between peptides during a 5.0 ns MD simulations of three Glu-Asn-Gly peptides in solution. The total life time of the hydrogen bonds is measured from the total number of frames (saved every 20 fs) that the donor atom displayed a hydrogen bond interaction.

Six and ten peptides

The Ramachandran maps for six and ten peptides in solution are shown in Figures 4.8d and 4.8e respectively. The distribution of ϕ and ψ dihedral angles shows that the

majority of the configurational space of the dihedral angles is sampled. The most visited conformations of the six and ten peptides in solution show a similar behaviour to the simulations with three peptides in solution. Single peptides not interacting with the other organic molecules in the system are observed in the upper left quadrant. The right quadrant shows the ϕ and ψ dihedral angles of those peptides that interact with each other and form hydrogen bonds. In the case of six and ten peptides, multiple clusters of interacting peptides are observed throughout the simulations. These clusters of agglomerated peptides consist either of two or three peptides each (Figure 4.12). The strong hydrogen bonding between the peptides allow these clusters to stay intact throughout the simulations and the hydrogen bonds in the simulation of six peptides are shown in Table 4.4. The data for ten peptides in solution is similar, and is shown in Appendix D.
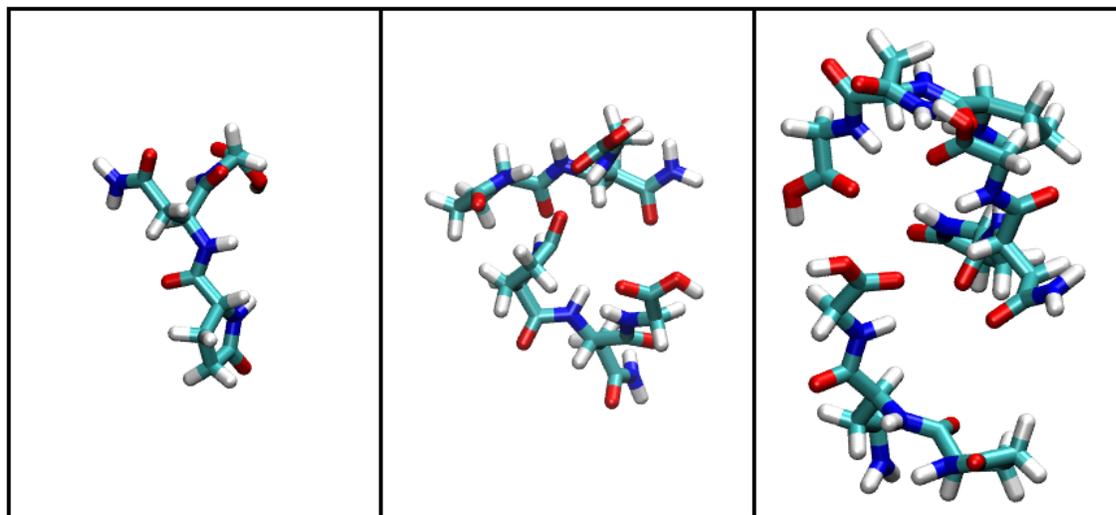
Figure 4.12: Hydrogen bond interactions between multiple peptides in solution cause the peptides to agglomerate and form clusters. Throughout the simulations, single peptides, two peptide clusters and three peptide clusters are observed.

Aggregation and hydrogen bonding

The simulations of multiple peptides in solution all show a preference for peptide - peptide interactions and the agglomeration of the peptides into small clusters. The main constituents of the interactions between the peptides is seen to be the amide

Donor Peptide-1	Total life time of H-bond (ns)	Donor Peptide-2	Total life time of H-bond (ns)	Donor Peptide-3	Total life time of H-bond (ns)
N-H$_1$	0.10	N-H$_1$	0.42	N-H$_1$	1.41
N-H$_2$	0.14	N-H$_2$	0.95	N-H$_2$	0.68
N-H$_3$	0.26	N-H$_3$	0.73	N-H$_3$	0.48
N-H$_4$	0.36	N-H$_4$	3.18	N-H$_4$	1.17
N-H$_5$	0.27	N-H$_5$	3.83	N-H$_5$	1.44
Average	0.23	Average	1.82	Average	1.03

Donor Peptide-4	Total life time of H-bond (ns)	Donor Peptide-5	Total life time of H-bond (ns)	Donor Peptide-6	Total life time of H-bond (ns)
N-H$_1$	0.86	N-H$_1$	2.34	N-H$_1$	0.22
N-H$_2$	0.53	N-H$_2$	0.12	N-H$_2$	0.20
N-H$_3$	0.58	N-H$_3$	1.41	N-H$_3$	0.76
N-H$_4$	1.60	N-H$_4$	2.51	N-H$_4$	1.53
N-H$_5$	2.13	N-H$_5$	4.16	N-H$_5$	3.92
Average	1.14	Average	2.11	Average	1.33

Table 4.4: Hydrogen bonding between peptides during a 5.0 ns MD simulations of six Glu-Asn-Gly peptides in solution. The total life time of the hydrogen bonds is measured from the total number of frames (saved every 20 fs) that the donor atom displayed a hydrogen bond interaction.

functional groups of the peptide bonds between the residues, as seen in Tables 4.2 to 4.4). As seen in the Ramachandran maps from the simulations, the interactions that occur between the peptides affect their conformation. conformations of the peptides. When the peptides agglomerate into small clusters the hydrogen bonds between the peptides force them in conformations otherwise not accessible.

The average life times of the hydrogen bonds within these clusters also differ depending on the cluster size. Over all the simulations the average lifetime of hydrogen bonding for peptides in a cluster of two is 1.90 ns. Within these clusters amides $N - H_1$ and $N - H_3$ are primarily involved with the hydrogen bonding. These two donors are on the main backbone of the peptide (Figure 4.10), causing the peptides to interact whilst parallel to one another. The average life time of hydrogen bonds within clusters of three peptides decreases to 1.06 ns and the involvement of donors $N - H_1$ and $N - H_3$ decreases. Within three peptide clusters the amides $N - H_4$ and $N - H_5$ are the primary donors, situated on the end the Asn side chain. Shifting the interactions from the main backbone to the sidechain of the peptide causes the peptides to interact head on, as seen in Figure 4.12.

4.2.3 Conformation on surface

Using the ϕ and ψ dihedral angles in a Ramachandran map the conformations of the Glu-Asn-Gly peptides can be mapped when adsorbing onto the $(10\bar{1}4)$ surface of calcite.

Single peptide

The adsorption of a single peptide on the surface of calcite (Figure 4.13a) does not alter the conformation with respect to a free peptide in solution. The major conformation that can be observed is at $\phi = $ -160° and $\psi = $ 70°, the same β-sheet domain as observed before. However, the data in the Ramachandran map show clearer, sharper edges for

the angle distribution, similar to data seen before for i.e. two peptides in solution (Figure 4.8b), where a large peptide - peptide interaction is present. The well defined data suggests that the conformation of the peptide is more rigid than in solution and that upon binding to the calcite surface the peptide undergoes a transition to more rigid, structured molecule. Additionally, in contrast to the peptide in solution, the conformations of the peptide that are sampled throughout the simulation are limited. Where in solution the majority of the ϕ/ψ conformations were visited, in the adsorbed state the peptide is kept in a rigid conformation, and only a narrow set of angles is observed, much like a disorder-to-order transition seen in IDPs.

Two peptides

A distinct difference in the conformation of multiple peptides can be seen upon binding to the surface of calcite. Whereas in solution the interactions between two peptides caused the conformation of the peptides to lie within the lower right quadrant of the Ramachandran map, upon binding to the calcite surface this conformational behaviour is lost. The peptides behave as though they are single peptides binding to the surface and the ϕ and ψ angles observed for the peptides (Figure 4.13b) replicate those seen in a single peptide on the surface of calcite (Figure 4.13a). There is, however, still a large number of interactions present between the two peptides in the form of hydrogen bonding. The conformational behaviour that is observed can be attributed to the interaction between the peptide and the calcite surface, which has a more pronounced effect than the peptide - peptide interaction. This surface effect on the peptide conformation can be compared to that observed for larger peptides, such as n16-N, where the peptide - surface interaction induces an ordering of the peptide, similar to the disorder-to-order transitions that can be observed in the case of IDPs.

(a) One peptide.

(b) Two peptides.

(c) Three peptides.

(d) Six peptides.

(e) Ten peptides.
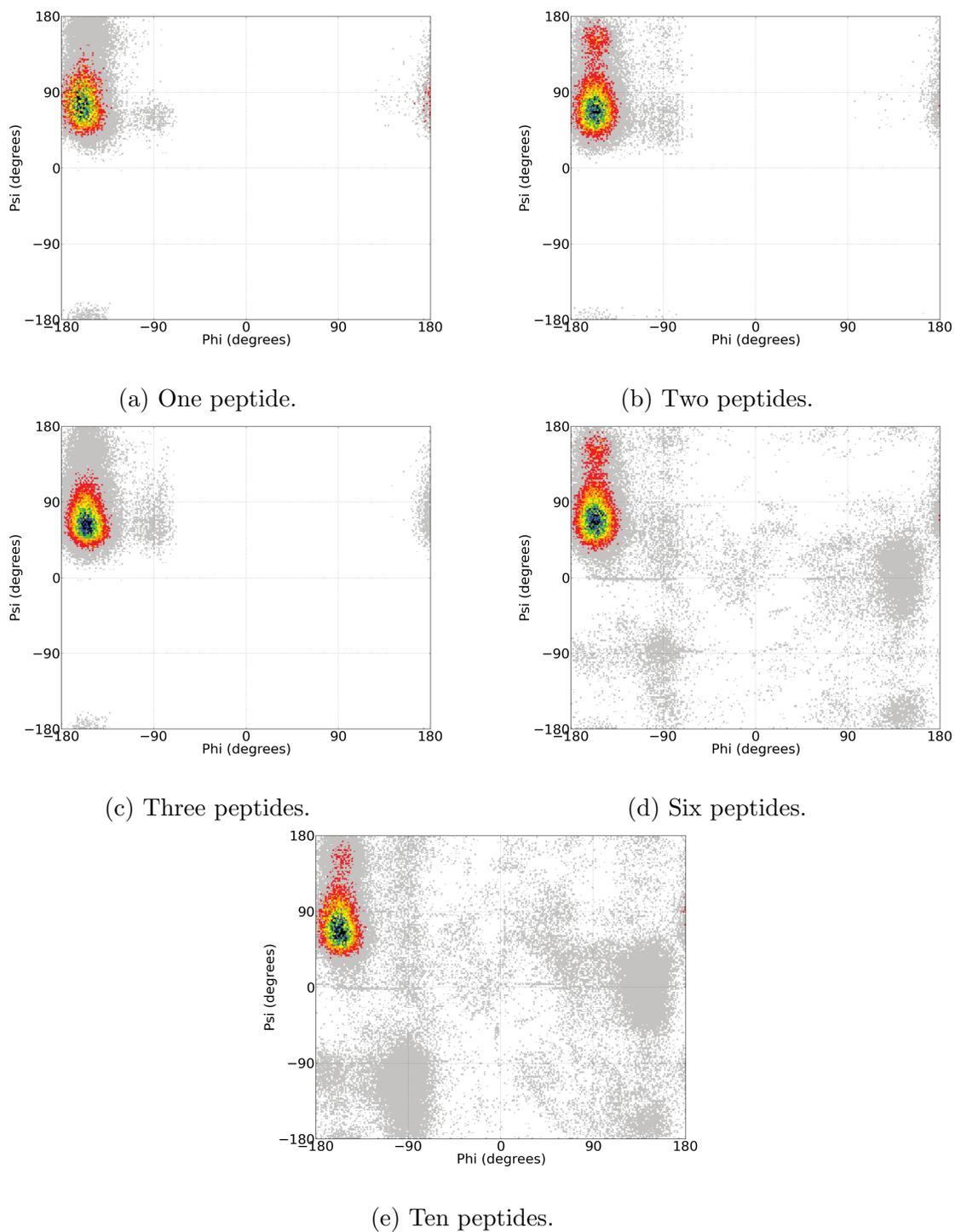
Figure 4.13: Ramachandran maps obtained with MD simulations for the Glu-Asn-Gly peptide(s) on the ($10\bar{1}4$) surface of calcite. The colour coding represent a linear scale for the number of counts in each bin, with increasing count from grey, red, orange. yellow, green, blue, black.

Three, six and ten peptides

With multiple peptides in solution the peptides show a very uniform behaviour. When analysing the ϕ and ψ dihedral angles for the simulations with three, six and ten peptides interacting with the surface, similar Ramachandran maps are obtained (Figures 4.13c, 4.13d and 4.13e for three, six and ten peptides respectively) although more conformations are sampled in the simulations with six and ten peptides, most likely due to the number of peptides in these simulations.

These results suggest that when the peptide is adsorbing to the calcite surface it is restricted to a specific conformation. This restricted conformation is likely to be the lowest energy conformation that enables the peptide to bind to the surface with minimal disruption to the water on top of the surface. Such restricted binding conformations are also seen in the case of IDPs upon binding to their target. Throughout the simulations the peptides still agglomerate together in clusters and the hydrogen bonds between the peptides are still intact. However, the peptide - surface interactions are seen to more important and have a more pronounced effect than the peptide - peptide interactions.

Binding conformation

Although the peptides were inserted in a random orientation during the system setup, in all the simulations the peptides are seen to be adsorbing onto the calcite ($10\bar{1}4$) surface in the same manner, as seen in Figure 4.14. As discussed in Chapter 1.2.2, the highly structured water layer on top of the calcite surface creates an energy barrier for adsorbing molecules, and the most favourable binding energies arise from conformations that minimise the disruption to this water layer. The conformation of the peptides on the surface of calcite are seen to be penetrating the water layer with minimal disruption and binding to the surface with the carboxyl functional groups, considered to generate stronger interactions than the amide functional groups [83].

The Z-density profile of the water oxygen and the carboxyl oxygen of the Glu-Asn-Gly peptide (Figure 4.15) shows the ordered water layers on top of the calcite surface and the carboxyl oxygen situated the same distance from the surface as the first water layer oxygen.
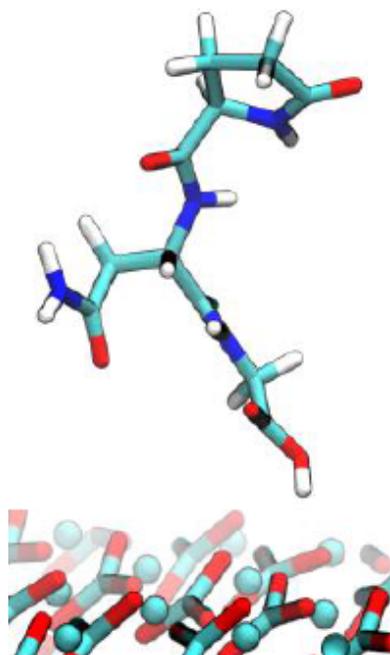
Figure 4.14: The Glu-Asn-Gly peptides are seen to be binding onto the surface of calcite in the same manner, with the carboxyl functional group binding to the surface and the majority of the molecule extending away from the surface. Water molecules are omitted for clarity.

4.3   Adsorption of tripeptides on calcite surface

4.3.1   Energetics calculations from simulations

The adsorption energies of the Glu-Asn-Gly peptides on the surface of calcite were calculated from the simulations described earlier. Additional simulations of a water-only solvated $(10\bar{1}4)$ calcite surface with 20,000 TIP3P water molecules and a bulk water system were performed in order to calculate these energies. The same simulation protocols were used as for the peptide simulations (Chapter 4.2.1). Using Equation 4.1 the
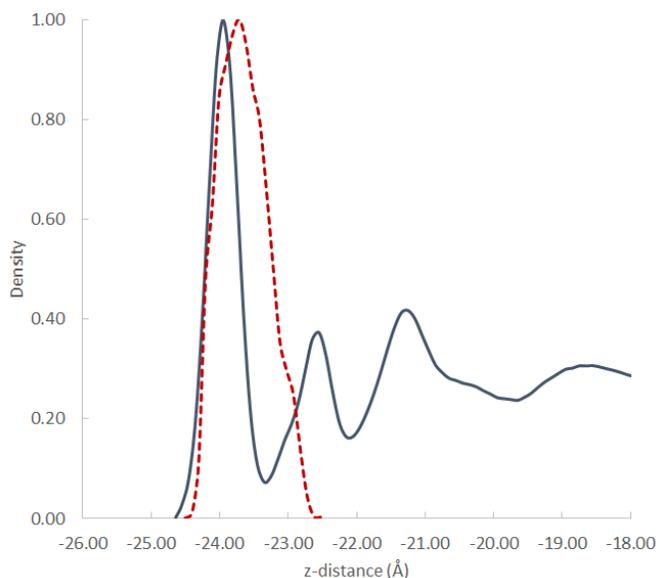
Figure 4.15: The Z-density profile of a single Glu-Asn-Gly peptide adsorbing onto the calcite (10$\bar{1}$4) surface. The profile shows the water oxygen (blue solid line) and peptide carboxyl oxygen (red dashed line). The carboxyl functional group is seen to bind to the surface at the same distance as the first water layer.

adsorption energies of the peptides ($E_{adsorption}$) with respect to the water-only solvated surfaces ($E_{water-calcite}$) and the solvated peptides ($E_{water-peptide}$) were calculated,

$$E_{adsorption} = E_{water-calcite-peptide} - E_{water-calcite} - E_{water-peptide} \qquad (4.1)$$

where $E_{water-calcite-peptide}$ is the energy of the solvated calcite - peptide system.

Adsorption energies

The adsorption energies for the peptides are shown in Table 4.5. The energy of adsorption of a single peptide is comparable to other small molecules on the surface of calcite, e.g. -64.1 kJ mol^{-1} for methanoic acid [83], -22.9 to -60.6 kJ mol^{-1} for various polysaccharides [82], -73.6 kJ mol^{-1} for mannose [81] and -46.5 kJ mol^{-1} for acrylic acid functional groups (this work, Chapter 5). Despite the strong binding of the water to the calcite surface the peptides are still seen to penetrate through the water and bind via a strong bond between the carboxyl functional group. The data in Table 4.5

119

shows no clear trend in the adsorption energy per peptide with an increasing amount of peptides in the system. It is, however, clear that the system with two peptides has a significant lower adsorption energy per peptide (-57.1 kJ mol^{-1}) than in the case of a single peptide (-69.5 kJ mol^{-1}). This effect can be attributed to the strong binding between the two peptides, influencing the binding to the surface. With the adsorption of the peptides in the simulations of three, six and ten peptides, a combination of binding modes can be observed, with peptides binding as single molecule or in clusters of two or three peptides.

Number of peptides	Adsorption energy (kJ mol^{-1})	Standard error (kJ mol^{-1})	Adsorption energy per peptide (kJ mol^{-1})
One	-69.5	2.3	-69.5
Two	-114.1	1.9	-57.1
Three	-184.6	3.4	-61.5
Six	-374.9	1.6	-62.5
Ten	-538.6	2.4	-53.9

Table 4.5: Adsorption energies of Glu-Asn-gly peptides on the surface of $(10\bar{1}4)$ calcite. The energy in kJ mol^{-1} refers to the total energy of the simulation box. Where the adsorption energy is calculated per peptide this is done by dividing the total energy by the number of peptides in the system.

Thermodynamic cycle

The association energy for the peptide - peptide interaction between two peptides in solution and on the surface of calcite can be calculated using Equations 4.2 and 4.3:

$$E_{association-solution} = (E_{water-two-peptides} - E_{water}) \tag{4.2}$$

$$-2(E_{water-one-peptide} - 2E_{water})$$

$$E_{association-surface} = (E_{water-calcite-two-peptides} - E_{water}) \tag{4.3}$$

$$-2(E_{water-calcite-one-peptide} - 2E_{water})$$

where E_{water} is the energy of the water only and $E_{water-one-peptide}$, $E_{water-two-peptides}$, $E_{water-calcite-one-peptide}$ and $E_{water-calcite-two-peptides}$ are the energies of the solvated peptides systems and solvated calcite-peptide systems respectively.

With the energies of the surface - peptide interactions from Equation 4.1 and the energies of the peptide - peptide interactions both in solution and on the surface of calcite from Equations 4.2 and 4.3 respectively, a thermodynamic cycle can be constructed. This cycle is shown in Figure 4.16 and shows the energetic behaviour of two Glu-Asn-Gly peptides in solution and on the surface of calcite.
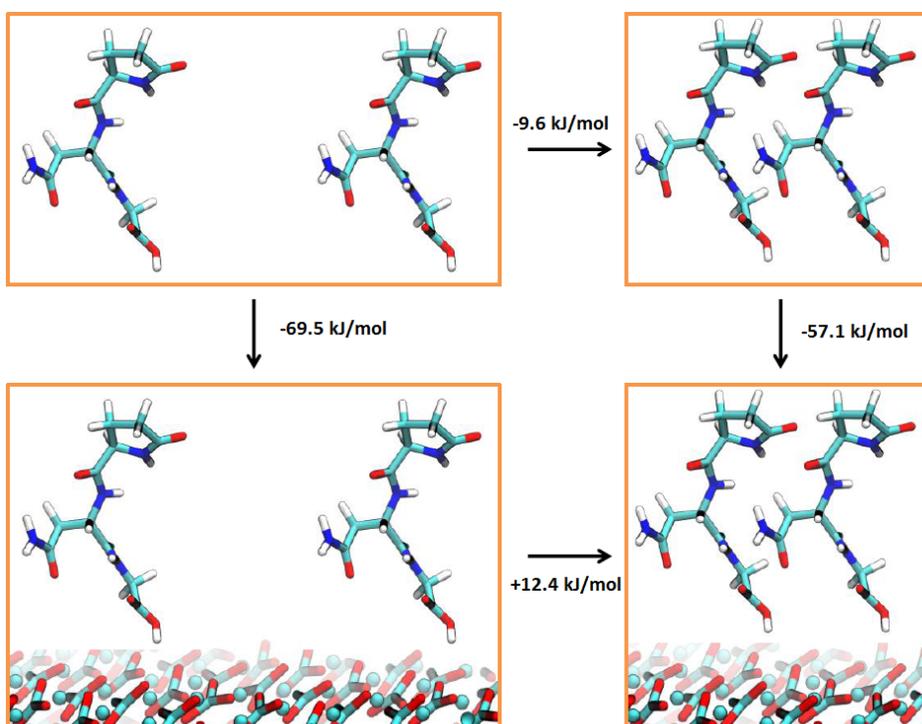
Figure 4.16: From the adsorption energies and the interaction energies between the Glu-Asn-Gly peptides a thermodynamic cycle can be constructed that describes the energetic behaviour of multiple peptides. The negative association energy in solution indicates that the aggregation of these peptides is a favourable process, unlike their interactions on the surface of calcite.

The negative association energy between the peptides in solution indicates that the aggregation of these peptides in solution is a favourable process, with an energy of -9.6 $\pm$ 1.9 kJ mol^{-1} associated with the aggregation of the two peptides. *Vice versa*, the association of these peptides when adsorbed onto the surface of calcite is not a favourable

process, with an interaction energy of 12.4 ± 1.8 kJ mol^{-1}. These results are an indication that the aggregation of the peptides in solution is likely, as has been calculated from previous simulations [171] and experiments [170] of larger peptides.

4.3.2 Precipitation experiments

A preliminary study was performed to analyse the effect of the Glu-Asn-Gly tripeptide on the growth of $CaCO_3$ using experimental techniques. The ready availability of Glu-Asn-Gly enabled this peptide to be used within constant composition precipitation experiments as described by Lakshtanov and Stipp [63]. The quantitative effect on the growth of calcite using these constant composition precipitation experiments was measured at the NanoGeoScience laboratory at Copenhagen University.

Experimental setup

Calcium carbonate (reagent grade $CaCO_3$, 99.95 %, supplied by Merck) was recrystallised by exposure to deionised water and 100 % CO_2 gas, using the method of Stipp and Hochella [174], to remove any organic impurities and to increase the homogeneity of the crystal seeds. Using N_2 adsorption, the BET surface area[4] of the calcite seeds was determined to be 0.2 m^{2}g^{-1}. Reagent grade $CaCl_2 \cdot H_2O$ and double distilled MQ water[5] was used to prepare a solution of $CaCl_2$. The Ca^{2+} content of the solution was analysed using an atomic absorption spectrometer (Perkin Elmer AAS Analyst 800). A 0.66 M peptide solution was prepared by solvating 1.0 mg Glu-Asn-Gly (Sigma) in 5.0 mL 0.1 M NaCl solution. All other solutions, with ionic strength $I = 0.1$ M, were prepared by mixing stock solutions of $NaHCO_3$ and NaCl with MQ water.

In a double walled glass vessel, kept at a constant temperature of 298 K, a weighed

---

[4]The BET surface area is derived from the Brunauer-Emmett-Teller theory which describes the physical adsorption of gas molecules on surfaces [175].

[5]MQ or Milli-Q water is a trademark from Merck Millipore who produce 'ultra pure' Type 1 water which has an ion resistivity of >18.0 MΩcm at 298 K [176].

amount of calcite seeds was added to a 2 mM Ca^{2+} HCO_3^- supersaturated solution made up with NaCl. A pH-stat titration system (Metrohm 780 meter and Metrohm NTC/glass/3M KCl electrode) was used to maintain steady state and super saturation conditions. The precipitation of calcite occurs via the growth of the existing calcite crystals and via the reaction in Equation 4.4:

$$Ca^{2+} + HCO_3^- \longrightarrow CaCO_3 \downarrow + H^+ \tag{4.4}$$
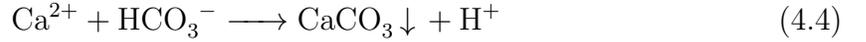
where $CaCO_3$ precipitates by consuming the Ca^{2+} and CO_3^{2-} ions in the solution, causing a decrease in the pH of the system. A decrease below a pH of 8.3 triggers the influx of 0.09 M $CaCl_2$ and 0.09 M Na_2CO_3 via a syringe pump, replenishing the Ca^{2+} and CO_3^{2-} ions in the solution and raising the pH. In this equilibrium of constant supersaturated solution the precipitation rate of calcite can be calculated via the slope of the plot of added volume $CaCl_2/Na_2CO_3$ versus time. The precipitation rate R can be written as Equation 4.5:

$$R = R' \frac{[Ca]_{titrant}}{A_{seeds}} = R' \frac{[Ca]_{titrant}}{m_{seeds} S_{seeds}} \tag{4.5}$$

where $[Ca]_{titrant}$ is the Ca^{2+} concentrations of the added titrant solution, A_{seeds} is the total surface area of the calcite seeds, m_{seeds} is the mass of calcite seeds, S_{seeds} the specific BET surface area of the calcite seeds and R' denotes the titrant addition rate. When the precipitation rate is constant (measured by a constant titrant addition rate), a specific amount of Glu-Asn-Gly peptide solution is added to the reaction vessel. By measuring the change in precipitation rate before (R'_0) and after (R') the addition of the peptide to the solution, the quantitative inhibition effect of the peptide on the growth of calcite can be observed. The initial precipitation rates, R_0, for all experiments were measured and are seen to be similar in all experiments, independent of the amount of calcite seeds added. All details of the experimental setup are shown in Table 4.6.

Seed mass (mg)	$C_{peptide}$ x10^{-3} (g L^{-1})	R'_0 x10^{-4} (L s^{-1})	R_0 x10^{-6} (mol m^{-2}s^{-1})	R' x10^{-4} (L s^{-1})	R x10^{-6} (mol m^{-2}s^{-1})	Relative growth
33.5	-	1.41	0.88	-	-	-
33.4	-	1.61	1.01	-	-	-
26.5	-	1.38	1.09	-	-	-
33.3	0.26	1.71	1.08	1.68	1.06	0.98
26.7	0.26	1.35	1.06	1.34	1.05	0.99
31.5	0.66	1.53	1.02	1.47	0.99	0.96
32.6	0.66	1.71	1.10	1.65	1.06	0.96
28.0	0.66	1.44	1.08	1.34	1.00	0.93
27.6	1.32	1.43	1.08	1.27	0.96	0.88
26.6	1.32	1.28	1.01	1.26	0.99	0.98
27.5	2.64	1.30	0.99	1.29	0.98	0.99
27.0	2.64	1.30	1.01	1.21	0.94	0.93
26.5	2.64	1.12	0.89	1.16	0.92	1.03
26.5	3.30	1.24	0.98	1.19	0.94	0.96

Table 4.6: Details for all experimental runs where $C_{peptide}$ is the peptide concentration added; R'_0 is the initial titrant addition rate; R_0 is the initial precipitation rate; R' is the titrant addition rate after addition of the peptide, R is the precipitation rate after addition of the peptide and the relative growth rate is the ratio of R/R_0.

Inhibition of calcite growth

The constant composition precipitation experiments showed a constant precipitation rate of calcite, as measured by the added volume of Na_2CO_3 and $CaCl_2$, as seen in Figure 4.17a. After an initial constant precipitation rate it is seen that the precipitation rate steadily decreases after approximately 10,000 s (Figure 4.17b). This decrease occurs at the same time as calcite crystals are observed to adhere to the glass surface of the experiment chamber. As the adhesion of the calcite crystals on the walls of the vessel directly affects the available surface area of calcite seeds, this growth decline is expected. Since this adhesion of the crystals to the glass vessel affects the results, the data after the first 2,000 steps of the experiment were discarded.

The effect of the Glu-Asn-Gly peptide on the growth of calcite was measured over a peptide concentration range of 0 - 3.3 mgL^{-1} and the results are shown in Figure 4.18.

Whereas previous experiments [37,63,65,66] using similar small molecules saw a range of effects on calcite growth, the data for the Glu-Asn-Gly peptide shows no discernible effect of the tripeptide on the growth of calcite. The lack of effect on calcite growth seen in these preliminary precipitation experiments is counter to the binding observed in the MD simulations. Further precipitation experiments with the Glu-Asn-Gly peptide using a larger range of peptide concentration are desirable. The inclusion of Langmuir adsorption isotherm experiments would be able to shed light on binding behaviour of these peptides. Additionally, the effect of the peptide on other polymorphs of $CaCO_3$, such as aragonite, could be analysed.

(a) Constant precipitation rate. (b) Tail-off for a precipitation experiment.
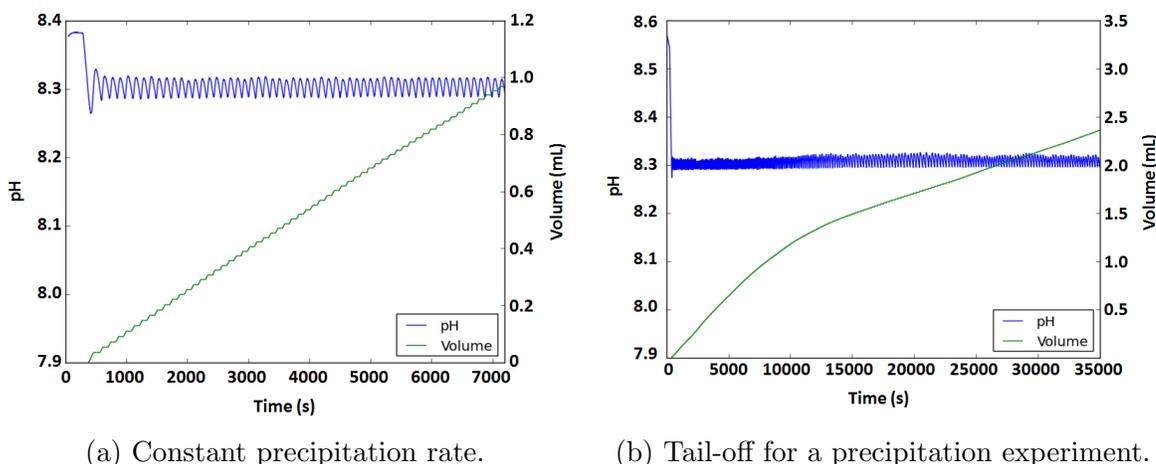
Figure 4.17: Examples of the constant composition precipitation experiments results. The precipitation rate of calcite stays constant until ca 10,000 s, after which the adhesion of the calcite crystals on the walls of the vessel affects this precipitation rate.
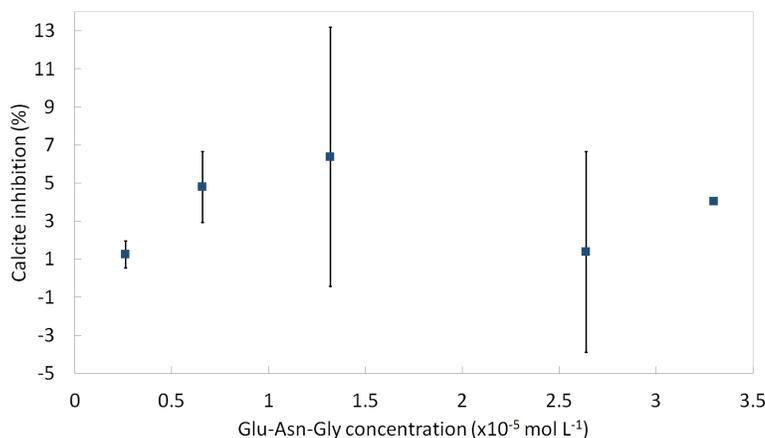
Figure 4.18: Inhibition of the calcite growth rate by the Glu-Asn-Gly peptide in constant composition precipitation experiments.

4.4 Conclusion

The conformational behaviour of the Glu-Asn-Gly peptides in solution depends strongly on the presence of other biomolecules. Single peptides show a single conformation as shown by the ϕ and ψ dihedral angles in the Ramachandran maps and the peptide exhibits a flexible behaviour. This single behaviour of the peptide changes when interacting with another peptide. The strong peptide - peptide interactions, which occur due to a large number of hydrogen bonds between the molecules, force the peptides into a more rigid conformation with a different set of ϕ/ψ angles. The Ramachandran maps of multiple peptides in solution show dihedral angle conformations that exhibit sharp, detailed edges that correspond to this rigid structure. A similar behaviour can be observed for IDPs, where inter-molecular interactions between the IDP and a target molecule or surface induce a disorder-to-order transition, transforming the flexible peptide or protein into a rigid molecule.

The simulations of three or more peptides in solution show a mixture of the above behaviours where some peptide molecules are seen to be behaving as single, free peptides in solution and the appearance of two- and three-peptide clusters. The conformation of the peptides within these clusters depend on the size of the clusters. Within two-peptide clusters, the hydrogen bonding occurs predominantly between the backbone amides and causes the peptides to interact parallel to one another. In the case of three-peptide clusters the side chain amide is involved with the majority of the hydrogen bonding, causing the conformation of the peptides to shift slightly. The ability of these peptides to agglomerate is also seen with IDPs involved with biomineralisation processes, such as the n16 peptide [170, 171].

The binding of a single peptide to the (10$\bar{1}$4) calcite surface induces a large disorder-to-order transition, as the peptide is seen to change from a flexible to a rigid structure. The binding of the peptide occurs via the carboxyl functional group situated on the C-terminus of the peptide. The Z-density profile of the simulations show that the carboxyl

oxygen is situated at the same distance from the surface as the water oxygen in the first surface layer. The rest of the peptide is extended away from the surface into the aqueous environment in order to minimise the disruption to the structured water layers on top of the mineral surface.

The binding of multiple peptides occurs in the same manner as a single peptide. The binding conformations all show a distinct similar behaviour, with the same ϕ and ψ backbone dihedral angles as for a single peptide adsorbing onto the surface. The interactions between the peptides are still intact and the hydrogen bonds between the peptides are observed for extended time scales. The conformational behaviour that is observed can be attributed to the relative strong interactions between the peptide and the calcite surface compared to the peptide - peptide interactions. The effect that the calcite surface has on the conformation of the peptide can be compared to that of larger peptides, such as n16-N, where the peptide - surface interaction induces an ordering of the peptide, much like a disorder-to-order transition seen in IDPs.

Similar changes in conformational rotation can be observed in simulation work on conformational kinetics of butane in slit pores by Travis and Searles [177] and Brookes *et al.* [178], which show that a change of conformation can be observed when n-butane is in proximity to the surface of a confinement. A layering effect can be seen to occur at the surface, with the conformation of the molecules changing to allow for closer packing, resulting in a higher molecular density. In addition, the diffusion of these flexible molecules also becomes diminished by the addition of rigidity within the molecule, as seen by the work on torsional flexibility of linear alkanes by Braga and Travis [179].

Within the constant composition precipitation experiments performed using the Glu-Asn-Gly peptides, no effect of the peptide on the growth of calcite was observed. Previous experiments using small molecules [37, 63, 65, 66] show a range of calcite growth inhibition, and similar results were expected. The absence of any inhibition can be explained by several factors. Although the setup of the precipitation experiments and

in particular the peptide concentrations used was similar to previous experiments [63], duplicate experiments should be performed to rule out any experimental error. Another explanation for the lack of inhibition could be the strong agglomeration seen between the tripeptides. Throughout the simulations the majority of the peptides reside in clusters with strong, long lasting hydrogen bonds keeping these clusters intact. Within the experimental setup this agglomeration might have impeded the interactions with the calcite surface.

The binding energy obtained from the MD simulations for a single Glu-Asn-Gly peptide of -69.5 kJ mol^{-1} shows the favourable binding of the peptide to the calcite surface and compares well with other simulations of similar biomolecules. With the adsorption of two peptides to the surface a decrease in this adsorption energy can be observed from -57.1 to -61.5. For multiple peptides a mixture of these binding behaviours is observed, with average binding energies of -62.5 and -53.9 kJ mol^{-1} for six and ten peptides respectively.

These results show that in solution hydrogen bonding between unstructured peptides or proteins induce a strong transition from a flexible to ordered structure. The calcite surface has a similar influence of the conformation of the peptides and upon binding to the surface the peptides exhibit a rigid behaviour. These results help to show that the paradigm in structural biology that the function of peptides or proteins is linked to their three-dimensional structure does not hold for all molecules. Instead, the three-dimensional structure of these biomolecules is not a fixed arrangement. The presence of a surface or indeed other biomolecules can induce structure to these flexible molecules, allowing them to remain versatile and able to bind to many targets.

Chapter 5

Adsorption of poly acrylic acid onto calcite surfaces

5.1 Introduction

Previous chapters (1,4) have described the attachment of biomolecules to mineral surfaces such as calcite. One of the research areas in which this attachment of biomolecules is important is in the study of microbial cells. Using organic biomolecules these microbial cells attach to mineral surfaces to form biofilms. In this chapter the polymer poly acrylic acid is used as a model system for these biomolecules and molecular dynamics simulations are performed to analyse the behaviour of this polymer in solution and on the surface of calcite. The results are compared to experimental studies performed in close collaboration with the computational simulations.

5.1.1 Biofilms and bacterial growth on surfaces

Microbial cells, or prokaryotes, are microscopic unicellular life forms such as bacteria that are invisible to the naked eye yet make up a large portion of the Earth's life forms. It is estimated that there are a total of $4 - 6 \times 10^{30}$ prokaryotic life forms on earth [180].

Instead of existing as individual floating cells these microbial cells tend to agglomerate and attach to surfaces. Through translocation, aggregation and surface attachment they can position themselves in niches on the surface as dynamic and structurally complex biological systems, which are referred to as biofilms [85, 181]. The formation of these biofilms occurs in three distinct steps (Figure 5.1). Firstly, single prokaryote cells attach themselves to the surface. Secondly, multiple cells on the surface aggregate into microcolonies and thirdly, through the excretion of extracellular polymer substance (EPS) by the prokaryotic cells, these microcolonies form an organic matrix binding the microbial cells together. The EPS consists of organic material excreted by the microbial cells and is made up of macromolecules such as polysaccharides, proteins and lipids [67]. This organic matrix shows a classic viscoelastic behaviour and can withstand rapid changes in shear stress and can dissipate long lasting load stresses through the viscosity of the material, and can therefore rapidly adapt to its environment.

The aggregation and attachment to surfaces has several advantages. The surface on which these cells agglomerate provides a stable growth environment for the cells to gather nutrients. The biofilms can concentrate nutrients and dispose of toxic metabolites through permeable water channels in the EPS matrix. The close proximity of these microbial cells allows for inter-cellular interactions and communication to adapt to the environment, and adapt to changes in cell density [182]. The EPS matrix also serves as protection against the environment and shields the cells from harsh environmental factors such as extreme pH values, extreme temperatures, UV radiation, osmotic shocks and desiccation. The formation of these biofilms also has the advantage of increasing the resistance of these microbial cells to antibiotics, and studies have reported up to a 1,000-fold increase in resistance of bacteria in biofilms compared to free floating bacteria [183]. This resistivity is thought to be partly caused by the EPS barrier itself through the direct interaction between the organics in the EPS and antibiotic molecules.

Because of the adaptation of these biofilms, their formation is observed to be widespread throughout nature, even in extreme environments such as acid mine drainages at pH=0

[184] and thermal hot springs [185]. Biofilms are also present in the human body and it is thought that more than 60 % of all hospital infections are due to biofilm formation [181]. Examples of diseases in which biofilms play a major role are cystic fibrosis, a condition where the lungs get filled with mucus; chronic ear infection caused by bacterial biofilms; tuberculosis, a bacterial disease where nodules grow within the lung tissue; periodontitis, a chronic inflammation of the gums around the teeth, caused by biofilm formation around cavities in the teeth; and bacterial endocarditis, inflammation of heart or blood vessel tissue do to the formation of bacterial biofilms on already damaged tissue [85, 181, 186, 187]. Costs associated with these diseases are high (£931 million per annum in England [188]) and mortality rates for some can be severe (up to 70% for bacterial endocarditus [187]).
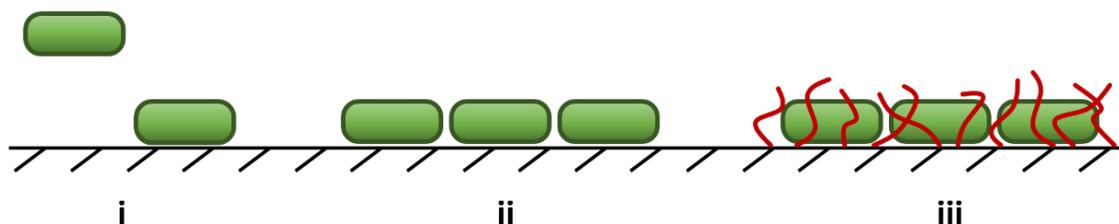
Figure 5.1: Schematic representation of bacterial biofilm formation in three distinct steps: i) attachment of prokaryote cells to the surface; ii) aggregation of the cells on the surface; iii) excretion of organic macromolecules to form EPS matrix.

5.1.2   Poly acrylic acid as model system for biofilms

The design of the experiments associated with the simulations described here is focused on the interaction between the EPS macromolecules and the surfaces it attaches itself to. Previous studies [189, 190] have shown the importance of functional groups such as carboxyl groups in the attachment of macromolecules such as lipopolysaccharides, mycolic acid and alginate that are found in the EPS. The polymer poly acrylic acid (PAA, Figure 5.2) has been suggested as a suitable system to model the interactions between EPS and surfaces, due to its carboxylic acid functional groups [189–191].

The polymer PAA is a well-studied polymer and several experimental and computational groups have studied the interactions between the polymer and surfaces and counter ions. The behaviour of the polymer in solution has been studied by Laguecir *et al.* [192] who found that increasing the polymer chain length also increased the total electrostatic potential on each monomer, making it harder to ionise the polymer. The diffusivity of the polymer in solution decreased with increasing chain length, which can be explained by either a conformational change of the polymer in solution or the aggregation of multiple polymers, and this decrease is also seen with increasing pH values. Bulo *et al.* [124] examined the interactions between short chain PAA (Mw = 1200 g mol^{-1}) and calcium ions in solution using computational simulations and found that at high ionic strengths the PAA chains exhibit an extended conformation stabilised by calcium ions. The same research group also performed computational studies on the complexation of PAA with calcium and carbonate ions, showing that PAA chains strongly bind to calcium carbonate species [193]. Geffroy *et al.* [194] used microcalorimetry to analyse the binding of PAA on the surface of calcite and found a positive enthalpy of binding ($+2$ kJ mol^{-1}) and suggested that the mechanism in which the polymer would bind onto the surface would have an entropic driving force. The same group also suggested that small molecular weight polymers will have a greater affinity for the surface than large molecular weight polymers, as the positive counter ions in solution reduces the electrostatic repulsion of the short chains rapidly.

Previous computational studies have been performed on the adsorption of PAA on calcite, and a wide range of adsorption energies have been obtained from these simulations. Liu *et al.* [195] performed molecular dynamics simulations of PAA chains (10 monomers) in a vacuum. The adsorption process of the polymer onto calcite resulted in a strong binding to the surface and they found that the coulombic interaction was the most important contribution to the interaction energy between the PAA and surface. The interaction energies of these simulations varied depending on the calcite surface used, and they calculated an interaction energy of -917 kJ mol^{-1} for a ($10\bar{1}4$) surface.

However, as these simulations were performed in vacuum, these values can not be considered relevant as they do not take into account the effect of the solution. Aschauer *et al.* [196] later performed molecular dynamics simulations with a PAA chain consisting of 10 monomer units using SPC water as solvent. The simulations they performed showed a lack of binding of the PAA on a flat ($10\bar{1}4$) calcite surface and a binding energy of $+395$ kJ mol^{-1} on acute stepped calcite. The authors explain this surprising positive binding energy as the result of a large energy barrier to break through the strongly layered water structure on top of the calcite surface and the formation of complexes that stabilises the PAA in a coiled configuration on the surface of calcite. As the PAA polymer is still observed to bind to the surface of calcite, they hypothesise that a large gain in entropy by the disruption of the water layers on the surface will result in a negative free energy of adsorption. Using molecular dynamics simulations of a PAA dimer in solution and on the surface of calcite, Zhu *et al.* [190] calculated a binding energy of -45 kJ mol^{-1} on the surface of ($10\bar{1}4$) calcite. Because of the widely different methods used throughout these calculations, no direct comparison can be made between the simulations.

5.1.3 Aims and objectives

The computational studies in this work were designed to study the conformational behaviour of PAA macromolecules in solution at different pH values, analyse the adsorption of these polymers to the surface of calcite and compare these computational results to experimental studies. The difficulty with studying large macromolecules in computer simulations is the number of degrees of freedom that the macromolecule has and the large energy barriers between the different conformations that it can adopt. It is difficult to sample all the conformational configurations that such macromolecules can adopt using straightforward molecular dynamics simulations, and in order to increase the efficiency of sampling this phase space a different approach must be used, as discussed in Chapter 2. The metadynamics method developed in 2002 by Laio and Par-

rinello [121] has been utilized in this work to enhance the sampling of the system. By using the history dependent metadynamics method and suitable variables to describe to conformation of the polymer, a free energy landscape of the polymer conformation can be obtained. A 28 monomer polymer of PAA with molecular weight of ca. 2000 g mol^{-1} was used in the simulation work and results were compared to experimental studies by the research group of Romero-Gonzalez[1], who used a PAA chain with a molecular weight of $Mn = 19580$ g mol^{-1}. The behaviour of the different lengths of PAA chain in the experiments and simulations is expected to slightly differ, with larger chains showing a decrease in diffusivity and a decrease in the degree of ionisation as described above.

The aim of this work is to gain a better understanding of the conformational behaviour of macromolecules when interacting with the surface of calcite, and to assess what functionalities of these macromolecules are important when adsorbing onto the surface of $(10\bar{1}4)$ calcite. Molecular dynamics studies that focus on the conformational behaviour of PAA in solution and the adsorption of the polymer chain onto the $(10\bar{1}4)$ surface of calcite for a range of pH values will give an insight into the behaviour of these macromolecules.

5.2 Methods

5.2.1 System setup

To model the PAA chain, a polymer consisting of 28 monomer units (Figure 5.2) was used throughout all the simulation work. An atactic polymer with a random stereochemical configuration was chosen to mimic experimentally obtained PAA. In order to simulate the polymer at different pH values, the degree of ionisation of the chain was adjusted by removing hydrogen atoms from the carboxylic acid functional groups

---

[1]Department of Civil and Structural Engineering, Kroto Research Institute, University of Sheffield.

throughout the polymer. The pH values were chosen as pH 6.5 and pH 9.5, a range that would reflect the environment found in biological systems and that shows a distinct difference in the degree of ionisation of the polymer. These values are also in line with experimental protocols of PAA adsorption. To model the environmental conditions of pH 6.5, 20 randomly selected functional groups were ionised to produce a polymer with degree of ionisation $\alpha = 0.75$, in line with experimental values obtained via titration techniques (Figure 5.3) [192], and a net charge of -20. All 28 functional groups were ionised ($\alpha = 1.00$) to model the polymer at pH 9.5 with a net charge of -28. The atomic charges of the polymer were calculated using `AMBER 12 Antechamber` [138] which uses the AM1-BCC method, a semi-empirical method to calculate atomic charges for organic molecules, as described in Appendix B. The polymers were simulated within an aqueous environment to obtain initial configurations. A purpose built program (`buildcalcite.jar`[2]) was used to create a slab of calcite with a perfect ($10\bar{1}4$) surface with a depth of 32 Å (12 $CaCO_3$ layers) and surface dimensions of 104 x 108 Å. The program `Packmol` [137] was used to add the PAA polymer, Ca^{2+} counter ions and a waterbox containing 40,000 TIP3P [104] water molecules with a density of 1.0 g ml^{-1} on top of the calcite surface in a similar fashion as seen in Figure 4.4. The polymer was started either in solution or on the surface of calcite (<3 Å from the surface). Eight different setup configurations were generated, as shown in Table 5.1, to explore the configurational behaviour of the polymer in solution and on the surface of calcite and to analyse this behaviour in the presence and absence of nearby counter ions.

5.2.2 Metadynamics

The metadynamics method developed by Laio *et al.* [121] was implemented to enhance the sampling of the system and explore all configurations of the PAA polymer. The metadynamics method is explained in detail in Chapter 2.5.2. The method relies on a

---

[2]Program developed by Robert Darkins at The London Centre for Nanotechnology, University College London.
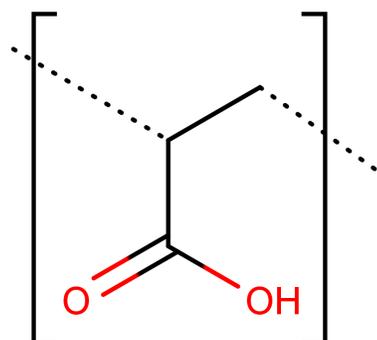
Figure 5.2: Chemical structure of poly acrylic acid (PAA) which consists of a carbon backbone with carboxylic acid functional groups.
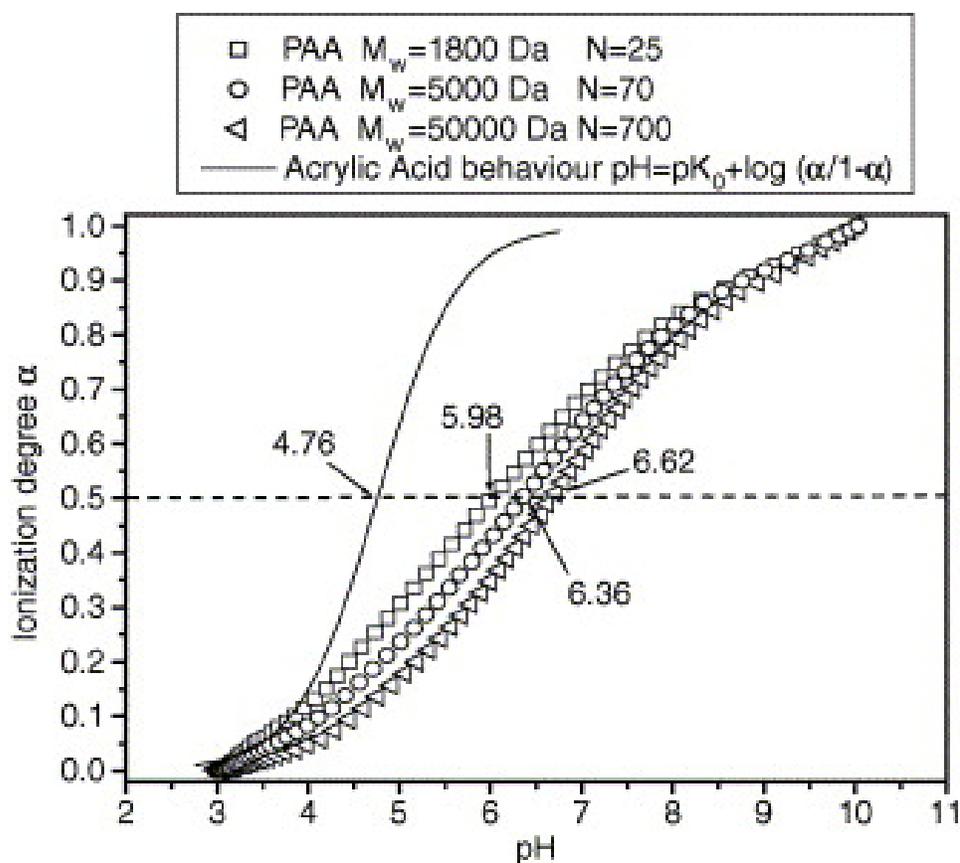
Figure 5.3: Experimental titration for PAA of various chain lengths. Reprinted from [192] with permission from Elsevier.

Config	pH	Number of functional groups ionised	Complexation	Polymer environment
1	6.5	20	Ca^{2+} close (<5 Å)	surface
2	6.5	20	Ca^{2+} close (<5 Å)	solution
3	6.5	20	Ca^{2+} far (>10 Å)	surface
4	6.5	20	Ca^{2+} far (>10 Å)	solution
5	9.5	28	Ca^{2+} close (<5 Å)	surface
6	9.5	28	Ca^{2+} close (<5 Å)	solution
7	9.5	28	Ca^{2+} far (>10 Å)	surface
8	9.5	28	Ca^{2+} far (>10 Å)	solution

Table 5.1: Details of the different poly acrylic acid configurations used for the metadynamic simulations.

biased potential that is added to the total potential in the form of added Gaussians at a given point on the trajectory described using a set of collective variables (CVs). The addition of this history dependent bias in combination with properly chosen Gaussians leads to the generation of a free energy surface as a function of the chosen CVs. As the aim of this work was to gain a better understanding of the conformational behaviour of the polymer, the shape of the polymer was used to define the CV that control the biased potential to be added the the total potential. In polymer chemistry the radius of gyration can be used to describe the size and shape of a polymer, and can be used to differentiate between polymer configurations - in particular between coiled, globular configurations and extended, open configurations. The radius of gyration (R_{gyr}) is written as the mass weighted rmsd of all n atoms to the centre of mass of the polymer (Equation 5.1):

$$R_{gyr} = \left(\frac{\sum_i^n |r_i - r_{cm}|^2}{\sum_i^n m_i} \right)^{\frac{1}{2}} \tag{5.1}$$

where n is the number of atoms in the polymer, r_i is the three dimensional position of atom i, m_i is the mass of atom i and r_{cm} is the centre of mass of the polymer and is defined as Equation 5.2

$$r_{cm} = \frac{\sum_{i}^{n} r_i m_i}{\sum_{i}^{n} m_i} \tag{5.2}$$

To sample the whole configurational space of the polymer and escape the local energy minima of certain configurations, the radius of gyration of the polymer was used as the CV for the metadynamics simulations.

5.2.3 Modelling protocols

The molecular dynamics package `DL POLY Classic` [140] in combination with the metadynamics plug-in `Plumed v. 1.3.0` [197] was used for all simulations. The calcite potentials developed by Raiteri *et al.* [26] were used for the mineral phase, the organic phase was described by the `AMBER ff12SB` potentials [111] and the method developed by Freeman *et al.* [103] was used to produce cross-terms between the organic and mineral phases, as described in Chapter 2.2.4. The details of the poly acrylic acid potentials used are described in Appendix G. To relax the water on the surface of calcite and around the PAA molecule and counter ions, all simulations were subjected to an equilibration stage, in which the system was simulated using an NVT ensemble, with the metadynamic bias potential set to zero to avoid any biased configurations at this stage. The system was equilibrated at a temperature of 295 K for 1.0 ns with a timestep of 0.1 fs, with a Nosé-Hoover thermostat [119] with a relaxation time of 0.5 ps. The density of the waterbox was monitored during the equilibration period to ensure the correct density was maintained. To ensure the system was fully equilibrated, the configurational energy was analysed during the simulations and observed to level out.

Subsequent metadynamics simulations were run for 2.0 ns in an NVT ensemble with a timestep of 0.1 fs, using a Nosé-Hoover thermostat with a relaxation time of 0.5 ps. The radius of gyration of the PAA polymer (Equation 5.1) was used as CV to bias the simulation. The timestep, width and height of added Gaussians were chosen such that the system explored a large configurational space efficiently whilst ensuring that the

energy of the system stayed consistent. The Gaussians were added every 0.5 ps and the Gaussian height (the weight of the added bias) and width were chosen to be $\omega = 0.05$ kJ mol^{-1} and $\delta s_\alpha = 0.35$ Å respectively. As only a single CV was used during the metadynamics the free energy surface of the system was reconstructed by summing the Gaussians added during the simulation trajectory.

Four configurations were taken from the metadynamics simulations to calculate the adsorption energy of the PAA polymer on the surface of calcite: a coiled configuration in solution; a coiled configuration adsorbed on the surface of calcite; an extended conformation in solution; and an extended conformation on the calcite surface. Whereas metadynamics simulations provide a free energy landscape as a function of the CV, these simulations were performed to retrieve a configurational energy of the system. Molecular dynamics simulations without metadynamics were performed with these starting configurations. The simulations were run in an NVT ensemble with a timestep of 0.1 fs and a Nosé-Hoover thermostat with a relaxation timestep of 0.5 ps. The system was equilibrated for 0.5 ns, after which a data production simulation of 1.0 ns was performed.

5.2.4   Experimental methods

Experimental studies of the PAA polymer were carried out by the research group of Romero-Gonzalez in close collaboration with the simulations [198]. Using time-resolved anisotropy measurements (TRAMS) the conformational behaviour of a long PAA chain was analysed. The basis of TRAMS is the observations of time dependent intensities of fluorescence emitted in planes parallel and perpendicular to a polarised light source used to excite the sample (Equation 5.3):

$$r(t) = \frac{i_\parallel(t) - G i_\perp(t)}{i_\parallel(t) + 2G i_\perp(t)} \qquad (5.3)$$

where $r(t)$ is the decay of anisotropy of a molecule, $i_\parallel(t)$ and $i_\perp(t)$ are the intensities in the parallel and perpendicular planes respectively, and G is a correction factor [199]. The rotation of a fluorophore attached to a macromolecule is linked to the anisotropy of this macromolecule. The anisotropy decays via a mono exponential function in Equation 5.4:

$$r(t) = r_o \exp(-t/\tau_c) \tag{5.4}$$

where r_o is the intrinsic anisotropy and τ_c is the correlation time which characterizes the motion of the fluorophore. The correlation time relates back to the conformational behaviour of the macromolecule, with a short τ_c being associated with an expanded flexible chain while a large τ_c would be consistent with a collapsed slow moving globular structure [200–202].

5.3 Results and Discussion

5.3.1 Poly acrylic acid conformation in solution

The free energy surface of the PAA polymer in solution was reconstructed by summing the bias Gaussians that were added to the total potential during the trajectory of the simulation. This free energy surface is shown in (Figure 5.4) and demonstrates the influence of the Ca^{2+} counter ions present in the solution on the conformation of the polymer. It is evident from the radius of gyration shown in the free energy surface plot that when the Ca^{2+} counter ions are in close proximity to the polymer, the polymer adopts an open and extended conformation. When these counter ions are moved further away from the polymer, the preferred conformation of the polymer is a random-coiled, closed conformation, evident from the deep well of ca. 6 kJ mol^{-1} in Figure 5.4. The presence of the positively charged counter ions in proximity to the polymer forces the

polymer into a more open and extended conformation due to the strong interactions between the counter ions and the carboxyl groups of the polymer, an interaction that induces a rigidity to the polymer that is supported by the high density of charge on backbone of the negatively charged polymer.
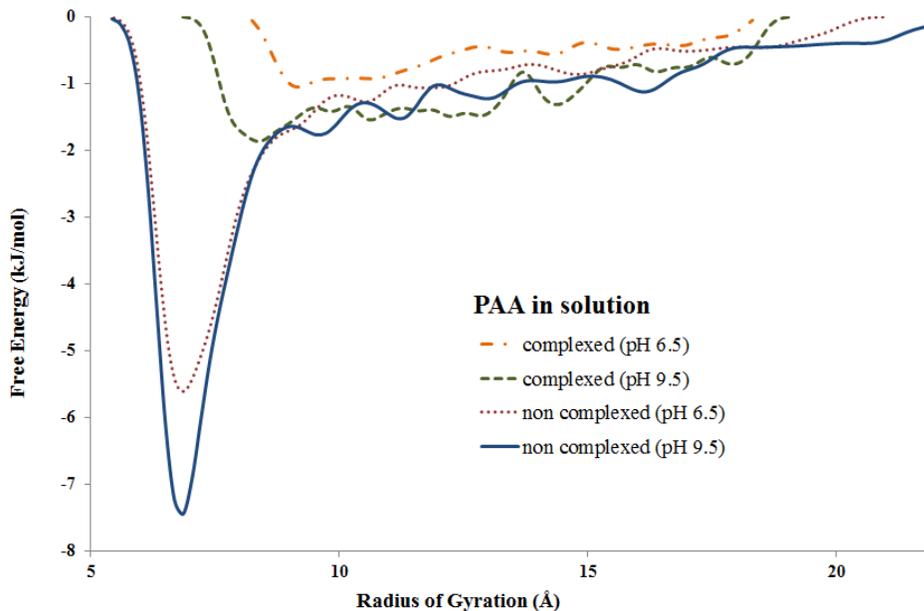
Figure 5.4: Free energy surface of the radius of gyration of the polymer in solution, as calculated with the metadynamics simulations.

The complexation of the polymer throughout the simulations is found to be constant. The Ca^{2+} counter ions that are bound to the carboxylate groups stay associated to the same groups throughout the simulation and do not dissociate. Ca^{2+} ions that have been placed further away from the polymer in the starting configuration ($>10Å$) tend not to associate with the carboxylate groups during the simulation time. The radial distribution function of the Ca^{2+} to carboxylate-oxygen distance is plotted in Figure 5.5 and shows a strong association of the carboxylate groups to the Ca^{2+} ions during the simulations of the complexed polymer (solid blue line). In the simulations where the counter ions are moved away from the polymer, there is no interaction observed between the two (dashed red line). The strong binding of the Ca^{2+} ions to the carboxylate groups that is observed throughout these simulations constrains the polymer backbone and promotes an extended, less flexible conformation of the polymer. Molecular dy-
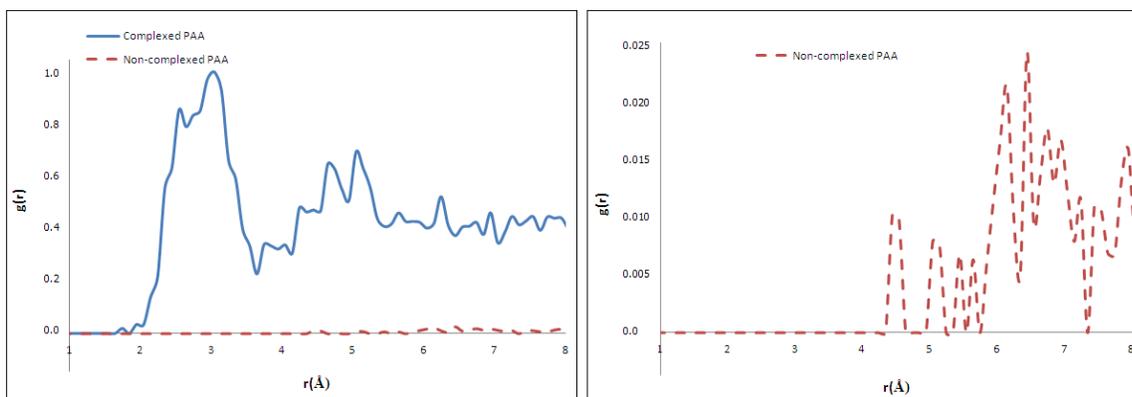
Figure 5.5: Radial distribution function of the calcium to carboxylate-oxygen in solution. The solid blue line shows the rdf for the PAA with Ca^{2+} ions complexed to the polymer; the dashed red line shows the rdf for the Ca^{2+} not complexed to PAA. The radial distribution function for the non-complexed system is expanded on the right.

namics simulations of Ca^{2+} ions and PAA chains in water by Bulo *et al.* [124] showed a similar results with strong interactions between the carboxylate groups and the counter ions.

The difference in configurational between the polymer complexed by counter ions and not complexed with counter ions can be calculated from the MD simulations that were run without metadynamics. When comparing the conformational energies of the two systems, an energy difference of -23.7 ± 16.6 kJ mol^{-1} and -37.5 ± 20.9 kJ mol^{-1} at pH 6.5 and pH 9.5 respectively can be observed. These results show that the complexation of the polymer in solution by Ca^{2+} counter ions is an energetically favourable reaction.

The changes observed in the conformational behaviour of the PAA polymer in solution from a coiled conformation with a small radius of gyration to a more extended, open conformation is also visible in experimental TRAMS measurements. Previous work by Ebdon *et al.* [203] and Soutar *et al.* [200] show that PAA that was tagged with an acenapthylene (ACE, Figure 5.7a) fluorescence label collapses into a partially coiled conformation at low pH values, whereas it will adopt an extended chain conformation at high pH values due to the repulsion of negative charges on the carboxylate functional groups. The experimental work used in the work of Sparks *et al.* [198] used PAA poly-
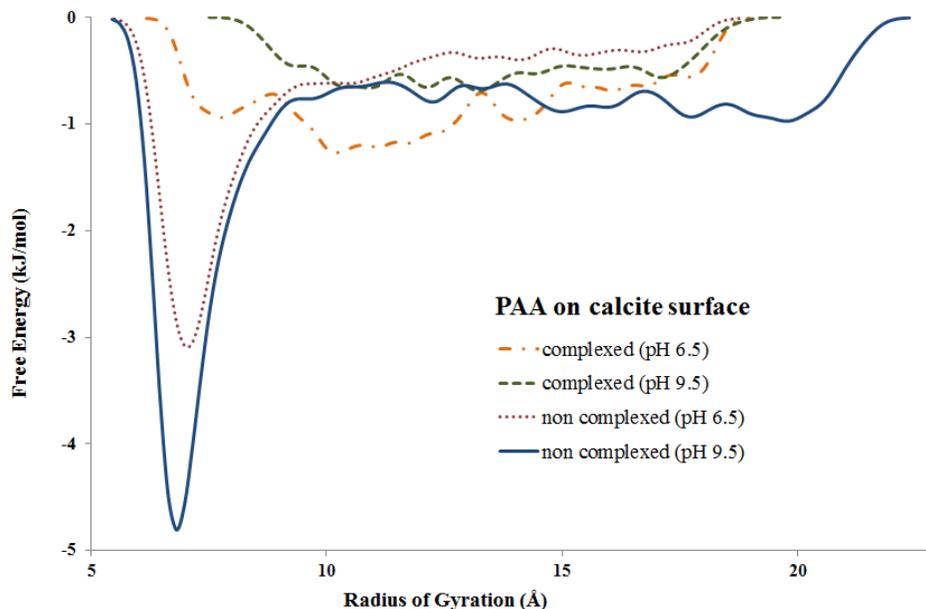
Figure 5.6: Free energy surface of the radius of gyration of the polymer on the surface of calcite, as calculated with the metadynamics simulations.

mers labelled with ACE and 4-amino napthalene-1-sulfonic acid (AmNS, Figure 5.7b) and is shown in Figure 5.8. The conformational behaviour observed in these experiments is similar to previous studies [200, 203]. The data with the AmNS fluorescence label shows the same trend as the ACE-labelled PAA, although the behaviour is less pronounced. This difference can be explained by the differences in the method of attachment of the fluorescence label to the polymer. The ACE label is rigidly attached to the polymer via two bonds and can not rotate with respect to the polymer backbone. AmNS on the other hand is only attached through one bond and can therefore rotate freely through this bond. The effect of this rotation is that the data obtained for AmNS not only describes the motion of the backbone polymer, but also the rotation of the fluorescence tag. Due to this extra rotation the data obtained from experiments will exhibit shorter τ_c values, although these data are still of value in revealing the conformational behaviour of the polymer since it is likely to be sensitive to the microviscosity of the environment.

Previous studies [201, 204] with polymethacrylic acid showed that various fluorescence labels could distinguish between conformational changes of the polymer, irrespective

143

of their mode of attachment to the polymer. The data presented in Figure 5.8 show that above pH 8 the polymer exhibits a short τ_c due to the free rotation of both the ACE- and AmNS-labels, caused by the extended conformation of the polymer. At pH values below 6, the polymer contracts into a partially coiled structure which inhibits the free movement of the fluorescence labels and causes a longer τ_c. These anisotropy measurements follow the results obtained by Laguecir *et al.* [192], who observed a polymer expansion at high pH values and a contraction of the polymer at low pH values with a PAA polymer tagged with a Rhodamine 123 fluorescence tag.
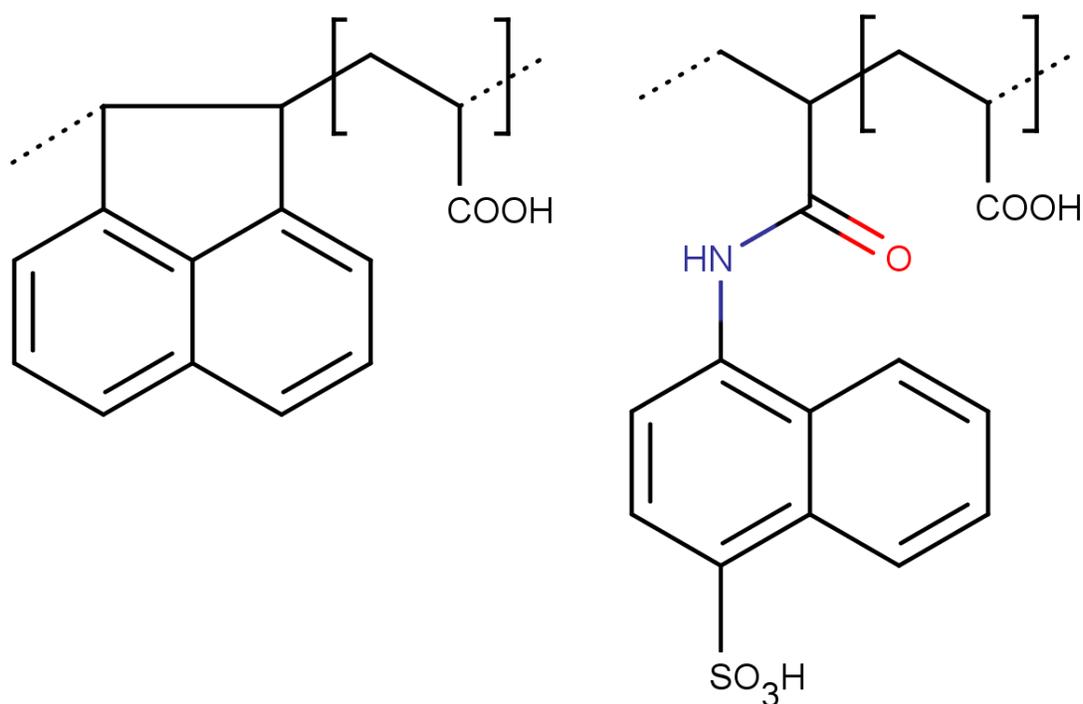
(a) Acenapthylene. (b) 4-amino napthalene-1-sulfonic acid.

Figure 5.7: Chemical structure of the fluorescence tags used in the experimental studies.

5.3.2 Poly acrylic acid conformation on calcite surface

The reconstructed free energy surface of the polymer on the surface of calcite is shown in Figure 5.6. The behaviour of the PAA chain on the surface of calcite during the metadynamics simulations shows a similar behaviour to the polymer in solution. The
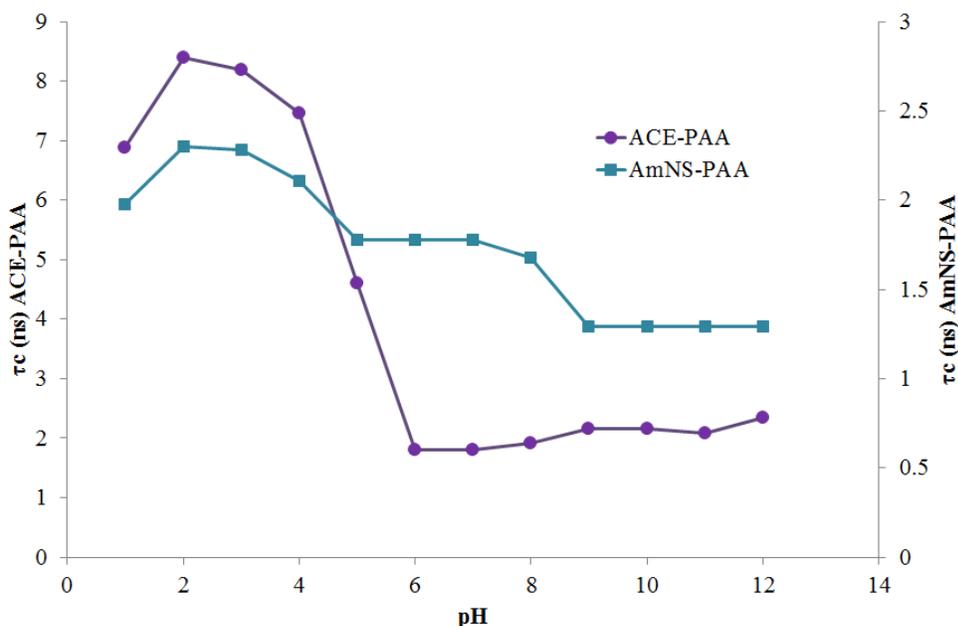
Figure 5.8: Conformation of PAA as a function of pH using time resolved anisotropy measurements (TRAMS). From Sparks *et al.* [198].

radius of gyration of the PAA polymer is seen to be dependent on both the pH of the system and the presence of counter ions near the polymer. The complexation of the polymer induces an open, extended conformation of the polymer whereas a coiled, closed conformation is observed when the counter ions are absent. Throughout the simulations the polymer stays in close contact with the surface and the strong interaction between the calcite surface and the polymer inhibits the polymer from desorbing.

Experimental data for the AmNS-labelled PAA polymer on the surface of calcite has been produced by Sparks *et al.* [198] using TRAMS measurements, and can be seen in Figure 5.9. From the TRAMS data the correlation time can be observed to increase at pH 7, 9 and 11. Previous studies have reported extremely long correlation times for polymers upon binding to surfaces. Soutar *et al.* [202] studied a fluorescence-labelled polydimethylacrylamide polymer and reported a change in τ_c from ca. 2 ns for the polymer in solution to ca. 1 μs when adsorbed on the surface of silica. The increase in correlation time in the studies by Sparks *et al.* indicates that the polymer has a smaller radius of gyration, indicating that the conformation of the polymer exhibits a partially collapsed, coiled structure on the surface of calcite. However, these data may

also reflect the fact that motion of the macromolecular chain is restricted as adsorption proceeds and PAA binds to the mineral. At 1.0 wt% added calcite, the correlation time decreases for both pH values. This drop off in correlation time (or increase in radius of gyration of the polymer) can be explained by a saturation of the calcite surface by a full layer of PAA chains. Additional polymer chains added to the solution cannot adsorb to the surface and are either floating in solution or are loosely bound to other polymer chains in the first adsorption layer. The same effect can be observed when analysing the absorbed amount of PAA onto the calcite surface using steady state fluorescence spectroscopy (Figure 5.10) where the amount of adsorbed PAA on calcite reaches a plateau at a PAA concentration of ca. 3 mM. Computational studies performed by Molnar *et al.* [205] have shown a similar effect, where Ca^{2+} ions associate with single PAA chains, shielding the negative charges on the polymer chains and facilitating the agglomeration of polymers on the mineral surface to form biofilms.
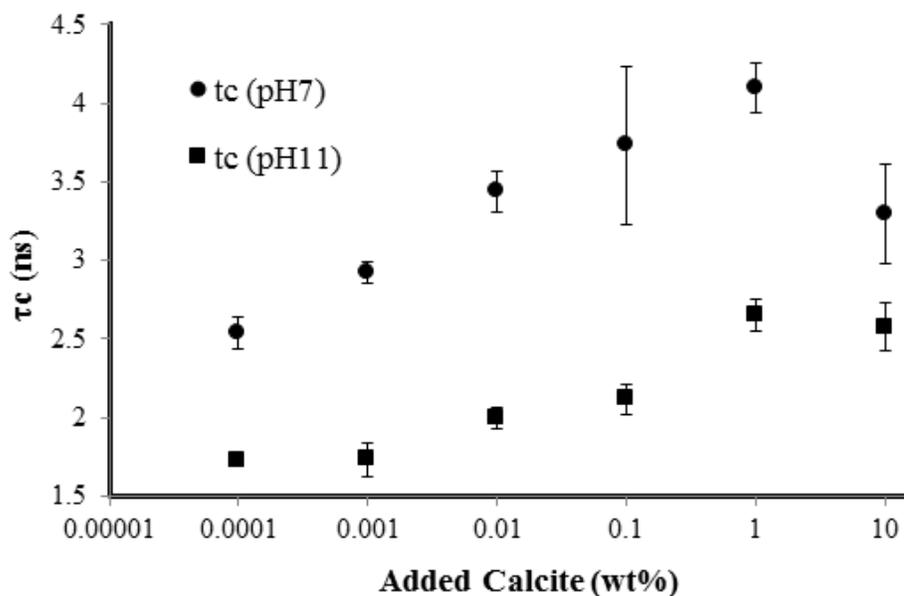
Figure 5.9: Conformation of PAA adsorbed to the surface of calcite as measured by TRAMS. From Sparks *et al.* [198].
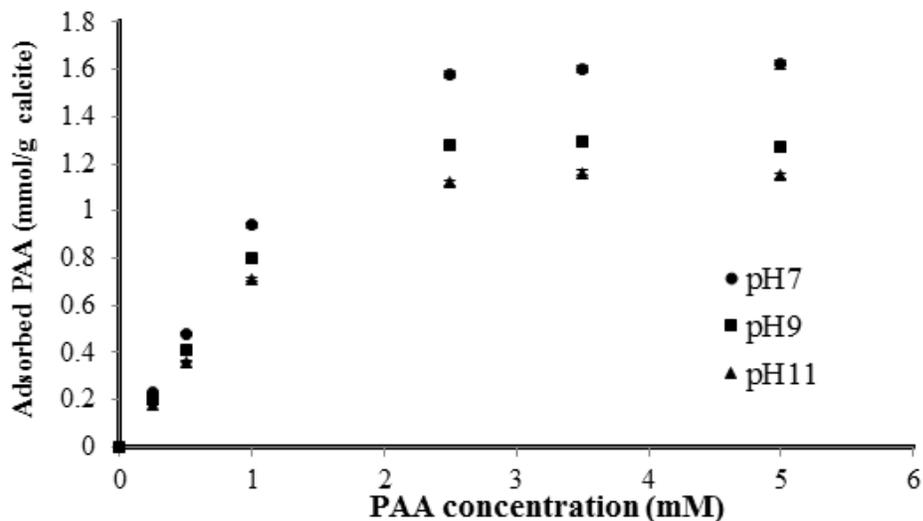
Figure 5.10: Experimental adsorption of AmNS-labelled PAA onto calcite surface at pH 7 and pH 11, measured using solid state fluorescent spectroscopy. From Sparks *et al.* [198].

5.3.3 Adsorption energies

The adsorption energy of the PAA polymer on the surface of calcite can be calculated using the energetics from the molecular dynamics simulations of the polymer in solution and on the surface of calcite. Using Equation 5.5 the adsorption energies of the polymer ($E_{adsorption}$) with respect to the water-only solvated surfaces ($E_{water-calcite}$) and the solvated polymer ($E_{water-polymer}$) were calculated:

$$E_{adsorption} = (E_{water-calcite-polymer} - E_{water-calcite} - E_{water-polymer}) \qquad (5.5)$$

where $E_{water-calcite-polymer}$ is the configurational energy of the solvated calcite-polymer system. This method calculates the adsorption energy for the PAA polymer taking into account only the enthalpic contribution of the system. However, using this method the entropic contribution to the binding is ignored. The adsorption of the PAA macromolecules to the surface of calcite will be a positive entropic process, as the binding process will disrupt the highly ordered water layers on the surface of calcite. In order to calculate an estimate of the free energy of adsorption (which takes into account

pH	$E_{adsorption}$[a] (kJ mol^{-1})		n^b		TnS^c (kJ mol^{-1})		A^d (kJ mol^{-1})	
	ext'ed	coiled	ext'ed	coiled	ext'ed	coiled	ext'ed	coiled
6.5	-325±28.9	-203±40.2	30.1	15.4	53.8	27.5	-118.1	-30.2
9.5	-317±33.0	-178±35.5	30.7	14.8	54.9	26.5	-112.1	-15.6

Table 5.2: Calculated values for the energy of adsorption of extended and coiled PAA chains onto the surface of (10$\bar{1}$4) calcite. [a] enthalpy of adsorption from simulations; [b] number of displaced water molecules; [c] entropic contribution due to displaced water molecules; [d] estimated free energy of adsorption.

the entropy of the system), the entropic energy of the release of these water molecules needs to be taken into account. An estimate of the free energy of adsorption, A, can be made by assuming the most important entropy term arises from the displacement of water molecules from the surface. As our simulations show a similar conformational behaviour of the PAA polymer in solution and on the surface of calcite, it is a reasonable assumption that the conformation entropy of the molecules in both cases will be similar. Therefore the estimate of the free energy of adsorption can be written as Equation 5.6:

$$A = U - TnS \tag{5.6}$$

where U is the internal energy of adsorption, T is the temperature, n is the number of water molecules displaced from the calcite surface during adsorption and S is the entropic contribution arising from the displacement of a single water molecule. Freeman and Harding [81] estimated this entropic contribution for a water molecule on the surface of (10$\bar{1}$4) calcite to be ca. 6 J mol^{-1} K^{-1}. From the PAA simulations performed in this work both the configurational energy of adsorption as calculated by Equation 5.5 and the estimated free energy of adsorption as calculated by Equation 5.6 on the (10$\bar{1}$4) calcite surface were calculated. The data for these energies are shown in Table 5.2. The predicted binding energy decreases with increasing pH values, indicating that the polymer binds more strongly to the calcite surfaces at lower pH values.

An adsorption energy can also be calculated from experimental data by extracting the data from the steady state fluorescence experiments (Figure 5.10). By fitting an adsorption isotherm to the experimental data, an adsorption constant can be obtained, with which the energy of adsorption can be calculated. The Langmuir isotherm, developed in the early 1900s [206], is one of the most used adsorption isotherms and has the functional form of Equation 5.7:

$$q_e = \frac{q_m K_a C_e}{1 + K_a C_e} \tag{5.7}$$

where q_e is the equilibrium adsorption capacity (mg g^{-1}), q_m is the maximum adsorption capacity (mg g^{-1}), K_a is the adsorption equilibrium constant (L mg^{-1}) and C_e is the equilibrium concentration (mg L^{-1}). With the experimental data the equilibrium adsorption constant K_a can be obtained, and from the relationship between ΔG^0 and K_a in Equation 5.8 the Gibbs free energy change ΔG^0 of the system can be calculated:

$$\Delta G^0 = -RT ln K_a \tag{5.8}$$

For the calculation of the adsorption constant with a Langmuir isotherm, the following assumptions are made: The surface to which the molecules adsorb is flat and homogeneous; all adsorption sites have the same energy; no more than a single mono layer can be adsorbed onto the surface; and adjacent molecules adsorbed onto the surface do not interact with each other. It should be pointed out that the saturation seen in experiments does not imply that all the binding sites are individually occupied. Large molecules can occupy more than one binding site. Indeed, jamming [207], whereby the molecules block large number of possible sites, is almost inevitable in a system of this kind. A Langmuir isotherm energy may include mediated bonding through calcium ions and even a contribution from PAA molecules that are loosely attached to an absorbed

PAA layer. Sparks *et al.* [198] fitted a Langmuir isotherm to the experimental data and obtained the adsorption free energies of -2.12, -1.41 and -1.18 kJ mol^{-1} at pH 7, 9 and 11 respectively. The negative values for ΔG indicates a favourable binding of PAA to the surface of calcite, with the polymer binding more strongly to the surface at lower pH values, when less functional groups are ionised. These experimental adsorption energies are within the same range of those obtained by Geffroy *et al.* [194], who measured an adsorption enthalpy of $+2$ kJ mol^{-1} for PAA adsorption onto calcite.

The difference between the free energy obtained from the calculated energy of absorption and the Langmuir isotherm analysis of the experimental absorption curve for the PAA molecule can be explained by the fact that these two values refer to different quantities. Due to the small chain length of the polymer simulated in the computational studies, the calculated free energy of adsorption refers to the binding of a small molecule on the surface of calcite, in which a large number of the functional groups are involved with the direct binding to the surface. For the experimental obtained free energy, this is calculated with respect for a much larger polymer system, where parts of the macromolecule can exhibit different modes of binding (Figure 5.11), such as loops (part of the polymer extends into the solution), trains (extended parts of the polymer attached to the surface) and tails (end of the polymer chain extended into solution). The longer polymer chain in the experimental studies will exhibit these behaviours upon binding, whereas the shorter polymer used in simulation studies (by necessity) will not be able to exhibit all these conformations. As the computational simulations will only represent part of the binding process of the polymer (indicated by the red box in Figure 5.11), this will result in a higher effective binding energy per unit length of the polymer than seen in the experimental studies.
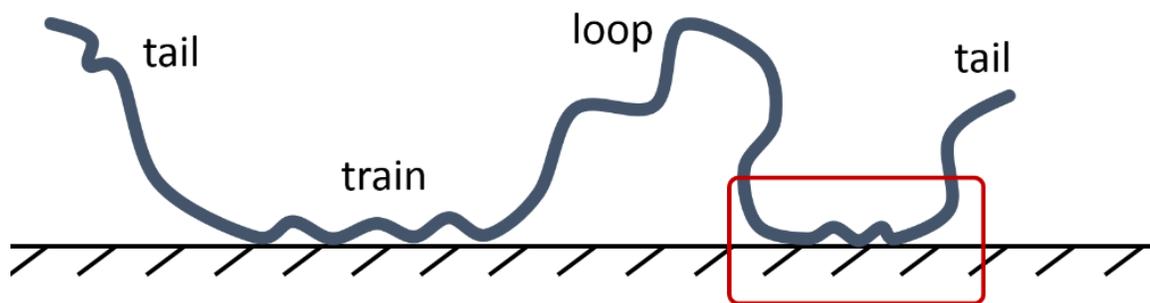
Figure 5.11: Polymer attachment mechanism to surfaces, showing trains, loops and trails. The smaller polymer chain used in the computational simulations will be able to represent only a small section of the polymer (indicated by the red box) and thus the calculated binding energy will differ to that measured experimentally.

Conformation	pH 6.5	pH 9.5	Average	Percentage of total functional groups
Extended	7.1±0.9	6.7±0.8	6.9	25%
Coiled	4.0±0.6	3.9±0.7	4.0	14%

Table 5.3: Number of functional groups of PAA interacting with the calcite surface.

5.3.4 Binding mechanism

The number of functional groups interacting with the calcite surface upon binding can be calculated from the computational simulations and are displayed in Table 5.3. The number of displaced water molecules depends strongly on the conformation adopted by the PAA chain, with twice as many water molecules displaced when the polymer adopts an extended conformation than for a coiled conformation. Similarly, the number of functional groups involved with the binding process changes depending on the adopted conformation. In an extended conformation (Figure 5.12) the average number of carboxylic acid groups interacting with the surface is 6.9, corresponding to 25% of the total number of functional groups on the polymer. This number drops to 4.0 (14%) for a more coiled conformation (Figure 5.12). The configurational binding energies per functional group interacting with the surface at pH 9.5 are -45.7 ± 17.3 kJ mol^{-1} (coiled) and -47.3 ± 10.6 kJ mol^{-1} (extended) are not significantly different and suggest that there is a much larger entropic gain with the adsorption of the extended conformation on the surface of calcite than with the coiled conformation.
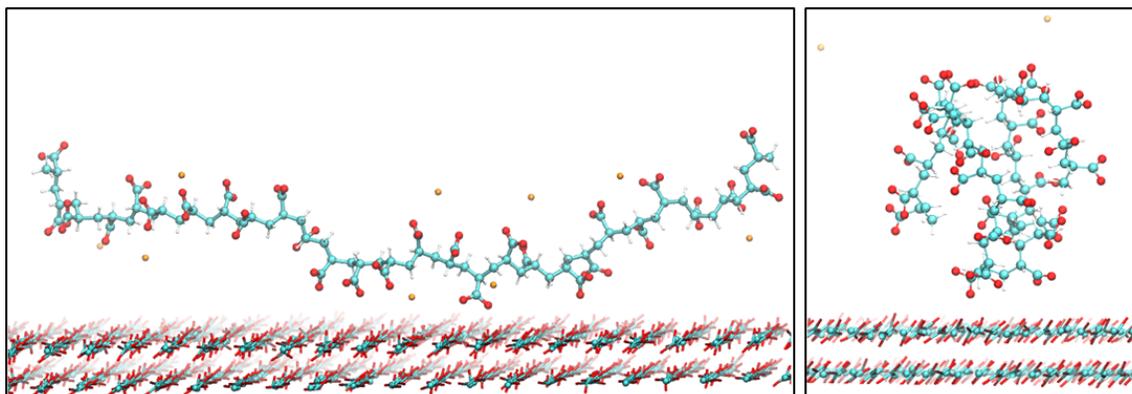
Figure 5.12: Typical binding motif of an extended PAA chain (left) and a coiled PAA chain (right) on the surface of 10.4 calcite. Oxygen, carbon, hydrogen atoms and Ca^{2+} counter ions are depicted in red, cyan, white and orange respectively. Water molecules are omitted for clarity.

Table 5.3 shows that only up to a quarter of the functional groups present in the polymer are involved with the binding process, and that this number does not drastically increase with more charged functional groups at higher pH values of the environment. When analysing a typical binding motif of an extended polymer chain on the surface of calcite (e.g. Figure 5.12) it can be seen that only the central part of the polymer stays in constant contact with the surface, much the same as seen in the train like conformation of longer polymeric chains (Figure 5.11). Additionally, only alternating functional groups in the central section of the polymer are seen to be involved in the binding process. This behaviour suggests that not only the number of charged functional groups present will dictate the binding of a molecule to surfaces, but also the spacing between these groups. On the surface of $(10\bar{1}4)$ calcite there are a lot of exposed CO_3^{2-} ions present which could hinder the adsorption of molecules with a high density of negative charges. The complexation of the polymer with positively charged calcium ions might therefore be aiding the adsorption of the polymer to the calcite surface by shielding these negative charges. The counter ions are complexed to the polymer for extended periods of time and continue to be associated with the polymer when bound to the surface, indicating this complexation could be a mechanism of calcium transport to the surface.

5.4 Conclusion

Whereas previous computational studies have focussed on the binding energies of PAA on the surfaces of minerals and on the binding of free calcium ions in solution, the work presented here is a study of both the adsorption of longer polymer chains on the surface of calcite and the conformational behaviour of these chains. Using molecular dynamics simulations in combination with metadynamics, a good comparison could be made with the experimental results obtained by the research group of Romero-Gonzalez.

The role of calcium counter ions in the system is of great importance for the conformation and binding of the polymer. The polymer complexed with Ca^{2+} counter ions will prefer an extended conformation in which the counter ions are strongly associated with the carboxylate functional group throughout the simulations, whereas the conformation adopted by the polymer when not complexed by counter ions is a coiled, sphere like structure. The extended conformation has a higher adsorption energy on the calcite surface than the adopted conformation of the polymer when not complexed. The long time-scales of the polymer - calcium association, even when binding to the surface of calcite, provides an insight of a possible mechanism of how the calcium ions could be transported through the tightly bound water layers to the surface of calcite.

The computational methodology developed in this work to study the conformation and binding of large macromolecules in solution and on the surface of calcite can be used in future to study similar sized macromolecules. The large differences in adsorption energies between experiment and simulation can be explained by the differences in chain length of the polymer used as the short chained polymers used within the simulation work are expected to have a higher affinity for the surface compared to those used experimentally. The adsorption experiments performed using steady state fluorescence spectroscopy and the computational simulations of adsorption both indicate that PAA is readily adsorbed onto the surface of calcite due to the functional groups present in the molecule. The strength of this adsorption is dependent on the pH of the system

and the presence of counter ions in solution and the polymer binds more strongly to the surface of calcite in a lower pH environment.

The strong binding of the PAA polymers and the extended conformations of the polymer chains that maximise the points of contact with the surface in the simulations provided here and the experimental work by the research group of Romero-Gonzalez shows that flexible macromolecules with the right functional groups can bind strongly to calcite surfaces. The binding strength of these molecules is dependent on both the pH and counter ions, and can be related to the binding of bacteria via cell-wall macromolecules such as mycolic acid to surfaces and the formation of biofilms.

The enhanced binding of the extended conformation PAA on the surface of calcite shows that flexible chains, such as intrinsically disordered proteins, could be beneficial to the binding of these macromolecules. The fact that only 14 to 25% of the functional groups present on the polymer are involved in the binding process suggests that the total number of functional groups is not the only factor that plays a role in the adsorption of biomolecules on the surface of minerals. In addition the the total charge, the spacing between these charged groups to match the underlying charge pattern of the mineral surface plays a role, a fact that is important for the future design of biomolecules to interact with these mineral surfaces.

Chapter 6

Adsorption of amino acids onto amorphous surfaces

6.1   Introduction

Amorphous calcium carbonate (ACC) is the non crystalline phase of $CaCO_3$ and is highly unstable. Due to this instability the amorphous phase will transform into one of the crystalline phases, a process both thermodynamically favoured and kinetically fast [22, 208, 209]. Biogenic ACC, however, is observed to last for longer time scales. Within many different organisms ACC has been observed to be present either as a precursor for the crystalline $CaCO_3$ phases or for longer periods of time as amorphous material [4, 210].

The difficulty with identifying ACC within organisms is the close association between the amorphous and other crystalline phases within the same material and as such, the use of ACC as a material within biology is probably underestimated. There are, however, a good few examples of the use of ACC within organisms, as shown in Table 6.1 and Figure 6.1. The fact that the ACC is seen to be stabilised within a variety of organisms shows that the crystallisation process of ACC can be inhibited by these

organisms, and several experiments have shown that an organic phase within these organisms is in part responsible for this stabilisation [208, 211].

Organism	Location	Function
Crustaceans	Crab cuticle	Mechanical strength
Echinoids	Larval spicules	Precurser phase
Porifera	Spicules	Strength/protection
Ascidiacea	Spicules	Mechanical strength
Plants	Leaves	Calcium store

Table 6.1: Occurrence of amorphous calcium carbonate in nature.

6.1.1 Earthworms

One of the organisms that has been seen to produce calcium carbonate particles is the earthworm, which excretes $CaCO_3$ granules in a process observed as early as 1880 by Darwin [212]. Earthworms such as the *Lumbricus terrestris* and *Lumbricus rubellus* are responsible for the production of calcium carbonate nanoparticles in soils, which are important in the biogeochemical cycles of soils [213–217]. These calcium carbonate granules range from single crystals to particles of 2.5 mm in diameter [213] and are produced in the calciferous glands of earthworms and subsequently excreted and consist mainly of calcite [214]. However, other phases of calcium carbonate are seen to be present: vaterite, aragonite and anhydrous amorphous calcium carbonate regions are seen to exist for long time scales in these granules (Figure 6.2) [215]. The unusually high stability of this amorphous state, existing for >2 years, is of great interest for a range of applications, including for example the prevention of scaling in pipes used for oil recovery.

As the structure of anhydrous amorphous calcium carbonate is inherently unstable, it is abnormal to see this phase is such high quantities for long periods of time. Recent work by Demarchi *et al.* [218] shows that several amino acids, including glutamic acid and glutamine, are present in high concentrations in the vicinity of the ACC granules
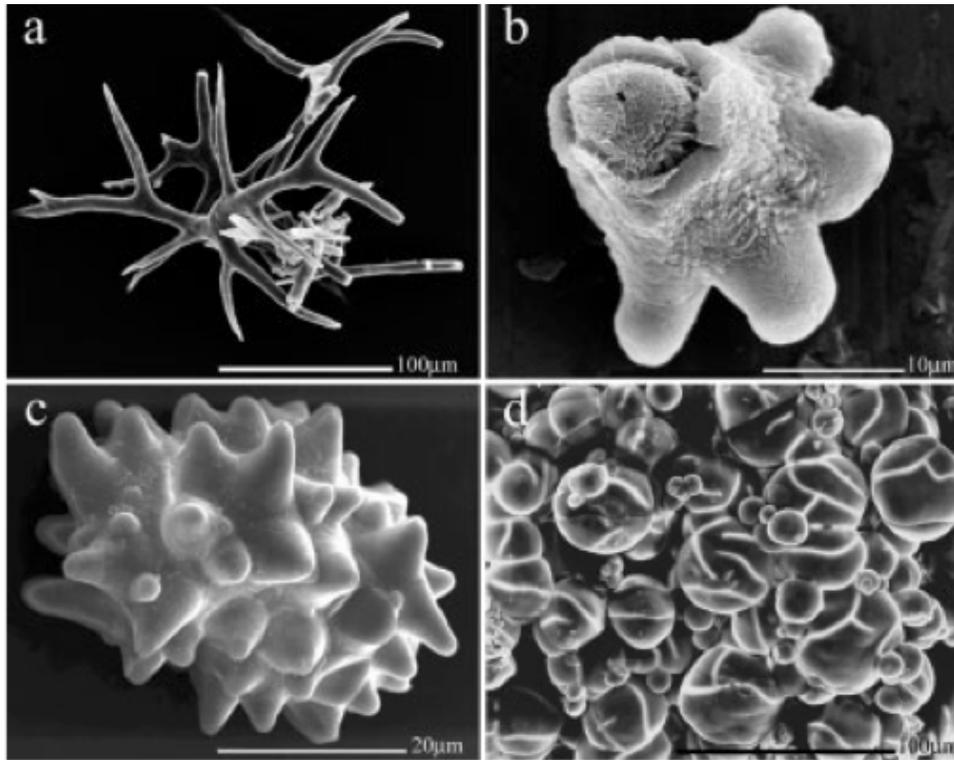
Figure 6.1: SEM images of examples of the use of amorphous calcium carbonate in biology. a) Body spicule from the sea tulip *Pyura pachydermatina* made up completely of ACC. b) Cross section of a tunic spicule of the *Pyura pachydermatina* which shows an ACC core separated from an outer calcite layer by an organic membrane. c) A cystolith, a small agglomeration of mineral ions, from the leaves of a the *Ficus microcarpa*. d) ACC granules from the crustacean *Orchestia cavimana*, a temporarily calcium store. From [5]. Reprinted with permission from the Wiley Company.
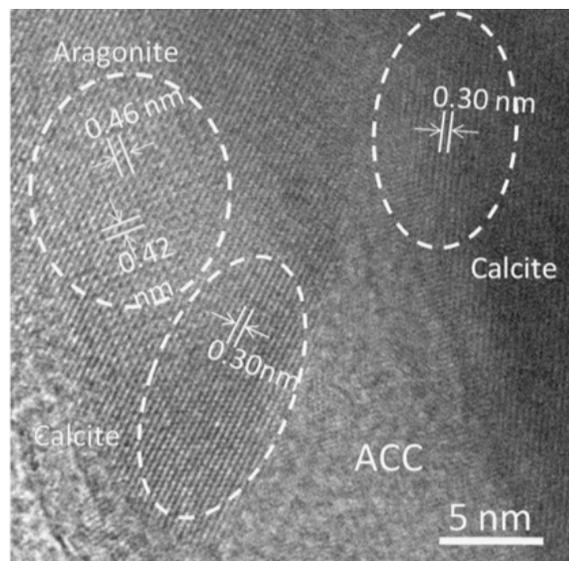
Figure 6.2: High resolution transmission electron microscope image of a granule produced by the earthworm *Lumbricus friendi* shows areas of calcite, aragonite and ACC. Reprinted from [215] with permission from Elsevier.

excreted by earthworms. The unusual high presence of these small amino acids might give an insight into the increased stability of the ACC particles.

6.1.2 Simulation studies

Several simulation studies have looked at the interactions between ACC and biomolecules. Freeman *et al.* [72] performed computational studies of the chicken eggshell protein ovocleidin-17 in the presence of small amorphous calcium carbonate (ACC) nano particles and found that the protein catalyses the transformation of ACC to calcite. Recently, Raiteri *et al.* [219] have performed computational experiments on the interactions between organic molecules and ACC. Using metadynamics simulations with the distance between the organic molecules and separate calcium and carbonate ions as collective variable, they resolved a free energy landscape of the association between $CaCO_3$ and Asp, acetate and citrate. Throughout their simulations they observed that the calcium ions strongly associate with the carboxyl functional groups on the organic molecules. When simulating the adsorption of these organics on the surface of small ACC nano particles they observed a weak solvent separated adsorption of the organics on the ACC surface.

6.2 Aims and objectives

The aim of this preliminary project is to gain an understanding of the interactions between small biomolecules and anhydrous ACC and gain insight into the binding strength of these small biomolecules to amorphous surfaces. Within this work molecular dynamics simulations were performed to analyse the interactions between a range of amino acids and the amorphous surface. Several analysis techniques were used to gain a better understanding of the interactions involved with the binding of the amino acids to the surface, and to gain an understanding of the strength of these interactions.

6.3 Methods

6.3.1 System setup

The charged state of amino acids changes depending on the pKa of the amino acid functional groups and the environment the amino acid is exposed to. The pH values considered in this work were chosen to reflect those normally used within experimental work. Four amino acids were used within this work (Figure 6.3): aspartic acid (Asp, D), glycine (Gly, G), glutamic acid (Glu, E) and arginine (Arg, R). The pKa values for the amine and carboxyl functional groups and the overall charges of these amino acids are detailed in Table 6.2. In order to charge balance the simulations, Ca^{2+} and CO_3^{2-} counter ions were added to the simulations. One Ca^{2+} counter ion was added to the simulation of Asp, one Ca^{2+} counter ion was added to a simulation with two Glu and one CO_3^{2-} counter ion was added to the simulation with two Arg amino acids.

| (a) Asp. | (b) Glu. | (c) Gly. | (d) Arg. |
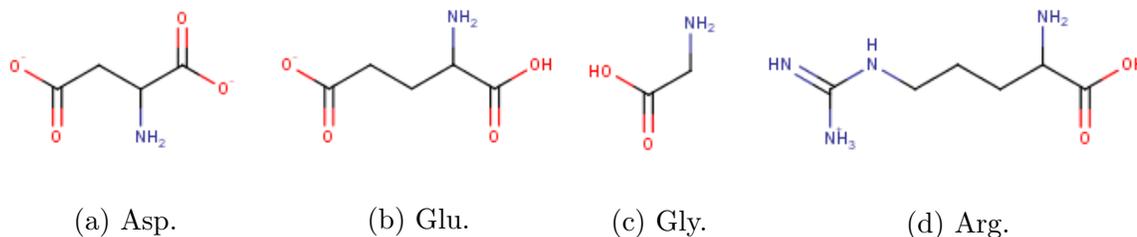
Figure 6.3: The chemical structures of the amino acids used within this work.

Amino acid	pKa(NH_2)	pKa(COOH)	pKa(COOH sidechain)	Total charge
Asp	9.82	2.10	3.86	2-
Gly	9.60	2.34	-	0
Glu	9.47	2.10	4.07	1-
Arg	9.04	2.01	12.48	1+

Table 6.2: Details of the pKa values and the total charge of the amino acids used within this work: Asp, Gly, Glu and Arg.

To model the anhydrous amorphous calcium carbonate a slab of calcite of 27 Å x 28 Å x 29 Å and containing 294 $CaCO_3$ formula units was melted and subsequently quenched

to produce an amorphous $CaCO_3$ surface. Subsequently a water box containing 3000 TIP3P water molecules [104] was placed on top of the amorphous surface using the program `Packmol` [137], ensuring a water density of of 1.0 g ml^{-1} in a setup similar to that as seen in Figure 4.4. The amino acids were placed at a distance of 4 Å from the surface. Two configurations for each amino acid were simulated, in which the starting positions were different.

6.3.2 Modelling protocols

The models of the amino acids in this work were prepared using the `Amber 12 Antechamber` package [138]. The atomic charges of the molecules were calculated using the AM1-BCC method [139] which uses a semi-empirical method with charge corrections to calculate the electronic structure of a molecule, as detailed in Appendix B. The inter- and intra-molecular interactions of the calcium carbonate were described by the potentials developed by Raiteri *et al.* [26] and the potentials from the `AMBER` *ff12SB* forcefield [111] were used for the amino acids and water molecules. The cross-terms between the organic and inorganic ($CaCO_3$) phases were calculated using the method described by Freeman *et al.* [103] (Chapter 2.2.4). Details of the specific potentials for this work are detailed in Appendix H.

All simulations were carried out using the molecular dynamics package `DL POLY Classic` [140]. The calcite slab was simulated at 3000 K for for 3.0 ns with a timestep of 1.0 fs in an NVT ensemble with a Nosé-Hoover thermostat [119] with a relaxation time of 0.5 ps. Subsequently the system was quenched in steps of 300 K with simulations of 1.0 ns at each temperature step using the same parameters. A final simulation was performed at 300 K for 3.0 ns to properly equilibrate the system. After the addition of the water, the systems were equilibrated at a temperature of 300 K for 100 ps with a timestep of 0.1 fs in an NVT ensemble with a Nosé-Hoover thermostat [119] with a relaxation time of 0.5 ps. A second equilibration simulation for 1.0 ns with a timestep of 1.0 fs

was performed. The density of the waterbox was monitored during the equilibration period to ensure the correct density was maintained. The configurational energy of the system during the equilibration simulations was analysed to ensure a fully equilibrated system. Subsequently, a 2.0 ns simulation using the same parameters was carried out for data analysis.

Additional simulations of the water-only solvated amino acids with 3000 TIP3P water molecules and a bulk water system were performed in order to calculate the energies of adsorption. The same simulation protocols were used as described above, and the simulations were performed for 1.0 ns. Using Equation 6.1 the adsorption energies of the amino acids ($E_{adsorption}$) with respect to the water-only solvated surfaces ($E_{water-acc}$) and the solvated amino acids ($E_{water-aminoacid}$) were calculated:

$$E_{adsorption} = (E_{water-acc-aminoacid} - E_{water-acc} - E_{water-aminoacid}) \qquad (6.1)$$

where $E_{water-acc-aminoacid}$ is the conformational energy of the solvated ACC - amino acid system.

6.4 Results and discussion

6.4.1 Amorphous slab

The radial distribution function (RDF) of the calcium to calcium distances for both the calcite slab and the amorphous slab are shown in Figure 6.4. From this RDF it is clear that the crystalline structure of the calcite has disappeared and that the slab is fully amorphous. The short and medium range order that is still present within ACC [5, 220] is also seen within the structure produced here. The long range order, however, has completely disappeared. In order to analyse the surface of the amorphous phase, the density of the water and calcium carbonate in the z-direction throughout the simulation

box has been measured and is displayed in Figure 6.5. It can be seen that the surface between water and calcium carbonate is not well defined and that water penetrates slightly into the structure of the amorphous calcium carbonate.
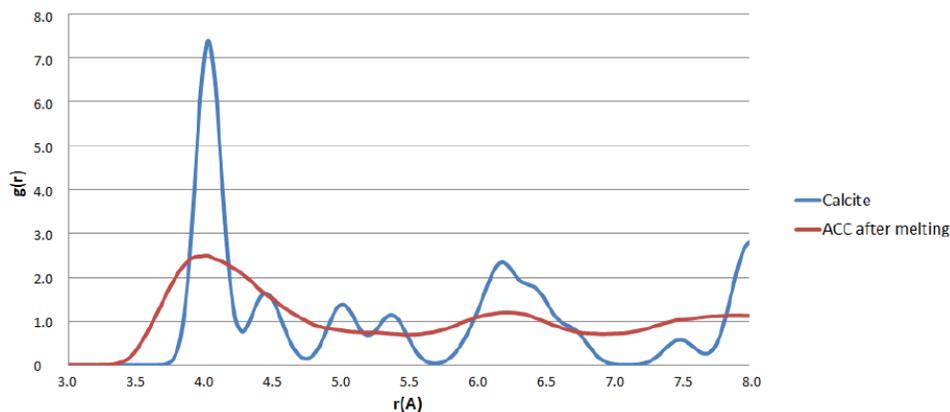
Figure 6.4: Radial Distribution Function of Ca-Ca distances in the system before (calcite) and after (ACC) melting at 3000 K.
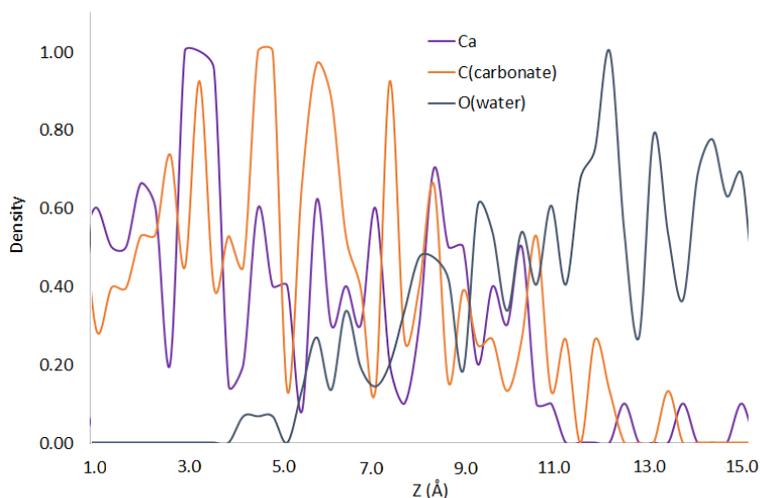
Figure 6.5: Normalised Z-density profile of Calcium, C(carbonate) and O(water) in a simulation with an ACC slab with water. The surface of the amorphous phase is not well defined, and this surface can be seen between 5 and 11 Å.

6.4.2 Adsorption energies

The adsorption energies of the amino acids on the surface of the ACC slab were calculated with Equation 6.1 and these energies and the standard error can be found in

Table 6.3. From these adsorption energies a clear distinction can be made between the adsorption of Arg and the adsorption of Asp, Gly and Glu. The former amino acid exhibits a relatively high, favourable adsorption energy, whereas the latter amino acids show a positive energy. For the three amino acids with a positive energy of adsorption, this energy is seen to decrease with an increase in negative charge of the system, with average adsorption energies of 63.15 kJ mol^{-1} for neutral Gly and 27.53 and 37.50 kJ mol^{-1} for the negatively charged Asp and Glu respectively.

When comparing the modes of binding for all three amino acids, the positive nature of these adsorption energies becomes clear. In Figure 6.6 the binding modes of Asp, Gly and Glu are shown.

Amino acid	Energy config 1 (kJ mol^{-1})	Standard error (kJ mol^{-1})	Energy config 2 (kJ mol^{-1})	Standard error (kJ mol^{-1})
Asp	24.62	4.60	30.44	4.20
Gly	68.54	4.49	57.75	4.62
Glu	44.69	5.23	30.31	4.68
Arg	-139.70	4.32	-121.04	4.74

Table 6.3: Binding energies of the amino acids Asp, Gly, Glu and Arg on the surface of ACC.

Throughout the simulations the amino acids Asp and Glu are seen to hover above the surface at a distance of approximately 2 to 4 Å, with the carboxyl functional groups complexed to a Ca^{2+} ion throughout the majority of the simulations. This strong complexation is common with carboxyl functional groups, as shown in previous computational studies [124]. The amino acids are seen to be inhibited from adsorbing nearer the surface, an effect due to an energy barrier caused by either a charge repulsion or the water structure on top of the surface of ACC. For the amino acid Gly a different interaction is observed, one between the amine functional group and a CO_3^{2-} ion. In the simulations of Gly the amino acid is seen to distance itself approximately 4 Å from the rest of the ACC surface, associating with a carbonate ion situated on top of the ACC surface.

In the case of the amino acid Arg, a different mode of adsorption is observed altogether (Figure 6.7). Throughout the simulations strong hydrogen bonds are developed between the three nitrogen atoms of the Arg side chain and three separate carbonate ions on the ACC surface. A fourth carbonate ion becomes associated with the amine group on the C-terminus of the amino acid, causing the amino acid to fully adsorb onto the surface of the ACC. This strong interaction is reflected in the highly negative adsorption energies observed for the Arg amino acid.

The Z-density profiles for all four amino acids on the surface of ACC are shown in Figure 6.8. A few observations can be made from these plots. Firstly, the surface of the ACC is rough, with the water and mineral interface not well defined. Secondly, some of the Ca^{2+} and CO_3^{2-} ions are seen to be dissociating from the surface. The distances between the ACC and the amino acids is also distinctly different between in the case of Asp, Gly and Glu versus the Arg amino acid. The Z-density plot of the latter shows a closer adsorption to the surface than the other three amino acids, due to the strong interactions with the nitrogen atoms in the side chain of the Arg molecule.
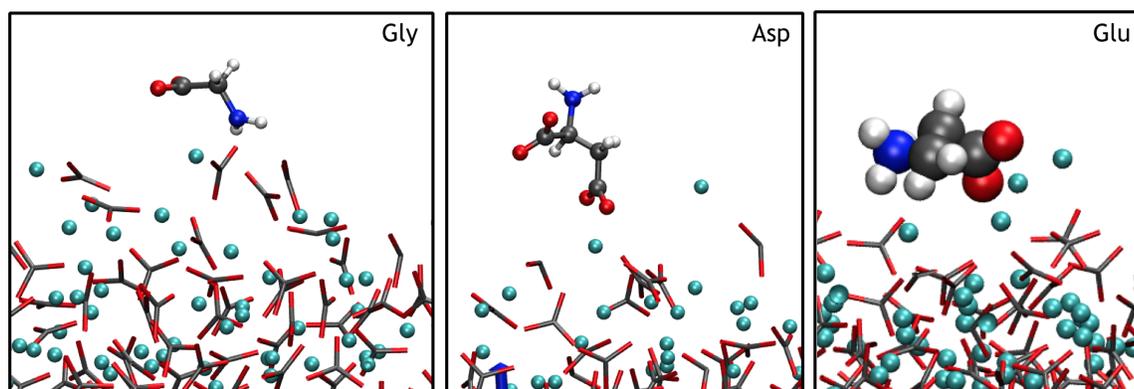
Figure 6.6: The binding configurations of the amino acids Asp, Gly and Glu on the surface of ACC.
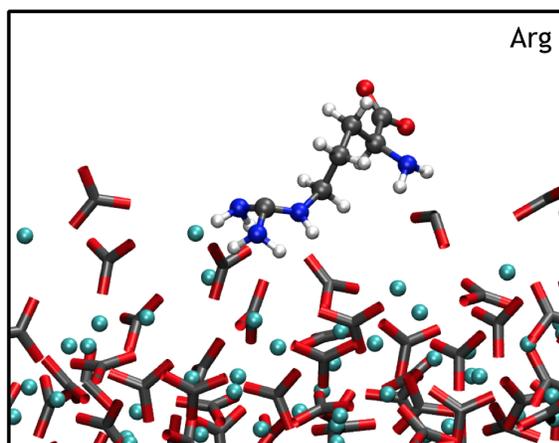
Figure 6.7: The binding configuration of the amino acids Arg on the surface of ACC.

(a) Asp.

(b) Gly.

(c) Glu.
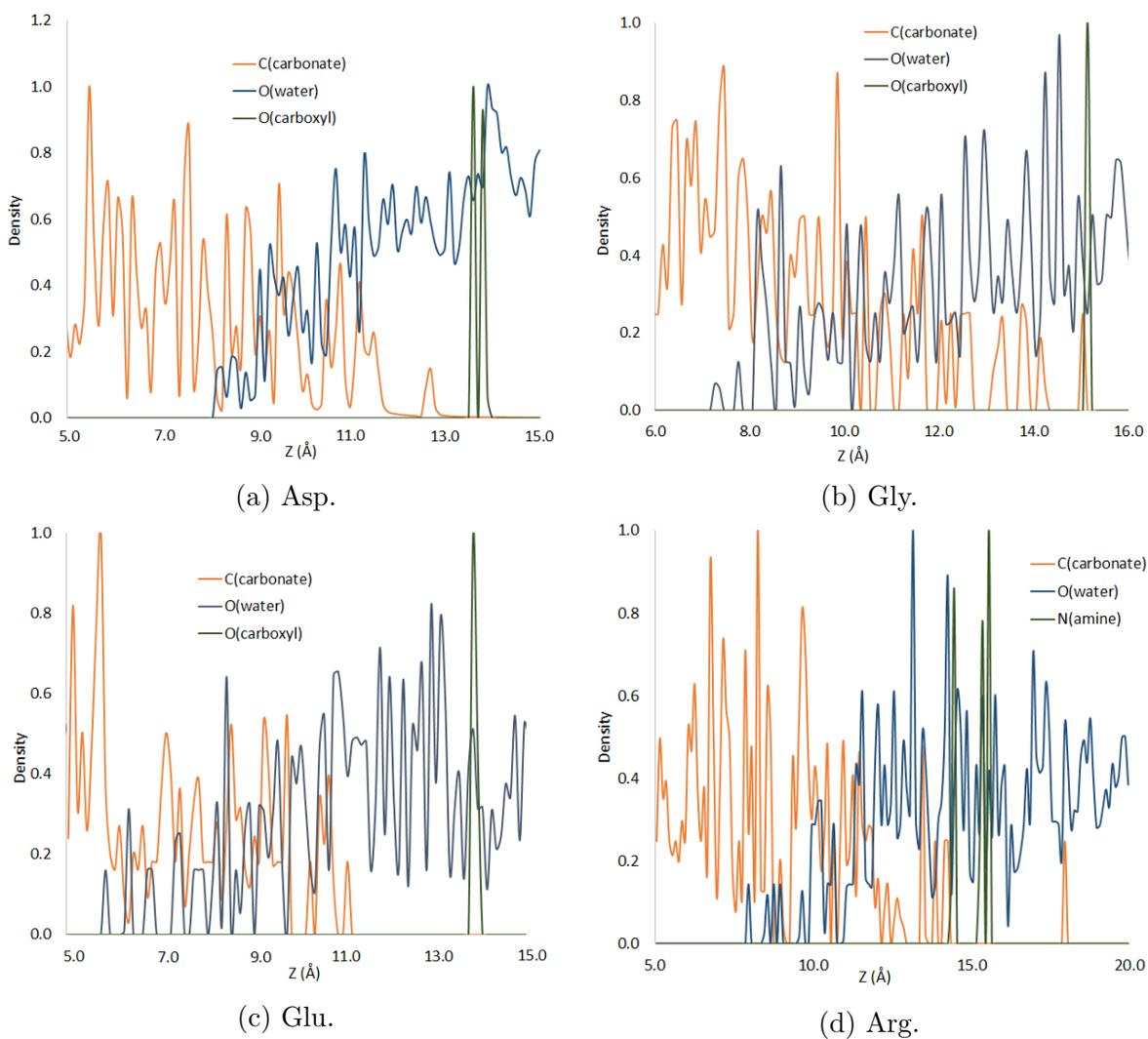
(d) Arg.

Figure 6.8: Normalised Z-density profiles of the amino acids Asp, Gly, Glu and Arg on the surface of ACC. The surfaces between the ACC and water are not well defined, with these surfaces starting at 8 - 10 Å from the centre of the simulation box.

6.5 Conclusion

The binding energies calculated for the four amino acids on the surface of ACC show that there is a weak interaction of the amino acids Asp, Gly and Glu with the surface that involves the interaction between the carboxyl functional groups and calcium ions in the case of Asp and Glu, and an interaction between the amine functional group and a carbonate ion with the adsorption of Gly. The amino acids are separated from the ACC surface by a water layer, a similar behaviour as seen in the recent simulation study by Raiteri *et al.* [219], who show that the binding of Asp, acetate and citrate is mediated through a solvent layer.

The average adsorption energies of 27.53, 63.15 and 37.50 kJ mol^{-1} for the amino acids Asp, Gly and Glu respectively in this work are positive and show that the organic and mineral phases are separated by a water layer. The average adsorption energy of -130.37 kJ mol^{-1} for Arg, however, is negative and is indicative of a highly favourable interaction between the amino acid and $CaCO_3$. The differences in the energies of adsorption seem to rise from the highly favourable interaction between the amine functional groups on the Arg side chain and the carbonate ions in solution.

The simulations performed in this work show that there are a number of interactions present between amino acids and the amorphous surface of calcium carbonate. The fact that strong interactions can be found to exist between both the carboxyl functional group and the mineral cations and the amino functional groups and mineral anions shows that although the adsorption energies for these molecules are positive, there are still several interactions available that may well be able to induce a stabilisation of ACC nano particles. Indeed, the experimental results from Raiteri *et al.* [219] show that the organic additives used in their experiments were able to inhibit nucleation events, stabilise amorphous particles and stabilise pre-nucleation clusters.

Computational simulations give a good indication of the interactions between the organic molecules and the mineral phase, and can shed light on the amorphous nature

of calcium carbonate, such as the nano particles excreted by earthworms in soil samples. The preliminary simulations performed within this work, however, would have to be extended to confirm any of the conclusions drawn from the differences in binding energies. Potential of mean force calculations such as those performed by Raiteri *et al.* [219] would be able to offer more rigorous results than have been calculated to date. Additionally, the simulation of the interactions between a larger range of amino acids and the surface of these ACC nanoparticles could shed light on the relative strengths of these interactions.

Chapter 7

Conclusion and future work

7.1   Inclusion of amino acids into calcite crystals

This work presents the first simulation study of the inclusion of amino acids within calcite crystals. Molecular dynamics simulations show that the inclusions of these biomolecules are associated with a high energy of inclusion. The mode of these inclusions must be via a kinetic rather than a thermodynamic pathway. The high vacancy defect energies for these systems also supports that claim. The simulations of these inclusions show that a 'goodness of fit' principle is maintained, where molecules will get incorporated within the crystal if the disruption to the lattice can be minimised. This is apparent when looking at the structural data for the inclusions of Asp and Gly within the calcite crystal. The amino acid Asp is more readily incorporated within the crystal whilst causing a minimal disruption to the crystal lattice, whereas Gly causes more pronounced lattice distortions. The geometry of Asp is such that the distances between both carboxyl functional groups of the amino acid can mimic both the distances and positions of vacant carbonate ions, with the amine functional group occupying a vacant calcium position.

The anisotropic behaviour of the calcite crystal can also clearly be seen from the simu-

lations performed in this work, and the expansion of the crystal in the c-direction is an order of magnitude larger than the expansion in the a/b-directions. The simulations show that the included biomolecules are positioned in between the planes perpendicular to the c-axis of the crystal, and show a similar behaviour to biogenic calcite with intra-crystalline proteins, suggesting that the inclusion of both small and large biomolecules result in similar structures. The computational simulation results show a good agreement with experimental studies that have been performed in close collaboration.

The work presented here shows that the key feature for the inclusion of biomolecules within the calcite mineral phase is the minimisation of the disruption to the crystal lattice. Those molecules that exhibit a good fit within the lattice will be more readily incorporated. These simulation studies together with the experimental data shows that the mechanical properties of these biocomposite materials can be altered depending on the amino acid content, and shows a way in which composite materials with improved mechanical and physical properties can be produced within synthetic materials design.

The scope for future work within this project is extensive, as the simulations can be extended to include a whole range of biomolecules such as amino acids and larger peptides or proteins, and there is a plethora of experimental work available on the inclusion of these molecules. In addition to energetics and structure analysis the mechanical properties such as the fracture toughness of these composite materials can also be analysed using computational techniques to create a complete picture of the effect of these inclusions.

7.2   Conformational behaviour of biomolecules

The conformational behaviour of both small and large biomolecules in solution and on the $(10\bar{1}4)$ surface of calcite was investigated using computational simulations.

7.2.1 Glu-Asn-Gly

Whereas many computational studies focus on the analysis of single molecules in solution or on the surface of minerals, this study has analysed the conformational behaviour of multiple small Glu-Asn-Gly tripeptides. In solution these peptides show a strong dependence on the presence of other biomolecules in close proximity. Throughout the simulations the flexible structure of single peptides is seen to change when interacting with other peptides. The strong peptide - peptide interactions are caused by extensive hydrogen bonding between the molecules. The Ramachandran maps of the dihedral backbone angles of these molecules show that the peptide - peptide interactions force the peptides into a more rigid conformation with a different set of ϕ/ψ angles.

The adsorption of the peptides on the $(10\bar{1}4)$ surface of calcite show a favourable interaction between the organic and mineral phases. A distinct binding conformation is observed for these peptides via the carboxyl functional group situated on the C-terminus of the peptide. Upon binding the oxygen of this functional group is situated at the same distance from the surface as the first water layer on top of the calcite surface. The rest of the peptide is extended into the aqueous environment away from the surface. This conformation minimises the disruption to the highly ordered water layers on top of the surface of calcite, and maximises the binding energy of the peptides.

The hydrogen bonds between the peptides that are present throughout the molecular dynamics simulations induce the aggregation of two or three peptides into small clusters. The simulations show that this aggregation is a favourable process, a similar result as seen in other simulation and experimental studies of flexible, intrinsically disordered peptides [170, 171]. The data from the simulations performed here show that both the presence of the calcite surface and other biomolecules in solution induce a disorder-to-order transitions within the peptides. This transition alters the three-dimensional structure of the peptide from a flexible molecule to a rigid one with a fixed configuration. This behaviour is similar to that seen for intrinsically disordered proteins, in which the

flexible nature of the molecules allow them to remain versatile and able to bind to many targets.

The simulations show that the key interaction between the Glu-Asn-Gly peptide and the calcite surface is the carboxyl functional group binding to the surface, together with the minimisation of the disruption to the water structure on the surface. It is clear that the mineral surface also has a great effect on the conformational behaviour of the peptides with a flexible to rigid transformation occurring within the peptide.

As the use of intrinsically disordered peptides/proteins throughout the biomineralisation area is widespread, future work on these systems is desirable. Accelerated molecular dynamics such as those performed on the n16N peptide [171] could help to gain a better understanding of the peptide - petide and peptide - mineral interactions.

7.2.2 Poly acrylic acid

The work on poly acrylic acid presented in Chapter 5 is a study on conformational behaviour of a longer polymer in solution and its adsorption on the surface of calcite. In the presence of Ca^{2+} counter ions the polymer will prefer an extended conformation in which the counter ions are strongly associated with the carboxylate functional groups on the polymer. A coiled, sphere like conformation is observed in the absence of these counter ions. When adsorbing onto the $(10\bar{1}4)$ surface of calcite the polymer favours the open, extended conformation over the closed, coiled conformation. The enhanced binding of this extended conformation on the surface of calcite shows that flexible chains, such as intrinsically disordered proteins, could be beneficial to the binding of these macromolecules.

The key features of the interactions between the soft, organic phase and the hard, mineral phase are the carboxyl functional groups. Interestingly, only 14 to 25% of the functional groups present on the polymer are involved in the binding process. This suggests that the total number of functional groups present on a molecule is not the

only factor that influences the strength of adsorption onto the mineral surface. The charge and spacing of these functional groups to match the underlying charge pattern is also of importance, supporting the notion that flexible molecules would be beneficial to binding to mineral surfaces and a fact that could aid the future design of biomolecules to interact with these surfaces. The strong binding of the polymer via the carboxyl functional groups shows the mechanism in which such biomolecules attach to surfaces to form biofilms.

The computational methodology developed in this work can be used in future to analyse the conformational behaviour of similar sized macromolecules. As the work on smaller tripeptides shows a strong aggregation of flexible chains in solution, future work in this area could include the simulation of multiple poly acrylic acid polymers. Additionally, macromolecules such as lipopolysaccharides, mycolic acid and alginate, which are found in bacterial extracellular polymeric substance, could be simulated.

7.3 Adsorption of amino acids on amorphous surfaces

The preliminary simulation study of amino acids on the surface of ACC shows a distinct difference in the mode of binding of different amino acids. The amino acids Asp, Gly and Glu are seen to have a weak interaction with the surface of ACC mediated with a water layer between the organic and inorganic phases, similar to that observed in a recent simulation study [219]. The amino acid Arg on the other hand shows a strong, direct interaction with the surface of calcite in which the amine functional groups on the side chain of the amino acid are bound to carbonate ions on the surface of ACC.

These preliminary simulations would have to be extended to confirm any of the conclusions drawn from the differences in binding energies. Simulations that are able to construct a free energy surface of the interactions between the amorphous phase and

172

the organic phase, such as potential of mean force calculations would be able to offer more accurate results on these interactions. From experimental results [208, 211, 215] it is clear that a range of biomolecules is able to stabilise the amorphous phase of $CaCO_3$, and as such an variety of biomolecules could be simulated in the presence of ACC to uncover the key mineral - organic interactions that lead to this stability.

Bibliography

[1] RM Hazen, D Papineau, W Bleeker, RT Downs, JM Ferry, TJ McCoy, DA Sverjensky, and H Yang. Mineral evolution. *American Mineralogist*, 93(11-12):1693–1720, 2008.

[2] RM Hazen. Evolution of minerals. *Scientific American*, 302(3):58–65, 2010.

[3] RM Hazen, A Bekker, DL Bish, W Bleeker, RT Downs, J Farquhar, JM Ferry, ES Grew, AH Knoll, D Papineau, et al. Needs and opportunities in mineral evolution research. *American Mineralogist*, 96(7):953–963, 2011.

[4] S Mann. *Biomineralization: principles and concepts in bioinorganic materials chemistry*, volume 5. Oxford University Press, 2001.

[5] 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.

[6] EN Maslen, VA Streltsov, and NR Streltsova. X-ray study of the electron density in calcite, CaCO3. *Acta Crystallographica Section B: Structural Science*, 49(4):636–641, 1993.

[7] S Mukherjee. *Applied mineralogy: applications in industry and environment*. Springer Science & Business Media, 2012.

[8] NH de Leeuw and SC Parker. Surface structure and morphology of calcium carbonate polymorphs calcite, aragonite, and vaterite: an atomistic approach.

The Journal of Physical Chemistry B, 102(16):2914–2922, 1998.

[9] NH de Leeuw and SC Parker. Atomistic simulation of the effect of molecular adsorption of water on the surface structure and energies of calcite surfaces. *Journal of the Chemical Society, Faraday Transactions*, 93(3):467–475, 1997.

[10] JPR De Villiers. Crystal structures of aragonite, strontianite, and witherite. *The American Mineralogist*, 56:758–767, 1971.

[11] EN Caspi, B Pokroy, PL Lee, JP Quintana, and E Zolotoyabko. On the structure of aragonite. *Acta Crystallographica Section B: Structural Science*, 61(2):129–132, 2005.

[12] JDC McConnell. Vaterite from Ballycraigy, Lame, Northern Ireland. *Mineral. Mag*, 32:534–544, 1960.

[13] SR Kamhi. On the structure of vaterite CaCO3. *Acta Crystallographica*, 16(8):770–772, 1963.

[14] HJ Meyer. Struktur und fehlordnung des vaterits. *Zeitschrift für Kristallographie-Crystalline Materials*, 128(1-6):183–212, 1969.

[15] R Demichelis, P Raiteri, JD Gale, and R Dovesi. The multiple structures of vaterite. *Crystal Growth & Design*, 13(6):2247–2251, 2013.

[16] L Kabalah-Amitai, B Mayzel, Y Kauffmann, AN Fitch, L Bloch, PUPA Gilbert, and B Pokroy. Vaterite crystals contain two interspersed crystal structures. *science*, 340(6131):454–457, 2013.

[17] H Pauly. Ikaite, a new mineral from Greenland. *Arctic*, 16(4):263–264, 1963.

[18] G Marland. The stability of CaCO3.6H2O (ikaite). *Geochimica et Cosmochimica Acta*, 39(1):83–91, 1975.

[19] KF Hesse, H Küppers, and E Suess. Refinement of the structure of ikaite, CaCO3.6H2O. *Zeitschrift für Kristallographie-Crystalline Materials*, 163(1-4):227–232, 1983.

[20] H Hull and AG Turnbull. A thermochemical study of monohydrocalcite. *Geochimica et Cosmochimica Acta*, 37(3):685–694, 1973.

[21] H Effenberger. Kristallstruktur und infrarot-absorptionsspektrum von synthetischem monohydrocalcit, CaCO3.H2O. *Monatshefte für Chemie/Chemical Monthly*, 112(8-9):899–909, 1981.

[22] Y Politi, RA Metzler, M Abrecht, B Gilbert, FH Wilt, I Sagi, L Addadi, S Weiner, and PUPA Gilbert. Transformation mechanism of amorphous calcium carbonate into calcite in the sea urchin larval spicule. *Proceedings of the National Academy of Sciences*, 105(45):17362–17366, 2008.

[23] FM Michel, J MacDonald, J Feng, BL Phillips, L Ehm, C Tarabrella, JB Parise, and RJ Reeder. Structural characteristics of synthetic amorphous calcium carbonate. *Chemistry of Materials*, 20(14):4720–4728, 2008.

[24] A Becker, U Bismayer, M Epple, H Fabritius, B Hasse, J Shi, and A Ziegler. Structural characterisation of X-ray amorphous calcium carbonate (ACC) in sternal deposits of the crustacea Porcellio scaber. *Dalton Transactions*, (4):551–555, 2003.

[25] D Quigley and PM Rodger. Free energy and structure of calcium carbonate nanoparticles during early stages of crystallization. *The Journal of chemical physics*, 128(22):221101, 2008.

[26] P Raiteri and JD Gale. Water is the key to nonclassical nucleation of amorphous calcium carbonate. *Journal of the American Chemical Society*, 132(49):17623–17634, 2010.

[27] R Becker and W Döring. The kinetic treatment of nuclear formation in supersaturated vapors. *Ann. Phys*, 24(719):752, 1935.

[28] JJ De Yoreo and PG Vekilov. Principles of crystal nucleation and growth. *Reviews in mineralogy and geochemistry*, 54(1):57–93, 2003.

[29] D Gebauer, M Kellermeier, JD Gale, L Bergström, and H Cölfen. Prenucleation clusters as solute precursors in crystallisation. *Chemical Society Reviews*, 43(7):2348–2371, 2014.

[30] D Gebauer, A Völkel, and H Cölfen. Stable prenucleation calcium carbonate clusters. *Science*, 322(5909):1819–1822, 2008.

[31] D Gebauer and H Cölfen. Prenucleation clusters and non-classical nucleation. *Nano Today*, 6(6):564–584, 2011.

[32] FC Meldrum and RP Sear. Now you see them. *Science*, 322(5909):1802–1803, 2008.

[33] MH Nielsen, D Li, H Zhang, S Aloni, T Han, C Frandsen, J Seto, JF Banfield, H Cölfen, and JJ De Yoreo. Investigating processes of nanocrystal formation and transformation via liquid cell tem. *Microscopy and Microanalysis*, 20(02):425–436, 2014.

[34] D Quigley, CL Freeman, JH Harding, and PM Rodger. Sampling the structure of calcium carbonate nanoparticles with metadynamics. *The Journal of chemical physics*, 134(4):044703, 2011.

[35] H Cölfen. Biomineralization: a crystal-clear view. *Nature materials*, 9(12):960–961, 2010.

[36] JR Young, JM Didymus, PR Brown, B Prins, and S Mann. Crystal assembly and phylogenetic evolution in heterococcoliths. *Nature*, 356(6369):516–518, 1992.

[37] K Henriksen, SLS Stipp, JR Young, and ME Marsh. Biological control on calcite crystallization: AFM investigation of coccolith polysaccharide function. *American Mineralogist*, 89(11-12):1709–1716, 2004.

[38] U Riebesell, I Zondervan, B Rost, PD Tortell, RE Zeebe, and FMM Morel. Reduced calcification of marine plankton in response to increased atmospheric CO2. *Nature*, 407(6802):364–367, 2000.

[39] HA Lowenstam and S Weiner. *On biomineralization*, volume 324. Oxford University Press New York, 1989.

[40] G Falini, S Albeck, S Weiner, and L Addadi. Control of aragonite or calcite polymorphism by mollusk shell macromolecules. *Science*, 271(5245):67–69, 1996.

[41] S Weiner, L Addadi, and HD Wagner. Materials design in biology. *Materials Science and Engineering: C*, 11(1):1–8, 2000.

[42] LT Kuhn-Spearing, H Kessler, E Chateau, R Ballarini, AH Heuer, and SM Spearing. Fracture mechanisms of the strombus gigas conch shell: implications for the design of brittle laminates. *Journal of materials science*, 31(24):6583–6594, 1996.

[43] ME Kunitake, LM Mangano, JM Peloquin, SP Baker, and LA Estroff. Evaluation of strengthening mechanisms in calcite single crystals from mollusk shells. *Acta biomaterialia*, 9(2):5353–5359, 2013.

[44] JS Evans. Tuning in to mollusk shell nacre-and prismatic-associated protein terminal sequences. implications for biomineralization and the construction of high performance inorganic-organic composites. *Chemical reviews*, 108(11):4455–4462, 2008.

[45] LA Estroff, editor. *Chemical Reviews: Biomineralization*, volume 108. ACS Publications, 2008.

[46] S Albeck, J Aizenberg, L Addadi, and S Weiner. Interactions of various skeletal intracrystalline components with calcite crystals. *Journal of the American*

Chemical Society, 115(25):11691–11697, 1993.

[47] G Fu, SR Qiu, CA Orme, DE Morse, and JJ De Yoreo. Acceleration of calcite kinetics by abalone nacre proteins. *Advanced materials*, 17(22):2678–2683, 2005.

[48] C Chen, J Qi, RN Zuckermann, and JJ DeYoreo. Engineered biomimetic polymers as tunable agents for controlling CaCO3 mineralization. *Journal of the American Chemical Society*, 133(14):5214–5217, 2011.

[49] JJJM Donners, RJM Nolte, and NAJM Sommerdijk. A shape-persistent polymeric crystallization template for CaCO3. *Journal of the American Chemical Society*, 124(33):9700–9701, 2002. PMID: 12175216.

[50] LA Estroff, CD Incarvito, and AD Hamilton. Design of a synthetic foldamer that modifies the growth of calcite crystals. *Journal of the American Chemical Society*, 126(1):2–3, 2004.

[51] J Aizenberg, AJ Black, and GM Whitesides. Control of crystal nucleation by patterned self-assembled monolayers. *Nature*, 398(6727):495–498, 1999.

[52] J Aizenberg, AJ Black, and GM Whitesides. Oriented growth of calcite controlled by self-assembled monolayers of functionalized alkanethiols supported on gold and silver. *Journal of the American Chemical Society*, 121(18):4500–4509, 1999.

[53] YJ Han and J Aizenberg. Face-selective nucleation of calcite on self-assembled monolayers of alkanethiols: Effect of the parity of the alkyl chain. *Angewandte Chemie*, 115(31):3796–3798, 2003.

[54] AM Travaille, JJJM Donners, JW Gerritsen, NAJM Sommerdijk, RJM Nolte, and H van Kempen. Aligned growth of calcite crystals on a self-assembled monolayer. *Advanced Materials*, 14(7):492, 2002.

[55] AM Travaille, L Kaptijn, P Verwer, B Hulsken, JAAW Elemans, RJM Nolte, and H van Kempen. Highly oriented self-assembled monolayers as templates for epi-

taxial calcite growth. *Journal of the American Chemical Society*, 125(38):11571–11577, 2003.

[56] Q Hu, Michael H Nielsen, CL Freeman, LM Hamm, J Tao, JRI Lee, Thomas Yong-Jin Han, U Becker, JH Harding, PM Dove, et al. The thermodynamics of calcite nucleation at organic interfaces: Classical vs. non-classical pathways. *Faraday Discussions*, 159(1):509–523, 2012.

[57] AM Belcher, XH Wu, RJ Christensen, PK Hansma, GD Stucky, and DE Morse. Control of crystal phase switching and orientation by soluble mullusc-shell proteins. *Nature*, 381:56–58, 1996.

[58] F Marin and G Luquet. Molluscan shell proteins. *Comptes Rendus Palevol*, 3(6):469–492, 2004.

[59] T Samata, N Hayashi, M Kono, K Hasegawa, C Horita, and S Akera. A new matrix protein family related to the nacreous layer formation of pinctada fucata. *Febs Letters*, 462(1):225–229, 1999.

[60] RA Metzler, JS Evans, CE Killian, D Zhou, TH Churchill, NP Appathurai, SN Coppersmith, and PUPA Gilbert. Nacre protein fragment templates lamellar aragonite growth. *Journal of the American Chemical Society*, 132(18):6329–6334, 2010.

[61] EC Keene, JS Evans, and LA Estroff. Matrix interactions in biomineralization: Aragonite nucleation by an intrinsically disordered nacre polypeptide, n16N, associated with a β-chitin substrate. *Crystal Growth & Design*, 10(3):1383–1389, 2010.

[62] EC Keene, JS Evans, and LA Estroff. Silk fibroin hydrogels coupled with the n16Nβ-chitin complex: An in vitro organic matrix for controlling calcium carbonate mineralization. *Crystal Growth & Design*, 10(12):5169–5175, 2010.

[63] LZ Lakshtanov, N Bovet, and SLS Stipp. Inhibition of calcite growth by alginate. *Geochimica et Cosmochimica Acta*, 75(14):3945–3955, 2011.

[64] JW Nielsen, KK Sand, CS Pedersen, LZ Lakshtanov, JR Winther, M Willemoës, and SLS Stipp. Polysaccharide effects on calcite growth: The influence of composition and branching. *Crystal Growth & Design*, 12(10):4906–4910, 2012.

[65] S Elhadj, EA Salter, A Wierzbicki, JJ De Yoreo, N Han, and PM 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.

[66] S Elhadj, JJ De Yoreo, JR Hoyer, and PM Dove. Role of molecular charge and hydrophilicity in regulating the kinetics of crystal growth. *Proceedings of the National Academy of Sciences*, 103(51):19237–19242, 2006.

[67] JS Andrews, SA Rolfe, WE Huang, JD Scholes, and SA Banwart. Biofilm formation in environmental bacteria is influenced by different macromolecules depending on genus and species. *Environmental Microbiology*, 12(9):2496–2507, 2010.

[68] JH Harding and DM Duffy. The challenge of biominerals to simulations. *Journal of Materials Chemistry*, 16(12):1105–1112, 2006.

[69] JH Harding, DM Duffy, ML Sushko, PM Rodger, D Quigley, and JA Elliott. Computational techniques at the organic- inorganic interface in biomineralization. *Chemical reviews*, 108(11):4823–4854, 2008.

[70] S Kerisit, DJ Cooke, D Spagnoli, and SC Parker. Molecular dynamics simulations of the interactions between water and inorganic solids. *Journal of Materials Chemistry*, 15(14):1454–1462, 2005.

[71] DJ Cooke and JA Elliott. Atomistic simulations of calcite nanoparticles and their interaction with water. *Journal of Chemical Physics*, 127(10):104706–104706, 2007.

[72] CL Freeman, JH Harding, D Quigley, and PM Rodger. Structural control of crystal nuclei by an eggshell protein. *Angewandte Chemie International Edition*, 49(30):5135–5137, 2010.

[73] DM Duffy and JH Harding. Modeling the properties of self-assembled monolayers terminated by carboxylic acids. *Langmuir*, 21(9):3850–3857, 2005.

[74] D Quigley, PM Rodger, CL Freeman, JH Harding, and DM Duffy. Metadynamics simulations of calcite crystallization on self-assembled monolayers. *The Journal of chemical physics*, 131(9):094703, 2009.

[75] CL Freeman, Q Hu, MH Nielsen, J Tao, JJ De Yoreo, and JH Harding. Surface selectivity of calcite on self-assembled monolayers. *The Journal of Physical Chemistry C*, 117(10):5154–5163, 2013.

[76] S Kerisit, SC Parker, and JH Harding. Atomistic simulation of the dissociative adsorption of water on calcite surfaces. *The Journal of Physical Chemistry B*, 107(31):7676–7682, 2003.

[77] S Kerisit and SC 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.

[78] D Spagnoli, S Kerisit, and SC Parker. Atomistic simulation of the free energies of dissolution of ions from flat and stepped calcite surfaces. *Journal of crystal growth*, 294(1):103–110, 2006.

[79] P Geissbühler, P Fenter, E DiMasi, G Srajer, LB Sorensen, and NC Sturchio. Three-dimensional structure of the calcite-water interface by surface X-ray scattering. *Surface Science*, 573(2):191–203, 2004.

[80] TD Perry, RT Cygan, and R Mitchell. Molecular models of alginic acid: Interactions with calcium ions and calcite surfaces. *Geochimica et cosmochimica acta*, 70(14):3508–3532, 2006.

[81] CL Freeman and JH Harding. Entropy of molecular binding at solvated mineral surfaces. *The Journal of Physical Chemistry C*, 118(3):1506–1514, 2014.

[82] M Yang, SLS Stipp, and JH Harding. Biological control on calcite crystallization by polysaccharides. *Crystal Growth & Design*, 8(11):4066–4074, 2008.

[83] CL Freeman, I Asteriadis, M Yang, and JH. Harding. Interactions of organic molecules with calcite and magnesite surfaces. *The Journal of Physical Chemistry C*, 113(9):3666–3673, 2009.

[84] CL Freeman, JH Harding, D Quigley, and PM Rodger. Simulations of ovocleidin-17 binding to calcite surfaces and its implications for eggshell formation. *The Journal of Physical Chemistry C*, 115(16):8175–8183, 2011.

[85] L Hall-Stoodley, JW Costerton, and P Stoodley. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Reviews Microbiology*, 2(2):95–108, 2004.

[86] SW Lee, SB Park, SK Jeong, KS Lim, SH Lee, and MC Trachtenberg. On carbon dioxide storage based on biomineralization strategies. *Micron*, 41(4):273–282, 2010.

[87] B Pokroy, AN Fitch, F Marin, M Kapon, N Adir, and E Zolotoyabko. Anisotropic lattice distortions in biogenic calcite induced by intra-crystalline organic molecules. *Journal of structural biology*, 155(1):96–103, 2006.

[88] YY Kim, L Ribeiro, F Maillot, O Ward, SJ Eichhorn, and FC Meldrum. Bio-inspired synthesis and mechanical properties of calcite-polymer particle composites. *Advanced Materials*, 22(18):2082–2086, 2010.

[89] LA Estroff and AD Hamilton. At the interface of organic and inorganic chemistry: bioinspired synthesis of composite materials. *Chemistry of Materials*, 13(10):3227–3235, 2001.

[90] F Nudelman and NAJM Sommerdijk. Biomineralization as an inspiration for materials chemistry. *Angewandte Chemie International Edition*, 51(27):6582–6596, 2012.

[91] JA Elliott. Novel approaches to multiscale modelling in materials science. *International Materials Reviews*, 56(4):207–225, 2011.

[92] AR Leach. *Molecular modelling: principles and applications.* Pearson Education, 2001.

[93] E O'Reilly. *Quantum theory of solids.* CRC Press, 2002.

[94] PP Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

[95] P Raiteri, JD Gale, D Quigley, and PM Rodger. Derivation of an accurate forcefield for simulating the growth of calcium carbonate from aqueous solution: a new model for the calcite- water interface. *The Journal of Physical Chemistry C*, 114(13):5997–6010, 2010.

[96] MT Dove, B Winkler, M Leslie, MJ Harris, and EKH Salje. A new interatomic potential model for calcite: applications to lattice dynamics studies, phase transition, and isotope fractionation. *The American mineralogist*, 77(3-4):244–250, 1992.

[97] RA Jackson and GD Price. A transferable interatomic potential for calcium carbonate. *Molecular simulation*, 9(2):175–177, 1992.

[98] A Pavese, M Catti, GD Price, and RA Jackson. Interatomic potentials for CaCO3 polymorphs (calcite and aragonite), fitted to elastic and vibrational data. *Physics and chemistry of minerals*, 19(2):80–87, 1992.

[99] A Pavese, M Catti, SC Parker, and A Wall. Modelling of the thermal dependence of structural and elastic properties of calcite, CaCO3. *Physics and chemistry of minerals*, 23(2):89–93, 1996.

[100] DK Fisler, JD Gale, and RT Cygan. A shell model for the simulation of rhombohedral carbonate minerals and their point defects. *American Mineralogist*, 85(1):217–224, 2000.

[101] NH de Leeuw and SC Parker. Surface–water interactions in the dolomite problem. *Physical Chemistry Chemical Physics*, 3(15):3217–3221, 2001.

[102] F Bruneval, D Donadio, and M Parrinello. Molecular dynamics study of the solvation of calcium carbonate in water. *The Journal of Physical Chemistry B*, 111(42):12219–12227, 2007.

[103] CL Freeman, JH Harding, DJ Cooke, JA Elliott, JS Lardge, and DM Duffy. New forcefields for modeling biomineralization processes. *The Journal of Physical Chemistry C*, 111(32):11943–11951, 2007.

[104] WL Jorgensen, J Chandrasekhar, JD Madura, RW Impey, and ML Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[105] P Raiteri and JD Gale. Water is the key to nonclassical nucleation of amorphous calcium carbonate. *Journal of the American Chemical Society*, 132(49):17623–17634, 2010.

[106] HJC Berendsen, JPM Postma, WF van Gunsteren, and J Hermans. Interaction models for water in relation to protein hydration. In *Intermolecular forces*, pages 331–342. Springer, 1981.

[107] AD MacKerell, D Bashford, MLDR Bellott, RL Dunbrack, JD Evanseck, MJ Field, S Fischer, J Gao, H Guo, S Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[108] TA Soares, PH Hünenberger, MA Kastenholz, V Kräutler, T Lenz, RD Lins, C Oostenbrink, and WF van Gunsteren. An improved nucleic acid parameter set

for the GROMOS force field. *Journal of computational chemistry*, 26(7):725–737, 2005.

[109] WL Jorgensen, DS 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.

[110] H Sun. COMPASS: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. *The Journal of Physical Chemistry B*, 102(38):7338–7364, 1998.

[111] V Hornak, R Abel, A Okur, B Strockbine, A Roitberg, and C Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[112] DJ Price and CL Brooks. Modern protein force fields behave comparably in molecular dynamics simulations. *Journal of computational chemistry*, 23(11):1045–1057, 2002.

[113] B Hess and NFA 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):17616–17626, 2006.

[114] MP Allen and DJ Tildesley. Computer simulation of liquids. *New York: Oxford*, 385, 1989.

[115] K Schröder, J Sauer, M Leslie, C Richard, A Catlow, and JM Thomas. Bridging hydroxyl groups in zeolitic catalysts: a computer simulation of their structure, vibrational properties and acidity in protonated faujasites. *Chemical Physics Letters*, 188(3):320–325, 1992.

[116] L Verlet. Computer experiments on classical fluids. I Thermodynamical properties of Lennard-Jones molecules. *Physical review*, 159(1):98, 1967.

[117] RW Hockney. *The potential calculation and some applications Methods in Computational Physics.* Academic Press, New York, vol. 9 edition, December 1970.

[118] Daan Frenkel and Berend Smit. Understanding molecular simulations: from algorithms to applications. Technical report, Academic Press, 1996.

[119] WG Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[120] GN Patey and JP Valleau. A monte carlo method for obtaining the interionic potential of mean force in ionic solution. *The Journal of Chemical Physics*, 63(6):2334–2339, 1975.

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

[122] A Laio and FL Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):126601, 2008.

[123] PJ Steinhardt, DR Nelson, and M Ronchetti. Icosahedral bond orientational order in supercooled liquids. *Physical Review Letters*, 47(18):1297, 1981.

[124] RE Bulo, D Donadio, A Laio, F Molnar, J Rieger, and M Parrinello. Site binding of Ca2+ ions to polyacrylates in water: A molecular dynamics study of coiling and aggregation. *Macromolecules*, 40(9):3437–3442, 2007.

[125] A Barducci, M Bonomi, and M Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):826–843, 2011.

[126] KM Towe and GR Thompson. The structure of some bivalve shell carbonates prepared by ion-beam thinning. *Calcified tissue research*, 10(1):38–48, 1972.

[127] H Li, HL Xin, ME Kunitake, EC Keene, DA Muller, and LA Estroff. Calcite prisms from mollusk shells (Atrina Rigida): Swiss-cheese-like organic–inorganic

single-crystal composites. *Advanced Functional Materials*, 21(11):2028–2034, 2011.

[128] KS Katti and DR Katti. Why is nacre so tough and strong? *Materials Science and Engineering: C*, 26(8):1317–1324, 2006.

[129] B Kahr and RW Gurney. Dyeing crystals. *Chemical reviews*, 101(4):893–952, 2001.

[130] H Li, HL Xin, DA Muller, and LA Estroff. Visualizing the 3D internal structure of calcite single crystals grown in agarose hydrogels. *Science*, 326(5957):1244–1247, 2009.

[131] F Marin, R Amons, N Guichard, M Stigter, A Hecker, G Luquet, P Layrolle, G Alcaraz, C Riondet, and P Westbroek. Caspartin and calprismin, two proteins of the shell calcitic prisms of the Mediterranean fan mussel Pinna nobilis. *Journal of Biological Chemistry*, 280(40):33895–33908, 2005.

[132] YY Kim, K Ganesan, P Yang, AN Kulak, S Borukhin, S Pechook, L Ribeiro, R Kröger, SJ Eichhorn, SP Armes, et al. An artificial biomineral formed by incorporation of copolymer micelles in calcite crystals. *Nature materials*, 10(11):890–896, 2011.

[133] S Borukhin, L Bloch, T Radlauer, AH Hill, AN Fitch, and B Pokroy. Screening the incorporation of amino acids into an inorganic crystalline host: The case of calcite. *Advanced Functional Materials*, 22(20):4216–4224, 2012.

[134] YY Kim, B Demarchi, M Kunitake, DJ Sparks, J Carloni, CC Tang, CL Freeman, S Baker, B Pokroy, K Pinkman, JH Harding, LA Estroff, and FC Meldum. Incorporation, structure and properties of calcite crystals occluding amino acids. 2015. In progress.

[135] J Ihli, P Bots, A Kulak, LG Benning, and FC Meldrum. Elucidating mechanisms of diffusion-based calcium carbonate synthesis leads to controlled mesocrystal

formation. *Advanced Functional Materials*, 23(15):1965–1973, 2013.

[136] WD Bischoff, SK Sharma, and FT MacKenzie. Carbonate ion disorder in synthetic and biogenic magnesian calcites; a Raman spectral study. *American Mineralogist*, 70(5-6):581–589, 1985.

[137] L Martínez, R Andrade, EG Birgin, and JM Martínez. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, 30(13):2157–2164, 2009.

[138] DA Case, TA Darden, Thomas E Cheatham III, CL Simmerling, J Wang, RE Duke, R Luo, RC Walker, W Zhang, KM Merz, et al. AMBER 12. *University of California, San Francisco*, 142, 2012.

[139] A Jakalian, DB Jack, and CI Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II parameterization and validation. *Journal of computational chemistry*, 23(16):1623–1641, 2002.

[140] IT Todorov, T Forester, and W Smith. The DL POLY molecular simulation package, 2010.

[141] S Melchionna, G Ciccotti, and B Lee Holian. Hoover NPT dynamics for systems varying in shape and size. *Molecular Physics*, 78(3):533–544, 1993.

[142] PE Wright and HJ Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, 293(2):321–331, 1999.

[143] AK Dunker, E Garner, S Guilliot, P Romero, K Albrecht, J Hart, Z Obradovic, C Kissinger, and JE Villafranca. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. In *Pac Symp Biocomput*, volume 3, pages 473–484, 1998.

[144] AK Dunker, JD Lawson, CJ Brown, RM Williams, P Romero, JS Oh, CJ Oldfield, AM Campen, CM Ratliff, KW Hipps, et al. Intrinsically disordered protein.

Journal of Molecular Graphics and Modelling, 19(1):26–59, 2001.

[145] VN Uversky. Natively unfolded proteins: a point where biology waits for physics. *Protein science*, 11(4):739–756, 2002.

[146] AB Robinson and LR Robinson. Distribution of glutamine and asparagine residues and their near neighbors in peptides and proteins. *Proceedings of the National Academy of Sciences*, 88(20):8880–8884, 1991.

[147] JL Crawford, WN Lipscomb, and CG Schellman. The reverse turn as a polypeptide conformation in globular proteins. *Proceedings of the National Academy of Sciences*, 70(2):538–542, 1973.

[148] JA Smith, LG Pease, and KD Kopple. Reverse turns in peptides and protein. *Critical Reviews in Biochemistry and Molecular Biology*, 8(4):315–399, 1980.

[149] E de Alba, MA Jiménez, and M Rico. Turn residue sequence determines β-hairpin conformation in designed peptides. *Journal of the American Chemical Society*, 119(1):175–183, 1997.

[150] T Boulikas. Putative nuclear localization signals (NLS) in protein transcription factors. *Journal of cellular biochemistry*, 55(1):32–58, 1994.

[151] GE Schulz. Nucleotide binding proteins. *Molecular mechanism of biological recognition*, pages 79–94, 1979.

[152] JS Evans. Apples and oranges: comparing the structural aspects of biomineral- and ice-interaction proteins. *Current opinion in colloid & interface science*, 8(1):48–54, 2003.

[153] RA Metzler, GA Tribello, M Parrinello, and PUPA Gilbert. Asprich peptides are occluded in calcite and permanently disorder biomineral crystals. *Journal of the American Chemical Society*, 132(33):11585–11591, 2010.

[154] M Ndao, E Keene, FF Amos, G Rewari, CB Ponce, L Estroff, and JS Evans. Intrinsically disordered mollusk shell prismatic protein that modulates calcium carbonate crystal growth. *Biomacromolecules*, 11(10):2539–2544, 2010.

[155] G Xu and JS Evans. Model peptide studies of sequence repeats derived from the intracrystalline biomineralization protein, SM50. I GVGGR and GMGGQ repeats. *Biopolymers*, 49(4):303–312, 1999.

[156] B Zhang, G Xu, and JS Evans. Model peptide studies of sequence repeats derived from the intracrystalline biomineralization protein, SM50. II Pro, Asn-rich tandem repeats. *Biopolymers*, 54(6):464–475, 2000.

[157] B Zhang, BA Wustman, D Morse, and JS Evans. Model peptide studies of sequence regions in the elastomeric biomineralization protein, Lustrin. the C-domain consensus PG-,NVNCT-motif. *Biopolymers*, 63(6):358–369, 2002.

[158] BA Wustman, R Santos, B Zhang, and JS Evans. Identification of a glycine-loop-like coiled structure in the 34 AA Pro, Gly, Met repeat domain of the biomineral-associated protein, PM27. *Biopolymers*, 65(5):362–372, 2002.

[159] M Michenfelder, G Fu, C Lawrence, JC Weaver, BA Wustman, L Taranto, JS Evans, and DE Morse. Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains. *Biopolymers*, 70(4):522–533, 2003.

[160] IW Kim, DE Morse, and JS Evans. Molecular characterization of the 30-AA N-terminal mineral interaction domain of the biomineralization protein AP7. *Langmuir*, 20(26):11664–11673, 2004.

[161] IW Kim, S Collino, DE Morse, and JS Evans. A crystal modulating protein from molluscan nacre that limits the growth of calcite in vitro. *Crystal Growth & Design*, 6(5):1078–1082, 2006.

[162] FF Amos and JS Evans. AP7, a partially disordered pseudo C-RING protein, is capable of forming stabilized aragonite in vitro. *Biochemistry*, 48(6):1332–1339, 2009.

[163] S Collino, IW Kim, and JS Evans. Identification and structural characterization of an unusual ring-like sequence within an extracellular biomineralization protein, ap7. *Biochemistry*, 47(12):3745–3755, 2008.

[164] C Söllner, M Burghammer, E Busch-Nentwich, J Berger, H Schwarz, C Riekel, and T Nicolson. Control of crystal size and lattice formation by starmaker in otolith biomineralization. *Science*, 302(5643):282–286, 2003.

[165] TM Kapłon, G Rymarczyk, M Nocula-Ługowska, M Jakób, M Kochman, M Lisowski, Z Szewczuk, and A Ozyhar. Starmaker exhibits properties of an intrinsically disordered protein. *Biomacromolecules*, 9(8):2118–2125, 2008.

[166] S Collino and JS Evans. Molecular specifications of a mineral modulation sequence derived from the aragonite-promoting protein n16. *Biomacromolecules*, 9(7):1909–1918, 2008.

[167] RA Metzler, JS Evans, CE Killian, D Zhou, TH Churchill, NP Appathurai, SN Coppersmith, and PUPA Gilbert. Nacre protein fragment templates lamellar aragonite growth. *Journal of the American Chemical Society*, 132(18):6329–6334, 2010.

[168] CB Ponce and JS Evans. Polymorph crystal selection by n16, an intrinsically disordered nacre framework protein. *Crystal Growth & Design*, 11(10):4690–4696, 2011.

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

[170] I Perovic, EP Chang, M Lui, A Rao, H Cölfen, and JS Evans. A nacre protein, n16.3, self-assembles to form protein oligomers that dimensionally limit and organize mineral deposits. *Biochemistry*, 53(16):2739–2748, 2014.

[171] AH Brown, PM Rodger, JS Evans, and TR Walsh. Equilibrium conformational ensemble of the intrinsically disordered peptide n16N: Linking subdomain structures and function in nacre. *Biomacromolecules*, 15(12):4467–4479, 2014.

[172] GN Ramachandran, CT Ramakrishnan, and V Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology*, 7(1):95–99, 1963.

[173] SC Lovell, IW Davis, WB Arendall, PIW de Bakker, JM Word, MG Prisant, JS Richardson, and DC Richardson. Structure validation by cα geometry: ϕ, ψ and cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3):437–450, 2003.

[174] SL Stipp and MF Hochella. Structure and bonding environments at the calcite surface as observed with X-ray photoelectron spectroscopy (XPS) and low energy electron diffraction (LEED). *Geochimica et Cosmochimica Acta*, 55(6):1723–1736, 1991.

[175] S Brunauer, PH Emmett, and E Teller. Adsorption of gases in multimolecular layers. *Journal of the American Chemical Society*, 60(2):309–319, 1938.

[176] S Mabic, B Gerion, E Castillo, and I Kano. High purity water for inorganic analysis. http://www.merckmillipore.com, July 2007.

[177] KP Travis and DJ Searles. Effect of solvation and confinement on the trans-gauche isomerization reaction in n-butane. *The Journal of chemical physics*, 125(16):164501, 2006.

[178] SJ Brookes, DJ Searles, and KP Travis. The effect of confinement and wall structure on the kinetics of isomerisation of n-butane. *Molecular Simulation*, 35(1-2):172–185, 2009.

[179] C Braga and KP Travis. Computer simulation of the role of torsional flexibility on mass and momentum transport for a series of linear alkanes. *The Journal of chemical physics*, 137(6):064116, 2012.

[180] WB Whitman, DC Coleman, and WJ Wiebe. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, 95(12):6578–6583, 1998.

[181] ME Davey and GA O'toole. Microbial biofilms: from ecology to molecular genetics. *Microbiology and molecular biology reviews*, 64(4):847–867, 2000.

[182] M Dworkin. Recent advances in the social and developmental biology of the myxobacteria. *Microbiological reviews*, 60(1):70, 1996.

[183] P Gilbert, J Das, and I Foley. Biofilm susceptibility to antimicrobials. *Advances in dental research*, 11(1):160–167, 1997.

[184] KJ Edwards, PL Bond, TM Gihring, and JF Banfield. An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. *Science*, 287(5459):1796–1799, 2000.

[185] DM Ward, MJ Ferris, SC Nold, and MM Bateson. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiology and Molecular Biology Reviews*, 62(4):1353–1370, 1998.

[186] AK Ojha, AD Baughn, D Sambandan, T Hsu, X Trivelli, Y Guerardel, A Alahari, L Kremer, WR Jacobs, and GF Hatfull. Growth of mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Molecular Microbiology*, 69(1):164–174, 2008.

[187] JA Hyde, RO Darouiche, and JW Costerton. Strategies for prophylaxis against prosthetic valve endocarditis: a review article. *The Journal of heart valve disease*, 7(3):316–326, 1998.

[188] R Plowman, N Graves, MAS Griffin, JA Roberts, AV Swan, B Cookson, and L Taylor. The rate and cost of hospital-acquired infections occurring in patients

admitted to selected specialties of a district general hospital in england and the national burden imposed. *Journal of hospital infection*, 47(3):198–209, 2001.

[189] C Mayer, R Moritz, C Kirschner, W Borchard, R Maibaum, J Wingender, and HC Flemming. The role of intermolecular interactions: studies on model systems for bacterial biofilms. *International journal of biological macromolecules*, 26(1):3–16, 1999.

[190] B Zhu, X Xu, and R Tang. Hydration layer structures on calcite facets and their roles in selective adsorptions of biomolecules: A molecular dynamics study. *The Journal of chemical physics*, 139(23):234705, 2013.

[191] A Ghimire, RM Kasi, and CV Kumar. Proton-coupled protein binding: Controlling lysozyme and poly (acrylic acid) interactions with pH. *The Journal of Physical Chemistry B*, 118(19):5026–5033, 2014.

[192] A Laguecir, S Ulrich, J Labille, N Fatin-Rouge, S Stoll, and J Buffle. Size and pH effect on electrical and conformational behavior of poly (acrylic acid): Simulation and experiment. *European polymer journal*, 42(5):1135–1144, 2006.

[193] GA Tribello, C Liew, and M Parrinello. Binding of calcium and carbonate to polyacrylates. *The Journal of Physical Chemistry B*, 113(20):7081–7085, 2009.

[194] C Geffroy, J Persello, A Foissy, B Cabane, and F Tournilhac. The frontier between adsorption and precipitation of polyacrylic acid on calcium carbonate. *Revue de l'institut français du pétrôle*, 52(2):183–190, 1997.

[195] Q Liu, Q Wang, and L Xiang. Influence of poly acrylic acid on the dispersion of calcite nano-particles. *Applied Surface Science*, 254(21):7104–7108, 2008.

[196] U Aschauer, D Spagnoli, P Bowen, and SC Parker. Growth modification of seeded calcite using carboxylic acids: Atomistic simulations. *Journal of colloid and interface science*, 346(1):226–231, 2010.

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

[198] DJ Sparks, ME Romero-Gonzalez, E El-Taboni, CL Freeman, SA Hall, G Kakonyi, L Swanson, SA Banwart, and JH Harding. Adsorption of poly acrylic acid onto the surface of calcite: an experimental and simulation study. *Physical Chemistry Chemical Physics*, 2015. Submitted.

[199] T Azumi and SP McGlynn. Polarization of the luminescence of phenanthrene. *The Journal of Chemical Physics*, 37(10):2413–2420, 1962.

[200] I Soutar, L Swanson, RE Imhof, and G Rumbles. Synchrotron-generated time-resolved fluorescence anisotropy studies of the segmental relaxation of poly (acrylic acid) and poly (methacrylic acid) in dilute methanolic solutions. *Macromolecules*, 25(17):4399–4405, 1992.

[201] I Soutar and L Swanson. Luminescence studies of polyelectrolyte behavior in solution. III time-resolved fluorescence anisotropy measurements of the conformational behavior of poly (methacrylic acid) in dilute aqueous solutions. *Macromolecules*, 27(15):4304–4311, 1994.

[202] I Soutar, L Swanson, SJL Wallace, KP Ghiggino, DJ Haines, and TA Smith. Fluorescence studies of the behavior of poly (dimethylacrylamide) in dilute aqueous solution and at the solid-liquid interface. In *ACS Symposium Series*, volume 598, pages 363–363, 1995.

[203] JR Ebdon, BJ Hunt, DM Lucas, I Soutar, L Swanson, and AR Lane. Luminescence studies of hydrophobically modified, water-soluble polymers. I Fluorescence anisotropy and spectroscopic investigations of the conformational behaviour of copolymers of acrylic acid and styrene or methyl methacrylate. *Canadian journal of chemistry*, 73(11):1982–1994, 1995.

[204] L Ruiz-Pérez, A Pryke, M Sommer, G Battaglia, I Soutar, L Swanson, and M Geoghegan. Conformation of poly (methacrylic acid) chains in dilute aqueous solution. *Macromolecules*, 41(6):2203–2211, 2008.

[205] F Molnar and J Rieger. Like-charge attraction between anionic polyelectrolytes: Molecular dynamics simulations. *Langmuir*, 21(2):786–789, 2005.

[206] I Langmuir. The adsorption of gases on plane surfaces of glass, mica and platinum. *Journal of the American Chemical society*, 40(9):1361–1403, 1918.

[207] MM Pradas, MS Sánchez, GG Ferrer, and JLG Ribelles. Thermodynamics and statistical mechanics of multilayer adsorption. *The Journal of chemical physics*, 121(17):8524–8531, 2004.

[208] 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.

[209] FC Meldrum. Calcium carbonate in biomineralisation and biomimetic chemistry. *International Materials Reviews*, 48(3):187–224, 2003.

[210] S Weiner, J Mahamid, Y Politi, Y Ma, and L Addadi. Overview of the amorphous precursor phase strategy in biomineralization. *Frontiers of Materials Science in China*, 3(2):104–108, 2009.

[211] S Raz, PC Hamilton, FH Wilt, S Weiner, and L Addadi. The transient phase of amorphous calcium carbonate in sea urchin larval spicules: the involvement of proteins and magnesium ions in its formation and stabilization. *Advanced Functional Materials*, 13(6):480–486, 2003.

[212] C Darwin. *The formation of vegetable mould: through the action of worms, with observations on their habits.* John Murray, 1881.

[213] MG Canti and TG Piearce. Morphology and dynamics of calcium carbonate granules produced by different earthworm species: The 7th international symposium

on earthworm ecology, Cardiff, Wales, 2002. *Pedobiologia*, 47(5):511–521, 2003.

[214] MR Lee, ME Hodson, and G Langworthy. Earthworms produce granules of intricately zoned calcite. *Geology*, 36(12):943–946, 2008.

[215] L Gago-Duport, MJI Briones, JB Rodriguez, and B Covelo. Amorphous calcium carbonate biomineralization in the earthworms calciferous gland: pathways to the formation of crystalline phases. *Journal of structural biology*, 162(3):422–435, 2008.

[216] DC Lambkin, KH Gwilliam, C Layton, MG Canti, TG Piearce, and ME Hodson. Production and dissolution rates of earthworm-secreted calcium carbonate. *Pedobiologia*, 54:S119–S129, 2011.

[217] DC Lambkin, KH Gwilliam, C Layton, MG Canti, TG Piearce, and Mark E Hodson. Soil pH governs production rate of calcium carbonate secreted by the earthworm Lumbricus terrestris. *Applied Geochemistry*, 26:S64–S66, 2011.

[218] B Demarchi, LG Benning, A Brown, JH Harding, CL Freeman, K Penkman, and ME Hodson. Earthworms produce highly stable amorphous calcium carbonate. *Mineralogical Magazine*, 2013.

[219] P Raiteri, R Demichelis, JD Gale, M Kellermeier, D Gebauer, D Quigley, LB Wright, and TR Walsh. Exploring the influence of organic species on pre-and post-nucleation calcium carbonate. *Faraday Discussions*, 159(1):61–85, 2012.

[220] S Weiner and PM Dove. An overview of biomineralization processes and the problem of the vital effect. *Reviews in Mineralogy and Geochemistry*, 54(1):1–29, 2003.

[221] MJS Dewar, EG Zoebisch, EF Healy, and JJP Stewart. Development and use of quantum mechanical molecular models. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107(13):3902–3909, 1985.

[222] A Jakalian, BL Bush, DB Jack, and CI Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I method. *Journal of Computational Chemistry*, 21(2):132–146, 2000.

Appendix A

List of amino acids

Amino acid	Three letter code	One letter code	Amino acid	Three letter code	One letter code
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophane	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Table A.1: List of amino acids with three and one letter codes.

Appendix B

Partial charges from a
semi-empirical method

As discussed in Chapter 2.1 several methods are available to arrive at an approximation
of the wave function, such as the Hartree-Fock method, and methods that rely on the
electron density, such as the Density Functional Theory. Other semi-empirical methods
are available that use simplifications and parameters to get an approximation of the
wave function. The advantage of semi-empirical methods is the significant computa-
tional speed up, especially for larger systems. The disadvantage is the less accurate
approximations that can be made with these methods. One of these semi-empirical
methods is the Austin Model-1 (AM1) method [221].

Within the AM1 method, as with many semi-empirical methods, only valence electrons
are considered, with the core electrons being represented by a repulsion function. The
AM1 uses repulsive Gaussian functions to describe the core-core repulsion, and adds sev-
eral attractive Gaussian functions to compensate for excessive repulsions. The variables
of the AM1 method were parametrised using experimental results of the heat of forma-
tion, ionisation energy and the dipole moments for a range of organic molecules.

The method utilised to calculate partial charges for the organic molecules in this work

is the AM1-BCC (bond charge correction) method developed by Jakalian et al. [222] as a fast method to generate high quality partial charges. The partial charges q_i for an atom i are constructed using two terms (Equation B.1):

$$q_i = q_i^{pre} + q_i^{corr} \tag{B.1}$$

where q_i^{pre} are the atomic charges calculated using the fast AM1 method, and q_i^{corr} is a correction term to reproduce charges calculated by sophisticated, high level quantum mechanical methods (HF/6-31G* basis set). The correction term is given by Equation B.2:

$$q_i^{corr} = \sum_{\alpha=1}^{\gamma} T_{i\alpha} p_\alpha \tag{B.2}$$

where the summation is over the total number of bond types in the molecule γ, T is the connectivity matrix and p_α is the bond charge correction. The bond charge corrections p_α are parametrised using high level quantum mechanical computations.

Appendix C

Glu-Asn-Glu peptide structure

The N-terminus of peptides refers to the end of the peptide with the free amine functional group. However, the Glu residue in an N-terminus is prone to an intra molecular cyclisation between amine functional group and the carboxyl side chain, as shown in Figure C.1. This reaction is known as a pyroglutamic acid formation and the resulting structure of Glu-Asn-Gly is used throughout this work.
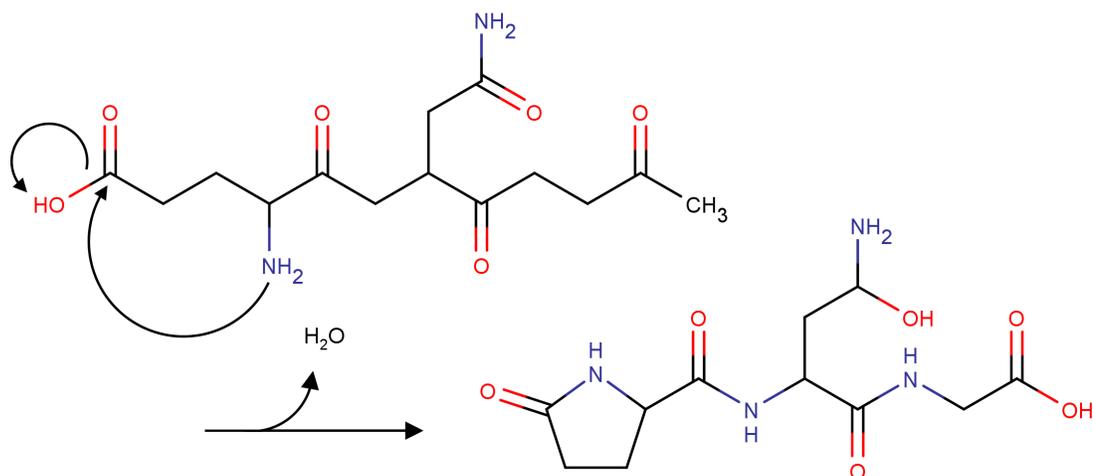
Figure C.1: The N-terminus amino acid Glu can undergo an intra molecular cyclisation reaction to form a pyroglutamic acid.

Appendix D

Hydrogen bond lifetime for ten tripeptides

Donor Peptide-1	Total life time of H-bond (ns)	Donor Peptide-2	Total life time of H-bond (ns)	Donor Peptide-3	Total life time of H-bond (ns)
N-H$_1$	0.12	N-H$_1$	0.32	N-H$_1$	0.75
N-H$_2$	1.97	N-H$_2$	0.52	N-H$_2$	0.20
N-H$_3$	0.58	N-H$_3$	0.91	N-H$_3$	0.91
N-H$_4$	0.56	N-H$_4$	4.43	N-H$_4$	2.99
N-H$_5$	0.14	N-H$_5$	0.15	N-H$_5$	0.13
Average	0.67	Average	1.27	Average	0.97

Table D.1: Hydrogen bonding between peptides during a 5.0 ns MD simulations of ten Glu-Asn-Gly peptides in solution.

Donor Peptide-4	Total life time of H-bond (ns)	Donor Peptide-5	Total life time of H-bond (ns)	Donor Peptide-6	Total life time of H-bond (ns)
N-H$_1$	0.60	N-H$_1$	0.73	N-H$_1$	0.26
N-H$_2$	0.05	N-H$_2$	0.36	N-H$_2$	0.15
N-H$_3$	0.71	N-H$_3$	2.48	N-H$_3$	0.59
N-H$_4$	2.02	N-H$_4$	4.09	N-H$_4$	1.07
N-H$_5$	0.35	N-H$_5$	0.32	N-H$_5$	0.39
Average	0.75	Average	1.60	Average	0.49

Donor Peptide-7	Total life time of H-bond (ns)	Donor Peptide-8	Total life time of H-bond (ns)	Donor Peptide-9	Total life time of H-bond (ns)
N-H$_1$	0.32	N-H$_1$	1.09	N-H$_1$	1.92
N-H$_2$	0.16	N-H$_2$	0.44	N-H$_2$	0.17
N-H$_3$	0.30	N-H$_3$	1.56	N-H$_3$	0.52
N-H$_4$	2.10	N-H$_4$	0.86	N-H$_4$	1.45
N-H$_5$	0.55	N-H$_5$	0.70	N-H$_5$	2.90
Average	0.69	Average	0.93	Average	1.39

Donor Peptide-10	Total life time of H-bond (ns)				
N-H$_1$	0.07				
N-H$_2$	0.18				
N-H$_3$	0.67				
N-H$_4$	2.71				
N-H$_5$	2.47				
Average	1.22				

Table D.2: Continued... hydrogen bonding between peptides during a 5.0 ns MD simulations of ten Glu-Asn-Gly peptides in solution.

Appendix E

Potential information for amino acid inclusions

Asp^{1-}	Charge (au)	Atom type
C1	-0.2014	backbone carbon
C1	-0.0075	sidechain carbon
C	0.8976	carboxyl carbon
C	0.9116	carboxyl carbon
O	-0.8083	carboxyl oxygen
O	-0.8333	carboxyl oxygen
N	-0.8656	amine nitrogen
H1	0.0737	backbone hydrogen
HC	0.0692	sidechain hydrogen
HN	0.4455	amino hydrogen

Asp^{2-}	Charge (au)	Atom type
C1	0.0785	backbone carbon
C1	-0.1559	sidechain carbon
C	0.8954	carboxyl carbon
C	0.9115	carboxyl carbon
O	-0.8929	carboxyl oxygen
O	-0.8812	carboxyl oxygen
N	-0.9059	amine nitrogen
H1	-0.0100	backbone hydrogen
HC	0.0355	sidechain hydrogen
HN	0.3317	amino hydrogen

Gly^0	Charge (au)	Atom type
C1	-0.0952	backbone carbon
C	0.9326	carboxyl carbon
O	-0.7553	carboxyl oxygen
N	-0.8356	amine nitrogen
H1	0.0887	backbone hydrogen
HN	0.4438	amino hydrogen

Gly^{1-}	Charge (au)	Atom type
C1	0.0118	backbone carbon
C	0.8806	carboxyl carbon
O	-0.8528	carboxyl oxygen
N	-0.8948	amine nitrogen
H1	0.0287	backbone hydrogen
HN	0.3258	amino hydrogen

Lennard-Jones	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
C1 C1	4367168.3069	2828.6533
C1 C	3872045.8800	2507.9582
C1 O	2712383.7593	2623.9516
C1 N	4167877.6150	3085.2839
C1 H1	284187.9963	444.1229
C1 HC	406834.7544	531.3850
C1 HN	10746.6001	86.3646
C C	3433057.3544	2223.6214
C O	2404870.5299	2326.4642
C N	3695349.5264	2735.4936
C H1	251968.5255	393.7710
C HC	360710.3555	471.1399
C HN	9528.2173	76.5731
O O	1590466.5073	2365.0646
O N	2540673.0890	2835.3883
O H1	154690.3401	385.6847
O HC	227871.2130	468.1074
O HN	4295.4573	64.2695
N N	3953566.9079	3354.9813
N H1	259860.4396	474.1637
N HC	375462.5897	569.9563

Lennard-Jones	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
N HN	8901.1862	87.7570
H1 H1	13647.6962	59.9032
H1 HC	20874.8341	74.0853
H1 HN	248.9753	8.0909
HC HC	31468.3113	90.9615
HC HN	448.7983	10.8629
HN HN	0.5860	0.3925
OC C1	55486.4674	183.3127
OC C	49195.7562	162.5298
OC O	26516.9115	149.1753
OC N	48654.7479	191.6599
OC H1	1966.0414	21.2421
OC HC	3217.3484	27.1727
CC O	2404870.529954	2326.464284
CC N3	3695349.526463	2735.493631
CC H1	251968.525521	393.771044
CC HC	360710.355574	471.139955
CC HN	9528.217324	76.573140

Buckingham	A (kJ mol^{-1})	ρ (Å)
CA O	104684.0984	0.2970
CA N	628351.1060	0.2530
OC HN	32641.4600	0.2170

Table E.1: Potential parameters for the interactions between the amino acids Asp and Gly and calcite.

Appendix F

Potential information for Glu-Asn-Gly peptide

Atom	Charge (au)	Atom type
O	-0.4025	carboxyl oxygen
O	-0.5077	carboxyl oxygen
O	-0.4516	carboxyl oxygen
O	-0.4622	carboxyl oxygen
O	-0.4424	carboxyl oxygen
OH	-0.4119	carbonyl oxygen
HO	0.3783	carbonyl hydrogen
N	-0.1993	amide nitrogen
N	-0.1352	amide nitrogen
N	-0.0260	amide nitrogen
N	-0.5808	amine nitrogen
C	0.0622	carboxyl carbon
C	0.2716	carboxyl carbon
C	-0.0302	carboxyl carbon
C	0.5332	carboxyl carbon
C	0.4300	carboxyl carbon
C1	-0.0537	backbone carbon
C1	-0.3799	backbone carbon
C1	-0.2201	backbone carbon
C1	-0.0167	backbone carbon
C1	-0.4121	backbone carbon
C1	-0.3037	backbone carbon

Atom	Charge (au)	Atom type
HN	0.3094	amide hydrogen
HN	0.3047	amide hydrogen
HN	0.3067	amide hydrogen
HN	0.3065	amine hydrogen
H1	0.1491	backbone hydrogen
H1	0.1930	backbone hydrogen
H1	0.2024	backbone hydrogen
HC	0.1826	sidechain hydrogen
HC	0.1722	sidechain hydrogen
HC	0.1834	sidechain hydrogen

Lennard-Jones	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
O O	1589402.8514	2363.4829
O OH	1970677.7509	2632.9941
O N	2538973.9689	2833.4921
O C1	2710569.8043	2622.1968
O C	2403262.2274	2324.9084
O H1	154586.8879	385.4267
O HC	227718.8198	467.7944
O HN	4292.5846	64.2266
OH OH	2434264.7108	2927.7406
OH N	3116979.0157	3140.9891
OH C1	3311820.7510	2899.8464
OH C	2936347.0744	2571.0799
OH H1	195360.3799	433.4913
OH HC	285677.9265	524.2042
OH HN	5877.1402	75.1874
N N	3950922.8850	3352.7376
N C1	4165090.2712	3083.2206
N C	3692878.1936	2733.6642
N H1	259686.6533	473.8466
N HC	375211.4922	569.5751
N HN	8895.2334	87.6983
C1 C1	4364247.6852	2826.7616
C1 C	3869456.3764	2506.2809
C1 H1	283997.9402	443.8259
C1 HC	406562.6762	531.0297
C1 HN	10739.4131	86.3068

Lennard-Jones	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
C C	3430761.4344	2222.1343
C H1	251800.0168	393.5077
C HC	360469.1239	470.8248
C HN	9521.8451	76.5219
H1 H1	13638.5691	59.8632
H1 HC	20860.8737	74.0358
H1 HN	248.8087	8.0855
HC HC	31447.2663	90.9006
HC HN	448.4982	10.8556
HN HN	0.5856	0.3922
O OW	1978737.9948	2432.8169
OH OW	2437216.7646	2701.2760
N OW	3106071.7699	2891.2019
C1 OW	3288207.0262	2664.3689
C OW	2915410.5271	2362.2993
H1 OW	199036.8403	403.4615
HC OW	289429.5515	486.5268
HN OW	6362.5509	72.1358
O OC	1982919.0241	2850.8057
OH OC	2455794.4937	3174.0117
N OC	3158085.5283	3412.4014
C1 OC	3366671.8764	3155.5780
C OC	2984979.5197	2797.8182
H1 OC	193985.5277	466.3015
HC OC	285116.0456	565.2947
HN OC	5522.7808	78.7189

Table F.1: Potential parameters for the interactions between the Glu-Asn-Gly peptide and calcite.

Appendix G

Potential information for poly acrylic acid

Atom	Charge (au)	Atom type
C1	-0.0737	terminal carbon
C1	-0.0409	CH2 carbon
C1	-0.1477	backbone carbon
CC	0.9150	carboxyl carbon
OC	-0.9130	carbonyl oxygen
H1	0.0171	Hydrogen of CT
H1	0.0262	Hydrogen of CH2
H1	0.0528	Hydrogen of CHCOO-

Lennard-Jones	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
C1 C1	4367168.3069	2828.6533
C1 CC	3872045.8800	2507.9582
C1 O	2712383.7593	2623.9516
C1 H1	406834.7544	531.3850
CC CC	3433057.3544	2223.6214
CC OC	2404870.5299	2326.4642
CC H1	360710.3555	471.1399
OC OC	1590466.5073	2365.0646
OC H1	227871.2130	468.1074
H1 H1	31468.3113	90.9615
C1 OW	3290574.2908	2664.06319
CC OW	2917509.4066	2362.0282

Lennard-Jones	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
OC OW	1979952.4366	2432.6108
H1 OW	289587.4783	486.5014
C1 O	55486.4674	183.3127
CC O	49195.7562	162.5298
OC O	26516.9115	149.1753
H1 O	3217.3484	27.1727

Buckingham	A (kJ mol^{-1})	ρ (Å)
CA OC	104684.0984	0.2970
OC OW	48241.6996	0.2300

Table G.1: Potential parameters for the interactions between poly acrylic acid and calcite.

Appendix H

Potential information for amorphous calcium carbonate

Asp	Charge (au)	Atom type
C1	0.0785	backbone carbon
C1	-0.1559	sidechain carbon
C	0.8954	carboxyl carbon
C	0.9115	carboxyl carbon
O	-0.8929	carboxyl oxygen
O	-0.8812	carboxyl oxygen
N	-0.9059	amine nitrogen
H1	-0.0100	backbone hydrogen
HC	0.0355	sidechain hydrogen
HN	0.3317	amino hydrogen

Gly	Charge (au)	Atom type
C1	-0.0952	backbone carbon
C	0.9326	carboxyl carbon
O	-0.7553	carboxyl oxygen
N	-0.8356	amine nitrogen
H1	0.0887	backbone hydrogen
HN	0.4438	amino hydrogen

Glu	Charge (au)	Atom type
C1	-0.0694	backbone carbon
C1	-0.1684	sidechain carbon
C1	0.0035	sidechain carbon
C	0.8876	carboxyl carbon
C	0.8956	carboxyl carbon
O	-0.6168	carboxyl oxygen
O	-0.8323	carboxyl oxygen
N	-1.1236	amine nitrogen
H1	0.0387	backbone hydrogen
HC	0.0642	sidechain hydrogen
HC	0.0627	sidechain hydrogen
HN	0.3931	amino hydrogen

Arg	Charge (au)	Atom type
C1	-0.1084	backbone carbon
C1	-0.0844	sidechain carbon
C1	-0.0285	sidechain carbon
C1	0.0163	sidechain carbon
C	0.5383	amide carbon
C	0.9106	carboxyl carbon
O	-0.7993	carboxyl oxygen
N	-0.8296	amide nitrogen
N	-0.3791	guanidine nitrogen
N	-0.4922	guanidine nitrogen
N	-0.4292	guanidine nitrogen
H1	0.0857	backbone hydrogen
HC	0.0657	sidechain hydrogen
HC	0.0737	sidechain hydrogen
HN	0.4548	amide hydrogen
HN	0.2987	guanidine hydrogen
HN	0.4818	guanidine hydrogen
HN	0.3117	guanidine hydrogen

Lennard-Jones	A (kJ mol^{-1} Å^{-12})	B (kJ mol^{-1} Å^{-6})
C1 C1	4367168.3069	2828.6533
C1 C	3872045.8800	2507.9582
C1 O	2712383.7593	2623.9516
C1 N	4167877.6150	3085.2839
C1 H1	284187.9963	444.1229
C1 HC	406834.7544	531.3850
C1 HN	10746.6001	86.3646
C C	3433057.3544	2223.6214
C O	2404870.5299	2326.4642
C N	3695349.5264	2735.4936
C H1	251968.5255	393.7710
C HC	360710.3555	471.1399
C HN	9528.2173	76.5731
O O	1590466.5073	2365.0646
O N	2540673.0890	2835.3883
O H1	154690.3401	385.6847
O HC	227871.2130	468.1074
O HN	4295.4573	64.2695
N N	3953566.9079	3354.9813
N H1	259860.4396	474.1637
N HC	375462.5897	569.9563
N HN	8901.1862	87.7570
H1 H1	13647.6962	59.9032
H1 HC	20874.8341	74.0853
H1 HN	248.9753	8.0909
HC HC	31468.3113	90.9615
HC HN	448.7983	10.8629
HN HN	0.5860	0.3925
OC C1	55486.4674	183.3127
OC C	49195.7562	162.5298
OC O	26516.9115	149.1753
OC N	48654.7479	191.6599
OC H1	1966.0414	21.2421
OC HC	3217.3484	27.1727
CC O	2404870.529954	2326.464284
CC N3	3695349.526463	2735.493631
CC H1	251968.525521	393.771044
CC HC	360710.355574	471.139955
CC HN	9528.217324	76.573140

Buckingham	A (kJ mol^{-1})	ρ (Å)
CA O	104684.0984	0.2970
CA N	628351.1060	0.2530
OC HN	32641.4600	0.2170

Table H.1: Potential parameters for the interactions between the amino acids Asp, Gly, Glu and Arg and amorphous calcium carbonate.

Fin.