COMPUTATIONAL INSIGHT INTO THE MOLECULAR MECHANISMS THAT CONTROL THE GROWTH OF INORGANIC CRYSTALS

ROBERT DAVID WARWICK DARKINS

A dissertation submitted for the degree of

Doctor of Engineering (EngD)

Department of Physics and Astronomy

University College London

September 2016

DECLARATION

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

R. D. W. Darkins
September 2016

ABSTRACT

After billions of years of evolution, nature has developed mechanisms for controlling the growth and assembly of materials right down to the nanoscale, an achievement that materials scientists hope to mimic. However, the underlying processes are extremely complex and depend on subtle behaviour at the molecular scale. In contrast to experimental methods, computer simulations can achieve the molecular resolution needed to investigate these mechanisms, and can therefore offer unique insight. Indeed, this dissertation employs a variety of state-of-the-art computational methodologies to investigate the molecular processes by which calcite, the most abundant biomineral on earth, grows, in addition to the role played by surfactants in soft templating technologically important inorganic materials.

Microsecond-long simulations are performed to reveal the behaviour of individual ions in the vicinity of calcite steps, providing new insight into the mechanisms responsible for kink nucleation.

Rare event methodologies are then used to study the dissolution process of kink sites in calcite crystals. It is discovered that this particular mineralisation process is too complex to be tamed by computational methods, which has far-reaching consequences for the development of highly predictive models of mineralisation.

A coarse-grained model for calcite precipitation is presented that displays the ability to connect molecular processes with both the kinetic and morphological characters of a crystal. However, the simulation is found to conflict with experimental observations regarding the dependence of step velocity on step length. The implication being that present models are unable to correctly describe step pinning, which is a major limitation.

Lastly, the role of surfactants in templating crystal growth via two very different mechanisms is investigated. In the one case, polymorph and orientation selection by self-assembled monolayers; and in the other, oriented heterogeneous nucleation of mesoporous organosilicas.

ACKNOWLEDGMENTS

I am extremely grateful to my primary supervisor, Dorothy Duffy, for her support and for allowing me free rein to pursue even my most hare-brained ideas.

I am also grateful for the collaborations with my secondary supervisor, Maria Sushko, who guided the soft templating research.

I wish to thank Alex Côté for his friendship and fruitful collaborations, not to mention the meticulous testing of my codes.

My work has benefited from input, at various stages, from Yongsoon Shin, Jim De Yoreo, David Quigley, Julian Gale, and the various members of the MIB consortium, notably Colin Freeman and Mark Rodger.

Thanks also to Paolo Raiteri for providing me with the LAMMPS implementation of his force fields. And to Fab Sidoli and Andy Gormanly for keeping the supercomputers alive.

Last but not least, I wish to thank my family for putting up with me. Oh, and for the lifts.

I acknowledge funding from EPSRC under the Molecular Modelling and Materials Science Industrial Doctorate Centre and the US Department of Energy (DOE), Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under Award KC020105-FWP12152.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

1.1 THE MATERIAL WORLD

Millions of years ago, the early members of the homo genus began to fashion tools out of materials such as flint, wood, and bone. This period of prehistory, spanning up to approximately ten thousand years ago, is known as the Stone Age. The Copper Age followed with the discovery that copper is produced, quite miraculousy, when malachite is added to hot embers. Copper, being a malleable metal, was well-suited for a wide range of applications. For example, it was crafted by the Egyptians into sickles, weapons, and the chisels used in constructing the pyramids, thus helping to give rise to the earliest civilisations. However, copper is soft, and so the Copper Age was quickly superseded by the Bronze Age when it was discovered that copper could be hardened by alloying it with other elements, such as tin and arsenic, to form bronze. The need for these rarer elements gave rise to new trade routes, and the spread of culture and technology throughout the world. In a similar vein, progress in smelting iron ores, and alloying the resulting iron with carbon, resulted in steel, the strongest pre-industrial material, and initiated the Iron Age.

While the Iron Age was the terminus of prehistory, our ensuing dependence on materials has not diminished. In fact, it is only within the past century, with the advent of quantum mechanics, that we have begun to really understand the inner workings of materials and achieve an entirely new level of material control, actualising the microcomputer revolution, among various other advances.

The next revolution in materials science will be in the development of nanomaterials: materials with features tailored to a scale below 100 nanometres in size. Progress has already been made in this direction, for example, with the synthesis of graphene, quantum dots, and nanowires, seeing early applications ranging from solar cells to medicine. However, we are a long way from achieving full control of the size, shape, phase, and orientation of materials to an arbitrary scale, along with the assembly of such components into larger, more complex systems. But we know that this ambition is achievable, for nature has already accomplished it.

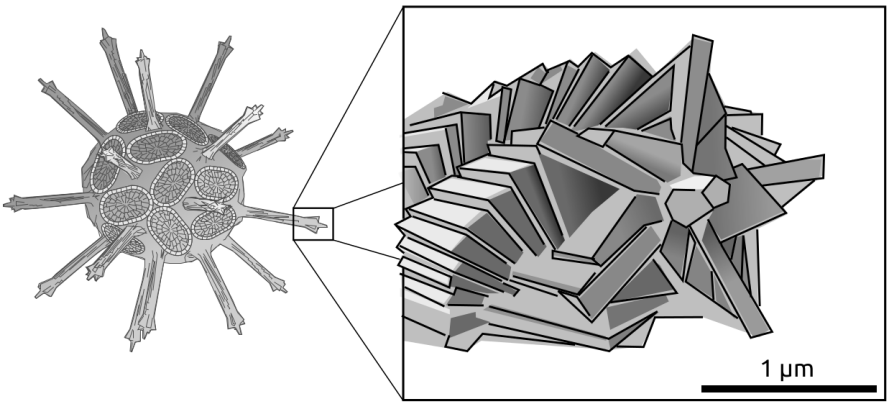


Figure 1.1: An illustration of Rhabdosphaera clavigera, a species of coccolithophore, which provides a beautiful example of biomineralisation. Original SEM images can be found in [1].

1.2 BIOMINERALISATION

Biological systems display an extraordinary level of control over materials right down to the nanoscale, with the most tantalising examples coming in the form of biominerals—biogenic inorganic crystals—which compose the hard structures in living organisms, such as skeletons and shells. As an example, Figure 1.1 depicts Rhabdosphaera clavigera, a species of coccolithophore, which is coated in plates of calcium carbonate with protruding spines. To quote Young et al. [1],

> "The spine structure is remarkably elegant, consisting of five interwound spirals of crystals. The component crystals of the spine show regular rhombohedral faces with perfectly regular alignment spiraling up the structure."

While the detailed mechanisms responsible for realising such structures are poorly understood, many of the broad control principles are well-established, and can be categorised as: spatial, chemical, morphological, structural, and constructional [9].

Spatial control typically involves growing the crystal within a confined space, such as a vesicle or an organic framework. Spatial confinement provides a few advantages: it limits the size and shape of the precipitating crystal, and it affords careful regulation of the concentration and composition of the internal solution, e.g. through the use of ion pumps in the membrane. Indeed, chemistry provides another means of control by adjusting the supersaturation, or by introducing impurities to modify the solubility of the crystal, as will be modelled in Chapter 5.

Crystal morphology can be controlled in various ways. One way is to grow the crystal in a volume with the desired shape, so that the boundary functions as a mould. Another method is to introduce impurities that inhibit growth in specific directions. A third mechanism, known as vectorial regulation, is to promote growth in only a localised section of the crystal, e.g. by increasing the supersaturation via a single ion pump, thus inducing growth in that one specific region.

Structural control generally refers to the preferential alignment of the growing crystal with respect to the organic template. An example of this will be seen in Chapter 6. And finally, there are the higher-order processes involved in assembling various crystals and organic components into a single structure, such as bone. However, we shall not address these mechanisms here.

The most abundant biomineral on earth is calcium carbonate ($CaCO_3$), for which calcite is the thermodynamically stable polymorph under ambient conditions. Calcite is therefore one of the most widely studied minerals, and we devote three chapters to investigating and modelling its growth and dissolution. Accordingly, we now provide a brief summary of the calcite crystal and its mode of growth.

1.3 CALCITE

Calcite ($CaCO_3$) has a hexagonal lattice with the space group $R\bar{3}c$. Its equilibrium morphology in pure solution under ambient conditions is the rhombohedron, as shown in Figure 1.2(b), with the $\{10\bar{1}4\}$ faces exposed. Calcite is usually grown from solution containing calcium Ca^{2+} and carbonate CO_3^{2-} ions, and it is useful to quantify the solution concentration with the so-called supersaturation

$$\sigma = \log \frac{a_{Ca^{2+}} a_{CO_3^{2-}}}{K_{sp}} \tag{1.1}$$

where $a_{Ca^{2+}}$ is the calcium activity (a measure of concentration), and similarly for $a_{CO_3^{2-}}$, and K_{sp} is the solubility product which equals $10^{-8.48}$ for calcite [10]. The solubility product is defined to equal the product $a_{Ca^{2+}} a_{CO_3^{2-}}$ when the solution is at equilibrium, and so, when $\sigma < 0$ the solution is undersaturated and the crystal dissolves; $\sigma = 0$ denotes saturation; and $\sigma > 0$ means the solution is supersaturated and the crystal will grow. Generally, growth involves the formation of two-dimensional islands on the facets that then grow to encompass the entire crystal face, increasing the size of the crystal by one

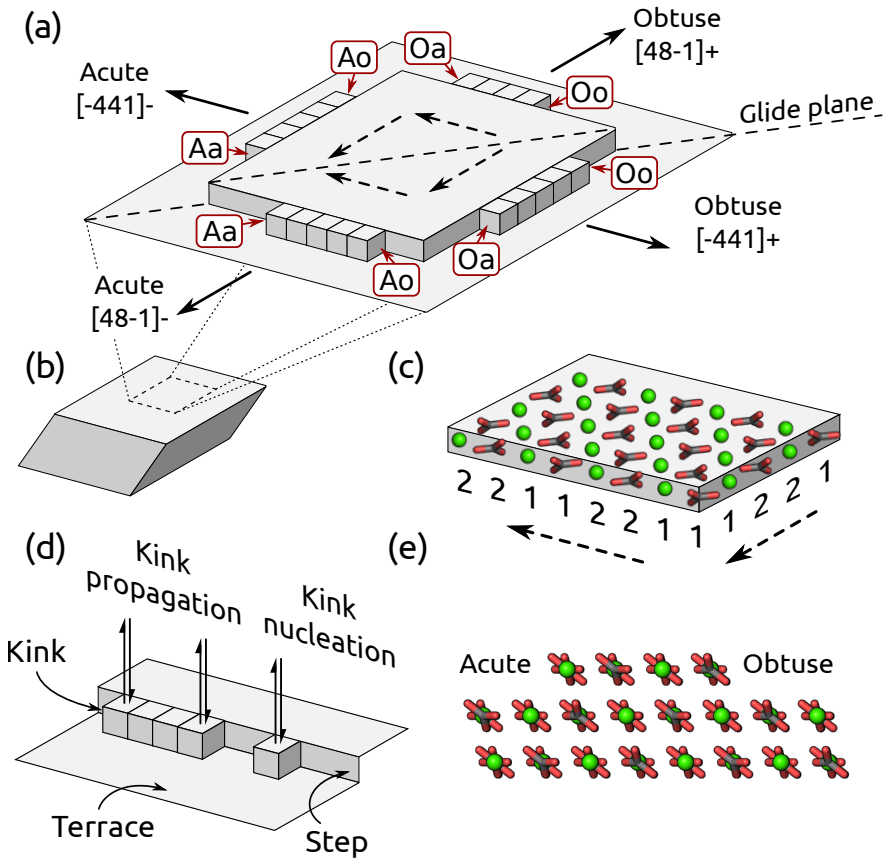


Figure 1.2: (a) A schematic representation of a 2D growth island on a calcite facet, showing the symmetry and kink labelling scheme. (b) The equilibrium rhombohedral morphology of calcite expresses the family of $\{10\bar{1}4\}$ faces. (c) The molecular structure of a growth island and how it relates to the step geometry. Each ion along a step is labelled with either 1 or 2 as described in the text. (d) A representation of the terrace-step-kink model of growth, described in the text. (e) The profile of the acute and obtuse steps.

layer at a time. However, the precise growth mechanism for calcite depends on the supersaturation state [10]. At high supersaturations, $\sigma > 1$, the dominant growth mechanism is two-dimensional surface nucleation, whereas at low supersaturations growth can only proceed from preexisting steps. These steps are continuously sourced from screw dislocations which are defects in the lattice that persist with growth [10, 11].

Step propagation in calcite is known [12] to be well-described by the terrace-step-kink model [13] which characterises the crystal surface in terms of features known as terraces, steps, and kinks, each of which is illustrated in Figure 1.2(d). Specifically, the large, featureless

facets are terraces, the one-dimensional rows that terminate a growth island are steps, and the sites with half the number of bonds as in the bulk crystal are kinks. Growth proceeeds by only two key processes: the nucleation of a pair of kink sites at a step, followed by the propagation of the kink sites along the step length due to the adsorption of new ions to the kink sites. It should be noted that these events also occur in reverse, with the continuous dissolution of kinks and kink nuclei from the steps.

Calcite naturally expresses two distinct step geometries: obtuse and acute, which are differentiated by the angle they make with the terrace, as illustrated in Figure 1.2(e). The obtuse steps grow in the directions $[\bar{4}41]+$ and $[48\bar{1}]+$, with the acute steps growing in the opposite directions, as depicted by the solid arrows in Figure 1.2(a). The low-energy steps consist of alternating units of calcium and carbonate ions, as shown in Figure 1.2(c). Note that the unit cell of calcite contains two $CaCO_3$ units and that the two carbonate ions have different orientations to each other. Consquently, there are two distinct carbonate ions exposed at the steps that differ by a rotation about the c-axis. Each carbonate ion may therefore be labelled with either 1 or 2 to denote its orientation, as depicted in Figure 1.2(c). The calcium ions may also be distinguished from each other based on their local carbonate configuration. Specifically, we introduce the simple rule that each calcium ion inherits the same index (1 or 2) from its neighbouring carbonate ion in accordance with the direction of the dashed arrows in Figure 1.2(a). Each ion may therefore be referred to as either Ca(1), Ca(2), CO_3(1), and CO_3(2).

The dashed line in Figure 1.2(a) represents glide plane symmetry, meaning that the crystal structures either side of the glide plane are mirror images of each other (within translation parallel to the plane). In the case of calcite, the plane is normal to the c-axis of the conventional cell, i.e. [001], and is therefore called the c-glide plane.

As can be seen in Figure 1.2(a), each step has two distinct kink geometries: for the acute step, we label them Aa and Ao, and for the obtuse step, Oa and Oo. Moreover, each of these kink geometries could be terminated by one of the four distinct ions listed above. These permutations sum to a total of 32 unique kink sites, or 16 when the glide plane symmetry is considered. To identify each, we introduce a nomenclature that combines the calcite geometry (Aa, Ao, etc) with the terminal ion (Ca(1), CO_3(1), etc). For example, we label the type (2) carbonate-terminated Ao kink site as Ao/CO_3(2). This notation will be used extensively in Chapter 4.

It is already well-established that the calcium and carbonate ions behave differently to each other at the active sites of calcite [4], and it will be shown in this thesis that the behaviour of the ions also depend on the precise geometry of each active site (e.g. which type of kink site) as well as on whether it is a $CO_3(1)$ or $CO_3(2)$ ion, and so on. When the atomic units of crystal growth are not identical to one another, the crystal is said to be non-Kossel [14].

1.4 BIOMIMETIC MATERIALS CHEMISTRY: SOFT TEMPLATING

Chemists have already begun to take inspiration from nature in devising strategies for synthesising inorganic materials, see Mann et al. [15] for a review. One such approach involves the use of organic matter to control the reaction field of a solution by serving as a template. This so-called soft templating approach can operate at two different levels: either at the molecular scale, bringing about heterogeneous nucleation, or at the supramolecular scale by serving as a meso-scale template. An example of the former mechanism observed in nature is the eggshell protein ovocleidin-17 which induces amorphous calcium carbonate to transition to calcite [16], while an example of a synthetic equivalent is the precipitation of oriented calcite crystals on functionalised self-assembled monolayers [17]. And an example of the latter mechanism in nature is the use of vesicles in templating the porous structure of the Radiolarian silica micro-skeleton [18], while a synthetic equivalent may be the use of surfactants to template high porosity in mesoporous organosilicas [19].

Soft templates are extremely versatile due to the range of two- and three-dimensional structures and nanopatterns that they exhibit. They are easily functionalised and, due to their flexibility, don't suffer from the high symmetry that inorganic templating through epitaxy does. Furthermore, once a crystalline phase has formed, soft templates are easily purged, e.g. through calcination.

A typical choice for a synthetic soft template is an organic amphiphile that aggregates non-covalently and self-assembles into supramolecular structures when added to a solvent. When such structures have long-range order they are called *liquid crystals* and the constituent amphiphiles are said to be *lyotropic*. Surfactants are common choices, notably sodium dodecyl sulphate (SDS), $CH_3(CH_2)_{11}[SO_4]^-Na^+$, and alkytrimethylammonium halides such as cetyltrimethylammonium chloride (CTAC),

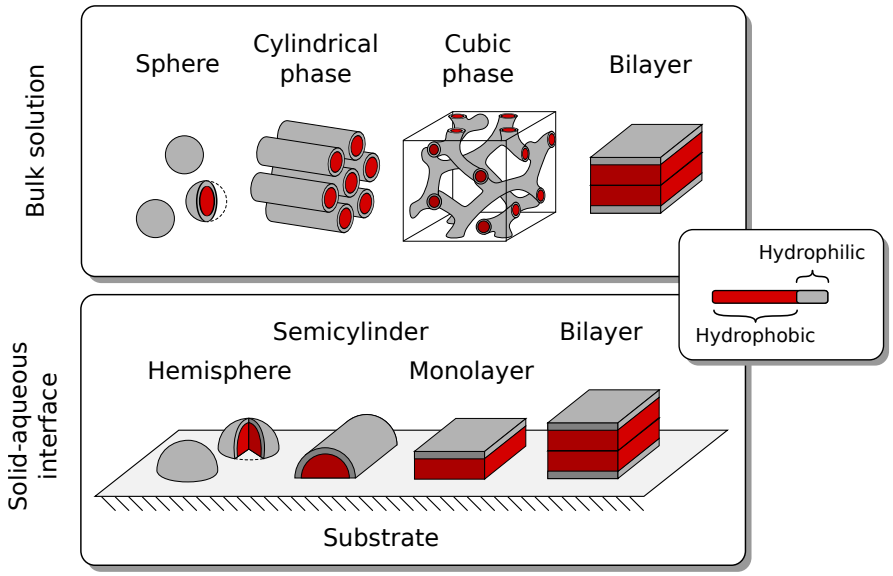


Figure 1.3: Common structures formed by lyotropic amphiphiles in bulk solution and at solid-aqueous interfaces.

$CH_3(CH_2)_{15}[N(CH_3)_3]^+Cl^-$. In these two examples, the sodium and chloride are counter-ions that bind ionically to the functional groups of the surfactants and serve to balance charge.

Amphiphile aggregation is usually driven by both a decrease in enthalpy and an increase in entropy (due to the release of water), with the resulting supramolecular structure depending on both the temperature and the concentration of the solution. As the concentration is increased above a critical value, known as the critical micelle concentration (CMC), a solution of lyotropic amphiphiles becomes saturated with monomers and spherical micelles form. Increasing the concentration further, the spherical micelles coalesce to form cylindrical micelles which may then pack together into a hexagonal phase (space group P6mm), as illustrated in Figure 1.3. Further still, some systems, including CTAC, transition to three-dimensionally periodic labyrinths with a cubic space group. These so-called cubic phases have mathematically well-defined structures that correspond to local minimal surfaces (e.g. the gyroid), meaning that any small variation in their structure will increase the surface area and thus the free energy. Lastly, increasing the concentration beyond the domain of cubic phases gives rise to bilayers.

Similar structures may form when amphiphiles aggregate at solid-aqueous interfaces, although the structure is now dependent not only on the amphiphile-amphiphile and amphiphile-solvent interactions,

but also on the amphiphile-substrate and solvent-substrate interactions (e.g. how hydrophobic the surface is). For example, under similar conditions, quaternary ammonium surfactants form semicylindrical micelles on graphite, hemispherical micelles on amorphous silica, and full cylinders on mica [20]. Figure 1.3 provides a summary of common structures that form both in bulk solution and at an interface, including hemispherical and semicylindrical micelles which are truncated so as to maximise coverage of the substrate. When soft templates are mixed with an inorganic precursor, the precursor may crystallise around the template, giving rise, for instance, to highly porous single crystals.

1.5 COMPUTATIONAL MATERIALS MODELLING

At present, computer simulations provide invaluable insight into processes that are difficult, if not impossible, to resolve experimentally. And this thesis will provide examples of that. However, one of the ultimate ambitions of computational science is to supplant, rather than merely augment, lab-based experimentation with computer simulations. Such *in silico* experiments promise many advantages over real-world experimentation. In particular, computer simulations offer an unparalleled resolution, potentially yielding a complete description of a system at any length- and time-scale required. *In silico* experiments could be setup, executed, and analysed automatically, only ever requiring access to relatively cheap and ubiquitous computer hardware, and not relying on scarce and expensive equipment for analysis. The types of materials available for investigation would no longer be limited by affordability or our ability to synthesise them. And as scientific experimentation becomes more computer-based, it opens up scientific progress more readily to the growing powers of artificial intelligence [21].

In the context of crystal growth, the subject of this thesis, the long-term ambition is to model the growth of crystals from a solution with a known composition; capturing all of the subtle effects that pH or the activity of certain impurities may have on the characteristics of crystal growth, such as morphology, kinetics, and composition. With such a detailed model, bespoke crystals could be designed on the computer, whether that be through an automated trial-and-error approach, e.g. with the aid of a genetic algorithm [22], or by working

backwards from the detailed mechanistic insight provided by such a model.

Calcite

In the case of calcite, work by Gray et al. [23, 24] has shown promise in the use of *de novo* peptides for bespoke control of calcite morphologies, and so there is great interest in developing a model of calcite growth that can accommodate the complex binding of biomolecules, such as peptides, to various surface sites and capture the growth consequences. However, there has been very little progress hitherto in developing such functionality.

Analytical models have been developed that can capture the impact that solution stoichiometry [2, 4] and impurity activities [3] have on calcite kinetics. These models, however, are extremely limited in their utility, being unable to describe morphologies. Instead, a model that includes molecular degrees of freedom is required.

In principle, standard molecular simulation could model the growth of calcite crystals from a solution. However, calcite growth from solution at typical supersaturations is an incredibly slow process, taking on the order of seconds for a step to advance only one nanometre. And so it cannot be modelled with the resolution of molecular simulation for which even the microsecond time-scale is seldom accessible.

Efforts in the literature to circumvent this problem have included simulating the growth of calcite directly from the melt, using either an elevated temperature or metadynamics to accelerate the growth. This approach has been used to identify the polymorph and orientation selection of self-assembled monolayer substrates [25, 26]; to measure the relative velocities of different calcite faces [27]; and identify the polymorph affinity of magnesium ions [28]. This approach, however, is obviously unsuited to modelling growth from solution which proceeds via ion-by-ion deposition.

Another approach to modelling calcite crystallisation is illustrated in the seminal work of Quigley et al. [29, 16, 30]. Here, metadynamics simulations were employed to generate the phase equilibria of calcium carbonate nanoparticles of different sizes, finding the amorphous phase to be thermodynamically stable below a size of 2 nm [30]. However, when in contact with ovocleidin-17, an eggshell protein, calcite became stable [16], revealing the role that this protein likely plays in eggshell formation. This approach, however, is restricted to the

nanometre length-scale and it cannot produce kinetically-determined morphologies, meaning that it is also unsuited for our purposes.

The most promising approach to modelling micrometre-scale crystallisation is to use a coarse-grained description of the crystal, typically representing the presence of a molecule at each site of a lattice with a simple 'yes' or a 'no', and then stochastically integrate the configuration over time with the kinetic Monte Carlo (KMC) algorithm. This technique was employed by Piana et al. [31, 32, 33] to simulate the growth of urea crystals and reproduce the experimental highly-faceted and elongated equilibrium morphology of urea. They did not proceed to investigate growth inhibition, however.

KMC algorithms have also been applied to calcite in the past [34, 35, 36, 37]. However, all of these studies were concerned with the dissolution process, such as etch pit formation, rather than growth. The one exception to this is a KMC simulation of calcite growth performed by De Yoreo et al. [38] which claimed to simulate the kinetic effect of kink-blocking on calcite (this mechanism is elaborate on in Section 5.3). However, this study had a number of limitations. Firstly, the requisite transition rates (discussed further below) were essentially guessed; the step was only one-dimensional and so the morphological impact was not investigated; and their simulated impurity had an infinite residence time, meaning that it was actually a step-pinner rather than a kink-blocker. Our research in Chapter 5 will therefore be the biggest step towards a coarse-grained model of the growth and inhibition of calcite crystals.

The big challenge in applying KMC to crystal growth is identifying all of the relevant atomistic processes and then obtaining the rates at which they occur. For example, at what rate do kink sites dissolve? And at what rate do the ions adsorb to the different step geometries? Moreover, at what rates do the impurities of interest attach to and dissolve from each active site of the crystal surface? Measuring these transition rates is currently beyond experimental capabilities, and computing them for rare processes is a notoriously difficult task. To date, virtually all investigations into the molecular processes at calcite surfaces have been limited to computing enthalpies, and in some cases free energies, of adsorption or incorporation [39, 40]. For example, Yang et al. [41] found that galactose binds more strongly to the acute step than the obtuse step; Freeman et al. [42] found ovocleidin-17 prefers the obtuse step; Cooke et al. [43] found ethanol to exhibit no step preference; and Elhadj et al. [6] found that the step preferred by polyaspartate depended on its chain length. Kinetic and morpho-

logical effects, when discussed, are then inferred qualitatively in a handwavy manner.

Nevertheless, with increasing computing power, more accurate force fields, and advancing methodologies, computing the rates of rare and complex events has become more accessible. Of particular significance to this thesis, Stack et al. [44] recently computed the adsorption and dissolution rates of barium ions to and from barite steps (in other words, kink nucleation). Repeating this calculation at different temperatures, they were able to obtain the activation energy for kink nucleation which was in excellent agreement with experimental measurements. The success of this study motivated us to apply a similar methodology to kink dissolution in calcite, which is the subject of Chapter 4. To my knowledge, the only other study that has applied rare event methods to kink dissolution is that of Liu et al. [45] where the dissolution pathway and associated free energy surface were generated for the kink sites of sodium chloride using *ab initio* molecular dynamics. However, their resulting free energy surface was much simpler than the one we report for calcite in Chapter 4, and they stopped short of computing rates.

In some fortunate instances, the transitions of interest are sufficiently quick that they can be sampled with direct molecular simulation. In other words, a very long trajectory without any form of biasing, from which the rates can be directly computed. This was the route taken by Piana et al. [33], discussed above, in their KMC simulation of urea. Urea grows quickly enough that the rates of the requisite molecular processes can be plucked directly from simulation. A similar approach was used recently by De La Pierre et al. [46] in obtaining the residence times of water molecules at the calcite steps. In the same vein, we analyse microsecond-long trajectories in Chapter 3 to obtain the transition rates of ions at the two calcite steps.

Soft templating

One of the main bottlenecks in understanding the mechanisms of soft templated crystal growth is the complex interplay between the template structure, chemistry, stability, and its interactions with the growing crystal. For example, it has been shown both experimentally [47] and theoretically [48] that the structure of organic templates is not retained during crystal growth. Instead, soft templates adjust to accommodate the growth of certain crystallographic planes that are not necessarily commensurate with the original template structure.

Such dynamical behaviour makes computer simulations particularly useful for studying soft templating.

The self-organisation of soft matter, such as surfactants, has been studied with molecular simulation for decades, particularly at liquid/liquid and liquid/air interfaces which are of great industrial importance [49]. And with advances in coarse-graining and computing power, the self-assembly of large biologically-significant systems has become accessible [50, 51].

The adsorption of surfactants at solid/liquid interfaces has also been widely investigated with molecular dynamics and, more recently, DFT and reactive MD. Most such studies have been concerned with mineral/aqueous interfaces, such as quartz [52], mica [53], silica [54, 55] and graphite [56, 57, 58, 59], due to their geochemical significance. These studies are generally concerned with structures, adsorption energies, surface tensions, frictional properties, and the impact of AFM tips. However, there have also been investigations into the interaction of surfactants with technologically important materials. Notably, the self-assembly of SDS on carbon nanotubes, where the resulting morphology was found to depend on the nanotube diameters [60]; SDS self-assembly on graphene, inspecting the interactions between the resultant graphene-surfactant assemblies [61]; and the self-assembly of alkanethiols on gold nanoparticles [62].

In the first soft templating study of this thesis, Section 6.1, the role that SDS surfactants may play in selecting polymorphs and orientations of titania is investigated by computing the adhesion energies of SDS to several titania faces. This approach was inspired by the work of Duffy et al. [63, 48] who investigated the selective traits of alkanethiols on calcite crystals in the same way. To our knowledge, there has been only one other simulation study of SDS surfactants at the titania/aqueous interface [64] which investigated morphologies at low coverages on only a few faces, and did not quantify the adhesions.

The second soft templating study, Section 6.2, is concerned with the morphology formed by CTAC surfactants when they self-assemble at silicon/aqueous interfaces, establishing that the reconstruction of different faces can dramatically alter the micelle morphology. To my knowledge, this is the first time that the orientation of a non-polar substrate has been found to alter the morphology of an adsorbed micelle. As described in Section 6.2, the motivation for this study was to rationalise experimental observations relating to the growth of periodic mesoporous organosilicas (PMOs) on silicon substrates. While we only infer qualitatively the significance of our findings to

PMO growth, there has been some recent progress in simulating PMO growth that is noteworthy.

Early modelling efforts involving lattice Monte Carlo were able to reproduce the formation of the periodic hexagonal PMO phase, MCM-41 [65, 66]. However, this method has greatly limited resolution and kinetic insight, and it excludes many molecular interactions that limit its utility. Instead, molecular simulations with explicit force fields were required for further progress, as has been pioneered by Jorge et al. [67, 68, 69, 67, 70, 68]. Starting with simulations of CTAB self-assembly in bulk solution [69, 67], the presence of siliceous species were found to engender micelle aggregation, producing rod-like rather than spherical micelles [70], which hinted at a so-called cooperative templating mechanism whereby the inorganic precursor plays a role in assembling its own template. This has further been supported by recent large-scale simulations, involving coarse-grained potentials and an explicit solvent, which produced the complete formation of the MCM-41 phase from a random starting solution [68].

1.6 THESIS OUTLINE

To summarise this chapter: future advances in materials science depend on our ability to control crystal growth at the nanoscale, and biomimicry is a promising way towards this goal. Achieving a detailed understanding of biomineralisation, however, will require an understanding of processes at a length- and time-scale that are inaccessible to experimentation but within the reach of computer simulations. This thesis reports on the application of various state-of-the-art computational methodologies to investigate two distinct processes of great significance in biomineralisation: the growth of calcite crystals, and the mechanisms of soft templating.

Chapter 2 provides an overview of the various computational and theoretical methods employed in this work. Starting with quantum mechanics and building up to molecular simulation, rare event methodologies, and kinetic Monte Carlo. The remaining chapters are results.

Chapter 3 investigates the behaviour of calcium ions, carbonate ions, and calcium carbonate ion pairs in the vicinity of the acute and obtuse calcite steps with an interest in establishing the kink nucleation mechanisms.

Chapter 4 employs a variety of rare event methodologies to compute the dissolution rate of calcite kink sites and provide new insight

into the dissolution mechanism. A major limitation to the rare event sampling of such complex transition pathways is established, with important consequences for future modelling efforts.

Chapter 5 presents a coarse-grained model for the growth and inhibition of calcite crystals, with parameters optimised by combining the results of Chapters 3 and 4 with experimental data. The impact of impurity mechanisms on growth kinetics and step morphology are then investigated, providing a major step towards predictive modelling of calcite precipitation.

Chapter 6 presents the study of two different instances of soft templating. The first, SDS surfactants self-assembling on titania, provides insight into the role that the surfactants may play in polymorph and orientation selection. The second study reveals the supramolecular morphology formed by CTAC surfactants on different orientations of a silicon substrate, from which we elucidate recent experimental findings.

Lastly, Chapter 7 provides a summary of the thesis and suggests future work.

2

THEORY AND METHODOLOGY

Starting with the equations of quantum mechanics, this chapter derives an arsenal of methodologies that are indispensable to molecular modelling and which form the basis for the work that follows.

Section 2.1 introduces the Schrödinger equation and then applies the Born-Oppenheimer approximation to separate the electronic and nuclear degrees of freedom, giving rise to a pair of coupled nuclear and electronic Schrödinger equations.

Section 2.2 shows how the electronic Schrödinger equation can be solved using the method of density functional theory (DFT) which reformulates the problem in terms of a 3-dimensional density functional. This formalism makes practical the application of quantum mechanics to hundreds of atoms.

In Section 2.3, the nuclear Schrödinger equation is shown to reduce to Newton's second law of motion in the classical limit $\hbar \to 0$, meaning that nuclei may be modelled approximately using classical mechanics. The electronic charges are then incorporated into the nuclei, and all residual electronic effects are approximated with a series of simple functional forms. This classical description of atoms, known as molecular mechanics, enables us to model tens of thousands of atoms.

Working within the framework of molecular (rather than quantum) mechanics, Section 2.4 introduces the key concepts of statistical mechanic, such as entropy and free energy, which bridge the theoretical divide between individual atoms and large collections of atoms.

Section 2.5 introduces molecular dynamics (MD) which integrates Newton's equations over time to generate molecular trajectories, making time-scales of $\sim 1\,\mu s$ accessible. These trajectories are then coupled with a thermostat and barostat to emulate an environment that enforces a constant temperature and pressure.

The problem of quasi-nonergodicity in molecular dynamics simulations is introduced in Section 2.6, and a variety of rare event methodologies (REMs) are presented to address it, providing both a thermodynamic and kinetic treatment of activated processes.

Lastly, Section 2.7 presents the kinetic Monte Carlo (KMC) algorithm for simulating the average behaviour of a large number of acti-

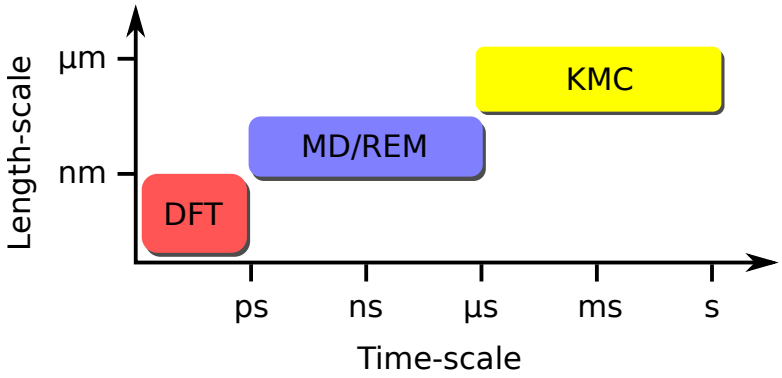


Figure 2.1: A list of the methods employed in this dissertation and the length- and time-scales to which they are applied.

vated processes. This method will be applied in Chapter 5 to model millions of atoms and reach time-scales of minutes.

These methods, and the length- and time-scales to which they are applied in this dissertation, are summarised in Figure 2.1. It can be seen that as the methods become coarser in their description of nature, larger length-scales and durations become accessible.

2.1 FIRST PRINCIPLES

The time-dependent Schrödinger equation provides a natural starting point for this dissertation

$$i\hbar\frac{\partial}{\partial t}\Psi(\{\mathbf{r}_i\},\{\mathbf{R}_i\};t) = \hat{\mathcal{H}}\Psi(\{\mathbf{r}_i\},\{\mathbf{R}_i\};t) \tag{2.1}$$

where $\hbar$ is the reduced Planck constant, i the imaginary unit, Ψ the wave function, and the standard non-relativistic Hamiltonian

$$\begin{aligned}\hat{\mathcal{H}} = &-\sum_i \frac{\hbar^2}{2M_i}\nabla^2_{\mathbf{R}_i} - \sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} \\ &+ \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} - \frac{1}{4\pi\epsilon_0}\sum_{i,j}\frac{Z_i}{|\mathbf{R}_i-\mathbf{r}_j|} \\ &+ \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{Z_iZ_j}{|\mathbf{R}_i-\mathbf{R}_j|}\end{aligned} \tag{2.2}$$

for the electronic $\{\mathbf{r}_i\}$ and nuclear $\{\mathbf{R}_i\}$ coordinates, and the nuclear masses $\{M_i\}$ and atomic numbers $\{Z_i\}$, where we assume $e = 1$. m_e is the electron mass and ϵ_0 the vacuum permittivity. The first two terms in Eq. (2.2) represent the kinetic energy of the nuclei and electrons, respectively. These are followed by the electrostatic electron-electron re-

pulsion, the electron-nucleus attraction, and lastly the nucleus-nucleus repulsion.

Solving Eq. (2.1) to obtain the many-body wave function Ψ is too large a task for virtually all systems of interest, even with the aid of modern supercomputers. Instead, a series of approximations are needed to make quantum mechanics practicable. An approximation that the remainder of this dissertation is implicitly built upon is that of Born and Oppenheimer [71] which assumes that the electronic and nuclear degrees of freedom can be separated, viz.

$$\Psi(\{\mathbf{r}_i\},\{\mathbf{R}_i\};t) = \Phi(\{\mathbf{r}_i\};\{\mathbf{R}_i\})\chi(\{\mathbf{R}_i\};t) \tag{2.3}$$

The rationale behind this separation is that electrons and nuclei both experience forces, and therefore changes in momenta, that are of the same order of magnitude. But since nuclei are orders of magnitude heavier than electrons, their velocities must be proportionally smaller. It is, therefore, reasonable to assume that the electrons relax instantaneously to their ground-state within a static array of nuclei, and so the electronic wave function must satisfy the time-independent Schrödinger equation

$$\begin{aligned}\Bigg(-\sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} \\ - \frac{1}{4\pi\epsilon_0}\sum_{i,j}\frac{Z_i}{|\mathbf{R}_i-\mathbf{r}_j|}\Bigg)\Phi(\{\mathbf{r}_i\};\{\mathbf{R}_i\}) = \mathcal{E}_e(\{\mathbf{R}_i\})\Phi(\{\mathbf{r}_i\};\{\mathbf{R}_i\})\end{aligned} \tag{2.4}$$

It can be shown [72] that, if Eq. (2.4) holds, then Eq. (2.3) is an approximate solution to Eq. (2.1) if the nuclear wave function satisfies

$$\begin{aligned}\Bigg(-\sum_i \frac{\hbar^2}{2M_i}\nabla^2_{\mathbf{R}_i} + \mathcal{E}_e(\{\mathbf{R}_i\}) \\ + \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{Z_iZ_j}{|\mathbf{R}_i-\mathbf{R}_j|}\Bigg)\chi(\{\mathbf{R}_i\};t) = i\hbar\frac{\partial}{\partial t}\chi(\{\mathbf{R}_i\};t)\end{aligned} \tag{2.5}$$

In summary, the Schrödinger equation of Eq. (2.1) can be reduced to a pair of coupled Schrödinger equations expressed in terms of electronic and nuclear wave functions of greatly reduced dimension. The next section will present a method for computing the ground-state properties of the system described by Eq. (2.4) without having to solve the many-electron wave function Φ. And in Section 2.3, Eq. (2.5) will be used to derive a classical description of nuclei in the limit $\hbar \to 0$, upon which the methods of statistical and molecular mechanics will be built.

2.2 DENSITY FUNCTIONAL THEORY

The time-independent Schrödinger equation of Eq. (2.4) describes the possible states that N electrons may occupy in the presence of a static array of nuclei $\{\mathbf{R}_i\}$. However, solving the many-electron wave function Φ is a highly formidable task due to the complexity of the many-body problem and the large $(3N)$ dimension of the wave function. Fortunately, the ground-state properties of the system can be found without knowing Φ by reformulating the problem in terms of a 3-dimensional density functional. This section presents only a brief account of density functional theory (DFT) due to the very brief application of it in this work. The reader is referred to [73, 74, 75] for a more thorough introduction.

2.2.1 *The density functional*

We begin by defining the electronic Hamiltonian

$$\hat{\mathcal{F}} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}_i}^2 + \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{2.6}$$

and the external potential

$$\hat{V} = -\frac{1}{4\pi\epsilon_0}\sum_{i,j}\frac{Z_i}{|\mathbf{R}_i - \mathbf{r}_j|} \tag{2.7}$$

such that the Hamiltonian of Eq. (2.4) equals $\hat{\mathcal{F}} + \hat{V}$. We also define the ground-state single-electron density in terms of the ground-state wave function Φ_0

$$n_0(\mathbf{r}) = \int \overline{\Phi}_0(\{\mathbf{r}_i\};\{\mathbf{R}_i\})\left[\sum_i \delta(\mathbf{r}-\mathbf{r}_i)\right]\Phi_0(\{\mathbf{r}_i\};\{\mathbf{R}_i\})\,\Pi_j\,d\mathbf{r}_j \tag{2.8}$$

$$= N\int |\Phi_0(\mathbf{r},\{\mathbf{r}_i\}_{i>1};\{\mathbf{R}_i\})|^2\,\Pi_{j>1}\,d\mathbf{r}_j \tag{2.9}$$

where the wave function is assumed to be normalised.

Hohenberg and Kohn [76] proved that the external potential $\hat{V}$ is uniquely determined, to within an additive constant, by the ground-state density n_0. Note also that the electronic Hamiltonian $\hat{\mathcal{F}}$ is the same for all systems of $N \equiv \int n_0(\mathbf{r})\,d\mathbf{r}$ electrons. It follows that the Hamiltonian, and therefore all properties of the system (including the excited states), are uniquely determined by the ground-state density n_0. This remarkable result is the bedrock of density functional theory.

2.2.2 *Kohn-Sham formulation*

In order to compute the ground-state density n_0, Kohn and Sham [77] made the *ansatz* that any interacting system has the same ground-state density as some auxiliary non-interacting system, specifically

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 - \sum_i \frac{Z_i}{|\mathbf{r}-\mathbf{R}_i|} + \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + V_{xc}^{\sigma}[n](\mathbf{r})\right)\phi_i^{\sigma}(\mathbf{r}) = \epsilon_i \phi_i^{\sigma}(\mathbf{r}) \tag{2.10}$$

where σ denotes the spin state (up or down). The functional $V_{xc}[n](\mathbf{r})$ is the so-called exchange-correlation potential. It is chosen to reproduce the non-classical electron-electron interactions that the auxiliary system would otherwise lack, and will be described further in the next section. The electrons fill the N lowest energy eigenfunctions, and the density is computed from

$$n(\mathbf{r}) = \sum_{\sigma}\sum_{i} |\phi_i^{\sigma}(\mathbf{r})|^2 \tag{2.11}$$

The ground-state density may then be obtained iteratively:

1. Guess the density $n(r)$
2. Diagonalise [78] the Hamiltonian in Eq. (2.10) to obtain the eigenfunctions ϕ_i^{σ}
3. Construct a new density $n(r)$ from Eq. (2.11)
4. Repeat from step 2 until self-consistency is achieved. The converged density will be that of the ground-state.

2.2.3 *Exchange-correlation potential*

The exchange-correlation potential is typically the main source of error in DFT calculations since its precise functional form is unknown and must therefore be approximated. The simplest form that is widely applied is the local density approximation (LDA) which assumes that the exchange-correlation energy at some point $\mathbf{r}$ is the same value it would have if the whole of space were filled with a homogeneous electron gas (jellium) with a density $n(\mathbf{r})$, i.e.

$$V_{xc}^{\sigma}[n](\mathbf{r}) = \epsilon_{xc}(n(\mathbf{r})) + n(\mathbf{r})\frac{\partial \epsilon_{xc}(n(\mathbf{r}))}{\partial n(\mathbf{r})} \tag{2.12}$$

where $\epsilon_{xc}(n(\mathbf{r}))$ is the exchange-correlation energy of a single electron in jellium of density $n(\mathbf{r})$, usually obtained from quantum Monte Carlo simulations.

The LDA is surprisingly accurate at predicting structural properties but it usually fails at predicting binding energies and modelling systems with inhomogeneous electron densities, such as surfaces. It can be improved upon, however, by incorporating non-local corrections that involve the density gradient ∇n, giving rise to a class of functionals called generalised-gradient approximations (GGA) [73].

2.2.4 *Periodicity, the plane wave basis, and k-point sampling*

Many systems of interest, such as bulk crystals and surfaces, are effectively infinite in extent but consist of a periodic repetition of some finite supercell. By exploiting this periodicity, it is possible to model an infinite system with a finite calculation. Mathematically, such systems must satisfy the Born-von Karman boundary condition

$$|\Psi(\mathbf{r}+\mathbf{T})|^2 = |\Psi(\mathbf{r})|^2 \tag{2.13}$$

for any lattice translation vector $\mathbf{T}$.

Through the application of Bloch's theorem and the Fourier series, a wave function ϕ in a periodic potential may be written in terms of a plane wave basis set

$$\phi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{2.14}$$

for any symmetry point $\mathbf{k}$ within the first Brillouin zone, and where the summation is over the reciprocal lattice vectors $\mathbf{G}$.

At this stage, the summation in Eq. (2.14) consists of an infinite number of contributions. Furthermore, the wave function must be integrated over an infinite number of k-points. In order to reduce the calculation to a finite size, we make two observations. First, each plane wave contributes $\frac{\hbar^2}{2m_e}|\mathbf{k}+\mathbf{G}|^2$ to the kinetic energy of the system, and so all of the plane waves that contribute energies greater than a sufficiently large cut-off energy may be omitted from the summation. Second, wave functions vary smoothly over reciprocal space, and so the integration over $\mathbf{k}$ may be approximated with a finite sum over a sufficiently dense or well-chosen set of k-points. In the case of large supercells, which have small Brillouin zones, it is often adequate to sample only the Γ-point, $\mathbf{k} = 0$.

The above periodic formulation may also be applied to aperiodic systems by imposing a false periodicity. For example, a single molecule in a sufficiently large periodic supercell will be almost indistinguishable from the same molecule without periodicity. One advantage for imposing an artificial periodicity and employing a plane wave basis is that many operations, such as evaluating the Hamiltonian, scale better in reciprocal space than in real space.

2.2.5 *Pseudopotentials*

In the Kohn-Sham approach to DFT, the electronic orbitals are orthogonal to each other. However, since the core electrons in an atom are localised near to the nucleus, the valence states must oscillate rapidly in the core region in order to achieve orthogonality. This increases the width of the Fourier spectrum which consequently requires a larger number of plane waves in the basis set. A way of overcoming this problem, as well as reducing the number of orbitals that need to be computed, is to replace the core electrons with an effective potential, known as a pseudopotential, that emulates their effect on the valence states.

2.3 MOLECULAR MECHANICS

2.3.1 *From quantum to classical mechanics*

Upon taking the classical limit $\hbar \to 0$, the Schödinger equation can be shown [79] to reduce to Newton's second law of motion

$$\frac{d\mathbf{P}_i}{dt} = -\nabla_{\mathbf{R}_i} U(\{\mathbf{R}_i\}) \tag{2.15}$$

where $\mathbf{P}_i$ are the conjugate momenta of $\mathbf{R}_i$, and U is the potential energy component of the Hamiltonian,

$$U(\{\mathbf{R}_i\}) = \mathcal{E}_e(\{\mathbf{R}_i\}) + \frac{1}{4\pi\epsilon_0} \sum_{i<j} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} \tag{2.16}$$

where $\mathcal{E}_e$ is the potential energy of the electronic subsystem as defined in Eq. (2.4).

It follows that nuclei can be approximated as classical point particles moving in a force field described by a potential energy U.

Molecular mechanics makes one further simplification: it incorporates the electronic charges into the nuclei so that the entire atom

is treated as a single classical point moving on the potential energy surface

$$U(\{\mathbf{R}_i\}) = \mathcal{E}'_e(\{\mathbf{R}_i\}) + \frac{1}{4\pi\epsilon_0}\sum_{i<j}\frac{q_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|} \tag{2.17}$$

where $\{q_i\}$ are the partial atomic charges and $\mathcal{E}'_e$ captures all of the residual electronic interactions, such as bonding, Pauli repulsion, and van der Waals attraction.

2.3.2 *Force fields*

It is possible to evaluate the electronic component $\mathcal{E}_e$ (or $\mathcal{E}'_e$) of the potential energy U using quantum mechanical methods. Indeed, this is the approach of *ab initio* molecular modelling [79]. However, we adopt the much faster classical molecular mechanics approach in which this electronic component is approximated with a series of empirically-optimised functionals

$$\begin{aligned}\mathcal{E}'_e(\{\mathbf{R}_i\}) = &\sum_{i,j} u_2^{ij}(\mathbf{R}_i, \mathbf{R}_j) \\ &+ \sum_{i,j,k} u_3^{ijk}(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k) \\ &+ \sum_{i,j,k,l} u_4^{ijkl}(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k, \mathbf{R}_l)\end{aligned} \tag{2.18}$$

where u_2^{ij}, u_3^{ijk}, and u_4^{ijkl} describe two-, three-, and four-body interactions, respectively, and the subscripted indices are used to identify the type of interaction (if any at all) between the various permutations of atoms. More complex functional forms than those in Eq. (2.18) are also in common use, such as Tersoff potentials for modelling covalent bonding, and the various metal potentials, which employ parameters unique to each atom that depend on the atom's local environment. Such potentials, however, are not used in this dissertation.

All of the atomic interactions can be classified as either intra- or inter-molecular. In the former case, the electronic structure has the effect of limiting various conformational properties of a molecule, such as bond lengths,

$$R_{ij} \equiv |\mathbf{R}_{ij}| = |\mathbf{R}_i - \mathbf{R}_j| \tag{2.19}$$

bond angles,

$$\theta_{ijk} = \arccos\left(\frac{\mathbf{R}_{ij}\cdot\mathbf{R}_{kj}}{R_{ij}R_{kj}}\right) \tag{2.20}$$

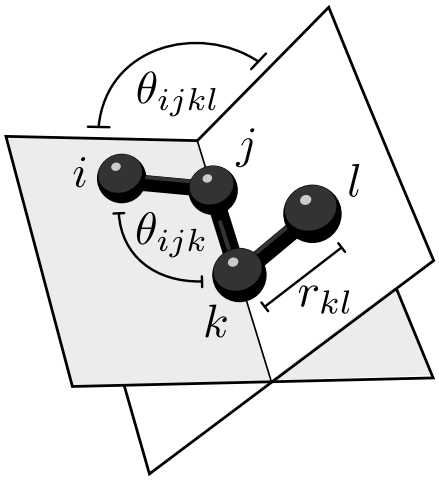


Figure 2.2: Electronic structure has the effect of restraining the bond lengths, angles, and dihedral angles in molecular conformations.

and dihedral angles,

$$\theta_{ijkl} = \arccos\left(\frac{(\mathbf{R}_{ij} \times \mathbf{R}_{jk}) \cdot (\mathbf{R}_{jk} \times \mathbf{R}_{kl})}{|\mathbf{R}_{ij} \times \mathbf{R}_{jk}||\mathbf{R}_{jk} \times \mathbf{R}_{kl}|}\right) \tag{2.21}$$

all of which are depicted in Fig. 2.2. The contributions in Eq. (2.18) are therefore chosen to add energy penalties when these conformational properties vary from the desired equilibrium values.

The non-electrostatic inter-molecular forces, on the other hand, produce just two phenomena: the Pauli repulsion between atoms when their electronic shells overlap, and van der Waals forces. These are invariably short-ranged, meaning that they decay more rapidly than r^{-3} in the limit $r \to \infty$. The significance of this is that such interactions only need to be evaluated for atoms that fall within some cut-off distance r_c of each other since the contribution to the total potential energy from beyond r_c may be approximated with a well-defined expression that is proportional to $\int_{r_c}^{\infty} u(r) 4\pi r^2 \, dr$.

Once a suitable set of functional forms have been identified, the various implicit parameters are obtained by optimising them to reproduce a set of data obtained either experimentally or through quantum mechanical (e.g. DFT) calculations, such as lattice parameters, elasticity constants, vibrational spectra, and enthalpies.

2.3.3 *Supercells and long-range electrostatics*

In Section 2.2.4 it was explained that many systems, such as bulk crystals and surfaces, can be treated as an infinitely periodic (or pseudo-

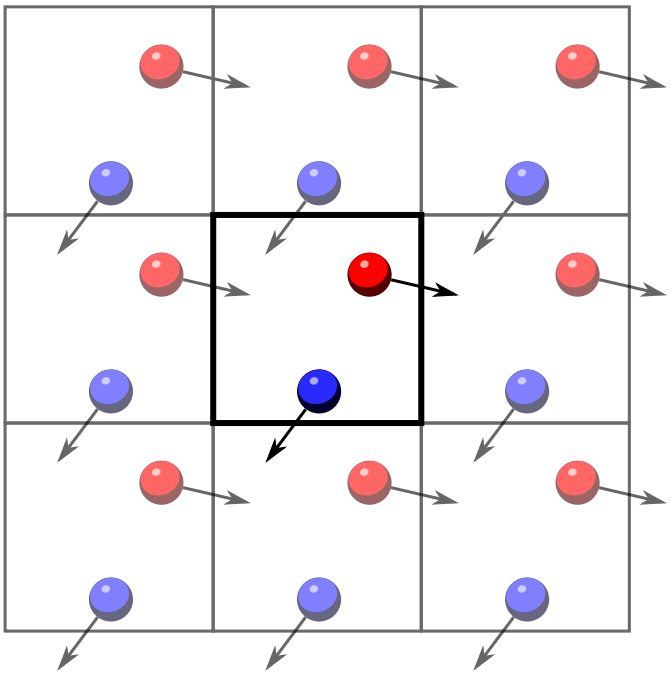


Figure 2.3: A two-dimensional illustration of periodic boundary conditions. The supercell (middle) is repeated infinitely in all directions and each atom interacts with all atoms within the supercell plus their periodic images. If an atom exits the supercell it will re-emerge through the opposite face.

periodic) repetition of a finite supercell. Here we elaborate on this idea in the context of molecular mechanics.

In a periodic system, the infinite number of atoms can be approximated as a finite motif of atoms, contained within a supercell of known dimension, that is repeated infinitely in all directions. If a moving atom exits the supercell then it will instantly reappear through the opposite face, as illustrated in Fig. 2.3.

The total potential energy of the supercell, cf. Eq. (2.17) and Eq. (2.18), now takes the form

$$\begin{aligned} U(\{\mathbf{R}_i\}) = &\sum_{\mathbf{A}} \sum_{i,j} u_2^{ij}(\mathbf{R}_i, \mathbf{R}_j + \mathbf{A}) \\ &+ \sum_{\mathbf{A},\mathbf{B}} \sum_{i,j,k} u_3^{ijk}(\mathbf{R}_i, \mathbf{R}_j + \mathbf{A}, \mathbf{R}_k + \mathbf{B}) \\ &+ \sum_{\mathbf{A},\mathbf{B},\mathbf{C}} \sum_{i,j,k,l} u_4^{ijkl}(\mathbf{R}_i, \mathbf{R}_j + \mathbf{A}, \mathbf{R}_k + \mathbf{B}, \mathbf{R}_l + \mathbf{C}) \\ &+ \frac{1}{4\pi\epsilon_0} \sum_{\mathbf{A}} {\sum_{i,j}}' \frac{q_i q_j}{|\mathbf{R}_i - \mathbf{R}_j - \mathbf{A}|} \end{aligned} \tag{2.22}$$

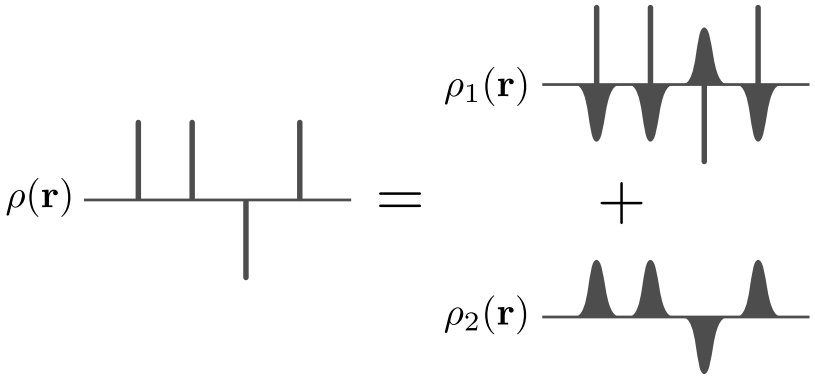


Figure 2.4: The Ewald method involves rewriting the charge density $\rho(\mathbf{r})$ of a collection of point charges as a sum of two different distributions: $\rho_1(\mathbf{r})$ is the same as ρ but screened by a collection of oppositely-charged Gaussians, and $\rho_2(\mathbf{r})$ is the negative of the screening charges.

where **A**, **B**, and **C** are each summed over the infinite number of cell translation vectors, and the prime on the last summation denotes that $i < j$ in the case $\mathbf{A} = 0$. Evaluating the first three lots of sums, i.e. the short-range terms, is straightforward since only bonded groups and non-bonded groups that fall within a finite cut-off distance need to be evaluated, as discussed in the previous section. The final (electrostatic) term, however, poses a challenge.

In principle, a finite cut-off could also be applied to the electrostatic contribution, but the cut-off would have to be prohibitively large in order to reduce the unphysical boundary effects to an acceptable level. Instead, the standard solution is to apply the method of Ewald [80] to evaluate the infinite sum.

The charge density of a set of charges $\{q_i\}$ at positions $\{\mathbf{R}_i\}$ is

$$\rho(\mathbf{r}) = \sum_{\mathbf{A}} \sum_{i} q_i \delta(\mathbf{r} - \mathbf{R}_i - \mathbf{A}) \tag{2.23}$$

where **A** is summed over the lattice translation vectors and $\delta(\cdot)$ is the Dirac delta function. This may be rewritten as

$$\rho(\mathbf{r}) = \rho_1(\mathbf{r}) + \rho_2(\mathbf{r}) \tag{2.24}$$

where

$$\rho_1(\mathbf{r}) \equiv \rho(\mathbf{r}) - \rho_2(\mathbf{r}) \tag{2.25}$$

$$\rho_2(\mathbf{r}) = \sum_{\mathbf{A}} \sum_{i} q_i \mathcal{G}(\mathbf{r} - \mathbf{R}_i - \mathbf{A}) \tag{2.26}$$

and $\mathcal{G}$ is a normalised Gaussian

$$\mathcal{G}(\mathbf{r}) = (\alpha/\pi)^{3/2} \exp(-\alpha r^2) \tag{2.27}$$

with a width parametrised by $\alpha > 0$.

The charge density $\rho_1(\mathbf{r})$ consists of the point charges from $\rho(\mathbf{r})$ screened by a set of opposite charges smeared in the form of Gaussians, and $\rho_2(\mathbf{r})$ is the negative of these screening charges. See Fig. 2.4. The potential energy of the charge configuration ρ_1 may be found via application of the Poisson equation. Similarly for ρ_2 except the Poisson equation is applied in Fourier space. The potential energy of ρ may then be written as the sum of the potential energies for ρ_1 and ρ_2, minus the artificial interaction energy between the point charges and Gaussian charges in ρ_1. The result is the following identity

$$\begin{aligned}
&\frac{1}{4\pi\epsilon_0}\sum_{\mathbf{A}}\sum_{i,j}{}' \frac{q_i q_j}{|\mathbf{R}_i - \mathbf{R}_j - \mathbf{A}|} \\
&\quad\equiv \frac{1}{4\pi\epsilon_0}\sum_{\mathbf{A}}\sum_{i,j}{}' \frac{q_i q_j}{|\mathbf{R}_i - \mathbf{R}_j - \mathbf{A}|}\operatorname{erfc}(|\mathbf{R}_i - \mathbf{R}_j - \mathbf{A}|\sqrt{\alpha}) \\
&\qquad + \frac{1}{2V\epsilon_0}\sum_{\mathbf{k}\neq 0}\frac{\exp(-k^2/4\alpha)}{k^2}\left|\sum_i q_i \exp(i\mathbf{k}\cdot\mathbf{R}_i)\right|^2 \\
&\qquad - \frac{1}{4\pi\epsilon_0}\sqrt{\frac{\alpha}{\pi}}\sum_i q_i^2
\end{aligned} \tag{2.28}$$

where V is the volume of the supercell, and $\operatorname{erfc}(x)$ is the complement of the error function

$$\operatorname{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt \tag{2.29}$$

The entire purpose of this reformulation is that the first term on the right-hand side of Eq. (2.28) is now a short-range term which converges quickly in real space and may therefore be truncated; the second term converges rapidly in k-space, meaning that it may also be truncated; and the third term is a finite sum.

In Eq. (2.28), the k-space sum is the most expensive term to compute, taking $\mathcal{O}(N^2)$ time. In this dissertation, we employ an optimisation known as PPPM [81] method which divides the supercell into a discrete grid (a mesh) that approximates the charge density ρ_2. The Poisson equation is then applied via discrete Fourier transforms to obtain a mesh-based approximation for the potential ϕ_2 which may then be interpolated. Lastly, a direct sum over the charges allows the energy to be computed. Through the use of fast Fourier transform methods, PPPM reduces the k-space sum to $\mathcal{O}(N \log N)$ time.

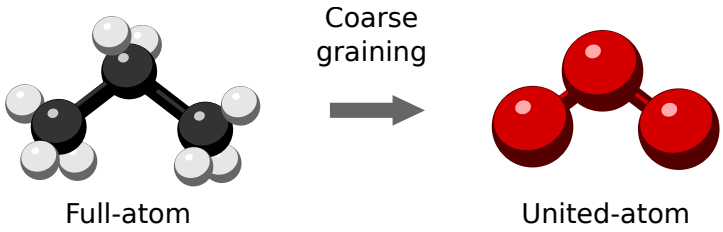


Figure 2.5: In the united-carbon model of propane, the eleven atoms are reduced to just three interaction sites.

2.3.4 *Coarse graining*

In moving from a quantum to a classical model of atoms, the nuclei and electrons have been united to form a single interaction site, with the electrons modelled implicitly via the force field. This reduction in resolution, known as coarse graining, comes at the cost of accuracy but makes larger systems more accessible by eliminating certain degrees of freedom.

It is common to extend this idea to groups of atoms. For example, when modelling alkane groups we will employ a united-carbon representation in which the methyl groups (-CH_3) and methylene bridges (-CH_2-) are treated as single particles. This process is illustrated in Fig. 2.5 for a propane molecule. Notice that the hydrogen atoms are fused with the carbon atoms and modelled implicitly instead.

Chapter 5 will develop this coarse graining idea further by treating space as a discrete lattice and using the kinetic Monte Carlo method (Section 2.7) to model on-lattice transitions.

2.4 STATISTICAL MECHANICS

The previous section introduced molecular mechanics which treats atoms, and sometimes molecules, as single points moving in a continuous space under the guidance of Newton's second law. In this model, the complete configuration, and all past and future configurations, of a collection of N particles may be uniquely defined by 6N coordinates: the position and momentum of each particle. In other words, the state of any particle system is fully described by a single point within a 6N-dimensional phase space which we shall call Γ-space. When considering large numbers of particles and long timescales, however, it becomes much more practical and intuitive to de-

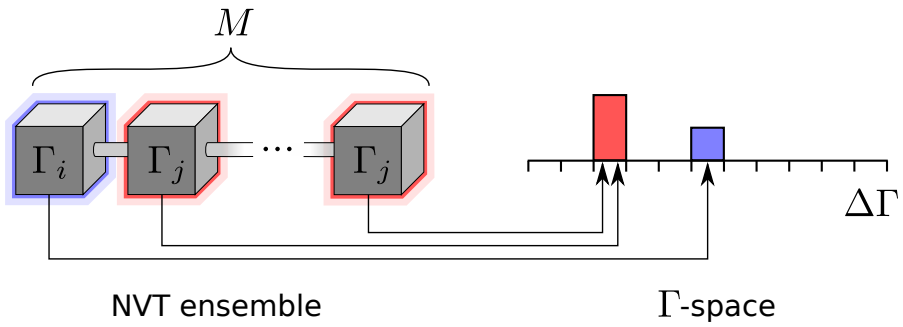


Figure 2.6: To derive the canonical distribution, we consider an ensemble of M boxes each of volume V and containing N particles. The systems are connected diathermally so that only heat may be exchanged between them. The state of each box is then mapped into a discretised phase space.

scribe the system in the language of thermodynamics, using concepts such as temperature and entropy, rather than with Γ-points.

Building on the accounts given in references [82, 83, 84, 85], this section will present the connection between Γ-space and thermodynamics, known as statistical mechanics, by considering the system shown schematically in Figure 2.6. The system consists of M boxes, each of volume V and containing N particles. The boxes are connected diathermally so that only heat may be exchanged between them. The entire system is considered in isolation and so the total energy ME remains constant over time, where E is the average energy per box.

The state of each box corresponds to some point x in Γ-space, and we define the function $\rho(x)$ such that $\rho(x)dx$ is the probability of finding a box within the differential element dx centered on x.

To enable us to count states, we shall discretise Γ-space into a finite number of cells of volume $\Delta\Gamma$. The probability of finding a system in the region $\Delta\Gamma$ centered on Γ_i may therefore be approximated as

$$P_i = \rho(\Gamma_i)\Delta\Gamma \tag{2.30}$$

It follows that each box is mapped to a discrete state Γ_i, as illustrated in Figure 2.6, with a probability P_i.

2.4.1 *Entropy and the second law of thermodynamics*

First, consider the case of a single box, $M = 1$. The box will have a volume V, contain N particles, and have a constant energy E. Statistical mechanics makes the postulate that the corresponding $6N$-dimensional Γ-point will visit all regions of Γ-space on the constant

energy E-hypersurface with equal probability. It follows that the most likely macrostate for the entire collection of particles to have is that with the highest number of corresponding Γ-points.

To illustrate this, imagine dividing the box into 2K cells, where $K \gg N$. Now place N non-interacting particles in the cells (without overlaps). The ratio of the number of possible arrangements where the particles occupy the entire box (Ω_2) versus only half of the box (Ω_1) will be

$$\Omega_2/\Omega_1 = \frac{(2K)!}{(2K-N)!N!} \Big/ \frac{K!}{(K-N)!N!} \to 2^N \tag{2.31}$$

in the limit $K \to \infty$. For a single mole of gas, $\Omega_2/\Omega_1 \sim 2^{10^{23}}$. Such an astronomically large number explains why gases fill their entire volume and never spontaneously compress.

In general, ordered macrostates correspond to far fewer Γ-points than disordered states, and so random sampling of phase space will naturally drive the system into a state of disorder. In other words, the macrostate will evolve so as to maximise the number of corresponding Γ-points, Ω. It will be convenient to define the so-called entropy

$$S = k_B \log \Omega \tag{2.32}$$

where k_B is Boltzmann's constant. Since the logarithm is monotonic then a maximisation of Ω will correspond to a maximisation of the entropy. It follows that $\Delta S \geqslant 0$ over time, which is the second law of thermodynamics.

2.4.2 *Temperature*

Consider just two boxes, $M = 2$. The second law of thermodynamics gives the following inequality

$$\Delta S = \left(\frac{\partial S}{\partial E}(E_1) - \frac{\partial S}{\partial E}(E_2) \right) \Delta E_1 \geqslant 0 \tag{2.33}$$

from which it follows that

$$\frac{\partial S}{\partial E}(E_1) > \frac{\partial S}{\partial E}(E_2) \quad \Rightarrow \quad \Delta E_1 > 0 \tag{2.34}$$

$$\frac{\partial S}{\partial E}(E_1) < \frac{\partial S}{\partial E}(E_2) \quad \Rightarrow \quad \Delta E_1 < 0 \tag{2.35}$$

In other words, energy flows from boxes with small $\partial S/\partial E$ to those with large $\partial S/\partial E$. It is therefore natural to define a statistical temperature T as

$$(k_B T)^{-1} = \frac{\partial S}{\partial E} \tag{2.36}$$

where k_B is introduced for dimensional purposes.

It turns out that when this definition is applied to the ideal gas, it produces the standard ideal gas law. And so the statistical temperature T indeed corresponds to the temperature measured by ideal gas thermometers.

2.4.3 *Canonical distribution*

Now consider a large number of boxes, $M \gg 1$. Each box will correspond to some discrete state in Γ-space, with $M_i \geqslant 0$ boxes mapping to Γ_i, where

$$1 = M^{-1} \sum_i M_i = \sum_i P_i \tag{2.37}$$

and $P_i = M_i/M$, with a possibility of degeneracy ($M_i > 1$).

The total number of ways that a particular distribution $\{M_i\}$ could arise is given by

$$\Omega = \frac{M!}{\Pi_i M_i!} \tag{2.38}$$

and the most likely distribution will be that which maximises Ω.

Intuitively, in the absence of any energy constraints, the least likely distribution would be that with all boxes in the same state, i.e. $M_i = M$ for some i, and the most likely would be that with a uniform distribution of M_i. However, when the system is subject to the energy constraint

$$E = M^{-1} \sum_i M_i E_i = \sum_i P_i E_i \tag{2.39}$$

the uniform distribution no longer maximises Ω. To see this, we switch from maximising Ω to the equivalent task of maximising the average entropy per box

$$\begin{aligned} S &= M^{-1} k_B \log\left(\frac{M!}{\Pi_i M_i!}\right) \\ &\approx -k_B \sum_i P_i \log P_i \end{aligned} \tag{2.40}$$

where Stirling's formula has been applied.

Maximising Eq. (2.40) subject to the constraints Eq. (2.37) and Eq. (2.39) can be achieved by introducing the Lagrange multipliers μ and β, and solving

$$\nabla_{\{P_i\},\mu,\beta} \left[S - \mu \left(\sum_i P_i - 1 \right) - \beta \left(\sum_i P_i E_i - E \right) \right] = 0 \tag{2.41}$$

which results in the Boltzmann distribution

$$P_i = Z^{-1} e^{-\beta E_i} \tag{2.42}$$

where $Z \equiv e^{1+\mu}$ is a normalisation constant.

Now taking the limits $M \to \infty$ and $\Delta\Gamma \to 0$, a continuous probability distribution is recovered

$$\rho(x) = Z^{-1} e^{-\beta E(x)} \tag{2.43}$$

where

$$Z = \int_\Gamma e^{-\beta E(x)} \, dx \tag{2.44}$$

follows from the normalisation of $\rho(x)$. A continuous definition of entropy is also obtained

$$S = -k_B \int_\Gamma \rho(x) \log \rho(x) \, dx \tag{2.45}$$

$$= k_B (\beta E + \log Z) \tag{2.46}$$

To interpret the multiplier β, note that

$$dS = k_B (\beta \, dE + E d\beta + \frac{\partial \log Z}{\partial \beta} d\beta) \tag{2.47}$$

and

$$E = -\frac{\partial \log Z}{\partial \beta} \tag{2.48}$$

from which it follows that

$$\beta = \frac{\partial S}{\partial E} = (k_B T)^{-1} \tag{2.49}$$

where the connection to temperature comes from Eq. (2.36).

2.4.4 *Ensemble averages*

Experimental measurements are typically performed over some finite time interval, and so a measured quantity Q is effectively averaged over an ensemble of configurations. Mathematically, such quantities can be computed as

$$\langle Q \rangle \equiv \int_\Gamma Q(x) \rho(x) \, dx \tag{2.50}$$

$$= \frac{\int_\Gamma Q(x) e^{-\beta E(x)} \, dx}{\int_\Gamma e^{-\beta E(x)} \, dx} \tag{2.51}$$

In some cases we may be interested in an average while some condition $\theta(\Gamma) = 0$ holds true,

$$\langle Q \rangle_\theta = \frac{\int_\Gamma Q(x)\delta(\theta(x))e^{-\beta E(x)}\,dx}{\int_\Gamma \delta(\theta(x))e^{-\beta E(x)}\,dx} \tag{2.52}$$

$$= \frac{\langle Q(x)\delta(\theta(x))\rangle}{\langle \delta(\theta(x))\rangle} \tag{2.53}$$

2.4.5 *Helmholtz free energy and metastability*

Let us define the so-called Helmholtz free energy

$$F \equiv E - TS \tag{2.54}$$

$$= -\beta^{-1} \log Z \tag{2.55}$$

$$= -\beta^{-1} \log \int_\Gamma e^{-\beta E(x)}\,dx \tag{2.56}$$

where the second line follows from Eq. (2.46) and the third line from the definition of the partition function, Eq. (2.44).

Now consider some function $\lambda : \Gamma \to \mathbb{R}^m$ of the system; it may project out atomic coordinates, bond angles, or something more abstract. The free energy of the system constrained to the hypersurface $\lambda(x) = \lambda_0$, measured with respect to the unconstrained system, is

$$F(\lambda_0) = -\beta^{-1} \log \int_\Gamma e^{-\beta E(x)}\delta(\lambda(x) - \lambda_0)\,dx \tag{2.57}$$

$$- \left(-\beta^{-1} \log \int_\Gamma e^{-\beta E(x)}\,dx\right) \tag{2.58}$$

$$= -\beta^{-1} \log \frac{\int_\Gamma e^{-\beta E(x)}\delta(\lambda(x) - \lambda_0)\,dx}{\int_\Gamma e^{-\beta E(x)}\,dx} \tag{2.59}$$

$$= -\beta^{-1} \log \langle \delta(\lambda(x) - \lambda_0)\rangle \tag{2.60}$$

and the corresponding gradient is

$$-\nabla_{\lambda_0} F(\lambda_0) = \beta^{-1} \nabla_{\lambda_0} \log \frac{\int_\Gamma e^{-\beta E(x)}\delta(\lambda(x) - \lambda_0)\,dx}{\int_\Gamma e^{-\beta E(x)}\,dx} \tag{2.61}$$

$$= \beta^{-1} \frac{\int_\Gamma (\nabla_{\lambda_0} E)\,\frac{\partial}{\partial E} e^{-\beta E(x)}\delta(\lambda(x) - \lambda_0)\,dx}{\int_\Gamma e^{-\beta E(x)}\delta(\lambda(x) - \lambda_0)\,dx} \tag{2.62}$$

$$= \frac{\int_\Gamma (-\nabla_{\lambda_0} E)\, e^{-\beta E(x)}\delta(\lambda(x) - \lambda_0)\,dx}{\int_\Gamma e^{-\beta E(x)}\delta(\lambda(x) - \lambda_0)\,dx} \tag{2.63}$$

$$= \langle -\nabla_{\lambda_0} E\rangle_{\lambda_0} \tag{2.64}$$

i.e. the ensemble average of the force constrained at λ_0.

This result is of fundamental importance. It reveals that, while the *instantaneous* force acting on λ will always attempt to decrease the

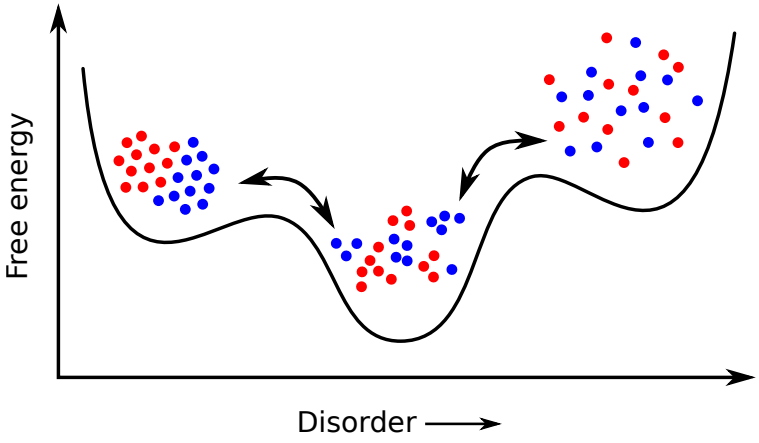


Figure 2.7: A hypothetical free energy curve for a toy system described in the text. It illustrates the role of entropy and metastability at a constant temperature.

potential energy of the system, the *average* force will attempt to minimise the free energy F. In other words, when the system is coupled to a heat bath that fixes the temperature, there is a tendency to minimise the potential energy while simultaneously maximising the entropy. The resulting equilibrium state, which corresponds to a free energy minimum, will therefore result from a balance between these two competing forces, with the relative proportion determined by the temperature T.

To give an example, consider a system of blue and red particles (one could imagine water and oil). Suppose that all of the particles are attracted to each other, but like-coloured particles attract more strongly than opposite-coloured particles. As the temperature is reduced to zero, the system will move to the (assume global) minimum of E, adopting an ordered structure with the red and blue particles segregated. As the temperature is increased, the colours will begin to mix and the system will become less well-ordered, but there will remain significant local clustering of like-colours. As the temperature is increased further, the bonds will break and the particles will move to a thoroughly-mixed state. These three states are shown in Figure 2.7 along with a hypothetical free energy curve as a function of some order parameter at one particular temperature.

In Figure 2.7 there exist three free energy minima. The minimum with the lowest free energy is called the thermodynamically stable state, whereas the other two are said to be metastable. While the metastable states are not thermodynamically favourable at this par-

ticular temperature, the system must overcome a free energy barrier to transition away from them which can take a long time (see Section 2.6 for a discussion of rare events). Changing the temperature will change the free energy curve and the relative stability of these three states.

For a more comprehensive discussion of free energies and metastability, see [84].

2.4.6 *Isothermal-isobaric ensemble*

In many real-world experiments, the system under study not only has a constant temperature T but is also free to change its volume V so as to maintain a constant pressure P. The above equations can be applied directly to such systems but with one minor alteration: the product PV must be added to all of the energy terms to account for the work done by the system on its surroundings. Of particular importance, the Helmholtz free energy becomes the Gibbs free energy

$$G = E + PV - TS \tag{2.65}$$

where the component $E + PV$ is called the enthalpy. The minima of G then determine the (meta)stable states of an isothermal-isobaric system.

2.5 MOLECULAR DYNAMICS

It is possible to numerically integrate the equations of motion, Eq. (2.15), to generate particle trajectories, thus allowing the dynamics of a system to be studied. This involves repeatedly advancing the positions and velocities by a small time increment, from t to $t + \Delta t$, as will be described in Section 2.5.1, until a sufficiently long time has been reached.

One of the properties of any Hamiltonian system is that it conserves energy over time. As a result, any potential energy changes that occur as a system evolves will result in kinetic energy being absorbed or released, thus modifying the temperature. However, many systems of interest interact with a heat bath which imposes a fixed temperature on the system. Section 2.5.2 will show how the equations of motion can be modified to maintain a fixed temperature. Maintaining a constant pressure will also be discussed in Section 2.5.3.

2.5.1 *Symplectic integration*

Summing the third-order Taylor expansions for $\mathbf{R}_i(t+\Delta t)$ and $\mathbf{R}_i(t-\Delta t)$, and rearranging, one obtains

$$\mathbf{R}_i(t+\Delta t) = 2\mathbf{R}_i(t) - \mathbf{R}_i(t-\Delta t) + \frac{1}{2}\frac{d^2\mathbf{R}_i}{dt^2}(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (2.66)$$

which is known as the Verlet scheme of symplectic integration.

The acceleration is computed from Newton's second law (Eq. (2.15))

$$\frac{d^2\mathbf{R}_i}{dt^2}(t) = -m_i^{-1}\nabla\mathcal{U}(\{\mathbf{R}_i(t)\}) \quad (2.67)$$

where m_i is the i-th atomic mass and $\mathcal{U}$ is the potential energy given in Eq. (2.17).

In the form of Eq. (2.66), the Verlet scheme requires a pair of consecutive coordinates, i.e. $\mathbf{R}_i(t-\Delta t)$ and $\mathbf{R}_i(t)$, in order to compute the subsequent coordinates $\mathbf{R}_i(t+\Delta t)$. This makes initiating the algorithm somewhat inelegant. Another shortcoming of Eq. (2.66) is that a calculation of the velocities at time t

$$\mathbf{v}_i(t) \equiv \frac{d\mathbf{R}_i}{dt}(t) = \frac{\mathbf{R}_i(t+\Delta t) - \mathbf{R}_i(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \quad (2.68)$$

requires that the subsequent positions $\mathbf{R}_i(t+\Delta t)$ be known, which can greatly complicate the implementation. To address these problems, variations of the Verlet scheme exist.

In this dissertation we employ the velocity Verlet method which is implemented in three stages:

1. The velocities are integrated by a half-time-step

$$\mathbf{v}_i(t+\frac{1}{2}\Delta t) = \mathbf{v}_i(t) + \frac{1}{2}\frac{d^2\mathbf{R}_i}{dt^2}(t)\Delta t^2 \quad (2.69)$$

   where the forces are computed at the current position of time t.

2. The positions are integrated

$$\mathbf{R}_i(t+\Delta t) = \mathbf{R}_i(t) + \mathbf{v}_i(t+\frac{1}{2}\Delta t)\Delta t \quad (2.70)$$

3. The velocities are integrated by a second half-time-step

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t+\frac{1}{2}\Delta t) + \frac{1}{2}\Delta t\frac{d^2\mathbf{R}_i}{dt^2}(t+\Delta t) \quad (2.71)$$

This method generates an identical trajectory to the original Verlet scheme of Eq. (2.66) but provides both the positions and velocities synchronously, and requires only the initial positions and velocities to commence.

In the limit $\Delta t \to 0$, the Verlet method has an important property of area-preservation [86], meaning that any constant-energy hypersurface in Γ-space will be mapped isomorphically back onto itself. If this were not the case then regions of Γ-space may become inaccessible, i.e. a failure of ergodicity (see Section 2.5.4 below), or states of a different energy may become accessible, giving rise to energy drift.

2.5.2 *Thermostatting*

Consider the kinetic energy of just one particle in one dimension, $E_i = p_i^2/2m_i$. The ensemble average of E_i is

$$\langle E_i \rangle = \frac{1}{2} k_B T \tag{2.72}$$

which leads naturally to the following definition for the instantaneous temperature

$$T(t) = \frac{1}{3Nk_B} \sum_{i=1}^{N} \frac{p_i^2}{m_i} \tag{2.73}$$

Under isothermal conditions, $T(t)$ will fluctuate about the desired temperature T_0. A naïve way of enforcing the temperature T_0 might then be to simply rescale all of the velocities at regular time intervals

$$v_i(t) \leftarrow v_i(t) \sqrt{\frac{T_0}{T(t)}} \tag{2.74}$$

However, this method fails to produce natural temperature fluctuations which, for instance, help to activate rare processes. Instead, the most popular method of thermostatting is that of Nosé and Hoover [87], which is deterministic and correctly samples the canonical ensemble. It involves introducing an additional degree of freedom to the system, with an effective mass Q, which represents a surrounding heatbath. The temperature is then regulated through the transfer of kinetic energy between the system and the heatbath, with the strength of coupling determined by Q. The result is a frictional force

$$\frac{d^2\mathbf{R}_i(t)}{dt^2} \leftarrow \frac{d^2\mathbf{R}_i(t)}{dt^2} - \chi(t)\mathbf{v}(t) \tag{2.75}$$

where the coefficient behaves akin to Newton's law of cooling

$$\frac{d\chi(t)}{dt} = \frac{3Nk_B}{Q}(T(t) - T_0) \tag{2.76}$$

2.5.3 *Barostatting*

To produce isothermal-isobaric conditions, we employ the Melchionna [88] modification of the Nosé-Hoover method which incorporates a barostat with the thermostat. The barostat functions in a similar way to the thermostat, by introducing an additional (tensorial) degree of freedom that modifies the equations of motion and the cell vectors in response to pressure deviations about the desired value.

2.5.4 *Ergodicity*

Ergodicity is a dynamical property that means a trajectory in Γ-space will eventually visit all regions of the space that belong to the correct ensemble. Put another way, any property f averaged over an infinitely long trajectory $x(t)$ will equal the corresponding ensemble average

$$\lim_{t\to\infty} t^{-1} \int_0^t f(x(t'))\, dt' = \int_\Gamma f(x)\rho(x)\, dx \tag{2.77}$$

This has two important consequences. The first is that statistically averaged properties may be obtained from molecular trajectories. The second is that time-averaged properties of trajectories eventually become independent of their initial conditions.

While there exists no general proof, it is widely believed that virtually all systems of chemical interest are ergodic, and so ergodicity is taken as a hypothesis.

2.6 RARE EVENT SAMPLING

The dynamics of a complex molecular system may exhibit behaviour on multiple characteristic time-scales, ranging from petahertz vibrations to rare events such as conformational transitions and phase nucleation. These rare events involve overcoming enthalpic and/or entropic barriers and are activated by improbable thermal fluctuations. Consequently, they usually fail to occur during the time-scales accessible to molecular simulation, or occur too infrequently for sound statistical sampling. This practical failure of ergodicity, known as quasi-nonergodicity, demands the development of new techniques in order to study activated processes.

2.6.1 *Collective variables*

For many systems, activated processes take place in a relatively low-dimensional space, and it is useful to project the molecular configurations $x \in \Gamma$ onto these reduced coordinates $\lambda(x) \in \mathbb{R}^m$ known as *collective variables* (CVs). Doing so simplifies the description of the transition and gives rise to a free energy surface $F(\lambda)$, cf. Eq. (2.59), which is generally smoother than the potential energy surface.

To provide a couple of exampes of collective variables used in this thesis: in chapters 3 and 4, the free energy is constructed for the dissolution of a single ion from step and kink sites. In these instances, λ projects out the (x, y, z) coordinates of the dissolving ion only, while all other coordinates are integrated out. As another example, in chapter 3, we briefly generate the free energy surface for the conformational space of alanine dipeptide. In this case, we assume that the conformations can be described entirely by a pair of dihedral angles.

The suitability of a particular set of collective variables will be discussed in the next section.

2.6.2 *Reaction pathways*

A reaction involves transitioning from one (meta)stable region of configuration space, the reactant states $A \subset \Gamma$, to another, the product states $B \subset \Gamma$. In λ-space, this involves traversing a path $\phi : [0, 1] \to \mathbb{R}^m$ that connects the two regions: $\phi(0) \in \lambda(A)$ and $\phi(1) \in \lambda(B)$.

The probability that a configuration in $\lambda(A)$ will evolve to the region $\lambda(B)$ along a path no longer than τ (in CV-space) can be written as a path integral

$$P(\tau) = \int_{\phi(0)\in\lambda(A),\, \phi(1)\in\lambda(B),\, \ell(\phi)\leqslant\tau} \tilde{P}[\phi]\, \mathcal{D}\phi \tag{2.78}$$

where $\ell(\phi)$ is the length of the path ϕ, and ϕ is integrated over all possible paths satisfying the constraints.

By applying the Kolmogorov backward equation to $P(\tau)$ and lifting any constraint on the length of the path ($\tau \to \infty$) it can be shown [89] that the path integral of Eq. (2.78) is dominated by a single trajectory: the so-called minimum free energy path (MFEP) which passes along the direction of steepest descent of F via a saddle point, i.e.

$$\nabla F(\phi(0)) = \nabla F(\phi(1)) = 0, \quad \text{and} \tag{2.79}$$

$$\nabla F(\phi(t)) = |\phi'(t)|^{-2} [\nabla F(\phi(t)) \cdot \phi'(t)] \phi'(t) \quad \forall t \in (0, 1) \tag{2.80}$$

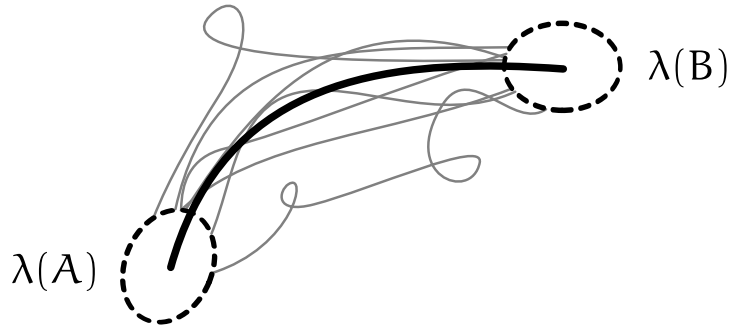


Figure 2.8: A number of possible trajectories from $\lambda(A)$ to $\lambda(B)$ in CV-space are shown (grey) along with the MFEP (black).

The MFEP is the most likely transition pathway, see Figure 2.8, and knowing it facilitates analysis of both the thermodynamics and kinetics of the transition, as detailed below.

It will also be convenient to introduce a single reaction coordinate $\xi : \Gamma \to \mathbb{R}$ that separates the reactant and product states and measures progress between them. In this thesis we use the length along the appropriate MFEP as our reaction coordinate. The *true* reaction coordinate for any process is (bijective with) the committor function $p(x)$ which gives the probability that the state $x \in \Gamma$ will evolve to (say) a product state in B before reaching a reactant state in A [90]. In fact, a particular set of collective variables $\lambda(x)$ will only be suitable for describing a transition if this committor function can be well-approximated as a function of the CVs alone

$$p(x) \approx f(\lambda(x)) \tag{2.81}$$

for some function $f : \mathbb{R}^m \to \mathbb{R}$.

There is no rigorous procedure for selecting CVs. Generally, intuition is relied upon to generate a list of coordinates that are likely to be relevant to the transition under consideration. However, once a list has been generated, there are steps one can take to eliminate certain CVs. Firstly, any CV that fluctuates on a much shorter time-scale than the mean reaction time can be neglected. Secondly, if a reconstruction of the free energy curve or of the committor function is sufficiently unaffected by the omission of certain coordinates, then those coordinates may be discarded. For example, the water-hydrogen coordination number of the terminal kink ion was found to be irrelevant in predicting the free energy barrier of kink dissolution for sodium chloride [45].

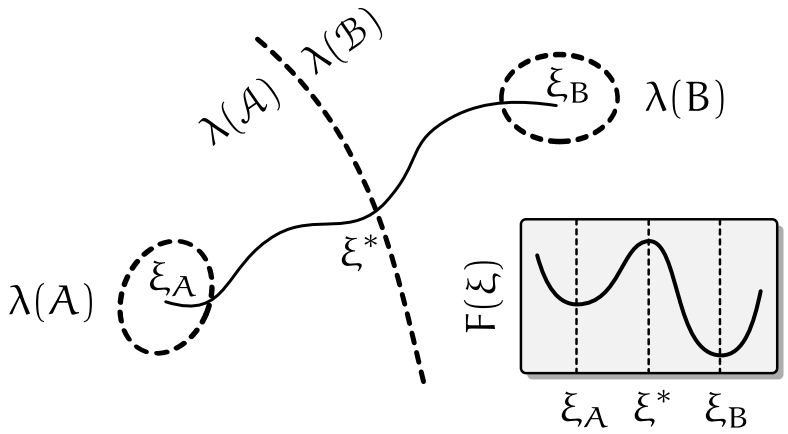


Figure 2.9: Space is divided by the maximum of the free energy curve into (calligraphic) regions $\mathcal{A}$ and $\mathcal{B}$.

In this dissertation, the suitability of the CVs will be taken as self-evident. As for the reaction coordinate, any function that can map bijectively with the minimum free energy path, at least in the vicinity of the transition states, is usually sufficient.

2.6.3 *Reactive flux*

This section is concerned with computing the flux (or rate) of transitions from region A to region B within a bistable system, where we define the *flux* as the probability of the event occurring per unit time.

The first step is to partition space into two sets, denoted calligraphically as $\mathcal{A}$ and $\mathcal{B}$, i.e. $\Gamma = \mathcal{A} \cup \mathcal{B}$. These sets must include the (non-calligraphic) (meta)stable regions A and B, i.e. $A \subset \mathcal{A}$ and $B \subset \mathcal{B}$, as well as their respective basins. We then define a reaction coordinate ξ that partitions space into the respective basins:

$$\mathcal{A} = \{x \in \Gamma \,|\, \xi(x) < \xi^*\} \tag{2.82}$$

$$\mathcal{B} = \{x \in \Gamma \,|\, \xi(x) \geqslant \xi^*\} = \mathcal{A}^c \tag{2.83}$$

where ξ^* corresponds to the (transition) maximum of the free energy curve $F(\xi)$, as depicted in Figure 2.9. Each element $x \in \Gamma$ can then be identified with either the set $\mathcal{A}$ or $\mathcal{B}$ via the indicator functions

$$h_{\mathcal{A}}(x) = \begin{cases} 1 & x \in \mathcal{A} \\ 0 & x \notin \mathcal{A} \end{cases} = \theta(\xi^* - \xi(x)) \tag{2.84}$$

$$h_{\mathcal{B}}(x) = \begin{cases} 1 & x \in \mathcal{B} \\ 0 & x \notin \mathcal{B} \end{cases} = \theta(\xi(x) - \xi^*) \tag{2.85}$$

where

$$\theta(x) \equiv \int_{-\infty}^{x} \delta(s)\, ds \tag{2.86}$$

is the Heaviside step function.

Now we define the so-called correlation function

$$c(t) = \frac{\langle h_{\mathcal{A}}(x(0)) h_{\mathcal{B}}(x(t))\rangle}{\langle h_{\mathcal{A}}(x)\rangle} \tag{2.87}$$

where the ensemble average $\langle\cdot\rangle$ is defined in Eq. (2.50). From the definition of the indicator functions we see that the term $h_{\mathcal{A}}(x(0))h_{\mathcal{B}}(x(t))$ in the numerator is 1 for any state that is in $\mathcal{A}$ and, at a time t in the future, is in $\mathcal{B}$. When averaged over the entire ensemble, the numerator therefore counts the fraction of states from the entire system that satisfy this condition. The denominator, on the other hand, simply counts the fraction of states that are in $\mathcal{A}$. The ratio, i. e. the correlation function $c(t)$, therefore equals the probability that any state chosen from $\mathcal{A}$ will be in $\mathcal{B}$ a time t in the future.

The flux from $\mathcal{A}$ to $\mathcal{B}$ is then the time-derivative of $c(t)$

$$\begin{aligned}
k_{\mathcal{A}\to\mathcal{B}}(t) &= \dot{c}(t) && (2.88)\\
&= \frac{\partial_t \langle h_{\mathcal{A}}(0) h_{\mathcal{B}}(t)\rangle}{\langle h_{\mathcal{A}}\rangle} && (2.89)\\
&= -\frac{\langle \dot{h}_{\mathcal{A}}(-t) h_{\mathcal{B}}(0)\rangle}{\langle h_{\mathcal{A}}\rangle} && (2.90)\\
&= -\frac{\langle \dot{h}_{\mathcal{A}}(0) h_{\mathcal{B}}(t)\rangle}{\langle h_{\mathcal{A}}\rangle} && (2.91)\\
&= \frac{\langle \delta(\xi^* - \xi(x(0)))\dot{\xi}(x(0))\theta(\xi(x(t)) - \xi^*)\rangle}{\langle \theta(\xi^* - \xi(x))\rangle} && (2.92)
\end{aligned}$$

where we have used the time-invariance of the ensemble average, $\langle h_{\mathcal{A}}(x(0))h_{\mathcal{B}}(x(t))\rangle = \langle h_{\mathcal{A}}(x(-t))h_{\mathcal{B}}(x(0))\rangle$, and the relation $\dot{\theta}(\xi^* - \xi(x)) = -\delta(\xi^* - \xi(x))\dot{\xi}(x)$.

The expression of Eq. (2.92) can be rewritten as the product

$$k_{\mathcal{A}\to\mathcal{B}}(t) = k^{\mathrm{TST}}_{\mathcal{A}\to\mathcal{B}}\kappa(t) \tag{2.93}$$

where

$$k^{\mathrm{TST}}_{\mathcal{A}\to\mathcal{B}} = \frac{\langle \delta(\xi^* - \xi(x(0)))\rangle}{\langle \theta(\xi^* - \xi(x))\rangle}\langle \dot{\xi}(x(0))\theta(\dot{\xi}(x(0))\rangle_{\xi(x(0))=\xi^*} \tag{2.94}$$

is the rate predicted by transition state theory (TST) [91], and

$$\begin{aligned}
\kappa(t) &= \frac{\langle \delta(\xi^* - \xi(x(0)))\dot{\xi}(x(0))\theta(\xi(x(t)) - \xi^*)\rangle}{\langle \delta(\xi^* - \xi(x(0)))\rangle\langle \dot{\xi}(x(0))\theta(\dot{\xi}(x(0))\rangle_{\xi(x(0))=\xi^*}} && (2.95)\\
&= \frac{\langle \dot{\xi}(x(0))\theta(\xi(x(t)) - \xi^*)\rangle_{\xi(x(0))=\xi^*}}{\langle \dot{\xi}(x(0))\theta(\dot{\xi}(x(0))\rangle_{\xi(x(0))=\xi^*}} && (2.96)
\end{aligned}$$

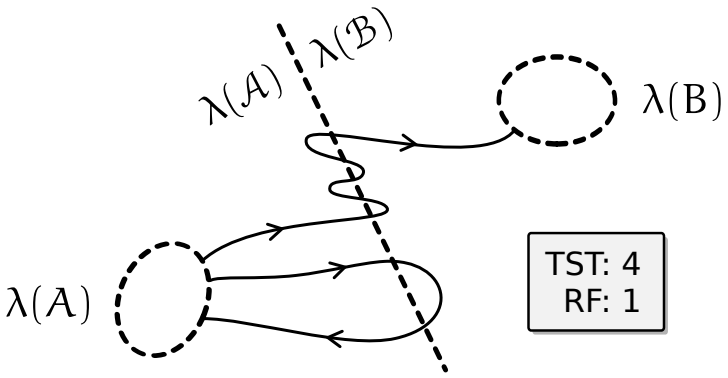


Figure 2.10: Trajectories may cross the transition boundary several times before reaching the reactant state, or they may return to the product state first. In the trajectories shown, the TST method would count four transitions whereas RF only counts one.

is the time-dependent component.

The rate predicted by TST counts all crossings that the trajectories make over the boundary from $\mathcal{A} \to \mathcal{B}$, including reactive trajectories that diffuse multiple times across the boundary, and non-reactive trajectories that momentarily cross the boundary. The time-dependent function $\kappa(t)$ accounts for this over-counting by subtracting any trajectories that pass backward from $\mathcal{B} \to \mathcal{A}$. This difference in the TST and reactive flux (RF) counting is illustrated in Figure 2.10.

The time-dependent behaviour of $\kappa(t)$ (and therefore $k_{\mathcal{A}\to\mathcal{B}}(t)$) is shown in Figure 2.11. Over an initial period comparable to the transition time-scale, $\kappa(t)$ will vary while the states in $\mathcal{A}$ relax to either A or B, temporarily plateauing to a constant value κ which we call the *transmission coefficient*. The corresponding time-independent flux

$$k_{A\to B} = k_{\mathcal{A}\to\mathcal{B}} = k^{\mathrm{TST}}_{\mathcal{A}\to\mathcal{B}}\kappa \tag{2.97}$$

is the reactive flux we wish to compute. Eventually, as time reaches a scale $\sim k^{-1}_{A\to B}$, the location of a trajectory becomes decorrelated with its initial position, and so the flux decreases to zero.

The TST expression, Eq. (2.94), may be rewritten in terms of the free energy

$$k^{\mathrm{TST}}_{\mathcal{A}\to\mathcal{B}} = \langle \dot{\xi}(x(0))\theta(\dot{\xi}(x(0))\rangle_{\xi(x(0))=\xi^*} \frac{e^{-\beta F(\xi^*)}}{\int_{-\infty}^{\xi^*} e^{-\beta F(\xi)}\, d\xi} \tag{2.98}$$

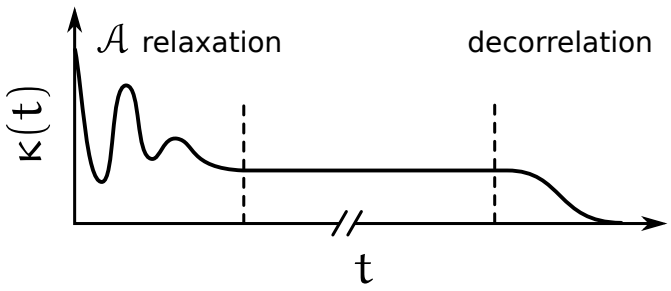


Figure 2.11: The time-dependent behaviour of $\kappa(t)$ and thus the transition rate $k_{\mathcal{A}\to\mathcal{B}}(t)$. The states relax to either A or B which causes a plateau. Eventually they become decorrelated with their initial positions and the average flux becomes zero.

where $\langle\delta(\xi^* - \xi(x))\rangle = e^{-\beta F(\xi^*)}$ follows from Eq. (2.60), and

$$\langle\theta(\xi^* - \xi(x))\rangle = \left\langle \int_{-\infty}^{\xi^* - \xi(x)} \delta(s)\, ds \right\rangle \tag{2.99}$$

$$= \int_{-\infty}^{\xi^*} \langle\delta(s - \xi(x))\rangle\, ds \tag{2.100}$$

$$= \int_{-\infty}^{\xi^*} e^{-\beta F(\xi)}\, d\xi \tag{2.101}$$

follows from the definition of the Heaviside function, Eq. (2.86).

This provides the standard route for computing the reactive flux of a transition:

1. construct the MFEP ϕ
2. compute the free energy curve $F(\xi)$
3. compute the TST flux, $k^{TST}_{\mathcal{A}\to\mathcal{B}}$, from Eq. (2.98)
4. compute the plateau of $\kappa(t)$, κ, by averaging Eq. (2.96) over a series of trajectories initiated at the top of the transition barrier, $\xi(x(0)) = \xi^*$
5. the flux is $k_{A\to B} = \kappa k^{TST}_{\mathcal{A}\to\mathcal{B}}$.

The details of steps 1 and 2 will be given in the following sections.

Lastly, we note from the approximation

$$\int_{-\infty}^{\xi^*} e^{-\beta F(\xi)}\, d\xi \approx e^{-\beta F(\xi_A)} \tag{2.102}$$

that the transition rate is proportional to the Boltzmann factor

$$k^{TST}_{A\to B} \propto e^{-\beta \Delta F} \tag{2.103}$$

where $\Delta F = F(\xi^*) - F(\xi_A)$ is the free energy barrier associated with the transition. This is known as the Arrhenius equation.

2.6.4 *Constructing minimum free energy pathways*

2.6.5 *Mean forces*

Some of the methods described in this section will require the mean force

$$f(\lambda_0) = -\nabla_\lambda F(\lambda_0) = \langle -\nabla_\lambda E \rangle_{\lambda_0} \tag{2.104}$$

to be evaluated, with the system constrained at $\lambda = \lambda_0$.

We achieve this using molecular dynamics by restraining (rather than constraining) the trajectory $x(t)$ to the point $\lambda = \lambda_0$ with a harmonic potential

$$U_{\lambda_0}(x(t)) = \frac{k}{2}|\lambda(x(t)) - \lambda_0|^2 \tag{2.105}$$

added to the Hamiltonian. The mean force is then obtained from

$$f(\lambda_0) = -\langle -\nabla_\lambda U_{\lambda_0} \rangle \approx t^{-1} \int_0^t \nabla_\lambda U_{\lambda_0}(x(t'))\, dt' \tag{2.106}$$

where the time-average follows from the ergodic hypothesis (cf. Section 2.5.4) and equality is achieved in the limits $t, k \to \infty$.

The spring constant k is restricted to a finite value in order to mitigate numerical integration errors. However, k must not be too small since the trajectory will mostly sample a region slightly downhill from the intended point λ_0 in CV-space. One should therefore decide the largest tollerable error in λ_0 and then choose an appropriate spring constant k.

2.6.5.1 *String method*

The string method of Weinan et al. [92] provides a way of constructing MFEPs in CV-space. It begins with a linear set of points $\{\lambda_i\}$ in CV-space, called *images*, that form a *string*. The mean force $-\nabla_\lambda F$ is evaluated at each point along the string and the positions integrated downhill using the forward Euler method

$$\lambda_i \leftarrow \lambda_i - h\nabla_\lambda F(\lambda_i) \tag{2.107}$$

for some small step size h. The string is then reparametrised so that the points are uniformly spaced between the two end-points, which we achieve by fitting a cubic spline to the new points and then interpolating. The process is illustrated in Figure 2.12 and is repeated until the maximum force acting on the string is below some threshold.

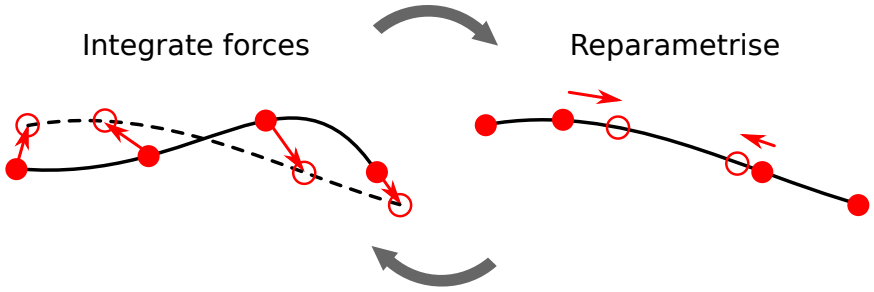


Figure 2.12: The string method begins with a string of points between the product and reactant basins. It incrementally integrates the forces, reparametrising the string after each iteration so as to maintain uniformity.

A popular and similar method is the nudged elastic band (NEB) method [93] which uses a harmonic potential between the λ_i images to maintain uniformity rather than the spline reparametrisation of the string method. The string method is the inextensible limit of the NEB method (i.e. the limit of an infinite spring constant), thus requiring one less parameter. It is also more stable, supporting a step size that scales as $h \sim N^{-1}$ compared to $h \sim N^{-2}$ for NEB, where N is the number of images [92, 94].

2.6.6 *Constructing free energy maps*

2.6.6.1 *Thermodynamic integration*

Once a minimum free energy path $\{\lambda_i\}$ has been constructed, the associated free energy curve can be obtained by integrating the mean forces along it

$$F(\lambda_i) - F(\lambda_0) \approx \sum_{j<i} \nabla_\lambda F(\lambda_j) \cdot (\lambda_{j+1} - \lambda_j) \tag{2.108}$$

where the evaluation of $\nabla_\lambda F$ is described in Section 2.6.5.

2.6.6.2 *Umbrella sampling*

While thermodynamic integration is easy to apply, a much more efficient and accurate method for reconstructing free energy curves is umbrella sampling which builds on the identity

$$\exp(-\beta F(\lambda_0)) = \langle \delta(\lambda - \lambda_0) \rangle \tag{2.109}$$

$$\equiv \frac{\exp(\beta V(\lambda_0)) \langle \delta(\lambda - \lambda_0) \rangle_V}{\langle \exp(\beta V(\lambda)) \rangle_V} \tag{2.110}$$

where the first line follows from Eq. (2.59), and $\langle\cdot\rangle_V$ denotes the addition of an external potential $V(\lambda)$ to the system. This identity therefore expresses the free energy of a system in terms of ensemble averages evaluated in the same system but with an auxiliary potential.

The idea of umbrella sampling is to divide the reaction coordinate into windows and use the auxiliary potential (typically harmonic) to restrain the system to each window, one at a time, and reconstruct the free energy there using the above identity.

Actually evaluating the relevant ensemble averages and piecing the free energy curve together is a complex process for which we use the weighted histogram analysis method (WHAM) [95] as implemented by Grossfield [96].

2.6.6.3 *Metadynamics*

The idea of metadynamics [97] is to modify the dynamics of a system by depositting small Gaussian-shaped clumps of energy to λ-space at regular time intervals. Mathematically, this involves augmenting the Hamiltonian with the time-dependent bias potential

$$V(\lambda, t) = w \sum_{k=0}^{\lfloor t/\tau \rfloor} \exp\left(-\frac{|\lambda - \lambda(x(k\tau))|^2}{2\delta\lambda^2}\right) \tag{2.111}$$

where w and $\delta\lambda$ are the height and width of the Gaussians, respectively, and τ is the time interval between depositions.

As a trajectory samples λ-space, V leaves a trail of energy along the way. The result is that if the system becomes kinetically trapped in some (meta)stable region, the potential will accumulate until it is large enough to drive the state over the free energy barrier and on to other regions of λ-space. This provides a solution to the problem of quasi-nonergodicity.

Furthermore, as the simulation progresses, V will fill up the free energy basins until the state is uniformly sampling λ-space. At this stage, V will have converged to the negative of the free energy map

$$F(\lambda) = -\lim_{t\to\infty} V(\lambda, t) \tag{2.112}$$

with a resolution determined by the Gaussian parameters. This process is illustrated in Figure 2.13.

One of the advantages of the metadynamics method is that it only requires an initial configuration and a suitable set of collective variables λ in order to sample the phase space and construct a free energy map. On the other hand, detecting convergence can be challenging,

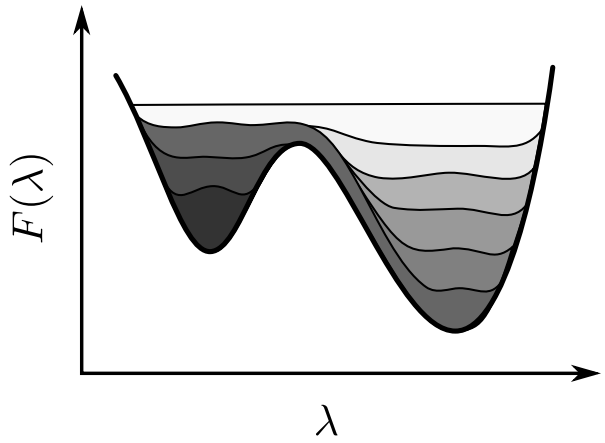


Figure 2.13: Metadynamics deposits energy in the regions of CV-space visited by a trajectory. If the trajectory becomes kinetically trapped then the auxiliary energy will accumulate and drive it over the free energy barrier. This tackles the problem of quasi-nonergodicity and allows the free energy to be mapped.

and choosing an efficient Gaussian height w requires prior knowledge of the approximate free energy barriers.

The convergence of metadynamics can be improved using well-temperating [98], whereby the Gaussian heights decrease with simulation time according to

$$w(t) = w_0 \exp\left(-\frac{V(\lambda(x(t)), t)}{k_B \Delta T}\right) \tag{2.113}$$

where w_0 is the initial Gaussian height, and ΔT is chosen to regulate the exploration of space and is typically specified in terms of the *bias factor* $\gamma = (T + \Delta T)/T$.

A trivially parallel means of accelerating metadynamics sampling is by allowing multiple independent simulations, known as multiple walkers, to contribute to a shared free energy reconstruction [99].

2.6.6.4 *Single-sweep sampling*

The single-sweep method [100] begins with a discrete set of approximately-uniform points $\{\lambda_i\}$ in λ-space and computes the mean forces $\{f_i\}$ acting at each point (cf. Section 2.6.5). The free energy surface $F(\lambda)$ is then approximated with a radial basis representation

$$F(\lambda) \approx \tilde{F}(\lambda) \equiv \sum_i \alpha_i \phi_\sigma(|\lambda - \lambda_i|) \tag{2.114}$$

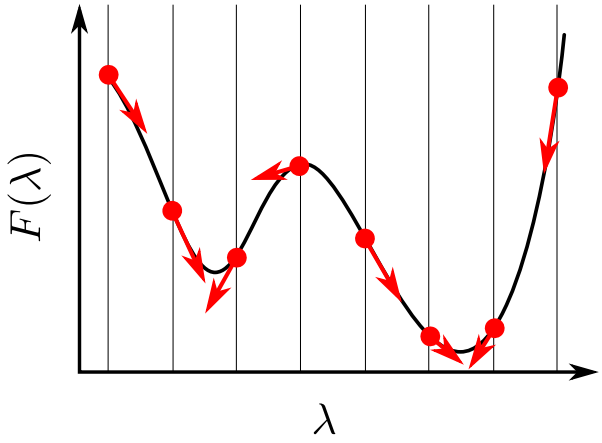


Figure 2.14: The single-sweep method fits a free energy surface, using a radial basis representation, to a discrete set of force measurements.

where $\phi_\sigma(r) = \exp(-r^2/2\sigma^2)$, and the coefficients are optimised to reproduce the forces f_i by minimising the functional

$$E_\sigma(\{\alpha_i\}) = \sum_j \left(f_j + \sum_k \alpha_k \nabla_\lambda \phi_\sigma(|\lambda_j - \lambda_k|) \right)^2 \tag{2.115}$$

as is illustrated in Figure 2.14. The Gaussian width σ is then chosen to minimise E_σ.

There are many ways to choose the points $\{\lambda_i\}$. Maragliano et al. [100] used an elevated temperature to drive the system to sample the space of interest, and then selected points from the trajectory that were separated by a minimum distance. In Chapter 3 we present an alternative version of the single-sweep method which employs a regular hexagonal close-packed lattice.

The single-sweep method is generally more efficient than metadynamics [100] and is much easier to implement than WHAM. However, it requires the user to generate, in advance, configurations that correspond to each λ_i point, which is not always easy to do.

2.7 KINETIC MONTE CARLO

Suppose that the important dynamics of a system may be reduced to a discrete set of random processes that occur at known rates $\{r_i\}(t)$. After waiting for some time interval $\Delta t(t)$, one of the events will occur in accordance with the probability distribution

$$P_i(t) = \frac{r_i(t)}{A(t)} \tag{2.116}$$

where

$$A(t) = \sum_i r_i(t) \tag{2.117}$$

is the *activity* of the system. As a consequence of this first event, the list of all possible events may be altered. A second event will then occur after some other time interval, and the sequence repeats. This is the kinetic Monte Carlo (KMC) algorithm [101]. The only remaining detail is how to compute the time interval $\Delta t(t)$ between subsequent events.

In the following identity

$$\Delta t \equiv -A^{-1} \log(e^{-A\Delta t}) \tag{2.118}$$

the exponential term $e^{-A\Delta t}$ is the (Poisson) probability that no event with a rate A will occur during the interval Δt. During the actual evolution of a system, this exponential term will stochastically adopt specific values in the range $(0, 1]$. Although the specific values are unknown to us, the same *average* dynamics may be obtained with

$$\Delta t(t) = -A(t)^{-1} \log(\mathcal{R}) \tag{2.119}$$

where $\mathcal{R} \in (0, 1]$ is a uniform random variable.

3

TRANSITIONS OF Ca^{2+}, CO_3^{2-}, AND $CaCO_3^0$ AT CALCITE STEPS

This will be the first of three chapters concerned with the modelling of calcite mineralisation. The motivation for studying calcite, as well as an introduction to the terrace-step-kink model upon which our work is based, was provided in Chapter 1. To briefly summarise the relevant points: calcite expresses two types of step, the obtuse and the acute steps. Growth proceeds by the attachment of units, whether they be calcium or carbonate ions, calcium carbonate pairs, or possibly even larger oligomers, to the steps to nucleate kink sites which then propagate. Modelling the growth of calcite therefore depends on knowing the attachment and detachment rates of the various species to these two different types of step.

This chapter begins with the introduction and validation of a new hybrid method, called dynamic single-sweep sampling, for constructing minimum free energy pathways and their associated free energy curves. This method is then applied to the two calcite steps to obtain the barriers of attachment and detachment for the individual Ca^{2+} and CO_3^{2-} ions. The results reveal that the individual ions exhibit relatively low stability at the two steps, suggesting that the system would be amenable to direct molecular simulation and therefore averting the need for further rare event sampling.

The second part of this chapter then employs direct microsecond-long molecular simulations to study the behaviour of Ca^{2+} and CO_3^{2-} ions, as well as $CaCO_3^0$ pairs, in the vicinity of the two steps. Analysis of the data reveals the metastable states, the relative attachment rates, and the absolute dissolution rates.

3.1 DYNAMIC SINGLE-SWEEP SAMPLING

Recall from Section 2.6.2 that an MFEP is a curve $\phi : [0, 1] \mapsto \mathbb{R}^n$ in an n-dimensional collective variable space that connects a pair of free energy minima along the direction of steepest descent via a saddle point. Starting with an initial guess for ϕ, the purpose of this section is to present a method that converges this curve towards the

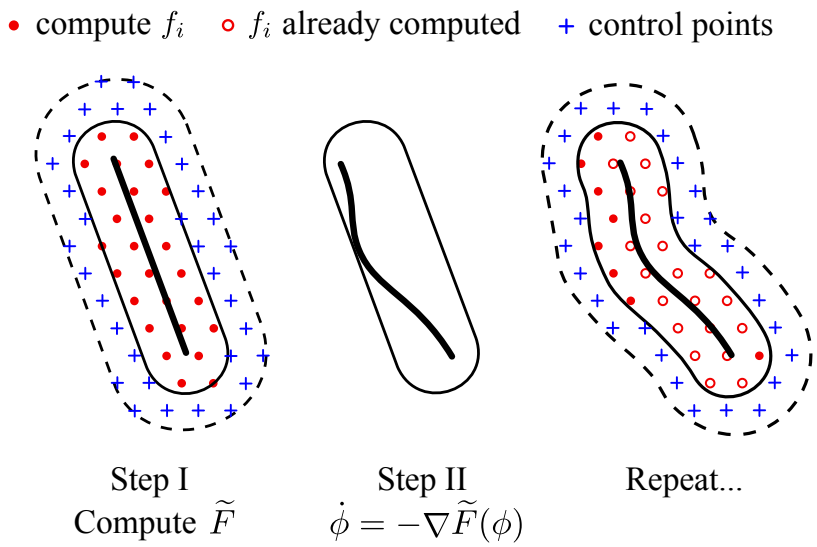


Figure 3.1: A schematic representation of our method. Step I: a local free energy map is constructed using single-sweep sampling. Step II: The string is minimised until it converges or exits the tube. The process is repeated. The circles and pluses represent control nodes. Mean forces need to be evaluated at the solid circles; forces have already been evaluated at the open circles; and the pluses serve as additional degrees of freedom.

local MFEP. It should be noted that, due to the scaling of this method, we only consider $n = 2$ and $n = 3$ dimensions.

The basic idea of how it works is to combine the string method (Section 2.6.5.1) with the single-sweep sampling method (Section 2.6.6.4) so that a free energy map is generated in the local vicinity of the string, and the string is converged within that dynamically-updated map. The advantage of this method over the standard single-sweep method is that the free energy is only sampled where it is needed. And the advantage of this method over the standard string method is that, by dynamically constructing the local free energy map, there is no need for duplicate force sampling, plus the force interpolation requires fewer force evaluations. We now proceed to outline the algorithm.

3.1.1 *Initialisation*

A discrete set of points $\{z_i\} \subset \mathbb{R}^n$ must be constructed within the volume of collective variable space that may be explored; it is over these grid points that the single-sweep sampling will be performed. We use a hexagonal close-packed (HCP) grid with lattice parameter a for

these points since it is the most dense packing of spheres in both two and three dimensions. Close-packing is desirable since it maximises the coverage and smoothing effect of the single-sweep method. Note, however, that a regular grid is not essential. A partially random array of points generated by temperature-accelerated molecular dynamics [102], as applied in [103], would also be suitable.

The MFEP ϕ is approximated by a discrete set of points $\{s_i\} \subset \mathbb{R}^n$ which we refer to as the string. As an initial choice, it is standard to construct a linear string that connects the basins of the two minima of interest. It is not necessary for these endpoints to exactly fall on the minima since their positions will be locally minimised during the process.

3.1.2 *Step I: local free energy map*

The first step of the algorithm is to construct the free energy map within the vicinity of the string. This will involve taking a subset of the grid within some radius of the string and then using the single-sweep method to generate a local free energy map.

More rigorously, given a string $\{s_i\}$ and grid $\{z_i\}$, we begin by defining the following sets for the two indices $k = 1, 2$:

$$I_k = \{i : \min_j |z_i - s_j| \leqslant R_k\} \tag{3.1}$$

The set I_k indexes the grid points that fall within a distance R_k of the string. We discuss the choice of R_1 and R_2 below but note for now that $R_1 < R_2$.

The next step is to compute the mean force f_i, or load a previously computed f_i, associated with each point in I_1. Evaluating the mean force at some point in collective variable space is described in Section 2.6.5

Utilising the single-sweep technique, the true free energy map F may be approximated within the vicinity of the string with a radial basis representation, denoted by a tilde

$$\widetilde{F}(z) = \sum_{i \in I_2} \lambda_i \psi_\sigma(|z - z_i|) \tag{3.2}$$

where $\psi_\sigma(r) = \exp(-r^2/2\sigma^2)$. In order to reconstruct F, the λ_i coefficients are optimised to reproduce the measured forces f_i by minimising the functional

$$E(\{\lambda_k\}) = \sum_{i \in I_1} \left(f_i + \sum_{j \in I_2} \lambda_j \nabla \psi_\sigma(|z_i - z_j|) \right)^2 \tag{3.3}$$

In principle, the Gaussian width σ should also be optimised, but we invariably found $\sigma = 1.5a$ to be suitable for an HCP grid.

Minimising E involves solving an underdetermined ($|I_1| < |I_2|$) system of linear equations, $\nabla_{\{\lambda_k\}} E = 0$. This is a standard numerical optimisation problem for which we employ the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno [104] algorithm in our implementation.

Note that the inner-sum in Eq. (3.3) is over the I_2 index set. The reason for including the I_2 grid points in this part of the analysis is that the I_1 grid points alone provide too few degrees of freedom to attain an accurate reconstruction of the free energy map within that volume. The outer radius, for which we recommend $R_2 = R_1 + \sigma$, is therefore used as a source of unconstrained control points that improve the reconstruction.

The optimal choice for R_1 depends not only on the free energy map and the initial string but also on the number of processors. However, we recommend, as a rule of thumb, that the smallest viable value of $R_1 = 1.5a$ be used. And therefore $R_2 = 3a$.

3.1.3 *Step II: string minimisation*

Having constructed the local free energy map $\widetilde{F}$, the string must be minimised using one of the various string minimisations procedures. We opt for the string method (SM from here on), described in Section 2.6.5.1 and summarised here in the new context.

At the outset, a backup of the string is to be made: $s_i^0 = s_i$ for all i. The SM proceeds by repeated application of the following two steps:

1. Each point s_i is integrated downhill using the forward Euler method,

$$s_i \leftarrow s_i - h\nabla\widetilde{F}(s_i), \tag{3.4}$$

   for a sufficiently small step size h, where $\nabla\widetilde{F}$ is evaluated using Eq. (3.2).

2. The string is reparametrised. We do this by fitting a cubic spline [105] to the new points s_i and then interpolate them uniformly along the curve. This eliminates the need for a tangential force to retain the integrity of the string. Indeed, the redistribution of the points may reverse the movement of the points tangentially to the spline but not normally.

These two steps are repeated until the string either moves by more than R_1

$$\max_i |s_i - s_i^0| > R_1 \tag{3.5}$$

or achieves convergence. To detect convergence, we compute the maximum gradient acting on the string, $\max_i |\nabla \tilde{F}(s_i)|$, at each iteration. If the smallest value obtained for this quantity is not improved upon within 100 consecutive iterations then convergence is achieved and we return to Step I. This may seem an uneconomical convergence criterion, but recall that the forces are calculated directly from our reconstructed free energy map and are therefore inexpensive.

3.1.4 *Termination*

Steps I and II are repeated successively until no new mean forces f_i need to be evaluated, i.e. all of the required forces have already been computed in previous iterations. In this event, the string will have converged to the MFEP and the algorithm terminates. The free energy along the MFEP may then be recovered without any further sampling by evaluating Eq. (3.2) along the MFEP.

3.1.5 *Müller-Brown test system*

The Müller-Brown potential [106] (MB2 for short) is a two-dimensional analytic surface that is widely employed to test the accuracy and efficiency of minimum energy path algorithms. It has the functional form

$$\begin{aligned} V(x,y) = \sum_{k=1}^{4} A_k \exp[a_k(x-x_k^0)^2 + \\ b_k(x-x_k^0)(y-y_k^0) + \\ c_k(y-y_k^0)^2] \end{aligned} \tag{3.6}$$

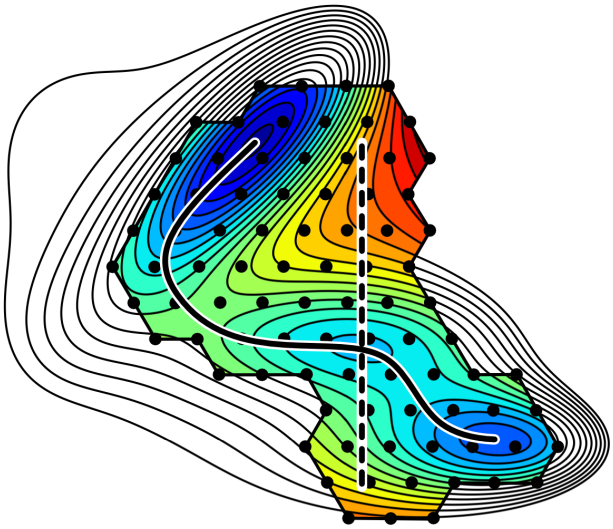


Figure 3.2: A contour plot of the Müller-Brown potential (MB2) with our initial string (dashed) and converged MFEP (solid). The circles represent the points at which the mean forces are evaluated (for $a = 0.2$) and the colour shows the region in which the free energy map is reconstructed.

where

$$A = (-200, -100, -170, 15) \tag{3.7}$$
$$a = (-1, -1, -6.5, 0.7) \tag{3.8}$$
$$b = (0, 0, 11, 0.6) \tag{3.9}$$
$$c = (-10, -10, -6.5, 0.7) \tag{3.10}$$
$$x^0 = (1, 0, -0.5, -1) \tag{3.11}$$
$$y^0 = (0, 0.5, 1.5, 1) \tag{3.12}$$

Since we are interested in both $n = 2$ and $n = 3$, we have constructed a three-dimensional analogue of the the Müller-Brown potential (called MB3) which takes the form

$$\begin{aligned} V(x,y,z) = \sum_{k=1}^{4} A_k \exp[a_k(x-x_k^0)^2 + \\ b_k(x-x_k^0)(y-y_k^0) + \\ c_k(y-y_k^0)^2 + \\ d_k(z-z_k^0)^2] \end{aligned} \tag{3.13}$$

where

$$d = (-8, -8, -8, 0.7) \tag{3.14}$$
$$z^0 = (0, 0.5, 0, 0) \tag{3.15}$$

and the other parameters are as above.

To test the accuracy of our method we constructed a linear string of 50-images from $(0, -0.2)$ to $(0, 1.5)$ and evolved it in MB2 using our dynamic single-sweep method to obtain the MFEPs, $\{s_i(\alpha)\}$, for a range of grid densities, $0.2 \leqslant \alpha \leqslant 0.4$. To quantify the accuracy of each string we defined the following dimensionless metric

$$\epsilon_s(\alpha) = \alpha^{-1} \max_i \min_{t\in[0,1]} |s_i(\alpha) - \phi(t)| \tag{3.16}$$

which expresses the distance between $\{s_i(\alpha)\}$ and the exact MFEP ϕ. We also defined

$$\epsilon_F(\alpha) = \frac{\max_{i,j} |\Delta\tilde{F}_{ij}(\alpha) - \Delta F_{ij}(\alpha)|}{\max_{i,j} |\Delta F_{ij}(\alpha)|} \tag{3.17}$$

where we use the shorthand

$$\Delta\tilde{F}_{ij}(\alpha) = (\tilde{F}(s_i(\alpha)) - \tilde{F}(s_j(\alpha))) \tag{3.18}$$

and similarly for $\Delta F_{ij}(\alpha)$. This metric expresses the largest error in the free energy differences of the reconstructed free energy curve as a fraction of the largest free energy difference. To put it another way, ϵ_F will typically provide an upper bound on the fractional error of any computed free energy barriers. Figure 3.3(a) shows the measured errors, ϵ_s and ϵ_F, as a function of the grid parameter α. Identical curves were obtained for $n = 3$ (MB3). It can be seen that for a sufficiently dense grid (e.g. $\alpha = 0.2$), our method is able to construct a highly accurate MFEP ($\epsilon_s = 4\%$, $\epsilon_F = 2\%$).

To gauge the computational efficiency of our method, we have pitted it against the SM. To do this, we first picked a grid parameter $0.2 \leqslant \alpha \leqslant 0.4$ and then chose the following three variables so as to give the SM the best possible chance against it:

- The step size h was invariably set to the largest stable value, $h = 4 \times 10^{-4}$.
- The fewest number of images were used that could match the accuracy (the ϵ_F error) of our method at the present α value. For example, 30 images were required for $\alpha = 0.2$ but only 8 for $\alpha = 0.4$.
- The string converged as soon as its coordinates were within a reasonable 10^{-3} of the exact string. In MB2 (MB3), the strings typically took 60 (100) iterations to converge. By comparison, the convergence criterion of the SM employed by [107] required 100 iterations in MB2. Note also that stochastic systems (which this is not) tend to require even longer to detect convergence.

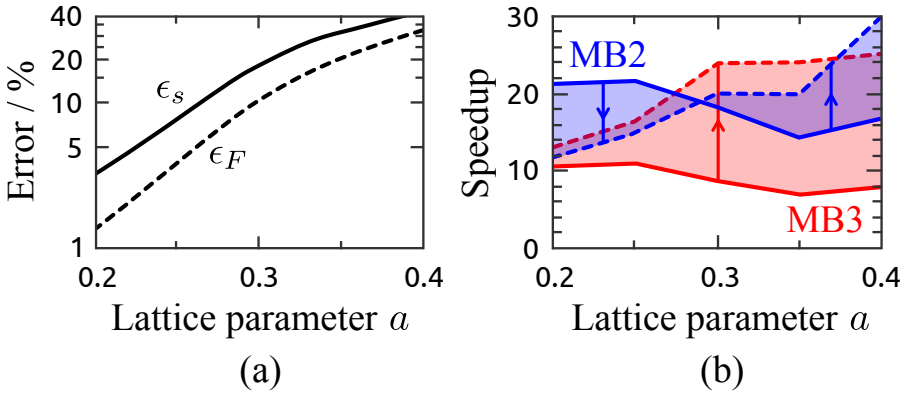


Figure 3.3: (a) The two error terms, ϵ_s and ϵ_F, for a range of grid densities. These curves apply equally to MB2 and MB3. (b) The speedup that our method offers over the standard string method for the two- and three-dimensional Müller-Brown surfaces, MB2 and MB3. The solid lines show the speedup achieved on a single processor and the dashed line on an infinite number of processors.

In order to achieve a fair comparison between the run-times of the two methods, a relative wall time $T(P;a)$ was assigned to each method as a function of both the grid parameter a and the (hypothetical) number of processors P available to it. P is relevant since both methods parallelise differently. This function is given by

$$T(P) = \sum_{i=1}^{M} \lfloor m_i/P \rfloor t(1) + t(\lfloor P/\mathrm{mod}(m_i, P) \rfloor) \tag{3.19}$$

where M is the total number of iterations, m_i the number of mean force evaluations at the i-th iteration, and the second term is only evaluated if $\mathrm{mod}(m_i, P) \neq 0$. The $t(p)$ function is the (relative) time required to evaluate a single mean force on p processors. To makes this representative of real molecular dynamics force calculations, we limit p to between 1 and 8 processors, and assume that the force evaluations scale as $t(p) = 100/(8 + 96p - 4p^2)$ which reproduces the scaling of the MPI-based DL_POLY Classic molecular dynamics package.

The speedup of our method over the SM is shown in Figure 3.3(b) as a function of a, ranging from $P = 1$ processor (solid line) to $P = \infty$ (dashed line), for both MB2 (blue) and MB3 (red). Evidently our method is an order of magnitude faster, with the speedup ranging from $7\times$ to $30\times$. This equates to the difference between one day and one-to-four weeks.

On a single processor ($P = 1$) the relative speed equals the ratio of the number of mean force evaluations required by the two methods.

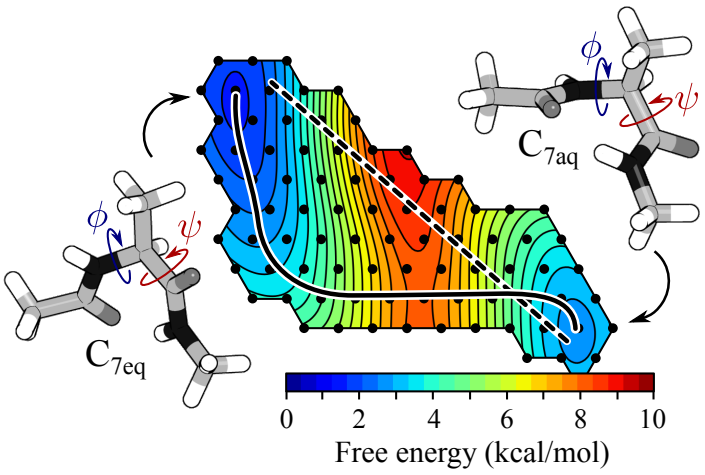


Figure 3.4: The free energy surface constructed for alanine dipeptide in vacuum. The circles show the points at which the mean forces were computed ($a = 15°$). The dashed line is the initial string and the solid line the converged MFEP which connects the two isomers shown, C_{7eq} and C_{7aq}.

As P increases, the speedup approaches the ratio of the number of iterations performed by the two methods. These two ratios therefore determine the relative scaling of the methods and explain why, in the case of MB2, the speedup for $a \lesssim 0.3$ decreases as the number of processors increases.

3.1.6 *Alanine dipeptide isomerisation validation*

Alanine dipeptide (ACE-ALA-NME) is a simple biomolecule of scientific importance since its conformational properties influence peptide and protein secondary structure. Having a relatively large barrier between its two main metastable states, C_{7eq} and C_{7aq}, it is widely employed (e.g. [89, 107, 108, 109, 110]) as a model for benchmarking methods that enhance rare event sampling. It is a more realistic test model than the Müller-Brown model above since the forces must be evaluated from molecular simulations and are therefore stochastic. We have therefore applied our dynamic single-sweep method to compute the MFEP between these two isomers in vacuum.

As collective variables we used the two Ramachandran dihedral angles, ϕ and ψ, which are the dihedral angles depicted in Figure 3.4, although they are known to provide an inexact transition mechanism [89]. The simulations were performed with the DL_POLY Classic 1.9 [111] molecular dynamics code and the all-atom parm99 AMBER po-

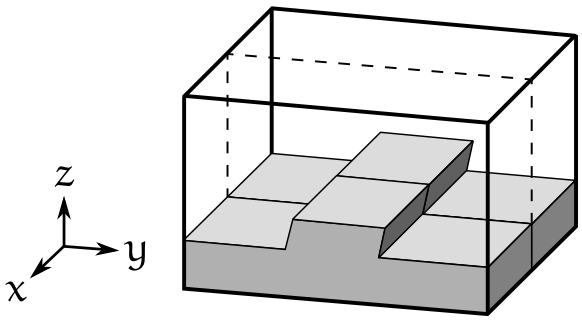


Figure 3.5: A schematic of the simulation cell used in the application of the dynamics single-sweep method to calcium and carbonate ions at the calcite steps. The calcite slab is shown in grey; water filled the volume above.

tentials [112]. We used a time-step of 1 fs and a velocity-rescaling thermostat to maintain a temperature of 300 K. The mean forces were obtained by imposing a harmonic potential with a force constant $k = 1000$ kcal/mol/rad^2 to restrain the dihedrals, and then averaging the restraining force over 250 ps. The initial string was a linear path connecting $(-70, 60)$ to $(60, -40)$ and the grid parameter was $a = 15°$.

The algorithm converged after only 5 iterations and 85 force evaluations. By comparison, the string methods implemented in [89, 107, 108] required at least 100 iterations and 2,000 mean force evaluations, despite starting with very similar initial strings. Our method is therefore at least $20\times$ faster, in line with the speedups reported in Figure 3.3.

To ensure that $a = 15°$ produced a sufficiently accurate MFEP, we repeated the method with a much denser grid ($a = 5°$). The coordinates of the resulting MFEP agreed to within $4°$ and the free energy barrier to within 4%.

3.2 Ca^{2+} AND CO_3^{2-} FREE ENERGIES AT THE STEPS

Having introduced and validated our dynamic single-sweep method, the next step is to apply it to the dissolution of calcium and carbonate ions from the acute and obtuse steps of calcite. More specifically, this section aims to compute the dissolution pathway and associated free energy curve for the two ions moving from the fully adsorbed states along the two steps into the bulk solution.

Each simulation cell consisted of a slab of calcite with the $(10\bar{1}4)$ faces exposed and two partial layers on either side, each exposing

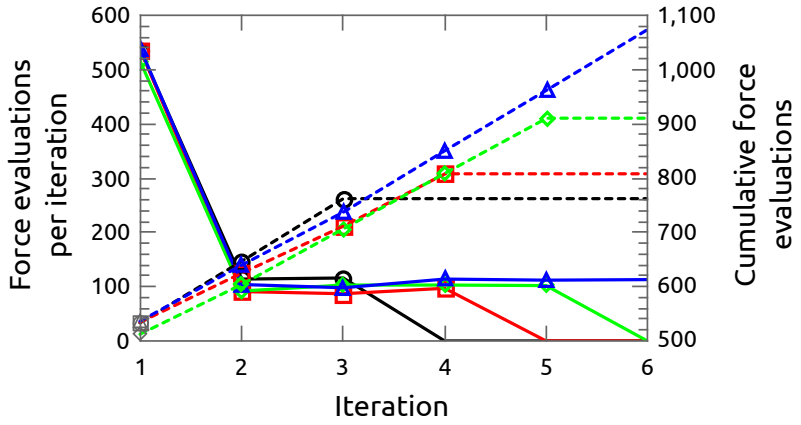


Figure 3.6: The solid lines show the number of force evaluations performed at each iteration of our dynamic single-sweep method, while the dashed lines show the cumulative number of force evaluations after each iteration. Black circles corresponds to A/Ca, red squares to A/CO_3, green diamonds to O/Ca, and blue triangles to O/CO_3.

the acute and obtuse steps, as depicted in Figure 3.5. The dimensions of the simulation cell were $2.55 \times 3.12 \times 3.40$ nm, with 112 $CaCO_3$ units in the calcite slab, and 460 water molecules filling the void between the periodic images. The height of the cell was adjusted to achieve a pressure of 1 atm at a temperature of 300 K. The force fields and simulation parameters are detailed in Appendix A.

We began with each ion in a fully adsorbed state, i.e. bound directly to both the step and the underlying layer, for each type of step, giving a total of four systems. We denote the system with the Ca^{2+} ion adsorbed to the acute and obtuse steps as A/Ca and O/Ca, respectively, and similarly for the carbonate CO_3^{2-} ion, A/CO_3 and O/CO_3.

Recall that the dynamic single-sweep method samples forces on a predefined grid. In this case, the grid was HCP with a parameter of 0.2 Å. It filled the entire simulation cell, except any grid points that were within a distance of 2.5 Å from the calcium carbonate step were culled to avoid sampling regions that would involve either atomic overlap or the destruction of the slab.

The initial string in each case was linear and passed from the initial adsorbed state directly into the bulk solution, aligned normal to the $(10\bar{1}4)$ surface. Each of the two ends of the string were fixed in place. This was of particular importance for the string end placed in the bulk solution where the free energy will have plateaued; without freezing this end, it would drift randomly under the stochastic mean forces. The dynamic single-sweep method was then applied to converge the

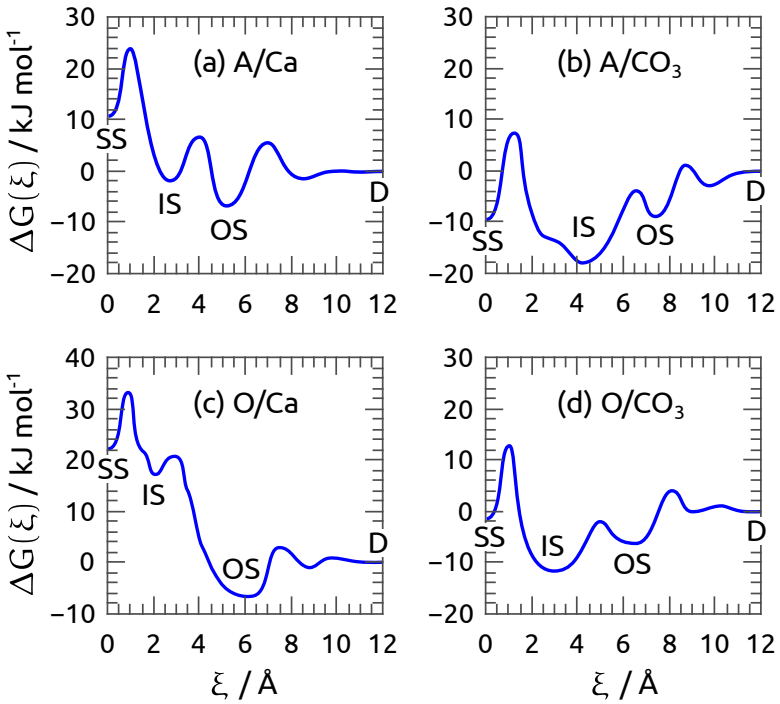


Figure 3.7: The free energy curves for the dissolution of the Ca^{2+} and CO_3^{2-} ions from the acute and obtuse steps, generated using our dynamic single-sweep sampling method. The reaction coordinate ξ is the length along the MFEP.

string towards the local MFEP. For each of the four simulations, the number of force evaluations performed per iteration are shown in Figure 3.6 along with the cumulative number of force evaluations. The number of iterations ranged from 3 to 6, with a total of between 765 and 1,079 force evaluations. We note that most of the force evaluations are performed on the first iteration, which is characteristic of this method.

The free energy curves along the resulting MFEPs are shown in Figure 3.7. In each case, three metastable states are obtained. These three states are similar to one another and can be characterised as follows:

- Step site (SS): the ion is bound directly to the step as well as to the layer below. This is the state of maximal adsorption which corresponds closely to the equilibrium grid site in the bulk calcite.
- Inner-sphere (IS): the ion has broken the direct bond with the underlying layer but remains directly bound to the step, i.e. the ion is within the step's solvation shell.

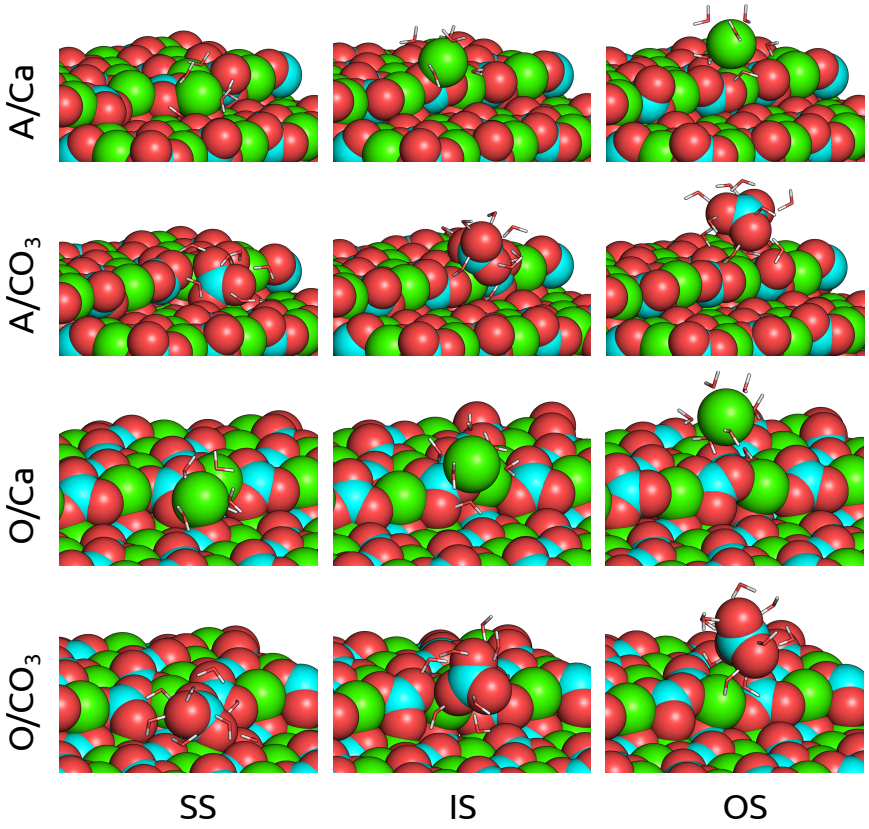


Figure 3.8: The SS, IS, and OS states for the Ca and CO_3 ions at the acute (A) and obtuse (O) steps with only the water molecules in the first solvation shell of the adsorbate shown.

- Outer-sphere (OS): the ion has broken all direct bonds with the calcite but remains attached to the step via its solvation shell.

All three of these states are shown for each of the four systems in Figure 3.8.

In the case of the calcium ion, the free energy is at a minimum, albeit very slightly, when bound indirectly to the step in the outer-sphere (OS) adsorption state. However, this state exhibits very limited kinetic stability. The SS and IS adsorption states, on the other hand, are associated with a substantial increase in free energy and will seldom be visited. But even when they are, the calcium ion will not remain for long due to the limited kinetic stability. The reason that the calcium ion averts direct adsorption to the step is because it has a high charge density and therefore a strong solvation shell which makes it naturally favour states of increased hydration.

In contrast to the calcium ion, the carbonate ion minimises the free energy when bound directly to the step. However, it favours the IS

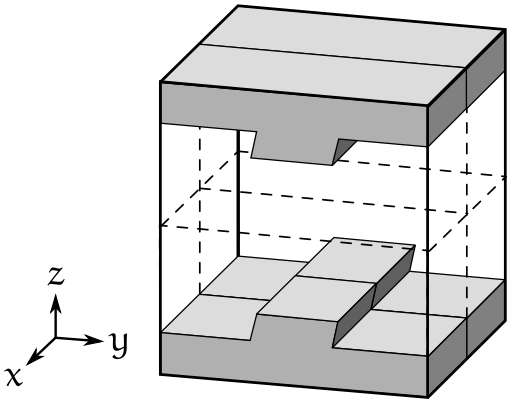


Figure 3.9: The simulation cell used for long duration MD simulations, consisting of a pair of calcite surfaces, with steps exposed, separated by water. The four quadrants shown are symmetrically equivalent.

rather than the SS state, presumably also due to the increased solvation afforded to it in this state. Similar to the calcium ion, however, the carbonate is presented with considerably large barriers to fully adsorb to the SS state suggesting that it will be a rare occurrence. Moreover, even when it does reach SS, the dissolution barrier is little greater than that in the IS state suggesting that the dissolution rates from SS and from IS will be the same order of magnitude.

The next step is to establish the actual dissolution rates. Since the free energy curves computed in this section reveal that the problem of computing transition rates may be amenable to direct simulation rather than requiring the application of rare event methodologies, this is the approach taken in the next section.

3.3 DIRECT MICROSECOND-SCALE SIMULATIONS

To gain insight into the behaviour of the calcium carbonate ions at the calcite steps, we performed molecular dynamics simulations in which a molecular unit, either a calcium ion Ca^2, a carbonate ion CO_3^{2-}, or a calcium carbonate pair $CaCO_3^0$, were placed in the centre of the simulation box and then allowed to naturally diffuse through the volume: diffusing, adsorbing to one of the various sites, desorbing, and repeating. In the case of the $CaCO_3^0$ pair, the distance between the two ions was monitored and they were found to remain bonded throughout the simulations without the need for an auxiliary potential to hold them together. The simulation cell is depicted in Figure 3.9 and

consisted of an inverted (and translated) version of the previous simulation cell. The cell could be divided into four quadrants that were symmetrically identical.

During each simulation, the position of each subject unit was recorded every picosecond. The simulations involving the individual Ca^{2+} or CO_3^{2-} ions were each terminated after 1 µs, while the $CaCO_3^0$ unit was afforded 4 µs due to it diffusing at a lower rate.

A histogram representing the probability distribution for each of the three units could then be generated from the recorded trajectories. Recall that in the simulation cell, shown in Figure 3.9, the cell exhibits four-fold symmetry. As such, the trajectory may be mapped into just one quadrant of the cell by application of the appropriate symmetry operation. This provides a smoother and more thoroughly sampled histogram than would otherwise be produced.

Observing that the probability of the unit being in a position (x, y, z) is related to its free energy map $G(x, y, z)$ by $P(x, y, z) \propto \exp(-\beta G(x, y, z))$, an approximate free energy map may be constructed from the histogram. To visualise it we project the map onto a two-dimensional plane by integrating out the third coordinate, e.g.

$$G(x, y) = -\beta^{-1} \log \int \exp(-\beta G(x, y, z))\, dz \tag{3.20}$$

In the case of the projection onto the xz-plane, we actually provide two projections, one for $y < 0$,

$$G_{y<0}(x, z) = -\beta^{-1} \log \int_{y<0} \exp(-\beta G(x, y, z))\, dy \tag{3.21}$$

and an analogous function for $y > 0$. The reason for this division is that the $y < 0$ volume contained the acute step while $y > 0$ contained the obtuse step. To present each projection, we imagine unfolding the net of the sampled volume as shown in Figure 3.10. These projections are then presented in Figures 3.11 and 3.12 for the Ca^{2+} and CO_3^{2-} ions, respectively. To visualise the $CaCO_3^0$ pair, we rendered the free energy maps twice, from the point of view of the calcium ion and the carbonate ion separately, as shown in Figures 3.13 and 3.14, respectively. These plots provided much more structure and clarity than simply plotting the free energy map as a function of the centre of mass of the pair.

It is important to note that two coordinates are not enough to capture the state of the subject units and therefore the actual free energy differences and barriers cannot be read from these plots. Nevertheless, the plots do provide insight into the structure of the free energy

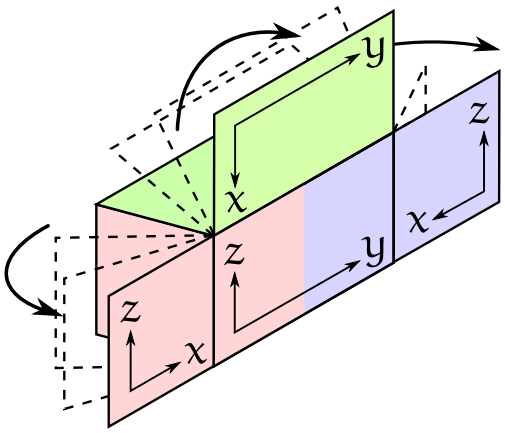


Figure 3.10: The correspondence between the two-dimensional free energy projections and the three-dimensional quadrant as utilised in the images that follow.

maps, as detailed below. It should also be noted that the features visible in these histograms are also visible if only half of each trajectory is binned. In other words, they are not merely noise but reflect real features of the free energy landscape.

The novelty of performing direct simulations is that, in addition to generating the histograms, the various transition rates may be obtained directly by assigning the subject ion to one of the states of interest (either acute or obtuse, either SS, IS, OS, or dissolved) at each point along its trajectory, based on its position, and then measuring the mean time between each transition. Furthermore, the relative attachment rate for each adsorbate to each of the four sites is given by the relative probability P that the adsorbate will bind to those sites. This was straightforwardly computed by simply counting the number of adsorption events to each site and then normalising. The results are reported in Table 3.1.

by counting the number of times the adsorbate visits each site, we can compute the relative attachment rate for the various sites; these are given by the

In the free energy map shown in Figure 3.11, it can be seen that the calcium ion is predisposed to avoid the calcite surface, with a slight reduction in free energy on approach, even several Ångström away from the calcite, due to the presence of water structure. Upon reaching the calcite, the calcium ion visits the OS site for both steps multiple times, and in the case of the acute step also transitions to the IS state. As predicted in the previous section, however, the barrier to the IS state at the obtuse step is simply too large, and the ion does not

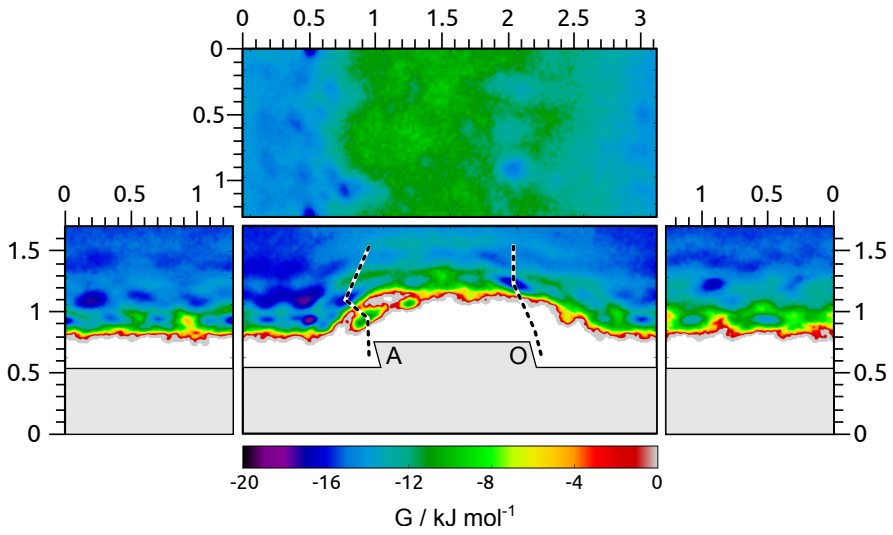


Figure 3.11: Two-dimensional projections of the approximate free energy map for the Ca^{2+} ion, in the vicinity of the calcite steps. See Figure 3.10 for an explanation of what each panel corresponds to. The axes along the perimeter show the scale in nanometres.

visit it once. We computed the transition rates between the adsorbed states and the dissolved states. In the case of the acute step, we treated the IS and OS states as a single state. As reported in Table 3.1, the calcium ion spends on the order of 100 ps adsorbed to either step before dissolving again, and shows a slight preference for adsorbing to the acute step over the obtuse.

The carbonate ion results in a vastly different free energy map (Figure 3.12) to the calcium ion. Firstly, the carbonate ion prefers to be in the vicinity of the calcite rather than in bulk solution. This makes sense because, in contrast to the calcium ion, the carbonate has a lower charge density and therefore weaker solvation shells. It visits both the outer- and inner-sphere adsorption states of both steps, dozens of times. The stability of the carbonate ion turns out to depend quite strongly on which site along the step it adsorbs to, characteristic of a non-Kossel crystal. The consequence is that one of the sites along the acute step has a highly stable IS state while the other has a much less stable state. The obtuse step, on the other hand, displays very similar stability at both available sites. Regarding the transition rates, we measure the mean time for the carbonate ion to completely dissolve from the IS state (in other words, back and forth transitions between the IS and OS states are discounted). It can be seen in Table 3.1 that the carbonate ion adsorbs for around 2 ns to the obtuse step while it remains for an average of around 12 ns when adsorbed to one of the

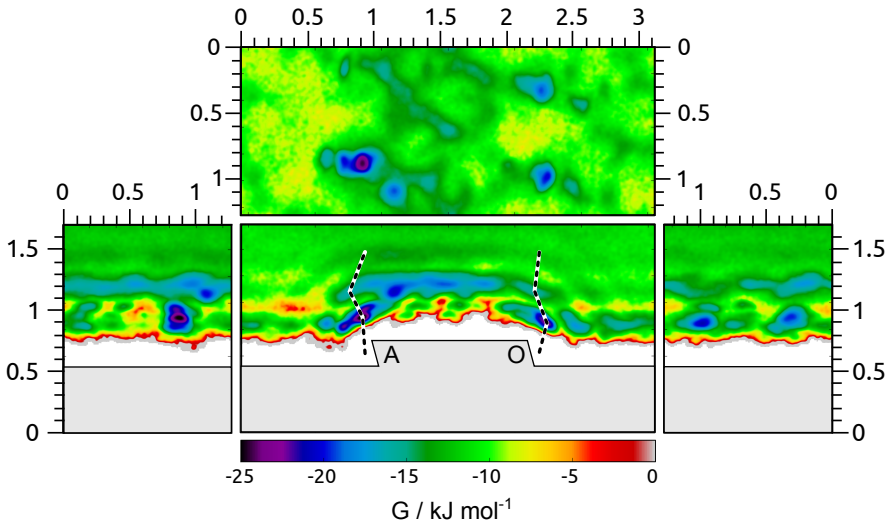


Figure 3.12: Two-dimensional projections of the approximate free energy map for the CO_3^{2-} ion, in the vicinity of the calcite steps. See Figure 3.10 for an explanation of what each panel corresponds to. The axes along the perimeter show the scale in nanometres.

acute sites. Interestingly, however, it shows no statistically significant preference to adsorb to any particular site.

Lastly, we consider the $CaCO_3^0$ pair, for which the free energy maps are shown in Figures 3.13 and 3.14, plotted from the point of view of the calcium ion and carbonate ion, respectively. Interestingly, the free energy map plotted from the perspective of the carbonate ion in the pair (Figure 3.14) is almost identical to that for the lone carbonate ion CO_3^{2-} shown in Figure 3.12, suggesting that the calcium carbonate

Table 3.1: The mean times r^{-1} for dissolution, and the relative probabilities P of the molecules adsorbing to each site. The 1 and 2 refer to the two different sites on each step as labelled in Figures 3.11-3.14.

		Obtuse		Acute	
		1	2	1	2
Ca^{2+}	r^{-1} (ns)	0.17	0.10	0.27	0.07
	P ($\pm$0.02)	0.21	0.18	0.19	0.42
CO_3^{2-}	r^{-1} (ns)	2.05	2.76	12.46	1.01
	P ($\pm$0.04)	0.27	0.23	0.24	0.25
$CaCO_3^0$	r^{-1} (ns)	2.27	26.14	29.21	3.27
	P ($\pm$0.03)	0.28	0.18	0.33	0.21

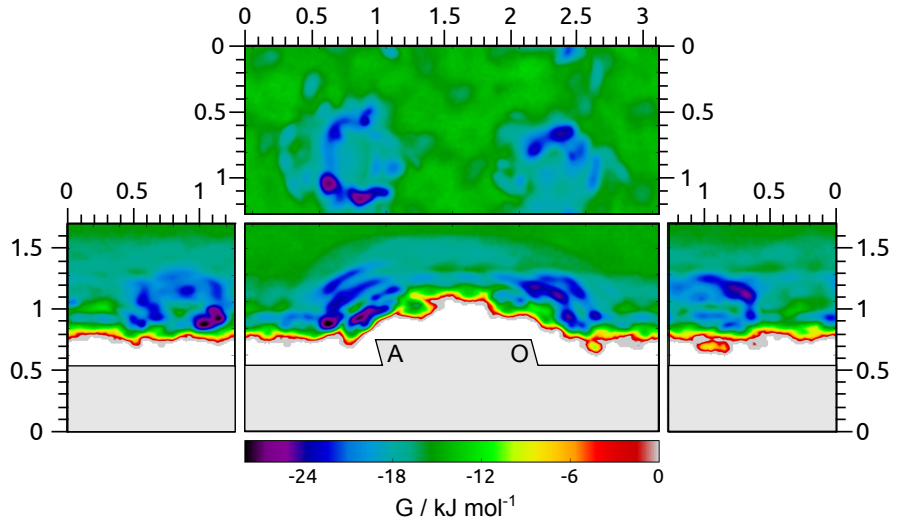


Figure 3.13: Two-dimensional projections of the approximate free energy map for the $CaCO_3^0$ pair, from the perspective of the calcium ion, in the vicinity of the calcite steps. See Figure 3.10 for an explanation of what each panel corresponds to. The axes along the perimeter show the scale in nanometres.

pair behaves much like a lone carbonate ion. However, there are a few key differences that arise from having the calcium ion attached:

1. The stability is increased for all adsorption sites.
2. While the CO_3^{2-} ion exhibits about the same stability at the two IS states on the obtuse step, the $CaCO_3^0$ pair breaks the symmetry and exhibits a much greater stability at just one of the sites on the obtuse step.
3. Consequently, the $CaCO_3^0$ pair displays almost the exact same stability at the acute and the obtuse steps.
4. However, the relative attachment rates suggest that the pair is slightly more likely to attach to the highly stable site on the acute step, while it is more likely to attach to the less stable site on the obtuse step.

The calcium component of the $CaCO_3^0$ pair, on the other hand, was not as disciplined as the carbonate. Even when the carbonate component was attached to a step, the calcium component spent considerable time pointing away from the step in a solvated state. However, it did also exhibit metastability attached to both steps in an OS state, and also in an IS state for the acute step.

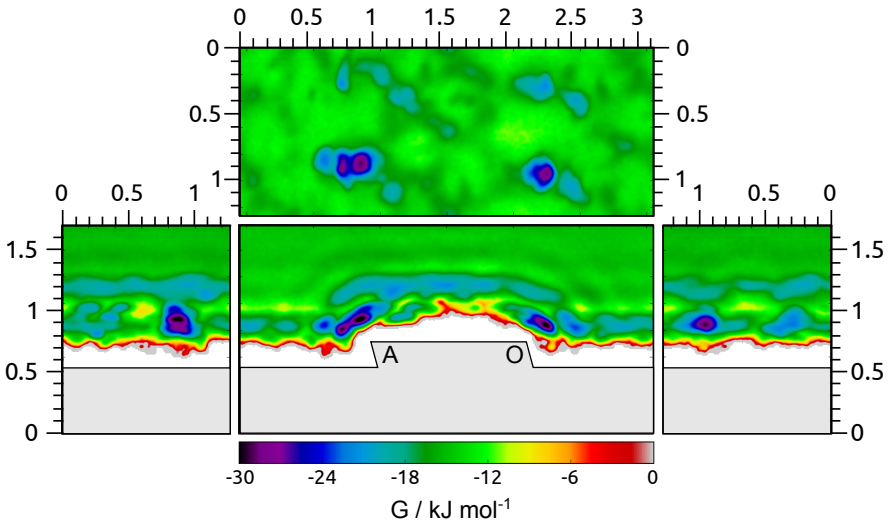


Figure 3.14: Two-dimensional projections of the approximate free energy map for the $CaCO_3^0$ pair, from the perspective of the carbonate ion, in the vicinity of the calcite steps. See Figure 3.10 for an explanation of what each panel corresponds to. The axes along the perimeter show the scale in nanometres.

In all cases, the ions failed to visit the SS state due to the large kinetic barriers. This suggests that either the final IS→SS transition limits the step nucleation rate, or that step nucleation requires multiple ions to aggregate in the IS state before being able to make the IS→SS transition.

3.4 SUMMARY

This chapter began by introducing a new hybrid method, named dynamic single-sweep sampling, for generating minimum free energy pathways and the associated free energy curves. It involves performing an exact string minimisation within an approximate free energy map that is dynamically constructed within the vicinity of the string. Standard validation tests found it to perform an order of magnitude faster than the standard string method.

This new method was then employed to model the dissolution of calcium Ca^{2+} and carbonate CO_3^{2-} ions from the fully-adsorbed sites at the acute and obtuse steps. In all four systems, three metastable states were observed, corresponding to the fully-adsorbed step site (SS), inner-sphere (IS) adsorption to the step, and outer-sphere (OS) adsorption to the step. The free energy curves revealed large barriers going from IS to SS suggesting that the ions will seldom reach

the SS sites. Moreover, in the case of the calcium ion, the OS state was found to be thermodynamically preferred, while the IS state was thermodynamically prefered for the carbonate ion.

We then turned our attention to performing direct, microsecond-long simulations of each ion moving in the vicinity of the two steps. In fact, in addition to the Ca^{2+} and CO_3^{2-} ions, we also simulated the behaviour of the $CaCO_3^0$ pair. The resulting probability histograms and measured transition rates were in accord with the predictions made from the dynamic single-sweep study. Specifically, none of the ions visited the SS sites at any point. The calcium ions mostly attached via outer-sphere adsorption and dissolved after very short intervals, on the order of 100 ps. The carbonate ions were most stable in the IS states, exhibiting greater stability at the acute step than the obtuse. They were found to dissolve from the steps with mean times on the order of 1-2 ns except for one particular site on the acute step where they would remain for an average of 12 ns.

In the case of the neutral $CaCO_3^0$ pair, the resulting probability histograms revealed it to behave a great deal like a single carbonate ion. The main effect of the attached calcium ion was to increase the stability of the various metastable states. But in doing so, it equalised the acute and obtuse steps such that, while the lone carbonate ion is much more stable at the acute step, the ion pair shows equal stability at the two steps.

Based on these results, it will be argued in Chapter 5 that kink nucleation during calcite growth will be dominated by the adsorption of the $CaCO_3^0$ pairs to the steps, and that the activity of the individual calcium and carbonate ions may be neglected from models of kink nucleation.

4

CALCITE KINK DISSOLUTION

Chapter 1 provided an account of calcite growth in terms of kink nucleation and kink propagation. The previous chapter, Chapter 3, investigated kink nucleation, while kink propagation will be the subject of this chapter. More specifically, the rate at which the calcite kink sites dissolve and the underlying mechanisms.

However, since it would not be practical to apply the requisite methodologies to all 16 kink sites, the first step will be to compute the dissolution enthalpies instead, and then identify just two kink sites that will likely limit the dissolution process. An assortment of rare event methods will then be applied to those two ostensibly rate-limiting kink sites to obtain their dissolution rates.

Each kink site in this chapter was modelled with a periodic slab of calcite that was oriented with respect to the unit cell such that an incomplete top layer, with a pair of steps exposed, necessarily formed a zigzag pattern in order to achieve periodicity. Kink sites were therefore naturally exposed and separated from each other by the width of the unit cell. This setup can be seen in Figure 4.1, and a total of 8 such models were constructed, exposing all 16 unique kink sites. In our simulations, each slab consisted of 143 $CaCO_3$ units. The unit cells had dimensions of $2.19 \times 2.91 \times 3.54$ nm, with the remaining non-calcite volume filled with 480 water molecules. The total momentum of the slab was fixed to zero to prevent it from drifting. The cell height had been adjusted to achieve an average pressure of 1 atm at a temperature of 300 K. This was achieved by applying a barostat (see Section 2.5.3) with a time-constant of 1 ps but to the cell height only and averaging the cell height. The configurations were then equilibrated for 10 ns. The force fields and simulation parameters are detailed in Appendix A.

4.1 DISSOLUTION ENTHALPIES

In order to compute the change in enthalpy during dissolution of the target kink site, two configurations were modelled. The first involved the terminal ion in the adsorbed kink state, and the second with the ion dissolved. This second configuration was created by freezing the

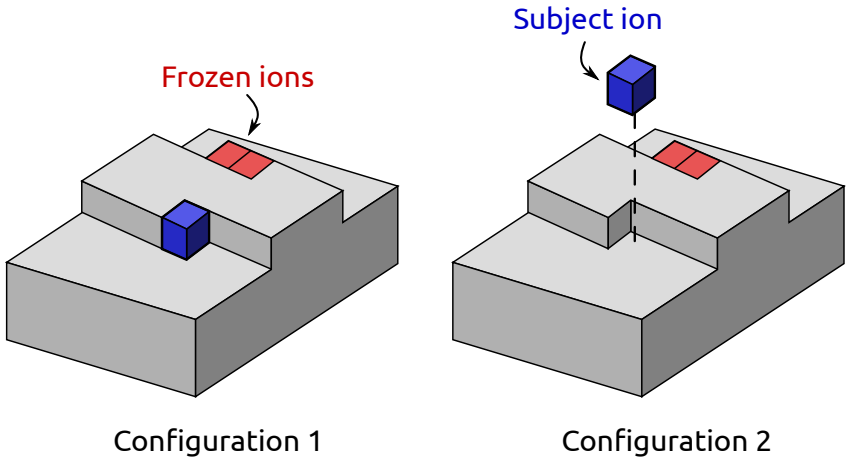


Figure 4.1: The two configurations used to compute dissolution enthalpies. In both configurations, the two ions that form the kink opposite the subject kink are frozen. The two configurations are distinguished by the position of the subject ion: either in an adsorbed kink state, or dissolved in solution.

first configuration, slowly dragging the subject ion normal to the surface and fixing it in the centre of the solution, unfreezing the slab (but keeping the subject ion fixed), and then equilibrating for 10 ns. In both configurations, the two ions that terminate the opposite kink site had their centres of mass fixed in their average location. This was because, while those sites were unlikely to dissolve during the timescale of our simulations, it was likely that some of them would transition to different states, for example, breaking a bond with the terrace below. Such random conformational changes would have affected the potential energy calculations. The setup is depicted in Figure 4.1.

After equilibration, the potential energy of each configuration was averaged over a period of 10 ns. The dissolution enthalpy is then simply the change in potential energy between the two configurations. Note that we monitored the displacement of the calcite slab to ensure that there were no substantial conformational changes during the averaging process that could have affected the potential energy calculations. The change in enthalpy during the dissolution of a single terminal ion is presented in Table 4.1 for all 16 kink sites.

Surprisingly, the dissolution enthalpies vary wildly amongst the various kink sites, ranging from -19.1 kJ/mol to $+40.5$ kJ/mol. And this difference is not just between the calcium and carbonate species, but also between the different lattice sites. Note that a negative dissolution enthalpy favours dissolution, whereas a positive enthalpy

Table 4.1: The dissolution enthalpies (kJ/mol) of all 16 kink sites of calcite, for the dissolution of the single terminal ion. The labelling is explained in Section 1.3. The standard errors are approximately 2 kJ/mol.

	Ca(1)	CO_3(1)	Ca(2)	CO_3(2)
Aa	4.2	−7.0	−19.1	34.4
Ao	25.2	−6.5	0.5	21.2
Oa	28.9	36.4	−14.2	−11.1
Oo	−18.1	40.5	6.2	7.0

favours adsorption. The large variation in the dissolution enthalpies can be attributed to the many different orientations that the carbonate ions exhibit in calcite. When a carbonate ion terminates a kink site, the direction of the carbonate, as well as the number of oxygen atoms exposed to the solvent, depends on which kink site it is (Aa, Oa, etc). Not ony does this variation in geometry affect the solvation energy of that CO_3-terminate kink site, but it also affects the bonding character of the subsequent calcium ion in a Ca-terminated kink site.

While the dissolution rate of each kink will depend on the relevant free energy landscape, and not merely on the dissolution enthalpy, one would expect a close relationship between the two. For example, a large positive enthalpy implies that the terminal ion will either have to overcome one large barrier, or a large succession of smaller barriers. In either case, dissolution will be inhibited. On the other hand, a negative enthalpy does not demand the existence of any barrier to dissolution whatsoever (although one will certainly exist for other reasons). In summary, one would reasonably expect larger enthalpies of dissolution to correspond to slower dissolution rates.

In Table 4.1, it can be seen that for the acute step, the slowest kink to dissolve will likely be Aa/CO_3(2) since it has the largest dissolution enthalpy, and we would expect this kink to dissolve via the terminal ion rather than the terminal pair. For the obtuse step, we would predict Oo/CO_3(1) to be the slowest to dissolve.

4.2 DISSOLUTION MECHANISMS

Having identified two kink sites that are likely to limit the rate of dissolution, the next task was to establish whether they actually dissolve via the single terminal ion or via the terminal pair functioning as a

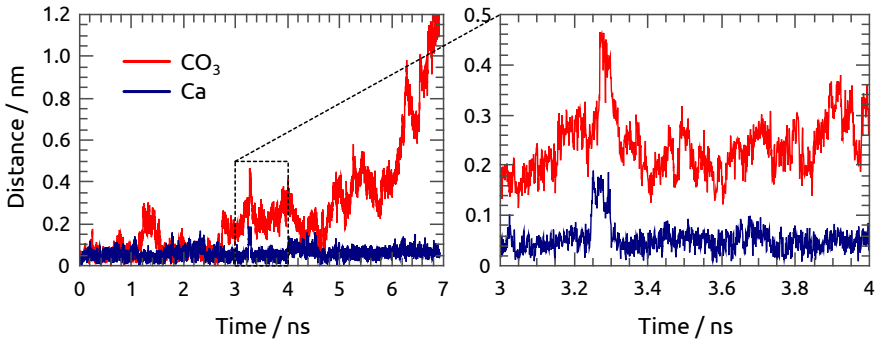


Figure 4.2: For the Aa/CO_3(2) kink site, the red curve shows the distance moved by the terminal carbonate ion from its initial kink-adsorbed position, and the blue for the neighbouring calcium ion, during a metadynamics-accelerated simulation of dissolution. The carbonate ion was found to dissolve on its own, leaving the calcium ion behind. See the text for discussion.

single unit. In reality, both mechanisms, and possibly some complex intermediate processes, will occur, albeit with one likely dominating over the others. However, it would not be possible to know the relative probabilities of these various pathways without first computing the rate of every possible dissolution pathway – an obviously impractical calculation.

Instead, in an attempt to establish the most likely dissolution mechanism, we ran a series of molecular simulations of the dissolution process, but accelerated by metadynamics (see Section 2.6.6.3 for a summary of this method). The acceleration was, of course, needed because a direct simulation would take prohibitively long.

The centre of mass of the calcium ion, and the centre of mass of the carbonate ion, were taken as the collective variables for the metadynamics, giving a total of six collective variables. Note that driving just the centre of mass of the calcium carbonate pair would have likely biased the pair to stick together and therefore the additional degrees of freedom were necessary.

Since the purpose of these simulations was to accelerate the trajectory rather than construct a free energy map, and also since we were sampling a large six-dimensional phase space, the Hamiltonian could be augmented with unusually large Gaussians. The Gaussian depositions had a width of 0.2 Å and were deposited at a rate of 2.0kT/ps (for $T = 300$ K). The temperature distribution was computed both with and without the metadynamics acceleration to ensure that the high deposition rate was not breaking adiabacity. Once either one of

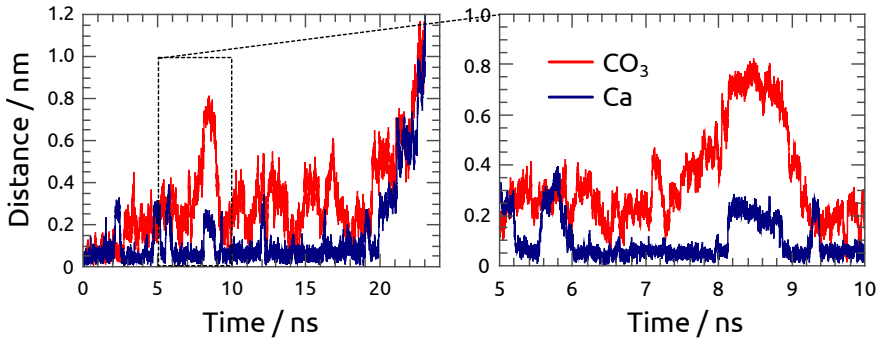


Figure 4.3: The description for this plot is the same as for Figure 4.2 except that, during this particular simulation, the calcium and carbonate pair were found to dissolve as a single unit. See the text for discussion.

the two terminal ions of the kink site had moved 1.2 nm from their initial positions, dissolution was assumed to have occurred and the simulation would terminate. This was repeated twenty times for each of the two kink sites.

The simplest way to analyse the mechanism in each simulation was to plot the distance that each of the two terminal ions had moved during the course of the simulation. The plot in Figure 4.2, obtained for the Aa/CO_3(2) kink site, was typical of virtually all of the simulations and for both kink sites. The terminal carbonate ion was the most animated, with the neighbouring calcium ion occasionally jumping to a new position, but quickly (~ 10 ps) returning to its previous position, as can be seen around 3.2 ns into the simulation. Eventually the terminal carbonate ion breaks away and dissolves, leaving the calcium ion in its equilibrium position. The acute kink took an average of 8.5 ns to dissolve compared to 12.9 for the obtuse kink. However, it is important to note that these numbers do not correspond to *real* time due to the metadynamics augmentation.

In a couple of the simulations, the terminal pair did in fact dissolve as a single unit, as shown in Figure 4.3. However, this only ever occurred after close but failed attempts by the carbonate ion to break away on its own, as occurred around 8 ns, and it is not until around 20 ns that the pair dissolve. This suggests that pair dissolution only ever occurred as a consequence of a heavily oversaturated bias potential.

From these simulations, therefore, it seems reasonable to conclude that both of the kink sites, Aa/CO_3(2) and Oo/CO_3(1), dissolve via the terminal ion only, with the neighbouring ion remaining in place. This is presumably because dissolving as a pair would require a more

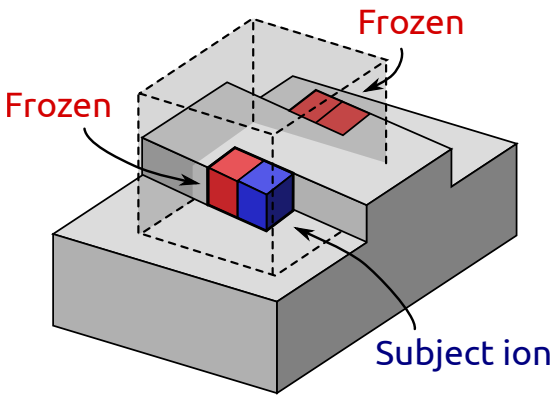


Figure 4.4: The description for this plot is the same as for Figure 4.2 except that, during this particular simulation, the calcium and carbonate pair were found to dissolve as a single unit. See the text for discussion.

complicated and therefore rarer sequence of events, such as the terminal carbonate breaking one or two bonds, followed by the calcium ion breaking a bond but without the carbonate ion first reforming those previous bonds. The ions would have to 'collaborate' to dissolve as a pair.

4.3 DISSOLUTION PATHWAYS

Having established that the two kink sites dissolve via the single terminal ion, the next task is to establish the precise pathway taken. For example, does the carbonate ion break away directly into the bulk solution, or diffuse away from the site along the step or terrace? Does it have to overcome a single activation barrier, or a complex sequence of barriers? To address these questions, we generated the three-dimensional free energy map for the terminal carbonate ion. From this, the minimum free energy pathway could be approximated and then converged using the string method, as detailed next.

4.3.1 *Free energy maps*

As the terminal carbonate ion breaks away from its kink site and explores the local volume, there will be an associated change of free energy. An accurate measure of this free energy difference depends on a suitable choice of collective variables which may include: the centre of mass of the carbonate ion, the orientation of the carbonate ion,

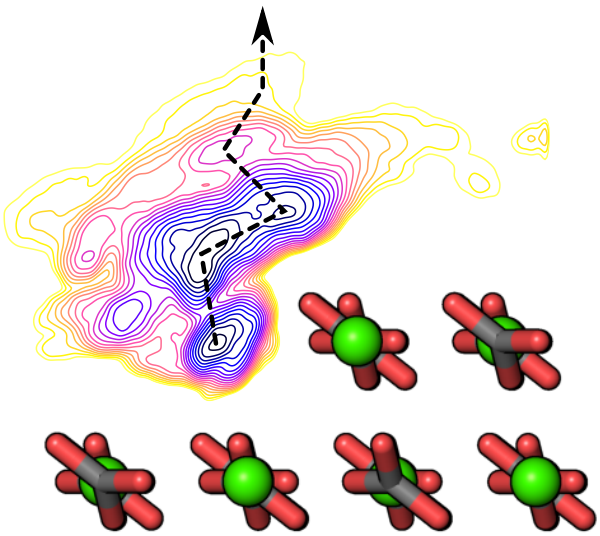
the local water structure (e.g. the density of the first hydration shell), as well as the configuration of the local calcite slab. However, constructing such a high-dimensional free energy map is unrealistic, so we make the ansatz that the free energy can be adequately described by only three numbers: the (x, y, z) coordinates of the centre of mass of the terminal carbonate ion. The hope is that the other coordinates are either redundant or present only very small barriers that do little to impede ergodicity.

Having selected the collective variables, we employed well-tempered metadynamics with multiple walkers to generate the free energy map. During the sampling, the carbonate ion was confined to a cube of length 1.2 nm, as shown in Figure 4.4, which was aligned such that the kink site was centered in the xy-plane, and the underlying calcite layer was aligned with the base of the cube. The neighbouring calcium ion was frozen in its equilibrium position by omitting it from the integration sequence, as justified in the previous section. All other atoms were unconstrained, except for the total momentum of the slab being fixed to zero to prevent drift. Note, however, that this calcium ion will be unfrozen during the string minimisation in the next section. The Gaussians had a height of 0.5 kJ/mol and were deposited every 500 steps into a bias potential shared by eight simultaneous walkers, each of which was initiated with the carbonate ion in a randomly selected position within the sampling volume. To enhance convergence, the well-tempered scheme was employed with a bias factor of 25.

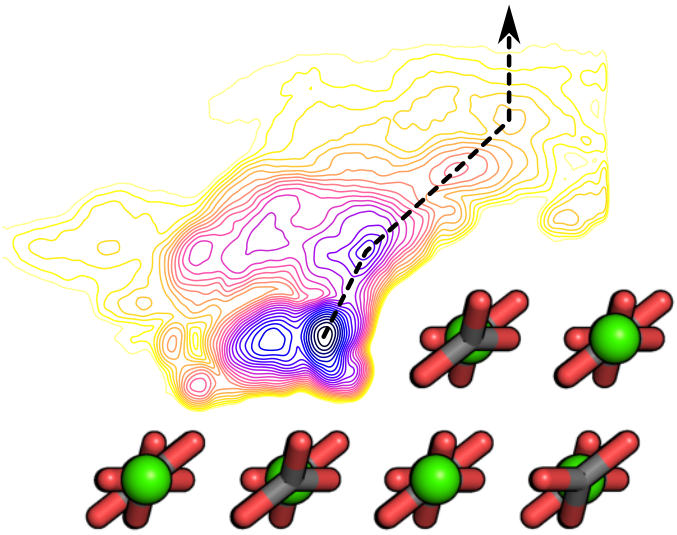
For each kink, the metadynamics accrued 3×10^6 Gaussians, the equivalent of 1.5 µs, during which it fully explored the sampling volumes. To visualise the resulting free energy maps, we project them onto the two-dimensional yz-plane of the centre of mass of the carbonate ion by integrating out the x coordinate:

$$F(y, z) = -\beta^{-1} \log \int e^{-\beta F(x,y,z)} \, dx \tag{4.1}$$

The resulting plots are shown in Figure 4.5. It should be noted that, while the contour lines in the figures correspond to differences of $1\,k_B T$, the actual free energy barriers cannot be accurately read directly from these plots for at least two reasons: all three degrees of freedom may be needed to correctly characterise the transitions, and the integration process may mix distinct free energy minima into one. Nevertheless, the general structure of each free energy map is discernible. In particular, it is apparent that there are multiple metastable states for the carbonate ion within the immediate vicinity of the kink



(a) Acute kink: Aa/CO_3(2)



(b) Obtuse kink: Oo/CO_3(1)

Figure 4.5: Two-dimensional projections of the three-dimensional free energy maps constructed for the terminal carbonate ion in the vicinity of the (a) Aa/CO_3(2) kink and the (b) Oo/CO_3(1) kink. Each contour line is separated by $1 k_B T$. The dashed lines represent the approximate minimum free energy pathways for dissolution.

site, and so dissolution will involve a complex pathway of multiple transitions. The next step is to identify the location of these minima and construct the minimum free energy pathway that connects them.

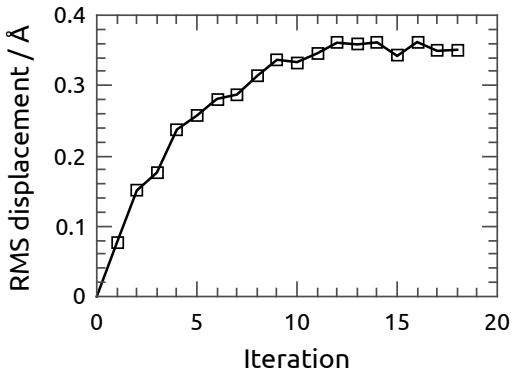


Figure 4.6: The root-mean-square displacement of the images in a string for successive iterations of the string method.

4.3.2 *Locating the minima*

Once the three-dimensional free energy maps have been generated, it is necessary to identify the locations of the various local minima. These minima, where the free energy derivatives are zero, $\nabla F = 0$, correspond to the metastable states of the carbonate ion, and the dissolution process will involve a sequence of transitions between them.

To identify the minima, we divided the space into a regular grid with cells spanning 0.1 Å in each dimension. A toy particle that experienced the free energy map was placed in the centre of each cell and then integrated downhill using the steepest descent algorithm [105]. All of the particles were thus moved to the minimum of whichever basin they were initially in.

In a few instances the particles would become trapped in minima that were the product of statistical noise. These particles could be identified and removed by computing the smallest barrier needed to be overcome in order to move to another cell with a lower free energy. This was computed by moving outward in concentric cubes until a cell with a lower free energy was found, and then recording the largest increase in free energy within that cube. A barrier smaller than a threshold of 1.0 kJ/mol would be registered as noise.

All points that ended up in the same cell after the minimisation step would be clustered into a single point and outputted. The result was a set of vectors corresponding to the various local minima (intermediate states).

4.3.3 *Constructing the pathways*

For the carbonate ion to transition from the initial kink site to the dissolved state in bulk solution, it must hop between the various intermediate metastable states. The MFEP of dissolution may therefore be approximated by the piecewise linear curve that connects the minima obtained from the metadynamics calculations into a chain, as depicted by the dashed lines in Figure 4.5.

The true MFEP, however, would not be piecewise linear. Instead, it would be curved so that there are no forces acting normally to it (by definition). To converge the piecewise linear curve to this true MFEP, we employed the standard string method as described in Section 2.6.5.1. In principle, the string could be converged within the metadynamics-generated free energy map; the requisite forces would simply be computed directly from the free energy map which would have the advantage of not requiring any further simulations. However, we instead opted to run molecular simulations to evaluate the mean forces for each image in the string at each iteration. In other words, to find the approximate MFEP in the exact free energy landscape rather than the exact MFEP in an approximate free energy landscape. There were two reasons for this:

1. The minor statistical noise present in the metadynamics-generated free energy maps was too difficult to control for and could engender quite substantial deformations in the strings.

2. During the metadynamics, the calcium ion neighbouring the carbonate ion was frozen (for pragmatic reasons described above). By evaluating the mean forces from scratch, the calcium ion could be released which would potentially introduce a small correction to the MFEP.

Regarding the string method, each path connecting a pair of minima was converged independently from the others. Each string was composed of images uniformly separated by around 0.25 Å, and the step size was

$$h = \min\left(1, \frac{0.1\,\text{Å}}{\max_i |f_i|}\right) \tag{4.2}$$

where f_i is the mean force evaluated for the i-th image. The forces were averaged for 250 ps and the images tethered with a spring constant of $k = 100$ eV/Å. The effect of the step size, Eq (4.2), was to cap

the largest displacement of any individual image to 0.1 Å. Convergence was manually detected, with the average string taking around 10-20 iterations to converge, as exemplified by Figure 4.6 which shows the root-mean-square (RMS) displacement of the images in one of the strings as a function of the number of iterations.

A snapshot from each converged metastable state is shown in Figure 4.7 for both of the kink sites. The dissolution pathway for Aa/CO_3(2) involves the carbonate ion moving vertically upward, breaking the bond with the underlying layer (B); reorienting itself and making room for a water molecule to wedge itself below (C); breaking the bond with one of the calcium ions, leaving it directly attached to only one and attached only indirectly to the other through the shared solvent (D); breaking the bond with the final calcium ion and moving into an outer-sphere adsorption state where it is bound to the kink through the shared solvent (E). From here the carbonate ion may break away into solution, completing the dissolution process.

The process is similar for the obtuse kink site except the carbonate doesn't have the reorientation step (acute B): the carbonate ion breaks the bond with the lower layer (B); it breaks the bond with the now-terminal calcium ion, sharing the solvent with it (C); it breaks the final calcium bond but remaining attached via the shared solvation shell (D). The final event is complete dissolution.

4.4 DISSOLUTION RATES AND FREE ENERGIES

The strings produced so far for each dissolution pathway had a density of roughly one image per 0.25 Å. However, for the analysis that follows, denser strings were required. This was straightforwardly achieved by fitting a cubic spline to each string and then inserting additional points along the curve and reparametrising to make them uniformly spaced again. The updated strings had a density of one image every 0.05 Å, a five-fold increase in density. What follows is the construction of the free energy curves and transition rates associated with each dissolution pathway.

4.4.1 *Free energy curves*

Taking the length along the string, ξ, as our reaction coordinate, the associated free energy change, $\Delta G(\xi)$, was computed using umbrella sampling. Specifically, the carbonate ion in each image along the

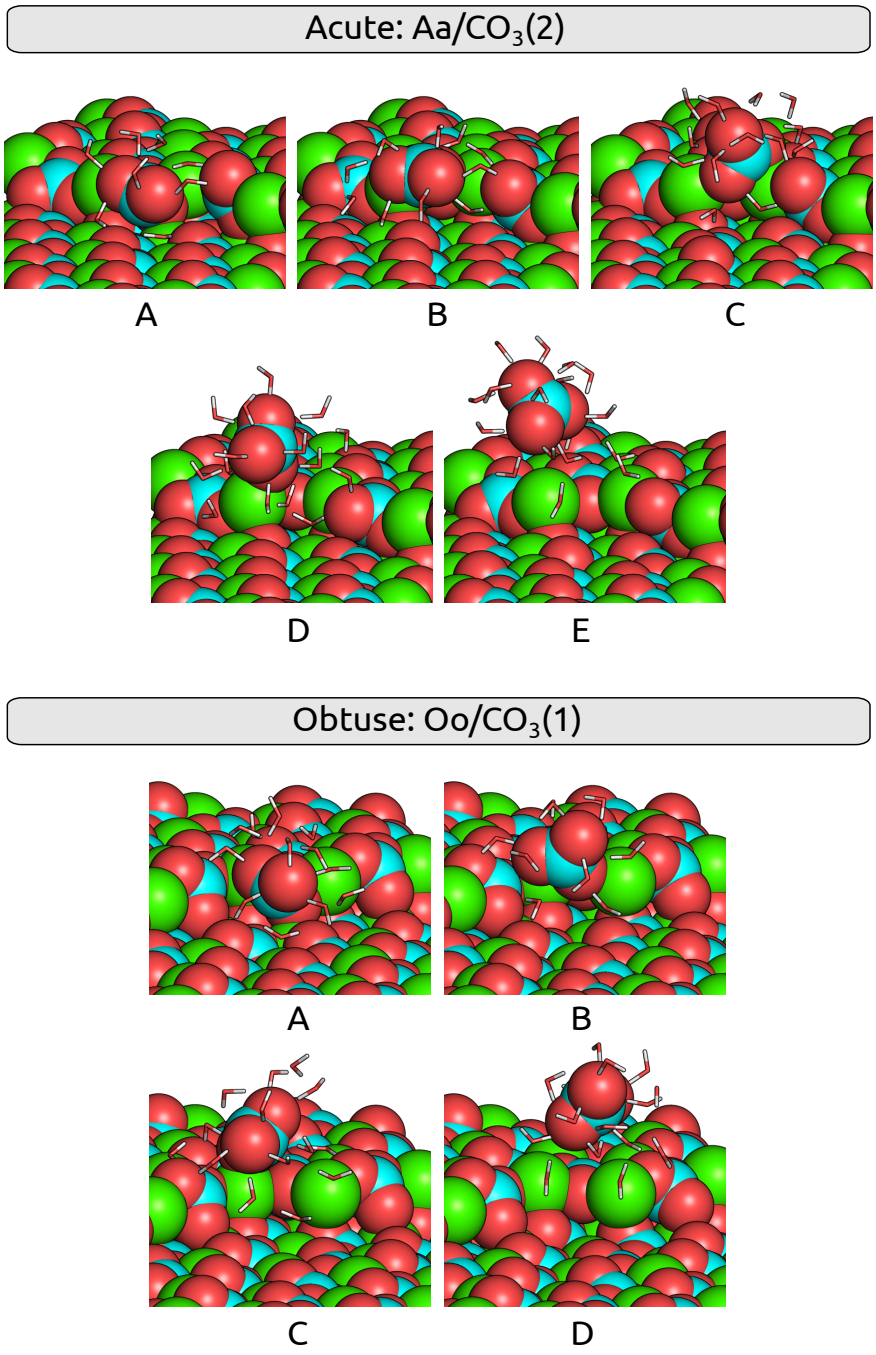


Figure 4.7: Starting with the kink state (A), these are snapshots of the various intermediate metastable states visited prior to dissolution for both the acute ($Aa/CO_3(2)$) and obtuse ($Oo/CO_3(1)$) kinks. Oxygen atoms are red, carbon teal, calcium green, hydrogen grey. The calcium carbonate is rendered as van der Waals surfaces while the water in the hydration shell of the terminal carbonate ion is shown in stick form. See the text for a description of each snapshot.

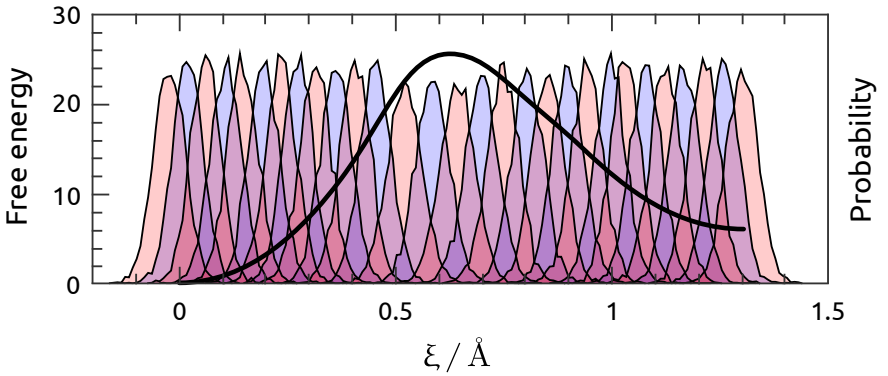


Figure 4.8: The alternating red-and-blue histograms are the probability distributions for the carbonate ion along the reaction coordinate ξ, when tethered to each of the sampling windows. From these histograms, WHAM reconstructs the underlying free energy curve which is shown in black. This particular interval corresponds to the A $\leftrightarrow$ B transition for the acute Aa/CO_3(2) kink.

string would be tethered to its window using a harmonic potential with a spring constant of 15 eV/Å. This would force the system to explore only that local region of the reaction coordinate. For each window, a histogram would be compiled by recording the reaction coordinate of the carbonate ion every 100 fs for a total of 500 ps. The various sampling windows are shown in Figure 4.8 for one of the transitions (the A $\leftrightarrow$ B transition of the Aa/CO_3(2) kink) along with the corresponding free energy curve that was then reconstructed from the histograms using the weighted histogram analysis method (see Section 2.6.6.2).

Repeating the above analysis for all of the transitions along each pathway, and assembling them together, resulted in the complete free energy curve for each dissolution pathway, as shown in Figure 4.9. Included are the 68% (give or take one standard deviation) and 95% (give or take two standard deviations) confidence intervals which were computed using bootstrap resampling with 10^3 Monte Carlo trials. Each metastable state is labelled with a letter that corresponds to a snapshot in Figure 4.7.

Note that the final barrier to dissolution is shown with a dashed line. This is because the final barrier was computed by dragging the carbonate ion vertically upwards into the solution, which is not necessarily the pathway taken. Therefore the barrier itself should not be interpreted physically, although the total free energy change will be accurate.

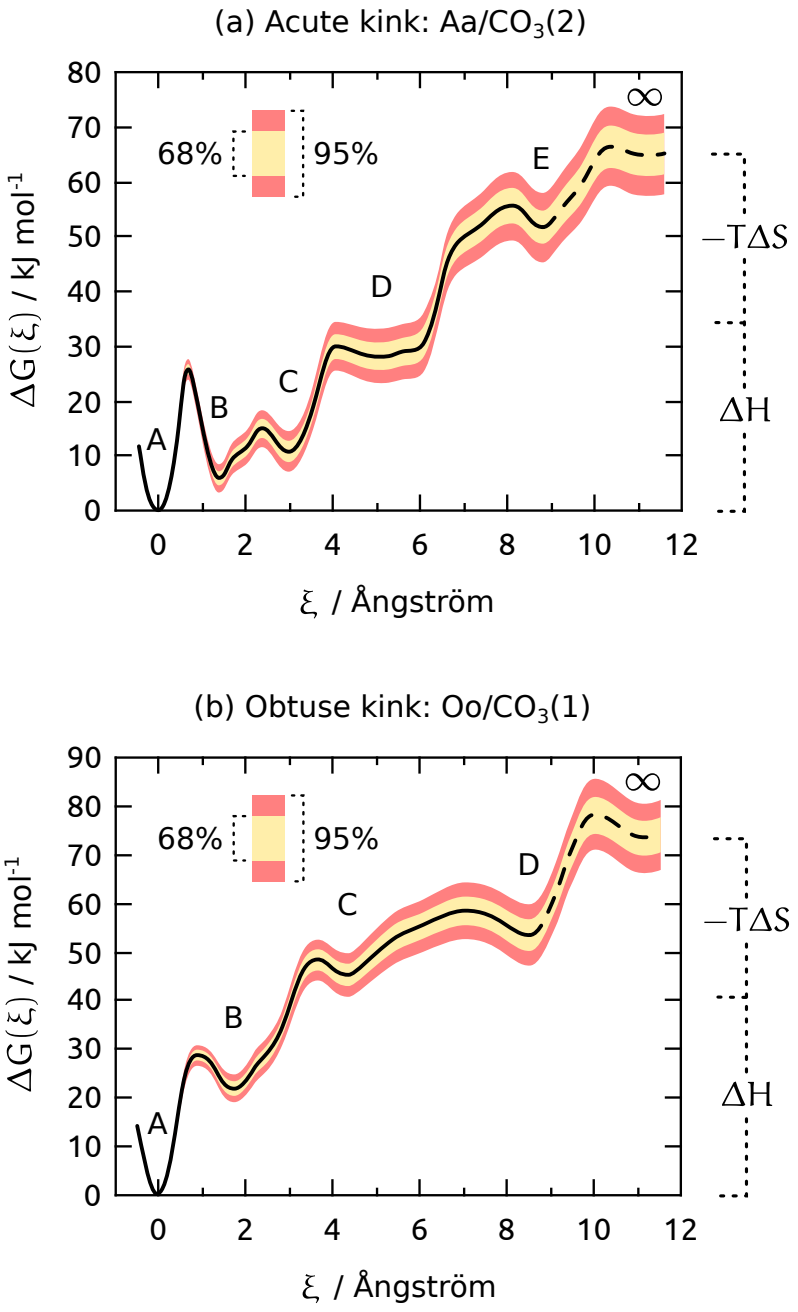


Figure 4.9: The free energy curves for the dissolution of the (a) $Aa/CO_3(2)$ and (b) $Oo/CO_3(1)$ kink sites. The 68% and 95% confidence intervals are shown for the free energy differences with respect to the initial kink states $\xi = 0$. Each metastable state is labelled with a letter that corresponds to a snapshot in Figure 4.7. The enthalpic ΔH and entropic $-T\Delta S$ contributions to the total dissolution free energy is shown.

Also shown in Figure 4.9 are the enthalpic ΔH and entropic $-T\Delta S$ contributions to the total dissolution free energies. These were ob-

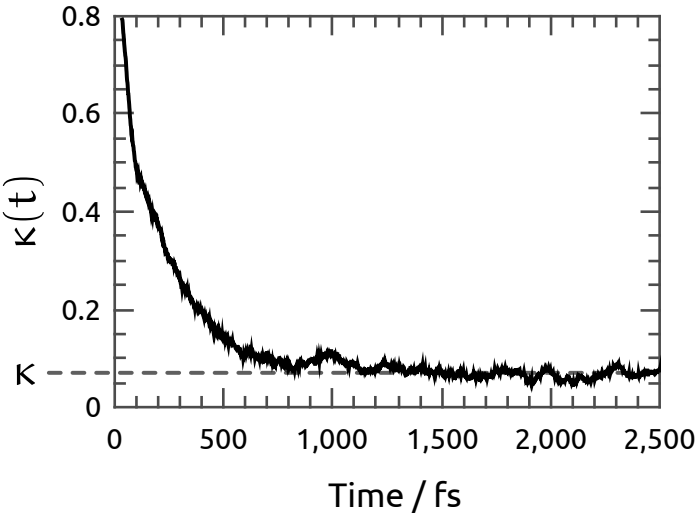


Figure 4.10: The transmission coefficient $\kappa(t)$ defined in Eq. (2.96) and sampled across the $A \leftrightarrow B$ transition barrier of the obtuse kink.

tained directly from the dissolution enthalpies computed previously in Section 4.1. It is interesting to see that the entropic contribution, $-T\Delta S$, which specifically measures the solvation entropy, makes almost as large a contribution as the enthalpy does, approximately 30 kJ/mol in both cases. To see why the solvation entropy decreases ($-T\Delta S > 0$ implies $\Delta S < 0$), consider the water structure before and after dissolution: Before dissolution, there will be water structure in the vicinity of the kink site. After dissolution, there will still be water structure in the vicinity of the (new) kink site but also water structure surrounding the desorbed carbonate ion. This overall increase in water structure comes at the cost of a reduced entropy.

The free energy curves will be discussed further below, in the context of the overall dissolution rates.

4.4.2 *Individual transition rates*

The penultimate step in our analysis was to employ the reactive flux method of Section 2.6.3 to compute the transition rates for all of the individual transitions. To briefly summarise the method: through the application of Eq. (2.98), transition state theory (TST) provides an estimate for the transition rate, k_{TST}, based solely on the free energy curve. The true transition rate is then $k = \kappa k_{TST}$ where κ is the transmission coefficient which accounts for the diffusive nature of the barrier crossings.

To compute κ for a particular transition, a configuration would be produced for the transition state which is the image along the string, between the two minima of interest, where the free energy is a maximum. The carbonate ion was then restrained in this state using a harmonic potential with a spring constant of 100 eV/Å, and the configuration equilibrated for 1 ns after which a simulation of 10 ns was performed, with the positions and velocities of the atoms harvested every 5 ps. For each one of these 2,000 harvested configurations, the carbonate ion had the restraint removed and the trajectory was simulated both forward and backward (reversing the signs of the velocities) for 2.5 ps each, meanwhile recording which side of the transition barrier the carbonate ion was on as a function of time. Applying Eq. (2.96) to this data gives the function $\kappa(t)$.

We plot $\kappa(t)$ in Figure 4.10 for the A $\leftrightarrow$ B barrier of the obtuse kink. The transmission coefficient, κ, is whatever value this function plateaus to; 0.07 in this particular example. Repeating the above analysis for all of the transition barriers, the results are listed in Table 4.2. Also included are the free energy barriers $\Delta G^{\ddagger}$ associated with each of the transitions, along with their standard errors. Since the transition rates obey the relation $k \propto \exp(-\beta \Delta G^{\ddagger})$, the uncertainty in each free energy barrier translates to be the primary source of error in the corresponding transition rate. It follows that the standard errors for the free energy barriers (multiplied by β) also serve as the standard errors for $\log k$.

The very final step in the dissolution sequence was studied differently to the other transitions. This was because, during the metadynamics simulations in Section 4.2, it was observed that the carbonate ions did not necessarily break away directly into the bulk solution but may have briefly diffused along the terrace or step before breaking away into the bulk. A more appropriate method for studying this final transition would therefore be a large series of direct molecular simulations starting from the final state and observing where the carbonate ends up going and how long it takes. Indeed, for each of the final metastable states (acute E and obtuse D), a series of 1,000 simulations were performed. If the carbonate ion moved more than 5 Å away from its starting position then it was registered as a dissolution event (i.e. the carbonate has moved sufficiently far from the kink site that it's unlikely to return). Note that the original kink site was well within the 5 A radius. On the other hand, if the z coordinate dropped below that of the preceding metastable state then it would be registered as a retreat to that previous state. Once either of these

Table 4.2: A list of all transitions. The ∞ state represents diffusion away from the kink site (whether that be along a step/terrace or via bulk diffusion). The activation barrier for each transition $\Delta G^{\ddagger}$ with the standard error is given, along with the mean time for the transition k_{TST}^{-1} predicted by TST, the transmission coefficient κ, and the true mean transition time k^{-1}.

Transition	$\Delta G^{\ddagger}$ (kJ/mol)	k_{TST}^{-1} (ns)	κ	k^{-1} (ns)
Acute kink: $Aa/CO_3(2)$				
A→B	25.4 ± 0.9	9.1	0.08	113.8
B→A	19.7 ± 0.9	1.3	0.08	16.4
B→C	9.0 ± 1.1	0.02	0.06	0.33
C→B	4.3 ± 0.8	0.004	0.06	0.07
C→D	18.6 ± 1.1	1.1	0.04	27.5
D→C	1.8 ± 1.2	0.003	0.04	0.075
D→E	27.2 ± 1.9	92.1	0.06	1,535.0
E→D	-	-	-	0.22
E→ ∞	-	-	-	0.67
Obtuse kink: $Oo/CO_3(1)$				
A→B	28.7 ± 1.1	29.3	0.07	418.6
B→A	7.8 ± 0.9	0.012	0.07	0.17
B→C	26.9 ± 1.6	29.4	0.08	367.5
C→B	3.7 ± 0.9	0.004	0.08	0.05
C→D	13.7 ± 1.8	0.19	0.04	4.75
D→C	-	-	-	0.57
D→ ∞	-	-	-	1.7

two events occured, the time taken was recorded and the simulation terminated.

4.4.3 *Overall dissolution rates*

The final task is to compute the overall dissolution rate, viz. the rate at which the carbonate ion transitions from the kink site (A) to the fully dissolved state (∞). It is clear from the above analysis that the entire process will involve multiple transitions between successive states until the dissolved state is finally reached. To find the average time taken for dissolution, we first consider a more general problem.

Suppose there exist n states, labelled 1, 2..., n. A random walk, starting at state 1, will make random transitions between consecutive states until eventually reaching state n. Let $P_{i\to i+1}$ be the probability that the state i will transition to state $i+1$, and thus $P_{i\to i-1} = 1 - P_{i\to i+1}$, where we have the boundary condition $P_{1\to 2} = 1$. Suppose also that the mean time spent in state i before making a transition is t_i. Now let τ_i be the mean time taken for state i to reach state n, in other words, τ_i is the time averaged over every possible pathway from i to the final n-th state. Then the mean times obey the difference equations

$$\tau_i = t_i + P_{i\to i+1}\tau_{i+1} + P_{i\to i-1}\tau_{i-1} \tag{4.3}$$

which is just a linear system of equations that can be solved simultaneously to obtain the τ_i values. The mean time (rate) to get from state 1 to state n is then τ_1 (τ_1^{-1}).

In the case of our transition pathways, the probabilities can be computed from the relevant transition rates as

$$P_{i\to i+1} = \frac{k_{i\to i+1}}{k_{i\to i+1} + k_{i\to i-1}} \tag{4.4}$$

and the mean time spent in state i is

$$t_i = \frac{1}{k_{i\to i+1} + k_{i\to i-1}} \tag{4.5}$$

Solving Eqs. (4.3) then provides us with the overall dissolution rate, $\nu \equiv \tau_1^{-1}$, for each pathway. Going a step further, errors can be straightforwardly incorporated into this analysis. Recall from the previous section that the primary source of error in the transition rates is the uncertainty in the free energy barriers; one may therefore replace each of the individual transition rates k_i with numbers from the distributions

$$\log k_i \sim \mathcal{N}(\log k_i, \beta^2\sigma_i^2) \tag{4.6}$$

where σ_i is the uncertainty in the i-th free energy barrier height, reported in Table 4.2. Repeating this random generation of rates 10^6 times and recording the overall dissolution rate ν for each set provides a histogram for ν that represents our confidence interval for it. This histogram is plotted in Figure 4.11(a) for each of the two kink sites. The peak of each plot represents our best guess for the dissolution rate with the width corresponding to the uncertainty. It must be stressed that the errors shown are purely those that arise from the umbrella sampling, specifically from the fact that only a finite number of

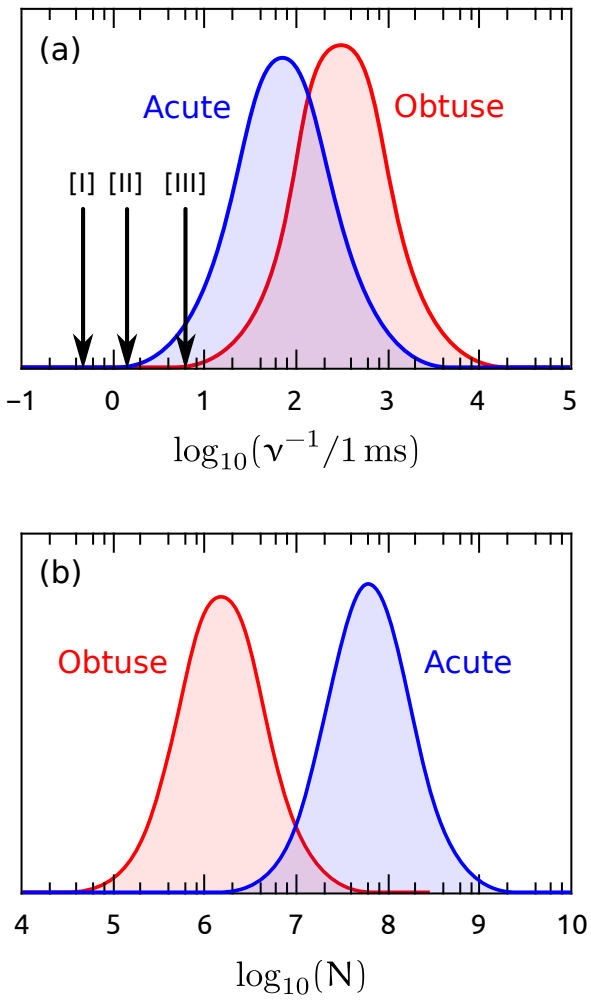


Figure 4.11: (a) The confidence interval computed for the dissolution rate of the Aa/CO_3(2) and Oo/CO_3(1) kink sites For comparison, estimates for the same numbers are provided from the literature: [I] Wolthers et al. [2], [II] Nielsen et al. [3], and [III] Stack et al. [4]. (b) The confidence interval for the mean number of transitions prior to dissolution. The huge number of transitions are the source for the vast errors

windows are sampled and for a finite period of time. In reality there will be other sources of error, most of which are entirely unquantifiable. For example, errors from the classical treatment of the system: the approximate functional forms and semiempirical parameters, the neglect of quantum and charge effects; errors from a coarse choice of collective variables; errors from the string not being perfectly converged; errors from the integration and the use of a thermostat. Note that each overall dissolution rate may be written in terms of a single activation barrier, $k_0 \exp(-\Delta E \beta)$, where ΔE is the total range covered

by the free energy curve. The error in ΔE is much smaller than the reported error in the overall rate. The errors may therefore be attributed to our inability to accurately gauge the kinetic prefactor for such complex free energy landscapes.

But even with our underestimation of the errors, each of the confidence intervals span an enormous three orders of magnitude. The relatively small errors in our individual transition rates have therefore been propagated in such a way that has greatly magnified them. To see why this is, we computed the average number of individual transitions that would occur prior to dissolution, for each kink. Using the same logic as that for deriving Eqs. (4.3), one obtains the following difference equations that describe the mean number of transitions N_i from state i to the final dissolved state,

$$N_i = 1 + P_{i\to i+1} N_{i+1} + (1 - P_{i\to i-1}) N_{i-1} \tag{4.7}$$

The resulting confidence interval for N_1 is shown in Figure 4.11(b) for both kinks. The conclusion is that each kink undergoes millions of transitions before successfully dissolving. The huge errors result from the fact that the (RMS average) error for the individual rates will scale as $\sqrt{N_1}$ and so this is the source of magnification.

A question that remains is why so many transitions are required for dissolution to occur. Or put another way, if the slowest transition barrier is only 1.5 μs (0.4 μs) for the acute (obtuse) kink, then why does it take tens or even hundreds of milliseconds for dissolution? This is best explained by revisiting the free energy curves in Figure 4.9. While none of the individual barriers are especially large, there are multiple metastable states that are of very low stability, presented with the option of either overcoming a large barrier and advancing forward, or overcoming a very small barrier and retreating backward. This asymmetry in the stability of the intermediate states is responsible for the huge number of failed dissolution attempts that must occur before success. And such an asymmetry seems inevitable for any complex pathway that exhibits a large change in free energy between the initial and final states.

Having addressed the issue of precision, we finally consider the question of accuracy. For comparison, the average kink dissolution rate has been estimated in previous studies by fitting simple analytical models to experimental data. More specifically, the analytical models describe calcite growth as consisting of just two events: kink nucleation and kink propagation (which is assumed rate-limiting). The step velocity may then be expressed in terms of the individual attachment

rates of the calcium and carbonate ions which are a function of the solution stoichiometry. The attachment rates my then be optimised by fitting to the experimental step velocities as a function of solution stoichiometry. The dissolution rates are then obtained straightforwardly from the attachment rates, knowing the solubility product of calcite. The estimated values are shown in Figure 4.11(a) with arrows that refer to the papers [I] Wolthers et al. [2], [II] Nielsen et al. [3], and [III] Stack et al. [4]. Our best estimates are one-to-two orders of magnitude larger than the rest, and therefore seem likely to be an overestimation of the true dissolution rates. It is worth noting, however, that while the cited papers measure the *average* rate at which the various kink sites dissolve, we have measured the rate of the ostensibly rate-limiting kink sites only, and therefore we expect it to be larger. A fairer comparison might therefore be to divide our dissolution rates by four (the number of ions in the primitive calcite unit cell). Doing so shifts the histograms towards the arrows such that our estimate for the acute dissolution rate is only a factor of 2 larger than the estimate from reference [III]. The obtuse, however, remains an order of magnitude too large.

4.5 SUMMARY AND DISCUSSION

This chapter began by presenting the dissolution enthalpies for all 16 of the unique kink sites for calcite crystals, mediated by the single terminal ion dissolving, and also the terminal pair dissolving as a single unit. The results revealed an enormous variability, not just between the calcium and carbonate species, but also between the same species that occupy different lattice sites. This variability points strongly towards calcite being a highly non-Kossel crystal.

From the dissolution enthalpies, an educated guess could be made as to which kink site for the acute and obtuse steps would likely be rate-limiting. Using metadynamics to drive the dissolution of these two kink sites revealed a preference for dissolving via the terminal ion only, rather than the terminal pair.

A three-dimensional free energy map was then generated for the terminal ion of each kink site, revealing the presence of multiple transition states between the adsorbed kink site and the solution. Connecting these intermediate metastable states with a string, and converging the string with the string method, resulted in the dissolution pathway for each kink.

Umbrella and reactive flux sampling of each pathway provided us with the free energies and rates for all of the transitions along the pathways. These individual rates, along with their errors, were then combined to find the overall dissolution rate for each kink. The confidence intervals for our dissolution rates were found to span three orders of magnitude each, an enormous error, and this was based only on the readily quantifiable sampling errors. Regarding accuracy, we estimated mean dissolution times of 80 ms and 344 ms for the acute and obtuse kink sites, respectively. These numbers are one-to-two orders of magnitude larger than existing estimates in the literature.

The poor precision and accuracy reported in this chapter has potentially far reaching consequences. One of the ongoing projects of the computational sciences is to compute the rates of rare molecular processes that can then be plugged into coarse grained models of some large-scale molecular system, as in the next chapter. But if any of the requisite transitions are complex, like those studied in this chapter, then they will necessarily be susceptible to poor sampling precision, which calls into question the utility of this bottom-up approach to anything other than trivial systems. While it is certainly true that force fields will only get more accurate, computers more powerful, and methodologies more adept, it would appear that these *ab initio* ambitions are many years away from being realised.

Despite these inadequacies, molecular simulations remain able to provide unique insight into these complex processes. For example, we can draw the following conclusions with reasonable confidence:

- Calcite kink dissolution involves a complex pathway of multiple transitions.
- Dissolution takes orders of magnitude longer than any of the individual transitions, due to the asymmetrical instability of some of the intermediate states.
- The barriers to *adsorption* are very small, suggesting that the adsorption of carbonate ions, at least to these particular kink sites, will be diffusion-limited. This is in agreement with the conclusions of Sand et al. [113].
- The cost of water structure is roughly half of the driving force that inhibits dissolution in these kink sites. One of the mechanisms by which impurities affect calcite solubility may therefore be by modifying the water structure in the vicinity of the kinks.

5

A COARSE-GRAINED MODEL OF CALCITE GROWTH AND INHIBITION

Chapter 1 described how calcite crystals grow via step propagation, where the steps are sourced from screw dislocations at low supersaturations. As the steps propagate across the crystal face, they eventually reach the edges of the crystal, meaning that whatever morphology the steps exhibit while growing, the same morphology will be adopted by the entire crystal face [12]. It follows that both the kinetics and morphologies of entire calcite crystals can be understood in terms of the kinetics and morphologies of just these propagating steps.

In this chapter, we combine the insight obtained in the previous two chapters with experimental data to construct a coarse-grained model of calcite growth. The model will then be employed to investigate the effects that impurities have on growth.

5.1 MODEL AND IMPLEMENTATION OVERVIEW

Since we are interested in modelling the propagation of steps across a two-dimensional surface, the crystal may be described by a two-dimensional grid. We use a square grid of size $n \times n$, where n is a power of two, for reasons described below. In all simulations the grid will have periodic boundary conditions.

Each grid cell corresponds to a single ion within the growing (or dissolving) top layer of the crystal, and may therefore store either a calcium ion, a carbonate ion, an impurity, or, by default, the solvent. We shall refer to the value of each cell as its *state*; the surrounding eight cells as the *neighbourhood*; and the combination of state and neighbourhood, i.e. the local 3×3 grid, as the *configuration* of the cell. Based on the configuration of each cell, a list of possible events may be generated. For example, if a particular cell is in the solvent state with a kink neighbourhood, it may have the option of transitioning to (say) a calcium ion or to an impurity. A list should therefore be generated to store every possible event that may occur, along with the corresponding transition rate for each event (details below). In addition to having a particular state, each cell may be *frozen*, meaning that its state is fixed in time by omitting its events from the event list.

The system is then integrated over time using the kinetic Monte Carlo (KMC) method described in Section 2.7. Specifically, the algorithm proceeds by randomly selecting an event from the even list in accordance with the probability distribution given in Eq. 2.116, and then executing said event. Once the state of the corresponding cell has been changed, the event list of that cell as well as its eight nearest neighbours must be updated. Lastly, the time is incremented by the stochastic variable given in Eq. 2.119, and the whole process repeats.

With regards to implementation, we store the event list in a 4-ary tree structure, where the entire grid (the root) is spatially divided into four quadrants (nodes), each of which is further divided into four quadrants (nodes), and so on until the smallest event domain (leaf) spans a single cell. Each leaf then contains the event list for the corresponding cell. There are a few advantages to storing the data in such a structure. Firstly, all of the key calculations, such as adding and removing events from the list can be performed in $\mathcal{O}(\log n)$ time, where n is the cell size, compared to $\mathcal{O}(N)$ for a linear list, where N is the total number of possible events. Furthermore, dividing the events into spatial domains allows the algorithm to be straightforwardly parallelised (although we do not do this). It should now be clear why the grid must be a square and the size n a power of two.

In the case of a pure system (i.e. no impurities), each cell has up to 512 possible configurations. However, during the evolution of the system, depending on the event table, only a subset of these configurations will actually be accessible. Moreover, of those that are accessible, most will be related to several other configurations via an appropriate symmetry operation, thus greatly reducing the representation of our event table. In the end, for our calcite model, we found that all of the available configurations could be generated by applying the D_8 dihedral group to the configurations shown in Figure 5.1. Note that the configurations in the top row are the intended, operative configurations that can be characterised as kinks, steps, corners, and gaps. To clarify: corners are similar to kink sites but they typically form at the corners of a growing layer where two steps meet each other; and the gaps typically form when either a pair of kinks propagate towards each other, shortly before annihilating, or when a step dissolves to form a pair of kinks. The bottom row, on the other hand, shows the undesired artefacts that may occasionally be generated by particular sequences of the operative events. Our code explicitly searches for these configurations and rectifies them without any time increment,

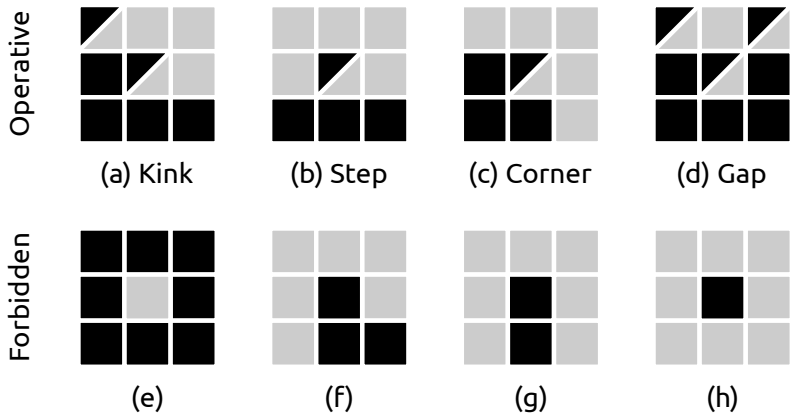


Figure 5.1: Every local cell configuration accessible in our model may be generated by applying the D_8 symmetry group to one of these configurations. The black cells denote calcium or carbonate ions, the grey denotes solvent, and the two-tone squares denote either a calcium ion, a carbonate ion, or solvent. The intended, canonical configurations are displayed in the top row, and the unwanted artefacts that must be explicitly forbidden in the bottom row.

viz. by adsorbing an ion in the case of (e) and dissolving the central unit for (f)-(h).

Whether each cell corresponds to, say, an Oo kink, or an Oa kink, depends on the orientation of the operative configuration. For example, Figure 5.1(a), as shown, corresponds to the Oo kink site. Whereas, if it were flipped horizontally, it would correspond to the Oa kink site. In order to build the event list, the appropriate site must be identified, along with the state of the central cell. But a third piece of information is also needed: what we refer to as the *basis index*. Recall from Chapter 3 that calcite is non-Kossel, meaning that the transition rates for a cell will depend not only on whether the cell houses a calcium or a carbonate ion, but which calcium or carbonate ion in the basis it corresponds to. We resolve the basis type by mapping each cell to a basis index, $(i, j) \mapsto \mathrm{mod}(i + j, 4)$, where 4 is the number of ions in the primitive unit cell for calcite.

In total, all of the possible configurations give rise to 160 possible events, each of which must be assigned a transition rate. Needless to say, we make a series of approximations to greatly reduce this number, as detailed in Section 5.2 below where a scheme is outlined for optimising the parameters.

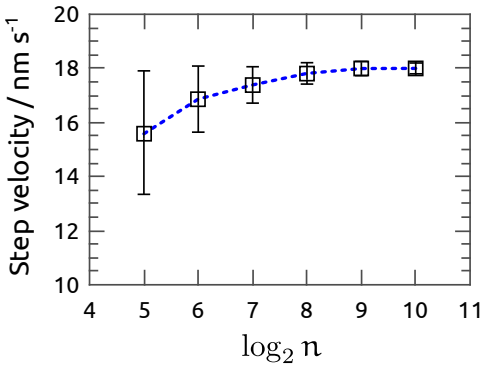


Figure 5.2: The measured step velocity in our KMC simulations as a function of the grid size n.

5.1.1 *Measuring step velocities and morphologies*

This chapter is concerned entirely with the kinetics and morphologies of calcite crystals under various conditions, and so it was necessary to have a means for measuring both.

In the case of step velocities, within a square grid of dimension $n \times n$, half of the grid was initially occupied with a calcite crystal, exposing two opposing steps that were infinitely long due to the periodicity of the grid. One of the steps would be frozen to prevent it from propagating, and the unfrozen step would then be integrated over time. In each case, an initial equilibration was performed to allow the step to achieve its natural kink density, and involved performing a maximum of 10^6 iterations and terminating if the unfrozen step advanced, in either direction, by more than five layers (1.6 nm). A production period was then performed until the step either reached one end of the grid or the maximum of 10^7 iterations had been reached. The net number of ions added to the crystal during the production, divided by the number of ions per layer n, multiplied by the length of each layer 0.32 nm, and divided by the simulation time gave the step velocity. Of course, the accuracy of the measured step velocity depends on the size of the grid n, and we show in Figure 5.2 the average and standard deviation for the step velocity of the obtuse calcite step (with the optimised parameters below) for various grid sizes. It can be seen that by $n = 2^9$ the average step velocity measurements are converged with respect to n and the standard errors are acceptable, and so this was the choice of n employed in this work.

To investigate step morphologies, we would begin with a $2^{10} \times 2^{10}$ grid with a small 21×21 square of calcite in the centre. The system

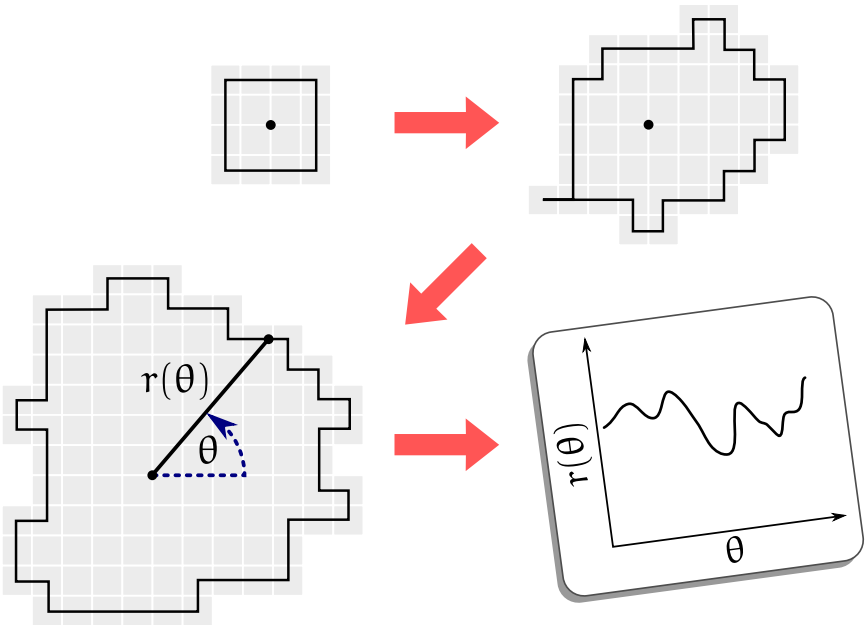


Figure 5.3: Crystal morphology is computed by allowing a small layer to grow into a larger one and then mapping the edges to a radial plot which may be averaged over many repeated runs.

would then be integrated through time so that the square grew to become a much larger crystal. The simulation would terminate once either 10^7 iterations had been performed, or the crystal reached the edge of the grid. In principle, the resulting crystal could then simply be rendered to produce an image. However, the crystals would typically exhibit small random variations about some true, underlying morphology, due to the stochastic nature of growth. We therefore

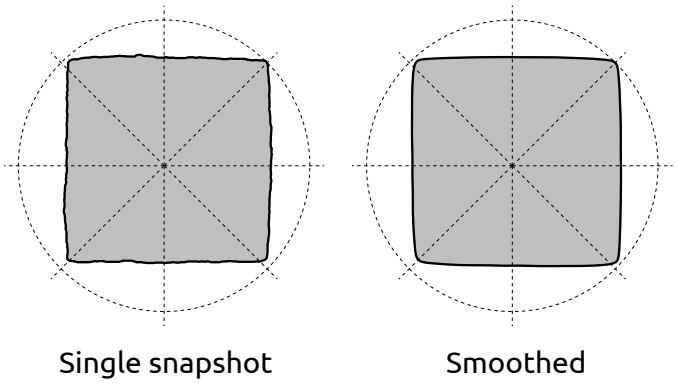


Figure 5.4: A snapshot of a simulated island after it has grown to approximately 300 nm in width, and the same island but smoothed as described in the text. The smoothing has a purely local effect.

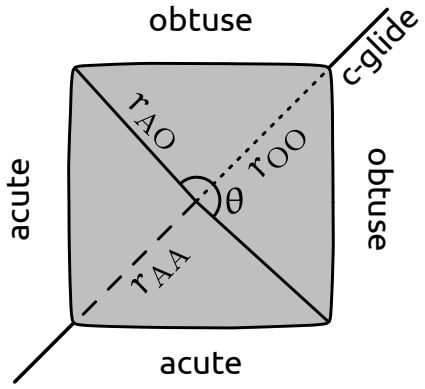


Figure 5.5: The orientation shared with all of the step morphologies that follow. Also shown are a few morphology metrics that measure the distances and angles between the origin and the various corners of the morphology.

sought to generate an average morphology to mitigate these fluctuations. To this end, we would repeat the previous simulations ten times and construct a polar plot $r(\theta)$ for the morphology of each, as illustrated in Figure 5.3. Each polar plot would then be normalised such that the average of $r(\theta)$ was unity, and then all of the radial plots could be straightforwardly averaged. The resulting, average morphology could then be produced by rendering the averaged polar plot. The smoothing effect is illustrated in Figure **??** where a single snapshot and its smoothed equivalent of a grown island are shown (the simulation parameters are those described in Section [**?**] below). The smoothing can seen to have a purely local effect, preserving the overall morphology. All of the morphologies will be rendered with the orientation shown in Figure 5.5, with the obtuse steps pointing north and east. We also define the metrics shown: r_{AA} being the distance between the origin and the corner where the acute steps meet, and similarly for r_{AO} and r_{OO}, and θ is the angle between the AO and OA segments, as shown.

It should be noted that our renderings of the morphology treat the calcite geomtry as orthogonal, whereas it is actually skewed. So a square morphology in our simulations would correspond to a non-square rhombus in an experimental image. Therefore, in order to compare our morphologies to those observed experimentally, we will trace the experimental morphologies and apply the appropriate affine transformation to make the underlying geometry orthorgonal.

5.2 PARAMETER OPTIMISATION

At this stage, we have 160 transition rates that must be specified, consisting of the adsorption and dissolution of each species to and from each site. First, we shall consider the nature of these transitions, then impose a series of approximations to greatly reduce this number, and finally optimise the remaining parameters.

Our model of growth proceeds by the nucleation of kink sites followed by their propagation. In the case of the nucleation, molecules must adsorb to the step, as modelled in Chapter 3. These molecules may be individual calcium or carbonate ions, or oligomers such as the ion pair $CaCO_3^0$. Now suppose one of the single ions were to adsorb to a step. In order to nucleate growth, the next event would be for an ion of the other species to adsorb next to it, thus forming an ion pair on the step. However, in the vast majority of cases, the first ion will dissolve before the second ion arrives. On the other hand, preformed ion pairs are already abundant in the solution. Specifically, the equilibrium constant for the reaction

$$Ca^{2+} + CO_3^{2-} \rightleftharpoons CaCO_3^0 \tag{5.1}$$

is $K_{CaCO_3^0} = 10^{3.22}$ [114], from which it follows that at a typical experimental supersaturation given by the chemical activities $a_{Ca^{2+}} = a_{CO_3^{2-}} = 10^{-4}$ M, the activity of the ion pair will be $a_{CaCO_3^0} = 2 \times 10^{-5}$ M, i.e. the same order of magnitude as the individual ions. It follows that the adsorption of preformed ion pairs to the steps will dominate the kink nucleation process.

In the case of kink propagation, it has been experimentally established that step velocities significantly decline for highly non-stoichiometric solutions at fixed supersaturation (more on this below). In other words, when the ratio of calcium to carbonate ions is far from unity, the step velocities become very small even if the product of the calcium and carbonate concentrations are kept fixed. This would not be the case if kink propagation were dominated by ion pairs because the ion pair concentration depends only on the product of the calcium and carbonate concentrations and should therefore remain unchanged. We may therefore assume that kinks propagate via the attachment of individual ions. This assumption is also supported by our work in Chapter 4, where it was concluded that the adsorption of carbonate ions to kink sites is diffusion-limited, and therefore the attachment of carbonate ions should be no slower than that of ion pairs. We further assume

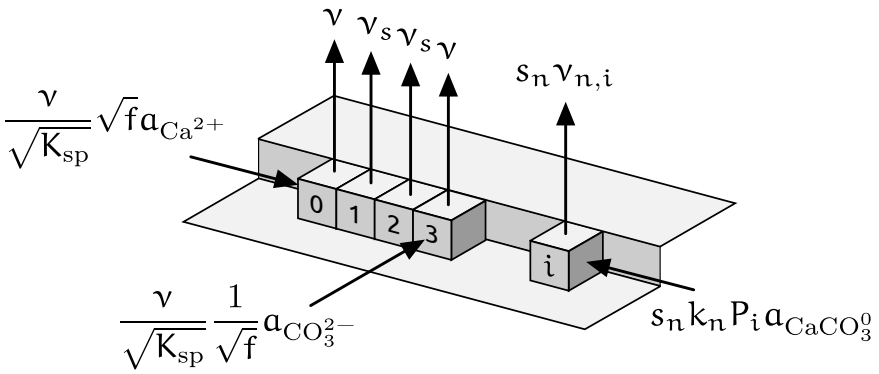


Figure 5.6: All of the unique parameters in our model for step growth, as described in the text.

that the kink sites dissolve one ion at a time, as is supported by the results of Chapter 4.

Now we turn our attention to reducing the number of parameters. Firstly, calcite has c-glide symmetry which immediately halves the number of parameters. Secondly, for each step type, the calcium and carbonate ions each have four unique kink sites that they can adsorb to. We make the assumption that the ions adsorb to, and desorb from, these four kink types at the same rate. In other words, that the crystal is only non-Kossel with respect to the different species. We further assume that the calcium- and carbonate-terminated kink sites dissolve at the same rate, ν. The difference in the abundance of calcium- and carbonate-terminated kink sites is therefore modelled by allowing the species to attach at different rates, i.e the relative attachment rates are constrained by $k_{Ca^{2+}}/k_{CO_3^{2-}} = f$. We also assume that all of the kink sites have identical solubilities, so that for each step the kink attachment rates can be expressed in terms of the kink dissolution rates

$$k_{Ca^{2+}} = \frac{\nu}{\sqrt{K_{sp}}} \sqrt{f} a_{Ca^{2+}} \tag{5.2}$$

$$k_{CO_3^{2-}} = \frac{\nu}{\sqrt{K_{sp}}} \frac{1}{\sqrt{f}} a_{CO_3^{2-}} \tag{5.3}$$

where $K_{sp} = 10^{-8.48}$ is the experimental solubility product. To see where these equations come from, consider first the case of $f = 1$. At equilibrium the chemical activities would be $a_{Ca^{2+}} = a_{CO_3^{2-}} = \sqrt{K_{sp}}$, by definition of K_{sp}, and we would have $k_{Ca^{2+}} = k_{CO_3^{2-}} = \nu$. Then, away from equilibrium, we have $k_{Ca^{2+}} = \nu a_{Ca^{2+}}/\sqrt{K_{sp}}$ and similarly for CO_3^{2-}. Finally, the f factors are the only terms that satisfy the constraint $k_{Ca^{2+}}/k_{CO_3^{2-}} = f$ while preserving the solubility product $k_{Ca^{2+}} k_{CO_3^{2-}}/\nu^2$.

We identified above that the kink nucleation is mediated by ion pairs, $CaCO_3^0$. However, our KMC code only supported the modification of single cells, and so we treated the nucleating pairs as single units in the simulation. We express their adsorption rates as $s_n k_n P_i a_{CaCO_3^0}$, where s_n is a constant that we shall set to unity for now (and may therefore be ignored, until later). To explain P_i, recall from Chapter 3 that the ion pairs do not adsorb to both steps, nor to each step site, with an equal probability. Rather, the free energy landscape makes adsorption to certain sites more likely to occur. The P_i terms are therefore the probabilities reported in Table 3.1 that we obtained from molecular simulation. The ion pair activity, $a_{CaCO_3^0}$, was defined above. And the remaining k_n term is the free parameter which measures the rate at which the pairs attach to one of the four sites, in other words, $k_n/4$ is the mean attachment rate (per molarity) of ion pairs to the step sites. The nuclei are assumed to dissolve at the rates $\nu_{n,0}$ and $\nu_{n,1}$ obtained from our earlier molecular simulations and also reported in Table 3.1.

In Figure 5.1, two other surface features were identified: corners and gaps. In our model we treat the corner sites like ordinary kink sites. We also attach ions to the gap sites at the same rate as ordinary kink sites, but we dissolve them at a different (lower) rate ν_s.

In summary, each step has three free parameters which we label with A or O superscripts for the acute or obtuse step, respectively, plus k_n which is shared between the steps:

$$\begin{aligned} &\nu^A, \quad f^A, \quad \nu_s^A, \\ &\nu^O, \quad f^O, \quad \nu_s^O, \\ &k_n \end{aligned} \tag{5.4}$$

In other words, there is a total of fifteen parameters for the model, eight of which were determined by our molecular simulations in Chapter 3, leaving seven free parameters to optimise. We now outline a procedure for optimising them with respect to experimental data. The experimental data is the dependence of step velocity on solution stoichiometry (the ratio of calcium to carbonate ions in the solution) at a fixed supersaturation of $\sigma = 0.9$ as reported in Hong et al. [5]. These data points are shown in Figure 5.7(a), and the optimisation procedure is as follows:

1. Pick ν^A.
2. Pick f^A and f^O.

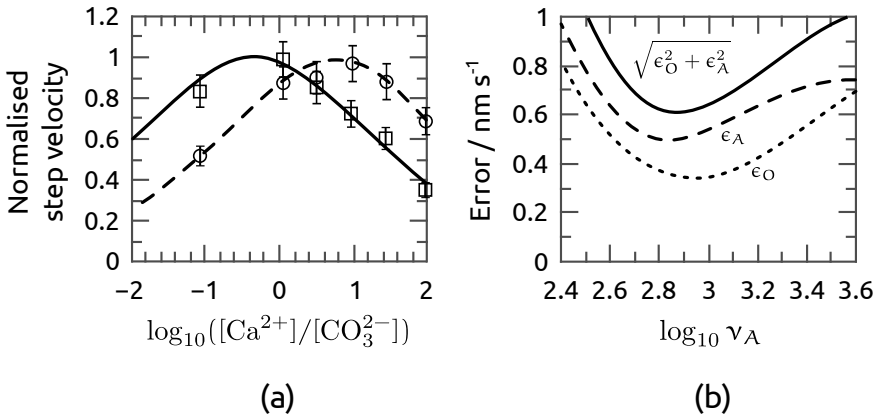


Figure 5.7: (a) The step velocity as a function of solution stoichiometry as computed by our simulations for the acute (solid line) and obtuse (dashed line) steps. The experimental points from Hong et al. [5] are also shown for acute (square) and obtuse (circle). (b) During the fitting procedure, the acute dissolution rate ν^A is optimised to maximise agreement with the experimental points in (a). ϵ_A is the RMS error between our simulations and the experimental points, and similarly for ϵ_O.

3. Optimise k_n and ν_s^A to simultaneously give the experimental acute step velocity under stoichiometric conditions, and zero velocity at saturation ($\sigma = 0$, where σ was defined in Eq. (1.1)).

4. The nucleation rate is then imposed on the obtuse step by k_n, leaving ν^O and ν_s^O to be optimised to reproduce the experimental obtuse step velocity and zero velocity at saturation.

5. Modify f^A and f^O and repeat from step 3 so as to optimise agreement with the experimental data points.

6. Modify ν^A and repeat from step 2, also until agreement with the experimental data is maximised.

The variation of the RMS error between our simulations and the experimental data as a function of the choice of the first parameter ν^A is shown in Figure 5.7(b). The velocity profiles that result from the optimal fit are shown in Figure 5.7(a). Note that the f parameters shift the velocity profiles along the abscissa, while the ν parameters modify the widths.

We note in passing that, if the four kink sites are given different dissolution rates, $\nu_0, \ldots, \nu_3$, then this has no observable effect on the velocity profile of Figure 5.7 as long as the mean dissolution rate, $\left(\frac{1}{4}\sum_i \nu_i^{-1}\right)^{-1}$, is preserved.

Table 5.1: The optimised growth parameters for our KMC model of calcite growth. See Figure 5.6 for the relation of each to the growth.

	ν	f	ν_s	P_0	P_1	$\nu_{n,0}$	$\nu_{n,1}$	k_n
Acute	$10^{2.87}$	1.5	3.2	0.33	0.21	$10^{7.53}$	$10^{8.49}$	$10^{10.8}$
Obtuse	$10^{3.04}$	0.2	2.7	0.28	0.18	$10^{8.64}$	$10^{7.58}$	

The final optimised parameters are all listed in Table 5.1. The kink sites are found to dissolve on the order of 10^3 s^{-1}, the same order of magnitude as predicted in previous work [4, 3, 2]. Note that our parametrisation suggests that ion pairs attach to the step sites on the order of 10^{10} $s^{-1}M^{-1}$. By comparison, the individual ions attach to the *kink* sites on the order of $\nu/\sqrt{K_{sp}} \approx 10^7$ $s^{-1}M^{-1}$. Putting aside errors in our model, this discrepancy could be attributed to the kink sites presenting larger adsorption barriers, coupled with the ion pairs being able to diffuse further across the calcite terraces, thus increasing the delivery rate.

Such a high attachment rate is needed due to the very rapid dissolution rate of the pairs from the steps. Indeed, the vast majority of the attaching ion pairs will dissolve before any subsequent ions attach. This actually proves to be a source of great inefficiency in the KMC simulations. For this reason, we adjust the s_n parameter introduced earlier to a value other than unity. Specifically, we set

$$s_n = \frac{10^4\ s^{-1}}{\min(\nu_{n,0}, \nu_{n,1})} \tag{5.5}$$

for each step. This means that the kink nuclei dissolve no quicker than 10^4 s^{-1}, and the attachment rates are scaled proportionately to maintain step velocities. Note that if a numerator much smaller than 10^4 s^{-1} is used then the attachment rate no longer scales linearly due to the morphological impact of having long-lasting nuclei.

We finally consider the morphology of our simulated crystal layer and compare it to the experimental morphologies traced from the AFM images of Stack et al. [4] and shown in Figure 5.8. Recall that the morphology has undergone a smoothing process described in Section 5.1.1. Under conditions of supersaturation and stoichiometric solution, the calcite step morphology is square (highly polygonal with a low kink density) with respect to the calcite lattice, in agreement with the experimental morphology. When the solution is made highly non-stoichiometric, with a much higher concentration of carbonate ions than calcium ions, the corner where the two obtuse steps

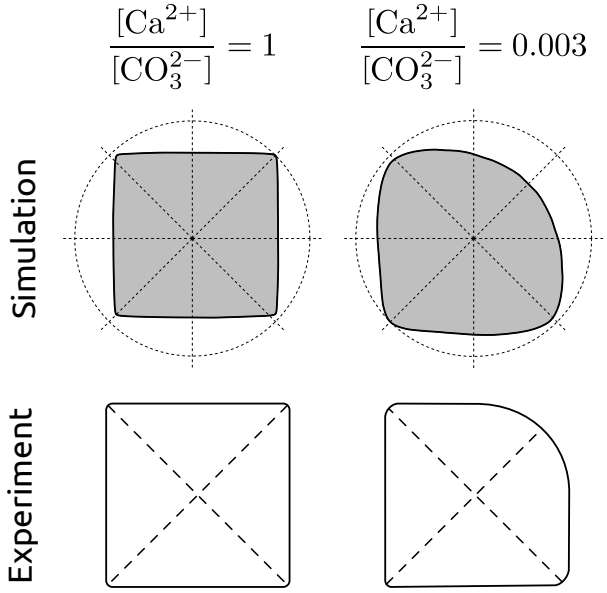


Figure 5.8: The morphology of our simulation, and a trace of the experimental morphology [4] under both stoichiometric and highly non-stoichiometric conditions. The widths of the simulated islands are approximately 300 nm.

meet becomes highly curved, as is also observed experimentally. This is a consequence of the obtuse step having a higher abundance of carbonate-terminated kink sites; the scarcity of calcium ions inhibits the propagation of kink sites to a greater extent than it does the acute step. This causes the kink sites to bunch up where the two steps meet, leading to the observed curvature. This demonstrates that the KMC model can provide insight into morphological processes in pure systems. We now turn our attention to impure systems.

5.3 KINK BLOCKING

Some impurities will bind to the surface features of a growing crystal for long enough to disrupt the growth process, but not long enough to become incorporated into the bulk. Aspartic acid is a well-known example of such an impurity [12, 115, 116]. If the impurity binds transiently to a kink site and completely halts the propagation of the kink, it is said to have a *kink blocking* effect, and this is the mechanism studied in this section.

Within a KMC simulation, kink blocking may be straightforwardly modelled by allowing a new species, the *impurity*, to attach to the kink and corner sites identified in Figure 5.1. By not introducing any

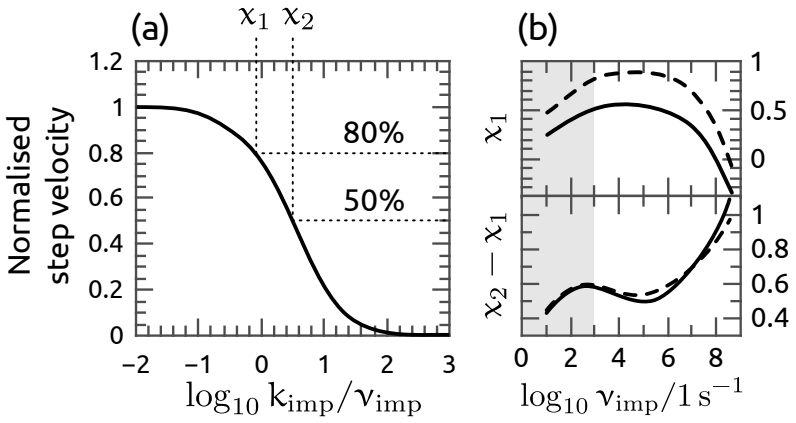


Figure 5.9: (a)A typical velocity profile for the calcite steps grown in the presence of a kink blocking impurity at various concentrations. (b) The precise offset and gradient of the profile displays a subtle dependence on the absolute rates.

further events into the event table, the impurity will automatically block any further growth in its vicinity. In this section we only model the attachment of impurities to calcium-terminated kink sites and, for now, we shall assume that the impurity attaches to these kink and corner sites at the same rate k_{imp}, with dimension per unit time, and dissolves at a rate ν_{imp}.

To study the kinetic effect of kink blocking, we computed the step velocity of both the acute and obtuse steps for the range $10 < \nu_{imp} < 10^9\ s^{-1}$ and $10^{-2} < k_{imp}/\nu_{imp} < 10^3$. The ratio k_{imp}/ν_{imp} gave rise to a similar (but subtly different) velocity profile irrespective of the absolute rate constants, and is shown in Figure 5.9(a). This profile reveals that kink blocking has no effect until the ratio reaches $k_{imp}/\nu_{imp} \approx 10^{-1}$ after which there is a rapid decline in step velocity with an increased attachment rate (i.e. increased impurity concentration). This same curve was previously predicted by the analytical model of Nielsen et al. [3]. However, there is a nuance that arises in our KMC model that cannot be reproduced with such analytical modelling, and that is the precise dependence of the curve on the absolute transition rates. To characterise this dependence, we define two numbers: x_1 and x_2, which correspond to the values of $\log_{10} k_{imp}/\nu_{imp}$ when the velocity is reduced by 80% and 50%, respectively, as illustrated in Figure 5.9(a). The value of x_1 therefore captures the offset of the curve along the abscissa, while the difference $x_2 - x_1$ captures the gradient of the curve. We plot in Figure 5.9(b) the variation of these two numbers, for the acute (solid) and obtuse (dashed) steps,

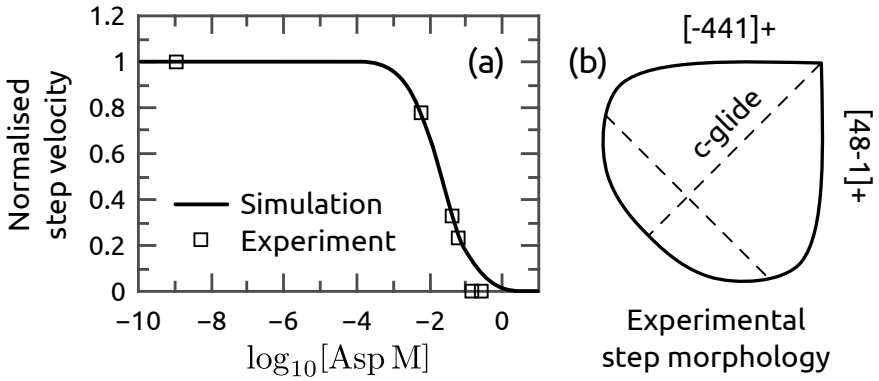


Figure 5.10: (a) The experimental [6] step velocities (squares) for the acute step of calcite grown in the presence of aspartic acid, compared with our optimised kink blocking profile from the simulations. (b) The experimental [7] step morphology at [Asp]=0.01 M.

with respect to the absolute dissolution rate ν_{imp}. Note that the calcite kink sites dissolve at a rate $\sim 10^3$ s^{-1}, and so we grey out the interval $\nu_{imp} < 10^3$ s^{-1} because here the impurities will reside at the kink sites for long enough to affect the growth via a step pinning mechanism (Section 5.4 below) and is therefore of minor relevance here.

It was previously mentioned that aspartic acid is believed to have a kink blocking effect on calcite. Fortunately, there exist experimental measurements of the calcite acute step velocities as a function of the aspartic acid concentration [Asp] for comparison [6]. These are reproduced in Figure 5.10(a) along with the simulated profile corresponding to the optimised rates:

$$\begin{aligned} k_{Asp} &= (10^{8.1}\ s^{-1}M^{-1})[Asp\,M] \\ \nu_{Asp} &= 10^{5.8}\ s^{-1} \end{aligned} \tag{5.6}$$

Clearly there is excellent agreement between the experiment and simulation. The optimised rates are also entirely plausible. For example, we predict the aspartic acid to attach to the kink sites approximately $10\times$ quicker than the calcium and carbonate ions do, per molarity. This is likely since aspartic will have a lower dehydration frequency than either ion. Aspartic acid was also found in molecular simulations by Nada et al. [117, 118] to exhibit strong indirect binding to the calcite terrace, suggesting that it will likely have a larger surface diffusion coefficient and thus be delivered to the kink sites at an increased rate. We now turn our attention to the morphological impact of kink blocking, returning to the case of aspartic acid afterward.

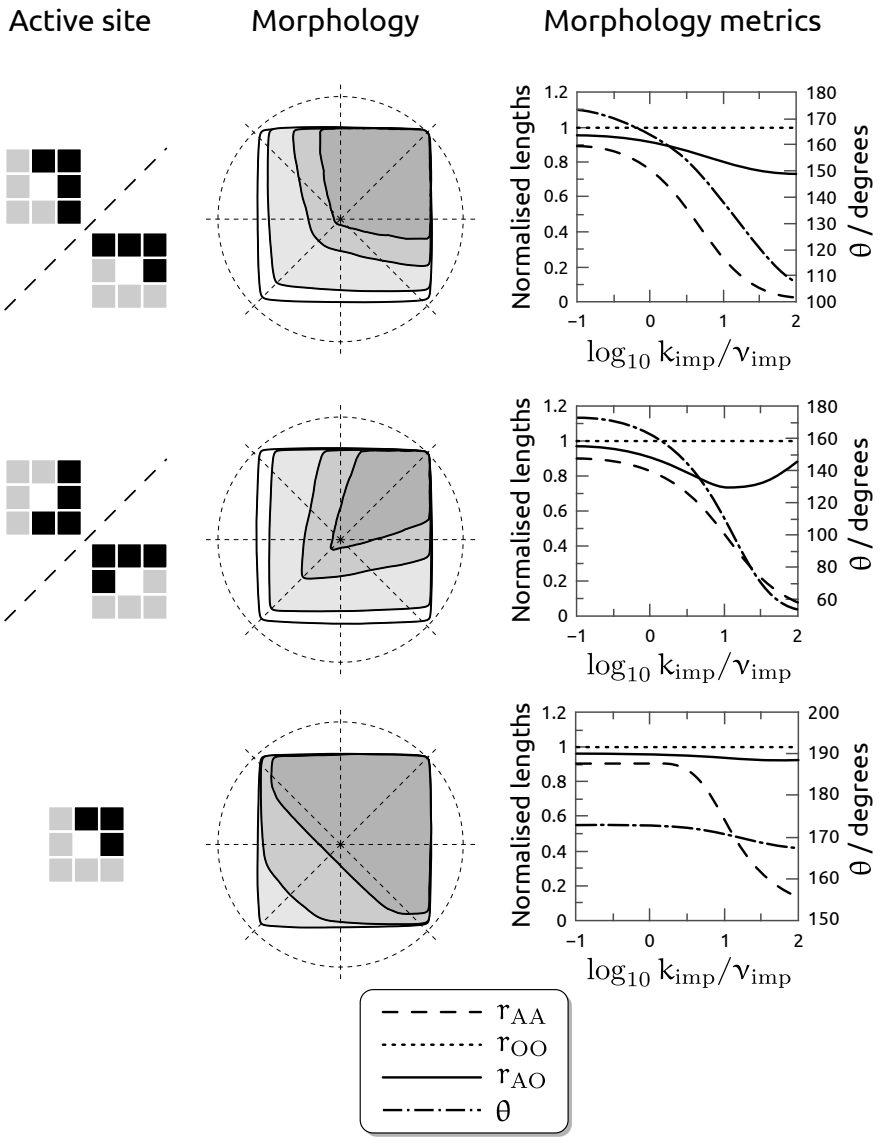


Figure 5.11: The relationship between the calcite site at which a kink blocking impurity is active and its morphological impact. The impurity will occupy the white squares in the active sites shown. The renderings, going from white to dark grey, correspond to $k_{imp}/\nu_{imp} = -1, 0, 1, 2$.

To investigate the morphological effect of kink blocking, we considered the case of a kink blocking impurity attaching preferentially to individual surface features. Specifically, to the Aa, Ao, Oa, and Oo kink sites, as well as the OO and AA corner sites, each in turn. Note that we treat the impurity as achiral, so it actually binds to two sites (the mirror image sites). These mirror image sites are depicted in Figures 5.11 and 5.12. We choose the same dissolution rate predicted for aspartic acid, $\nu_{imp} = \nu_{Asp} = 10^{5.8}\ s^{-1}$, and compute the morphologi-

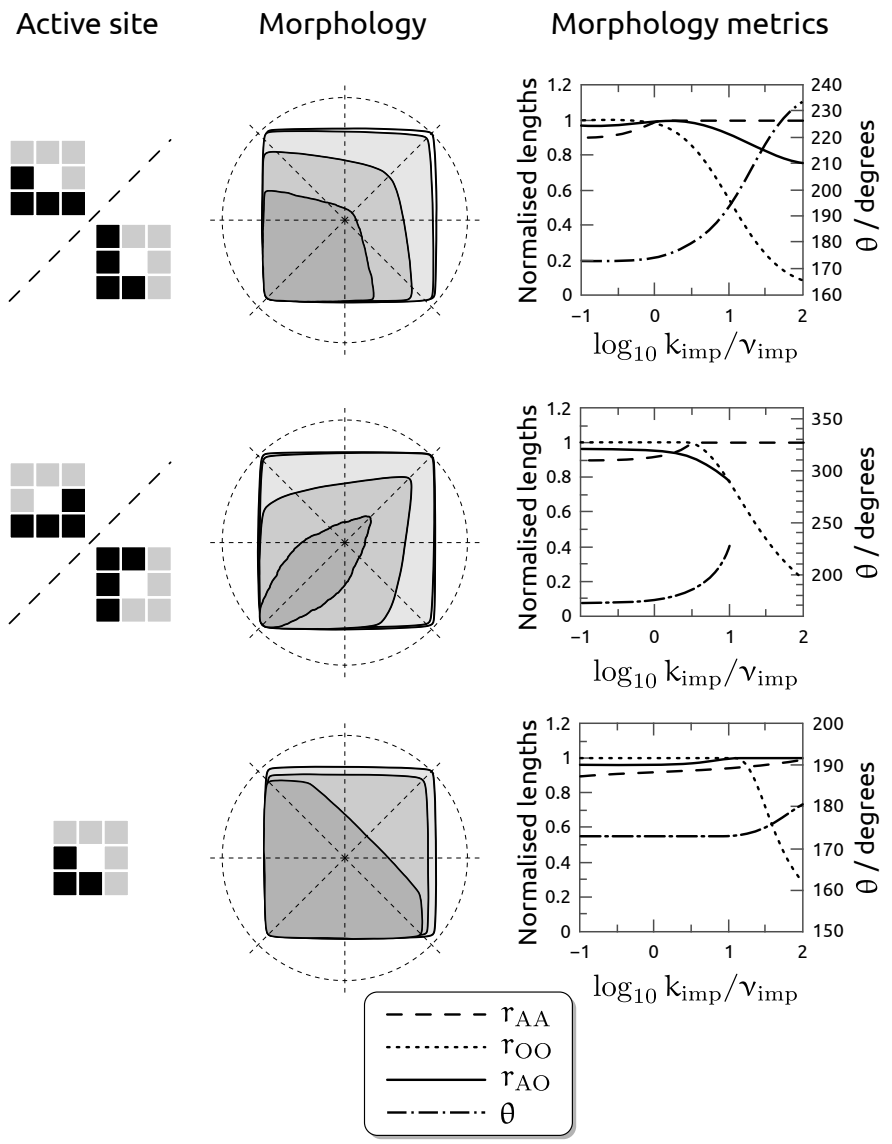


Figure 5.12: The relationship between the calcite site at which a kink blocking impurity is active and its morphological impact. The renderings, going from white to dark grey, correspond to $k_{imp}/\nu_{imp} = -1, 0, 1, 2$.

cal impact for attachment rates in the range $10^{-1} \leqslant k_{imp}/\nu_{imp} \leqslant 10^{2}$. For each morphology, we also compute the morphology metrics defined in Figure 5.5. The results are shown in Figure 5.11 for sites active at the acute steps, and Figure 5.12 for the obtuse steps.

There are a few observations to make. Firstly, when the impurities stabilise a particular kink direction, the steps maintain linearity but become angled in a predictable way. Surprisingly, however, once the steps have achieved a particular incline (the maximum being around 10°), an increased attachment rate does not further increase the gra-

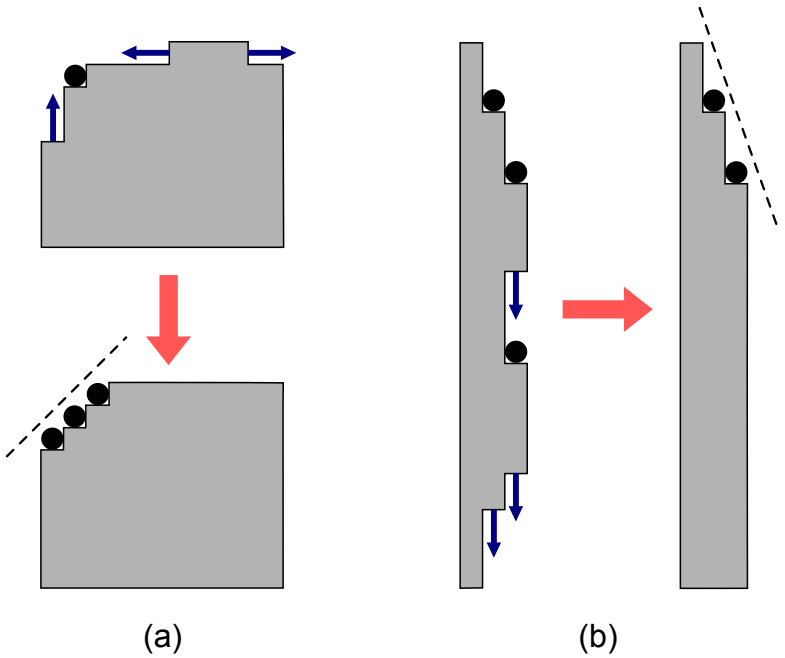


Figure 5.13: Schematic rationale for why the (a) corner- and (b) kink-blocking impurities (black circles) have the morphological effects reported in the previous two figures. (a) Kinks propagate until they reach a pinned corner which causes them to also become pinned corners, inviting the attachment of additional corner-blocking impurities which stabilise the diagonal step. (b) When kinks are only pinned in one direction, the opposing kinks grow out, resulting in an inclined step.

dient but simply reduces the step velocity. In the case of the corner sites being active, they have the effect of stabilising and expressing the polar faces that run perpendicular to the c-glide plane. One final observation is that the morphological effects on the obtuse step begin at a higher impurity attachment rate, with a more abrupt onset, as a consequence of the acute step having a higher abundance of calcium-terminated kink sites to which the impurity binds.

A rationalisation for the above morphological effects is provided schematically in Figure 5.13. In the instance of corner-blocking impurities, Figure 5.13(a), kink sites propagate until they reach a pinned corner, resulting in them also becoming transiently-pinned corners. These kink sites have now become active corner sites to which additional impurities may adsorb, thus stabilising the diagonal step. In the case of kink-blocking impurities, Figure 5.13(b), the kinks are only pinned in one direction. As a result, when the opposing kinks prop-

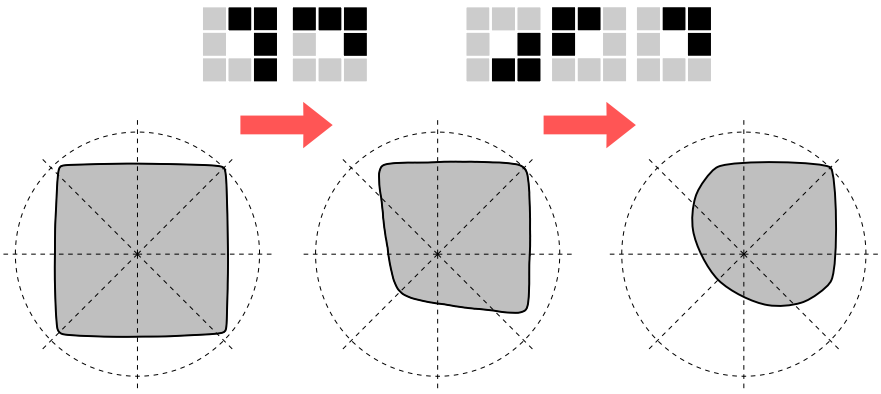


Figure 5.14: By stabilising kink blockers at particular sites (shown) the step morphology can be modified to match that observed experimentally (Figure 5.10(b)).

agate, the step transitions from a rough step to a slightly inclined step.

Returning to the case of aspartic acid. A trace of the calcite step morphology in the presence of [Asp]=0.01 M as observed under AFM imaging by Orme et al. [7] is shown in Figure 5.10(b). More accurately, the AFM images revealed a slight asymmetry about the c-glide plane due to the chiral nature of aspartic acid, but for our purposes we enforce the symmetry and consider just one half of the morphology. Notice that the obtuse steps remain flat and seemingly unaffected by the aspartic acid, while the acute steps exhibit a dramatic inhibition in the AA direction and a substantial curvature that is not seen in any one of our morphology snapshots in Figures 5.11 and 5.12. However, as will be shown next, it is possible to generate the same effect by activating the right combination of sites.

To reproduce the morphological effect of aspartic acid, we began by adsorbing the impurity to all calcium-terminated kink sites with the same rate, equivalent to the experimental concentration. The dissolution rates from all of the sites were also initially equal, and matched our predicted dissolution rate for aspartic acid. We then increased the dissolution rate (reduced the stability) of the aspartic acid at the Ao sites by a factor of 7; this number was chosen through trial-and-error, comparing by eye the resulting morphology to the experimental contour. This had the effect of tucking in the AA corner, reducing r_{AA}. The stability of the aspartic acid at the AA and AO (OA) corner sites was then increased by reducing the dissolution rate by a factor of 5. This gave rise to the stabilisation of the corners and a smooth curvature, as depicted in Figure 5.14, in accord with the experimental find-

ings (Figure 5.10). It should be noted, however, that this parametrisation is not unique. Each parameter could be adjusted in various ways to produce a similar morphology. Moreover, we cannot deduce whether the dissolution rates are responsible for the various stabilisations or the attachment rates. In fact, the molecular simulations of Nada et al. [117, 118] suggest that aspartic acid shows a strong reluctance to bind to particular sites in the first place due to the local water structure.

5.4 STEP PINNING AND GIBBS-THOMSON VIOLATION

Another mechanism by which impurities may inhibit growth is by attaching to the surface (whether that be the kinks, steps, or terraces) with a high density and for a relatively long duration, thus forcing the calcite to grow in the space between the impurities. If the impurities are spaced closely together then the calcite can only propagate if it can achieve a sufficiently high kink density which increases its solubility and thus effectively reduces its local supersaturation. As a result, if the impurity spacing is below some critical length $L_c(\sigma)$, which is a function of the supersaturation, the calcite cannot grow between them. The impurities are said to *pin* the growth.

We demonstrate this effect in Figure 5.15 where a calcite step is confronted with a spread of permanently fixed impurities. If the step is grown at a low supersaturation of $\sigma = 0.3$ then it is unable to proceed and the impurities completely pin the growth. However, at an increased supersaturation of $\sigma = 0.9$ the calcite is able to penetrate. In doing so, the step becomes highly roughened with the same features observed when calcite is grown in the presence of abalone shell proteins that are believed to function as step pinners [119].

Despite this apparent success, there is a fundamental discrepancy between our simulations and the experiments. To correctly capture the kinetic and morphological impact of step pinning impurities, the calcite step segment that grows between a pair of impurities separated by a distance L must exhibit the correct speed dependence on L. Theoretically, one would expect it to obey the Gibbs-Thomson relation [120]

$$\frac{v(L)}{v(\infty)} = 1 - \frac{\exp(\sigma L_c(\sigma)/L) - 1}{\exp(\sigma) - 1} \tag{5.7}$$

and indeed, this relation agrees precisely with our KMC measurements, as shown in Figure 5.16(a) for the obtuse step at a supersatu-

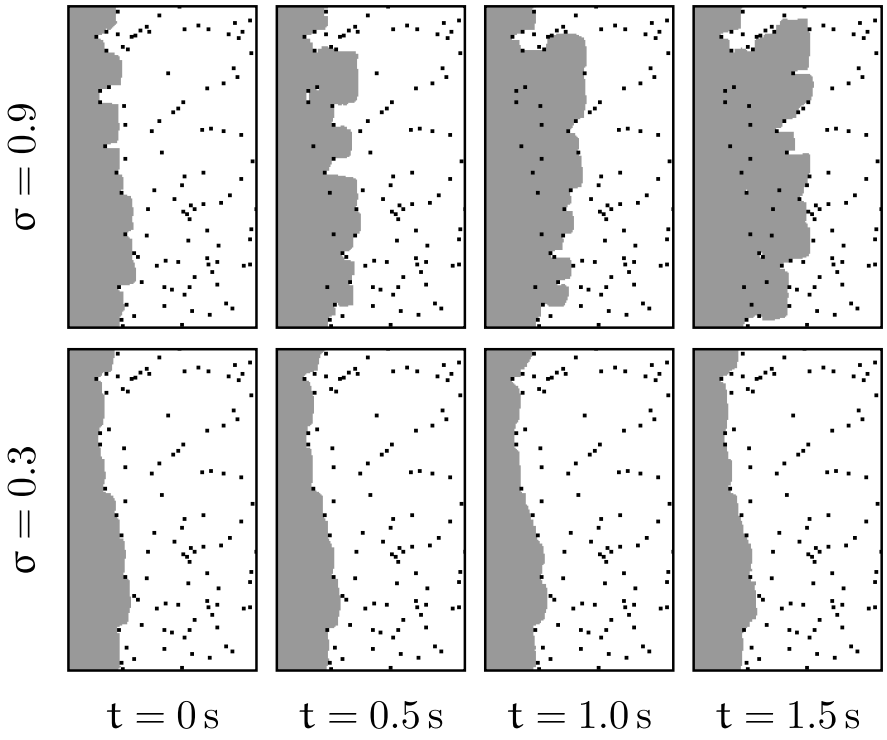


Figure 5.15: Two-dimensional projections of the growth of calcite at high and low supersaturations in the presence of step pinning impurities. The cell heights are 82 nm.

ration of $\sigma = 0.668$. However, this is not what is found experimentally. Instead, the step velocity achieves maximum velocity by $L \approx 2L_c$, and even more strangely, this same surge in growth is observed for all supersaturations, in contrast to the prediction of Eq. (5.7). Moreover, the experimental critical lengths are approximately $10\times$ greater than those produced in our simulations. It should be noted that this same phenomenon has been observed in other low kink-density crystals and has been identified by Chernov [121] as one of the major unresolved problems in crystal growth kinetics.

The conventional wisdom has been that this effect is something that naturally arises when a crystal has a sufficiently low kink density, due to the diffusion and annihilation of kinks generated at the corners [122]. However, such a mechanism should arise in our KMC simulations if it were in fact responsible, and yet it does not. We have also investigated and ruled out two other possible mechanisms, as described next.

In all of the experiments from which the velocity profiles $v(L)$ are obtained, the L-sized step segments are generated at screw dislocations. Such defects naturally generate strain fields that presumably

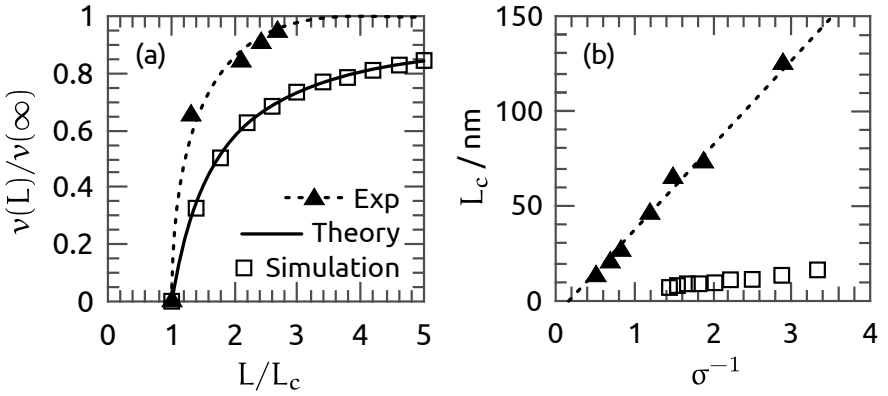


Figure 5.16: (a) Step velocity dependence on step length. Our KMC simulation matches the theoretical Gibbs-Thomson relation but not the experiments. (b) The dependence of critical length on supersaturation found experimentally and in our simulations. The former are approximately $10\times$ larger than the latter.

increase the local solubility of the crystal. This led us to propose that only once the step segment has grown large enough to escape the undersaturation imposed by the strain field is it able to nucleate growth. Once these nuclei become established far away from the dislocation centre, the kinks may then diffuse along the step and into the undersaturated region, giving rise to growth. We modelled this by increasing the dissolution rate of the kink sites by a factor $\exp(A/r)$, where r is the distance between the kink and the edge of the step segment (the centre of the screw dislocation), and A is a parameter chosen to reproduce the experimental critical lengths. This mechanism was able to give rise to the large critical lengths as well as the sudden onset of step velocity. However, the crystal was unable to grow into the inner, undersaturated region, resulting in a hole with a radius comparable to the critical length. Needless to say, this is not what is seen experimentally.

A second idea was to recognise that it is highly unlikely that the kink sites that propagate in opposite directions along a step (i.e. Aa and Ao, or Oa and Oo) have the exact same solubilities as each other. As a consequence, the saturation state for a step will actually correspond to, say, the Aa kink site growing at the same rate that the Ao kink site is dissolving. This means that the features of the steps must slowly diffuse across the step length. Now, when a nucleus forms on a step, it must reach a certain critical size before it is likely to become established as a permanent part of the crystal, which will take time. However, if the surface features are drifting to the edge of the

segment, whereupon they are annihilated, then this places a length-dependent limit on how long the nuclei have to become established. We modelled this process by adding a constant to the growth rate of the Aa kink site, and subtracting the same constant from the Ao kink site, which preserves the solubility of the step but gives rise to drift. By increasing this constant, we could indeed increase the critical length, also giving rise to a more abrupt surge in step propagation. However, doing so came at a cost. Firstly, it broke the linear dependence of L_c on σ^{-1} seen in Figure 5.16(b), and secondly, it caused the step to become angled to an extent that would be easily resolved under AFM. So we can also rule out this mechanism.

Unfortunately, until this problem has been resolved, step pinning, and possibly other inhibitory mechanisms, will remain beyond the reach of accurate, predictive modelling.

5.5 SUMMARY

We have developed a coarse-grained model for the growth of calcite crystals, whereby the underlying processes and relevant rates have been obtained through a combination of computational and experimental data.

This chapter proves that even a fairly rudimentary description of calcite growth has great potential for capturing both the kinetic and morphological processes. Both in the case of a pure system, as illustrated by the agreement between our simulations and the experiments for non-stoichiometric solutions, but also for impure systems. For example, our systematic study of kink blocking revealed the relation between individual molecular mechanisms and their broader morphological consequences. Also of interest is that our model captures kinetic nuances that analytical modelling cannot, due to the effect that absolute transition rates have on morphology. By accounting for these nuances, we were able to provide an estimate for the dissolution rate of aspartic acid from the calcite kink sites.

Lastly, we reveal that while our model can exhibit the step pinning phenomenon, it is incapable of providing an accurate account of the kinetic and morphological effects of real impurities. This is because our measured dependence of the velocity of a step segment on the length of the segment conforms with the theoretical Gibbs-Thomson relation, whereas the experimental crystals violate it. And until the responsible mechanism has been identified and incorporated into our

model, step pinning will remain beyond predictive modelling. This is a major setback that must be resolved.

6

SOFT Ttemplating

This chapter presents two separate simulation studies of surfactant self-assembly at solid-aqueous interfaces, with the aim of better understanding the role that the surfactants play in soft templating at two very different scales. In the first section we study the self-assembly of sodium dodecyl sulphate (SDS), $CH_3(CH_2)_{11}[SO_4]^-Na^+$, monolayers at the titania-water interface for various polymorphs and orientations of the substrate, and quantify the titania-monolayer adhesion. This allows one to predict and rationalise the influence that the monolayers may have on polymorph and orientation selection during the nucleation and growth of titania. The second section identifies the structures formed by cetyltrimethylammonium chloride (CTAC), $CH_3(CH_2)_{15}[N(CH_3)_3]^+Cl^-$, aggregates on two different orientations of silicon, the results of which provide insight into recent experimental findings that periodic mesoporous organosilicates may be grown on only a specific orientation of silicon. The atomistic structure of each surfactant can be seen in Figure 6.1.

6.1 SDS self-assembly on titania substrates

6.1.1 *Introduction*

Titania (TiO_2) is a technologically important material with a wide range of applications, notably in solar cells [123, 124], photocatalysis [125, 126], biomaterials [127], pigments [128], and sunscreens [129]. It has two primary polymorphs, the equilibrium *rutile* phase and the metastable *anatase*. The soft-templating approach to synthesising meso-structured titania usually involves a co-assembly process of a

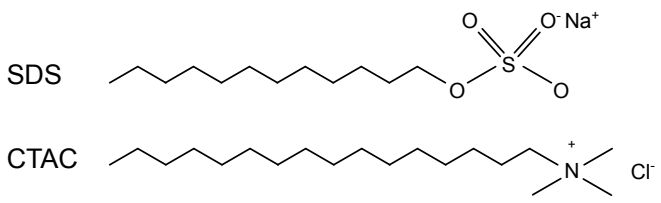


Figure 6.1: The atomistic structures of the SDS and CTAC surfactants.

titania precursor with a surfactant template, a common choice being SDS. For example, Mitra et al. [130] used SDS micelles self-assembled in an oil/water microemulsion as a template to produce titania with high surface area mesoporosity. Similarly, Wang et al. [131, 132] produced highly crystalline mesoporous metal oxides by using SDS, in the form of a functional surfactant matrix, as an organic template. Chen et al. [133] adsorbed SDS on anatase nanocrystalline building blocks which served to direct the assembly of the blocks into a three-dimensional mesoporous material.

The properties of nanoparticles, e.g. their photocatalytic properties, are sensitive to the size and morphology of the nanoparticle, and both of these attributes can be controlled with surfactants. For instance, Liao et al. [134, 135] have grown titania nanoparticles in the presence of various surfactants, generating a range of morphologies. Notably, when grown in the presence of SDS, the particles were found to take on cubic morphologies, irrespective of the precursor used.

The SDS surfactants have also been used to functionalise graphene sheets and facilitate the self-assembly of nanocrystalline titania, resulting in titania-graphene hybrid materials with the potential of being used as electrodes in Li-ion batteries [136].

To provide insight into the SDS-titania interactions, we simulate the self-assembly of SDS monolayers at both the titania-vacuum and titania-water interfaces for two different polymorphs, four different orientations, and eleven different monolayer densities, totalling 176 simulations, and compute the relevant adhesion energies. Both the nucleation and morphological effect that surfactants have on a precipitating crystal may be understood by the energy with which they bind to the various faces; those to which the surfactants bind most strongly are more likely to be expressed in the growing crystal. The surfactant structure that forms at the interface, which is responsible for titania mesoporosity, will also be influenced by the strength of binding.

6.1.2 *Methodology*

We have employed molecular dynamics to simulate the adsorption of SDS monolayers on the (100), (110), (011), and (001) faces for both the anatase and rutile polymorphs, see Figure 6.2. The notation A(hkl) and R(hkl) shall be used to refer to the (hkl) surface of anatase and rutile, respectively. For each surface the simulations were repeated

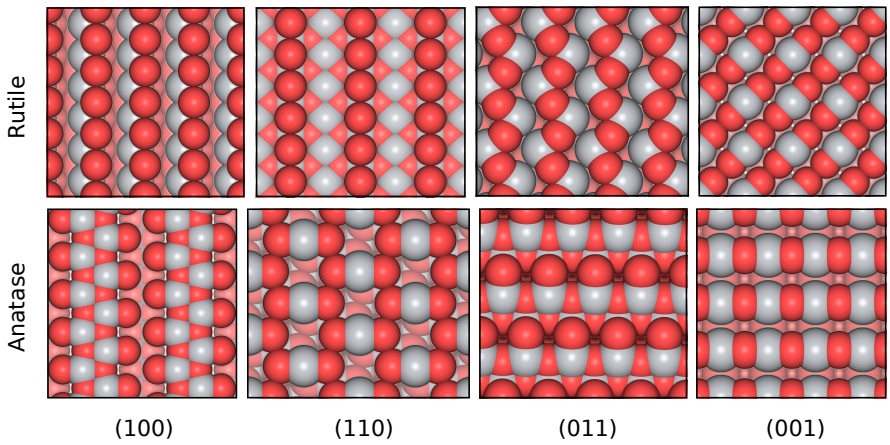


Figure 6.2: A top-down view of the eight titania surfaces upon which the SDS monolayers are adsorbed. Each atom is represented by its van der Waals surface; oxygen in red and titanium in grey.

for eleven surfactant densities in the range $\sim 2.2\ \mathrm{nm}^{-2}$ to $\sim 4.4\ \mathrm{nm}^{-2}$. For a density much below this range the surfactants are too sparse, leading to behaviour uncharacteristic of monolayers, such as local aggregation of surfactants and tails interacting with the surface. A density above this range is too high and results in the ejection of surfactants from the surface. All stable monolayer configurations are therefore expected to fall within this range.

Each titania surface had dimensions as close as possible to 3 nm $\times$ 3 nm $\times$ 1.5 nm, with periodic boundary conditions. A vacuum slab of 7 nm was added to attenuate self-interaction between the images in the z-dimension. In the initial configuration of each simulation, the surfactants were placed 0.4 nm clear of the surface in all-trans configuration with their tails aligned normal to the surface and each sodium cation positioned between the sulphate and the surface. They were randomly distributed across the surface (using a Poisson disc sampling algorithm to prevent overlap).

In addition to the above, two further interfaces were constructed, each with an initially stable coverage rather than a random configuration. These were a hollow c(2x2) coverage on an A(001) surface and a bridged (2x1) coverage on a R(110) surface. The relaxed configuration of the latter is illustrated in Figure 6.3. These stable coverages enabled us to determine the effect, if any, that the initial monolayer configuration had on the final adhesion.

In the present work we are interested in computing the adhesion energies between the various monolayers and the various surfaces. The adhesion energy between a titania surface and some component

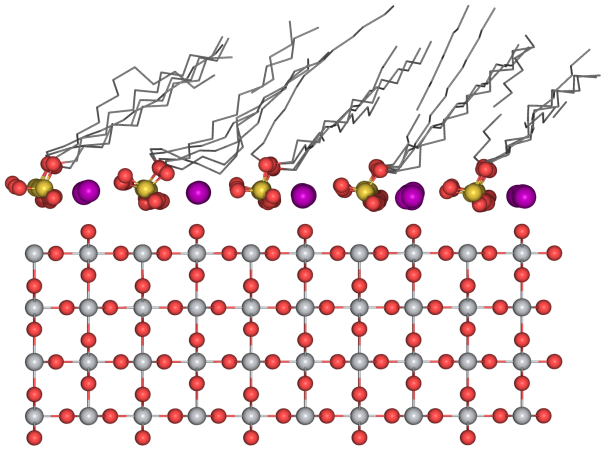


Figure 6.3: A monolayer with a bridged (2x1) coverage relaxed on the rutile (110) surface at 300 K. Oxygen in red, titanium in light grey, sodium in purple, sulphur in yellow, and the united carbon chains in dark grey.

X of the monolayer (e.g. a particular atomic species), at one particular time-point t, is computed as follows

$$\beta(X,t) = A^{-1}[E(X+TiO_2,t) - (E(X,t) + E(TiO_2,t))], \tag{6.1}$$

where $E(TiO_2,t)$ and $E(X,t)$ are the configurational energies of the titania surface and the X component at time t, each in isolation, and $E(X+TiO_2,t)$ is that when they are combined. A is the surface area. The adhesion energies $\beta(X)$ reported below were obtained by equilibrating the system for 60 ps and then averaging $\beta(X,t)$ over a production period of 40 ps.

With this method the adhesion energy between the surface and the individual components of the monolayer, e.g. $\beta(Na)$, were computed. The total adhesion energy between the surface and the monolayer, denoted β, must equal the sum over all components of the monolayer

$$\beta = \sum_X \beta(X) \tag{6.2}$$

In contrast to the above simulations, an experimental setting would include water. A more relevant energy would therefore be the difference between the monolayer-crystal and water-crystal adhesion energies, which quantifies the preference for each surface to bind to a monolayer rather than water. To this end, a second batch of simulations was carried out, analogous to those performed in a previous study of carboxylic acids on calcite surfaces [63]. For each surface,

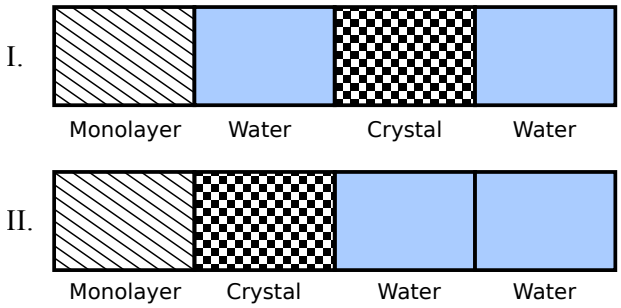


Figure 6.4: Schematic representation of the two simulations performed to account for the displacement of water during nucleation.

and for each monolayer density, the two simulations depicted in Figure 6.4 were performed. Simulation I consisted of a slab of water sandwiched between the surface and monolayer in both directions. In Simulation II the monolayer was adsorbed directly onto the crystal while the water was repositioned on the other side of the crystal. Since both simulations had an equal number of ions, the difference in configurational energies, E_I and E_{II}, gives

$$\beta' = A^{-1}(E_{II} - E_I) = \beta_{cm} - \beta_{mw} - \beta_{cw} \tag{6.3}$$

where β_{cm}, β_{mw}, and β_{cw} are the crystal-monolayer, monolayer-water, and crystal-water adhesion energies respectively, and A is the surface area. Note that $\beta_{cm} \equiv \beta$ as defined in Eq. (6.2). This modified adhesion energy, β', accounts for the displacement of water during nucleation and was averaged over a 40 ps production period having been equilibrated for 100 ps.

6.1.2.1 *Simulation details*

The SDS molecules and their interaction with the titania surfaces were described using the force field of Domínguez et al. [137, 64]. This model consists of 12 united carbon atoms attached to an explicitly modelled sulphate headgroup, plus the sodium counterion. The bond lengths and the angles are constrained by harmonic potentials while the torsional angles in the tail and the headgroup are described by the Ryckeart and Bellemans potential and the cosine-form potential, respectively. The SDS and titania interact via a Lennard-Jones potential. The water interactions were also taken from the work of Domínguez et al. and consist of SPC-water with Lennard-Jones potentials acting between the water and all other ions in the system.

To model the titania we diverge from the work of Domínguez and use the force field of Matsui and Akaogi [138] which reproduces sur-

face energies, lattice energies, and polymorphic structures more accurately than any alternative force field [139, 140, 141]. The electric charges assigned to the titanium and oxygen ions in this force field, however, are 1.91 times larger than in the model of Domínguez. In order to retain the correct interatomic spacing, the ϵ-LJ parameters assigned to the titanium and oxygen ions in Ref [64] had to be scaled by 1.91^2.

The molecular dynamics simulations were carried out using the `DL_POLY Classic 1.9` [142] software package in the canonical ensemble (for the vacuum simulations) and isobaric-isothermal ensemble (simulations containing water). The smooth particle mesh Ewald summation method was employed to handle long-range electrostatics with a precision of 10^{-5}. The Nosé-Hoover algorithm maintained the desired temperature (pressure) with a relaxation constant of 0.5 ps (2.0 ps). The Verlet algorithm with a time-step of 1 fs was used to integrate the equations of motion.

6.1.3 *Results and discussion*

The crystal-monolayer adhesion energies in vacuum, β, were computed as previously described and are presented in Figure 6.5(a). It can be seen that they range from -0.7 to -1.6 Jm^{-2}. The three main contributions to β were found to come from the interaction between the titania and the sodium cations, the sulphur cations, and the three peripheral oxygen anions in the OSO_3 functional group of the surfactants. These three contributions are also plotted in Figures 6.5(c)-(e).

The adhesion energies for the two systems that were initially constructed to have stable configurations, viz. R(110)/(2x1) and A(001)/c(2x2), are also shown in Figure 6.5(a), denoted by squares. Both data points correspond to an increased adhesion energy of approximately 0.05 Jm^{-2} over the corresponding initially-random structures. We conclude that the starting monolayer configurations in our simulations have had a negligible effect on the final energies.

The components of the adhesion energies reveal that the sodium cations, which bind to the undercoordinated oxygen anions on the surfaces, invariably dominate the adhesion. It follows that the stronger the sodium cations bind to a surface, the stronger the entire monolayer binds. This role of the cations in binding indicates that the titania surfaces are mostly anionic in character, in agreement with the

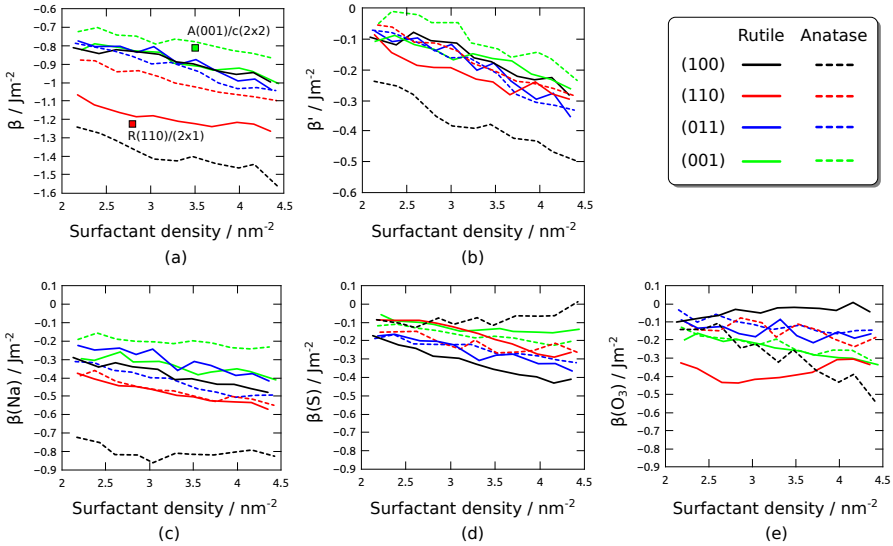


Figure 6.5: Various adhesion energies as a function of monolayer density: (a) the monolayer-crystal adhesion energies in vacuum, where the squares correspond to the labelled coverages (see text), (b) the energetic preference for binding to the monolayer rather than water, (c) (d) (e) the adhesion energies between components of the monolayer and the titania substrate.

experimentally observed linear uptake in SDS adsorption on titania films when the pH is decreased below 7 [143].

The monolayers bind more strongly to the surfaces A(100) and R(110) than to the others by a significant amount, approximately 0.4 Jm^{-2} and 0.2 Jm^{-2}, respectively, for all densities. The components of the adhesion energies reveal the features that make these two surfaces distinct from the rest. The first point to note is that the sodium cations bind more strongly to these two surfaces than to the rest; this is especially true for A(100). The second point is that the O_3 anions bind to these two surfaces more strongly than the sulphur cations do, which is not the case for any of the other surfaces. It follows that there are both anionic- and cationic-friendly regions on these two surfaces, in contrast to the predominantly cationic-friendly nature of the other surfaces. This lateral polarity of A(100) and R(110) is evident in the x-density plots of the sodium and sulphur ions for each surface, as shown in Figure 6.6. It can be seen in each case that the sodium cations are concentrated around the undercoordinated oxygen anions while the sulphur ions (sulphate groups) bestride the undercoordinated titanium cations. In summary, the monolayers organise themselves to achieve electrostatic complementarity with the surfaces. Moreover, they are able to match the charge periodicity of

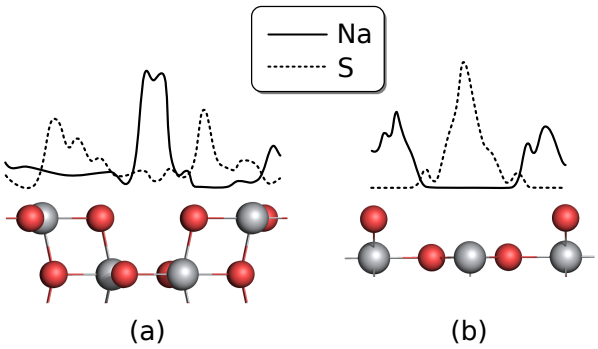


Figure 6.6: The normalised x-density of the sodium (solid) and sulphur (dashed) ions on the (a) A(100) and (b) R(110) surfaces with respect to the surface features shown. Compare to Figure 6.2. These distributions were computed for a surfactant density of ~ 3.3 nm^{-2}.

these two surfaces. The matching of charge periodicity is known to maximise adhesion [144].

On the A(100) surface, two rows of sulphates are forced next to each other with negligibly few sodium cations between them to act as an adhesive. This is an undesirable configuration and reveals that the driving force for electrostatic complementarity with the substrate dominates over the intra-monolayer interactions in directing the assembly.

The modified adhesion energies β', which account for water displacement, are presented in Figure 6.5(b). Comparing them to Figure 6.5(a) reveals that the process of displacing water has little qualitative effect on the conclusions: A(100) remains the strongest-binding surface (-0.2 to -0.5 Jm^{-2}) with R(110) mostly second (-0.1 to -0.3 Jm^{-2}) and A(001) the weakest (0 to -0.2 Jm^{-2}). The water does, however, reduce the difference between the adhesion energies of all of the surfaces quite significantly. This is presumably because the strength with which the water and the monolayers bind to each surface are roughly proportional. The fact that all of the adhesion energies β' are negative suggests that the monolayers will indeed enhance nucleation irrespective of the surface that forms.

The findings of this section help to rationalise those of Núñez-Rojas and Domínguez [64]. In their simulations, it was found that when SDS was adsorbed at the titania/water interface, it formed micelles on R(001) and hemi-micelles on R(100); in each case the SDS was bound to the surface mainly through the hydrophobic tails that were shielded from the water. However, on R(110), they found that the

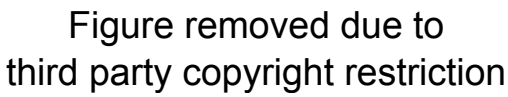


Figure 6.7: SEM images of pyramidal sodium silicate cubic mesophases grown on the Si(100) substrate. Shown after two hours, and then a further two hours in fresh solution. Taken from [8].

headgroups bound predominantly to the surface while the hydrophobic tails were exposed to the water. The favourable binding of SDS to R(110) found here explains this result and we would predict similar behaviour for A(100) while the other anatase surfaces will likely give rise to (hemi)micelles.

6.2 CTAC SELF-ASSEMBLY ON SILICON SUBSTRATES

6.2.1 *Introduction*

Periodic mesoporous organosilicates are glasses that exhibit a periodic network of mesopores, with applications in catalysis, sensing, and drug delivery [19]. They are typically synthesised from an aqueous solution of structure-directing surfactants and a silicate precursor, and have been grown on substrates such as graphite [145], mica [146], and glass [147]. In these cases, the surfactants are believed to self-assemble upon the substrate to form a mesostructured template that subsequently aids nucleation and may control the orientation of the overgrowing crystal.

The present study is motivated by recent experimental findings [8] involving a solution of 1,2-Bis(trimethoxysilyl)ethane (BTME) precursor, CTAC surfactants, sodium hydroxide, and water, in the molar ratio of 1:0.6-0.9:2.10:353, respectively. The solution underwent vigorous stirring at room temperature and was aged overnight. The aged solution was then transferred to a Teflon container and a small silicon wafer suspended vertically in it. The solution was subsequently

heated to, and maintained at, a temperature of 100 °C. This was carried out for two different silicon wafers, one with Si(100) faces and the other with Si(111). Within two hours, the Si(100) wafer was found to be covered with cubic-phased ($Pm\bar{3}n$) mesoporous amorphous sodium silicate pyramids, whereas Si(111) gave rise to no discernible growth. The pyramids grew from myriad nucleation sites and were all oriented in the same direction. As they encountered each other they would merge, resulting in a single crystal (*crystal*, in this context, refers to the periodic mesoporous structure). These pyramids on Si(100) are shown in Figure 6.7 after 2 hours in the oven, and after the resulting silicon wafer underwent a further 2 hours but in a fresh solution.

In this section we aim to explain (i) why the crystals grow on Si(100) but not Si(111), and (ii) why the crystals are unidirectional. Since oligomerisation occurs on a much longer time-scale than micellisation [148], it is expected that the silicon surface will be in direct contact with a surfactant layer. Furthermore, it is difficult to imagine that the amorphous sodium silicate itself should have any preference regarding the substrate orientation, and so it seems likely that the structures formed by the surfactants on the two different surfaces will be responsible for the experimental findings. We therefore simulate the aggregation of CTAC surfactants at the silicon-aqueous interface for the non-oxidised, reconstructed Si(100) and Si(111) surfaces. The drivers for adsorption on these two surfaces will differ substantially from those previously discussed. For instance, while adsorption on metal oxides are governed predominantly by ionic bonding, hydrogen bonding, and hydrophobic interactions [149], adsorption on silicon will be directed by soft epitaxial forces: van der Waals and induced image charges. Modelling image charges [150], however, is very challenging and we expect it to play a negligible role in our system due to the mostly-localised covalent bonding in silicon and the predominance of the neutral alkane group in the surfactants.

6.2.2 *Methodology*

Two simulations were performed. The initial configuration of each consisted of a silicon surface, Si(100) or Si(111), above which a layer of CTA^+ surfactants were aligned normal to the surface in all-trans configuration with the hydrophilic headgroups pointing towards the surface. The chloride counterions were placed between the surface

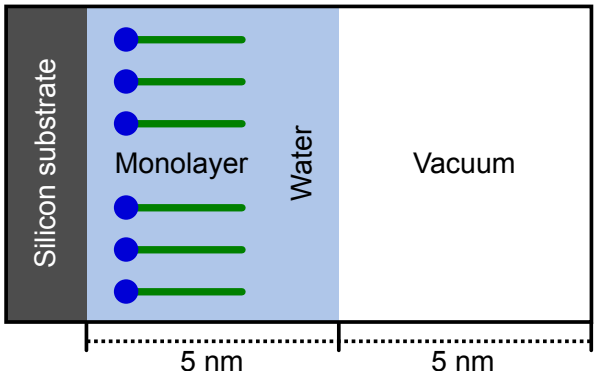


Figure 6.8: A schematic representation of our initial simulation configurations. The surfactants are randomly scattered within the water with their hydrophilic headgroups (blue circles) facing the silicon substrate.

and the headgroups. The structure was then surrounded by a lattice of water molecules, with any overlapping molecules removed. The amount of water was chosen such that the thickness of the water layer, upon relaxation, was approximately 5 nm. A vacuum slab of 5 nm was then added above the water; in addition to segregating the periodic images, the vacuum slab enabled the system to maintain zero pressure while sampling the canonical ensemble. The water does not fill the vacuum due to the energy cost associated with the water vaporising or with the surface becoming too non-planar. The purpose of aligning the surfactants in such an unfavourable initial configuration was to be sure that we did not bias the final configurations by trapping the monolayer in a local free energy minimum. The initial simulation configuration is summarised schematically in Figure 6.8.

The concentration of surfactants at an interface is known to vary non-linearly with the bulk concentration, as described by the adsorption isotherm. However, while the bulk concentration of CTAC used in the experiments is known to be 2.0 mM, we are unaware of any published CTAC/silicon isotherms. We instead used the CTAC/silica isotherms [151] as an approximate guide which gave an adsorption density of 1.9 nm^{-2} for the desired bulk concentration.

6.2.2.1 *Silicon surfaces*

Both Si(100) and Si(111) are known to undergo significant reconstruction, as depicted in Figure 6.9. In the case of Si(100) we constructed the c(4 × 2) buckled configuration using density functional theory (DFT), details of which can be found in the next section. We note

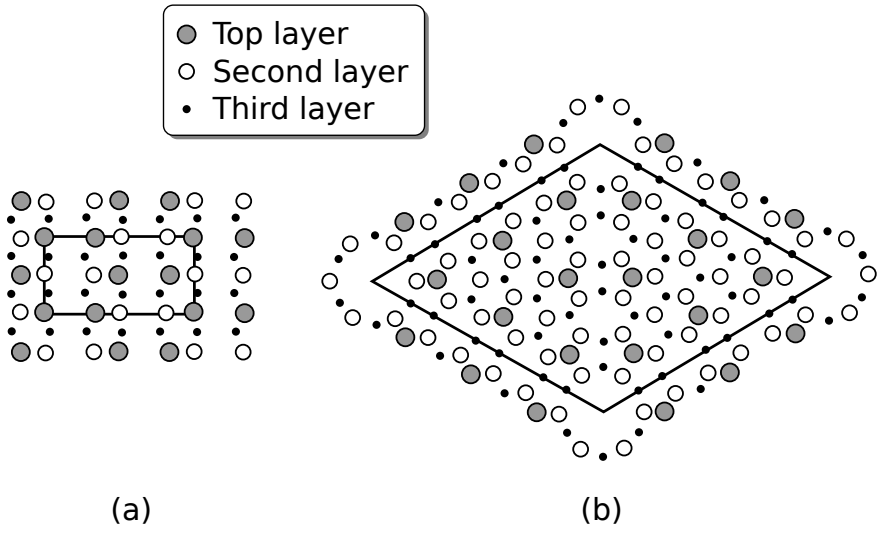


Figure 6.9: Structure of the reconstructed (a) Si(100)-c(4×2) and (b) Si(111)-p(7×7) surfaces with the conventional unit cells outlined.

that while there remains some debate regarding whether or not the surface dimers buckle experimentally [73, p. 268], the buckling appears to be irrelevant to our findings. In the case of Si(111), the surface undergoes a very elaborate reconstruction that requires a large p(7 × 7) unit cell, making it impractical to build this surface with DFT as we did for Si(100). Instead, the configuration was constructed by reproducing the geometry determined from low-energy electron diffraction analysis [152].

In our simulations, surfaces larger than a single unit cell were used. For Si(100), the slab was composed of 4 × 8 unit cells which had dimensions of approximately 6.2 × 6.2 nm. The Si(111) slab consisted of 2 × 2 unit cells and had cell vectors of length 8.1 nm and at an angle of 54.7° to each other. These slabs were specifically chosen to be large enough to support both semicylindrical and hemispherical micelles so as not to bias the final micelle structure.

6.2.2.2 *Interatomic potentials*

The CTA^+ surfactants and the chloride ions were described using the GROMOS96 G45a3 [153, 154] force field which is a united carbon atom model. These potentials have been employed in several recent studies of alkytrimethylammonium halide micellisation [69, 70, 155, 156, 157, 67]. The water was described by the rigid SPC/E model [158], and the silicon surfaces were frozen so no Si-Si interactions were required. Freezing the silicon surfaces is likely to be a safe ap-

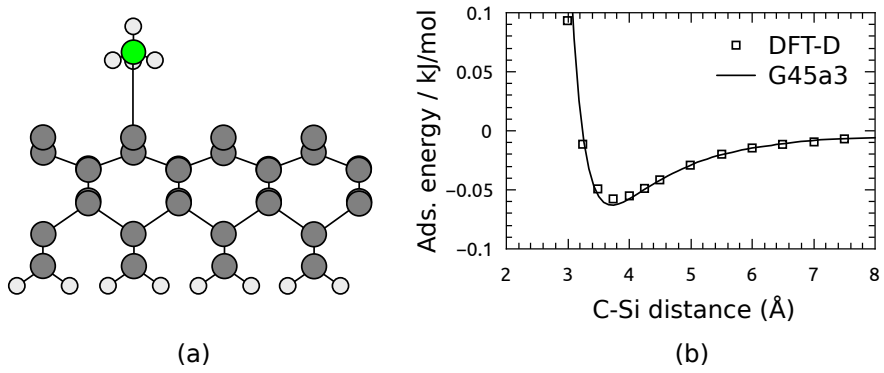


Figure 6.10: (a) A schematic representation of the atomic model used to validate the silicon potentials. A methane molecule is positioned above a silicon atom on the top layer of Si(100) and the energy-distance curve is computed. (b) The resulting adsorption curve for the DFT-D (squares) and the classical G45a3 (line) calculations.

proximation because the reconstruction of the Si(111) and Si(100) surfaces involve large reductions in surface energy of 51 and 70 meV/Å^2, respectively [159], and so they are unlikely to further reconstruct during the experiments.

The interactions between the surface Si atoms and all other atoms were modelled using the Lennard-Jones potential parameters also provided by G45a3. However, it was not clear from the outset whether these potentials would be suitable for modelling silicon surfaces. The next section describes our attempt to validate them.

6.2.2.3 *Force field validation*

In order to confirm that the G45a3 potentials were suitable for the current study we compared the adsorption energy curves for a CH_4 molecule on a Si surface calculated using the classical potentials with those computed using DFT. The CH_4 molecule is assumed to be representative of the -CH_2- methylene bridges within the surfactant alkane chains. This is because the two additional hydrogen atoms saturate the bonds that would otherwise be formed with the carbon atoms, resulting in a molecule with the same tetrahedral configuration.

The DFT calculations were performed using a five layer slab of Si(100) composed of 2x1 lateral unit cells, with dimensions of 15.47×15.47 Å, constructed using the procedure in Ref [160]. Each silicon atom in the bottom layer was passivated with a pair of hydrogen atoms. The slab was periodic in three dimensions with a vacuum slab of 20 Å added to attenuate self-interaction between the images

in the z-dimension. A methane (CH_4) molecule was then positioned directly atop an uppermost silicon atom and the carbon-silicon distance varied, as illustrated in Figure 6.10(a). The difference in energy with respect to an isolated surface and methane molecule was then recorded. The calculations were performed using DFT as implemented in Quantum Espresso 4.1 [161]. The generalised gradient approximation based Perdew, Burke, and Ernzerhof (PBE) exchange-correlation functional was employed. The core electrons were modelled with Troullier-Martins [162] norm-conserving pseudopotentials with the configurations [$1s^2$] $2s^2$ $2p^2$ and [Ne] $3s^2$ $3p^2$ for carbon and silicon, respectively. Only the Γ-point was sampled and a kinetic energy cutoff of 30 Ry was used. Due to the locality of the exchange-correlation functional, intermolecular dispersion effects are not directly reproduced. To account for this we used the DFT-D dispersion correction of Grimme et al. [163] which consists of a parametrised pairwise r^{-6} energy contribution. The computed adsorption energies are plotted in Figure 6.10(b).

The adsorption energy was computed in an analogous way atomistically for the same silicon surface (minus the hydrogen atoms) and for the G45a3 model of the CH_2 molecule which is a constituent of the alkane tails. The resulting energy curve is also plotted in Figure 6.10(b). There is very good agreement between the two sets of data. Both give the same energy minimum and equilibrium spacing to within 0.01 kJ/mol and 0.2 Å, respectively. We conclude that the G45a3 silicon parameters are suitable for modelling silicon surfaces.

6.2.2.4 *Simulation details*

This work employed the same simulation setup as the previous section, see Section 6.1.2.1, except only the canonical ensemble was sampled and a larger integration time-step of 2 fs was used.

6.2.3 *Results*

In the initial configurations, the hydrophobic tails were exposed to the water above. Within the first 200 ps many of the surfactants had inverted and aggregated in an attempt to shield their tails from the water. In the case of Si(100), this resulted in an elongated micelle that was partially attached to the surface and which increased its contact over the subsequent 300 ps, as shown in Figure 6.11. This process of aggregation, followed by a gradually increasing surface contact, also

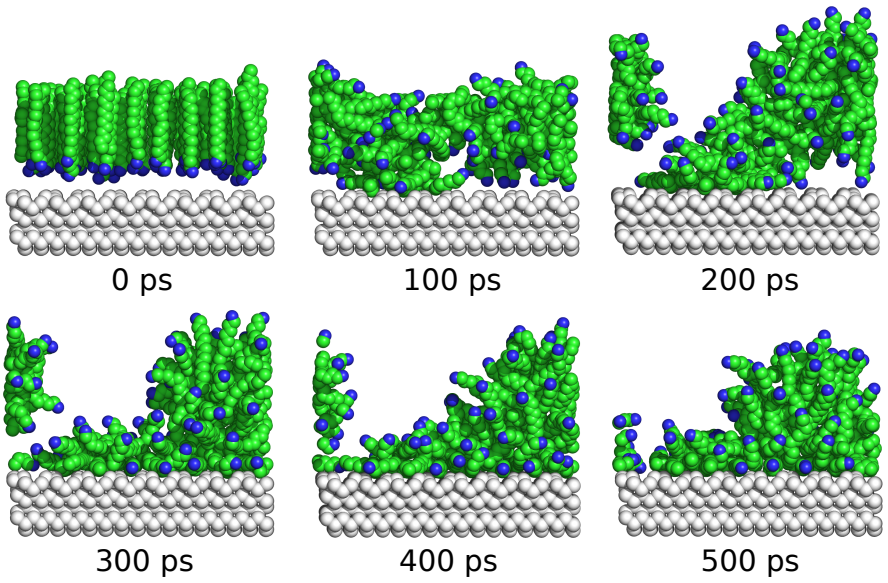


Figure 6.11: Time-evolution of the surfactants at the Si(100)/water interface. A rudimentary semicylindrical micelle is formed. The silicon surface is shown in grey, the alkane tail in green, and the nitrogen atom from the headgroup in blue. All other atoms are hidden for clarity.

unfolded at the Si(111) interface except that, due to the larger unit cell and therefore the larger number of surfactants, three separate micelles were created (composed of 13, 35, and 58 surfactants).

The aggregation of the surfactants and their subsequent adsorption to the surfaces would have been driven partly, if not mostly, by entropy. This is because both processes are accompanied by the release of confined water molecules. Such a discharge increases the degrees of freedom of the system and therefore the entropy. In fact, in bulk CTAC micellisation, over 65% of the free energy change is due to an increase in solvation entropy [164].

A crucial difference between the micellisation at the Si(100) and Si(111) interfaces was the organisation of the surfactants as they came into contact with the surfaces. On Si(100) the surfactants successively aligned in parallel, driven partly by their interaction with the preceding surfactants, resulting in a linear array that naturally gave rise to a semicylindrical structure. This final structure, shown in Figure 6.12, is slightly incomplete suggesting that our (tentative) choice of surfactant concentration was indeed too low.

By contrast, the surfactants on Si(111) exhibited no such linear organisation and instead dispersed radially outward as they increased their coverage of the surface. This resulted in a pair of hemispher-

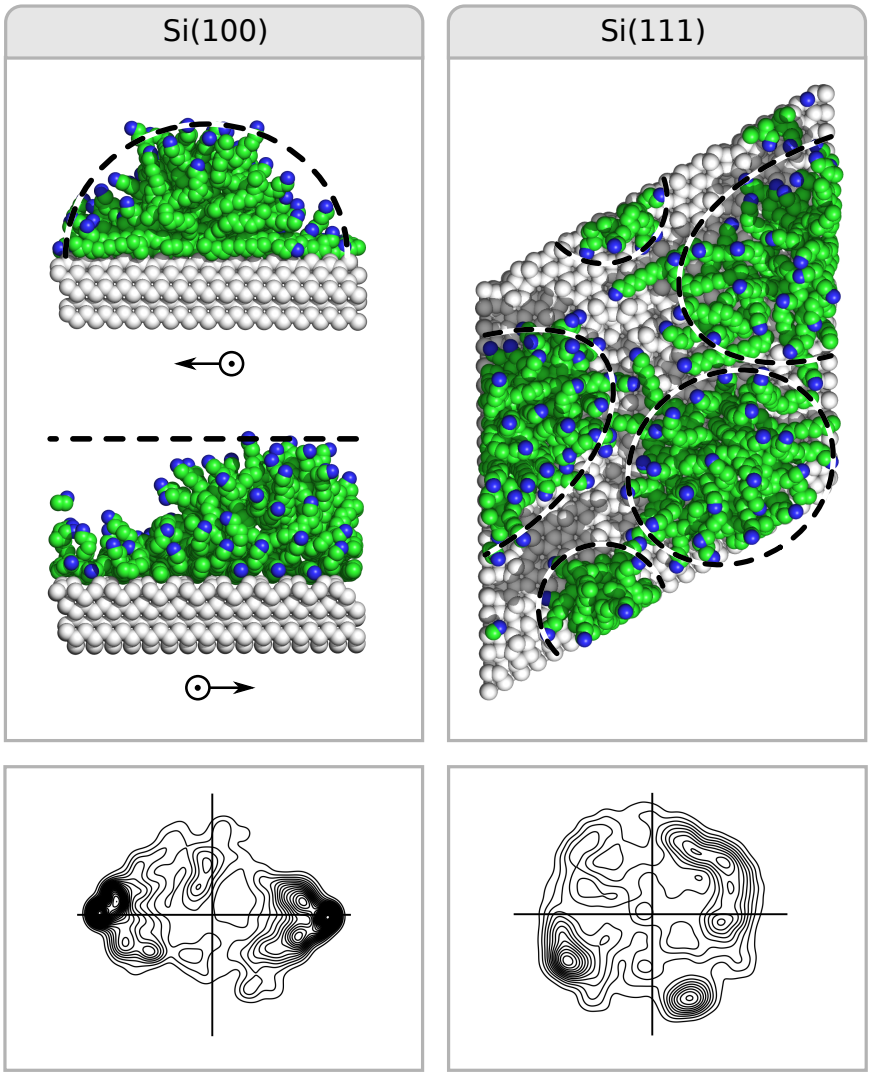


Figure 6.12: Side-views of the semicylindrical micelle formed on Si(100), and an overhead view of the hemispherical micelles (highlighted for clarity) formed on Si(111). These are the final configurations after 5 ns of simulation. See Figure 6.11 for the colour scheme. Density in the xy-plane of the vector between the headgroup and tail of each surfactant on the two substrates, Si(100) and Si(111), also shown. The respective densities are characteristic of semicylindrical and hemispherical micelles.

ical micelles (the smallest of the three micelles remained detached from the surface) as shown in Figure 6.12. The densities in the xy-plane of the vector between the headgroup and tail of each surfactant on the two substrates, Si(100) and Si(111), are also shown in Figure 6.12. The respective densities are characteristic of semicylindrical and hemispherical micelles.

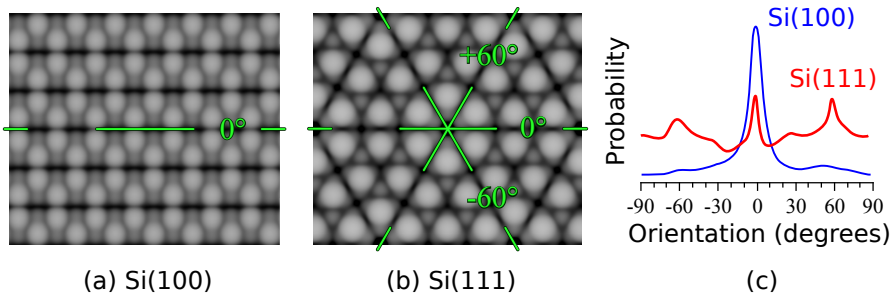


Figure 6.13: The topography of (a) Si(100) and (b) Si(111) from the perspective of a -CH_2- unit in the surfactant tail. The long surface crevices are highlighted in green. (c) The angle distribution of the surfactants that fall within 1 nm of each surface.

6.2.3.1 *Role of surface topography*

To explain the different micelle structures that form on Si(100) and Si(111) we have analysed the topographies of the two surfaces. Figures 6.13(a) and (b) show the variation in adhesion energy between each silicon surface and a single methylene bridge (-CH_2-) from the alkane tails, where dark regions correspond to strong adhesion. This was computed by sweeping the -CH_2- probe across a two-dimensional grid and then, for each (x, y) coordinate, minimising the energy in the z direction. We have also computed the orientation of the surfactants near to the surfaces. More specifically, for each unit $\mathbf{x}_i$ of the hydrocarbons within 1 nm of the surface, we projected $\mathbf{x}_{i+1} - \mathbf{x}_{i-1}$ onto the xy-plane and computed the angle with respect to the $+x$ axis. The time-averaged distribution of these angles is shown in Figure 6.13(c) for both surfaces.

Si(100) displays a distinct corrugation with rows of crevices. The surfactants evidently align themselves with these features and thus adopt a commensurate structure. A similar attempt is made by the surfactants on Si(111) to match the long crevices, as evidenced by the peaks at -60°, 0°, and +60° in Figure 6.13(c). By attempting to match the topography and symmetry of the surfaces, the hemispherical and semicylindrical structures naturally ensue.

When a semicylindrical micelle forms at the *graphite*-aqueous interface, it aligns with one of the three equivalent symmetry axes of the crystal plane. This alignment is believed to result from the fortuitous match between the distance of the centres of the hexagons (2.46 Å) and the distance between alternate units of the hydrocarbon chains (2.51 Å) giving rise to efficient chain packing and strong adhesion [165]. We have discovered a similar fortuity regarding hy-

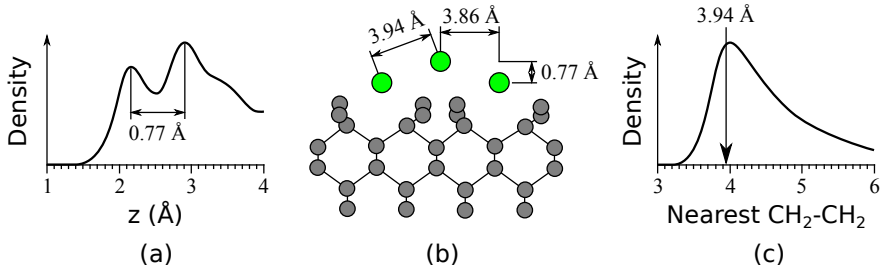


Figure 6.14: The structure of the surfactants adsorbed to Si(100). (a) Two distinct layers visible in the z-density of the surfactants. (b) The geometry of the surfactants (green, aligned normal to the page) with respect to the Si(100) surface. (c) A histogram showing the intermolecular distance between each CH_2 and its nearest counterpart. The arrow indicates the average spacing between neighbouring surfactants adsorbed on Si(100), as shown in (b).

drocarbons on Si(100), albeit inter- rather than intra-molecular in nature. Figure 6.14(a) shows the z-density of the hydrocarbons at the Si(100)-water interface. The peaks reveal two distinct layers on the surface, vertically separated by 0.77 Å. By measuring the features of the Si(100) surface, we find that the hydrocarbons must be horizontally separated by an average of 3.86 Å. This arrangement is illustrated in Figure 6.14(b) and it follows that the average distance between the hydrocarbons in contact with the surface is 3.94 Å. This is less than 0.1 Å short of the mode separation of all of the hydrocarbons in the micelle, as shown in Figure 6.14(c). The hydrocarbons are therefore able to achieve registry with the substrate without straining.

The applicability of our findings to an experimental setting will, of course, depend on how the silicon surfaces are prepared. When silanol groups form on the surfaces they prevent direct interaction between the silicon and the surfactants. This can reduce the anisotropy and thus eliminate any influence from the surface topography on the micelle structure [166].

6.2.4 *Discussion*

The above results reveal that CTAC surfactants form hemispherical micelles on Si(111) but semicylindrical micelles on Si(100). To the best of our knowledge, this is the first reported instance of micelle morphology being determined by the topography of a non-polar substrate, and it leads straightforwardly to a plausible account for the

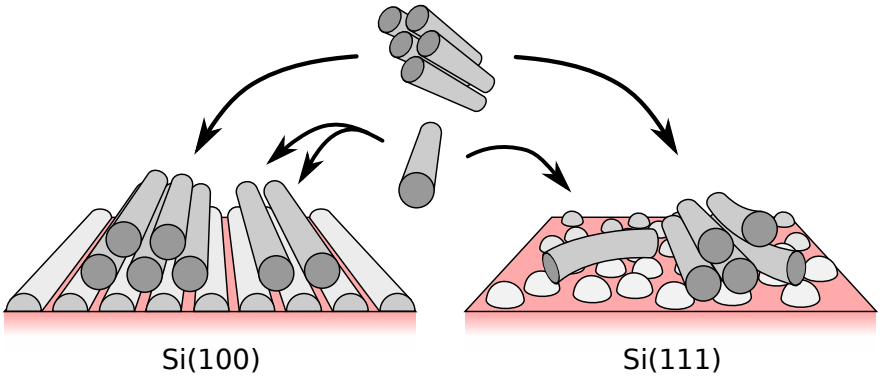


Figure 6.15: On Si(100), we propose the micelles align with the semicylindrical micelles that pattern the surface. This provides strong binding and thus begets heterogeneous nucleation. Sub-critical homogeneous nuclei may also adsorb to the substrate, becoming stabilised before dissolving. On Si(111), however, there is no unidirectional alignment and no congruence with the substrate, resulting in no heterogeneous nucleation.

growth of mesoporous silicates on Si(100) substrates, which we now expound.

The concentration of CTAC in the experimental solution was 2.0 mM, slightly above the (pure) CMC of 1.3 mM [164]. Computer simulations [167] and experiment [168] have revealed that spherical CTAC micelles transition to cylinders in the presence of siliceous species and anionic salts, such as sodium salicylate, and so the bulk solution will likely contain cylindrical micelles. These micelles, which will be coated in sodium ions and siliceous species, may then meet the surfactant-coated Si(100) substrate. By aligning with the rows of hemicylinders, the adsorbing micelle will achieve a stable configuration. It may also be the case that cylindrical micelles aggregate in bulk solution but, unable to reach the critical nucleus size, will dissolve unless they adsorb to the substrate. The surface therefore begets heterogeneous nucleation, and all nuclei will necessarily be aligned with the same axis. These two mechanisms, heterogeneous nucleation and sub-critical homogeneous nucleation with oriented attachment, are shown in Figure 6.15. Note that the sodium and siliceous species, which would serve as an adhesive between the micelles, are not depicted in the figure.

So far it is easy to see how cylindrical mesophases may form on Si(100). To explain the experimentally-observed cubic phases, however, we suggest that the cylindrical mesophase undergoes a transition to the cubic phase. Such a phase transition has been observed

[169] to occur and is initiated by branching between the cylinders, leading to a perforated lamellar phase and then to the cubic phase. We note that this transition has only been observed to the Ia$\bar{3}$d space group and, moreover, that CTAC liquid crystals in pure solution only form Ia$\bar{3}$d, not the observed Pm$\bar{3}$n space group. While it is possible that the Ia$\bar{3}$d crystal undergoes a further transition to Pm$\bar{3}$d, it seems more likely that the hexagonal mesophase transitions directly to Pm$\bar{3}$n, and that this space group has become stabilised by the sodium silicate which has demonstrably had a dramatic effect on the phase equilibrium.

This account agrees with the prevailing concesus that PMOs form via cooperative templating, in which the silicate precursor alters the phase stability and therefore plays a role in constructing its own template, as opposed to precipitating within a pre-formed liquid crystal (although examples of this latter mechanism are known).

In contrast to the above narrative, cylindrical micelles that adsorb on the hemispherical-decorated Si(111) substrate will not be able to achieve congruence with the structure and will therefore not bind strongly. There is also no unique direction to the pattern and therefore no driving force for long-range ordering.

6.3 CHAPTER SUMMARY

The flexibility of soft templates, the multitude of structures that they can adopt at various scales, and their dynamical co-assembly with other components of a system, pose a theoretical challenge in predicting and rationalising their effect on inorganic crystallisation. With advances in computing power and methodologies, however, these problems will ultimately yield to simulation.

In Section 6.1, monolayers of the anionic SDS surfactant have been adsorbed on the (100), (110), (011), and (001) surfaces for both the rutile and anatase polymorphs, and for densities ranging from about 2.2 to 4.4 nm^{-2}. The adhesion energies have been computed in each case, revealing that these monolayers bind most strongly to anatase (100), with an adhesion energy of up to -1.56 Jm^{-2}, while rutile (110) is in second place with an adhesion energy of up to -1.27 Jm^{-2}. These findings were rationalised based on the lateral charge polarity exhibited by these particular surfaces. The monolayer configurations formed on the A(100) and R(110) surfaces would suggest that surface-monolayer electrostatic complementarity primarily directs the assembly.

In Section 6.2, CTAC surfactants at the silicon-water interface self-assembled to produce two different morphologies depending on the silicon orientation: hemispherical micelles on Si(111) and semicylindrical micelles on Si(100). Comparing the structure and orientations of the adsorbed surfactants to the surface topographies revealed that the surfactants adopted their respective morphologies in an attempt to match the symmetry of the substrates. We also observed that, in the case of Si(100), the surfactants were fortuitously able to achieve perfect registry with the surface features without straining, suggesting a strong driving force for organisation.

These results demonstrate that epitaxy can play a crucial role in determining micelle structure even in the case of non-polar surfaces with which the surfactants interact weakly via van der Waals forces. Moreover, they allowed us to explain experimental observations that cubic mesophases form on the Si(100) surface but not Si(111).

7

CONCLUSIONS

7.1 THESIS SUMMARY

Chapter 1 outlined the major ambition of materials science: to achieve nanoscale control of the size, shape, phase, morphology, and assembly of materials. Nature sets a precedent for what can be achieved, along with clues as to the requisite machinary. However, the mechanisms require a molecular-scale resolution that cannot be achieved experimentally. But this is where computer simulations prevail. This thesis has showcased some state-of-the-art computational methodologies and applied them to investigate the molecular processes responsible for controlling inorganic crystal growth, revealing both the utility and present limitations of the computational sciences.

In Chapter 3, microsecond-long simulations were performed to reveal the behaviour of Ca^{2+}, CO_3^{2-}, and $CaCO_3^0$ adions in the vicinity of the two calcite steps. The calcium ions were found to display limited metastability in the vicinity of the steps, instead favouring bulk solution, while the carbonate ions were stable when adsorbed to the steps, but exhibited a preference for the acute step over the obtuse. The $CaCO_3^0$ ion pair was found to behave much like a lone carbonate ion, producing a very similar free energy map. However, the presence of the calcium ion increased the stability of the pair at the obtuse step, resulting in the ion pair spending almost equal amounts of time at the two steps. This would suggest that kink nucleation occurs at a similar rate at the two step geometries.

Chapter 4 focused on kinks, the other active feature of calcite surfaces, employing a variety of rare event methodologies to compute the dissolution rate at the two step geometries. The primary finding of this chapter was that calcite kink dissolution is a long, complex process involving millions of sub-transitions. As a result, the errors in the individual transitions are magnified to such an extent that the predicted transition rate for the entire dissolution process spans a few orders of magnitude. Such precision can be matched with much simpler analysis, revealing that even the most advanced rare event methodologies have little utility in studying certain rare events of great importance in mineralisation. This will have far reaching conse-

quences in how molecular simulation is applied in the future. Nevertheless, this research also revealed some interesting features of calcite kinks: that the adsorption of carbonate ions appears to be diffusion-limited, and that water structure is responsible for approximately half of the driving force that inhibits dissolution.

Chapter 5 combined the results of Chapters 3 and 4, along with experimental data, to construct and optimise a coarse-grained model of calcite precipitation. This model is integrated through time using kinetic Monte Carlo and supports the introduction of impurities. It displayed great potential in capturing both the kinetic and morphological attributes of calcite. This was illustrated both by the behaviour of the crystal under stoichiometric and non-stoichiometric conditions, but also in the presence of impurities. Specifically, the kink blocking impurity mechanism was systematically studied, with the individual molecular mechanisms linked directly to their morphological impact. Interestingly, our KMC model captured kinetic nuances that analytical kinetic models fail to. Aspartic acid was confirmed to exhibit behaviour characteristic of a kink blocker with regards to both its kinetic and morphological effects, and we were able to predict that aspartic acid dissolves at a rate on the order of microseconds. Finally, we considered step pinning and, in particular, the dependence of the velocity of a step segment on its length. The simulation was found to perfectly match the theoretical Gibbs-Thomson predictions, but this is in stark contrast to the experimental findings which we regard as a major obstacle that must be resolved.

Moving away from calcite, Chapter 6 investigated two forms of soft templating. The first was concerned with molecular-scale templating whereby monolayers of SDS surfactants were allowed to self-assemble on various titania faces. The aim of this section was to characterise the binding to help rationalise polymorph and orientation selection. The second study computed the preferred micelle morphologies of CTAC surfactants when adsorbed on two different orientations of silicon. We predicted that CTAC should form hemispherical micelles on Si(111) without long-range order, but linear rows of semicylindrical micelles on Si(100), with long-range order. These predictions have since been confirmed experimentally and are responsible for the selective growth of oriented mesoporous organosilicas on the Si(100) substrate.

7.2 OUTLOOK

A major legacy of this thesis will likely be the conflict between Chapters 4 and 5. On the one hand, our coarse-grained model of calcite growth (Chapter 5) provides a major step towards predictive simulation of mineralisation processes; showcasing both the capacity and limitations of such a model to capture the kinetic and morphological characters of growth. On the other hand, this progress is undermined by the conclusion of Chapter 4, that certain important mineralisation processes are currently beyond the reach of rare event methodologies, and will remain so for the foreseeable future. This raises the question of how the requisite rates of a KMC model will be obtained if not through rare event sampling.

To give a specific example of where this presents a problem, consider the growth of calcite in the presence of magnesium ions, which is of huge importance in geology and biomineralisation. Small quantities of the magnesium ions are incorporated into bulk calcite which increases the solubility of the resulting crystal. However, the precise mechanisms by which magnesium affects the solubility are unknown. For instance, is the solubility affected by the lattice strain that results from magnesium ions being much smaller than calcium ions? Or is solubility primarily affected by the impact of the magnesium ions in the immediate vicinity of an active site, such as a kink? If the latter, then which active sites are most susceptible? Just kink propagation or also kink nucleation? All kink sites, or only specific kink sites? Is it the attachment or the detachment rates, or both, that are affected? Clearly there are a huge number of unknown variables, each of varying significance. In an ideal world, we would analyse the calcite kink dissolution rates for various configurations of incorporated magnesium, but we now know that this not tenable. Instead, progress may require a more oblique approach. For example, the following information is obtainable from molecular simulation:

- The local stress and enthalpy costs associated with magnesium incorporation.
- The attachment rates of calcium, carbonate, and magnesium ions to the active sites for various magnesium-bearing configurations.

Information gleaned from such data may provide clues as to the relevant mechanisms, with the prospect of greatly reducing the number

of unknown parameters. After which, unique kinetic or morphological effects may be mapped to the parameter space, potentially allowing the parameters to be uniquely optimised. This would likely require a close collaboration between computational and experimental researchers, where the computational scientists would identify the required experimental data points.

Finally, we reiterate the problem presented by the experimental observations that calcite step velocities violate the Gibbs-Thomson relation, whereas our simulations comply with it (see Section 5.4). The consequence is that our model is unable to correctly describe step pinning, a very important inhibitory mechanism. While the significance of this Gibbs-Thomson violation has been discussed before, no notable progress has been made in resolving it. We therefore emphasise the need for a directed effort to resolve it. A first step in this direction may be for a computational/theoretical group to explicitly link every plausible molecular process with the step velocity profile as a function of step length.

Once this problem has been resolved and accounted for in a KMC simulation, future work of importance would be to incorporate surfactant effects into the model, whereby large molecules, such as peptides, bind preferentially to particular local kink configurations.

A

CALCITE SIMULATION DETAILS

Unless stated otherwise, the reader may assume that all of the calcite simulations used the following setup.

The calcium carbonate was described by the force field of Raiteri et al. [170] which was parametrised by fitting against experimental thermodynamic and structural data. The water was SPC/Fw [171].

The simulations were performed using LAMMPS (7 Dec 2015 release). The integration time-step was 1 fs. The canonical ensemble was sampled using the Nosé-Hoover thermostat to maintain a temperature of 300 K with a damping parameter of 100 fs. Long-range electrostatics were handled with the PPPM method and an accuracy of 10^{-4} (which means the force errors will be less than 10^{-4} times the force magnitude between a pair of monovalent charges separated by 1 Å).

BIBLIOGRAPHY

[1] Jeremy R Young and Karen Henriksen. Biomineralization within vesicles: the calcite of coccoliths. *Reviews in mineralogy and geochemistry*, 54(1):189–215, 2003.

[2] Mariëtte Wolthers, Gernot Nehrke, Jon Petter Gustafsson, and Philippe Van Cappellen. Calcite growth kinetics: Modeling the effect of solution stoichiometry. *Geochimica et Cosmochimica Acta*, 77:121–134, 2012.

[3] Laura C Nielsen, James J De Yoreo, and Donald J DePaolo. General model for calcite growth kinetics in the presence of impurity ions. *Geochimica et Cosmochimica Acta*, 115:100–114, 2013.

[4] Andrew G Stack and Meg C Grantham. Growth rate of calcite steps as a function of aqueous calcium-to-carbonate ratio: independent attachment and detachment of calcium and carbonate ions. *Crystal Growth & Design*, 10(3):1409–1413, 2010.

[5] Mina Hong and H Henry Teng. Implications of solution chemistry effects: Direction-specific restraints on the step kinetics of calcite growth. *Geochimica et Cosmochimica Acta*, 141:228–239, 2014.

[6] 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.

[7] CA Orme, A Noy, A Wierzbicki, MT McBride, M Grantham, HH Teng, PM Dove, and JJ DeYoreo. Formation of chiral morphologies through selective binding of amino acids to calcite surface steps. *Nature*, 411(6839):775–779, 2001.

[8] Yongsoon Shin, Jinhui Tao, Bruce W Arey, Chongmin Wang, Gregory J Exarhos, James J De Yoreo, Maria L Sushko, and Jun Liu. Double Epitaxy as a Paradigm for Templated Growth of Highly Ordered Three-Dimensional Mesophase Crystals. *ACS nano*, 2016.

[9] Stephen Mann. *Biomineralization*. Oxford.: Oxford University Press, 2001.

[10] H Henry Teng, Patricia M Dove, and James J De Yoreo. Kinetics of calcite growth: surface processes and relationships to macroscopic rate laws. *Geochimica et Cosmochimica Acta*, 64(13): 2255–2266, 2000.

[11] AJ Gratz, PE Hillner, and PK Hansma. Step dynamics and spiral growth on calcite. *Geochimica et Cosmochimica Acta*, 57(2): 491–495, 1993.

[12] James J De Yoreo and Patricia M Dove. Shaping crystals with biomolecules. *Science*, 306(5700):1301–1302, 2004.

[13] W-K_ Burton, N Cabrera, and FC Frank. The growth of crystals and the equilibrium structure of their surfaces. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 243(866):299–358, 1951.

[14] Gernot Nehrke, Gert-Jan Reichart, P Van Cappellen, C Meile, and Jelle Bijma. Dependence of calcite growth rate and Sr partitioning on solution stoichiometry: non-Kossel crystal growth. *Geochimica et Cosmochimica Acta*, 71(9):2240–2249, 2007.

[15] Stephen Mann. *Biomimetic materials chemistry*. John Wiley & Sons, 1995.

[16] Colin L Freeman, John H Harding, David Quigley, and P Mark Rodger. Structural control of crystal nuclei by an eggshell protein. *Angewandte Chemie International Edition*, 49(30):5135–5137, 2010.

[17] Joanna Aizenberg, Andrew J Black, and George M Whitesides. Control of crystal nucleation by patterned self-assembled monolayers. *Nature*, 398(6727):495–498, 1999.

[18] Stephen Mann. Biomineralization: the form (id) able part of bioinorganic chemistry! *Journal of the Chemical Society, Dalton Transactions*, (21):3953–3962, 1997.

[19] Frank Hoffmann, Maximilian Cornelius, Jürgen Morell, and Michael Fröba. Silica-Based Mesoporous Organic–Inorganic Hybrid Materials. *Angewandte Chemie International Edition*, 45 (20):3216–3251, 2006.

[20] Srinivas Manne and Hermann E Gaub. Molecular organization of surfactants at solid-liquid interfaces. *Science*, 270(5241):1480–1482, 1995.

[21] Nick Bostrom. *Superintelligence: Paths, dangers, strategies*. OUP Oxford, 2014.

[22] David M Deaven and Kai-Ming Ho. Molecular geometry optimization with a genetic algorithm. *Physical review letters*, 75(2): 288, 1995.

[23] David L Masica, Sarah B Schrier, Elizabeth A Specht, and Jeffrey J Gray. De Novo Design of Peptide- Calcite Biomineralization Systems. *Journal of the American Chemical Society*, 132(35): 12252–12262, 2010.

[24] Sarah B Schrier, Marianna K Sayeg, and Jeffrey J Gray. Prediction of calcite morphology from computational and experimental studies of mutations of a de novo-designed peptide. *Langmuir*, 27(18):11520–11527, 2011.

[25] Colin L Freeman, Q Hu, Michael H Nielsen, J Tao, JJ De Yoreo, and John H Harding. Surface selectivity of calcite on self-assembled monolayers. *The Journal of Physical Chemistry C*, 117 (10):5154–5163, 2013.

[26] Alexander S Côté, Robert Darkins, and Dorothy M Duffy. Modeling calcite crystallization on self-assembled carboxylate-terminated alkanethiols. *The Journal of Physical Chemistry C*, 118 (33):19188–19193, 2014.

[27] Robert Darkins, Alexander S Côté, Colin L Freeman, and Dorothy M Duffy. Crystallisation rates of calcite from an amorphous precursor in confinement. *Journal of Crystal Growth*, 367: 110–114, 2013.

[28] Hiroki Nada, Tatsuya Nishimura, Takeshi Sakamoto, and Takashi Kato. Heterogeneous growth of calcite at aragonite {001}-and vaterite {001}-melt interfaces: A molecular dynamics simulation study. *Journal of Crystal Growth*, 450:148–159, 2016.

[29] David Quigley and P Mark Rodger. Free energy and structure of calcium carbonate nanoparticles during early stages of crystallization, 2008.

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

[31] Stefano Piana, Manijeh Reyhani, and Julian D Gale. Simulating micrometre-scale crystal growth from solution. *Nature*, 438 (7064):70–73, 2005.

[32] Stefano Piana and JD Gale. Three-dimensional kinetic Monte Carlo simulation of crystal growth from solution. *Journal of crystal growth*, 294(1):46–52, 2006.

[33] Stefano Piana, Franca Jones, and Julian D Gale. Assisted desolvation as a key kinetic step for crystal growth. *Journal of the American Chemical Society*, 128(41):13568–13574, 2006.

[34] Inna Kurganskaya and Andreas Luttge. Kinetic Monte Carlo Approach To Study Carbonate Dissolution. *The Journal of Physical Chemistry C*, 120(12):6482–6492, 2016.

[35] Jacques Schott, Susan Brantley, David Crerar, Christophe Guy, Maria Borcsik, and Christian Willaime. Dissolution kinetics of strained calcite. *Geochimica et Cosmochimica Acta*, 53(2):373–382, 1989.

[36] Rick E Williford, Donald R Baer, James E Amonette, and Alan S Lea. Dissolution and growth of (1014) calcite in flowing water: estimation of back reaction rates via kinetic Monte Carlo simulations. *Journal of crystal growth*, 262(1):503–518, 2004.

[37] JM McCoy and John P LaFemina. Kinetic Monte Carlo investigation of pit formation at the CaCO 3 (1014) surface-water interface. *Surface science*, 373(2):288–299, 1997.

[38] JJ De Yoreo, LA Zepeda-Ruiz, RW Friddle, SR Qiu, LE Wasylenki, AA Chernov, GH Gilmer, and PM Dove. Rethinking classical crystal growth models through molecular scale insights: consequences of kink-limited kinetics. *Crystal Growth & Design*, 9(12):5135–5144, 2009.

[39] John H Harding, Dorothy M Duffy, Maria L Sushko, P Mark Rodger, David Quigley, and James A Elliott. Computational techniques at the organic- inorganic interface in biomineralization. *Chemical reviews*, 108(11):4823–4854, 2008.

[40] Musa Ozboyaci, Daria B Kokh, Stefano Corni, and Rebecca C Wade. Modeling and simulation of protein–surface interactions: achievements and challenges. *Quarterly reviews of biophysics*, 49: e4, 2016.

[41] Mingjun Yang, John Harding, and Susan Louise Svane Stipp. Simulations of monosaccharide on calcite surfaces. *Mineralogical Magazine*, 72(1):295–299, 2008.

[42] Colin L Freeman, John H Harding, David Quigley, and P Mark 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.

[43] David J Cooke, Richard J Gray, Karina Krarup Sand, Susan Louise Svane Stipp, and JA Elliott. Interaction of ethanol and water with the {1014} surface of calcite. *Langmuir*, 26(18):14520–14529, 2010.

[44] Andrew G Stack, Paolo Raiteri, and Julian D Gale. Accurate rates of the complex mechanisms for growth and dissolution of minerals using a combination of rare-event theories. *Journal of the American Chemical Society*, 134(1):11–14, 2011.

[45] Li-Min Liu, Alessandro Laio, and Angelos Michaelides. Initial stages of salt crystal dissolution determined with ab initio molecular dynamics. *Physical Chemistry Chemical Physics*, 13(29): 13162–13166, 2011.

[46] Marco De La Pierre, Paolo Raiteri, and Julian D Gale. Structure and dynamics of water at step edges on the calcite {1014} surface. *Crystal Growth & Design*, 16(10):5907–5914, 2016.

[47] Jonathan RI Lee, T Yong-Jin Han, Trevor M Willey, Dongbo Wang, Robert W Meulenberg, Joakim Nilsson, Patricia M Dove, Louis J Terminello, Tony van Buuren, and James J De Yoreo. Structural development of mercaptophenol self-assembled monolayers and the overlying mineral phase during templated $CaCO_3$ crystallization from a transient amorphous film. *Journal of the American Chemical Society*, 129(34):10370–10381, 2007.

[48] Dorothy M Duffy, A Markus Travaille, Herman van Kempen, and John H Harding. Effect of bicarbonate ions on the crystal-

lization of calcite on self-assembled monolayers. *The Journal of Physical Chemistry B*, 109(12):5713–5718, 2005.

[49] Beena Rai et al. Modeling self-assembly of surfactants at interfaces. *Current Opinion in Chemical Engineering*, 15:84–94, 2017.

[50] John MA Grime, James F Dama, Barbie K Ganser-Pornillos, Cora L Woodward, Grant J Jensen, Mark Yeager, and Gregory A Voth. Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nature communications*, 7, 2016.

[51] Michael L Klein and Wataru Shinoda. Large-scale molecular dynamics simulations of self-assembling systems. *Science*, 321 (5890):798–800, 2008.

[52] Li Wang, Runqing Liu, Yuehua Hu, Jiapeng Liu, and Wei Sun. Adsorption behavior of mixed cationic/anionic surfactants and their depression mechanism on the flotation of quartz. *Powder Technology*, 302:15–20, 2016.

[53] Joshua L Bradley-Shaw, Philip J Camp, Peter J Dowding, and Ken Lewtas. Molecular Dynamics Simulations of Glycerol Monooleate Confined between Mica Surfaces. *Langmuir*, 32(31): 7707–7718, 2016.

[54] Kunal Shah, Patrick Chiu, Mayank Jain, José Fortes, Brij Moudgil, and Susan Sinnott. Morphology and Mechanical Properties of Surfactant Aggregates at Water- Silica Interfaces: Molecular Dynamics Simulations. *Langmuir*, 21(12):5337–5342, 2005.

[55] Kunal Shah, Patrick Chiu, and Susan B Sinnott. Comparison of morphology and mechanical properties of surfactant aggregates at water–silica and water–graphite interfaces from molecular dynamics simulations. *Journal of colloid and interface science*, 296(1):342–349, 2006.

[56] Maria Sammalkorpi, Athanassios Z Panagiotopoulos, and Mikko Haataja. Structure and dynamics of surfactant and hydrocarbon aggregates on graphite: A molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 112(10):2915–2921, 2008.

[57] Hector Domínguez. Self-aggregation of the SDS surfactant at a solid- liquid interface. *The Journal of Physical Chemistry B*, 111 (16):4054–4059, 2007.

[58] Goundla Srinivas, Steven O Nielsen, Preston B Moore, and Michael L Klein. Molecular Dynamics Simulations of Surfactant Self-Organization at a Solid- Liquid Interface. *Journal of the American Chemical Society*, 128(3):848–853, 2006.

[59] Hector Domínguez. Structure of the sodium dodecyl sulfate surfactant on a solid surface in different NaCl solutions. *Langmuir*, 25(16):9006–9011, 2009.

[60] Zhijun Xu, Xiaoning Yang, and Zhen Yang. A molecular simulation probing of structure and interaction for supramolecular sodium dodecyl sulfate/single-wall carbon nanotube assemblies. *Nano letters*, 10(3):985–991, 2010.

[61] Shangchao Lin, Chih-Jen Shih, Michael S Strano, and Daniel Blankschtein. Molecular insights into the surface morphology, layering structure, and aggregation kinetics of surfactant-stabilized graphene dispersions. *Journal of the American Chemical Society*, 133(32):12810–12823, 2011.

[62] Pradip Kr Ghorai and Sharon C Glotzer. Molecular dynamics simulation study of self-assembled monolayers of alkanethiol surfactants on spherical gold nanoparticles. *The Journal of Physical Chemistry C*, 111(43):15857–15862, 2007.

[63] D.M. Duffy and J.H. Harding. Simulation of organic monolayers as templates for the nucleation of calcite crystals. *Langmuir*, 20(18):7630–7636, 2004.

[64] E. Núñez-Rojas and H. Domínguez. Computational studies on the behavior of Sodium Dodecyl Sulfate (SDS) at TiO2(rutile)/water interfaces. *Journal of colloid and interface science*, 364(2):417–427, 2011.

[65] Flor R Siperstein and Keith E Gubbins. Phase Separation and Liquid Crystal Self-Assembly in Surfactant- Inorganic- Solvent Systems. *Langmuir*, 19(6):2049–2057, 2003.

[66] Lin Jin, Scott M Auerbach, and Peter A Monson. Simulating the formation of surfactant-templated mesoporous silica materials: a model with both surfactant self-assembly and silica polymerization. *Langmuir*, 29(2):766–780, 2012.

[67] Miguel Jorge. Structure of cationic surfactant micelles from molecular simulations of self-assembly. *Journal of Molecular Structure: THEOCHEM*, 946(1):88–93, 2010.

[68] Germán Pérez-Sánchez, Szu-Chia Chien, José RB Gomes, M Natália DS Cordeiro, Scott M Auerbach, Peter A Monson, and Miguel Jorge. Multiscale Model for the Templated Synthesis of Mesoporous Silica: The Essential Role of Silica Oligomers. *Chemistry of Materials*, 28(8):2715–2727, 2016.

[69] Miguel Jorge. Molecular dynamics simulation of self-assembly of n-Decyltrimethylammonium Bromide micelles. *Langmuir*, 24 (11):5714–5725, 2008.

[70] Germán Pérez-Sánchez, José RB Gomes, and Miguel Jorge. Modeling Self-Assembly of Silica/Surfactant Mesostructures in the Templated Synthesis of Nanoporous Solids. *Langmuir*, 29 (7):2387–2396, 2013.

[71] Max Born and R Oppenheimer. On the quantum theory of molecules. *Ann. Phys*, 84(4):457–484, 1927.

[72] John M Ziman. *Principles of the Theory of Solids*. Cambridge university press, 1972.

[73] Richard M Martin. *Electronic structure: basic theory and practical methods*. Cambridge university press, 2004.

[74] Wolfram Koch and Max C Holthausen. *A chemist's guide to density functional theory*. John Wiley & Sons, 2015.

[75] David Sholl and Janice A Steckel. *Density functional theory: a practical introduction*. John Wiley & Sons, 2011.

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

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

[78] Mike C Payne, Michael P Teter, Douglas C Allan, TA Arias, and JD Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Reviews of Modern Physics*, 64(4):1045, 1992.

[79] Dominik Marx and Jürg Hutter. *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press, 2009.

[80] Paul Peter Ewald. Ewald summation. *Ann. Phys*, 369:253, 1921.

[81] Roger W Hockney and James W Eastwood. *Computer simulation using particles*. CRC Press, 1988.

[82] Brian Peter Cowan. *Topics in statistical mechanics*, volume 54. World Scientific, 2005.

[83] Roger Bowley and Mariana Sanchez. *Introductory statistical mechanics*. Clarendon Press Oxford, 1999.

[84] E Darve. Free Energy Calculations: Theory and Applications in Chemistry and Biology. *Springer series in Chemical Physics*, 86: 119–170, 2007.

[85] Lev Davidovich Landau and EM Lifshitz. Statistical physics, part I, 1980.

[86] Daan Frenkel and Berend Smit. Understanding molecular simulations: from algorithms to applications. *Academic, San Diego*, 1996.

[87] Denis J Evans and Brad Lee Holian. The nose–hoover thermostat. *The Journal of chemical physics*, 83(8):4069–4074, 1985.

[88] Simone Melchionna, Giovanni Ciccotti, and Brad Lee Holian. Hoover NPT dynamics for systems varying in shape and size. *Molecular Physics*, 78(3):533–544, 1993.

[89] Luca Maragliano, Alexander Fischer, Eric Vanden-Eijnden, and Giovanni Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *The Journal of chemical physics*, 125(2):024106, 2006.

[90] Eric Vanden-Eijnden. Transition path theory. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, pages 453–493. Springer, 2006.

[91] Antonio C Lasaga. Transition state theory. *Rev. Mineral.;(United States)*, 8, 1981.

[92] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *The Journal of Chemical Physics*, 126(16):164103, 2007.

[93] Graeme Henkelman, Blas P Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics*, 113(22):9901–9904, 2000.

[94] Baron Peters, Andreas Heyden, Alexis T Bell, and Arup Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of chemical physics*, 120(17):7877–7886, 2004.

[95] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry*, 13(8):1011–1021, 1992.

[96] Alan Grossfield. WHAM: the weighted histogram analysis method. http://membrane.urmc.rochester.edu/content/wham.

[97] Alessandro Laio and Francesco L 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.

[98] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters*, 100(2):020603, 2008.

[99] Paolo Raiteri, Alessandro Laio, Francesco Luigi Gervasio, Cristian Micheletti, and Michele Parrinello. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *The Journal of Physical Chemistry B*, 110(8):3533–3539, 2006.

[100] Luca Maragliano and Eric Vanden-Eijnden. Single-sweep methods for free energy calculations. *The Journal of chemical physics*, 128(18):184110, 2008.

[101] Arthur F Voter. Introduction to the kinetic Monte Carlo method. In *Radiation Effects in Solids*, pages 1–23. Springer, 2007.

[102] Mads R So, Arthur F Voter, et al. Temperature-accelerated dynamics for simulation of infrequent events. *The Journal of Chemical Physics*, 112(21):9599–9606, 2000.

[103] Luca Maragliano and Eric Vanden-Eijnden. On-the-fly string method for minimum free energy paths calculation. *Chemical Physics Letters*, 446(1):182–190, 2007.

[104] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

[105] William H Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[106] Klaus Müller and Leo D Brown. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica chimica acta*, 53(1):75–93, 1979.

[107] Michael A Bellucci and Bernhardt L Trout. Bézier curve string method for the study of rare events in complex chemical systems. *The Journal of chemical physics*, 141(7):074110, 2014.

[108] Davide Branduardi and José D Faraldo-Gómez. String Method for Calculation of Minimum Free-Energy Paths in Cartesian Space in Freely Tumbling Systems. *Journal of chemical theory and computation*, 9(9):4140–4154, 2013.

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

[110] Davide Branduardi, Francesco Luigi Gervasio, and Michele Parrinello. From A to B in free energy space. *The Journal of chemical physics*, 126(5):054103, 2007.

[111] Ilian T Todorov, William Smith, Kostya Trachenko, and Martin T Dove. DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. *Journal of Materials Chemistry*, 16(20):1911–1918, 2006.

[112] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry*, 24(16):1999–2012, 2003.

[113] Karina Krarup Sand, Dominique Tobler, Sören Dobberschütz, Kirsten Kolbjørn Larsen, Emil Makovicky, Martin Peter Andersson, Mariette Wolthers, and Susan LS Stipp. Calcite growth

kinetics: Dependence on saturation index, Ca2+: CO32-activity ratio and surface atomic structure. *Crystal Growth & Design*.

[114] L Niel Plummer and Eurybiades Busenberg. The solubilities of calcite, aragonite and vaterite in CO 2-H 2 O solutions between 0 and 90 C, and an evaluation of the aqueous model for the system CaCO 3-CO 2-H 2 O. *Geochimica et Cosmochimica Acta*, 46(6):1011–1040, 1982.

[115] James J De Yoreo and Peter G Vekilov. Principles of crystal nucleation and growth. *Reviews in mineralogy and geochemistry*, 54(1):57–93, 2003.

[116] James J De Yoreo, Andrzej Wierzbicki, and Patricia M Dove. New insights into mechanisms of biomolecular control on growth of inorganic crystals. *CrystEngComm*, 9(12):1144–1152, 2007.

[117] Hiroki Nada. Difference in the conformation and dynamics of aspartic acid on the flat regions, step edges, and kinks of a calcite surface: A molecular dynamics study. *The Journal of Physical Chemistry C*, 118(26):14335–14345, 2014.

[118] Hiroki Nada. Importance of water in the control of calcite crystal growth by organic molecules. *Polymer Journal*, 2014.

[119] Deron A Walters, Bettye L Smith, Angela M Belcher, George T Paloczi, Galen D Stucky, Daniel E Morse, and Paul K Hansma. Modification of calcite crystal growth by abalone shell proteins: an atomic force microscope study. *Biophysical Journal*, 72(3):1425, 1997.

[120] H Henry Teng, Patricia M Dove, Christine A Orme, and James J De Yoreo. Thermodynamics of calcite growth: baseline for understanding biomineral formation. *Science*, 282(5389):724–727, 1998.

[121] AA Chernov. Notes on interface growth kinetics 50 years after Burton, Cabrera and Frank. *Journal of Crystal Growth*, 264(4): 499–518, 2004.

[122] AA Chernov, JJ DeYoreo, and LN Rashkovich. Fluctuations and Gibbs-Thomson law-the simple physics. *Journal of Optoelectronics and Advanced Materials*, 9(5):1191–1197, 2007.

[123] G. Phani, G. Tulloch, D. Vittorio, and I. Skryabin. Titania solar cells: new photovoltaic technology. *Renewable energy*, 22(1):303–309, 2001.

[124] M. Gratzel. Molecular photovoltaics that mimic photosynthesis. *Pure and Applied Chemistry*, 73(3):459–468, 2001.

[125] DA Tryk, A. Fujishima, and K. Honda. Recent topics in photoelectrochemistry: achievements and future prospects. *Electrochimica acta*, 45(15):2363–2376, 2000.

[126] A. Fujishima and K. Honda. Photolysis-decomposition of water at the surface of an irradiated semiconductor. *Nature*, 238(5385): 37–38, 1972.

[127] T. Kokubo, H.M. Kim, and M. Kawashita. Novel bioactive materials with different mechanical properties. *Biomaterials*, 24(13): 2161–2175, 2003.

[128] S. Farrokhpay. A review of polymeric dispersant stabilisation of titania pigment. *Advances in colloid and interface science*, 151 (1):24–32, 2009.

[129] G. Wakefield, M. Green, S. Lipscomb, and B. Flutter. Modified titania nanomaterials for sunscreen applications–reducing free radical generation and DNA damage. *Materials science and technology*, 20(8):985–988, 2004.

[130] A. Mitra, A. Bhaumik, and B.K. Paul. Synthesis and characterization of mesoporous titanium dioxide using self-assembly of sodium dodecyl sulfate and benzyl alcohol systems as templates. *Microporous and Mesoporous Materials*, 109(1):66–72, 2008.

[131] D. Wang, Z. Ma, S. Dai, J. Liu, Z. Nie, M.H. Engelhard, Q. Huo, C. Wang, and R. Kou. Low-temperature synthesis of tunable mesoporous crystalline transition metal oxides and applications as Au catalyst supports. *The Journal of Physical Chemistry C*, 112(35):13499–13509, 2008.

[132] D. Wang, D. Choi, Z. Yang, V.V. Viswanathan, Z. Nie, C. Wang, Y. Song, J.G. Zhang, and J. Liu. Synthesis and Li-ion insertion properties of highly crystalline mesoporous rutile TiO2. *Chemistry of Materials*, 20(10):3435–3442, 2008.

[133] C.H. Chen, C.H. Liu, Y.C. Su, and C.M. Yang. Surface-modified anatase nanocrystalline building blocks for constructing catalytically highly active nanoporous titania materials. *Applied Catalysis B: Environmental*, 2012.

[134] DL Liao and BQ Liao. Shape, size and photocatalytic activity control of TiO2 nanoparticles with surfactants. *Journal of Photochemistry and Photobiology A: Chemistry*, 187(2):363–369, 2007.

[135] DL Liao, GS Wu, and BQ Liao. Zeta potential of shape-controlled TiO2 nanoparticles with surfactants. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 348(1):270–275, 2009.

[136] D. Wang, D. Choi, J. Li, Z. Yang, Z. Nie, R. Kou, D. Hu, C. Wang, L.V. Saraf, J. Zhang, et al. Self-assembled TiO2–graphene hybrid nanostructures for enhanced Li-ion insertion. *ACS nano*, 3(4): 907–914, 2009.

[137] H. Domínguez. Structural Transition of the Sodium Dodecyl Sulfate (SDS) Surfactant Induced by Changes in Surfactant Concentrations. *The Journal of Physical Chemistry B*, 115(43):12422–12428, 2011.

[138] M. Matsui and M. Akaogi. Molecular dynamics simulation of the structural and physical properties of the four polymorphs of TiO2. *Molecular Simulation*, 6(4-6):239–244, 1991.

[139] DR Collins and W. Smith. Evaluation of TiO2 Force Fields. Technical report, CCLRC Technical Report DL-TR-96-001, Daresbury Laboratory, UK, 1996.

[140] V. Swamy, JD Gale, and LS Dubrovinsky. Atomistic simulation of the crystal structures and bulk moduli of TiO2 polymorphs. *Journal of Physics and Chemistry of Solids*, 62(5):887–895, 2001.

[141] V. Swamy and J.D. Gale. Transferable variable-charge interatomic potential for atomistic simulation of titanium oxides. *Physical Review B*, 62(9):5406, 2000.

[142] W. Smith and T.R. Forester. *The DL_POLY Classic User Manual*.

[143] K.D. Dobson, A.D. Roddick-Lanzilotta, and A.J. McQuillan. An in situ infrared spectroscopic investigation of adsorption of sodium dodecylsulfate and of cetyltrimethylammonium bromide surfactants to TiO2, ZrO2, Al2O3, and Ta2O5 particle

films from aqueous solutions. *Vibrational spectroscopy*, 24(2):287–295, 2000.

[144] C. Jin, Y. Bai, A. Jagota, and C.Y. Hui. Adhesion selectivity by electrostatic complementarity. II. Two-dimensional analysis. *Journal of Applied Physics*, 110(5):054903–054903, 2011.

[145] Hong Yang, Neil Coombs, Igor Sokolov, and Geoffrey A Ozin. Registered growth of mesoporous silica films on graphite. *J. Mater. Chem.*, 7(7):1285–1290, 1997.

[146] Hong Yang, Alex Kuperman, Neil Coombs, Suzan Mamiche-Afara, and Geoffrey A Ozin. Synthesis of oriented films of mesoporous silica on mica. *Nature*, 379(6567):703–705, 1996.

[147] Hirokatsu Miyata and Kazuyuki Kuroda. Formation of a continuous mesoporous silica film with fully aligned mesochannels on a glass substrate. *Chemistry of Materials*, 12(1):49–54, 2000.

[148] Alon V McCormick, Alexis T Bell, and CJ Radke. Quantitative determination of siliceous species in sodium silicate solutions by ^{29}Si NMR spectroscopy. *Zeolites*, 7(3):183–190, 1987.

[149] Siddharth V Patwardhan, Fateme S Emami, Rajiv J Berry, Sharon E Jones, Rajesh R Naik, Olivier Deschaume, Hendrik Heinz, and Carole C Perry. Chemistry of aqueous silica nanoparticle surfaces and the mechanism of selective peptide adsorption. *Journal of the American Chemical Society*, 134(14): 6244–6256, 2012.

[150] Hendrik Heinz, Kshitij C Jha, Jutta Luettmer-Strathmann, Barry L Farmer, and Rajesh R Naik. Polarization at metal–biomolecular interfaces in solution. *Journal of The Royal Society Interface*, page rsif20100318, 2010.

[151] R Atkin, VSJ Craig, EJ Wanless, and S Biggs. The influence of chain length and electrolyte on the adsorption kinetics of cationic surfactants at the silica–aqueous solution interface. *Journal of colloid and interface science*, 266(2):236–244, 2003.

[152] H Huang, SY Tong, WE Packard, and MB Webb. Atomic geometry of Si (111) 7×7 by dynamical low-energy electron diffraction. *Physics Letters A*, 130(3):166–170, 1988.

[153] Wilfred F van Gunsteren, SR Billeter, AA Eising, Philippe H Hünenberger, PKHC Krüger, Alan E Mark, WRP Scott, and Ilario G

Tironi. Biomolecular simulation: The {GROMOS96} manual and user guide. 1996.

[154] Walter RP Scott, Philippe H Hünenberger, Ilario G Tironi, Alan E Mark, Salomon R Billeter, Jens Fennen, Andrew E Torda, Thomas Huber, Peter Krüger, and Wilfred F van Gunsteren. The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A*, 103(19):3596–3607, 1999.

[155] Miguel Jorge, José RB Gomes, M Natália DS Cordeiro, and Nigel A Seaton. Molecular dynamics simulation of the early stages of the synthesis of periodic mesoporous silica. *The Journal of Physical Chemistry B*, 113(3):708–718, 2008.

[156] Miguel Jorge, José RB Gomes, M Natália DS Cordeiro, and Nigel A Seaton. Molecular simulation of silica/surfactant self-assembly in the synthesis of periodic mesoporous silicas. *Journal of the American Chemical Society*, 129(50):15414–15415, 2007.

[157] Ryusuke Futamura, Miguel Jorge, and José RB Gomes. Role of the organic linker in the early stages of the templated synthesis of PMOs. *Physical Chemistry Chemical Physics*, 15(17):6166–6169, 2013.

[158] HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91 (24):6269–6271, 1987.

[159] Guang-Hong Lu, Minghuang Huang, Martin Cuma, and Feng Liu. Relative stability of Si surfaces: A first-principles study. *Surface science*, 588(1):61–70, 2005.

[160] A Ramstad, G Brocks, and PJ Kelly. Theoretical study of the Si (100) surface reconstruction. *Physical Review B*, 51(20):14504, 1995.

[161] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, 2009.

[162] Norman Troullier and José Luriaas Martins. Efficient pseudopotentials for plane-wave calculations. *Physical Review B*, 43(3): 1993, 1991.

[163] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics*, 132:154104, 2010.

[164] Jitendra Mata, Dharmesh Varade, and Prashant Bahadur. Aggregation behavior of quaternary salt based cationic surfactants. *Thermochimica Acta*, 428(1):147–155, 2005.

[165] Fredrik Tiberg, Johanna Brinck, and Lachlan Grant. Adsorption and surface-induced self-assembly of surfactants at the solid–aqueous interface. *Current opinion in colloid & interface science*, 4 (6):411–419, 1999.

[166] Hirokatsu Miyata and Kazuyuki Kuroda. Preferred alignment of mesochannels in a mesoporous silica film grown on a silicon (110) surface. *Journal of the American Chemical Society*, 121(33): 7618–7624, 1999.

[167] Ashish V Sangwai and Radhakrishna Sureshkumar. Coarse-grained molecular dynamics simulations of the sphere to rod transition in surfactant micelles. *Langmuir*, 27(11):6628–6638, 2011.

[168] Sharon Ruthstein, Judith Schmidt, Ellina Kesselman, Yeshayahu Talmon, and Daniella Goldfarb. Resolving intermediate solution structures during the formation of mesoporous SBA-15. *Journal of the American Chemical Society*, 128(10):3366–3374, 2006.

[169] Liora Omer, Sharon Ruthstein, Daniella Goldfarb, and Yeshayahu Talmon. High-resolution cryogenic-electron microscopy reveals details of a hexagonal-to-bicontinuous cubic phase transition in mesoporous silica synthesis. *Journal of the American Chemical Society*, 131(34):12466–12473, 2009.

[170] Paolo Raiteri, Raffaella Demichelis, and Julian D Gale. Thermodynamically consistent force field for molecular dynamics simulations of alkaline-earth carbonates and their aqueous speciation. *The Journal of Physical Chemistry C*, 119(43):24447–24458, 2015.

[171] Yujie Wu, Harald L Tepper, and Gregory A Voth. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of chemical physics*, 124(2):024503, 2006.