



University of  
**Strathclyde**  
**Glasgow**

# The Effects of Interfaces on Heterogeneous Nucleation of Urea: Experiments and Simulations

Samira Anker

Department of Chemical and Process Engineering  
University of Strathclyde

A thesis submitted for the degree of  
Doctor of Philosophy  
2024

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

# Abstract

The formation of crystals from solutions is a natural phenomenon, which is frequently used in industrial separation and purification processes. Crystal nucleation commonly occurs heterogeneously at interfaces, however, the processes are difficult to control since the behaviour is not well understood. This work uses a combination of nucleation experiments and atomistic molecular dynamics (MD) simulations to investigate the mechanisms influencing heterogeneous nucleation in aqueous urea solutions.

Quiescent, cooling crystallisation experiments were performed and two high-throughput methods were compared to determine the optimal experimental conditions and setup. Concentrations of 1800-2280 g kg<sup>-1</sup> and temperatures from -5-25 °C were tested, corresponding to moderate and high supersaturations. 15-20 °C provided the best induction time measurements with low nucleation probabilities at 25 °C. Heterogeneous nucleation experiments were performed for aqueous urea solutions in contact with air, glass, diamond and polytetrafluoroethylene (PTFE) surfaces. 1 ml solution samples were contained in glass vials with a solution-air interface, which was taken as the control system. The presence of a PTFE surface was found to significantly increase the nucleation probability, up to  $\times 8$ , in comparison to the control samples, indicating that heterogeneous nucleation occurred at the PTFE surface. The presence of a diamond surface also showed a slight increase of nucleation probability, up to  $\times 4$ .

MD simulations were used to study urea solutions in contact with PTFE, diamond,  $\alpha$ -quartz, and vacuum interfaces. Firstly, a force field validation protocol was developed, since the performance of MD simulations is highly dependent on the underlying

models. Four GAFF and five OPLS force fields were compared, and a urea-specific charge-optimised GAFF force field and the original all-atom OPLS force field were found to be the best for simulating bulk crystals and aqueous solutions of urea. MD simulations of urea solutions at the solid interfaces found that strong solute-surface dispersion interactions enhanced the solute concentration near the interface. The interfacial urea concentration was enhanced by 32-59% and 13-29% for the PTFE and diamond surfaces respectively. The enhanced interfacial concentration contributes to the explanation for the observed heterogeneous nucleation rates and provides a new basis for guiding materials selection to improve crystallisation processes and to reduce fouling.



Figure 1: Artistic impression of the essence of this research, which was included in the Images of Research outreach exhibition at the University of Strathclyde in 2022.

# Acknowledgements

I would like to thank all the people who have played an important part during my PhD journey.

Firstly, I am ever so grateful to my supervisors Karen Johnston, Jan Sefcik, and Paul Mulheran, for providing me with this experience. I have particularly appreciated your numerous interesting ideas and insights, helpful feedback and our many detailed discussions. I am also thankful to David McKechnie for his patience when showing me the ropes and all his help with troubleshooting along the way.

No research is done in isolation, so I want to thank all my peers and colleagues for helping to create a varied, vibrant, and valuable workplace. I really appreciate both the direct and indirect support from my friends, thank you to each of you for joining me on this journey in your own way.

I am very thankful to my family, especially my parents, for their continuous love, guidance, and encouragement in following my dreams. Finally, I want to thank my partner Andreas who has supported me throughout, you are an invaluable friend, thank you for everything.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Layout of the Thesis . . . . .	2
1.2 Previously Published Works . . . . .	3
<b>2 Background</b>	<b>4</b>
2.1 Crystallisation . . . . .	4
2.2 Model System: Urea from Aqueous Solution . . . . .	17
2.3 Research Challenges, Hypothesis and Aims . . . . .	21
<b>3 Methodology</b>	<b>24</b>
3.1 Molecular Dynamics . . . . .	24
3.2 Force Fields . . . . .	34
3.3 Nucleation Experiments and Measurements . . . . .	41
3.4 Nucleation Rate Models . . . . .	44

<b>4</b>	<b>Validation of Force Fields</b>	<b>49</b>
4.1	Methodology . . . . .	49
4.2	Results and Discussion . . . . .	53
4.3	Conclusions . . . . .	72
<b>5</b>	<b>Urea Nucleation and Solution Structure at Interfaces</b>	<b>76</b>
5.1	Methodology . . . . .	77
5.2	Results and Discussion . . . . .	86
5.3	Conclusions . . . . .	107
<b>6</b>	<b>Temperature Cycling and Urea Nucleation</b>	<b>109</b>
6.1	Methodology . . . . .	110
6.2	Results and Discussion . . . . .	113
6.3	Conclusions . . . . .	126
<b>7</b>	<b>Conclusions</b>	<b>128</b>
7.1	Findings and Outlook . . . . .	128
7.2	Limitations and Future Research Directions . . . . .	132
<b>A</b>	<b>Additional Force Field and Validation Details</b>	<b>136</b>
A.1	Force Field Parameters and Equations . . . . .	136
A.2	Tabulated Results . . . . .	145
A.3	Further Radial Distribution Functions . . . . .	152
<b>B</b>	<b>Additional Details and Analyses for Interfacial Studies</b>	<b>157</b>
B.1	Force Field Parameters and Equations . . . . .	157

B.2 Simulation Results . . . . .	160
B.3 Experimental Results . . . . .	169
<b>C Additional Temperature Cycling Analyses</b>	<b>173</b>
C.1 Breakdown of Experiments A-C . . . . .	174
C.2 Mass Change . . . . .	177
C.3 Unconditional Probability Distributions . . . . .	179
C.4 Fitted Nucleation Rates . . . . .	180
<b>Bibliography</b>	<b>182</b>

# Chapter 1

## Introduction

Crystallisation is the formation of an ordered solid from a liquid or solution, and it proceeds through the birth of a new crystal, known as nucleation, followed by the growth of this crystal nucleus. Crystallisation is common in the natural world with examples including the freezing of water to ice, and the formation of minerals such as salt or stalagmites and stalactites. Industrially, crystallisation is commonly used as a separation technique with many applications including in the chemicals, food and drink, and pharmaceutical sectors. The nucleation process influences many of the desired quality attributes of the crystalline products such as solid form, size distribution, morphology, and purity.<sup>1,2</sup> It is important to have a good understanding of nucleation behaviours to enable the development and operation of well controlled crystallisation processes.

Nucleation is a stochastic process, consequently, experimental studies are based on producing many replicates which can be combined to obtain reliable rate data for the process. Small samples are often used in nucleation studies, with volumes typically being on micro- or millilitre scale. There is a large surface area to volume ratio for small samples. Therefore, surface effects which influence nucleation are more prominent in small scale experiments and can influence the results if they are scaled to larger systems with significantly smaller surface area to volume ratios. This work focuses on improving our understanding of surface induced nucleation, which is known as heterogeneous

nucleation.

There is a growing range of computational tools which can be used to study crystal properties and crystallisation processes. Molecular simulations are commonly used to model crystal growth and dissolution processes, and can even model rare nucleation events for some systems,<sup>3,4</sup> allowing these behaviours to be studied on a molecular level which is not possible experimentally. This work uses a combination of nucleation experiments and molecular simulations to explore heterogeneous nucleation across multiple scales.

## **1.1 Layout of the Thesis**

The layout of the thesis is as follows:

Chapter 2 provides a general introduction to crystallisation and nucleation, including the influence of solid, liquid, and gaseous interfaces on nucleation. The use of molecular simulations to model crystallisation processes and interfacial influences is discussed. There is a discussion of the merits and properties of the aqueous urea model system which has been selected for this work. The research challenges, hypothesis and aims of this work are outlined.

Chapter 3 outlines the general methodologies for this work. The principles of molecular dynamics simulations are described and the available models (force fields) are compared. The stochasticity of nucleation experiments and need for high-throughput experiments is discussed, with an overview of the available experimental equipment and set-ups. The statistical models for analysing nucleation kinetics are considered.

Chapter 4 focuses on the comparison and validation of force fields for modelling crystallisation processes. The specific simulation details and system set-up are given for both bulk crystals and solutions. The crystal properties considered include lattice parameters, density, and cohesive energy. The solution properties considered are density, diffusion coefficients, and radial distribution functions. The many urea force fields short-listed in Chapter 3 are compared, and those best suited for future crystallisation work

are identified. Additional simulation details and analyses are provided in Appendix A. Chapter 5 considers the influence of control (glass and air), polytetrafluoroethylene (PTFE) and diamond interfaces on urea nucleation. The specific simulation details and experimental set-up is detailed. The different nucleation behaviours at these interfaces is discussed considering the impact on induction time and nucleation rates. The molecular level influence of these interfaces on the solution structure is simulated providing additional explanations for the experimental observations. Additional simulation details and analyses are provided in Appendix B.

Chapter 6 explores the use of automated thermal cycling to increase experimental throughput whilst minimising set-up time. The influence of temperature, concentration, and stirring on the quality of results is discussed. The variation in nucleation induction time with cycle number and between individual vials is discussed. Additional analyses are provided in Appendix C.

Chapter 7 finishes with a discussion of the overall conclusions that can be drawn from this work. The limitations of this work are discussed and suggestions are given for how the results obtained could be used to direct future research.

## 1.2 Previously Published Works

The force field discussion and validation work presented in Sections 2.2 and 3.2, Chapter 4 and Appendix A has been previously published as:

Anker, S.; McKechnie, D.; Mulheran, P.; Sefcik, J.; Johnston, K. Assessment of GAFF and OPLS Force Fields for Urea: Crystal and Aqueous Solution Properties. *Crystal Growth & Design* **2024**, *24*, 143–158 <https://doi.org/10.1021/acs.cgd.3c00785>

## Chapter 2

# Background

### 2.1 Crystallisation

The crystallisation of a solid from a solution does not usually occur instantaneously, since it requires the spontaneous formation of a crystalline nucleus either in the bulk solution or at an interface with the solution. Nucleation is a stochastic process, which requires concentrations above the solubility limit to occur. The rate of nucleation is dependent on the concentration, temperature and volume of the solution and can be influenced by surface effects, agitation, the presence of other crystals, additives, pH, etc. The mechanisms of the nucleation process and the direct influence of experimental conditions are not well understood, which makes it difficult to predict the behaviour and develop robust and efficient crystallisation processes.

Nuclei that reach the critical nucleus size will grow to form a detectable crystal, however if the nucleus does not reach the critical size it will dissolve, this is illustrated in Figure 2.1. As the crystallisation process proceeds the solution concentration decreases, crystal growth can continue to occur at lower concentrations, still above the solubility limit, where nucleation becomes negligible. The crystal growth rate is dependent on the transport of new molecules to crystal surface and on the ease of attaching those molecules, which is related to the shape and type of surface. Processes with high nu-

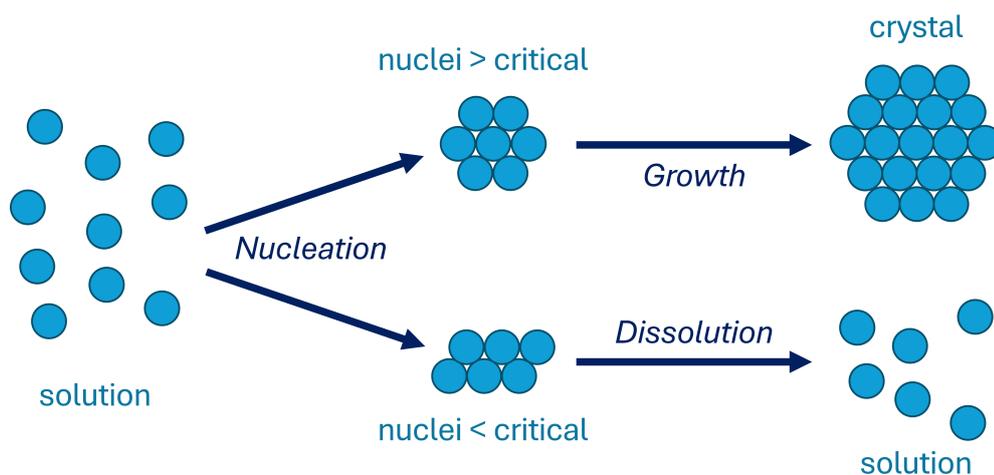


Figure 2.1: The outcome of a nucleation event depends on the size of the nuclei compared to the critical size.

creation rates will typically lead to the presence of many, small crystals, whereas the crystals are expected to grow larger if there is only a limited number of nuclei available.

Most molecules have more than one crystal form, these different structures are known as polymorphs. There are many theoretical polymorphic structures possible for a molecule, however, only a few are typically obtainable experimentally.<sup>5-8</sup> The physical properties of different polymorphs can vary significantly, affecting both the manufacturability and the product qualities. It is very important to be able to accurately and consistently crystallise particles with the desired size and shape and of a specific polymorph for the pharmaceutical industry.<sup>8-11</sup> Therefore, it is important to understand how the nucleation process may affect the crystallisation outcome.

### 2.1.1 Supersaturation

For nucleation to occur a solution must be supersaturated. A saturated solution is where the chemical potential of the solute is equal to the chemical potential of the solid phase, this is at the solubility limit of the solute, which is dependent on the solvent and temperature and is highly system dependent. Solutions that are at concentrations lower than the saturation concentration are undersaturated, whilst supersaturated solutions

are at concentrations greater than the saturation concentration.

There are many definitions used to quantify supersaturation, for example by calculating the relative degree of supersaturation,  $S$ , using Eq. 2.1, where  $c$  is the solution concentration, and  $c_{\text{equil}}$  is the solubility limit or equilibrium concentration:

$$S = \frac{c}{c_{\text{equil}}} \quad (2.1)$$

From this definition, values of  $S = 1$ ,  $S < 1$  and  $S > 1$  indicate saturated, undersaturated and supersaturated solutions, respectively. Supersaturation can be achieved by either increasing the concentration of the solute or by decreasing the solubility of the solute. Means of increasing the solute concentration include evaporating the solvent or in specific cases a more soluble metastable polymorph can be added, where the target concentration is above the solubility of the less soluble polymorph but below the solubility of the more stable polymorph. Means of decreasing the solute solubility include changing the temperature, adding a secondary solvent, an additive or changing the pH.

Supersaturated solutions are not thermodynamically stable or equilibrated. However, solutions with moderate supersaturations can exist in a metastable state, this metastable zone is system dependent, and is influenced by experimental conditions. Most crystallisation processes are performed in the metastable zone where nucleation is highly system dependent and occurs stochastically. The solution will become unstable if higher supersaturations are reached, this will lead to instantaneous nucleation without any control over the process. This crystallisation phase diagram is summarised in Figure 2.2, for a typical organic molecule where temperature significantly influences solubility. A crystal placed in an undersaturated solution will dissolve until the solution becomes saturated achieving an equilibrium between the crystal and the solution. However, the crystal will dissolve completely if the concentration is so low that an equilibrium is never reached and the solution remains undersaturated. Conversely a crystal placed in a supersaturated solution will grow (alongside any other crystals and nuclei) until the saturated concentration is reached and the system becomes equilibrated.

Thermodynamically, once a solution is supersaturated, crystallisation should occur,

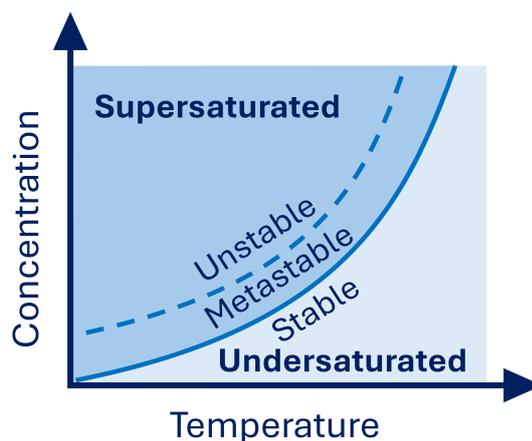


Figure 2.2: A typical crystallisation phase diagram for an organic molecule.

since the supersaturated solution is metastable and crystallisation would lead to the production of stable crystal and solution phases.<sup>12</sup> However, there are free energy barriers associated with the formation of a nucleus, which must be overcome, this often results in slow nucleation kinetics, making the process stochastic.<sup>12</sup> The formation of a crystal nucleus leads to the formation of a new surface and the surface free energy of this interface with the surrounding solution is unfavourable. This needs to be balanced by the favourable interactions between molecules within the nucleus. Therefore, a nucleus may begin to form but be unable to overcome the barriers and so break up. When the nucleus has reached a sufficiently large size, the favourable interactions surpass the unfavourable surface free energy costs, and the nucleus is stabilised. This size is the critical nucleus size, which is system dependent, but is typically on the order of nanometres containing 10-1000 molecules, although it can be larger.<sup>13</sup>

### 2.1.2 Nucleation Mechanisms and Theories

There are several theories available to describe the formation of crystal nuclei, the most common theory is the Classical Nucleation Theory<sup>1,14</sup> (CNT). According to CNT, stochastic density fluctuations lead to the formation of small ordered clusters (crystal nuclei) of the constituent molecules (or atoms).<sup>1,14</sup> Further molecules can attach to this crystal nucleus, but molecules can also detach freely. The stability of the nucleus is

dependent on the bulk and surface free energies, which depend on the crystal nucleus size. The interactions between the molecules in the bulk crystal lattice are energetically favourable. However, the creation of a new interface, between the new crystal nucleus and solution, results in an unfavourable surface free energy. The unfavourable surface free energy is dominant in small nuclei, however as the nucleus size increases above the critical nucleus size the bulk free energy becomes dominant, stabilising the nucleus.<sup>1,14</sup> CNT assumes that the nucleus has the same bulk properties, such as density, structure and composition, as the stable macroscopic crystal, and that there is a clear nucleus-solution boundary.<sup>1</sup>

CNT is a compelling qualitative model, but due to the many simplifications involved CNT struggles to predict quantitative nucleation properties such as nucleation rate.<sup>15,16</sup> Therefore, alternative non-classical nucleation models have been developed, including Two-Step Nucleation<sup>15</sup> (TSN) and Pre-Nucleation Clusters<sup>16,17</sup> (PNC).

The Two-Step Nucleation Theory was developed for protein systems but has also been found to be applicable to small organic molecules.<sup>1,15</sup> TSN involves density fluctuations bringing together the solute molecules to form a new dense phase, this is a liquid-like cluster, which is typically metastable.<sup>1,15</sup> Nucleation occurs within this dense cluster and the nuclei can grow by the addition of other molecules from outwith the cluster.<sup>1,15</sup> There are energy barriers associated both with the formation of the dense clusters and with the formation of the nucleus within the formed clusters, the latter is typically the rate limiting step.<sup>1,15</sup> Here the fluctuations of density and fluctuations of molecular order act independently.<sup>1,15</sup>

Pre-Nucleation Clusters are typically observed in inorganic systems,<sup>1</sup> but mesoscale clusters have also been observed and linked to nucleation in organic systems such as glycine and valine solutions.<sup>18-20</sup> The PNCs consist of the molecules (or atoms or ions) that will make up the crystal but can also include other species too.<sup>17</sup> PNCs are small, thermodynamically stable groupings of solute molecules, which may resemble the crystalline structure but may also be amorphous, and act as precursors to the nucleation phase change.<sup>17</sup> However, PNCs are not the formation of a second phase as they are

a constituent of the mesostructured solution phase. There is not necessarily a clear boundary between the cluster and the solution and the PNCs can change rapidly.<sup>16,17</sup>

There are two types of nucleation: primary and secondary. Primary nucleation is when the first crystal nucleus forms in a supersaturated solution in the absence of crystals of the solute. Primary nucleation can be either homogeneous or heterogeneous. Homogeneous nucleation occurs in the bulk of the solution and scales with volume, the more solution there is, the larger the probability of nucleation occurring at any point in time. Heterogeneous nucleation occurs at the solution boundary such as the container wall, an agitator blade, the air above the solution, or particles suspended in the solution, and scales with the area of the relevant surfaces. Secondary nucleation occurs when crystals of the solute are already present in the solution. Secondary nucleation often makes use of seed crystals placed in the solution on purpose at the desired crystallisation point, this helps to control crystal size and polymorphism and reduce the significance of primary nucleation. Seeding has been commonly used in industrial crystallisation processes, in particular in the pharmaceutical industry to achieve the control of polymorphism.<sup>10,21-23</sup> High product quality is ensured by the implementation of carefully developed seeding stages with tightly controlled temperature profiles and schemes for the addition of seeds. This requires substantial resources to be spent on process development and the preparation of seeds, which is unique to each process and possibly even the stage of development. However, seeding is not always feasible and it does not necessarily eliminate the possibility of uncontrolled nucleation either during the crystallisation process or fouling on other surfaces such as agitators, probes and pipes. Therefore, there is still plenty of scope for a better fundamental understanding of primary nucleation mechanisms to significantly improve both current and future crystallisation processes.

The most prevalent type of primary nucleation is heterogeneous nucleation. The surface of a nucleus that forms at an interface will both be in contact with the solution and the interface, which reduces the contact area with the solution. If the crystal-surface interactions are more favourable than the crystal-solution interactions, then this helps

to stabilise the nucleus by reducing the overall interfacial free energy barrier and reduces the critical nucleus size, leading to increased rates of heterogeneous nucleation. Heterogeneous nucleation can be induced by the presence of any interfaces in the solution. Heterogeneous nucleation may occur on surfaces which are inextricably part of the set-up such as the walls of the vessel or piping, or on the surfaces of agitators and probes. Despite its prevalence in many important crystallisation processes, the detailed understanding heterogeneous nucleation is still limited, which makes these processes difficult to control.<sup>1,2,21</sup>

There are many distinct factors which may be involved in heterogeneous nucleation, which can relate to both the topography and the functionality of the surfaces. The roughness of a surface is related to nucleation for many compounds and nano-pores offer large surface areas for nucleation as well as confinement effects which can be varied to control the polymorphic structures that are able to form.<sup>21</sup> The chemical interactions between the surface and solution may directly or indirectly assist the formation of nuclei and can help to control the polymorphic outcome. These surfaces may enhance nucleation by fixing the molecule in a location or even aligning it in an orientation, which is favourable for the formation of the crystal nuclei. These heterogeneous nucleation behaviours can be utilised by the manipulation of surface templates to enhance crystal nucleation and may offer selectivity for obtaining the desired polymorph.<sup>1,15,21</sup>

Laboratory studies involving nucleation are often conducted using either cooling or evaporation crystallisation. Evaporative studies rely on evaporation of the solvent to supersaturate the solution for nucleation to occur. Typical set-ups consist of many solution droplets spread over a solid surface, which creates significant solution-air and solution-solid interfaces. Cooling experiments are typically performed in sealed vials or under a layer of inert oil to prevent evaporation of the solution, which creates significant solution-container and solution-oil interfaces. Although these types of interfaces do not necessarily have any specific chemical interactions they may still have significant effects on the nucleation rate.<sup>24-27</sup> In laboratory experiments, the effects of heterogeneous interfaces are often enhanced, since most work is carried out with a large number of

small samples, which have a high surface area to volume ratio.

Small organic molecules are typically used as model compounds for conducting academic research into the crystallisation of pharmaceuticals. Common molecules used include small drug molecules such as paracetamol, carbamazepine, and aspirin, as well as other easily accessible molecules such as lactose, glycine, and other amino acids. Heterogeneous nucleation can be related to solid, liquid, or gaseous interfaces, which are all considered here.

### **Solid Heterogeneous Interfaces**

Crystallisation experiments are based on the liquid phase with a minimum experimental requirement for a vessel or container that will hold the solution, which introduces at least one solid surface, and additional solid surfaces are easy to introduce. The consideration given to these surfaces ranges from minimal, where they are just an inherent part of a practical setup, to being carefully designed to optimise nucleation. The influence of surfaces will change with the scale of the experiment as the surface area to volume ratio changes. However, even smaller surfaces which may not be considered for their nucleating properties, such as thermocouples and other in situ probes, can have surprising effects on nucleation.

The influence of solid surfaces on nucleation includes both chemical and physical features, and has been widely studied, with many types of surface materials and techniques available. The focus here is on the chemical interactions with surface materials such as glass, polymers and metals, and the surface treatments used to obtain specific surface functional groups.

The presence of a platinum surface is suggested to influence molecular alignment based on solution-surface interactions, which has been shown to enhance the nucleation of unstable glycine polymorphs.<sup>28</sup> The presence of a polytetrafluoroethylene (PTFE) interface in glycine solutions has been found to increase the nucleation rate in quiescent conditions, without influencing the polymorphic outcome.<sup>25</sup> Crystallisation experiments in quiescent carbamazepine solutions have found that the addition of PTFE

and glass interfaces promoted the formation of the metastable polymorph, compared to solutions with no additional interfaces present.<sup>29</sup> In the same system additional tin interfaces led to an increase in the concomitant formation of both the metastable and stable polymorphs.<sup>29</sup>

The influence of specific chemical groups can be tested using functionalised surfaces or self-assembled monolayers. The nucleation of carbamazepine from glass vials had a slow nucleation rate with several polymorphs forming concomitantly, whereas functionalising the vials with cyano-, mercapto- or fluoro-terminated molecules both increased the nucleation rate and allowed for selective polymorph formation.<sup>27</sup> This indicates that the polymorph outcome might be dependent on the polarity of the surface groups, interestingly this dependence on the surface did not apply at very low or high concentrations.<sup>27</sup> Self-assembled monolayers of small molecules with desired head-groups are typically built on metal surfaces and produce a very homogeneous array of the exposed surface head-group. Specific interactions with the surface can also aid nucleation by keeping the molecules in orientations favourable to the crystal structure.<sup>30</sup> This has been used to steer nucleation experiments by helping to control experimental conditions and stabilise alternative polymorphs.<sup>31,32</sup>

Another development is the addition of solid nucleation substrates to the solution where crystals form and grow attached to the substrate surface, which remains part of the final crystal product.<sup>33,34</sup> This technique has been effectively used to control parameters including particle size distribution and product yield.<sup>33,34</sup> This requires the material to be carefully selected to avoid any adverse effects on the product, for example for pharmaceuticals this will need to be an accepted excipient or inactive ingredient.

### **Liquid Heterogeneous Interfaces**

Liquid interfaces may be more difficult to implement in a larger scale crystallisation process but are frequently introduced to small scale nucleation experiments. Nucleation studies make use of many repeated experiments to get statistically representative results, which are important for understanding the nucleation kinetic effects. The use of

microwell plates allows large arrays of small droplets to be tested simultaneously, with minimal material and time requirements, and to prevent evaporation the solution is kept under a layer of oil.<sup>35</sup> Microfluidic chips consist of narrow channels where small droplets of liquids can be studied as they flowed through the channels, separated by a secondary medium, which is typically an oil.<sup>36</sup> This can be used to study crystal nucleation while optimising the use of materials and space.

The small scale of these experiments mean that the relative surface area to volume ratio of the solution is very large, it is therefore important to consider any effects this may have on the results. Oil layers have been considered to influence nucleation in various ways including by affecting the diffusion properties of the system,<sup>37</sup> or by containing additional impurities that may enter solution and act as heterogeneous nucleant,<sup>38</sup> and it can also absorb or dissolve some of the solute from the solution. The effect of the oil layer on nucleation rate has been included in some quantitative analyses through the inclusion of a surface energy term.<sup>36,39</sup>

Recent work has directly investigated the influence of oil on nucleation rates by performing controlled nucleation experiments using aqueous glycine solutions as a model system. This found that the nucleation rate was significantly increased by the addition of an oil layer, compared to vials without it.<sup>26</sup> These experiments were conducted in sealed vials to minimise evaporative effects. The experiments were coupled with simulations that indicated that the oil increased local concentration in the interfacial solution layer.<sup>26</sup> This indicates that there are interactions between the oil and the solution which influence nucleation.

### **Gaseous Heterogeneous Interfaces**

Gas interfaces are also an inherent part of many crystallisation processes. Evaporative studies must have a gas interface to facilitate the evaporation and there are some studies which consider the effect of this on nucleation. Microfluidic and microdroplet experiments have also been designed to analyse the impact of gas interfaces. The deliberate addition of gas bubbles has been explored as a new technique to influence nucleation

rates.<sup>40</sup>

Evaporative nucleation experiments are common since large arrays of small droplets can easily be prepared and monitored to observe nucleation as evaporation proceeds. The solution concentration, droplet size and evaporation rates are typically varied to understand the influence on the polymorphic outcomes, more detailed studies have also considered the location within the droplet where nucleation occurs.<sup>28,31,32</sup> For example, in glycine solutions it was found that nucleation at the air-solution-solid contact point promoted the formation of less stable  $\beta$ -phase.<sup>28</sup> This was attributed to the solution-solid interface and will have been aided by the changes in the local concentration related to the ongoing evaporation, but still leaves the interesting question regarding any solution-air interface effects.

Microfluidic experiments have also been devised to test the effect of gas bubbles on nucleation. Experiments conducted both in stagnant and flow systems indicated that the gas-solution interface promoted nucleation.<sup>40</sup> For example, visual analysis of paracetamol crystallisation with added gas ( $N_2$ ) bubbles found that crystals formed at both the gas-solution interface and in the bulk solution.<sup>40</sup> Interestingly, the crystals that nucleated in the bulk solution later attached themselves to the gas-solution interface, this indicates that the free energy of the crystal was lowered by surface tension effects, which implies that the gas-solution interface provided favourable conditions for crystallisation.<sup>40</sup>

Experiments using hanging microdroplets of lysozyme solution found that additional air bubbles within the droplet both reduced the required nucleation time and increased the crystal mass yield.<sup>41</sup> Analysis of microscope images of the droplets showed that crystals formed at the edge of the bubbles, which even influenced the shape of some of the crystals.<sup>41</sup>

‘Gassing crystallisation’ is an emerging technique which involves the generation of many nano- or micro-sized bubbles in the range of 50 nm to 2500  $\mu\text{m}$ . These bubbles appear to act as an effective nucleant for small organic molecules (including paracetamol, lactose, amino acids, and carboxylic acids), proteins, and salts using gasses such as

N<sub>2</sub>, O<sub>2</sub> and CO<sub>2</sub>. There are many possible mechanisms which may be involved in increasing the nucleation rates.<sup>42</sup> Dissolved gasses from the bubbles may change the pH of the solution which can affect the solubility, impacting the supersaturation. Solvent molecules may evaporate from the solution and enter the bubbles thus increasing the local solution supersaturation. The addition of bubbles can lower the temperature of the surrounding solution if pressurised gas is used, since the gas temperature drops during depressurisation, which will affect local supersaturation. Increased agitation and mass transfer by the movement of the bubbles may also influence the nucleation rate. The processes can be tuned to increase or decrease the crystal growth rate independently of the nucleation rate since these mechanism are affected by various factors.<sup>42</sup>

### 2.1.3 Simulations of Nucleation and Growth

Molecular simulations can provide additional insights into molecular behaviour and interactions, for this, molecular dynamics (MD) simulations are commonly used. These simulations are based on mathematical models of the interactions of atoms and molecules, known as force fields. There are many force fields available, some have broad applications whilst others have been developed to model specific properties of one material. Simulations are only as reliable as the models they are based on and slight differences in the models used and their implementation can lead to significant differences in the results. The scope of molecular simulations is restricted by computational power, one of the main limitations is the small time- and length-scales that are computationally achievable, this means that kinetic data is often either inaccessible or inaccurate. Nevertheless, molecular simulations still provide many interesting observations and new insights on a molecular level which is very difficult to obtain experimentally. It is important that results are analysed carefully and critically and considered in the context of the physical phenomenon simulated. Simulations have been performed to investigate various aspects of crystal growth and dissolution, and enhanced simulation techniques have enabled the direct study of nucleation mechanisms.

MD simulations can be used to illustrate the effects of experimental conditions on crys-

tal growth and structure, which have previously been assumed based on experimental observations. Crystal growth and dissolution can be studied relatively simply by placing crystals in solution (or just solvent), these crystals are often larger than those used for seeding simulations. This has been used for many purposes including to study growth mechanisms and the growth of particular crystal faces,<sup>43,44</sup> and dissolution rates and mechanisms.<sup>44,45</sup>

Nucleation is difficult to study directly using MD simulations due to the long induction times of nucleation experiments compared to the very short timescales obtainable computationally, meaning that the chance of observing nucleation events in simulations is low. Nucleation can still be studied in MD simulations using techniques such as seeding<sup>46</sup> or enhanced sampling techniques such as well-tempered metadynamics.<sup>4,47–50</sup> Seeding involves placing a small crystallite, the seed, in a saturated solution and this can be used to study properties such as critical nucleus size and structure, and even estimate nucleation rates.<sup>46</sup> These seeds can be made up of between tens to thousands of molecules, depending on the system investigated.<sup>46</sup> The insight gained from the seeding approach is limited since it bypasses the kinetic challenges of nucleation and does not address the initial nuclei formation steps. Well-tempered metadynamics allows increased and improved sampling of the system which enhances the likelihood of nucleation events occurring, allowing this to be studied directly.<sup>4,47–50</sup>

### **Interfacial Effects**

To the best of our knowledge, MD studies of heterogeneous nucleation from bulk solutions have not been performed. Instead, recent work studying interfaces which are known to influence nucleation has found a link between solution-interface dispersion interactions, molecular scale interfacial solution layers and the experimentally observed nucleation behaviour.<sup>25,26,51</sup>

MD simulations of supersaturated aqueous glycine solutions in contact with a film of oil found that the local glycine concentration was enhanced in the solution near the interface.<sup>26</sup> Conversely, if the oil was replaced by air this led to a decrease in the in-

terfacial glycine concentration.<sup>26</sup> These simulations were designed to investigate the previously discussed experiments which found that an oil layer significantly increased the nucleation rates.<sup>26</sup> The higher local solution concentration near the oil can explain the enhanced nucleation rates experienced with the additional oil interfaces, since nucleation rate is directly related to concentration. This work proposed that the interfacial concentration effect was related to van der Waals interactions, specifically that the enhanced glycine concentration was due to stronger oil-glycine interactions compared to oil-water interactions.<sup>26</sup> Further simulations, where the van der Waals dispersion interactions were modelled in isolation, led to similar interfacial phenomena, which was also found to be significant for materials such as graphite and PTFE.<sup>51</sup>

This simulation approach to studying the causes of heterogeneous nucleation is very interesting since it provides an insight into what surface interactions are important, and why they are effective. Despite differences in the type of data obtained, links can be made between experimental observations and simulated results. These results also act as a potential starter for more detailed simulations of heterogeneous nucleation. It would be interesting to apply the techniques that were used to simulate homogeneous nucleation in the presence of interfaces, such simulations can be used to investigate previously overlooked phenomena, such as significantly increased interfacial layer concentrations, which are applicable to the heterogeneous systems in general.

## 2.2 Model System: Urea from Aqueous Solution

Many different small drug molecules and other organic molecules have been regularly used in nucleation experiments. However, molecular simulations of crystallisation processes and particularly nucleation has only been performed using a small number of molecules. This is due to very small molecules with simple crystal structures being favoured due to limitations in computational power. Urea is the only organic molecule that has been modelled extensively with MD simulations to study crystal growth, dissolution and nucleation behaviour.<sup>3,43-49,52</sup>

Urea is an interesting molecule with very high aqueous solubility and rapid crystal growth, and is used for a wide range of applications. The uses of urea include as a fertiliser,<sup>53</sup> as a component of resins,<sup>53</sup> as a reagent in chemical synthesis<sup>54</sup> (including as an additive to vehicles to help reduce harmful emissions), and as an additive in creams to help rehydrate skin. There is only one known polymorph of urea at standard conditions which simplifies experiments since different polymorphs need not be considered. Experimentally, urea has previously been studied to get an understanding of the kinetics and mechanisms of nucleation and growth,<sup>55–59</sup> and has been used as a model compound for testing various new crystallisation techniques including laser induced nucleation,<sup>60–62</sup> and the use of solid-state NMR to monitor crystallisation *in situ*.<sup>63</sup>

Urea, in aqueous solutions, has been selected as the model crystallisation compound for this work. The previous nucleation simulations with urea have been very insightful and indicate that the molecule is amenable to various simulation techniques, whilst there is still scope for further work such as looking at heterogeneous effects. Additionally, the use of urea will complement the previous interfacial simulations performed with glycine and help determine the generalisability of those results.

Urea is highly soluble in water with a solubility of  $1200 \text{ g}_{\text{urea}} \text{ kg}_{\text{water}}^{-1}$  (referred to from here onward as  $\text{g kg}^{-1}$ ) at normal conditions (300 K and 1 atm).<sup>64</sup> The reported aqueous solubility is very consistent between different sources as shown in Figure 2.3. Urea is also soluble in several organic solvents such as alcohols.<sup>49,65</sup>

Urea is a small, symmetric organic molecule, with four known crystal structures that have been observed experimentally, denoted here as forms I, III, IV and V, where only form I exists at ambient conditions. Lattice parameters are only available for forms I, III and IV and these three structures are shown in Figure 2.4. Form I and IV have a similar structure, with the same number of  $\text{NH}\cdots\text{O}$  hydrogen bonds, the structure of form III differs more and forms one fewer  $\text{NH}\cdots\text{O}$  hydrogen bond.<sup>68</sup>

Forms I and IV have only two molecules in the unit cell, although four are shown here to give a better visualisation of the packing structure. Form I has the space group

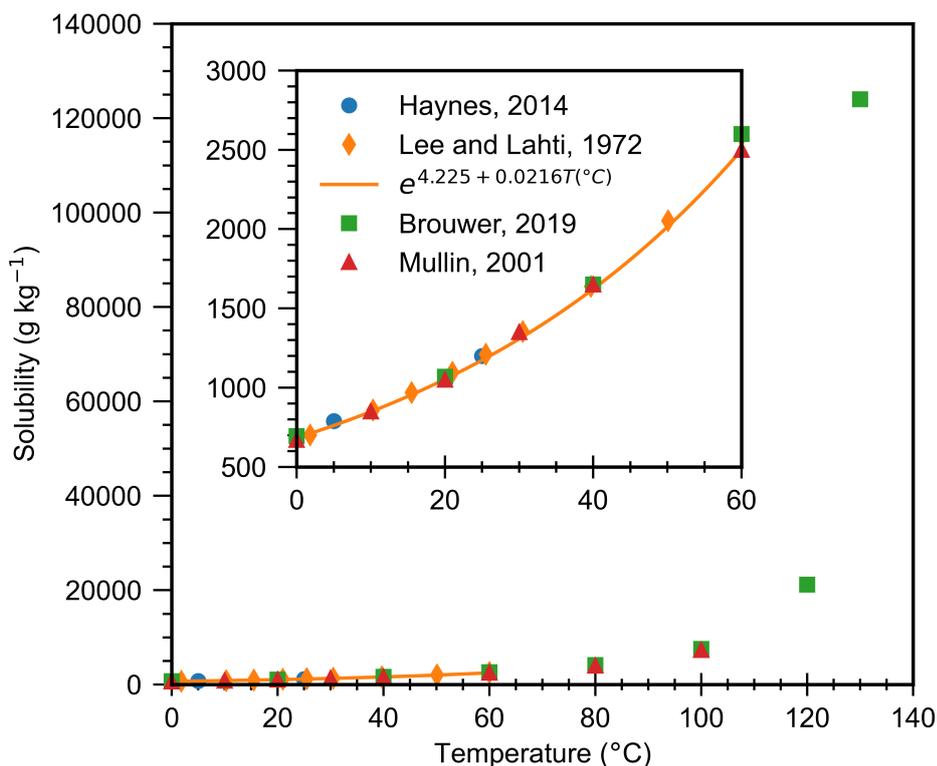


Figure 2.3: Solubility of aqueous urea across a wide temperature range, high urea concentrations causes boiling point elevation of water allowing data to be measured above 100 °C. References are Haynes,<sup>66</sup> Lee and Lahti,<sup>65</sup> Brouwer,<sup>67</sup> and Mullin.<sup>64</sup>

P4<sub>2</sub>m (no. 113), which is a tetragonal structure with lattice parameters  $a = b \neq c$  and  $\alpha = \beta = \gamma = 90^\circ$ ,<sup>69</sup> as seen in Figure 2.4a. The form I unit cell is composed of two urea molecules that are perpendicular to each other, when viewed along the  $z$  direction. A form II structure was initially discovered, but it has not been observed since and it is thought that form II corresponds to form IV.<sup>68,70</sup> Form III has the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (no. 19), which is an orthorhombic structure with lattice parameters  $a \neq b \neq c$  and  $\alpha = \beta = \gamma = 90^\circ$ , as seen in Figure 2.4b. The form III unit cell is composed of four urea molecules, this is a high-pressure form, observed experimentally above 0.48 GPa.<sup>71</sup> Form IV has the space group P2<sub>1</sub>2<sub>1</sub>2 (no. 18), which is also an orthorhombic structure with lattice parameters  $a \neq b \neq c$  and  $\alpha = \beta = \gamma = 90^\circ$ , as seen in Figure 2.4c, this structure is also a high-pressure form observed experimentally above 2.80 GPa.<sup>71</sup> The

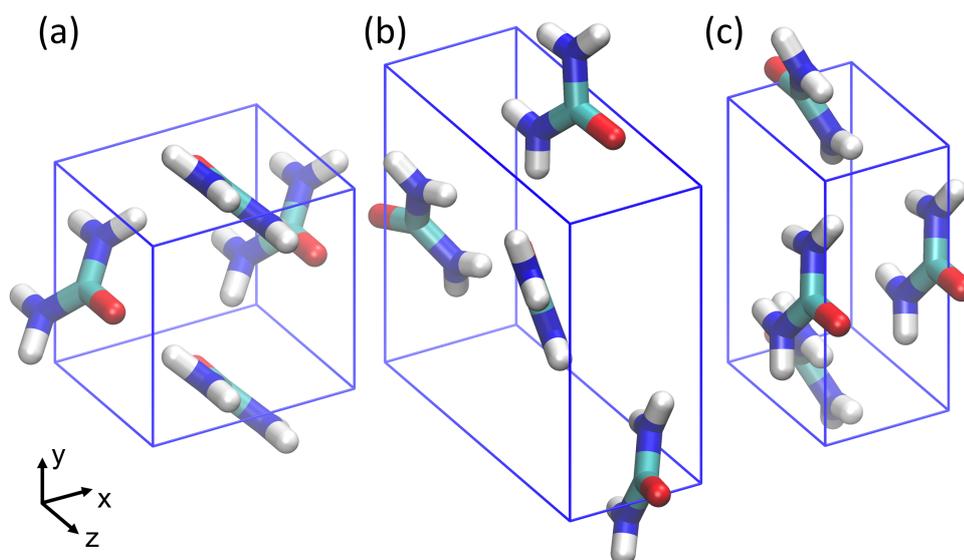


Figure 2.4: Urea unit cells for (a) form I, (b) form III and (c) form IV.

form IV unit cell is also composed of two urea molecules, it is similar to form I, except that the molecules are aligned in a herringbone pattern along the  $z$  direction instead of being perpendicular to each other, and the overall structure is more compressed. Form V is only observed at pressures above 7.8 GPa, and no lattice parameters have been obtained for this structure.<sup>70–72</sup> MD simulations of urea crystals in solution have identified an additional crystal form which is only relevant in nuclei smaller than the critical nucleus size.<sup>47–49</sup>

Experimentally, urea crystals grown from aqueous solutions have only two facets,  $[001]$  and  $[110]$ , the first of these grows much faster than the second, leading to needle shaped crystals. MD studies of the  $[001]$  urea crystal face shows that this grows via rough growth, not using a layer-by-layer mechanism, and the removal of surface defects is the rate limiting step.<sup>3,44</sup> The growth of the  $[110]$  crystal face is based on the birth and spread mechanisms.<sup>3</sup> Similarly the rate limiting step of dissolution, based on studies of spherical crystals without any particular faces, is the detachment of molecules from the surface and this is not affected by a significant diffusion layer.<sup>45</sup> Experimentally, it is known that additives, like biuret, influence the crystal growth of urea resulting

in fewer needle-like crystals.<sup>56</sup> Insights from MD simulations showed that the additive only interacts with and bonds to the [001] surface, slowing down the growth of this facet, whilst not affecting growth on the [110] face.<sup>3</sup> If the growth rates of the two face types are more similar, this will lead to crystal morphologies that are more cubic and less needle-like, which aligns with the experimental observations. Furthermore, MD simulations of urea crystal growth with and without additives, in different solvents, has been able to reproduce the related experimental morphologies, and explain the effect of different growth mechanisms and defects.<sup>3,52,73</sup>

Homogeneous urea nucleation has been simulated from both aqueous and organic solutions as well as a urea melt, using well-tempered metadynamics with enhanced sampling.<sup>47-49</sup> It has been found that fluctuations in local solution concentrations lead to the formation of transient and disordered clusters of urea molecules. These clusters act as precursors in which it is possible for stable crystal nuclei to form. These nuclei had an initial crystal form that has not been observed experimentally, and only once the nuclei had grown to a certain size did they spontaneously transform into the known crystal form I.<sup>47-49</sup> This alternative crystal form for small nuclei was also observed in seeded dissolution studies, which started with the experimental form I but after partially dissolving, the nuclei changed form during the simulation due to its small size.<sup>46</sup> Increases and decreases of the nucleus size, caused by growth and dissolution, leads to spontaneous transitions between the two forms at a critical nucleus size. These nuclei are too small to exhibit the experimental needle-like shapes but do provide insight into the nucleation mechanisms and early growth behaviours, in a way which is currently inaccessible by experimental observation.

## 2.3 Research Challenges, Hypothesis and Aims

Heterogeneous nucleation is very common and can occur in a supersaturated solution in contact with any type of material. Solid interfaces are important since they are an inherent part of any reaction setup and fouling is a common problem. In practice, it is relatively simple to introduce a solid with specific surface properties to achieve a

desired nucleation effect. Liquid interfaces are frequently used in small scale experiments and can have a significant effect on nucleation processes but will require thorough separation processes to ensure high product purity. Gas interfaces have some effect on nucleation, but it is unclear whether this is directly related to surface interaction, or other phenomena facilitated by the presence of the gas phase. The ability to manipulate the influence of gases on nucleation processes is advantageous since gases can easily be introduced to a solution without being present in the final products.

Academic research, industrial process development and scale-up of crystallisation are all very system dependent and based around extensive empirical experimentation.<sup>21</sup> Extensive research efforts have identified many types of surfaces that influence heterogeneous nucleation and has illustrated how specific properties can be manipulated to achieve desired crystallisation outcomes, but the understanding of the mechanisms behind the behaviour is still limited. High-throughput experimental methods are needed to produce good nucleation kinetics data, however, current methods are not suited to studying selected surfaces in isolation, and only allow for macroscopic observations. Molecular level insights can be obtained using simulations, although heterogeneous nucleation is still unexplored, these methods rely on carefully selected and validated models to provide physically relevant results.

The current research challenge is to understand the interactions between solutions and surface materials and how this affects nucleation mechanisms to enable heterogeneous nucleation to be applied in a more predictive and efficient manner. This work is based on the hypothesis that surface-solution dispersion interactions lead to changes in the solution structure and concentration near an interface.<sup>26,51</sup> The aims of this work are to:

- investigate how different interfaces impact urea nucleation kinetics from quiescent aqueous solutions
- combine experiments and molecular simulations to improve the understanding of heterogeneous nucleation mechanisms

- investigate high-throughput methods for heterogeneous nucleation experiments
- investigate force fields for modelling crystallisation properties

## Chapter 3

# Methodology

This chapter discusses the underlying theory of the research methods and approaches used. Firstly, molecular dynamics are discussed, considering the underlying algorithms, important considerations, typical simulation and analysis procedures, and limitations. Then there is a thorough discussion of the force fields models, which are used in molecular dynamics simulations, covering both how they work and some of the models which are available. Finally, the setup and equipment used for the nucleation experiments are discussed, and an overview of nucleation rate models is given. The detailed methodology varies for each of the experimental and simulation studies, therefore, the specific details are given in the relevant chapters.

### 3.1 Molecular Dynamics

Molecular dynamics (MD) models the behaviour of a system of atoms or molecules, to gain information about the movement and interaction of these particles. MD simulations are based on a series of iterative calculations, which can be used to obtain equilibrium structural data and dynamic properties.

### 3.1.1 Algorithm

MD involves calculating the forces on and solving Newton's equations of motion for all the particles in the system, throughout the duration of the simulation. MD systems are prepared by defining the simulation space (typically a 3D box), identifying the required particles for the simulation and assigning coordinate positions to these particles within this space. At the start of a simulation, all particles are given initial velocities, typically a distribution of velocities is used which is related to the system temperature.

The particles interact with each other through bonded forces (within molecules), and van der Waals dispersion and electrostatic forces between particles, these all contribute to the potential energy of the system. These interaction forces are particle dependent and are defined by the force field, which is discussed further in Section 3.2.

The kinetic energy of the full system is related to the number of degrees of freedom (DOF) and temperature according to  $E_{\text{kinetic}} = \frac{1}{2}k_B T \times \text{DOF}$ , where particles have thermal, translational and rotational degrees of freedom. On an atomic level the kinetic energy is calculated according to  $E_{\text{kinetic}} = \frac{1}{2}mv^2$ , using the mass and velocity of each particle. The particle velocities are adjusted throughout the simulation as the particles move around and interact with each other.

MD simulations are based on iterative simulations. Starting with the initial positions and velocities Newton's equations of motions are solved for all the particles, using a numerical integration algorithm. The potential and kinetic energy is calculated for each particle along with any other desired properties. The particles are moved and the velocities are redetermined. All these calculations are repeated for each time step, until the required simulation time has been achieved. The integration algorithms used are described in Section 3.1.1.

The time step is the frequency that the iterative calculations are performed at. The numerical integration methods require short time steps to be used, this ensures that the particles do not move too far between each integration step, which would make the simulation unstable. However, if the time steps are too short then the required number

of calculations, and thus the computational power and time, will become infeasibly long. Typically time steps are on the order of femtoseconds (fs). Simulations continue until the desired number of time steps have been performed. For example, if a 1 fs time step is used, calculations will be performed every 1 fs (simulation time) and a 1 picosecond (ps) simulation will involve 1000 time steps.

The coordinate positions of all the particles is tracked throughout the simulation and is saved in a trajectory file, this typically leads to large file sizes, therefore this is usually only recorded once for every couple of thousand time steps.

### Integration Algorithms

Several algorithms are available for integrating Newton's equations of motion and proceeding the simulation, some of the most commonly used are the Verlet and Leapfrog methods.<sup>74,75</sup>

The Verlet algorithm is derived from a Taylor expansion around position  $r(t)$  as follows:<sup>75</sup>

$$\begin{aligned}
 r(t \pm \delta t) &= \sum_{n=0}^{\infty} \left( \frac{1}{n!} r^{(n)}(t) ((t \pm \delta t) - t)^n \right) \\
 r(t \pm \delta t) &= r(t) + r'(t)((t \pm \delta t) - t) + \frac{1}{2!} r''(t)((t \pm \delta t) - t)^2 + \dots \\
 r(t \pm \delta t) &= r(t) \pm r'(t)\delta t + \frac{1}{2} r''(t)\delta t^2 \pm \dots \\
 r(t \pm \delta t) &= r(t) \pm v(t)\delta t + \frac{1}{2} a(t)\delta t^2
 \end{aligned} \tag{3.1}$$

Adding together the two resultant equation gives the final form which is independent of velocity

$$\begin{aligned}
 r(t + \delta t) + r(t - \delta t) &= r(t) + v(t)\delta t + \frac{1}{2} a(t)\delta t^2 + r(t) - v(t)\delta t + \frac{1}{2} a(t)\delta t^2 \\
 r(t + \delta t) &= 2r(t) - r(t - \delta t) + a(t)\delta t^2
 \end{aligned} \tag{3.2}$$

Here  $r$  is the  $xyz$  position vector,  $v$  is velocity,  $a$  is acceleration,  $t$  is time and  $\delta t$  is the

time step.

The Verlet algorithm does not involve calculating the velocity of the particles. The velocity can be calculated separately, but with a greater error, when it is needed to calculate the kinetic energy.

The Velocity-Verlet algorithm is an alternative form of the Verlet algorithm, which involves the calculation of the velocity. Despite the additional velocity calculation, the computational and memory requirements are similar since one less particle position is required. The Velocity-Verlet method calculates the velocity at two half-steps, using the following algorithm:<sup>75,76</sup>

Firstly, the Velocity-Verlet algorithm calculates the velocity after a half-step at time  $t + \frac{\delta t}{2}$ :

$$v(t + \frac{1}{2}\delta t) = v(t) + \frac{1}{2}a(t)\delta t \quad (3.3)$$

This velocity is used to determine the particle position at time  $t + \delta t$ :

$$r(t + \delta t) = r(t) + v(t + \frac{1}{2}\delta t)\delta t \quad (3.4)$$

Then a force evaluation is required to determine the acceleration at time  $t + \delta t$ , which allows the velocity at time  $t + \delta t$  to be calculated:

$$v(t + \delta t) = v(t + \frac{1}{2}\delta t) + \frac{1}{2}a(t + \delta t)\delta t \quad (3.5)$$

Newton's equations of motion are time reversible, meaning that if the movement of the particles is reversed at any point in time, then the resulting trajectory will match the original trajectory. Non-reversible calculations can lead to significant problem with long-term drifts in energy. The Verlet and Velocity-Verlet algorithms are both time reversible, but the implementation is affected by the accuracy of the computer.<sup>75</sup> The Velocity-Verlet algorithm is used for all the simulations performed in this work,<sup>77</sup> this algorithm is the same as the 'kick-drift-kick' form of the Leapfrog method.<sup>78</sup>

### 3.1.2 Ensembles

MD simulations are performed for very small systems, therefore, the system boundaries and any interactions with surroundings are very significant. The system can be open, closed or partially open/closed. The energy ( $E$ ) and number of moles ( $N$ ) are constant in a closed system, where neither energy or matter is exchanged with the surroundings. The temperature ( $T$ ) and chemical potential ( $\mu$ ) are constant in an open system, where both matter and energy is exchanged with the surroundings. Simulations are most commonly performed in a partially open system with constant number of moles and temperature. In addition to these variables either the volume ( $V$ ) or pressure ( $P$ ) is kept constant during simulations.

The combination of these factors that are kept constant are known as the system ensemble. The choice of ensemble is dependent on the simulation requirements, however, the constant NVT and NPT ensembles, are most commonly used, since they are comparable to experimental conditions. Constant NVT conditions, known as the canonical ensemble, allow vacuum (or gas) spaces to be simulated. Constant NPT conditions, known as the isothermal-isobaric ensemble, often provide more realistic conditions and are useful to obtain equilibrium properties such as density. The number of moles in and the volume of the system are very easy to keep constant. The temperature and pressure are controlled by algorithms known as thermostats and barostats respectively, and similarly other integration algorithms can be used to maintain a constant energy. The chemical potential is more difficult to keep constant since this requires the number of molecules to change during the simulation, which requires enhanced simulation techniques.

### 3.1.3 Thermostats and Barostats

Thermostats are used to control the temperature in a simulated system, which is generally achieved by exchanging heat energy with an external heat bath through adjusting the velocities of the particles. There is a wide range of thermostats available, with

varying degrees of simplicity and accuracy. The Velocity Rescaling and the Berendsen thermostats<sup>79</sup> are simple and efficient but do not generate or sample the canonical ensemble correctly so are not suitable for comparing to experimental conditions, but can be used to speed up equilibration. The Andersen thermostat<sup>80</sup> uses random collisions between simulated particles and a heat bath, and thus samples the canonical ensemble correctly, but the dynamics of the system are unphysical since the particle velocities decorrelate quickly. The Nosé-Hoover thermostat<sup>81,82</sup> is more complex since it interacts with the heat bath by adding a friction term to the equations of motion, which correctly generates the canonical ensemble. The Nosé-Hoover thermostat, requires a carefully selected relaxation time to ensure that it is both stable and efficient. The Nosé-Hoover thermostat has been used for all the simulations in this work.

Barostats are used to control the pressure in a simulated system, which typically involves adjusting the volume, through rescaling the distances between the particles and as such changing the system size. The change in volume can be isotropic or anisotropic. Isotropic means that the relative size change is the same in all dimensions. Anisotropic means the size changes independently in all dimensions, so the overall shape of the system may change, which is useful when simulating solid structures. There is a range of barostats available, with varying degrees of simplicity and accuracy. The Berendsen barostat<sup>79</sup> is a simple algorithm that weakly couples the system to a pressure bath, similarly to the Berendsen thermostat it does not correctly sample the physical ensemble, but can be used to speed up equilibration. The Andersen barostat<sup>80</sup> adds an additional degree of freedom to the equations of motion, this is also coupled to a pressure bath and correctly samples the physical ensemble but only works isotropically. The Nosé-Hoover<sup>82,83</sup> and Parrinello-Rahman<sup>84</sup> are similar barostats, which are developed from the Andersen barostat to work anisotropically. An anisotropic barostat is desirable when simulating crystal structures, therefore, the Nosé-Hoover barostat was selected (for the simulations requiring pressure control) since this was the only anisotropic barostat available in LAMMPS.

The use of thermostats and barostats does not lead to constant temperature and pres-

sure values, but when the algorithms are implemented correctly, there will only be small fluctuations around the desired values with time.

### 3.1.4 Periodic Boundary Conditions

In simulated systems, particles will collide with a fixed boundary, at the edge of the simulation box, and bounce back towards the bulk, similarly to at vessel walls. This behaviour occurs much more frequently in simulations, due to the significantly smaller system size, compared to experiments, and the effects of these interactions are exaggerated in a manner which is not representative of real solutions. For example, a small molecule in the centre of a 10 nm long simulation box, may pass only 10 molecules before interacting with the edge, whereas in a 10 mm long experimental sample this distance would be about 10 million molecules.

The influence of these surfaces on the behaviour of the system can be minimised by applying periodic boundary conditions (PBC). PBCs do not impede the movement of particles at a boundary, instead the particle passes unhindered into an image of the simulation box. In effect this means that a particle which is moving towards and leaving at the top-right corner of the box will re-enter in the bottom-left corner of the box and keep moving in the same direction. Although the coordinate position of the particle will be even further up and to the right. PBCs gives the effect of the system being infinite and can be applied in all three dimensions or only in some. To prevent particles from interacting with themselves, PBCs require that the minimum system size (in the dimensions that PBCs are being applied) is at least twice the length of interaction cutoff distances.

### 3.1.5 Setup and Equilibration

MD simulations involve four main stages: system setup, energy minimisation, equilibration run, and production run. System setup involves obtaining the required force field parameters and starting molecular structures and formatting these as required by

the selected MD software. There are various tools available to assist with system setup for various force fields and simulation software.

Typically simulations start with an energy minimisation, which is important to allow the simulation to run smoothly. Depending on how the starting positions for the particles are generated, it is possible to end up with particles which are too close to each other or even with overlapping molecules. This unphysical positioning is associated with very high energies and can lead to simulations crashing instantly. These configurations can typically occur if molecules are inserted randomly, such as in solutions but should not occur with crystal structures which are based on experimental positions. Minimisation allows these misplaced particles or molecules to move into physically realistic positions, which will allow the simulation to run successfully. Minimisation is also important to allow crystal structures to relax, since the exact optimal structure can vary between force fields. Different minimisation algorithms are available.

The equilibration and production simulations are performed at the same conditions, using the same or different ensembles depending on the requirements. The equilibration run is required to ensure that the system has reached a stable state from the initial positions and energy minimisation. The establishment of equilibrium can be determined by monitoring the system temperature, density and other properties of interest, which should be stable without any significant fluctuations or drift. The time required to reach equilibrium is very system dependent. Once equilibrium has been reached, the production run allows the desired properties to be sampled over a longer period of time.

### **3.1.6 Analysis**

A large amount of data is generated during the simulations, which includes trajectory data (the coordinates of all the particles with time), and calculated global and local properties. The particular thermodynamic properties of interest might include the system temperature, pressure, energy (both constituent parts and the combined value), density and system size.

There are many ways to analyse the data generated from MD simulations. Typical approaches involve monitoring bulk properties such as density and interaction energies, this can both involve checking for fluctuations or drifts with time, and calculating time averaged values to compare with other work. More enhanced analyses include monitoring the packing of specific particles or their movement with time, examples of these are discussed below. All the numerical analysis in this work was performed using Python programming. Visualisation of the trajectory is a very powerful form of analysis, this is an easy way to gain familiarity with the data. Snapshots of the whole trajectory can be viewed in sequence or it is possible to focus on specific regions or molecules. This can be performed with specific software such as Visual Molecular Dynamics<sup>85</sup> (VMD) or Open Visualisation Tool<sup>86</sup> (OVITO).

Radial distribution functions (RDF) provide information about the packing of particles in a system. RDFs are a profile of the likelihood of finding any particles at increasing distances from a reference particle, as shown in Figure 3.1. RDFs can be calculated for particles of the same type or two different types. Analysis of the RDFs can help to provide information about the bonds and interactions present in the system. RDFs are experimentally measurable and easy to calculate for simulated systems, which allows comparison between the two methods.

The mean square displacement (MSD,  $M(t)$ ) is a measure of the displacement of a particle from its original location, according to:

$$M(t) = (x(t) - x(t_0))^2 + (y(t) - y(t_0))^2 + (z(t) - z(t_0))^2 \quad (3.6)$$

This uses the  $x$ ,  $y$ , and  $z$  positions at time  $t$  and compares these to the original positions at time  $t_0$ . The MSD of a particle is related to the diffusion coefficient,  $D$ , by the Einstein equation:<sup>87</sup>

$$M(t) = 6Dt \quad (3.7)$$

The diffusion coefficient is another useful property, which allows for comparison between experiments and simulations, and influences the dynamics of the simulated system.

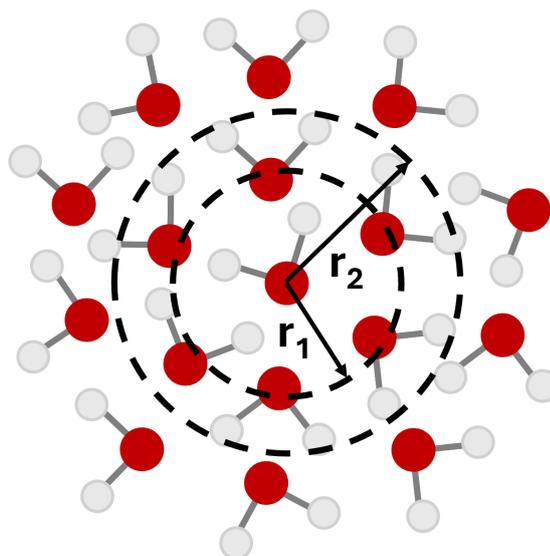


Figure 3.1: The probability of finding a red atom is significantly higher at a distance of  $r_1$  from the central atom compared to the probability at  $r_2$ .

### 3.1.7 Limitations of Molecular Dynamics

MD can be a very powerful tool, however it is important to understand the workings and limitations of the method to ensure that it is used appropriately and meaningfully.

Many calculations are performed every time step to evaluate the forces between all the particles, determine the desired properties, control the conditions, and perform time step integration to proceed the simulation. This is very computationally demanding, so only small system sizes and short simulation times can be obtained. Typically MD simulations are on the order of nanometres, involving hundreds or thousands of molecules, with the simulated time being on the order of nanoseconds. Small system sizes can lead to problems with finite-size effects where the sampled properties are not representative of real bulk properties, this needs to be considered when planning and analysing simulations.

The performance of MD simulations is very dependent on the force field model used to accurately reproduce the desired physical behaviours, Section 3.2 gives a detailed discussion of force fields. Most MD models are not polarisable, specific polarisable force fields have been developed but these are limited and often only applicable to a small

number of molecules.<sup>88–90</sup> MD simulations were not developed for studying chemical reactions, however this is possible through the use of specially designed reactive force fields,<sup>91,92</sup> and other new methods.<sup>93</sup>

### 3.1.8 Molecular Dynamics Software

There are multiple softwares available for performing MD simulations including: Large-scale Atomic/Molecular Massively Parallel Simulator<sup>77</sup> (LAMMPS); Groningen Machine for Chemical Simulations<sup>94</sup> (GROMACS); Assisted Model Building with Energy Refinement<sup>95</sup> (AMBER); and Nanoscale Molecular Dynamics<sup>96</sup> (NAMD). LAMMPS is an open-source software in which any force field can be specified, it is suitable for modelling a large range of materials including metals, inorganics, polymers, and biomolecules. GROMACS is an open-source software which contains many pre-prepared force fields and allows others to be specified, it is intended for modelling biomolecular systems but it is commonly used for other applications. The AMBER software requires a licence, it is intended for modelling biomolecular systems and it uses a family of in-built AMBER force fields. The NAMD software is free for non-commercial use, a range of force fields are supported, and it is intended for modelling large biomolecular systems. Both LAMMPS and GROMACS are used for simulating crystallisation systems. GROMACS is advantageous since it is generally very fast and efficient in running simulations compared to other software and it may be more user friendly for new users. However, LAMMPS was chosen for this work, due to its advantages of being highly flexible and its applicability to a larger range of systems.

## 3.2 Force Fields

Classical molecular simulations such as Monte-Carlo and molecular dynamics (MD) as well as metadynamics simulations are reliant on the use of force fields to emulate the material properties of the systems and phases investigated. Force fields are models for describing the interaction energy and behaviours of particles in a simulated system

using a selection of equations and parameters. Both intramolecular and intermolecular potential energies can be modelled, these are related to the movements within individual molecules, and the interactions and attractions between nearby atoms, respectively. The force field parameters vary depending on the types of atoms involved and different forms of equations can even be used too. There are many force fields available where the parameters have been carefully selected to produce certain properties in the system. Therefore, the force fields used need to be carefully selected and tested to ensure that they work well for the intended applications.

Intramolecular interactions occur within molecules, since the atoms are not completely fixed in a certain position, but can move a little within the molecular structure. The bond lengths and angles can stretch and contract, which is described by the bond and the angle interaction potentials respectively. The molecules can also twist and turn around the bonds, which is described by the dihedral and the improper dihedral interactions. The degree to which this movement occurs varies between molecules, and is defined by the force field.

In MD, two types of intermolecular interactions are considered relating to the electrostatic and dispersion forces. Partial charges are given to all the atoms, with charges assigned based on the atom type, these charges form the basis of the calculated electrostatic interaction energies. Some atoms are positively charged whilst others are negatively charged, which should add up across a molecule to form a neutral molecule. The electrostatic interaction energy between two atoms increases with the difference in partial charges and proximity, according to the Coulomb potential. The weak, short-distance dispersion interactions between atoms are typically modelled with a Lennard-Jones potential, which is parameterised by the maximum interaction energy and associated distance for each atom. The electrostatic and dispersion interactions exist between atoms in the same molecule and atoms in different molecules. However, the interactions are typically zeroed for atoms separated by one or two bonds and reduced for atoms separated by three bonds. These intermolecular interactions are often only calculated up to a certain distance around the atoms to reduce computational requirements, how-

ever, additional long range corrections are typically applied. Again, the parameters and equations required to implement this in MD is specified by the force field.

A wide variety of force fields are available in the literature ranging from commonly used general force fields that can be applied to many different molecules to niche force fields optimising a few parameters for one specific molecule. These force fields can also be combined in various ways; for aqueous solutions it is particularly common to use different force fields for solute and solvent (water) molecules, but there are also examples of intermolecular parameters from one force field being paired with the intramolecular parameters of another. Regardless, the choice of the force field is important and needs to be validated for the intended application. For studies of crystallisation processes, it is important for the force field to be able to reproduce both crystal and solution behaviours well. We are not aware of any standardised procedure for validating force field performance using both crystal and solution properties. Therefore, in this work we propose a set of simple tests that can be used for crystal and solution force field validation and we apply them to several different force fields using urea as a model system.

Some of the most common general force fields for modelling organic molecules are: Optimised Potentials for Liquid Simulations (OPLS) developed for modelling liquids and aqueous solutions of organic molecules;<sup>97</sup> Assisted Model Building with Energy Refinement (AMBER) developed for modelling proteins and nucleic acids;<sup>98</sup> Generalised AMBER Force Field (GAFF) developed for modelling small organic molecules and to be compatible with AMBER;<sup>99</sup> Chemistry at Harvard Molecular Mechanics (CHARMM) developed for modelling proteins, nucleic acids and lipids;<sup>100</sup> and Groningen Molecular Simulation (GROMOS) developed for modelling biomolecular systems such as proteins and nucleotides.<sup>101</sup> Although several of these are intended for larger molecules it is relatively easy to apply the parameters to smaller organic molecules. Several versions are available for all of these force fields, some of which differ significantly from previous versions.

### 3.2.1 Force Fields Available for Urea

A large number of force fields are available for urea, but only a few of these have been extensively validated. An overview of the force fields that have been tested for urea, in one way or another, is given in Table 3.1. Out of the force fields that have been used to model urea crystals and solutions only OPLS and GAFF have been widely used. Only the solution phase has been tested for OPLS, and for GAFF only crystal phase tests have been performed, however subsequent studies involve both the crystal and solution phases, which implies that both the OPLS and GAFF force fields can reliably reproduce both phases to some extent.

The OPLS-GROMOS force field was used in the earlier simulations of urea crystal growth and dissolution.<sup>44</sup> However, the GAFF force field has been favoured in more recent studies, due to the broad range of other molecules that can also be modelled with GAFF.<sup>3</sup> The GAFF force field has been used to study the effects of additives<sup>3</sup> and solvents<sup>49,52</sup> on urea crystallisation and to simulate homogeneous nucleation using well-tempered metadynamics with enhanced sampling.<sup>47-49</sup> The dissolution of small nuclei-like crystals has been studied using both the GAFF<sup>46</sup> and urea optimised GAFF<sup>45</sup> force fields.

The original GAFF force field,<sup>99,116</sup> denoted here as GAFF1, was developed for use with most organic and pharmaceutical molecules. There have been some updates to the GAFF1 parameters and a second generation of the force field, GAFF2, has been developed. GAFF2 includes both updated bonded and non-bonded parameters compared to GAFF1.<sup>99,116,117</sup> There are no charges directly associated with GAFF, and these need to be calculated on a molecule by molecule basis. The Antechamber tool,<sup>116</sup> used to obtain the force field parameters, includes a default option for calculating charges based on the AM1-BCC charge model, which does not require any further inputs. These two force fields are referred to as GAFF1 (version 1.81, AM1-BCC charges<sup>99,116</sup>) and GAFF2 (version 2.11, AM1-BCC charges<sup>99,116</sup>) in the rest of this work.

An alternative version of GAFF was specifically developed for urea.<sup>111</sup> This version geometrically optimised for the bonded potential parameters, but did not alter the non-

Table 3.1: Overview of the force fields that have been tested for urea.

Urea model	Water model	Application	Tested properties
OPLS* <sup>102</sup> (urea specific)	TIP4P	Solution	Absolute free energy of hydration, solution structural correlations <sup>102</sup>
OPLS* <sup>102</sup> (urea specific)	TIP3P	Solution	Density, diffusion coefficients <sup>103</sup>
CHARMM <sup>100</sup>	TIP3P	Dimers and solution	Diffusion coefficients, solution structural correlations <sup>104</sup>
CHARMM <sup>100</sup>	TIP3P	Crystal and solution	Solution structural correlations, diffusion and solvation free energy; bulk crystal density and enthalpy of sublimation; solubility <sup>105</sup>
OPLS* <sup>102</sup> +CHARMM <sup>106</sup>	SPC/E	Solution and cosolvent	Density, solution structural correlations, diffusion and dielectric properties <sup>107</sup>
OPLS* <sup>102</sup> +GROMOS <sup>108,109</sup>	SPC	Solution	Density, energy of solution, heat of solvation, free enthalpy of desolvation and urea diffusion <sup>110</sup>
GAFF <sup>99</sup>	TIP3P	Crystal and solution	Bulk crystal lattice parameters and melting point temperature (from a solid liquid interface) <sup>3</sup>
GAFF <sup>111</sup> (urea optimised)	N/A	Dimers and crystal	Cohesive energy, sublimation and melting point temperatures <sup>112</sup>
KBFF <sup>113</sup> (Kirkwood-Buff, urea specific)	SPC/E, SPC, TIP3P	Solution	Solution structural correlations, partial molar volumes, isothermal compressibility, activity derivatives and coefficients, density, relative permittivity, diffusion constant and crystal lattice parameters <sup>113</sup>
KBFF <sup>113</sup> (urea specific)	TIP3P	Solution	Density, diffusion coefficients <sup>103</sup>
SAPT-FF <sup>105</sup> (polarised)	SWM4-NDP (polarised)	Crystal and solution	Solution structural correlations, diffusion and solvation free energy; bulk crystal density and enthalpy of sublimation; solubility <sup>105</sup>
COMPASS <sup>114</sup> (polarised)	N/A	Crystal	Crystal lattice parameters <sup>114</sup>
AMOEBA <sup>88,90</sup> (polarised)	N/A	Crystal	Crystal lattice parameters <sup>115</sup>

\* The urea specific OPLS<sup>102</sup> force field only has intermolecular parameters, therefore, simulations have either been done without intramolecular interactions or by taking these parameters from the secondary force field listed after the + sign.

bonded Lennard-Jones parameters from version 1. Here, the RESP charge model was used and seven sets of charges were calculated for different orientations of urea dimers (D1-D7) and from these D1 and D3 were chosen for this work, since D1 was based on the crystal structure and D3 was recommended as the most suitable overall. These two force fields are referred to as GAFF-D1 (optimised GAFF1, RESP-D1 charges<sup>111</sup>) and GAFF-D3 (optimised GAFF1, RESP-D3 charges<sup>111</sup>).

The OPLS force field was developed as a series of intermolecular parameters for different types of organic molecules. There were no associated intramolecular parameters, the molecules were simply kept rigid throughout the simulations, with the structure being based on experimental parameters. The original OPLS force fields were not all-atom force fields but included united-atom terms for carbon atoms where all hydrogen atoms bonded to carbon atoms were implicitly included in the carbon atom parametrisation. Versions were developed for liquid hydrocarbons,<sup>118</sup> peptides and amides (OPLS-Amide),<sup>119,120</sup> liquid alcohols,<sup>121</sup> proteins,<sup>122,123</sup> and nucleotide bases.<sup>124</sup> Later, the general all-atom OPLS force field (OPLS-AA) was developed for both liquid and solid simulations<sup>97,125</sup> and we have chosen to test this as it is widely used. OPLS-AA consists of the bond and angle parameters from AMBER,<sup>98,126</sup> newly calculated dihedral and improper parameters<sup>97</sup> and OPLS intermolecular parameters. The parameters for OPLS-AA were obtained from tables in publications by Jorgensen et al.<sup>97</sup> and Weiner et al.<sup>98</sup> The LigParGen software has been created, by the developers of OPLS, to more easily obtain the OPLS force field parameters from an input structure file.<sup>127-129</sup> However, when tested LigParGen produced different parameters compared to OPLS-AA, and these are also tested and denoted as OPLS-AA-N. OPLS-AA and OPLS-AA-N differ by the charges, Lennard-Jones parameters of the carbon atom, and OPLS-AA-N has one additional angle parameter. Prior to the parameterisation of OPLS-AA a urea-specific version was developed<sup>102</sup> (OPLS-Urea), based on OPLS-Amide<sup>119,120</sup>, versions of this continue to be used. We also tested the OPLS-AA as above but with the intermolecular parameters of OPLS-Urea, called OPLS-AA-D.

OPLS-Urea has only intermolecular parameters, this has been used as is for some solu-

tion simulations<sup>103,130</sup> and has been combined with the intramolecular parameters from other force fields, including CHARMM22<sup>107</sup> and GROMOS96<sup>110</sup> both of which were only used for solutions. The combination with GROMOS96, which was implemented with rigid bond lengths, has been extensively validated by Smith et al.<sup>110</sup>, with comparisons of density, enthalpy of mixing, free enthalpy of urea hydration and urea diffusivity properties to experimental data. This led to the subsequent use of this force field by Piana et al.<sup>44,73</sup> in their work on crystal growth and dissolution. Therefore, we also test this force field, referring to it as OPLS-S, but implement it without rigid bonds. We note that there is some discrepancy between the dihedral parameters in the GROMOS96 source<sup>101</sup> and those cited by Smith et al.<sup>110</sup> We have not been able to access the manual<sup>108</sup> with the parameters used by Smith et al.<sup>110</sup>, however, the GROMOS 53A5 and 53A6 parameter set<sup>101</sup> was identified as being the relevant parameter set as it is the first published set which contains the parameters used by Smith et al.<sup>110</sup> There appears to be some inconsistency between the GROMOS source<sup>101</sup> and the parameters used by Smith et al.<sup>110</sup> First, the O–C–N–H dihedrals are applied to only two out of the four instances of these dihedrals, without any explanation of this choice. In addition, the parameters chosen are taken from the X–C–C–X (6-ring) example and not the X–C–N–X example. Therefore, we also tested a version that uses the original GROMOS96 dihedrals (OPLS-G).

A summary of the urea force fields investigated in this work is given in Table 3.2. The associated force field equations and parameters are given in Appendix A.1.

### 3.2.2 Force Fields Available for Water

A variety of water force fields have been used in combination with various urea force fields including SPC/E,<sup>45,46</sup> SPC,<sup>44,45,73,110,131</sup> TIP3P,<sup>3,43,45,48,52,102</sup> and TIP4P/2005.<sup>45</sup> GAFF was developed with TIP3P water and OPLS with TIP4P water, however, both are compatible with and have been successfully used with most of these water force fields. While all models reproduce the density of pure water well, SPC/E is best for reproducing bulk dynamics and structures, including self-diffusion coefficients, followed

Table 3.2: Nine selected force fields, summarising the source of bonded, Lennard-Jones and electrostatic parameters.

Force field	Bonded	Lennard-Jones	Electrostatic
GAFF1	GAFF1 <sup>99,116</sup>	GAFF1 <sup>99,116</sup>	Calculated AM1-BCC <sup>116</sup>
GAFF2	GAFF2 <sup>99,116</sup>	GAFF2 <sup>99,116</sup>	Calculated AM1-BCC <sup>116</sup>
GAFF-D1	Optimised GAFF1 <sup>111</sup>	GAFF1 <sup>99,111</sup>	Optimised RESP <sup>111</sup>
GAFF-D3	Optimised GAFF1 <sup>111</sup>	GAFF1 <sup>99,111</sup>	Optimised RESP <sup>111</sup>
OPLS-AA	AMBER <sup>98,126</sup> & OPLS-AA <sup>97</sup>	OPLS-AA <sup>97</sup>	OPLS-AA <sup>97</sup>
OPLS-AA-N	OPLS-AA-N <sup>97,127</sup>	OPLS-AA-N <sup>97,127</sup>	OPLS-AA-N <sup>97,127</sup>
OPLS-AA-D	OPLS-AA-N <sup>97,127</sup>	OPLS-Urea <sup>102</sup>	OPLS-Urea <sup>102</sup>
OPLS-S	GROMOS based <sup>101,110</sup>	OPLS-Urea <sup>102,110</sup>	OPLS-Urea <sup>102,110</sup>
OPLS-G	GROMOS <sup>101</sup>	OPLS-Urea <sup>102,110</sup>	OPLS-Urea <sup>102,110</sup>

by TIP4P, SPC and TIP3P.<sup>132–134</sup> The SPC/E model<sup>133,135</sup> was chosen for this work due to its ability to reproduce pure water properties well and its good performance in previous works with other small organic molecules (modelled with several different force fields including GAFF and OPLS).<sup>134,136</sup>

### 3.3 Nucleation Experiments and Measurements

Nucleation is a time dependent process, therefore, experiments are focused on determining the precise time of crystallisation in a large number of samples. Nucleation experiments involve preparing samples from an undersaturated solution and then changing the conditions to supersaturate the samples. The nucleation time is the time at which nucleation occurs (crystals are detected) with reference to the time that the desired supersaturation is reached (for example the beginning of the isothermal phase for cooling crystallisation). Temperature control is very important during nucleation experiments, since solubility and supersaturation often varies significantly with temperature. Crystals can be detected visually by analysing images or changes in light transmissivity, for some systems the heat of crystallisation leads to detectable temperature changes which can also be used. This work uses different experimental setups with different detection methods, firstly, a custom setup with webcams was used, which was followed by the commercial Crystal16 equipment.

### 3.3.1 Custom Nucleation Setup

In our custom setup the Polar Bear equipment was firstly used to heat the prepared vials to ensure complete dissolution, and then to cool the samples to the crystallisation temperature. The Polar Bear Plus Crystal is a programmable, precision heating and cooling platform produced by Cambridge Reactor Design. The Polar Bear uses interchangeable plate attachments to allow for accurate temperature control ( $-40-150 \pm 0.1$  °C) for a range of vessels from vials to round bottom flasks. Up to 80 1.5 ml vials can be prepared at the same time.

The prepared samples were manually transferred from the Polar Bear to racks in a temperature controlled incubator. For this project, vial racks were specially designed and 3D printed to optimise the arrangement of vials within the incubator. The samples were monitored by webcams which are placed within the incubator and take pictures at regular intervals. To optimise the image quality the incubator was blacked-out, with a small window to let in light from an external lamp.

The series of pictures of the vials were analysed to determine the nucleation time, when the crystals first appeared. The accuracy of this detection method depends on the quality of the images and the capabilities of the analysis tools. An example of the images obtained in this work is shown in Figure 3.2. The fast growth and secondary nucleation of urea mean that when nucleation occurs the vials change from being clear and uncrystallised to being filled with crystals in the time interval between two images being taken. Therefore, in this instance it was very easy to accurately determine the nucleation time, without needing any advanced analysis techniques.

### 3.3.2 Crystal16 Equipment

The Crystal16 is a programmable multi-vessel tool from Technobis Crystallization Systems. It has capacity for 16 vials of 1.5 ml volume and accurate control of temperature ( $-15-150 \pm 0.1$  °C), with variable heating/cooling rates ( $0-20$  °C  $\text{min}^{-1}$ ), and optional stirring ( $0-1250$  rpm). The 16 vials are split between four zones (A–D), which are con-



Figure 3.2: Example webcam image from this work. Crystallisation has taken place in nine vials and PTFE stirrer bars are seen at the bottom of most vials.

trolled separately, and the vials are numbered accordingly (A1–D4). The equipment measures the transmissivity of the solution in each vial, every second, using a beam of light, part way up the vial, which passes through the solution to a detector. A clear solution without any crystals will have 100% transmissivity, this decreases as crystals nucleate and grow, and transmissivity can fall to 0% in systems where crystals are sufficiently numerous. The Crystal16 is commonly used to determine the solubility and metastable zone width of the system as well as nucleation induction times.

The repeated temperature cycling possible with automated equipments such as the Crystal16 allows for greatly enhanced data collection with minimal setup time and material requirements. However, some potential problems related to performing many cycles are solvent evaporation, encrustation, and thermal history effects. Solvent evaporation may occur over time (as vial caps may not be entirely gas tight), especially if the solution is repeatedly being heated to high temperatures. This can lead to changes in composition and supersaturation, and the overall mass should be checked at the end of experiments. Encrustation or crowning is an effect where crystals form at the edge of the vial near the meniscus, this can prevent all the material from being redissolved between the cycles. Crowning can be particularly problematic if the solution level drops due to evaporation, or vials are partially filled and agitation leads to liquid splashing. The

thermal history, which is the previous exposure to different temperatures, can influence nucleation behaviours.<sup>137–140</sup> Repeated heating and cooling cycles lead to increasingly complex thermal histories, therefore, this may lead to a shift in behaviour with time.

To avoid possible problems with repeated cycling a relatively low number of cycles is often used in nucleation experiments. Five cycles appears to be the most common number of cycles across a range of crystallisation applications,<sup>141–147</sup> although this varies.<sup>148</sup> In the study of ice nucleation 12 cycles have been used regularly and even 24 cycles were tested, without there being any significant cyclical influence on the nucleation temperature.<sup>149,150</sup> Crystal16 experiments are typically agitated,<sup>141–148</sup> this ensures that crystals are moved throughout the vial and will be detected soon after nucleation, even if only a low number of crystals are formed. However, for certain systems, such as the nucleation of ice,<sup>149,150</sup> it is possible to obtain accurate nucleation times for unagitated systems.

A large range of cooling (and heating) rates are possible with the Crystal16 equipment, which is useful since the choice of cooling rate can influence the crystallisation behaviour. Fast cooling rates can lead to problems with insufficient heat transfer, where the temperature in the solution lags behind programmed temperature. Cooling rates down to  $-5\text{ °C min}^{-1}$  have been found to work reliably, whereas  $-10$  and  $-15\text{ °C min}^{-1}$  have led to unsatisfactory heat transfer.<sup>148</sup> Fast cooling rates can also enhance early crystallisation in some systems by crash cooling effects. Slow cooling rates increase the overall length of the experiments, which can also lead to crystallisation during the cooling stage for quickly nucleating systems. A large range of cooling rates have successfully been used for studying different systems ranging from  $-0.2$  to  $-5.0\text{ °C min}^{-1}$ .<sup>141–150</sup>

### 3.4 Nucleation Rate Models

Nucleation is a stochastic process, meaning that a distribution of nucleation times will be obtained for a set of identical samples. The probability ( $P(t)$ ) of a successful nucleation event occurring increases with time ( $t$ ) and tends toward 100% given sufficient

time. Therefore, nucleation experiments typically involve many identical samples, tens or even hundreds of replicas, to obtain a good distribution of nucleation times. This distribution allows the time dependent nucleation probability to be determined, which can be used to calculate the nucleation rate.

The time dependent nucleation rate ( $j(t)$ ) can be defined with the differential equation:

$$\frac{dP(t)}{dt} = j(t)(1 - P(t)) \quad (3.8)$$

If the nucleation rate is constant, then the nucleation probability is simply given by:

$$P(t) = 1 - \exp(-j \times t) \quad \text{Exponential} \quad (3.9)$$

This is a simple exponential function, which assumes that the nuclei form according to a Poisson time distribution. The time independent nucleation rate ( $j$ ) is dependent on the volume ( $V$ ) but can also be expressed independently of volume using ( $J$ ) according to:

$$\tau = \frac{1}{j} = \frac{1}{JV} \quad (3.10)$$

Here,  $\tau$  is the timescale of nucleation and is often used in nucleation models instead of  $j$  or  $J$ . The nucleation time is the time that a stable nucleus is formed, this nucleus is nanosized and needs to grow before it is large enough to be detected experimentally. This growth time varies depending on the system studied and it is also influenced by the detection method used, therefore, the detection lag time ( $t_d$ ) is important to include in the rate models. The exponential nucleation model, in Eq. 3.9 can also be expressed as:

$$P(t) = 1 - \exp\left(-\frac{t - t_d}{\tau}\right) \quad \text{Exponential} \quad (3.11)$$

The exponential model assumes that there is only one nucleation mechanism with one nucleation rate, which is the same for all the samples and that all the samples nucleate, thus the cumulative distribution function should tend to unity for long enough times. This model is suitable for describing many classically nucleating systems including a range of small organic and drug molecules.<sup>141,148,151,152</sup>

However, not all nucleation processes are well described by the exponential model, this can be due to the existence of multiple nucleation mechanisms with different rates within the same samples, or due to the nucleation rate changing with time. In addition to this, with small organic molecules it is commonly observed that cumulative distributions for induction times tend to reach a plateau value less than one.<sup>25,26,38,141,148,152–156</sup> In these cases, some samples never nucleate during an experiment, despite long experiment times and a significant proportion of the samples nucleating within a relatively short time.

Changing nucleation rates can be described by a modified exponential, known as the Weibull function:

$$P(t) = 1 - \exp\left(-\left(\frac{t-t_d}{\tau}\right)^\beta\right) \quad \text{Weibull} \quad (3.12)$$

Here, the exponent  $\beta$  changes relative to the nucleation rate. A stretched exponential function, with  $\beta < 1$ , describes a nucleation rate which is decreasing with time. A compressed exponential function, with  $\beta > 1$ , describes a nucleation rate which is increasing with time. Decreasing nucleation rates have been observed for different kinds of systems, including several drug molecules,<sup>152–154</sup> this includes both systems where full nucleation is achieved and those where  $P(t)$  plateaus below unity. Increasing nucleation rates only occurs very rarely,<sup>152</sup> but have been observed for glycine nucleation in a microfluidic setup,<sup>157</sup> however, nucleation in one droplet induced nucleation in neighbouring droplets, which makes the making the results less clear.

Alternatively, the presence of multiple, but constant nucleation rates, can be modelled by combining several exponential functions together:

$$P(t) = A \left[1 - \exp\left(-\frac{t-t_d}{\tau_1}\right)\right] + (1-A) \left[1 - \exp\left(-\frac{t-t_d}{\tau_2}\right)\right] \quad (3.13)$$

Biexponential

Here,  $A$  and  $(1-A)$  are the fraction of samples in nucleation regimes 1 and 2, respectively, and  $\tau_1$  and  $\tau_2$  are the timescales of the two nucleation regimes. This model may

be applicable where multiple mechanisms are competing such as concurrent heterogeneous and homogeneous nucleation, or to a two-step nucleation process. This has been proven to be applicable to a number of experiments with small organic molecules for both heterogeneous nucleation,<sup>153,154</sup> and two-step nucleation processes.<sup>155</sup>

More specifically, heterogeneous nucleation, is often considered to be caused by the impurities, therefore, models have been developed where the nucleation rate is dependent on the number of impurities present:

$$P(t) = 1 - \exp \left[ -m \left( 1 - \exp \left( -\frac{t - t_d}{\tau} \right) \right) \right] - \exp[-m] \left( \exp \left( -\frac{t - t_d}{\tau_0} \right) - 1 \right) \quad (3.14)$$

Pound – La Mer

Here,  $m$  is the mean number of impurities in each sample, nucleation caused by the impurities is related to the timescale  $\tau$ , and  $\tau_0$  is for nucleation in the absence of impurities. This can be simplified, if the nucleation rate is assumed to be zero for samples without impurities (and  $\tau_0$  is infinite):

$$P(t) = 1 - \exp \left[ -m \left( 1 - \exp \left( -\frac{t - t_d}{\tau} \right) \right) \right] \quad \text{Gompertz} \quad (3.15)$$

These models are expected to be suitable for a range of nucleating systems,<sup>152</sup> and the Pound-La Mer have been found to fit well for the heterogeneous nucleation of the drug molecule effucimibe.<sup>158</sup>

Some of these nucleation models are only suitable or intended for specific types of nucleation behaviour, whilst others can be applied more generally. This means that for some systems, there are several nucleation models which may have similar qualities of fit, especially as the number of fitting parameters are increased. Therefore, the choice of which nucleation model(s) to apply is important, and the results obtained need to be carefully interpreted.

In our analysis, the fitting to these equations has been performed using the *lmfit*

*Model*<sup>159</sup> function in Python, which is similar to the Python *scipy.optimize.curve\_fit* function. We found *lmfit* to be advantageous since it allows the implementation of bounds on the fitting parameters, whilst still calculating standard error values for each of the fitted parameters. In order to ensure that the fitted results are representative, fittings were only performed for systems with five or more nucleation events.

## Chapter 4

# Validation of Force Fields

This chapter outlines a series of general tests that can be used for the validation of force fields for crystallisation studies by considering both the crystal and solution properties predicted using these force fields. A range of OPLS and GAFF force fields that have been used to study urea will be considered. This will enable the most suitable force field to be identified for future work on urea crystal nucleation and growth. Further development and improvement of the force fields will enable MD simulations to be used for more insightful studies of crystallisation phenomena.

### 4.1 Methodology

#### 4.1.1 Molecular Dynamics Simulations

Molecular dynamics simulations were performed using the LAMMPS software.<sup>77,160</sup> All simulations were performed in the isothermal-isobaric ensemble (NPT) using a time step of 2.0 fs, with thermodynamic and structural properties sampled every 2000 fs. The temperature and pressure were controlled by a Nosé-Hoover thermostat and barostat. Damping parameters of 0.2 and 2.0 ps were used for the temperature and pressure, respectively. The size of the simulation cell is rescaled independently for each of the three axes, however, the cell angles are constrained to the initial value of 90°.

Periodic boundary conditions were applied in all dimensions for both the crystal and the solution simulations.

The nine force fields selected in Chapter 3 are all used, the force field parameters are given in Appendix A.1. A cutoff of 9.0 Å was used for both the Lennard-Jones and short range electrostatic interactions, which is the default value used with both GAFF and OPLS. Long range electrostatics were calculated using a particle-particle-particle-mesh with a relative error in forces of  $1 \times 10^{-4}$ . Long range Lennard-Jones interactions and their effects on energy and pressure are corrected for using equation 5 of Sun<sup>161</sup>.

The strength of the Lennard-Jones interactions between 1 – 4 bonded atoms was set to 0.5 of the full interaction strength, and set to 0 for 1 – 2 and 1 – 3 bonds. The strength of the electrostatic interactions between 1 – 4 bonded atoms was set to 0.83333333 and 0.5 of the full interaction strength for the GAFF and OPLS force fields, respectively, and set to 0 for 1 – 2 and 1 – 3 bonds. These scale factors were set according to the defaults of GAFF<sup>99</sup> and OPLS.<sup>97,110</sup>

#### 4.1.2 Crystal Setup and Analysis

Urea crystals were set up in the form I and form IV unit cells. Form I was selected as it is the ambient form. Form III was not considered since it is a high-pressure form, and we are interested only in crystallisation at ambient conditions. Despite the high pressure required to obtain form IV experimentally, this form was considered due to the similarities between form IV and distorted forms seen in some of the form I simulations. For each force field, the unit cell was energy minimised and the optimised unit cell was used to build a crystal supercell of  $5 \times 5 \times 5$  unit cells, which was also energy minimised. The energy minimisation was performed with the Polak-Ribiere version of the conjugate gradient algorithm. The size and shape of the simulation box were allowed to vary independently in all dimensions during the minimisation with a maximum allowed fractional volume change of 0.0001 per iteration. NPT simulations were performed with an anisotropic barostat which allowed the crystal to independently change its  $a$ ,  $b$  and  $c$  lattice parameters. The NPT simulations were performed at temperatures of

300, 400, 450 and 500 K, all at 1 atm, and for 10 ns.

The cohesive energy ( $E_{\text{cohesive}}$ ) was calculated as follows:

$$E_{\text{cohesive}} = \frac{E_{\text{crystal}}}{N} - E_{\text{molecule}} \quad (4.1)$$

Here,  $E_{\text{crystal}}$  is the potential energy of the crystal,  $N$  is the number of molecules in the bulk crystal, and  $E_{\text{molecule}}$  is the potential energy of one molecule in vacuum for the same conditions. Reference simulations of one urea molecule in vacuum were performed for each force field, these were run for 1 ns and the potential energy over this period was averaged.

### 4.1.3 Solution Setup and Analysis

Ten different concentrations of urea solutions were tested for each urea force field, as well as a single simulation with pure water. The solutions contained 1000 water molecules and an increasing number of urea molecules from 0 to 1000, with corresponding concentrations and percentage mass of urea (% mass) given in Table 4.1. The solution concentrations range from very dilute at 33.34 g kg<sup>-1</sup> (3.23% mass) to highly supersaturated at 3334 g kg<sup>-1</sup> (76.9% mass), where the experimental solubility is 1200 g kg<sup>-1</sup> (54.5% mass) at 300 K.<sup>64</sup>

Table 4.1: The ten selected solution compositions.

Molecules		Mass (Da)		Concentration	
Urea	Water	Urea	Water	(g kg <sup>-1</sup> )	(% mass)
0	1000	0.00	18 016	0.00	0.00
10	1000	600.62	18 016	33.34	3.23
50	1000	3003.1	18 016	166.7	14.29
150	1000	9009.3	18 016	500.1	33.34
200	1000	12 012	18 016	666.8	40.00
300	1000	18 018	18 016	1000	50.00
400	1000	24 024	18 016	1334	57.15
500	1000	30 031	18 016	1667	62.50
600	1000	36 037	18 016	2000	66.67
1000	1000	60 062	18 016	3334	76.93

The systems were set up by random insertion of the urea molecules into a simulation

box followed by random insertion of the water molecules. An energy minimisation was performed to ensure that there were no overlapping atoms or molecules. The energy minimisation was performed using a steepest decent algorithm with 0.0001 stopping tolerances for both the energy and forces. NPT simulations were performed at 1 atm and 300 K, using an isotropic barostat. Simulations were equilibrated for 2 ns, followed by a 20 ns production run. The SHAKE algorithm was used to keep the bond lengths and angles fixed in the SPC/E water model.

The mean square displacement (MSD) of urea was calculated using:

$$M(t) = \left\langle \frac{1}{N} \sum_{n=1}^N (x(t) - x(t_0))^2 + (y(t) - y(t_0))^2 + (z(t) - z(t_0))^2 \right\rangle_{t_0} \quad (4.2)$$

Eq. 4.2 is an expanded version of Eq. 3.6, where the MSD is averaged over the number urea molecules ( $N$ ) and multiple time origins are used. Here the MSD is denoted as  $M(t)$ , and  $x(t)$ ,  $y(t)$  and  $z(t)$  are the coordinates of the centre of mass of the urea molecule at time step  $t$ , and  $x(t_0)$ ,  $y(t_0)$  and  $z(t_0)$  are the initial positions. Multiple time origin MSDs, also known as windowed MSDs, are used to maximise the use of the available data, by also using all but the last  $t$  value as  $t_0$  values. The diffusion coefficient ( $D$ ) was calculated using the Einstein equation given in Eq. 3.7.  $M(t)$  was calculated using data sampled every 10 ps and for calculating the  $D$  the gradient of the  $M(t)$  in the time interval between 1 and 10 ns was used.

#### 4.1.4 Force Field Validation Protocol

The bulk crystal structure, for each relevant polymorph, is simulated at ambient conditions and the following properties were tested:

- Crystal lattice parameters
- Crystal density
- Cohesive energy

Additional bulk crystal simulations are performed at higher temperatures to obtain

insight into the crystal stability and melting behaviour.

Aqueous solutions are simulated for both undersaturated and supersaturated concentrations and the following properties were tested:

- Solution density
- Solution radial distribution coefficients
- Diffusion coefficient of urea

## 4.2 Results and Discussion

In this section, we first present the results of the bulk crystal simulations and then present the studies of the aqueous solutions.

### 4.2.1 Crystal Properties

In this section, we investigate the bulk urea crystal properties under ambient conditions. We present the crystal lattice parameters, densities and cohesive energies at 300 K. We also present the crystal lattice parameters and densities at temperatures of 400, 450 and 500 K, along with a discussion of how the higher temperatures affect the crystal form favoured by particular force fields.

#### Crystal Structure

The lattice parameters of the energy minimised form I structure for each force field are shown in Table 4.2. The experimentally measured lattice parameters at 12 K<sup>69,162</sup> are also shown for reference. All force fields slightly underestimate  $a$  and overestimate  $c$  and the density ( $\rho$ ) is overestimated by all of the force fields. OPLS-S and OPLS-G are relatively close to the experimental values, overestimating the density by less than 1%, the worst performing force field GAFF-D1 overestimated this by 11%. In general, energy minimised lattice parameters from the OPLS force fields are in better agreement with experiment than those obtained using the GAFF force fields.

Table 4.2: Lattice parameters and density for the form I crystal structure after energy minimisation.

Force field	$a = b$ (Å)	$c$ (Å)	$\rho$ (g cm <sup>-3</sup> )
Experiment*	5.565	4.684	1.375
GAFF1	5.324	4.820	1.460
GAFF2	5.321	4.774	1.476
GAFF-D1	5.221	4.810	1.521
GAFF-D3	5.328	4.811	1.460
OPLS-AA	5.412	4.795	1.420
OPLS-AA-N	5.350	4.830	1.442
OPLS-AA-D	5.415	4.786	1.421
OPLS-S	5.493	4.785	1.382
OPLS-G	5.490	4.775	1.386

\* Measured at 12 K and ambient pressure.<sup>69</sup>

We also investigated the variation of the crystal structure at 300 K and the average lattice parameters are shown in Figure 4.1, and exemplar crystal structures are shown in Figure 4.2. Tabulated values of our results with the standard deviation are given in Appendix A.2.1. All systems started with the  $5 \times 5 \times 5$  supercell corresponding to the energy minimised form I structure. All four GAFF force fields and OPLS-AA retained the form I structure at 300 K, as shown in Figure 4.2a. The  $a$  and  $b$  lattice parameters fluctuated around an average value where  $a = b$ . The fluctuations were greatest for GAFF1 and GAFF2 where the standard deviation of the fluctuations was 4% of the mean  $a$  and  $b$  values, this was only 2% for OPLS-AA, 1.5% for GAFF-D3 and 0.6% for GAFF-D1. These fluctuations could be an effect of the Nosé-Hoover barostat and thermostat.

For OPLS-S, OPLS-G and OPLS-AA-D a different crystal structure was obtained, with the lattice parameters  $a < b$ , as shown in Figure 4.2b. Despite the changes in the  $a$  and  $b$  lengths, the molecules in the distorted form retain their perpendicular alignment to each other, so that this structure is similar to form I. The non-equal  $a$  and  $b$  lengths are similar to form IV, however, in form IV the molecules are not aligned perpendicularly. Therefore, this distorted structure can be seen as being similar to both forms I and IV. Similar distortions have been reported by Piana and Gale<sup>44</sup> in their work on urea crystal dissolution and growth, using the OPLS-S force field and by Weerasinghe and

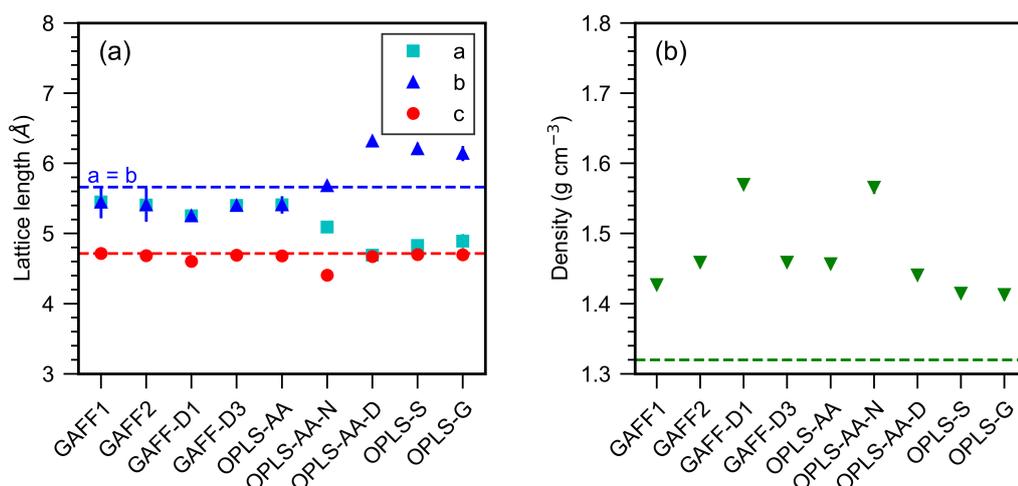


Figure 4.1: Average (a) lattice parameters and (b) density for form I urea crystals at 300 K. Error bars, representing the standard deviation are in most cases smaller than the symbols. Tabulated values with the standard deviation are given in Appendix A.2.1. Horizontal lines represent experimental values at 301 K.<sup>162</sup>

Smith<sup>113</sup> in the development of the urea KBFF force field. OPLS-AA-N does not retain the form I crystal structure, and instead becomes amorphous with  $a \neq b \neq c$  and only very small fluctuations in these parameters, as shown in Figure 4.2c.

The snapshot shown in Figure 4.2a is simply the configuration of GAFF1 at the end of the simulation, where the  $a : b$  ratio is approximately opposite to that of Figure 4.2b. The two structures are significantly different since the structure in (a) changes shape continually whereas the structure in (b) does not. The shape in (a) fluctuates between  $a = b$ ,  $a > b$ ,  $a = b$ ,  $a < b$  ... with an average value of  $a = b$ , whereas the fluctuations to the shape in (b) always maintain  $a < b$ .

All of the force fields overestimate the density significantly compared to the experimental crystal density of  $1.33 \text{ g cm}^{-3}$ .<sup>64</sup> For form I the density performance is best for OPLS-G and OPLS-S with densities of  $1.413 \pm 0.008$  and  $1.415 \pm 0.007 \text{ g cm}^{-3}$ , and worst for GAFF-D1 and OPLS-AA-D with densities of  $1.570 \pm 0.007$  and  $1.566 \pm 0.008 \text{ g cm}^{-3}$ .

Similar performance has been obtained in other simulations of urea crystals. Density

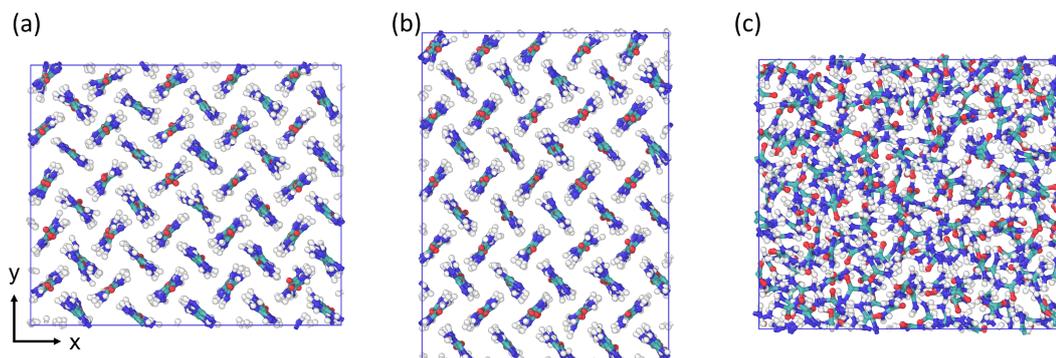


Figure 4.2: Snapshots of the crystal structure at the end of the 10 ns simulation at 300 K, (a) form I (GAFF1 also representative of GAFF2, GAFF-D1, GAFF-D3 and OPLS-AA), (b) distorted form I/IV (OPLS-AA-D also representative of OPLS-S and OPLS-G) and (c) amorphous (OPLS-AA-N).

values of  $\sim 1.382 \text{ g cm}^{-3}$  (this density is calculated from the reported lattice parameters)<sup>3</sup> and later  $\sim 1.46 \text{ g cm}^{-3}$ ,<sup>7</sup> were obtained using GAFF1. These two values were obtained from two different studies within the same research group, showing the effect differences in application has on the results obtained with the same force field. Simulations with other force fields have obtained values of  $1.30 \text{ g cm}^{-3}$  using a specially developed polarisable force field,<sup>105</sup>  $1.38 \text{ g cm}^{-3}$  with CHARMM,<sup>105</sup> and  $\sim 1.512 \text{ g cm}^{-3}$  with KBFF at a reduced temperature of 123 K (this density is calculated from the reported lattice parameters).<sup>113</sup> Our results fit well within the range of previously obtained results, highlighting the importance of both the force field and simulation conditions on the results obtained.

The higher densities, compared to experiment, obtained in the current study is an indication that the force field intermolecular interactions are too strong. For example, there are various parameter differences between GAFF1, GAFF-D1 and GAFF-D3, the most significant difference between these is in the partial charges. For these three force fields, the partial charges are lowest for GAFF1 and highest for GAFF-D1, corresponding to the ranking of the crystal densities. Similarly, for the OPLS force fields OPLS-AA, OPLS-AA-N and OPLS-AA-D have very similar parameters with different partial charges. OPLS-AA-N has some very strong charges, which results in a very dense

solid (the crystal structure is lost). Comparing OPLS-AA to OPLS-AA-D, OPLS-AA has slightly larger partial charges and correspondingly a greater crystal density.

Due to the observed distortion in form I, we also studied the properties of the form IV structure. The lattice parameters and density of the energy minimised form IV unit cell are shown in Table 4.3. The experimentally measured lattice parameters at 296 K and 2.96 GPa are also shown,<sup>71</sup> although we note that the energy minimised structure would correspond to a 0 K, low pressure condition. However, there are only two reported form IV structures<sup>68,71</sup> on the Cambridge Structural Database, so there is not sufficient data available to extrapolate this to 0 K or 1 atm.

Table 4.3: Lattice parameters and density for the form IV crystal structure after energy minimisation.

Force field	$a$ (Å)	$b$ (Å)	$c$ (Å)	$\rho$ (g cm <sup>-3</sup> )
Experiment*	3.408	7.362	4.648	1.711
GAFF1	3.505	7.523	4.847	1.561
GAFF2	3.499	7.102	4.796	1.673
GAFF-D1	3.512	7.052	4.813	1.673
GAFF-D3	3.507	7.454	4.817	1.584
OPLS-AA	3.576	7.614	4.883	1.500
OPLS-AA-N	3.646	7.181	5.222	1.459
OPLS-AA-D	3.698	7.569	4.882	1.460
OPLS-S	3.660	7.688	4.886	1.451
OPLS-G	3.649	7.696	4.888	1.453

\* Measured at 296 K and 2.96 GPa.<sup>71</sup>

All of the force fields overestimate  $a$  and  $c$ .  $b$  is underestimated by GAFF1, GAFF-D1 and OPLS-AA-N and overestimated by the remaining force fields. The density is underestimated by all of the force fields, but GAFF2 and GAFF-D1 are relatively close to the experimental values, underestimating the density by less than 3%, OPLS-AA is underestimated by 12% and the remaining OPLS force fields performed equally badly underestimating this by 15%. In general, the energy minimised lattice parameters from the GAFF force fields are in better agreement with experiment than those obtained using the OPLS force fields. This is opposite to the behaviour obtained for form I.

We also investigated the variation of the form IV crystal structure using the  $5 \times 5 \times 5$  supercell at 300 K. The average lattice parameters and density are shown in Figure 4.3,

with exemplar crystal structures shown in Figure 4.4. Tabulated values of our results with the standard deviation are given in Appendix A.2.1.

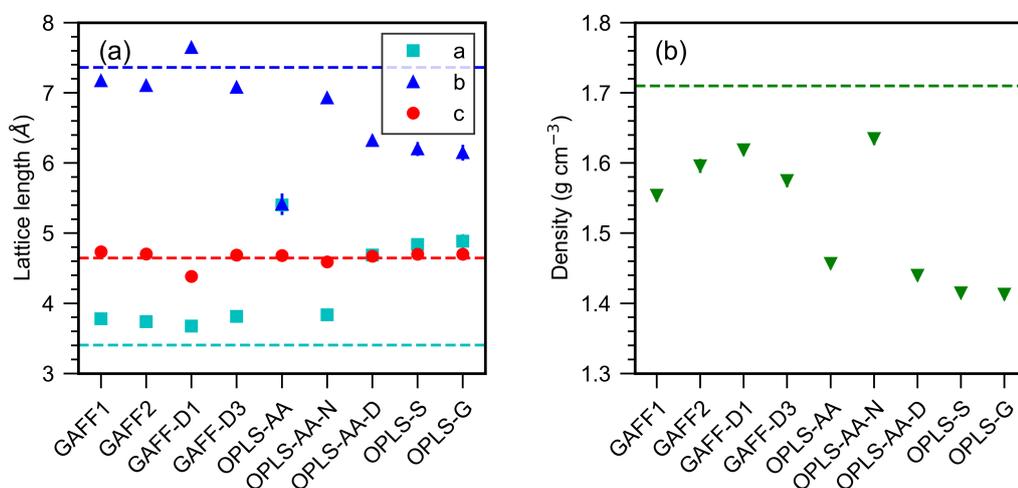


Figure 4.3: Average (a) lattice parameters and (b) density for form IV urea crystals at 300 K. Error bars, representing the standard deviation are in most cases smaller than the symbols. Tabulated values with the standard deviation are given in Appendix A.2.1. Horizontal lines represent experimental values at 296 K and 2.96 GPa<sup>71</sup>.

All of the GAFF force fields retained the form IV structure at 300 K as shown in Figure 4.4a. The fluctuations in the lattice parameters were much less than for the form I structure. The standard deviation in  $a$  and  $b$  is seven times smaller for form IV than form I for GAFF1 and GAFF2, for GAFF-D3 it is three times smaller and it is unchanged for GAFF-D1. The standard deviation in the  $c$  length is much less significant for all of the force fields in both forms I and IV. OPLS-AA reverted to the form I structure, as shown in Figure 4.4b, indicating that this force field is most stable in the form I structure. The lattice parameters for OPLS-AA-D, OPLS-S and OPLS-G differ from the experimental form IV structure and instead take on the distorted structure also obtained from the form I simulations using these force fields, this structure is shown in Figure 4.2b and also 4.4c, starting from forms I and IV respectively. Despite the lattice parameters being similar to the experimental values, the OPLS-AA-N force field does not retain the form IV structure, instead adopting a different crystal form, as shown in Figure 4.5. This new form may be one of the new structures found by the

recent work carried out to predict the polymorphs of urea.<sup>7,163</sup>

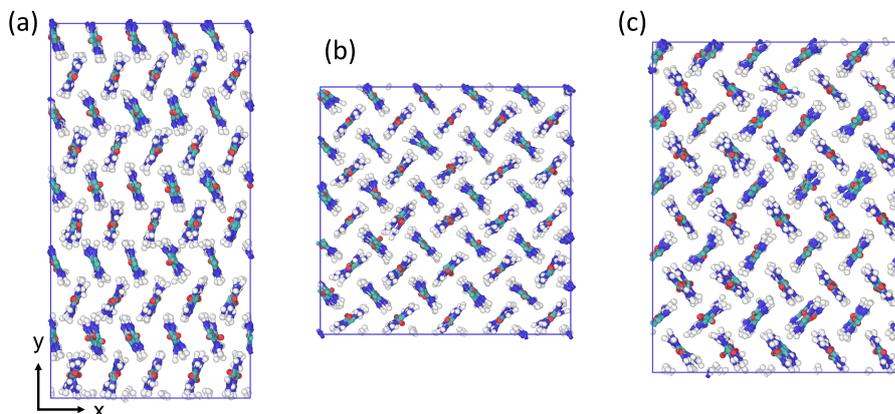


Figure 4.4: Snapshots of the crystal structure at the end of the 10 ns simulation at 300 K, (a) form IV (GAFF1 also representative of GAFF2, GAFF-D1 and GAFF-D3), (b) form I OPLS-AA and (c) distorted form I/IV (OPLS-AA-D also representative of OPLS-S and OPLS-G).

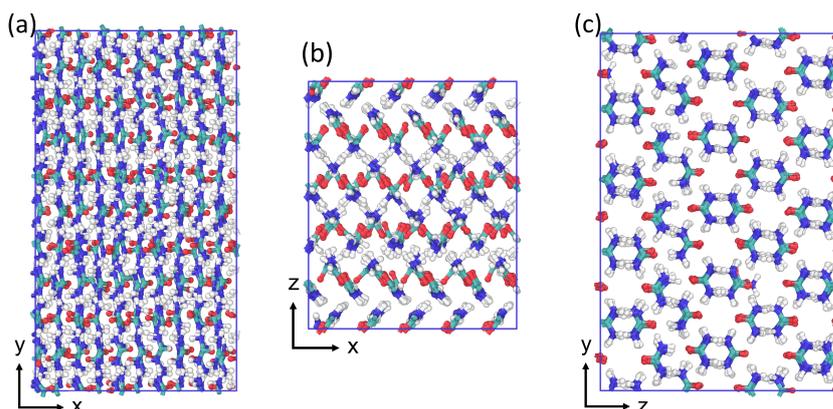


Figure 4.5: Snapshots of the new crystal structure attained by OPLS-AA-N at the end of the 10 ns simulation at 300 K viewed along the (a)  $x$ - $y$  plane, (b)  $x$ - $z$  plane and (c)  $z$ - $y$  plane.

For form IV, a comparison is made to the experimentally measured lattice parameters, and the corresponding density calculated from this. This is unlikely to compare directly to the values obtained from the simulations, since there is a significant pressure difference between these two. The simulations were done at 1 atm whereas the experiments are done above 2.8 GPa ( $\sim 27,000$  atm). However, since form IV is seen in the simulations, comparison to the measured parameters is relevant.

### Crystal Cohesive Energy

Table 4.4 shows the cohesive energy of forms I and IV with the various force fields averaged over the NPT simulation. Tabulated values of the crystal potential energy are given in Appendix A.2.2. For all of the GAFF force fields form IV is the lower energy structure, albeit with only a slight difference between the two forms. This indicates that form IV is more stable than form I for these force fields. This is consistent with the findings of Francia et al.<sup>7</sup> who used Crystal Structure Prediction methods alongside molecular dynamics using GAFF1 to study the relative energy rankings of different polymorphs of urea. OPLS-AA-N has a significant difference in cohesive energy between the form I simulations (which were amorphous solids) and the new crystal structure found from the form IV simulation. There was no difference in cohesive energy for the remaining OPLS force fields, which is expected since they reverted to the same crystal structures.

Table 4.4: Cohesive energy per molecule during 300 K NPT simulation.

Force field	$E_{\text{cohesive}}$ (kJ mol <sup>-1</sup> )		
	Starting in form I	Starting in form IV	Energy difference
Experiment <sup>164-168</sup>	-87.65 to -98.58		
GAFF1	-81.3 ± 0.3	-82.6 ± 0.3	1.3 ± 0.6
GAFF2	-80.0 ± 0.4	-81.6 ± 0.4	1.6 ± 0.8
GAFF-D1	-113.5 ± 0.4	-118.2 ± 0.4	4.7 ± 0.8
GAFF-D3	-91.9 ± 0.4	-92.7 ± 0.4	0.8 ± 0.8
OPLS-AA	-87.5 ± 0.3	-87.5 ± 0.3	0.0 ± 0.7
OPLS-AA-N	-98.8 ± 0.3*	-117.5 ± 0.3**	18.7 ± 0.6
OPLS-AA-D	-83.7 ± 0.4	-83.7 ± 0.4	0.0 ± 0.8
OPLS-S	-79.3 ± 0.3	-79.4 ± 0.3	0.1 ± 0.6
OPLS-G	-80.2 ± 0.3	-80.2 ± 0.3	0.0 ± 0.6

\*Amorphous. \*\*New form.

\*\*\* The mean and standard error are calculated with a sampling frequency of 0.1 ns.

Cohesive energies can be compared to experimental sublimation enthalpies, at the same temperature. Experimental sublimation enthalpies of urea range from -87.65 to -98.58 kJ mol<sup>-1</sup> at 298 K.<sup>164-168</sup> GAFF-D3 is the only force field to obtain cohesive energies within the experimental range, which it does for both forms I and IV. The OPLS-AA and OPLS-AA-D form I and IV simulations produce cohesive energies which are

Table 4.5: Crystal cohesive energy for the NPT simulations.

Force field	$E_{\text{cohesive}}$ (kJ mol <sup>-1</sup> )		
	Starting in form I	Starting in form IV	Energy difference
GAFF1	-81.3 ± 0.3	-82.6 ± 0.3	1.3 ± 0.6
GAFF2	-80.0 ± 0.4	-81.6 ± 0.4	1.6 ± 0.8
GAFF-D1	-113.5 ± 0.4	-118.2 ± 0.4	4.7 ± 0.8
GAFF-D3	-91.9 ± 0.4	-92.7 ± 0.4	0.8 ± 0.8
OPLS-AA	-87.5 ± 0.3	-87.5 ± 0.3	0.0 ± 0.7
OPLS-AA-N	-98.8 ± 0.3	-117.5 ± 0.3	18.7 ± 0.6
OPLS-AA-D	-83.7 ± 0.4	-83.7 ± 0.4	0.0 ± 0.8
OPLS-S	-79.3 ± 0.3	-79.4 ± 0.3	0.1 ± 0.6
OPLS-G	-80.2 ± 0.3	-80.2 ± 0.3	0.1 ± 0.6

\* The mean and standard deviation are calculated with a sampling frequency of 0.1 ns.

within 5% of the experimental values, as does the amorphous OPLS-AA-N simulation although since this is amorphous it is not appropriate to compare it to the crystal cohesive energy.

The cohesive energy of GAFF-D3 was previously calculated for different structures in vacuum including  $8 \times 8 \times 8$  and  $20 \times 5 \times 5$  supercells.<sup>112</sup> It was found that the cohesive energy of the  $8 \times 8 \times 8$  supercell was within the experimental range and the cohesive energy for the  $20 \times 5 \times 5$  supercell was close to the experimental values. This is in agreement with our work. Elsewhere, in a comparison of the CHARMM and the polarised SAPT-FF force fields, the cohesive energies were found to be 101.2 and 99.0 kJ mol<sup>-1</sup> for these respectively.<sup>105</sup>

### Crystal Stability and the Effect of Temperature

Simulations of the bulk urea crystal were carried out at higher temperatures of 400, 450 and 500 K to investigate the effect of temperature on the crystal structure. We considered what happens to urea crystals that start in the form I structure, which is summarised in Table 4.6, and that start in the form IV structure, summarised in Table 4.7. The structure of both the amorphous solid and the melt consists of disordered molecules as is shown in Figure 4.2c, the difference between the two is that the molecules have translational and rotational mobility in the melt but not the amorphous solid, this

was determined by visual inspection of the trajectory using VMD.<sup>85</sup>

Table 4.6: Crystal forms and transitions of bulk crystal at various temperatures starting from form I.

Force field	300K	400K	450K	500K
GAFF1	Form I	Form IV	Form IV	Form IV
GAFF2	Form I	Form IV	Form IV	Form IV
GAFF-D1	Form I	Melt	Melt	Melt
GAFF-D3	Form I	Form I	Form I	Melt
OPLS-AA	Form I	Form I	Form I	Form I
OPLS-AA-N	Amorphous	Amorphous/Melt	Melt	Melt
OPLS-AA-D	Distorted I/IV	Distorted I/IV	Distorted I/IV	Melt
OPLS-S	Distorted I/IV	Distorted I/IV	Distorted I/IV	Melt
OPLS-G	Distorted I/IV	Distorted I/IV	Distorted I/IV	Melt

Table 4.7: Crystal forms and transitions of bulk crystal at various temperatures starting from form IV.

Force field	300K	400K	450K	500K
GAFF1	Form IV	Form IV	Form IV	Form IV
GAFF2	Form IV	Form IV	Form IV	Form IV
GAFF-D1	Form IV	Form IV	Melt	Melt
GAFF-D3	Form IV	Form I	Form I	Melt
OPLS-AA	Form I	Form I	Form I	Form I
OPLS-AA-N	New form	New form	New form	New form
OPLS-AA-D	Distorted I/IV	Distorted I/IV	Distorted I/IV	Melt
OPLS-S	Distorted I/IV	Melt	Distorted I/IV	Melt
OPLS-G	Distorted I/IV	Distorted I/IV	Distorted I/IV	Melt

There is some vibration of the molecules around the C=O axis (looking down the  $z$  direction) for all of the force fields, this can be seen in Figures 4.6a and 4.6b, the extent of this depends on the force field and also increases with temperature. This vibration is not the same for all of the molecules in the crystal, this results in the molecules in different layers not being perfectly aligned with the molecules above and below them.

Starting from form I, both the GAFF1 and GAFF2 crystals transformed to form IV at all higher temperatures. When starting from form IV, both GAFF1 and GAFF2 remain in the form IV structure even at higher temperatures. The alignment of the form IV herringbone pattern inverts spontaneously during simulations at 400 K, changing between the structure in Figures 4.6a and 4.6b, this applies to GAFF1 and GAFF2 structures

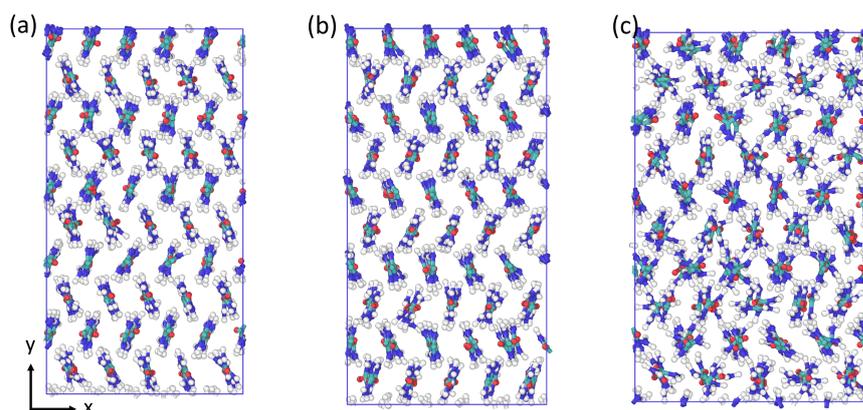


Figure 4.6: Snapshots of simulations run starting with the form IV crystal structure at higher temperatures. (a) ‘left’ herringbone alignment at 400 K for GAFF1 also representative of GAFF2, (b) ‘right’ herringbone alignment at 400 K for GAFF1 also representative of GAFF2 and (c) some molecules are spinning at 500 K for GAFF1 also representative of GAFF2.

starting both from form I and IV. At the higher temperatures (450 and 500 K) the molecules start to spin around their centre, as illustrated in Figure 4.6c, for both structures starting from forms I and IV, which may indicate that this is close to melting.

Starting from form I, the GAFF-D1 crystal melted in the simulations at 400, 450 and 500 K. An additional simulation at 350 K was carried out, at which temperature GAFF-D1 did not melt and the molecules just vibrated around the C=O axis similarly to at 300 K. Starting from form IV the GAFF-D1 molecules showed some spinning at 400 K and the melted at 450 and 500 K. These much lower melting points compared to the GAFF1 and GAFF2 force fields, indicate that this force field is stable for a smaller range of conditions.

When starting from form I, GAFF-D3 remained in form I but melted at 500 K. When starting from form IV, it transitioned to form I at 400 and 450 K, and melted at 500 K. It is interesting that GAFF-D3 seems to be more stable in form I, while the other GAFF force fields are not. There is a smaller difference between the cohesive energies of GAFF-D3 in the two structures than for the other GAFF force fields. This indicates that this behaviour may be due to an entropic effect and that entropy dominates over enthalpy.

OPLS-AA reverts from form I to IV at all temperatures, for the higher temperature simulations there are significant and increasing fluctuations of the crystal size around an average  $a = b$  value. Starting from form I the OPLS-AA-N system at 400 K is somewhere between the amorphous solid described above and a melt, and is a melt at 450 and 500 K. Starting from form IV the OPLS-AA-N system retains the new crystal form at all of the higher temperatures. The OPLS-S, OPLS-G and OPLS-AA-D crystals all take on the distorted form I/IV structure, as described above, for 400 and 450 K starting from both form I and IV. The only exception to this is OPLS-S at 400 K starting from form IV which melts despite the simulation at 450 K not melting. This may indicate that some small instability led to melting which has not been observed in any of the other simulations due to short simulation times. All of the OPLS-S, OPLS-G and OPLS-AA-D crystals melt in the simulations at 500 K.

We note that the experimental melting point of urea is 406 K. We observe that most of the force fields did not result in melting of the bulk urea crystal at 400 or 450 K, with some not melting even at 500 K. However, periodic boundaries make the crystal an infinite lattice with no edges, and therefore there is a superheating phenomenon leading to significant overestimation of the melting point.<sup>169</sup> For this reason, melting points are not accurately determined by simply heating a bulk crystal. The purpose of these simulations at increased temperatures is to gain insight into the relative stabilities of the force fields compared. However, other methods such as studying the crystal-melt interface can be used to more accurately gauge the melting point of a system, this has been used for GAFF1 urea where the melting point was found to lie between 400 and 420 K.<sup>3</sup> Crystallites can also be used to test melting properties instead of the bulk crystal, this is likely to lead to an underestimation of the melting point. This has been carried out with GAFF-D3 for an  $8 \times 8 \times 8$  unit cell cubic crystal and found that crystal melted completely at 385 K.<sup>112</sup>

From the above results of the behaviour at increased temperatures, it appears that GAFF-D3 and OPLS-AA are the only force fields that are stable in form I at most temperatures. OPLS-AA is the most consistent force field, with the same structure and

cohesive energy obtained regardless of the starting structure. It also retains the same structure at all of the temperatures, however, there are significant fluctuations around the average lattice lengths. At 300 K GAFF-D3 is more stable in form IV based on the cohesive energy, however, it is only in simulations at 400 K and above that a spontaneous change from form I to IV is observed. OPLS-G, OPLS-S and OPLS-AA-D are stable in a distorted form I/IV structure. GAFF1 and GAFF2 crystals are more stable in form IV than form I, this is consistent with the cohesive energies obtained and the findings of Francia et al.<sup>7</sup> GAFF-D1 melts at higher temperatures and OPLS-AA-N also exhibits melting or a new crystal form, not observed for urea.

Based on all of the results for the crystal structures, we conclude that GAFF-D3 and OPLS-AA are the most suitable force fields for modelling the form I urea crystal. The performance of form I density and lattice parameters is very similar between these two force fields, however GAFF-D3 experiences smaller fluctuations in the size of the supercell. The cohesive energy of GAFF-D3 is closer to the experimental values than that of OPLS-AA, but GAFF-D3 has a lower cohesive energy for form IV instead of the expected form I. GAFF-D3 can be used to model urea in the high-pressure form IV form at ambient conditions, but has a preference for the form I structure at increased temperatures, whereas OPLS-AA converts from form IV to I even at ambient conditions.

### 4.2.2 Solution Properties

In this section we investigate the properties of urea aqueous solutions over a large concentration range, compared to experiments and simulations from the literature, which are limited to undersaturated concentrations. Our simulations extend into the supersaturated concentration region, as this is relevant to studies of urea crystallisation. The highest concentration studied was 1000 urea molecules in 1000 water molecules, which is 3334 g kg<sup>-1</sup>, which corresponds to a supersaturation of 2.78 at 300 K, based on the experimental solubility.<sup>64</sup> We present the solution density, radial distribution

functions and diffusion properties.

### Solution Density

Time averaged solution densities are shown in Figure 4.7 and compared to experimental data and simulation results from the literature. Tabulated values of our results with the standard deviation are given in Appendix A.2.3. For the pure urea density we used crystal form I densities at 300 K. The solution densities obtained are close to the experimental values at low concentrations but deviate increasingly from experimental values at higher concentrations. This is not surprising, since the SPC/E water force field used here reproduces the experimental density of water well, which gives a density of  $0.9993 \pm 0.0008 \text{ g cm}^{-3}$  at 300K, compared to the experimental density of  $0.997 \text{ g cm}^{-3}$ . However, the urea crystal density is overestimated by all of the force fields leading to the deviations at high urea concentrations.

For the solution phase OPLS-S, OPLS-G, and GAFF1 perform the best. Note that it is not possible to distinguish the OPLS-S and OPLS-G curves from each other, since these force fields are very similar. The most concentrated urea solution for which experimental data is available is 50 mass %, at this concentration all force fields except GAFF-D1 and OPLS-AA-N perform well, overestimating the density by less than 5%. GAFF1, OPLS-S and OPLS-G perform very well overestimating the concentration by less than 3%, GAFF2 and OPLS-AA-D overestimate by less than 4%, and GAFF-D3 and OPLS-AA overestimate by less than 5%. Only GAFF-D1 and OPLS-AA-N significantly overestimate the density by almost 10 and 12% respectively.

Data from Smith et al.<sup>110</sup> is based on the OPLS-S force field coupled with the SPC water model, which underestimated the density of pure water at 300 K. Thus the SPC water model causes the density of the low concentration solutions to be underestimated, while for higher concentrations their OPLS-S simulations reproduce the experimental density values more closely than our work. However, the gradient of the density with concentration in our work is lower than that of Smith et al.<sup>110</sup> and more similar to the experimental gradient. Chitra and Smith<sup>107</sup> use a combination of the non-bonded

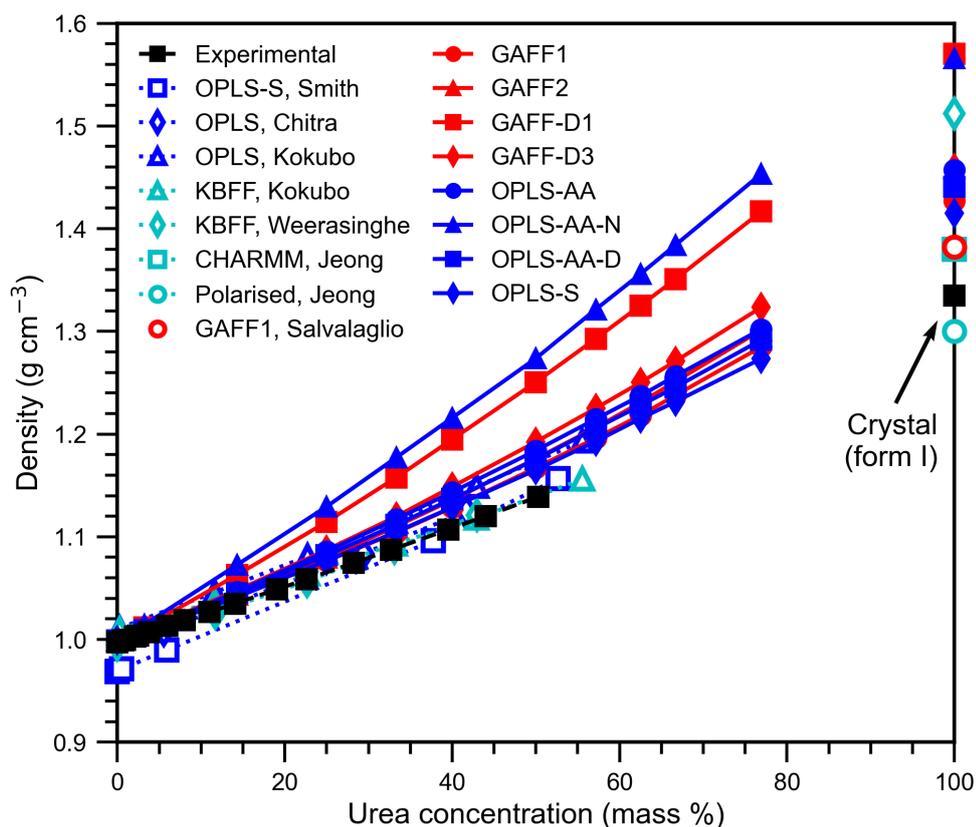


Figure 4.7: Aqueous urea solution density with concentration. OPLS-S overlaps OPLS-G (not shown). Error bars, representing the standard deviation are smaller than the symbols. The key gives the urea force field and the literature source where appropriate. The literature references are the experimental solution density,<sup>170</sup> and the crystal form I density,<sup>64</sup> and simulations from Smith et al.<sup>110</sup> (with SPC water), Chitra and Smith<sup>107</sup> (SPC/E water), Kokubo and Pettitt<sup>103</sup> (TIP3P water), Weerasinghe and Smith<sup>113</sup> (SPC/E water), Jeong et al.<sup>105</sup> (TIP3P water with CHARMM, SWM4-NDP water with polarised SAPT-FF), and Salvalaglio et al.<sup>3</sup>

OPLS urea parameters with the bonded CHARMM parameters, with the SPC/E, which leads to a very close reproduction of the experimental density. Kokubo and Pettitt<sup>103</sup> use only the non-bonded OPLS urea parameters which lead to density values similar to those obtained in this work, whereas their use of KBFF leads to the most accurate density reproduction.

This shows that the combination of water and urea (solvent and solute) force fields influences the solution density, as expected. Apart from GAFF-D1 and OPLS-AA-N,

all of the other GAFF and OPLS force fields tested here, as well as the additional ones from Chitra and Smith<sup>107</sup> and Kokubo and Pettitt<sup>103</sup>, reproduce well the density of aqueous urea solutions.

### Solution Radial Distribution Functions

Now we turn to the solution structure. Two different atomic radial distribution functions (RDFs) are discussed here: O–H<sub>W</sub>, and O–N, where the subscript ‘W’ indicates that the hydrogen belongs to a water molecule and atoms without subscripts belong to urea. Additional RDFs for the H–O<sub>W</sub>, O–O<sub>W</sub>, N–H<sub>W</sub>, N–O<sub>W</sub> and C–O<sub>W</sub> interactions are presented in Appendix A.3. These are compared to RDF curves from the literature, obtained from both experiment and simulation. Two RDFs are shown for each atom pair, a dilute one (3.23% mass) and a more concentrated one (50.0% mass). The dilute reference RDFs are from Duffy et al.<sup>102</sup> at 1.24% mass and Ishida et al.<sup>171</sup> where the concentration was just referred to as ‘dilute’. The concentrated reference RDFs are at 43.0% mass, 45.5% mass and 58.8% mass for Weerasinghe and Smith<sup>113</sup>, Soper et al.<sup>130</sup> and Burton et al.<sup>172</sup>, respectively. Two of the literature RDFs were obtained experimentally, Burton et al.<sup>172</sup> used neutron scattering and Soper et al.<sup>130</sup> used neutron diffraction with empirical potential structure refinement with the OPLS-Urea urea and SPC/E water models to process the results. The other RDFs were obtained from molecular dynamics simulations with the following force fields: Duffy et al.<sup>102</sup> used OPLS-Urea with the TIP4P water model; Ishida et al.<sup>171</sup> used the RISM-SCF method (reference interaction site model - self-consistent-field) for urea with the SPC water model; and Weerasinghe and Smith<sup>113</sup> used the KBFF urea model with the SPC/E water model.

The O–H<sub>W</sub> RDFs are shown in Figure 4.8 and the shapes are similar for both the dilute and concentrated solutions. There is a first peak just below 2 Å, indicating that strong hydrogen bonding between urea and water is present. There are also weaker second and third peaks appearing at around 3 and 5 Å, respectively. The O–H<sub>W</sub> RDFs are similar to those obtained by Duffy et al., but differ from Ishida et al., which have a second

peak at around 3.5 Å. For the concentrated solution, the main difference between the different force fields is that the first peak is significantly higher for the four GAFF force fields. The OPLS RDFs are similar to Soper et al. but differ from Burton et al., which have a weak first peak at 2.5 Å and a barely noticeable second peak.

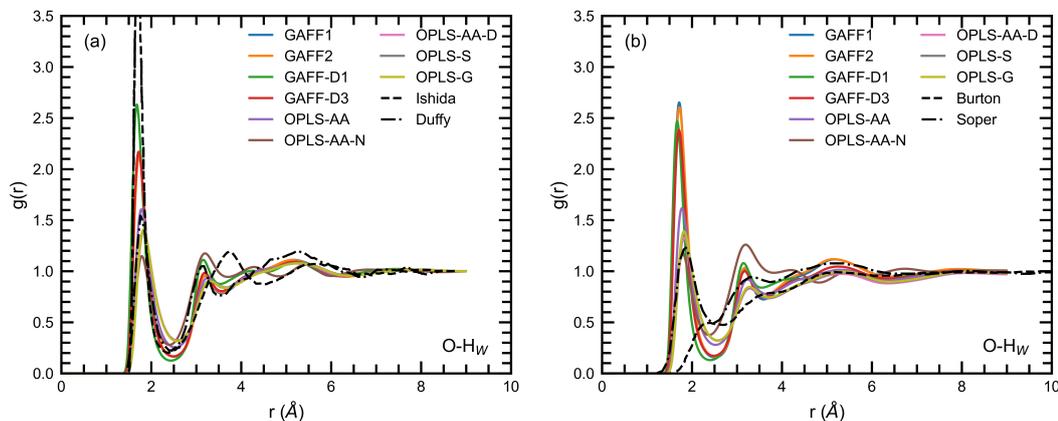


Figure 4.8: O–H<sub>W</sub> RDFs for (a) the dilute solutions and (b) the concentrated solutions. Literature data is taken from Ishida et al.<sup>171</sup>, Duffy et al.<sup>102</sup>, Burton et al.<sup>172</sup> and Soper et al.<sup>130</sup>

The O–N RDFs, shown in Figure 4.9, provide insight into the urea-urea interactions both within the same molecule and between different molecules. The RDFs have two peaks at around 3 and 5 Å. The location of the first peak corresponds to the interaction between O and N molecules within the same molecule, so this should be similar for all of the solutions regardless of concentration. Our RDFs of these materials are similar to those of Burton et al. and Weerasinghe and Smith. The first peak for Burton et al. fell much more slowly in the concentrated solution. Unusually the first peak is much lower for OPLS-AA-N compared with the rest of the force fields. However, there is an extra peak between the first and second for OPLS-AA-N at > 3.5 Å and the second peak is shifted forward to > 4.5 Å. This may be related to the partial charge difference between the O and N atoms, which is significantly greater for OPLS-AA-N compared to all of the other force fields.

Overall our results compare well with the literature sources, with slight variation between the different force fields, with the exceptions of GAFF-D1 and particularly OPLS-

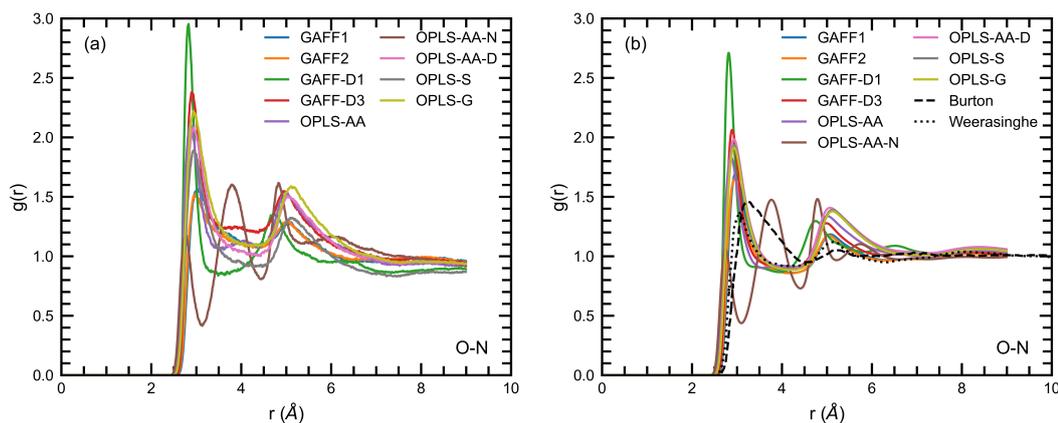


Figure 4.9: O–N RDFs for (a) the dilute solutions and (b) the concentrated solutions. Literature data is taken from Burton et al.<sup>172</sup> and Weerasinghe and Smith<sup>113</sup>.

AA-N which do not reproduce the urea solution structure well. In general, the partial charges and the charge differences between two atoms, have a small, but noticeable, effect on the RDF structure. GAFF1 and GAFF2 have a shared charge set, as do OPLS-G, OPLS-S and OPLS-AA-D and where the height and position of the peaks vary slightly between the different force fields, the peaks are generally very similar within each of these two groups. In general, when comparing the first peak of each RDF for the different force fields, the peaks at the lowest  $r$  values are taller and narrower than the corresponding peaks at slightly larger  $r$  values.

### Diffusion Coefficients

Finally, we compare how the various force fields describe the solution dynamics. Calculated diffusion coefficients for urea in aqueous solution, at 300 K, are shown in Figure 4.10 and compared to experimental data and simulation results from the literature. Tabulated values of our results with the standard deviation are given in Appendix A.2.4. The diffusion coefficients decrease with increasing concentration as the solution becomes more densely packed with urea molecules. Generally, the GAFF force fields more closely reproduce the experimental urea diffusion coefficient than the OPLS force fields do. GAFF-D3 most closely matches the experimental diffusion coefficients followed by

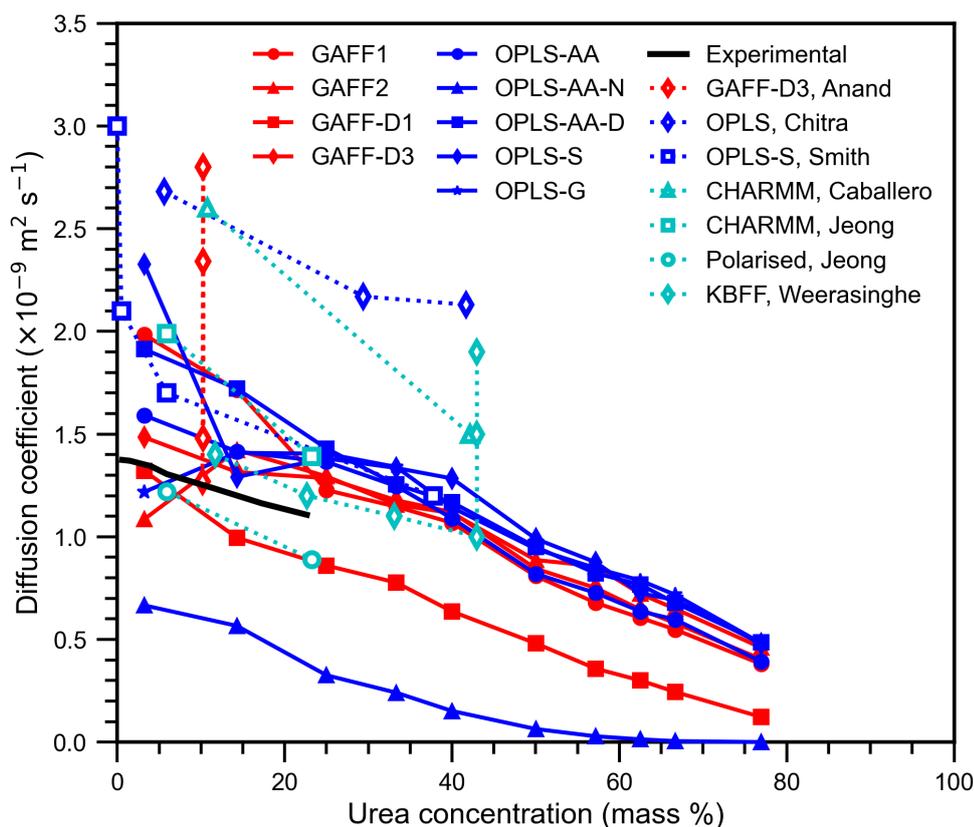


Figure 4.10: Aqueous urea solution diffusion coefficient with concentration. Error bars, representing the standard error are smaller than the symbols. The key gives the urea force field and the literature source where appropriate. The literature references are experimental,<sup>173</sup> and simulated by Anand and Patey<sup>45</sup> (with TIP3P, SPC, SPC/E and TIP4P water from top to bottom), Chitra and Smith<sup>107</sup> (SPC/E water), Smith et al.<sup>110</sup> (SPC water), Caballero-Herrera and Nilsson<sup>104</sup> (TIP3P water), Jeong et al.<sup>105</sup> (TIP3P water with CHARMM, SWM4-NDP water with polarised SAPT-FF), and Weerasinghe and Smith<sup>113</sup> (SPC/E water and also TIP3P and SPC water (top and middle) at their highest concentration).

GAFF-D1 and OPLS-AA. Most force fields studied here overestimate the diffusion coefficient. However, GAFF-D1 and OPLS-AA-N underestimate the diffusion coefficient, these two force fields predicted the highest solution densities, meaning that the solution is more closely packed and reduces the molecular diffusion in the solution.

Our results are within the range of the literature results from other simulations. The results from Anand and Patey<sup>45</sup> used the GAFF-D3 urea force field and four different water force fields (TIP3P, SPC, SPC/E and TIP4P/2005) and show how significantly

the choice of water force field can affect the diffusion coefficient of a solution. Their GAFF-D3 and SPC/E force field combination matched very well with our values obtained for the same force field combination at similar concentrations. Similarly the results from Smith et al.<sup>110</sup> for OPLS-S and SPC are very close to our results from OPLS-S and SPC/E. The large variation in the performance of diffusion coefficients from the literature, can arise from simulation size, exact calculation method and the region of MSD data used to extract the diffusion coefficient. For more accurate estimation of diffusion coefficient a series of simulations at different system sizes should be used.<sup>174</sup>

Overall, looking at the solution results all of the force fields perform relatively well, with the exception of GAFF-D1 and OPLS-AA-N. The OPLS force fields perform slightly better for solution density, the GAFF force fields perform slightly better for urea diffusion coefficients and there is no clear distinction when looking at the RDFs. The GAFF-D1's poor performance follows on from its high crystal density and OPLS-AA-N can reproduce neither realistic crystal nor solution structures and behaviours. For OPLS-AA-N this is disappointing since it is much more user friendly to use the LigParGen software to obtain the force field parameters than to manually go through the lists of parameters for OPLS-AA which are published across several articles. We note that this may be specific for urea, possibly due to its small size, since other smaller organic molecules have successfully been parametrised using LigParGen.<sup>136</sup>

### 4.3 Conclusions

In this work we have compared the bulk crystal and solution properties of urea for four GAFF force fields and five OPLS force fields. Parametrisation of partial charges was done using the Antechamber software<sup>116</sup> for GAFF1 and GAFF2, and using LigParGen<sup>127-129</sup> for OPLS-AA-N and manually taking published parameters from Özpınar et al.<sup>111</sup> for GAFF-D1 and GAFF-D3, from Jorgensen et al.<sup>97</sup> and Weiner et al.<sup>98</sup> for OPLS-AA, from Smith et al.<sup>110</sup> for OPLS-S, also consulting Oostenbrink et al.<sup>101</sup> for OPLS-G and Duffy et al.<sup>102</sup> for OPLS-AA-D. The SPC/E model was selected for

water since other studies have already concluded that it is good for modelling solution properties including density and diffusion and it can successfully be paired with a range of other force fields including both GAFF and OPLS.<sup>45,134,136</sup>

The bulk crystal simulations were carried out at 300, 400, 450 and 500 K for each force field, starting from both forms I and IV, with one additional simulation at 350 K for GAFF-D1 form I, leading to 73 bulk crystal simulations. Starting from crystal form I, at 300 K all four GAFF force fields and OPLS-AA retain the form I structure. The OPLS-AA-N crystal collapses into an amorphous solid, while the remaining OPLS force fields form a distorted form I/IV crystal structure, which has previously been observed with OPLS-S<sup>44</sup> (this structure also has similarities to the form IV urea crystal structure). Starting from crystal form IV, which is obtainable only experimentally at high pressures, the GAFF force fields retain this crystal structure. OPLS-AA transforms to form I, OPLS-AA-N transforms into a new crystal form, while the remaining OPLS force fields take on the same distorted crystal form that was observed when starting from form I.

At 300 K all of the force fields overestimate the experimental form I crystal density by 7–19%, in contrast, the high-pressure form IV density is underestimated by 4–17%. The form I density is best reproduced by OPLS-S and OPLS-G in the distorted form I/IV structure with the worst performance from GAFF-D1 and OPLS-AA-N. The cohesive energies of forms I and IV are very close to each other, with form IV being marginally more stable at 300 K. The crystal cohesive energy, compared to the experimental sublimation, was accurately reproduced by GAFF-D3 with OPLS-AA and OPLS-AA-D also performing well. The stability of the bulk crystals was tested at higher temperatures of 400, 450, and 500 K, although these simulations were not intended to give accurate estimates of the melting point. This found that GAFF1 and GAFF2 have a preference for the form IV structure (with form I to IV transitions), contrastingly GAFF-D3 and OPLS-AA prefer form I (with form IV to I transitions), OPLS-AA-D, OPLS-S and OPLS-G retain the distorted form I/IV crystal structure and GAFF-D1 melts since it is not stable at higher temperatures. These tests on the

bulk crystal indicate that GAFF-D3 and OPLS-AA are the most suitable force fields for modelling urea crystals. GAFF-D3 accurately reproduces the crystal form at 300 K and the cohesive energy as well as favouring form I compared to form IV at higher temperatures. OPLS-AA performs very similarly to GAFF-D3, it has a slightly lower cohesive energy and is most stable in form I for all simulation conditions.

Ten different solution concentrations were studied for each force field, with one additional simulation of pure water, leading to 91 simulations. The range of concentration was varied from very dilute ( $33.34 \text{ g kg}^{-1}$ ) to highly supersaturated solutions ( $3334 \text{ g kg}^{-1}$ ). All of the force fields reproduced the aqueous solution density well, apart from GAFF-D1 and OPLS-AA-N, which significantly overestimated the density even at low urea concentrations. Radial distribution functions showed that all of the force fields, with the exception of OPLS-AA-N, and to some extent GAFF-D1 give the structure of urea solutions in good agreement with the literature. The diffusion coefficients of the solution were reproduced reasonably well by all of the force fields, again with the exception of OPLS-AA-N, which significantly underestimated the diffusion coefficient, indicating that there was very little to no diffusion taking place in any of the supersaturated solutions. Based on the properties we have studied GAFF1, GAFF2, GAFF-D3, OPLS-AA, OPLS-AA-D, OPLS-S and OPLS-G all perform similarly.

OPLS-AA-N is essentially a newer version of the OPLS-AA force field, based on the same bonded and Lennard-Jones parameters, but with a few small changes. However, where OPLS-AA has predetermined partial charges for each atom type, OPLS-AA-N instead calculates these charges based on the molecular structure. For the case of urea, the partial charges and charge dipoles within the molecule differ significantly between the two force fields. The partial charges make OPLS-AA-N unsuitable for modelling urea, either in the crystal or solution state.

We conclude that the best overall performing force fields are GAFF-D3 and OPLS-AA, which have good properties in both the crystal and the solution phases. GAFF-D3 accurately reproduces the crystal cohesive energy and high temperature behaviour of the crystal including predicting the stability of the form I crystal structure, this is also

the best performing force field based on the diffusion coefficients calculated. OPLS-AA has good overall crystal properties with a preference for the stable form I crystal structure in all conditions tested and is the best performing force field based on the solution density. The better performance of GAFF-D3 compared to the other GAFF force fields shows that for GAFF a molecule specific charge optimisation is worthwhile, which has also been noted for other drug like organic molecules.<sup>134,136</sup> Conversely, for the OPLS force fields, the standard OPLS-AA force field performed well and adding the urea-specific charges of OPLS-AA-D or OPLS-AA-N was not advantageous. This highlights the sensitivity of systems to small changes in force fields and the importance of validating the force field for intended applications before use.

More generally we have discussed the importance of performing force field validation tests at the outset of new studies. For the application to crystallisation processes both crystal and solution properties should be tested. We have suggested the use of simple bulk crystal simulations to test the crystal structure, density and cohesive energy as well as bulk solution simulations to test solution density and diffusion coefficients. All of these properties can easily be obtained from short test simulations and compared to experimental data. We also note that clearly reporting the force field and simulation parameters used is important to enable the reproducibility of published work. By continuing to develop the use of MD to model real systems and improve the force fields available for use, we can begin to use MD to carry out more complex studies, including predictive experiments in areas such as drug design, which will enable better overall use of resources in the research and development field.

## Chapter 5

# Urea Nucleation and Solution Structure at Interfaces

Crystallisation from solution is a widely used separation technique and the nucleation of crystals commonly occurs through heterogeneous mechanisms at the interface of the solution with a vessel, impurity or another crystal. Heterogeneous nucleation is dominant at low and moderate supersaturations, although at very high supersaturations homogeneous nucleation becomes significant. Despite the prevalence of crystallisation in natural and industrial processes, a limited understanding of heterogeneous nucleation is still a barrier in the development of efficient crystallisation processes.<sup>1,21</sup>

This work aims to further understanding of the heterogeneous nucleation mechanism by exploring urea crystallisation from aqueous solution at different interfaces. The interfaces used in this work were polytetrafluoroethylene (PTFE), diamond, glass and air. PTFE is commonly used as a coating in reaction vessels, probes and stirrers, as it is strongly hydrophobic and temperature resistant, and despite being assumed to be inert it has previously been shown to enhance the nucleation rate of small organic molecules.<sup>25,29</sup> Alkanes such as heptane and tridecane, which were shown to increase the nucleation rate in glycine,<sup>26</sup> are unsuitable for use with urea, since urea reacts with alkanes to form adducts.<sup>175,176</sup> Diamond is typically used as the sensor material in the

probes for ATR-FTIR spectroscopy<sup>177–180</sup> and is inert with hard, well defined surfaces. ATR-FTIR spectroscopy is frequently used to monitor crystallisation processes, however fouling of the probe is a common issue,<sup>181–183</sup> and therefore it is interesting to understand the effect that the diamond surface may have on this. Glass and air interfaces were also considered as the experiments were performed in glass vials with air above the solution.

In this work, we first present MD simulations to investigate the concentration enhancement in urea aqueous solutions at different interfaces. Second, we present nucleation experiments of urea crystallisation from aqueous solutions at these interfaces and determine the nucleation induction times and hence mean nucleation rates. Finally, we discuss the correspondence between the enhanced interfacial concentration observed in the simulations and the experimentally obtained nucleation rate.

## 5.1 Methodology

In this section, we first present the molecular dynamics simulation details and setup, describe the force fields used, and give an overview of the parametrisation procedure of the model interfaces. We then give an overview of the experimental procedures for the heterogeneous nucleation experiments.

### 5.1.1 Molecular Dynamics Simulations

#### Simulation Setup and Procedure

Molecular dynamics simulations were performed using the LAMMPS software.<sup>77,160</sup>

Urea aqueous solutions were set up between a Lennard Jones (LJ) 9-3 wall potential on one side and a vacuum interface on the other side. The LJ wall potential was parametrised to simulate PTFE, diamond, or  $\alpha$ -quartz ( $\alpha$ -quartz is used to simulate glass) interfaces. The vacuum layer is a good approximation for an air interface since less than one molecule of air ( $N_2$  or  $O_2$ ) would be expected to occupy the vacuum volume,

at standard temperature and pressure. A snapshot of the simulation setup is shown in Figure 5.1. Two different solution concentrations and three different simulation cell sizes

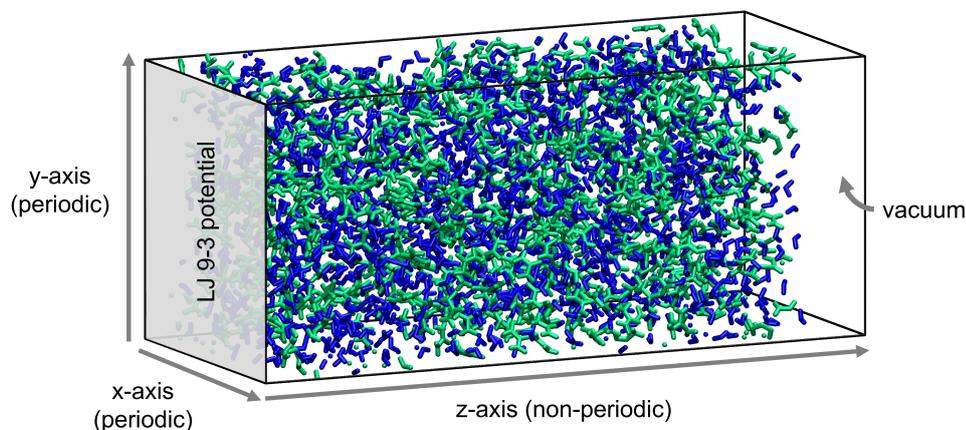


Figure 5.1: Snapshot of the setup of the solution with the LJ interface.

were used for each interface, giving a total of six simulations for each of the PTFE, diamond and  $\alpha$ -quartz interfaces. The two urea solution concentrations were  $1200 \text{ g kg}^{-1}$ , which corresponds to the experimental solubility concentration<sup>64</sup> at  $25 \text{ }^\circ\text{C}$ , and  $2280 \text{ g kg}^{-1}$ , which corresponds to the highest concentration used in our experiments. For all interface systems, the  $x_{\text{cell}}$  and  $y_{\text{cell}}$  lengths were kept at the same value of  $3.45 \text{ nm}$  and periodic boundary conditions were applied in the  $x$  and  $y$  dimensions. The LJ 9-3 wall potential was applied at  $z = 0 \text{ nm}$  and the  $z$  boundaries were fixed and non-periodic. The small, medium, and large systems had  $z_{\text{cell}}$  lengths of  $5.0$ ,  $8.0$  and  $15.0 \text{ nm}$ , respectively, except for the large PTFE system of concentration  $1200 \text{ g kg}^{-1}$ , which had a smaller  $z_{\text{cell}}$  length of  $14.0 \text{ nm}$ . This smaller length did not affect the solution as only the vacuum layer was shorter at  $\sim 0.5 \text{ nm}$  compared to  $\sim 1.0\text{--}1.5 \text{ nm}$  for the other systems. The solution compositions for each system are given in Table 5.1. The urea molecules were added to the simulation box using random insertion, followed by random insertion of the water molecules. Molecules were only inserted into the portion of the simulation box the solution was expected to occupy, based on the estimated solution density. An energy minimisation was performed to ensure that there were no overlapping atoms or molecules. The energy minimisation was performed using

Table 5.1: Solution compositions for each interface system.

Cell size	$z_{\text{cell}}$ (nm)	1200 g kg <sup>-1</sup>		2280 g kg <sup>-1</sup>	
		$N_{\text{urea}}$	$N_{\text{water}}$	$N_{\text{urea}}$	$N_{\text{water}}$
Small	5.0	252	700	342	500
Medium	8.0	504	1400	684	1000
Large	15.0*	1008	2800	1368	2000

\*  $z_{\text{cell}} = 14.0$  for PTFE

a steepest decent algorithm with 0.0001 stopping tolerances for both the energy and forces.

All simulations were performed in the canonical ensemble (NVT), since fixed system boundaries is a requirement for implementing a LJ wall potential. However, we note that the vacuum layer allows the density of the solution to fluctuate naturally. A time step of 2 fs was used, with thermodynamic and structural properties sampled every 2 ps. The simulations were performed at 298 K with the temperature controlled by a Nosé-Hoover thermostat, using damping parameters of 0.2 ps. Simulations were equilibrated for 10 ns followed by a 30 ns production run.

### Force Field Models and Parameters

A urea specific, geometrically optimised version of the GAFF force field with calculated RESP charges<sup>111</sup> and the SPC/E water model<sup>133,135</sup> were used. These force field models were chosen based on the results in Chapter 4, which tested the properties of the urea crystal as well as the properties of urea aqueous solution using the SPC/E water model.

The strength of the Lennard-Jones interactions between 1 – 4 bonded atoms were set to 0.5 of the full interaction strength, and set to 0 for 1 – 2 and 1 – 3 bonds. The strength of the electrostatic interactions between 1 – 4 bonded atoms was set to 0.83333333 of the full interaction strength, and set to 0 for 1 – 2 and 1 – 3 bonds. These scale factors were set according to the defaults of GAFF.<sup>99</sup> The SHAKE algorithm was used to keep the bond lengths and angle fixed in the SPC/E water model.

A cutoff of 9.0 Å was used for both the Lennard-Jones and the short range electrostatic

interactions, which is the default value for GAFF.<sup>99</sup> Long range electrostatics were calculated using a particle-particle-particle-mesh method with a relative error in forces of  $1 \times 10^{-4}$ . Long range Lennard-Jones interactions and their effects on energy and pressure can be corrected for using equation 5 of Sun<sup>161</sup>, however, this was not intended for NVT systems. This correction was applied for the 1200 g kg<sup>-1</sup> solutions but not the 2280 g kg<sup>-1</sup> solutions, with no obvious effects on the results.

The interfaces were represented by LJ 9-3 potentials, which were parametrised using atomistic models as described in McKechnie et al.<sup>51</sup>, significantly reducing the computational cost of the interfacial simulations. The LJ 9-3 potential is implemented within the LAMMPS code as:

$$E_{\text{LJ } 9-3} = \varepsilon_{iw} \left[ \frac{2}{15} \left( \frac{\sigma_{iw}}{z_{iw}} \right)^9 - \left( \frac{\sigma_{iw}}{z_{iw}} \right)^3 \right] \quad (5.1)$$

For interactions between a given atom ( $i$ ) and the wall ( $w$ ).  $\varepsilon_{iw}$  is related to the minimum energy between the atom and the wall,  $\sigma_{iw}$  is related to the distance required for the energy to be zero, and  $z_{iw}$  is the perpendicular distance from the wall to the atom.

The  $\sigma_{iw}$  and  $\varepsilon_{iw}$  parameters are defined differently for the LJ 9-3 potential compared to the LJ 12-6 potential (which is used for the atom-atom interactions). The commonly used arithmetic mean is not suitable for  $\sigma_{iw}$  due to the differences between the LJ 12-6 and 9-3 potentials, as discussed by McKechnie et al.<sup>51</sup>, instead the adjusted form in Eq. 5.2 is used. Despite small differences in the  $\varepsilon_{iw}$  parameters, the geometric mean, given in Eq. 5.3 (and Eq. A.14) remains suitable.

$$0.715\sigma_{iw} = \frac{\sigma_{ii} + 0.715\sigma_{ww}}{2} \quad (5.2)$$

$$\varepsilon_{iw} = (\varepsilon_{ii}\varepsilon_{ww})^{\frac{1}{2}} \quad (5.3)$$

All the force field equations and parameters are given in Appendix B.1, and the wall parametrisation method is described in the section below.

## Parametrisation of the LJ Walls

The interfaces that were parametrised are PTFE, diamond and  $\alpha$ -quartz. First, we describe the steps taken to set up the atomistic interface used in the parametrisation procedure. Then we describe the energy calculations needed between the atomistic interface and a single atom (for all the atom types present in the solution). Then we explain how the simulated interaction energies are used to calculate the wall-interface interaction parameters, still for each individual solution atom type. Finally, we use all the individual parameters to obtain average wall-wall interaction parameters for the interface.

The parametrisation procedure used a 3D, fully atomistic setup of the interface material, with a length of at least 30 Å in each dimension. PTFE was modelled using a backbone chain of length 13-carbon atoms<sup>51</sup> and with chains aligned with the  $z$ -axis, which were packed to give the desired system size. Diamond was modelled using the known lattice structure.<sup>184</sup> The glass interface was modelled with silica, since this is the main component, using the crystalline  $\alpha$ -quartz structure<sup>185</sup> for simplicity. (Borosilicate glass was used in the experiments, this has an amorphous structure and consists of > 70% silica (SiO<sub>2</sub>), < 10% boron trioxide (B<sub>2</sub>O<sub>3</sub>), and smaller amounts of other metal oxides.) The diamond and  $\alpha$ -quartz unit cells were replicated to achieve the desired system size.

The PTFE and diamond surfaces used the GAFF<sup>99</sup> force field, the  $\alpha$ -quartz (glass) surface used a silica specific model<sup>186</sup> that was designed to be combined with many force fields, including GAFF. The LJ parameters that were used from these force fields are given in Appendix B.1. The exposed faces of the diamond and  $\alpha$ -quartz surfaces were simply the (001) faces exposed along the  $z$ -axis from replicating the unit cells, and PTFE had the heads of the chains exposed. Different facets were not considered, since the exposed surface groups are the same and the influence of packing differences is expected to be minimal when the details of the atomistic surface are lost during the reduction to the LJ 9-3 potential, which has only two tunable parameters.

Once the surfaces are prepared, the interaction energy between a single atom with the

interface is simulated. This atom is one of the atom types that will be interacting with the surface, all solution atom types from the solute and solvent were used iteratively. The previous work with glycine used 6 different solute atom types (c, c3, o, n (n4), hn, hx from GAFF) and 1 solvent atom type (only O from SPC/E is used since the LJ values for H are zero). This selection of atom types covers all the atom types in urea, with two additional types, therefore, to remain consistent and interchangeable with the previous work, these 7 atom types were used for the parametrisation in this work.

The LJ interaction energy between the atom and the interface,  $E_{LJ\ 9-3}$ , was calculated for a range of  $z$  separation distances to enable the energy profile to be determined.

Since the atom arrangement varies across the interface, a range of  $xy$  positions across the interface were used to get more representative results. 140  $z$  coordinates spanning from 0.1 to 14 Å above the surface were used to obtain the potential energy profile. For each  $z$  coordinate, 36 distinct  $xy$  positions were used and averaged. The  $\sigma_{iw}$  and  $\varepsilon_{iw}$  values are obtained by fitting Eq. 5.1 to the calculated energy profile. All the calculated  $\sigma_{iw}$  and corresponding  $\sigma_{ii}$  parameters were fitted to Eq. 5.2 to obtain a single  $\sigma_{ww}$  for the interface material. Similarly, all the  $\varepsilon_{iw}$  and corresponding  $\varepsilon_{ii}$  parameters are fitted to Eq. 5.3 and used to obtain a single  $\varepsilon_{ww}$  for the interface material.

For example, Figure 5.2 shows the LJ interaction energy between the diamond interface and a C atom (for one  $xy$  position). Using a Python script, Eq. 5.1 is fitted to this energy profile, which estimates  $\varepsilon_{iw}$  and  $\sigma_{iw}$  to be  $\varepsilon_{iw} = 0.7076$  kcal mol<sup>-1</sup> and  $\sigma_{iw} = 3.6994$  Å. The LJ 9-3 potential corresponding to these fitted parameters is also shown in Figure 5.2. The known  $\varepsilon_{ii} = 0.08860$  kcal mol<sup>-1</sup> and  $\sigma_{ii} = 3.3997$  Å for C can be used with Eqs. 5.2 and 5.3 to estimate the wall parameters as  $\varepsilon_{ww} = 5.8221$  kcal mol<sup>-1</sup> and  $\sigma_{ww} = 2.6440$  Å. These values differ from the final wall parameters, since the final parameters were obtained by fitting all the  $\varepsilon_{iw}$  and  $\sigma_{iw}$ , for all the 36 positions and 7 atom types, to the relevant equations using another Python script.

The calculated  $\sigma_{ww}$  and  $\varepsilon_{ww}$  values are given in Table 5.2. These parameters were then used to calculate the  $\sigma_{iw}$  and  $\varepsilon_{iw}$  values between the parametrised wall and each of the solution atoms, which was used in the simulations. We see that PTFE has strong

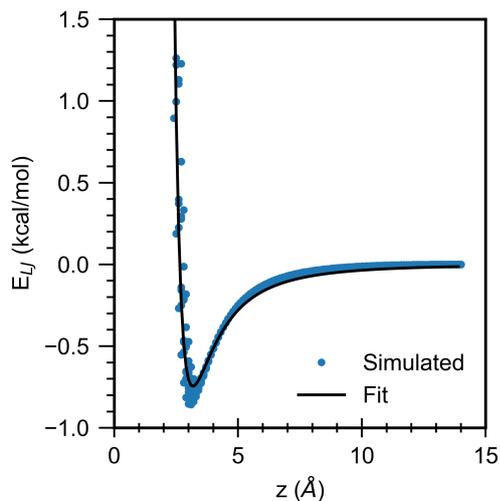


Figure 5.2: Simulated and fitted LJ interaction energy for a C atom at the diamond interface.

interactions, diamond has moderate interactions and the interactions with  $\alpha$ -quartz are weak.

Table 5.2: Fitted  $\varepsilon_{ww}$  and  $\sigma_{ww}$  parameters for the selected interface materials.

Material	$\varepsilon_{ww}$ (kcal mol <sup>-1</sup> )	$\sigma_{ww}$ (Å)
PTFE	12.5	3.7
Diamond	4.5	2.9
$\alpha$ -quartz	0.76	2.7

### 5.1.2 Experimental Setup and Procedure

For the nucleation induction time experiments vials of urea solution were prepared, images were taken by webcam and induction times determined. Urea stock solutions with concentrations of 1800, 2040 and 2280 g kg<sup>-1</sup> were prepared. The concentrations were selected to give solutions with supersaturations of 1.5, 1.7 and 1.9 at 25 °C, based on the literature urea solubility<sup>64</sup> of 1200 g kg<sup>-1</sup> according to Eq. 2.1. Urea powder (Sigma-Aldrich, for electrophoresis  $\geq 98\%$ ) and ultra pure water (Fluka Analytical, LC-MS CHROMASOLV) were used. The required amount of urea and water was weighed into a glass bottle, which was then sealed. The bottle was placed on a hot plate, with a

temperature probe, in an incubator, both set to 60 °C, and left until the urea had fully dissolved. No stirrer bar was used, instead the bottle was gently shaken from time to time to aid dissolution.

To prepare the vials for the induction measurements, 1.0 ml of solution was pipetted into 1.5 ml borosilicate glass vials (VWR 548-0018), to prevent crystallisation this was performed in an incubator at 60 °C. New vials and pipette tips were used for each sample, these were used without any pre-cleaning, and were preheated to 60 °C in the incubator. Each vial was sealed immediately after the solution was added and was transferred to a Polar Bear Plus Crystal, which had been preheated to 60 °C. For the experiments with the additional interfaces, the PTFE stirrer bars or diamonds were added to the vials before preheating. Once all the vials had been filled, the temperature of the Polar Bear was increased to 70 °C. The vials were held at 70 °C for 1 hour to ensure that any crystals formed during the preparatory steps fully redissolved. The vials were cooled to 25 °C at a rate of  $-1.5\text{ °C min}^{-1}$ . Finally, the vials were carefully transferred to vial racks in an incubator set to 25 °C for temperature control. Some vials nucleated during the preparatory steps, either during the cooling ramp in the Polar Bear or on transfer to the vial racks in the incubator. The experiments were repeated until there were 40 vials for each concentration-interface combination, not counting the vials that crystallised during the preparation. This means that several different solutions were prepared for each of the systems.

The diamonds used in the nucleation experiments were previously used to study crystallisation of pharmaceutical materials in high pressure anvils, but had developed defects that made them unsuitable for continued use. Before their first use in the nucleation experiments, the diamonds were cleaned with multiple washes of acetone to remove traces of the araldite glue, which was previously used to attach the diamonds to anvils. The diamonds were then washed with copious amounts of deionised water and left to dry. One of the larger diamonds had a slight cut along one of the faces, and the other five diamonds had small cracks through which some residue material from the high pressure experiments had entered. Despite our cleaning efforts, it was not

possible to remove the residue from inside the diamonds, but as this material entered the diamond under very high pressure, it was thought to be unlikely for the residue to escape in the ambient conditions of the nucleation experiments. The same six diamonds were reused for all the nucleation experiments, and these were cleaned between each use with copious amounts of deionised water and left to dry. A selection of about fifty PTFE stirrer bars were reused in the experiments with PTFE bars, these were washed with copious amounts deionised water and left to dry before each use. The glass vials were taken straight from the packaging and no pre-cleaning was performed.

Experiments with only the air and glass (vial) interfaces were taken as the control, to find the baseline nucleation rate for the systems. Heterogeneous nucleation increases with surface area, so this must be considered when comparing different interfaces. The wetted area of the glass vials was approximately  $490 \text{ mm}^2$ . PTFE magnetic stirrer bars of size  $7 \times 2 \text{ mm}$  were used, with an estimated surface area of  $44 \text{ mm}^2$ . Two different sizes of diamonds were used, two smaller and four larger ones, each with estimated surface areas of  $25$  and  $35 \text{ mm}^2$ , respectively. The addition of the PTFE bar or diamond increased the liquid level in the vial, which increased the wetted glass surface area. The volume of the PTFE bar was  $\sim 2\%$  of the volume of the solution, which led to  $\sim 2\%$  increase in the wetted glass area ( $\sim 500 \text{ mm}^2$ ). Compared to the PTFE bars, the change in the wetted surface area caused by the diamonds was less due to the smaller sizes. The air surface area was approximately  $70 \text{ mm}^2$ , this was estimated assuming a flat meniscus.

The vials were kept in the incubator for 3 days, during this time they were monitored with webcams, which took pictures of all the vials every 5 minutes. These images were used to determine the nucleation time in each of the vials, with an accuracy to within  $\pm 5$  minutes, based on the imaging interval. The obtained crystals were not isolated or characterised, since this study focuses on the nucleation time and there are not several urea polymorphs to consider at ambient conditions.

## 5.2 Results and Discussion

We first present the results of the simulations at the interfaces, including a detailed analysis of the composition and ordering of molecules in the interfacial regions. Next, we present the results from the nucleation experiments, including nucleation rates fitted to the biexponential model, and discuss the effects of the individual nucleants used and any variability between them. Finally, we discuss the combined experimental and computational insights.

### 5.2.1 MD Simulations of Urea Solutions at Interfaces

We first present the simulations with the LJ wall representing the PTFE interface. Figure 5.3 shows a snapshot of the  $1200 \text{ g kg}^{-1}$  solution at the PTFE surface and the corresponding density profile, this is for the medium sized system where  $z_{\text{cell}} = 8.0 \text{ nm}$ , where we see that the solution has a thickness of  $\sim 7 \text{ nm}$ . Next to the PTFE surface,

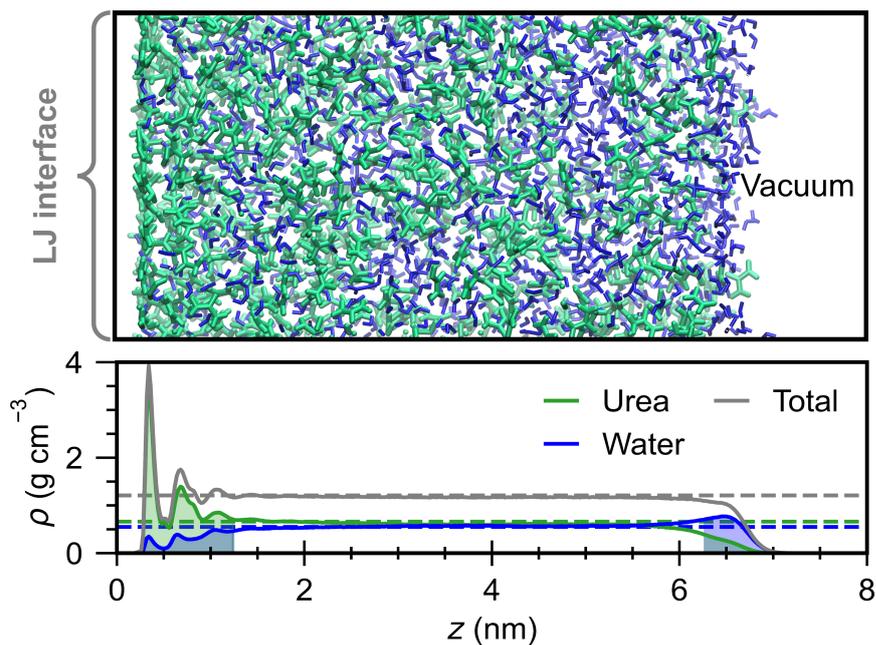


Figure 5.3: Urea solution structure at LJ interfaces for  $1200 \text{ g kg}^{-1}$  PTFE. Urea molecules are green, and water molecules are blue. The upper panel shows a snapshot of the equilibrated solution structure and the lower panel shows the corresponding density profile along the  $z$ -axis.

on the left of the snapshot, two layers can be easily distinguished in the solution, and these mainly consist of urea molecules. It is thus clear that the solution near the PTFE interface has a higher ratio of urea to water molecules, compared to that of the bulk solution in the centre of the film. On the other side of the film there is a region of vacuum, representing air. The solution next to the vacuum does not exhibit layering and it can be seen from the snapshot, that the solution in this interfacial region mainly consists of water.

Figure 5.3 also shows the density profiles of urea and water from the PTFE to vacuum interfaces and the layering and interfacial effects described above can be clearly seen. From the density profiles, a very weak third layer can be identified at the PTFE interface in both the urea and water density profiles. The three layers are within the first 1 nm of solution bordering the interface, and after this the densities level off to their bulk densities in the centre of the film. The average bulk densities, shown by the dashed horizontal lines, were interpolated from the results in Chapter 4, where simulations of bulk urea solutions were performed under NPT conditions using fully periodic boundaries. There is a region of enhanced water concentration and corresponding urea depletion, in the solution at the vacuum interface, which stretches  $\sim 1$  nm into the bulk solution. The layering effect is well known for liquids in contact with surfaces, and has also been observed in glycine solutions in contact with both a fully atomistic frozen, crystalline tridecane surface and a tridecane LJ 9-3 surface.<sup>51</sup>

These results show that when an aqueous urea solution is in contact with an interface the dispersion interactions between the surface, the urea and the water molecules cause a significant change in the interfacial solution density and composition. The increase in the urea concentration at the PTFE interface and the corresponding decrease in the water concentration is related to the urea molecule having stronger dispersion interactions than the water molecule. Conversely, the absence of dispersion forces at the vacuum interface mean that it is more favourable for urea to stay within the solution, maximising the dispersion interactions with other molecules, which results in the more weakly interacting water molecules being pushed to the solution edge.

Three different solution film widths were tested, approximately 3.5, 7.0 and 13.5 nm, with the number of molecules differing by a factor of two between both the small and medium, and medium and large system sizes. The film width and solution concentration did not influence the interfacial behaviour, as shown in Figure 5.4. The density peaks overlap perfectly for the first peak and very closely for the second peak regardless of concentration and system size, and by the third peak, as the bulk level is approached, the profiles split according to their concentration. In the middle part of the film, the

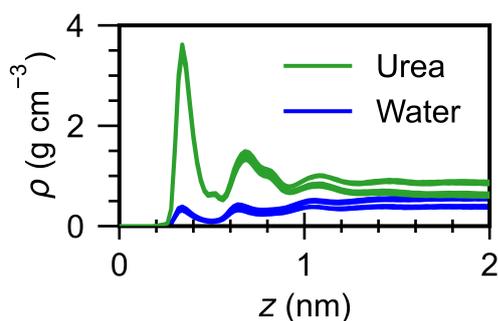


Figure 5.4: Solution structure at the PTFE LJ 9-3 interface. All six simulations, from the three system sizes and the two concentrations, have been overlaid.

solution is not influenced by the interactions with the interfaces, instead bulk behaviour is achieved. The region of bulk solution is significantly smaller in the small systems, nonetheless, the concentration profiles level off in the middle section between the two interfaces. The enhanced interfacial concentration does not significantly deplete urea from the bulk solution, which may in part be due to the opposing effect at the vacuum interface. The bulk concentration of urea is very similar to the expected value, as can be seen in Figure 5.3. This applies to both concentrations and all three film thicknesses, which means that even small simulations can achieve representative bulk behaviour. This was not the case in similar simulations of glycine solutions where significant depletion of glycine in the centre of the films was noted.<sup>51</sup> This may be due to the much lower glycine solution concentration compared to urea, or due to the stronger clustering of glycine.

The individual density profiles for all the interfaces and at both of the concentrations

are shown in Figure 5.5 for the medium sized films. Similar density profiles for the small

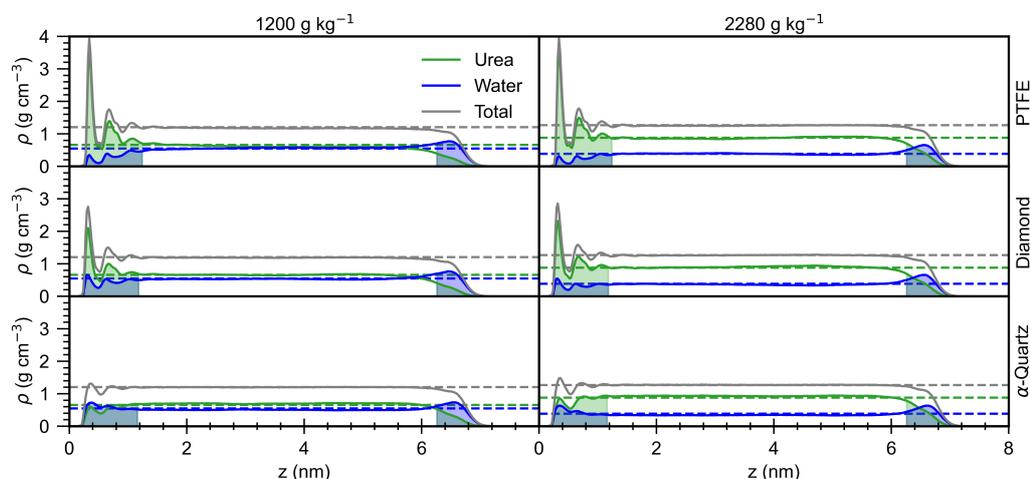


Figure 5.5: Urea density profiles for all systems studied, with 1 nm regions highlighted at both the LJ and vacuum interfaces. The dashed horizontal lines show the average concentration from the bulk NPT simulations in Chapter 4.

and the large simulations are available in Appendix B.2.1. It can clearly be seen that the largest overall interfacial density enhancement is obtained at the PTFE interface followed by diamond and then  $\alpha$ -quartz, while the vacuum interface shows density reduction. The same trend applies when considering only the interfacial density of urea. Conversely, water is most depleted in the order of PTFE, diamond, and  $\alpha$ -quartz, and is enhanced at the vacuum interface. These trends are the same for both the 1200 and 2280 g kg<sup>-1</sup> solution concentrations. The magnitudes of the density peaks for the two concentrations at each interface are very similar, this can be seen in Figure 5.4 for the PTFE interface and similar plots are available for the diamond,  $\alpha$ -quartz, and vacuum interfaces in Appendix B.2.2. This indicates that the molecular layering and packing at the surface depends more strongly on the interfacial interaction than the solution concentration. Therefore, surface materials with surface interactions stronger than that of PTFE, are expected to induce even higher interfacial densities.

The diamond interface also induces urea concentration enhancement, again with three clear peaks. The shape of the concentration profiles for PTFE and diamond are very

similar, except that the magnitude of the urea peaks are lower for diamond, which allows the water peaks to be slightly higher. The lower urea peaks at the diamond interface, compared to PTFE, corresponds to the interface having weaker dispersion interactions. The urea peaks are slightly higher for the 2280 g kg<sup>-1</sup> solution compared to the 1200 g kg<sup>-1</sup> solution, showing that for weaker interface dispersion interactions the concentration affects the interfacial concentration enhancement more in urea solutions. In both solutions the initial water peak is slightly higher than the bulk concentration, meaning the interfacial water concentration is larger for the 1200 g kg<sup>-1</sup> than the 2280 g kg<sup>-1</sup> solution. The  $\alpha$ -quartz interface does not induce any significant concentration enhancement of urea. There are still three peaks in the concentration profile, but these are all approximately at the expected bulk concentration.

The water concentration profile has also got three sets of peaks and troughs at the LJ interfaces. The water concentration near the PTFE interface is significantly reduced in the 1200 g kg<sup>-1</sup> solution, but is only slightly reduced in the 2280 g kg<sup>-1</sup> solution as the overall water concentration in the bulk is lower. This results in the interfacial water concentrations being very similar for both solution concentrations, similarly to the urea concentrations. At the diamond interface the first peak in the concentration profile is slightly above the expected bulk value, with the subsequent peaks being just below the expected value. For both solution concentrations at the  $\alpha$ -quartz interface, the first two water peaks are higher than the bulk concentration. This suggests that the weak LJ wall potential is not able to induce concentration enhancement of urea but instead has a stronger interaction with the water.

To quantify the interfacial effects, we have calculated the densities in the first 1 nm of solution bordering the interface ( $\rho_{\text{int}}$ ) as indicated by the shaded regions in Figure 5.5. The corresponding interfacial densities of urea ( $\rho_{\text{urea}}$ ) and water ( $\rho_{\text{water}}$ ) in the 1 nm interfacial region are given in Table 5.3. There is a gap between the LJ interface and the solution molecules, this is caused by the highly repulsive part of the potential, and relates to the  $\sigma_{ww}$  parameter and the combining rules. This region has been excluded from the interfacial density calculations, therefore, the interfacial densities have been

calculated for the 1 nm region between  $0.715\sigma_{ww}$  nm and  $1 + 0.715\sigma_{ww}$  nm.

The bulk values presented in Table 5.3 were obtained from the density equilibrated bulk NPT simulations of urea solutions performed in Chapter 4, as different solution concentrations were used, linear interpolation has been performed to obtain the bulk values presented here. The bulk density values were 1.192, 1.225, 1.227 and 1.323 g cm<sup>-3</sup> for 1000, 1334, 2000 and 3334 g kg<sup>-1</sup>.

Table 5.3: Average interfacial density. The number of samples ( $n$ ) for each interface is given and the  $\pm$  values on densities are the standard errors.

Interface	$c$ (g kg <sup>-1</sup> )	$n$	$\rho_{\text{urea}}$ (g cm <sup>-3</sup> )	$\rho_{\text{water}}$ (g cm <sup>-3</sup> )	$\rho_{\text{int}}$ (g cm <sup>-3</sup> )
PTFE	1200	3	$1.05 \pm 0.05$	$0.34 \pm 0.01$	$1.39 \pm 0.06$
Diamond	1200	3	$0.85 \pm 0.03$	$0.42 \pm 0.01$	$1.28 \pm 0.04$
$\alpha$ -quartz	1200	3	$0.55 \pm 0.01$	$0.55 \pm 0.01$	$1.09 \pm 0.02$
Vacuum	1200	9	$0.16 \pm 0.01$	$0.41 \pm 0.01$	$0.57 \pm 0.02$
Bulk*	1200	1	$0.66 \pm 0.01$	$0.55 \pm 0.01$	$1.20 \pm 0.01$
PTFE	2280	3	$1.16 \pm 0.05$	$0.26 \pm 0.01$	$1.42 \pm 0.05$
Diamond	2280	3	$0.99 \pm 0.03$	$0.33 \pm 0.01$	$1.32 \pm 0.04$
$\alpha$ -quartz	2280	3	$0.75 \pm 0.02$	$0.41 \pm 0.01$	$1.16 \pm 0.02$
Vacuum	2280	9	$0.24 \pm 0.01$	$0.34 \pm 0.01$	$0.58 \pm 0.02$
Bulk*	2280	1	$0.88 \pm 0.01$	$0.39 \pm 0.01$	$1.27 \pm 0.01$

\* Bulk values are interpolated from the bulk simulations in Chapter 4.

The results presented here are averaged over the different system sizes, since this had little effect on the interfacial concentrations. The interfacial concentration for each individual system is provided in Appendix B.2.3.

The calculated average densities at the vacuum interface are also based on a 1 nm width of solution, starting at the solution-vacuum interface. The edge of the solution is determined separately for each simulation. The maximum  $z$  positions achieved by any urea molecule is determined for 1 ns bins of the equilibrated trajectory. The average value of these maximum  $z$  values is then taken to be the position of the solution-vacuum interface. The values presented in Table 5.3 are an average over all the nine simulations that were performed for each concentration (three LJ interfaces each with three system sizes). The vacuum interfacial densities were calculated individually for all the systems studied and the spread of values are similar. This is a further indication that bulk behaviour is obtained in all the systems and no significant urea depletion is caused by

the interfaces. All the values are given in Appendix B.2.3.

Figure 5.6 illustrates how the interfacial urea density is related to the strength of the wall potential. Initially, the interfacial density increases significantly with  $\varepsilon_{ww}$  at low  $\varepsilon_{ww}$  values, but at higher  $\varepsilon_{ww}$  this effect is decreasing. This indicates that a much

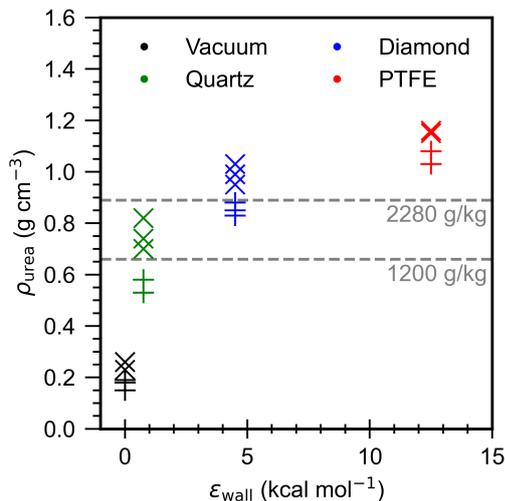


Figure 5.6: Interfacial urea density compared to the wall strength, for all the systems studied. Three symbols are shown for each interface and concentration, 1200 g kg $^{-1}$  (+) and 2280 g kg $^{-1}$  (×), for each of the small, medium and large systems. The horizontal lines show the average concentration calculated from the bulk NPT simulations in Chapter 4.

stronger wall (such as  $\varepsilon_{ww} = 17.2$  kcal mol $^{-1}$  for graphite<sup>51</sup>) would not significantly increase the interfacial urea density. Therefore, it seems that there might be a system-dependent critical wall strength required for the interface to interact with the solute, but significantly exceeding this strength will not significantly increase the interfacial effects. We note that, although the strongly interacting surfaces result in a significant concentration enhancement of urea, the interface is at equilibrium, so the higher concentration does not mean that the supersaturation is higher. In addition, we do not know the solubility of urea for the force field, which is likely to differ from the experimental value. Nevertheless, we would expect to see an enhancement regardless of the actual value of the solubility.

Throughout the simulations the molecules move between the interfacial layers and the

bulk solution. Each molecule moves between the interface and bulk regions at different rates: some molecules stay predominantly either at the interface or in the bulk, while other molecules are constantly exchanging between these. Whilst molecules are in an interfacial layer, they continue to move across the  $xy$  plane. This has been determined by visual inspection of the trajectories using VMD.<sup>85</sup>

The change in the interfacial urea density is tracked throughout the equilibration and production run, this is shown in Figures B.7-B.9 in Appendix B.2.4. There is a clear increase in the interfacial concentration during the first 10 ns (equilibration period) for the PTFE and diamond simulations, after which the interfacial concentration fluctuates around the average value. This is an indication that the system has equilibrated, this happens faster for diamond systems than for PTFE, and is hardly distinguishable for the  $\alpha$ -quartz systems. There is no significant influence of system size on this equilibration behaviour.

The changes in interfacial concentration are independent of system size, which indicates that the diffusion of molecules during the equilibration run is sufficient for equilibrium to be reached. Over the equilibration period, the molecules are on average expected to diffuse by 7.9 and 5.4 nm in the low and high concentration systems, respectively, based on interpolation of the diffusion coefficients previously calculated for bulk urea solutions in Chapter 4. This corresponds to  $\sim z_{\text{cell}}$  for the small systems and  $\sim \frac{1}{3} z_{\text{cell}}$  for the large systems, and is greater than both  $x_{\text{cell}}$  and  $y_{\text{cell}}$  for all the systems.

### Ordering at the Interface

The enhanced urea concentration in the interfacial region may reduce the barrier to nucleation by reducing surface energy and improving the kinetics, it is possible that these effects may be further enhanced by changes in the solution structure. Therefore, we have analysed the ordering of the urea molecules near the interfaces and in the bulk solution. To do this, we calculated the  $P_2$  bond orientation parameter for both the C–O and N–N vectors within the urea molecules. Urea is a flat molecule and the C–O and N–N vectors are perpendicular to each other, so this combination provides

good information about the orientation of the full molecule.  $P_2$  was calculated using:

$$P_2 = \frac{3}{2} \langle \cos^2 \theta \rangle - \frac{1}{2} \quad (5.4)$$

Here,  $\theta$  is the angle between the normal of the  $z$ -axis (which is normal to the interface) and the bond vector. The angular brackets indicate that average values are obtained considering all the molecules at all times, for each  $z$  bin.  $P_2$  values can range from  $-0.5$  to  $1.0$ . A value of  $-0.5$  indicates that all the bond vectors lie parallel to the interface, whereas  $1.0$  indicates that all the bond vectors are perpendicular. For a mixture of randomly oriented molecules, the average  $P_2$  value will be  $0.0$ . Here, the  $P_2$  value is calculated for each molecule at each time step, which is averaged, for bins along the  $z$ -axis, for all the molecules for the desired period of time.

Figure 5.7 shows how the  $P_2$  values of the C–O and the N–N vectors vary with the distance from the interface for the  $1200 \text{ g kg}^{-1}$  medium sized PTFE system. Similar plots for all the other systems are available in Appendix B.2.5. The molecules are

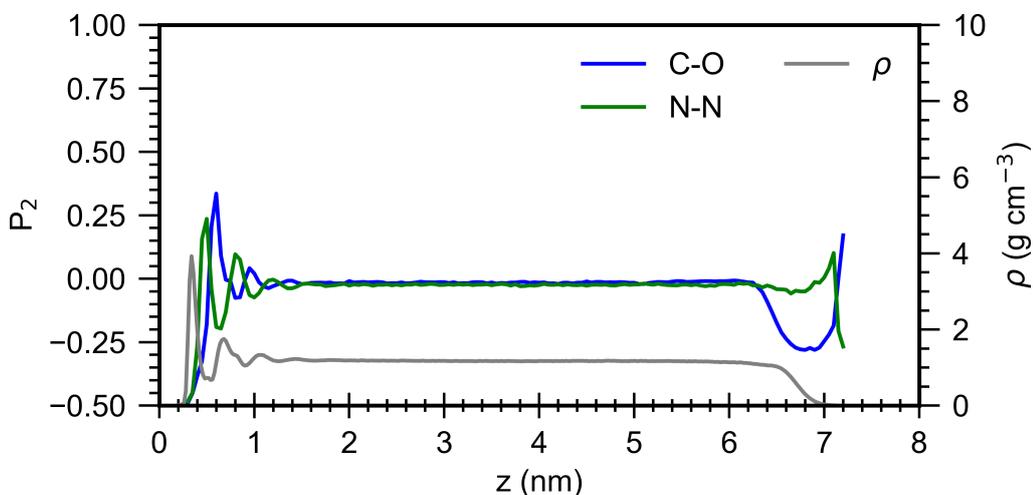


Figure 5.7:  $P_2$  bond orientations for  $1200 \text{ g kg}^{-1}$  medium PTFE system, the total solution density is given for reference.

clearly oriented at the interface with the  $P_2$  values starting at  $-0.5$  and rapidly increasing to above zero, then falling below zero, but not as low as previously, followed

by two similar oscillations around zero before levelling off in the bulk solution. These oscillations are related to the layering in the solution near the interface: the troughs in the  $P_2$  values are related to the peaks in the density profile and vice versa. In the initial part of the concentration peak ( $\sim 0.2-0.4$  nm), the urea molecules are tightly packed and highly ordered, so that both the C–O and the N–N vectors are parallel to the interface ( $\theta = 90^\circ$ ), indicating that urea is lying flat against it. This maximises the interactions between the C, O and N urea atoms with the interface by bringing them as close as possible. The  $P_2$  values above or near zero correspond to the troughs in the concentration profile, where there are fewer urea molecules present, due to the layering in the solution. Here the molecules fill the space that is available and the most favourable ordering is lost. In the second and third solution layer, the average  $P_2$  value shows that only a small proportion of the molecules achieve the orientation of the first layer. This behaviour is similar for both the C–O and the N–N bond vectors, however from  $\sim 0.4$  nm the  $P_2$  profiles separate slightly, with the N–N peaks and troughs occurring at slightly lower  $z$  values than those for C–O.

The  $P_2$  profiles are similar for all the PTFE and diamond systems. The profile for the solutions with the  $\alpha$ -quartz interface are simpler, as is shown in Figure 5.8. The

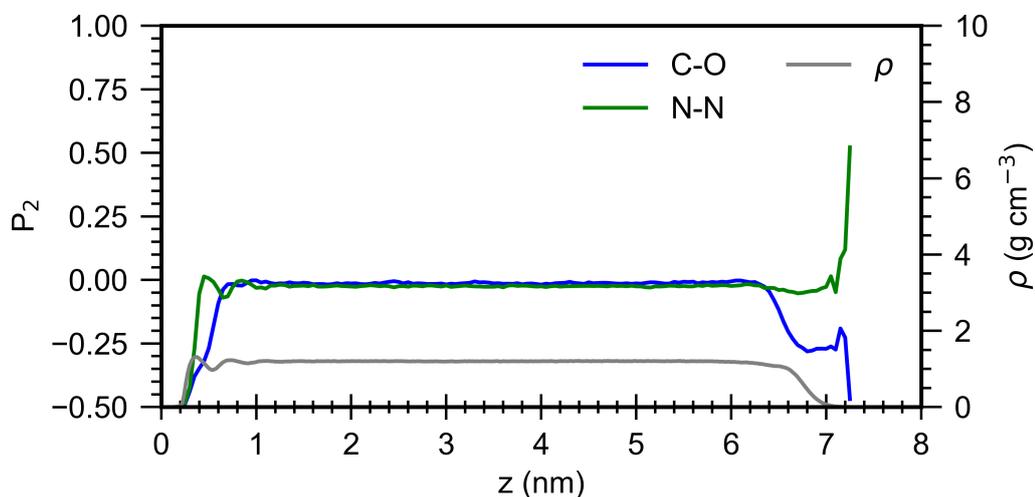


Figure 5.8:  $P_2$  bond orientations for  $1200 \text{ g kg}^{-1}$  medium  $\alpha$ -quartz system.

$P_2$  profile starts at -0.5, increases to just above zero and then settles around zero. The fluctuations are slightly stronger for the N–N than the C–O vector. The weak  $\alpha$ -quartz interactions led to ordering of the molecules immediately next to the interface, but was not strong enough to induce significant layering or concentration enhancement of urea molecules or lead to ordering in the molecules further away.

There is a slight difference between the interfaces in how close the urea molecules can get to the interface, which is related to the wall distance parameter  $\sigma_{ww}$ . The urea molecules are closest to the interface for the  $\alpha$ -quartz interface and furthest away for the PTFE interface.

It is worth noting, for all the systems, that the  $P_2$  value for the bulk part of the solution is slightly below zero. This indicates that the solution, even in the large systems, is not fully randomly orientated, even in the bulk solution. The same discrepancy was found when calculating the  $P_2$  values in the  $x$ ,  $y$  and  $z$  directions for the bulk solution systems studied in Chapter 4.

Towards the vacuum interface, for all the systems, in the region where the concentration of urea molecules is decreasing, the C–O  $P_2$  value decreased from the near zero bulk value, however, the N–N  $P_2$  value increased. This indicates that there are more molecules with the C–O bond parallel to the interface but that the N–N vector is perpendicular. Interestingly, these curves tend to invert themselves at the very edge of the solution, which indicates that at the vacuum the preferential orientation is likely to be for the N–N bond to be parallel and for the C–O bond to be perpendicular to the interface. It is worth noting that this part of the curve corresponds to very few urea molecules compared to the rest of the solution, which explains the discrepancy apparent in Figures 5.7 and 5.8.

### 5.2.2 Heterogeneous Nucleation Induction Times

The results of the nucleation induction time experiments for the different supersaturations and interfaces are summarised in Table 5.4. In some vials nucleation occurred

during the cooling ramp or whilst being moved from the Polar Bear and into the vial racks in the incubator and these are recorded as "Discounted". Vials where crystallisation occurred during the isothermal holding period are recorded as "Crystallised". There was also a significant proportion of vials where crystallisation did not occur, within the duration of the experiment. The overall probabilities of nucleation for all the systems tested are relatively low, with  $\leq 32.5\%$  of vials nucleating for all the systems, except the high concentration PTFE system with a probability of 62.5%.

Table 5.4: Summary of the nucleation time experiments performed for different supersaturations ( $S$ ) and interfaces. The total number of vials is shown along with the number discounted and crystallised, and the overall nucleation probability for each system.

$S$	Interface	Experiments	Discounted	Crystallised	Probability
1.5	Control	40	0	1	0.025
1.5	Diamond	40	0	4	0.100
1.5	PTFE	41	1	8	0.200
1.7	Control	40	0	0	0.000
1.7	Diamond	40	0	2	0.050
1.7	PTFE	45	5	8	0.200
1.9	Control	42	2	13	0.325
1.9	Diamond	43	3	12	0.300
1.9	PTFE	58	18	25	0.625

As expected, crystallisation was most probable for the systems with the higher supersaturations. There were similar proportions of vials nucleating at 1800 and 2040  $\text{g kg}^{-1}$ , but the probability of crystallisation is significantly higher at 2280  $\text{g kg}^{-1}$ , indicating a non-linear dependence of supersaturation on nucleation rate in these solutions.

Urea solutions are shear and agitation sensitive, and the movement of the vials during the preparation steps provided enough agitation to induce nucleation in some of these vials, this behaviour increased significantly with supersaturation. It is plausible that these discounted vials would have had some of the fastest nucleation rates, if they had been successfully transferred to the incubator without nucleating. Therefore, our experimental setup may have been unable to capture nucleation times of the vials with the greatest nucleation potential, possibly underestimating the nucleation rates of the affected systems. For this reason, we present conditional probabilities where the number

crystallised is calculated based only on the 40 vials that did not crystallise during the preparation.

The conditional cumulative probability distribution function of the induction times are presented in Figure 5.9, where the results are plotted separately for each concentration. Additional plots with the unconditional cumulative probability distribution are given in Appendix B.3.1. These plots also include the vials that were discounted due to their nucleation being associated with the cooling or moving stages of the setup, where all the discounted vials are taken to have nucleated at time zero. Although these unconditional plots reach slightly higher nucleation probabilities, it does not change the overall trends described above.

Given the low nucleation probabilities for 1800 and 2040 g kg<sup>-1</sup>, it would have been interesting to study higher supersaturations by changing the concentrations or temperatures used, however, this was not possible due to limitations with our experimental setup. The urea solutions were prepared at 60 °C, with additional heating of the final samples at 70 °C. The solutions were dissolved and pipetted at temperature in an incubator, which limited the maximum temperature for solution preparation to 60 °C. These temperatures should also avoid problems with the decomposition of urea at higher temperatures. The incubator which was used to keep the vials at a constant temperature, during the monitoring stage, had a minimum operating temperature of 25 °C, therefore, it was not possible to go to lower temperatures to achieve higher supersaturations.

### 5.2.3 Heterogeneous Nucleation Rates

The biexponential model has been fitted to all the experimental cumulative probability distributions discussed above, and the resulting parameters are given in Table 5.5. This allows a quantitative discussion of nucleation rates, since the timescales of nucleation ( $\tau$ ) is inversely proportional to the nucleation rate. In addition to these tabulated results, plots of these fitted parameters and the experimental nucleation times are available in Appendix B.3.2.

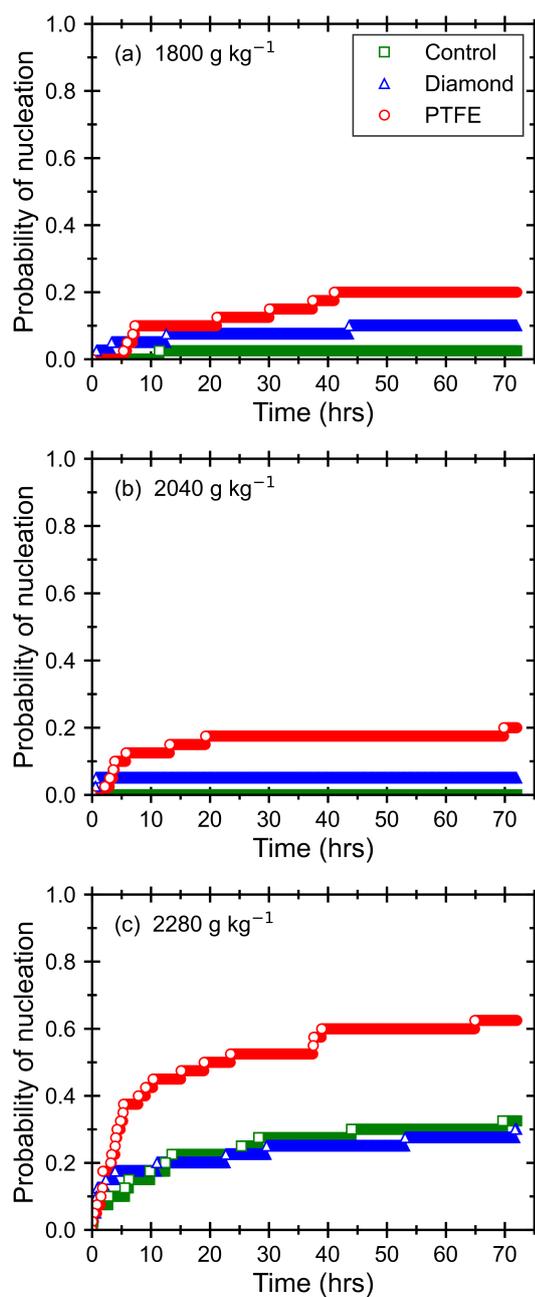


Figure 5.9: Cumulative probability distribution function of induction times of urea crystallisation from aqueous solution with heterogeneous interfaces PTFE, diamond and glass for (a) 1800 g kg<sup>-1</sup>, (b) 2040 g kg<sup>-1</sup> and (c) 2280 g kg<sup>-1</sup>.

This fitting procedure was not used for systems where less than five nucleation events occurred, to avoid over-fitting the data. Instead,  $A$  could be estimated as the fraction

Table 5.5: Fitted parameters for the biexponential nucleation model.

Interface	$c$ (g kg <sup>-1</sup> )	$\tau_1$ (hrs)	$\tau_2$ (hrs)	$A$	$R^2$
PTFE	1800	27.0 ± 3.0	$\infty$	0.23 ± 0.03	0.936
PTFE	2040	7.26 ± 0.15	11 000 ± 5000	0.1736 ± 0.0016	0.942
PTFE	2280	4.65 ± 0.07	167 ± 3	0.444 ± 0.002	0.979
Diamond	2280	1.44 ± 0.05	472 ± 6	0.1753 ± 0.0009	0.943
Control	2280	8.23 ± 0.19	600 ± 20	0.230 ± 0.003	0.963

of vials nucleated at the end of the experiment.  $\tau$  was taken to be infinity ( $\infty$ ), where the time scale was greater than  $10^8$  and the standard error values were significantly larger.  $t_d$  was taken to be zero, since the first observed nucleation event happened very quickly relative to the time scale of other events.  $t_d$  can also be included as an extra fitting parameter, this was tested and has no significant effect on the results. (Even for the low concentration PTFE system, where the first nucleation event seen on Figure 5.9 occurs after 5 hrs, the calculated  $t_d$  was less than 1 hr for all the models, indicating that with increased sample size nucleation is predicted to occur earlier to change the initial shape of the curve.)

Generally, the biexponential model provided a very good fit. The two  $\tau$  parameters describe a fast and a slower nucleation regime with different nucleation mechanisms. The fast nucleation regime, relating to  $\tau_1$ , is likely to represent heterogeneous nucleation and could be related to the enhanced local concentration near the surfaces. The slower nucleation regime, relating to  $\tau_2$ , could be a homogeneous nucleation process, or a slower heterogeneous process with different nucleation mechanisms.

The effect of the solution supersaturation on the nucleation rate followed a similar trend for each of the three interfaces. Considering the three PTFE systems, as the concentration was increased the fast nucleation time scale decreased by less than a factor of six overall, whereas the slower nucleation time scale decreased by several orders of magnitude between each concentration. The fraction of vials nucleating in the fast regime decreased slightly from the low to the medium concentration but increased significantly for the high concentration. The fast nucleation regime is expected to be linked to the increase in the local concentration. The slow nucleation time scale changed

significantly with solution concentration, it went from being insignificant for the low concentration to only differing to the fast regime by one to two orders of magnitude for the high concentration system.

The effects of the different interfaces can best be determined by considering the three systems at the high concentration. The PTFE system had a higher fraction of nucleation events in the fast regime, which was also shorter compared to the control system. The presence of the PTFE led to an increase of both the number and the frequency of nucleation events. Interestingly, diamond system had the lowest fraction of vials nucleating in the fast regime, but still had a faster nucleation rate than the control system. This may be affected by the low number of nucleation events, since it is counterintuitive that diamond would simultaneously speed up and reduce nucleation.

It is worth noting that the surface area of the diamonds used is smaller than the area of the PTFE bars used, and they are both significantly smaller than the surface area of the glass vials. The heterogeneous nucleation fraction ( $A$ ) from Table 5.5 has been divided by the surface area associated with the dominant surface, to try to account for the differences in surface area, this is shown in Table 5.6. Considering the effect of surface area it is clear that both diamond and PTFE are much more effective at inducing heterogeneous nucleation. The PTFE surface remains more effective than diamond, differing by approximately a factor of two. Therefore, the influence of the change in the wetted glass surface area, when a PTFE bar or a diamond is added to a vial, can be assumed to have a negligible effect on nucleation rate.

Table 5.6: Comparison between the interfacial surface area and the associated nucleation fraction.

Interface	$c$ (g kg <sup>-1</sup> )	$area$ (mm <sup>2</sup> )	$A$	$\frac{A}{area}$	$\frac{A}{area} \div \left(\frac{A}{area}\right)_{control}$
PTFE	1800	44	0.23	0.0052	-
PTFE	2040	44	0.17	0.0039	-
PTFE	2280	44	0.44	0.010	21
Diamond	2280	35*	0.18	0.0051	11
Control	2280	490**	0.23	0.00047	1

\* The largest area for the diamonds are used to avoid overestimating the effect.

\*\* For simplicity, only the glass surface area is considered for the control system.

#### **5.2.4 Effect of Individual Nucleants**

There have been many discussions about whether heterogeneous nucleation is due to impurities and imperfections present in the solution or interface or if it is due to significant interactions between specific surfaces and the solution. Therefore, we include a discussion of the different known interfaces in our experiments and the expected variability within each of these.

New glass vials were used for each experiment, these were taken straight from the packaging they were supplied in. There was some variability between each vial but it is assumed to be as small as possible, and would have been the same between each vial. Greater variability is expected to have arisen, if the vials were reused and cleaned.

Most of the PTFE experiments, for each concentration, were run simultaneously. This required a large number of PTFE stirrer bars, therefore, the effect of the differences between individual PTFE stirrer bars is not discernible. The PTFE stirrers used were of a variety of ages, some had been used in previous nucleation experiments, whilst others were new. All the PTFE stirrers were cleaned and then visually inspected by eye at the beginning and there were no obvious defects.

The diamonds that were used, all had their own unique imperfections, containing different cracks and a small core of encased mixed chemicals, which was not able to be characterised. The diamonds were extensively cleaned prior to use and it was assumed that the solids contained within the core did not escape into the solutions they were immersed in. The number of diamond experiments that led to nucleation was low, and since the same diamonds were repeatedly used many times it is relevant to confirm whether or not the nucleation is related to individual diamonds.

Across all the concentrations, the individual diamonds were tracked for 51 out of 120 experiments. Overall, nucleation took place in 18 of the 120 diamond experiments and only 7 of these were tracked with regards to identifying the individual diamonds involved. There is no evidence of any diamond being a super-nucleant by consistently inducing nucleation in repeated runs. Therefore, we assume that the nucleation is not

due to specific defects or impurities related to an individual diamond. A diamond-by-diamond overview of nucleation is given in Table 5.7 and a more detailed breakdown is available in Appendix B.3.3.

Table 5.7: Breakdown of the individual diamonds and how frequently they induced nucleation. A-F represent the six individual diamonds, X represents the experiments where the diamonds were not individually tracked.

Diamond	Nucleation	No nucleation	% nucleated
A	1	8	11.1
B	1	8	11.1
C	0	9	0.0
D	2	7	22.2
E	2	6	25.0
F	1	6	14.3
X	11	58	15.9
Total	18	102	15.0

### 5.2.5 Combined Experimental and Computational Insights

The experiments show that the PTFE surfaces significantly increased the nucleation rate in aqueous urea solutions compared to the control system with only glass and air surfaces present. The diamond surfaces increased the overall nucleation probability at shorter times and lower urea concentrations. These results align well with the findings from the simulations of urea solutions at these interfaces. The simulated PTFE surface induced strong concentration enhancement (32-59%) of the urea molecules in the nearby solution, this was the same with the diamond surface but to a lesser extent (13-29%), however, the much weaker  $\alpha$ -quartz (glass) and vacuum (air) surfaces did not lead to this effect, as can be seen in Table 5.3. The increased local concentration near the PTFE and the diamond surfaces can help to explain the increased experimental nucleation rates.

The experimental results indicate that heterogeneous nucleation happens at nucleation sites that are specific to each individual nucleation system. The additional PTFE surfaces appear to have a significantly higher probability of possessing an effective nucleation site than the glass and air surfaces also present and even in relation to the diamond

surface. There was also a significant increase in nucleation with increased concentration, so solution impurities can not be excluded as possible nucleation sites, this is discussed later. The simulations indicated that the interactions between the solid surfaces and the solution led to local concentration enhancement of urea, which improves the nucleation kinetics. We conclude that the interactions (or lack of interactions as is the case with air) between surfaces and solution will affect local solution concentrations. Strong interactions between the surface and the solute will lead to local interfacial concentration enhancement, which results in favourable nucleation conditions.

However, the enhanced concentration is not sufficient to significantly increase homogeneous nucleation rates in aqueous urea solutions. The nucleation sites associated with the surfaces also have an important influence on the nucleation rate. It is possible that not all nucleation sites act in the same way and that some sites require higher local concentrations to be effective compared to other sites. No nucleation events were observed in the simulations, this is expected because of the short time and size scales of the simulations. It would be interesting to perform similar tests with imperfect interfaces to further study the effects of these nucleation sites.

Experimentally, supersaturated solutions are in a pseudo-equilibrium without a net change in the chemical potential of the solution to act as a driving force for nucleation. The simulated concentration gradients across the interfacial systems are also in equilibrium across system as is the chemical potential, and the molecules move freely between the inner and outer interfacial layers and the bulk solution. However, the increased local urea concentration improves the nucleation kinetics, and interactions with the interface further aid nucleation by inducing local molecular ordering and by reducing the surface energy of any nuclei formed. Our combined results indicate that heterogeneous surfaces enhance nucleation by reducing the surface energy and by improving the kinetics for nucleation by enhancing the concentration in the region near the interface.

It is interesting to try to make direct comparisons between the results from the experiments and simulations, despite their different scales and reimits. The experimental nucleation probability for the low (and medium) concentration PTFE solution is only

slightly lower compared to the control system at the high concentration. However, the predicted interfacial supersaturation in the lower concentration PTFE solutions is significantly higher than the supersaturation in the high concentration control solutions. This could indicate that the interactions between the solution and each interface is different between the experiments and the simulations. However, it could also be another indication that the nucleation behaviour is not simply linked linearly to supersaturation.

Heterogeneous nucleation is often considered to be caused by solution impurities. Our results indicate that interactions with the significant surfaces present also have a significant role in enhancing heterogeneous nucleation rates. We added clean interfaces into the solution and observed significantly enhanced nucleation rates and in addition to this the MD simulations illustrate how these surfaces make the conditions for nucleation more favourable. Therefore, we do not think that the increased nucleation observed when introducing additional interfaces is only related to an increase in the number of impurities in the system in the traditional sense.

The relative effect of the additional interfaces on nucleation, compared to the control system, depends on the solution concentration. The increase in nucleation probability is most significant at the lower concentrations, which indicates that this effect is not simply caused by increased numbers of impurities. Assuming each additional interface contributes the same number of impurities to each vial, then even higher nucleation probabilities would be expected for the high concentration systems, which may also have more impurities originating from the solution.

The impurities originating from the solution are also important to consider. Urea has a very high solubility in water of  $1200 \text{ g kg}^{-1}$  at  $25 \text{ }^\circ\text{C}$ ,<sup>64</sup> therefore, for a fixed volume of solution, the concentration significantly affects the masses of urea and water required. In the simulations, both the number of urea and water molecules were adjusted, to obtain the same volume of solution at different concentrations. Similarly, in the experiments the mass of water in each vial is significantly less at the high concentration compared to the low concentration, the opposite applies to the mass of urea. This means that

experimentally increasing the urea concentration increases the mass of urea in each vial and reduces the mass of water. Therefore, as the highest concentration favoured the highest nucleation rates, this could indicate that more impurities are coming from urea than water.

It is interesting that our simulations at the air (vacuum) interface indicates that air surfaces are likely to reduce the probability of surface enhanced heterogeneous nucleation. Air interfaces are typically associated with nucleation in other applications such as in microfluidic experiments and in ‘gassing crystallisation’ where gas bubbles (such as air) are added to increase the nucleation rate.<sup>40,42</sup> In microfluidic experiments nucleation is attributed to the ongoing evaporation, and in addition to this ‘gassing crystallisation’ also has several other potential mechanisms involved such as agitation and changes in temperature and pH. Our results indicate that it is entirely these secondary factors that are significant in influencing the nucleation processes, rather than the direct presence of the air interface.

If the results from our simulations also apply to air bubbles in solution and there are sufficiently many bubbles, which reduce the interfacial solute concentration, then this will lead to the solute favouring a smaller volume of the solution, thus increasing the local supersaturation there. Solute depletion at the bubble surfaces will only be noticeable if the solute molecules diffuse faster than the gas bubbles move around. It is possible that the effect of a vacuum (or air) interface varies significantly for different solute and solvent combinations.

It is satisfying that the relative performance of the experimental nucleation rates in the different systems studied aligns with the simulated interfacial enhancement for all of the surfaces tested. This indicates that with further development and testing, these simple simulations have the potential to be used as screening tools to predict the relative heterogeneous nucleation behaviour of different systems.

### 5.3 Conclusions

We have performed experiments to determine the heterogeneous nucleation induction time of aqueous urea solutions in three different systems: a control system consisting of a glass vial with air above the solution, and systems with added PTFE or diamond interfaces. We have also performed MD simulations of solutions in contact with the above surfaces, to determine the molecular effects of these surfaces on the solution.

The MD simulations made use of specifically parametrised Lennard-Jones 9-3 potentials to model the interfaces, which only represent the dispersion forces. The interactions between the PTFE and the diamond surfaces and the solution, led to a significantly increased concentration of urea and decreased concentration of water immediately next to the interface.  $\alpha$ -quartz was used instead of glass, this had a much weaker potential than the other surfaces, and as such did not significantly affect the concentration in the surrounding solution. Vacuum interfaces, which are very similar to air in these conditions, were present in all the simulations, these led to a significant reduction in the nearby concentration of urea. Both the LJ and vacuum interfaces led to increased ordering of the urea molecules in the interfacial regions. The  $P_2$  bond order parameter indicated that the urea molecules preferentially lie flat at the LJ interface with both C–O and N–N vectors parallel to the interface. In the region of decreased concentration near the vacuum the preferential orientation is for the C–O to be parallel and the N–N to be perpendicular to the surface, however, this is inverted at the vacuum edge. This highlights the importance of not assuming that bulk properties apply near surfaces. As expected, no nucleation events were observed in the MD simulations, this is due to the very short simulation times and small system sizes resulting in a very low probability for nucleation.

The experimentally observed nucleation rates were relatively low with many vials not nucleating during the hold time, despite the high supersaturations used (1.5, 1.7 and 1.9) and long experiment durations (3 days). The nucleation rate was significantly increased by the addition of the PTFE interface compared to the control system. However, while the addition of the diamond interface increased the initial nucleation rate, the

overall nucleation probability was not significantly increased. The nucleation rates were significantly higher for the high supersaturation compared to the low and medium, this is likely to be due to improved nucleation kinetics related to the higher concentration.

The glass ( $\alpha$ -quartz) systems show only weak interfacial effects with slow nucleation, indicating classical heterogeneous nucleation is taking place. The PTFE surfaces cause strong interfacial effects leading to fast nucleation rates, indicating that there is a form of kinetically enhanced heterogeneous nucleation. The diamond surfaces show simulated interfacial effects but have less significant experimental impact, this indicates that there may be a barrier for the kinetically enhanced nucleation to become significant.

The same overall result was obtained from the experiments and the simulations, since both identified low heterogeneous nucleation related to glass and air surfaces and significantly increased nucleation at PTFE surfaces. These results show that laboratory scale experiments and atomistic scale simulations can be used together to improve our understanding of complex phenomena, using heterogeneous nucleation as our example.

## Chapter 6

# Temperature Cycling and Urea Nucleation

Nucleation is a stochastic event, so representative nucleation rate data can only be obtained from having sampled a sufficiently large amount of solution, which often requires many individual samples of small volumes. The results from Chapter 5 showed that it may be challenging to identify the suitable conditions needed to obtain good nucleation statistics, since high supersaturations do not ensure a high nucleation probability in unagitated systems. Therefore, in this Chapter, we further explore the crystallisation phase diagram for unagitated urea solutions, with the aim of identifying conditions with a high nucleation probability, which will allow more reliable estimation of nucleation rates, without requiring hundreds of samples. The ability to efficiently perform nucleation experiments with many different surface materials will be invaluable for extending the validation of the interfacial simulation method applied in Chapter 5. This could help to develop a predictive tool for heterogeneous nucleation, by efficiently allowing more materials to be compared using both approaches.

Here, we develop an experimental protocol to increase the throughput and generate useful data by using automated equipment that performs consecutive nucleation experiments in the same vials using temperature cycling over a wide range of crystallisation

temperatures. While simple temperature cycling is a commonly used approach in crystallisation experiments, to minimise time and material requirements, it has not been applied to unagitated experiments, which presents some unique challenges in comparison with agitated systems.

The Crystal16 equipment is used which uses changes in transmissivity (or turbidity) as the indicator for nucleation. The nature and location of the transmissivity measurements usually requires nucleation experiments to be agitated to accurately allow the crystals to be detected at the time of nucleation. However, urea has fast growth and secondary nucleation rates, with needle crystals which quickly spread throughout the whole vial, therefore, it should be possible to achieve accurate nucleation data for unagitated experiments.

This work is designed to explore the feasibility of adapting current automated methodologies to study unagitated nucleation. In Chapter 5, the presence of stirrer bars have been shown to induce heterogeneous nucleation, therefore, unagitated experiments are needed to study homogeneous nucleation and the independent influence of other heterogeneous nucleants. Firstly, a range of temperatures are studied to investigate the urea crystallisation phase diagram and the optimal conditions for obtaining higher overall nucleation probabilities, with faster nucleation rates and shorter experimental times are identified. These optimal conditions are used in further experiments to obtain a larger set of nucleation induction times, which is used to study the variability within vials and between cycles.

## **6.1 Methodology**

### **6.1.1 Experimental Setup and Procedures**

To enable lower temperatures to be tested, compared to the setup used in Chapter 5, the Crystal16 equipment was used. Urea stock solutions were prepared at the desired concentration for each experiment. The concentrations of 1800 and 2280 g kg<sup>-1</sup> were selected based on the low and high concentrations of the previous experiments. Urea

powder (Sigma-Aldrich, for electrophoresis  $\geq 98\%$ ) and ultra pure water (Fluka Analytical, LC-MS CHROMASOLV) were used. The required amount of urea and water was weighed into a glass bottle, which was then sealed. The bottle was placed in a water bath, at 60 °C, and left until the urea had fully dissolved. No stirrer bar was used, instead the bottle was gently shaken from time to time to aid dissolution.

16 vials were prepared for each experiment using the same stock solution. New, 1.5 ml glass vials (VWR 548-0018) were used for all the experiments, without any pre-cleaning. 1.0 ml of stock solution was pipetted from the bottle into each vial, and the lids were used to seal the vials. New pipette tips, preheated to 50 °C, were used for each vial. The vials were not preheated and were kept in a metal vial holder at room temperature. Once all the vials had been filled and sealed, each individual vial was weighed. In addition to lids the vials in Exp D–F were sealed with a layer of Parafilm to further reduce any evaporation during the experiment. The vials were then transferred to the Crystal16, which was at ambient temperature, and the pre-programmed temperature cycles were started. At the end of the temperature cycles the vials were removed and reweighed.

### 6.1.2 Experimental Conditions

Six sets of experiments were performed, which all consisted of repeated temperature cycles to maximise the experimental throughput. The temperature cycles consisted of a heating ramp and hot hold period ( $\tau_H$ ) to allow the solution to dissolve fully, followed by a cooling ramp to the desired crystallisation temperature ( $T_C$ ) and a hold period ( $\tau_C$ ) to study crystal nucleation, before the cycle was restarted. Figure 6.1 illustrates the different stages of the temperature cycles, at a given  $T_C$ .

The first three experiments were designed to explore the temperature effects on the crystallisation phase diagram, cycling across a wide range of crystallisation temperatures. The last three experiments were designed with longer  $\tau_C$  hold periods to obtain nucleation times at the three best performing  $T_C$  values. Table 6.1 give an overview of the variable experimental conditions. The hot hold temperature  $T_H = 70$  °C, and the

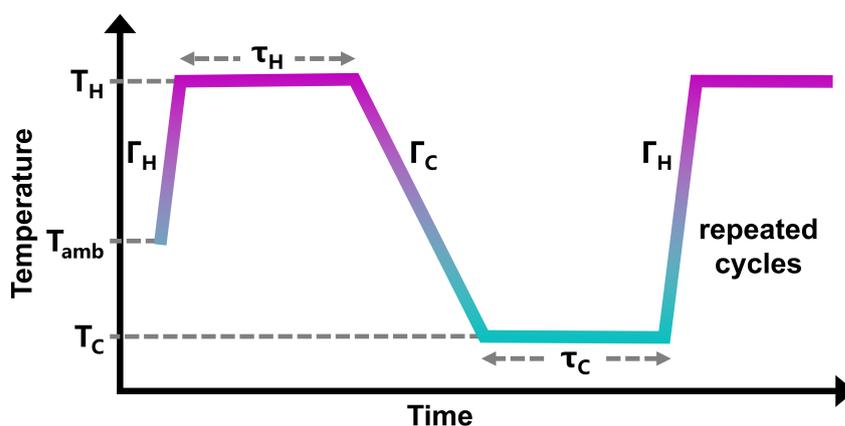


Figure 6.1: Schematic showing the stages involved with the temperature cycling which are repeated sequentially, at a given  $T_C$ .

heating and cooling rates  $\Gamma_H = 10.0 \text{ }^\circ\text{C min}^{-1}$  and  $\Gamma_C = -1.5 \text{ }^\circ\text{C min}^{-1}$  respectively, were constant for all the experiments.

Table 6.1: Experimental conditions. Replicates are the number of repeats at each  $T_C$ , and Cycles is the total number of temperature cycles that were performed across all the replicates.

Exp.	$c \text{ (g kg}^{-1}\text{)}$	$\tau_H$	$\tau_C$	$T_C \text{ (}^\circ\text{C)}$	Replicates	Cycles
A	2280	4 hrs unstirred	1 hr	-5, 0, 5, 10, 15, 20, 25	3 (4 for 15 $^\circ\text{C}$ )	22
B	2280	1 hr stirred	1 hr	-5, 0, 5, 10, 15, 20	3	18
C	1800	4 hrs unstirred	1 hr	-5, 0, 5, 10, 15, 20, 25	3	21
D	2280	1 hr stirred	4 hrs	15	13	13
E	2280	1 hr stirred	4 hrs	20	15	15
F	2280	1 hr stirred	4 hrs	25	15	15

In experiments A–C, three identical heating and cooling cycles were performed with  $T_C = -5 \text{ }^\circ\text{C}$ , then three cycles are performed with  $T_C = 0 \text{ }^\circ\text{C}$ , this continued until all the experiments outlined in Table 6.1 have been performed. Experiments D–F are all performed with a single  $T_C$  value, therefore, all the temperature cycles within each experiment are identical. The overall duration of the experiments were similar to those in Chapter 5, despite the many cycles that are performed here, since each cycle is relatively short.

In some of the experiments, stirring was used during the heating ramp and hot hold

period to aid dissolution, the reason for this is discussed in Section 6.2.1. In the experiments with stirred heating, PTFE stirrer bars ( $7 \times 2$  mm) were placed in the empty vials before the solution was pipetted in. Stirrer bars were not added to the vial if no stirring was performed. In Table 6.1 the  $\tau_H$  column also distinguishes whether the experiments were stirred or unstirred, since the stirring only applied to this phase of the temperature cycle. All of the nucleation stages were performed under unstirred, quiescent conditions, however, some of the experiments had an additional static heterogeneous PTFE stirrer in the vial.

The temperature range of  $-5$ – $25$  °C was selected with  $5$  °C increments.  $-5$  °C was selected as the lowest temperature, since freezing point depression prevents the water in aqueous urea solutions from freezing at this temperature.<sup>187</sup>  $25$  °C was selected as the highest temperature, to allow comparison to the experiments in Chapter 5, from which it is known that the nucleation rate is low. The solubility concentrations and supersaturations for all the conditions studied are given in Table 6.2.

Table 6.2: Solubility and supersaturation across the range of experimental conditions.

Temperature (°C)	-5	0	5	10	15	20	25
Solubility <sup>65</sup> (g kg <sup>-1</sup> )	610	680	760	850	950	1100	1200
S <sub>1800</sub> g/kg	3.0	2.6	2.4	2.1	1.9	1.6	1.5
S <sub>2280</sub> g/kg	3.7	3.4	3.0	2.7	2.4	2.1	1.9

## 6.2 Results and Discussion

In this section, we first discuss the influence of temperature and repeated cycling on primary urea nucleation behaviour. We identify conditions suitable for performing unagitated induction time measurements and present further results based on this. We discuss the overall nucleation behaviour and vial-to-vial variability including rate data obtained using the biexponential nucleation model.

### 6.2.1 Variable Temperature Cycles

Experiments A–C investigated the nucleation behaviour of urea at a range of temperatures. Nucleation often occurred during the cooling ramp due to the low temperatures and associated high supersaturations tested. The variation in nucleation temperature with cycle is given in Figure 6.2, for simplicity all four vials from each zone have the same colour, which means that these cannot be distinguished from each other. Additional figures subdividing each experiment into four graphs (one for each zone) are available in Appendix C.1. The fraction of nucleation events that occurred during the cooling ramp, isothermal hold and vials that never nucleated are shown in Figure 6.3. Overall, there was a low number of vials nucleating during the isothermal hold, therefore, it was not meaningful to analyse the induction times for these initial experiments.

There is a range of nucleation temperatures across the three experiments, which is related to the variability between vials and stochasticity of nucleation phenomena. The temperature range is broader for the highly concentrated solutions compared to the less concentrated solution. The range of nucleation temperatures is higher for the solutions with the PTFE bars present, due to the enhanced heterogeneous nucleation these induce. The cycle number also has an effect on the nucleation temperature particularly across the first six or so cycles, where there was a steady shift in the nucleation temperature, which can be seen in Figure 6.2. In Experiment A, the nucleation temperatures steadily dropped for the first six cycles to reach a constant level. Conversely, in Experiment C, the nucleation temperatures steadily increased to reach a constant level. While there seems to be a continuing gradual variation of nucleation temperatures across all the cycles in Experiment B. The reasons for this behaviour is unknown, but it is one of many examples of the complicated influences of thermal history on nucleation.<sup>26,137,138,140</sup> The decrease in nucleation temperature could be an indication that the solution became more stabilised as the experiment progressed, whereas the increase in nucleation temperature could be an indication of a problem with incomplete dissolution. In all cases, when high enough temperatures were reached, the nucleation started to change from occurring during the cooling ramp to the isothermal hold period.

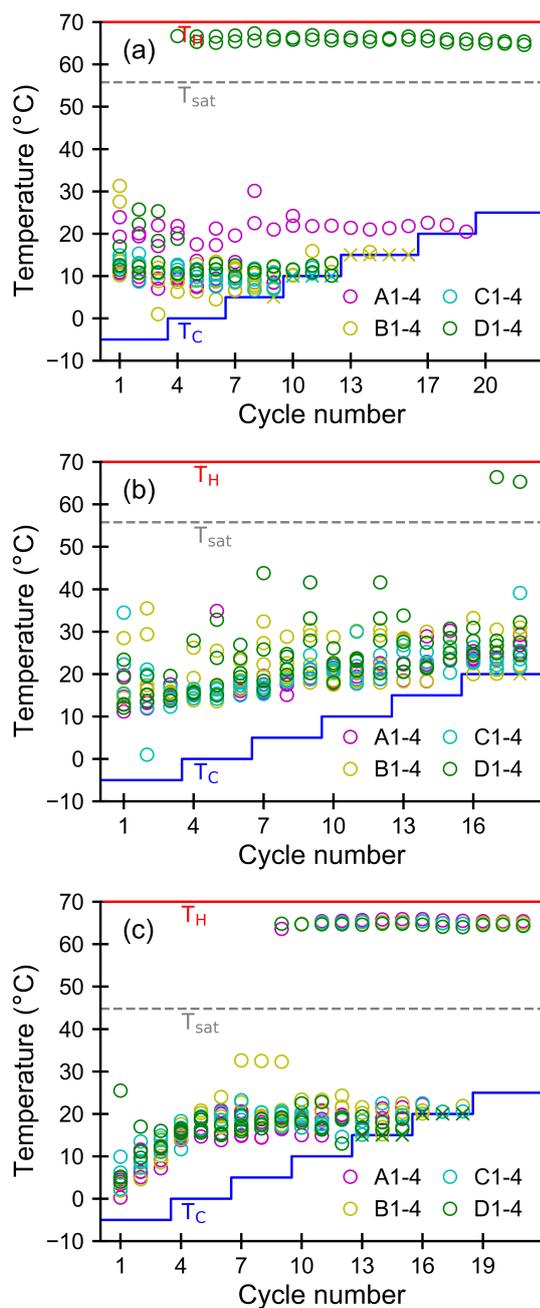


Figure 6.2: Nucleation conditions and outcomes for the variable temperature cycles in Experiments (a) A ( $2280 \text{ g kg}^{-1}$  without stirred heating), (b) B ( $2280 \text{ g kg}^{-1}$  with stirred heating), and (c) C ( $1800 \text{ g kg}^{-1}$  without stirred heating). Circles represent nucleation during the cooling ramp, crosses represent nucleation during the hold period, and no symbol indicates no nucleation, this can be seen more clearly in Figures C.1-C.3 in Appendix C.1.

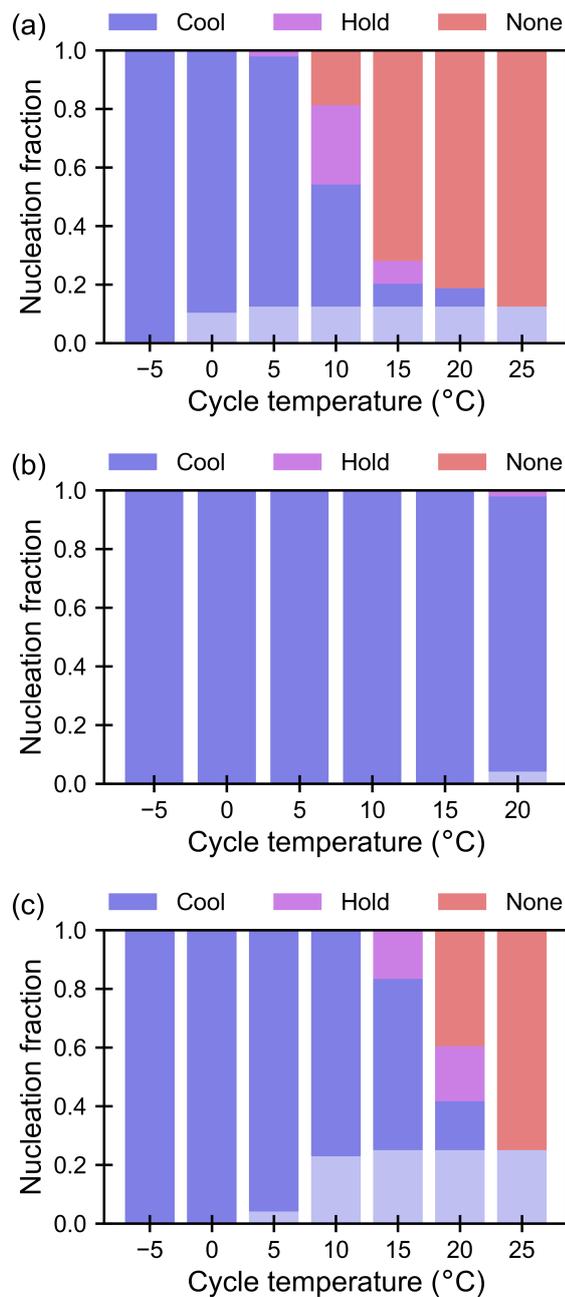


Figure 6.3: Nucleation temperature and outcomes for the variable temperature cycles in Experiments (a) A (2280 g kg<sup>-1</sup> without stirred heating), (b) B (2280 g kg<sup>-1</sup> with stirred heating), and (c) C (1800 g kg<sup>-1</sup> without stirred heating). The lighter blue indicates where crystallisation occurred above the saturation temperature.

In Experiment B there was frequently nucleation happening at or above 25 °C, which was less common for Experiments A and C. This aligns with the previous experiments,

in Chapter 5, where a significant proportion of the vials, with conditions corresponding to Experiment B, nucleated during the cooling ramp in the Polar Bear or on the transfer to the incubator. This shows the effect the presence of the stirrer bar has on the probability of nucleation in this aqueous urea system.

A few of the vials appeared to nucleate at very high temperatures, above the saturation temperature ( $T_{\text{sat}}$ ), which is indeed unexpected. This occurred for 12.5%, 6.25% and 25% of the vials in Experiments A, B and C respectively. Once a vial appeared to nucleate at a high temperature this continued until the end of the experiment. We consider that this behaviour may be related to incomplete dissolution of previously formed crystals. This anomalous behaviour may also have been influenced by solvent evaporation during the experiment, which could influence the relative concentration in the vials, this is discussed later in relation to Figure 6.4. It is also possible that the solubility (and so saturation conditions) may have been influenced by impurities in the solution. The anomalous behaviour was more frequent in Experiments A and C, where the hot hold was four times longer for in Experiment B. Therefore, there may have been some relationship to possible urea decomposition (which generated impurities), which links to the problem developing during the experiment rather than being present from the beginning.

Manual checks established that it is possible for some crystals to remain at the bottom of the vial without being detected by the transmissivity sensor, which is located part-way up the vial. These undissolved crystals would then have been eventually suspended and detected to give rise to what was observed as apparent nucleation. The apparent nucleation at very high temperatures was much more significant in the unstirred experiments, regardless of the solution concentration, and despite the long duration of the hot holds. This could be related to problems with the heat transfer between the equipment and the vials and solution, and agitation will have improved the heat transfer and dispersion during the heating stage. Therefore, we conclude that it is not practical to perform experiments without using stirring to aid dissolution and that it is not possible to use this experimental methodology to compare the effects of homogeneous

and heterogeneous nucleation.

Between 18 and 22 short cycles were performed for Experiments A–C, this is significantly more than what is often done to avoid problems with evaporation and crowning.<sup>141–148</sup> No sign of crowning was visible when examining the uncrystallised vials at the end of the experiment, therefore, this was not thought to be a significant problem. Undissolved crystals in vials were easily identified, due to the tendency for nucleation to occur at the high temperatures above  $T_{\text{sat}}$ , which allowed these vials to be discounted from further analysis. Evaporation was found to be significant for some vials but not all, the evaporation will have led to the concentration of the solution increasing slightly with each cycle. Still, the end nucleation temperatures correspond to supersaturated conditions, based the estimated final solution concentration, with the exception of the vials that were discounted due to dissolution problems.

It is also possible that the results at the end of the cycles were influenced by the thermal history, whether this be the changes in cycles or cooling to temperatures significantly below the crystallisation point, although not enough experiments were performed to study this closely.

The change in mass of the vials from the beginning to the end of the experiment was measured, and it was found to be negligible for most of the vials. However, there was significant mass loss for some of the vials, this is assumed to be entirely due to the evaporation of water, which will have increased the solution concentrations. The changes in mass are illustrated in Figure 6.4 for Experiments A–C, and tabulated values for all the experiments are available in Appendix C.2. There was no clear relationship between mass loss and the tendency for vials to nucleate above the saturation temperature. Furthermore, even when the changes in solution concentration relating to the observed mass changes are considered, then the high temperature nucleation events still correspond to understaturated conditions.

These initial experiments indicate that there is quite a narrow set of temperature conditions to obtain reasonably short nucleation times whilst performing cooling crystallisation experiments in quiescent aqueous urea solutions. At or below 15 °C, almost all the

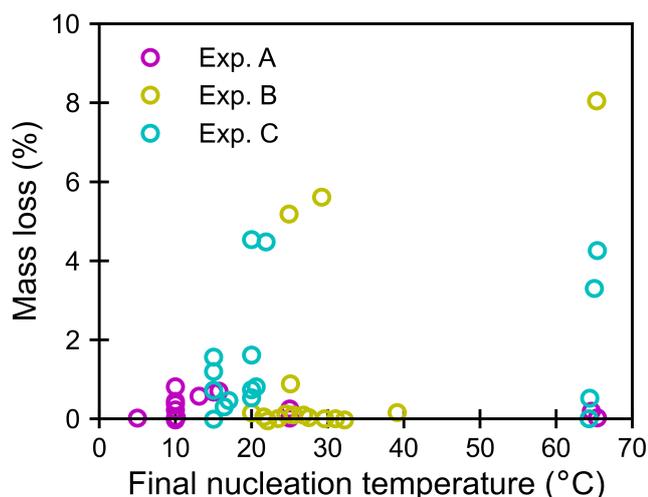


Figure 6.4: Mass loss in comparison to the final nucleation temperature of each vial in Experiments A–C.

nucleation occurs during the cooling ramp, the cycles at 15 and 20 °C have a combination of nucleation during the cooling ramp and during the isothermal hold period, and nucleation was very limited at 25 °C. A moderate cooling rate of  $\Gamma_C = -1.5 \text{ }^\circ\text{C min}^{-1}$  was used for all of the experiments, which is consistent with the experiments in Chapter 5. It is expected that a higher cooling rate is likely to have increased nucleation due to crash cooling, and whilst a lower cooling rate might have reduced this it would also increase the time before isothermal nucleation measurements were possible.

Our results indicate that nucleation is dependent on the interplay of temperature and concentration and is not simply dependent on supersaturation. For example, Experiment A at 25 °C and Experiment C at 15 °C reach the same supersaturation, however only in Experiment C does nucleation occur during the cooling ramp.

### 6.2.2 Repeated Temperature Cycles

Experiments D–F investigated nucleation across repeated, identical temperature cycles. Many repeated cycles were performed to obtain a large data set. The number of cycles and  $\tau_H$  was reduced, compared to the initial experiments, to reduce evaporation. Experiments D, E and F had  $T_C = 15, 20$  and  $25 \text{ }^\circ\text{C}$  respectively and  $\tau_C = 4 \text{ hrs}$ , which

was extended compared the initial experiments to obtain longer nucleation induction time data.

Profiles of the nucleation temperature with cycle number are given in Figure 6.5, with the fractional nucleation behaviours overlaid at the top. Nucleation during the cooling ramp was significant at 15 and 20 °C, which was expected from the initial experiments. There was only one instance of nucleation during the cooling ramp for the 25 °C experiment, which was less than expected based on the results in Experiments A–C and in Chapter 5. There was no obvious shift in the range of nucleation temperatures as the experiments proceeded. This could be related to starting at higher temperatures compared to the initial experiments, and that there were more cycles with identical temperature profiles to compare. Figure 6.5 shows that there is no correlation between the cycle number and nucleation outcome, which is good since this indicates that thermal history does not influence the results in these experiments. Therefore, up to 15 repeated temperature cycles can reliably be used to perform nucleation experiments for aqueous urea solutions in these conditions.

The nucleation induction times are given in Figure 6.6, where the variability between vials can be determined. Generally, the results appear to be stochastic with the variability in nucleation times spread across the vials. However, there were some specific vials which had a tendency to mainly exhibit either short or long nucleation times. For example, at 15 °C 11 nucleation events occurred between 0.5–4.0 hrs, but these were split between only 5 vials out of a total of 16 vials. Similarly, at 25 °C 21 nucleation events occurred in total, this was spread across only 9 of the 16 vials, the number of nucleation events in each vial ranged from 0 to 4. The largest number of induction time measurements were obtained at 20 °C, with an average of almost 11 nucleation events per vial, but interestingly nucleation never occurred in the vial in position C2. This highlights the importance of having many separate samples to take account of vial-to-vial variability, even if the experimental throughput is scaled by recycling samples.

The vial in position C2 was one of the most slowly nucleating vials in Experiment D and no nucleation occurs in the vial placed in this position in Experiment E. However,

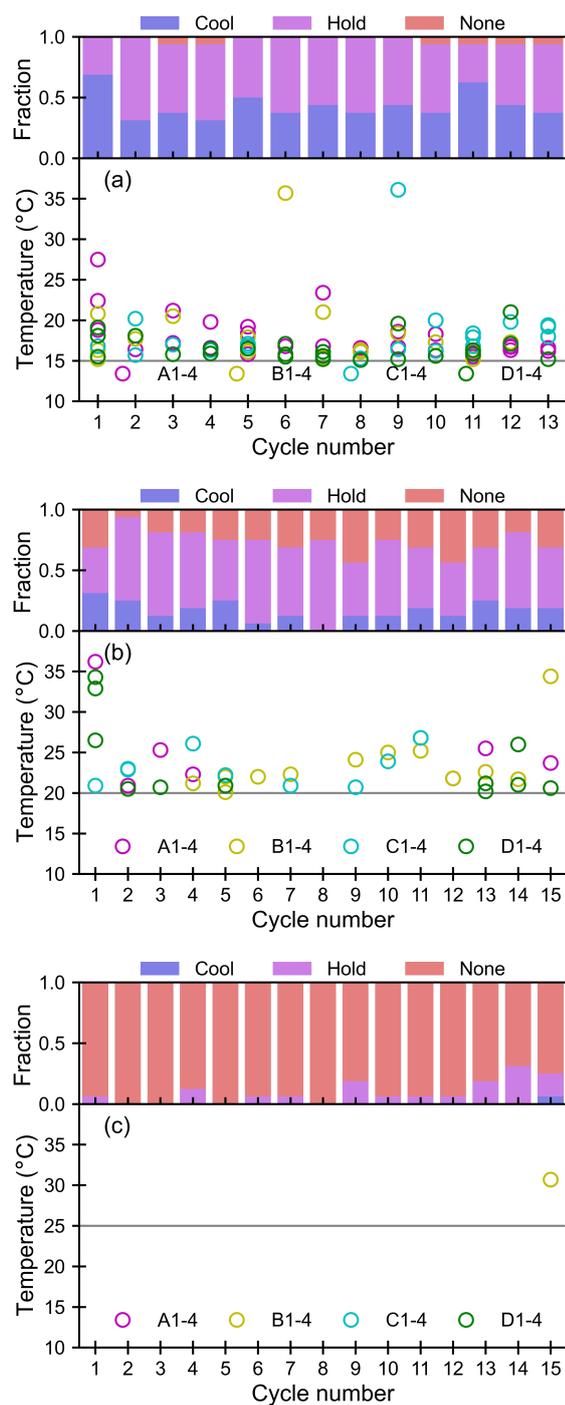


Figure 6.5: Nucleation fractions and temperatures for the repeated cycles in Experiments (a) D (15 °C), (b) E (20 °C), and (c) F (25 °C). The grey line gives the minimum temperature for each cycle.

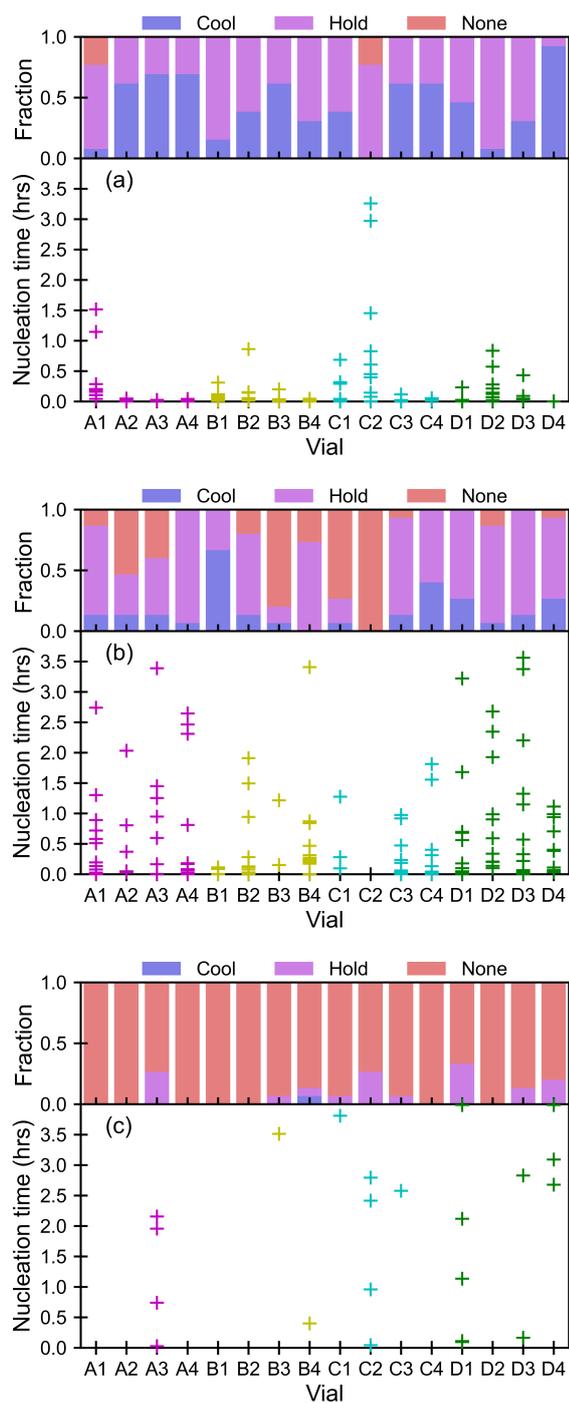


Figure 6.6: Nucleation fractions and times within each vial in Experiments (a) D (15 °C), (b) E (20 °C), and (c) F (25 °C).

there is no correlation between these two results since the vials are independent and Experiments D and E were even performed concurrently using two different Crystal16 machines. Furthermore, in Experiment F the C2 position contained one of the most nucleating vials.

### 6.2.3 Induction Times

The conditional cumulative probability distribution function of the induction times are presented in Figure 6.7. Only the nucleation events that occurred during the isothermal hold are considered, and all the nucleation events that occurred during the cooling ramp have been removed from the dataset. An additional plot with the unconditional cumulative probability distribution, including the samples that nucleated during the cooling stage, is given in Figure C.4 in Appendix C.3.

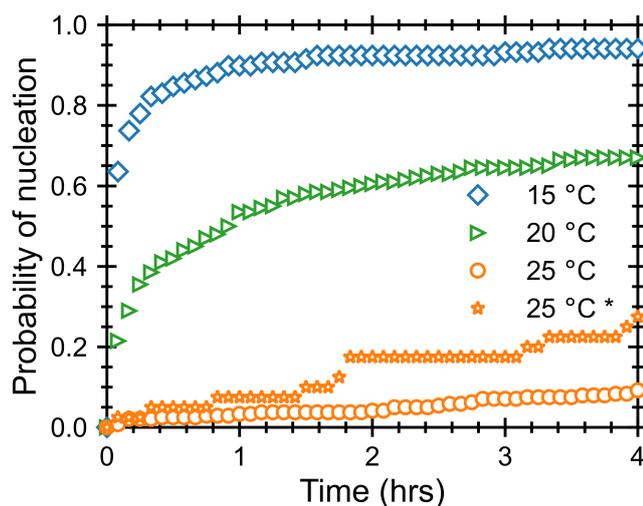


Figure 6.7: Cumulative probability distribution of the nucleation induction times for Experiments D (15 °C), E (20 °C), F (25 °C) and the equivalent results from Chapter 5 denoted by \* (25 °C, 2280 g kg<sup>-1</sup>, PTFE). For clarity, the symbols are only plotted every five minutes, however, for Experiments D–F measurements were made every second.

The nucleation rate increases significantly with decreasing temperature and the associated increasing supersaturation. At 15 °C, the initial probability of nucleation increases very fast and the probability curve reaches above 90% within the 4 hr period. At 20 °C,

the probability of nucleation still increases fast and reaches around 70% at the end of the experiment. The unconditional probability distributions at 15 and 20 °C are very similar to these conditional distributions, as can be seen in Appendix C.3.

The nucleation probability is very low at 25 °C, but is steadily increasing over the 4 hr period. The 2280 g kg<sup>-1</sup> PTFE probability curve from Chapter 5 is also included in Figure 6.7, since this was performed at the same conditions as the 25 °C experiment performed here. Despite the composition, temperature, cooling rate, and heterogeneous interfaces being equivalent for these two experiments the nucleation probability is significantly higher for the previous experiments, which were performed using the Polar Bear and incubator setup instead of with the Crystal16 equipment. The difference between the two methods at 25 °C is even more apparent for the unconditional probability, which is due to the significant number of samples nucleating during the preparation stage of the Polar Bear experiments unlike with the Crystal16 methodology. This highlights the sensitivity of nucleation rate data to the exact methodologies and other experimental variabilities, such as differences in heat transfer or temperature calibration between the different equipments.

The biexponential model, described in Section 3.4, was used to estimate nucleation rates, which are presented in Table 6.3. The proportion of fast nucleating vials (A) increases with decreasing temperature, although there is no change in the content of the vials, this indicates a dependence on supersaturation to activate the nucleation sites present. The fast nucleation timescale ( $\tau_1$ ) decreases with temperature from 20 to 15 °C, as the higher supersaturation drives faster nucleation rates. However,  $\tau_1$  cannot be reliably estimated at the higher temperature of 25 °C, where the overall nucleation is relatively slow and only the slow nucleation mechanism is observed.

Table 6.3: Fitted nucleation parameters for the biexponential model.

Exp.	$\tau_1$ (hrs)	$\tau_2$ (hrs)	A	R <sup>2</sup>	Samples	Nucleation
D (15 °C)	0.065 ± 0.003	2.6 ± 0.2	0.835 ± 0.008	0.988	117	111
E (20 °C)	0.198 ± 0.014	6.7 ± 0.4	0.452 ± 0.010	0.983	199	134
F (25 °C)	0.05 ± 0.05 *	53.1 ± 1.7	0.0125 ± 0.0014	0.964	239	22

\*  $\tau_1$  is not statistically significant.

To further consider the variability between vials, the biexponential model is also applied to each of the vials individually. Two sets of fits were calculated. Firstly, each vial was treated as an individual system and all the parameters were fitted. Secondly, the vials were assumed to have the same slow nucleation timescale ( $\tau_2$ ) which was constrained to the value given in Table 6.3. To avoid over-fitting the data, the model was only applied to vials with a minimum of five nucleation events. The results for Experiment E (20 °C) are shown in Figure 6.8, and the results for Experiments D and F are available in Appendix C.4.

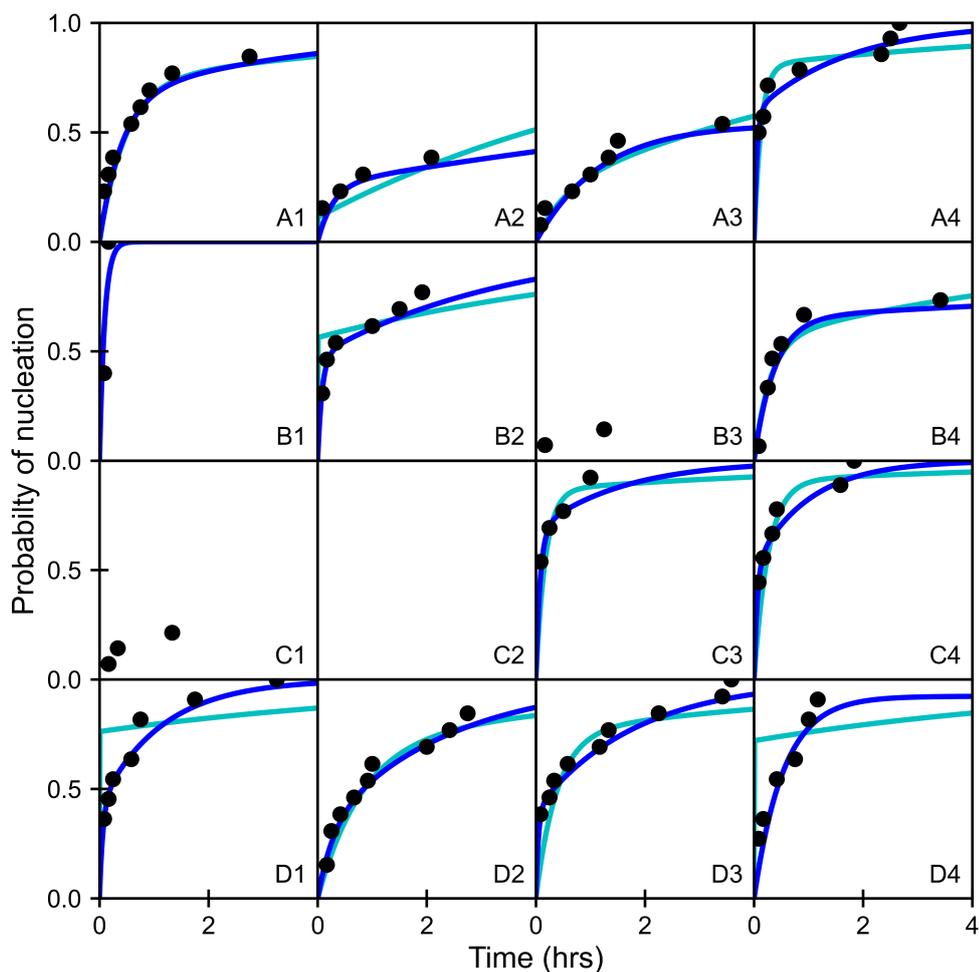


Figure 6.8: Fitted nucleation rates for individual vials in Experiment E (20 °C). The blue line is the unconstrained fit and the cyan line is the fit with  $\tau_2$  constrained.

The shape of the probability curve varies between the vials. The majority of vials

demonstrate an expected spread of nucleation times. Some vials have dominantly fast nucleation times with higher overall nucleation probabilities, the vial in position B1 in Figure 6.8 is an extreme example of this. Other vials, such as those in positions B3 and C1, have a low overall nucleation probability, with only a few nucleation events, although the nucleation times tend to be relatively quick. This variability between vials is an indicator that differences in heterogeneous surfaces or nucleation sites between vials have a significant influence on the nucleation outcome. This aligns with the findings of Deck and Mazzotti<sup>150</sup>, where series of 12 or 24 repeated cycles were used to study ice nucleation temperatures, in the Crystal16 apparatus. They found a significant variability in the nucleation temperatures of different vials but there was no significant variability between cycles or influence of vial positioning in the equipment.

In Figure 6.8 it is clear that the unconstrained fits for Experiment E fit the data better than if  $\tau_2$  is kept constant. This could be an indication that  $\tau_1$  and  $\tau_2$  are not simply representing the heterogeneous and homogeneous nucleation respectively, since a homogeneous term would be expected to apply equally to all of the vials. There is less of a difference between these two fits for Experiments D and F, as can be seen in Figures C.5 and C.6 in Appendix C.4.

### 6.3 Conclusions

The crystallisation phase diagram of aqueous urea was explored for unagitated solutions using temperature cycling in the Crystal16 equipment. Nucleation during the cooling ramp was found to be significant at temperatures of 20 °C and below. Isothermal induction times were still obtainable at 15 and 20 °C, since crystallisation during cooling only occurred for part of the samples. The nucleation times were very short at 15 °C with 95% of the samples nucleating during the 4 hrs of the experiment. A larger spread of nucleation times was obtained at 20 °C with 67% of the samples nucleating over the experiment, possibly providing the best conditions for further studies. Although several short nucleation times were observed at 25 °C, the overall probability of nucleation was very low at only 9%.

Reliable induction time measurements from unagitated solutions were obtainable due to the fast secondary nucleation and growth rates of urea, despite the equipment used typically requiring samples to be agitated for other systems. However, anomalous nucleation behaviour was observed in Experiments A–C, where nucleation occurred at temperatures above  $T_{\text{sat}}$ . This behaviour is thought to be linked to dissolution (and possible heat transfer) problems since it did not occur in Experiments D–F, where the samples were agitated throughout the dissolution stages. The requirement for agitation during dissolution necessitates the presence of a stirrer bar in the vials at all times, which will lead to the generation of heterogeneous nucleation data, since the stirrers are known to act as a heterogeneous nucleant.

The extensive use of repeated dissolution and crystallisation cycles did not affect the induction times, where the same temperature profile was used throughout the experiment. This allowed a large set of induction times to be collected efficiently, including sufficient data to produce nucleation probability profiles for individual vials. The nucleation rates were also analysed with the biexponential nucleation model, this indicated that the heterogeneous nucleation observed is dependent on high supersaturations to activate the nucleation sites in a system. Analysis of individual vials found that the majority had a stochastic range of nucleation times, but some were characterised by consistently fast or slow nucleation times. The nucleation rates calculated for individual vials had better fits when both fast and slow nucleation timescales were optimised, compared to keeping the slow nucleation timescale constant. This indicates that the slow nucleation behaviour which was observed is unlikely to be entirely related to homogeneous behaviour.

## Chapter 7

# Conclusions

This chapter summarises the work and findings contained in this thesis and its significance. This is followed by a discussion of the limitations of the work and suggestions for future research.

### 7.1 Findings and Outlook

This work has been performed to improve the understanding of how solid surfaces influence heterogeneous nucleation from solution. This work supports the hypothesis that surface-solution dispersion interactions influence the surrounding solution concentration.<sup>26,51</sup> This has been examined using a combination of molecular dynamics simulations and nucleation experiments. Urea in aqueous solutions was identified as a suitable model system due to its amenability for molecular dynamics simulations and its experimentally favourable fast growth rate without any competing polymorphs. The PTFE and control (glass and air) surfaces were selected due to their prevalence in nucleation experiments and availability and diamond was included due to its use in probe tips.

An extensive set of MD simulations were performed to compare and validate the performance of four GAFF and five OPLS force fields for modelling urea. The performance

of the force fields varied significantly, which highlights the importance of using representative conditions to validate models before they are taken into use. The GAFF-D3 and OPLS-AA force fields were found to be the best overall with similarly good performances. This work outlined a set of simple validation tests that are relevant for crystallisation work.

The GAFF-D3 force field was used to study urea solutions at surfaces. The dispersion interactions between a surface and solution were simulated using a LJ potential that was specifically parametrised to represent the  $\alpha$ -quartz (glass), PTFE and diamond surfaces. The PTFE surface had the strongest dispersion forces, followed by diamond, and  $\alpha$ -quartz had the weakest potential. The interfacial simulations showed that strong dispersion interactions with an interface causes a significant concentration enhancement of urea in the interfacial region, but that this does not occur if the interface only has weak dispersion interactions. For medium strength interactions the interfacial urea concentration increases with solution concentration, however, for very strong interactions, the bulk concentration does not significantly affect the interfacial concentration. Conversely, the water concentration is reduced in the interfacial region but increases as the dispersion interactions weaken. There was also a vacuum (air) interface present in all the simulations, which is the absence of any dispersion interactions, this led to a decrease in the nearby urea concentration. The effect of the vacuum interface was very similar for all the systems studied, which indicates that bulk conditions are obtained in the middle region of the simulation, and that the interfacial concentration enhancement does not cause significant depletion effects in the solution.

To measure the effect of the different surfaces on urea crystallisation, small scale nucleation experiments were performed using two different experimental setups. The first setup involved heating and cooling the samples using the Polar Bear equipment in preparation for the nucleation stage. Then the vials were placed on racks in a temperature controlled incubator for the nucleation stage and monitored using webcams, which allowed visual inspection of the samples. The second setup used the Crystall16 equipment, where both the heating and cooling of the samples and the nucleation stage

are performed within the equipment. The Crystal16 equipment uses a laser and transmissivity sensor to detect the onset of nucleation, but it is not possible to see the vials during the experiment.

Experiments were performed using the webcam setup at three concentrations, to determine the nucleation times in glass vials (control system) and with an additional PTFE bar or diamond inserted in the vials. The overall nucleation probabilities were relatively low when only the control surfaces were present, despite the supersaturation being as high as practicable. The addition of a PTFE surface significantly increased the nucleation probability. The addition of the diamond surfaces led to the sharpest initial increase in nucleation probability, but there was no significant difference in the final nucleation probability compared to the control system. Further experiments were performed using the Crystal16 setup with the highest solution concentration and PTFE bars present. In these experiments the nucleation probabilities were further increased by going to lower temperatures in the experiments with the Crystal16 equipment.

The solution preparation and experimental conditions were kept as similar as possible for the two setups, nevertheless, the nucleation probabilities differed in comparative experiments at 25 °C. 27.5% of the PTFE samples nucleated within 4 hrs for custom setup with the webcam, compared to only 9.2% of vials with the Crystal16 equipment. It is possible that the nucleation probability was underestimated for the custom setup, since a significant proportion of vials nucleated during the preparation stages. However, due to the lower nucleation probability in the Crystal16 experiments, these do not offer additional insight on this front. This indicates that small methodological differences can quantitatively influence nucleation behaviour.

The large number of repeated cycles that were performed during the Crystal16 experiments indicated that the throughput can be significantly increased through the use of temperature cycling, since no significant correlation was found between nucleation behaviour and the cycle number. Nonetheless, this work highlights the importance of having a large number of individual samples in nucleation experiments, to ensure that the results encompass the influence of vial-to-vial variability, which is especially impor-

tant when the overall nucleation probabilities are low.

The changes in experimental nucleation probability, which are dependent on the surfaces present in the system, compare well to the results obtained from the interfacial simulations. The low nucleation rate in the control system corresponds to interfacial urea concentrations which are at the bulk concentration or lower for the simulated  $\alpha$ -quartz and vacuum surfaces. Whereas the PTFE surface significantly increases the nucleation probability, which corresponds to the significantly enhanced interfacial urea concentration observed in the simulations. Interestingly, the lesser interfacial concentration enhancement for the diamond surface combined with the increase in initial but not overall nucleation probability indicates that there is a transition point where the interfacial effects have diminishing effects on nucleation behaviour. This indicates that the simulated interfacial concentration enhancement is a likely explanation for the causes of heterogeneous nucleation in these systems and these simulations may be able to be used to predict whether enhanced heterogeneous nucleation is likely to occur in a system.

This work suggests that heterogeneous nucleation is influenced by dispersion interactions between a surface and the solution, which can have important influences on crystallisation processes. The simulated interfacial concentration enhancement only stretches over  $\sim 1$  nm but still has a significant effect on nucleation probabilities in 1 ml experiments. Therefore, this work shows that bulk properties cannot be assumed to apply near interfaces and this may influence nucleation behaviour even at larger scales. This improved understanding of how surface-solution dispersion interactions influence nucleation can be applied to the selection of surface coatings and inserts during equipment design to optimise crystallisation processes by enhancing nucleation and preventing fouling as required.

## 7.2 Limitations and Future Research Directions

The scope of this work was limited to considering only one system: urea in aqueous solutions, this allowed for a highly detailed study, but does not enable general conclusions to be drawn. However, the effect on solution structure from simulated interfacial dispersion interactions and the corresponding relation to the enhanced experimental nucleation rates, aligns with previous nucleation studies with aqueous glycine solutions.<sup>25,26,51</sup> Therefore, it is possible to conclude that these results are generalisable for small organic molecules in aqueous solutions, although further work will be required to study the influence of larger molecules and other solvents.

There are practical limitations associated with both of the experimental setups that were tested. The throughput and crystallisation temperatures were restricted in the custom setup method (with the Polar Bear, incubator and webcams). The throughput was limited to 80 vials per experiment, which was determined by the capacity of the Polar Bear. The crystallisation temperatures were limited by the use of the incubator, which had a minimum operating temperature of 25 °C, which resulted in low overall nucleation probabilities.

The Crystal16 equipment can reach lower temperatures compared to the incubator, however, the number of individual samples and problems with unagitated dissolution constrained the use of this method. The Crystal16 equipment can only hold 16 vials at a time, however, repeated temperature cycles can be performed, which enables the same vials to be easily sampled multiple times during one experiment. This improves the overall experimental throughput and increases the amount of data obtainable with the same amount of material and preparation time. Although the throughput is still a compromise between the number of repeated cycles and the length of induction times considered. Difficulties with the dissolution of the crystals between cycles meant that stirring was required for the temperature cycling approach to be effective, therefore, this approach makes it more difficult to compare the effect of heterogeneous surfaces, if PTFE stirrer bars are required in all samples.

It was not possible to obtain statistically significant nucleation probabilities and rates for all of the experimental conditions tested, due to the low proportion of samples that nucleated in some of the experiments in both Chapters 5 and 6. The systems with few nucleation events would have benefited from increased sampling so that the number of nucleation events, rather than the number of samples, was comparable across the range of experiments. This was not possible due to the time constraints of this work, but is recommended for future work.

The influence of cooling rate and heat transfer on nucleation behaviour has not been studied in this work. The same, moderate cooling rate of  $-1.5\text{ }^{\circ}\text{C min}^{-1}$  was used for all the nucleation experiments and was assumed to provide sufficient heat transfer. This was selected based on other work, which has found that cooling rates of  $-5\text{ }^{\circ}\text{C min}^{-1}$  and slower provides good heat transfer for nucleation experiments on the 1 ml scale,<sup>148</sup> and since rates ranging from  $-0.2$  to  $-5.0\text{ }^{\circ}\text{C min}^{-1}$  typically provide reliable results.<sup>25,26,141–150</sup> It is, however, still possible for the cooling rate to influence the nucleation behaviour. A faster cooling rate brings the solution to the desired nucleation temperature quicker, which allows the induction time measurements to start sooner. Therefore, in this work, it is possible that a faster cooling rate will have the desired effect of reducing the number of samples which need to be discarded due to nucleation during the cooling ramp. However, undesirable crash crystallisation can occur if the solution is cooled too quickly and the solution becomes thermodynamically unstable. Therefore, it is also possible that a slower nucleation rate will have the desired effect of reducing the number of vials which nucleate during the cooling ramp by ensuring that the solution stays in the metastable zone. In these experiments, a notable proportion of vials nucleated during the cooling stages, therefore, it would be valuable for future work to test whether this can be reduced by changing the cooling rate. The heat transfer from the equipment to the vials and solution is important to ensure that the specified heating and cooling rates are obtained. It is possible that slower cooling and heating rates would have improved the heat transfer in the vials. This might have helped to prevent the anomalous behaviour in the early Crystal16 experiments where nucleation occurred above the saturation temperature. In the future it would be interesting to test

if this can be used to improve the comparability between the experiments using the custom setup and Crystal16 equipment.

The PTFE, diamond, glass and air materials tested in this work were selected due to their prevalence in crystallisation experiments and the availability in the laboratory. These surfaces have a diverse range of properties, both in terms of the chemical composition, surface groups, and the surface roughness. PTFE was found to significantly increase the nucleate rate, and whilst the simulations indicated that this was attributable to dispersion interactions, it remains possible that other factors such as surface roughness may also have an influence. In the future, it would be interesting to use a more targeted range of surface materials. Firstly, it would be interesting to test different surface functional groups, such as amine groups which are similar to urea, to understand how chemical interactions or arrangements may influence nucleation rate. Secondly, it would be useful to compare materials with similar properties (e.g. polymers such as polyethylene and polypropylene) across both experiments and simulations to determine how well these methods compare for more chemically similar materials, with more similar dispersion interactions. Moving forward, it is important to understand how the impact of these surface effects changes with scale. Therefore, it would be useful to conduct a set of experiments with a wide range of surface areas and solution volumes, as well as different surface roughnesses. This will help to determine how the interfacial effects can be used to optimise the design of crystallisation processes on an industrial scale.

No pre- or post-characterisation was performed for any of the surfaces, however, this might have revealed impurities or abrasions on the surfaces, which would further inform the conclusions drawn from the experiments. The surface areas were estimated by assuming that the surfaces were smooth, since the specific surface roughness of the materials was not known. In the future, more detailed characterisation of the surfaces may enable further quantitative discussion, however, it will not change the overall conclusions. The condition of both the PTFE and glass surfaces was representative of real conditions, and it is highly relevant to know that the addition of a typical PTFE

bar will influence nucleation rate.

The interfacial simulations only considered the influence of dispersion forces between the interface and solution. This omitted other important factors including electrostatic interactions and surface topography such as the three dimensional nature of surfaces and the variation of surface groups across the surface. However, these factors were outwith the scope of this work, and a thorough understanding of the influence of the dispersion forces provides a valuable baseline for future investigations.

Furthermore, it would be interesting to apply the combined experimental and simulation approach to a wide range of surface materials and crystallisation systems. The LJ wall parametrisation process could be expanded to consider a much larger array of materials to better understand what surface properties influence the wall strength. This should ideally be compared with experimental nucleation data to help to determine how accurately the LJ wall parameters can be used to predict the influence on nucleation rate. This could be further expanded to consider the influence of multicomponent systems containing a mixture of solvents or solution additives.

Heterogeneous nucleation has not been directly simulated in this work, but the interfacial systems that were used can act as a good starting point for this. The influence of the enhanced interfacial concentration on the stability and growth of crystal nuclei could be simulated by comparing small seed crystals placed in the interfacial region and bulk solution. Furthermore, enhanced simulation techniques such as well-tempered metadynamics,<sup>4,47-50</sup> which have been used to model homogeneous nucleation, could be applied to systems with LJ interfaces.

## Appendix A

# Additional Force Field and Validation Details

### A.1 Force Field Parameters and Equations

This section provides all the force field equations for bonded and non-bonded interactions used by all the force fields studied. This is followed by tables with the parameters that have been used for each of the force fields.

#### A.1.1 Force Field Equations

The same force field equations are used for all of the GAFF force fields (GAFF1, GAFF2, GAFF-D1 and GAFF-D3). The same non-bonded equations are used for all of the OPLS force fields but the bonded equations differ. OPLS-AA, OPLS-AA-N and OPLS-AA-D use the bonded AMBER equations, where as OPLS-S and OPLS-G use the bonded GROMOS equations.

### Bond Interactions

GAFF, OPLS-AA, OPLS-AA-N and OPLS-AA-D use a harmonic bond potential, given in Eq. A.1, which is implemented with *bond\_style harmonic* in LAMMPS. This is also used by the SPC/E water model. OPLS-S and OPLS-G use the gromos bond potential, given in Eq. A.2, which is implemented with *bond\_style gromos* in LAMMPS.

$$E_{\text{bond}} = K_r (r - r_0)^2 \quad (\text{A.1})$$

$$E_{\text{bond}} = K_r (r^2 - r_0^2)^2 \quad (\text{A.2})$$

$K_r$  is the bond constant,  $r_0$  is the equilibrium distance and  $r$  is the distance between the two atoms.

### Angle Interactions

GAFF, OPLS-AA, OPLS-AA-N and OPLS-AA-D use a harmonic angle potential, given in Eq. A.3, which is implemented with *angle\_style harmonic* in LAMMPS. This is also used by the SPC/E water model. OPLS-S and OPLS-G use a squared cosine angle potential, given in Eq. A.4, which is implemented with *angle\_style cosine/squared* in LAMMPS.

$$E_{\text{angle}} = K_{\theta} (\theta - \theta_0)^2 \quad (\text{A.3})$$

$$E_{\text{angle}} = K_{\theta} (\cos(\theta) - \cos(\theta_0))^2 \quad (\text{A.4})$$

$K_{\theta}$  is the angle constant,  $\theta_0$  is the equilibrium angle and  $\theta$  is the angle between the three atoms.

### Dihedral Interactions

GAFF uses a fourier dihedral potential, given in Eq. A.5, which is implemented with *dihedral\_style fourier* in LAMMPS. OPLS-AA, OPLS-AA-N and OPLS-AA-D use the opls dihedral potential, given in Eq. A.6, which is implemented with *dihedral\_style opls*

in LAMMPS. OPLS-S and OPLS-G use a harmonic dihedral potential, given in Eq. A.7, which is implemented with *dihedral\_style harmonic* in LAMMPS.

$$E_{\text{dihedral}} = K_{\phi,i} [1.0 + \cos(n_i\phi - \delta_i)] \quad (\text{A.5})$$

$$\begin{aligned} E_{\text{dihedral}} = & \frac{1}{2}V_1 [1.0 + \cos(\phi)] + \frac{1}{2}V_2 [1.0 - \cos(2\phi)] \\ & + \frac{1}{2}V_3 [1.0 + \cos(3\phi)] + \frac{1}{2}V_4 [1.0 - \cos(4\phi)] \end{aligned} \quad (\text{A.6})$$

$$E_{\text{dihedral}} = K_{\phi} [1.0 + d \cos(n\phi)] \quad (\text{A.7})$$

$K_{\phi,i}$  and  $V_{1-4}$  are the force constants,  $\phi$  is the dihedral angle,  $n$  is the periodicity of torsion,  $\delta$  is the phase angle and  $d = \cos(\delta)$ . These three equations are describing the same potential in different ways: A.7 is simply a re-arranged version of Eq. A.5 and Eq. A.6 is a re-arranged and expanded version of Eq. A.5.

### Improper Dihedral Interactions

GAFF and OPLS-AA, OPLS-AA-N and OPLS-AA-D use a fourier improper potential, given in Eq. A.8, which is implemented with *improper\_style cvff* in LAMMPS. This is the potential that was used for the GAFF dihedrals. OPLS-S and OPLS-G uses a harmonic improper potential, given in Eq. A.9, which is implemented with *improper\_style harmonic* in LAMMPS.

$$E_{\text{improper}} = K_{\xi} [1.0 + d \cos(n\xi)] \quad (\text{A.8})$$

$$E_{\text{improper}} = K_{\xi} (\xi - \xi_0)^2 \quad (\text{A.9})$$

$K_{\xi}$  is the force constant,  $\xi$  is the improper angle,  $n$  is the periodicity of torsion and  $d = \cos(\delta)$  where  $\delta$  is the phase angle.

## Non-Bonded Interactions

All the force fields use the Coulomb potential for the non-bonded electrostatic interactions, this is given in Eq. A.10.

$$E_{\text{Coul}} = \sum \frac{q_i q_j}{4\pi\epsilon_0 r} \quad (\text{A.10})$$

$q$  is the charge,  $i$  and  $j$  are the two different atoms,  $r$  is the distance between atoms  $i$  and  $j$  and  $\epsilon_0$  is the dielectric constant with a default value of 1.0 for a vacuum.

All the force fields use the Lennard-Jones potentials for the non-bonded dispersion interactions, this is given in Eq. A.11.

$$E_{\text{LJ}} = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - \left( \frac{\sigma_{ij}}{r} \right)^6 \right] \quad (\text{A.11})$$

Again,  $i$  and  $j$  are the two atoms interacting and  $r$  is the distance between these atoms.  $\epsilon$  is the minimum energy between the two atoms and  $\sigma$  is the distance required between the two atoms for the energy to be zero.

The  $\epsilon$  and  $\sigma$  values are specified for the interactions between atoms of the same type (e.g. C–C interactions). Mixing rules are used to calculate these values for any atom pairing (e.g. C–O and C–N). For this, all the GAFF force fields use the arithmetic (Lorentz Berthelot) mixing rules given in Eqs. A.12 and A.14 and all the OPLS force fields use the geometric mixing rules given in Eqs. A.13 and A.14. For water, the mixing rules were chosen to match the other force field being used.

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \quad (\text{A.12})$$

$$\sigma_{ij} = (\sigma_i \sigma_j)^{\frac{1}{2}} \quad (\text{A.13})$$

$$\epsilon_{ij} = (\epsilon_i \epsilon_j)^{\frac{1}{2}} \quad (\text{A.14})$$

### A.1.2 GAFF Parameters for Urea

The GAFF1 and GAFF2 parameters are taken from GAFF version 1.81 and 2.11 respectively, these were obtained using the Antechamber software which is part of AmberTools21<sup>116</sup>. The GAFF-D1 and GAFF-D3 parameters were taken from Özpınar et al.<sup>111</sup>

Table A.1: GAFF mass and charge parameters.

Atom type	Mass (Da)	Charge (e)		
	GAFF <sup>99,116</sup>	GAFF1, <sup>99,116</sup> GAFF2 <sup>116</sup>	GAFF-D1 <sup>111</sup>	GAFF-D3 <sup>111</sup>
C	12.01	0.7261	1.172	0.884
O	16.00	-0.6391	-0.795	-0.660
N	14.01	-0.6420	-1.098	-0.888
H	1.008	0.2990	0.454	0.388

Table A.2: The GAFF Lennard-Jones parameters.

Atom type	GAFF1, <sup>99,116</sup> GAFF-D1, <sup>111</sup> GAFF-D3 <sup>111</sup>		GAFF2 <sup>116</sup>	
	$\epsilon$ (kcal mol <sup>-1</sup> )	$\sigma^*$ (Å)	$\epsilon$ (kcal mol <sup>-1</sup> )	$\sigma^*$ (Å)
C	0.0860	3.3997	0.0988	3.3152
O	0.2100	2.9599	0.1463	3.0481
N	0.1700	3.2500	0.1636	3.1809
H	0.0157	1.0691	0.0100	1.1065

\* The  $\sigma$  (Lennard-Jones diameter) values were given as  $R_{\min}$  (van der Waals radius) values,  $\sigma = 2^{\frac{5}{6}} R_{\min}$  was used for the conversion (from  $2R_{\min} = 2^{\frac{1}{6}} \sigma$ ).

Table A.3: The GAFF bond and angle parameters.

Bonds	GAFF1 <sup>99,116</sup>		GAFF2 <sup>116</sup>		GAFF-D1, GAFF-D3 <sup>111</sup>	
	$K_r$	$r_0$	$K_r$	$r_0$	$K_r$	$r_0$
C - O	637.7	1.2183	652.57	1.218	656	1.250
C - N	427.6	1.3789	356.21	1.379	424	1.383
N - H	403.2	1.0129	527.31	1.013	434	1.010

Angles	GAFF1 <sup>99,116</sup>		GAFF2 <sup>116</sup>		GAFF-D1, GAFF-D3 <sup>111</sup>	
	$K_\theta$	$\theta_0$	$K_\theta$	$\theta_0$	$K_\theta$	$\theta_0$
C - N - H	48.33	117.55	48.691	117.550	30	120.0
N - C - O	74.22	123.05	113.811	123.050	80	120.9
N - C - N	72.90	113.56	112.428	113.560	70	118.6
H - N - H	39.56	117.95	39.011	117.950	35	120.0

\*  $K_r$  in kcal mol<sup>-1</sup> Å<sup>-2</sup>,  $r_0$  in Å,  $K_\theta$  in kcal mol<sup>-1</sup> rad<sup>-2</sup> and  $\theta_0$  in °.

Table A.4: The GAFF dihedral and improper dihedral parameters.

Dihedrals	$K_\phi$ (kcal mol <sup>-1</sup> )	$n$ (-)	$d$ (°)
H - N - C - O	2.50	2	180.0
H - N - C - O	2.00	1	0.0
H - N - C - N*	2.50	2	180.0

Improper dihedrals	$K_\phi$ (kcal mol <sup>-1</sup> )	$n$ (-)	$d$ (°)
C - N - H - H	1.1	2	180.0
O - C - N - N	10.5	2	180.0

\* These parameters are the same for all the force fields<sup>99,111,116</sup> except the H-N-C-N dihedral which is not used in GAFF-D1 and GAFF-D3<sup>99,111</sup>.

### A.1.3 OPLS Parameters for Urea

The OPLS-AA parameters were obtained from Jorgensen et al.<sup>97</sup> with reference to Weiner et al.<sup>98,126</sup> The OPLS-AA-D parameters were obtained from Duffy et al.<sup>102</sup> with reference to Jorgensen et al.<sup>97</sup>, Jorgensen and Tirado-Rives<sup>127</sup> The OPLS-AA-N parameters were obtained using the LigParGen server.<sup>128,129</sup> The OPLS-S and OPLS-G parameters are taken from Smith et al.<sup>110</sup> with reference to Oostenbrink et al.<sup>101</sup>

Table A.5: OPLS mass and charge parameters.

Atom type	Mass (Da)		Charge (e)	
	OPLS <sup>127-129</sup>	OPLS-AA <sup>97</sup>	OPLS-AA-D, <sup>102</sup> OPLS-S, <sup>110</sup> OPLS-G <sup>110</sup>	OPLS-AA-N <sup>127-129</sup>
C	12.01100	0.500	0.142	0.95929
O	15.99940	-0.500	-0.390	-0.51455
N	14.00670	-0.760	-0.542	-1.30043
H	1.00800	0.380	0.333	0.53904

Table A.6: The OPLS Lennard-Jones parameters.

Atom type	OPLS-AA, <sup>97</sup> OPLS-AA-D, <sup>102</sup> OPLS-S, <sup>110</sup> OPLS-G <sup>110</sup>		OPLS-AA-N <sup>127-129</sup>	
	$\epsilon$ (kcal mol <sup>-1</sup> )	$\sigma$ (Å)	$\epsilon$ (kcal mol <sup>-1</sup> )	$\sigma$ (Å)
C	0.105	3.750	0.070	3.550
O	0.210	2.960	0.210	2.960
N	0.170	3.250	0.170	3.250
H	0.000	0.000	0.000	0.000

Table A.7: The OPLS bond and angle parameters.

Bonds	OPLS-G, <sup>101,110</sup> OPLS-S <sup>101,110</sup>		OPLS-AA, <sup>97,98</sup> OPLS-AA-N, <sup>127-129</sup> OPLS-AA-D <sup>127-129</sup>	
	$K_r$	$r_0$	$K_r$	$r_0$
C - O	78.3	1.2650	570.0	1.2290
C - N	61.5	1.3500	490.0	1.3350
N - H	111.7	1.0000	434.0	1.0100
Angles	$K_\theta$	$\theta_0$	$K_\theta$	$\theta_0$
N - C - O	82.5	121.4	80.0	122.9
N - C - N*	76.0	117.2	70.0	114.2
C - N - H	46.6	120.0	35.0	119.8
H - N - H	53.2	120.0	35.0	120.0

\* The N - C - N bond is not used in OPLS-AA.

\*\*  $K_r$  in kcal mol<sup>-1</sup> Å<sup>-4</sup>,  $r_0$  in Å,  $K_\theta$  in kcal mol<sup>-1</sup> rad<sup>-2</sup> and  $\theta_0$  in °.

\*\*\*  $K_r$  and  $K_\theta$  have been multiplied by  $\frac{1}{4}$  and  $\frac{1}{2}$  respectively, to match LAMMPS formatting.

Table A.8: The OPLS dihedral and improper dihedral parameters.

Dihedrals	OPLS-AA, <sup>97</sup> OPLS-AA-N, <sup>127-129</sup> OPLS-AA-D <sup>127-129</sup>			
	$V_1$	$V_2$	$V_3$	$V_4$
H - N - C - N	0.000	4.900	0.000	0.000
H - N - C - O	0.000	4.900	0.000	0.000
Improper dihedrals	$K_\xi$	$n$	$d$	
O - C - N - N	10.5	2	-1	
C - N - H - H	2.5	2	-1	

\*  $V$  in kcal mol<sup>-1</sup>,  $K_\xi$  in kcal mol<sup>-1</sup> rad<sup>-2</sup> and  $n$  and  $d$  are unit less.

Table A.9: The OPLS dihedral and improper dihedral parameters.

Dihedrals	OPLS-S <sup>110</sup>			OPLS-G <sup>101</sup>		
	$K_\phi$	$d$	$n$	$K_\phi$	$d$	$n$
H - N - C - O	9.99	-1	2	8.01	-1	2

Improper dihedrals	$K_\xi$	$\xi_0$	$K_\xi$	$\xi_0$
O - C - N - N	20.0	180	20.0	180
C - N - H - H	20.0	180	20.0	180

\*  $K_\phi$  in kcal mol<sup>-1</sup>,  $d$  and  $n$  are unit less,  $K_\xi$  in kcal mol<sup>-1</sup> rad<sup>-2</sup> and  $\xi_0$  in °.

\*\*  $K_\xi$  has been multiplied by  $\frac{1}{2}$  and 180° has been added to  $\xi_0$  to match LAMMPS formatting.

### A.1.4 SPC/E Water Parameters

The SPC/E parameters were obtained from Mark and Nilsson<sup>133</sup>, Berendsen et al.<sup>135</sup>

Table A.10: The SPC/E mass, charge and Lennard-Jones parameters.

Atom type	Mass (Da)	Charge (e) <sup>133,135</sup>	$\epsilon$ (kcal mol <sup>-1</sup> ) <sup>133,135</sup>	$\sigma$ (Å) <sup>133,135</sup>
O	16.00	-0.8476	0.1553	3.1656
H	1.008	0.4238	0.0000	0.0000

 Table A.11: The SPC/E bond and angle parameters.<sup>133,135</sup>

Bond	$K_r$ (kcal mol <sup>-1</sup> Å <sup>-2</sup> )	$r_0$ (Å)
O - H	1000.0	1.000

Angle	$K_\theta$ (kcal mol <sup>-1</sup> rad <sup>-2</sup> )	$r_0$ (°)
H - O - H	100.0	109.47

## A.2 Tabulated Results

### A.2.1 Crystal Lattice Parameters

Table A.12: Crystal lattice parameters after minimisation of the experimental form I crystal structure.

Force field	a (Å)	b (Å)	c (Å)	$\rho$ (g cm <sup>-3</sup> )
GAFF1	5.324	5.324	4.820	1.460
GAFF2	5.321	5.321	4.774	1.476
GAFF-D1	5.221	5.221	4.810	1.521
GAFF-D3	5.328	5.328	4.811	1.460
OPLS-AA	5.412	5.412	4.795	1.420
OPLS-AA-N	5.350	5.350	4.830	1.442
OPLS-AA-D	5.415	5.415	4.786	1.421
OPLS-S	5.493	5.493	4.785	1.382
OPLS-G	5.490	5.490	4.775	1.386

Table A.13: Crystal lattice parameters after minimisation of the experimental form IV crystal structure.

Force field	a (Å)	b (Å)	c (Å)	$\rho$ (g cm <sup>-3</sup> )
GAFF1	3.505	7.523	4.847	1.561
GAFF2	3.499	7.102	4.796	1.673
GAFF-D1	3.512	7.052	4.813	1.673
GAFF-D3	3.507	7.454	4.817	1.584
OPLS-AA	3.576	7.614	4.883	1.500
OPLS-AA-N	3.646	7.181	5.222	1.459
OPLS-AA-D	3.698	7.569	4.882	1.460
OPLS-S	3.660	7.688	4.886	1.451
OPLS-G	3.649	7.696	4.888	1.453

Table A.14: Crystal lattice parameters from the NPT simulations starting in form I.

Force field	a (Å)	b (Å)	c (Å)	$\rho$ (g cm <sup>-3</sup> )
GAFF1	5.451 ± 0.230	5.447 ± 0.239	4.715 ± 0.011	1.427 ± 0.007
GAFF2	5.405 ± 0.237	5.409 ± 0.238	4.685 ± 0.010	1.459 ± 0.007
GAFF-D1	5.253 ± 0.030	5.252 ± 0.030	4.605 ± 0.012	1.570 ± 0.007
GAFF-D3	5.398 ± 0.082	5.400 ± 0.082	4.691 ± 0.011	1.459 ± 0.006
OPLS-AA	5.409 ± 0.123	5.408 ± 0.125	4.681 ± 0.011	1.457 ± 0.006
OPLS-AA-N	5.091 ± 0.042	5.680 ± 0.031	4.405 ± 0.033	1.566 ± 0.009
OPLS-AA-D	4.691 ± 0.069	6.317 ± 0.078	4.673 ± 0.011	1.441 ± 0.007
OPLS-S	4.830 ± 0.082	6.207 ± 0.090	4.701 ± 0.010	1.415 ± 0.007
OPLS-G	4.893 ± 0.102	6.141 ± 0.108	4.699 ± 0.010	1.413 ± 0.008

\* The mean and standard deviation are calculated with a sampling frequency of 0.1 ns.

Table A.15: Crystal lattice parameters from the NPT simulations starting in form IV.

Force field	a (Å)	b (Å)	c (Å)	$\rho$ (g cm <sup>-3</sup> )
GAFF1	3.779 ± 0.033	7.174 ± 0.039	4.735 ± 0.010	1.554 ± 0.009
GAFF2	3.740 ± 0.034	7.104 ± 0.037	4.705 ± 0.011	1.596 ± 0.010
GAFF-D1	3.675 ± 0.024	7.646 ± 0.042	4.385 ± 0.019	1.619 ± 0.008
GAFF-D3	3.814 ± 0.037	7.080 ± 0.049	4.689 ± 0.013	1.575 ± 0.009
OPLS-AA	5.402 ± 0.146	5.416 ± 0.152	4.682 ± 0.012	1.457 ± 0.006
OPLS-AA-N	3.835 ± 0.016	6.929 ± 0.023	4.591 ± 0.030	1.635 ± 0.009
OPLS-AA-D	4.688 ± 0.070	6.322 ± 0.081	4.673 ± 0.012	1.440 ± 0.008
OPLS-S	4.837 ± 0.087	6.200 ± 0.100	4.701 ± 0.012	1.415 ± 0.007
OPLS-G	4.885 ± 0.103	6.149 ± 0.112	4.700 ± 0.012	1.413 ± 0.008

\* The mean and standard deviation are calculated with a sampling frequency of 0.1 ns.

## A.2.2 Crystal Potential Energy

Table A.16: Crystal potential energy after minimisation of the experimental structure.

Force field	$E_{\text{crystal}}$ (kJ mol <sup>-1</sup> )		
	Starting in form I	Starting in form IV	Energy difference
GAFF1	-543.20	-540.86	-2.34
GAFF2	-541.36	-543.23	1.87
GAFF-D1	-1172.58	-1171.01	-1.57
GAFF-D3	-807.82	-804.61	-3.21
OPLS-AA	-332.00	-319.02	-12.98
OPLS-AA-N	-572.47	-560.49	-11.98
OPLS-AA-D	-215.92	-205.53	-10.39
OPLS-S	-213.68	-201.32	-12.36
OPLS-G	-213.96	-200.80	-13.16

Table A.17: Crystal potential energy for the NPT simulations.

Force field	$E_{\text{crystal}}$ (kJ mol <sup>-1</sup> )		
	Starting in form I	Starting in form IV	Energy difference
GAFF1	-510.26 ± 1.22	-511.52 ± 1.12	1.26 ± 2.34
GAFF2	-507.22 ± 1.26	-508.81 ± 1.28	1.59 ± 2.54
GAFF-D1	-1143.52 ± 1.14	-1148.22 ± 0.97	4.70 ± 2.10
GAFF-D3	-775.11 ± 1.22	-775.90 ± 1.12	0.79 ± 2.34
OPLS-AA	-301.24 ± 1.16	-301.22 ± 0.85	-0.02 ± 2.00
OPLS-AA-N	-558.30 ± 0.90	-577.03 ± 0.87	18.73 ± 1.78
OPLS-AA-D	-184.46 ± 1.18	-184.40 ± 0.92	-0.06 ± 2.10
OPLS-S	-181.97 ± 1.21	-182.01 ± 0.87	0.04 ± 2.08
OPLS-G	-182.00 ± 1.29	-181.96 ± 0.88	-0.04 ± 2.08

\* The mean and standard deviation are calculated with a sampling frequency of 0.1 ns.

### A.2.3 Solution Density

Table A.18: Solution density.

Concentration		Density (g cm <sup>-3</sup> )		
Urea molecules	Mass %	GAFF1	GAFF2	GAFF-D1
0	0.00	0.999 ± 0.008	0.999 ± 0.008	0.999 ± 0.008
10	3.23	1.008 ± 0.008	1.008 ± 0.008	1.012 ± 0.008
50	14.29	1.041 ± 0.007	1.044 ± 0.008	1.063 ± 0.008
100	25.00	1.076 ± 0.007	1.080 ± 0.007	1.115 ± 0.007
150	33.34	1.104 ± 0.007	1.110 ± 0.008	1.158 ± 0.007
200	40.00	1.128 ± 0.007	1.136 ± 0.007	1.195 ± 0.006
300	50.00	1.167 ± 0.006	1.176 ± 0.007	1.251 ± 0.005
400	57.15	1.196 ± 0.006	1.206 ± 0.005	1.293 ± 0.005
500	62.50	1.218 ± 0.005	1.231 ± 0.006	1.325 ± 0.005
600	66.67	1.238 ± 0.005	1.251 ± 0.006	1.351 ± 0.005
1000	76.93	1.285 ± 0.004	1.301 ± 0.004	1.417 ± 0.004

\* The mean and standard deviation are calculated with a sampling frequency of 0.1 ns.

Table A.19: Solution density .

Concentration		Density (g cm <sup>-3</sup> )		
Urea molecules	Mass %	GAFF-D3	OPLS-AA	OPLS-AA-N
0	0.00	0.999 ± 0.008	0.999 ± 0.008	0.999 ± 0.008
10	3.23	1.009 ± 0.008	1.009 ± 0.008	1.014 ± 0.008
50	14.29	1.048 ± 0.008	1.047 ± 0.008	1.073 ± 0.007
100	25.00	1.087 ± 0.007	1.086 ± 0.007	1.129 ± 0.007
150	33.34	1.120 ± 0.007	1.118 ± 0.006	1.177 ± 0.007
200	40.00	1.149 ± 0.007	1.144 ± 0.006	1.216 ± 0.007
300	50.00	1.192 ± 0.007	1.184 ± 0.005	1.274 ± 0.006
400	57.15	1.225 ± 0.006	1.215 ± 0.006	1.321 ± 0.005
500	62.50	1.250 ± 0.006	1.238 ± 0.005	1.356 ± 0.005
600	66.67	1.271 ± 0.005	1.256 ± 0.005	1.384 ± 0.004
1000	76.93	1.323 ± 0.004	1.303 ± 0.004	1.453 ± 0.004

\* The mean and standard deviation are calculated with a sampling frequency of 0.1 ns.

Table A.20: Solution density.

Concentration		Density (g cm <sup>-3</sup> )		
Urea molecules	Mass %	OPLS-AA-D	OPLS-S	OPLS-G
0	0.00	0.999 ± 0.008	0.999 ± 0.008	0.999 ± 0.008
10	3.23	1.009 ± 0.008	1.007 ± 0.008	1.007 ± 0.007
50	14.29	1.045 ± 0.008	1.041 ± 0.008	1.042 ± 0.008
100	25.00	1.082 ± 0.008	1.077 ± 0.007	1.076 ± 0.007
150	33.33	1.112 ± 0.007	1.105 ± 0.007	1.105 ± 0.007
200	40.00	1.137 ± 0.007	1.129 ± 0.006	1.129 ± 0.007
300	50.00	1.175 ± 0.006	1.165 ± 0.006	1.165 ± 0.006
400	57.15	1.206 ± 0.005	1.193 ± 0.005	1.193 ± 0.005
500	62.50	1.228 ± 0.005	1.215 ± 0.006	1.214 ± 0.005
600	66.67	1.246 ± 0.005	1.232 ± 0.005	1.231 ± 0.005
1000	76.93	1.292 ± 0.004	1.273 ± 0.004	1.274 ± 0.004

\* The mean and standard deviation are calculated with a sampling frequency of 0.1 ns.

## A.2.4 Solution Diffusion Coefficients

The solution diffusion coefficients were calculated using the gradient over the first 10 ns of the MSD. Trajectory data was used every 0.01 ns for calculating the multi-time-origin MSD.

Table A.21: Solution diffusion coefficient.

Concentration		Diffusion coefficient ( $\times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ )		
Urea molecules	Mass %	GAFF1	GAFF2	GAFF-D1
10	3.23	$1.9834 \pm 0.0028$	$1.0856 \pm 0.0029$	$1.3211 \pm 0.0021$
50	14.29	$1.7137 \pm 0.0018$	$1.4226 \pm 0.0005$	$0.9952 \pm 0.0013$
100	25.00	$1.2271 \pm 0.0007$	$1.2938 \pm 0.0004$	$0.8596 \pm 0.0010$
150	33.34	$1.1464 \pm 0.0005$	$1.1569 \pm 0.0004$	$0.7770 \pm 0.0004$
200	40.00	$1.0678 \pm 0.0008$	$1.1146 \pm 0.0002$	$0.6362 \pm 0.0002$
300	50.00	$0.8090 \pm 0.0001$	$0.8867 \pm 0.0004$	$0.4812 \pm 0.0003$
400	57.15	$0.6790 \pm 0.0002$	$0.8568 \pm 0.0002$	0.3583
500	62.50	$0.6055 \pm 0.0001$	$0.7220 \pm 0.0002$	0.3004
600	66.67	$0.5482 \pm 0.0001$	$0.6492 \pm 0.0002$	0.2446
1000	76.93	0.3806	$0.4602 \pm 0.0001$	0.1227

\* The uncertainty values is the standard error.

\*\* Where no uncertainty value is given, this rounded to 0.0000.

Table A.22: Solution diffusion coefficient.

Concentration		Diffusion coefficient ( $\times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ )		
Urea molecules	Mass %	GAFF-D3	OPLS-AA	OPLS-AA-N
10	3.23	$1.4841 \pm 0.0017$	$1.5908 \pm 0.0034$	$0.6659 \pm 0.0033$
50	14.29	$1.3152 \pm 0.0004$	$1.4147 \pm 0.0006$	$0.5662 \pm 0.0001$
100	25.00	$1.2854 \pm 0.0005$	$1.3645 \pm 0.0009$	$0.3257 \pm 0.0002$
150	33.34	$1.1776 \pm 0.0006$	$1.2464 \pm 0.0003$	$0.2409 \pm 0.0001$
200	40.00	$1.1197 \pm 0.0008$	$1.0898 \pm 0.0004$	0.1517
300	50.00	$0.8446 \pm 0.0005$	$0.8198 \pm 0.0004$	0.0641
400	57.15	$0.7509 \pm 0.0001$	$0.7266 \pm 0.0001$	0.0287
500	62.50	$0.6460 \pm 0.0001$	$0.6370 \pm 0.0001$	0.0139
600	66.67	$0.5798 \pm 0.0001$	$0.5970 \pm 0.0001$	0.0052
1000	76.93	$0.4044 \pm 0.0001$	$0.3924 \pm 0.0001$	0.0005

\* The uncertainty values is the standard error.

\*\* Where no uncertainty value is given, this rounded to 0.0000.

Table A.23: Solution diffusion coefficient.

Concentration		Diffusion coefficient ( $\times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ )		
Urea molecules	Mass %	OPLS-AA-D	OPLS-S	OPLS-G
10	3.23	$1.9129 \pm 0.0024$	$2.3257 \pm 0.0059$	$1.2183 \pm 0.0039$
50	14.29	$1.7211 \pm 0.0009$	$1.2912 \pm 0.0013$	$1.4087 \pm 0.0023$
100	25.00	$1.4297 \pm 0.0011$	$1.3815 \pm 0.0008$	$1.4043 \pm 0.0014$
150	33.34	$1.2541 \pm 0.0004$	$1.3348 \pm 0.0004$	$1.3397 \pm 0.0003$
200	40.00	$1.1676 \pm 0.0001$	$1.2838 \pm 0.0009$	$1.1421 \pm 0.0004$
300	50.00	$0.9507 \pm 0.0005$	$0.9901 \pm 0.0002$	$0.9408 \pm 0.0002$
400	57.15	$0.8224 \pm 0.0002$	$0.8760 \pm 0.0002$	$0.8471 \pm 0.0003$
500	62.50	$0.7665 \pm 0.0003$	$0.7247 \pm 0.0002$	$0.7873 \pm 0.0001$
600	66.67	$0.6778 \pm 0.0001$	$0.6922 \pm 0.0003$	$0.7174 \pm 0.0002$
1000	76.93	$0.4859 \pm 0.0001$	0.4789	$0.4827 \pm 0.0001$

\* The uncertainty values is the standard error.

\*\* Where no uncertainty value is given, this rounded to 0.0000.

### A.3 Further Radial Distribution Functions

The O–H<sub>W</sub> and O–N RDFs have been discussed in the main paper, here we also provide RDFs and discussion for the following interactions H–O<sub>W</sub>, O–O<sub>W</sub>, N–H<sub>W</sub>, N–O<sub>W</sub> and C–O<sub>W</sub>. The RDFs have been calculated up to 9.0 Å, at which point the curves are tailing off to a value of 1.0 for all the different RDF interactions and force fields studied. The only exception is the O–N RDFs for the dilute solutions, which is also tailing off but more slowly.

#### A.3.1 H–O<sub>W</sub>

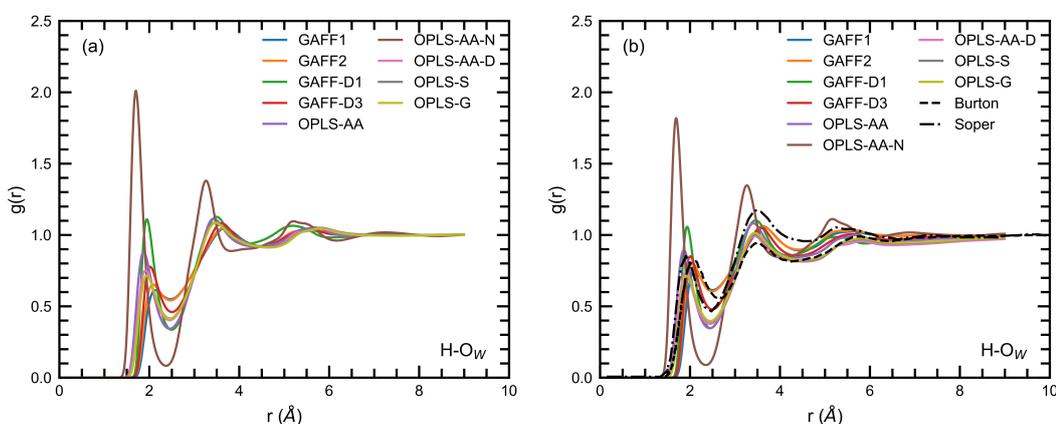


Figure A.1: H–O<sub>W</sub> RDFs for (a) the dilute solutions and (b) the concentrated solutions. Literature data is taken from Burton et al.<sup>172</sup> and Soper et al.<sup>130</sup>

The H–O<sub>W</sub> RDFs in Figure A.1 are similar for all the force fields and concentrations, except that OPLS-AA-N has greatly extended peaks and troughs, which also appear earlier. Like with O–H<sub>W</sub> the first peak is at  $< 2$  Å, indicating strong hydrogen bonding. There are also weaker second and third peaks appearing at 3.5 Å and  $> 5.0$  Å respectively. The H–O<sub>W</sub> RDFs are similar to those from Burton et al.<sup>172</sup> and Soper et al.<sup>130</sup>, and also from Ishida et al.<sup>171</sup> and Duffy et al.<sup>102</sup> but these had separate RDFs for the urea H molecules in both syn and anti arrangements, which were not shown for clarity. The RDFs of Duffy et al.<sup>102</sup> indicate that the broader third peak is made up of two pairs of smaller peaks from these (syn-peak, anti-peak, syn-peak, anti-peak).

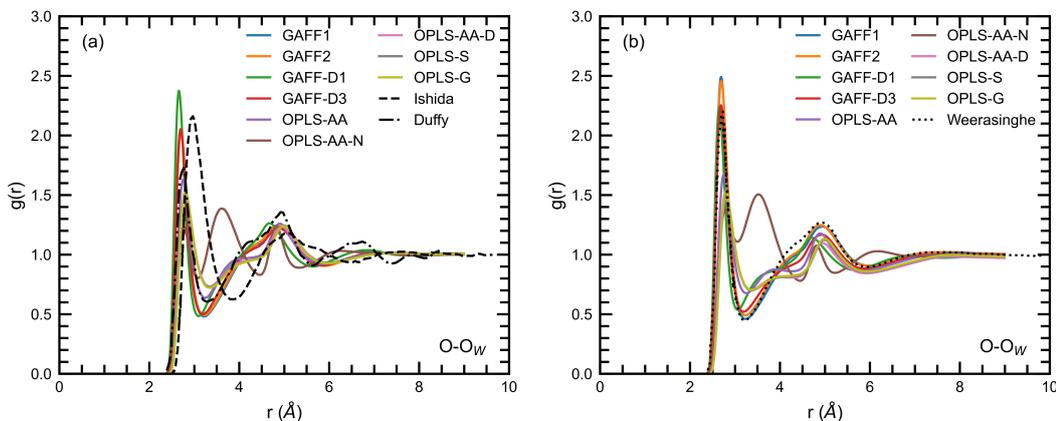
A.3.2 O–O<sub>W</sub>

Figure A.2: O–O<sub>W</sub> RDFs for (a) the dilute solutions and (b) the concentrated solutions. Literature data is taken from Ishida et al.<sup>171</sup>, Duffy et al.<sup>102</sup>, and Weerasinghe and Smith<sup>113</sup>.

For the O–O<sub>W</sub> RDFs, shown in in Figure A.2, there is a strong first peak at  $> 2.5$  Å followed by broader second and third peaks for all the force fields apart from OPLS-AA-N. For the remaining OPLS force fields a small second peak occurs at  $> 3.5$  Å, which plateaus before the third peak appears at 5 Å. For the GAFF force fields a slow rise leads to the second peak which appears later at  $> 4.0$  Å, this peak is higher than that for the OPLS force fields. This later, taller second peak runs into the third peak at 5.0 Å. This shape is similar to that obtained by Duffy et al.<sup>102</sup> and Weerasinghe and Smith<sup>113</sup>. These peaks are very similar for both the dilute and concentrated systems. The overlapping second and third peaks, could be a combination of water molecules not directly bonded to the O on the urea molecule and water molecules hydrogen bonded to the other parts of the urea molecule. OPLS-AA-N has first and third peaks very similar to those of the other OPLS force fields. However, the second OPLS-AA-N peak is much stronger and more well defined than those of the other force fields, and it peaks earlier at 3.5 Å. This is not just due to the O–O<sub>W</sub> partial charges (and also those of O–H<sub>W</sub>), since the charge differences for OPLS-AA-N lies between those of the other OPLS force fields and GAFF force fields.

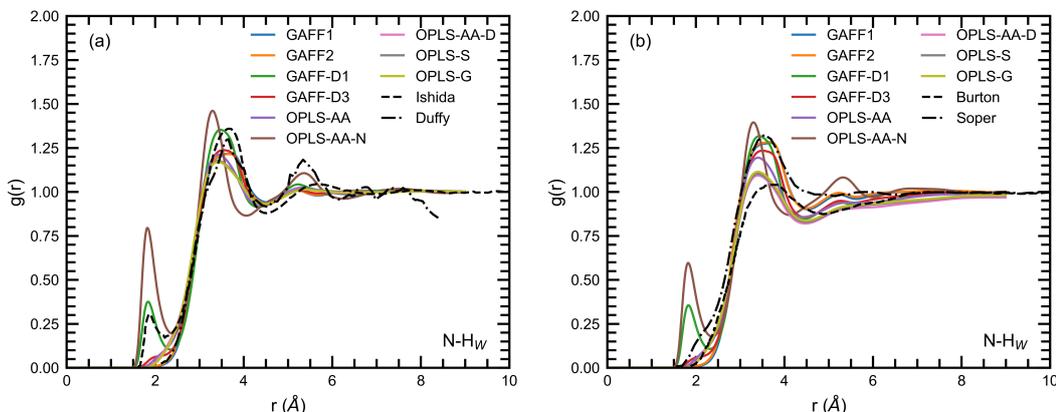
A.3.3 N–H<sub>W</sub>

Figure A.3: N–H<sub>W</sub> RDFs for (a) the dilute solutions and (b) the concentrated solutions. Literature data is taken from Ishida et al.<sup>171</sup>, Duffy et al.<sup>102</sup>, Burton et al.<sup>172</sup>, and Soper et al.<sup>130</sup>

The N–H<sub>W</sub> RDF, in Figure A.3, has an extra peak at  $< 2 \text{ \AA}$  for GAFF-D1 and OPLS-AA-N, and which is very weak for GAFF-D3 and absent from the other force fields. This matches a very clear peak in Ishida et al.<sup>171</sup>, whilst there is only a small bump in Soper et al.<sup>130</sup> and no peak in Duffy et al.<sup>102</sup> and Burton et al.<sup>172</sup>. This indicates that there are only strong N $\cdots$ H<sub>W</sub> hydrogen bonds presents in GAFF-D1 and OPLS-AA-N, with a few present in GAFF-D3. These additional hydrogen bonds can be related to the higher density of these force fields. The first main peak appears at  $> 3.0 \text{ \AA}$ , which also appears in all the reference RDFs. There is a second peak at  $> 5.0 \text{ \AA}$ , this is not noticeable in the higher concentrations of OPLS or reference solutions, with the exception of OPLS-AA-N where this is enhanced in the higher concentrations.

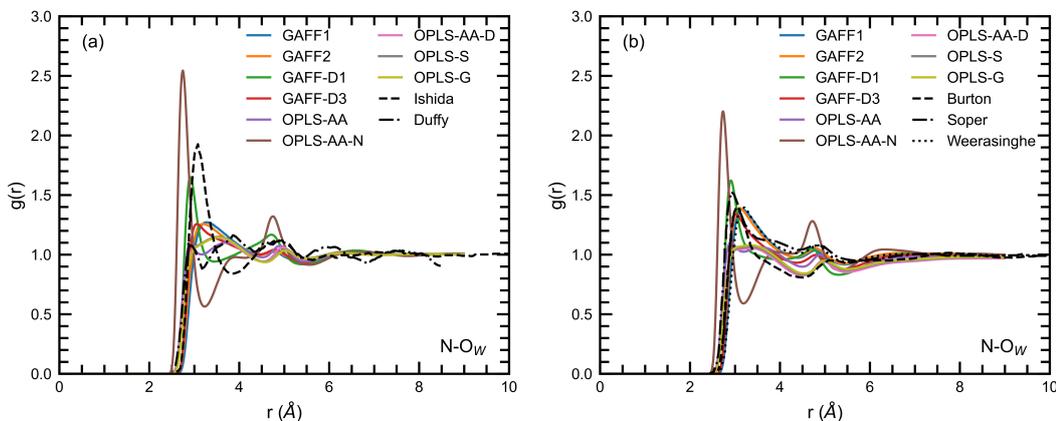
A.3.4 N–O<sub>W</sub>

Figure A.4: N–O<sub>W</sub> RDFs for (a) the dilute solutions and (b) the concentrated solutions. Literature data is taken from Ishida et al.<sup>171</sup>, Duffy et al.<sup>102</sup>, Burton et al.<sup>172</sup>, Soper et al.<sup>130</sup>, and Weerasinghe and Smith<sup>113</sup>.

Interestingly, the charge on the N atom is more negative than that of the O<sub>W</sub> atom for GAFF-D1, GAFF-D3 and OPLS-AA-N, but this is the opposite for the remaining force fields. Despite this there is no significant difference between the N–O<sub>W</sub> RDF of GAFF-D3 (for which the N and OW charges are very similar) compared to GAFF1 and GAFF2, as shown in Figure A.4. The OPLS RDF peaks are generally lower and shifted to the right compared to those of the GAFF force fields. Again, OPLS-AA-N provides the exception, with significantly taller first and second peaks, and the first peak further to the left than that of the other force fields. The GAFF-D1 RDF sits somewhere between that of OPLS-AA-N and the other GAFF force fields, getting more like the GAFF RDFs at higher concentrations, where the water-to-urea ratio decreases. In the dilute solutions neither Ishida et al.<sup>171</sup> or Duffy et al.<sup>102</sup> match the first peaks obtained here. However, at the higher concentrations, the Weerasinghe and Smith<sup>113</sup> curve matches the GAFF RDFs well, the Burton et al.<sup>172</sup> curve has a taller first peak but otherwise is similar to the OPLS RDFs and Soper et al.<sup>130</sup> has similarities to GAFF-D1.

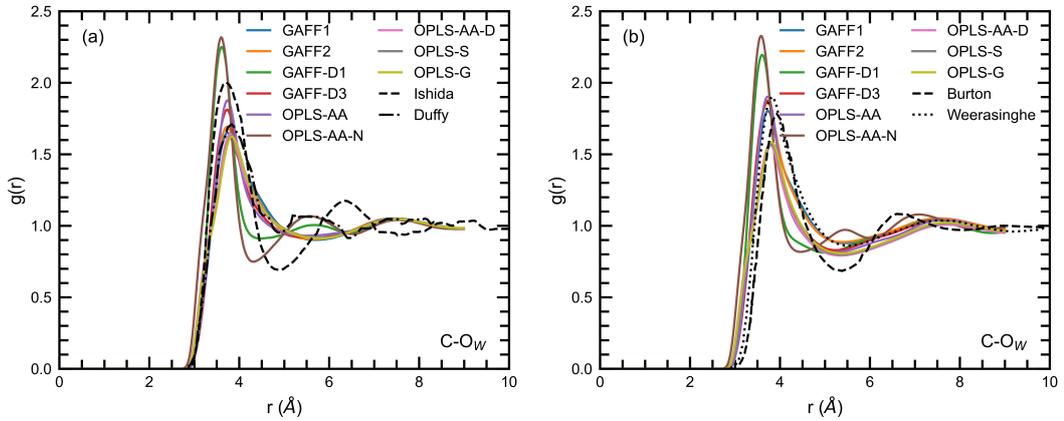
A.3.5 C–O<sub>W</sub>


Figure A.5: C–O<sub>W</sub> RDFs for (a) the dilute solutions and (b) the concentrated solutions. Literature data is taken from Ishida et al.<sup>171</sup>, Duffy et al.<sup>102</sup>, Burton et al.<sup>172</sup>, and Weerasinghe and Smith<sup>113</sup>.

The C–O<sub>W</sub> RDFs in Figure A.5 have one main peak at  $> 3.5 \text{ \AA}$  and a lower broad peak at  $\geq 7.5 \text{ \AA}$  as it is levelling off, the RDFs are similar for both the dilute and concentrated solutions. OPLS-AA-N and GAFF-D1 have significantly taller first peaks than the other force fields. In the dilute systems OPLS-AA-N and GAFF-D1 both have an additional peak at  $\sim 5.5 \text{ \AA}$ , this also occurred for Duffy et al.<sup>102</sup> this is also present at the higher concentrations for OPLS-AA-N only. The RDFs of Weerasinghe and Smith<sup>113</sup> match well, so does Burton et al.<sup>172</sup> although the second peak is shifted slightly earlier and this is significantly earlier in Ishida et al.<sup>171</sup>

## Appendix B

# Additional Details and Analyses for Interfacial Studies

### B.1 Force Field Parameters and Equations

This section lists all the equations and parameters describing the bonded and non-bonded interactions for the GAFF-D3 (urea) and SPC/E (water) force fields used in the interfacial simulations in Chapter 5.

The force field equations have already been defined in Appendix A.1 as follows: The harmonic bond and angle potentials given in Eqs. A.1 and A.3 are used together with the fourier potential for both dihedral and improper dihedral interactions given in Eqs. A.5 and A.8. The Coulomb potential given in Eq. A.10 is used for the non-bonded electrostatic interactions and the Lennard-Jones 12-6 potential given in Eq. A.11 is used for the non-bonded dispersion interactions between all the individual atoms.

The GAFF-D3 urea parameters are taken from Özpınar et al.<sup>111</sup> and the SPC/E water parameters are taken from Mark and Nilsson<sup>133</sup>, Berendsen et al.<sup>135</sup>

Table B.1: Bonded force field parameters.

Bonds	$K_r$ (kcal mol <sup>-1</sup> Å <sup>-2</sup> )	$r_0$ (Å)	
C - O	656	1.250	
C - N	424	1.383	
N - H	434	1.010	
O - H (water)	1000.0	1.000	
Angles	$K_\theta$ (kcal mol <sup>-1</sup> rad <sup>-2</sup> )	$\theta_0$ (°)	
C - N - H	30	120.0	
N - C - O	80	120.9	
N - C - N	70	118.6	
H - N - H	35	120.0	
H - O - H (water)	100.0	109.47	
Dihedrals	$K_\phi$ (kcal mol <sup>-1</sup> )	$n$	$d$ (°)
H - N - C - O	2.50	2	180.0
H - N - C - O	2.00	1	0.0
Improper dihedrals	$K_\phi$ (kcal mol <sup>-1</sup> )	$n$	$d$ (°)
C - N - H - H	1.1	2	180.0
O - C - N - N	10.5	2	180.0

Table B.2: Non-bonded force field parameters.

Atom type	Mass (Da)	Charge (e)	$\epsilon$ (kcal mol <sup>-1</sup> )	$\sigma$ (Å)
C	12.01	0.884	0.0860	3.3997
O	16.00	-0.660	0.2100	2.9599
N	14.01	-0.888	0.1700	3.2500
H	1.008	0.388	0.0157	1.0691
O (water)	16.00	-0.8476	0.1553	3.1656
H (water)	1.008	0.4238	0.0000	0.0000

GAFF typically present the  $\sigma$  (Lennard-Jones diameter) values as  $R_{\min}$  (van der Waals radius) values,  $\sigma = 2^{\frac{5}{6}} R_{\min}$  was used for the conversion (from  $2R_{\min} = 2^{\frac{1}{6}}\sigma$ ).

The LJ parameters for the interfacial materials are given in Table B.3. PTFE and diamond use parameters from GAFF<sup>99</sup> and  $\alpha$ -quartz uses specially parametrised parameters.<sup>186</sup>

Table B.3: Non-bonded force field parameters for the interface materials.

Material	Atom type	$\epsilon$ (kcal mol <sup>-1</sup> )	$\sigma$ (Å)
PTFE	C	0.1094	3.40
PTFE	F	0.061	3.12
Diamond	C	0.1094	3.3997
$\alpha$ -quartz	Si	0.093	4.15
$\alpha$ -quartz	O	0.054	3.47

## B.2 Simulation Results

### B.2.1 Density Profile Plots

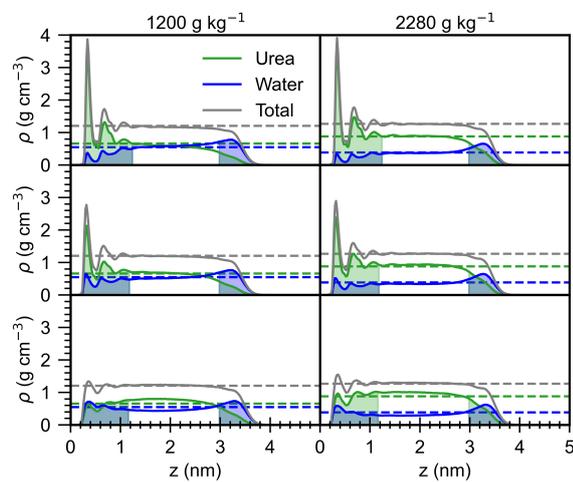


Figure B.1: Simulated density profiles for the small interfacial systems.

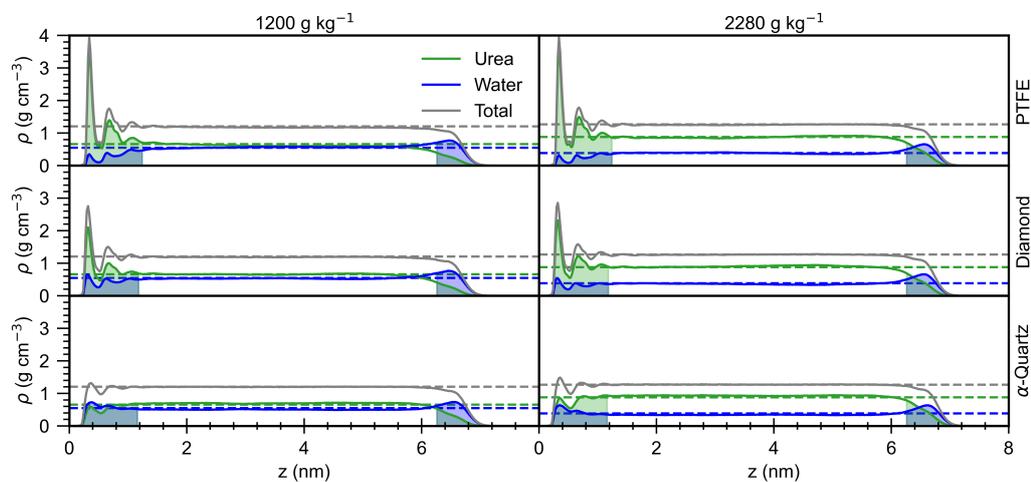


Figure B.2: Simulated density profiles for the medium interfacial systems.

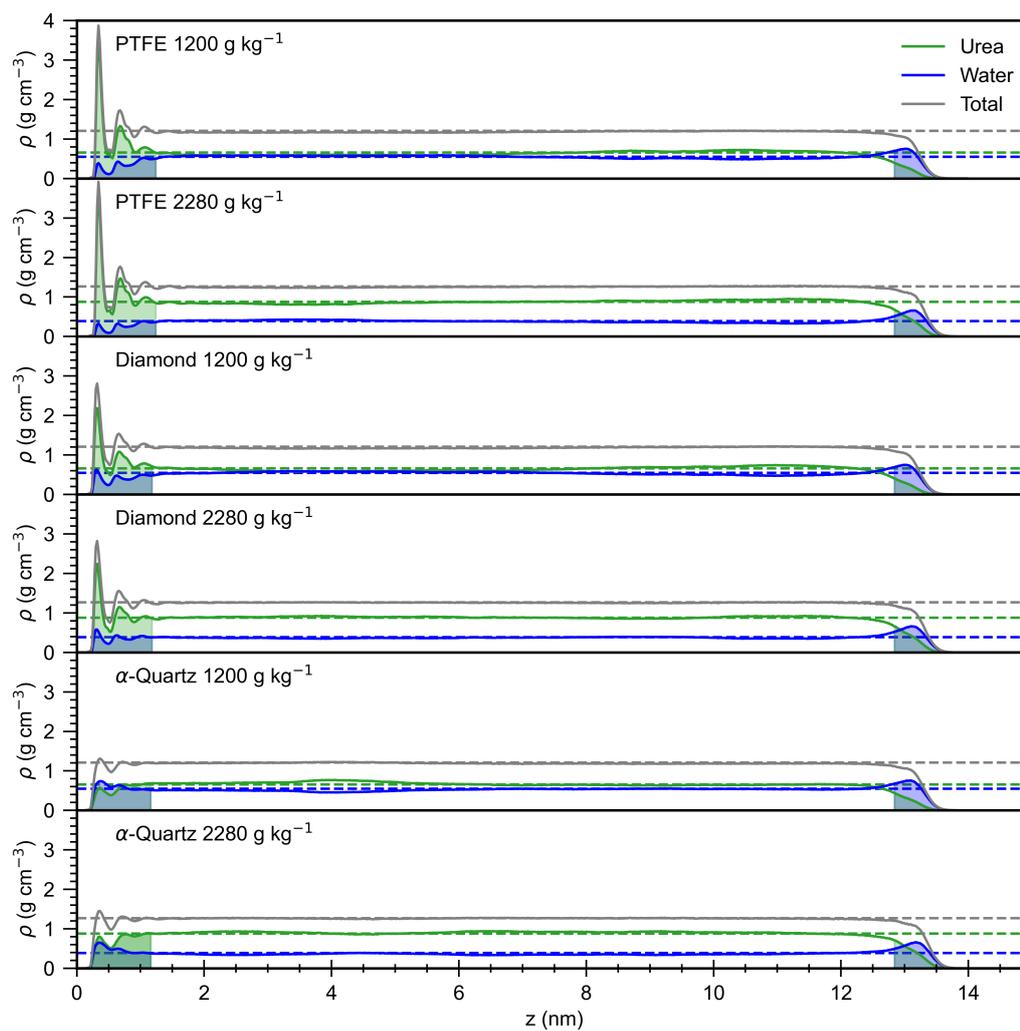


Figure B.3: Simulated density profiles for the large interfacial systems.

## B.2.2 Interfacial Density Profiles

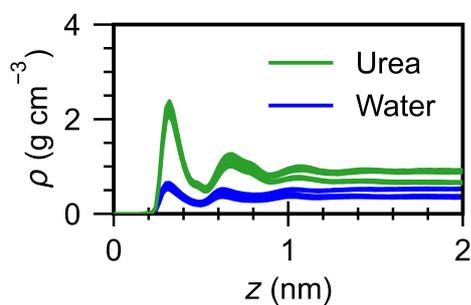


Figure B.4: Solution structure at the diamond LJ interface. All six simulations, from the three system sizes and the two concentrations, have been overlaid.

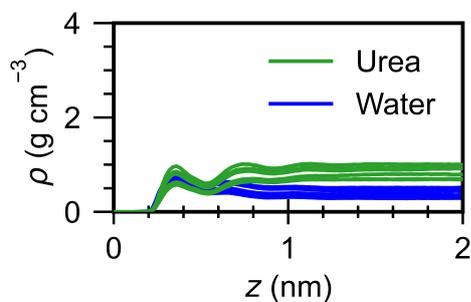


Figure B.5: Solution structure at the  $\alpha$ -quartz LJ interface. All six simulations, from the three system sizes and the two concentrations, have been overlaid.

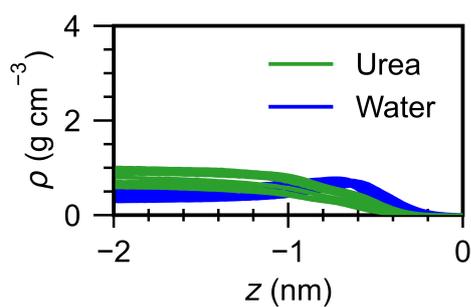


Figure B.6: Solution structure at the vacuum interface. All eighteen simulations, from the three system sizes and the two concentrations, for each of the interfaces have been overlaid.

### B.2.3 Average Interfacial Density Tables

Table B.4: Interfacial urea and water concentrations at the Lennard Jones and vacuum interfaces.

Interface	$c$ (g kg <sup>-1</sup> )	Size	LJ interface (g cm <sup>-3</sup> )		Vacuum (g cm <sup>-3</sup> )	
			$\rho_{\text{urea}}$	$\rho_{\text{water}}$	$\rho_{\text{urea}}$	$\rho_{\text{water}}$
PTFE	1200	small	1.03 ± 0.09	0.35 ± 0.02	0.19 ± 0.03	0.40 ± 0.04
PTFE	1200	medium	1.08 ± 0.10	0.32 ± 0.02	0.18 ± 0.03	0.40 ± 0.04
PTFE	1200	large	1.03 ± 0.09	0.35 ± 0.02	0.15 ± 0.02	0.39 ± 0.04
PTFE	2280	small	1.16 ± 0.09	0.26 ± 0.01	0.26 ± 0.04	0.32 ± 0.03
PTFE	2280	medium	1.16 ± 0.09	0.26 ± 0.01	0.26 ± 0.04	0.33 ± 0.04
PTFE	2280	large	1.15 ± 0.09	0.27 ± 0.01	0.23 ± 0.03	0.35 ± 0.04
Diamond	1200	small	0.85 ± 0.06	0.42 ± 0.02	0.14 ± 0.02	0.42 ± 0.05
Diamond	1200	medium	0.83 ± 0.06	0.44 ± 0.02	0.15 ± 0.02	0.42 ± 0.04
Diamond	1200	large	0.88 ± 0.06	0.41 ± 0.02	0.16 ± 0.02	0.40 ± 0.04
Diamond	2280	small	1.03 ± 0.06	0.30 ± 0.01	0.24 ± 0.03	0.35 ± 0.04
Diamond	2280	medium	0.99 ± 0.06	0.33 ± 0.01	0.23 ± 0.03	0.34 ± 0.04
Diamond	2280	large	0.95 ± 0.06	0.36 ± 0.01	0.23 ± 0.03	0.34 ± 0.04
Quartz	1200	small	0.58 ± 0.03	0.53 ± 0.02	0.16 ± 0.02	0.42 ± 0.04
Quartz	1200	medium	0.53 ± 0.02	0.55 ± 0.02	0.16 ± 0.02	0.42 ± 0.04
Quartz	1200	large	0.53 ± 0.02	0.56 ± 0.02	0.16 ± 0.02	0.40 ± 0.04
Quartz	2280	small	0.82 ± 0.04	0.37 ± 0.02	0.26 ± 0.04	0.34 ± 0.04
Quartz	2280	medium	0.74 ± 0.03	0.42 ± 0.02	0.24 ± 0.03	0.35 ± 0.04
Quartz	2280	large	0.70 ± 0.03	0.44 ± 0.02	0.23 ± 0.03	0.35 ± 0.04

### B.2.4 Interfacial Concentration during Equilibration

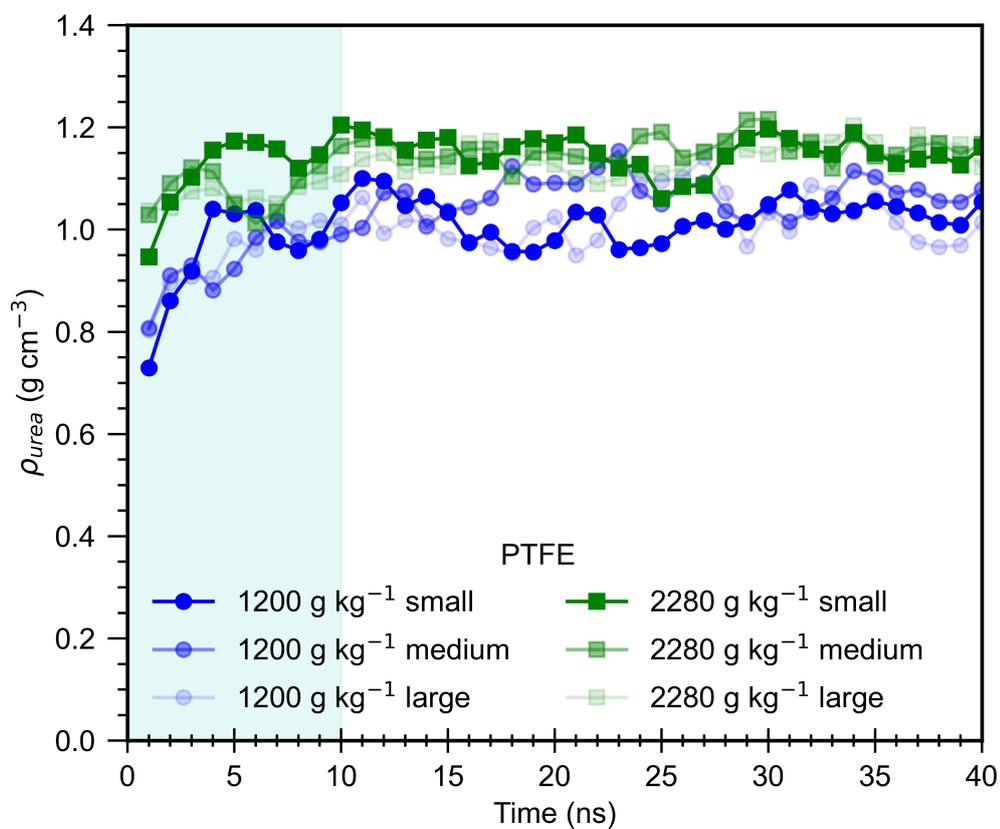


Figure B.7: Changes in interfacial urea concentration with time at the PTFE surface. The shaded area from 0-10 ns is the equilibration period and the production run is 10-40 ns.

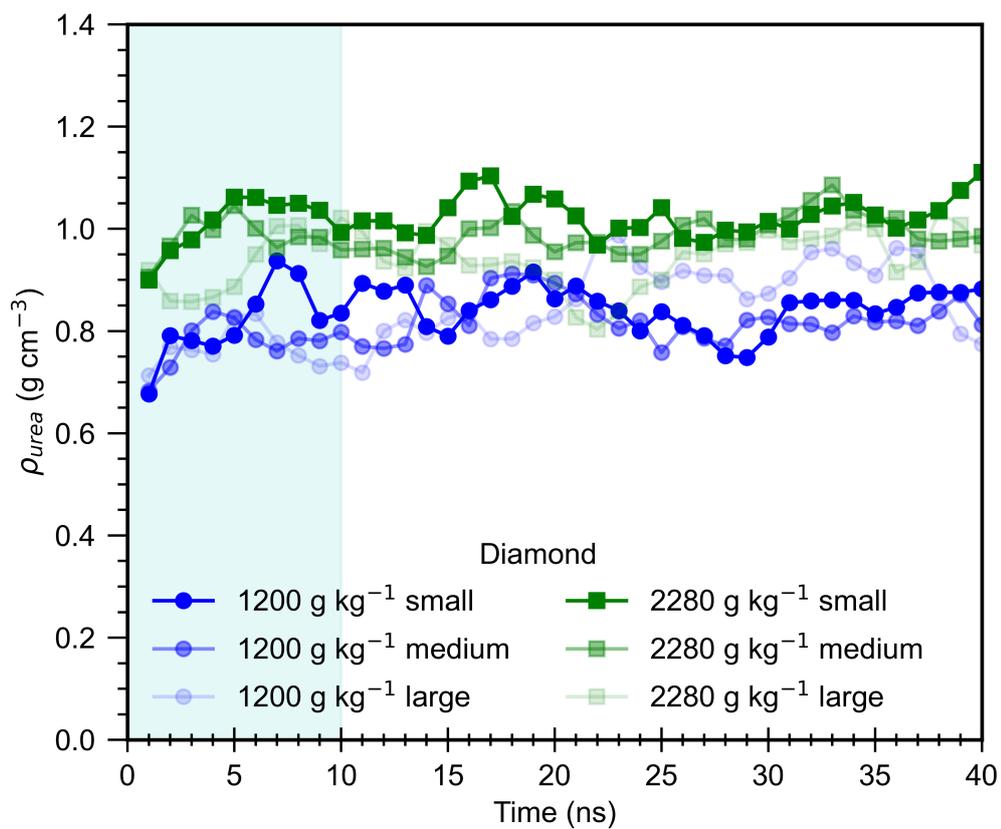


Figure B.8: Changes in interfacial urea concentration with time at the diamond surface. The shaded area from 0-10 ns is the equilibration period and the production run is 10-40 ns.

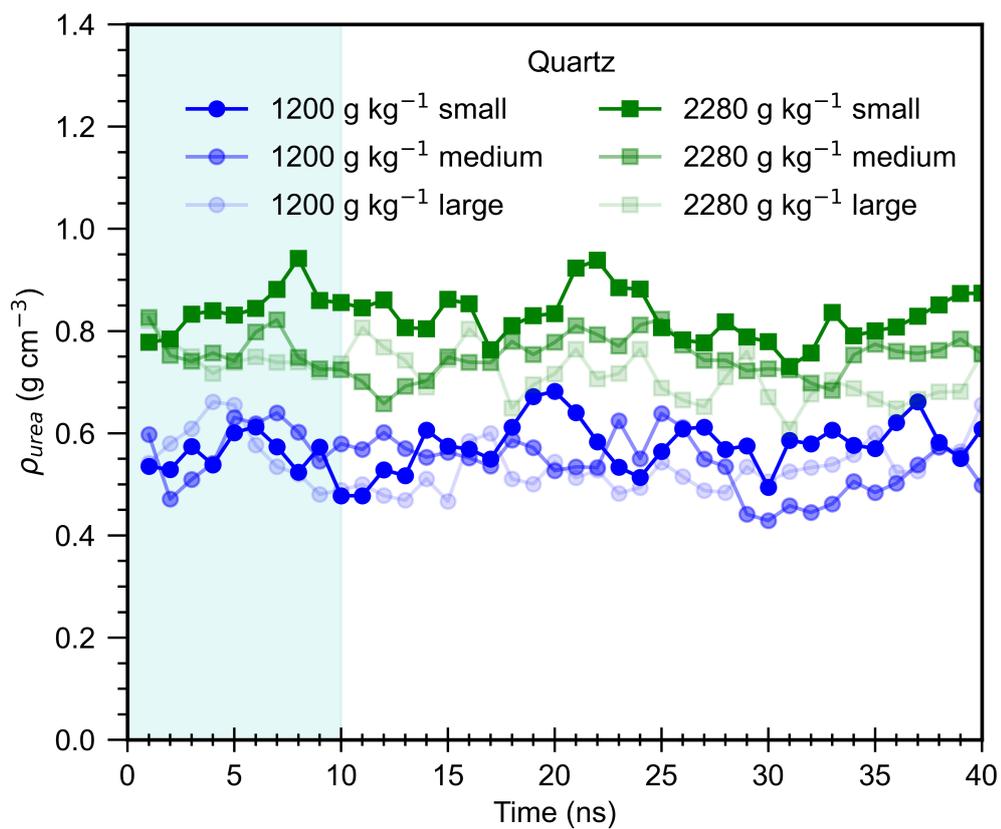


Figure B.9: Changes in interfacial urea concentration with time at the  $\alpha$ -quartz surface. The shaded area from 0-10 ns is the equilibration period and the production run is 10-40 ns.

### B.2.5 $P_2$ Profile Plots

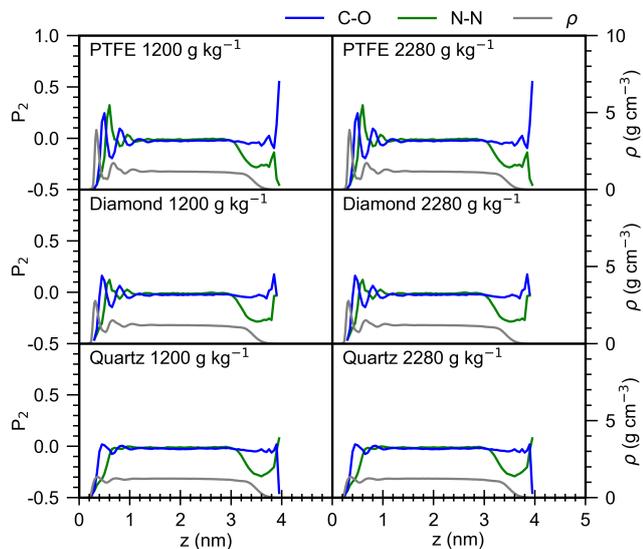


Figure B.10:  $P_2$  profiles for the small interfacial systems, with total solution density profile for comparison.

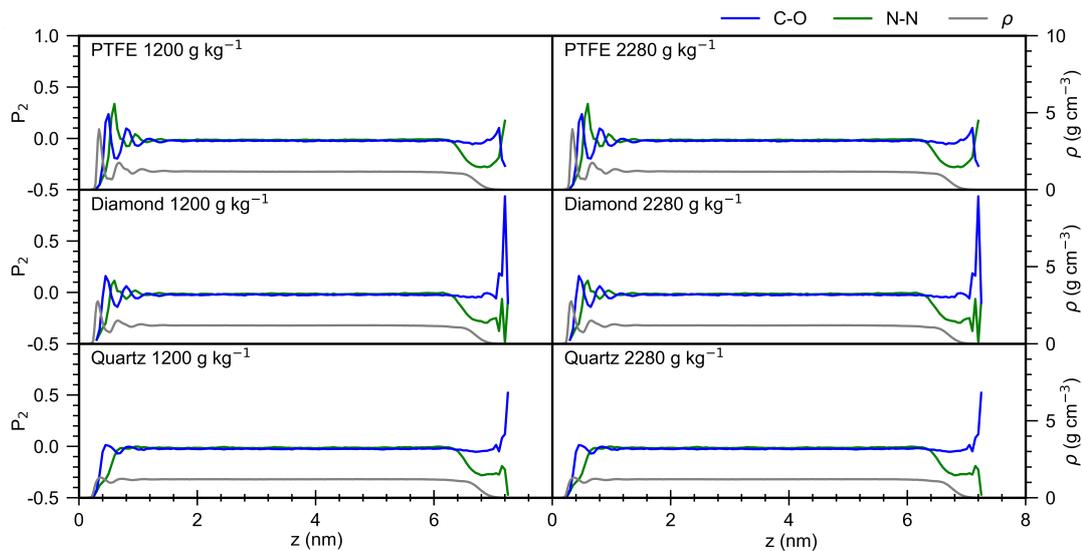


Figure B.11:  $P_2$  profiles for the medium interfacial systems, with total solution density profile for comparison.

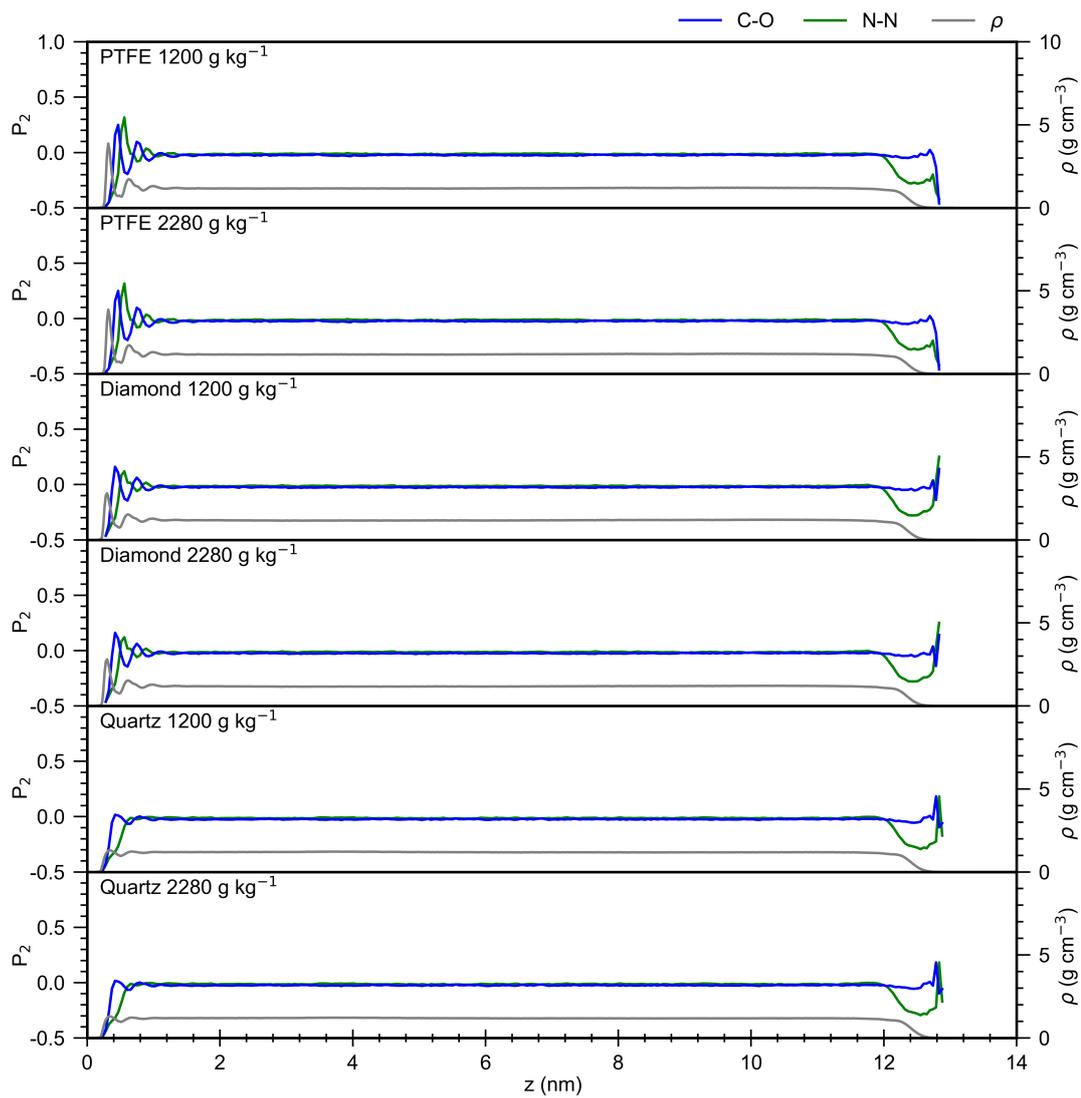


Figure B.12:  $P_2$  profiles for the large interfacial systems, with total solution density profile for comparison.

## **B.3 Experimental Results**

### **B.3.1 Unconditional Cumulative Distribution**

The unconditional cumulative probability nucleation distribution includes the samples which nucleated before the beginning of the data collection stage of the experiment. This is shown in Figure B.13, where all the early nucleation events are considered to have occurred at time zero, since no further details are available.

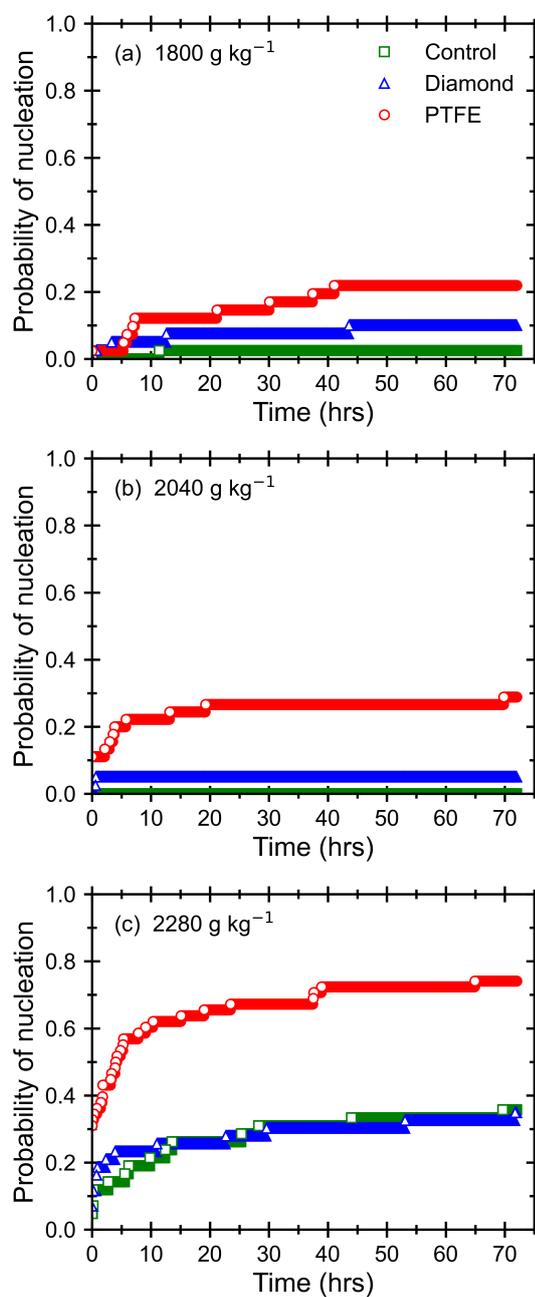


Figure B.13: Cumulative, unconditional probability distribution function of induction times of urea crystallisation from aqueous solution with heterogeneous interfaces PTFE, diamond and glass for (a) 1800 g kg<sup>-1</sup>, (b) 2040 g kg<sup>-1</sup> and (c) 2280 g kg<sup>-1</sup>.

### B.3.2 Fitted Nucleation Rates

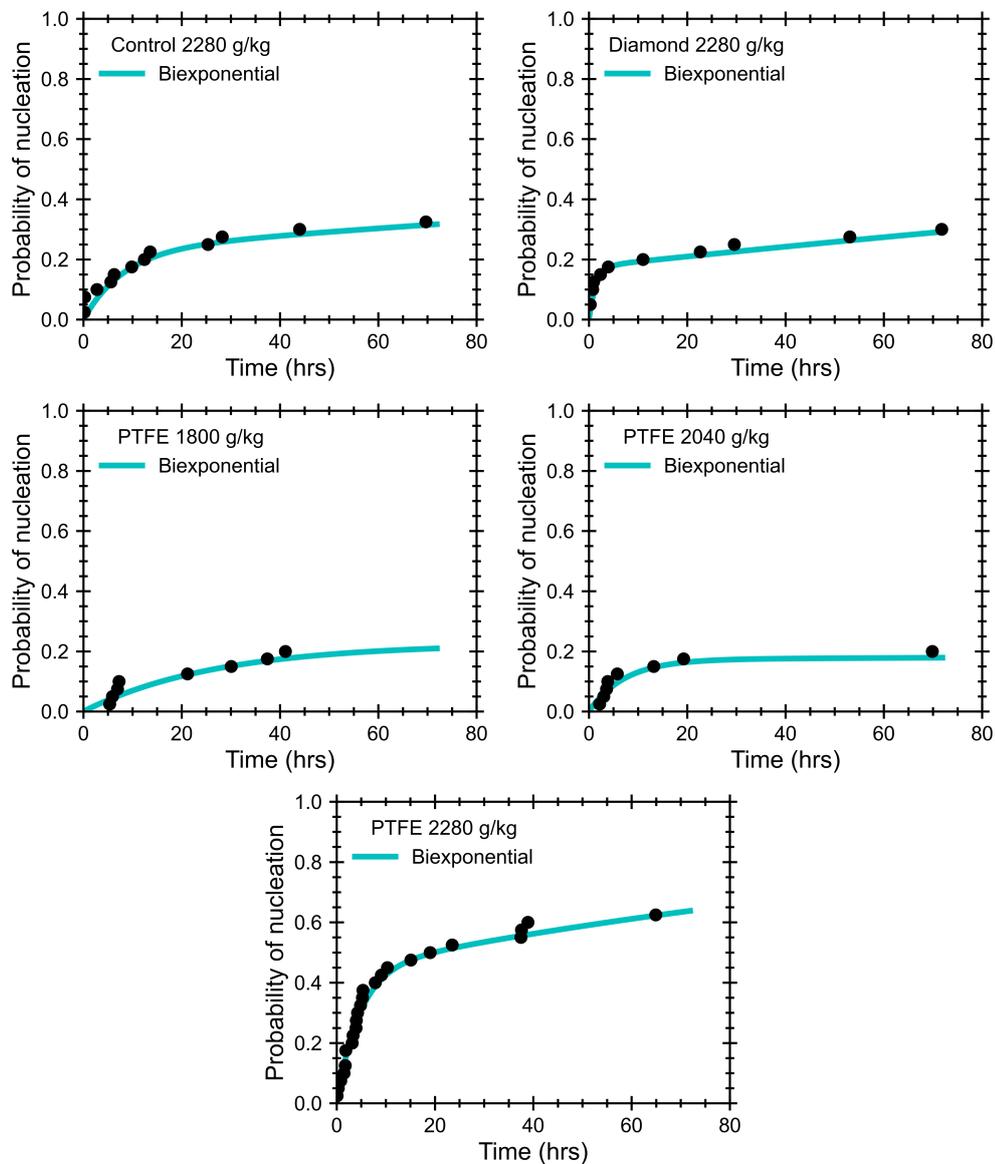


Figure B.14: Fitted nucleation rates with experimental nucleation times.

### B.3.3 Detailed Breakdown of Individual Diamonds

Table B.5: Detailed breakdown of the individual diamonds and how frequently they induced nucleation, for each concentration. A-F represent the six individual diamonds, X represents the experiments where the diamonds were not individually tracked.

System		A	B	C	D	E	F	X
1800 g kg <sup>-1</sup>	Nucleation	0	0	0	0	0	0	4
	No nucleation	1	1	1	1	0	0	32
2040 g kg <sup>-1</sup>	Nucleation	0	1	0	0	1	0	0
	No nucleation	6	5	6	6	4	5	6
2280 g kg <sup>-1</sup>	Nucleation	1	0	0	2	1	1	7
	No nucleation	1	2	2	0	2	1	20

## Appendix C

# Additional Temperature Cycling Analyses

### C.1 Breakdown of Experiments A-C

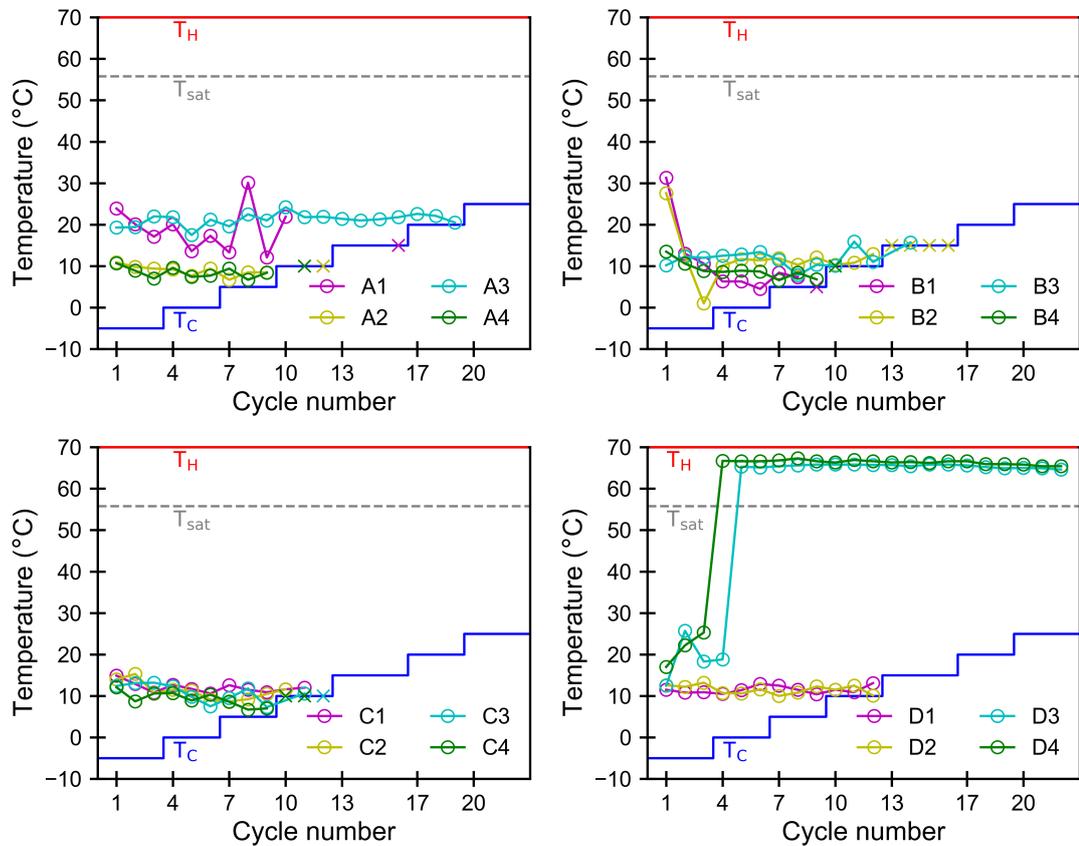


Figure C.1: Nucleation conditions for the screening Experiment A with variable temperature cycles for each zone individually. Circles represent nucleation during the cooling ramp, crosses represent nucleation during the hold period, and no symbol indicates no nucleation.

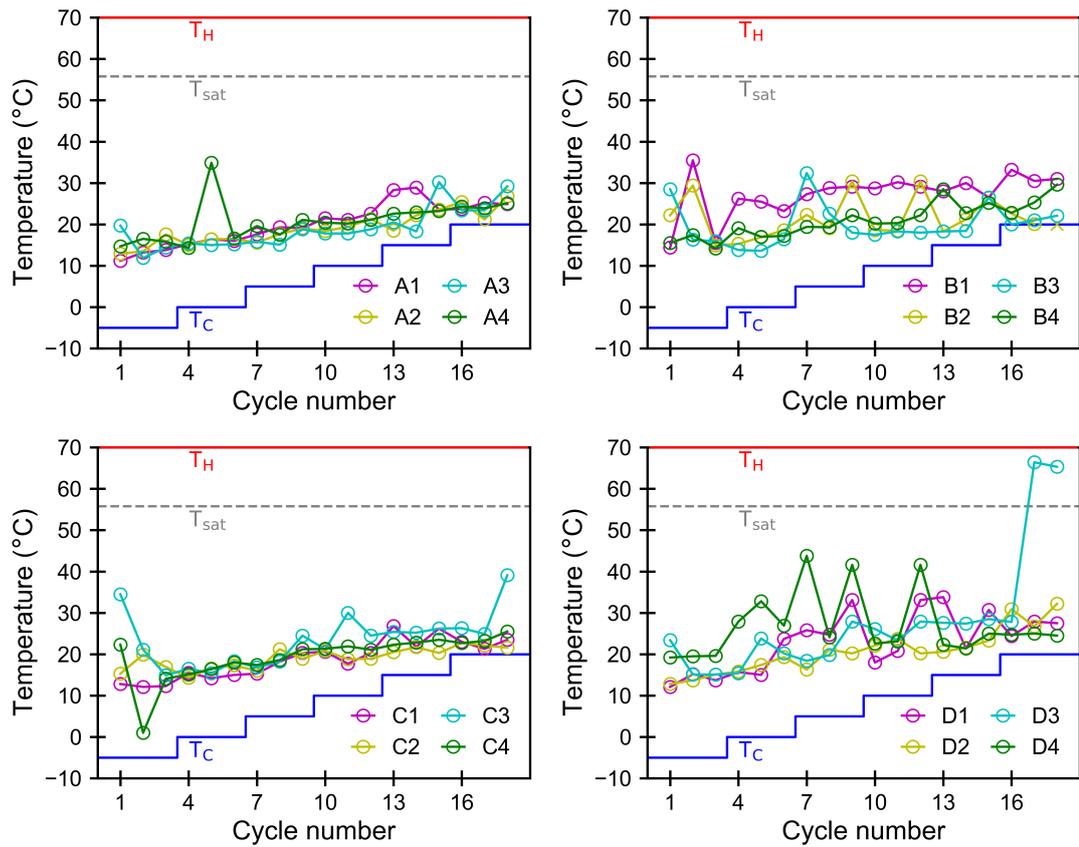


Figure C.2: Nucleation conditions for the screening Experiment B with variable temperature cycles for each zone individually. Circles represent nucleation during the cooling ramp, crosses represent nucleation during the hold period, and no symbol indicates no nucleation.

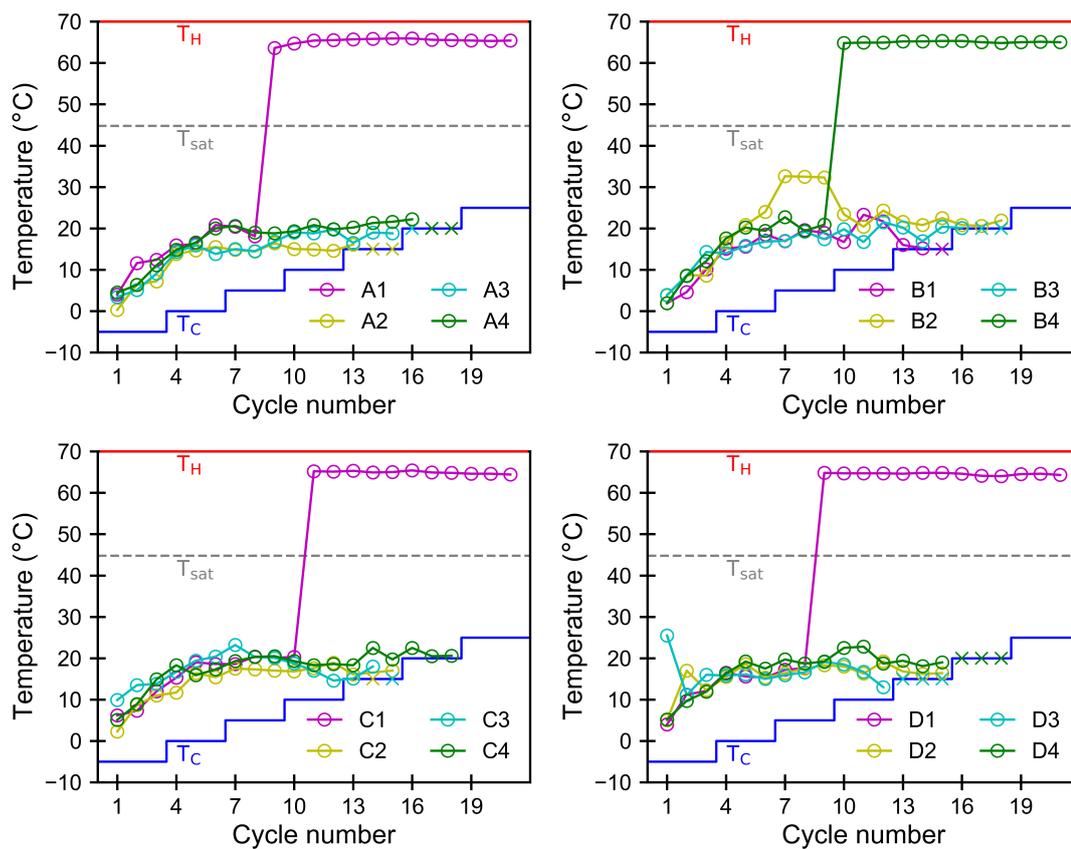


Figure C.3: Nucleation conditions for the screening Experiment B with variable temperature cycles for each zone individually. Circles represent nucleation during the cooling ramp, crosses represent nucleation during the hold period, and no symbol indicates no nucleation.

## **C.2 Mass Change**

The change in mass of the vials over the duration of each experiments is recorded in Table C.1. Only two mass measurements was made for each vial, the mass of the vial with solution and cap, at the beginning and end of the experiment. The mass change in most vials was negligible, ranging from a small loss, no change to a small increase, which could be attributed to the uncertainties involved in weighing. Some vials had a significant mass reduction over the duration of the experiment, this was reduced but not eliminated by the additional Parafilm seal in Experiments D–F. We assume that the mass loss is entirely due to the evaporation of water, which will have led to an increase in the concentration of the urea solution in the affected vials.

It is useful to estimate the mass of solution in each vial, to put the change in mass into context, since it can be assumed that this is entirely due to loss of solution. Estimates based on extrapolated urea solution density based on data from Gucker et al.<sup>170</sup>, using the recommended quadratic relationship and assumes that exactly 1.00 ml of solution was pipetted into each vial. The estimated solution mass in each vial is 1.200 g for Experiments A, B, D–F, and 1.184 g for Experiment C.

Table C.1: Mass change of vials, in g, for each experiment. Vials with significant mass loss (greater than 0.02 g) have been highlighted in boldface.

Vial	A	B	C	D	E	F
Total mass	1.200	1.200	1.184	1.200	1.200	1.200
A1	-0.0081	<b>-0.0622</b>	<b>-0.0504</b>	-0.0014	0.0002	0.0001
A2	-0.0003	-0.0011	0.0001	-0.0015	-0.0001	-0.0004
A3	-0.0030	<b>-0.0673</b>	-0.0191	-0.0011	-0.0002	<b>-0.0374</b>
A4	-0.0045	-0.0106	-0.0087	-0.0010	-0.0003	-0.0097
B1	-0.0002	0.0000	-0.0185	-0.0013	-0.0033	-0.0005
B2	-0.0002	-0.0018	<b>-0.0530</b>	-0.0004	-0.0011	-0.0006
B3	-0.0084	0.0006	<b>-0.0537</b>	-0.0007	0.0002	<b>-0.0217</b>
B4	-0.0097	0.0000	<b>-0.0391</b>	-0.0012	-0.0004	-0.0009
C1	0.0003	-0.0001	-0.0062	-0.0003	0.0000	-0.0002
C2	-0.0027	-0.0006	-0.0055	-0.0015	-0.0001	-0.0099
C3	-0.0002	-0.0019	-0.0086	-0.0016	-0.0002	-0.0002
C4	-0.0053	-0.0012	-0.0096	-0.0004	-0.0066	-0.0006
D1	-0.0069	-0.0004	0.0000	<b>-0.0222</b>	-0.0001	-0.0002
D2	-0.0004	0.0003	-0.0036	-0.0016	-0.0001	0.0001
D3	-0.0023	<b>-0.0966</b>	-0.0142	-0.0010	0.0000	-0.0003
D4	-0.0003	-0.0015	-0.0063	-0.0067	-0.0003	-0.0004

### C.3 Unconditional Probability Distributions

The unconditional probability distributions are shown here in Figure C.4. In Experiments D–F, the conditional and unconditional probability distributions are relatively similar, which can be seen when comparing to the conditional probability distribution plot in Figure 6.7. However, the difference between two experiments at 25 °C is significantly greater when considering the unconditional probability, due to significantly increased number of vials nucleation during cooling in using the experimental method of Chapter 5.

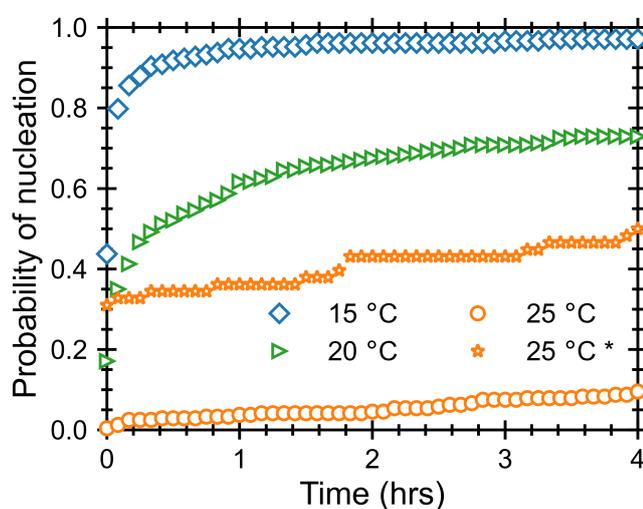


Figure C.4: Unconditional nucleation induction times for Experiments D (15 °C), E (20 °C), F (25 °C) and the equivalent results from Chapter 5 denoted by \* (25 °C, 2280 g kg<sup>-1</sup>, PTFE). For clarity, the symbols are only plotted every five minutes, however, for Experiments D–F measurements were made every second.

### C.4 Fitted Nucleation Rates

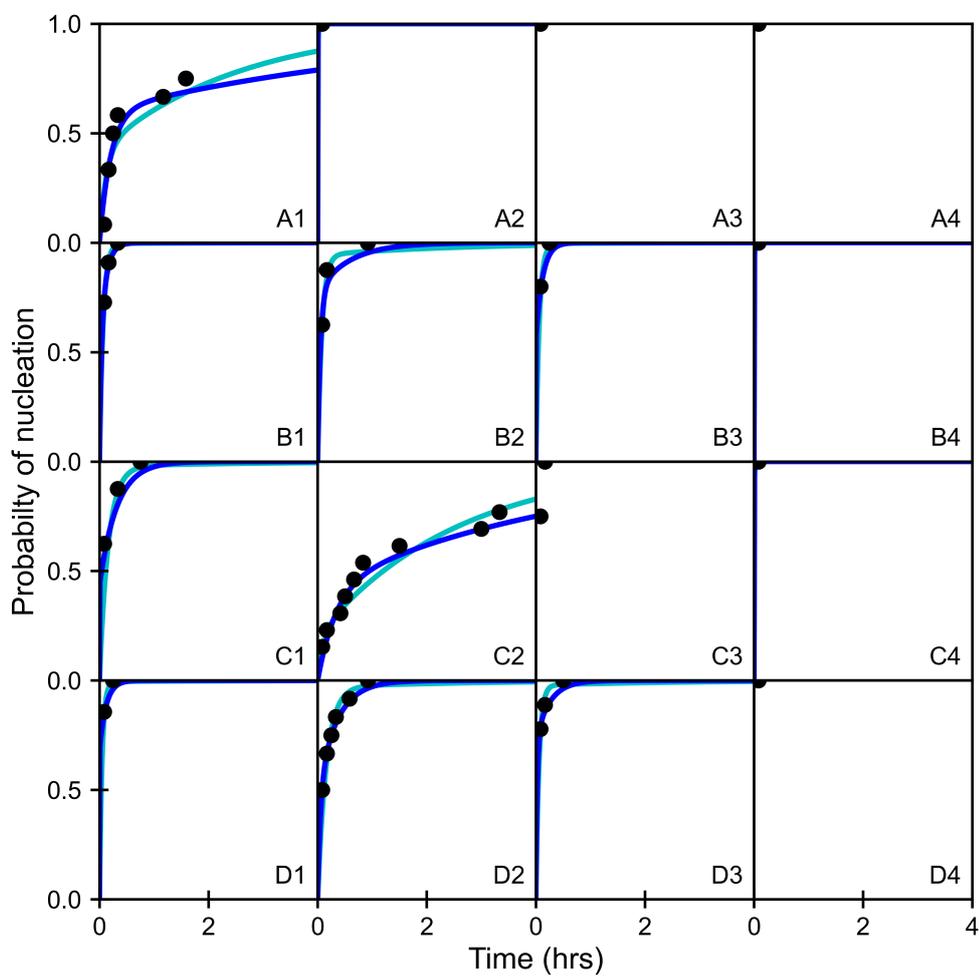


Figure C.5: Fitted nucleation rates for individual vials in Experiment D (15 °C). The blue line is the unconstrained fit and the cyan line is the fit with  $\tau_2$  constrained.

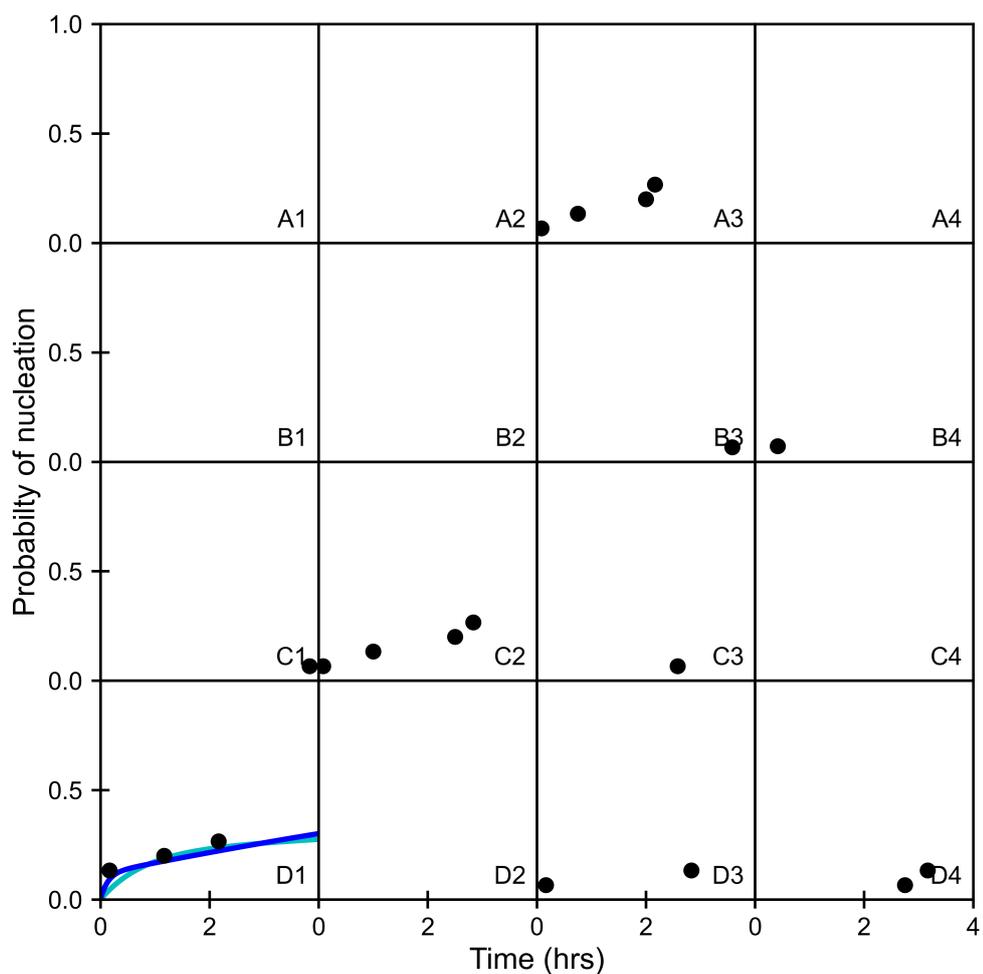


Figure C.6: Fitted nucleation rates for individual vials in Experiment F (25 °C). The blue line is the unconstrained fit and the cyan line is the fit with  $\tau_2$  constrained.

# Bibliography

- [1] Davey, R. J.; Schroeder, S. L. M.; ter Horst, J. H. Nucleation of Organic Crystals-A Molecular Perspective. *Angew. Chem. Int. Ed.* **2013**, *52*, 2166–2179.
- [2] Chen, J.; Sarma, B.; Evans, J. M. B.; Myerson, A. S. Pharmaceutical Crystallization. *Crystal Growth & Design* **2011**, *11*, 887–895.
- [3] Salvalaglio, M.; Vetter, T.; Giberti, F.; Mazzotti, M.; Parrinello, M. Uncovering Molecular Details of Urea Crystal Growth in the Presence of Additives. *J. Am. Chem. Soc.* **2012**, *134*, 17221–17233.
- [4] Finney, A. R.; Salvalaglio, M. Molecular Simulation Approaches to Study Crystal Nucleation from Solutions: Theoretical Considerations and Computational Challenges. *WIREs Comput Mol Sci* **2024**, *14*, e1697.
- [5] Price, S. L. Why Don't We Find More Polymorphs? *Acta Crystallogr B Struct Sci Cryst Eng Mater* **2013**, *69*, 313–328.
- [6] Neumann, M. A.; van de Streek, J.; Fabbiani, F. P. A.; Hidber, P.; Grassmann, O. Combined Crystal Structure Prediction and High-Pressure Crystallization in Rational Pharmaceutical Polymorph Screening. *Nature Communications* **2015**, *6*, 7793.
- [7] Francia, N. F.; Price, L. S.; Nyman, J.; Price, S. L.; Salvalaglio, M. Systematic Finite-Temperature Reduction of Crystal Energy Landscapes. *Crystal Growth & Design* **2020**, *20*, 6847–6862.

- [8] Lee, E. H. A Practical Guide to Pharmaceutical Polymorph Screening & Selection. *Asian Journal of Pharmaceutical Sciences* **2014**, *9*, 163–175.
- [9] Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: An Extraordinary Example of Conformational Polymorphism. *Pharm Res* **2001**, *18*, 8.
- [10] Bauer, J. F. Polymorphism—A Critical Consideration in Pharmaceutical Development, Manufacturing, and Stability. *Journal of validation technology* **2008**, *14*, 15–23.
- [11] Bučar, D.-K.; Lancaster, R. W.; Bernstein, J. Disappearing Polymorphs Revisited. *Angew. Chem. Int. Ed.* **2015**, *54*, 6972–6993.
- [12] Zahn, D. Thermodynamics and Kinetics of Prenucleation Clusters, Classical and Non-Classical Nucleation. *ChemPhysChem* **2015**, *16*, 2069–2075.
- [13] Karthika, S.; Radhakrishnan, T. K.; Kalaichelvi, P. A Review of Classical and Nonclassical Nucleation Theories. *Crystal Growth & Design* **2016**, *16*, 6663–6681.
- [14] Mazzotti, M.; Vetter, T.; Ochsenbein, D. R.; Maggioni, G. M.; Lindenberg, C. In *Polymorphism in the Pharmaceutical Industry*, 1st ed.; Hilfiker, R., Raumer, M. V., Eds.; Wiley, 2018; pp 261–283.
- [15] Vekilov, P. G. Nucleation. *Crystal Growth & Design* **2010**, *10*, 5007–5019.
- [16] Gebauer, D.; Cölfen, H. Prenucleation Clusters and Non-Classical Nucleation. *Nano Today* **2011**, *6*, 564–584.
- [17] Gebauer, D.; Kellermeier, M.; Gale, J. D.; Bergström, L.; Cölfen, H. Prenucleation Clusters as Solute Precursors in Crystallisation. *Chem. Soc. Rev.* **2014**, *43*, 2348–2371.
- [18] Jawor-Baczynska, A.; Moore, B. D.; Lee, H. S.; McCormick, A. V.; Sefcik, J. Population and Size Distribution of Solute-Rich Mesospecies within Mesostructured Aqueous Amino Acid Solutions. *Faraday Discuss.* **2013**, *167*, 425.

- [19] Jawor-Baczynska, A.; Moore, B. D.; Sefcik, J. Effect of Mixing, Concentration and Temperature on the Formation of Mesoscaled Solutions and Their Role in the Nucleation of DL -Valine Crystals. *Faraday Discuss.* **2015**, *179*, 141–154.
- [20] Zimbitas, G.; Jawor-Baczynska, A.; Vesga, M. J.; Javid, N.; Moore, B. D.; Parkinson, J.; Sefcik, J. Investigation of Molecular and Mesoscale Clusters in Undersaturated Glycine Aqueous Solutions. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2019**, *579*, 123633.
- [21] Thakore, S. D.; Sood, A.; Bansal, A. K. Emerging Role of Primary Heterogeneous Nucleation in Pharmaceutical Crystallization. *Drug Dev Res* **2020**, *81*, 3–22.
- [22] O’Sullivan, B.; Smith, B.; Baramidze, G. *Recent Advances for Seeding a Crystallization Process*; 2012.
- [23] Anwar, J.; Khan, S.; Lindfors, L. Secondary Crystal Nucleation: Nuclei Breeding Factory Uncovered. *Angew Chem Int Ed* **2015**, *54*, 14681–14684.
- [24] Qian, R.-Y.; Botsaris, G. D. Nuclei Breeding from a Chiral Crystal Seed of NaClO<sub>3</sub>. *Chemical Engineering Science* **1998**, *53*, 1745–1756.
- [25] Vesga, M. J.; McKechnie, D.; Mulheran, P. A.; Johnston, K.; Sefcik, J. Conundrum of  $\gamma$  Glycine Nucleation Revisited: To Stir or Not to Stir? *CrystEngComm* **2019**, *21*, 2234–2243.
- [26] McKechnie, D.; Anker, S.; Zahid, S.; Mulheran, P. A.; Sefcik, J.; Johnston, K. Interfacial Concentration Effect Facilitates Heterogeneous Nucleation from Solution. *J. Phys. Chem. Lett.* **2020**, *11*, 2263–2271.
- [27] Parambil, J. V.; Poornachary, S. K.; Tan, R. B. H.; Heng, J. Y. Y. Template-Induced Polymorphic Selectivity: The Effects of Surface Chemistry and Solute Concentration on Carbamazepine Crystallisation. *CrystEngComm* **2014**, *16*, 4927–4930.
- [28] Seyedhosseini, E.; Ivanov, M.; Bystrov, V.; Bdikin, I.; Zelenovskiy, P.; Shur, V. Y.; Kudryavtsev, A.; Mishina, E. D.; Sigov, A. S.; Kholkin, A. L. Growth and Non-

- linear Optical Properties of  $\beta$ -Glycine Crystals Grown on Pt Substrates. *Crystal Growth & Design* **2014**, *14*, 2831–2837.
- [29] Yang, H.; Song, C. L.; Lim, Y. X. S.; Chen, W.; Heng, J. Y. Y. Selective Crystallisation of Carbamazepine Polymorphs on Surfaces with Differing Properties. *CrystEngComm* **2017**, *19*, 6573–6578.
- [30] Frostman, L. M.; Bader, M. M.; Ward, M. D. Nucleation and Growth of Molecular Crystals on Self-Assembled Monolayers. *Langmuir* **1994**, *10*, 576–582.
- [31] Lee, A. Y.; Lee, I. S.; Myerson, A. S. Factors Affecting the Polymorphic Outcome of Glycine Crystals Constrained on Patterned Substrates. *Chem. Eng. Technol.* **2006**, *29*, 281–285.
- [32] Lee, I. S.; Kim, K. T.; Lee, A. Y.; Myerson, A. S. Concomitant Crystallization of Glycine on Patterned Substrates: The Effect of pH on the Polymorphic Outcome. *Crystal Growth & Design* **2008**, *8*, 108–113.
- [33] Testa, C. J.; Shvedova, K.; Hu, C.; Wu, W.; Born, S. C.; Takizawa, B.; Mascia, S. Heterogeneous Crystallization as a Process Intensification Technology in an Integrated Continuous Manufacturing Process for Pharmaceuticals. *Org. Process Res. Dev.* **2021**, *25*, 225–238.
- [34] Arribas Bueno, R.; Crowley, C. M.; Hodnett, B. K.; Hudson, S.; Davern, P. Influence of Process Parameters on the Heterogeneous Nucleation of Active Pharmaceutical Ingredients onto Excipients. *Org. Process Res. Dev.* **2017**, *21*, 559–570.
- [35] Chayen, N. E.; Shaw Stewart, P. D.; Blow, D. M. Microbatch Crystallization under Oil — a New Technique Allowing Many Small-Volume Crystallization Trials. *Journal of Crystal Growth* **1992**, *122*, 176–180.
- [36] Candoni, N.; Grossier, R.; Lagaize, M.; Veessler, S. Advances in the Use of Microfluidics to Study Crystallization Fundamentals. *Annu. Rev. Chem. Biomol. Eng.* **2019**, *10*, 59–83.

- [37] Bolanos-Garcia, V. M. The Use of Oil in a Counter-Diffusive System Allows to Control Nucleation and Coarsening during Protein Crystallization. *Journal of Crystal Growth* **2003**, *253*, 517–523.
- [38] Little, L. J.; Sear, R. P.; Keddie, J. L. Does the  $\gamma$  Polymorph of Glycine Nucleate Faster? A Quantitative Study of Nucleation from Aqueous Solution. *Crystal Growth & Design* **2015**, *15*, 5345–5354.
- [39] Ildefonso, M.; Candoni, N.; Veessler, S. Heterogeneous Nucleation in Droplet-Based Nucleation Measurements. *Crystal Growth & Design* **2013**, *13*, 2107–2110.
- [40] Fatemi, N.; Dong, Z.; Van Gerven, T.; Kuhn, S. Microbubbles as Heterogeneous Nucleation Sites for Crystallization in Continuous Microfluidic Devices. *Langmuir* **2019**, *35*, 60–69.
- [41] Tian, W.; Rielly, C.; Yang, H. Protein Crystallisation with Air Bubble Templates: Case of Gas–Liquid–Solid Interfaces. *CrystEngComm* **2021**, *23*, 8159–8168.
- [42] Tegladza, I. D.; Lin, G.; Liu, C.; Gu, X. Control of Crystal Nucleation, Size and Morphology Using Micro-/Nanobubbles as Green Additives – a Review. *Separation and Purification Technology* **2023**, *311*, 123232.
- [43] Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular Dynamics Simulations of Solutions at Constant Chemical Potential. *The Journal of Chemical Physics* **2015**, *142*, 144113.
- [44] Piana, S.; Gale, J. D. Understanding the Barriers to Crystal Growth: Dynamical Simulation of the Dissolution and Growth of Urea from Aqueous Solution. *J. Am. Chem. Soc.* **2005**, *127*, 1975–1982.
- [45] Anand, A.; Patey, G. N. Mechanism of Urea Crystal Dissolution in Water from Molecular Dynamics Simulation. *J. Phys. Chem. B* **2018**, *122*, 1213–1222.
- [46] Mandal, T.; Larson, R. G. Nucleation of Urea from Aqueous Solution: Structure, Critical Size, and Rate. *J. Chem. Phys.* **2017**, *146*, 134501.

- [47] Giberti, F.; Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Insight into the Nucleation of Urea Crystals from the Melt. *Chemical Engineering Science* **2015**, *121*, 51–59.
- [48] Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. Molecular-Dynamics Simulations of Urea Nucleation from Aqueous Solution. *Proc Natl Acad Sci USA* **2015**, *112*, E6–E14.
- [49] Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Urea Homogeneous Nucleation Mechanism Is Solvent Dependent. *Faraday Discuss.* **2015**, *179*, 291–307.
- [50] Karmakar, T.; Piaggi, P. M.; Parrinello, M. Molecular Dynamics Simulations of Crystal Nucleation from Solution at Constant Chemical Potential. *J. Chem. Theory Comput.* **2019**, *15*, 6923–6930.
- [51] McKechnie, D.; Mulheran, P. A.; Sefcik, J.; Johnston, K. Tuning Interfacial Concentration Enhancement through Dispersion Interactions to Facilitate Heterogeneous Nucleation. *J. Phys. Chem. C* **2022**, *126*, 16387–16400.
- [52] Salvalaglio, M.; Vetter, T.; Mazzotti, M.; Parrinello, M. Controlling and Predicting Crystal Shapes: The Case of Urea. *Angew. Chem. Int. Ed.* **2013**, *52*, 13369–13372.
- [53] Bennett, R. C.; Van Buren, M. *Crystallization from Solutions and Melts*; Springer US: Boston, MA, 1969; pp 44–49.
- [54] Smith, M.; March, J. *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, 5th ed.; J. Wiley & sons: New York, 2001.
- [55] Lodaya, K. D.; Lahti, L. E.; Jones, M. L. An Investigation into the Nucleation Kinetics of Urea Crystallization in Water by Means of Crystal-Size Distribution Analysis. *Ind. Eng. Chem. Proc. Des. Dev.* **1977**, *16*, 294–297.
- [56] Davey, R.; Fila, W.; Garside, J. The Influence of Biuret on the Growth Kinetics of Urea Crystals from Aqueous Solutions. *Journal of Crystal Growth* **1986**, *79*, 607–613.

- [57] Bisker-Leib, V.; Doherty, M. F. Modeling the Crystal Shape of Polar Organic Materials: Prediction of Urea Crystals Grown from Polar and Nonpolar Solvents. *Crystal Growth & Design* **2001**, *1*, 455–461.
- [58] Chen, X.; Sun, C.; Wu, S.; Yu, Y.; Xue, D. Molecular Paradigm Dependent Nucleation in Urea Aqueous Solution. *Crystal Growth & Design* **2017**, *17*, 2594–2599.
- [59] Xu, S.; Chen, Y.; Gong, J.; Wang, J. Interplay between Kinetics and Thermodynamics on the Probability Nucleation Rate of a Urea–Water Crystallization System. *Crystal Growth & Design* **2018**, *18*, 2305–2315.
- [60] Garetz, B. A.; Aber, J. E.; Goddard, N. L.; Young, R. G.; Myerson, A. S. Nonphotochemical, Polarization-Dependent, Laser-Induced Nucleation in Supersaturated Aqueous Urea Solutions. *Phys. Rev. Lett.* **1996**, *77*, 3475–3476.
- [61] Yoshikawa, H. Y.; Hosokawa, Y.; Masuhara, H. Explosive Crystallization of Urea Triggered by Focused Femtosecond Laser Irradiation. *Jpn. J. Appl. Phys.* **2006**, *45*, L23.
- [62] Liu, Y.; Ward, M. R.; Alexander, A. J. Polarization Independence of Laser-Induced Nucleation in Supersaturated Aqueous Urea Solutions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 3464–3467.
- [63] Hughes, C. E.; Williams, P. A.; Keast, V. L.; Charalampopoulos, V. G.; Edwards-Gau, G. R.; Harris, K. D. M. New in Situ Solid-State NMR Techniques for Probing the Evolution of Crystallization Processes: Pre-Nucleation, Nucleation and Growth. *Faraday Discuss.* **2015**, *179*, 115–140.
- [64] Mullin, J., Ed. *Crystallization (Fourth Edition)*; Butterworth-Heinemann: Oxford, 2001; pp 478–535.
- [65] Lee, F.-M.; Lahti, L. E. Solubility of Urea in Water-Alcohol Mixtures. *J. Chem. Eng. Data* **1972**, *17*, 304–306.
- [66] Haynes, W. M. *CRC Handbook of Chemistry and Physics: A Ready-Reference*

- Book of Chemical and Physical Data*; CRC Press: Boca Raton; London; New York, 2014.
- [67] Brouwer, M. *Kirk-Othmer Encyclopedia of Chemical Technology*; American Cancer Society, 2019; pp 1–19.
- [68] Olejniczak, A.; Ostrowska, K.; Katrusiak, A. H-Bond Breaking in High-Pressure Urea. *J. Phys. Chem. C* **2009**, *113*, 15761–15767.
- [69] Swaminathan, S.; Craven, B. M.; Spackman, M. A.; Stewart, R. F. Theoretical and Experimental Studies of the Charge Density in Urea. *Acta Crystallogr B Struct Sci* **1984**, *40*, 398–404.
- [70] Dziubek, K.; Citroni, M.; Fanetti, S.; Cairns, A. B.; Bini, R. High-Pressure High-Temperature Structural Properties of Urea. *J. Phys. Chem. C* **2017**, *121*, 2380–2387.
- [71] Roszak, K.; Katrusiak, A. Giant Anomalous Strain between High-Pressure Phases and the Mesomers of Urea. *J. Phys. Chem. C* **2017**, *121*, 778–784.
- [72] Mazurek, A.; Szeleszczuk, Ł.; Pisklak, D. M. Can We Predict the Pressure Induced Phase Transition of Urea? Application of Quantum Molecular Dynamics. *Molecules* **2020**, *25*, 1584.
- [73] Piana, S.; Reyhani, M.; Gale, J. D. Simulating Micrometre-Scale Crystal Growth from Solution. *Nature* **2005**, *438*, 70–73.
- [74] Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Computational Science Series 1; Academic Press: San Diego, 2002.
- [75] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 2nd ed.; Oxford university press: Oxford, 2017.
- [76] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of

- Physical Clusters of Molecules: Application to Small Water Clusters. *The Journal of Chemical Physics* **1982**, *76*, 637–649.
- [77] LAMMPS <http://lammps.sandia.gov>.
- [78] Quinn, T.; Katz, N.; Stadel, J.; Lake, G. Time Stepping N-body Simulations. 1997.
- [79] Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.
- [80] Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *The Journal of Chemical Physics* **1980**, *72*, 2384–2393.
- [81] Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.
- [82] Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- [83] Hoover, W. G. Constant-Pressure Equations of Motion. *Phys. Rev. A* **1986**, *34*, 2499–2500.
- [84] Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.
- [85] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.
- [86] Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with OVITO—the Open Visualization Tool. *Modelling Simul. Mater. Sci. Eng.* **2010**, *18*, 015012.
- [87] Berg, H. C. *Random Walks in Biology*; Princeton University Press, 1984; pp 5–16.

- [88] Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- [89] González, M. Force Fields and Molecular Dynamics Simulations. *JDN* **2011**, *12*, 169–200.
- [90] Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.
- [91] Van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
- [92] Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; Van Duin, A. C. T. The ReaxFF Reactive Force-Field: Development, Applications and Future Directions. *npj Comput Mater* **2016**, *2*, 15011.
- [93] Gissinger, J. R.; Jensen, B. D.; Wise, K. E. Modeling Chemical Reactions in Classical Molecular Dynamics Simulations. *Polymer* **2017**, *128*, 211–217.
- [94] GROMACS <https://manual.gromacs.org/current/index.html>.
- [95] AMBER <https://ambermd.org/>.
- [96] NAMD <http://www.ks.uiuc.edu/Research/namd/>.
- [97] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [98] Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids: An All Atom Force Field. *J. Comput. Chem.* **1986**, *7*, 230–252.

- [99] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- [100] MacKerell, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- [101] Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- [102] Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. Urea: Potential Functions, Log P, and Free Energy of Hydration. *Isr. J. Chem.* **1993**, *33*, 323–330.
- [103] Kokubo, H.; Pettitt, B. M. Preferential Solvation in Urea Solutions at Different Concentrations: Properties from Simulation Studies. *J. Phys. Chem. B* **2007**, *111*, 5233–5242.
- [104] Caballero-Herrera, A.; Nilsson, L. Urea Parametrization for Molecular Dynamics Simulations. *J. Mol. Struct: THEOCHEM* **2006**, *758*, 139–148.
- [105] Jeong, K.-j.; McDaniel, J. G.; Yethiraj, A. A Transferable Polarizable Force Field for Urea Crystals and Aqueous Solutions. *J. Phys. Chem. B* **2020**, *124*, 7475–7483.
- [106] MacKerell, A. D. et al. Self-Consistent Parameterization of Biomolecules for Molecular Modeling and Condensed Phase Simulations. *FASEB Journal* **1992**, *6*, A143–A143.
- [107] Chitra, R.; Smith, P. E. Molecular Dynamics Simulations of the Properties of Cosolvent Solutions. *J. Phys. Chem. B* **2000**, *104*, 5854–5864.
- [108] van Gunsteren, W. F. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Biomos ; Vdf, Hochschulverlag AG an der ETH Zürich: Zürich, Groningen, 1996.

- [109] Daura, X.; Mark, A. E.; Van Gunsteren, W. F. Parametrization of Aliphatic CHn United Atoms of GROMOS96 Force Field. *J. Comput. Chem.* **1998**, *19*, 535–547.
- [110] Smith, L. J.; Berendsen, H. J. C.; van Gunsteren, W. F. Computer Simulation of Urea-Water Mixtures: A Test of Force Field Parameters for Use in Biomolecular Simulation. *J. Phys. Chem. B* **2004**, *108*, 1065–1071.
- [111] Özpınar, G. A.; Peukert, W.; Clark, T. An Improved Generalized AMBER Force Field (GAFF) for Urea. *J. Mol. Model.* **2010**, *16*, 1427–1440.
- [112] Özpınar, G. A.; Beierlein, F. R.; Peukert, W.; Zahn, D.; Clark, T. A Test of Improved Force Field Parameters for Urea: Molecular-Dynamics Simulations of Urea Crystals. *J Mol Model* **2012**, *18*, 3455–3466.
- [113] Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Mixtures of Urea and Water. *J. Phys. Chem. B* **2003**, *107*, 3891–3898.
- [114] Sun, H.; Kung, P. W.-C. Urea: An Ab Initio and Force Field Study of the Gas and Solid Phases. *J. Comput. Chem.* **2005**, *26*, 169–174.
- [115] Masunov, A. E.; Tannu, A.; Dyakov, A. A.; Matveeva, A. D.; Freidzon, A. Y.; Odionokov, A. V.; Bagaturyants, A. A. First Principles Crystal Engineering of Nonlinear Optical Materials. I. Prototypical Case of Urea. *The Journal of Chemical Physics* **2017**, *146*, 244104.
- [116] Case, D. A. et al. AmberTools21. University of California, San Francisco, 2021.
- [117] He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A Fast and High-Quality Charge Model for the next Generation General AMBER Force Field. *J. Chem. Phys.* **2020**, *153*, 114502.
- [118] Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106*, 6638–6646.
- [119] Jorgensen, W. L.; Swenson, C. J. Optimized Intermolecular Potential Functions for Amides and Peptides. Structure and Properties of Liquid Amides. *J. Am. Chem. Soc.* **1985**, *107*, 569–578.

- [120] Jorgensen, W. L.; Swenson, C. J. Optimized Intermolecular Potential Functions for Amides and Peptides. Hydration of Amides. *J. Am. Chem. Soc.* **1985**, *107*, 1489–1496.
- [121] Jorgensen, W. L. Optimized Intermolecular Potential Functions for Liquid Alcohols. *J. Phys. Chem.* **1986**, *90*, 1276–1284.
- [122] Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- [123] Tirado-Rives, J.; Jorgensen, W. L. Molecular Dynamics of Proteins with the OPLS Potential Functions. Simulation of the Third Domain of Silver Pheasant Ovomuroid in Water. *J. Am. Chem. Soc.* **1990**, *112*, 2773–2781.
- [124] Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. OPLS Potential Functions for Nucleotide Bases. Relative Association Constants of Hydrogen-Bonded Base Pairs in Chloroform. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
- [125] Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- [126] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- [127] Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665–6670.
- [128] Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14\*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864–3870.

- [129] Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucleic Acids Research* **2017**, *45*, W331–W336.
- [130] Soper, A.; Castner, E.; Luzar, A. Impact of Urea on Water Structure: A Clue to Its Properties as a Denaturant? *Biophysical Chemistry* **2003**, *105*, 649–666.
- [131] Bergensträhle-Wohlert, M.; Berglund, L. A.; Brady, J. W.; Larsson, P. T.; Westlund, P.-O.; Wohlert, J. Concentration Enrichment of Urea at Cellulose Surfaces: Results from Molecular Dynamics Simulations and NMR Spectroscopy. *Cellulose* **2012**, *19*, 1–12.
- [132] van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. A Systematic Study of Water Models for Molecular Simulation: Derivation of Water Models Optimized for Use with a Reaction Field. *The Journal of Chemical Physics* **1998**, *108*, 10220–10230.
- [133] Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- [134] Cheong, D. W.; Boon, Y. D. Comparative Study of Force Fields for Molecular Dynamics Simulations of  $\alpha$ -Glycine Crystal Growth from Solution. *Crystal Growth & Design* **2010**, *10*, 5146–5158.
- [135] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- [136] Vasseti, D.; Pagliai, M.; Procacci, P. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1983–1995.
- [137] Black, S.; Davey, R. Crystallisation of Amino Acids. *Journal of Crystal Growth* **1988**, *90*, 136–144.

- [138] Qi, S.; Avalle, P.; Saklatvala, R.; Craig, D. Q. An Investigation into the Effects of Thermal History on the Crystallisation Behaviour of Amorphous Paracetamol. *European Journal of Pharmaceutics and Biopharmaceutics* **2008**, *69*, 364–371.
- [139] Nordström, F. L.; Svärd, M.; Malmberg, B.; Rasmuson, Å. C. Influence of Solution Thermal and Structural History on the Nucleation of *m*-Hydroxybenzoic Acid Polymorphs. *Crystal Growth & Design* **2012**, *12*, 4340–4348.
- [140] Kuhs, M.; Zeglinski, J.; Rasmuson, Å. C. Influence of History of Solution in Crystal Nucleation of Fenoxycarb: Kinetics and Mechanisms. *Crystal Growth & Design* **2014**, *14*, 905–915.
- [141] Jiang, S.; Ter Horst, J. H. Crystal Nucleation Rates from Probability Distributions of Induction Times. *Crystal Growth & Design* **2011**, *11*, 256–261.
- [142] Sullivan, R. A.; Davey, R. J.; Sadiq, G.; Dent, G.; Back, K. R.; Ter Horst, J. H.; Toroz, D.; Hammond, R. B. Revealing the Roles of Desolvation and Molecular Self-Assembly in Crystal Nucleation from Solution: Benzoic and *p*-Aminobenzoic Acids. *Crystal Growth & Design* **2014**, *14*, 2689–2696.
- [143] Brandel, C.; Ter Horst, J. H. Measuring Induction Times and Crystal Nucleation Rates. *Faraday Discuss.* **2015**, *179*, 199–214.
- [144] Tang, S. K.; Davey, R. J.; Sacchi, P.; Cruz-Cabeza, A. J. Can Molecular Flexibility Control Crystallization? The Case of *Para* Substituted Benzoic Acids. *Chem. Sci.* **2021**, *12*, 993–1000.
- [145] Kaskiewicz, P. L.; Downie, R.; Dowding, P. J.; George, N.; Roberts, K. J. Data for Crystallisation, Dissolution and Saturation Temperatures of a Model Fuel Comprising Eicosane Crystallising from Supersaturated Toluene Solutions in the Presence of a Cold-Flow Improver Additive. *Data in Brief* **2022**, *43*, 108455.
- [146] Flannigan, J. M. Optical Tweezer and Particle Control of Nucleation and Growth at the Microscale. Ph.D. thesis, University of Strathclyde, Glasgow, UK, 2023.

- [147] Powell, D. Analysis of Nucleation Kinetics in pH-shift Crystallization Using Agitated and Microfluidic Systems. Ph.D. thesis, University of Strathclyde, Glasgow, UK, 2023.
- [148] Kulkarni, S. A.; Kadam, S. S.; Meekes, H.; Stankiewicz, A. I.; Ter Horst, J. H. Crystal Nucleation Kinetics from Induction Times and Metastable Zone Widths. *Crystal Growth & Design* **2013**, *13*, 2435–2440.
- [149] Deck, L.-T.; Wittenberg, L.; Mazzotti, M. Thermodynamics Explains How Solution Composition Affects the Kinetics of Stochastic Ice Nucleation. *J. Phys. Chem. Lett.* **2023**, *14*, 5993–6000.
- [150] Deck, L.-T.; Mazzotti, M. Characterizing and Measuring the Ice Nucleation Kinetics of Aqueous Solutions in Vials. *Chemical Engineering Science* **2023**, *272*, 118531.
- [151] Quon, J. L.; Chadwick, K.; Wood, G. P. F.; Sheu, I.; Brettmann, B. K.; Myerson, A. S.; Trout, B. L. Templated Nucleation of Acetaminophen on Spherical Excipient Agglomerates. *Langmuir* **2013**, *29*, 3292–3300.
- [152] Sear, R. P. Quantitative Studies of Crystal Nucleation at Constant Supersaturation: Experimental Data and Models. *CrystEngComm* **2014**, *16*, 6506–6522.
- [153] Diao, Y.; Helgeson, M. E.; Siam, Z. A.; Doyle, P. S.; Myerson, A. S.; Hatton, T. A.; Trout, B. L. Nucleation under Soft Confinement: Role of Polymer–Solute Interactions. *Crystal Growth & Design* **2012**, *12*, 508–517.
- [154] Diao, Y.; Whaley, K. E.; Helgeson, M. E.; Woldeyes, M. A.; Doyle, P. S.; Myerson, A. S.; Hatton, T. A.; Trout, B. L. Gel-Induced Selective Crystallization of Polymorphs. *J. Am. Chem. Soc.* **2012**, *134*, 673–684.
- [155] Javid, N.; Kendall, T.; Burns, I. S.; Sefcik, J. Filtration Suppresses Laser-Induced Nucleation of Glycine in Aqueous Solutions. *Crystal Growth & Design* **2016**, *16*, 4196–4202.

- [156] Dela Cruz, I. J. C.; Perez, J. V.; Alamani, B. G.; Capellades, G.; Myerson, A. S. Influence of Volume on the Nucleation of Model Organic Molecular Crystals through an Induction Time Approach. *Crystal Growth & Design* **2021**, *21*, 2932–2941.
- [157] Toldy, A. I.; Badruddoza, A. Z. M.; Zheng, L.; Hatton, T. A.; Gunawan, R.; Rajagopalan, R.; Khan, S. A. Spherical Crystallization of Glycine from Monodisperse Microfluidic Emulsions. *Crystal Growth & Design* **2012**, *12*, 3977–3982.
- [158] Teychené, S.; Biscans, B. Crystal Nucleation in a Droplet Based Microfluidic Crystallizer. *Chemical Engineering Science* **2012**, *77*, 242–248.
- [159] Newville, M.; Stensitzki, T.; Allen, D. B.; Ingargiola, A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. Zenodo, 2014.
- [160] Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **1995**, *117*, 1–19.
- [161] Sun, H. COMPASS: An Ab Initio Force-Field Optimized for Condensed-Phase Applications Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102*, 7338–7364.
- [162] Sklar, N.; Senko, M. E.; Post, B. Thermal Effects in Urea: The Crystal Structure at  $-140^{\circ}\text{C}$  and at Room Temperature. *Acta Cryst* **1961**, *14*, 716–720.
- [163] Piaggi, P. M.; Parrinello, M. Predicting Polymorphism in Molecular Crystals Using Orientational Entropy. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 10251–10256.
- [164] Suzuki, K.; Onishi, S.-i.; Koide, T.; Seki, S. Vapor Pressures of Molecular Crystals. XI. Vapor Pressures of Crystalline Urea and Diformylhydrazine. Energies of Hydrogen Bonds in These Crystals. *BCSJ* **1956**, *29*, 127–131.
- [165] Ferro, D.; Barone, G.; Della Gatta, G.; Piacente, V. Vapour Pressures and Sublimation Enthalpies of Urea and Some of Its Derivatives. *The Journal of Chemical Thermodynamics* **1987**, *19*, 915–923.

- [166] Emel'yanenko, V. N.; Kabo, G. J.; Verevkin, S. P. Measurement and Prediction of Thermochemical Properties: Improved Increments for the Estimation of Enthalpies of Sublimation and Standard Enthalpies of Formation of Alkyl Derivatives of Urea. *J. Chem. Eng. Data* **2006**, *51*, 79–87.
- [167] Zaitsau, D.; Kabo, G.; Kozyro, A.; Sevruk, V. The Effect of the Failure of Isotropy of a Gas in an Effusion Cell on the Vapor Pressure and Enthalpy of Sublimation for Alkyl Derivatives of Carbamide. *Thermochimica Acta* **2003**, *406*, 17–28.
- [168] De Wit, H.; Van Miltenburg, J.; De Kruif, C. Thermodynamic Properties of Molecular Organic Crystals Containing Nitrogen, Oxygen, and Sulphur 1. Vapour Pressures and Enthalpies of Sublimation. *The Journal of Chemical Thermodynamics* **1983**, *15*, 651–663.
- [169] Zhang, Y.; Maginn, E. J. A Comparison of Methods for Melting Point Calculation Using Molecular Dynamics Simulations. *The Journal of Chemical Physics* **2012**, *136*, 144116.
- [170] Gucker, F. T.; Gage, F. W.; Moser, C. E. The Densities of Aqueous Solutions of Urea at 25 and 30° and the Apparent Molal Volume of Urea<sup>1</sup>. *J. Am. Chem. Soc.* **1938**, *60*, 2582–2588.
- [171] Ishida, T.; Rossky, P. J.; Castner, E. W. A Theoretical Investigation of the Shape and Hydration Properties of Aqueous Urea: Evidence for Nonplanar Urea Geometry. *J. Phys. Chem. B* **2004**, *108*, 17583–17590.
- [172] Burton, R. C.; Ferrari, E. S.; Davey, R. J.; Hopwood, J.; Quayle, M. J.; Finney, J. L.; Bowron, D. T. The Structure of a Supersaturated Solution: A Neutron Scattering Study of Aqueous Urea. *Cryst. Growth Des.* **2008**, *8*, 1559–1565.
- [173] Albright, J. G.; Mills, R. A Study of Diffusion in the Ternary System, Labeled Urea-Urea-Water, at 25° by Measurements of the Intradiffusion Coefficients <sup>1</sup> of Urea <sup>2</sup>. *J. Phys. Chem.* **1965**, *69*, 3120–3126.

- [174] Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.
- [175] Zimmerschied, W. J.; Dinerstein, R. A.; Weitkamp, A. W.; Marschner, R. F. Crystalline Adducts of Urea with Linear Aliphatic Compounds. *Ind. Eng. Chem.* **1950**, *42*, 1300–1306.
- [176] Gilson, D.; McDowell, C. Nuclear Magnetic Resonance Studies of Urea and Thiourea Adducts. *Molecular Physics* **1961**, *4*, 125–134.
- [177] Guide to FT-IR Spectroscopy. <https://www.bruker.com/en/products-and-solutions/infrared-and-raman/ft-ir-routine-spectrometer/what-is-ft-ir-spectroscopy.html>.
- [178] FTIR Spectroscopy. [https://www.mt.com/gb/en/home/applications/L1\\_AutoChem\\_Applications/ftir-spectroscopy.html](https://www.mt.com/gb/en/home/applications/L1_AutoChem_Applications/ftir-spectroscopy.html).
- [179] Attenuated Total Reflectance (ATR) | Anton Paar Wiki. <https://wiki.anton-paar.com/en/attenuated-total-reflectance-atr/>.
- [180] ATR Precautions. [https://www.shimadzu.eu/service-support/technical-support/ftir/essential\\_knowledge/atr.html](https://www.shimadzu.eu/service-support/technical-support/ftir/essential_knowledge/atr.html).
- [181] Brown, C. J. et al. Integrated Continuous Process Design for Crystallisation, Spherical Agglomeration, and Filtration of Lovastatin. *J Pharm Innov* **2024**, *19*, 9.
- [182] Zhang, F.; Du, K.; Guo, L.; Xu, Q.; Shan, B. Comparative Study of Preprocessing on an ATR-FTIR Calibration Model for In Situ Monitoring of Solution Concentration in Cooling Crystallization. *Chem Eng & Technol* **2021**, *44*, 2279–2289.
- [183] Lewiner, F.; Klein, J.; Puel, F.; Févotte, G. On-Line ATR FTIR Measurement of Supersaturation during Solution Crystallization Processes. Calibration and Applications on Three Solute/Solvent Systems. *Chemical Engineering Science* **2001**, *56*, 2069–2084.

- [184] Shikata, S.; Tanno, T.; Teraji, T.; Kanda, H.; Yamada, T.; Kushibiki, J.-i. Precise Measurements of Diamond Lattice Constant Using Bond Method. *Jpn. J. Appl. Phys.* **2018**, *57*, 111301.
- [185] Levien, L.; Prewitt, C. T.; Weidner, D. J. Structure and Elastic Properties of Quartz at Pressure. *American Mineralogist* **1980**, *65*, 920–930.
- [186] Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chem. Mater.* **2014**, *26*, 2647–2658.
- [187] Chadwell, H. M.; Politi, F. W. The Freezing Points of Concentrated Aqueous Solutions of Urea, Urethan, and Acetamide <sup>1</sup>. *J. Am. Chem. Soc.* **1938**, *60*, 1291–1293.