X-ray Studies of Molecular Structure during the Crystallisation of Organic Salts

Bethan Evans

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

EPSRC Centre for Doctoral Training in Complex Particulate Products & Processes

School of Chemical and Process Engineering

February 2022

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 3 of the thesis is based on a paper that has been submitted for publication in Physical Chemistry Chemical Physics, as follows:

The Electronic Structure of Imidazolium Cations in Aqueous Solution: X-ray Raman Scattering Studies of the C and N K-edges Combined with XPS and DFT Calculations.

Bethan Evans, Laila H Al-Madhagi, Anuradha R Pallipurath, Sin-Yuen Chang, Blanka Detlefs, Elizabeth A Willneff, Robert B Hammond, Helen P Wheatcroft and Sven LM Schroeder

I was involved in planning and carrying out the beamtime experiment, and was responsible for interpreting the results, performing the calculations, synthesising the salt, planning and running the NAP-XPS experiments and writing the manuscript.

The contributions of the other authors were:

Laila H Al-Madhagi and Sven LM Schroeder conceived the original idea, planned the beamtime experiment and helped carry out the beamtime experiment and the interpret the results. Sin-Yuen Chang and Blanka Detlefs helped carry out the beamtime experiment. Anuradha R Pallipurath provided guidance with performing and interpreting the calculations and helped with the NAP-XPS experiments and sample preparation. Elizabeth A Willneff provided guidance and helped run the NAP-XPS experiments and interpret the data. Sven LM Schroeder helped interpret NAP-XPS results and write the manuscript. Sven LM Schroeder, Robert B Hammond and Helen P Wheatcroft supervised the project. All authors provided feedback and contributed to the final manuscript.

Acknowledgements

This work was made possible by funding from the Engineering and Physical Sciences Research Council and AstraZeneca.

I wish to thank my academic and industrial supervisors—Prof. Sven Schroeder, Dr. Robert Hammond and Dr. Helen Wheatcroft—who have provided me with expert guidance throughout. Special thanks go to Dr. Anuradha Pallipurath for patiently repeating herself when I asked the same questions over and over, and tirelessly reading my chapters, right up to the final possible moments before going on maternity leave. Thank you, Anu.

Data collection at beamtime would not have been possible without the hands-on help and moral support from friends and colleagues in the Molecular Processes and Material Engineering group, especially those who attended beamtimes with me. I would not have known where to start without the guidance of Dr. Laila Al-Madhagi, and I owe particular thanks to Anu, Arturs and Gunjan for helping with beamtime preparation and operation.

I received help and support from the staff and my fellow PGRs at the Centre for Doctoral Training in Complex Particulate Products and Processes, for which I am most grateful. I was lucky to be sat with the team in 2.12, who made pre-Covid studies so much more enjoyable!

Thank you to everyone who has given me the benefit of their technical knowledge, time, kindness or support, and enabled me to complete this thesis.

Finally, to my wonderful friends and family, who have provided motivation, tea, and a sounding board when I have needed it, and gave me encouragement to keep going—thank you! I could not have done it without you.

Abstract

Increasing numbers of pharmaceutical drugs are being produced in salt form to improve bioavailability, and handling and processing properties. Effective control of organic salt crystallisation from solution requires an improved understanding of the molecular-level interactions present in the solution before and during crystallisation.

State-of-the-art X-ray techniques probe solution-phase systems at the short time- and length-scales required to reveal information about the chemical environment and local molecular structure of the solution. Core-level X-ray spectroscopy and X-ray diffraction, combined with computational analysis, have been used to characterise an organic salt in solution and the phase from which an organic salt crystallises.

The near-edge X-ray absorption fine structure (NEXAFS) spectra of the C and N K-edges of aqueous imidazole (Imid) and imidazolium ($ImidH^+$) are acquired using a novel combination of X-ray Raman scattering (XRS), near-ambient pressure X-ray photoelectron spectroscopy (NAP-XPS) and time-dependent density functional theory (TDDFT) calculations. C and N 1s core-level binding energies (from NAP-XPS) were used to assign relative ionisation potential energies. N-atom pseudoequivalence in $ImidH^+$, defined by the single $1s\rightarrow1\pi^*$ transition peak, contrasts with the two N moieties observed in Imid. TDDFT calculations revealed $1s\rightarrow3\pi^*$ resonances in the $ImidH^+$ C K-edge spectrum that are not present in the Imid spectrum.

X-ray Pair Distribution Function (XPDF) patterns collected during the *in situ* cooling crystallisation of an aqueous solution of guanidine hydrochloride (GuHCl) indicate three distinct solution phases prior to crystallisation. For the first time, Empirical Potential Structure Refinement (EPSR) structural models, refined to experimental data, have been used to suggest the possible structural motifs that may dominate in the pre-crystallisation phases.

The combined XPDF/EPSR approach was applied to the structures of 2, 4 and 6 M aqueous GuHCl solutions and suggested medium-range structural differences in Gdm^+-Gdm^+ interactions due to a critical solvation change between the 2 M solution, and higher-concentration solutions.

Contents

List of Tables

List of Tables in the Appendices

List of Figures

List of Figures in the Appendices

Key abbreviations

API	Active pharmaceutical ingredients
CIP	Contact ion pair
CNT	Classical nucleation theory
CoordN	Coordination number
CSP	Crystal structure prediction
DFT	Density functional theory
DLS	Diamond Light Source
DSCS	Differential scattering cross-section
E_B	Binding energy
E_K	Kinetic energy
EP	Empirical potential
E_P	Photon energy
EPSR	Empirical Potential Structure Refinement
ESRF	European Synchrotron Radiation Facility
FWHM	Full width half maximum
Gdm^+	Guanidinium cations
GuHCl	Guanidine hydrochloride
IL	Ionic liquid
Imid	Imidazole
$ImidH^+$	Imidazolium
ImidHCl	Imidazole hydrochloride
IP	Ionisation potential
ISEELS	Inner shell electron energy loss
Low-Z	Low-atomic number
MD	Molecular dynamics
MM	Molecular modelling
MO	Molecular orbital
MSZ(W)	Metastable zone (width)
NAP-XPS	Near-atmospheric pressure XPS
NEXAFS	Near Edge X-Ray Absorption Fine Structure
OPLS	Optimised potentials for liquid simulations
PCA	Principle component analysis
PCM	Polarisable continuum model
PDF	Pair distribution function
PXRD	Powder X-ray diffraction
q	Momentum transfer
QM	Quantum mechanics
RP	Reference potential
SCXRD	Single crystal XRD
SDF	Spatial density functions
SPD	Spatial probability density
SShIP	Solvent-shared ion pair
SSIP	Solvent-separated ion pair
TDDFT	Time-dependent density functional theory
UHV	Ultra-high vacuum
XA(S)	X-ray Absorption (Spectroscopy)
XPDF	X-ray Pair Distribution Function
XPS	X-ray Photoelectron Spectroscopy
XRD	X-ray diffraction
XRS	X-ray Raman scattering

Chapter 1 Introduction

1.1 Crystal engineering of solid drugs

Drug products are formulated to be bioavailable once they arrive at the required site of action on or within the body [1]. Active pharmaceutical ingredients (APIs), which are frequently formulated and delivered as solid dosage forms [2], need to be water-soluble to facilitate absorption across the intestinal membrane [3–6], or for administration via injections, infusions, or nasal or eye drops [1]. Solid APIs may exist in crystalline or amorphous phases [2,7]. Crystalline materials are generally easier to isolate and purify than amorphous solids, and have longer shelf lives as they are less prone to oxidation [1,3,6]. Typically, crystalline solids have better solid-state handling properties, although some can have processing issues, e.g. poor flowability and compressibility, or low bulk densities, which may lead to storage issues [1].

Hydrates/solvates, salts, or co-crystals, which may be formed from a salt and a secondary molecule, are commonly used to form multi-component compounds [4]. Although co-crystal drug compounds have been discussed in the literature since Wöhler first presented quinhydrone in 1844 [8], the concept of crystal engineering in its current form only came about in the 1980s and 1990s [9].

Crystal engineering applies our understanding of intermolecular interactions in crystal structures to the design of new solids, which presents an opportunity to fine-tune new materials including APIs to optimise manufacturability and performance characteristics [2,10,11]. Improving the therapeutic effect of a drug formulation may also modify its pharmacokinetic or physicochemical properties, such as stability, bioavailability, or crystal size, shape and morphology [1,4,12–14].

A trend towards the discovery and development of increasingly target-specific and potent APIs, possibly due to the implementation of *in silico* lead molecule identification and optimisation [5,9,15], often results in lipophilic drug entities with high molecular weights and low solubilities [5,16]. In a 2006 classification of the 200 most commonly-used drug products in the USA, UK, Spain and Japan, around 20% of the considered drug products were highly water-insoluble [15]. As a result, there is a real driver to develop drugs in salt form.

1.1.1 Salts as drugs

If an API is insoluble or difficult to crystallise, salt screening—the search for a suitable salt formulation—may be conducted [17]. Around 70% of APIs have the requisite functionality for salt formation [17]. The selected counterion will influence physical properties such as the melting point, thermal and chemical stability, crystallinity [18,19] and hygroscopicity, which should be minimised (rule of thumb < 2% for the salt form to pass through screening [20]).

The International Union of Pure and Applied Chemistry (IUPAC) defines a salt as 'a chemical compound consisting of an assembly of cations and anions' [21], yet the classification of multi-component crystals can be inconsistent, with ambiguous or contradictory systems proposed in the literature [4,7,22,23].

To form a salt or co-crystal, a molecule should contain a proton-donating or proton-accepting moiety. As crystal engineering has developed, a formalism for the classification of salts and co-crystals has been established, and the literature suggests that there is a continuum between the these compound types based on the extent of proton transfer [17,24,25], with partial transfer (hydrogen-bonding) forming a co-crystal, and complete proton transfer forming a salt [24].

Salts are formed in an acid-base reaction [26], and the relative pK_a values[1] of each component predicts whether a salt or co-crystal will be formed [24,26]. As a rule of thumb, a reaction where $\Delta pK_a > 3$ will produce a salt, and if $\Delta pK_a < 0$, a co-crystal will be formed [24,27]. Whilst most acid-base complexes can be defined by this parameter, the classification is less clear when ΔpK_a values fall in the 1–3 range [24]. If ΔpK_a is small, and the energy barrier separating the hydrogen-bond-donor/acceptor potential wells is low, the hydrogen is quasi-central and a short, strong hydrogen bond is formed [25,28]. The pK_a of the ionisable group is often established early on, and potential counterions are selected for further screening based on this $\Delta pK_a > 3$ rule [19]. It should be noted that pK_a values quoted for a given temperature (often 25°C) normally relate to aqueous solutions, so if other solvents are used, the pK_a will vary [29].

1.1.2 API screening

When selecting the appropriate salt form of a drug molecule, factors including solubility, pH-solubility profiles, the feasibility of full-scale production and solid-state properties for production and processing are considered [1,18]. Ensuring that a drug candidate is

[1] pK_a = -log (dissociation constant, K_a)

sufficiently water-soluble early on in the drug screening process is essential for pharmacological testing and preclinical trials [1].

Although rapid salt selection is crucial to reduce the time to market of the new drug [11], solid-form screening of potential new salt compounds is time- and labour-intensive [4,30].

Salt screens are very sensitive to crystallisation conditions [18]. Drug solubility and supersaturation is assessed by developing a calibration curve with ultraviolet-visible light (UV-vis) or infrared (IR) spectroscopy [1]. Deciding when a solution equilibrium is achieved can be problematic, as the solid and liquid phases may need to be in contact for a long time to achieve the equilibrium state [31]. Crystallisation systems such as Crystal16$^{(TM)}$ [32] allow efficient small-scale solubility screening, which is useful if limited material is available for screening. However, industrial-scale process conditions, such as temperature control and crystalliser flow characteristics, are difficult to replicate in the laboratory, meaning some level of empirical process design is required [30].

Computational methods are being developed as a complementary screening tool for organic molecules [33]. Although not all plausible structures identified by crystal structure prediction (CSP) would form in practice [34], methods have been developed to determine normally the most thermodynamically-stable crystal structures from the chemical diagram [33]. In a 2010 blind study using CSP methods, some previously unpublished crystal structures were found to be easy to predict, but some could not be solved by computational methods [33]. The field of computational crystallisation and CSP is rapidly evolving, with continual development of new methods and approaches, and better integration of CSP studies with experimental work is possible as the experience-base increases [33]. Gaining a better physicochemical understanding of crystallisation will bring about improvements to computational analysis methods [34], resulting in economic and time-saving benefits to the salt screening process.

1.2 Nucleation theory

Crystallisation is the process by which a solid phase is formed from a liquid, solution or vapour phase [35]. Despite being an essential step in the pharmaceutical [17], food production, bulk and fine chemicals [36] and agrochemicals [37] sectors, the molecular-level understanding of how this process occurs is incomplete [3,38].

Conventionally, crystallisation processes are modified by empirical changes to temperature, supersaturation levels, or fluid dynamics, e.g. agitation [39]. However, it is generally the

molecular-level kinetics, which are influenced by intermolecular interactions [10], that govern crystallisation processes [34] and thus determine the properties of the resultant crystals, such as size distribution and polymorph type [10,40–44]. The lack of an atomistic understanding of the structure-forming processes that occur before and during crystal nucleation and growth limits our ability to effectively and efficiently control the crystallisation process [41,45]. Consequently, control of the crystalline product, which is required for the synthesis of some new materials [46] and production of a consistent pharmaceutical product, is reduced [47].

Nucleation can be classified as a primary (homogeneous or heterogeneous) or secondary process (Figure 1.1). Primary nucleation occurs in the absence of seed crystals [31]. Homogeneous crystallisation occurs if nuclei form spontaneously and there are no foreign particles in the solution to facilitate it [4,31], whereas heterogeneous crystallisation is induced by impurities present in the solution, e.g. dust [31]. Secondary nucleation, which is induced by seed crystals or attrition fragments in the solution [31,40], affects the particle size distribution due to different generations of crystals forming in the solution.

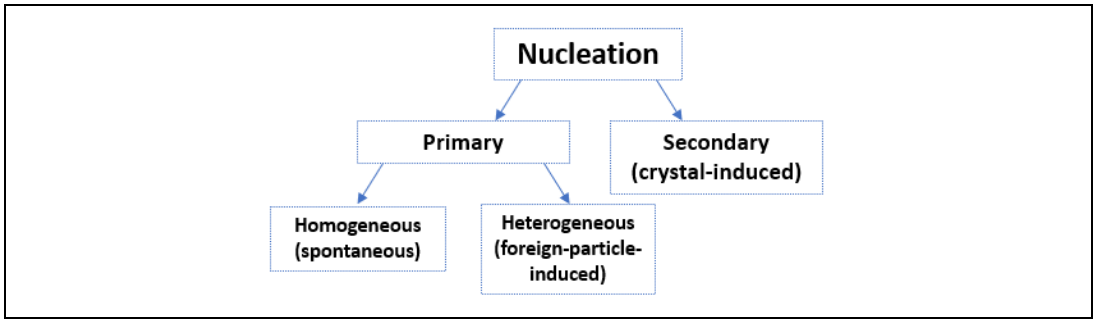


Figure 1.1 Nucleation classification, based on ref. [40]

Seeded secondary nucleation is often the pathway that is followed in industry, as it enables greater control of the product morphology and size distribution [40]. Heterogeneous nucleation on equipment surfaces is unwanted, as this results in cleaning issues, reduced yield and potentially unwanted polymorph growth [38].

The classical and non-classical nucleation theories introduced in this section are detailed elsewhere [45,48–50].

1.2.1 Supersaturation

Supersaturation, a prerequisite for nucleation, is the driving force of crystallisation and dominates the kinetic processes that occur during nucleation [37,51,52]. Supersaturation is achieved when the crystallising species is dissolved in a solvent system at a concentration above the equilibrium (saturated) concentration [31].

The supersaturation of a solution σ, can be expressed thermodynamically in terms of the activity coefficients of the solute in saturated and unsaturated conditions [1,42,53]. Because requisite specific-system activity values are frequently unavailable, a measure of supersaturation based on concentration is used instead (Equation 1):

$$\sigma = \frac{(c - c*)}{c*} \qquad \textbf{Equation 1}$$

where c is the solution concentration and c* the equilibrium concentration of the solution at a given temperature [53]. The supersaturation ratio S, can be described as Equation 2 [54]:

$$S = \frac{c}{c*} \qquad \textbf{Equation 2}$$

A solubility curve describes the mass of solute that can completely dissolve in a solvent at a given temperature (Figure 1.2). Along the solubility curve (solid line), the saturated system is in equilibrium, so the further addition of solute will give rise to the appearance of undissolved crystals in the solution [54].

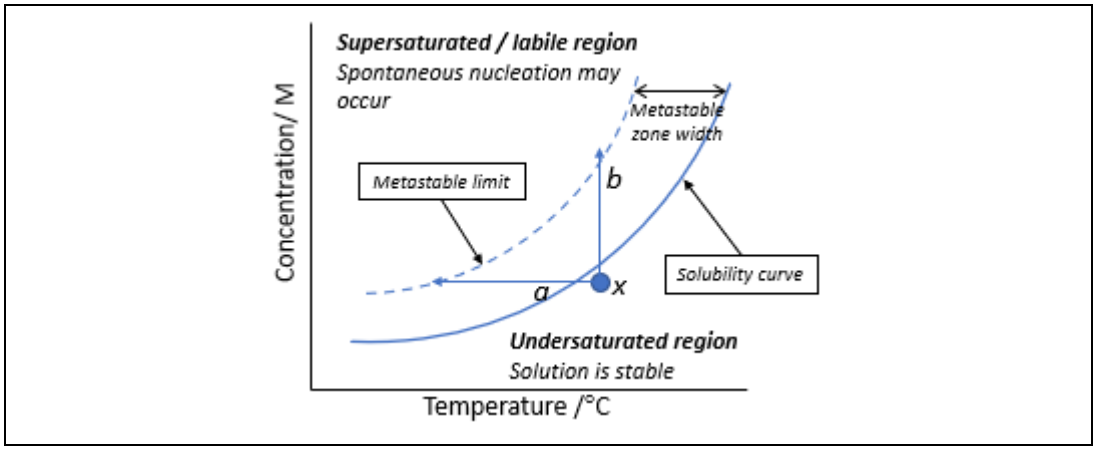


Figure 1.2 Solubility curve showing the metastable zone

Below the solubility curve, the solution is undersaturated, so the mass of solute present is below the equilibrium saturation value. Above the metastable limit (dotted line), the supersaturated solution is described as labile, meaning that new crystals can form spontaneously [37,42]. For a solution at point *x*, supersaturation can be achieved by either decreasing the temperature (along line *a* on Figure 1.2) or by evaporating the solvent to increase the concentration (along line *b*) [55]. The addition of an anti-solvent can also generate a supersaturated solution resulting in crystallisation [56].

The metastable zone (MSZ) between the undersaturated and supersaturated regions is characterised by the lack of spontaneous primary crystal nucleation [55,57]. The metastable

zone width (MSZW) describes the maximum supercooling before spontaneous nucleation occurs [58]. The MSZW is determined by measuring the temperature at which nucleation occurs, for example using a turbidity probe, for solutions of increasing supersaturation [55,58,59]. The MSZW is system-specific, as it is sensitive to environmental factors such as the agitation rate, the presence of impurities and the surface properties of the crystallising vessel [37,55]. If the solute is held in a supersaturated solution, the excess solute will crystallise out to leave a saturated solution after what is known as the induction time [54].

1.2.2 Classic nucleation theory (CNT) and its limitations

Whilst we are able to comprehensively characterise bulk crystal structures, and our understanding of solution chemistry is improving, the significant challenge of deriving a physicochemical understanding critical stages of nucleation and crystal growth remains [34].

Classic Nucleation Theory (CNT), which was initially developed in the 1920s by Volmer and Weber to describe the condensation of vapour to a liquid, has been extended and refined over time to explain crystal nucleation within melts and solutions [38,45,60,61].

In CNT, nucleation is considered a first-order phase transition [43] based on the thermodynamic relationship between the bulk energy of the solid phase and the interfacial energy between the solute and solvent in the crystallising system [62]. CNT assumes that if the bulk metastable phase is sufficiently supersaturated, the system's free energy barrier can be overcome to induce nucleation [31,51]. During nucleation, a small number of crystallising units join to form an interface between the solution and the new phase [42], thus reducing the free energy of the system [46].

These symmetrical crystal nuclei undergo a series of elementary single-particle (atom, molecule or ion) attachment or detachment events [31,63]. If the molecular cluster radius r, is below the critical size for nucleation r_c, it is energetically favourable for the cluster to dissolve, so the molecules return to solution. If r_c is achieved, growth is thermodynamically favourable, as it lowers the free energy of the system [42]. These critical nuclei are considered the transition state between supersaturation and crystal growth [49]. The number of crystallising units in a critical nucleus ranges from 10 to a few thousand, depending on the crystallising substance [40].

Gibbs' thermodynamic description of CNT [45,64,65] states that the overall excess free energy ΔG, is equal to the sum of the free energy change for the phase transformation ΔG_V, and the excess free energy between the bulk and surface of the particle ΔG_S, i.e. formation of the solute-solvent interface (Equation 3 [40,45], expanded to Equation 4):

$$\Delta G = \Delta G_V + \Delta G_S \quad \textbf{Equation 3}$$

$$= \frac{4}{3}\pi r^3 \Delta g_v + 4\pi r^2 \gamma \quad \textbf{Equation 4}$$

where r is the radius of the nucleus, Δg_v is the free energy change for the phase change per unit volume between the phase in which nucleation is occurring and the phase that is nucleating, and γ is the surface tension between the solute surface and the supersaturated solution [40]. The competing relationship between these two factors (Figure 1.3) shows that as r increases, the increasingly-negative ΔG_V term dominates the surface term ΔG_s, which favours dissolution, and nucleation becomes energetically favourable [45]. Additionally, as supersaturation increases, the nucleation barrier and r_c decrease until eventually, nucleation is spontaneous [42].

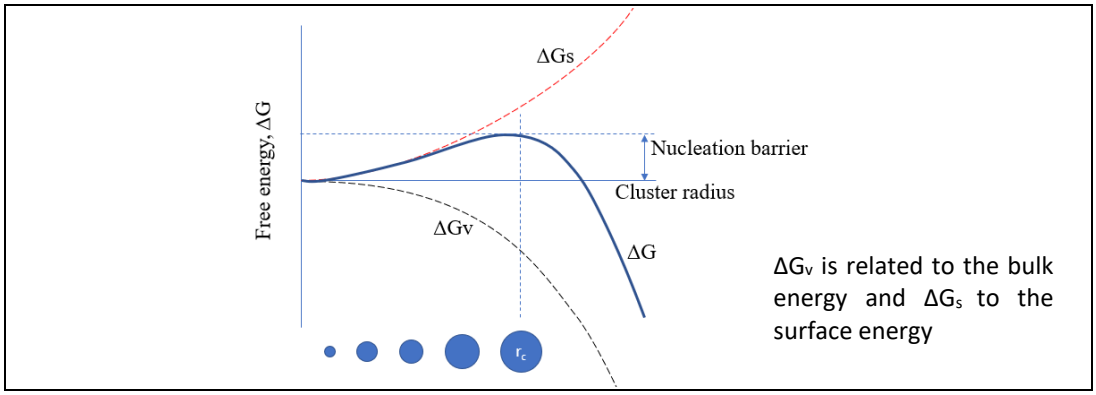


Figure 1.3 Change in free energy with cluster radius r

Heterogeneous nucleation is thermodynamically more favourable, with a nucleation rate that can be many orders of magnitude greater than homogeneous nucleation [38]. When a nucleus forms on a surface, the surface area of the nucleus-solution interface must be less than in the homogeneous case, reducing the surface term and lowering the nucleation energy barrier (Equation 3) [31,38].

The nucleation rate J (Equation 5), is a measure of the number of crystalline clusters that overcome the nucleation barrier per unit volume and unit time [40].

$$J = A\, exp\left(-\frac{B}{ln^2\, S}\right) \quad \textbf{Equation 5}$$

In this expression, constant A relates to the frequency of unit attachment to the nucleus, which is affected by kinetic factors, e.g. diffusion coefficient of the attachment unit and solvent selection, and constant B describes the thermodynamic free energy barrier for nucleation [43]. The degree of supersaturation S influences the rate of crystallisation, as a small increase in S causes a proportionally much larger increase in J [40].

The importance of solvent-solute interactions, which is not specifically considered by CNT, has been demonstrated in theoretical and experimental studies [66–68]. Molecular dynamics simulations have shown that solvent selection influences the kinetics and thermodynamics of molecular assembly events [69]. A computational modelling study of NaCl nucleation [66] and an *in situ* X-ray Raman scattering (XRS) study of the crystallisation of imidazole (Imid) [67] both suggest that solute desolvation is the rate-limiting step of nucleation, rather than diffusion. A study of two-dimensional barite growth showed that desolvation of the cation, and the surface on which crystal growth was occurring, were the rate-limiting factors for crystal nucleation and growth [68].

CNT does not however adequately explain observed nucleation processes [44], and over time, computational and experimental results have not matched CNT-based theoretical predictions [44,45,70]. The nucleation rate for the condensation of water, the system on which CNT was based, was found to be one to two orders of magnitude lower than estimated using CNT, despite the input parameters of CNT being known and highly accurate [17]. A review by *Erdemir et al.* [45] identified that whilst CNT correctly predicted what was happening on a qualitative level during the gas-liquid transition of single-component nonpolar fluids, the calculated values did not agree with the observed nucleation rates.

CNT assumes that the nucleation clusters form as spherical droplets, with the same density and molecular arrangement as the final crystals [45]. To assume the composition of the nucleus would be the same as the final bulk crystal also assumes that molecule attachment takes place in an ordered, ideal manner, with single identical monomer growth units attaching to the growing cluster [17,45]. This assumption does not allow for any cluster collisions leading to attachment, or to sections of the clusters breaking off [45]. For salt crystallisation, the attachment units could alternate between individual ions and ion pairs that form in solution, thereby changing the attachment mechanism, which is not captured by CNT.

Although small observed deviations from CNT may be empirically parametrised in some situations [49], some experiments and simulations have started to indicate multi-stage crystallisation processes may be involved, which are too complicated even for modified versions of CNT [71].

1.2.3 Alternatives to CNT

During the last decade, evidence pointing towards the existence of crystallisation pathways via observed metastable intermediate phases [50], rather than through the stochastic formation of critical nuclei, has increased [72].

Nucleation theories have been developed for pre-nucleation clusters or aggregates [49,69,73,74], critical density clusters [75], and amorphous precursors [47], even if the mechanisms for their formation and transition cannot yet be fully explained [70,76].

Multiple-step nucleation models, such as the two-step model in Figure 1.4, have been proposed to explain inconsistencies with CNT, such as significant differences in experimental and predicted nucleation rates, and unexpected relationships between the level of supersaturation and nucleation kinetics, for organic and inorganic small-molecule systems, proteins and colloids [43,45,47,77,78].

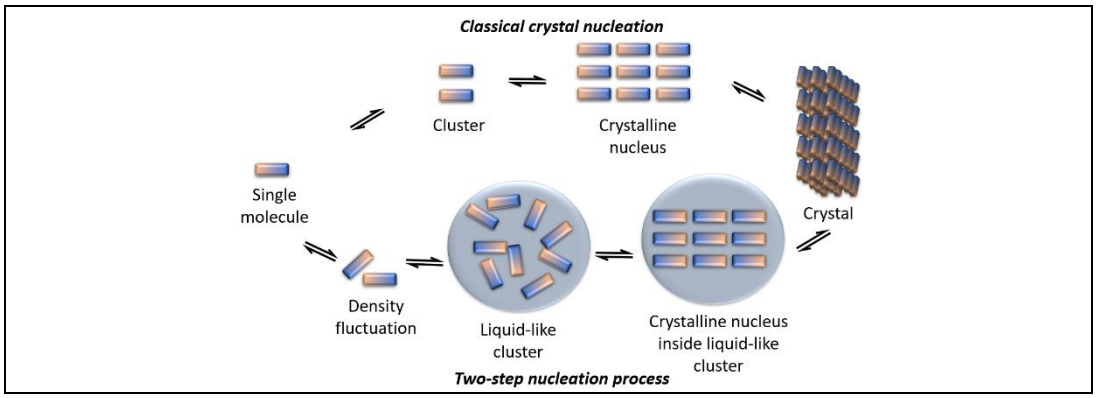


Figure 1.4 Proposed pathway for two-step nucleation, compared with the classical nucleation pathway. Based on ref. [43]

One proposed pathway is that highly-disordered pre-nucleation clusters, several hundred nanometres in size, form in regions of dense liquid [43,47]. Crystallising units assemble as a crystalline nucleus inside these mesoscopic clusters [47]. This assertion is supported by experimental data indicating a spinodal region and two-step mechanism for crystallisation in protein solutions [47], and computer simulations indicating the initial formation of amorphous pre-nucleation, from which the crystalline material will nucleate [10,78].

Until recently, the scale of detection of the available experimental techniques has been insufficient to probe the precise morphology of pre-nucleation clusters, but the development of synchrotron X-ray techniques should provide information about these structures, informing nucleation theory, and revealing molecular self-assembly processes during crystallisation [40,74]. Molecular modelling of crystallisation, combined with machine learning, will complement the advances in experimental techniques and accelerate our molecular-level understanding of crystallisation [34,71]. However, there are still significant sources of uncertainty in the modelling process [71]. Despite this recent progress in crystallisation theory and molecular modelling, predictive design of crystallisation processes to tune the physicochemical properties of the crystallised product will require a further step

change in our fundamental, molecular-level understanding of crystal nucleation and growth [34,67].

Direct *in situ* monitoring of a system undergoing crystallisation is necessary to probe the structure of pre-crystalline phases, which are not described by CNT [34,39]. The availability of higher-energy fourth generation synchrotron sources provides the potential for experimental techniques with the time- and length-scales necessary to examine nucleation kinetics with *in situ* crystallisation experiments [52]. Insights into the nature of the intermolecular interactions in the pre-crystallisation phases may be key to unlocking this new level of understanding [39].

It is hoped that computer simulations will be a helpful, complementary tool alongside experimental techniques for identifying critical nucleation events [34], and ultimately improving our understanding of the molecular-level mechanisms of crystallisation [38].

1.2.4 Industrial-scale crystallisation

Crystallisation is usually the final stage in the synthesis of a crystalline API [37]. Control of crystallisation and growth is necessary to produce the required crystal polymorph and to meet the particle size, size distribution and morphology specifications that influence the downstream processing properties and API efficacy [1,30,79]. On an industrial scale, controlling the form and stability of a crystal product during crystallisation is a balance between the physical chemistry of the system (kinetics and thermodynamics) and chemical engineering principles (residence time, agitation), which influence the local supersaturation [31,42].

If the product to be crystallised reaches the crystallisation stage as a melt, progressive freezing (fractional crystallisation) allows the purified product to be extracted from the melt using a surface cooled below the melting point [42]. In solution crystallisation, there are three methods used to modify the supersaturation and induce nucleation: cooling the solution; evaporating the solvent; and adding an anti-solvent to change the solubility of the solute in the binary solvent [31]. The required scale of production, yield and product characteristics will influence the crystallisation process to be selected for a particular product. The crystallisation study in this thesis uses cooling crystallisation.

Crystallisation technology can be classified as batch or continuous production. Batch crystallisation processing is still prevalent in the pharmaceutical industry, despite higher production costs and potential variation between batches [80]. More precise control

systems and strategies using process analytical technology (PAT) could improve the reliability and production rate of batch crystallisation production [30].

There is growing interest in continuous crystallisation processes in the pharmaceutical industry [81] to reduce operating costs, increase throughput, offer more straightforward scale-up, and decrease attrition from agitation [30]. Specific continuous processes bring particular benefits, e.g. plug flow crystallisation has low rates of secondary nucleation, leading to a smaller particle size distribution [30].

For the industrial crystallisation of organic salts, whether batch or continuous flow, control of the nucleation and crystal growth is key [82]. For this, our understanding of the changing solute-solvent, solvent-solvent and solute-solute intermolecular interactions throughout the crystallisation process must be developed to improve the performance and quality of crystallisation units [30].

1.3 Organic salt systems

1.3.1 Salts in solution

During the process of salt solvation, the interaction energies between the ions in a lattice, which are dominated by electrostatic energy, must be overcome for solute-solvent interactions to form [1,83]. In aqueous solutions, the ion-dipole interaction dominates the dissolution of a salt [1]. Dissolution of a salt may also be accompanied by dissociation into independent cations and anions, the extent of which is defined by the dissociation constant and influenced by the solution pH [1,84].

The energy change due to the changes in interactions is described by the enthalpy of solution (ΔH_{sol}) in Equation 6:

$$\Delta H_{sol} = \frac{Z^2}{2r}(1-\varepsilon) \qquad \textbf{Equation 6}$$

where Z = charge number, r = ionic radius and dielectric constant of the solvent = ε [83].

The solubility of a salt is directly related to the dielectric constant ε [83]. The greater the value of ε, the lower the Coulombic attraction F, between oppositely-charged ions of separation distance d (Equation 7):

$$\mathrm{F} = \frac{e^2}{\varepsilon . d^2} \qquad \textbf{Equation 7}$$

where e = elementary charge and d = the interionic separation [83].

The simple diffusion model describes the dissolution of a pharmaceutical solid [1]. The concentration of the diffusion layer at the solid–liquid interface decreases from a saturated solution c_s, adjacent to the solid to the concentration of the bulk c_b, at a distance h, from the solute (Figure 1.5).

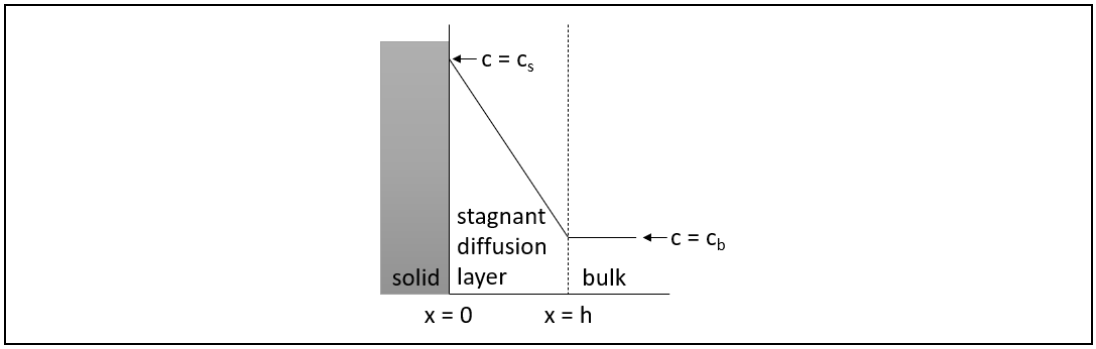


Figure 1.5 Schematic of the diffusion layer model

The ability of the molecules to diffuse through this film into the bulk solution is the limiting factor for dissolution, and can be described by Fick's first law (assuming no reaction between the solute and solvent):

$$J = \frac{dm}{A.dt} = \frac{D}{h}(c_s - c_b) \approx \frac{D}{h} c_s \; (if \; c_b \ll c_s) \qquad \textbf{Equation 8}$$

where J is the mass of material *m*, passing through the boundary layer, per unit time *t*, per unit surface area *A* (the flux). *D* is the diffusion coefficient of the solute, and *h* is the boundary layer thickness.

Establishing the solubility of a salt can be complicated. The salt form of an API will have different chemical and physical properties from the neutral parent compound and from other salt forms, as counterion selection influences salt solubility [1]. Although computational models have been developed to calculate solubility, they are not yet able to simulate systems containing ions [1,85].

1.1.1.1 Ions in solution

Ions in aqueous solution are of interest in fields as diverse as geology [86], the biological environments of DNA and proteins [87,88], biological ion channel properties [89] and electrochemistry [90]. The study of ions in biological systems is important, as concentrated aqueous salt solutions cause the salting-in and salting-out of proteins and other biomolecules, and affect enzyme activity [88,91–95]. Studying the structure of the solvent in solutions is useful, especially in aqueous systems, as changes to the structure of bulk water may provide information about the mechanisms driving biological processes [96].

Ion-specific effects on protein structures and protein-folding are attributed to a number of potential factors: the ion's solvation shell structure and its ability to desolvate to interact with the proteins; interactions between proteins and ligands, which occur despite the electrostatic forces present; the propensity of the cation to bind to its counter-anion; and the interaction of the ions with the water molecules, which may in turn affect the hydrogen bonding network of water [91,92,97]. The addition of charged solutes to water can have an electrostrictive effect on the water around the ion [98], the extent of which depends on the ion's charge, size, structure, and hydration and hydrophobic properties [99], and is proposed as the first step in protein folding [100].

In the literature, there is limited information on the study of aqueous organic salt solutions for the purpose of crystallisation. However, ions in aqueous solutions—monatomic ions in particular—play a crucial role in biological systems and have been the subject of extensive research [88,101–103]. Despite this, the intermolecular interactions of water and ions with biomolecules, which could affect protein stability and ion transportation, are not well understood [101].

Over time, experimental data and computational modelling to demonstrate how ions disrupt water's long-range hydrogen-bonded networks has been contradictory [97,104,105]. Indeed, some evidence points towards only local disruption of the first hydration shell by monovalent ions [88,106,107], whereas the effect of divalent ions on coordination shells might be different [105].

The Hofmeister series empirically orders cations and anions by the extent to which they cause proteins to salt-out [94]. The ion specificity principles of the Hofmeister series have also been applied in other fields such as polymer science, colloids and the ordering of metal–organic frameworks in solution [108]. The ions in the Hofmeister series are classified as "structure makers" (kosmotropes) or "structure breakers" (chaotropes), referring to the strengthening of or disruption to the water structure beyond the first hydration shell [102,104].

Studying the specific ion effects of a single ionic species, such as its influence on the local solvent structure or apparent solvent density, is incomplete without also considering the effect of the counterion and the solution concentration, which impacts an ion's classification as structure maker or breaker [105]. The electrostatic forces that dominate in low concentration solutions are increasingly screened at higher concentrations [109], where most water molecules will be hydrating the ions [109].

The concentration of a salt in solution affects the crystallisation properties of the solution. If the concentration is near the solubility limit, there is insufficient solvent available to fully solvate the ions. A highly concentrated salt solution is said to be more viscous [105,110]. The effect on crystallisation from a concentrated salt solution is that the higher viscosity limits the diffusion of particles to the surface of the crystal nucleus, which could affect the particles' ability to correctly orientate and be incorporated into the crystal structure.

1.1.1.2 Ion pairing

In the early 20th century, it was observed that some organic salt solutions did not have the conductive properties that would be expected if the solute had fully dissolved and dissociated in solution [84]. It was proposed that partial ionic dissociation was occurring, and the non-dissociating salt molecules remained as an ion pair [84]. The formation of ion pairs occurs when oppositely charged ions in solution associate, likely due to long-range electrostatic forces [111]. The attractive electrostatic forces between oppositely-charged ions must overcome the ion separation due to thermal energy in the system, allowing the free ions to interact and form a neutral species [110]. The ion pairing lifetime ($< 10^{-5}$ s) is greater than would be expected if diffusion alone were considered [84].

The Bjerrum theory of ion pairing suggests that ion pairing is greatest when [111]:

- Concentration is high: insufficient solvent molecules to form solvation shells around the ions.
- Both ions are small and oppositely charged: with strong reciprocal attraction, water molecules in the water spheres around the ions are expelled [109].
- Temperature is low: insufficient thermal energy in the system to break the ion pair apart [110].
- Solvent permittivity (dielectric constant) is low [110]. Ion pairing is lowest in aqueous electrolyte solutions, increasing in organic solvents with low dielectric constants.

For monovalent ion pairs to be unambiguously present, a solvent of relative permittivity < 30 is required, but if either ion has a charge greater than 1, ion pairing can occur in any solvent [111]. The phenomenon of charge shielding or masking due to the close proximity of an ion pair can result in behaviour unexpected for an ionic species, such as an increased solubility in apolar solvents [112]. As such, it is suggested that ion pairing may explain some unpredicted observations of electrolyte systems, e.g. unexpected vibrational resonances [111].

A range of spectroscopic techniques and molecular modelling, e.g. molecular dynamics (MD), have been used to study ion-water interactions and water self-interactions in the presence of ions [113]. Far-infrared (THz) absorption spectroscopy, X-ray absorption spectroscopy and MD indicated that even at relatively low (physiologically-relevant) concentrations, ion pairing occurs in, and affects the functionality of, some biological systems [113].

A multi-stage process for the formation of an ion pair is proposed (Figure 1.6) [111]. Ion pairs transition from solvent-separated ion pairs (SSIP) to solvent-shared ion pairs (SShIP), then on to form contact ion pairs (CIP) [111].

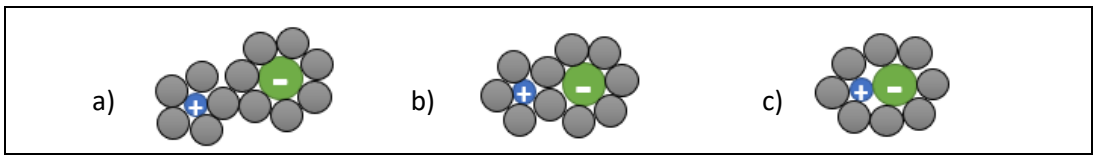


Figure 1.6 Types of ion pairs

a) solvent-separated ion pair (SSIP), b) solvent-shared ion pair (SShIP) and c) contact ion pair (CIP). Based on [111].

Although often occurring between oppositely-charged ions, it is also suggested that liked-charged ion pairing is possible. Contact Cl^--Cl^- ion pairs are found in solid crystal structures [114], where their strong polarisation effect on their local environment can exceed their repulsion [115]. The ability of the protein arginine to penetrate across cellular membranes is said to be due at least in part to the like-charge ion pairing of its Gdm^+ side chains [93], which are stabilised in solution [115].

New core-level X-ray spectroscopy techniques are being used to explore the dynamics of nucleation, such as measuring the local structure of pre-nucleation clusters in solutions. The application of high-energy X-ray total scattering diffraction techniques combined with computational modelling, where the model can be validated with experimental data (Section 2.5.4), could also reveal more information about the extent of ion pairing and pre-nucleation aggregation.

1.3.2 Intermolecular interactions

Atoms, molecules, or ions are considered to be bonded when they are brought together in a way that can be detected experimentally or computationally [116]. Closed-shell atoms and molecules may interact weakly, due to dispersion or Van der Waals interactions, or strongly because of a chemical bond with a high stabilisation energy [117]. Between these two

extremes, there are many types of interactions with sometimes complex origins involving multiple energy terms [117].

In a crystalline solid, the constituent atoms, ions or molecules are arranged in a fixed, rigid pattern, forming a two- or three-dimensional lattice [40]. These lattices exhibit long-range order, meaning there is an internal regular pattern, with defined crystal facets [40]. The regularity of the structure results in anisotropic crystals with directionally-dependent physical, mechanical and optical properties[2] [40,42] that influence the processability of the crystal.

Aakeröy described a crystal structure as 'a subtle balance between a multitude of non-covalent forces' [118], and the same can be said of a solution structure, although the dynamic nature of a solution means that the contributions of each will change over time. Characterising the solution structure will facilitate the development of a physicochemical understanding of crystallisation by revealing the relative importance of the interactions within the crystallising phase [119]. The evolving molecular interactions and thermal effects in a crystallising solution influence the eventual crystal properties [10].

In a liquid or solution, molecules experience a combination of Van der Waals, hydrogen-bonding, π-π and Coulombic interactions [10]. Interactions can broadly be defined as intramolecular interactions, e.g. covalent, ionic, or metallic bonds, or weaker, intermolecular interactions (Van der Waals interactions).

Van der Waals interactions are caused by the permanent or induced distortion of the electron density of one atom due to proximity of another's electron density [117], including dipole-dipole, dipole-induced dipole and induced dipole-induced dipole (London dispersion forces) [1]. As London dispersion forces are due to instantaneous dipole fluctuations [120], they tend to be weak, short-range and temporary. Dipole-dipole interactions, of which the hydrogen bond is an example, can occur between molecules, or between functional groups on the same molecule.

IUPAC defines hydrogen bonds as interactions occurring between a strongly electronegative atom with a lone pair of electrons, and a hydrogen atom that is covalently-bonded to a second relatively electronegative atom [21]. However, experimental and theoretical studies have shown that hydrogen bonds are not purely electrostatic interactions [121], as they may have electrostatic, polarisation, Van der Waals and covalent characteristics [116]. Hydrogen bonds are directional interactions [1] and have a large influence on the structure and

[2] With the exception of optically isotropic cubic systems.

properties of materials such as ionic liquids [116,122,123] and biological molecules [28], e.g. protein structures [124–127].

Many properties of water, e.g. unusually high melting and boiling points and a high dielectric constant, are due to oxygen's two lone electron pairs, making water highly polar and giving it its defining tetrahedral structure [124]. There is a similar mechanism for other compounds with a hydroxyl group, e.g. alcohols, sugars, and organic acids (and thus DNA and proteins) [128].

The strength of hydrogen bonds varies depending on the participating elements, ranging from being similar in energy to Van der Waals interactions to stronger than a covalent bond [116]. Weak hydrogen bonds (3-7 kcal mol^{-1}) play a role in DNA base pairing and the stabilisation of biomolecules, such as proteins and peptides [28,129]. It is claimed that some hydrogen bonds in solution have energies up to 20 kcal mol^{-1}, and understanding the intermolecular mechanism behind these interactions is of interest [28,129]. Hydrogen bonds also form between salt ions and solvent in solution. In this work, hydrogen bonds are considered to act over a range up to 2.5 Å.

In the crystallisation of small organic molecules, π-stacking is an important attractive force in addition to hydrogen bonding [130]. π-stacking is more likely in molecules where there are electron-withdrawing and electron-donating groups, typically heterocyclic aromatic rings. Three principal geometries of aromatic π-stacking have been identified, with parallel-stacking molecule separation in the region of the Van der Waals separation (3.3–3.8 Å) (Figure 1.7). In the case of phenyl rings, the T-shaped geometry with the C–H···π interaction was identified to be more stable than π-stacking [130]. Solvent polarity can affect the extent of π-π stacking, as cation π-π stacking of the ionic liquid C2mim[X] (where X = Cl^-, Br^- or I^-)[3] was reported for non-polar solvents at the expense of anion-cation ion pairing [131], whereas in propionitrile, long-lived anion-cation pairs formed [131].

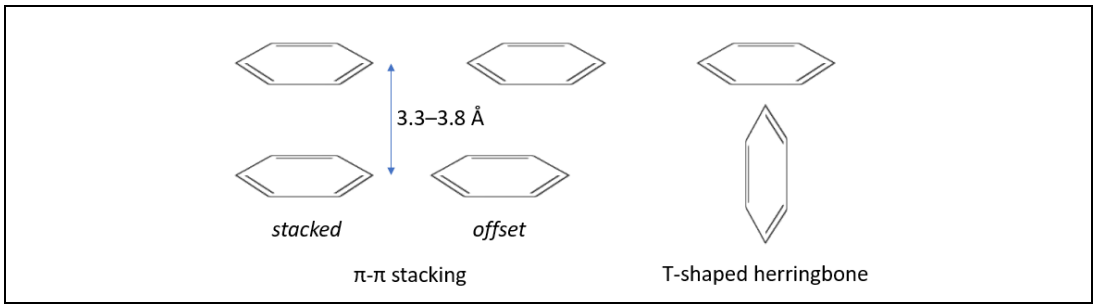


Figure 1.7 Geometries of aromatic interactions: π-π and herringbone T-shaped stacking motifs of phenyl rings, based on ref. [130]

[3] C2mim: 1-alkyl-3-methylimidazolium cation (n = carbon chain length)

The electrostatic (or Coulombic) contribution includes interactions between permanent charges and multipoles, e.g. dipole-dipole interactions, and may be attractive or repulsive [117]. It is the primary contributor to the lattice energy of an ionic compound [83]. Even the bonding of highly-charged ions will have some covalent nature to it, meaning that electron density will be shared to an extent between the ions [83].

Although the Coulombic interaction is strong and long-ranged, solvent shielding will reduce the distance over which the electrostatic force will operate. Due to the attraction between oppositely-charged ions, it is likely that anions will be located close to cations, and vice versa, in solution [132]. A salt solution will be neutral overall, but there will be local charge fluctuations, as multiple counterions surround a central, oppositely-charged ion [132].

1.3.3 Systems to be investigated

1.1.1.3 Imidazole hydrochloride

Imid is an organic heterocyclic compound with high solubility in polar solvents (Figure 1.8a). The lone pair of electrons on N1 (pK_a = 14.9 [133]) contributes to the delocalised π system, while the lone pair on N3 (pK_a = 7.0 [133]) fills the hybridised sp^2 orbital [134]. Imid and Imid derivatives play an important role in biological systems and are present in many natural and synthetic drug molecules [135]. Imid is a component of purine [136], and as a side-chain of histidine, Imid plays an important part in enzyme catalysis [137]. Imid acts as a proton relay due to its ability to form hydrogen bonds with water via the two N moieties [136–139]. A study of histidine residue identified that at physiological pH, where the protonation state of Imid is flexible, dimers of Imid and its protonated form imidazolium ($ImidH^+$, Figure 1.8b) form proton relay pathways [140]. The tuneable photophysical and electrical properties of Imid derivatives are also of interest for electronic applications, including organic light-emitting diodes (OLEDs) [141].

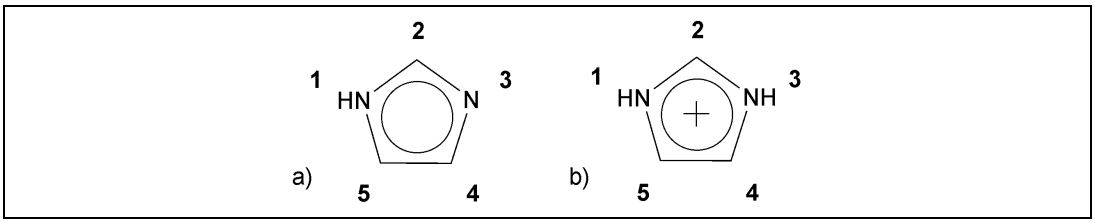


Figure 1.8 Molecular structures of (a) imidazole & (b) imidazolium – International Union of Pure and Applied Chemistry (IUPAC) numbering

Ionic liquids (IL), the class of salts comprising an organic cation and an organic or inorganic counterion with a melting point below 100°C [142–144], are commonly composed of an N-based heterocyclic cation, e.g. Imid- or pyridine-based cations [145], with various anions.

The properties of IL are tuned by the ionic composition [145,146] or the alkyl chain length on the cation [147]. IL have become popular as 'green' solvents [148], being highly-thermally stable, non-volatile and non-flammable. Imidazolium-based ILs are increasingly being used as environmentally more benign reaction media, replacing volatile molecular solvents [148–150], for biomass hydrolysis [151] and drug delivery [152].

The structures of IL and the existence of cation-anion ion pairing in IL solutions have been analysed using IL-compatible techniques such as neutron diffraction with isotopic substitution (NDIS) [147,153], infrared (IR) spectroscopy [146], nuclear magnetic resonance (NMR) [146,154,155] and molecular modelling [156–158]. The formation of supramolecular aggregates of IL molecules is due to hydrogen-bonding between the molecules. When diluted in low-polar solvent, contact ions pairs have been shown to be the dominant species, with solvent-separated ion pairs appearing in high polar solvents [157].

In this study, aqueous solutions of Imid and imidazole hydrochloride (ImidHCl) and their solid-phases are analysed by core-level X-ray spectroscopic techniques to compare their electronic structures. The speciation diagram for Imid is in Figure 1.9.

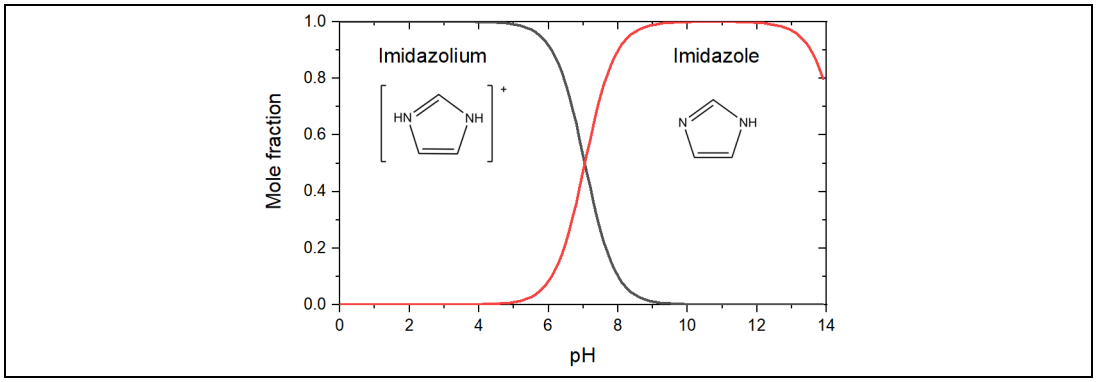


Figure 1.9 Speciation diagram for imidazole in water

The relative concentrations of protonated (black line) and neutral (red line) Imid are shown.

With a melting point of 158–161°C (source: Sigma Aldrich MSDS), ImidHCl is not an IL per se, but this study of the dissolved cation may be pertinent to IL research.

Imid in solution has previously been characterised by: near-edge X-ray absorption fine-structure (NEXAFS) to quantify changes in the unoccupied density of states of Imid with changing solution concentration [159], which found that Imid self-associates at low concentration (0.5 M); and NEXAFS detected by XRS [67] during a *in situ* cooling crystallisation [67], which showed that average Imid solvation shell structure stayed constant through the MSZ until the point of crystallisation.

Liquid-jet X-ray Photoelectron Spectroscopy (XPS) characterised and contrasted Imid and $ImidH^+$ in aqueous solution, finding pseudoequivalence of the N core-level binding energies (E_B) in $ImidH^+$ [160]. A combined X-ray Emission Spectroscopy (XES)/Resonant Inelastic X-ray Scattering (RIXS) study [161] indicated a limited influence of solute-solvent hydrogen-bonding on the electronic environment of the compounds [161].

In this work, a novel, combined near-ambient pressure (NAP-) XPS-NEXAFS-XRS approach is applied to obtain C and N K-edge fine-structure spectra and explore the unoccupied density of states of $ImidH^+$.

1.1.1.4 Guanidine hydrochloride

Guanidine (Gu), and its protonated form guanidinium, Gdm^+, are of chemical and biological interest [162,163]. The speciation diagram for guanidine is in Figure 1.10.

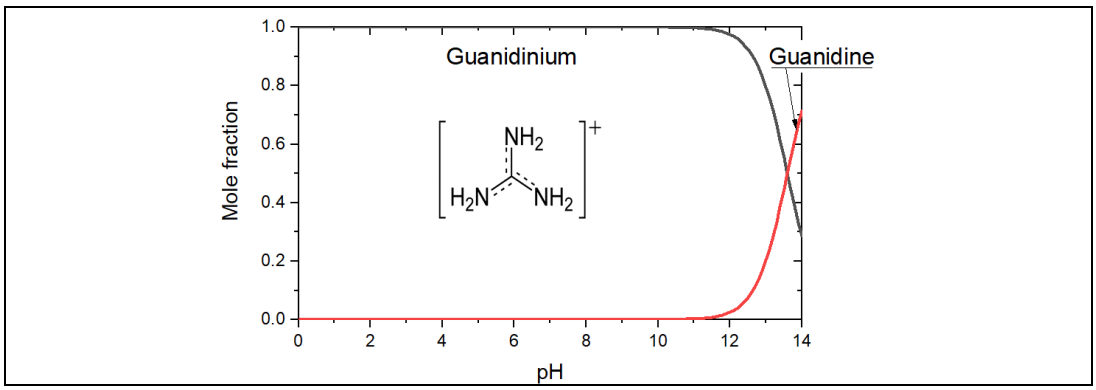


Figure 1.10 Speciation diagram for guanidine in water

The relative concentrations of protonated Gdm^+ (black line) and neutral Gu (red line) are shown

Gu derivatives are used in the production of melamine plastics [163], and explosives [164], and research into the development of drugs with a Gu core for diverse applications is ongoing [165]. Gdm^+ is found as a fragment or side chain of larger biological molecules, for example arginine, creatine and pyrimidine bases of DNA [163,166]. Gdm^+ salts are often used in biomolecular research to study ion specific effects [98].

Gu is a stronger organic base than most other amines (pK_a = 13.6) [167]. It readily accepts a proton to form Gdm^+ ($C(NH_2)_3^+$), as it can readily delocalise the charge over the three $-NH_2$ groups (Figure 1.11) [162]. The efficient charge distribution across the planar Gdm^+ ion provides electrostatic stability, with the highly-positive C centre attached to highly-negative N moieties with their highly-positive H atoms [162]. This stability enables Gdm^+ to form hydrogen bonds with water via the amino groups (N-H···O) in the plane of the molecule, but not via the central C atom (C···O-H) [168].

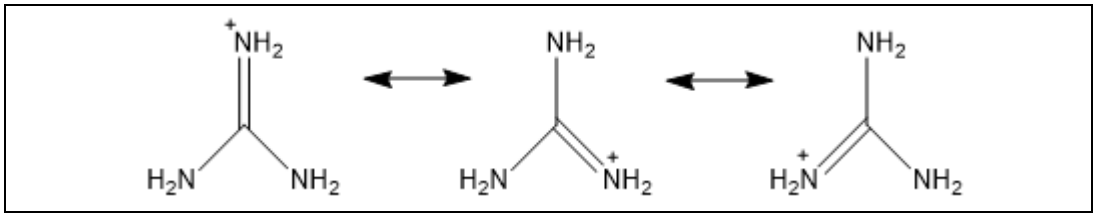


Figure 1.11 Guanidinium structure schematic, with its three resonance structures

In Gdm^+, each equivalent C-N bond takes on a double bond character due to the delocalised orbital resulting in the 'Y aromaticity' of the molecule [162]. Self-consistent field (SCF) calculations indicate that the rotation of the C-N results in a weakening of the π bond and increases the bond length of the rotating $-NH_2$ group [162]. As a consequence, the bond lengths of the other C-N bonds decrease, which strengthens the bonds and increases the energy barrier for the rotation of a second or third $-NH_2$ group [162], thus minimising rotation of the $-NH_2$ groups [169].

1.3.3.1.1 GuHCl – crystal structure

A search of the Cambridge Structural Database (CSD) revealed three deposited crystal structures of GuHCl (GANIDC [170], GANIDC01 [171], GANIDC02 [172]), although the crystallographic information file for GANIDC [170] is incomplete for X-ray pair distribution function (XPDF) analysis, so it will not be included further. All are space group Pbca (61) and the structures of GANIDC01 [171] and GANIDC02 [172] are shown in Figure 1.12. The calculated powder X-ray diffraction (PXRD) patterns of the two complete structures from Mercury [173] indicate a single polymorphic form (App. Figure 1).

1.3.3.1.2 Solvation of Gdm^+ ions

GuHCl has long been known to be a protein denaturant, especially at concentrations above 5 M [93,174,175]. One proposed mechanism is that the faces of the Gdm^+ ion associate with the hydrophobic surfaces on the protein at the same time as forming in-plane hydrogen bonds with water and polar functional groups on the protein [97,176]. Gdm^+ can inhibit salt bridging in protein structures, affecting the stability of protein folding [98,174].

The denaturing capacity of Gdm^+ depends on the counterion [168,177]. It is suggested that GuHCl is a strong denaturant owing to the absence of in-plane hydrogen bonding of Cl^- to Gdm^+, indicating weak ion pairing [168]. Proteins are stabilised by guanidinium sulphate (Gu_2SO_4) because of the strong interactions between the sulphate (SO_4^{2-}) ions and water [177]. It was previously proposed that SO_4^{2-} forms two linear hydrogen bonds with the Gdm^+ ions, creating nanoscale aggregates and preventing protein-Gdm^+ interactions [178], although there is debate about the strength of experimental data supporting this hypothesis

[177]. It has been calculated that Gdm^+ does not cause electrostriction, which may further highlight the importance of the counterion in the denaturant properties of Gdm^+ salt solutions [98]. It is worth noting that some of the experimental and theoretical analysis techniques discussed in the literature do not consider the stabilising (or otherwise) effect of the counterion in solution.

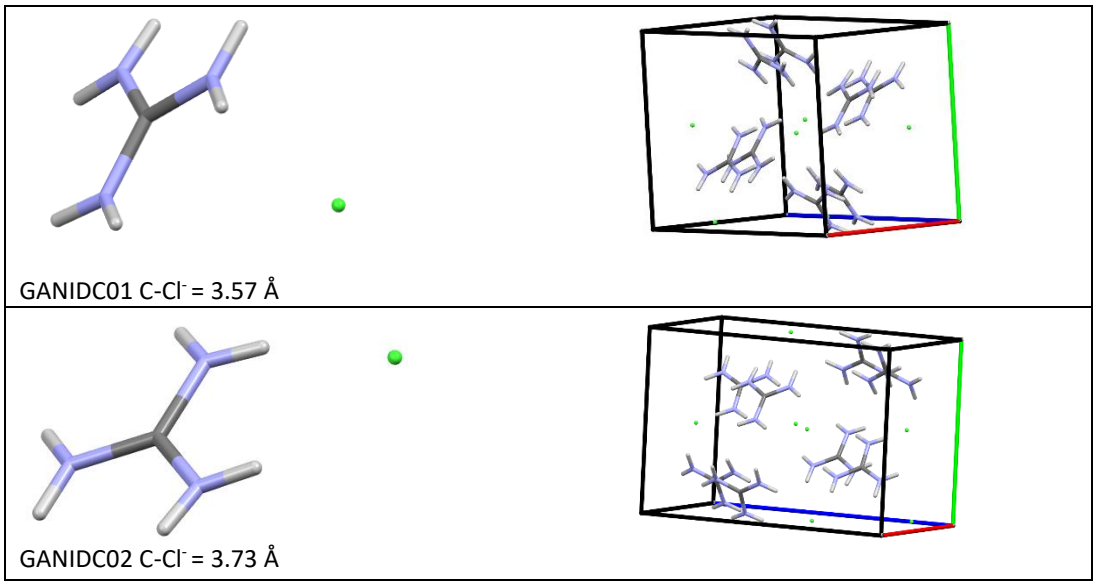


Figure 1.12 Solid structures of GuHCl, deposited in the CCDC: GANIDC01 [171] **(top), GANIDC02** [172] **(bottom), as depicted in Mercury**

The solvation of Gdm^+ has been described as 'bimodal' because the first solvation shell around Gdm^+ is anisotropic [97,168]. Water molecules interact around the plane of the molecule via the $-NH_2$ groups, but there is little interaction with the faces of Gdm^+ [97,168]. Several studies indicate that Gdm^+ ions have little effect on the structure of bulk water, as they are weakly hydrated and do not compete effectively with water molecules for hydrogen bonds [91,92,179,180].

A combined MD/NDIS study identified 4.5 hydrogen-bonded waters in the plane of the Gdm^+, with ~5 more in a diffuse cloud above and below the C atom [97]. Two infrared photodissociation (IRPD) spectroscopic studies indicate that the structure of the hydration shell of Gdm^+ changes with number of water molecules in its local environment [92,181]. Up to three water molecules form two hydrogen bonds with the $-NH_2$ groups, with each water molecule stretching between two amine groups [92]. The addition of further water molecules (n = 4 or 5) resulted in the formation of a second hydration shell, rather than there being interaction between water molecules and the central C (experimental and computational limitations prevented the identification of explicit hydrated structures beyond this) [92]. Molecular orbital analysis was in agreement, finding insufficient orbital

density around the C for hydrogen-bonding with water [92]. The second IRPD study, which was more representative of interactions of Gdm^+ in bulk water, clustered gaseous Gdm^+ with up to 100 water molecules [181]. It was more energetically favourable for the inner shell water molecules to accept only single, linear hydrogen bonds with the Gdm^+ ($NH_2\cdots OH_2$) to optimise the number of water-water hydrogen bonds where n ≥ 8 [181].

1.3.3.1.3 Like-charge ion pairing of Gdm^+

Understanding the like-charge ion pairing of Gdm^+ in solution could have an impact on the development of therapeutic molecules that can be delivered directly into cells [93]. There are extensive computational studies to confirm that Gdm^+, as solvated ions and as a protein side-chain, stack in either parallel or offset formation, form T-shaped dimer structures (Figure 1.7), or form solvent-shared ion pairs (Figure 1.6) [93,175,182,183].

As close interaction of two like-charged ions in the gas phase results in a very high repulsion energy (> 100 kJ/mol), the idea of like-charge ion pairing in aqueous solutions seems counterintuitive [93]. However, computational studies of Gdm^+ ions in aqueous solution indicate that the high dielectric constant of water attenuates the effect of the ionic charges in solution [91,93,181,182,184]. This attenuation enables 'cavitation effects' and stabilises the like-charge ion pairing through quadrupole-quadrupole and dispersion interactions.

In solutions containing ionic moieties, the electrostatic energy is the principal contributor to the solvation free energy, along with cavitation and dispersion-repulsion components [185]. In the polarisable continuum model (PCM) approach, there is a solvent exclusion effect, whereby the solvent dielectric is excluded from a volume (cavity) around the solute molecule, and the electrostatic free energy is provided by the interface of the solute's charge distribution and the dielectric medium [185]. It is the overlapping of these cavities upon ion complexation that can lead to stabilisation of dimers [185,186]. In an explicitly solvated model, dimer stabilisation was attributed to the hydrogen-bonding between the Gdm^+ ion and the surrounding water molecules, and between water molecules around the adjacent Gdm^+ ions [186].

Quantum chemical (QC) and MD modelling suggests that Gdm^+ ion pairs are (weakly) thermodynamically stable in water, including as side-chains of larger molecules [91,93,186]. Water stabilises the electrostatic repulsion, and whilst a stable dimer was achieved with a cluster of three water molecules, a 12-molecule water cluster provided maximum stability with a C-C atomic separation of 3.37 Å [91]. The addition of further water molecules in the second solvation shell had a negligible effect on stability [91].

When investigating the pairing of Gdm^+ as an amino acid side-chain using MD [91], five different Gdm^+ force fields were tested to see how they influenced the ion separation of contact ion pairs (CIP) and solvent-shared ion pairs (SShIP). Two of the force fields[4] resulted in CIP separations of r = 4 Å and SShIP at = 7 Å [91]. No CIP was found in the simulations with the other three potential parameters[5], which had C-C separations at around r = 12 Å according to the radial distribution function (RDF) peak [91]. However, further analysis showed that it was the parametrisation of the backbone dihedral angle that affected the Gdm^+ pairing interactions, rather than the force field applied to the Gdm^+ ions, and concluded that Gdm^+ ions in water interacted if they were not sterically prevented from doing so [91].

MD analysis of 2, 4 and 6 M GuHCl aqueous GuHCl solutions showed significant intensity at r = 3.85 Å, indicating Gdm^+-Gdm^+ CIP [182]. The intensity decreased with increasing concentration (although this does not relate directly to coordination number CoordN , which was not given, as CoordN is a density-weighted integration of the area under the peak) [182]. Conversely, the peak at ~7.7 Å indicating SShIP increased in intensity with concentration [182]. In this analysis, there was no mention of the T-type dimers that were identified in *ab initio* modelling [183], although the shoulder in the C-C RDF at ~4.5 Å could indicate their presence in the higher (4 & 6 M) concentrations [182].

Experimental observations of the liked-charged complexes in real solution conditions have been limited to NDIS combined with MD [97], and cryo-ion mobility mass spectrometry [187]. The cryo-ion mobility−mass spectrometry study investigated the size and shape of the solvation shell around Gdm^+ with reducing numbers of water molecules in the cluster [187]. The stabilisation of the parallel-stacked Gdm^+-Gdm^+ ions pairs by water bridging was explored [187]. The arrival-time distribution vs mass/charge analysis indicated a significant change in the hydration shell structure between n = 6 and 9 [187]. The orientation of the hydration shell changed from a dome-like hemispherical structure (n > 9) to a planar structure, with interstitial hydrogen bonding in the Gdm^+ plane occurring only when the cluster was sufficiently dehydrated [187].

The like-charge ion pairing behaviour of Gdm^+ has been compared with other cations in solution. *Ab initio* MD (AIMD) calculations compared an explicitly-hydrated Gdm^+ dimer to a set of control simulations (a hydrated NH_4^+ dimer) [93,179]. It was found that the dominant attractive interaction between Gdm^+ ions (Van der Waals) was not present in the NH_4^+

[4] nonpolarizable parm99 and polarizable pol-parm99
[5] nonpolarizable parm99SB, parm03 and parm10

simulations, where no contact like-charge ion pairs formed [93,179]. Although there is a risk that ion pairs may form in simulations as a result of artefacts of the applied force fields, *ab initio* and empirical approaches in the literature give rise to evidence of ion pair stacking [183].

Applying a PCM to represent the water solvent around like-ion pairs (Gdm^+, NH_4^+, Na^+ and NO_3^-) with a separation of 3.32 Å showed that the Gdm^+ pair was the only pair with an attractive potential in water (free energy of association -2.1 kcal/mol) [186]. The geometrically-similar NO_3^- ion had a lower repulsion energy than the more spherical ions (2.3 kcal/mol, compared to 7.4 and 5.6 kcal/mol) [186]. MD modelling of Gdm^+ as a side-chain in di-arginine identified intermolecular interactions above and below the Gdm^+ group, whilst water molecules and Cl^- ions were positioned around the plane of Gdm^+ [186].

Guanidine hydrochloride (GuHCl) has a relatively simple structure, and interesting thermodynamic[6] and cation-pairing behaviour in solution [91,93,168,186,188]. As GuHCl has previously been studied in a combined neutron diffraction / MD analysis [97], it was selected as a suitable organic salt candidate for the X-ray total scattering and computational modelling analysis in this work.

1.4 Liquid and solution system characterisation

1.4.1 Experimental X-ray techniques

Structural correlations in amorphous materials, including liquids or solutions, exist over the short- to medium-range (< 15 Å). Until recently, probing a transforming system to acquire molecular-level information about the pre-nucleation or intermediate phases that might exist, and deriving the intermolecular interactions at play, has been limited by the time- and length-scales of the experimental techniques available [46,52,189]. A suite of experimental methods, often applied in conjunction with computational modelling, are now available to probe the structure of solutions [190].

Temperature-induced phase changes in a crystalline solid with long-range, periodic structure can be derived from *in situ* X-ray diffraction (XRD) [191], but the technique cannot be applied to liquid or solution systems that only display local structure. *In situ* monitoring techniques

[6] Unlike some other salts, the activity and osmotic coefficients of aqueous solutions of Gdm^+ salts decrease with increasing concentration across the whole concentration range rather than plateauing before the maximum concentration is achieved [168].

have been developed to identify pre-crystallisation aggregation and growth stages for a range of material types including pharmaceuticals, e.g. a combined nonlinear optical/microscopy technique [44], but spatial resolution was limited thus individual molecular interactions were not identified.

The main experimental techniques that have been used in this thesis to probe the solutions of interest and provide some insight into the electronic and molecular structure of liquids and solutions fall into two categories, X-ray spectroscopy and X-ray scattering, an overview of which is given here. The techniques that have been used in this study are detailed further in Chapter 2.

1.4.1.1 Spectroscopic techniques

Spectroscopic techniques characterise materials by measuring the energy that is absorbed or scattered by the probed material, as a function of wavelength (energy) [192].

Laboratory-based spectroscopic techniques used to probe liquid and solution structures include IR, Raman, ultraviolet/visible (UV/vis), NMR and X-ray photoelectron spectroscopy (XPS).

Molecules store energy in translational, vibrational and rotational modes, and the energy required for the molecules to transition between these energy states can be used to infer information about their structure [132,193]. The intensity of the spectra will change according to adjacent atoms or the chemical environment [194].

IR detects the energies at which less light is transmitted through the sample, and has therefore been absorbed by the vibrational modes of molecular bonds, typically in the region 100-5000 cm^{-1} [195]. Optical resonant Raman spectroscopy is a light-scattering techniques that characterises molecular vibrational (phonon) and rotational modes, and electronic excitations in the meV range [196]. THz dielectric relaxation (THz-DR) and polarisation-resolved femtosecond infrared (fs-IR) pump–probe spectroscopy has also been used to study the dynamics of the solvent in a solution [180].

NMR, which detects the resonant frequency at which the nuclei in a sample respond to an applied magnetic field [197], has been used to study pre-nucleation aggregation in crystallising solutions [198], and *in situ* NMR has been developed to track solute-solute interactions during crystallisation from solution [198,199]. Information about the changing nature and relative probabilities of different solvent-solvent and solute-solvent interactions has also been detected using *in situ* NMR, providing limited information about the solution structure before and during crystallisation [200]. Although NMR is used to probe local atomic

arrangement, it does not provide all-atom structural information simultaneously and interatomic spacing is only yielded over a small range (5–6 Å) [201].

Core-level XPS provides information on the local electronic structure and the bonding of the probed element [159], and as such, has been used to distinguish salts from co-crystals in the solid state [202,203]. Until recently, XPS of solution systems has only been possible using liquid-jet systems [160,204]. However, the development of near-ambient pressure XPS has allowed the measurement of solution samples, as has been used in Chapter 3.

Over the last decade, synchrotron-based X-ray core-level spectroscopies have emerged as sensitive probes of the chemical state and local structure around molecular solutes and solvent molecules in solution [67,159,160,205,206]. X-ray absorption fine structure (XAFS) comprises NEXAFS and extended X-ray absorption fine structure (EXAFS). The section of the spectrum close to the absorption edge is the NEXAFS spectrum, whereas the EXAFS region can be 50-100 eV above the edge. Like NMR, the structural information derived from EXAFS is only over a short range [201].

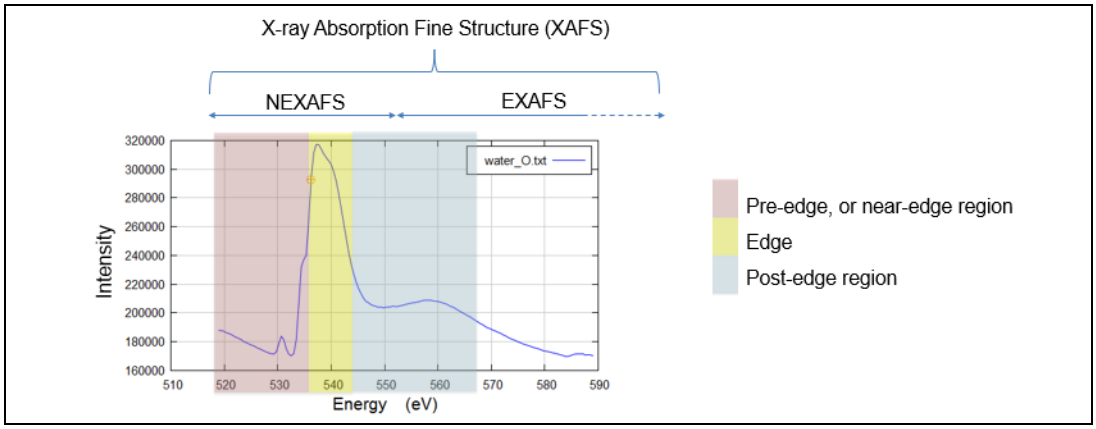


Figure 1.13 Example NEXAFS spectrum: O K-edge for a water sample measured in Kapton

The typical regions used in NEXAFS and EXAFS analysis are identified, along with the pre-edge, edge and post-edge regions

NEXAFS is particularly effective for molecules with containing low-Z molecules [207], such as drug molecules. As NEXAFS probes the unoccupied valence orbitals, the technique is highly sensitive to changes to an atom's chemical and physical environment, including protonation [67,159]. Its use to investigate pharmaceutical molecules in the solid state is relatively recent, and the study of solutions even more so.

X-ray Raman scattering (XRS), also known as non-resonant inelastic X-ray scattering (NIXS), is a hard X-ray technique that also probes the unoccupied density of states of atomic species to generate spectra that are proportional to NEXAFS [208]. XRS has negligible self-absorption effects and so produces a more reliable spectrum than can be achieved using soft X-ray

techniques [209], and as a bulk sensitive technique, it is more useful for studying the solution structure than XAS [210]. XRS is used to look at samples that are too complex to study under vacuum using soft X-rays due to the high absorbance of the sample window or ambient environment, or the high vapour pressure of the sample, and is ideal to measure *in situ* crystallising solution samples [211].

Like XAS, XRS is sensitive to hydrogen bonds, and as the time scale of the excitation and scattering process is orders of magnitude shorter than hydrogen bond dynamics, it is a suitable method to study the structure of liquid water [210,212–214]. NEXAFS spectra provide a direct probe of the molecular electronic structure, which is influenced by the local structure around the water molecules [215]. XRS was found to be sensitive to the interactions between acetonitrile and water, which was validated by density functional theory (DFT) calculations [216].

The effect of ions on water structure has relevance in many chemical and biological processes [217]. There are competing ideas on how ion hydration enhances or breaks down the structure and properties of the bulk water [215]. A study of the O K-edge spectra for aqueous solutions of sulphuric acid identified that increasing the concentration of ions increased the number of donated hydrogen bonds per water molecule [218]. It was also noted that XRS presents the opportunity to use backscattering geometry to study non-dipole transitions, which can give additional information on the local structure and sensitivity to speciation [218,219].

An *in situ* time-resolved XRS study of the solid-state dimerisation of cinnamic acid, combined with direct tomography to show the bond breaking and formation during the reaction process, captured the dimerisation and subsequent disintegration of the product [220].

Beam damage of liquid-phase samples can be reduce by agitating or circulating the sample in the beam [218,221,222]. For the *in situ* cooling crystallisation study of Imid [67], a flow cell was designed to minimise beam damage and allow temperature-controlled cooling of the solution. The insignificant change in spectral features through the MSZ indicated that desolvation is the rate-determining step in crystallisation and provided no evidence for the existence of pre-nucleation clusters [67]. The role of hydrogen bonding in the solvation of Imid in solution derived from the N K-edge spectra [67] was supported in a later neutron diffraction study [67,223].

The self-association of Imid has been studied extensively in the literature. Synchrotron X-ray techniques have been used to understand the effect of solution concentration [159] and the solvent type [224] on the electronic and chemical environment of Imid (XAS) and to

investigate the evolution of the structure of an aqueous Imid solution during cooling crystallisation (XRS) [67]. The study in Chapter 3 contrasting the XRS-derived N and C K-edge spectra of Imid and $ImidH^+$ complements and extends previous studies undertaken by the group [67,159,223], and includes the application of DFT calculations and NAP-XPS to support XRS data analysis.

1.4.1.2 Diffraction techniques

Diffraction techniques measure the scattering of the incident quanta (X-rays, neutrons or electrons), which is processed to infer structural information about a material [225]. The wavelength of the incident probe limits the resolution of the measurements.

Energy dispersive X-ray diffraction (EDXRD) uses an energy-discriminating detector to measure the Bragg reflections of diffracted polychromatic (white) synchrotron light [226]. EDXRD is a suitable technique for the *in situ* monitoring of crystal formation, but it does not provide information on solution structure [46]. To fully understand phase transformation mechanisms during crystallisation, molecular-level local structure data collected in real-time are required [227] and can be achieved with X-ray total scattering, which will also capture diffuse scattering to provide information about local structures when used in combination with XPDF analysis (Section 2.5.4).

X-ray diffraction methods are not sensitive to the position of hydrogen, with its single electron localised away from the nucleus, and the signal is dominated by scattering of heavier atoms. However, computational analysis of the data allows the hydrogen atom positions to be inferred. Tracking structural changes during *in situ* crystallisation experiments requires the high signal-to-noise ratio [186] of a synchrotron source and often a relatively complex sample environment, which can be accommodated at many synchrotron beamlines [186]. For structural analysis, X-rays require a wavelength comparable to atomic separation, which is feasible using synchrotron light. XPDF provides time-averaged information of the probed system, so faster acquisition times will make the technique increasingly suitable for phase change analysis.

NDIS is a powerful tool for the detection of hydrogen, and is useful for understanding the structure of a liquid or solution in equilibrium [100,147,153,223,228–230], but the time-scale of the data collection is insufficient for *in situ* crystallisation analysis. The use of neutron scattering as a probe for organic and biological materials has been limited by the low flux and long data acquisition times [231].

The structure of water has been explored with neutron diffraction and X-ray total scattering studies, combined with pair distribution function (PDF) analysis [232–237]. XPDF analysis has been used to probe liquids, and organic solutes and nanoparticles in solution [238–242]. NDIS and XPDF analysis of Imid in aqueous solution has shown the techniques to be comparable [223].

Key to investigating phase transitions in the crystallisation process will be the acquisition of time-resolved data at a sufficiently short time-scale for structural determination, which requires short pulses of X-rays, electrons or neutrons [243]. The X-ray free-electron laser (XFEL) generates a monoenergetic, relativistic electron bunch that is first passed through a linear accelerator (Linac) before interacting with magnetic undulators, resulting in the emission of sub-femtosecond pulses [244]. Techniques that will benefit from XFEL include XAS, X-ray diffraction and X-ray coherent diffractive imaging [243].

Due to the strong scattering of electrons, reliable PDFs have been obtained from laboratory electron microscope electron-diffraction data [245]. Electron diffraction is being developed for the study of nanostructures, but its application is limited to thin film samples (below 10 nm) by attenuation and low penetration [245].

Ultrafast electron diffraction (UED) uses femtosecond or even attosecond [244] electron pulses to generate real-time diffraction imaging for the study of ultrafast phase transformations and chemical reactions [246]. With a larger elastic scattering cross section than X-rays, data with a higher spatial resolution can be acquired with UED [246]. Radio-frequency acceleration-based photoemission electron guns produce a sufficiently high-brightness electron beam to study the ultrafast dynamics of biological systems and solid-state chemistry [246]. Atomic-resolution direct imaging to track a nuclear wave packet of gas-phase molecules during nonadiabatic processes has been achieved using UED [247].

A study combining the ultrafast X-ray and electron diffraction of small gas-phase organic molecules found the two techniques gave rise to complementary molecular structure measurements [248].

The HEIMDAL instrument at the new spallation neutron source (European Spallation Source) will combine thermal neutron powder diffraction (TNPD), small angle neutron scattering (SANS) and neutron imaging (NI) so that different probes will generate data for different length scales [231], and the high-flux neutron source has been designed for the *in situ* study of phase transitions [249]. The small-angle neutron scattering (SANS) and low-resolution total scattering will be used to probe the pre-nucleation phase, with diffraction to provide structural information on the solid phase [249]. At present, the available sample

environment would not accommodate a laboratory-scale flow cell, as used in the XPDF experiments described in this thesis, but the suitability of a neutron probe for low-Z elements [249] means that future advances in the neutron diffraction technique could provide great insight into nucleation and crystallisation mechanisms.

The XPDF study of a crystallising system described in this work indicates the potential of the technique to understand the evolving structure of a crystallising solution from *in situ* X-ray total scattering measurement.

1.4.2 Computational techniques

1.4.2.1 Quantum mechanics methods

Quantum mechanics (QM) methods, e.g. *ab initio* and density functional theory (DFT) calculations, are used to describe the electronic structures of atoms, ions or molecules. As the calculations can be computationally expensive, generally only a small number of atoms are considered (up to a few hundred) [250]. QM methods are used for the geometry optimisation (or energy minimisation) of molecules, which identifies the structure of the lowest energy configuration. Conformational analysis verifies that the global energy minimum, not a local minimum, has been identified [251]. Reasonable energy approximations for small, neutral molecules may be obtained from calculations of individual molecules in the gas phase. However, calculations for a molecule in solution, especially where polar solvents are present, require the use of either an implicit PCM to represent the dielectric constant of the surrounding molecules, or for solvent molecules to be explicitly placed around the solute molecule [251].

Electronic transitions (excited-state systems) can be modelled using QM methods [250]. Spectroscopic spectra, e.g. IR or X-ray absorption, can be generated from the QM excited-state calculations to inform the interpretation experimental data [250]. It is possible to predict the excitation energy of an electron when an atom or molecule is exposed to an incident beam by considering the molecular polarisability [251]. When there is divergence in the estimation of the molecular polarisability, which occurs when the quantum of energy absorbed by the molecule is equal to the energy gap between the ground and excited states, the excitation energy can be estimated [251]. This method is known as time-dependent DFT (TDDFT) (or TDHF when applied to restricted Hartree Fock calculations) [251]. TDDFT is a popular quantum chemistry method, as it calculates electronic excited states in a computationally economic way [252,253].

1.4.2.2 Molecular mechanics

Computer modelling is used for *in silico* experiments across a range of time- and length-scales at different points in the drug discovery process [250]. The molecular modelling of chemical and biological systems at the atomic and molecular scales enables analysis such as the adsorption of small molecules to organic and inorganic surfaces to provide insight into drug delivery mechanisms [250].

For large molecular systems, molecular mechanics (MM) simulations are performed, requiring classical force fields, as opposed to QM force fields. MD computational tools simulate the physical basis of a system and are used to describe natural phenomena and to design new compounds [250,254]. MD models can include > 10,000 molecules [255]. Starting with a system at equilibrium at t = 0 s, the components in an MD model move in accordance to Newton's equations of motion, and the trajectories of the components are followed as the system evolves [250]. The distance of each atomic movement is based on the system temperature and the defined timestep [251]. MD models have been used to reproduce the structure of proteins Id other biological molecules based on experimental single crystal X-ray diffraction (SCXRD) and high-resolution XRD data [250].

The Monte Carlo (MC) method is a statistical probabilistic technique to model classical many-body systems in equilibrium (as opposed to dynamic systems) [256]. MC simulations are based on the random motion of components, including molecular rotation or translation, or dihedral motion, i.e. not Newton-based physical movements [257]. The component movement will be accepted if the potential energy of the system decreases.

Force fields are a critical aspect of classical MM simulations, as they define how the components interact with each other [254] (Section 1.4.2.3). Force field selection for the simulations in the EPSR studies is discussed in Section 2.9.2.2.2. Molecular modelling of water has been extensively researched in the literature, and a discussion on the water model selection for the simulations in this study is in Section 2.9.2.2.2.

1.4.2.3 Force fields

Force field parameterisation is a specialist area of research, as the quality of the force field parameters will impact the accuracy of the molecular model, with inconsistency in force fields leading to artefacts in the data [183,258]. The interaction potentials considered in the force field include the Lennard-Jones potentials and Coulombic potential.

The Lennard-Jones potential (or 6-12 potential) describes the potential energy of interactions between two nonbonded atoms, as a function of their separation r (Equation

9). The Lennard-Jones potential is a combination of the short-range repulsion felt by atoms/molecules as they approach (12-potential), and the attractive dispersion force that exists when the separation distance increases (6-potential) [251]. The Lennard-Jones potential (Figure 1.14) is given by

$$\mathrm{U}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad \textbf{Equation 9}$$

where U is the particle-particle potential energy, ε is the well-depth parameter (kJ mol^{-1}), and σ is the range parameter (Å).

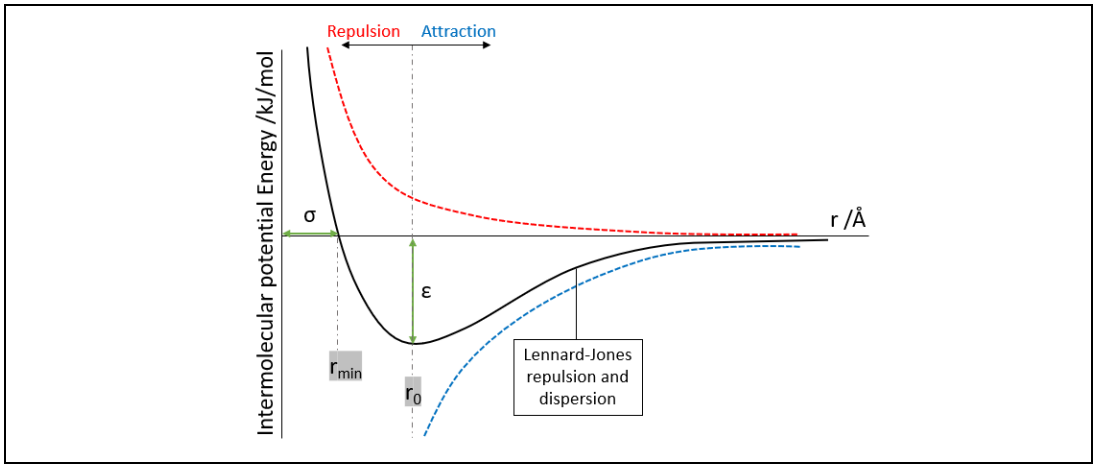


Figure 1.14 Lennard-Jones potential: describing the attractive and repulsive interactions between simple atoms and molecules

Ab initio force fields are derived from QM calculations using theoretical principles. The challenge of *ab initio* parameterisation is that the computationally-expensive basis sets needed to acquire accurate Lennard-Jones parameters preclude the analysis of larger molecules, whilst lower levels of theory exclude electron correlation, and thus would not treat dispersion interactions sufficiently to produce reliable parameters [255]. As *ab initio* calculations are based on a single small molecule, or a small molecular cluster, the long-range effects of other molecules in the system, for example solvent molecules, are not captured.

Empirical force fields are developed by reproducing the experimental condensed-phase properties of a compound, such as heats of vaporisation or molecular volumes [255]. Depending on the applied calibration procedure, e.g. degrees of freedom considered, target properties, optimisation strategy, different sets of parameters can generate similar results in empirical force field development, the so-called 'parameter correlation problem' [254,255]. By using additional target data, such as *ab initio* interaction geometries and

energies to assess the parameterisation, a combined empirical/*ab initio* method for force field development can be employed [255].

Different empirical force fields have been developed to suit the type of simulation and required parameter granularity. With all-atom (AA) force fields, parameters are allocated to each atom (interaction site), while the coarser-grained united-atom (UA) force fields consider a group of atoms to be an interaction site [254]. UA force fields result in less computationally-intensive simulations, but are appropriate where the H atom motion is of little interest and the key intermolecular interactions do not involve H atoms [254].

There are many examples of optimised potentials for liquid simulations (OPLS) UA and AA force field datasets providing Lennard-Jones parameters for series of molecules or components [259–263]. Whilst some Lennard-Jones parameters are said to be transferrable between similar atom types, for example the parameters for an alkane may also be used for an alkane sub-group on another molecule, it is not necessarily the case for the defined partial charges [254,260].

Atomic partial charges can be derived from population analyses in *ab initio* calculations and are indicative of the charge distribution in a molecule, although the Löwdin and Mulliken methods may be unreasonable where diffuse basis sets are applied [264,265]. Alternative calculation techniques apply an electrostatic surface potential methodology, e.g. CHELPG (Charges from electrostatic potentials) or RESP (restrained electrostatic potentials), and although they are more time-consuming, they may produce better approximations, especially in charged systems [251,266,267].

The phase of the investigated system influences force field selection [267]. Lennard-Jones parameters that have been applied to chlorine in the literature change with its environment, with σ ranging from 3.5 Å (organic chlorine atom) to 4.4 Å (chloride ion in aqueous solution), with 3.77 Å applied to an ionic liquid Cl^- anion [267]. Charge reduction, or scaled-charges, is an empirical technique that has been applied to molecular models in the literature [107,136,268–270] to crudely represent dielectric screening effects [271]. Reduced charges were applied in an EPSR simulation [269] on the basis that the charges had been derived for an isolated gas-phase molecule, rather than one that is interacting with other molecules like in the probed system. Charge parameterisation, including the application of reduced charges, for the EPSR modelling of aqueous solutions of GuHCl is considered in Appendix F.

1.4.2.4 Water models

The structure of water has been of continued interest over the years due to its ubiquity, complex behaviour, and interesting properties [272]. Its importance to life sciences means that there is a large body of research into the structure of and self-interactions of water [213,272–274]. There are competing theories about its structure even at ambient conditions, let alone before structural modifications, e.g. extreme temperatures and pressures, or the addition of solutes are considered [272]. As new analytical and computational techniques emerge, and the resolution of data is improved, new hypotheses about the network of hydrogen-bonded water molecules are proposed [234,275].

Many models representing the structure and behaviour of water molecules have been developed empirically or parameterised by *ab initio* calculations [276]. An empirical model generally represents some, but not all, properties of water successfully, or may describe water well macroscopically, but not microscopically [276,277]. Rigid water molecules constrain the intramolecular degrees of freedom that are relaxed in more flexible molecules [276].

Polarisable water models describe the many-body inductive effects in water [276]. Although they may generate more realistic and robust molecular simulations, incorporating polarisation effects increase the complexity and the computational cost of the model [278]. Effective pair potential models are often simplified to exclude polarisability occurring beyond the pair interactions, even though polarisation in a polar liquid strengthens intermolecular bonding, and so ideally should be considered [279].

Two families of commonly-used water models used in the simulation of aqueous systems are TIPnP [137,169,178,280–283] and single point charge (SPC) [178,183,280,284–287] models. The TIPnP family of empirical water models was specifically developed for biomolecular simulations and attracts a low computational cost [276]. The SPC/E model is an extended SPC model, which incorporates the self-energy due to polarisation overlooked by many effective pair potential models [279].

Whilst the behaviour of the water molecules is an important feature of a solvation study, there is no single method to follow in the literature. Indeed, mixture models of water were first suggested by Röntgen in 1892 [288] and have been proposed elsewhere, on the basis that the structure of water is non-uniform. X-ray scattering and X-ray absorption and emission data identified density fluctuations in ambient water, indicating that water is inhomogeneous at the nanometre length-scale [275]. Soper [289] initially suggested that the observed fluctuations were consistent with fluctuations expected of any liquid, due to the

random and dynamic nature of the molecular structure, and that water molecules are arranged in a roughly tetrahedral configuration, However, his later research, a Monte Carlo simulation of a two-component water model validated with X-ray and neutron total scattering experimental data, captured the structure and thermodynamics of water at ambient conditions [232]. The water molecules only formed strong hydrogen bonds with 'unlike' water molecules, although 'like' and 'unlike' water molecules appeared in the first coordination shell [232]. The coexistence of two types of local structures in water (ordered (tetrahedral) and disordered (non-tetrahedral) structures) was proposed in another, classical MD modelling of X-ray and neutron scattering experiment data [290]. In contrast to the mixture model idea is the continuum model, which describes the structure of water as a network of continuously distorted hydrogen bonds [290].

Details of water model selection for the EPSR simulations are in Section 2.9.2.2.2.

1.5 Aims and objectives

Recent developments in experimental and computational methods provide an opportunity to understand more about the structure of a crystallising solution and the speciation of its components. The momentum transfer of high-energy X-rays, as they are scattered before and during crystallisation, can be analysed to reveal insights into the composition and structure of the crystallising solution phase.

The aim of this project is to develop experimental X-ray techniques, along with the associated computational calculations and modelling, to characterise the phase from which organic salts crystallise. X-ray techniques, including XRS, XPS and X-ray total scattering, are applied to identify the speciation of the salt in solution and derive the structure of the crystallising solution. Using this information, the intermolecular interactions between the ions, solvent molecules and the crystal product during the cooling crystallisation of an organic salt can be evaluated, including the propensity of ions to form ion pairs in solution.

Measuring the evolution of the intermolecular interactions in a crystallising solution is a key step to understanding the molecular-scale processes that occur during the crystallisation of pharmaceutical organic salts.

The main aims of this project will be met by achieving the following objectives:

1. Probe the speciation of organic salt ions in aqueous solution using hard X-rays and compare the protonated ion with the neutral form of the organic molecule.

- Collect C and N K-edge spectra of aqueous Imid and $ImidH^+$ in real-world conditions using hard X-ray XRS.
- Develop and apply a protocol to measure the C 1s and N 1s core-level binding energies of aqueous Imid and $ImidH^+$ using NAP-XPS for the interpretation of the XRS spectra.
- Undertake TDDFT calculations to understand the expected electronic 1s→π* transitions in Imid and $ImidH^+$, and assign transition peaks of the experimental C and N K-edge XRS spectra.

2. Use a combined XPDF/modelling approach to reveal the intermolecular interactions between the ions and solvent molecules in a concentrated aqueous solution of GuHCl during cooling crystallisation.
   - Collect a series of time- and temperature-resolved, *in situ* X-ray total scattering data during the cooling crystallisation of GuHCl from aqueous solution, with sufficient counting statistics for the acquisition of high-resolution XPDF patterns.
   - Analyse the XPDF patterns derived from X-ray total scattering data to evaluate the structure of GuHCl in aqueous solution prior to nucleation.
   - Apply a combined XPDF/EPSR approach to generate structural models of the system at different points in the cooling crystallisation process.
   - Propose the molecular structures that could be present in the pre-crystallisation phases to account for the intermolecular interactions identified in the structural models.
3. Use a combined XPDF/modelling approach to study the effects of concentration on the intermolecular interactions in aqueous GuHCl solutions.
   - Generate structural models of three different concentration aqueous GuHCl solutions.
   - Identify concentration-related changes to the solution structure, including anion-cation and like-charged (Gdm^+-Gdm^+) ion pairing and ion solvation.

The advancement of experimental X-ray and complementary theoretical computational techniques is essential to identify the optimum nucleation and growth conditions for crystalline products with the desired properties. A fundamental understanding of the crystallisation pathways of organic salts is required to inform predictive design, and improve the control of crystallisation processes. A transformative development like this would result in more efficient and reproducible industrial-scale salt production, which could be applied in many industrial sectors.

1.6 Outline of thesis chapters

Chapter 1 introduces the drivers for the development of drugs in salt form. Classic and multistage nucleation theories, issues specific to ions in solution, and characterisation techniques that can be used to probe solution systems are outlined. The compounds studied in this work (ImidHCl and GuHCl), including previous investigations into the compounds in solution, e.g. solvation structure and the like-charged ion pairing of Gdm^+, are presented.

Chapter 2 describes the theoretical background for the experimental and computational techniques used in this thesis including XRS, TDDFT calculations, and X-ray total scattering and EPSR modelling for XPDF analysis. In addition, the complementary analytical techniques that are used to characterise the chemical composition and physical form of the crystallised organic salt samples, including PXRD, SCXRD and helium pycnometry, are described.

Chapter 3 explores the use of core-level spectroscopic and X-ray diffraction techniques to characterise organic salts in solution and the phase from which the salts crystallise. The C and N K-edges of aqueous Imid and $ImidH^+$ were acquired using XRS and analysed using NAP-XPS measurements of the aqueous solutions and TDDFT calculations, to compare the speciation of $ImidH^+$ in aqueous solution with that of aqueous Imid.

The focus of Chapter 4 is the collection and analysis of X-ray total scattering data to study the evolving structure of a 9 M aqueous GuHCl solution during cooling crystallisation. Data were collected during an *in situ* cooling crystallisation experiment at the XPDF beamline at Diamond Light Source (DLS). Structural changes that take place as the crystallising solution cools through the MSZ are identified using XPDF pattern analysis. To interpret the X-ray total scattering data further, statistical structural models of the concentrated aqueous GuHCl solutions are developed using EPSR simulations, which are refined to the experimental data. These models are interrogated to identify the intermolecular interactions present in the pre-crystallisation phases during the cooling crystallisation process.

Chapter 5 reports on the variations in the aqueous GuHCl solution structure as a function of concentration, using the combined XPDF/EPSR technique developed in Chapter 4. Finally, Chapter 6 summarises the key findings of the work described in this thesis, and suggests interesting future research that could be conducted to further develop the work undertaken here.

Chapter 2 Research techniques

2.1 Introduction

X-ray techniques have been used to characterise organic salts in solution and the phase from which the salt crystallises. This chapter describes the materials, including sample preparation, experimental and computational methods, and analytical techniques that have been applied in this research.

X-ray core-level spectroscopy and TDDFT calculations have been used to investigate the speciation of ImidHCl in water. The evolving structure of an aqueous GuHCl solution during cooling crystallisation has been measured in an *in situ* X-ray total scattering study, and the structure of aqueous GuHCl solutions at different concentrations have also been analysed. EPSR (Monte Carlo) simulations were used to develop statistical structural models of the systems based on the X-ray total scattering data, which could be interrogated to identify the intermolecular interactions present in the solution structures. In addition, complementary analytical techniques, including powder and single crystal X-ray diffraction (PXRD/SCXRD), helium pycnometry and scanning electron microscopy with energy dispersive X-ray spectroscopy (SEM-EDX), have been used to characterise the chemical composition and physical form of the solid samples.

2.2 Materials and sample preparation

2.2.1 Imidazole and imidazole hydrochloride

Imidazole (Sigma Aldrich, puriss. p.a. grade, purity ≥ 99.5%) and hydrochloric acid (HCl, Fluka, puriss. p.a. grade, purity ≥ 32%) were used without further purification.

5 M aqueous Imid solution for XRS analysis was prepared with laboratory grade purified deionised water (MilliQ-POD, resistivity = 18.2 MΩ.cm), resulting in a solution of pH 10. For the ImidHCl solution, HCl was added dropwise to the aqueous Imid solution, resulting in an aqueous solution at pH 3, indicating that protonation of Imid to form $ImidH^+$ had occurred.

For XPS analysis, a ~2.5 M aqueous Imid solution was prepared and put directly in the sample chamber. The solid ImidHCl sample dissolved in the humid atmosphere in the sample

chamber, which was provided by laboratory grade purified deionised water in an open bottle (Suez Select Fusion, 18 MΩ.cm).

ImidHCl salt was required for XPS, SCXRD and XPDF analysis. Two methods were used to synthesise the colourless compound:

Preparation of solid ImidHCl by rapid evaporation: A concentrated solution of ImidHCl was prepared from Imid and conc. hydrochloric acid (10% stoichiometric excess). The solution was heated on a thin aluminium foil to evaporate the water, and the remaining molten salt was then cooled.

Preparation of solid ImidHCl by evaporation under vacuum: Concentrated HCl (Sigma Aldrich, puriss. p.a. grade, purity ≥ 37%) (12.49 g) was added dropwise into Imid (8.56 g, without modification or purification – stoichiometric ratio), with magnetic stirring to augment the dissipation of heat during the addition process to produce a 6.9 M ImidHCl aqueous solution. The solution was placed dropwise on a glass petri dish, which was put in a desiccator containing desiccating material (3 Å molecular sieves, 4–8 mesh). A vacuum pump applied a vacuum until the salt was dry.

Storage of solid ImidHCl: The salt was stored under vacuum, as it is highly hygroscopic.

2.2.2 Guanidine hydrochloride

GuHCl (VWR, ≥98%, Technical Grade) was used without further purification. Aqueous GuHCl solutions for the capillary and *in situ* analyses were prepared using laboratory grade purified deionised water (Suez Select Fusion, 18 MΩ.cm). For analysis of the crystallised product by helium pycnometry, PXRD, SCXRD, SEM-EDX and XPDF, the GuHCl was recrystallised in the jacketed vessel by reducing the temperature of a 9 M solution from ~65–~20°C. GuHCl crystallises as a colourless crystalline compound, which was vacuum filtered using a Buchner flask and funnel and rinsed with chilled deionised water. The sample was then dried in an oven at ~60°C for approx. 6 h.

2.2.3 XPDF sample preparation

2.2.3.1 Capillaries: powder samples

A powder average should be obtained in the XPDF measurement, so powder samples should be ground to produce a small particle size (ideally < 40 μm) [291]. The powder is then packed into a 1.5 mm outside diameter (OD) borosilicate capillary to a level of around 1.0–1.5 cm—to ensure that the sample would be in the beam when mounted on the capillary spinner—

and sealed with clay. As ImidHCl is highly hygroscopic, the sample was dried under vacuum after grinding before filling the capillary.

2.2.3.2 Capillaries: liquid samples

A 1.5 mm capillary of deionised water (18.2 MΩ.cm Suez Select Fusion) for the comparison study related to the concentration study (see 2.5.4.1.2 and Chapter 4) was clay-sealed.

Samples to be measured in the high-throughput beamtime were in 2.0 mm OD borosilicate capillaries and mounted in the static (non-rotating) sample holder (see 2.5.4.1.2).

Stock solutions (50 ml) of 2, 4 and 6 M GuHCl samples were prepared in volumetric flasks with GuHCl (as received) and laboratory grade purified deionised water (Suez Select Fusion, 18 MΩ.cm), as described in Table 2.1. The solubility of GuHCl at 20°C is 6.0 M. This was the maximum concentration selected for analysis in the GuHCl concentration study, as capillary XPDF measurements were taken at room temperature.

Table 2.1 Aqueous GuHCl solution sample preparation (50 ml samples prepared)

Sample concentration /M	Mass of GuHCl /g	Mass of water /g	Density of solution[7] /g.cm^{-1}
2	9.69	42.24	1.04
4	19.18	35.35	1.09
6	28.72	28.35	1.14

GuHCl solution capillaries were heat-sealed. For this, the capillary was dipped in liquid nitrogen to freeze the sample and prevent sample evaporation and mounted in a heat-proof holder (Figure 2.1a&b).

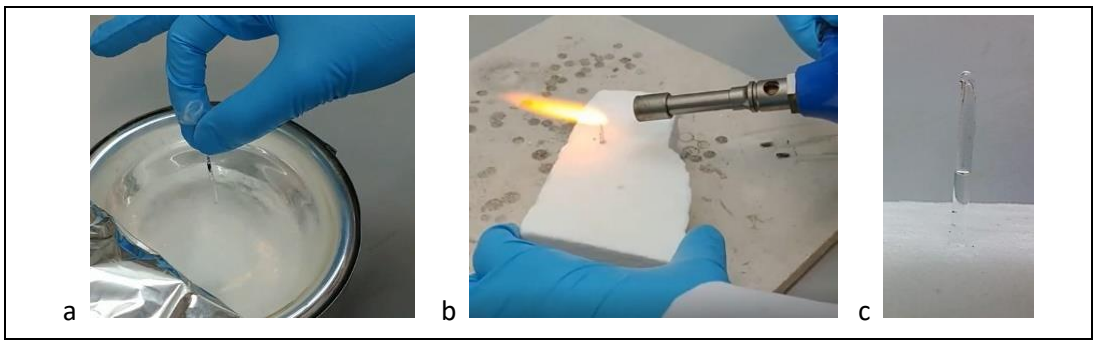


Figure 2.1 Production of sealed borosilicate glass capillaries

a) sample is frozen in liquid nitrogen, b) top of capillary is melted using a Bunsen burner and c) glass is sealed with tweezers

The top of the capillary was heated using a Bunsen burner and the top of the capillary closed using tweezers (Figure 2.1b&c). This process was undertaken in a fume hood. For protection,

[7] Density was calculated from the total measured mass (g) of water and GuHCl / 50 (ml)

whilst allowing safe manipulation of equipment, cotton gloves were worn between an inner and outer nitrile glove. The 2.0 mm deionised water capillary (Suez Select Fusion, 18 MΩ.cm) was also heat-sealed following the process outlined above.

2.2.3.3 *In situ* crystallisation of GuHCl

The *in situ* crystallisation experiment needed to take place within a temperature range of 5–70°C, due to the use of the water bath for solution temperature control. Experiments were conducted to identify the solution concentration that would dissolve < 70°C and reproducibly crystallise in range 10–25°C.

The initial selection of the starting concentration for *in situ* analysis was based on the outcome of the Crystal16 analysis (see Section 2.8 and Appendix A for details). A 9 M solution of GuHCl was made using 127.68 g GuHCl (as received) and 54.64 g deionised water. The solution was stirred in the jacketed vessel whilst it was heated to ~65°C to fully dissolve the solute. A solution volume of 149ml was measured.

The 9 M GuHCl solution crystallised at 13.6–14.4°C in Crystal16. At a larger scale, in the jacketed vessel with magnetic stirring, but without the Marprene tubing, crystallisation occurred at < 10°C due to the different system parameters, e.g. volume:surface ratio, relative stirrer size, and heating and cooling rates. The addition of the Marprene tubing and Kapton tubing with metal Swagelok fittings changed the surfaces that the crystallising solution contacted, and the mixing regime and temperature differentials in the jacketed vessel, and crystallisation occurred at a solution temperature of ~21°C. The solution pH ranged from 4.6–4.9 across a temperature range of 45.0–23.5°C.

2.3 Synchrotron science

Accelerating a charged particle, either in a straight line or along a curved path, results in the emission of electromagnetic radiation. Synchrotrons are circular particle accelerators, where charged particles are accelerated to generate short-wavelength, high-energy radiation. Although any charged particle could be used, radiation emission intensity is indirectly proportional to the mass of the particle hence electrons are used [292].

X-ray radiation wavelengths typically vary from 0.01 to 1 nm [293], or even shorter with high-energy accelerators [293], making X-rays suitable for the study of the structure and dynamics of matter with interatomic distances in the order of angstroms (Å) [294].

The layout of a typical synchrotron facility is shown in Figure 2.2 [295,296]. Bunches of electrons are accelerated to relativistic speeds in the linear accelerator (Linac) using electric

fields. The electrons are further accelerated in the booster synchrotron, after which they enter the storage ring. Focusing magnets act as a magnetic lens to focus the electron beam [297].

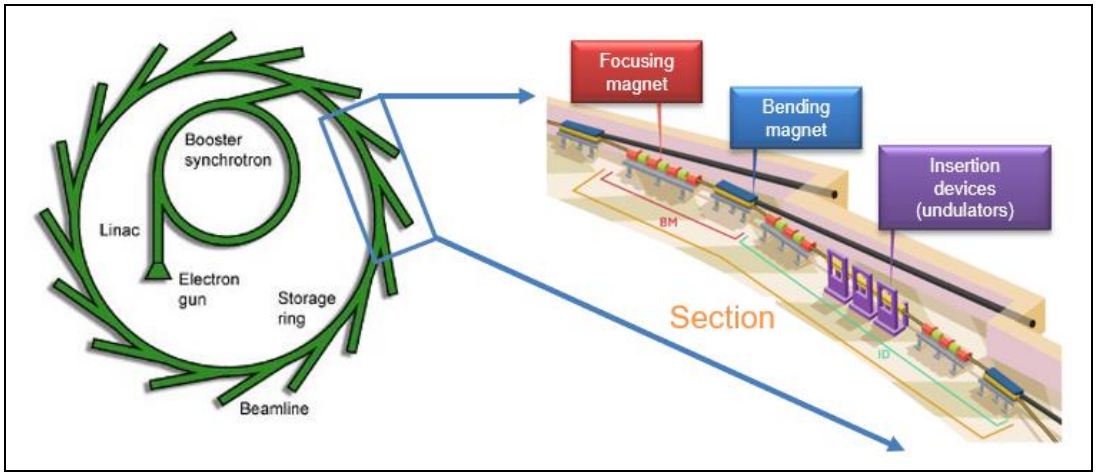


Figure 2.2 Layout of a storage ring (from ref. [295]**) and the layout of magnetic devices within it (from ref.** [296]**)**

The storage ring is surrounded by beam lines that supply radiation to experimental hutches, with instruments and detectors specific to the requirements of application of the radiation. The source of the radiation is either a bending magnet or an insertion device.

The electrons lose energy, which is emitted tangentially as X-rays, when the electron beam changes direction. Bending magnets force the electrons through a single bend, causing the release of a broad-spectrum beam of electromagnetic radiation ('white' radiation), with wavelengths from ~10^{1} to ~10^{-9} m [297].

A wiggler is a type of insertion device comprising arrays of magnets of opposite polarity, positioned to cause the electrons moving between them to oscillate and emit high-energy radiation [292]. The brilliance of X-rays emitted from wigglers is approximately two orders of magnitude greater than for bending magnets [297].

An undulator is an insertion device located in the straight section of the storage ring and comprises arrays of opposite-polarity magnets [297]. The many, small deflections that the electrons undergo cause constructive interference, resulting in an intense beam at the energy selected by the undulator [297]. The brilliance of the beam can be four orders of magnitude greater than the beam emitted by a bending magnet [292,297]. X-rays inserted into the experimental hutch via an undulator are hard X-rays (> ~10 keV) [298]. In soft X-rays systems, samples are required to be under vacuum to reduce attenuation and to maximise the signal that is measured.

To fully exploit the radiation beam properties for the required user, the beam is further modified by optical components located in the optics hutch [294].

The monochromator eliminates photons outside of the required energy range using crystal diffraction [297]. The crystals' orientation and geometries (there are often two or more crystals used) define the energy range that will be incident upon the sample. The mirrors focus the beam to reduce the spot size, and slits are used to collimate the beam [294].

The closer to the speed of light that the electron is travelling, the greater the collimation of the ejected photons of energy and the greater the brightness of the beam produced. Beam brightness is a measure of the spatial dependency of the beam's intensity, or the flux per unit solid angle per unit area [299]. Focusing the beam to increase flux affects radial divergence.

Brilliance describes the number of photons per second per square millimetre of area per square milliradian of solid angle within a bandwidth of 0.1% of any given energy (photons.sec^{-1}.0.1% bandwidth^{-1}.mm^{-2}.mrad^{-2}) [297]. High-brilliance beams have a higher useable flux density and can be suitable for low concentration samples that are resistant to beam damage, but could cause beam damage and degradation of biological or organic samples [297].

Brilliance is often the characteristic used to compare radiation sources. Figure 2.3 shows the development of synchrotron brightness since 1945, and compares it to the growth in leading-edge computing speed [300].

The recent upgrade at the European Synchrotron Radiation Facility (ESRF), Grenoble, from the third generation beamline to the world's first fourth generation synchrotron ESRF-EBS (Extremely Brilliant Source), has increased the brilliance by around 10^2 [301]. Faster sources, such as X-ray Free-Electron Lasers (XFEL), will enable ultra-fast probing of materials, resulting in *in situ* experiments with a higher time resolution [302,303]. XFEL combine particle accelerator and laser technology to produce a very brilliant source of radiation [304].

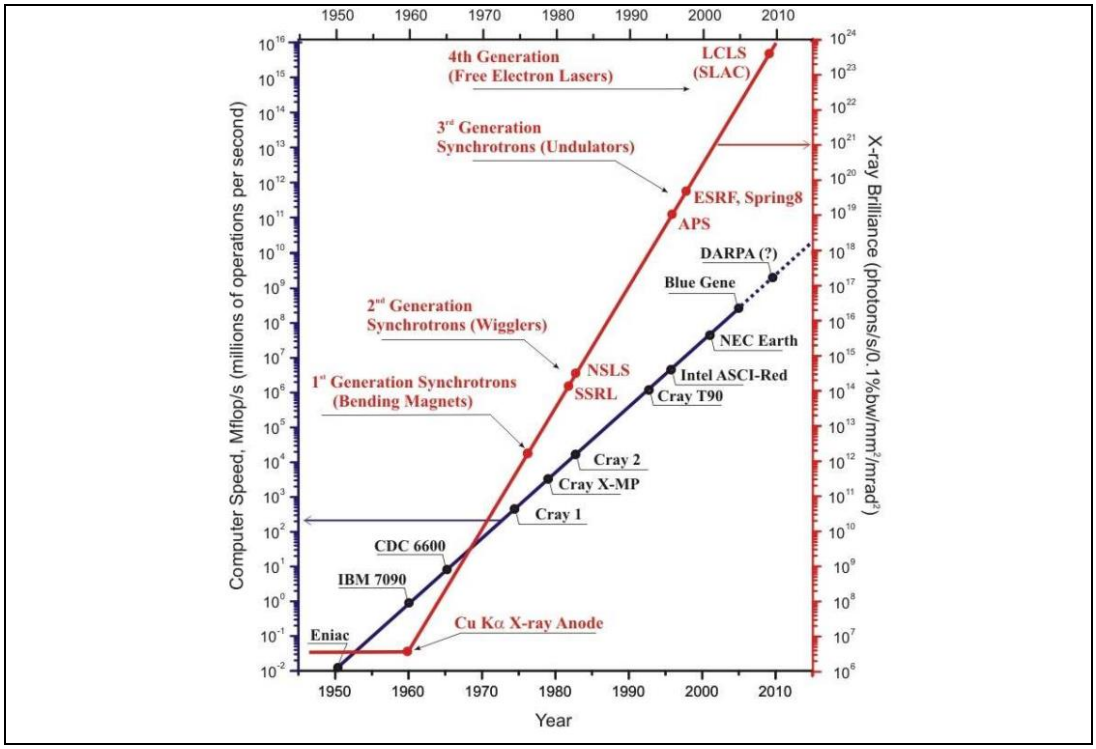


Figure 2.3 Synchrotron radiation development, based on X-ray source brilliance, from 1945–2015 (from ref. [300]**)**

2.4 Core-level X-ray spectroscopy

This section outlines the theory and application of the core-level X-ray spectroscopic techniques that have been used in this thesis. Core-level characterisation involves the excitation of core-level (K-edge, or 1s) electrons to create a core-hole. Depending on the technique, the energy or intensity of emitted or scattered photons or electrons provides information on the chemical state, local geometric structure or nature of the chemical bonds of the probed element [305].

Before the development of suitable synchrotron X-ray sources, inner-shell electron energy-loss spectroscopy (ISEELS), which could be undertaken in the laboratory, was key to the development of K-edge spectroscopy and dominated as a technique for element-specific analysis of electronic structure [207].

Figure 2.4 illustrates the core-level excitation and the resulting emissions that can be measured.

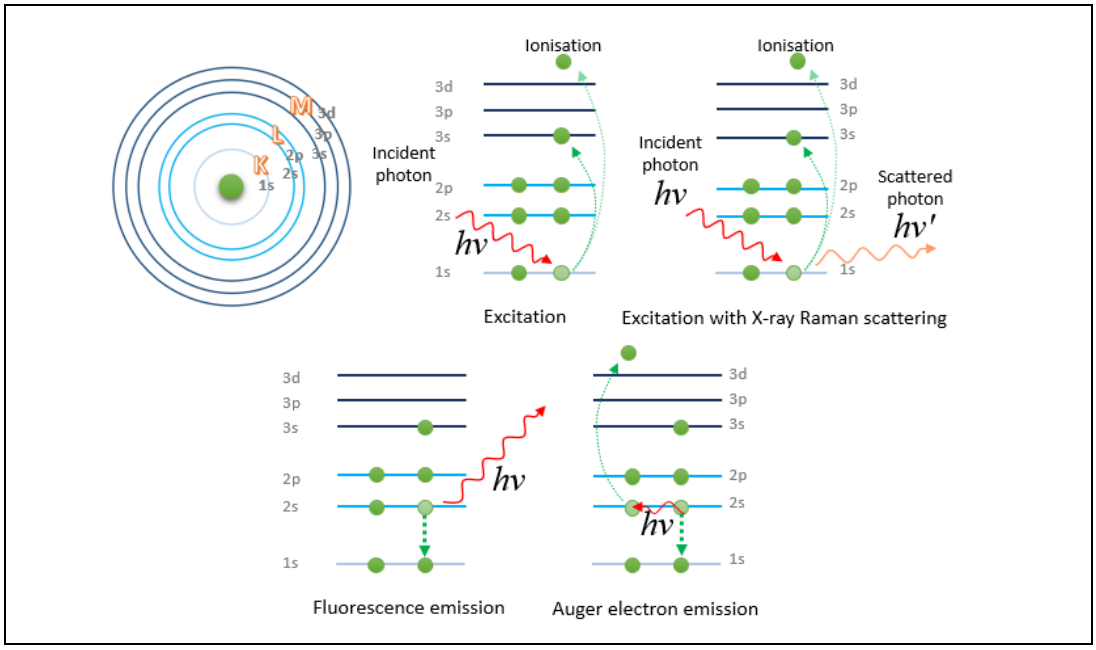


Figure 2.4 The photon and electron emission processes in core-level spectroscopy

2.4.1 X-ray absorption spectroscopy

X-ray Absorption Spectroscopy (XAS) experiments to acquire near-edge X-ray absorption fine structure (NEXAFS) spectra determine the local chemical and geometrical structure of matter [207]. In the NEXAFS technique, as the incident energy is scanned across the required range, absorbed incident photons excite core-level electrons to the lower unoccupied molecular orbitals (LUMO), described as π* or σ* transitions, or to continuum states, where the core-level electron is ejected from the probed atom (Figure 2.4) [208]. Secondary excitation events may also be measured, e.g. the emission of a fluorescence photon (radiative process) or Auger electron (non-radiative process) due to the release of energy when electrons drop to fill the core-hole left by the initial excitation (Figure 2.4) [208].

XAS is a resonant process, meaning the quantised photon is fully absorbed, and is performed in the soft X-ray range (< 1 keV). Soft X-ray studies often require samples to be in high or ultra-high vacuum (UHV) chambers, as these X-rays have a short path length [306,307]. With submicron absorption path lengths, XAS is a surface-sensitive analysis technique [306,307].

The monochromator used to tune the incident X-rays can introduce features in the spectrum across the 275–325 eV energy range, in the region of the C K-edge spectrum [207].

Liquid-jet set-ups have allowed XAS analysis of some liquid systems [306,308,309], although the sample temperature and pressure conditions are difficult to control and there are limitations on the sample environment. As such, these systems are less suitable for phase changes studies, such as where crystallisation is being induced.

2.4.2 X-ray Raman Scattering

Inelastic light scattering was first characterised in the 1920s, when it was seen that samples illuminated with visible, ultraviolet and infrared light emit secondary radiation that was characteristic of the atoms or molecules present in the bulk [310]. Whilst optical resonant Raman spectroscopy characterises molecular vibrational (phonon) and rotational modes and electronic excitations in the meV range [196], X-ray Raman scattering (XRS) is an emerging technique that uses hard X-rays (6–10 keV) to probe the unoccupied density of states of atomic species [208,311,312]. The excitation of core-level electrons occurs as a result of the energy-loss scattering event of the high-energy incoming photons [313].

XRS can measure both dipole and higher-order transitions, as a function of momentum transfer, q. In the low-q region (the dipole limit), where q is small compared to the inverse radial distribution of the ground-state electrons, XRS produces spectra that are proportional to the NEXAFS acquired from XAS [209,210,306,314–316]. An XRS spectrum that is dominated by dipole allowed transitions can be analysed using the same tools and techniques that are used for XAS [317].

With an X-ray penetration depth in the millimetre range, XRS is a bulk-sensitive technique [210,220]. XRS has negligible self-absorption effects, producing more reliable spectra than can be achieved using soft X-ray techniques. The real advantage of XRS lies in its power to probe materials in complex sample environments, including time-resolved processes (reactions or phase changes) [210,211,219]. As a consequence, XRS studies span many fields, including geosciences [318–322], water and ice [213,323,324], glasses [325], historical materials [219] and batteries [326–328].

With a low scattering cross-section, XRS experiments require high-energy, brilliant synchrotron radiation[8] to provide a high-flux, monochromatic, collimated X-ray beam of known polarisation for a reasonable signal-to-noise ratio [219]. A laboratory-based source would not provide high enough resolution or signal-to-noise ratio [329].

[8] ISEELS is equivalent to XRS at low-q values because the transition operators are similar at this momentum transfer range [506]. The limitations of measuring samples in the condensed state mean that whilst ISEELS measurements can support the interpretation and validation of XRS measurements, the electronic characterisation of samples in more complex sample environments is not possible with ISEELS [507].

2.4.2.1 Background theory

Most light scattering from a surface undergoes elastic Rayleigh scattering, where the light changes direction without change in energy or frequency [195]. Approximately 1/1000th of the scattered photons experience Raman scattering, where the final photon wavelength λ', is different to the incident wavelength λ, due to its interaction with the electrons in the scattering material (inelastic scattering) [195]. The resultant photon is scattered at a scattering angle 2θ, from the incident path. Inelastic scattering is 1/1000th of the intensity of the Rayleigh elastic scattering, and Raman is a small (< 5%) proportion of the inelastic (Compton and Raman) scattering (Figure 2.5). The Compton scattering is due to the interaction of the X-rays with the valence electrons.

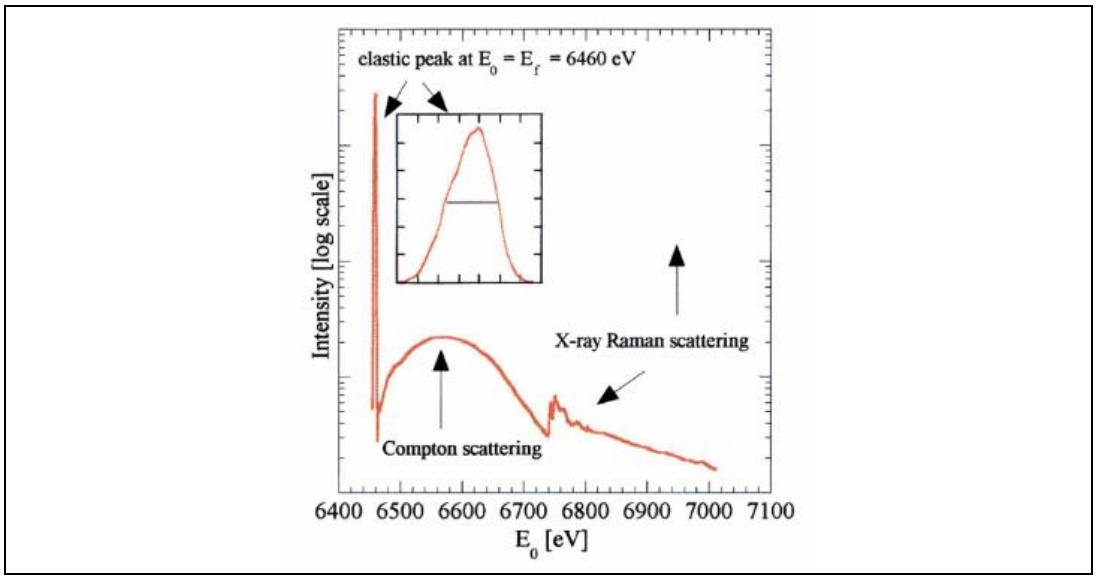


Figure 2.5 A comparison of the ratio of Rayleigh (elastic), Compton and Raman scattering from graphite. Taken from ref. [307]

Inelastic X-ray scattering (IXS) experiments examine the correlated motion of electrons of an excited multi-particle system by measuring both the energy and the change in momentum of scattered photons. There are broadly three non-resonant inelastic X-ray scattering (NIXS) techniques[9] for electronic excitation: Compton scattering, valence electron excitation and core-level excitation (XRS).

[9] The branches of NIXS are differentiated according to the Hamiltonian operator (the quantum mechanical operator related to the kinetic and potential energies of all particles in the system), which describes the interactions of the electromagnetic field of the incident X-rays with the electrons [210].

The near-edge fine structure is determined by scanning the beamline pre-monochromator across the required energy range. An incident photon with energy ω_1 and wave vector k_1, interacts with an electron, transferring a small proportion of its energy ω, during the excitation process causing a core-level absorption event [196,211]. The photon scatters inelastically with reduced energy ω_2 ($=\omega_1 - \omega$) and wave vector k_2 [196,211]. The intensity of the residual scattered photons is detected as a function of energy loss (ω) and momentum transfer (q) (Figure 2.6). The energy loss term here, ω, is comparable to the absorbed photon energy, and the momentum-dependence is equivalent to the absorbed photon polarisation vector, in soft X-ray XAS [210,219,330].

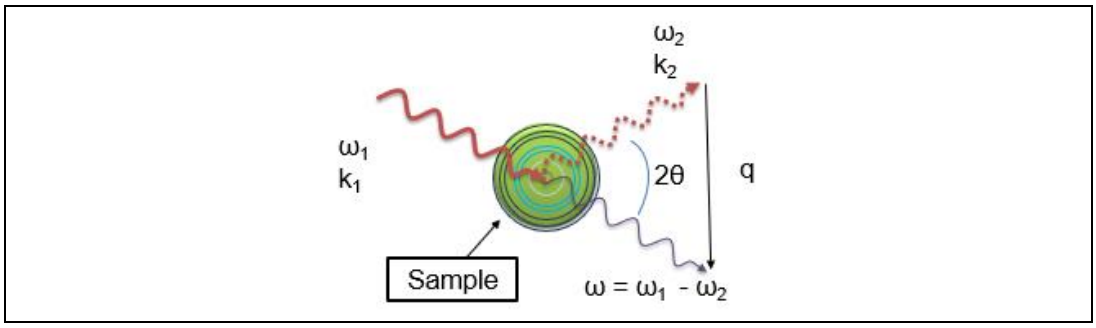


Figure 2.6 Energy loss during XRS process

In the XRS technique, the measured quantity is proportional to the double differential scattering cross-section (DSCS), which represents the probability that a particle of energy ω_1, will scatter into the solid angle element $[\Omega, \Omega + \delta\Omega]$ and into the energy range $[\omega_2, \omega_2 + \delta\omega_2]$, due to an inelastic scattering event [196]. When calculated according to lowest order perturbation theory, double DSCS is a product of the Thomson DSCS and the dynamic structure factor (Equation 10) [329].

$$\frac{\delta^2\sigma}{\delta\Omega\delta\varpi_2} = \left(\frac{\delta\sigma}{\delta\Omega}\right)_{Th} S(q,\varpi) \qquad \textbf{Equation 10}$$

The Thomson DSCS evaluates the radiation-electron interaction strength, and is described in terms of the energy and polarisation vectors of the incident and scattered photons [317]. The polarisation of the scattered light can also change, which is indicative of a surface molecule's orientation [207]. Due to the directionality of covalent bonds, which are common in molecules containing lighter elements, K-edge near-edge fine structure spectra of low-Z elements are strongly polarisation-dependent for molecules of fixed orientation [207].

The dynamic structure factor S(q,ω), is a function of the momentum and energy transfers and compares the initial and final states of a many-particle system, as described by Fermi's golden rule [211]. S(q,ω) can be expanded to

$$S(q,\varpi) = \sum_{i,f} p_i \left|\langle f \left| \sum_j \exp\left(iq.r_j\right) \right| i\rangle\right|^2 \delta(E_1 - E_2 + \varpi) \qquad \textbf{Equation 11}$$

The transition operator, *exp(iq.r_j)* can be expanded in a Taylor series (Equation 12).

$$exp\left(iq.r_j\right) = 1 + iq.r + \frac{1}{2}(iq.r)^2 + \ \ldots \qquad \textbf{Equation 12}$$

The first important term is *q.r*, which is the dipole operator [196]. There are no transitions as a result of the unity operator [196]. At low-q, dipole transitions dominate the series. For higher momentum transfer, the Taylor expansion increases to include more terms, reducing the importance of the dipole transitions. For certain materials with d- and f- electrons, such as insulators and electronic materials, probing high q multipole transitions (beyond the dipole limit) can have useful applications, but for the purposes of this study, the region of interest is limited to the dipole allowed low-q region, where the change in orbital angular momentum δl = ±1 [213].

Spectral resolution may range between 0.01 and 2 eV and is a convolution of the resolutions of the incident beam bandwidth, which is affected by the beam-side monochromator, and analyser crystal properties [209,312,331]. As an increase in energy or momentum transfer resolution is offset by a reduction in flux reaching the sample, the selection of the monochromator will depend on the sample type and the information sought in the experiment [210]. The acquired spectra will be a good approximation of the density of states in the atom's ground state from the perturbation of the core-electron excitation [196,317].

A full description of XRS theory can be found in detail in the literature [196,210,329].

2.4.2.2 XRS instrumentation

Many spectrometers are developed to measure a fixed scattered photon energy ω_2 whilst scanning across the energy range so that core-level excitation events at each incident energy can be detected [210,329]. The intensity of the scattered radiation, which has inelastically scattered from the sample through a constant scattering angle and a large solid-angle $\delta\Omega$, is collected and energy-analysed [210,329]. It is noted that increasing the collection efficiency of the spectrometer can impact on the resolution that can be achieved with a specific spectrometer configuration. Spectrometer design should optimise the energy resolution so that atomic species can be differentiated from the XRS spectra, but with a sufficiently high intensity for meaningful measurements [312].

An NIXS spectrometer, sometimes operating in a near-back-scattering configuration, has silicon crystal analysers to focus the scattered X-rays back onto the detector. Diced crystal

analysers improve the energy resolution and they perform significantly better than conventional spherically bent analysers, as the dicing reduces deformations due to lattice stress [312]. The analyser crystals are often housed in vacuum, or helium atmosphere to reduce stray scattering from the air [209,332]. The collimation of the scattered radiation as it enters the detector housing further increases the signal-to-noise ratio [209]. An analyser module at the ID20 beamline at the ESRF, containing 12 analyser crystals, is shown in Figure 2.7 (taken from ref. [209]). The precise details of the XRS beamline instrumentation available at different synchrotrons can be found in the literature [209,211,312,332–334].

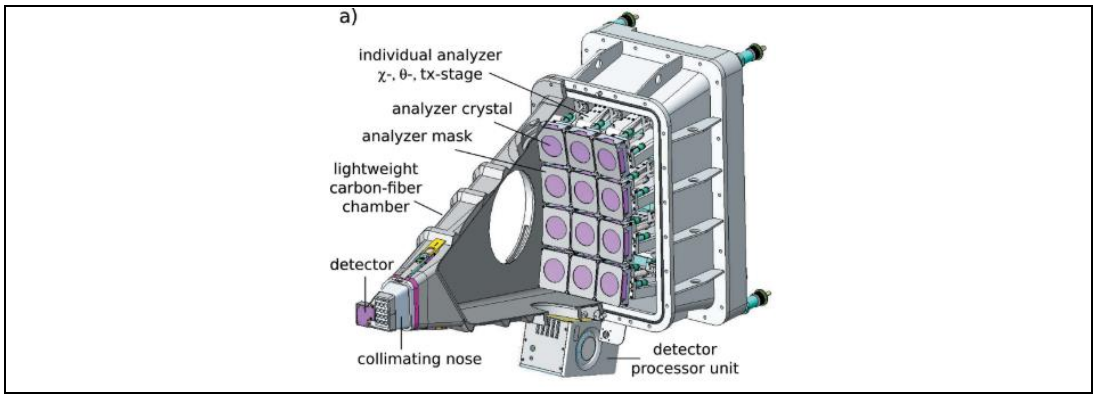


Figure 2.7 An analyser module housing 12 Si(nn0) analyser crystals on a 1 m Rowland circle: one of 6 modules at ID20, ESRF: from ref. [209]

The time to acquire an XRS scan is dependent on factors such as the beam flux, the sample environment, and type and number of detectors. The sample will be irradiated for an extended period, from minutes to potentially hours if multiple scans are taken for several edge spectra. During an experimental session, repeated spectra measurements over time identify chemical changes that may occur from beam damage, which can be reduced by moving the beam between scans or using automatic repositioning of the beam spot [335,336].

To measure under complex sample environments, Diamond Anvil Cells (DAC) have been developed to achieve a high-pressure environment for X-ray spectroscopy and diffraction studies of water and ice, geological samples and glasses [337–341], and application-specific flow cells have been developed for the XRS analysis of liquids [67,221,327].

2.4.2.3 European Synchrotron Radiation Facility (ESRF), ID-20

The XRS measurements of aqueous solutions of Imid and ImidHCl (presented in Chapter 3) were collected at the inelastic scattering beamline ID-20 at ESRF, Grenoble, France [183]. ID-20 is an undulator beamline, operated with four 26 mm-period undulators. The beam is collimated by a white-beam mirror before passing through a liquid-nitrogen cooled double-

crystal Si(111) pre-monochromator and an Si(311) channel-cut post-monochromator to operate in a high-energy resolution mode. The overall energy resolution, comprising the instrumental resolution of the monochromator and the analyser crystal, was found to be ~0.7 eV during these experiments, based on the full width half maximum (FWHM) of the elastic peak. The beam was focused to a spot size of ~27 µm (horizontal) x 31 µm (vertical) using a mirror system in Kirkpatrick–Baez geometry.

Six Maxipix detector modules, three in each of the horizontal and vertical scattering planes, collect the scattered photons at q values between ~1.5 $Å^{-1}$ and 9.5 $Å^{-1}$ [209]. Each detector module contains 12 spherically-bent Si(660) analyser crystals. Details on the dimensions and arrangement of the analyser crystals, and the resulting benefit to the energy resolution of this arrangement, is described elsewhere [209]. The elastic energy was set to 9.68 keV. The X-ray Raman spectrum was obtained by scanning the beamline monochromator energy at a fixed analyser setting. The measured energy difference (incident beamline monochromator energy minus measured analyser energy) corresponds to the energy loss. To obtain NEXAFS information, the XRS spectra were collected in the low-q regime, which is dominated by dipole allowed transitions [315]. For the work described in Chapter 3, measurements were performed in forward-scattering (low-q) geometry, collected by the VD detector module positioned at an angle of 42° and q = 3.6 $Å^{-1}$ for the study of dipole allowed transitions [312].

2.4.2.3.1 Sample flow cell

The *in situ* jacketed flow crystalliser (Figure 2.8) used on the XRS beamline was designed to replicate the conditions of experimental crystallisers in the laboratory. The crystalliser was commissioned for *in situ* crystallisation experiments, as described in ref. [67].

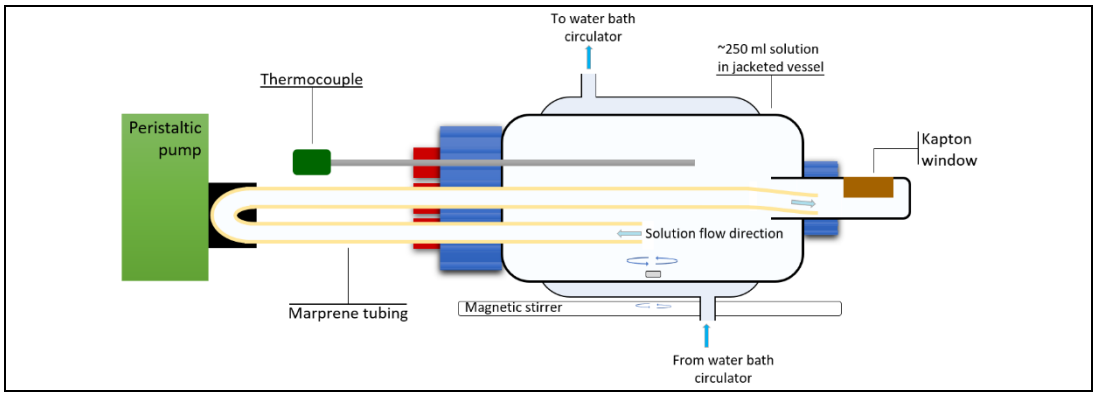


Figure 2.8 Schematic representation of the flow cell used in the *in situ* crystallisation study

The flow crystalliser design reduces the potential for radiation damage by circulating the solution exposed to the beam. The beam was aligned to pass through a 75 µm thick Kapton (polyimide) window, which is transparent to X-rays, located in the crystalliser neck. A

thermocouple measures the temperature of the solution, which is temperature-controlled via the crystalliser jacket using a Julabo chiller, although the data detailed in Chapter 3 were acquired at room temperature (~23°C).

2.4.2.3.2 Data processing

The background needs to be subtracted from the high-resolution scans. The lighter-element K-edges are generally located where there is a large Compton scattering background. This effect is shown clearly in Figure 2.9, where increasing magnification of the long scan reveals further information on the elastic peak (ω_2, monochromator energy) (Figure 2.9a) and Compton peaks (Figure 2.9b) and the carbon, nitrogen and oxygen K-edges (Figure 2.9c).

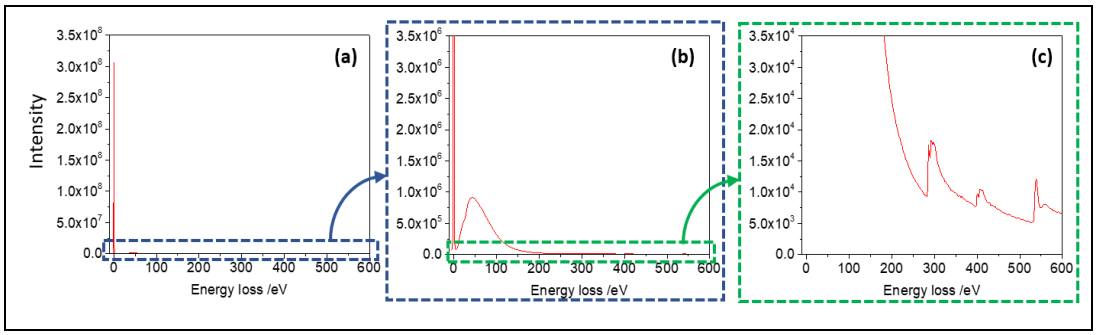


Figure 2.9 Elastic scan across the energy range of interest

(a) with a maximum intensity value of $3.5x10^8$, showing the intensity of the elastic peak, (b) with a maximum intensity value of $3.5x10^6$, showing the Compton scattering and (c) with a maximum intensity value of $3.5x10^4$, showing the carbon, nitrogen and oxygen K-edges.

The user selects specific pixels on the 2D detectors (regions of interest, RoI) for evaluation, excluding significant background interference or measured intensity that is out of the required momentum transfer range [211]. The background subtraction is performed using the beamline-specific processing code XRStools, the Python-based open source software developed at ESRF [209]. Features in the background spectra should be fully removed, which is particularly challenging if the signal-to-background ratio is small [207]. Using a suitable spectrum normalisation procedure is essential for reliable and comparable spectra [207].

Where there is more than one atom type of a probed species, the spectra can be decomposed using methods such as multi-component fitting, or non-negative matrix factorisation [222,342]. The near-edge spectra peaks can be assigned using binding energy measurements by XPS, and X-ray absorption (XA) spectra from TDDFT calculations, and fitted using peak fitting software.

2.4.3 X-ray photoelectron spectroscopy (XPS)

2.4.3.1 Overview of XPS

X-ray photoelectron spectroscopy (XPS) is a widely-used surface analytical technique that provides chemical and elemental information about a material, to a depth of ~10 nm [343,344]. The technique, also known as Electron Spectroscopy for Chemical Analysis (ESCA), is based on measuring the kinetic energy E_K, of photoelectrons that are emitted from the surface of a material [345].

The schematic of an X-ray photoelectron spectrometer is shown in Figure 2.10. XPS analysis involves the irradiation of a surface with monochromatic X-ray photons, which if sufficiently high in energy, may be absorbed by electrons causing their ejection from the surface of the material [345].

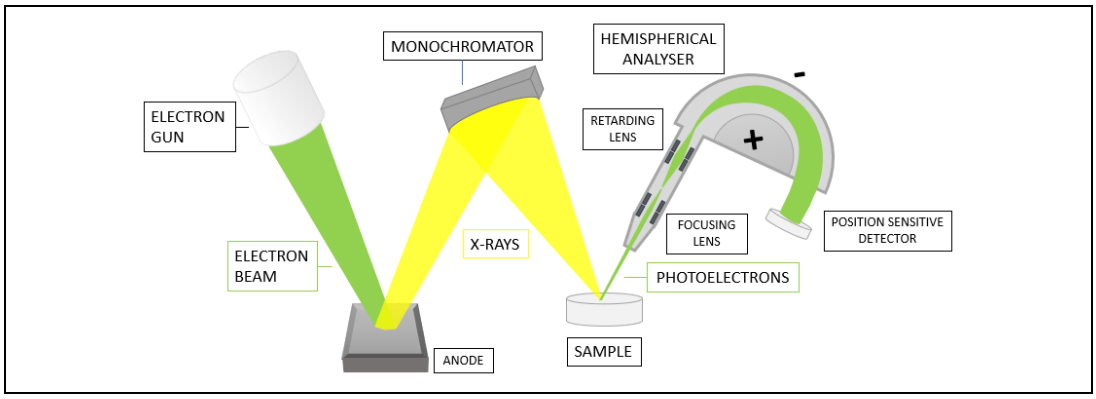


Figure 2.10 Schematic representation of XPS

The photon energy E_p, of the incident X-ray must exceed the electron binding energy E_B to bring the electron to vacuum and provide sufficient additional E_K to move the electron to the surface of the sample and into the analysis chamber. As energy is conserved, E_P can be related to the E_B and E_K by

$$E_P = h\nu = E_B + E_K \qquad \textbf{Equation 13}$$

where h = Planck's constant,[10] and ν = the frequency of the X-ray (Hz). The E_B of an electron will vary depending on the probed atom type and speciation, as different chemical environments will change the electron density and therefore the E_B [346].

A small proportion of the ejected photoelectrons will enter the analyser nozzle of the spectrometer [346]. The number of electrons emitted per E_K interval (the pass energy), over

[10] h = 6.62606957 × 10^{-34} m^2kgs^{-1}

a specified energy range, is measured in the hemispherical analyser [345]. Reducing the pass energy increases the resolution of the scan, but this is offset by a lower count rate [346].

Traditionally, samples are measured *in vacuo* at pressures in the region of 10^{-8} mbar [347]. This UHV environment gives the photoelectrons a satisfactory free mean path when travelling from the sample to the detector by minimising their interactions with other particles.

Charging of the sample can occur during measurement, as the sample is losing photoelectrons causing a build-up of positive charge. Charge compensation in UHV-XPS is managed using flood-guns, a source of low-energy electrons [346]. Whilst spectra of non-conductive materials can be measured with the use of charge compensation, it does not overcome the problem of energy calibration, as the surface potential of the sample is still unknown [346].

Energy calibration is required to determine the actual E_B from the measured E_B of the photoelectron, and various techniques are summarised in the literature [346]. Each method requires identification of known peak, and adjustment of the energy scale accordingly. For samples that are poorly conducting (non-metals), that do not have a good contact with the spectrometer or are not sufficiently charge-compensated, the Fermi Level cannot be used as a reference level [346]. Often, the C 1s adventitious carbon (C_{adv}) layer is used for energy calibration, as it will be present on any sample that is exposed to air, although there is debate in the literature about the suitability of this approach [346].

UHV-XPS is unsuitable for some materials, such as organic compounds and liquids, which are too volatile, or for biological samples that may be altered when frozen for UHV-XPS analysis. Near-ambient pressure XPS (NAP-XPS) allows sample measurement at higher pressures and in environments more representative of their real-world conditions [347].

2.4.3.2 Near-ambient pressure XPS (NAP-XPS)

Near-ambient pressure XPS (NAP-XPS) enables samples to be measured at pressures in the order of a few mbar [347]. At this low sample chamber pressure, there is a substantial pressure differential between the X-ray gun, the analysis chamber and the analyser [347]. To overcome this difference, the EnviroESCA used in the work described here operates a differential pumping system, with four stages from the analysis chamber to the Phoibos 150 NAP hemispherical analyser.

With the gaseous environment in the NAP analysis chamber, the electrons have a short free mean path, which means that the small-aperture nozzle through which photoelectrons are

removed from the chamber needs to be close to the sample being measured. In the case of the EnviroESCA, pressure in this analysis chamber is maintained using a roughing pump for evacuation and gas dosing [347]. The three subsequent pumping stages use UHV pumps to bring the pressure down, each separated by small apertures with electron focussing at each stage to maximise the count rate [347].

Samples can be temperature controlled via the sample stage using the resistance heating plate or the Peltier cooling plate. The samples can also be dosed with gas. Organic samples are insulating materials and are subject to significant charging. In the EnviroESCA, charge compensation is provided by electron-ion pair formation of gas-phase molecules present in the chamber during sample analysis, induced by absorption of the photo- and Auger electrons emitted from the sample, alongside direct ionisation by X-ray absorption [348].

Energy calibration was performed based on the allocation of a C_{adv} peak at 285 eV. Whilst the difficulties and potential inaccuracies using this method are acknowledged [346], and alternatives were explored, it is considered to be the most appropriate for the organic samples measured under NAP-XPS. It was likely that there would be greater contamination when measuring samples at NAP compared with samples that are measured at UHV pressures.

2.4.3.3 SPECS EnviroESCA NAP-XPS

The EnviroESCA used to collect X-ray photoelectron spectra of Imid and ImidHCl is equipped with a monochromatic Al Kα X-ray source (1486.71 eV). The beam (300 μm diameter) is separated from the analysis chamber by a silicon nitride window. The hemispherical Phoibos NAP 150 analyser operates in small-area mode with a 1D delay-line detector.

In the work described here, Imid and ImidHCl samples were measured when dry and in aqueous solution. For the solid sample analysis, Imid was ground to a powder before being pressed into an aluminium differential scanning calorimetry pan or a moulded foil pan and mounted on an aluminium scanning electron microscopy stub. ImidHCl is hygroscopic and grinding was ineffective, so the solid material was just pressed into the pan.

Once the dry samples were in the analysis chamber, 3 ml/min argon was introduced to maintain pressure. A humid environment was required in the analysis chamber for the ImidHCl solution sample, as water vapour adsorbed onto the hygroscopic ImidHCl to form an aqueous solution. To dose the water vapour, an open glass bottle of deionised water was placed in the sample chamber (Figure 2.11a). The Imid solution was put in a Petri dish, which was placed directly in the sample chamber.

To further minimise solvent evaporation, the use of the Peltier cooler was trialled. A thermocouple was placed between the pan and the Peltier cooling plate and when transferred to the analysis chamber, the current was initially set at 1 A and then adjusted to maintain the temperature at 278–283 K. There are clips on the Peltier cooler to hold down the sample, but it was found that the clips could compress the sides of the pan. Therefore, a foil cover over the pan, which could be held down by the clips, was used (Figure 2.11b).

Figure 2.11 XPS data collection solution sample chamber setup

a) Sample chamber with deionised water b) sample pan secured with foil

When the analysis work was started, an Imid solution sample was measured in the chamber through a range of pressures (1, 5, 7, 11 and 15 mbar) to gauge the signal-to-noise ratio and the stability of the sample. 11 mbar allowed a measurement with an acceptable signal-to-noise ratio, and at this pressure, water vapour in the chamber condensed and adsorbed onto the sample, and minimised evaporation. 11 mbar had also been used in the analysis of hydrated bacteria in the literature [343].

Samples containing water (or other solvents) must be carefully handled. If for example a solid is not completely dry before being transferred to the analysis chamber for measurement, or if a solution evaporates, the sample could move or 'erupt' due to degassing and material could enter the nozzle due to the small working separation between the sample and nozzle. The pressure was reduced slowly to allow residual water to leave the solid sample. The pressure in the sample chamber was reduced by pumping down to 50 mbar (fast), then from 50 to 25 mbar (medium), where the pressure was held until it was seen that water degassing was complete and the sample was stable, after which the sample was transferred to the analysis chamber. The pressure was then reduced to 11 mbar.

Measuring the aqueous solutions was challenging, as the working separation of the nozzle and solution sample varied. ImidHCl formed a solution when the solid sample was placed in the humid environment and did not require the Peltier cooler to help to maintain the sample in solution. Refinement of measuring solutions in the NAP-XPS showed that was preferable to measure aqueous Imid in a Petri dish without additional water in a bottle or the Peltier

cooler, as the water in the solution provided sufficient humidity. It was decided to collect spectra of the solid samples in a dry (argon) environment and solutions in a humid environment at 11 mbar at ambient temperatures.

2.5 X-ray diffraction

2.5.1 Background and basic theory

X-ray diffraction (XRD) methods have been used to extract structural information from materials for over 100 years [349]. The first X-ray diffraction photograph of a periodic crystal was acquired in 1912 and confirmed that the wavelength of X-rays is in the order of interatomic distances [294]. The analysis of what is today known as Bragg reflections solved the crystal structure in 1913 [350,351]. Crystalline material structures were derived at first, but more complex structures such as DNA (1952) and proteins were also established using crystallographic techniques [294,349].

XRD describes the scattering of the coherent incident radiation when it interacts with the electron density of the atoms in a sample, and the detection of the scattered quanta irrespective of any energy exchange [225,351].

If the parallel lattice planes in a crystalline material are separated by a distance d_{hkl}, (where *hkl* are the Miller indices of the set of planes) and d_{hkl} is similar in magnitude to the wavelength of the incident monochromatic X-rays λ, the lattice planes act as a diffraction grating (Figure 2.12) [201]. Diffracted waves in the same spatial region superimpose, resulting in constructive interference of in phase waves to produce a diffraction pattern of sharp diffraction spots, or Bragg peaks. Waves that are out of phase will cancel out through destructive interference. When the X-rays superimpose constructively, the difference in path length is nλ, where n is an integer (distance aO' + O'b = nλ in Figure 2.12) [351].

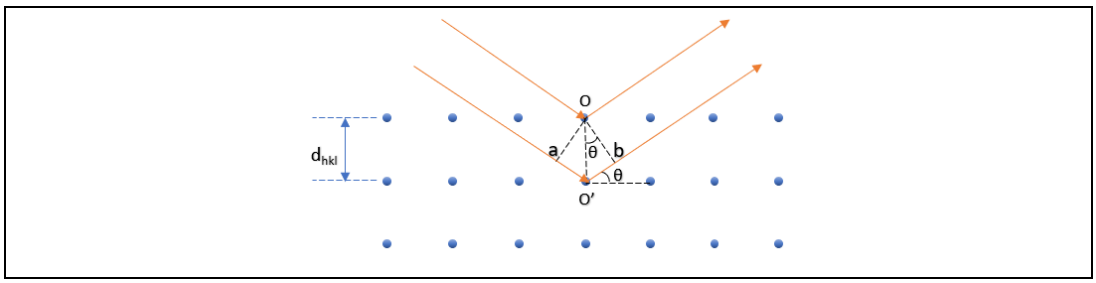


Figure 2.12 Schematic representation of Bragg's law for X-ray diffraction

The path difference nλ, is dependent on the Bragg angle θ, and the separation d, which are related by Bragg's law (Equation 14):

$$n\lambda = 2d \sin\theta \qquad \textbf{Equation 14}$$

The diffraction pattern in reciprocal-space (Q-space) is the Fourier transform of the atomic spacing in real-space (r-space) [201,225].

With no long-range periodic structure, non-crystalline material structures cannot be analysed from the Bragg peaks. X-ray methods to analyse the structure of liquids using X-ray total scattering were developed as early as 1916 [352]. Initially, X-ray total scattering diffraction patterns were used for the classification of liquid types until calculation methods, such as Debye and Menke's Fourier transform method in 1930 (ref. [353]), were developed to interpret the diffraction patterns [352].

Distortions in the average crystalline structure can produce desirable real material properties [354]. Traditional crystallography cannot determine the atomic-scale structure of complex materials, but differences in the local structure can be identified through pair distribution function (PDF) analysis [201,354]. The PDF method determines local and thermal disorder in crystalline materials and, importantly, reveals short- and medium-range order in disordered materials, including amorphous solids or liquids, or mixtures of these phases, e.g. solutes in solution [349].

The PDF is the weighted probability of finding atoms a certain distance apart, where the peak positions correspond to bond lengths and the integrated area relates to the coordination number, so PDF analysis gives a more complete picture of a material's structure across the length scales [355,356]. The PDF is derived from the Fourier transform of the experimentally-measured reciprocal-space scattering function [357].

Early neutron diffraction studies replicated X-ray diffraction patterns, e.g. the study of liquid mercury in 1954, although the resolution was not as good as the X-ray method [358]. The development of spallation neutron sources, and the use of neutron diffraction with isotopic substitution (NDIS) to obtain multiple measurements, renewed interest in the total scattering technique for determining the local structure of amorphous solids and liquids, as well as the local disorder in crystalline materials, which was not possible with Bragg peak analysis alone [356].

In the NDIS technique, the neutron diffraction of isotopically-enriched, but chemically-equivalent, samples are measured. If the neutron scattering lengths of the isotopes are sufficiently different, the partial structure factors can be extracted and Fourier transformed to give the real-space partial distribution functions [225,355]. NDIS is a complementary technique to X-ray total scattering, with the methods being combined to determine the

partial structure factors in some structure analysis studies [225,356,359,360]. A series of NDIS diffraction experiments on a system with multiple chemical species with varying scattering power can accurately determine the partial pair-distribution functions [225]. NDIS has the benefit of being sensitive to the position of hydrogen atoms, whereas X-ray total scattering is not [223] However, whilst NDIS can distinguish between hydrogen and deuterium, there is little contrast between some elements commonly found in organic samples, e.g. oxygen and carbon. The timescale for NDIS data collection can be in the order of hours, making NDIS unsuitable for the measurement of *in situ* phase change processes. Furthermore, deuteration of glycerine for NDIS analysis of crystallisation resulted in the formation of the γ polymorph, rather than the expected α polymorph, indicating that the technique could impact on the nucleation pathway being observed [361].

X-ray total scattering simultaneously captures the Bragg scattering and diffuse scattering, the scattering intensity between and underneath the Bragg peaks [357]. The information on two-body atomic correlations and thus local structure is in the diffuse scattering, and is weak in comparison to the Bragg scattering [225,349].

The X-ray diffraction technique has developed with improvements in the measurement of the diffracted X-ray intensity, the generation of high energy X-rays and improved computing power, which enabled the introduction of computational techniques for the analysis of complex diffraction data (Section 2.9.2) [351]. Synchrotron sources generate much higher-energy photons with shorter wavelengths than is possible with laboratory-scale XPDF instruments thus producing higher-resolution data. At the Diamond Light Source (DLS) I15-1 XPDF beamline, the X-ray energy E = 76.6 keV ≈ λ = 0.163 Å, which is in the order of atomic spacing of condensed matter. However, X-ray total scattering is not sensitive to hydrogen position, and there is little contrast between atoms that are adjacent on the periodic table, which should be considered during experiment design.

X-ray total scattering measurements can be acquired under process conditions thus is a suitable technique to probe the structure of a solution during cooling crystallisation. XPDF acquisition time is relatively short compared to NDIS and other spectroscopic techniques. Scanning times can be in the order of seconds for the acquisition of diffraction patterns with good counting statistics, depending on the material [362]. High-flux synchrotron beams further reduce measurement times, allowing for the *in situ* measurement of phase change processes. The development of fourth generation beamlines with more brilliant, higher flux beams, such as the recently upgraded ESRF-EBS in Grenoble and the proposed Diamond-II upgrade, will yield higher resolution data and quicker data acquisition times.

2.5.2 Single crystal XRD

SCXRD analysis allows the three dimensional atomic geometry of the crystal to be solved from the electron distribution determined from the diffraction pattern [363]. The position of H atoms can also be estimated based on the electron density distribution.

SCXRD was carried out on a sample of the recrystallised GuHCl, and a sample of the synthesised ImidHCl. Measurements were carried out at 130 K on an Agilent SuperNova diffractometer (Department of Chemistry, University of Leeds). The diffractometer was equipped with an Atlas CCD detector and connected to an Oxford Cryostream low-temperature device, using mirror monochromated Cu K_α radiation (λ = 1.54184 Å) from a Microfocus X-ray source. The structure was solved by intrinsic phasing using SHELXT [364] and refined by a full matrix least squares technique based on F^2 SHELXL2014 [364]. The structures were solved by Dr. A. R. Pallipurath (School of Chemical and Process Engineering) and Dr. Chris Pask (Department of Chemistry). The results are in Appendix A.

2.5.3 Powder X-ray diffraction

Powder X-ray diffraction (PXRD) is a long-established technique for the phase identification of a crystalline material by providing the average long-range structure and unit cell dimensions [365]. Rietveld refinement of Bragg reflections in XRD patterns overcomes any deviation from the ideal crystal structure due to dynamic thermal vibrations [366].

The PANalytical X'Pert PRO (X-ray diffraction facility, SCAPE, University of Leeds) was used to identify the crystal structure of the *as received* and recrystallised GuHCl to identify potential differences in the polymorphic form. Room-temperature XRD scans in the range of $5° < 2\theta < 40°$ were recorded over 12 min with step size 0.016°. The PXRD patterns are shown in Appendix A.

2.5.4 X-ray total scattering

Atomic pair distribution function (PDF) experiments, also known as X-ray total scattering experiments, can be performed with an Mo or Ag laboratory source or, preferably, a higher-energy, more brilliant synchrotron beam [291,367]. At the beamline, the coherent X-ray beam (via a single-crystal monochromator) that is incident on the sample scatters, with scattering angle 2θ, to a 2D detector (Figure 2.13).

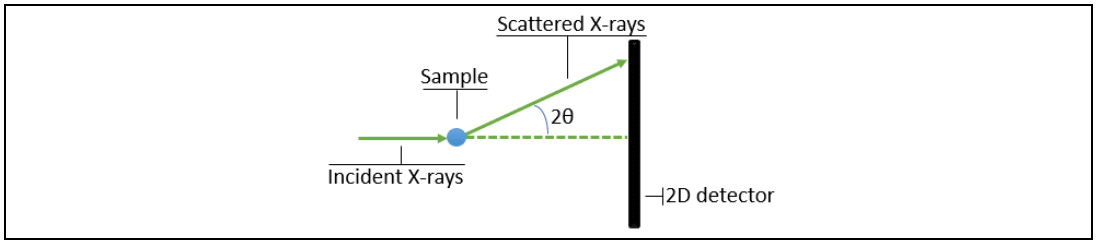


Figure 2.13 X-ray total scattering sample and detector arrangement

The scattering experiment is assumed to be an elastic process, meaning there is negligible change in the incident radiation energy E [201]. E is inversely proportional to the wavelength and related to Q, the wave vector change in reciprocal space ($Å^{-1}$), by Equation 15, where θ is half the scattering angle 2θ.

$$Q = \frac{4\pi \sin(\theta)}{\lambda} = \frac{4\pi \sin(\theta) \, . \, E}{hc} \qquad \textbf{Equation 15}$$

X-rays interact with and scatter from the spatially-extended electron density of an atom, and it roughly describes the location of the atom centre [225]. The rate at which X-rays scatter is described by the scattering cross-section σ [11]. The incident flux Φ, of the incident beam is the number of incident quanta per second per unit cross-sectional area. The XPDF technique measures the X-ray differential scattering cross-section (DSCS) for the diffraction the beam, as defined by Equation 16 [225] where Ω represents the unit solid angle.

$$\frac{d\sigma}{d\Omega} = \frac{no.\,quanta\ scattered\ per\ second\ towards\ detector\ into\ d\Omega}{\Phi d\Omega} \qquad \textbf{Equation 16}$$

The raw data collected on the 2D detector is corrected for the internal dark current by the generic data acquisition software at the beamline, before being integrated to produce a 1D scattering pattern, known as the structure factor S(Q) (Figure 2.14).

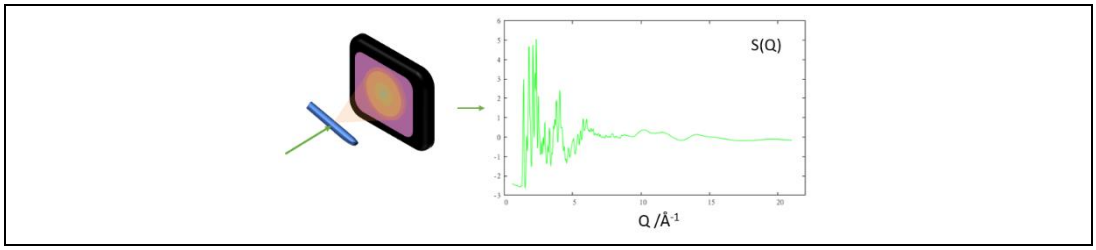


Figure 2.14 Schematic of X-ray total scattering 2D pattern data collection, which is integrated to acquire the structure factor for GuHCl

[11] (No. electrons excited per unit time) / (No. incident photons per unit time per unit area) [207]

S(Q), when corrected to account for backgrounds, absorption, multiple scattering and Compton scattering, gives the reduced structure factor F(Q) (=Q[S(Q)-1]) [368]. F(Q) relates to the Q-dependent X-ray scattering factors $f_i(Q)$ by Equation 17 [369]

$$F(Q) = \left[\frac{1}{N}\frac{d\sigma}{d\Omega} - \sum_{i=1}^{n} c_i f_i(Q)^2\right] / \left[\sum_{i=1}^{n} c_i f_i(Q)\right]^2 \qquad \textbf{Equation 17}$$

where the material contains *N* atoms of *n* chemical species of atomic fraction *c*.

The DSCS is essentially proportional to the structure factor in a monatomic system, or where there are multiple atom types, to a weighted sum of partial structure factors that take the varying atomic form factors into account [225]. The Q-dependence of X-ray scattering form factors means that the evaluation of X-ray pair correlation functions is non-trivial, as a 'sharpening factor' must be applied to account for the form factor reduction at high-Q [356,369].

For systems with multiple atom types, an approximation of the weighted average form factor is applied to handle the otherwise complex Q-dependence [245]. The DSCS comprises a total interference function describing the interference of waves between different atoms ('distinct' term) and a second ('self') term describing the interference due to atoms of the same chemical species, averaged across the whole sample (Equation 18) [225,369].

$$\frac{1}{N}\left[\frac{d\sigma}{d\Omega}(Q)\right] = F(Q) + \sum_{\alpha}^{n} c_\alpha \overline{b_\alpha^2} \qquad \textbf{Equation 18}$$

A highly-accurate F(Q) requiring careful data reduction is needed for the study of liquids [291].

A number of publications have tried to define the formalisms employed in neutron and X-ray total scattering analysis, as there is some variation across different research communities [291,349,356,369–371]. To summarise the terminology used here, G(r), the real-space differential reduced pair distribution function, is defined in ref. [371]:

$$G(r) = 4\pi r \rho_0 [g(r) - 1] \qquad \textbf{Equation 19}$$

where ρ_0 is the average atomic number density and g(r) (= $\rho(r)/\rho_0$) is the atomic pair distribution function [370] and $\rho(r)$ is the atomic pair density. As r⟶0 the slope of the function G(r) is $-4\pi r\rho_0$. G(r) tends to 0 with increasing r. This parameter is particularly

useful for the study of liquids and solutions, as it helps to emphasise medium-range interactions.

The Fourier transform of F(Q) yields G(r) (the total scattering PDF) by Equation 20 [245,367]

$$G(r) = \frac{2}{\pi}\int_{Q_{min}}^{Q_{max}} F(Q)\sin(Qr)\,dQ \qquad \textbf{Equation 20}$$

where the wave vector range Q_{max}–Q_{min} is limited by the experimental apparatus.

As partial structure factors $S_{\alpha\beta}(Q)$, cannot be accessed directly using X-ray total scattering, the structure models of the probed systems generate their real-space equivalents, which, when Fourier transformed and combined, make the total structure factor S(Q) [245]. In this way, the structure generated by the molecular model may be compared to the experimental data.

Although PDF peak resolution is directly proportional to Q_{max}, achieved by collecting over a wide scattering angle, a lower Q_{max} may be applied to remove high-Q data with a low signal-to-noise ratio [365,367]. Whilst the low-Q region represents long-range order, the diffuse scattering between and underneath the Bragg peaks gives information on the short and medium-range atomic interactions (Figure 2.15) [201].

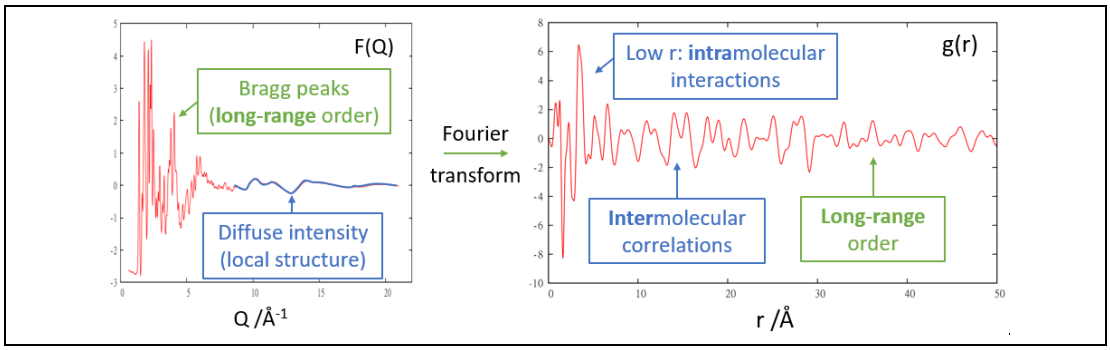


Figure 2.15 Features of the F(Q) and g(r) (derived from the Fourier transform of F(Q)) for solid GuHCl

The characteristics of the peaks provide a structural fingerprint of the material being analysed, with peak positions giving a representation of the material's average atomic distribution, and the area under the peaks indicating the coordination number at a given r-spacing [201]. Partial pair distribution functions are extracted from XPDF patterns using computational methods (detailed in Section 2.9.2).

The coordination of one atom around another in a material can be quantified by measuring the number of atoms that can be found in a volume around the central atom. The minimum

and maximum radii r_{min} and r_{max}, of the probed volume are related by $r_{max} = r_{min} + \Delta r$. This process can be repeated, by incrementally changing Δr, to determine the frequency distribution of the atomic pair.

To determine the atomic pair distribution function $g_{\alpha\beta}(r)$, the measured density distribution is normalised to take into account the expected (ideal) distribution if the atom types were equally distributed throughout the material [370]:

$$g_{\alpha\beta}(r) = \frac{measured\ density\ distribution\ \rho(r)}{ideal\ distribution\ \rho} \qquad \textbf{Equation 21}$$

In the low-r region, where there is repulsion between atoms, $g_{\alpha\beta}(r) = 0$. $g_{\alpha\beta}(r)$ tends to 1 with increasing r, as the measured structure approaches the ideal structure. The partial distribution function describes the correlation between a specific pair of atom types, giving understanding of the material structure [372]. The coordination number $n_{\alpha\beta}$, (atom type β around atom type α), due to steric or intermolecular interaction effects, is calculated by integrating under the partial distribution function [193]:

$$n_{\alpha\beta} = 4\pi\rho c_{\beta} \int_{r_{min}}^{r_{max}} r^2 g_{\alpha\beta}(r) dr \qquad \textbf{Equation 22}$$

where ρ is the atomic number density of the probed system and c_{β} is the fraction of atoms that are type β. The partial distribution function does not give any directional information on the position of the atomic pairs [201].

2.5.4.1 XPDF beamline at Diamond Light Source (DLS), I15-1

The XPDF experiments detailed in this thesis were conducted at beamline I15-1 at DLS (Figure 2.16) [373].

2.5.4.1.1 Beamline instrumentation

The 76.6 keV beam has a spot size of ~700 µm (horizontal) x 20 µm (vertical) via a horizontally focusing Si(311) Bent-Laue monochromator and a vertically focusing multilayer mirror [373]. The beam flux is tuneable to suit the probed material. The two detectors, for diffuse and high-resolution Bragg scattering data, are independently positioned to optimise the sample-to-detector distance. For the diffuse scattering needed for XPDF analysis, a large area Perkin Elmer detector (PE XRD 4343 CT detector) was used (active area of 432 × 432 mm^2, 150 µm pixel size).

The positioning of the detector and the high-energy beam gives access to $Q_{max} \approx 40$ $Å^{-1}$. However, the inherently diffuse scattering from materials with low structural coherence

does not require data collection to a high Q_{max} [201]. Organic materials are weakly scattering and have a poor signal due to thermal atomic motion [367]. It is therefore more difficult to acquire good counting statistics at a high Q_{max} for such samples, so using a Q_{max} that is lower than the experimental maximum when processing the data reduces unfavourable signal-to-noise ratio [367]. However, as organic materials tend to exhibit lower solid-state symmetry, it is possible to acquire good information on the changing bond lengths at a lower Q_{max}. The data for the organic samples considered here were processed using Q_{max} = 21 $Å^{-1}$.

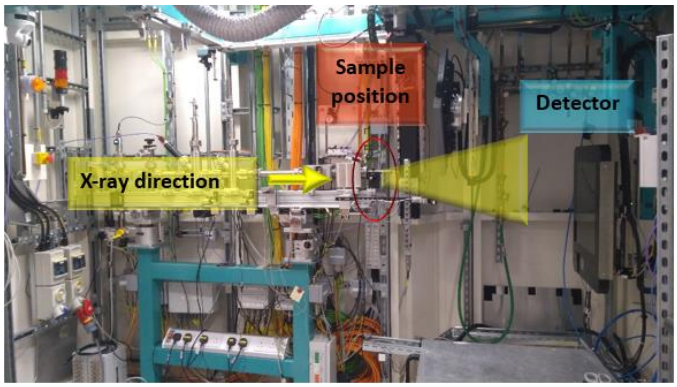


Figure 2.16 Layout of XPDF beamline I15-1, DLS

To post-process the X-ray total scattering data and extract the reduced structure factor, the following datasets should be acquired:

- Calibration material (silicon in the capillary or Kapton tube) for beamline alignment
- Sample environment (e.g. capillary or Kapton tube) or background material (e.g. solvent in capillary or Kapton)
- Sample data

For background subtraction, X-ray total scattering data of the sample environment, e.g. a blank capillary, should collected for (at least) the same statistical accuracy as the sample data [201].

The high-level workflow for the collection of X-ray total scattering data is summarised in Figure 2.17. The method employed will depend on, amongst other things, the type of experiment that you are running *(ex or in situ*), and the type and scattering potential of the sample (see the decision points highlighted in the workflow, Figure 2.17).

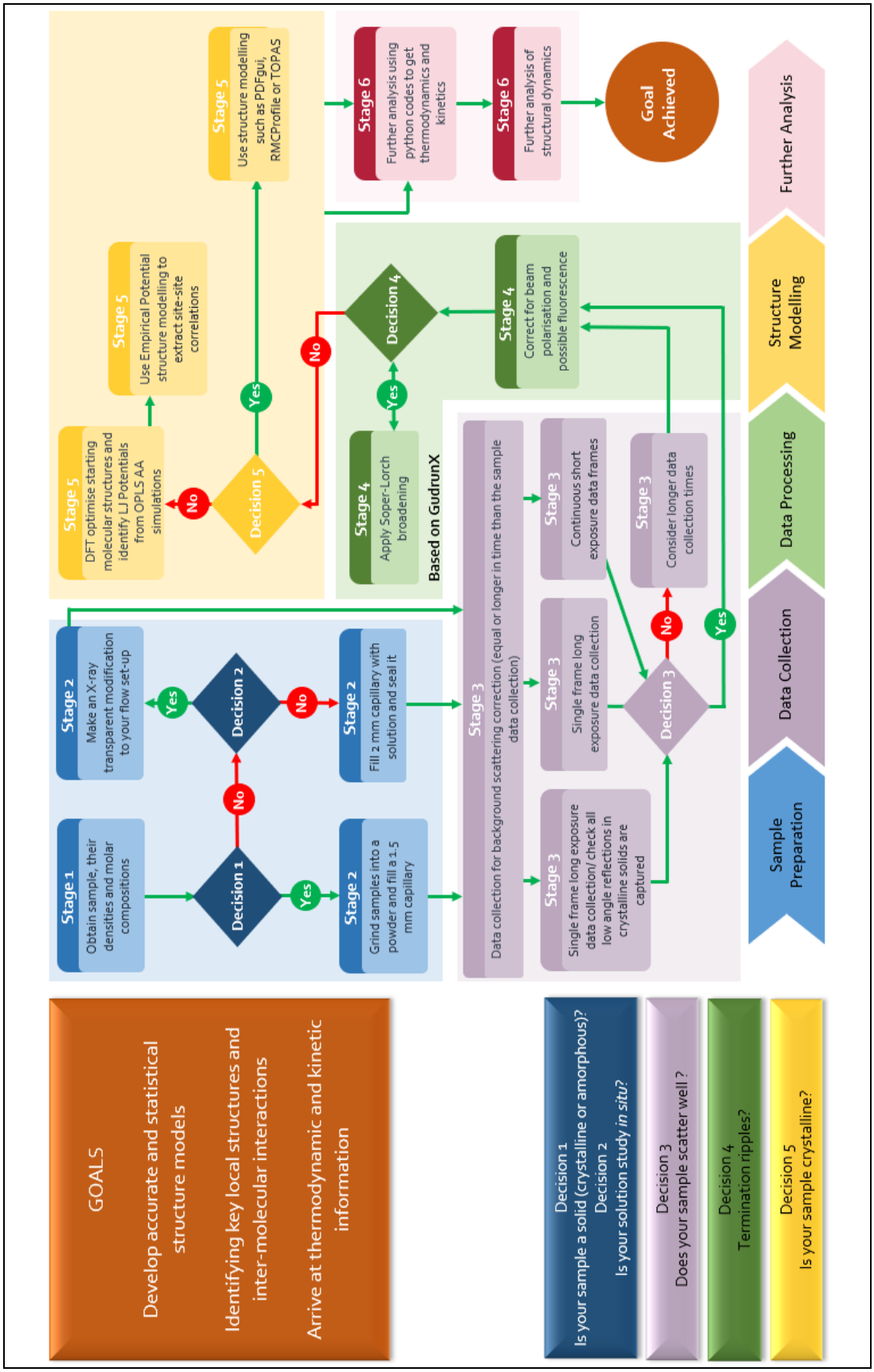


Figure 2.17 High-level workflow for X-ray total scattering data collection

Developed with Dr. A R Pallipurath and A Pugejs

2.5.4.1.2 Sample mounting

Three sample-mounting configurations have been used for the data collection outlined in this thesis: spinning capillaries; static capillaries; and a static Kapton tube.

The powder and some liquid and solution samples were measured in borosilicate glass capillaries. The samples in 1.5 mm OD (1.17 mm ID) borosilicate glass capillaries were sealed with clay (Section 2.2.3) and mounted on the multi-capillary spinner (max. 15 samples, Figure 2.18a). Spinning the sample improves powder averaging [356], and averages out the curvature of the capillary. Sample packing is estimated to be 0.5 for powder samples. The alignment of the capillaries is consistent, as they are held in place with (drill) chucks.

As one beamtime session involved a large capillary (*ex situ*) data collection, a high-throughput sample holder was commissioned by the beamline scientists to minimise sample change-over time (Figure 2.18b). The high-throughput holder holds up to 50 capillaries and can be filled *ex situ*, whereas the capillary spinner holds up to 15 capillaries (13 at the time of the experiment) and required the capillaries to be mounted individually in the hutch. The high-throughput sample holder requires borosilicate capillaries with a 2 mm OD, compared with 1.5 mm OD for the spinning capillary holder. Larger-diameter capillaries can be beneficial for organic samples, which are weakly scattering, to increase the acquired signal, although geometric effects can reduce the PDF resolution [291].

The 2 mm capillaries were heat-sealed (see Section 2.2.3.1). In this case, as the capillaries are static, there is no averaging of the curvature. Although it could be seen from the experimental hutch that the beam position was roughly central, only two sample positions were fully calibrated using silicon. Validation of the data from the high-throughput sample holder is addressed in Section 2.5.4.1.4.

The solutions and solvent samples in the *in situ* crystallisation experiment were measured in a 6 mm OD Kapton tube that was mounted on the bespoke holder (Figure 2.18c). With a much larger diameter than the capillaries, any inaccuracies due to the curvature of the static Kapton tube is much lower than for the capillaries. It was noted that the circulation of the solution or solvent systems using a peristaltic pump could introduce pulsation in the Kapton tube, especially at higher flowrates. This problem was considered in flowrate selection.

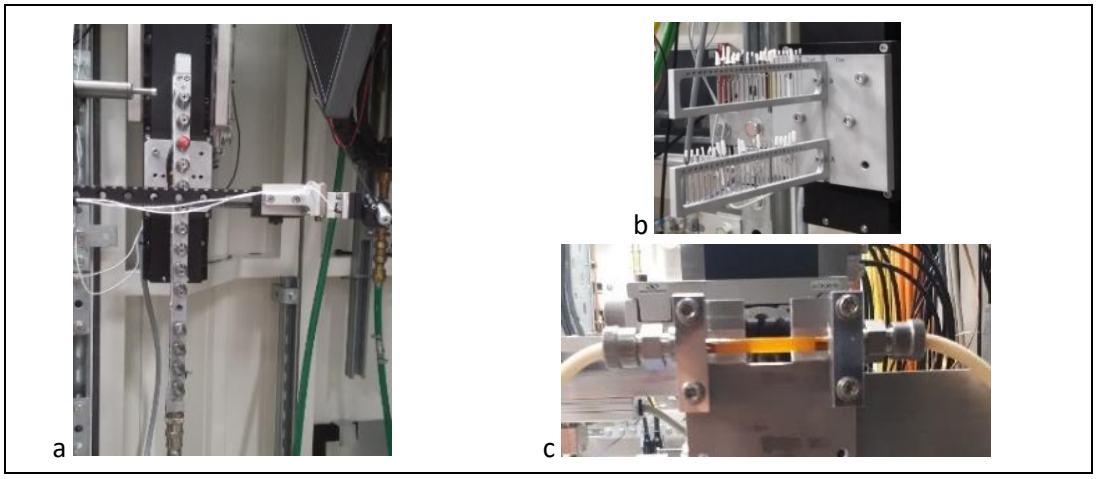


Figure 2.18 XPDF sample holders at beamline I15-1, DLS

a) multi-capillary spinner, b) high-throughput sample holder, and c) Kapton tube

2.5.4.1.3 Data processing

The generic data acquisition software at DLS automatically applies the initial corrections for the internal dark current to the raw data during data collection. These normalised structure factor data sets S(Q), are further reduced and normalised with the GudrunX software to account for the background contribution from the sample container and surrounding air, sample absorption, multiple scattering and Compton scattering events, fluorescence, and composition and geometry [238], to give the interference differential cross-section F(Q). Multiple scattering or absorption of the high-energy X-rays used in XPDF analysis is generally not significant [201]. The Q-dependent background is removed with a top hat function prior to Fourier transform [365,368]. GudrunX can use one or multiple frames of data to produce F(Q). The Fourier transform of F(Q) leads to the reduced PDF G(r) [365,368].

Errors in the final real-space PDF may arise from a number of sources, including: termination ripples that are introduced in the Fourier transform process of the data from reciprocal-space to real-space; low-resolution measurements that are limited by low Q_{max} or instrument resolution; insufficient counting statistics; or inaccurate data correction [374]. Termination ripples are minimised where $Q_{max} > 30$ Å^{-1} [374]. Q_{max} should be selected to ensure that broadening due to termination is insignificant compared with thermal broadening [374].

The atomic composition of the sample is applied in the GudrunX processing, along with the required top hat width, broadening factor (because of the finite Q_{max} applied [201]) and a broadening function of width 0.125 Å to reduce the appearance of termination ripples in the real-space distribution. A top hat function width of 3.5 Å^{-1} and r_{min} = 0.65 Å (to remove artefacts in the structure below this value) are applied in this analysis.

Analysis of the X-ray total scattering data using computational methods (Section 2.9.2) can be used to identify partial pair distribution functions $g_{\alpha\rho}(r)$ from intermolecular interactions.

2.5.4.1.4 Capillary sample holder validation

The high-throughput sample holder (described in Section 2.5.4.1.2) was used to mount the GuHCl solution and deionised water samples for the GuHCl concentration study (Chapter 5). Spinning the capillary averages out the curvature of the capillary in the region through which the X-rays pass. In a static holder, this averaging does not happen hence it is vital to validate whether the sample geometry, i.e. static or spinning capillary, affects the quality of the data acquired.

This section reports the results of the comparative studies of the XPDF patterns for water when measured in a 1.5 mm and 2.0 mm capillaries. In addition, the effect of the data collection format (single or multiple frames) and length of scan collection are considered.

Data collection: Data were collected as shown in Table 2.2. For background subtraction, X-ray total scattering data of blank 1.5 mm and 2.0 mm capillaries were collected for at least the same amount of time as the sample data.

X-ray total scattering data for the 6 M GuHCl sample were collected, twice with 20 frame x 30 s and twice with a single frame x 600 s, to confirm reproducibility in the high-throughput sample holder and to examine the effect of data collection time. The frames in the multiple-frame scans can be processed individually or in groups of up to 20.

Table 2.2 Background and sample scan times

Sample	Background scan (empty capillary)	Sample scan
1.5 mm water	1 frame x 600 s (600 s)	20 frame x 30 s (600 s)
2.0 mm water	3 frame x 600 s (1800 s)	20 frame x 30 s (600 s)
2.0 mm 6 M GuHCl solution	3 frame x 600 s (1800 s)	20 frame x 30 s (600 s)
6 M GuHCl repeat measurements	3 frame x 600 s (1800 s)	20 frame x 30 s (600 s)
6 M GuHCl single frame (plus repeat)	3 frame x 600 s (1800 s)	1 frame x 600 s (600 s)

Capillary spinner vs static high-throughput sample holder comparison: The capillaries in the high-throughput sample holder are not as accurately geometry optimised in the data collection position as the capillaries in the capillary spinner due to the lack of averaging of the cylindrical geometry. X-ray total scattering data for water samples were collected from a 1.5 mm capillary in the capillary spinner and a 2.0 mm capillary in the static high-throughput sample holder to verify that the similar data is generated. The data were processed using GudrunX and the resulting F(Q), g(r) and G(r) patterns are shown in Figure 2.19.

There is reasonable agreement between the 1.5 mm and 2.0 mm capillaries for each of the datasets. A degree of variability between repeat total scattering datasets of the same material is expected [375], all the more so for a liquid system where the molecular structure is dynamic. Only a subtle difference in g(r) and G(r) at 2 Å is noticed in this case (Figure 2.19b), where the data collected from the spinning 1.5 mm capillary show two interactions while data from the 2 mm capillary suggest one broad range interaction. This subtle difference may be due to errors introduced during the Fourier transformation. The EPSR molecular model refines simulated site-site partial structure factors to the experimental structure factor data, and so if the features in r-space are artefacts because of the Fourier transform process, they will not occur in the simulation.

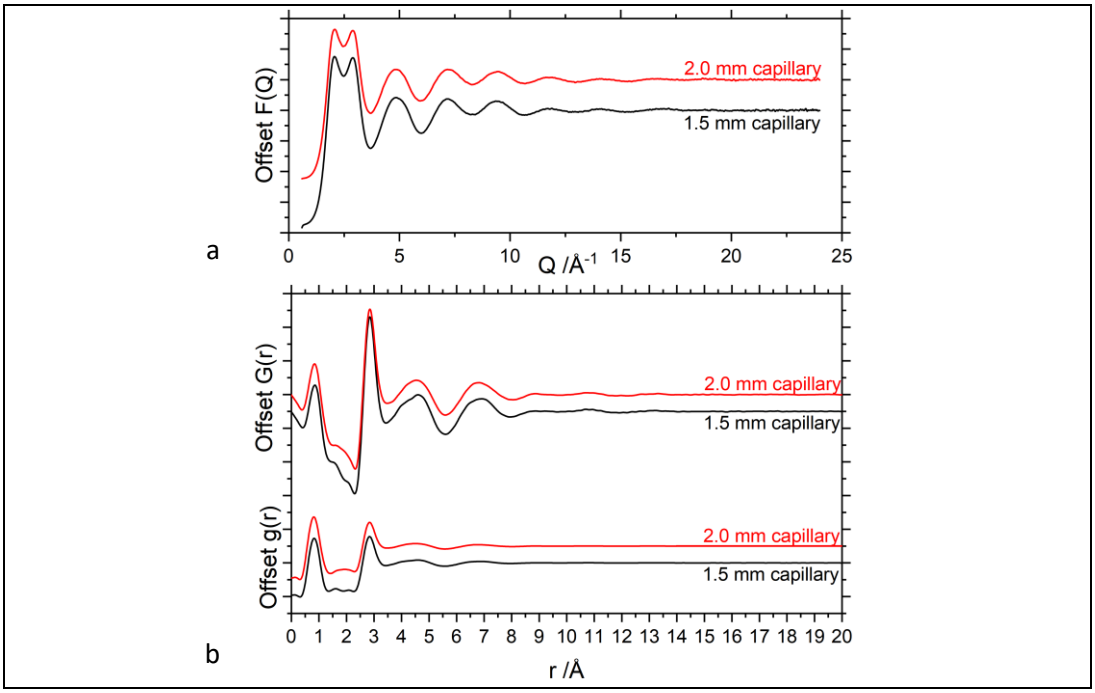


Figure 2.19 Comparison of a) F(Q) and b) G(r) and g(r) for water samples in 1.5 and 2.0 mm capillaries

Study of the length of data collection and repeatability of XPDF measurements: In this section, X-ray total scattering data collection for the 6 M GuHCl aqueous solution was repeated to see if the data sets are comparable (Table 2.3). The measurements were made with the same 2.0 mm capillary in the same position in the high-throughput sample holder, although the measurements were not all taken consecutively thus the holder position was moved between the repeated data collections.

Table 2.3 6 M GuHCl 2 mm capillary & background scans

Scan number	Scan name	Number of frames	Length of frames /s	Total scan length /s
26556	Empty capillary	3	600	1800
26599	Multi_1	20	30	600
26604	10min_1	1	600	600
26605	Multi_2	20	30	600
26615	10min_2	1	600	600

Datasets Multi_1 and Multi_2 were processed using identical composition and processing parameters (e.g. polarisation and sample background) in GudrunX (Figure 2.20 a & b). There are clear differences at Q > 14 $Å^{-1}$ and r ≈ 1 Å. Features at low-r (< 1 Å) may be affected by termination ripples due to the Fourier transform process. This effect can result in artefacts, e.g. features that are not real, or incorrect scattering intensities. The 10min_1 and 10min_2 data sets were more similar when processed with identical parameters (Figure 2.20 c & d).

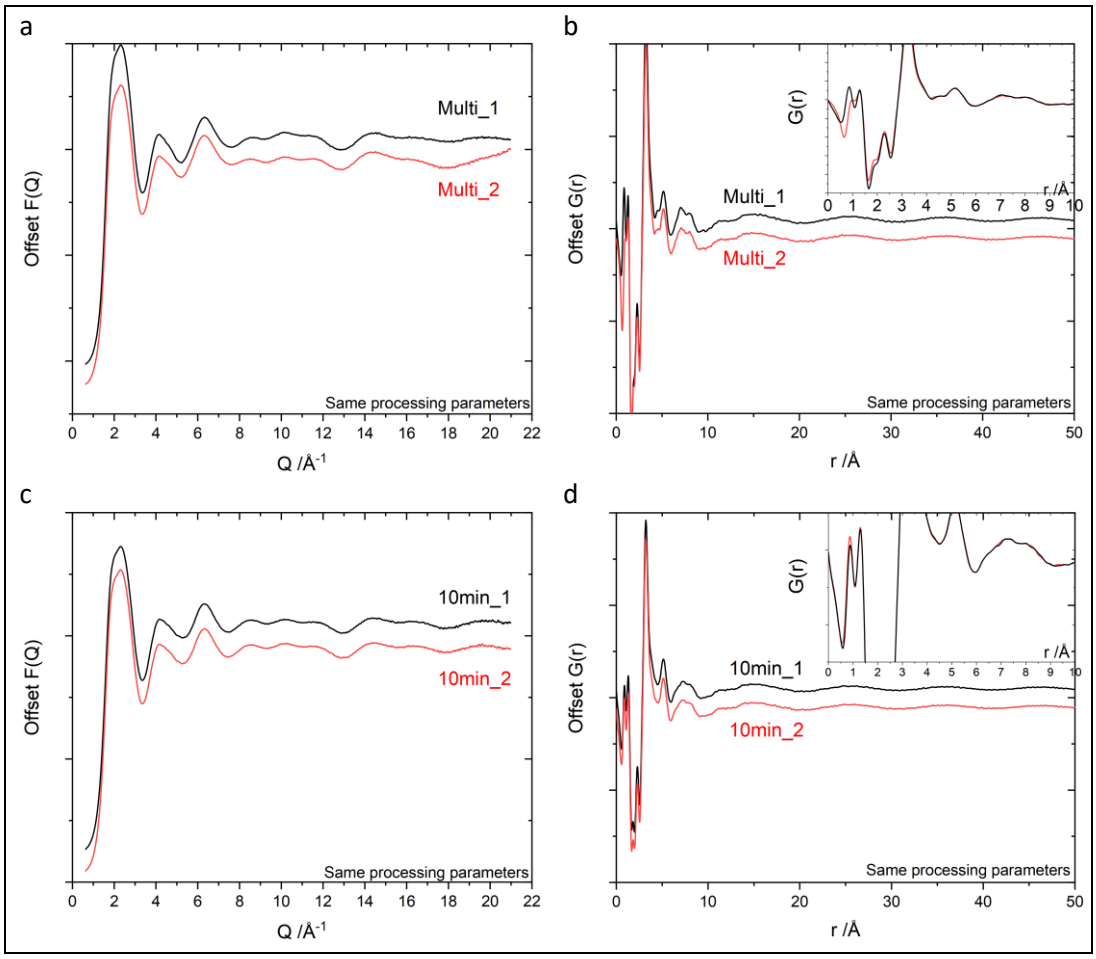


Figure 2.20 Comparison of F(Q) and G(r) with identical processing parameters

(a&b) multi_1 and multi_2 (averaged 20 frames of data, 30s each frame), and (c&d) 10min_1 and 10min_2 (single 10 minute frames of data)

These tests indicate that better results are achieved when the processing parameters set in GudrunX are refined for each dataset. The block processing of multiple sets of data using the

same parameters may be suitable if there are many scans to process, but each set of processed data would need to be checked so the best results would be provided be individual scan processing.

When each individual data set was refined separately, the F(Q) and G(r) of Multi_1 & 2 and 10min_1 & 2 are similar (Figure 2.21). Again, there are some low-r differences that are likely due to termination ripples. There are also differences between the single-frame and the multiple-frame G(r) at around 4.5, 7 and 8 Å (indicated with *). These features, like the features at r < 1 Å, are low intensity, so their position and area are heavily influenced by the termination ripples generated during the Fourier transform[12] that represent the larger peaks beside them. The XPDF patterns for the multiple-frame scans (Multi_1 and Multi_2) have clearer peaks indicating better resolution data than the single-frames scans (10min_1 and 10min_2), although both sets of data have similar counting statistics. This observation is potentially due to the point at which the data is averaged by the generic data acquisition system (before or after the integration of the 2D diffraction patterns).

Comparison of the four complete X-ray total scattering data sets (Multi_1 and Multi_2, and 10min_1 and 10min_2) shows that, although there are some slight differences in intensity, the two data collection strategies of one single 10 min frame vs 20 x 30 s frames result in acceptably similar F(Q) and G(r) features.

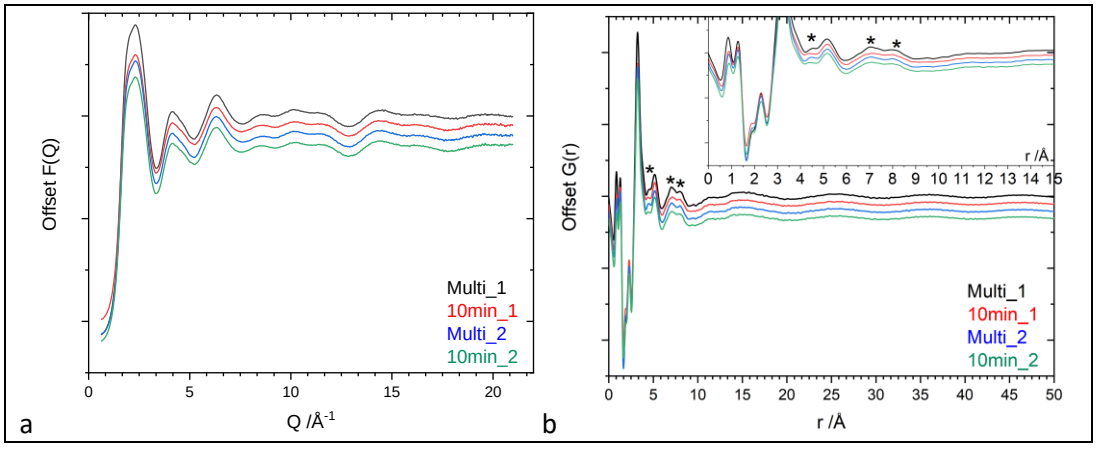


Figure 2.21 Comparison of a) F(Q) and b) G(r) for 6 M GuHCl X-ray total scattering data collections

Effect of collection time: Increasing the time that X-ray total scattering data is collected increases the signal-to-noise, making features in the scattering pattern clearer. To look at the effect of data collection time on the GuHCl solution data, different numbers of frames

[12] The Fourier transform of the top-hat function applied to the structure factor is a sinc function

of scan 26599 (sets of 1, 5, 10 and 20 scans) were processed together in GudrunX. There is significant noise in the high-Q region after a single frame (30 s data collection), which is significantly reduced with the combination of five frames (Figure 2.22a). Doubling the number of frames from five to 10 has little impact, but with 20 frames, the data is smoother at high-Q. Increasing the data collection statistics by a factor of n increases the signal-to-noise by a factor of √n [376], which is roughly in line with what is seen here. In the G(r) plot (Figure 2.22b (*)), there are two features between 9 and 10 Å with one and five frames, but just one peak with 10 and 20 frames. A 30 s scan does give a good indication of the structure factor and related of a GuHCl solution sample, but it should be noted that some of the smaller features should be treated with caution when interpreting the data.

In the *in situ* study (Chapter 4), data were collected in 30 s frames. This comparative study indicates that the structure factors collected in that time should have sufficient counting statistics to give broadly reasonable PDF patterns, although slight differences in features compared to a system that is measured for a longer time period could be expected. There is noise even at Q = 21 $Å^{-1}$, which is much lower than what can be measured experimentally and may cause some distortion at low-r (Figure 2.22b (inset)). However, in the case of the *in situ* study, the selection of the data collection time was a balance between capturing sufficient counting statistics for a reliable structure factor and recognising that solution structure will be changing with the decreasing temperature (cooling rate)[13].

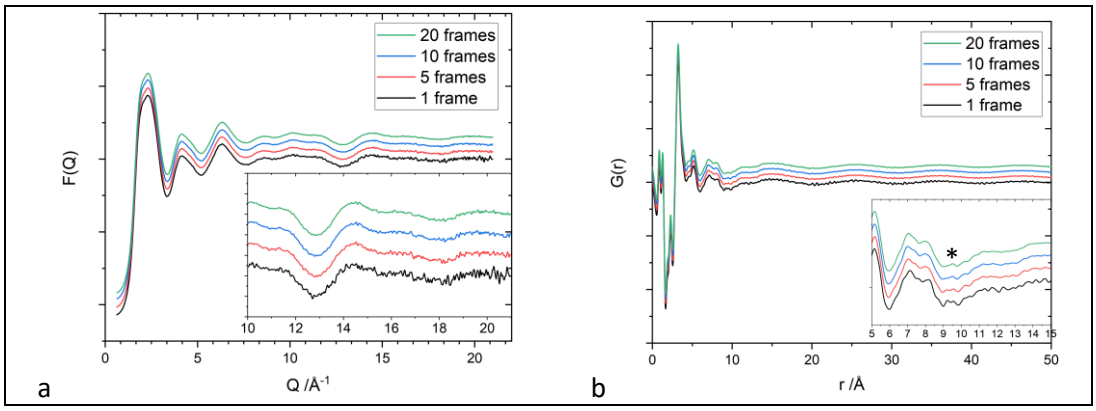


Figure 2.22 Comparison of the effect of collection time on data quality: a) F(Q) and b) G(r) for 6 M GuHCl X-ray total scattering data collections.

1, 5, 10 and 20 frames of data from the same data collection were averaged to see the effect on the data quality

[13] Future *in situ* experiments will benefit from better time/temperature resolution with faster data acquisition for example with X-ray Free-Electron Laser (XFEL)

As EPSR modelling is based on statistical averaging, it is likely that any small differences in the measured S(Q) and resultant g(r) would be averaged out. This comparison of data acquisition protocols has been useful to determine that there are no significant differences in the principle XPDF pattern features between the single 30 s frame of data, as is acquired in the *in situ* study, and a 600 s data set with higher statistics. In the comparison of the molecular structures of the different concentrations of GuHCl solutions (Chapter 5), the 600 s X-ray total scattering data sets are used, to maximise the available counting statistics.

2.6 Scanning electron microscopy (SEM) imaging with Energy Dispersive X-Ray (EDX)

Scanning electron microscopy (SEM) generates high-resolution images of a sample under vacuum by scanning the surface with a high-energy electron beam, which is focused using electromagnetic lenses [377]. The electrons interact with the sample, elastically or inelastically, to generate secondary electrons (SE), backscattered electrons (BSE) or X-rays. By scanning across an area of the sample surface, images of the sample's topography and underlying structure can be created.

Inelastic scattering results in energy transfer from the incident electron to the electrons in the sample. If the transfer of energy is sufficient to cause ionisation of sample atoms, SE are emitted from the sample [377]. SE are much lower energy (< 50 eV) than BSE and because SE are from the top few nanometres of the sample, they generate images of the sample topography. The brightness of the SE image is related to the energy of the detected SE. Although BSE produce lower-resolution images than the SE, they provide information from further below the material surface than SE [377].

If the incident electron is deflected with negligible energy transfer (elastically) and the deflection is > 90˚, BSE are produced [377]. When they escape the surface of the sample, BSE are higher energy than SE (> 50 eV), having undergone single or multiple scattering events. Backscattering is increased when electrons are deflected from higher atomic mass (high-Z) elements. Higher energy incident electrons penetrate further into the surface of the sample thus increasing the volume of the sample that is probed [377]. The penetration of the electron beam is greater with increasing accelerating voltage and when the sample contains low-Z elements.

Energy Dispersive X-Ray (EDX) spectrometry is a microanalysis method that identifies the elemental composition of a sample from emitted X-rays (H, He and Li cannot be identified by this method) [378]. When an atom is ionised due to the excitation of a core electron, an electron from a higher energy level will fill the vacant core-hole, resulting in the emission of a photon (see fluorescence emission in Figure 2.4). The energy of the emitted photon will be indicative of the emitting atom. The relative intensity of the photon emission peaks of the is used for elemental quantification [378].

A Hitachi TM3030Plus tabletop scanning electron microscope was used for SEM analysis of GuHCl samples (as received and recrystallised) and EDX analysis of the *as received* material. The results of this characterisation work are in Appendix A.

2.7 Helium pycnometer density measurement

For the processing of the *in situ* X-ray total scattering data, the density of the recrystallised GuHCl is required. The sample of recrystallised GuHCl was measured using a Thermo Pycnomatic ATC helium pycnometer. Helium pycnometry is a gas-displacement technique. Helium is used due to its small atomic size, which allows it to enter the pores of the probed sample, and its ideal gas behaviour at room temperature.

The dried sample, of known weight, is put into sample cell of known volume. Helium gas is introduced into the sample cell. The helium is then expanded into a second cell of known volume. As the volume of an ideal gas is inversely proportional to its pressure (Boyle's law), the volume of the gas in the sample—and therefore the volume of the solid material—can be determined. From this, the solid sample density can be calculated. The results of this characterisation work are in Appendix A.

2.8 Crystal16

Solubility measurements of eight concentrations of aqueous GuHCl solutions were carried out using two Technobis Crystallization Systems Crystal16 benchtop crystallisation units [32]. The data were analysed to inform the concentration of solution that should be used for the *in situ* XPDF crystallisation study.

Crystal16 uses turbidimetric detectors to monitor the dissolution ('clear point') and crystallisation ('cloud point') temperatures of a solute in a solvent. Initially, the solution turbidity is tuned to set the starting transmission to 100%, i.e. fully dissolved. A reduction in light transmission indicates that nucleation has occurred. 1.0 ml solution in 1.5 ml vials are

heated and cooled thorough a set range of temperatures at defined heating and cooling rates between 5 and 70°C , including holding the solutions at set temperatures for a time before continuing the temperature cycling, if required. The solutions are stirred (700 rpm) with micro-magnetic stirrers (7 mm x 2 mm OD).

Solutions prepared at 1.0, 2.0, 4.0 and 6.0 M (unit 1) and 7.0, 7.8, 9.3 and 9.8 M (unit 2) went through two sets of investigations at different heating and cooling rates. The details of the heating and cooling cycles and the results of this characterisation work are in Appendix A.

2.9 Computational methods

This section outlines the computational analysis methods applied in this thesis. Quantum chemistry calculations to geometry optimise structures for DFT/TDDFT molecular orbital analysis, to generate K-edge X-ray absorption spectra for XRS experimental data analysis (Chapter 3), and to establish starting geometries for the Empirical Potential Structure Refinement (EPSR) simulations (Chapter 4 and Chapter 5).

Computational techniques available for the modelling of XPDF data are outlined, with details of the application of PDFgui for the analysis of solid XPDF data and EPSR for the analysis of liquid and solution systems used in this study.

2.9.1 Quantum chemistry calculations

DFT and TDDFT calculations have been undertaken using ORCA (version 4.0.0) [379] on ARC, the high performance computing cluster at the University of Leeds. ORCA is a quantum chemistry program that allows *ab initio*, DFT and semiempirical SCF-MO (self-consistent field molecular orbital) calculations [379]. Local-density approximations (LDA), generalised gradient approximations (GGA), meta-GGA, hybrid, double-hybrid, and range-separated functionals are available in ORCA [380]. ORCA applies DFT code that uses a highly-optimised numerical integration algorithm for the exchange-correlation.

Basis set: In order to calculate the energies of molecular orbitals, the orbitals must be defined by mathematical functions [251]. Wavefunctions are used to describe the probability density of an electron in an atomic orbital. When considering molecular orbitals, a series of basis functions (the basis set) are used to approximate the linear combination of atomic orbitals (LCAO) [251].

The selection of the basis set for quantum mechanical calculations is a compromise between the required accuracy and the calculation cost [381]. As the complexity of the system

increases, so should the applied level of theory. For simple (single atoms) systems, a minimal basis set may be suitable, but they cannot describe the anisotropic distribution of electrons in molecular orbitals [251,382]. An increased number of basis functions are applied by using a double-zeta (DZ), triple-zeta (TZ) or quadruple-zeta (QZ) basis set [382]. A polarisable basis set allows for the deformation of the electron density [382], which is important in molecular systems.

Gas-phase calculations were initially carried out with a Pople 6-31G* basis set, a standard, split-valence double-zeta basis set modified for polarisation, but as it was originally optimised at a Hartree-Fock level, it is less suitable for the correlated calculations [383]. Instead, a minimally-augmented ma-def2-TZVP basis set[14] [384] has been applied in all ORCA quantum chemistry calculations.

Polarised basis sets developed by Dunning, which are optimised using correlated CISD (configuration interaction, single and double excitation) wavefunctions, provide a description of correlated systems, but are computationally costly [381]. Karlsruhe basis sets, which are modified Dunning basis sets and have been shown to reproduce interaction energy data that is competitive with the Dunning basis sets, have also been further developed for organic compounds, and so have been applied here [381,385]. The triple zeta TZVP, which was developed to be suitable for low-Z elements (H–Rn), is suited for quantitatively accurate DFT treatments (a larger, more costly, polarisation would be required for Hartree-Fock treatments). Examples of its use for the TDDFT analysis of K-edges are seen in the literature [67,386].

The Coulomb fitting auxiliary basis set[15] def2/J provides good geometry optimisation with low bond length and angle errors and is considered adequate for noncovalent interactions and ionisation potentials, and as such is suitable for the calculation of geometries and conformational energies of organic compounds [381,384,387].

The RIJCOSX (resolution of the identity (RI)) approximation [388] speeds up the hybrid TDDFT calculations[16] [379,389], and the diffuse basis functions needed for excited state calculations in the form of minimally-augmented diffuse basis functions (ma-def2-TZVP) provide a good

[14] Valence triple-zeta basis set with polarisation function

[15] Approximate Coulomb energies in conjunction with orbital basis sets of split valence, triple-zeta valence and quadruple zeta valence quality

[16] By expanding the electron density to allow faster calculation of the Coulomb integrals

cost-accuracy compromise and is sufficient for weakly bound systems [390–392]. The minimally-augmented basis set allows the use of a larger zeta set [392].

Dispersion forces should be considered, in particular for organic molecules, to help describe interactions such as hydrogen bonds [393,394]. The applied D3 semiempirical dispersion correction accounts for basis set superposition errors and London dispersion effects and is suitable for use where there are weak intramolecular interactions [395,396].

Functional: B3LYP, a hybrid (20% Hartree-Fock exchange) functional with local and gradient-corrected exchange and correlation terms, was applied for most calculations [383,397].

Frequency calculations: Vibrational analysis was performed on each geometry-optimised system to ensure no imaginary modes were present, which would indicate that the geometry was not at a local energy minimum.

TDDFT: TDDFT enables an analysis of the transition state and is used to calculate K-edge absorption spectra by restricting the donor orbitals for core level excitations [387].

EPSR geometry optimisation: The geometry optimisation of Gdm^+ for EPSR simulations of GuHCl was initially carried out using RKS BP86 RI SVP def2/J. A review using the functional and basis sets described above[17] resulted in a bond length difference of around 1%. As the molecule geometries are not fixed in the EPSR simulation, the EPSR simulations were not rerun based on this alternative optimised geometry. The geometries of the water molecules used in the EPSR simulations (TIP3P and SPC/E) were defined in the literature.

2.9.2 Computational modelling for X-ray total scattering analysis

In XPDF analysis, computational analysis is required to fit the structural models to the data and extract quantitatively-reliable information [371]. There are many tools available for the analysis of X-ray total scattering data and selecting the appropriate tool depends on the type and level of complexity of the probed system.

Modelling techniques can be divided into ‘small-box’ and ‘big-box’ methods. Small-box methods use Rietveld refinement to fit theoretical structure data to experimental data by varying crystallographic parameters [398], e.g. Diffpy-CMI [399], which can integrate data from a range of experimental inputs, such as EXAFS or TEM, to provide additional constraints, and its related graphical user interface, xINTERPDF [350,400], for the analysis of organic crystal structures, TOPAS [356,401], and PDFfit2/PDFgui, which is principally for

[17] B3LYP RIJCOSX D3 ma-def2-TZVP def2/J

nanoscale studies [402]. This type of analysis is suitable for materials with periodic structures, but not liquid or amorphous materials, or crystalline materials with significant local disorder [403]. Big-box methods involve moving the atoms around in a structural model to minimise the difference between the experimental and calculated PDF, often using the reverse Monte Carlo method [398], e.g. RMCprofile, with periodic boundary conditions, for modelling the local structure of crystalline to amorphous materials, or liquids [404,405], and DISCUS [354], which has been developed to build complex models and identify local disorder in crystalline materials.

EPSR combines these approaches, as the structural model is developed by applying a small number of parameters, before it is then refined to the experimental data [398,406].

A new code based on EPSR, Dissolve, is in development and is capable of handling multi-phase systems with up to $3x10^6$ atoms [407]. Dissolve will combine the modelling and data analysis of EPSR and dlputils [408] in one package, however it is still undergoing testing and so could not be applied in this study.

In this study, the structure of solid materials (as reference XPDF patterns) will be analysed using PDFgui, and the liquid and solution structures will be analysed by EPSR.

2.9.2.1 Solid XPDF analysis

PDFgui is the graphical user interface of PDFfit2 and refines crystallographic structural data (available from the crystallographic information file (.cif)) to experimental PDF data [402]. PDFfit2 is based on the real-space Rietveld refinement technique so as with crystallography, Rietveld refinement is applied to develop a crystal structure model consistent with diffraction data. The tool adjusts the crystal structure and instrumental parameters until the structural data matches the XPDF data [402]. Experimental factors, such as temperature and Q-range, are taken into account for the refinement, alongside lattice parameters (*a, b, c* and *α, β, γ*), atomic positions, a PDF peak broadening term related to correlated atomic motion, and displacement and scale factors [402].

2.9.2.2 Liquid and solution XPDF analysis

The two main methods for the refinement of 3D structural models to interpret total scattering data of non-periodic disordered (liquids and non-crystalline) materials, RMCprofile and EPSR [225], are both based on Monte Carlo methods. Monte Carlo methods are based on Boltzmann-like statistical mechanics distributions to generate probable structure model solutions [257].

Reverse Monte Carlo (RMC) methods are used to generate a molecular model, for which a reduced structure factor F(Q), and pair distribution function g(r), is calculated [403]. The ensemble of atoms are moved at random, with the movement being accepted if it results in the computed functions being closer to the experimental data, based on a χ^2 comparison, or according to a certain probability (analogous to a Boltzmann factor in MC simulations) if the fit is worse [225,403]. There is no requirement to apply a force field to the particles in the first instance [404]. One potential challenge to the robustness of the molecular structure generated by RMCprofile is that there is likely to be more than one configuration of a statistical ensemble of particles that could be consistent with diffraction data, although arguably, this would be expected in nature [409,410].

EPSR was initially created at ISIS Neutron Source for the interpretation of neutron scattering data and it is also suitable for the analysis of X-ray total scattering data. It is the only available software for the analysis of the liquid and solvated systems. EPSR evolved from the RMC method, but implements pairwise-additive interaction potentials (the reference potential RP, also referred to as the seed potential in the literature) to drive the interatomic interactions, and therefore the site-site-pair correlation functions, in an ensemble of particles [409,411]. A complete account of the development of EPSR can be found in the literature [406,409,411].

2.9.2.2.1 Overview of EPSR

EPSR is a Monte Carlo-based modelling tool for the visualisation of total scattering data by producing a statistically-probable molecular structure model of the probed system [257]. EPSR aims to produce a simulated differential scattering cross section that fits the experimental data through the random movement of components in the simulation box.

EPSR first captures distinctive features in the probed system, for example hydrogen-bonding between water molecules and water structure in aqueous systems, or ionic charges, using the RP (atomic potentials applied to each atom type) [105]. Perturbations (empirical potential EP) are then introduced to refine the fit to experimental data [105].

The simulation box starts with the same sample composition and sample properties, including temperature and atomic number density, as the real sample. Using the RP comprising 12–6 Lennard-Jones potentials and Coulomb charges, an atomic configuration is generated without reference to the experimental data. The atoms and molecules undergo small random moves through molecular translation, rotation of the whole molecule or a side chain, or atomic moves. Based on the Metropolis MC algorithm, if the move causes the

potential energy of the system U, to reduce, the new configuration is accepted [257]. If U increases, the move is accepted with a probability P, described by [257]

$$\mathrm{P} = \exp\left[\frac{-\Delta U}{kT}\right] \qquad \textbf{Equation 23}$$

where k is the Boltzmann constant and T the temperature in K. Once the system's potential energy has stabilised and the model is structurally equilibrated, the experimental reduced structure factor F(Q) is used to guide the molecular configuration in the EP stage [412].

The introduction of perturbations to the RP by way of the EP reduces the difference between calculated and experimental F(Q) values, increasing the goodness of fit to the experimental data. This stage of the simulation results in a more statistically probable configuration than would be achieved if the atomic force field data alone were applied. However, it is important that the initial RP data produces a configuration that gives a reasonable fit to the data before applying the perturbations via the EP. There are no long-range effects on the potential energy, in part due to the limitation of the range over which the EP can be applied [257], which may affect how well EPSR models a system with charged ions. The refinement of the EPSR model fit is based on comparison of the simulated to experimental structure factors, rather than the comparison of real-space data, to avoid fitting non-physical termination ripples in the PDF introduced during the Fourier transform process [235].

Discrepancies between the scattering data and the simulated structure will always be present and cannot be fully accounted for [235]. Errors may be introduced at low-Q (< 2.5 $Å^{-1}$) during the processing of the data, in particular when subtracting the background and estimating the inelasticity corrections (due to electron recoil in the case of X-ray total scattering) [224,235]. These errors are greater for low-Z systems [235]. In the case of these data, the background subtraction may have introduced errors due to the potentially imperfect geometries of the static capillaries in the high-throughput holder – the instrument calibration was undertaken on a couple of, rather than all, capillary positions. The Fourier transform process also introduces artefacts due to the finite counting statistics and experimental Q-range [235]. Whilst the EP phase of the EPSR simulation tries to refine the simulation to the experimental data, it should not be forced to fit errors and uncertainties in the data—the aim should be to quantitatively and qualitatively fit the main features [235].

The high-level workflow for molecular simulation using EPSR is summarised in Figure 2.23.

2.9.2.2.2 EPSR simulation setup

The simulation box is defined in terms of the components, temperature and atomic number density. The reference potential (RP) is defined by the Lennard-Jones and atomic partial charge parameters:

Force field selection: EPSR version 25 [257] used in this study supports the use of Lennard-Jones potentials to represent short-range repulsive forces and longer-range dispersion forces, and effective charges that are truncated at large distances, to define the force field.

To identify the most appropriate values to apply to Gdm^+ and Cl^-, a parametrisation study for force field assignment was undertaken to compare the Lennard-Jones potential values and partial charges from the literature, including the application of reduced charges (summarised in Appendix F).

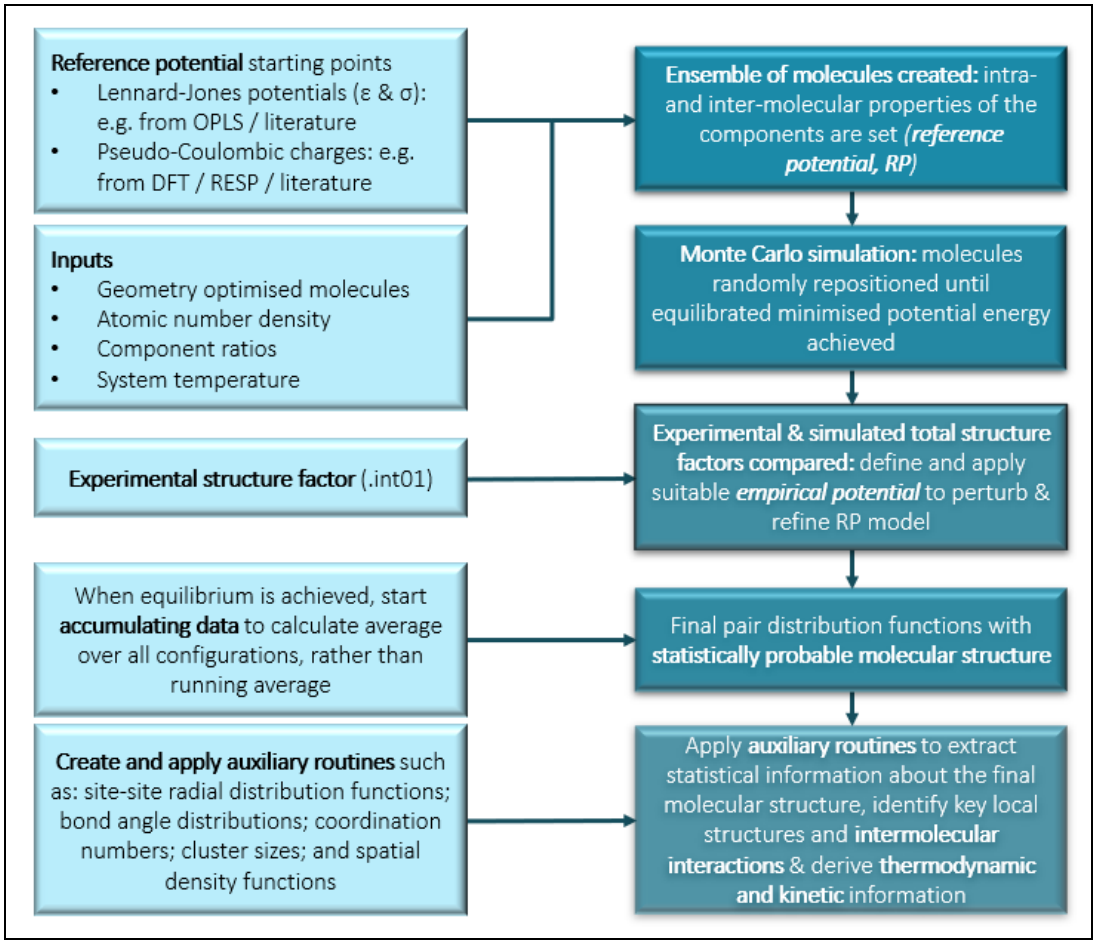


Figure 2.23 EPSR workflow diagram: a description of the stages of an EPSR simulation

The required inputs are outlined on the left-hand side, and a description of each stage is on the right-hand side

Water model selection: Although many previous EPSR models of water or aqueous systems have used various single component water models (TIP4P/2005 [375], SPC [413], SPC/E

[223,235,413–416], a recent XPDF/EPSR study used a combined water model, applying SPC/E and TIP3P water models, to achieve the best fit to total X-ray scattering data of acetone-water binary solvent samples [242]. A parametrisation study to select the water model for use in the simulation, which compared SPC/E, TIP3P and a 50%:50% combination, was undertaken. The TIP3P water model was selected for the simulations in Chapter 4 and Chapter 5. A summary of the parametrisation study, including the Lennard-Jones potentials, partial charges and bond angles of the water models used, is in Appendix F.

Each simulation starts with an expanded simulation box to allow the molecules to move past each other more freely in the initial stages. Once the system energy is below 0 kJ/mol, the size of the box gradually reduces to the correct volume. Each simulation was run with the RP until the system energy was at equilibrium (~1,500 iterations). The EP was then applied to perturb the reference model and to force the simulation to agree as well as possible with the experimental F(Q) data.

Atomic pair correlations (site-site partial structure factors and partial pair distribution functions) are generated, and additional quantities such as bond angle distributions, coordination numbers and distributions, and density fluctuations can be evaluated in EPSR [257]. Whilst the PDF gives a one-dimensional picture of the distribution of one atom or molecule around another, the spatial arrangement of the interactions are lost by the spherical averaging [97,257]. Angular g(r) and spatial distributions of one species around another were calculated using dlputils routines [408]. The routines used in the evaluation of the EPSR simulations in Chapters 4 and 5 are outlined in Appendix E.

Chapter 3 The electronic structure of imidazolium cations in aqueous solution: X-ray Raman scattering studies of the C and N K-edges combined with XPS and DFT calculations

3.1 Introduction

Dissociated ions in strongly interacting solvents have been shown to influence solvent structure beyond the first solvation shell [113,417]. Ions influence the thermodynamic and transport properties of the bulk solution [418], as electrostatic solute-solvent interactions, combining ionic effects and hydrogen bonding, can be locally stronger than solvent-solvent hydrogen bonding interactions [419], especially in highly soluble and highly ionic systems. Neutron diffraction studies of high-concentration aqueous salt solutions have shown that the presence of ions induces a change in the water-water interactions, with the degree of perturbation being dependent on the nature of the ion pair [95].

Ionic liquids (IL) comprise an organic cation and an organic or inorganic anion with a melting point below 100°C [142]. Their physical properties can be tuned by adjusting the species used [122]. The biological activity of IL make them useful in the evolving field of IL for active pharmaceutical ingredients, where IL are components of drugs, or in the drug synthesis process [420]. Imidazolium-based IL are increasingly being used as environmentally more benign reaction media, replacing volatile molecular solvents [148–150]. The properties of hydrophilic IL can be adjusted by varying the proportion of water added [421]. Preferential hydration of IL ions when dissociated in aqueous solution prevents water from interacting with the surfaces of biological molecules in solution, which changes enzyme activity or affects protein or peptide solubility [421].

Imidazole (Imid) is an organic heterocyclic compound with high solubility in polar solvents (Figure 1.8a). Imid groups play an important role in biological systems and are present in many natural and synthetic drug molecules [135]. Derivative compounds of Imid include purine, histamine and nucleic acids, and as a side-chain of histidine, Imid plays an important part in enzyme catalysis [135,137]. Imid acts as a proton relay through its two N moieties, which both form hydrogen bonds with water [137–139], with the pyrrole-type N1 acting as a proton-donor (pK_a = 14.9) [133], and the pyridine-type N3 as a proton-acceptor (pK_a = 7.0) [133].

The electronic structure of Imid in solution has previously been characterised in depth by core-level spectroscopies, including X-ray photoelectron spectroscopy (XPS) [160], near-edge X-ray absorption fine-structure (NEXAFS) [159], NEXAFS detected by X-ray Raman Scattering (XRS-NEXAFS) [67], and a combined X-ray Emission Spectroscopy (XES)/Resonant Inelastic X-ray Scattering (RIXS) method [161]. This XES/RIXS study was undertaken with soft X-rays using a flow cell.

The NEXAFS [159] and XRS-NEXAFS [67] studies quantified changes in the unoccupied density of states of Imid, including the effects of changing solution concentration [159] and phase change during crystallisation from aqueous solution [67]. The studies identified the sensitivity of Imid to its chemical environment, as the N K-edge features are different in the solid, solution and gas phases [67,159]. Further to this, the similarity in N K-edge spectra across the concentration range studied indicated that Imid self-association may occur even in dilute solutions [159]. The difference in ionisation potential (ΔIP) is an indication of the difference between the pre-edge 1s→1π* peaks. It was found that electron density redistribution in the crystal structure of Imid due to hydrogen bonding via the N centres reduces the 1s→1π* split (~1.5 eV in solid compared to ~1.7 eV in solution) [159].

Our previous XRS study of the structural evolution of an aqueous Imid solution through the metastable zone during cooling crystallisation identified minimal variation in the C and N K-edge spectra [67]. This finding suggested that desolvation is the rate-limiting step, as there is little change to the coordination, and therefore solvation, of the Imid molecules until the point of crystallisation.

The liquid-jet XPS and XES/RIXS studies also characterised the protonated form of Imid, cationic imidazolium ($ImidH^+$, Figure 1.8b), in aqueous imidazole hydrochloride (ImidHCl) solutions, and contrasted it to solvated Imid [160,161]. XPS provides valuable information for the interpretation of NEXAFS data, as the chemical shifts of the photoemission lines (core-level binding energy (E_B)) indicate the relative IP positions in the NEXAFS spectra. The XPS study identified the pseudoequivalence of the N binding energies E_B in $ImidH^+$ [160], and the XES/RIXS study indicated a limited influence of solute-solvent hydrogen bonding on the electronic environment of the compounds [161].

Here, it is demonstrated that hard X-ray techniques have developed sufficiently to acquire data of a sufficiently high resolution for the investigation of chemical speciation of a solute in solution. Furthermore, the use of a flow crystalliser designed to be more representative of a laboratory-type crystalliser is an important progression from narrow-channelled

microfluidic flow cells used previously [422], as the setup described here can be used for crystallisation experiments.

X-ray core-level spectroscopies are sensitive probes of the chemical state and local structure around molecular solutes and solvent molecules in solution [67,159–161,203,205,206,423,424]. NEXAFS, which probes the density of unoccupied bound states and shifts in the core IP, has previously been used to examine ionic molecular solutes [159,424–428]. Detailed information about chemical states and local bonding is determined from the energies of non-bonding and anti-bonding unoccupied molecular orbitals [67,203,319,424,428,429].

E_B shifts provide an insight into the solid-state structure of organic materials, including bond lengths [203], the extent of protonation [430], and hydrogen bonding [25], and characterise molecular ions by differentiating between protonated and deprotonated functional groups [160,423,430–434]. Recent developments of laboratory near-ambient pressure XPS (NAP-XPS) instruments now facilitate the examination of solution systems in the laboratory without microfluidics or microjets [348,435,436] .

At low momentum transfer (low-*q*), X-ray Raman (XR) spectra are equivalent to NEXAFS spectra [306]. XRS is a hard X-ray, non-resonant inelastic X-ray scattering technique allowing the use of complex sample environments for studies under extreme temperatures and pressures and avoiding the limitations of working under vacuum that are associated with soft X-ray techniques [211]. Previous work in the research group showed that a temperature-controlled jacketed flow cell enabled the *in situ* XRS-detected NEXAFS characterisation of Imid molecules in aqueous solutions, probing the evolution of molecular structure through the MSZ to crystallisation [67].

This chapter presents a combined XPS-NEXAFS-XRS approach to characterising solvated Imid and $ImidH^+$ systems by laboratory NAP-XPS and XRS-detected NEXAFS, and shows how these techniques, when combined with time-dependent density functional theory (TDDFT) calculations, lead to a comprehensive understanding of the local electronic structure of the ion in aqueous solution. A summary of the workflow is shown in Figure 3.1. The key experimental and computational methods and conclusions described in this chapter have been submitted to Physical Chemistry Chemical Physics for publication (July 2021)[18].

[18] Manuscript has been reviewed and responses to reviewers has been prepared.

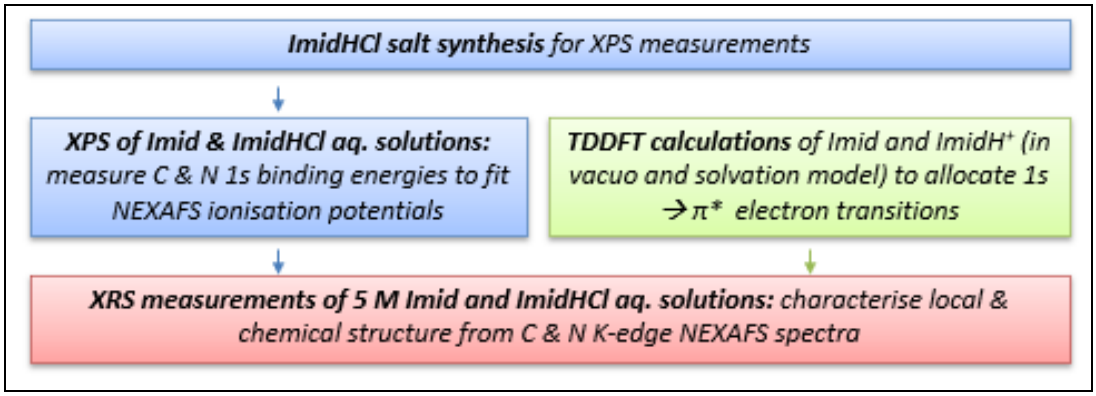


Figure 3.1 Combined NAP-XPS/XRS/TDDFT workflow

An overview of the experimental and computational techniques used to collect and interpret XRS spectra

3.2 Methodology

3.2.1 X-ray Raman Scattering (XRS)

The XR spectra of aqueous Imid and ImidHCl solutions were collected at the European Synchrotron Radiation Facility (ESRF), beamline ID20. A more complete description of the beamline can be found in Section 2.4.2.3.

Approx. 250 ml of aqueous Imid (pH 10) and aqueous ImidHCl (pH 3) solutions were prepared, as described in Section 2.2.1. The *in situ* jacketed flow crystalliser (Figure 2.8) was mounted on the sample stage. The stock solution was agitated in the vessel using a magnetic stirrer and circulated through Marprene tubing using a peristaltic pump. Recirculation allowed fresh solution to be delivered to the Kapton window to reduce the potential for radiation damage of the sample. Although the jacketed vessel could control the temperature of the solution, the spectra described here were collected at room temperature (~23°C).

The elastic energy E_0, was set to 9.68 keV. XR spectra were obtained by scanning the beamline monochromator energy at a fixed analyser setting. The measured energy difference (incident beamline monochromator energy minus analyser energy, E_0) corresponds to the energy loss. E_0 was monitored frequently throughout the data collection process to ensure that there was minimal monochromator drift.

The XRS-detected NEXAFS spectra were collected in the low momentum transfer (low-*q*) regime, which is dominated by dipole allowed transitions, in which the core-level electrons transition from the 1s (K-edge) to π* or σ* antibonding molecular orbitals [315]. Measurements presented here were performed in forward scattering (low-*q*) geometry and

collected by the VD detector module positioned at an angle of 42° and q = 3.6 Å^{-1}. The data post-processing is described in Section 2.4.2.3.2.

3.2.2 Near-Ambient Pressure XPS (NAP-XPS)

X-ray photoelectron (XP) spectra were collected with a SPECS EnviroESCA Near-Ambient Pressure X-ray Photoelectron Spectroscopy (NAP-XPS) system at the University of Leeds. A more complete description of NAP-XPS can be found in Section 2.4.3.2.

Imid (≥99.5%, Sigma-Aldrich) was used without further purification. An aqueous solution (approx. 2.5 M) was made by dissolving solid Imid in deionised water (18.2 MΩ.cm, Suez Select Fusion). The ImidHCl salt was synthesised by the evaporation processes described in Section 2.2.1 (vacuum evaporation except for initial SCXRD (rapid evaporation)).

The Imid solution was put in a Petri dish, which was placed directly into the sample chamber. Dry ImidHCl was pressed into an aluminium differential scanning calorimetry pan, as the salt is highly hygroscopic and will dissolve in the humid environment of the sample chamber. The pan was mounted on an aluminium scanning electron microscopy stub with copper tape (Figure 3.2). An open glass bottle of purified deionised water (18.2 MΩ.cm, Suez Select Fusion) was placed in the sample chamber to dissolve the ImidHCl and reduce solvent evaporation during measurements.

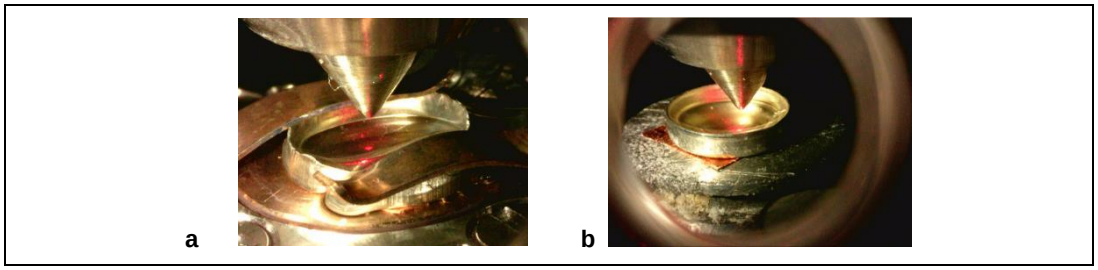


Figure 3.2 XPS samples: a) Imid and b) ImidHCl aqueous solution

Imid solution from an earlier measurement session using the Peltier cooler

XP spectra of dry and solution samples were collected at ambient temperature at a pressure of 11 mbar (see Section 2.4.3). This pressure was maintained throughout the analysis by argon gas (dry samples) or water vapour (solution samples). In this chapter, only the solution samples will be discussed. The XPS analysis of dry Imid and ImidHCl is in Appendix B.

The SPECS EnviroESCA NAP-XPS system used to collect the spectra was equipped with monochromatic Al Kα radiation source (1486.71 eV) operating at 42 W. The beam (300 µm diameter) was separated from the analysis chamber with a silicon nitride window.

The pressure in the sample chamber was reduced by pumping down to 50 mbar (fast pumping), then from 50 to 25 mbar (medium), where the pressure was held until it was seen that water degassing was complete. The sample was then transferred to the analysis chamber and the pressure was reduced to 11 mbar.

Survey scans were collected in a scan before and after the high-resolution data collection, with a step size of 1 eV, pass energy of 100 eV and dwell time of 0.1 s. High-resolution N and C 1s spectra were collected in 16 scans with a step size of 0.15/0.1 eV (aq. Imid/aq. ImidHCl), a pass energy of 50 eV and a dwell time of 0.1/0.25 s (aq. Imid/aq. ImidHCl). High-resolution O 1s spectra were collected in nine scans, with a step size of 0.15/0.1 eV (aq. Imid/aq. ImidHCl), a pass energy of 50 eV and a dwell time of 0.25 s.

Organic materials are insulating and are subject to significant charging. Charge compensation was provided by electron-ion pair formation in the surrounding water vapour atmosphere, induced by absorption of the photo- and Auger electrons emitted from the sample, alongside direct ionisation by X-ray absorption (XA) [348].

Data analysis was carried out with CasaXPS (version 2.3.18) [437]. High resolution core-level scans were deconvoluted with a U 2 Tougaard background and GL(30) lineshapes. An analyser transmission function and Scofield sensitivity factors were applied for quantitative analysis.

E_B were referenced to the primary C 1s adventitious carbon (C_{adv}) peak (285 eV) based on an established literature value [438,439]. Quantitative elemental analysis of the survey spectra indicated the amount of excess carbon in the analysis area. This information was used to deconvolute the C1s spectra by fixing the proportion of C_{adv} to that identified from the survey spectra for the initial fitting (Appendix B for details).The C_{adv} peaks were unconstrained and allowed to float as free parameters for the final refinement of peak fitting.

E_B are reported to the nearest 0.01 eV to reflect the reproducibility and precision achievable in a series of measurements on one sample under otherwise identical conditions. Absolute E_B values are generally reproducible to one significant digit, closer to 0.1 eV. Where data from the literature are referenced, they are reported with the published precision.

3.2.3 Density functional theory (DFT) and time-dependent DFT (TDDFT) calculations

Ground state molecular orbital energy calculations (DFT) and excited-state (TDDFT) calculations were carried out to predict theoretical XA spectra and inform the assignment of electronic transitions in the XR spectra.

Gas-phase optimisation: Imid and $ImidH^+$ structure geometries were optimised *in vacuo,* using DFT in ORCA quantum chemistry software (version 4.0.2) [380]. Frequency analysis was carried out after the geometry optimisation to check that the optimised structures corresponded to minima on the potential energy surface. To identify global minima, more sophisticated methods, such as molecular dynamics or metadynamics simulations, would be required to explore the conformational space. However, since Imid and $ImidH^+$ are planar, the simple optimisation carried out here is deemed sufficient.

The B3LYP functional and a minimally-augmented def2-TZVP basis set [384], with the RIJCOSX approximation [388], def2/J auxiliary basis set and D3 dispersion correction were applied [396], consistent with the previous Imid study [67]. The self-consistent field (SCF) convergence TightSCF criterion and integration grid Grid3 FinalGrid5 were applied. As $ImidH^+$ is the isoelectronic protonated counterpart of Imid, the same level of theory was applied to both structures [440]. The molecular orbitals were visualised in Avogadro (version 1.2.0) [441].

To investigate the sensitivity of transition energies to intermolecular interactions (short and long range), single point energy calculations were performed on a single ImidHCl molecular unit *in vacuo*, taken from the ImidHCl crystal structure derived from the first SCXRD analysis of ImidHCl (SCXRD details in Appendix A)[19], and the cluster of eight Imid molecules, taken from the crystal structure of Imid (IMAZOL04 from CCDC, ref. [138]). The geometric structures were extracted from the crystallographic information file from SCXRD (ImidHCl) or from CCDC (Imid). No frequency calculations were performed.

[19] SCXRD analysis of ImidHCl synthesised by rapid evaporation showed one N3 of eight in the asymmetric unit cell did not appear to be protonated, so synthesis of ImidHCl was attempted using vacuum evaporation technique (see Section 2.2.1). Refinement of this second fully-protonated SCXRD structure using XPDF/PDFgui was not possible due to disorder around two of the eight $ImidH^+$ rings (further details in Appendix A)

Solvation models

Solvation model selection: N and C K-edges are sensitive to solute-solvent interactions, which is not accounted for in the gas-phase models. In this analysis, the solvation model is required to understand the potential solvent effect on the core-level electron transitions taking place so that the pre-edge transition peaks in the XR spectra can be assigned. Although the excited-state calculations are not intended to be quantitative, consideration has been given to the type of solvation model used.

An implicit solvation model, where a dielectric constant based on the solvent is applied (known as a conductor-like polarisable continuum model, CPCM), accounts for the long-range effects of the solvent, but may be less suitable to replace a polarised solvent with charge fluctuations. An explicit solvation model, where molecules of the solvent species are positioned around the probed molecule, accounts for local solvation effects such as hydrogen bonding between the solute and solvent.

A combined liquid-jet XPS/*ab initio*/MD study of hydrated Imid showed that the calculated vertical IP of valence electrons approached the experimentally-measured value as the number of explicit water molecules was increased from 0 to 5 [442]. Increasing the number of water molecules may tend towards convergence of a bulk system in this case [442], but without the inclusion of other solute molecules (or ions and counter-ions in the case of ImidHCl), other interactions beyond water-solute interactions would not be accounted for.

A previous NEXAFS/computational study of aqueous Imid compared the effect of including different numbers of water molecules in the hydration shell on the ΔIP of N1 and N3 [159]. It was found that as the number of water molecules in the hydration shell increased above three, ΔIP decreased and was lower than the experimentally-observed energy shift between the N K-edge 1π* transition peaks ($\Delta E_{N1\text{-}N3}$ 1s→1π*), which is expected to be the same value as ΔIP [159]. Hydration models with two or three water molecules were in good agreement with experimental spectra, even though only one N centre was directly interacting with a water molecule in the solvation model in this case [159].

The possible short or long-range effects that would occur in the solution due to other species present—Imid molecules, or $ImidH^+$ and/or Cl^- ions in the case of ImidHCl—are not accounted for in the proposed simple explicit or implicit hydration models. Analysis of an MD model of an aqueous Imid solution was previously undertaken to recover the average N

and C K-edge XA spectra of all the Imid molecules.[20] It was found that there was more variation in the N1 1s→1π* transition peaks than the N3 1s→1π* transition peak due to interactions with adjacent molecules. The changes in electron density around N1 are more localised than those around N3, where the electron density is delocalised across the molecule. The MD study also identified that the C K-edge spectra were influenced by Imid-Imid interactions, but not so much by interactions with water due to the hydrophobic nature of the C moieties. This result strengthens the suggestion that increasing the number of water molecules in a microhydration cluster would not necessarily improve the results. The MD modelling of an aqueous ImidHCl solution would be more complex, as the effect of the counter-ion, the long-range ionic electrostatic effects and the solvent-solvent and the solvent-solute hydrogen bonding would be present.

A neutron diffraction pair distribution function analysis [223] showed that in aqueous solution, the mostly likely Imid self-interaction is via π-π stacking between the hydrophobic planes, at a range of ~3.5-3.8 Å. Most Imid molecules were found to be fully hydrated, resulting in few direct Imid-Imid interactions. Although π-π stacking of Imid does occur in solution, these interactions are secondary relative to Imid-solvent hydrogen bonding [223]. Due to the separation distance of the Imid molecules involved in stacking, the likely π-π effect on the core-level spectra will be less than the effect of hydrogen bonding with water, and therefore are unlikely to greatly affect 1s→π* transition energies. Previous modelling of Imid solutions has also shown a tendency of Imid to form chains [443]. Given the sensitivity of Imid to its conformational surroundings [159], the effect of these possible Imid self-interaction makes it difficult to define a model that will show good agreement with experimental data.

A similar stacking effect may occur in the aqueous ImidHCl solution. The $ImidH^+$ and Cl^- ions present in solution would generate local electrostatic effects on each other and on the (polar) solvent, which in turn would provide some charge shielding between the ions. In addition to possible oppositely-charged ion pair formation ($ImidH^+$-Cl^-), like-charged ion pairs ($ImidH^+$-$ImidH^+$) may form. Guanidinium, another planar cation with a low charge density, is known to form parallel stacks (dimers and larger n-mers) that are stabilised by water molecules [91]. The proposed water hydrogen-bonding arrangement that may stabilise guanidinium dimers could apply to $ImidH^+$ ions in solution. Investigating hydration

[20] MD modelling was undertaken by another researcher in the group and is currently not available for publication.

models to this extent is beyond the scope of this study, as it requires the modelling of the total scattering data.

In summary, increasing the number of components in the model would not necessarily improve the outcome. Whilst it is noted that a better outcome may be achieved when the longer-range solvent effects are considered [159], a systematic study comparing the use of CPCM, single and double-water molecule explicit hydration models, and a hybrid model (two water molecules and the CPCM solvent background), found that the two-water explicit model generated N and C K-edge XA spectra that were most comparable to the experimental data [67]. Therefore, it was considered pragmatic to adopt the same approach here.

Solvation model optimisation: In this study, the water molecules in the explicit models were initially positioned adjacent to the N atoms, as indicated in the literature for microhydrated gas-phase clusters for Imid [67], and $ImidH^+$ [442] (Figure 3.3). The position of the water molecules was geometry-optimised, after which single point energy molecular orbital calculations were carried out (ORCA version 4.0.2) [380]. Calculations for implicit hybrid solvation model structures were also performed. Frequency calculations showed imaginary modes were present in the $ImidH^+$ model, which were not eliminated after multiple geometry adjustments, so the hybrid models were discontinued.

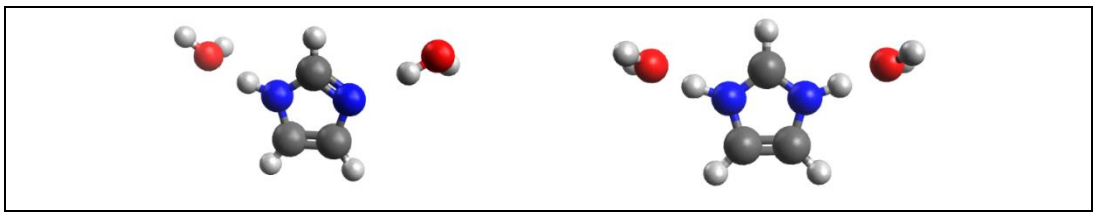


Figure 3.3 Explicit solvation models of imidazole (l) and imidazolium (r)

Visualisation of the positions of the water molecules during the TDDFT calculations

TDDFT calculations: TDDFT calculations were run on the gas-phase components, the implicit and explicit solvation models, the ImidHCl molecular unit and N atoms in a central and external molecule in the eight Imid cluster using ORCA version 4.0.2. [380].

A summary of the DFT / TDDFT calculations and the applied functional/basis sets is shown in Table 3.1.

Table 3.1 DFT and TDDFT calculations

• **Gas-phase Imid / ImidH$^+$** • **Explicit solvation model**: Imid or ImidH$^+$ with two water molecules • **Implicit solvation model**: Imid or ImidH$^+$ with water background (CPCM)	**Optimisation / single point energy / TDDFT:** B3LYP RIJCOSX D3 ma-def2-TZVP def2/J TIGHTSCF Grid3 FinalGrid5
• **Single ImidHCl molecular unit** • **Solid structure**: Imid (8 molecules from crystal structure in CCDC, ref. [138])	**Optimisation:** Crystallographic structures used to maintain the intra- and intermolecular distances, which is crucial to acquiring reliable core-level data **Single point energy / TDDFT:** B3LYP RIJCOSX D3 ma-def2-TZVP def2/J TIGHTSCF Grid3 FinalGrid5

3.3 Results and discussion

As Imid and ImidH$^+$ are aromatic molecules with significant π bond conjugations, significant pre-edge π* resonances on the near-edge spectra of Imid and ImidH$^+$ are expected [207]. Comparing the N K-edge spectra for Imid and ImidH$^+$ (Figure 3.4), there are two clear pre-edge peaks for Imid (399.8 eV and 401.6 eV) representing the two N species, but only one for ImidH$^+$ (401.3 eV), where the protonation of the Imid results in a symmetrical molecule (N pseudoequivalence). The N K-edge spectrum appears steeper after the pre-edge features in the Imid spectrum than for ImidH$^+$, and a small shoulder feature is visible ~404 eV in the ImidH$^+$ spectrum. Comparing the Imid and ImidH$^+$ C K-edge spectra, there is a small shift (+0.2 eV) in the pre-edge peak position, which below the instrument resolution, but the shapes of the spectra are broadly similar.

The TDDFT-calculated XA spectra are used to allocate the pre-edge electron transition peak positions (Section 3.3.3) and the XPS E_B values are used to define the relative IP positions (Section 3.3.2).

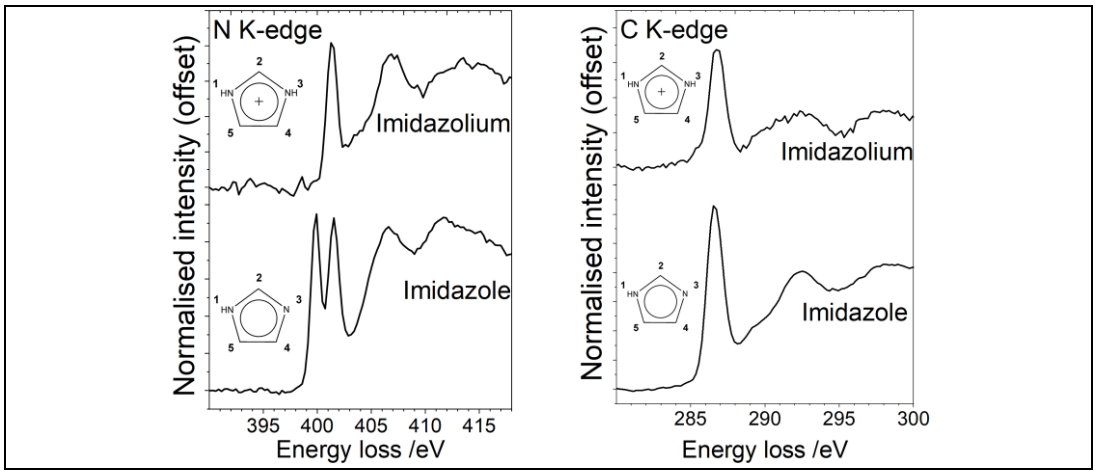


Figure 3.4 Comparison of the Imid and ImidH$^+$ XRS N and C K-edge spectra

3.3.1 DFT and TDDFT calculations

The ground state energies of atomic core levels, and the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbitals (LUMO) energies, were derived from DFT single point energy calculations. Schematic representations of molecular orbital energies of N and C K-edges are shown for gas-phase Imid and $ImidH^+$ (App. Figure 14), and for explicitly-solvated Imid and $ImidH^+$ (App. Figure 15).

In gas-phase Imid there are two low-lying virtual π* orbitals available for electronic transition from the core-level (1s). Visualisation of the two lowest unoccupied π* molecular orbitals (Figure 3.5a&b) indicate the electronic transitions that may be observed.

Analysis of the calculated XA spectra from TDDFT also indicate that in Imid, there is no density of states for a 1s→2π* N1 transition. In $ImidH^+$, there are no 1s→2π* C2 transitions, and only very small C5 and N1 transitions. The density of states also indicates that the 1s→1σ* electronic transition would be weak, and are not likely to be evident in the experimental spectrum [159].

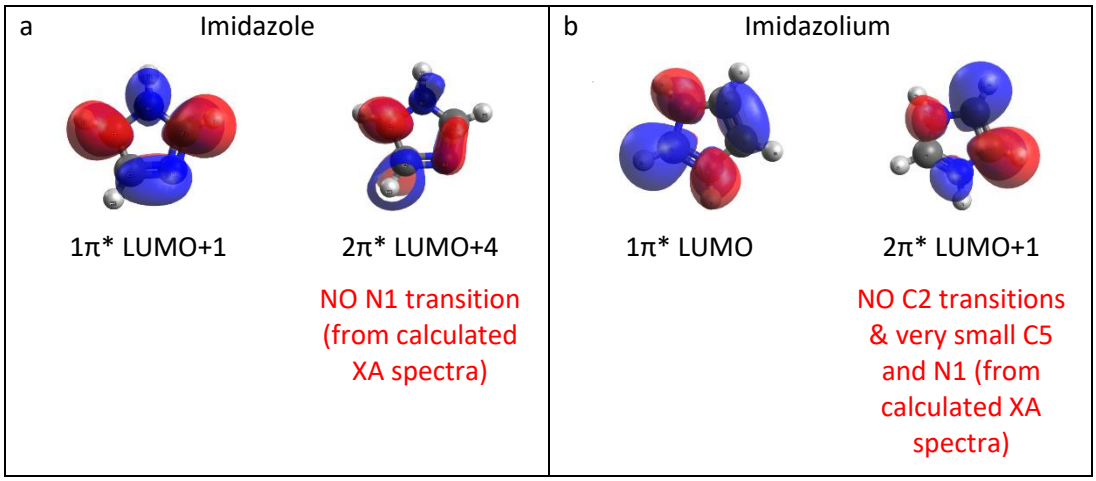


Figure 3.5 Molecular orbital structures of gas-phase a) Imid and b) $ImidH^+$

In explicitly-solvated Imid, there are two low-lying virtual π* orbitals available for 1s→π* transitions, compared with three for solvated $ImidH^+$. As seen in the gas-phase calculation, there is no density of states for an N1 1s→2π* transition in explicitly-solvated Imid. There are no C2 1s→2π*, or N1/N3 1s→3π* transitions in explicitly-solvated $ImidH^+$.

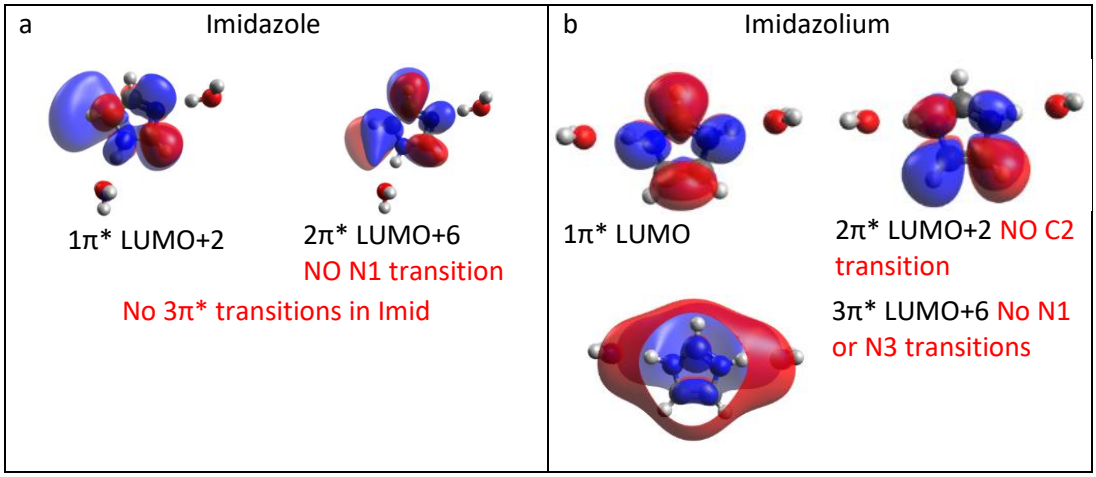


Figure 3.6 Molecular orbital structures of explicitly-solvated a) Imid and b) $ImidH^+$

The predicted XA spectra for the N and C K-edges of Imid and $ImidH^+$ are shown along with the experimental data (Figure 3.7 & Figure 3.8). The calculated spectra peaks have been aligned with the first peak of the experimental spectra.

ΔEN1-N3 1s→1π* is greater in gas-phase Imid than in solvated models and experimental data (Figure 3.7a), showing that hydrogen bonding of the Imid with the water via the N atoms reduce the difference in the π* energy levels. In solution, the lone pair on the N3 of Imid (-N) accepts a hydrogen bond from the polar water solvent [195]. In UV spectroscopy of solutions, this interaction has previously been observed to have the effect of lowering the electronic energy in the ground state, but not in the excited state [195]. This effect is observed here as well, where the N3 1s molecular orbital energy is lowered due to the removal of local electron density by the electrostatic attraction of the hydrogen from the water molecule (solvated ground state calculations in App. Figure 15). At N1, the hydrogen bond formation is due to proton donation to the oxygen moiety of water, increasing the local (N1) electron density. The polarisation of the N-H bond is not delocalised into the π system, so the change in electron density at the N1 centre is stronger than at the N3 centre [67]. As a result of these effects, the difference between the N1 and N3 1s→1π* transition energies (ΔE N1-N3 1s→1π*) is calculated as 2.05 eV for the gas-phase Imid, 1.68 eV for solvated Imid (Table 3.6), and 1.40 eV for solid-phase Imid [67], following the expected trend [67].

The gas-phase spectrum clearly shows the N3 2π* transition peak (~401 eV). This peak appears as a shoulder feature in the explicit model spectrum, and as a peak slightly below the experimental N3 1s→2π* transition in the implicit model.

In the C K-edge (Figure 3.7b), the C4 2π* transition in the explicit spectrum is under the first main peak, and the C5 2π* can be seen as a shoulder of the main peak (287.7 eV). In the implicit spectrum, C4 and C5 2π* transitions account for the shoulder feature (287.9 eV). In all models, the C2 2π* is ~2 eV above the main peak.

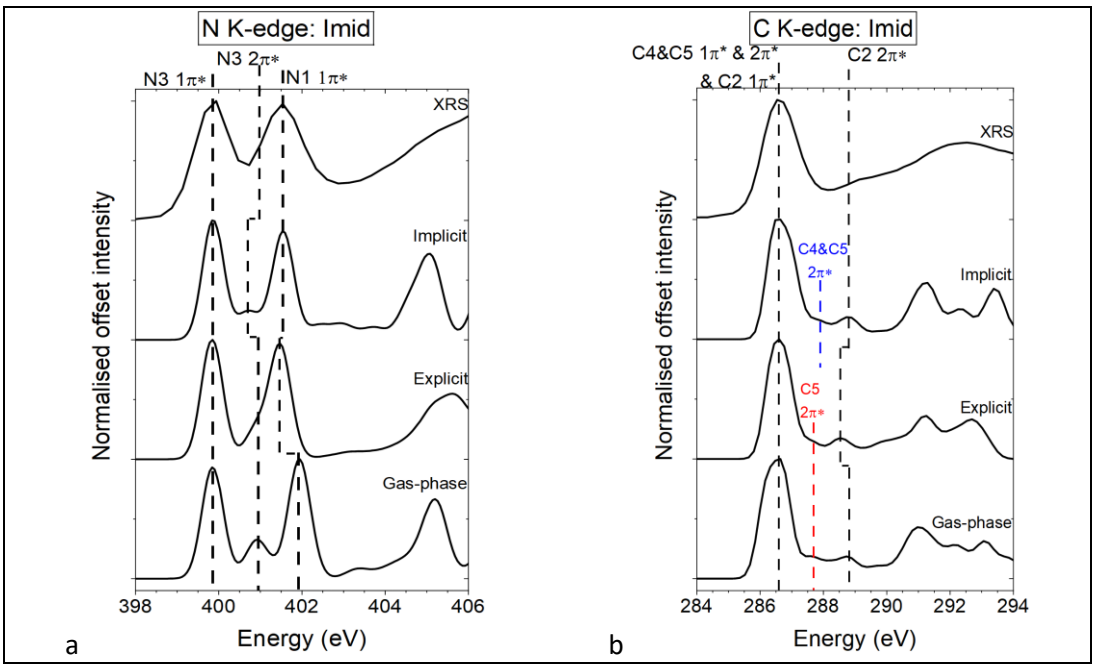


Figure 3.7 Imid: calculated a) N and b) C K-edge XA spectra of gas-phase, and explicit and implicit solvation models. FWHM of predicted XA spectra = 0.6 eV and dotted lines indicate the peak positions of the fitted experimental data.

As both N atoms in $ImidH^+$ are -NH, the N1 and N3 1s→1π* transitions are (almost) equivalent, at 389.22 eV and 389.22 eV (gas-phase) and 388.91 eV and 388.88 eV (in solution, Table 3.6 (note: 12.5 eV energy shift applied in the table to enable comparison with the experimental data)).

The explicit model has been used for peak assignment, as it was considered to give the best agreement for Imid [67], and has good agreement for ImidHCl. In the explicitly-solvated model for Imid, consecutive π* transitions occur at LUMO+2 and LUMO+6, with no transition for N1 in LUMO+6 (Figure 3.9). The second, smaller N3 peak at ~400.9 eV is consistent with the calculated spectrum shown in the literature (Fig. 4 in ref. [67]) for the Imid N K-edge ISEELS gas-phase spectrum (from ref. [444]) and XRS aqueous solution spectra.

The accuracy of the predicted XA spectra and the TDDFT calculations could be improved if it were possible to include the effect of the Cl^- ions that are also in solution, such as long-range solvent or electrostatic effects. The solvation shell around the Cl^- ions in solution could affect the polarisation of the surrounding solvent. To explore the effect of including a Cl^- ion, TDDFT calculations were performed on a single unit of ImidHCl (Figure 3.10).

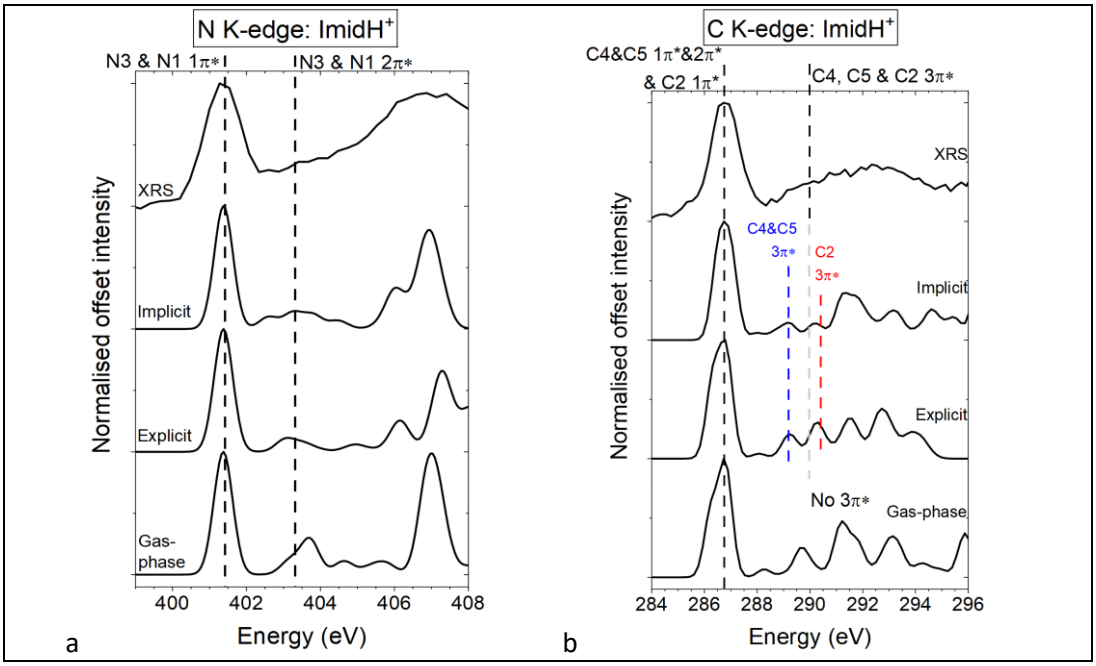


Figure 3.8 ImidH$^+$: calculated a) N and b) C K-edge XA spectra of the gas-phase, and explicit and implicit solvation models. FWHM of predicted XA spectra = 0.6 eV and dotted lines indicate the peak positions of the fitted experimental data.

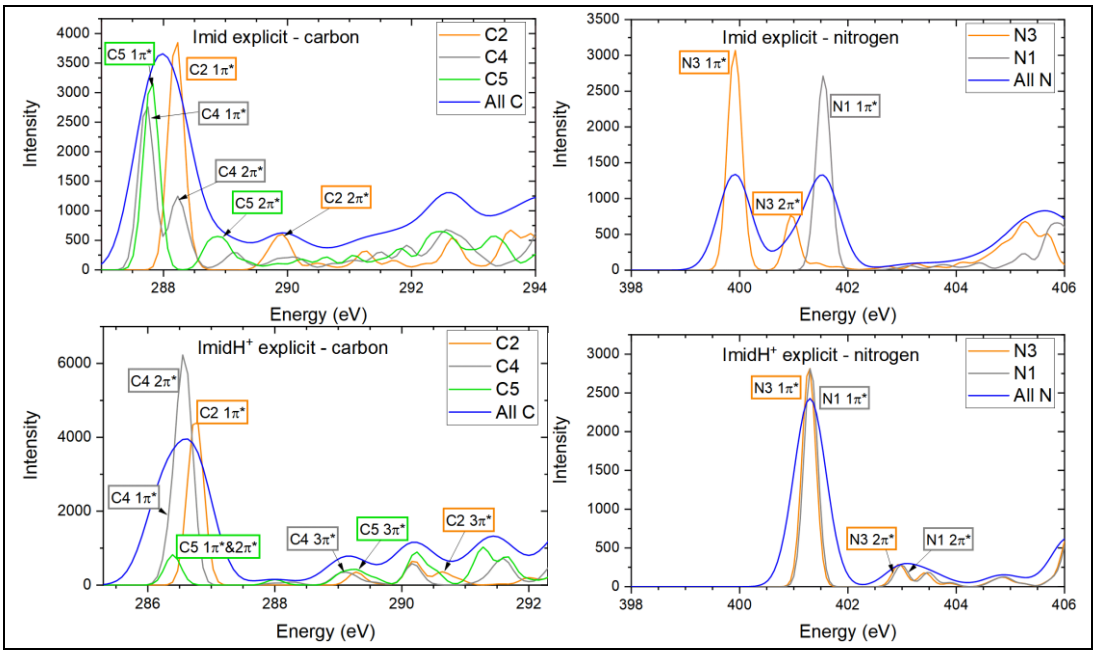


Figure 3.9 Explicit solvation model predicted XA spectra: Imid C K-edge (top left) and N K-edge (top right); and ImidH$^+$ C K-edge (bottom left) and N K-edge (bottom right)

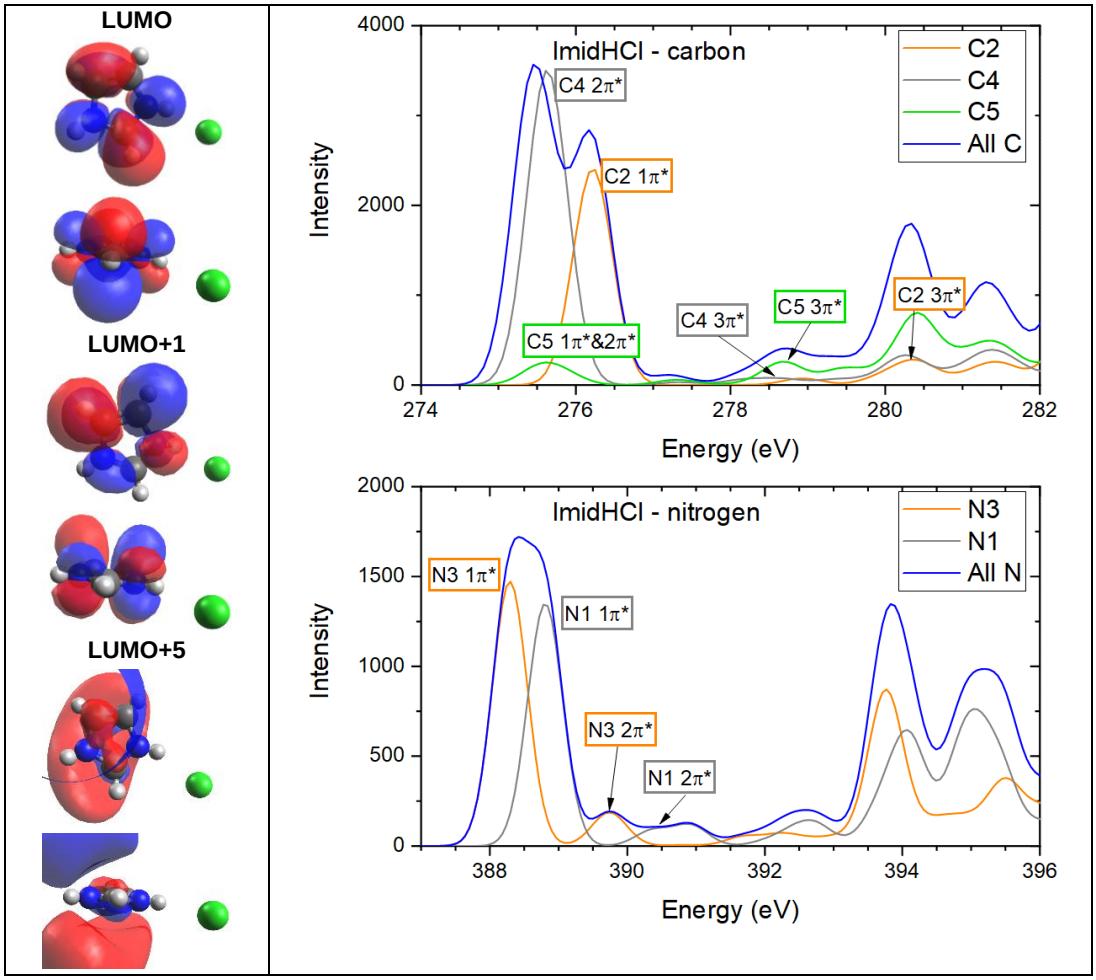


Figure 3.10 Predicted XA spectra C K-edge (top) and N K-edge (bottom) for ImidHCl (single unit, *in vacuo*)

The first C K-edge pre-edge peak in ImidHCl is broader (FWHM ~1.3 eV, Figure 3.10, top), with a greater ΔE between the C2 1π* and C4 2π* transitions (0.6 eV) than that for $ImidH^+$ (FWHM ~1.0 eV, ΔE (C2 1π* - C4 2π*) = 0.2 eV, Figure 3.9, bottom left). The C4 1π* transition present in $ImidH^+$ is not present in ImidHCl. Furthermore in the N K-edge spectrum, $\Delta E_{(N1-N3)}$ 1s→1π* transition peak separation is 0.5 eV in ImidHCl (Figure 3.10, bottom), resulting in a broader pre-edge peak (FWHM 1.0 eV) than seen either of the gas-phase models (Figure 3.11)[21].

[21] All three XA spectra were generated with a FWHM of 0.6 eV

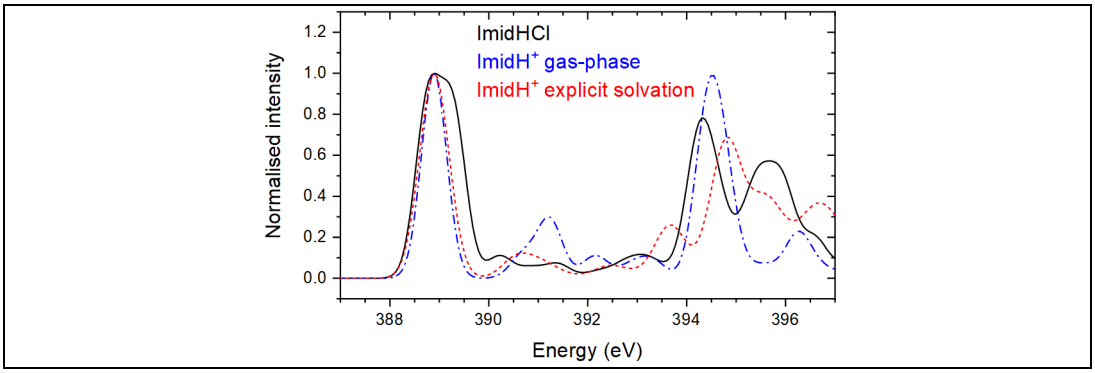


Figure 3.11 Predicted XA spectra N K-edge for $ImidH^+$ (gas-phase and explicit solvation models) and ImidHCl (molecular unit), aligned to the explicit model first peak

In the ImidHCl crystal structure, the $ImidH^+$ would be surrounded by other Cl^-, which may reduce the $\Delta E_{(N1\text{-}N3)}$ 1s→1π* seen in the single molecular unit, as observed in the solid Imid structure where $\Delta E_{(N1\text{-}N3)}$ 1s→1π* for the external Imid was 0.88 eV higher than for the central Imid (Appendix C).

It may be however that in solution, $\Delta E_{(N1\text{-}N3)}$ 1s→1π* of $ImidH^+$ could be indicative of the extent of ion pairing, as the solution is cooled through the MSZ prior to crystallisation. Molecular modelling to identify $ImidH^+/Cl^-$/water positions in larger clusters, followed by TDDFT, could be carried out to investigate this further. ImidHCl could not be crystallised from water by cooling crystallisation within the temperature range accessible by the experimental setup, so this study was not undertaken during the XRS beamtime. However, the characterisation of another organic (possibly imidazolium-based) salt system in solution passing through the metastable zone during cooling crystallisation, by applying the combined XRS-NEXAFS-XRS method with TDDFT described here, is something that could be explored further.

The core-level transitions for solvated Imid and $ImidH^+$ are detailed further in the following N and C K-edge XRS NEXAFS analysis (Section 3.3.3).

3.3.2 X-ray photoelectron spectroscopy

Imid and ImidHCl in solution were characterised with NAP-XPS to identify the chemical states of N and C atoms in Imid and $ImidH^+$. The relative chemical shifts between the photoemission lines were used to assign the IP energies in the XRS-derived NEXAFS spectra.

Survey scans for hydrated Imid and ImidHCl were measured before and after the high-resolution scans. Quantitative elemental analysis of the survey spectra was used to indicate the amount of excess carbon in the analysis area. This information was used to deconvolute the C1s spectra by fixing the quantity of C_{adv} (285 eV) to that identified from the survey

spectra. Variations between the two scans, indicating different proportions of C and N, may be due to the sample moving in the X-ray beam during data collection. Furthermore, the distance between the nozzle and the sample could change during data collection, as there could have been small changes in the analysed surface because of the nature of the sample.

Table 3.2 Measured contamination based on quantitative elemental analysis of the survey spectra

Compound	Survey No.	%C	%N	Area ratio C_{adv}:C:N			% area C_{adv}	
							Survey	High-res
Aq. Imid	1	60.51	39.49	0.07	3	2	2.1%	45.6
	2	22.57	9.12	1.95	3	2	39.4%	
Aq. ImidHCl	1	82.61	17.39	6.50	3	2	68.4%	65.3
	2	82.23	17.77	6.26	3	2	67.6%	

Measuring samples under ultra-high vacuum (UHV) decreases both signal attenuation, improving the signal/noise ratio, and the contamination layer on the sample [436]. A much higher proportion of C_{adv} in the C 1s signal is expected with NAP-XPS than with UHV-XPS.

C_{adv} represents a large proportion of the signal in both C 1s spectra[22]. Although the contamination (C_{adv}) area percentage was initially fixed in the peak fitting process according to the second survey scans (Table 3.2), it was unconstrainted in the final fit, recognising that the quantitative analysis suggested a changing contribution over the course of the data collection. This estimate based on excess C, over and above the stoichiometric ratio in $ImidH^+$, assumes that no N contamination is present.

The core-level C 1s and N 1s XP spectra and spectral features for aqueous Imid and ImidHCl are presented in Figure 3.12 and Table 3.3.

Table 3.3 Measured binding energies of aqueous Imid and ImidHCl - corresponding spectra are in Figure 3.12

Compound	Transition	Peak energy /eV	ΔE /eV	
Aqueous Imid	N 1s (N3)	399.30	(N1-N3) =	*1.65*
	N 1s (N1)	400.95		
	C 1s (C4)	285.44	(C5-C4) =	0.30
	C 1s (C5)	285.74	(C2-C5) =	0.87
	C 1s (C2)	286.61	(C2-C4) =	*1.17*
	C 1s adventitious	285.00		
Aqueous ImidHCl	N1 & N3 1s	401.57		*0.00*
	C 1s (C4)	286.37	(C5-C4) =	0.00
	C 1s (C5)	286.37	(C2-C5) =	1.31
	C 1s (C2)	287.68	(C2-C4) =	*1.31*
	C 1s adventitious	285.00		

[22] C 1s spectra collected by NAP-XPS indicated the presence of C_{adv} in a deionised water sample, which had been stored in the Suez unit. C_{adv} was also observed in a bottled mineral water sample in the literature [448].

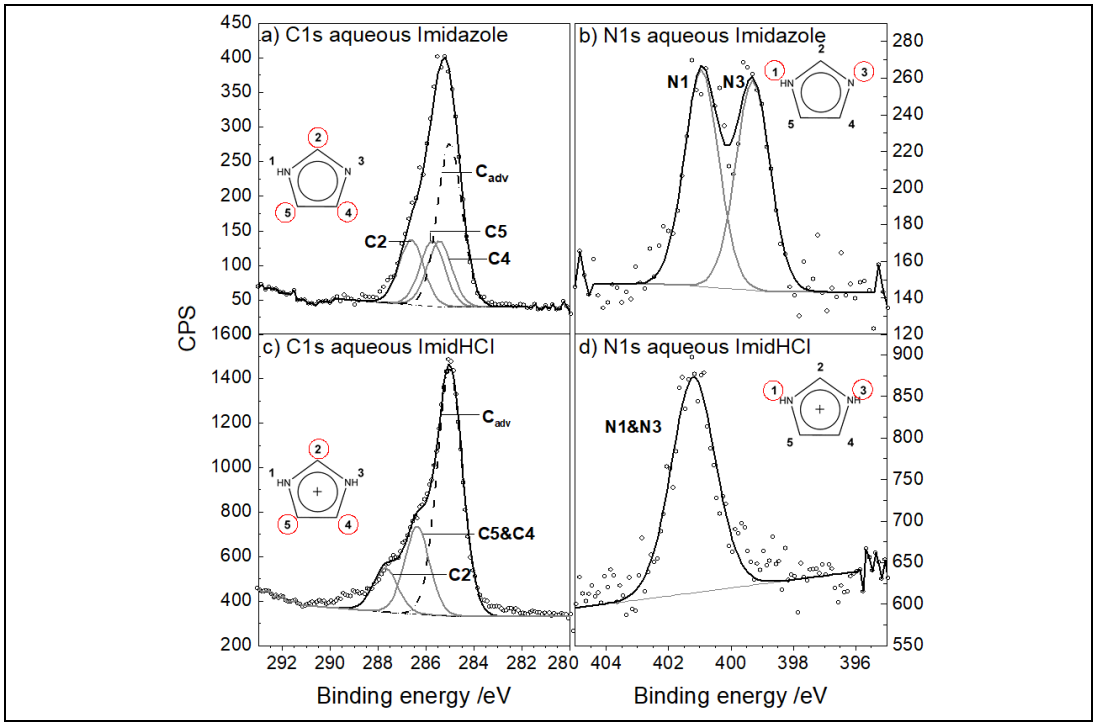


Figure 3.12 Fitted XP spectra of a) C 1s aq. Imid b) N 1s aq. Imid c) C 1s aq. ImidHCl and d) N 1s aq. ImidHCl

Open circles: Experimental data points, black line: fitted spectra, grey line: background

In the case of aqueous Imid, the protonated and unprotonated N atoms were identified in the N 1s spectrum (Figure 3.12b). These E_B peaks were separated by $\Delta E_{B(N1-N3)}$ = +1.65 eV, which is in agreement with a value previously reported from a liquid-jet XPS study of aqueous Imid (reported to 1 decimal place, as 1.7 eV [160]). The C 1s spectrum of aqueous Imid revealed a higher E_B attributed to C2 in each system, as it is bonded to two N atoms that are more electronegative [445]. The energy shifts of $\Delta E_{B(C2-C5)}$ = +0.87 eV and $\Delta E_{B(C5-C4)}$ = +0.30 eV (Figure 3.12a) deviate from the liquid-jet study, in which the E_B for C4 and C5 of hydrated Imid are separated by 0.4 eV, with C2 E_B 0.8 eV higher ($\Delta E_{B(C2-C4)}$ = +1.2 eV) [160]. This difference is likely due to superposition with the intense adventitious carbon peak, which was not present in the liquid-jet spectra, and it is expected to have some components from C bound to OH groups in the range of the C2, C5 and C4.

In the case of aqueous ImidHCl, one feature is observed in the N 1s spectrum (Figure 3.12d). When both N atoms are protonated, the net positive charge is delocalised across the aromatic system and increases the N 1s E_B of both N in $ImidH^+$. At 401.57 eV (Table 3.3), the combined N1/N3 peak in $ImidH^+$ is shifted by +0.62 eV relative to the protonated N1 (-NH) component (400.95 eV) and by +2.27 eV relative to the unprotonated N3 component (399.30 eV) in Imid. Similar E_B shifts upon protonation of Imid have been observed in a liquid-jet XPS study of aqueous Imid and $ImidH^+$ solutions (2.7 eV) [160]. An energy shift of +2.3 eV was

observed upon protonation of N3 in the Imid side chain of theophylline (-N) when it formed a salicylic-5-sulfonate monohydrate salt (-NH$^+$), whilst the E_B of the N1 moiety (-NH) shifted by +0.90 eV [446]. The C4 and C5 E_B of ImidH$^+$ occur at the same E_B, with $\Delta E_{B(C2-C4)}$ = +1.31 eV, whilst the liquid-jet XPS study also found the ImidH$^+$ C4 and C5 E_B to be equal, but with a smaller $\Delta E_{B(C2-C4)}$, at +1.2 eV [160].

Interaction of the water with the samples is evidenced in the O 1s spectra (Figure 3.13). Both spectra have been calibrated with reference to C_{adv}.

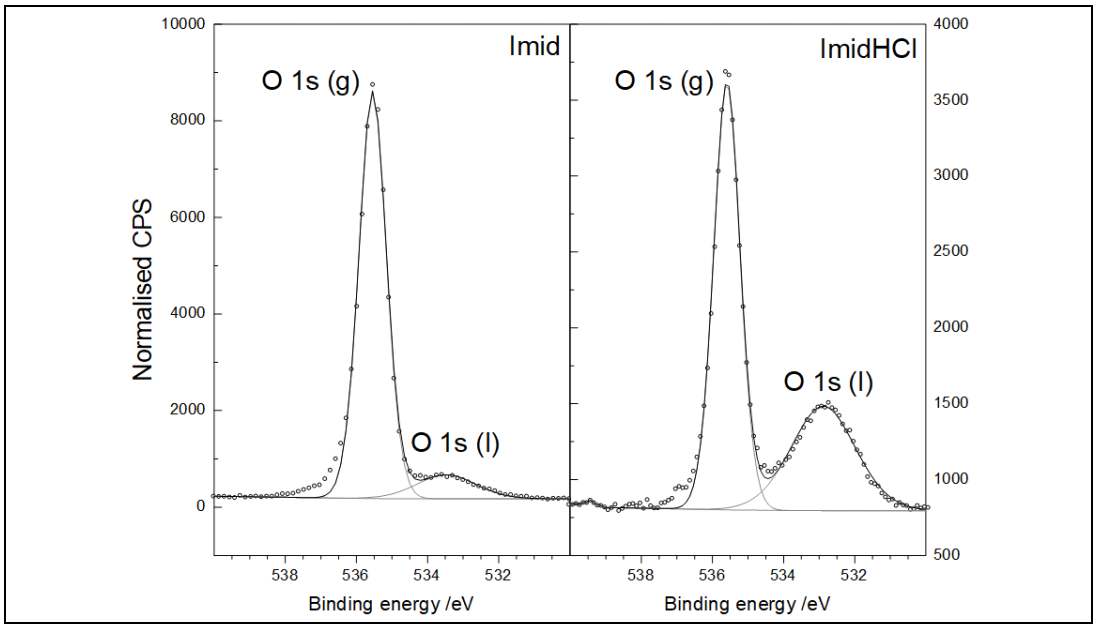


Figure 3.13 XP O 1s spectra a) aq. Imid b) aq. ImidHCl. (Note: Normalised CPS scales are not equal)

Table 3.4 Measured O 1s binding energies of aqueous Imid and ImidHCl - corresponding spectra are in Figure 3.13

Compound	Peak name	E_B /eV	FWHM*	% conc	ΔE /eV
Aqueous Imid	O 1s (g)	535.54	0.90	88.63	2.07
	O 1s (l)	533.47	1.96	11.37	
Aqueous ImidHCl	O 1s (g)	535.60	0.87	62.58	2.72
	O 1s (l)	532.88	2.13	37.42	

* full width half maximum

The lower intensity of the O 1s (l) peak for aq. ImidHCl could be due to fewer water molecules in the solution sample than in ImidH$^+$. Additionally, the nozzle working distance could affect the intensity ratio of the O 1s (g):O 1s (l) peaks. It was difficult to adjust the working distance with precision due to the changing level of the sample.

The smaller water peaks originate from a combination of water molecules that are solvating Imid, or ImidH$^+$ and Cl$^-$, that are hydrogen-bonded to other water molecules in the solution,

or that are interacting with contamination in the solution. The higher-energy peak is mainly due to water vapour. Any contribution from ionised water products is likely to be small.

These O 1s E_B values are compared to O 1s data in the literature in Table 3.5.

Table 3.5 Comparison of O 1s data

Sample (Ref.)	O 1s (g)		O 1s (l)		ΔE_B /eV
	E_B /eV	FWHM /eV	E_B /eV	FWHM /eV	
Imid	535.54	0.90	533.47	1.96	2.07
ImidH⁺	535.60	0.87	532.88	2.13	2.72
Liquid-jet [160,447] *Vacuum, 277 K*	-	-	538.1	-	-
Liq. water NAP-XPS [448] 3 mbar, 300 K	535.8	0.63	533.6	2.69	2.2
Gas-phase NAP-XPS [449] 6 mbar, 300 K	536.0	0.68	-	-	-
Bacteria NAP-XPS [343] 11 mbar	535.4	-	533.1	1.2	2.3
Fe_2O_3 [450] 1.3 mbar	535.5	-	533.2	-	2.3

The O 1s (g) peaks for aq. Imid and $ImidH^+$ fall within the range observed in the literature.

The difference between these E_B values, 0.06 eV for O 1s (l) and 0.59 eV for O 1s (g), suggest that the calibration method using C_{adv} is reasonably consistent for finding the relative energy shifts between the scans, although the absolute value may not be completely accurate.

It is worth noting here that the E_B values in the liquid-jet study of Imid and $ImidH^+$ solutions were calibrated using the O 1s photoelectron peak of water [160], stated to be 538.1 eV [447], ~5.2 eV above the values presented in this study (Table 3.3). Although the absolute E_B values in this study and the liquid-jet study are different, liquid-jet E_B data [160] are around +4.9 eV higher than the NAP-XPS data presented here, and $\Delta E_B(N3_{Imid}\text{-}N3_{ImidH+})$ = 2.27 eV compared to 2.8 eV for NAP-XPS. $\Delta E_{B(N3\text{-}C2)}$ for the Imid and $ImidH^+$ solutions are 112.2 eV and 114.2 eV (liquid-jet [160]), and 112.7 eV and 113.9 eV (with NAP-XPS). These findings suggest that further refinement of the NAP-XPS solution measurement procedure could result in a solution E_B measurement technique that is more accessible, but equally as effective, as synchrotron-based measurements.

3.3.3 X-ray Raman spectra

3.3.3.1 N K-edge XRS NEXAFS

In Figure 3.14, the N K-edge NEXAFS spectra of 5 M aqueous solutions of Imid (bottom spectrum) and ImidHCl (top spectrum) are compared. The peak positions of the fitted experimental spectra and the calculated TDDFT spectra for the explicitly solvated structure

models of aqueous solutions of Imid and ImidHCl are presented in Table 3.6 (an energy shift of 12.3 eV (Imid) and 12.5 eV (ImidHCl) have been applied to the calculated transitions).

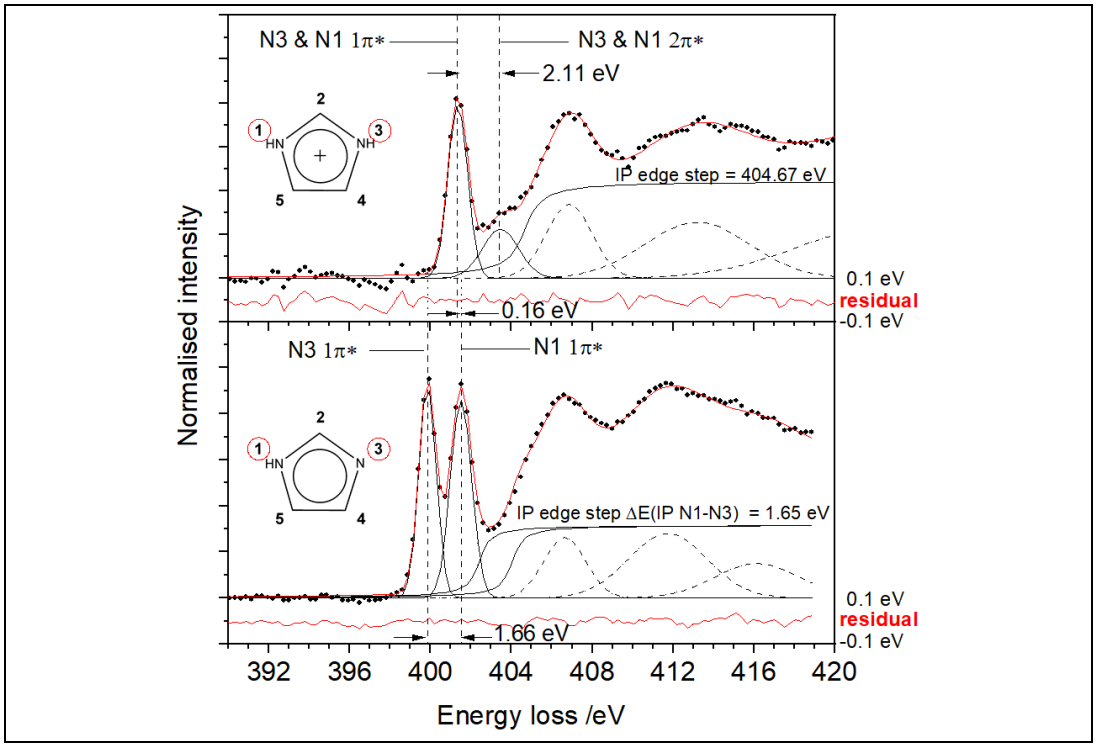


Figure 3.14 Experimental N K-edge XR spectra for 5 M aqueous Imid solution (pH10) – bottom; and 5 M aqueous ImidHCl solution (pH3) - top.

Experimental (black dots) overlaid with fitted spectrum (red). Gaussian pre-edge 1s→π* transition peaks (solid Gaussian), IP arctan function (solid) and post-edge σ*-shape resonances (dashed Gaussian).

In the Imid spectrum, the pre-edge region is dominated by two sharp Gaussian peaks, representing the 1s→π* transitions of the core-level electrons in the N moieties. TDDFT calculations show that the lowest energy peak (399.85 eV) represents the N3 1s→1π* transition, and the second peak is a convolution of the N1 1s→1π* and N3 1s→2π* transitions (401.53 eV) (see Section 3.3.1). The IPs are of equal step size due to the one-to-one ratio of nitrogen moieties in the Imid molecule, with a constrained energy shift according to the measured N3 and N1 XPS binding energies (see Section 3.3.2). The σ*-shape resonances after the IP are of increasing full width half maximum (FWHM), as expected [207].

There is no density of states at the N1 atom representing 2π* transition either in the gas-phase LUMO+2 (Figure 3.5) or the explicitly-solvated model LUMO+6 (Figure 3.6), which excludes a significant N1 1s→2π* transition. The measured energy difference between the N3 and N1 1s→π* centroids is 1.66 eV (Table 3.3). This energy shift value is consistent with

a previous value of ~1.7 eV for a 5 M aqueous solution of Imid, measured by optical luminescence NEXAFS (Table 3.7) [159].

Table 3.6 Peak position and assignment of the experimental and calculated N K-edge spectra. The corresponding experimental spectra are shown in Figure 3.14.

†The energy shifts of 12.3 eV (Imid) & 12.5 eV (ImidHCl) have been applied to the calculated transitions presented here. *Under N1 1s→1π* peak

	Imid solution			**ImidHCl solution**		
	Exp. spectra /eV	Calc.† /eV	Diff. (exp-calc) /eV	Exp. spectra /eV	Calc.† /eV	Diff. (exp-calc) /eV
E (N1 1s→1π*)	401.51	401.53	-0.02	401.35	401.38	-0.03
E (N3 1s→1π*)	399.85	399.85	0.00	401.35	401.35	0.00
ΔE (N1-N3 1π*)	1.66	1.68		0.00	0.03	
E (N1 1s→2π*)	-	-		403.46	403.06	0.40
E (N3 1s→2π*)	401.51*	400.90	0.607	403.46	403.02	0.44
ΔE (N1-N3 2π*)	-	-		0.00	0.04	
ΔE (N1 2π*-1π*)	-	-		2.11	1.68	
ΔE (N3 2π*-1π*)	1.66	1.05		2.11	1.67	
IP (N1)	404.05	403.31	0.74	404.67	408.23	-3.56
IP (N3)	402.40	401.56	0.84	404.67	408.19	-3.53
ΔIP (N1-N3)	1.65	1.75		0.00	0.04	
1σ* (C–N)	406.65			406.86		
2σ* (C–N)	411.75			413.20		
δ1 (1σ*-IP(N3))	4.25			2.19		
δ2 (2σ*-IP(N3))	9.35			8.54		
δ3 (1σ*-IP(N1))	2.60			2.19		
δ4 (2σ*-IP(N1))	7.70			8.54		

Comparison of the energy shift between the N K-edge 1π* transition peaks ($\Delta E_{N1\text{-}N3}$ 1s→1π*) in the spectra of solid Imid (1.425 eV, measured by XRS[67]), a 10 M aqueous Imid solution (1.624 eV, measured by XRS [67]) and gas-phase Imid (2.444 eV, measured by inner shell electron energy loss spectroscopy, ISEELS [444]) demonstrates the sensitivity of the N K-edge pre-edge region of Imid to molecular interactions, be that Imid-water or Imid-Imid interactions. Thus as the likely number and proximity of Imid-Imid interactions increases, $\Delta E_{N1\text{-}N3}$ 1s→1π* decreases.

Table 3.7 Comparison of measured 1s→1π* transition energy shifts.

Phase	**Technique**	**Energy shift /eV**
Gas-phase [444]	ISEELS	2.444
Solution (5 M) [159]	XAS	~1.7
Solution (5 M) (this study)	XRS	1.66
Solution (10 M) [67]	XRS	1.624
Solid [13]	XRS	1.425

Compared to the gas-phase (2.444 eV) [444], the hydrogen bonding of Imid solution with strongly polar water (N3···H-O or N1-H···O) would result in the increased core-level E_B of N3 and reduced E_B of N1, thus decreasing $\Delta E_{N1\text{-}N3}$ 1s→1π*(1.66 eV at 5 M), as shown in Table 3.7.

$\Delta E_{N1\text{-}N3}$ 1s→1π* in the 5 M solution (1.66 eV) is greater than in 10 M solution or solid Imid (1.62 eV and 1.42 eV) [65]. The number of water molecules per Imid molecule is 11 in the 5 M solution, compared to < 4 in the 10 M solution [65,221]. It was identified that there are two hydrogen-bonded water molecules in the inner hydration shell of each Imid molecule [65]. There are more free water molecules, and consequently a greater separation distance between Imid molecules, in the 5 M solution [65]. As such, the increased self-interaction of Imid molecules in the 10 M solution results in $\Delta E_{N1\text{-}N3}$ 1s→1π* between that of the 5 M solution and the solid. Based on this experimental evidence, it is proposed that self-association of Imid in aqueous solution is not favoured over hydration, because the Imid-Imid hydrogen-bonding interactions (N···H-N) are weaker than the solute-solvent interactions. The reduced $\Delta E_{N1\text{-}N3}$ 1s→1π* (1.42 eV) in solid-state Imid is not an indication of stronger Imid-Imid hydrogen-bonding, but instead could relate to N1-H hydrogen-bonding to a delocalised system via N3 in the solid (N1H···N3, with each partially delocalised Imid monomer in the chain acting simultaneously as H-donor, via N in sigma-bonded N1-H, and an H-acceptor, via delocalised N3), compared to the N moieties hydrogen-bonding to two separate (highly-electronegative) water molecules when in solution. As a result, the shifting but delocalised electron densities within the Imid chain approach parity and a lower 1s→1π* split than for Imid in solution.

The FWHM of the 1π* transition peaks of Imid in 5M solution are 1.0 eV (N3) and 1.2 eV (N1), which is slightly higher than the equivalent peaks in the 10 M solution (~0.9 eV for N3 and ~1.1 eV for N1, for 10 M solution at 20°C, cf. SI of ref. [67]). The difference in FWHM of the two sets of measurements are too small relative to the instrument broadening to draw any conclusions.

In contrast, the ImidHCl N K-edge spectrum is dominated by a single sharp Gaussian peak at 401.35 eV, which is a convolution of the 1s→1π* transitions of N3 and N1 (both -NH) (Figure 3.14, bottom). The protonation of N3 results in a symmetrical molecule, in which the environments of the two nitrogen atoms are similar due to resonance effects. Consequently, the N1 and N3 1s→1π* transition energies are equivalent, as was seen in the NAP-XPS N1 and N3 1s E_B measurements.

The peak position of the ImidHCl N3/N1 1π* transition is 0.16 eV lower than that of the (roughly equivalent) N1 1π* transition in Imid, as the electron density around N1 in ImidHCl has decreased slightly due to the protonation of N3. It should be noted that this measured energy shift (0.16 eV) is within the 0.7 eV resolution of the instrument. The FWHM of the

N3/N1 1s→1π* peak is 1.1 eV, which highlights the equivalence of the N moieties, as there is no significant broadening compared to the Imid N3/N1 1s→1π* peaks.

There is a single IP continuum step for ImidHCl, as the two nitrogen moieties are pseudoequivalent. At 403.46 eV, a second, broader Gaussian peak is identified. Analysis of the TDDFT calculations attributes this peak to the N3/N1 1s→2π* electron transitions.

3.3.3.2 C K-edge XRS NEXAFS

The C K-edge spectra of Imid and ImidHCl aqueous solutions are shown in Figure 3.15. The peak positions of the fitted XRS experimental spectra and TDDFT explicitly-solvated structure models are presented in Table 3.1.

Only one relatively sharp Gaussian peak in the pre-edge region (286.62 eV) is observed in the Imid spectrum (Figure 3.15, bottom). Analysis of the TDDFT calculations suggests this peak convolutes C4 and C5 1s→1π* and 1s→2π* transitions, along with some contribution from C2 1s→1π*. At 288.23 eV, a second broader, less intense Gaussian peak has been fitted, which represents C2 1s→2π* transitions (ΔE of the peaks is 1.61 eV).

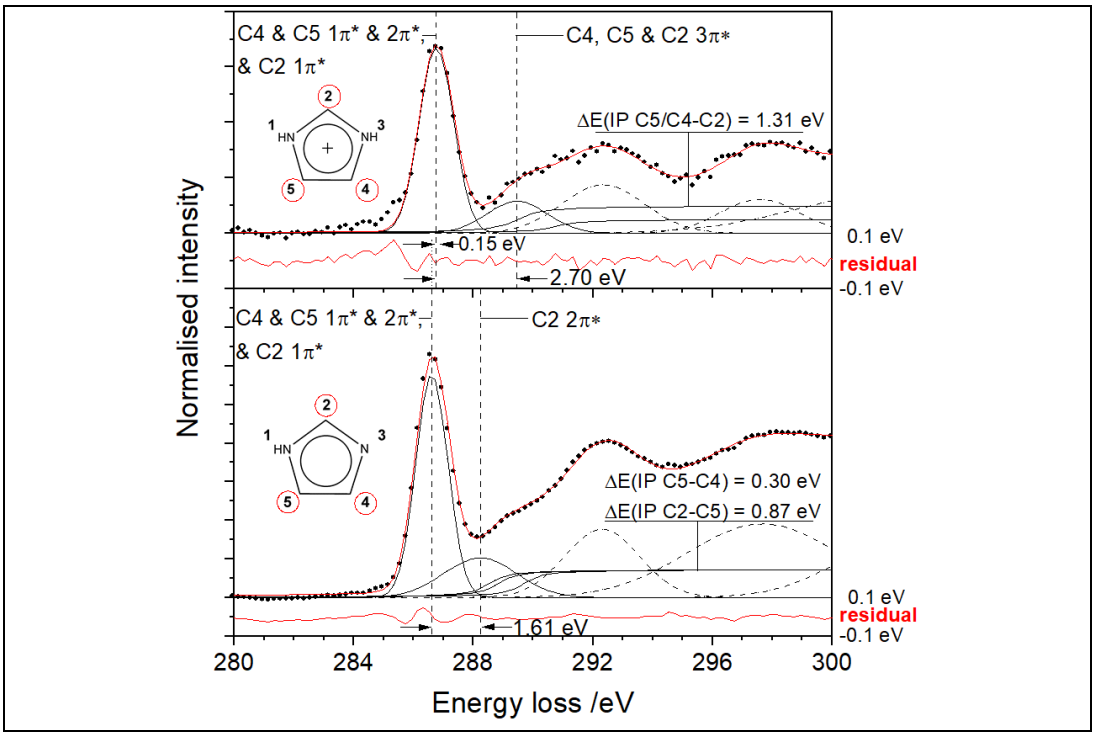


Figure 3.15 Experimental C K-edge XR spectra for 5 M aqueous Imid solution (pH10) – bottom; and 5 M aqueous ImidHCl solution (pH3) - top.

Experimental (black dots) overlaid with fitted spectrum (red). Gaussian pre-edge 1s→π* transition peaks (solid Gaussian), IP arctan function (solid) and post-edge σ*-shape resonances (dashed Gaussian). Residual shown below each spectrum.

As the 2π* (LUMO+2) transitions of the solvated $ImidH^+$ ion has no density of states at the C2 atom (Figure 3.6), a significant C2 1s→2π* transition can be excluded. However, the density of states at the C2, C4 and C5 atoms at LUMO+6, represent 1s→3π* electronic excitations.

Due to the redistribution of charge across the $ImidH^+$ ion, the electron density around all C moieties has decreased compared to Imid, which is evidenced by the increased C 1s E_B (Table 3.3). The greatest difference is around C2, where the E_B is 1.07 eV higher in $ImidH^+$ than in Imid, whereas the E_B increases are 0.93 eV and 0.63 eV for C4 and C5.

The significant difference between the Imid and $ImidH^+$ C K-edge spectra is the energy shift between the pre-edge Gaussian peaks. The first pre-edge feature of the $ImidH^+$ spectrum at 286.77 eV is 0.14 eV above that in the Imid spectrum and represents C4 and C5 1s→1π* and 1s→2π* transitions as before, with an increased contribution from C2 1s→1π* excitations. The second peak at 289.47 eV is 2.70 eV above the first. The DFT analysis shows that there are no 2π* states accessible for C 1s electrons from C2 in $ImidH^+$. The second peak is instead a result of C4, C5 and C2 1s→3π* transitions. This is a major difference between the $ImidH^+$ and Imid C K-edge spectra.

Table 3.8 Peak position and assignment of the experimental and calculated C K-edge spectra. The corresponding experimental spectra are shown in Figure 3.15.

†The energy shifts of ∼10.7 eV (Imid) and 10.3 eV (ImidHCl) have been applied to the calculated transitions presented here. ‡ There is no C2 1s→2π* transition, but C2,C4 & C5 1s→3π* transitions are observed

	Imid solution			ImidHCl solution		
	Exp. spectra /eV	Calc.† /eV	Diff. (exp-calc) /eV	Exp. spectra /eV	Calc.† /eV	Diff. (exp-calc) /eV
E (C2 1s→1π*)	288.23	287.04	1.18	286.77	286.77	0.00
E (C5 1s→1π*)	286.62	286.62	0.00	286.77	286.26	0.51
E (C4 1s→1π*)	286.62	286.43	0.18	286.77	286.24	0.53
E (C2 1s→2π*)	286.62	288.80	-2.18	‡	‡	
E (C5 1s→2π*)	286.62	288.84	-2.22	286.77	286.76	0.01
E (C4 1s→2π*)	286.62	287.77	-1.15	286.77	286.74	0.03
ΔE (C2 2π*-1π*)	1.61	1.76		2.70	3.88	
E (C2 1s→3π*)				289.47	290.64	-1.18
E (C5 1s→3π*)				289.47	289.31	0.15
E (C4 1s→3π*)				289.47	289.30	0.16
IP (C2)	289.71	288.98	0.73	290.78	293.96	-3.18
IP (C5)	288.84	288.29	0.55	289.47	292.55	-3.09
IP (C4)	288.54	287.89	0.64	289.47	292.53	-3.07
ΔIP (C2-C5)	0.87	0.69		1.31	1.43	
ΔIP (C5-C4)	0.30	0.39		0.00	0.00	
1σ* (C–H)	292.30			292.34		
2σ* (C–C)	297.68			297.58		
3σ* (C–N)	302.53			302.60		

This combined TDDFT/XRS-NEXAFS/XPS technique could in principle be applied to observe the electronic structure changes of organic salt during cooling crystallisation, allowing changes in local bonding and short-range structure to be evaluated and providing evidence of changing ion pairing during the cooling crystallisation process.

3.4 Summary and conclusion

X-ray Raman Scattering (XRS) has been used to obtain C and N K-edge fine-structure spectra for aqueous solutions of neutral Imid and the protonated $ImidH^+$ at ambient temperature and pressure. In combination with DFT and TDDFT calculations, and NAP-XPS data, these XRS-derived fine-structure spectra provide a quantitative description of the molecular orbitals in Imid and $ImidH^+$ in aqueous solution, including the influence of the water interaction on the electronic structure of the solvated molecules.

$ImidH^+$ exhibits a N K-edge fine-structure that is fundamentally and clearly different from that of Imid, due to the equivalence of the N moieties in the cation. The change in the C K-edge is less obvious without the complementary TDDFT calculations, which identified C $1s \rightarrow 3\pi^*$ resonances in the $ImidH^+$ spectrum that are not present in the Imid spectrum. This difference indicates a change in the character of the electronic state of the C moieties following the protonation of Imid.

Previously, XPS data of Imid and $ImidH^+$ has been acquired using liquid-jet spectroscopy using a synchrotron source. The NAP-XPS charge neutralisation should enable the C and N 1s emission spectra to be reported on an absolute core-level binding energy scale.

Further refinement of the method for collecting solution NAP-XPS data—for example modifying the sample container to allow the working distance to be evaluated, having greater control of the solution concentration, perhaps by using a flow cell, or identifying opportunities to reduce adventitious C contamination of the sample during handling or synthesis—could result in NAP-XPS becoming a solution characterisation technique that is more accessible, but equally as effective, as synchrotron-based measurements.

The only other study in the literature where XRS is used to acquire the NEXAFS spectra of organic solutes in solution is the *in situ* crystallisation of Imid from water, which was reported by our research group [67]. This study of Imid's C and N K-edges of though the MSZ to crystallisation found no significant changes to the average solvation of Imid until the point of crystallisation, suggesting that desolvation is the rate-limiting step of crystallisation. It was not possible to crystallise ImidHCl from water by cooling crystallisation within the

temperature range accessible by the experimental setup, so a cooling crystallisation XRS study was not undertaken during this beamtime. However, this XRS technique could in principle be applied to observe the electronic structure changes of organic salt during cooling crystallisation, allowing changes in local bonding and short-range structure to be evaluated. Initial TDDFT analysis of an ImidHCl molecular unit suggested that when the Cl^- and $ImidH^+$ ions form ion pairs in solution, the N species loses the pseudoequivalence observed in the solvated $ImidH^+$ cation, i.e. $\Delta E_{(N3\text{-}N1)}$ $1s \rightarrow 1\pi^* > 0$ where ion pairs are present. The sensitivity of the N and C K-edges to changes in local chemistry suggests that observing differences in the $ImidH^+$ N K-edge could provide evidence of the extent of ion pairing during cooling crystallisation, as the solution is cooled through the MSZ to crystallisation. More advanced molecular modelling of larger $ImidH^+/Cl^-$/water clusters, followed by TDDFT, could further investigate this observation. Further to this, the methodology described here could be developed to understand the extent of ionisation at surfaces, and in non-aqueous solvents where pH and pKa measurements are more difficult using standard techniques.

Chapter 4 Cooling crystallisation of guanidine hydrochloride: an *in situ* XPDF and EPSR study

4.1 Introduction

X-ray techniques, including energy dispersive X-ray diffraction (EDXRD), small-angle (SAXS) and wide-angle (WAXS) X-ray scattering have previously been used to observe the emergence of long-range order during crystallisation from amorphous materials [451,452] or melts [191,453], or the crystallisation of inorganic salts or nanoparticles [454,455]. However, until recently our ability to probe the local structure of a crystallising solution has been limited by the time- and length-scales of available experimental techniques [39]. The development of X-ray total scattering, which detects diffuse and Bragg scattering, has allowed the determination of short- to medium-range atomic structure in disordered materials, such as liquids or solutions. High-brilliance synchrotron X-ray sources now provide an opportunity to rapidly collect X-ray data in real-world sample conditions, allowing materials to be monitored *in situ*, even during phase changes [456].

Analysis of X-ray pair distribution function (XPDF) patterns of a solution, as it is cooled through the metastable zone (MSZ), provides information on structural changes from an undersaturated solution to a supersaturated solution, to the crystallised phase in solution. Extracting the atom-atom (partial) pair distribution functions (PDF) and identifying the real-space correlations in the probed systems requires computational structure models to be developed from the scattering data. Combining PDF analysis with Empirical Potential Structure Refinement (EPSR) simulations [257] has previously been used to reveal the structure of the hydration shells around nanoparticles [239], solvation and aggregation of organic materials in solution[23] [223,269,286,416], and ionic liquids[23] [119,147].

Here, XPDF's fast data acquisition time has been exploited to examine the *in situ* cooling crystallisation of guanidine hydrochloride (GuHCl). Guanidinium (Gdm^+, Figure 1.11) is a planar ion with three hydrogen-bonding amino groups surrounding a central carbon atom [93]. Gdm^+ is of biological interest as it can inhibit enzyme activity [457], is a known denaturant, and influences protein folding as a side chain of arginine [97], and as such, has been the subject of previous computational [91,92,179,181–183,458] or combined

[23] Using neutron diffraction isotopic substitution (NDIS)

experimental and computational [92,97,181,188] studies. Gdm^+ has been found be weakly-hydrating, due to its hydrophobic planar side [459], which could allow the formation of Gdm^+-Gdm^+ like-charged ion pairs [179].

This crystallisation study of GuHCl from aqueous solution aims to further our understanding how molecular interactions change during phase transition. Computational analysis using EPSR simulations, which refine the modelled system to experimental data, have been carried out with datasets from across the MSZ and after crystallisation. Statistical analysis of the modelled solution structures, using EPSR and dlputils [460], has also been undertaken.

4.2 Method

The cooling crystallisation of an aqueous solution of 9 M GuHCl has been studied using a combination of XPDF and EPSR modelling. Sample preparation of the solid GuHCl (as received and recrystallised) and 9 M GuHCl aqueous solution, is described in Section 2.2. An overview of the workflow for the collection of X-ray total scattering data is presented in Figure 2.17.

4.2.1 XPDF data collection and processing

X-ray total scattering data were collected at the X-ray pair distribution function (XPDF) beamline I15-1, at DLS (Figure 2.16), at 76.7 keV (0.163 Å) [373]. The large area Perkin Elmer detector (PE XRD 4343 CT detector) has an active area of 432 × 432 mm^2, with 150 µm pixel size. The instrument was calibrated with silicon. Further details of the operation of the beamline are in Section 2.5.4.1.

The detector positioning gave access to $Q_{max} \approx 40$ $Å^{-1}$. A higher Q_{max} provides better pair distribution function (PDF) peak resolution, although in the case of organic materials, a lower Q_{max} is used to process the data [201,367].

The arrangement of the *in situ* set up at beamline I15-1 for the cooling crystallisation experiment is shown in Figure 4.1a&b. The *in situ* closed-loop flow cell setup has been developed to resemble the equipment used in laboratory-style cooling crystallisation studies (Figure 4.1c). A water bath controls the temperature of the ~150 ml stock solution, which is stirred with a magnetic stirrer (5 mm x 30 mm) at 300 rpm to maintain a uniform bulk solution temperature. The solution was circulated by a peristaltic pump (~85 ml/min) through insulated Marprene tubing (6.4 mm OD, 3.2 mm ID) to the X-ray transparent 6 mm OD Kapton tube (Goodfellow). Continuously providing a fresh solution sample at the Kapton

window by circulating the solution minimised any possible beam damage. A K-type thermocouple recorded the solution temperature.

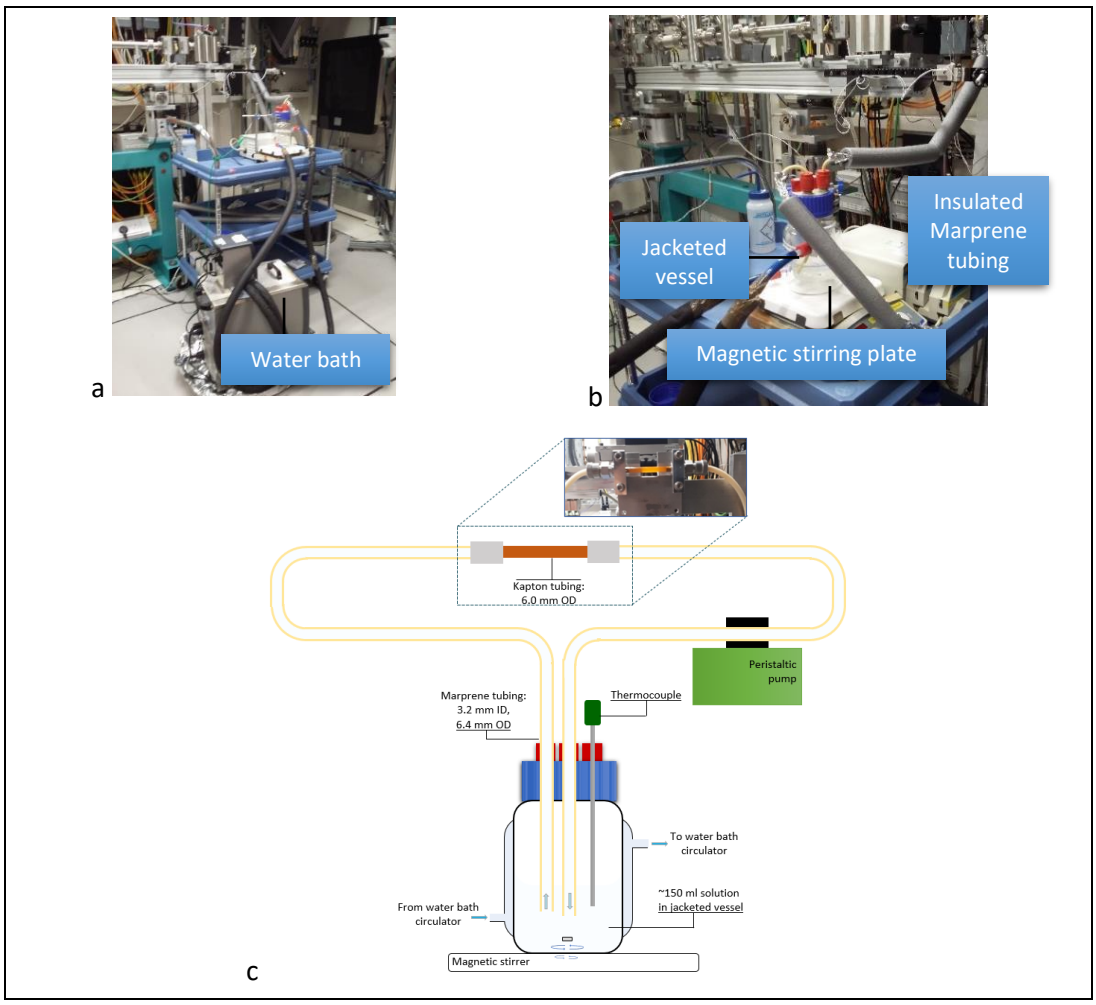


Figure 4.1 *In situ* experimental set up at beamline I15-1, DLS: (a) & (b) jacketed flow crystalliser arrangement with (c) X-ray transparent Kapton tubing for XPDF analysis

Images of the jacketed vessel and the Kapton tube were recorded and were visible from the control room so any problems with the experimental set up could be identified in real time. For example, significant air bubbles in the Kapton tube affecting the signal could be identified, and the onset of crystallisation could be observed so the experiment could be stopped, before crystals accumulate in and block the tubing.

The solution was heated to 65°C, after which the set point of the water bath to cool the solution in the jacketed vessel was set to 10°C and data collection was started[24]. Scattering patterns were collected continuously in 30 s frames. Every 120 frames, the data was stored in a new scan, resulting in a short pause (~20 s) in data collection.

[24] The cooling rate was not specified, but the solution cooled at a rate of 0.46°C/min until the solution temperature reached 45.7°C, after which the cooling rate was 0.79°C/min.

The data collected in the *in situ* experiments are listed in Table 4.1. The cooling crystallisation experiment was repeated twice. X-ray total scattering data were also collected whilst the solution was reheated (water bath set point 70°C) to capture the structural changes that occur during dissolution. The solution was cooled from 62.5°C (temperature of first data collection) until just after the appearance of crystals, which occurred around 20°C. X-ray total scattering data of deionised water were also collected with the acquisition time of 30 s per scans as the water was cooled from 68.2°C to 11.4°C. These data were collected for background correction, and was applied in the initial data processing (shown in Appendix D.1). The results in Section 4.3.2.1.2 were processed using an empty Kapton background instead. However, the XPDF patterns of cooling water are of interest in their own right and are shown in Section 4.3.2.1.1.

Table 4.1 List of scans and acquisition times for the *in situ* cooling crystallisation and dissolution study

Scan #	Sample	Comment	Acquisition time per frame /s	Temperature /°C
34306	Empty Kapton tube	Background scattering – single frame	600	Room temp.
34329–30	Run 1 - GuHCl cooling crystallisation	151 frames	30	62.5–19.6
34331	Reheating GuHCl solution	54 frames	30	18.8–62.8
34332–33	Run 2 - GuHCl cooling crystallisation	146 frames	30	62.6–20.2
34348	Empty Kapton tube	Background scattering – single frame	600	Room temp.
34349–50	Cooling of deionised water	218 frames	30	68.2–11.4

The solid samples of GuHCl were mounted on the capillary spinner along with an identical empty borosilicate glass capillary to collect background scattering. Data for each were collected with acquisition times of 600 s per scan.

4.2.1.1 Data processing

The X-ray total scattering data were processed to extract the structure factor F(Q), and pair distribution functions (PDF) g(r), also referred to here as the XPDF patterns. The generic data acquisition software at DLS automatically applied the initial corrections for the internal dark current to the raw data during data collection. GudrunX was used to correct the processed data for the instrument and sample holder backgrounds, absorption, multiple scattering and Compton scattering [365,368]. GudrunX produces interference differential cross sections F(Q), from which the real-space PDF patterns are generated by Fourier transform[25]

[25] The Fourier transform of F(Q) is G(r) (= $4\pi r\rho_0[g(r)-1]$)

[365,368]. There is an option in GudrunX to correct for fluorescence. Although the chloride ions are a possible source of X-ray fluorescence, trial and error showed that applying a fluorescence factor had negligible, if any, visible effect on the F(Q) or g(r) and so none were included.

For the solid samples, the background scattering from an empty borosilicate glass capillary is subtracted from the GuHCl powder X-ray total scattering data to acquire the XPDF pattern.

There are two ways in which the solution X-ray total scattering data will be used to analyse the evolution of the solution structure during cooling crystallisation:

1) ***Temperature-resolved study of the XPDF patterns***: the X-ray total scattering of a blank Kapton tube is background subtracted from each set of scan data in GudrunX to reveal changes in the intermolecular interactions between the solute-solute and solute-water molecules during the cooling crystallisation experiment.
2) ***EPSR modelling***: the X-ray total scattering of a blank Kapton tube is background subtracted from selected scan data, as the structure factor related to all components in the sample is required for molecular simulation.

4.2.2 Computational analysis

4.2.2.1 Solid-state GuHCl analysis

PDFgui was used for the refinement of the solid state GuHCl data, to generate a reference XPDF pattern for the analysis of the *in situ* cooling crystallisation data [402].

4.2.2.2 EPSR analysis for *in situ* crystallisation study

Structural models of the solution were generated using EPSR version 25 [257,406]. When the empirical potential (EP) was applied, the models were refined to increase the goodness of fit with the F(Q) data from the GudrunX processing of the X-ray total scattering data. Further details of EPSR are in Section 2.9.2.2.1.

Component structure: The Gdm^+ ion was geometry optimised in ORCA (RKS BP86 RI SVP def2/J) to provide the starting structure [380]. The Gdm^+ ion geometries vary during the simulation depending on the level of molecule flexibility set in EPSR, which is related to the temperature of the solution. An ion structure from the simulation of frame 150 shows the bond lengths variation, with the N-H bond lengths between 0.96 Å and 1.10 Å, and the C-N bond lengths between 1.36 Å and 1.41 Å.

Solution density: The EPSR simulation requires the atomic number density of the sample being modelled. Measuring the density of the sample from the crystalliser *in situ* at each

required temperature posed a challenge. Two approaches were tried—measuring the mass of a known volume of solution extracted from the stock solution using an autopipette and a glass graduated pipette—and both resulted in inaccurate repeat measurements. The GuHCl salt crystallised on the graduated pipette, so the complete sample mass was not measured. The volume of the sample removed by the autopipette was affected by the temperature of the solution (this issue had also been observed during the preparation of the Crystal16 samples, Section 2.8), resulting in significant variation in the measured sample mass at a given solution temperature, and so this method did not appear to be accurate or repeatable. Furthermore, the aliquots that were extracted were not put back into the stock solution, potentially altering the crystallisation process conditions and making the system unrepresentative of the *in situ* conditions. Instead, the volume of the solution, the mass of which was known, was tracked as it cooled in a measuring cylinder to provide data to be converted into the atomic number density. The present method seems to provide physically sensible structural models, although finding an alternative method for ascertaining the solution density more reliably for future system analysis should be identified (see Section 6.2).

Component Lennard-Jones potentials and partial charges: The atomic partial charges and LJ potentials used in these EPSR simulations are shown in Table 4.2. There are many different force field data sets in the literature for Gdm^+ [458], Cl^- [268], and water [267,268,458,461–463]. A parametrisation study was undertaken to identify the most appropriate parameters to apply to the model, including the water model type (SPC/E and/or TIP3P), the LJ potentials and partial charges. Details of this study are outlined in Appendix F. The application of charge reduction has been applied to empirically improve the fit of modelled data to experimental data, and is discussed in the literature review in Section 1.4.2.3.

Table 4.2 Lennard-Jones potentials and partial charges for Gdm^+ and Cl^- ions and TIP3P water

		q	σ /Å	ε /kJ/mol	Reference
Gdm^+	C	0.797	3.77	0.4170	Ref. [458] (with 20% charge reduction)
	N	-0.759	3.11	0.5000	
	H	0.380	1.58	0.0880	
Cl^-	Cl	-0.8	3.97	0.6190	Ref. [268] (with 20% charge reduction)
Water TIP3P	O_t	-0.834	3.1506	0.6364	Ref. [462][a]
	H_t	0.417	0	0	

[a] Lennard-Jones σ and ε parameter values for TIP3P 3-point water model from [463]

The bond angle for water was set at 104.5° [463]. As is extensively described in the literature, there is no consensus on the optimal way to model water, either alone or as part of a solution (Section 1.4.2.4). As part of the parametrisation study, EPSR solution models

comprising water components with 100% SPC/E, 100% TIP3P and 50:50 SPC/E and TIP3P molecules were compared and TIP3P was selected (Appendix F).

Solution structure model generation: For the EPSR molecular models, boxes of molecules were constructed with the dimensions, atomic number density and temperatures in Table 4.3. The solutions comprise 694 water molecules with 306 Gdm^+ and 306 Cl^- ions.

Table 4.3 EPSR simulation boxes

Frame number	Simulation temperature /K	Side length of cubic box /Å	Simulation box volume / Å^3	Atomic number density /atoms. Å^{-3}
GuHCl_0	335.7	38.70	57,970	0.09398
GuHCl_3	335.1	38.70	57,951	0.09401
GuHCl_4	334.7	38.70	57,941	0.09403
GuHCl_28	329.1	38.65	57,754	0.09433
GuHCl_29	328.9	38.65	57,747	0.09434
GuHCl_30	328.6	38.65	57,742	0.09435
GuHCl_60	322.0	38.60	57,517	0.09472
GuHCl_89	313.3	38.54	57,227	0.09520
GuHCl_114	303.3	38.46	56,880	0.09578
GuHCl_129	298.0	38.42	56,697	0.09609
GuHCl_145	293.3	38.38	56,532	0.09637

Each simulation was run with the reference potential (RP) for ~1,500 iterations until the system energy was at equilibrium. The empirical potential (EP) for the simulations was 4 kJ/mol. The EP simulations ran for 400-500 iterations before the simulations were set to accumulate. The accumulation and analysis routines were run for a further ~400-500 iterations.

Pair correlation functions: Gdm^+ has 10 atoms (three distinct atom types, see **Figure 4.2**). In the water molecule, there are two distinct atom types in water, labelled O_t and H_t (TIP3P water model). Along with Cl^-, the aqueous GuHCl system comprises six different atomic labels (N = 6) and so has N(N+1)/2 = 21 distinct partial PDFs [368].

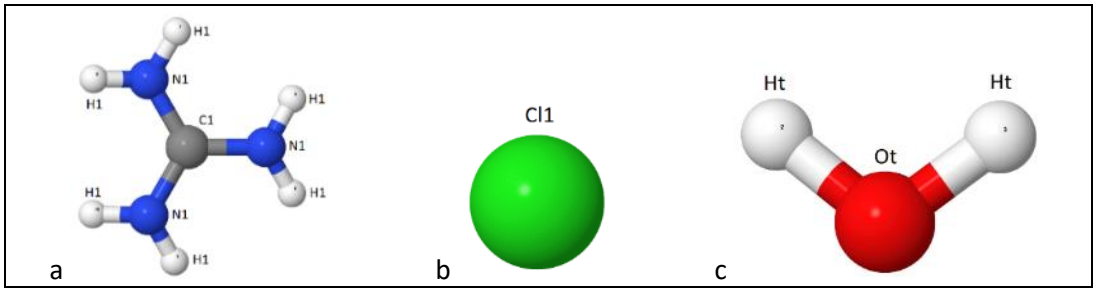


Figure 4.2 EPSR simulation components and their atom types

Statistical analysis: Statistical analysis to compare the simulations weas carried out with EPSR (Clusters) and dlputils (angular radial distribution functions ARDF, *intertorsion* for angle analysis, and spatial probability densities (SPD)) [460]. Data for statistical analysis were

collected after EP equilibrium had been achieved. Details of the auxiliary routine parameters are in Appendix E.

SPD, visualised with Aten [408], provide further insight into the local coordination of molecules around a central component, as the 3D histogram binning indicates the spatial arrangement of the interactions. SDF have been visualised for four solutions across the temperature range: GuHCl_3 (stage 1), GuHCl_60 (stage 2), GuHCl_129 (stage 3) and GuHCl_150 (crystallisation stage)—see Table 4.5 for Stages.

Ideal distribution of ions: Before comparing the $g_{\alpha\beta}(r)$ of atomic pairs, the ideal distribution of ions in solution has been evaluated. From the EPSR simulation box dimensions, which are based on experimental density measurements, the ion centres would be ~4.5 Å apart,[26] if evenly distributed. The geometrical separation of fully hydrated, monovalent solvent-separated ionic pairs (SSIP) in solution was considered by Marcus [464]. A rough estimation, based on the assumptions of the σ values applied to the atoms in EPSR, and a water molecule diameter being 2.78 Å [464], suggests a ~10 Å ion centre separation (assuming a planar interaction). With a water:GuHCl ratio of 2.3:1, Gdm^+ and Cl^- cannot be fully hydrated in this concentrated solution. Using the calculation specifically for concentrated salt solutions [465], the cation-anion separation in a symmetrical (1 anion:1 cation) salt solution d_{C-A}^{av} (Å) is

$$d_{C-A}^{av} = 10.(0.940.c)^{-1/3} \qquad \textbf{Equation 24}$$

where c is molar concentration, giving a separation of 4.9 Å. (Note: this calcualtion does not account for the non-spherical nature of Gdm^+.) EPSR simulation $g_{\alpha\beta}(r)$ indicate ion-ion interactions at shorter separation distances than these estimated average values.

4.3 Results and discussion

4.3.1 Solid GuHCl analysis

Single crystal XRD (SCXRD) of recrystallised GuHCl was collected for refinement against the XPDF data using PDFgui [402]. Details of the SCXRD analysis are available in Section 2.5.2 and Appendix A. GuHCl crystallises as a colourless compound in an orthorhombic cell. It was solved in the *Pbca* space group with one formula unit in the asymmetric unit (Figure 4.3).

[26] Based on simulations 0, 129 and 145

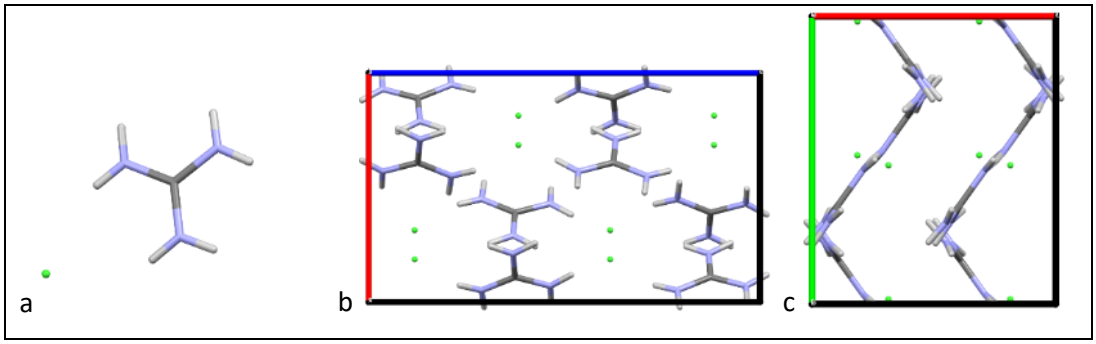


Figure 4.3 Recrystallised GuHCl, structure solved with SCXRD: a) unit cell and packing structure, viewed along the b) b-axis and c) c-axis

The atomic separation distances in the SCXRD-identified formula unit are shown in App. Figure 7 and Table 4.4. The cell coordinates of the measured structure are shown in App. Table 3 along with the three GuHCl structures from the Cambridge Structural Database (CSD). All structures belong to the space group P*bca*, which along with the XRD analysis (Section 2.5.3), indicate that there are no known GuHCl polymorphs. No examples of GuHCl polymorphs detected by PXRD or spectroscopic techniques were identified in the literature.

Table 4.4 Atomic separations in GuHCl unit cell

Atom pair	Average distance /Å
C – N	1.33
N – N	2.30
N – H	0.85
C – H	1.88
H – H (same / adjacent C)	1.48 / 2.26
C - Cl	3.73

The solid GuHCl XPDF pattern was measured at room temperature (~290 K) and so the refinement was carried out for this temperature. The refined parameters include the lattice parameters by space group (axis lengths *a, b, c,* and internal angles *α, β, γ*) and isotropic atomic displacement parameters, which were assigned to each atom type in the structure. The refinement was carried out over a range of 0.7-50.0 Å. As Q_{max} = 21 Å^{-1} was used in the experimental data processing, it was applied in the data fitting process. The Q_{broad} (broadening) and Q_{damp} (damping) terms, which arise from experimental resolution effects, were found to be 0.035 Å^{-1} and 0.037 Å^{-1}. A peak broadening term related to correlated atomic motions of the molecules was constant. The scale factor for optimised fitting was found to be 1.31.

The refined structure's PDF pattern is shown in Figure 4.4. The initial refinement with r_{min} = 0.7 Å (Figure 4.4, bottom) has a residual value R_w of 0.35. Achieving R_w in the range 0.20–0.30 is said to be favourable for organic crystals [367], but this range was based on a PDF refinement with r_{min} = 4.5 Å, to focus on intermolecular interactions. Further PDFgui analysis

of the GuHCl data with a refinement range of 2.5-50.0 Å (Figure 4.4, top), which removes some of the intramolecular interactions, results in a fit with $R_w = 0.26$, in agreement with the acceptable range in the literature [367].

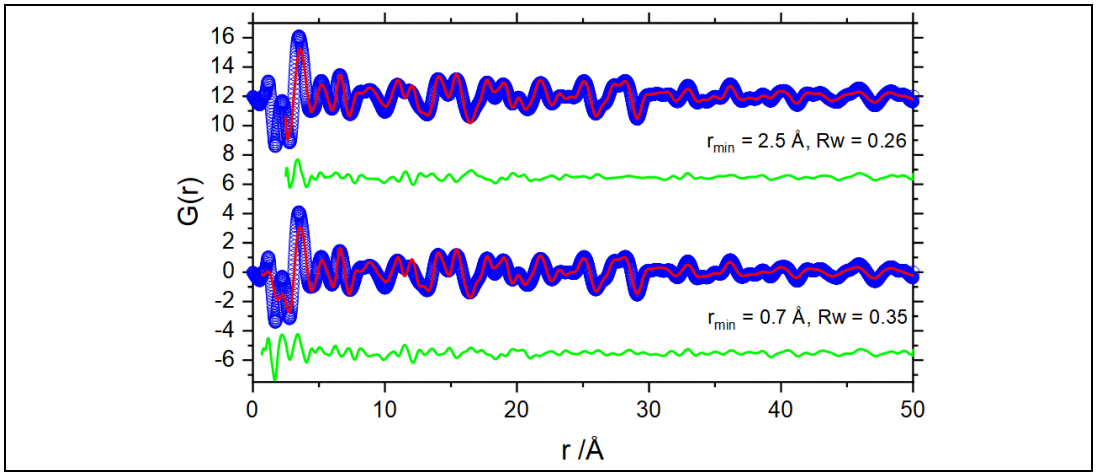


Figure 4.4 Refined structure of solid GuHCl PDFs using PDFgui [402]

Blue circles: experimental XPDF data; Red line: modelled PDF; Green line: difference

Atomic pair separations were extracted from the refined fits, up to a crystal length of 30 Å. The most frequent interactions are presented in Appendix A.3.2.1 and were used to identify interactions in the EPSR analysis.

4.3.2 *In situ* cooling crystallisation and dissolution

This section presents the analysis of XPDF data collection of the 9 M aqueous GuHCl crystallising solution during cooling crystallisation and its subsequent reheating. First the temperature-dependent XPDF pattern data are presented, followed by the computational structural modelling using EPSR in Section 4.3.2.2.

4.3.2.1 Structural evolution of solution structure through the analysis of temperature-resolved XPDF patterns

Summary of terminology used: F(Q) is the reduced structure factor, the corrected and normalised structure factor S(Q). The reduced atomic pair distribution function G(r) is the Fourier transform of F(Q), and g(r) is the atomic pair distribution function (see Equation 19). The partial pair distribution function $g_{\alpha\beta}(r)$ describes the probability of finding one atom type β, around another α, per unit volume, in the material [257]. The terminology is explained further in Section 2.5.4.

For the XPDF pattern study of the cooling crystallisation solution, the X-ray total scattering data were initially corrected for background scattering by subtracting the scattering data for water in Kapton at the equivalent temperature. It was anticipated that this process would

remove scattering from water self-interaction and make the solute-solvent and solute-solute interactions of interest clearer to see. Therefore X-ray total scattering data of water were collected whilst cooling (68.2–11.4°C), using the same setup as used for the *in situ* crystallisation (Figure 4.1).

Initial analysis indicated that there were distinct solution structural phases in the cooling crystallising solution in the MSZ. Similar phases were found in both cooling crystallisation runs 1 and 2. However, due to the low water to ion ratio in the 9 M solution, resulting in limited pockets of bulk water in the sample, and the relatively high scattering of the Gdm^+ and Cl^- ions, it was decided that it would be more appropriate to subtract an empty Kapton background to analyse the evolution of the solution structure (Section 4.3.2.1.2). The original analysis of the XPDF patterns, where water in Kapton was used for background subtraction, is shown in Appendix D.1. Nonetheless, analysis of the time-series of water X-ray total scattering data is still of interest.

4.3.2.1.1 Cooling water in Kapton

There are changes in the relative intensity and separation of the peaks at ~2.0 Å and ~2.9 Å, as illustrated by the dotted lines in Figure 4.5b. As the temperature of the water decreases, the second peak increases in intensity relative to the first and all subsequent peaks are higher in intensity than at high temperatures. The separation of the peaks moves from 0.61 $Å^{-1}$ (68.2°C) to 0.88 $Å^{-1}$ (11.4°C). These changes are indicative of increase structural ordering being observed at lower temperatures [466].

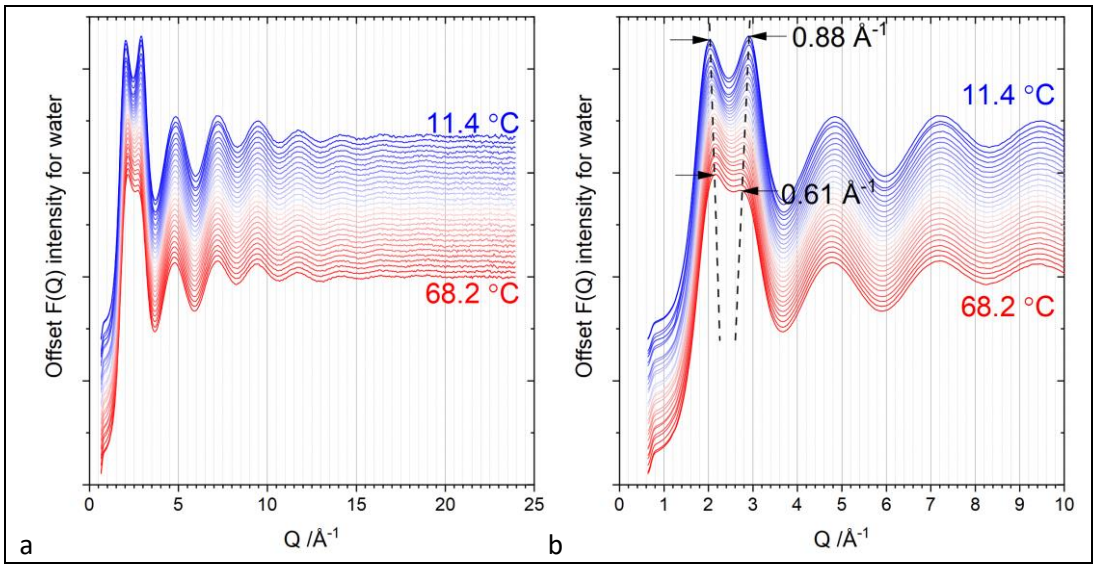


Figure 4.5 Selected F(Q) for water: 68.2–11.4°C

This increased structuring with decreasing temperature is also seen in the G(r) (Figure 4.6), as the intensity of the second and third O···O peaks increases indicating increased structural ordering. The peak positions also shift slightly to a lower-r. The feature that appears at ~3.5 Å in the contour plot of the G(r) data (highlighted by a purple box in Figure 4.7) highlights the increasing intensity difference between the first and second O···O shells as the local order in the water increases.

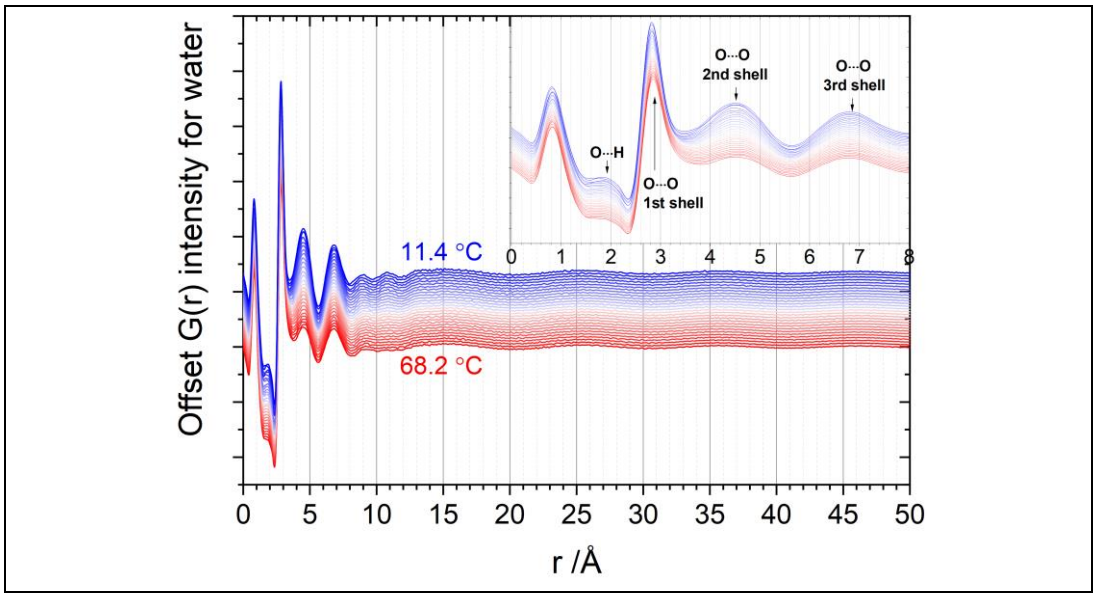


Figure 4.6 G(r) for water: 68.2-11.4°C: 0-50 Å (inset 0-15 Å)

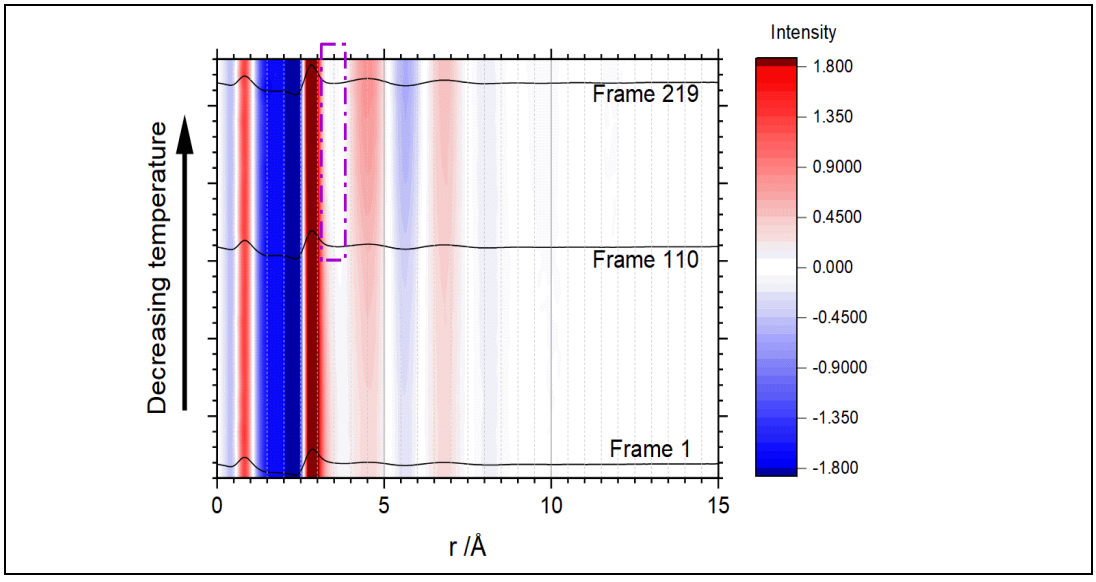


Figure 4.7 G(r) for water: 68.2-11.4°C: contour plot 0-15 Å

The differences are clearer when the pattern for each G(r) with the first (higher temperature) G(r) subtracted are compared (Figure 4.8). The changes that occur at 2.8 Å, 4.5 Å and 6.8 Å relate to the shifting to a lower-r of the successive O···O peaks. The peak at

1.9 Å (indicated with *) increases intensity with temperature, but only slightly, showing the strongly directional O···H hydrogen bond changes very little with temperature [235]. Changes to the H···H bonds cannot be identified from the XPDF pattern alone, and structural modelling to fit the data would be required.

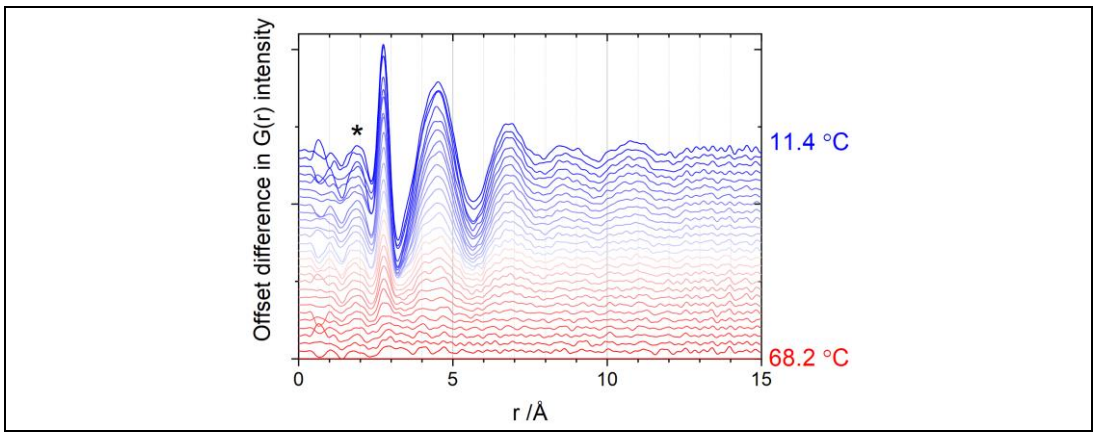


Figure 4.8 G(r) difference in for water: 68.2-11.4°C

4.3.2.1.2 Cooling crystallisation: Run 1

XPDF scattering data were collected in 30 s frames whilst the crystallising solution was cooled from 62.5°C (first data collection) until just after the appearance of crystals, which occurred around 20°C. In total 151 frames were collected in two scans, which means that after the first 120 frames, there was a short pause (~20 s) in data collection whilst the scan reset. The temperature of the solution was measured at the start of each frame (Figure 4.9). The top inset shows the crystallisation exotherm, with an increase in solution temperature between frames 147 and 148. The inset on the left indicates where the scan resets.

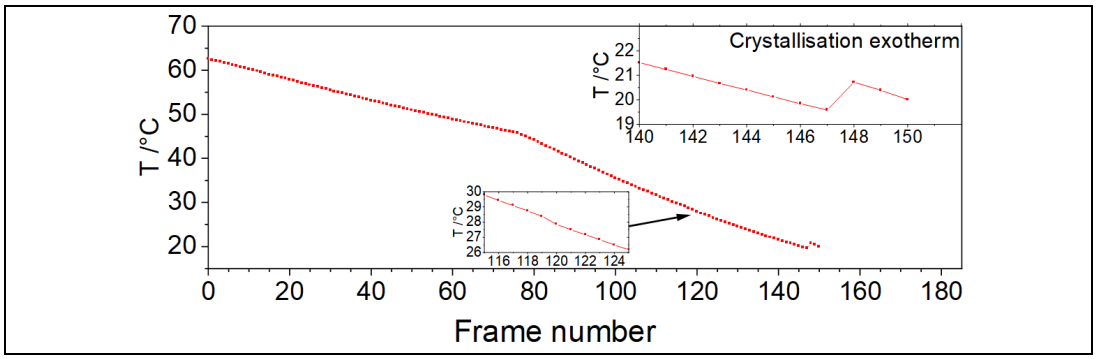


Figure 4.9 Run 1 crystallising solution temperature during data collection

Inset top-right: a crystallisation exotherm is observed to occur during the 147th frame hence an increase in temperature from the 147th to 148th frame. The central inset indicates the point in the data collection where the second scan starts after 120 frames.

Temperature is measured at the start of the 30 s data collection.

The change in gradient at frame 77 occurs at ~45°C, and this is where the cooling rate changes due to the automatic default settings of the water bath. This change in cooling rate at ~45°C is seen in cooling crystallisation Run 2 and in the cooling water measurements (Figure 4.10), indicating that this change is not related to any physical change within the solution. Ideally, the cooling rate would have been fixed at a set rate throughout, but unfortunately it was not the case during this data collection.

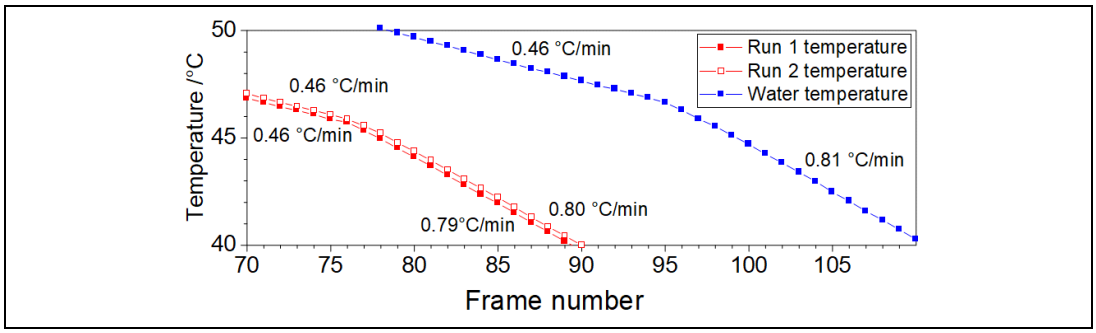


Figure 4.10 Runs 1 and 2, and cooling water experiments: temperatures and cooling rates

Sample temperature at the beginning of each data frame, and cooling rates either side of 45°C, where there is a change in the cooling rate due to the automatic settings of the chiller.

Images of the solution in the jacketed vessel during the cooling crystallisation are shown in Figure 4.11. The solution becomes visibly hazier at ~30°C (Figure 4.11b), and there is an obvious presence of crystals at 20.4°C (Figure 4.11c).

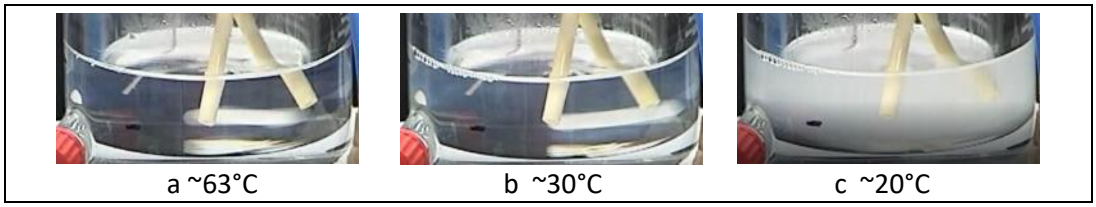


Figure 4.11 Solution in the jacketed crystalliser at around a) 63°C, b) 30°C and c) 20°C

Not all the solvated GuHCl recrystallised during the cooling crystallisation. Measuring the mass of filtered and dried recrystallised GuHCl indicated a yield of only ~2%. This yield is an underestimate, as the process of removing and filtering the crystals caused the dissolution of some of the solid that formed. However, it illustrates that a significant proportion of GuHCl was likely still in solution at the end of the *in situ* experiment.

The significant features at low-Q values in the normalised S(Q) data (before data reduction using GudrunX) indicates that there is long-range order, and therefore crystals are present, in the final four frames (Figure 4.12).

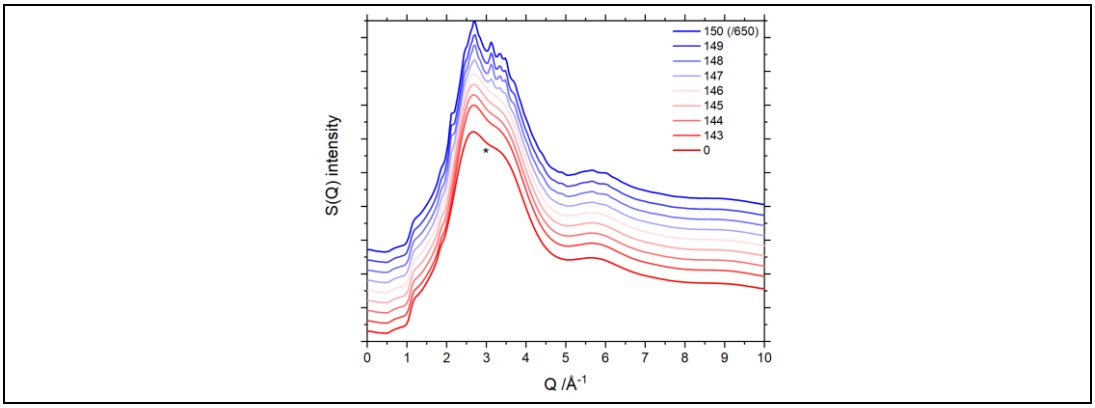


Figure 4.12 Run 1: offset S(Q) for the first frame (frame 0, 62.5°C) and the final eight frames.

The intensity of the final frame (150 20.0°C), where the most solid crystal product is present, has been reduced by a factor of 650 so that it can be plotted alongside the other data.

The trough between the peaks at 2.7 $Å^{-1}$ and 3.4 $Å^{-1}$ (indicated with * in Figure 4.12) is shallower in run 146, indicating some long-range structural change is occurring, although not to the extent seen in the following scattering patterns. Note, the first frame is assigned frame 0, as this is consistent with the data collection at the beam time. In total, 151 frames were collected, 0–150. Frame 150 intensity has been divided by 650, as the intensity was a factor of 650 greater due to the large number of crystals present in the Kapton window.

Looking at normalised and corrected F(Q) data for the cooling crystallisation run, again the final four frames indicate the presence of long-range structure (Figure 4.13a&b). The first principal peaks in the GuHCl solution are at 1.8 $Å^{-1}$ and 2.4 $Å^{-1}$, followed by peaks at 4.0 $Å^{-1}$, 6.3 $Å^{-1}$ and 8.4 $Å^{-1}$. In pure water, the peaks are positioned at 2.0 $Å^{-1}$, 2.9 $Å^{-1}$, 4.9 $Å^{-1}$, 7.2 $Å^{-1}$ and 9.5 $Å^{-1}$ (Figure 4.5), which indicates that water structure does not dominate in this solution, indicating that the majority if not all water is solvating the ions in solution.

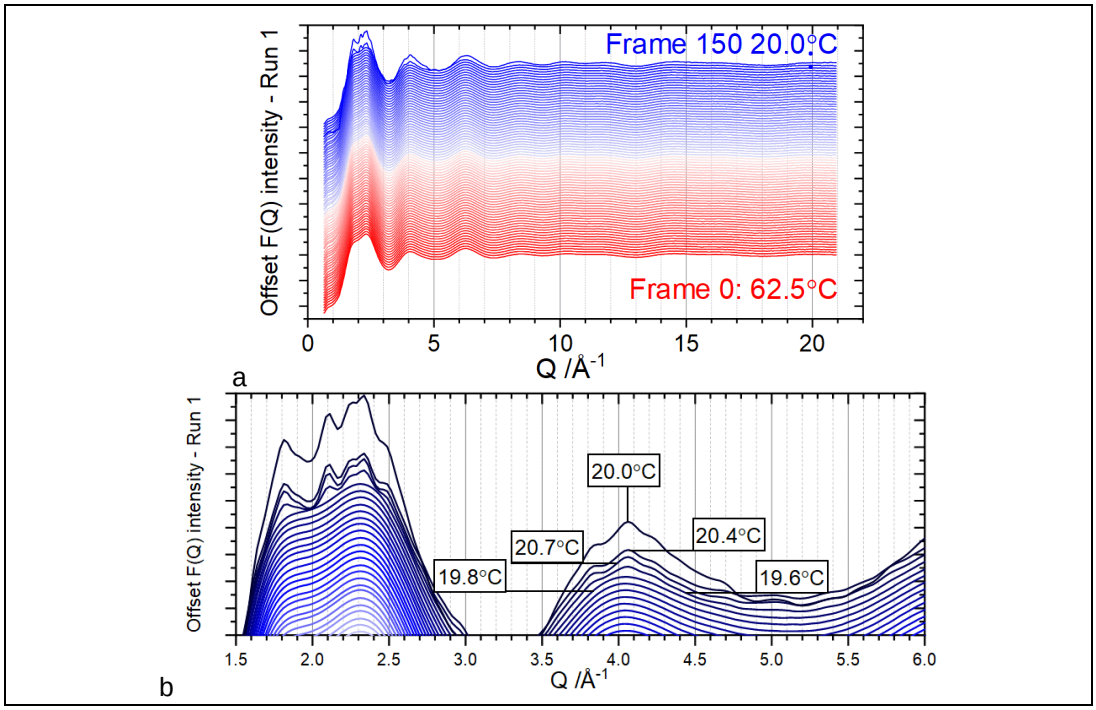


Figure 4.13 Run 1: a) offset alternate F(Q) and b) offset low-Q F(Q)

Long-range order can be seen emerging up to 50 Å in the final frames of the g(r) (Figure 4.14a). These features are clearer in the G(r) plots (Figure 4.15), where the features in the high-r are magnified. The difference between the first and last g(r) is shown in Figure 4.14b.

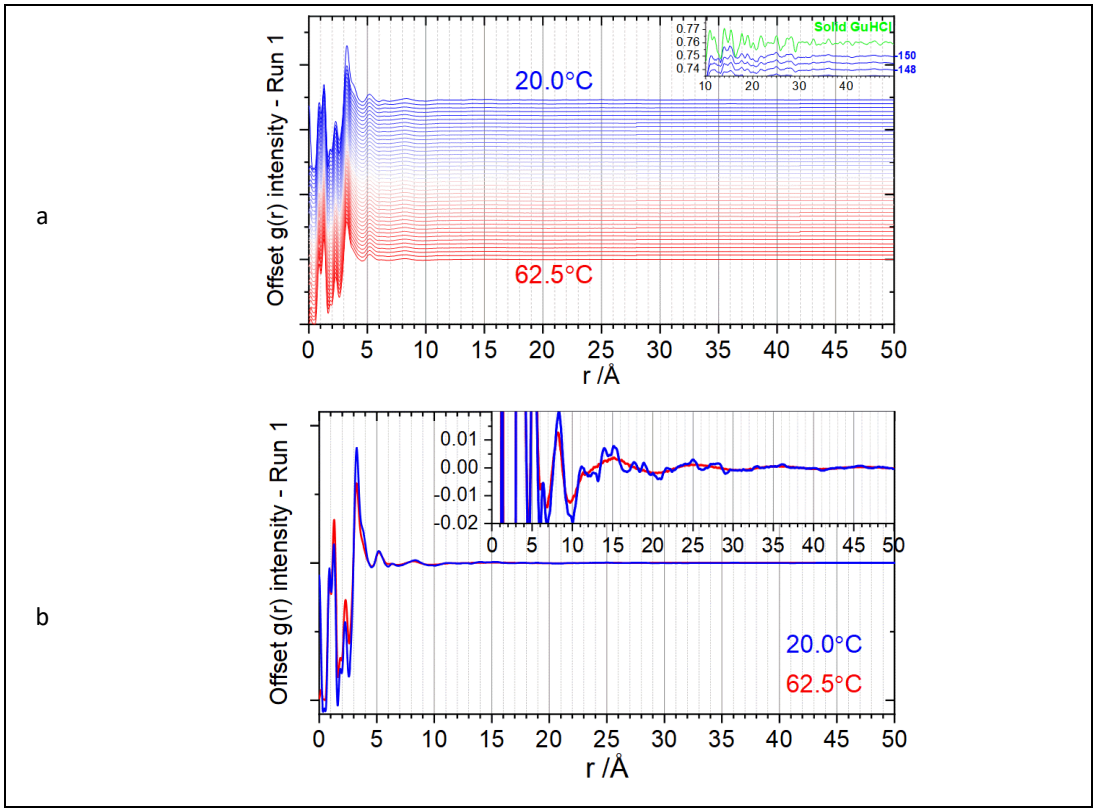


Figure 4.14 Run 1: a) offset g(r) (0–50 Å) and b) g(r) comparison of the first and last XPDF patterns

Some differences in the G(r) can be seen by eye, for example long-range order can be seen in the top four G(r) XPDF patterns in **Figure 4.15**a&b, with little obvious evidence of long-range interactions in the solution before this point. There is a peak forming at 6.5 Å (**Figure 4.15**b, ‡).

Figure 4.15c shows the long-range order of the XPDF patterns more clearly, and illustrates that there are no significant features indicating molecular interaction of GuHCl > 10 Å until the final four frames of data.

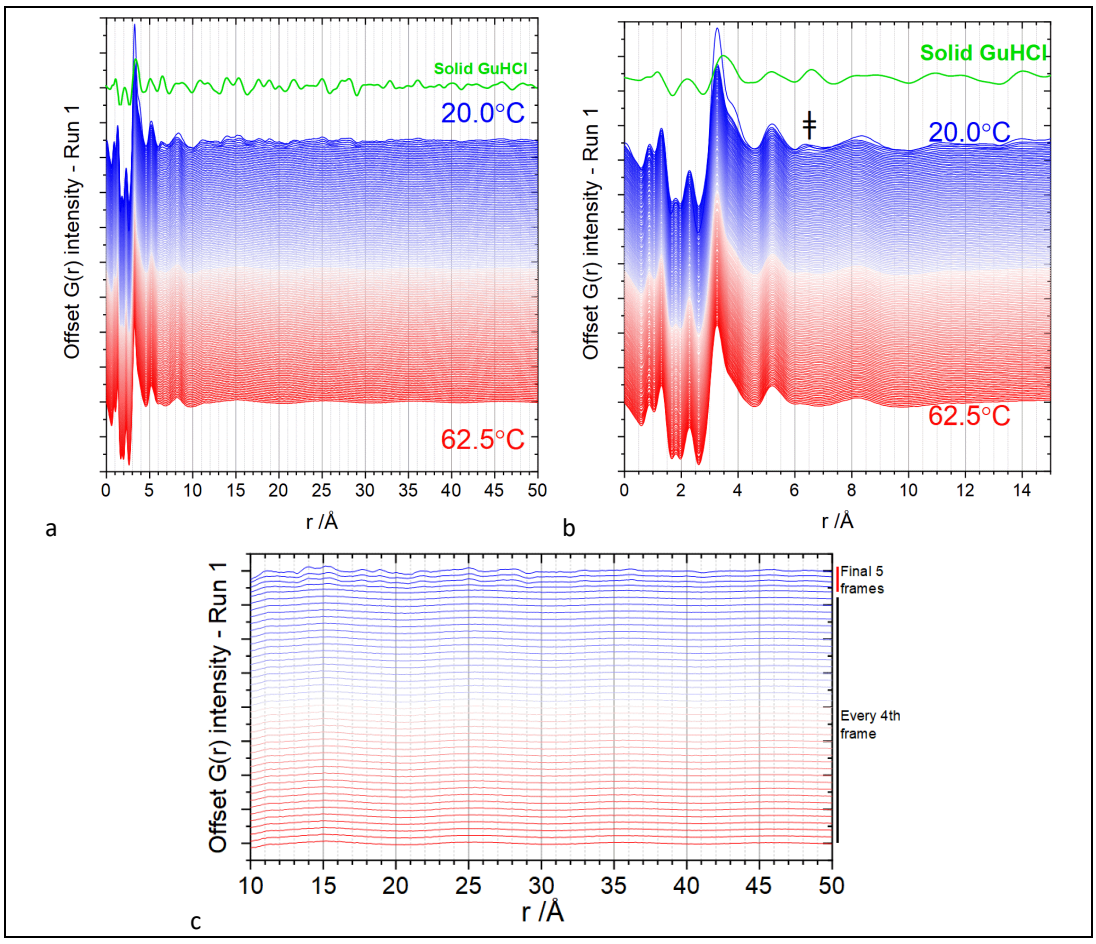


Figure 4.15 Run 1: a) offset G(r) (0–50 Å) b) offset G(r) data (0–15 Å) and c) every fourth G(r), 10–50 Å

The contour plots of the G(r) in Figure 4.16a&b show that there is some level of long-range order in the solution, illustrated by the faint blue and red bands, which match the peaks and troughs of the solution XPDF patterns. The emergence of long-range structure of GuHCl, matching the solid GuHCl XPDF pattern in green, can be seen in the final few frames (Figure 4.16a purple box, and shown in Figure 4.16d). There in an increased intensity at ~12 Å (orange box, Figure 4.16a). Figure 4.16c shows that there is an increasing trough at ~6 Å, with an increasing intensity at ~8 Å as the solution temperature moves through the metastable zone.

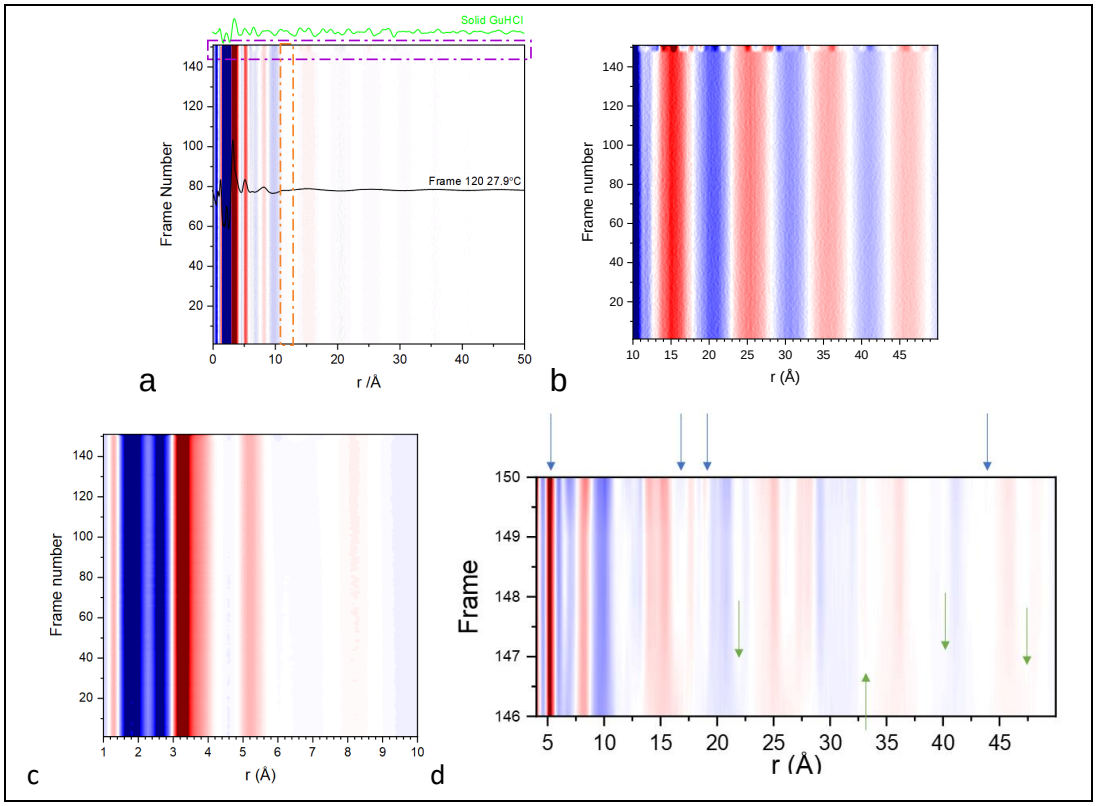


Figure 4.16 Run 1: G(r) contour plots a) 0–50 Å and b) 10–50 Å c) 1–8 Å and d) 4–50 Å (last five frames of data only)

Some intensity changes are clearer in the G(r) difference plots (Figure 4.17a&b). Aside from the emergence of long-range order in the final four frames, the obvious differences appear to be for short-range correlations, indicated by * (3.2 Å) and ‡ (~6 Å), where a feature forms upon crystallisation).

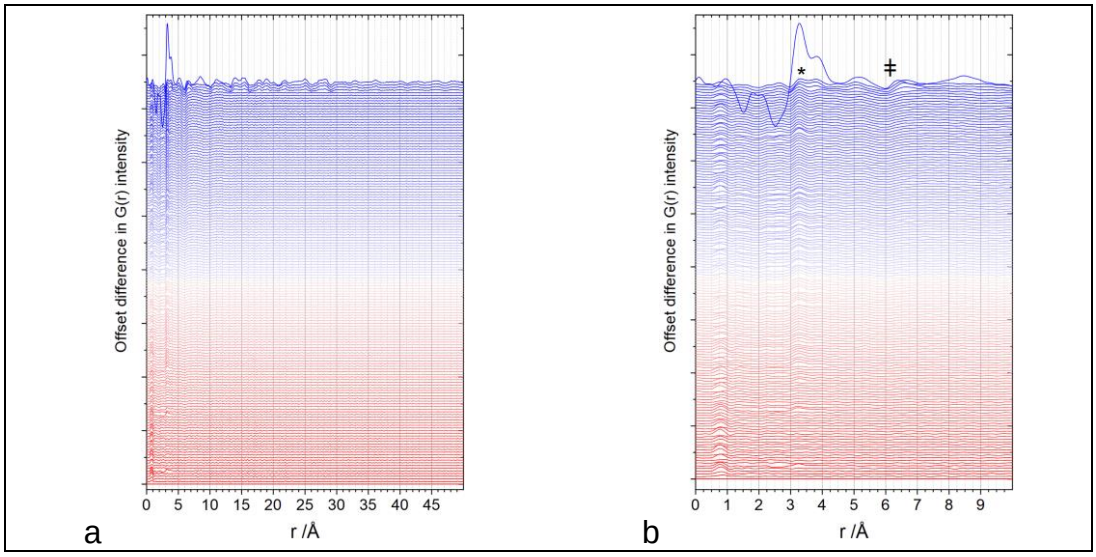


Figure 4.17 Run 1 - G(r) difference patterns a) 0–50 Å and b) 0–10 Å

G(r) – G(r) of frame 0 (62.5°C)

Contour plots of the difference plots further highlight that significant structural change occurs in the long-range order in the final frames only, but from around frame 32 to frame 43 onwards (53.6–52.5°C, indicated with *), there are changes in the medium-range order, ~5-15 Å (Figure 4.18a&b).

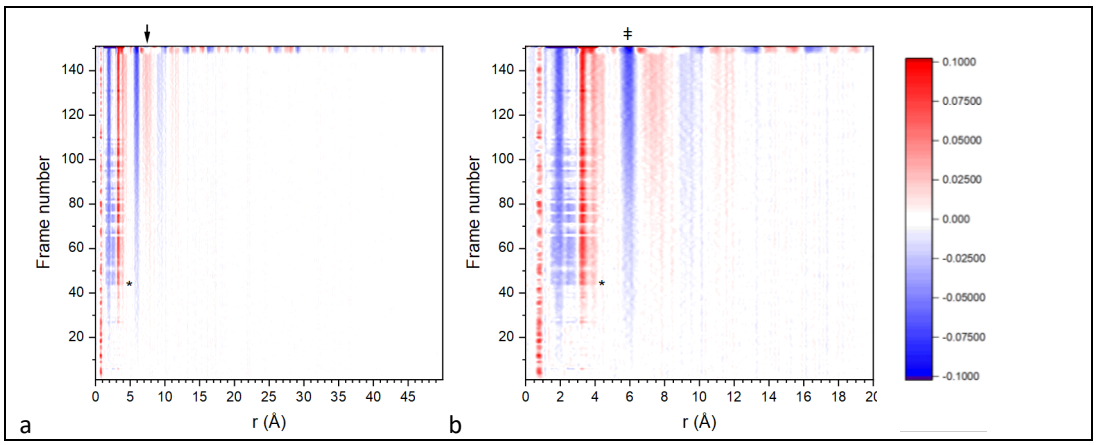


Figure 4.18 Run 1 – G(r) difference contour plot set between +0.1 (red) and -0.1 (blue): a) 0–50 Å and b) 0–20 Å

Principal component analysis (PCA): PCA is a useful tool for the analysis of data with many variables. By reducing the dimensionality of the dataset with minimum data loss, the data can be more easily analysed and interpreted [467]. The principal components (PC) are abstract variables that do not necessarily have any physical meaning and describe variation in the data sets [468].

With the XPDF patterns, PCA will map out variations in the interatomic interactions of the aqueous GuHCl system over the course of the *in situ* data collection. The data can be analysed across the whole r-range, or across specific r-ranges to identify where variation is most significant. To avoid artefacts appearing as a PC, only the PCs that describe a high percentage of variation will be considered (generally the first three PC, after which %variation can drop to < 1%) [468]. Covariance matrix PCA has been performed using OriginLab 2019.

PCA analysis has been carried out on the G(r) datasets, and on short-, medium- and longer-range r-values. In each case, the first and second PC describe the greatest variation, suggesting that PCA is an appropriate tool for analysing variation in the data.

The score plots of the first and second PCs are shown in Figure 4.19 for G(r) over the ranges 0–50 Å and 1–50 Å. Using G(r) instead of g(r) minimises the features < 1 Å, so the PC analysis over the ranges 0–50 Å and 1–50 Å are similar (Figure 4.19a&b). However, to exclude any potentially spurious and non-physical features below 1 Å, which may affect the analysis

[291], a minimum r-value of 1 Å will be applied in the analysis. Comparing Figure 4.19a&c shows the effect of retaining and excluding the final four frames, in which crystals are present, in the PCA. As I am interested in the solution structure in the phase prior to crystallisation, frames 147-150 will be exclude in the analysis, to increase the variation in the second PC.

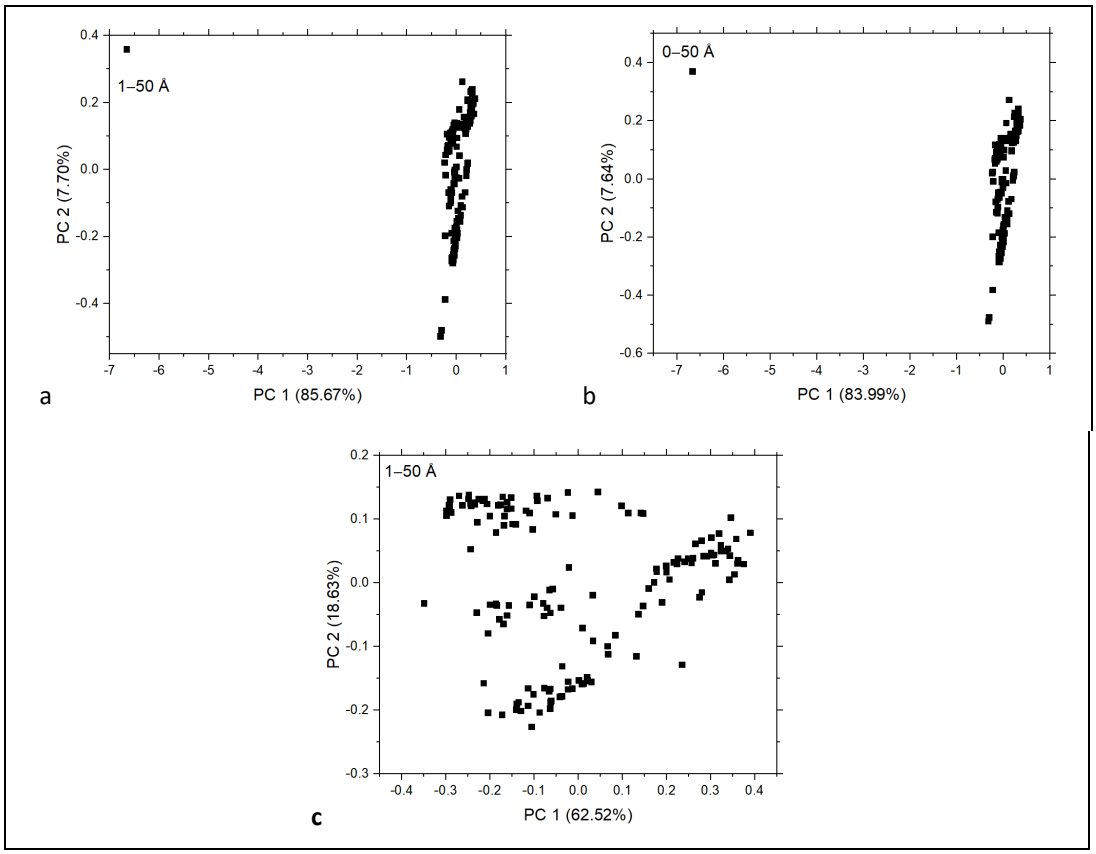


Figure 4.19 Cooling crystallisation run 1: PCA score plots for all G(r) XPDF patterns over the ranges 0–50 Å and 1–50 Å, and from 1–50 Å excluding the final 4 frames

PC1 v PC2 for a) 0-50 Å b) 1-50 Å and c) 1-50 Å for the solution only (final 4 frames excluded)

PCA has been carried out on the G(r) measured in the MSZ (frames 0–146, 62.5–19.8°C) to highlight short-, medium- and long-range intermolecular interactions (Figure 4.20). The variation in the short- and medium-range interactions (Figure 4.20a&b, where > 90% of the variation in the G(r) is in PC1&2) is greater than for the longer-range interactions (Figure 4.20c&d, where < 30% variation is captured by PC1&2). The spread of the points in the plots is greater in the ranges below 10 Å than above.

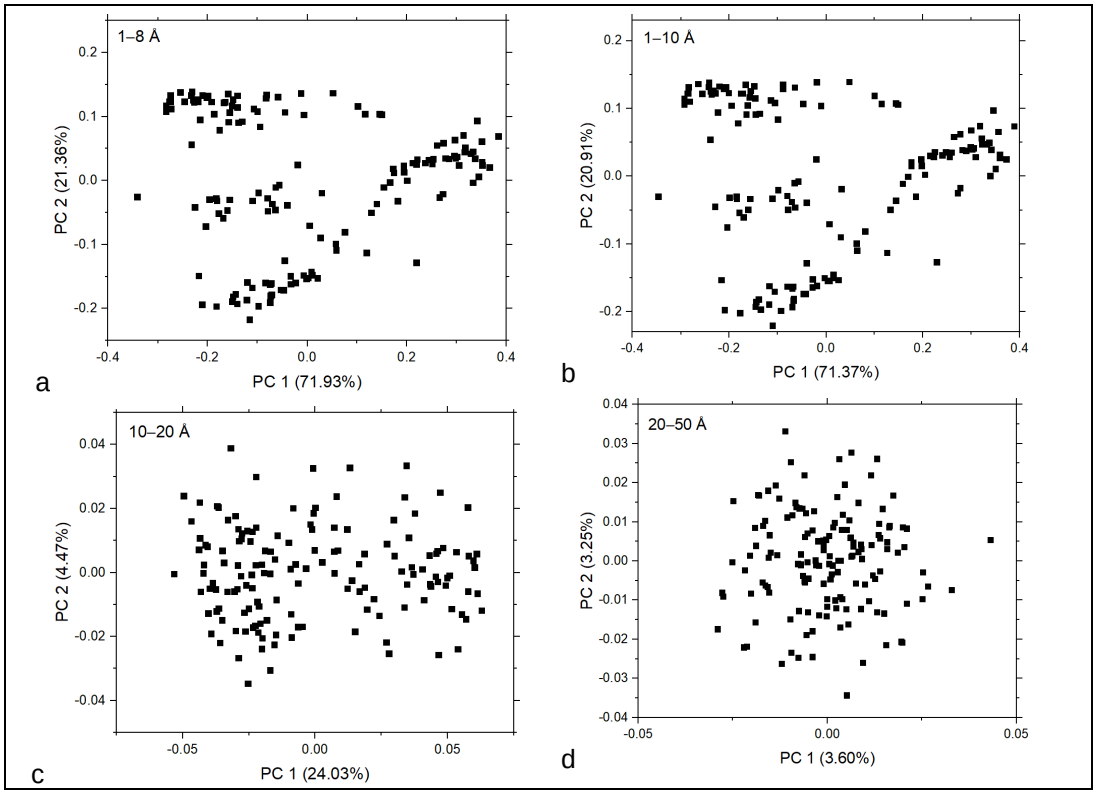


Figure 4.20 PCA: run 1 G(r): comparing the variation seen in shorter-range and longer-range interactions

PC1 v PC2 for a) 1–8 Å, b) 1-10 Å, c) 10–20 Å and d) 20-50 Å

Short- to medium-range interactions (< 15 Å) are key in the analysis of GuHCl solution structures, where changes in hydrogen bonding (< 2.5 Å), and in the interactions between ions, including Gdm^+ dimers, as well as changes in the solvation of ions, ion pairs or small clusters, will be observed.

The loading plots give information on the coefficients applied to each variable (in this case, r-values) so that the r-values that most influence the variation in each PC can be identified. There are key r-ranges that influence a PC more than others, as seen in the loading plots in Figure 4.21a&b.

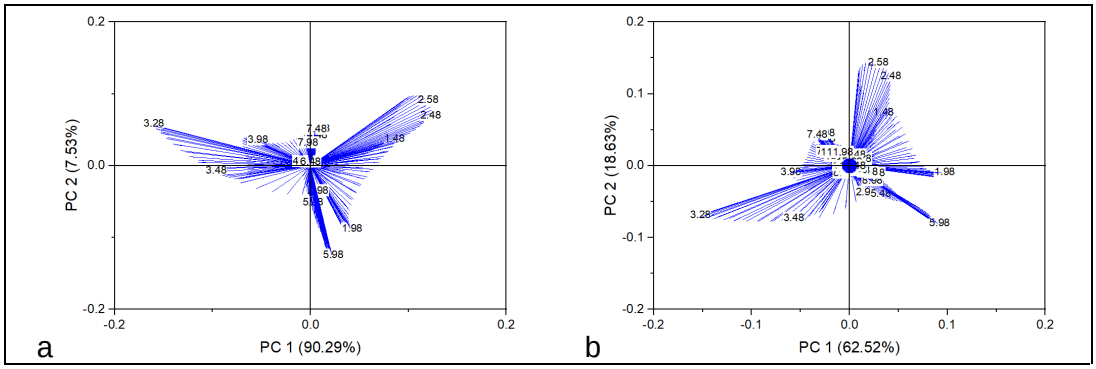


Figure 4.21 Loading plots for PC 1 and 2 for short-medium range interactions: a) 1–8 Å and b) 1-50 Å

The variation in PC1 is most heavily influenced by changes in structure at around 2.6 Å and 3.3 Å in the 1–8 Å range, and around 2.0 Å, 3.3 Å and 6.0 Å in the longer range. Variation in PC2 is greatest around 2.6 Å and 6.0 Å in the 1–8 Å range, and 2.6 Å, 3.5 Å and 6.0 Å at 1–50 Å.

The maximum variation in the datasets occurs in the range 1–8 Å (total variation = variation from PC1 (90.29%) + variation from PC2 (7.53%) = (97.82%), compared with 81.05% for 1–50 Å). Therefore, analysis of the data clustering has been carried out on the data r-range 1–8 Å. PCA is not in itself a clustering technique, but groupings can be indicative of the datasets with similar profiles [469].

Clustering of the datasets has been carried out using K-means cluster analysis, which groups the points by minimising the distance between the points and the centre of the cluster. To optimise the number of clusters, the sum of squares of the distance between the central point of the cluster to each of the points is calculated for 2 to n clusters. Plotting these values indicates an 'elbow' point (Figure 4.22), which corresponds to the optimal number of clusters [470].

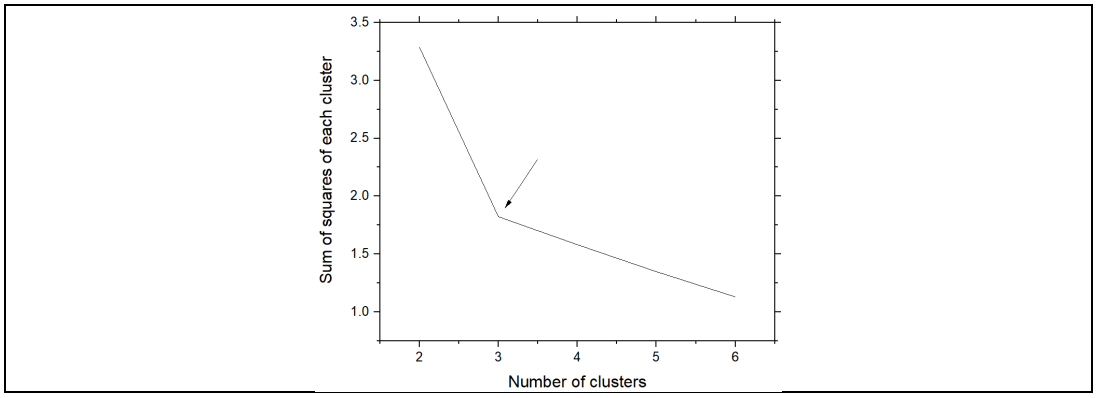


Figure 4.22 K-means analysis – finding the optimal number of clusters in the data

In this case, where frames 0–146 (solution phase only) are considered over the range 1–8 Å, the optimal number of clusters is 3, coloured black, red and gold in Figure 4.23.

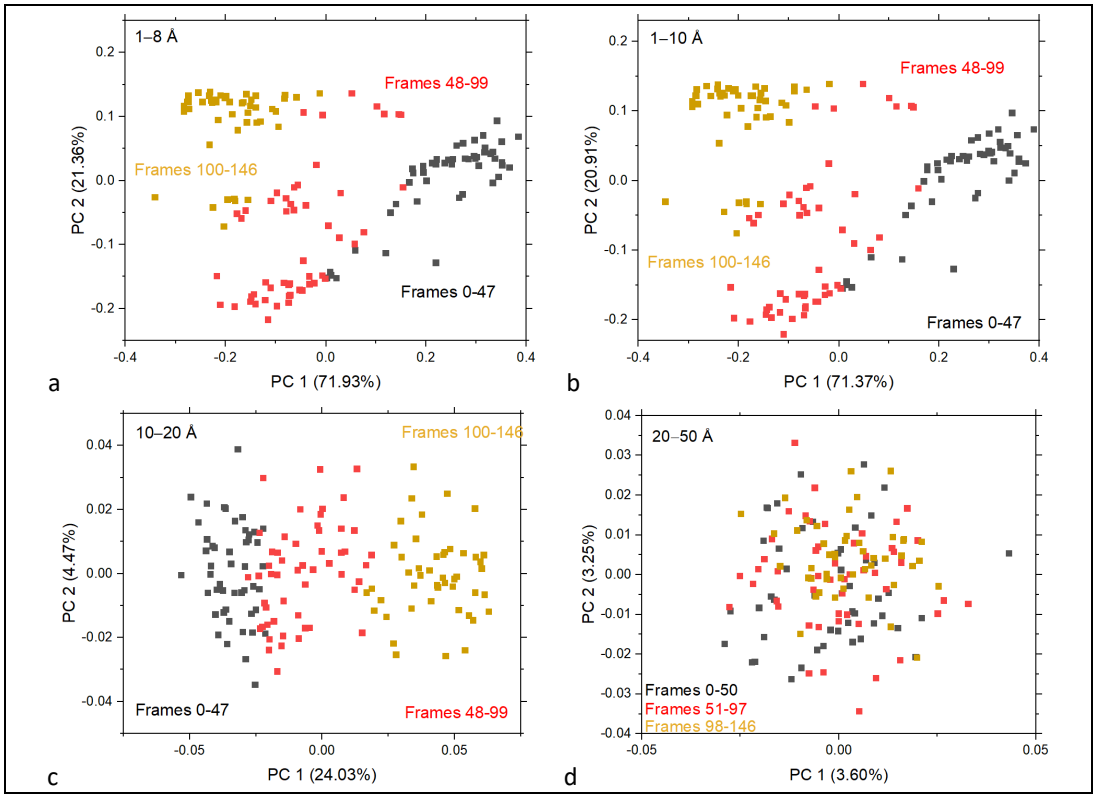


Figure 4.23 Run 1 G(r) PCA clustered score plots for a) 1–8 Å , b) 1–10 Å, c) 10–20 Å and d) 20-50 Å

The three clusters identified by K-means analysis of the solution-phase G(r) patterns over the range 1–8 Å are indicated in each score plot: stage 1 (black), stage 2 (red) and stage 3 (gold).

In addition to these 3 clusters, the patterns with distinct long-range order (frames 148-150) are shown as green, and the frame at which the transition to crystallisation takes place (147) is blue (Figure 4.24), as shown in Table 4.5.

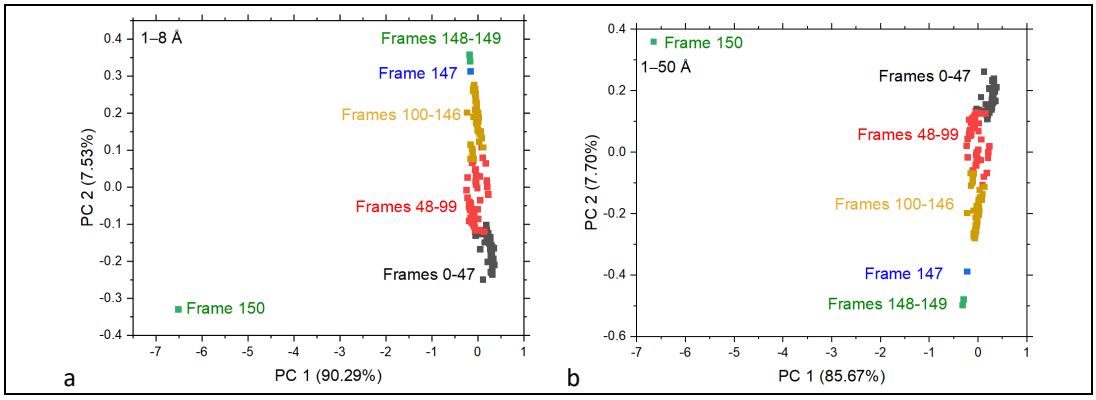


Figure 4.24 Run 1 G(r) PCA score plots for all *in situ* data a) 1–8 Å and b) 1–50 Å

Solution-phase clusters: stage 1 (black), stage 2 (red) and stage 3 (gold), plus the transition (blue) and crystallisation (blue) stages.

Table 4.5 PCA-derived stages for Run 1

	Frame numbers	Temperature /°C	Average cooling rate /°C min^{-1}
Stage 1	0-47	62.5–51.6	0.47
Stage 2	48-99	51.4–36.0	0.60
Stage 3	100-146	35.5–19.8	0.69
Transition	147	19.6	-2.26
Crystallisation	148-150	20.7–20.0	0.47

The groupings have been applied to the G(r) difference patterns (Figure 4.25a). The overlaid G(r) difference plot[27] (Figure 4.25b) identifies more clearly the position of the changing interactions as the solution cools. From stage 1 to stage 2, there is decreasing intensity around 2 Å. These changes in short-range interactions could represent changes to hydrogen bonding between Gdm^+ or Cl^- and water, or even between water molecules if small clusters of water are present. An increase in intensity is also seen at ~3.2 Å.

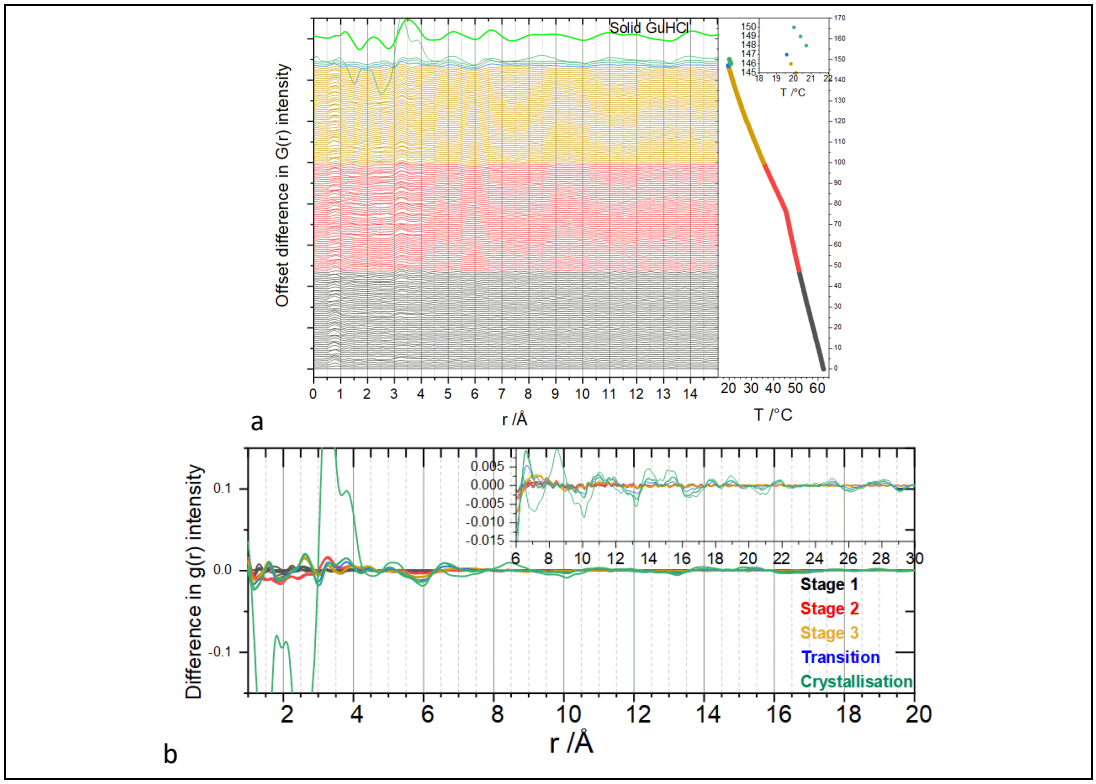


Figure 4.25 Run 1 – a) G(r) difference plots (offset), and b) overlaid G(r) difference plots (the central frames of date from stages 1-3, plus the transition and crystallisation stage G(r) difference patterns)

The clustering has been applied to the G(r) difference patterns (G(r) – G(r) of frame 0). a) shows the groupings of the offset difference patterns, and b) shows differences between the G(r) of frames of data from the middle of each stage at ~57.0°C, 46.5°C and 26.5°C. similarities between the G(r) difference patterns can be seen within each stage, whilst distinct differences between the stages can be seen.

[27] G(r) of a given frame – G(r) of the first frame

PDFgui analysis of solid GuHCl showed that there are N-Cl^- and C-Cl^- separations of 3.3 Å and 3.8 Å (Appendix A.3.2.1), which cause the change in the double peaked feature at 3–4 Å. The increasing intensity from Stage 2 could indicate the interaction of Gdm^+ ions (C-C or C-N).

From stage 2 to 3, there changes in intensity up to ~7 Å, suggesting some restructuring of molecular clusters or changes to solvation. The long-range structure starts to appear in the transition stage, along with increased intensity around 5 Å and 6.5 Å.

It is acknowledged that the time resolution of the data collection, with 30 s XPDF scattering frames, means that there could be a series of structural changes during the transition stage. However, this study also shows the potential to capture these changes with higher-brilliance X-rays, where shorter data collection with higher counting statistics will be possible.

The clustering of the XPDF patterns in Cooling Crystallisation Run 1 should be compared with similar analysis for Cooling Crystallisation Run 2 to establish if the stages are consistent (see Section 6.2)[28].

4.3.2.1.3 Cooling crystallisation: Run 2

In the second cooling crystallisation run, XPDF scattering data was collected in 30 s frames whilst the crystallising solution was cooled from 62.6°C (first data collection) until just after the appearance of crystals, which occurred around 20°C. In total 146 frames were collected in two scans (after the first 120 frames, there was a short pause (~20 s) in data collection whilst the scan reset). Run 2 was stopped more quickly than Run 1 after the appearance of crystals, with only one frame of data being collected after the crystal were visible from the control room.

The initial analysis of the XPDF patterns using a water in Kapton background (Appendix D.1) showed that the visual grouping of the frames of data in Run 2 gave similar temperature changes for the transition between stages 1 and 2, and stages 2 and 3. Therefore, it is likely

[28] Initially, the XPDF pattern analysis was carried out with a background of water in Kapton in GudrunX the processing step. In this analysis, solution temperature of the three proposed stages in the MSZ during Cooling Crystallisation runs 1 and 2 were consistent. It would therefore be probable that applying the empty Kapton background to the Cooling Crystallisation run 2 data would present similar clustering as seen here for run 1.

that the process of the Run 2 X-ray total scattering data using an empty Kapton background should give similar results to those found in the cooling crystallisation run 1 analysis.

4.3.2.1.4 Heating the solution: dissolution

Between Runs 1 and 2, X-ray total scattering data were collected whilst the solution was reheated (30 s frames, 18.8–62.7°C). 53 data frames were collected (the 54th has been omitted here as there was a lot of noise in the data, possibly due to an air bubble forming in the Kapton tube).

It is also suggested in Section 6.2 that the X-ray total scattering data collected during the reheating of the solution should be analysed in the same way to review the changes to the solution structure during dissolution.

4.3.2.1.5 Summary of the XPDF pattern analysis

Analysis of the XPDF patterns has identified that there are temperature-related structural changes that occur in the short to medium-range of intermolecular interactions in Run 1 when an empty Kapton background is used for background subtraction, and for both Run 1 and 2 when water in Kapton is used for the background subtraction (Appendix D.1) . PCA identified the variation of the solution-phase data that was captured by PC 1 and 2 when applied to different r-ranges. K-means clustering analysis identified that there were three clusters of data in the MSZ.

It is likely that stage 1 is the unsaturated solution phase,[29] as the starting temperature of the measurements was selected to be higher than the temperature at which the solute appeared to be dissolved, to ensure dissolution was complete[30]. The second stage may indicate the onset of supersaturation, which will be the driving force for crystallisation. The third stage would also therefore be in a supersaturated state, although analysis of the G(r) difference plots suggests there are structural differences between stages 2 and 3 (Figure 4.25). These differences may be due to the different clustering of the ions in solution or changes in ion hydration. These aspects will be explored in 4.3.2.2.

Acquiring the XPDF patterns over a shorter time interval could provide further information on the order in which different interactions emerge, especially during the transition stage.

[29] The extent of undersaturation should be measured to fully characterise the system

[30] Also based on analysis with Crystal16

Data collection with the necessary high-resolution may be possible with X-ray Free-Electron Laser (XFEL) or the ESRF-EBS (Section 2.3).

EPSR analysis may identify the changes in intermolecular interactions in the short to medium range that result in the changes in the G(r) patterns , shedding light on the structural changes that are occurring in the solution. The changes between 1.3 Å and 2.2 Å could relate to changes in solvation of the ions, as this interaction distance is in the hydrogen bond range.

4.3.2.2 *In situ* crystallisation EPSR analysis

EPSR simulations were constructed to compare molecular models of the solution structure at different points in the MSZ before crystallisation, identified as stages 1, 2 and 3 by PCA in Appendix D.1. The simulations of the last four data frames were undertaken to see how EPSR would handle the simulation. Inspecting the molecular models by eye revealed no clear crystalline structures. This outcome confirmed the limitation of EPSR with respect to representing a system containing crystallised phases, as EPSR has not been developed to handle Ewald summation, which is essential to model crystalline structures.

Crystalline materials in solution pose an added complexity with the presence of a solid phase suspended in a liquid phase [407]. It is not known if the sample passing through the Kapton tube and scattering the X-rays is representative of the bulk solution, for which the density has been measured. It is possible that a lower concentration solution, due to some of the solute having formed a solid crystal product, was being measured. EPSR simulations cannot describe the solution where the crystal phase is present, as the molecular structure in the EPSR model is not representative of the real solution after crystallisation due to the limitations of EPSR modelling crystalline material in solution. No simulations from the transition or crystallisation stages are included here.

The EPSR analysis described here was undertaken to model molecular structures for a range of temperatures in the MSZ (Table 4.6), with the output of frames of data within the MSZ stage being combined to calculate average g(r).

Table 4.6 EPSR simulations for structural modelling of XPDF data from the cooling crystallisation experiment

	Frame number	Simulation temp. /K	Stage (from PCA)
1	GuHCl_0	335.7	**Stage 1**
2	GuHCl_3	335.1	**Stage 1**
3	GuHCl_4	334.7	**Stage 1**
4	GuHCl_28	329.1	**Stage 1**
5	GuHCl_29	328.9	**Stage 1**
6	GuHCl_30	328.6	**Stage 2**
7	GuHCl_60	322.0	**Stage 2**
8	GuHCl_89	313.3	**Stage 2**
9	GuHCl_114	303.3	**Stage 3**
10	GuHCl_129	298.0	**Stage 3**
11	GuHCl_145	293.3	**Stage 3**

As described previously, the analysis of the XPDF patterns was initially undertaken using the water in Kapton X-ray total scattering data for the background subtraction in the data processing stage. This approach was later revised and undertaken using X-ray total scattering data of empty Kapton for background subtraction (Section 4.3.2.1). In both cases, three phases with different solution structures were identified, although the temperatures at which these changes occurred were different (Table 4.7).

Table 4.7 Comparison of the solution structure stage transitions

	Original analysis (water in Kapton background, Appendix D.1.1)		Revised analysis (empty Kapton background, Section 4.3.2.1)	
	Frame numbers	Transition temperature /°C	Frame numbers	Transition temperature /°C
Stage 1	0-29	62.5	0-47	62.5
Stage 2	30-113	55.5	48-99	51.4
Stage 3	114-146	30.1	100-146	35.5
Transition	147	19.6	147	19.6
Crystallisation	148-150	20.7	148-150	20.7

Although the original analysis placed simulation 6 (frame 30) in stage 2, the revised approach to clustering the data places simulation 6 in stage 1. However, the work presented here is based on the original clustering. Given the limitations on the EPSR modelling presented here (due to resources[31], the software's ability to model crystallising systems and the relatively low counting statistics of the X-ray total scattering data compared with what could be achieved with a fourth generation beamline today), it was felt that keeping the original simulation groupings would still allow the potential for EPSR as a modelling tool to be demonstrated, whilst accepting that improvements could be made to refine the modelling in future work.

Key parameters of an EPSR simulation include the system temperature and atomic number density. The simulations were set up based on the temperature recorded at the start of the data collection and the solution atomic number density based on this temperature. The simulation box parameters are in Table 4.3.

Following a discussion of the goodness of fit of the EPSR simulations, the evolution of the intermolecular interactions in the solution as it transitions between stages is discussed with

[31] Such as the technique used to measure the solution density at the required temperature and the computational power available for running EPSR, which limited the possible box size and number of iterations of the model, as there was limited access to laboratory and computational facilities as a result of the Covid lockdown.

accompanying partial pair distribution functions $g_{\alpha\beta}(r)$ and coordination number (CoordN) data. Some partial pair distribution function $g_{\alpha\beta}(r)$ plots showing the simulated $g_{\alpha\beta}(r)$ (averaged by stage) are included within the discussion. All $g_{\alpha\beta}(r)$ plots (grouped and averaged by stage) are available in Appendix D, along with all CoordN data tables[32]. Finally, the water structure in the solution and how it changes across the stages is analysed.

4.3.2.2.1 F(Q) and g(r) of simulations

The simulated structure factors F(Q) (Figure 4.26) and g(r) (Figure 4.27) for selected GuHCl solutions - one model is shown here in stages 1-3, as identified in the PCA of the XPDF patterns - are compared with the experimental data.

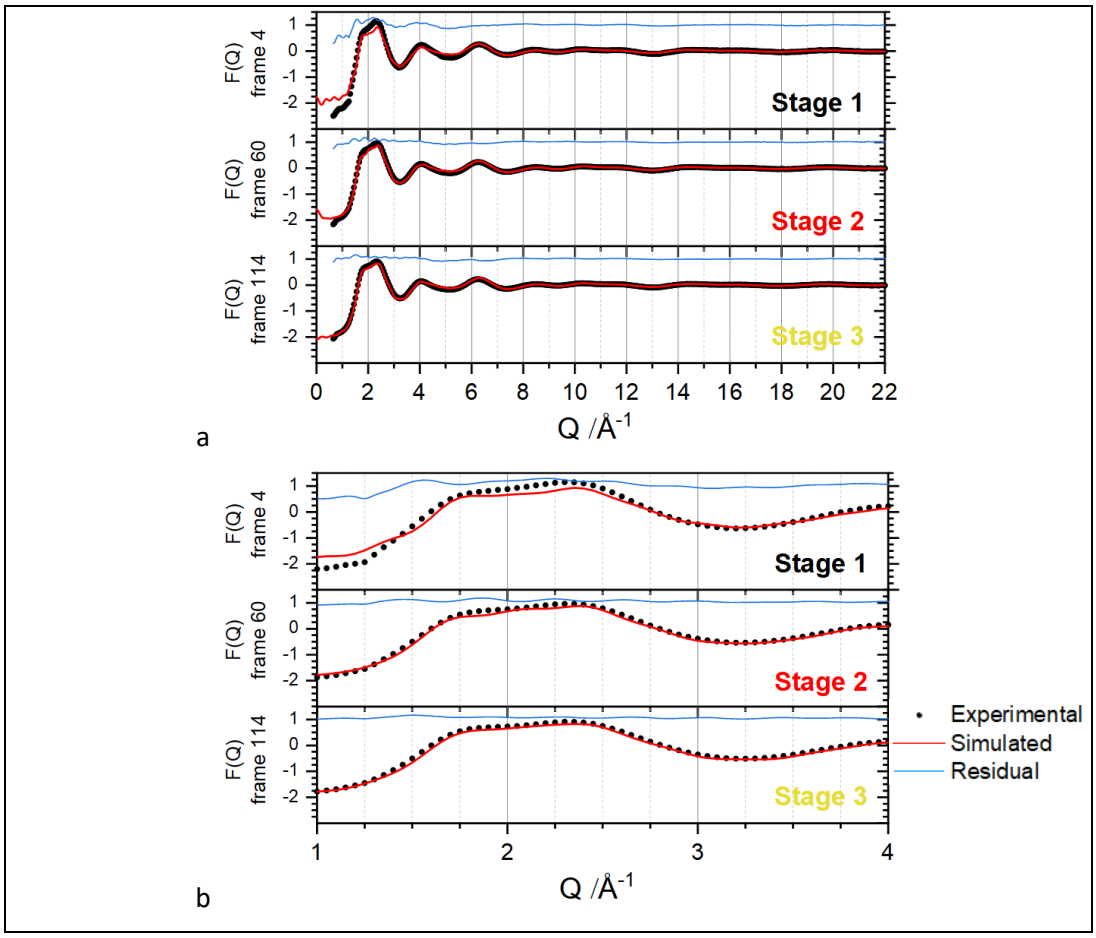


Figure 4.26 Modelled vs. experimental F(Q) for GuHCl EPSR models a) 0–22 Å^{-1} b) low-Q

The models of the solution phases show a good quality of fit agreement at low-r values (> 1 Å), so they may provide some insight into the local structural changes observed in the XPDF patterns across the MSZ (Section 4.3.2.1).

[32] The standard deviation SD, of the CoordN data is shown and has been calculated on a sample basis, as not all frames in each stage have been simulated.

The g(r) plots (Figure 4.27a&b) are up to r values of ~19 Å, as the size of the simulation box limits the range over which the PDF can be calculated. However, this range is sufficient to analyse the short and medium-range order interactions. The size of the simulation boxes used in the simulations was limited by the access to computational power to run the simulations[33]. It had been confirmed by the team maintaining the EPSR software that simulations of ~1,000 molecules should be sufficient to represent solution samples [471]

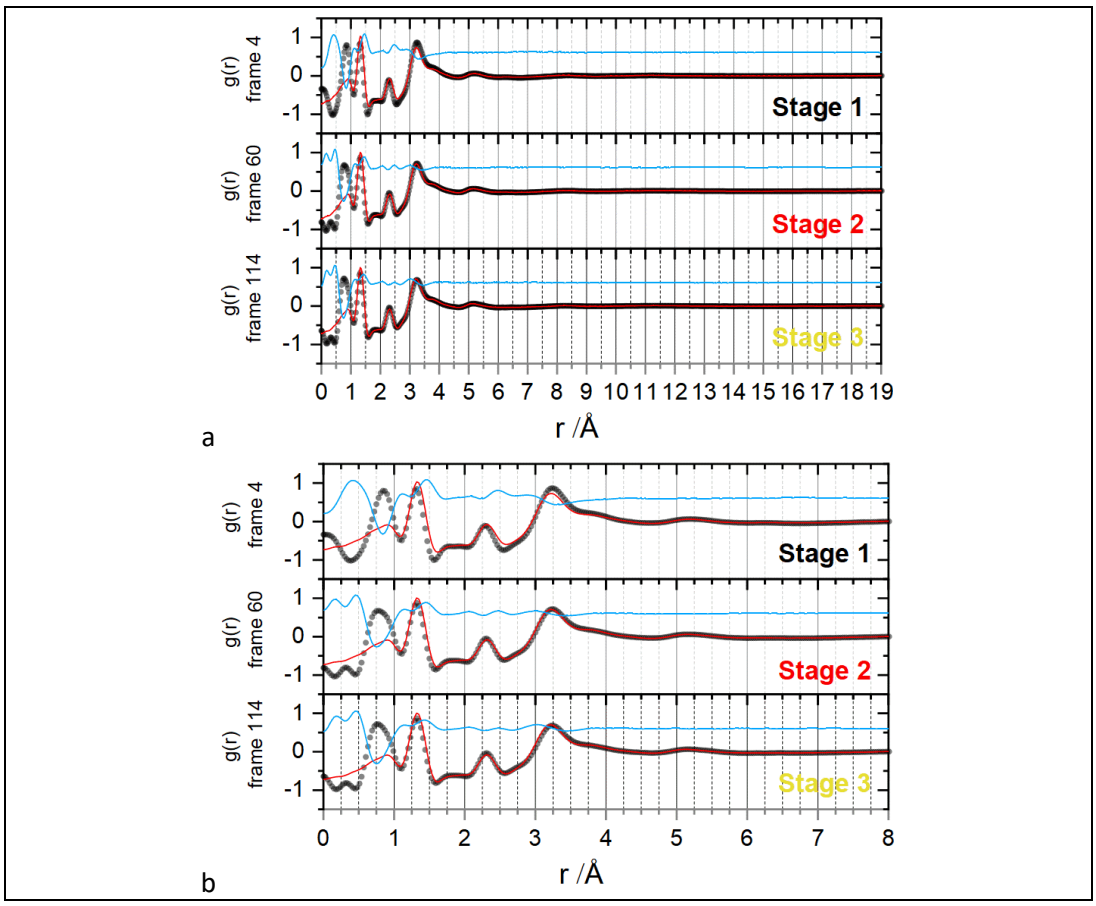


Figure 4.27 Modelled vs experimental g(r): a) 0–19 Å and b) 0–8 Å for GuHCl EPSR models

Intermolecular partial PDFs $g_{\alpha\beta}(r)$, have been extracted from each simulation. The $g_{\alpha\beta}(r)$ from simulations in the same stage, as defined by the original clustering analysis (Appendix D.1), have been averaged to derive a stage-averaged $g_{\alpha\beta}(r)$.

[33] The remote desktop connection to the PC on campus dropped out near the start of the lockdown in March 2020 and could not be reliably maintained. It was not possible to run the EPSR simulations on the high performance computers (HPC) at the University of Leeds, as EPSR required use of the interactive node, which was not available for long-term use.

4.3.2.2.1.1 Evolution of the aqueous GuHCl solution

PC analysis clustered the XPDF patterns in three stages prior to crystallisation[34]. It is proposed that stage 1 is the undersaturated solution that transitions to a supersaturated solution (stage 2) at 51.4°C. There is a change in the local structure of the supersaturated solution at 35.5°C (stage 3).

The local and medium-range structural motifs that may contribute to the XPDF patterns through the MSZ are proposed in Figure 4.28, and are based on EPSR simulation analysis (such as comparison of interatomic $g_{\alpha\beta}(r)$, coordination numbers and molecule clustering in the different stages) and studies of Gdm^+ hydration and like-charge contact ion pairs (CIP) in the literature (computational studies [91,92,179,181–183,458], or a combined experimental and computational approach [92,97,181,188]). It is not suggested that the motifs described are present exclusively in the stages to which they are associated in Figure 4.28, but that they may be a dominant molecular arrangement in the solution during that stage, leading to differences in the X-ray total scattering patterns.

4.3.2.2.1.2 Ion-ion interactions

PCA of the XPDF patterns indicate short and medium-range structural changes (< 20 Å) in the MSZ stages 1, 2 and 3. $g_{\alpha\beta}(r)$ for C-C, C-N, N-N and Cl^--Cl^- (figures in Appendix D) for all EPSR simulations show statistically-significant features up to 10–12 Å, after which the $g_{\alpha\beta}(r)$ oscillate close to 1.[35] In the case of $g_{Cl^-Cl^-}(r)$, three clear coordination shells are present. These observations suggest that small molecular clusters of solvated ions form in the solution, even in the undersaturated solution (stage 1), and it is the structures of these clusters that evolve through the MSZ.

The 3D distributions of water, Cl^- and Gdm^+ sites around a central Gdm^+ ion is revealed in the spatial probability density (SPD) plots (Figure 4.29a-c). The evolution of these sites is seen as the solution cools through the structural stages in the MSZ. At each stage, there is a strong tendency for water and Cl^- to compete for interaction sites around the plane of Gdm^+, mainly in the interstitial space between two $-NH_2$ groups. Although Gdm^+ principally interacts with the hydrophobic faces of the central Gdm^+ ion, the probability of interactions around the

[34] The original analysis, where a water in Kapton background was used in the data processing stage: Appendix D.1. Revised analysis using a blank Kapton background (Section 4.3.2.1.2).

[35] These clusters are also seen in the crystallisation stage simulation, as a significant quantity of GuHCl remains in solution after crystallisation.

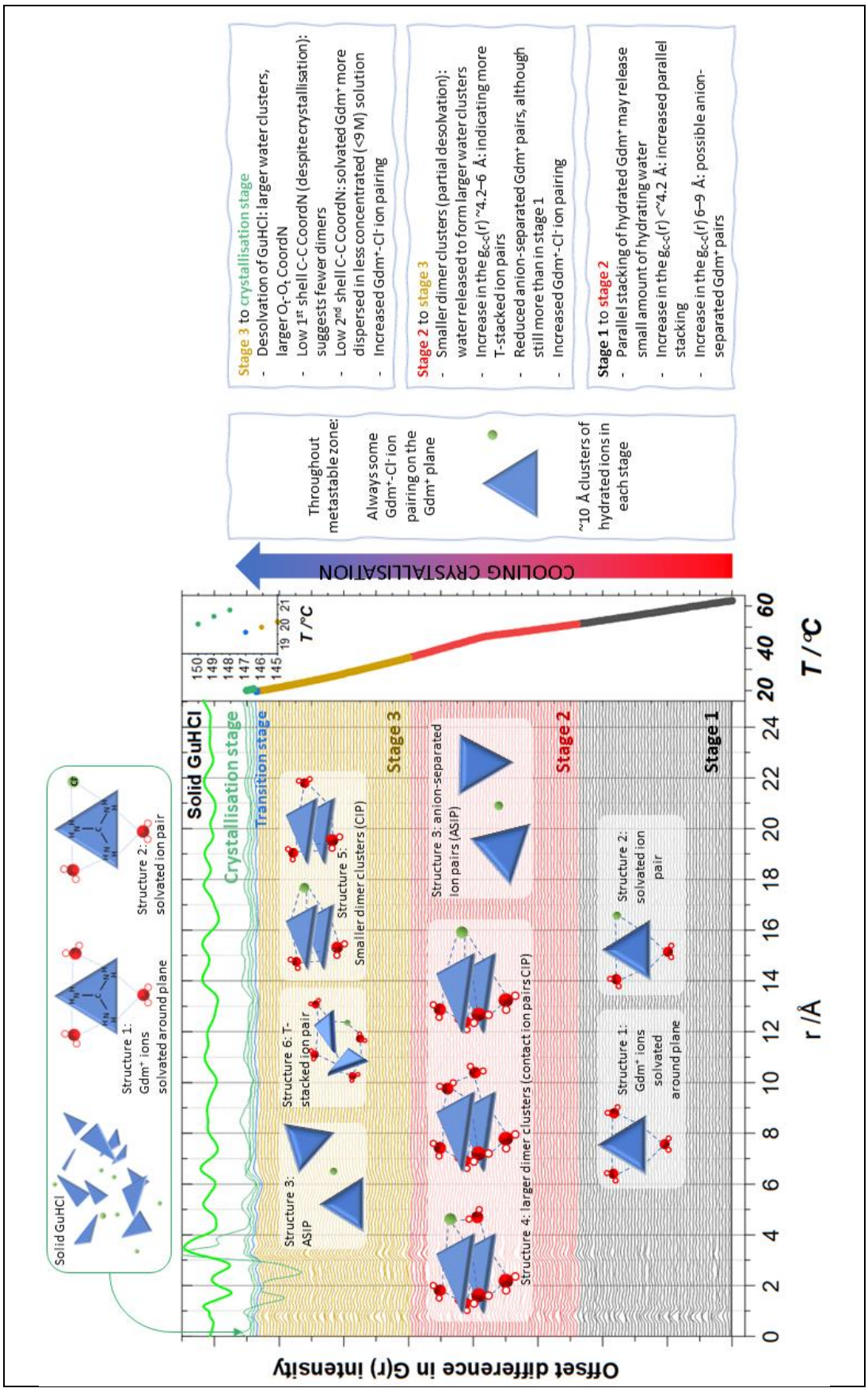


Figure 4.28 Summary of structural evolution of aqueous GuHCl solution during cooling crystallisation.

Schematics of the structural motifs that may be common in each stage.

plane of the molecule increases from stage 1 through to stage 3 (seen in 2.7x average molecular density). The SPD plots indicate that Gdm^+-Cl^- pairing is highly probable in all three stages and that Gdm^+-water hydrogen-bonding via the H atoms in the plane of the Gdm^+ is more probable than via the central carbon.

Gdm$^+$-Gdm$^+$ interactions: At 4x component average molecular density (middle-left in Figure 4.29a-c), the probability of parallel-stacked Gdm^+ ions being present increases between stages 2 and 3. The GuHCl crystal contains offset parallel-stacked and T-stacked Gdm^+ (Figure 4.30a&b).

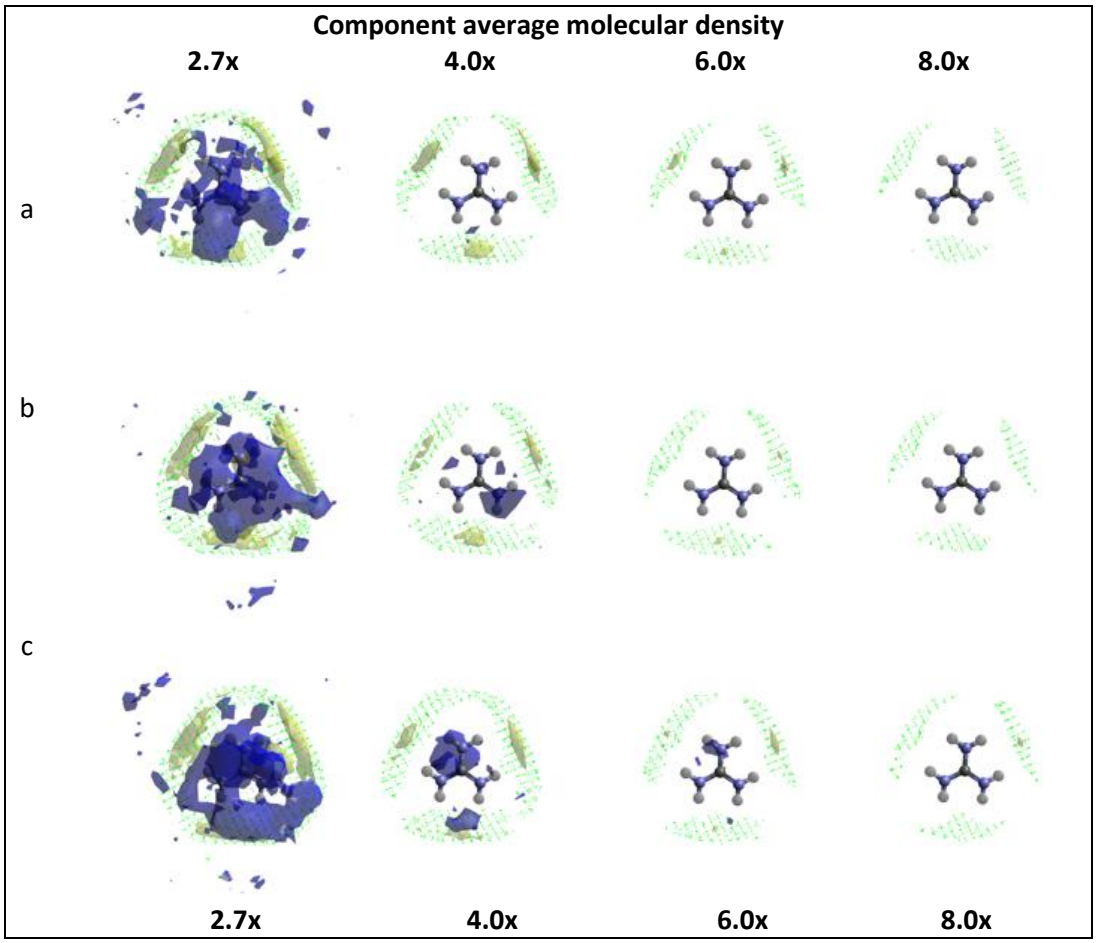


Figure 4.29 Spatial probability densities showing the distribution of Gdm^+ (blue), water (yellow) and Cl^- (green) sites around a central Gdm^+ ion

Left to right: 0.014, 0.021$^{\alpha}$, 0.032, 0.043$^{\beta}$ ions Å^{-3} and 0.048$^{\delta}$, 0.072$^{\varepsilon}$, 0.098$^{\zeta}$ water molecules Å^{-3} (2.7, 4.0, 6.0 and 8.0 times the average molecular density of the Gdm^+ ions, Cl^- ions and the water molecules)

$^{\alpha}$0.022 (GuHCl_129) $^{\beta}$0.042 (GuHCl_3) $^{\delta}$0.049 (GuHCl_129) $^{\varepsilon}$0.073 (GuHCl_129) $^{\zeta}$0. 096 (GuHCl_129)

a) **Stage 1**: GuHCl_3, b) **Stage 2**: GuHCl_60, c) **Stage 3**: GuHCl_129

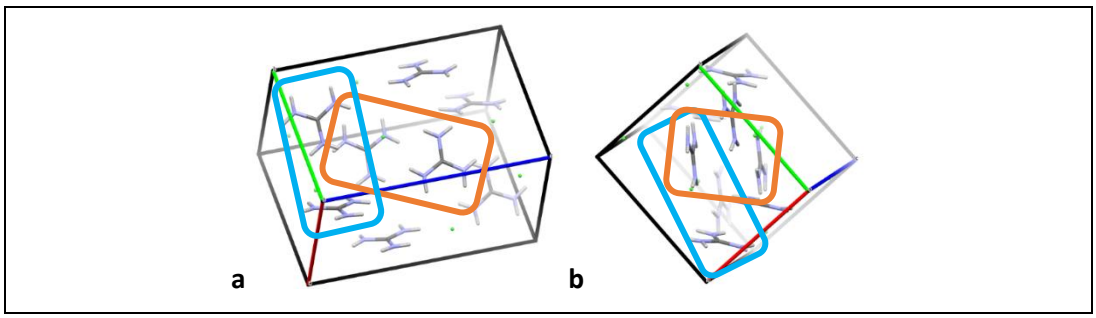


Figure 4.30 Solid GuHCl structure motifs

Structures derived from SCXRD: a) offset parallel stacking (orange) b) T-stacking (light blue)

C-C interactions: In the crystal structure, the shortest C-C interaction (4.7 Å) is between offset parallel-stacked Gdm^+ (Figure 4.30), but closer C-C interactions are observed in solution.

There are no solvent molecules between a contact ion pair (CIP) (Figure 4.31a). Like-charged Gdm^+ ions form parallel-stacked dimers, which can have C-C interactions ≤4 Å (structure 4 or 5 in Figure 4.28). These dimers could stabilised by surrounding water molecules, possibly as few as three water molecules per dimer according to an *ab initio* dimerisation study (the study did not include counterions) [91]. In aqueous GuHCl solutions, water and Cl^- compete for interactions sites around the plane of Gdm^+ and therefore Cl^- ions could form bonds with a pair of parallel-stacked Gdm^+ to stabilise the dimer in place of water [97].

Gdm^+ ions also form T-shaped CIP in solution with ~5 Å C-C separation (Figure 4.31b and structure 6 in Figure 4.28) [93,179].

It is proposed in the literature that solvent-shared ion pairs (SShIP), have separation of ~6.3 Å in a potentials of mean force study [472], ~7 Å (from DFT) [183] to 7.7 Å (MD study) [182] (Figure 4.31c&d). Inspection of the molecular models revealed examples of anion-separated Gdm^+-Gdm^+ ion pairs (ASIP) forming with a separation of ~8 Å (Figure 4.31e and structure 3 in Figure 4.28), which is a logical extension of the $g_{C\text{-}Cl}$-(r) first peak position being ~3.8 Å (Figure 4.37). Gdm^+-Cl^- ion pairs are formed either by Cl^- bonding with two amino groups, or by a longer, more linear hydrogen bond with one H on Gdm^+ (Figure 4.31f).

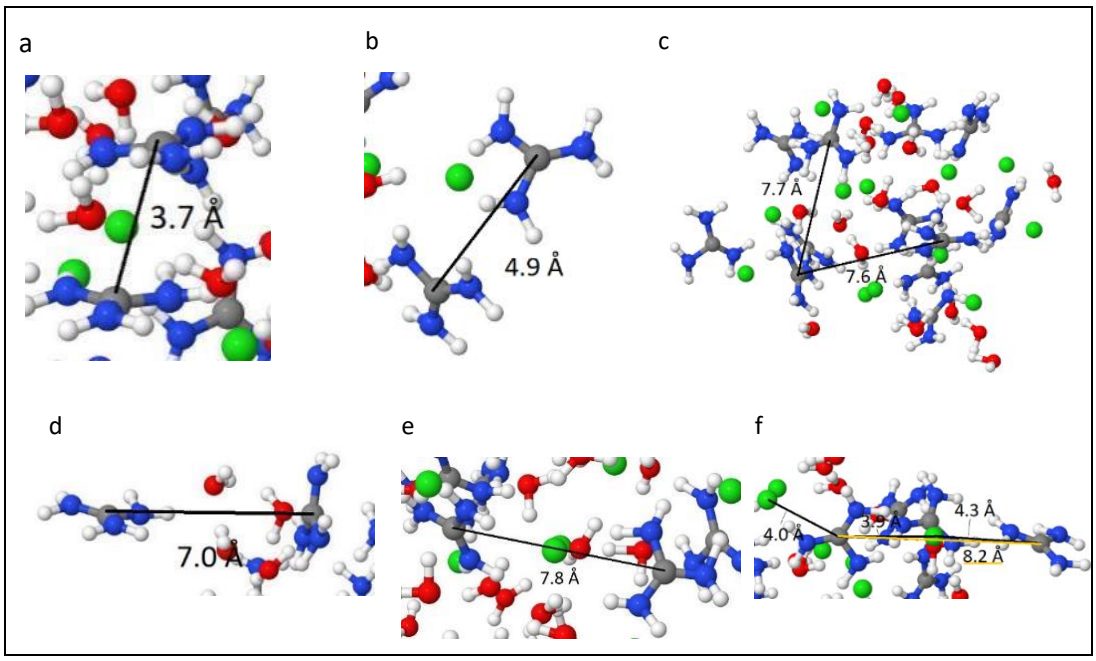


Figure 4.31 C-C interactions in the EPSR molecular models

a) parallel-stacked contact ion pairs (CIP), b) T-stacking, c&d) solvent-shared Gdm^+ ion pair (SShIP), e) anion-separated Gdm^+ ion pairs (ASIP) & f) C-Cl^- ion pairs

$g_{C\text{-}C}(r)$ (Figure 4.32) shows two distinct coordination shells in all solutions, with peaks at ~4.8 Å and ~7.5 Å. The position of the first peak decreases from ~4.8 Å to ~4.7 Å (stage 1 to stage 3 in Figure 4.32, #). In stages 2 and 3 the first peak becomes narrower and sharper, suggesting increased local order. The first peak shoulder feature at ~4 Å (Figure 4.32 inset (*)) is present to some extent in all simulated $g_{C\text{-}C}(r)$ but was more prominent in GuHCl_114.

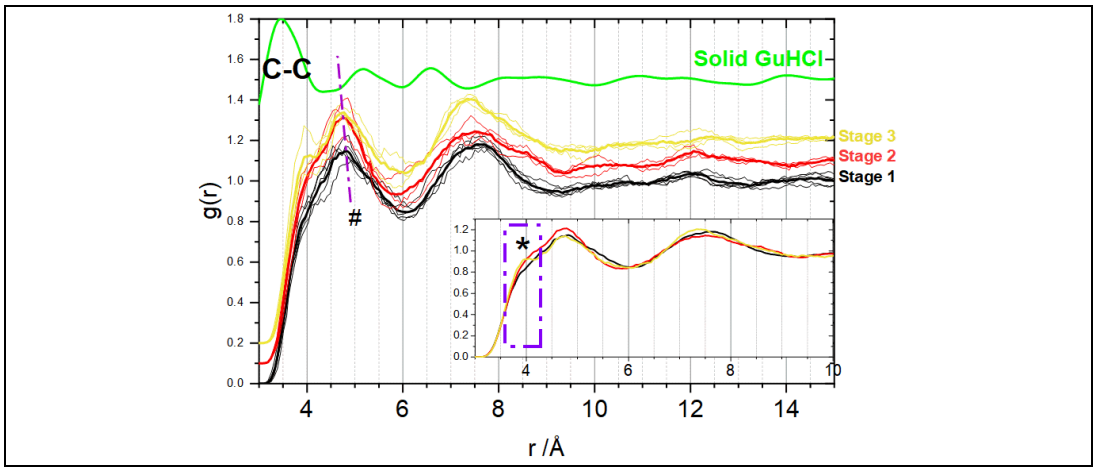


Figure 4.32 Offset C-C partial pair distribution functions $g_{C\text{-}C}(r)$

Grouped and averaged by stage (thick line), with the g(r) of solid GuHCl (light green)

The $g_{C\text{-}C}(r)$ second peak could be due to SShIP forming in the solution in the region 7–8 Å and ASIP at higher r-values (8–9 Å). The stage 2 and 3 peaks are more asymmetrical with increased intensity at ~8.5 Å. An increase in ASIP could account for these elongated peaks.

C-C CoordN has been evaluated for the first and second coordination shells (**Table 4.8**). Broadly speaking, the first and second peak CoordN are similar for stages 2 and 3, with lower coordination shell r_{max}, indicating potential structural differences, e.g. more ASIP and possibly SShIP, compared to stage 1.

Table 4.8 C-C CoordN – 1st and 2nd shells: averaged per stage

	C-C 1st shell			C-C 2nd shell		
	Ave. Range /Å	**Ave. CoordN**	***SD***	**Ave. Range /Å**	**Ave. CoordN**	***SD***
Stage 1	0–6.1	3.88	*0.04*	6.1–9.1	12.21	*0.08*
Stage 2	0–5.9	3.57	*0.08*	5.9–9.4	14.29*	*0.24*
Stage 3	0–6.0	3.73	*0.08*	6.0–9.3	13.95*	*0.08*

* CoordN up to 9.1 Å for stage 2 and 3 (same r_{max} as stage 1) = 12.66 and 12.80

The first peak CoordN has been further separated into the shoulder and main peak in **Table 4.9**. There will be some overlap between the shoulder and peak CoordN, but it could be used as estimate of the extent of parallel-stacking (shoulder) and T-stacking (peak) of Gdm^+ ions. Using this proxy suggests that the shoulder (< ~4.2 Å) has a high proportion of parallel-stacking, with more T-stacking or offset-parallel stacking for r = ~4.2—~6.0 Å. The T-:parallel-stacking CoordN ratio is highest in stage 1, reducing considerably in the supersaturated solutions (stages 2 and 3). This suggests close clustering of Gdm^+ in the supersaturated solutions occurs when the system thermodynamics make like-charged ion pairing energetically favourable. This difference will account for some of the structural changes detected between stages 1 and 2 in the XPDF patterns (Section 4.3.2.1).

Table 4.9 C-C CoordN – 1st shell shoulder and peak: averaged per stage

	C-C 1st peak shoulder (parallel-stacking)			C-C 1st peak, after shoulder (T-stacking)			T-:parallel-stacking CoordN ratio
	Ave. Range /Å	**Ave. CoordN**	***SD***	**Ave. Range /Å**	**Ave. CoordN**	***SD***	
Stage 1	0–4.1	0.45	*0.03*	4.1–6.1	3.43	*0.02*	7.68
Stage 2	0–4.2	0.58	*0.04*	4.2–5.9	2.99	*0.05*	5.18
Stage 3	0–4.2	0.59	*0.08*	4.2–6.0	3.14	*0.13*	5.39

The higher peak:shoulder CoordN ratio of stage 3 compared to stage 2 suggests greater T-stacking in stage 3. There are also more Gdm^+-Gdm^+ interaction sites around the plane of the central Gdm^+ observed in the stage 3 SPD plot (Figure 4.33c), compared to stage 2 (Figure 4.33b) support this idea.

The band of Gdm^+-water interaction sites around the plane of the Gdm^+ is broader in stages 2 and 3 (Figure 4.33b&c, highlighted with red box), indicating that the direction of the Gdm^+-water hydrogen bonds moves away from the plane of Gdm^+. These off-plane sites are likely

to represent the increased number of water molecules that are stabilising dimer formation by hydrogen-bonding with two Gdm^+.

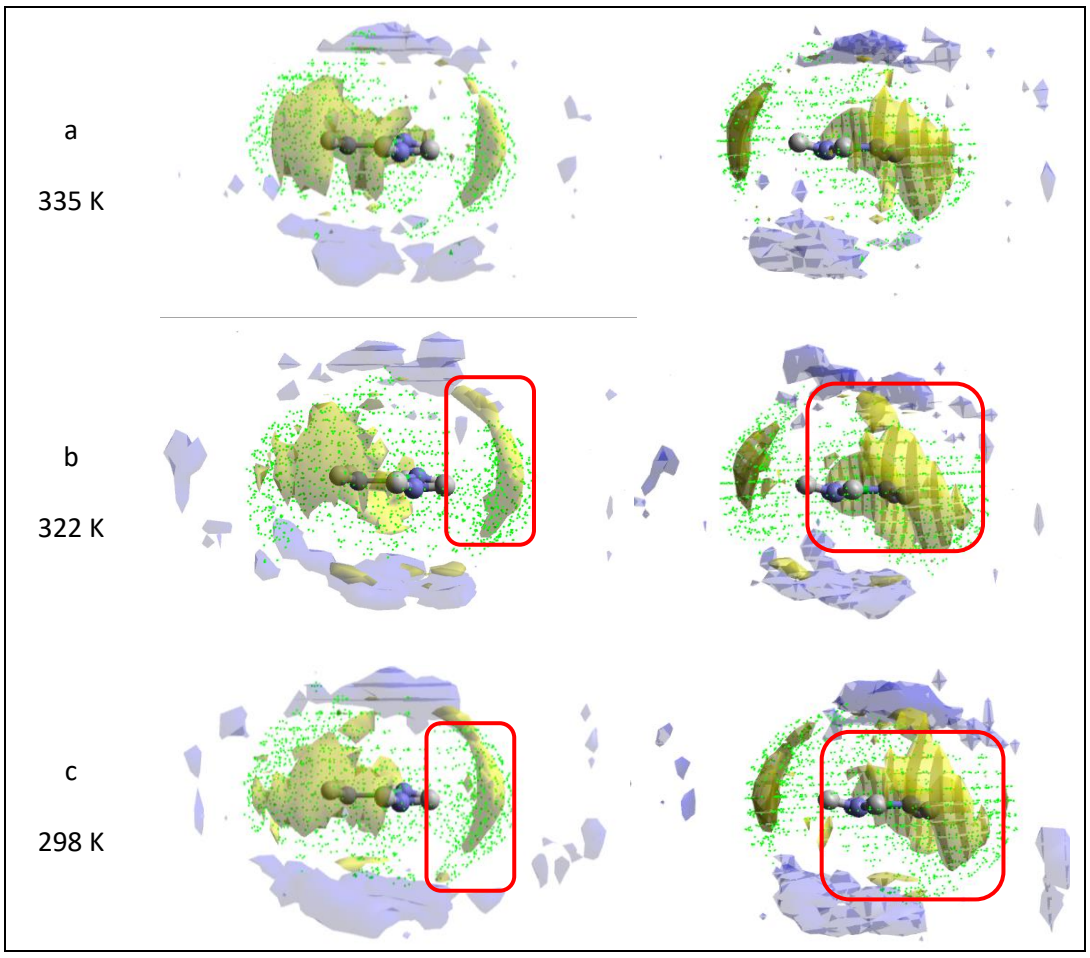


Figure 4.33 Spatial probability densities showing the distribution of Gdm^+ (blue), water (yellow) and Cl^- (green) sites around a central Gdm^+ ion: two projections per SPD

0.014 ions $Å^{-3}$ (2.7 times the average molecular density of the Gdm^+ and Cl^- ions) and 0.032 water molecules $Å^{-3}$ (2.7 times the average molecular density of the water)

a) **Stage 1**: GuHCl_3, b) **Stage 2**: GuHCl_60, c) **Stage 3**: GuHCl_129

C-N interactions: The main C-N interactions occur around 4.5 Å and 7.0 Å (Figure 4.34). The first peak position decreases from 4.6 Å (stage 1) to 4.4 Å (stages 2 and 3).

The interaction that intensifies in stage 3 (~5.75 Å, orange box Figure 4.34), but is less intense in stage 1, may be related to increased Gdm^+ T-stacking (structure 6 in Figure 4.28). The purple box in Figure 4.34 highlights the features between 8 and 10 Å. There is a single peak in stage 1 (9.0 Å) and double peaks in stage 2 (8.8 and 9.3 Å) and stage 3 (8.5 and 9.5 Å). Analysis of solid GuHCl indicates many C-N interactions at 8.6 Å and 9.2/9.3 Å, and between 9.8 Å and 10.0 Å (Appendix A.3.2.1).

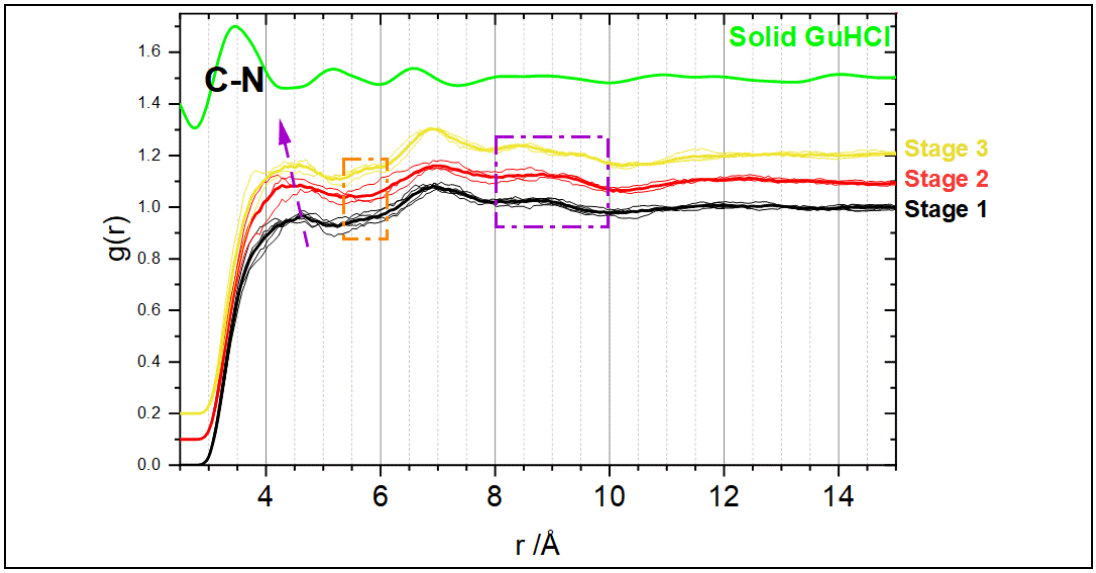


Figure 4.34 Average C-N partial pair distribution functions $g_{C\text{-}N}(r)$

Purple arrow indicates the reducing distance of the first interaction peak, orange box: highlights differing features between stages at ~5.75 Å, purple box: highlights differing features between stages at 8–10 Å, with g(r) of solid GuHCl (light green)

N-N interactions: There are distinct features in $g_{N\text{-}N}(r)$ (App. Figure 42) that are reasonably consistent with those in the solid GuHCl g(r) up to ~7 Å.

Short and medium-range features in $g_{N\text{-}N}(r)$ (purple box, App. Figure 42) could be key to differentiating between the two supersaturated solution structures (stages 2 and 3). The first peak position is at ~3.7 Å in stage 1, increasing slightly to ~3.8 Å in stages 2 and 3 (App. Figure 42, indicated with *). The peak intensities in stages 2 and 3 are higher (graph inset), which is also reflected in the increases N-N CoordN (App. Table 27), which could be due to dimer formation. The peak at ~6.5 Å has lower intensity in stage 2, increasing again in stage 3. This feature could occur if the local structure is dominated by dimerisation in stage 2, followed by increased T-stacking in stage 3.

Angular radial distribution functions (ARDF) of Gdm^+-Gdm^+: ARDF analysis splits the $g_{\alpha\beta}(r)$ into vector angles around a central molecule to show the local directional interactions of molecules. It has been applied in Figure 4.35 to visualise the probability of the plane of the surrounding Gdm^+ interacting with the central Gdm^+.

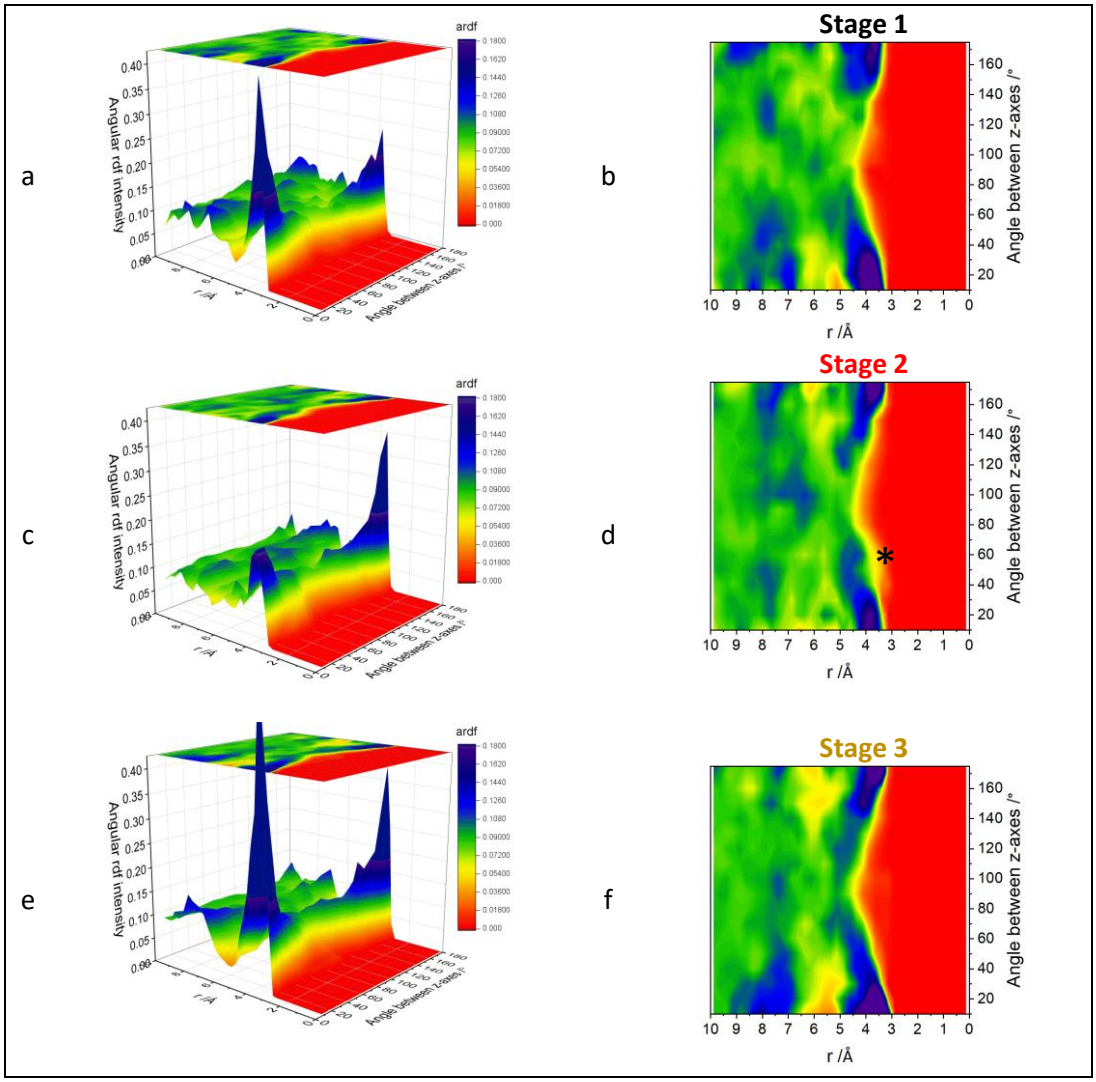


Figure 4.35 Gdm^+-Gdm^+ angular radial distribution function between the z-axes of Gdm^+ ions

a) **Stage 1**: GuHCl_3, b) **Stage 2**: GuHCl_60, c) **Stage 3**: GuHCl_129

ARDF are plotted as a function of the angle between the z-axes of the central Gdm^+ and the surrounding Gdm^+ ions for $0° \leq \theta \leq 180°$, where the z-axis of Gdm^+ is perpendicular to the plane of the ion (see Figure 4.36a for axis definitions).

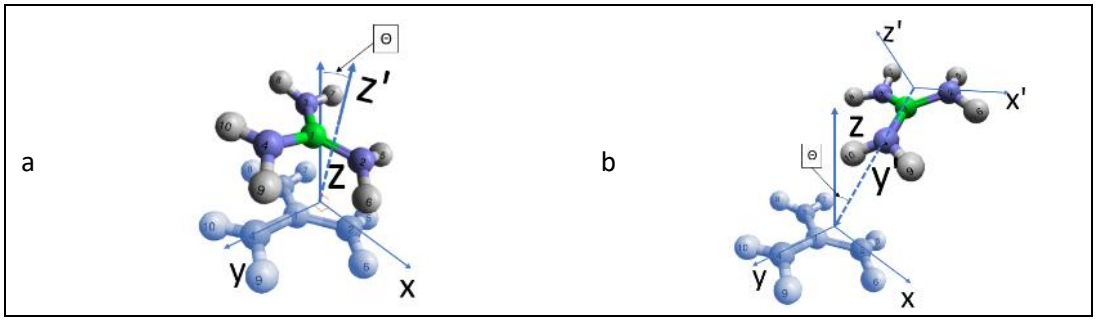


Figure 4.36 ARDF axis definitions for a) z-z axes interactions and b) z-y axes

In all stages, there are significant planar Gdm^+-Gdm^+ interactions, as each ARDF shows highest intensity around 0° and 180° (Figure 4.35a-f). In stage 2 (Figure 4.35c&d), the

highest-intensity regions (purple areas) around 0° and 180° appear over a reduced r-range (3.3–4.8 Å in stage 1 to 3.4–4.4 Å in stage 2) and angle-range, which could indicate increased ordering in the Gdm^+-Gdm^+ parallel stacking than in stage 1. There are interactions in stage 2 at ~3.5 Å over a greater angle range compared to stage 1 (Figure 4.35d(*)), which correlates with the $g_{C\text{-}C}(r)$ first peak shoulder.

In stage 2 (Figure 4.35d), there are increased T-stacking interaction sites (intensity around 90°) at r = 6–7 Å, which is closer in stage 3 (~4.5–5.5 Å) and is reflected in the increased Gdm^+-Gdm^+ interaction sites in the plane of Gdm^+ in **Figure 4.33**c .

ARDF for Gdm^+-Gdm^+ z-y axis interactions (-NH_2 group pointing towards the central Gdm^+ Figure 4.36b) are shown in App. Figure 51. An increase in T-stacking (interactions of the edge of Gdm^+ ions with the plane of the central Gdm^+) are seen in Stage 3 (App. Figure 51c).

<u>Gdm^+-Cl^- interactions:</u> Broad C-Cl^- interaction peaks at ~3.8 Å and ~4.3 Å are present in all $g_{C\text{-}Cl}(r)$ (Figure 4.37a&b). These interactions could relate to the interstitial position of Cl^- in the plane of Gdm^+ and the linear bond of Cl^- with -NH (3.7 Å and 4.1 Å in ref. [97]). The $g_{C\text{-}Cl}(r)$ and the probable interaction sites from the SPD plot (Figure 4.29) show that Gdm^+-Cl^- ion pairing occurs throughout the MSZ.

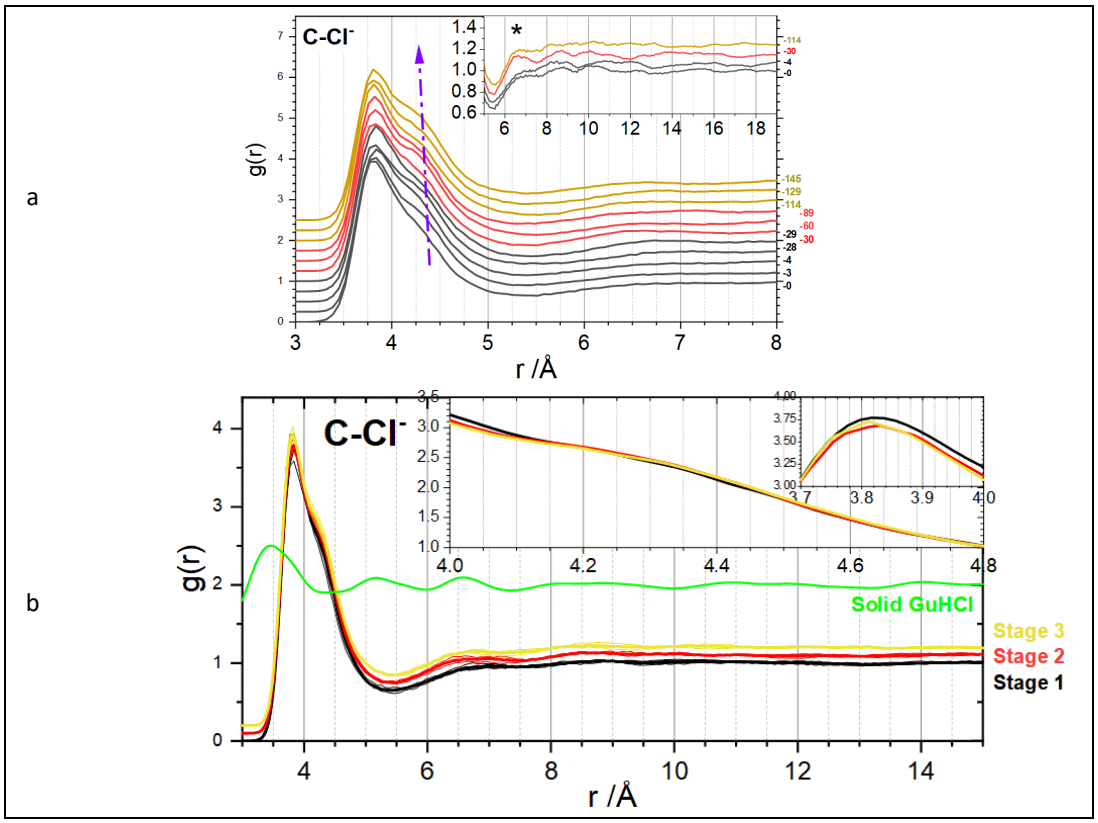


Figure 4.37 C-Cl^- partial pair distribution functions $g_{C\text{-}Cl}(r)$

a) offset $g_{C\text{-}Cl}(r)$ (purple arrow shows the reducing shoulder position) and b) stage-averaged $g_{C\text{-}Cl}(r)$, with the g(r) of solid GuHCl (light green)

The first $g_{C\text{-}Cl}$-(r) peaks are similar in height in the supersaturated solutions (stage 2 & 3), and the stage 1 and crystallisation stage. C-Cl^- CoordN in stage 1-3 (0–4.1 Å) are ~1.80, and 1.85 in the crystallisation stage (App. Table 22), suggesting that all Gdm^+ ions are paired with at least one Cl^- in solution and often more than one. The second peak becomes more prominent and shifts to a lower-r (~4.4–4.3 Å) across the MSZ (Figure 4.37a). The second shell CoordN do not vary significantly from stage 1 to stage 3 (2.59→2.66), staying at 2.66 in the crystallisation stage (App. Figure 23). The third coordination shell, which is much lower intensity, starts at 5.5 Å in all simulations, with r_{max} showing an evolving local structure through the MSZ as it decreases from 7.6 Å (stages 1 & 2), to 7.3 Å (stage 3), to 7.0 Å (crystallisation stage). This data suggests that the distance between adjacent ion pairs reduces during the crystallisation, probably due to reorganisation of the surrounding water, including a dehydration step between stages 2 and 3.

$g_{N\text{-}Cl}$-(r) (App. Figure 38) is dominated by the peaks at 3.2 Å and 5.1 Å. The narrower, first peak has an intensity almost twice that of the second, broader peak, representing the Cl^- being in the plane of Gdm^+ between two -NH_2 groups. From 8.5 Å, all $g_{N\text{-}Cl}$-(r) are close to 1, which indicates the clusters < 10 Å in size are forming in the solution.

<u>Cl^--Cl^- interactions:</u> $g_{Cl^-\text{-}Cl^-}$-(r) has clearly defined features up to 9 Å, after which perturbations are much smaller (Figure 4.38), again indicating ion clusters of < 10 Å. The shortest Cl^--Cl^- atomic separation in solid GuHCl is 4.2 Å and 4.7 Å (Appendix A.3.2.1). There is some evidence of features at these atomic separations in **Figure 4.38**.

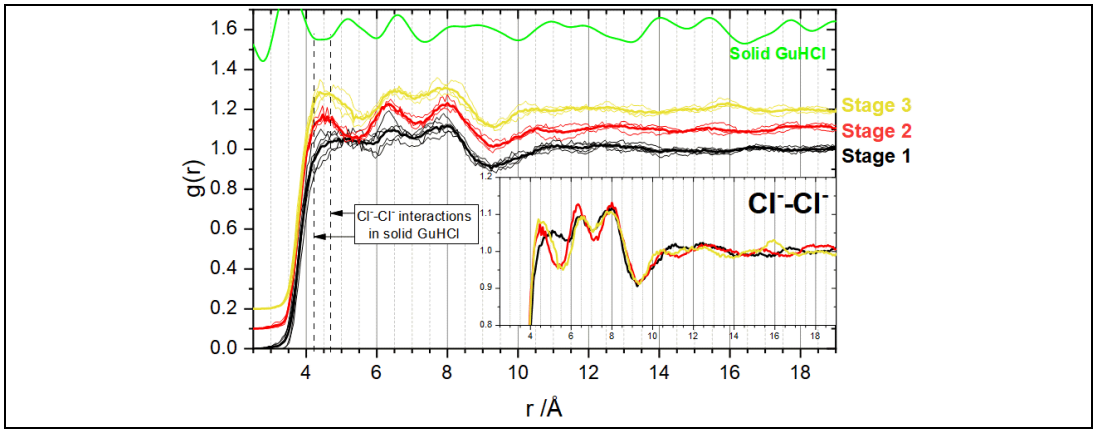


Figure 4.38 Cl^--Cl^- partial pair distribution functions $g_{Cl^-\text{-}Cl^-}$-(r)

With Cl^--Cl^- interaction distances in solid GuHCl indicated by the dash black lines and the g(r) of solid GuHCl (light green)

The first three coordination shells in stages 2 and 3 seem more defined than stage 1, with higher maxima and lower minima, and lower first peak positions at 4.5 Å (from ~5 Å in stage 1). It is particularly notable that first coordination shells peak narrows between stages 1 and

2, suggesting increased local structuring of Cl^- in the supersaturated solutions (stages 2 and 3). Possible interactions that could result in peak positions in the region of 4.5 Å, 6.5 Å and 8.0 Å were identified in a snapshot of the GuHCl_145 simulation box (Figure 4.39).

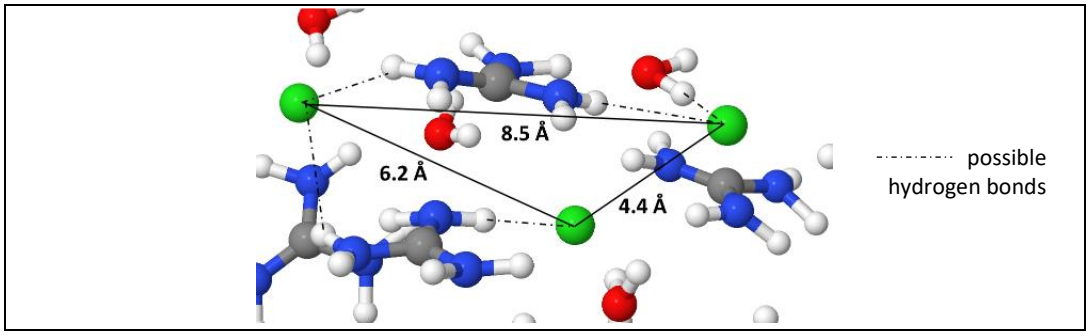


Figure 4.39 Molecular model from simulation GuHCl_145 to illustrate Cl^--Cl^- interactions

Cl^--Cl^- CoordN in the first shell fluctuates during the crystallisation experiment, decreasing from stage 1 to stage 2 (App. Table 34). CoordN over the three shells increases from stage 1 to stage 3. The number of Cl^- / A^3 (Cl^- density) in App. Table 34 accounts for the changing radius of the first coordination shell. Cl^- density decreases slightly from stage 1 to stage 2, which could be due to increased Cl^- interaction with water or Gdm^+, moving the Cl^- further apart. It then increases from stage 2 to stage 3, which could occur when desolvation to produce smaller dimer clusters occurs (structure 5, Figure 4.28), as Gdm^+-Cl^- pairing is high.

4.3.2.2.1.3 Ion solvation

Gdm^+ and Cl^- are known to be weakly hydrated, even in dilute solution [102,459]. Cl^- should have six water molecules in the hydration shell (in a low concentration solution, < 2 molal) [473,474], decreasing as the salt concentration increases [103], and Gdm^+ is reported to have an average of 9.9 water molecules in the hydration shell, of which 4.5 in the plane of the molecule [97]. The complete hydration shells are not achieved in this high concentration solution.

Gdm^+ solvation: A strongly-hydrated ion or molecule would be 'held' in the bulk solution and have less opportunity to interact with the protein thus stabilising it [459]. The strongly denaturing properties of Gdm^+ are attributed to it being a weakly hydrating cation [459], as it preferentially interacts with proteins in biological systems via the hydrophobic plane.

Ab initio calculations to analyse the hydration structure of stable Gdm^+ dimers showed that stabilisation is possible with as few as three water molecules (water molecule in the interstitial position between the amino groups, accepting two hydrogen bonds, one to each ion, Figure 4.40a, [91]). For dimers stabilised by 4–6 water molecules, the Gdm^+ ions only had three hydrogen bonds with the water molecules (Figure 4.40b) [91]. The *ab initio*

calculations exclude the counterion, which competes with water for the interstitial position and could explain how the ions are stable and in solution despite the low water:ion ratio.

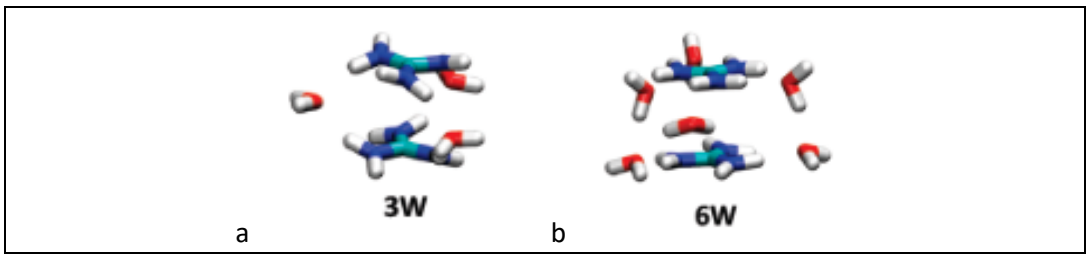


Figure 4.40 Optimised Gdm^+-Gdm^+ dimer structures from ref. [91]**, showing smaller (left: 3 water molecules) and larger (right: 6 water molecules) dimer clusters**

A water-Gdm^+ hydrogen bond has half the energy of a water-water hydrogen bond and as such are weaker [459]. Although water-Gdm^+ bonds have an electrostatic component that should strengthen it [475], the positive charge on H (+0.380) is less than that of H_t (+0.417), so O_t preferentially hydrogen-bonds with H_t, which could explain water cluster formation, if the Gdm^+ solvation is stable with as few as three water molecules [91].

The first H-O_t shell CoordN remains stable across the stages at ~2.5 waters per Gdm^+ ion (~0.41 per N-H···O_t, r_{max} = 2.5 Å: see App. Table 28), indicating that the number of water molecules interacting with the plane of the molecule stays constant. However, this data is non-directional, and it is not known whether the water molecules are coordinated with one or more -NH_2 groups on the Gdm^+ ion.

The structural changes between stages 1–3 identified by PCA could relate to changes in the complexation of the Gdm^+ dimer, going from a structure such as 6W (Figure 4.40 from ref. [91]), where 4, 5 or 6 water molecules or Cl^- stabilise the dimer, to a dimer requiring fewer stabilising surrounding molecules. The change from larger to smaller dimer clusters could explain the peak marked * in the stage 3 $g_{C\text{-}C}(r)$ (Figure 4.32), as there would be less separation between Gdm^+ in the smaller clusters. The evolution of the dimer structures, particularly from stage 2→3, but also to an extent from stage 1→2, results in the release of some water molecules that were hydrating the Gdm^+ (structures 1 and 2, Figure 4.28).

Although the average first shell CoordN stays constant for H-O_t (App. Table 28), the increase in C-O_t CoordN could indicate the formation of a diffuse water cloud of water molecules in a dome across the face of Gdm^+ as more water molecules become available after the larger to smaller water cluster desolvation step [92,97,181]. However, the increase in C-O_t CoordN observed from stage 2→3 (3.59→3.66, App**. Table 30**) is insignificant when the resolution of the EPSR simulations is considered. Further study with higher-resolution X-ray total

scattering data and larger EPSR simulation boxes would be of interest to see if a more significant difference could be observed.

Bond angle analysis: Analysis of the bond angles between water and Gdm^+ (N-H···O_t, r_{max} = 2.5 Å) was undertaken with dlputils. The analysis is based on atom number, not atom type, so data for each of the six -NH atomic combinations with atoms numbers 1 (O_t) & 2 (H_t) on the water molecule were extracted and averaged (see Appendix E.2 for details). Table 4.10 shows a summary of the average bond angle distributions.

Table 4.10 N-H···O_t bond angle distributions

Average N-H···O_t hydrogen bond	Stage 1	Stage 2	Stage 3
Mean /°	143.4	143.4	143.2
Mode /°	146.25	146.75	146.75
% 150–180°	39.1%	38.8%	38.9%
% 160–180°	19.3%	19.1%	18.8%
% 170–180°	5.3%	5.2%	5.1%

The differences in the mean and mode average bond distances between the stages are small (2.1° and 1.0°), but the data illustrate that on average, there is a low proportion of linear Gdm^+-water hydrogen bonds in each stage. The decrease in linear bond angles (160–180°) and the mean bond angles across the stages could support the idea of the evolution to smaller dimer clusters, but the change is small, and the mode value does not fit the trend. Further work should be carried to establish how robust this analysis is.

4.3.2.2.1.4 Water structure

Water clustering: Whilst the high solution concentration means that bulk water is not expected according to the ideal geometrical distribution of molecules in solution (Equation 24, ref. [465]), some water molecule clusters are observed on inspection of each molecular model. A desolvation step has been proposed between stages 2 and 3 due to changes from the large to small dimer clusters (Structures 4 to 5 in Figure 4.28), which would release water into the bulk solution.

Clusters auxiliary routine (Appendix E.1.2) has identified the proportion of water molecules in cluster size n, where O_t-O_t separation r = 2.3–2.9 Å (Figure 4.41)[36]. These data illustrate that > 65% water molecules in each simulation are in a cluster n = 1, i.e. more than 2.9 Å

[36] The analysis is based on this radius to exclude the shoulder feature at ~3.5 Å, where water molecules may be in proximity with, but not hydrogen-bonded to, another water molecule – see the following section: **O_t-O_t interactions**

from another O_t, which is not unexpected given the high concentration of the solution. More unclustered water molecules are present in stage 1 (6%, black bars) than in the stage 3 (gold bars). From n≥3, there is a higher proportion of water clusters in stage 3. The data also indicate a difference in the water clustering between stages 2 and 3 for n = 4–12, which could suggest a process to release some water (partial desolvation) between these stages. There is much less cluster size variation between stages 1 and 2.

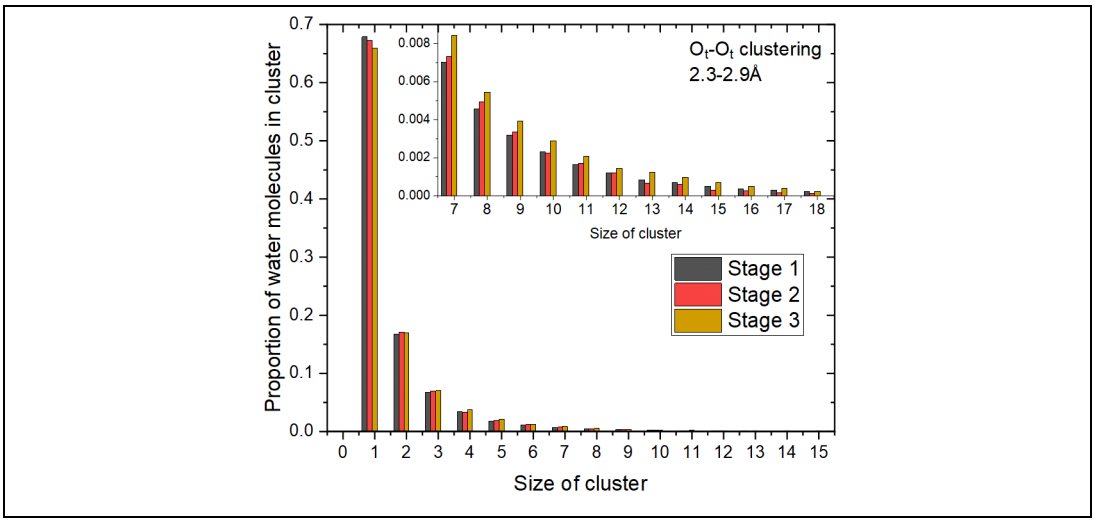


Figure 4.41 Clustering of water molecules, based on O_t-O_t interactions

The O_t-O_t, O_t-H_t and H_t-H_t g(r) and CoordN will indicate the extent of structuring of water molecules in the solution. The data indicate that there are changes in water interactions, and therefore the solvation of ions, in the stages before crystallisation.

O_t-O_t interactions: Figure 4.42a shows $g_{Ot\text{-}Ot}(r)$ for all simulations[37]. The first shell peak in $g_{Ot\text{-}Ot}(r)$ is in a similar position to that of pure water [375], suggesting that despite the low water:ion ratio, some hydrogen-bonded water molecules are present.

The intensity of the first peak in the average $g_{Ot\text{-}Ot}(r)$ is the same for stages 1 and 2 (black and red, inset Figure 4.42b), but increases in stage 3, indicating there is a release of solvating water molecules between stages 2 and 3. The first $g_{Ot\text{-}Ot}(r)$ peak narrows slightly from stage 1 through to stage 3, likely due to a reduction in thermal disorder, but as would be expected, this change is not significant [235].

Beyond the first solvation shell, the ordered, hydrogen-bonded network of pure water is disrupted, as the characteristic second and third shell peak positions (at 4.5 Å and 6.8 Å, ref. [375]) are not present. When water is heated, the second and third O_t-O_t coordination shells

[37] The subscript 't' represents the TIP3P water model, to distinguish the hydrogen in water (H_t) from that in Gdm^+ (H)

flatten due to the disorder [235], however in this case, as there is little second and third peak structure in the cooling solution, it indicates a lack of structure in the bulk water.

Instead, there is a shoulder feature at 3.5 Å, which may be due to the proximity of water molecules forming a solvation shell around an ion, e.g. two water molecules hydrating the same Cl^- ion, but are not directly hydrogen-bonded. Alternatively, the ions in solution may have an electrostrictive effect on the water structure [98]. It has been observed in aqueous solutions of NaOH [476], NaCl and KCl [105] that the position of the second peak in the O_t-O_t distribution moves to lower r-values, becoming a broadening of the first peak at higher concentrations. Here, the peak at 6 Å tends toward 5.3 Å in the crystallisation stage. This deviation from the characteristic tetrahedral structure of pure water is expected with the low water:ion ratio in the GuHCl solution. A similar feature at 3.5 Å is seen in an MD study of the structure of water in an undersaturated 3.1 M NaCl solution (water:ion ratio 9:1) [477]. However it is not present in an NDIS/EPSR study of water-glycerol [413], suggesting that the feature at 3.5 Å may be as a result of the presence of the ionic solutes.

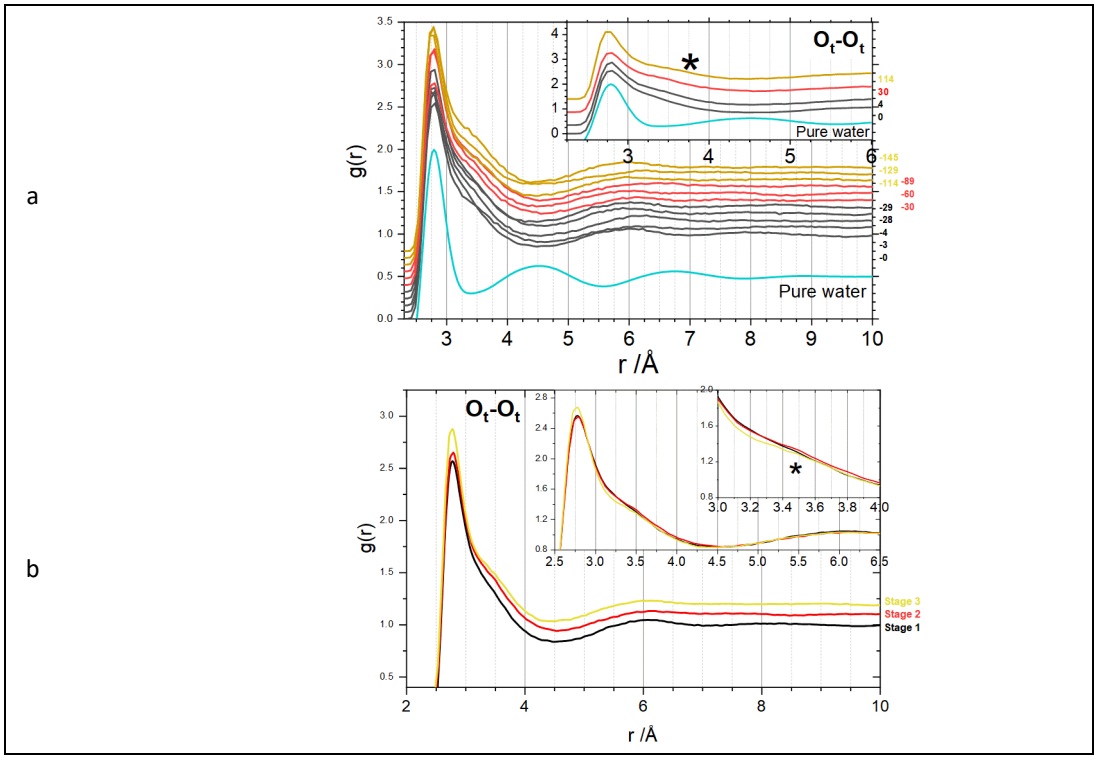


Figure 4.42 O_t-O_t partial pair distribution functions $g_{Ot\text{-}Ot}(r)$

a) offset $g_{Ot\text{-}Ot}(r)$, with pure water $g_{Ot\text{-}Ot}(r)$ in light blue, and

b) main plot: $g_{Ot\text{-}Ot}(r)$ averaged by stage (offset), inset: averaged $g_{Ot\text{-}Ot}(r)$, with * indicating shoulder position moving from ~3.5–3.7 Å

The shoulder at 3.5 Å becomes more defined and shifts to 3.7 Å between stage 1 and stage 3 (* on Figure 4.42a&b), which could be due to more water molecules becoming available

to form a second coordination shell around water after partial desolvation processes, e.g. dimer formation.

Table 4.11 O_t-O_t CoordN: averaged per stage

	O_t-O_t 1st shell (to shoulder)			O_t-O_t 1st shell shoulder		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0–3.1	1.39	*0.02*	3.1–5.5	3.33	*0.10*
Stage 2	0–3.1	1.39	*0.07*	3.1–5.5	3.41	*0.03*
Stage 3	0–3.1	1.44	*0.02*	3.1–5.5	3.35	*0.07*

O_t-O_t CoordN has been extracted for the first peak (0–3.1 Å) and shoulder (3.1–4.5 Å). An increasing average first shell O_t-O_t CoordN (Table 4.11) could suggest a desolvation step, but the increase between stages 2 and 3 (1.39→1.44) is very small, and therefore cannot be considered significant given the resolution of the EPSR simulation.

O_t-H_t interactions: There is a similar intensity increase between stages 2 and 3 in $g_{Ot\text{-}Ht}(r)$ (App. Figure 35). O_t-H_t CoordN (App. Table 17) indicate that although there is a slight decrease between stages 1 and 2, there is an overall increase in the first and second shell CoordN from Stage 1 to stage 3. Again, this could indicate the expulsion of water from between the clustering ions, but the change is too small to be conclusive.

H_t-H_t interactions: The first shell interactions of $g_{Ht\text{-}Ht}(r)$ (App. Figure 36) could include one or both H_t on an adjacent water molecule, depending on the relative orientation of the water molecules (Figure 4.43). The second peak could include an H_t from the adjacent water molecule, or H_t from a third water molecule hydrogen-bonded either directly to the second water molecule, or perhaps to the same ion as the second water molecule.

The first shell CoordN is around half the expected pure water value (~5.2 from ref. [375], where r_{max} = 2.95 Å), confirming the low numbers of water molecules in the bulk. The first shell CoordN increases between stages 1 and 3, indicating either more water molecules in the bulk, or a reorientation of the water molecules (Figure 4.43). Second shell H_t-H_t CoordN increases from Stage 1 to Stage 2, then decreases from Stage 2 to Stage 3 (App. Table 18), which could be due to a structure change from W1–W2 orientation to W2–W3 orientation, although the changes are again too small given the resolution of the simulations.

The $g_{Ht\text{-}Ht}(r)$ minimum between the first and second shell peaks is lowest in stage 3, indicating a more structured hydrogen-bond network develops, possibly after the release of water from the partial desolvation step between stages 2 and 3. The width of the second H_t-H_t peak decreases from stage 2 to stage 3, which could indicate greater ordering of the water hydrogen bonding structure.

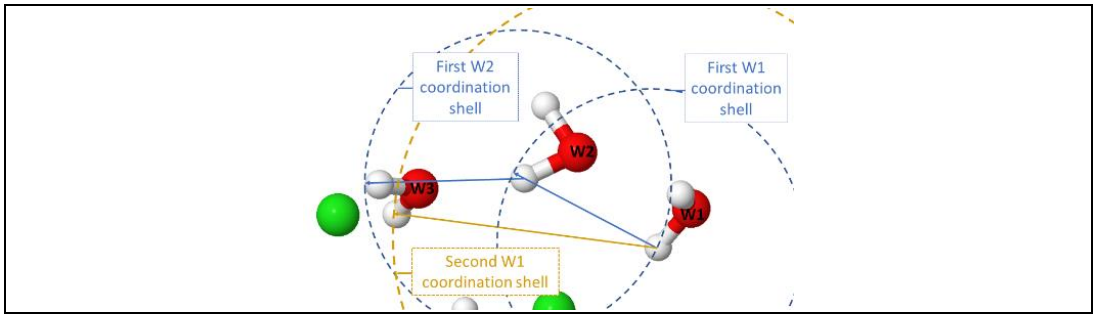


Figure 4.43 H_t-H_t coordination shells

The first coordination shell of H_t on water molecule 1 (W1) contains one H_t from W2 and the second shell contains both H_t from W2 and one from W3.

The first coordination shell from W2 contains one H_t from W1 and both H_t from W3.

Spatial probability densities: Figure 4.44 displays the 3D anisotropic spatial probability density distribution of water (centre of geometry) around a central water molecule[38]. Each SPD shows pronounced lobes opposite the H_t atoms of the central water molecule, with a band of intensity around the central O_t. Whilst the area over which interactions with the central O_t are similarly broad in each simulation, the lobes adjacent to the H_t atoms become smaller as the temperature decreases. There is a greater probability of finding interactions beyond the second hydration shell in the lower temperature solution (Figure 4.44c) than at higher temperatures (Figure 4.44a&b).

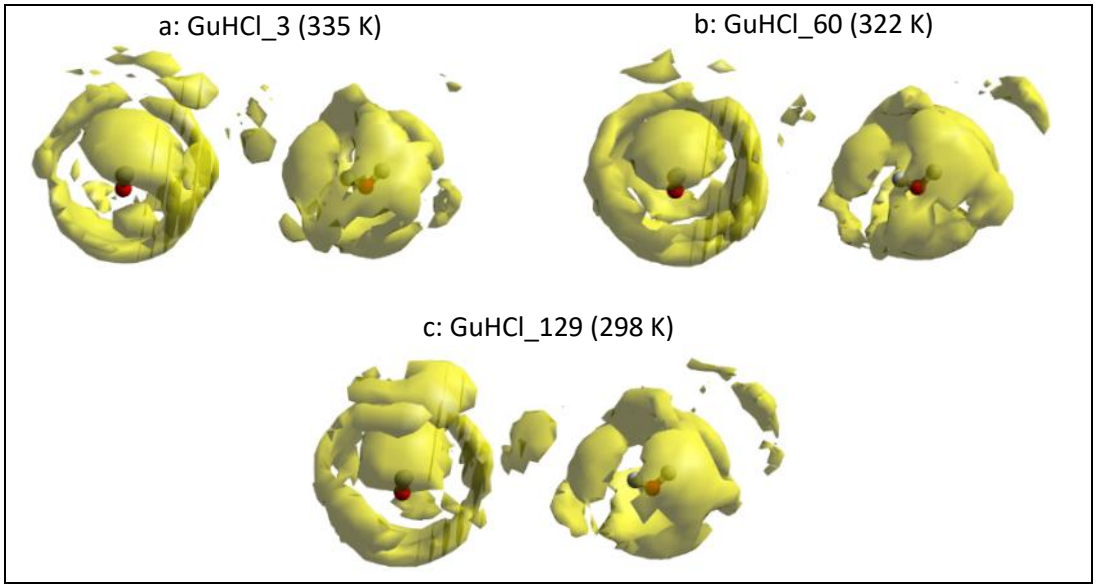


Figure 4.44 Spatial probability densities showing a-d) the distribution of water (yellow) sites (left: from side and right: face on) where a) Stage 1: GuHCl_3 at 335 K, b) Stage 2: GuHCl_60 at 322 K, c) Stage 3: GuHCl_129 at 298 K

0.0185 water molecules $Å^{-3}$ (1.55x average molecular density of the water)

[38] The SPD show the molecular density across the whole grid, not over a defined radius.

The hydration shells of water in solution are compared with the spatial density functions (SDF) of pure water at the most similar temperature from ref. [235] (Figure 4.45). The SDF, derived from Sharm calculations in EPSR, show interactions up to 5 Å, whilst the SPD show the molecular density across the whole grid, which explains why interactions are seen over a larger radius in the SPD plots. The characteristic lobes adjacent to the central H atoms are present in the lower temperature pure water SDF, with the band of intensity beneath O (**Figure 4.45**b).

The lobes adjacent to H_t appear wider for the solvating water (Figure 4.45d) than the pure water (Figure 4.45b) suggesting that the hydrogen bonds are less linear and less ordered in GuHCl solution (the temperatures are not identical). Combined with the intensity around O_t being broader in solution (Figure 4.45d) than around O in pure water (Figure 4.45b), it confirms that the solvating water in solution lacks tetrahedral local structure, even after crystallisation when water has been released. In the higher temperature structures, the first and second shells in pure water have merged (Figure 4.45a), whereas there is limited evidence of a second shell in Figure 4.45c. Figure 4.44 shows a partially-complete second shell in GuHCl_129, but this is less evident at high temperatures (GuHCl_3).

Similar water-water SPD analysis has been undertaken in the following chapter, which examines the effect of GuHCl concentration on solution structure.

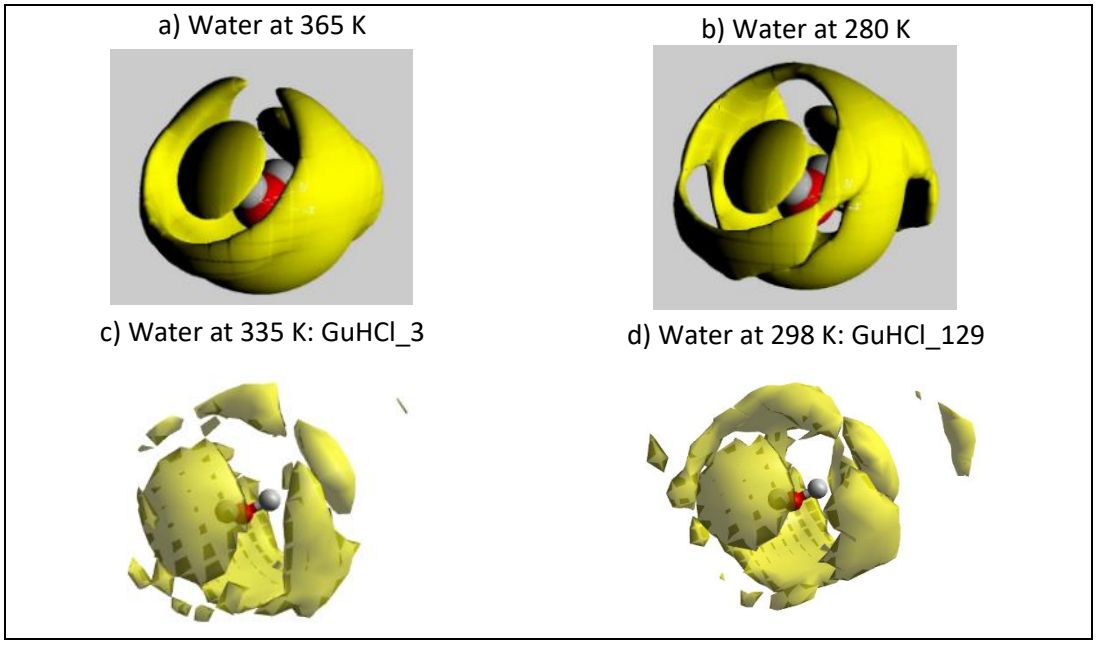


Figure 4.45 Spatial density functions (EPSR) for pure water (ref. [235]**) and spatial probability densities (dlputils from Figure 4.44) for solvating water**

Spatial density functions: top (most dense) 25% of water molecules within 5 Å from a central water molecule at a) 365 K and b) 280 K, from ref. [235].

Spatial probability densities: 0.018 water molecules Å-3.

4.4 Conclusions

XPDF pattern analysis and EPSR structural modelling have been used to reveal the changes to the structure of a 9 M aqueous GuHCl solution as it cools. The X-ray total scattering data successfully collected *in situ,* as the aqueous GuHCl solution cooled through the MSZ, to crystallisation, indicates that the emergence of long-range order correlates with the crystallisation exotherm measured during the experiment (Figure 4.9).

Principal component analysis (PCA) of the XPDF patterns from two sets of temperature-resolved data shows that there are three structural phases prior to the transition[39] and crystallisation stages: the high-temperature undersaturated solution (stage 1); a supersaturated solution appearing at 51.4°C (stage 2); and a further structural change at 35.5°C (stage 3). Crystallisation occurred at ~20°C.

To develop the EPSR model, a parameterisation study identified the force field data (Lennard-Jones potentials and partial charges) and water model type to give the best-fitting model when refined with and compared to experimental data.

Partial g(r) and CoordN of the molecular interactions were extracted from the EPSR molecular models and further structural information was acquired using dlputils analysis tools. $g_{C\text{-}Cl}$-(r) and spatial probability density plots indicate the presence of Gdm^+-Cl^- ion pairing on the plane of the Gdm^+ ion throughout the MSZ. Gdm^+-Gdm^+ CoordN data indicate an increase in Gdm^+-Gdm^+ planar stacking clusters (structure 4, Figure 4.28) in stage 2. It is proposed that a partial dehydration step occurs between stages 2 and 3, with the Gdm^+ dimers being stabilised by fewer water molecules in stage 3 (structure 5, Figure 4.28) than in the stage 2 dimers (structure 4). There is also a proposed increase in T-stacking of Gdm^+ ions (structure 6, Figure 4.28) in stage 3. Many partial g(r) indicate that local ordering up to ~10 Å is present in the solution across the MSZ, possibly due to the presence of parallel-stacked dimers, stabilised by water molecules and Cl^- ions, with possibly a third Gdm^+ T-stacked to the dimer structure.

The analysis presented in this chapter serves as proof of concept for the investigation of the evolving intermolecular interactions in a crystallising solution and demonstrates the potential power of the combined experimental XPDF/computational modelling technique.

[39] Transition phase: the description of the frame of X-ray total scattering data where long-range order appears

There are limitations to the application of EPSR modelling of the solution structure. EPSR is designed to model systems in equilibrium, but due to the nature of the *in situ* cooling crystallisation experiment, the temperature of the solution reduced during data collection by 0.5–0.7°C. For the local structure of the solution to be reliably represented in the structure factor, X-ray total scattering requires the collection of sufficient counting statistics with an acceptable signal-to-noise ratio. However, the development of high-brilliance synchrotron X-ray sources, e.g. XFEL or ESRF-EBS, will allow faster data acquisition with higher counting statistics. As a result, temperature and density variations in the solution per frame will be reduced and thus the structure factors will be more representative of the average solution structure at a given temperature.

Although a parameterisation study was undertaken to identify the appropriate Lennard-Jones potentials for the system, incorrect parameterisation could have introduced errors into the simulation. To overcome this, a machine learning approach to refine the parameters could be developed. The analysis of an aqueous GuHCl solution using NDIS instead of X-ray total scattering, and EPSR modelling, could also inform the selection of appropriate Lennard-Jones parameters for the *in situ* analysis.

The following chapter uses the combined XPDF/EPSR technique to investigate the influence of GuHCl solution concentration on the solution structure.

Chapter 5 The effect of concentration on the structure of guanidine hydrochloride in water: an XPDF and EPSR study

5.1 Context

With properties that could place it at either end of the Hofmeister scale [109], guanidinium (Gdm^+) has been the subject of many studies to understand the implications of its specific ion effects [98,478] on biological systems, e.g. the molecular mechanism of protein denaturation [176,182,457,479–481].

The hydration of Gdm^+ has been studied computationally [91,93,97,182,459,481] and experimentally [92,97,181,479]. Gdm^+ has been found to be weakly-hydrating [459], suggesting that its denaturant properties are due to Gdm^+ breaking the salt bridges that stabilise a folded protein [174]. Its hydrophobic faces interact directly with the surface of the protein thus destabilising the protein's hydrophobic core [93,482], rather than Gdm^+ disrupting the water structure to the extent that it impacts protein stability [479].

Despite the apparent electrostatic repulsion between a like-charged Gdm^+ pair [184], Gdm^+-Gdm^+ contact ion pairing in water has been predicted by many computer simulations [114,169,179,182,183]. Not all experimental data supports the existence of like-charged and oppositely-charged ion pairing, as dielectric relaxation spectroscopy (DRS) experiments found no evidence of any ion pairing (Gdm^+-Gdm^+ or Gdm^+-Cl^-) in solution [483,484]. However, neutron diffraction with isotopic substitution (NDIS), combined with MD simulations, has indicated that Gdm^+-Gdm^+ stacking and Gdm^+-Cl^- ion pairing does occur, with the chloride counterion contributing to the effectiveness of guanidine hydrochloride (GuHCl) as protein denaturant [97].

The behaviour of Gdm^+ ions has been observed to change depending on the concentration of the solution [485,486]. This chapter presents a combined X-ray pair distribution function (XPDF) and Empirical Potential Structure Refinement (EPSR) [257] analysis of the local structures of 2, 4 and 6 M aqueous solutions of GuHCl. Statistical analysis of the modelled solution structures, using EPSR and dlputils [460], has also been undertaken.

The development of a suitable force field is key to obtaining high-quality results from accurate modelling of molecular-scale interactions [487]. Various force fields have been developed in the literature for computational modelling of Gdm^+ [169,458,481] and

Cl^- [268,461,488]. The first stage of this work was a parameterisation study to refine and validate the force field selected for the EPSR simulations (Appendix F). These data were then applied in the development of molecular structure models that were validated with experimental data. Visualising the local structure in solution, including the extent of cation-anion association, will inform our understanding the evolution of intermolecular interactions during salt crystallisation, including the presence of pre-nucleation ionic aggregation.

5.2 Methodology

5.2.1 Experimental method

The X-ray total scattering data of GuHCl solution and water samples were collected at DLS I15-1. The high-throughput sample holder (Figure 2.18b and described further in Chapter 2) was used to mount 2, 4 and 6 M GuHCl aqueous solution and deionised water samples in 2 mm OD borosilicate capillaries. The capillary spinner (Figure 2.18a, described further in Chapter 2) was used for a deionised water sample in a 1.5 mm OD borosilicate capillary. Sample preparation, which allowed the calculation of density for the EPSR simulations, is described in Section 2.2.3.2.

X-ray total scattering data of deionised water in 1.5 mm and 2.0 mm capillaries were collected and analysed to ensure equivalence of the data from each type of sample geometry, i.e. capillaries in the 1.5 mm OD static and 2.0 mm OD spinning holders. This agreement was important to validate due to the cylindrical nature of the capillary. Spinning the capillary averages out the curvature of the region through which the X-rays pass. In a static holder, this averaging does not happen, and hence it is vital to validate whether the beam position and curvature of the capillary affects the quality of the data acquired. A study of data collection reproducibility and the effect of data collection time on the XPDF patterns (using 6 M aqueous GuHCl solutions) to validate the data acquisition strategy was carried out and is described in Section 2.5.4.1.4.

5.2.1.1 XPDF data collection and processing

X-ray total scattering data were collected at beamline I15-1 at Diamond Light Source at 76.7 keV (0.163 Å), which, with the detector positioning, gave access to $Q_{max} \approx 40$ Å^{-1}. However, in the case of organic materials, a lower Q_{max} is used to process the data [201,367]. The large area Perkin Elmer detector (PE XRD 4343 CT detector) has an active area of 432 × 432 mm^2,

with 150 µm pixel size. The instrument was calibrated with silicon. Further details of the operation of the beamline are in Section 2.5.4.1.

The X-ray total scattering data were processed to extract the reduced structure factor F(Q) and generate pair distribution functions (PDF) g(r), also referred to here as the XPDF patterns. The generic data acquisition software at DLS automatically applied the initial corrections for the internal dark current to the raw data during data collection. GudrunX was used to correct the processed data for the instrument and sample holder backgrounds, absorption, multiple scattering and Compton scattering [365,368]. GudrunX produces interference differential cross sections F(Q), from which the real-space PDF patterns are generated by Fourier transform [365,368]. There is an option in GudrunX to correct for fluorescence. Although the chloride ions are a possible source of fluorescence, trial and error showed that applying a fluorescence factor had negligible, if any, visible effect on the F(Q) or g(r) and so none were included. The appearance of termination ripples at low-r was reduced by applying a broadening function of width 0.125 Å on the r-space results.

For background subtraction, X-ray total scattering data of a blank capillary should be collected for at least the same amount of time as the sample data. Data collection times are shown in Table 5.1.

Table 5.1 Background and sample scan times

Sample	Background scan (empty capillary)	Sample scan
2.0 mm water	3 frame x 600 s (1800s)	20 frame x 30 s (600 s)
2.0 mm 2, 4 and 6 M GuHCl solutions	3 frame x 600 s (1800s)	20 frame x 30 s (600 s)

To study the interactions between the solute-solute and solute-water molecules with increasing GuHCl concentration, a background of deionised water in a capillary was subtracted from the X-ray total scattering data to remove the water self-interactions. An empty capillary background was subtracted from the X-ray total scattering data for the EPSR simulations, where all atomic pair correlations are considered.

Analysis of the 1.5 and 2.0 mm capillary water sample data and the additional 6 M GuHCl data for the reproducibility and data acquisition studies are in the Research Methods (Chapter 2). The analysis shows that the use of the high-throughput sample holder provided robust X-ray total scattering data to be used in this concentration study.

5.2.2 Computational method: EPSR analysis

Structural models of the solution were generated using EPSR version 25 [257,406]. When the empirical potential (EP) was applied, the models were refined to increase the goodness of fit with the F(Q) data from the GudrunX processing of the X-ray total scattering data. The calculation of radial distributions of atomic pairs and spatial density functions of one atom type around a central molecule, and statistical analysis of the molecular distributions were undertaken using routines within EPSR. Further details of EPSR are in Section 2.9.2.2.1.

Component structure: The Gdm^+ ion was geometry optimised in ORCA (RKS BP86 RI SVP def2/J) to provide the starting structure [380]. The Gdm^+ ion geometries in the model vary during the simulation, the extent of which depends on the level of molecule flexibility set in EPSR.

Component Lennard-Jones potentials and partial charges: The atomic partial charges and Lennard-Jones potentials used in the EPSR simulations of 2, 4 and 6 M GuHCl solutions are shown in Table 4.2. The parameters selected for this study were refined in the parameterisation study (Appendix F), as referenced in Chapter 4.

Solution structure model generation: For the EPSR molecular models, boxes of molecules were constructed with the compositions, dimensions and atomic number densities as shown in Table 5.2. All systems were modelled at 290 K.

Table 5.2 EPSR simulation box

Sample conc. /M	Side length of cubic box /Å	Simulation box volume /Å^3	# Gdm^+ and Cl^- ions /each	# water molecules	Ratio water:ion	Atomic number density /atoms.Å^{-3}
2	37.0642	50917.26	62	1438	11.6:1	0.098120
4	37.0620	50907.98	123	1203	4.9:1	0.097470
6	37.0576	50889.73	184	965	2.6:1	0.096660

The simulations started with an expanded simulation box to allow the molecules to move past each other more freely in the initial stages. Once the system energy was below 0 kJ/mol, the size of the box gradually reduced to the correct volume. After running with the RP until the system energy was at equilibrium (~1,500 iterations), the EP was applied to refine the model to experimental F(Q) data. The EP for the 2 and 4 M simulations was 5 kJ/mol, and 4 kJ/mol for the 6 M simulation. The EP simulations ran for ~1,000 iterations before the simulations were set to accumulate. The accumulation and analysis routines were run for a further ~1,000 iterations.

To refine the EPSR simulations, the effects of adjusting the molecule flexibility, the charge on the central carbon atom of Gdm^+ and the distance range applied to the potential truncation were explored and are detailed in Appendix F.

In summary, the study of molecule flexibility suggested that retaining dihedral angle constraints on the Gdm^+ molecule, but allowing some flexibility by setting the *vibtemp*, *angtemp*, and *dihtemp* parameters to 45/0.7/0.7, optimised the F(Q) and g(r) fits.

The experimental F(Q) data sets were further refined with GudrunX to improve the data processing after the parameterisation study. Therefore, the parameterisation study simulations cannot be directly compared to the simulation results in Section 5.3.2, but the study guided the setting up of the final simulations.

Statistical analysis: Statistical analysis to compare the simulations were carried out with EPSR (Clusters) and dlputils (angular radial distribution functions (ARDF), and geometry and angle analysis) [460]. Data for statistical analysis were collected after EP equilibrium had been achieved. Details of the auxiliary routine parameters are in Appendix E.

Spatial probability density (SPD) plotting, visualised with Aten [408], provides further insight into the local coordination of molecules around a central component, as the 3D histogram binning indicates the spatial arrangement of the interactions.

Repeat runs: To evaluate the molecular models produced by EPSR, some repeat simulations were carried out. For each concentration, the simulation was rerun from the beginning, i.e. the box of molecules was randomised from the previous setup at that concentration (so all other parameters were identical). The 4 M simulation was also repeated for the accumulation stage (after the EP run was complete), and the outputs of these simulations are compared.

Ideal distribution of ions: From the EPSR simulation box dimensions, which are based on experimental density measurements, the ideal distribution of components in solution (distances between the component centres if evenly spread in the simulation box) have been evaluated (Table 5.3).

Table 5.3 Average component separation

Sample conc. /M	Volume of simulation box		Average distance between ions /Å	Average distance between water molecules /Å
	per ion /Å^3	per water molecule /Å^3		
2	821.2	35.4	9.4	3.3
4	413.9	42.3	7.5	3.5
6	276.6	52.7	6.5	3.8

5.3 Results and discussion

This section presents the experimental XPDF patterns, and the EPSR simulations that have been undertaken to examine probable structures of the solutions, from the solution concentration study.

Summary of terminology used: F(Q), the reduced structure factor, is the corrected and normalised structure factor S(Q). The reduced atomic pair distribution function G(r) is the Fourier transform of F(Q), and g(r) is the atomic pair distribution function (see Equation 19). The partial pair distribution function $g_{\alpha\beta}(r)$ describes the probability of finding one atom type β, around another α, per unit volume, in the material [257]. The terminology is explained further in Section 2.5.4.

5.3.1 XPDF patterns

Figure 5.1 shows the G(r) patterns for 2, 4 and 6 M GuHCl in water. The shaded regions 1–7 highlight some features in the G(r) patterns.

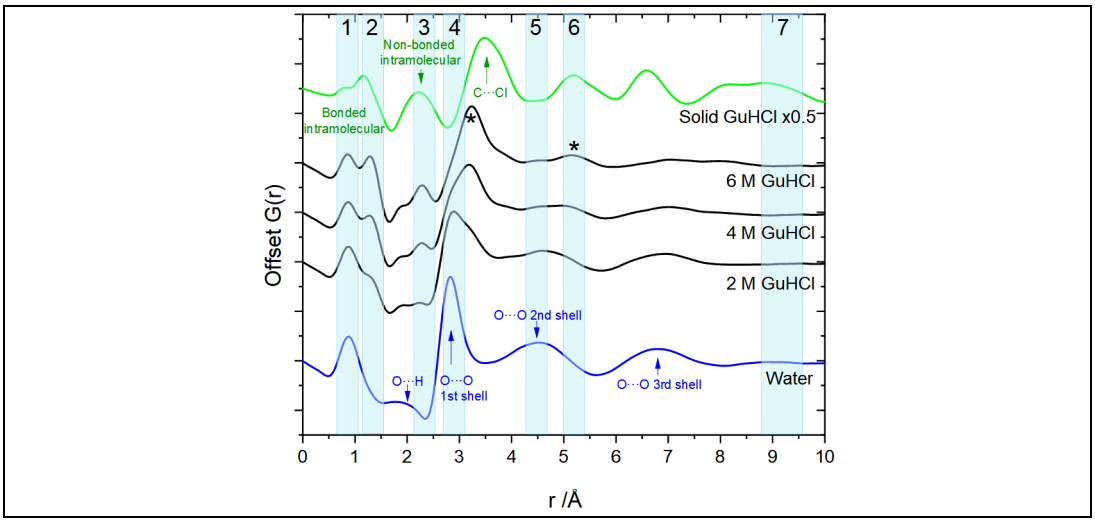


Figure 5.1 G(r) for 2, 4 and 6 M GuHCl X-ray total scattering data collections.

For comparison, G(r) of pure water and solid GuHCl are also shown.

Region 1 highlights a peak at ~1 Å, which represents the O-H bond in water and the N-H bonds in Gdm^+. As the concentration increases, the intensity of the peak decreases slightly. In the 2 M solution, there are slightly more ~1 Å interactions per $Å^3$ than in the 6 M (Table 5.4) and although this peak could be affected by termination ripples, the reduction in intensity corresponds with the expected number of intramolecular interactions. The increasing intensity due to the intramolecular C-N interactions at 1.3 Å (region 2) and N-N interactions at 2.3 Å (region 3) are seen as the concentration of the solution increases.

Region 4 shows the O-O interactions in the first shell at 2.8 Å in the water PDF. As the concentration of the GuHCl increases, this feature reduces as the water:ion ratio decreases (Table 5.2), which reduces the water available to form solvation shells and bulk water structure. The peak shifts to 3.2 Å with increasing concentration. This separation is consistent with a Cl^--N atomic separation of ~3.3 Å from the analysis of the solid GuHCl structure (Appendix A).

Table 5.4 Interactions at ~1 Å per Å^3 solution

Sample conc. /M	# Gdm^+ and Cl- ions /each	# water molecules	# N-H bonds	# O-H bonds	Total ~1 Å interactions	Simulation box volume /Å^3	~1 Å bond /Å^3
2	62	1,438	372	2,876	3,248	50,917.26	0.064
4	123	1,203	738	2,406	3,144	50,907.98	0.062
6	184	965	1,104	1,930	3,034	50,889.73	0.060

In region 5, the second O-O interaction at 4.5 Å decreases in intensity from 2 M to 6 M, whilst a feature in region 6 at 5.2 Å emerges. This feature could relate to the local structure of a hydrated ion or indicate ion-ion association. The third O-O interaction at 6.8 Å is barely present in the 6 M solution. This suggests that whilst there may be a small amount of water clustering in the solution, there would not be a 'bulk' water phase in the characteristic tetrahedral structure in the solution. The G(r) in region 7 are quite flat in water and in the solutions, although there is some intensity in the solid GuHCl G(r), which indicates that there is no significant ordering in the solution beyond 9 Å. The short-to-medium-range interactions can be explored more in the EPSR analysis.

5.3.2 EPSR analysis

Due to the random nature of a Monte Carlo simulation, and the fact that an ensemble of molecular structures could be consistent with a single X-ray total scattering pattern [489], it is not unreasonable that there may be more than one set of correlations that fit the data. This fact simply reflects the complexity of modelling a many-body system, where forces acting in the condensed phase are higher-order than simply pairwise-additive [489].

To analyse this possibility, the EPSR simulations for the 2, 4 and 6 M GuHCl solutions were rerun. Each molecular model was randomised and restarted, with the box expanded to allow the molecules to shift, in case they had been stuck in a configuration that was a local energy minimum, and following the same procedure as used the original models (Section 5.2.2). The summary of this analysis is that simulated F(Q) and g(r) for models refined to the same experimental data are almost identical, except from some low-Q deviation for the 4 M solution (App. Figure 56). For most partial g(r), there was reasonable agreement in peak

positions up to 5 Å for the original and rerun models, although the 2 M $g_{C\text{-}C}(r)$ differs at ~4 Å, which is a key atomic separation when considering the potential for Gdm^+-Gdm^+ stacking.

On the basis that both structural solutions are possible, the F(Q), g(r), partial g(r) and coordination numbers (CoordN) data from the original and rerun simulations have been averaged in the GuHCl concentration analysis in the following sections. The ARDF, SPD and angles data analysis using dlputils has been undertaken using the output from the original set of simulations.

The noise in the ion-ion g(r) increases the lower the concentration due to decreasing interactions occurring in the lower concentration solutions. Ideally, had better computational power been available[40], the 2 M solution in particular would have benefited from increased data accumulation. This effect is not significant in the ion-water g(r).

5.3.2.1 F(Q) and g(r) of simulations

The simulated reduced structure factors F(Q), for the 2, 4 and 6 M GuHCl solutions compare reasonably well with the experimental data (Figure 5.2a). The simulated F(Q) peaks at between 4 and 5 $Å^{-1}$ sit at a slightly higher Q-value than in the experimental data (Figure 5.2a *).

The structure factor of water has characteristic peaks at ~2.1 $Å^{-1}$ and 2.9 $Å^{-1}$ [235] and they are seen in the 2 M data (Figure 5.2a, indicated with *). Whilst the first peak remains unchanged from 2 to 6 M, the second peak shifts to a lower Q, and slightly increases in intensity, as the concentration increases. The geometry of the multiatomic planar Gdm^+ ions may influence how the principal diffraction peaks change with concentration, as it is different to that seen in the literature for aqueous solutions of LiCl salts, where the first peak strengthened in intensity whilst the second reduced [490]. The effect of the cation is seen in the F(Q) of the aqueous alkaline earth chloride salt solutions, where the first peaks in the larger strontium, barium and calcium cation solutions merge with the second peaks [491].

[40] The remote desktop connection to the PC on campus dropped out near the start of the lockdown in March 2020 and could not be reliably maintained. It was not possible to run the EPSR simulations on the high performance computers (HPC) at the University of Leeds, as EPSR required use of the interactive node, which was not available for long-term use.

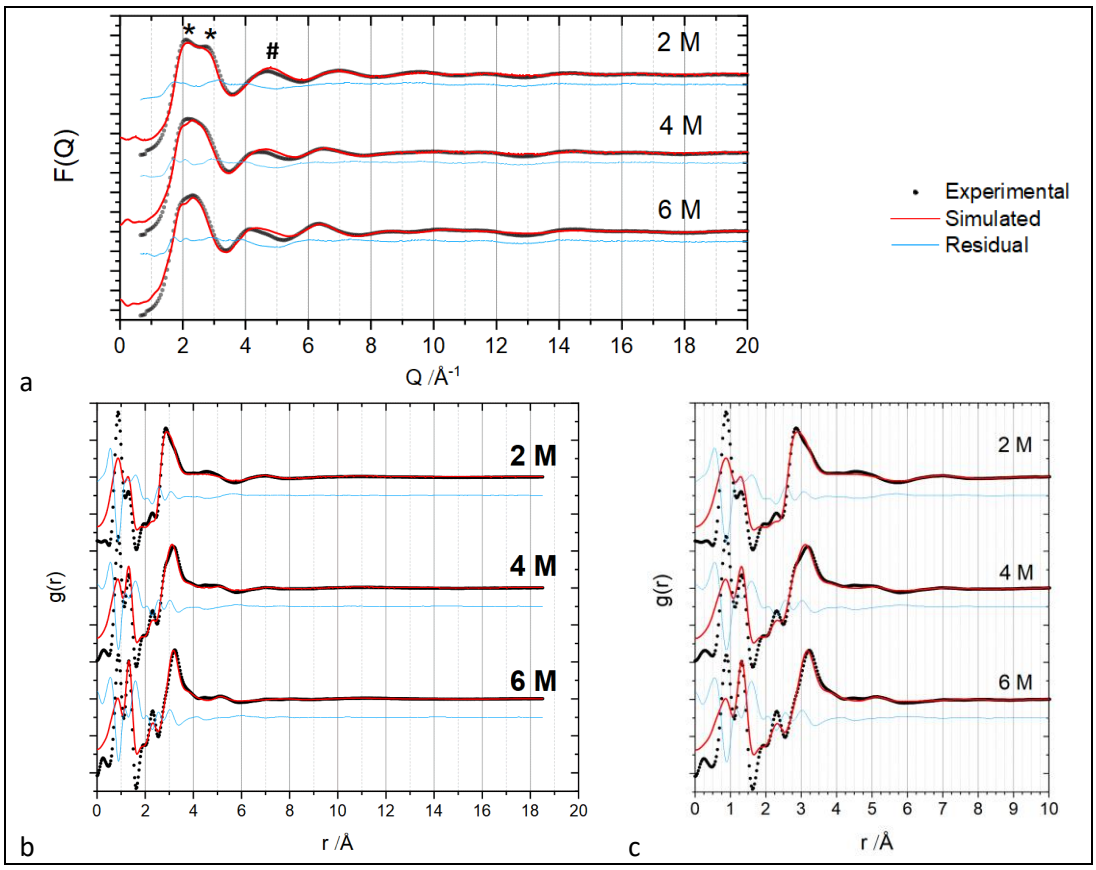


Figure 5.2 Experimental and simulated data: a) F(Q) and b&c) g(r) for 2, 4 and 6 M GuHCl

Residual (blue) – difference between experimental (black dots) and simulated data (red)

There are more obvious differences in the simulated and experimental g(r) plots (Figure 5.2b&c). The short bond (N-H, O-H) intensities are much greater in the experimental data g(r) than in the simulated g(r). However, it should also be noted that the peak heights of the 0.9 Å peak in the experimental g(r) calculated in EPSR is higher than the same peak in the GudrunX-processed g(r). The 0.9 Å peak is equivalent in intensity to the peak at 2.9 Å in the GudrunX-processed 2 M g(r), whereas it is ~0.3 Å greater in the EPSR g(r). For the 6 M solution, the peak at 0.9 Å is a similar height to the peak at ~3.3 Å in the EPSR g(r), but 0.28 lower in the GudrunX-processed g(r). The reason for this is unclear. Termination ripples are introduced during the Fourier transform process, which are particularly noticeable at low-r, but they would not cause such a difference in intensity. It does not appear to be related to the size of the box, i.e. the Q_{max} of the simulation, as the ratio of the 0.9 Å to 3.3 Å peak was similar when then simulation box volume was doubled in a test a simulation[41]. The reason for this intensity discrepancy should be explored further.

[41] Simulation of frame 30 of the cooling crystallisation run 1 (9 M aq. GuHCl solution, *in situ* analysis). The output of the larger simulation box is not included in this thesis.

For r > 2.5 Å, the modelled and experimental peak positions and intensities agree well, except for missing the small feature at ~4.5 Å (App. Figure 56). The reason for this could be a fundamental error in the applied reference potentials (RP), a mistake in the experimental data post-processing (although random processing errors having the same effect on all three datasets seems unlikely) or an alignment issue with the 2 mm capillary setup.

Changes made to RP values and Gdm^+ flexibility during the parameterisation study did not result in improving the fits at Q = 4–5 $Å^{-1}$ and r = 4.5 Å (Appendix G).The development of a machine learning method to refine the parameters could be useful if the XPDF/EPSR modelling approach were to be employed on a larger scale.

5.3.2.2 Ion pairing

The models have been interrogated in several ways to understand the molecular interactions between Gdm^+ ions. Partial g(r) of various atomic pair interactions have been extracted and are presented alongside CoordN, ARDF of Gdm^+ around Gdm^+, and SPD plots showing the most probable interaction sites of one component around another.

Gdm^+-Gdm^+ pairing

C-C- interactions: Although the $g_{C\text{-}C}(r)$ oscillations around 1 are small (Figure 5.3), the intensity fluctuations indicate that some short- and medium-range ordering of Gdm^+ ions is present in the solutions[42], including like-charged cation pairs from ~3.25 Å, as it is not possible for a water molecule or Cl^- to fit between two Gdm^+ at this separation. The 4 and 6 M $g_{C\text{-}C}(r)$ peak positions and intensities are quite similar up to ~9 Å, with small oscillations around 1 after 14 Å.

The fluctuations around 1 are small compared with $g_{C\text{-}Ot}(r)$ (Figure 5.13a), where the first peak intensities are > 1.5. However, compared with ideal component separations in solution (Table 5.3), the interactions in $g_{C\text{-}C}(r)$ and other partial PDF indicate molecular interaction distances that are significantly closer than the ideal distribution[43]. The 6 M ideal C-C separation (6.5 Å) is greater than the first peak at ~4.5 Å, demonstrating that like-charge parallel-stacking of Gdm^+ ions could be a real phenomenon and not just due to proximity of

[42] Much higher oscillations are achieved in in the g(r) that include Cl^-, due to the greater X-ray scattering power of Cl^-, whereas organic molecules are known to be much less scattering.

[43] 9.4 Å, 7.5 Å and 6.5 Å in 2, 4 and 6 M solutions

ions in solution because of concentration. These data make the presence of Gdm^+-Gdm^+ parallel-stacking in 2 M solution even more interesting.

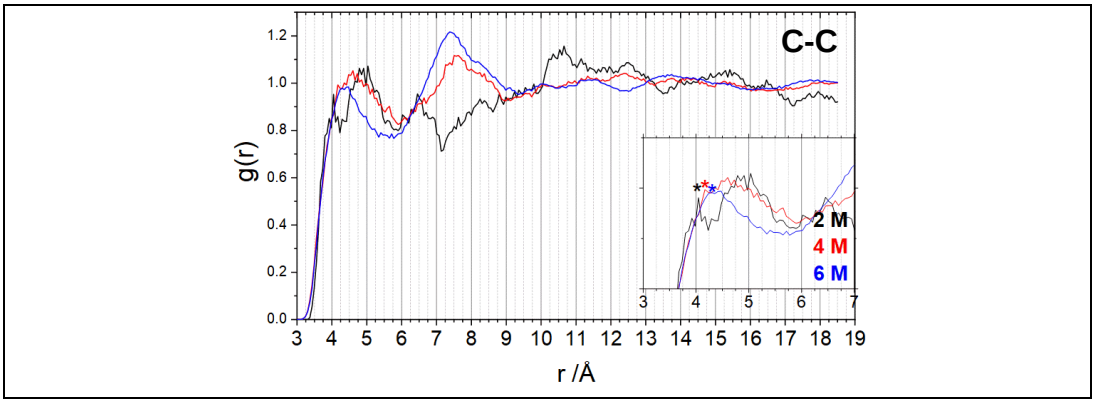


Figure 5.3 C-C partial pair distribution functions $g_{C\text{-}C}(r)$

The black, red and blue * in the inset indicate the positions of the first interaction peaks in the 2, 4 and 6 M solutions

There are two distinct peaks in the 2 M solution (4.0 Å and ~4.8 Å). The 4 M $g_{C\text{-}C}(r)$ has a broader, asymmetrical peak, centred at ~4.6 Å with a shoulder at ~4.2 Å, and the 6 M solution has an asymmetrical peak centred at ~4.3 Å. The first C-C interactions in solid GuHCl are at ~4.65 Å, so C-C interactions in solution occur at shorter distances than in the GuHCl salt, with subsequent peaks at 6.5 Å (2 M) and ~7.4–7.7 Å (4 and 6 M).

Gdm^+ and Cl^- are both weakly-hydrating ions [102,459], with up to six water molecules in the first hydration shell of Cl^- at low salt concentration (< 2 molal [473,474], decreasing as the salt concentration increases [103]), and an average of 4.5 water molecules hydrogen-bonding around the plane of Gdm^+, with a possible further ~5 water molecules forming a diffuse dome structure across the face of the ion [97].

Between the 2 M and 4 M solutions, the water:ion ratio drops from 11.6 (2 M) to 4.9 (4 M) (Table 5.2). There are insufficient water molecules to form hydration shells around Cl^- and the plane of Gdm^+ in the 4 M solution, so the ions may have to share solvent molecules. This observation is supported by a cryo-ion mobility–mass spectrometry study that investigated the size and shape of the solvation shell around Gdm^+ as the number of water molecules in the cluster n, reduces [187]. The stability of Gdm^+-Gdm^+ ions pairs, and specifically the water bridging between parallel-stacked Gdm^+ dimers, was also explored [187]. It was proposed that the discontinuity in the arrival-time distribution vs mass/charge analysis, i.e. a significant change in the hydration shell structure, occurred when n = 6–9 [187]. The orientation of the hydration shell changed from a dome-like hemispherical structure (n > 9) to a planar structure, with interstitial hydrogen bonding in the Gdm^+ plane occurring only

when the cluster is sufficiently dehydrated [187]. This finding is in agreement with infrared photodissociation (IRPD) spectra [92,181], which indicated that when the water:Gdm^+ ratio is < 5, three water molecules form the inner hydration shell, binding to the interstitial sites and accepting hydrogen bonds from adjacent NH_2 groups. Additional water molecules form a second solvation shell until n = 9, after which the water molecules in the first hydration shell rearrange to accept one hydrogen bond from an -NH_2 group, i.e. there are six water molecules in the first solvation shell [181]. The difference between the 2, and 4 and 6 M $g_{C\text{-}C}(r)$ for r = 6.5–9.0 Å is roughly consistent with the findings that there is a change in hydration shell structure between n = 6 and 9, supporting the idea of a critical solvation change between 2 and 4 M, with water:ion ratios of 11.6 and 4.9.

Although reproducible signals in the cryo-ion mobility−mass spectrometry study for parallel-stacked Gdm^+-Gdm^+ dimers occurred when n ≥ 15 [187], which is reasonably consistent with a MD study where the most stable parallel-stacked Gdm^+ dimers occurred when n = 12 [91], a locally-stable (gas-phase) Gdm^+ dimer was achieved with only three water molecules to overcome the electrostatic repulsion [91]. It therefore seems feasible that peaks at ~4 Å could be due to parallel-stacked Gdm^+-Gdm^+ contact ion pairs (CIP), and CIP closer than 4 Å are present in all three concentrations. Water-stabilised parallel-stacked Gdm^+ dimers are proposed as a motif in the solution structure in the 9 M GuHCl solution crystallisation study (Chapter 4).

A Gdm^+-Gdm^+ pair separated by 3.7 Å is described in the literature [179], in a model comprising two Gdm^+ ions in 62 water molecules (water:Gdm^+ = 31), and a previously published MD model of a 1.8 M solution showed a $g_{C\text{-}C}(r)$ first peak position of 3.9 Å, with a slight shoulder at 4.8 Å [188]. These data, combined with the data from this study, indicate that first C-C interaction peak positions (indicated by * in Figure 5.3) roughly increase with concentration (Table 5.5).

Table 5.5 Gdm^+-Gdm^+ separation: 1st principal $g_{C\text{-}C}(r)$ peak

Sample conc. /M	Ratio water:ion	1st peak position /Å
< 1.8 (Not specified - [179], no Cl^-)	31	3.7
1.8 [188]	16.6	3.9
2	11.6	4.0
4	4.9	4.2
6	2.6	4.3

However, this trend is inconsistent with two further MD studies. The first shows identical $g_{C\text{-}C}(r)$ peak positions for 2, 4 and 6 M GuHCl solutions at 3.85 Å (intensities of ~1.3–1.9) and 7.7 Å [182]. The second study indicated the first peak at ~5.5 Å for 2 and 6 M solutions (with

intensities of ~1.0 (2 M) and 1.2 (6 M)), with a second peak at ~7.5 Å [481]. Although it was concluded that there was little evidence of Gdm^+-Gdm^+ pairing, with the first peak being relatively high-r compared to other studies, some interactions are indicated from ~3.5 Å [481].

Peaks at ~5 Å are possibly due to T-shaped CIP (see Figure 4.31a&b), as the atomic spacing is too short for water or Cl^- to be between the Gdm^+ ions. The T-stacking intensities (~4.8–5.0 Å) that are clear in the 2 and 4 M solution $g_{C\text{-}C}(r)$ are less pronounced in the 6 M simulation distribution, but T-stacking could explain the asymmetrical nature of the 6 M $g_{C\text{-}C}(r)$ peak. Although T-stacking CIP interactions are not specifically identified in the aforementioned MD analyses [182,481], a shoulder feature of the first peak at around 5 Å is clearly seen in Mandal's study [182].

The CoordN up to ~ 6.0 Å have been extracted from these EPSR simulations, with CoordN < 4.3 Å being indicative of the number of parallel-stacked CIP, and 4.3–~5.9 Å indicating the number of T-stacked CIP (Table 5.6). There will be some overlap between the two 'shells', in particular as the trough between the two features is less clear with increasing concentration, but it could be used as proxy to estimate of the extent of Gdm^+ ion parallel-stacking (first feature: 'peak') and T-stacking (second feature: 'shoulder'). The T-:parallel-stacking CoordN ratios decrease with concentration, and whilst it seems rational that there are more parallel-stacked CIP in the 6 M solution, the similarity of the 2 and 4 M ratios is surprising given the proposed change in hydration structure between the concentrations.

Table 5.6 C-C CoordN – 1st shell shoulder and peak

Sample conc. /M	C-C 1st peak (parallel-stacked)		C-C 1st peak shoulder (T-stacking)		T-:parallel-stacking CoordN ratio
	Range /Å	Ave. CoordN	Range /Å	Ave. CoordN	
2	< 4.3	0.14	4.3–5.9	0.60	4.31
4	< 4.3	0.29	4.3–5.9	1.22	4.21
6	< 4.3	0.43	4.3–5.8	1.47	3.44

Not all Gdm^+ in the given ranges will be in the described CIP formations. To fully understand if there is a trend, more data points are required for other concentration solutions, and additional simulations could be run, potentially for a larger number of atoms and/or for a greater number of iterations, to acquire more robust data.

The SPD (Figure 5.4) show the interactions of Gdm^+ ions around a central Gdm^+, set to the 4% most likely interaction sites. The distributions of the interaction densities in the SPD confirm a high probability of parallel-stacked Gdm^+ dimers in all concentrations. As the

concentration increases, parallel-stacking and ordered interaction between Gdm^+ ions around the plane of the central Gdm^+ ion—indicative of T-stacked CIP—increases. Figure 5.4c in particular supports the suggestion that $g_{C\text{-}C}(r)$ intensities at ~4 Å relate to parallel-stacked CIP, with interactions ~5 Å due to T-stacked interactions (ratio of lengths of vertical:horizontal arrows = 0.625).

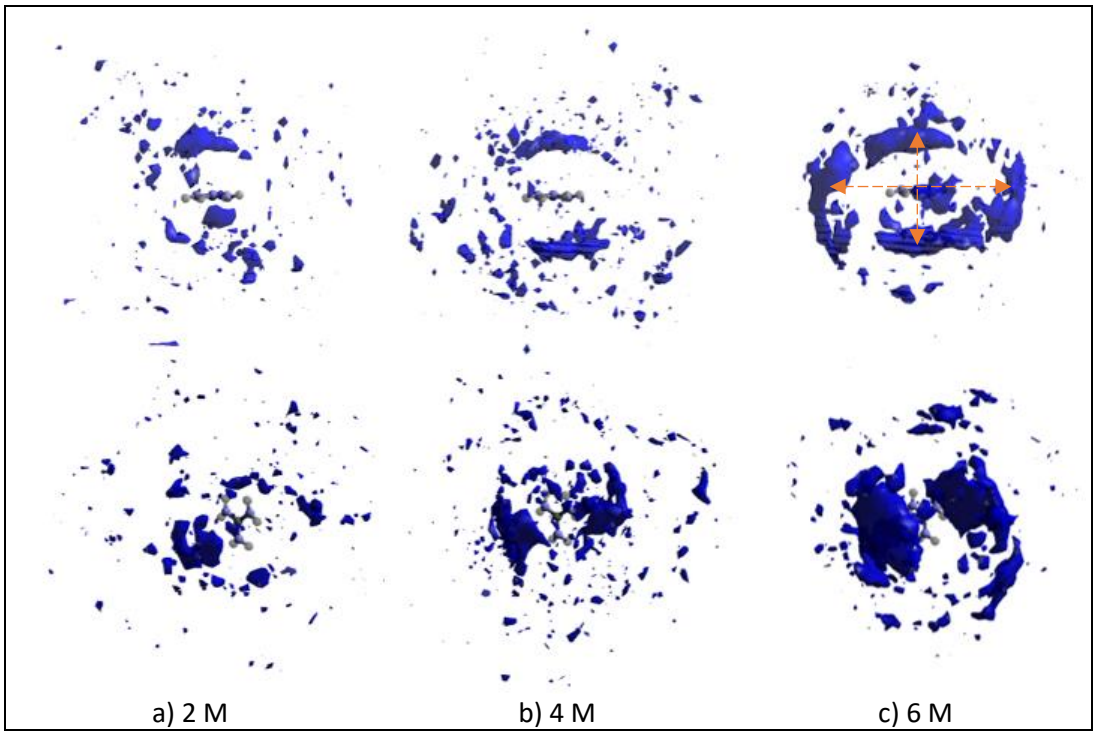


Figure 5.4 Spatial probability densities of Gdm^+ around Gdm^+ for a) 2, b) 4 and c) 6 M GuHCl solutions: top 4% most likely interaction sites

c) top: Ratio of lengths of vertical:horizontal lines = 0.625 to give an indication of the relative separation of parallel-stacked and T-stacked Gdm^+ ions

Interactions at ~7.5 Å in $g_{C\text{-}C}(r)$ could be related to solvent-shared ion pairs (SShIP), where a water molecule is positioned between the Gdm^+ ion pair (Figure 4.31c&d) [179,188]. In the literature, the C-C separation in SShIP have been identified from ~6.3 Å (potential of mean force (PMF) study [472]) to ~7.7 Å (MD study [182]). These separation distances may differ for a few reasons. The PMF model studied Gdm^+ as an arginine sidechain, not ions in solution, so the separation may have been influenced by non-bonding interactions and steric effects from the rest of the arginine molecule [472]. The medium to long-range effects of the electrostatic forces from other Gdm^+ or Cl^- ions on each other and on the solvent molecules included in the MD study [182] were not present in the PMF study, where the Gdm^+ ion pair was positioned in a cluster of water molecules [472].

Inspection of the model suggests that a C-C separation of ~6.5 Å, which creates a peak in the 2 M simulation, may be due to a SShIP where the Gdm^+ ions are in a T-shaped formation,

like Structure 6, Figure 4.28. Anion-separated Gdm^+ ions may have a similar C-C separation (C-Cl^- separation starts from ~3.5 Å (Figure 5.10a), so two Gdm^+ interacting side-by-side (Figure 5.6c) could be ~7 Å apart). These interactions seem less likely in the 2 M solution than in the 4 and 6 M solutions.

The first and second $g_{C\text{-}C}(r)$ peak CoordN compared in Table 5.7 are indicative of the number of CIP (first peak) and SShIP (second peak) interactions in each concentration[44]. The SShIP:CIP interaction ratio indicates a tendency of Gdm^+ to assemble in CIP clusters, even at low concentrations.

Table 5.7 C-C CoordN – 1st and 2nd shells

Sample conc. /M	C-C 1st peak (CIP)		C-C 2nd peak (SShIP)		SShIP:CIP CoordN ratio
	Range /Å	Ave. CoordN	Range /Å	Ave. CoordN	
2	< 5.9	0.74	5.9–7.2	0.72	0.98
4	< 5.9	1.50	5.9–7.2	1.56	1.04
6	< 5.8	1.90	5.8–7.2	2.64	1.39

Angular radial distribution functions (ARDF) of Gdm^+-Gdm^+: ARDF analysis splits the $g_{\alpha\beta}(r)$ into vector angles around a central molecule to show the directional interactions. ARDF in Figure 5.5a,c&e are plotted as a function of the angle between the z-axes of the central Gdm^+ and the surrounding Gdm^+ ions for $0° \leq \theta \leq 180°$. The z-axis of Gdm^+ is perpendicular to the plane of the ion (see Figure 4.36 for axis definitions), so the probability of the plane of the surrounding Gdm^+ interacting with the central Gdm^+ is visualised. ARDF for Gdm^+-Gdm^+ z-y axis interactions (-NH_2 group pointing towards the central Gdm^+ at 0° or side-by-side at 90° Figure 5.6b&c) are shown in Figure 5.5b,d&f.

z-z: The first interactions appear slightly closer in 4 and 6 M than in 2 M. The plots show parallel-stacked Gdm^+ in all three concentrations (pink * in Figure 5.5a,c&e), although in the 2 M solution, the stacking extends to 40–60° from parallel-stacked (offset parallel stacking Figure 5.6a) compared with regions of parallel-stacking closer to 0° and 180 ° in the 4 and 6 M solutions. There is less first shell (< ~5.9 Å) interaction intensity at θ = 90° in 6 M, whereas significant interactions are present > 6 Å (white * in Figure 5.5e). (This plot agrees with (Figure 5.5f), where the interactions towards 0° and 180° start at 6 Å.)

[44] Again, it is acknowledged that there may be Gdm^+-Gdm^+ interactions at these separation distances that may not fall into either classification, but Gdm^+ orientations cannot be distinguished from the CoordN analysis.

z-y: In first shell (< ~5.9 Å) the structures appear to change from zones of offset parallel-stacking in the 2 M solution, with the y-axis at ~60° from the z-axis of the central Gdm^+ (Figure 5.5b), to more side-by-side (θ = 90°) at 4 Å in 4 and 6 M (Figure 5.5d&f).

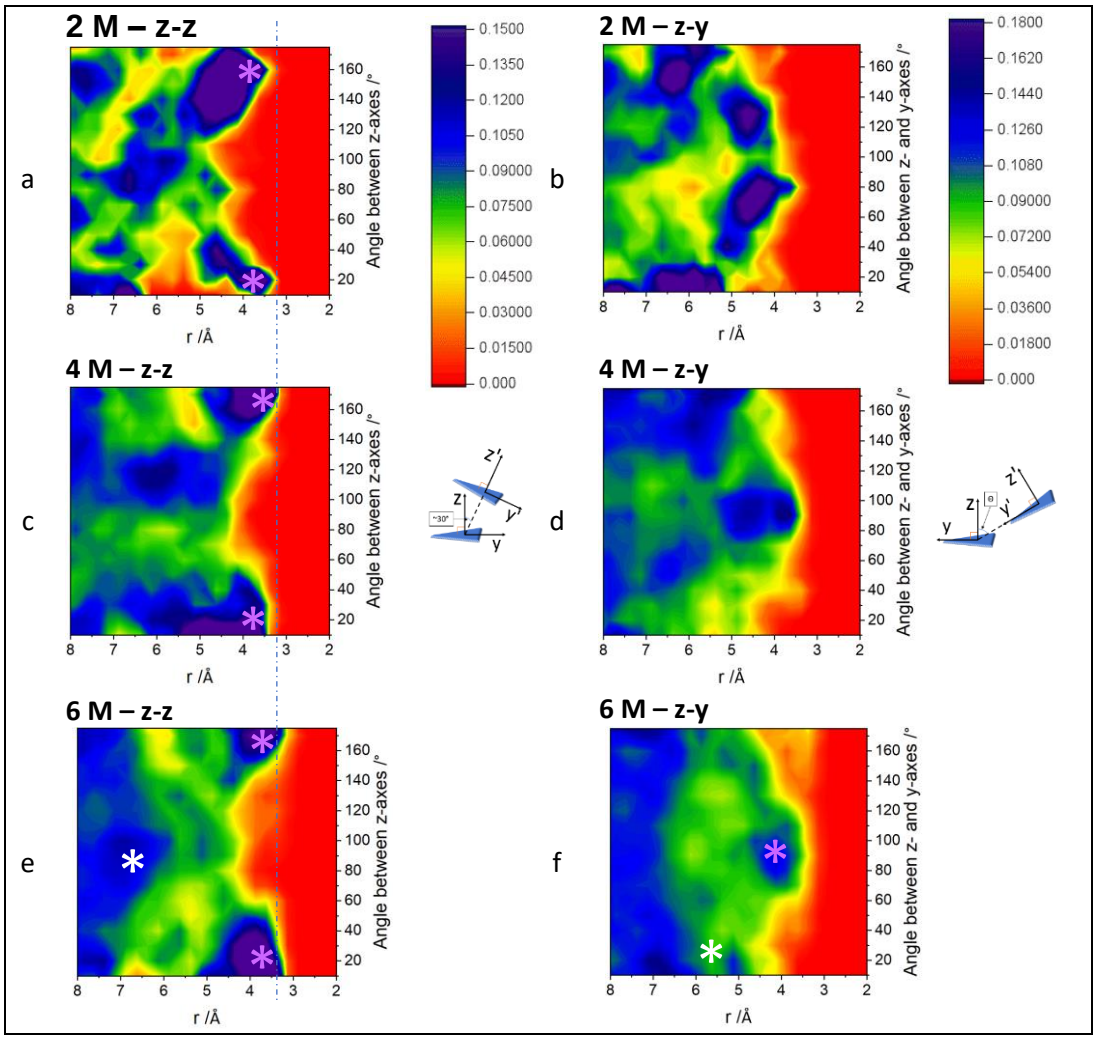


Figure 5.5 Gdm^+-Gdm^+ angular radial distribution function between the z-axes of Gdm^+ ions for a) 2, c) 4 and e) 6 M solutions, and the z-axis of the central Gdm^+ ion and the y-axis of the surrounding Gdm^+ for b) 2, d) 4 and f) 6 M solutions

The second shell structure in 6 M (6–7 Å) appears to be more intense around the plane central Gdm^+ (see white * in Figure 5.5e&f), as seen in the 6 M solution SPD (Figure 5.4c (bottom)). The same interactions are less clear in 4 M (* in Figure 5.5c&d, and Figure 5.4b), and are even less clear in the 2 M SPD (Figure 5.4a).

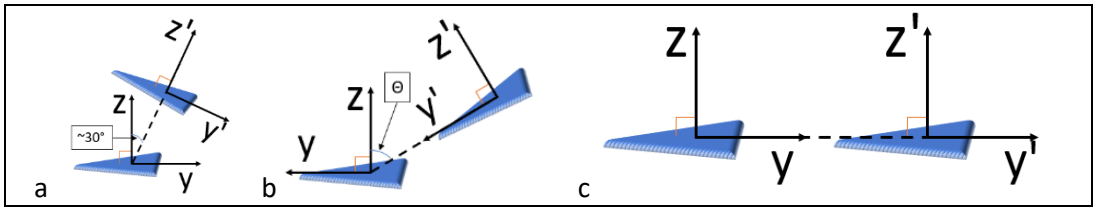


Figure 5.6 Gdm^+-Gdm^+ z-axis of the central Gdm^+ ion and the y-axis of the surrounding Gdm^+ a) offset parallel stacking, b) (non-orthogonal) T-stacked and c) side-by-side Gdm^+ ions

N-C interactions: As with $g_{C\text{-}C}(r)$, there appears to be a significant difference between the molecular structure of the 2 M solution and the higher concentration solution from 6 to 9 Å. The $g_{N\text{-}C}(r)$ peak at ~4.3 Å (* in Figure 5.7) is due to Gdm^+-Gdm^+ stacking. The asymmetrical nature of the 2 M peak could relate to the offset parallel-stacking proposed at this concentration, as the N atoms on one Gdm^+ will not be equidistant to the C on the second Gdm^+ in the dimer.

The peaks at ~6 Å and 7.5 Å are likely due to two solvent-shared Gdm^+ ions on the same plane, i.e. like Figure 5.6b where θ is close to 90°. This suggestion matches the z-y ARDF (Figure 5.5b), where there are C-C interactions at r = ~4.5 Å and θ = ~60° and ~120 °.

The peak at 7.5 Å in the 2 M solution seems to shift towards 7 Å and becomes more intense by 6 M (indicated with #), with a second peak emerging at 8.5 Å, which is consistent with the 9 M solution stage 3 $g_{N\text{-}C}(r)$ (Figure 4.34). The increasing intensity correlates with the z-y ARDF (Figure 5.5b,d&f), which increases in intensity where r > 6 Å (blue). It is proposed that the reducing separation occurs when the water molecule shared by two Gdm^+ in the 2 M solution is replaced by a Cl^- in the more concentrated solutions.

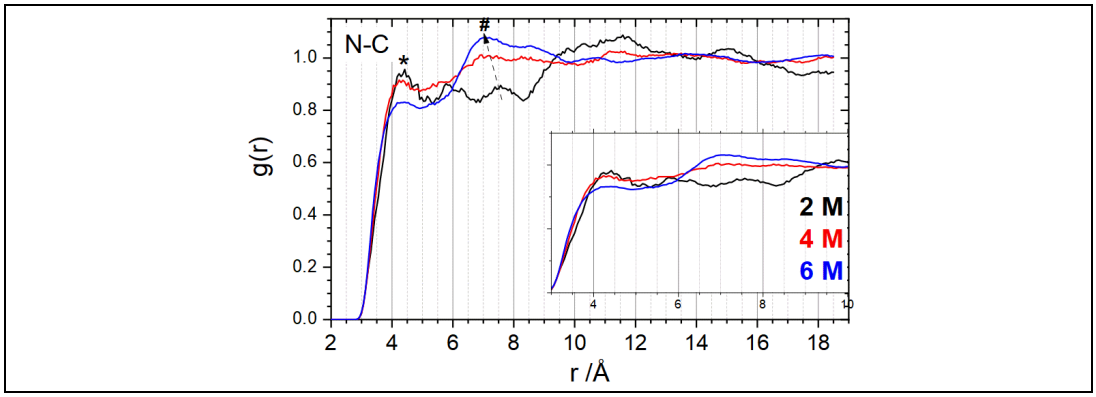


Figure 5.7 N-C partial pair distribution functions $g_{N\text{-}C}(r)$

* indicates the position of the first interaction peaks. The arrow indicated with # shows a possible change in interaction from 2 M to 6 M, with a decreasing separation between C and N

N-N interactions: There are some differences between the 2, 4 and 6 M solution $g_{N\text{-}N}(r)$ features. The 2 M solution does not have the shoulder feature at ~3.5 Å that is emerging in

the 4 and 6 M solutions and can be seen in the 9 M solutions (also included in Figure 5.8, peak indicated with *). In the *in situ* study, it is proposed that this peak is related to dimer formation, and the decreased water:ion ratio in the 9 M solution could simply make dimer formation significantly more frequent than in the lower concentration solutions.

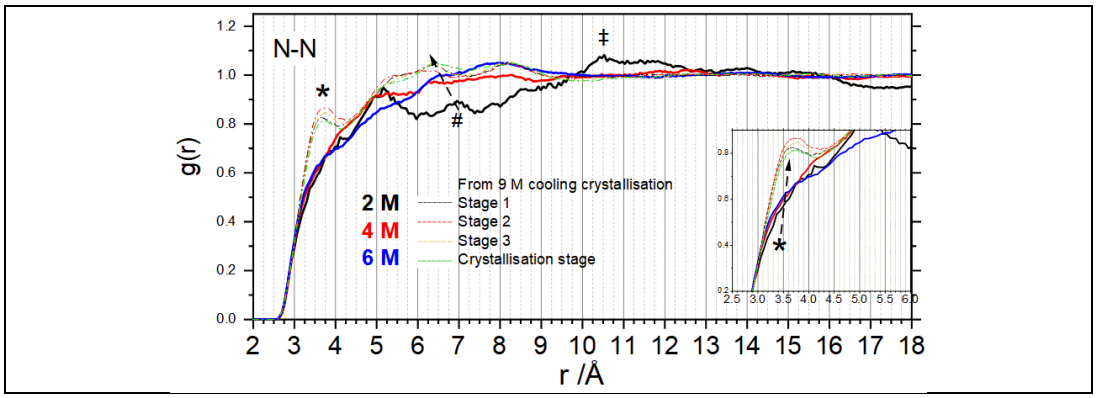


Figure 5.8 N-N partial pair distribution functions $g_{N\text{-}N}(r)$

The stage 1 (black), 2 (red), 3 (gold) and crystallisation stage (green) of the 9 M GuHCl solution $g_{N\text{-}N}(r)$ (dotted lines) are also shown

At 5.2 Å, there is a relatively sharp peak in the 2 M solution, which becomes flatter as the concentration increases. A similar feature is also present in the 9 M solution, although the intensity across the shell is greater in 9 M and the peak is shifted to 5.3 Å. As it has been proposed that the peak between 4 and 5 Å in the $g_{N\text{-}C}(r)$ could indicate offset parallel-stacked Gdm^+ with a greater separation distance, it is logical to propose that the peak at ~5.2 Å in the 2 M solution $g_{N\text{-}N}(r)$ also relates to these interactions.

A possible peak shift with concentration may occur from ~7 Å in 2 M to ~6.5 Å in the 9 M solution (highlighted # in Figure 5.8). This reducing N-N separation could occur because the number of water molecules binding to Gdm^+ reduces with increasing concentration, so solvent-separated Gdm^+ ions may become solvent-shared or even anion-shared ion pairs.

Finally, the peak seen in the 2 M solution at ~10.5 Å (‡ in Figure 5.8) is not present in the 4, 6 or 9 M solution. In the *in situ* study, many partial g(r) indicated that clustering of hydrated ions may occur up to r = ~10 Å. The change indicated with (‡) could relate to the decreasing size of the clusters with increasing concentration, e.g. if water- or Cl^--stabilised dimers, with a third Gdm^+ T-stacked with the dimer, form in the higher concentration solutions (Section 4.3.2.1).

Cl^--Cl^- pairing

$g_{Cl^-\text{-}Cl^-}(r)$ indicates that there are significant perturbations in the $g_{Cl^-\text{-}Cl^-}(r)$ up to around 11 Å (Figure 5.9). The oscillations may extend to a slightly higher r in the 2 M solution, suggesting

that larger clusters including Cl^- are present in the 2 M solution due to greater hydration of the ions at this concentration. If more solvent-separated ions are present, the water could provide charge shielding thus allowing larger clusters to form in the solution. However, the $g_{Cl^-Cl^-}(r)$ data are quite noisy at this concentration, so it is difficult to draw a conclusion.

The 4 M $g_{Cl^-Cl^-}(r)$ appears flatter than 2 and 6 M between 4.5 Å and 7.5 Å. The rerun $g_{Cl^-Cl^-}(r)$ was similar (App. Figure 60), although the reaccumulated $g_{Cl^-Cl^-}(r)$ shows a significant peak at r = 5.6 Å (App. Figure 66c). Further accumulation of the data may show whether these differences are significant.

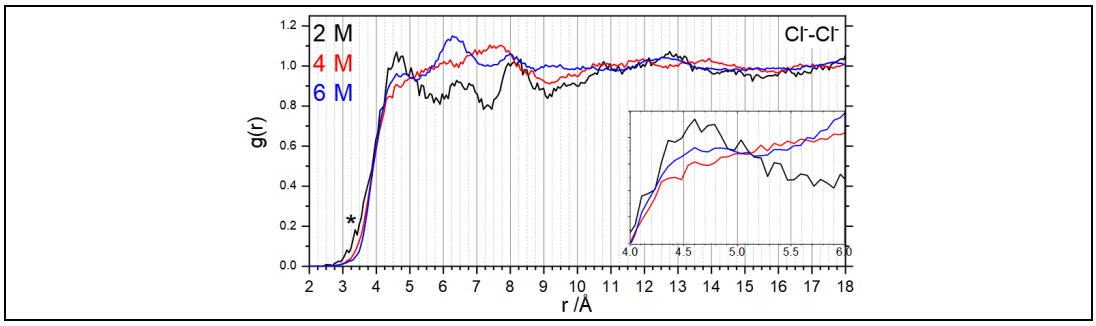


Figure 5.9 Cl^--Cl^- partial pair distribution functions $g_{Cl^-Cl^-}(r)$

Low-r interactions, in particular in the 2 M solution, are highlighted with *.

In this EPSR study, the first Cl^--Cl^- interaction peak at 4.6 Å suggests water-separated pairing of Cl^- ions. Although this is a shorter separation than suggested by Mason using NDIS/MD [97], where the water-separated Cl^--Cl^- ion pair peak was at 5.1 Å, it is in line with the proposed separation found by PMF in the literature (4.8 Å) [492]. The intensity of this peak is greater in the 2 M than 4 M or 6 M solutions, which could be a function of there being incomplete solvation of Cl^- in these higher concentration solutions due to the lower water:ion ratio. With smaller hydration shells, the solvent shielding effect will be less in the higher concentration solutions thus creating greater repulsion between anions [115].

In the 2 M solution, there is a shoulder feature[45] of the first main interaction peak at ~3.6 Å (* in Figure 5.9), which is a shorter Cl^--Cl^- interaction than is seen in the literature [182,481]. Although the parameterisation study indicated that better F(Q) and g(r) fits were generated in the EPSR model with charge reduction applied to Cl^- and Gdm^+, the comparison was based on the goodness of fit of the overall structure and did not drill down into specific atom-atom interactions. As there was a possibility that the charge reduction lessened the anion-anion repulsion resulting in shorter separation, the 2 M simulation was rerun with full charges on

[45] Further accumulation of the data may result in this feature being smoothed out.

the ions. It was found that although the intensities were lower r < 4.7 Å, Cl^--Cl^- interactions still occurred from ~3 Å.

We will see in the next section that C-Cl^- pairing occurs in all concentrations. The peak at 6.5 Å could relate to two Cl^- ions binding to -NH_2 groups in the plane of the same Gdm^+ ion [481], although it is suggested elsewhere that this separation would be greater (7.4 Å, [97]). It is interesting that this interaction appears in the 2 and 6 M $g_{Cl^-Cl^-}(r)$, but not 4 M $g_{Cl^-Cl^-}(r)$. As Cl^- and water compete for the same interaction sites around Gdm^+, it is reasonable to infer that such interactions are common in the 6 M solution, where there are insufficient water molecules to hydrate the Cl^- and Gdm^+ ions meaning that Gdm^+ may be more likely to directly interact with more than one Cl^-.

The $g_{Cl^-Cl^-}(r)$ from this XPDF/EPSR study are dissimilar to Cl^--Cl^- interactions identified from an MD study [182], where the distribution for each concentration has clear peak positions at ~5 Å and ~7.5 Å, but no interaction at 6.5 Å at any concentration. As the EPSR simulations described here have been refined to experimental data, it is possible that the MD alone may not have captured some of the interactions taking place in the solutions, especially as the characteristic $g_{O\text{-}O}(r)$ peaks also appear to be absent in the pure water simulation run alongside the GuHCl solutions [182].

The CoordN of Cl^- around Cl^- (Table 5.8) shows first shell coordination in the 6 M solution is almost double that of the 2 M solution. In the second shell, this increases four-fold in the 6 M solution compared to the 2 M solution, despite only a three-fold increase in the number of Cl^- ions in the 6 M solution, indicating that there may be more short-to-medium-range ion clustering in the 6 M solution.

Table 5.8 CoordN of Cl^--Cl^-

Sample conc. /M	Cl^--Cl^- 1st shell		Cl^--Cl^- 2nd shell	
	Range /Å	CoordN	Range /Å	CoordN
2	< 5.8	0.67	5.8–7.3	0.86
4	< 5.1	0.69	5.1–6.4	1.30
6	< 5.2	1.16	5.2–7.3	3.96

Cl^--Gdm^+ pairing

The $g_{Cl^-C}(r)$ show that ion-pair separation distances are similar across the concentrations (Figure 5.10a). The peak positions appear to be at ~3.9 Å and ~4.3 Å, representing Cl^- hydrogen-bonding to two -NH_2 groups, or a single hydrogen bond with one H. However, as these positions are higher than those in the literature (3.7 Å and 4.1 Å [97]), the peaks

were deconvoluted using Fityk [493] (Appendix G, App. Figure 67) to identify the first three peak positions, as there appears to be an additional feature on the 2 M $g_{Cl^-\text{-}C}(r)$ from ~4.8 Å (* on Figure 5.10a). Inspection of the molecular model indicates that Peak 3 could represent a solvent-separated interaction where Cl^- is positioned perpendicular to the face of Gdm^+. Three peaks were identified in each concentration, at 3.8 Å, 4.2 Å (closer to the expected values) and ~4.7 Å (Table 5.9).

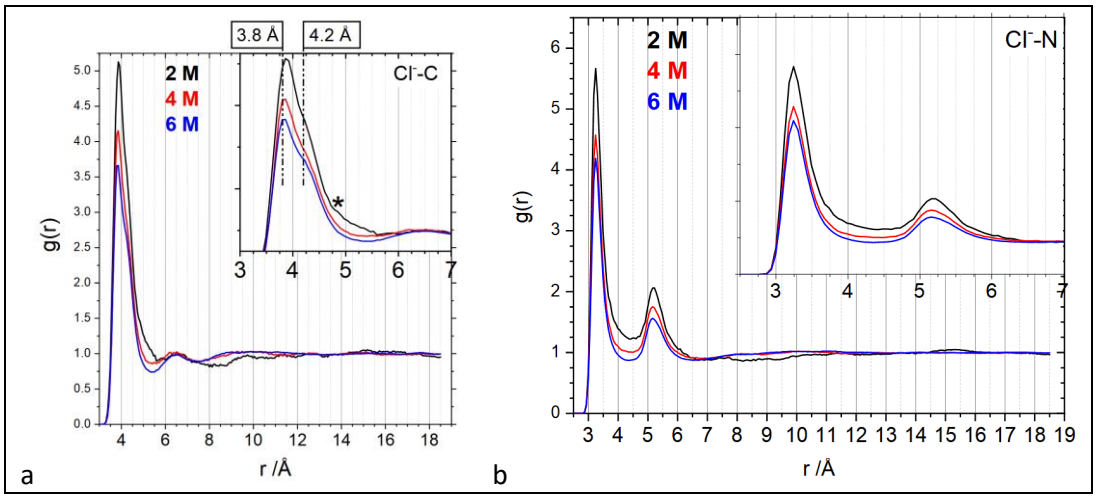


Figure 5.10 Partial pair distribution functions for a) Cl^--C ($g_{Cl^-\text{-}C}(r)$) and b) Cl^--N ($g_{Cl^-\text{-}N}(r)$)

In a), * highlights a feature in the 2 M solution that is not present in the 4 and 6 M solution g(r), where a clearer trough between the peaks is observed

Table 5.9 Deconvolution of $g_{Cl^-\text{-}C}(r)$

Sample conc. /M	Parameter	Peak 1	Peak 2	Peak 3
2	Position /Å	3.8	4.2	4.8
	Area	3.63	2.85	0.93
4	Position /Å	3.8	4.2	4.7
	Area	1.16	1.61	0.71
6	Position /Å	3.8	4.2	4.7
	Area	1.04	1.54	0.66

The area of the peak decreases with increasing concentration, suggesting that this arrangement is more common in the lower concentration solutions, possibly as part of the dome-like structure of water that may form across the face of the Gdm^+ ion, where sufficient water molecules are available. C-Cl^- CoordN are shown in Table 5.10, where the midpoints of the deconvoluted peaks have been selected as the r_{min} and r_{max} of the coordination shells. This CoordN data suggests that 50% of Gdm^+ in the 6 M solution are paired with Cl^- ions around the plane, compared with 20% in the 2 M solution.

Table 5.10 CoordN of Cl^- around C

Sample conc. /M	C-Cl^- 1st shell		C-Cl^- 2nd shell		C-Cl^- 3rd shell		Total of shells	
	Range /Å	CoordN	Range /Å	CoordN	Range /Å	CoordN	1–2	1–3
2	< 4.0	0.06	4.0–4.5	0.13	4.5–5.8	0.49	0.18	0.67
4	< 4.0	0.09	4.0–4.5	0.22	4.5–5.8	1.00	0.31	1.31
6	< 4.0	0.12	4.0–4.5	0.34	4.5–5.8	1.51	0.46	1.97

The XPDF analysis of the solid GuHCl reveals that the first two principal C-Cl^- interactions in solid are at ~3.7 Å and ~4.7 Å, so in solution, the shortest Gdm^+-Cl^- ion pairings separations are as those found in the final crystalline structure.

The spatial probability density (SPD) analysis of Cl^- (green) and water (yellow) around a central Gdm^+ ion (Figure 5.11), is displayed with probability set to the 5% most likely interaction sites for Cl^- and the top 1% for water in row 1, 4% (Cl^-) and 1% (water) in row 2 and 4% (Cl^-) and 2% (water) in row 3.

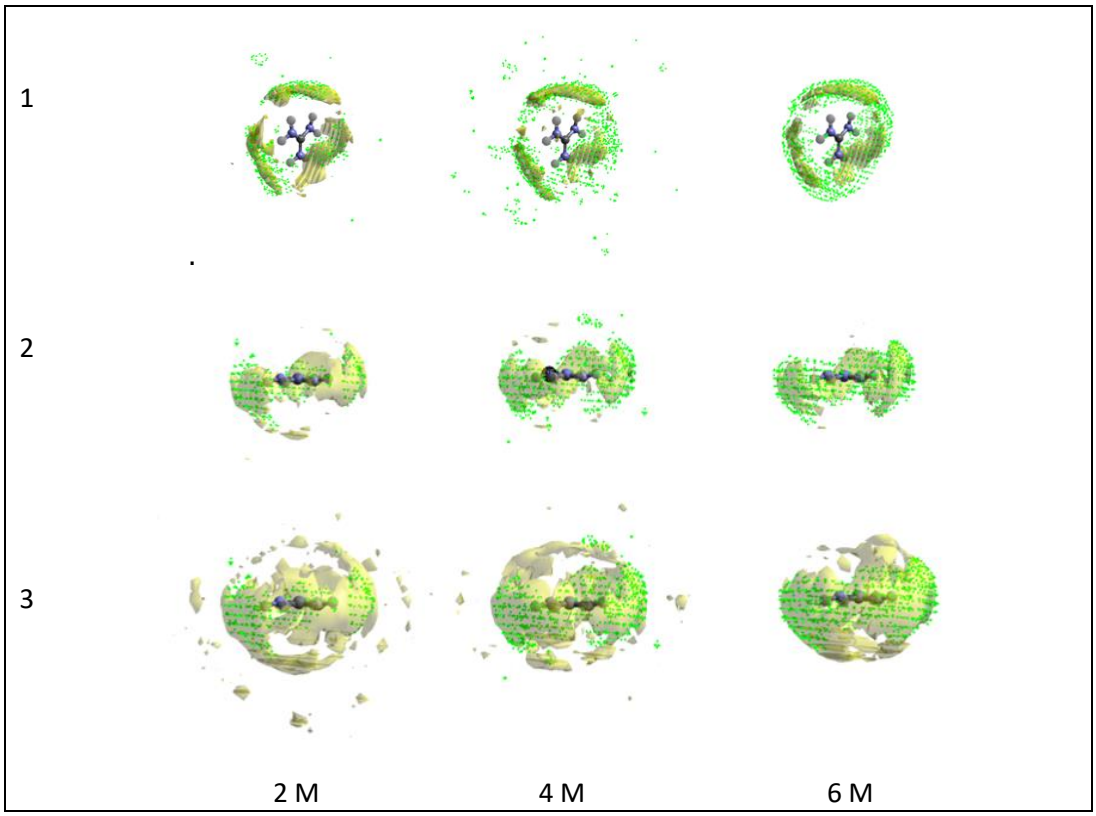


Figure 5.11 Spatial probability densities of Cl^- (green) and water (yellow) around Gdm^+ for a) 2, b) 4 and c) 6 M GuHCl solutions

Row 1) Top 5% most likely interaction sites for Cl^- and 1% for water, Row 2) Top 4% most likely interaction sites for Cl^- and 1% for water, Row 3) Top 4% most likely interaction sites for Cl^- and 2% for water.

The SPD show that Cl^- and water compete for positions around the plane of Gdm^+. It is not clear why the distribution of Cl^- around Gdm^+ in the 4 M solution is more scattered than for

the 2 or 6 M solutions. The bands of C-Cl^- interactions increasingly broaden away from the Gdm^+ plane as concentration increases (Figure 5.11, row 3). This broadening is also evident in Figure 5.12, where the location of Cl^- across the face of Gdm^+ can be observed. It is also interesting to note that the lower the concentration of the solution, the more likely it appears to be for the Cl^- to be in the interstitial position (Figure 5.12, orange arrow), whereas at higher concentrations, the Cl^- seems equally likely to be in the single hydrogen bond position (blue arrow).

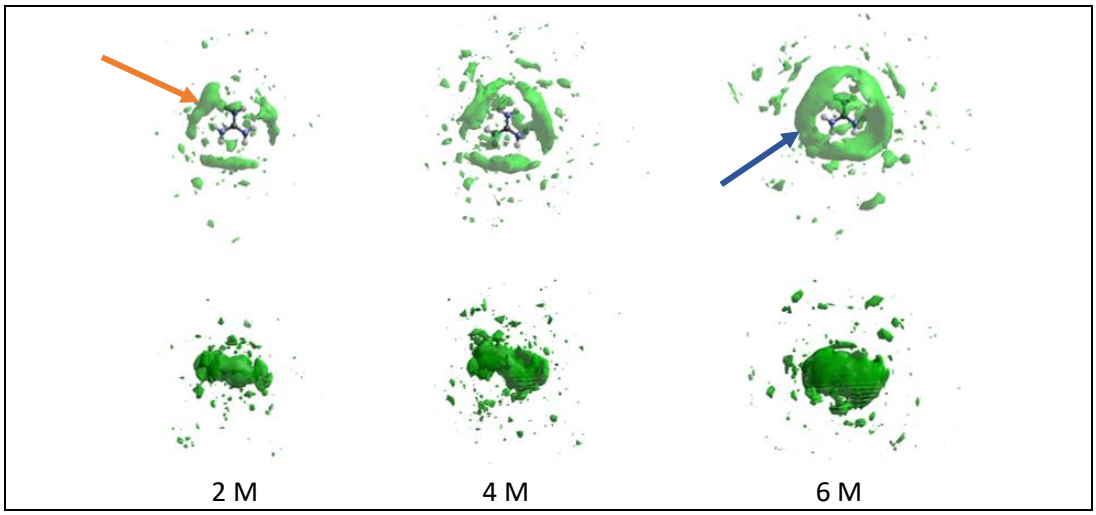


Figure 5.12 Spatial probability densities of Cl^- (green) around Gdm^+ for a) 2, b) 4 and c) 6 M GuHCl solutions: top 7% most likely interaction sites for Cl^-

Orange arrow: shows the interstitial position, blue arrow: single hydrogen bond position

The $g_{Cl^--N}(r)$ peaks at 3.2 Å and 5.2 Å are consistent with the features in the solution G(r) that emerge from 2 M to 6 M solutions (Figure 5.1, indicated with *). It is reasonable to assume that the features seen in the G(r) are related to this interaction, as chloride ions are highly scattering and N atoms form a high proportion of the scattering (non-H) atoms in the solution[46].

5.3.2.3 Ion solvation

C-water interactions: It is recognised in the literature that Gdm^+ does not have an isotropic hydration shell, instead water molecules interact around the plane of Gdm^+, with cavities forming around the hydrophobic faces due to the non-hydrogen-bonding central C [97,181,188]. In Figure 5.11, row 3, a shell of water is starting to form over the hydrophobic

[46] 11% (2 M) to 29% (6 M)

face of Gdm^+ (orange arrow), although this region is where Gdm^+ ions interact to form dimers [97,181,188].

This analysis supports the literature, as $g_{C\text{-}Ot}(r)$ indicates that there is no direct hydrogen bonding between the C and O_t, as interactions begin ~2.9 Å (Figure 5.13a). The hydrophobicity either side of the Gdm^+ plane, combined with the potential for hydrogen bonding with water around the plane, gives Gdm^+ the potential to destabilise proteins, with the Cl^- counterion increasing this effect [97]. Water hydrogen-bonding with the $-NH_2$ groups will dominate C-O_t interactions from ~3.0 Å up to ~4.6 Å [97], with shorter interactions due to the position of weakly bound water molecules above and below the Gdm^+ ion forming the shoulder at ~3.5 Å [47] (Figure 5.13a, indicated with *) and interactions at ~3.9 Å due to water hydrogen-bonding to H atoms on the amino ($-NH_2$) groups.

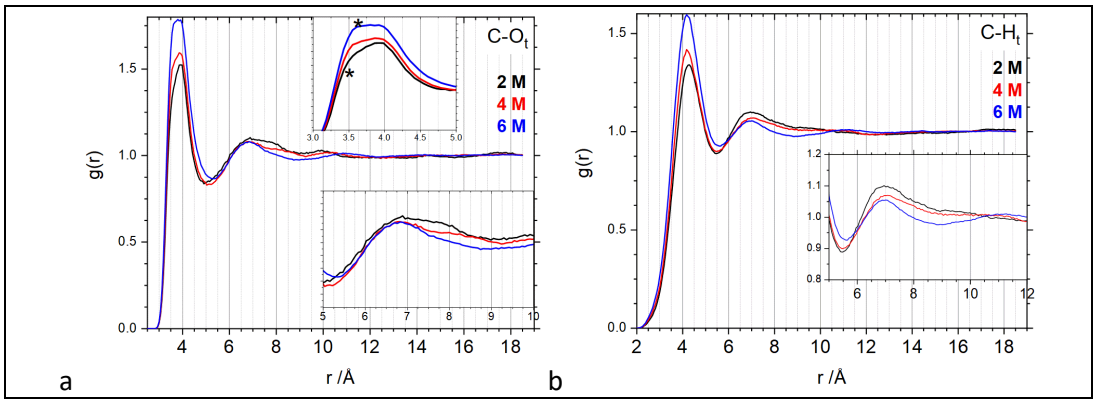


Figure 5.13 Partial pair distribution functions for a) C-O_t ($g_{C\text{-}Ot}(r)$) and b) C-H_t ($g_{C\text{-}Ht}(r)$)

Table 5.11 C-O_t peak positions

Sample conc. /M	C-O_t peak positions			
	1st peak shoulder position /Å	1st main peak position /Å	2nd peak position /Å	3rd peak position /Å
2	3.4	3.8	6.8	8.0
4	3.5	3.9	6.9	8.5
6	3.5	3.9	6.9	9.4

Table 5.12 shows the CoordN for O_t around the central C in Gdm^+ for the limit of direct C-O_t hydrogen-bonding (< 2.5 Å), for the first coordination shell from $g_{C\text{-}Ot}(r)$ (Figure 5.13a) and for the 4.6 Å interaction limit considered in the NDIS/MD study of the structure of a GuHCl solution with a water:GuHCl ratio of 18.4:1 (3 molal concentration) [97]. The 3 molal solution concentration modelled is between 2 and 4 M (although closer to 2 M) and found ~9.9 C-O_t interactions per Gdm^+ ion [97]. The number of water molecules associated with the Gdm^+

[47] Based on peak fitting in Fityk

ion within 4.6 Å reduces with concentration (Table 5.12) and is lower across the concentrations than was found in the 3 molal MD simulation [97].

Table 5.12 CoordN of water around central C in Gdm^+

Sample conc. /M	C-O_t hydrogen bonding limit		C-O_t 1st shell		C-O_t 1st shell (as described in [97])		C-H_t 1st shell	
	Range /Å	CoordN	Range /Å	CoordN	Range /Å	CoordN	Range /Å	CoordN
2	< 2.5	0.00	< 4.9	11.46	< 4.6	9.38	< 5.5	33.25
4	< 2.5	0.00	< 5.1	11.47	< 4.6	8.38	< 5.5	29.06
6	< 2.5	0.00	< 5.3	10.23	< 4.6	7.55	< 5.7	28.64

To distinguish between the hydrogen-bonding interactions around the plane of Gdm^+ and the weakly-bound water molecules above and below the Gdm^+ ion, the N-H···O_t interactions, which represent the water molecules hydrogen-bonding around the plane of the Gdm^+, are considered in Table 5.13. The H-O_t CoordN is calculated to include all interactions < 2.5 Å. It is possible that a water molecule could potentially be double-counted if the O_t is hydrogen-bonding with, or just within, 2.5 Å of a second H. In each solution, there will be a combination of interstitial N-H···O_t interactions and single-hydrogen-bonded N-H···O_t interactions, as described for Cl^- in Figure 5.11, although it is more likely that there will be a greater proportion of single hydrogen-bonded interactions in 2 M and interstitial-positioned water in 6 M. Therefore the total number of hydrating waters in the plane of Gdm^+ could range between three times and six times the CoordN, as indicated in Table 5.13. For 2 M, the data suggest there are on average 9.38 water in the shell r_{max} = 4.6 Å, with up to 4.83 of these being N-H···O_t interactions in the plane of the Gdm^+ thus 4.55 in the diffuse cloud across the face of the ion. These values suggest fewer waters directly hydrogen-bonding to Gdm^+ than in the 3 molal MD simulation [97] (~9.9 water molecules, of which 3 are weakly bound over the plane).

Table 5.13 H-O_t CoordN

Sample conc. /M	H-Ot hydrogen bonding limit		Water hydrogen-bonding between 2 -NH_2 groups (CoordN x 3)	Water hydrogen-bonding to single H (CoordN x 6)
	Range /Å	CoordN		
2	< 2.5	0.80	2.41	**4.83**
4	< 2.5	0.71	2.12	**4.24**
6	< 2.5	0.61	1.83	**3.66**

N-water interactions: The first peak in the $g_{N\text{-}Ot}(r)$ is the hydrogen-bonding interaction of a water molecule with H_t on an amino group on the Gdm^+ ion (Figure 5.14a). The second peak is due to O_t interacting with the N on the opposite side of the Gdm^+ ion (Figure 5.14c).

CoordN of N-H_t (Table 5.15) is more than double the N-O_t CoordN (Table 5.14), as H_t from water molecules that are not in the first coordination shell may be included [473]. Defining the precise r_{max} of the shells is also difficult from the partial g(r) for the non-spherical Gdm^+ ion.

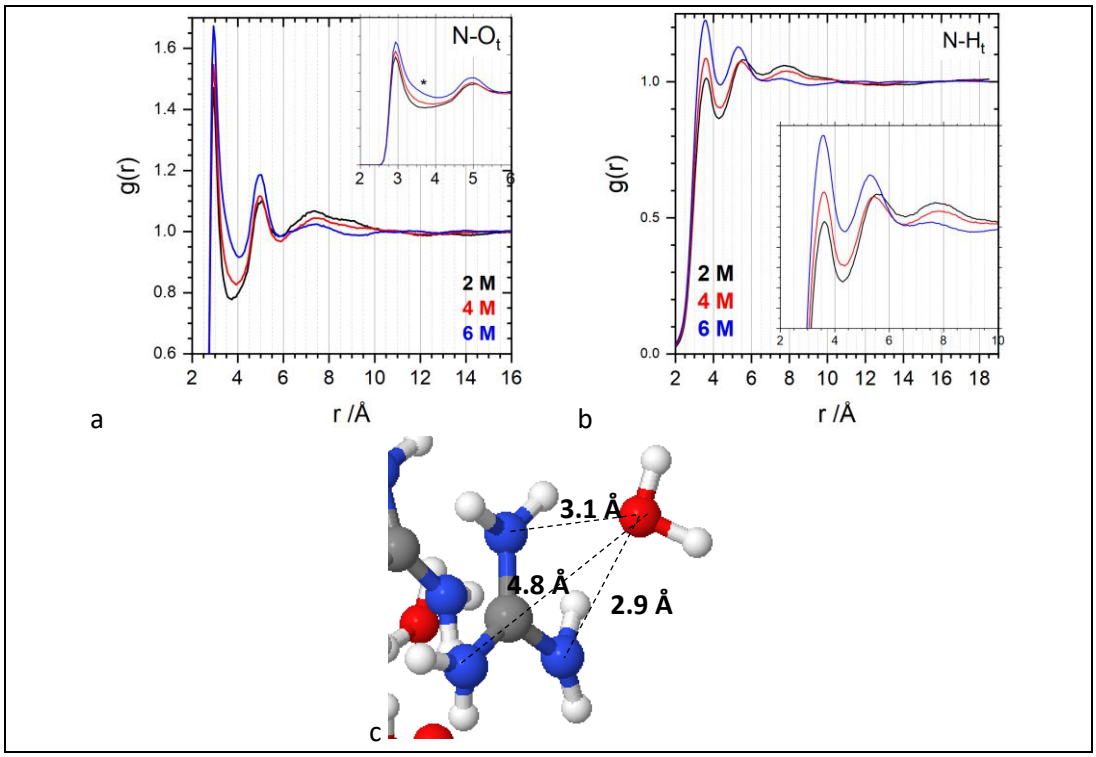


Figure 5.14 Partial pair distribution functions for a) N-O_t ($g_{N\text{-}Ot}(r)$) and b) N-H_t ($g_{N\text{-}Ht}(r)$). c) shows a possible arrangement of O_t around Gdm^+ to result in the peaks described in a)

The N-O_t CoordN decreases with concentration when considering hydrogen-bonding limit of 3.8 Å. Taking r_{max} of the first peak, the CoordN increases. The asymmetry of the first peak suggests that there is an additional and increasing interaction around 3.0 Å around the first shell. As this feature occurs more with increasing concentration, it could relate to water-stabilised dimer formation.

Table 5.14 CoordN of water (O_t) around N atoms

Sample conc. /M	N-O_t hydrogen bonding limit		N-O_t 1st shell		N-O_t 2nd shell	
	Range /Å	CoordN	Range /Å	CoordN	Range /Å	CoordN
2	< 3.8	4.07	< 3.7	3.59	3.7–5.8	16.64
4	< 3.8	3.68	< 3.9	4.01	3.9–5.8	13.34
6	< 3.8	3.33	< 4.1	4.37	4.1–5.8	10.55

Table 5.15 CoordN of water (H_t) around N atoms

Sample conc. /M	N-H_t 1st shell		N-H_t 2nd shell	
	Range /Å	CoordN	Range /Å	CoordN
2	< 4.4	13.84	4.4–6.6	49.22
4	< 4.4	12.40	4.4–6.6	41.05
6	< 4.3	10.10	4.3–6.6	35.06

An analysis of hydrogen bond interaction angles (N-H···O_t) was carried out using the Triangles EPSR auxiliary routine (Appendix E). It should be noted that the N-H···O_t hydrogen bond separation applied in the analysis was 2.8 Å, rather than 2.5 Å, the hydrogen bond distance that has been applied elsewhere for water molecules. Although this analysis is only based on the first set of simulations, it could still provide an indication of trends between the concentrations.

The probability of finding linear hydrogen bond angles appears slightly lower in the 6 M solution (Figure 5.15, solid lines for Gdm^+-water). A slight increase in the probability of interactions ~115° can also be seen in the 6 M solution (Figure 5.15, highlighted with *). If real, these bond angle differences may relate to smaller Gdm^+ dimer clusters being present in the 6 M solution due to fewer water molecules being available to stabilise the clusters, with fewer linear water-Gdm^+ interactions as a result (Table 5.16). However, the differences in Figure 5.15 are small, and so further analysis would be required to understand if these differences are repeatable.

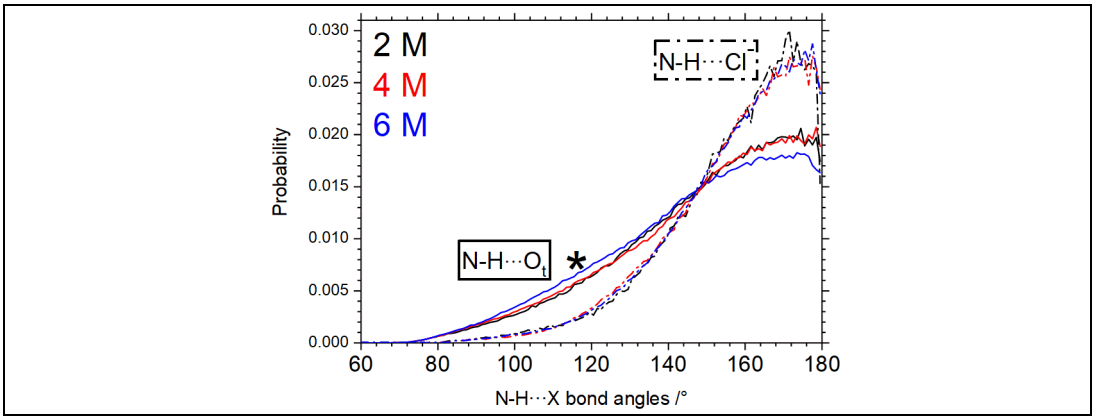


Figure 5.15 Angle distributions of hydrogen-bonding for Gdm^+-water (N-H···O_t) (solid lines) and Gdm^+-Cl^- (N-H···Cl^-) (dotted lines) interactions. (Data from original simulations.)

Table 5.16 Distribution of interaction angles of N-H···O_t hydrogen-bonding

Sample conc. /M	% total (N-H···O_t) interactions			
	80–150°	150–160°	160–170°	170–180°
2	42.9	16.7	18.9	21.4
4	42.8	16.7	18.8	21.6
6	46.7	16.1	17.5	19.4

The angle distributions of N-H···Cl^- (Figure 5.15, interactions up to 2.8 Å) indicate a higher probability of finding ion-pair hydrogen bonds > 160°. These data may suggest that water molecules are more likely to stabilise dimer formation than Cl^-. The N-H···Cl^- distributions in particular are noisy due to the relatively low number of ion-ion interactions in solution, so further accumulation of structures would have been beneficial.

Cl^--water interactions: Figure 5.16 shows the site-site interactions for water around chloride ions: a) Cl^--O_t and b) Cl^--H_t.

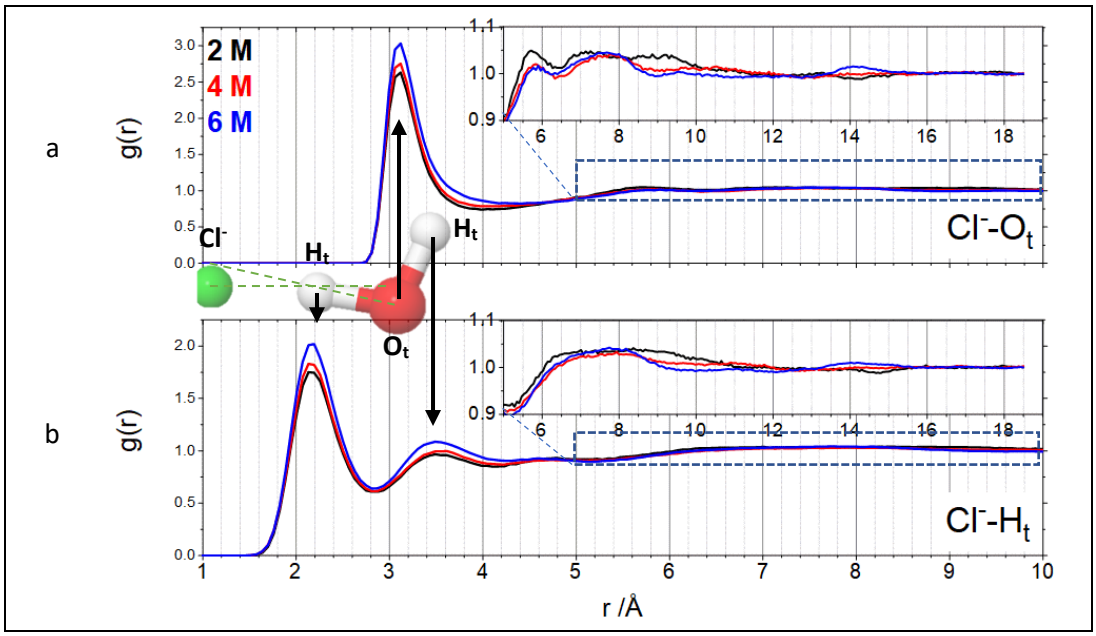


Figure 5.16 Partial pair distribution functions for water around Cl^-: a) $g_{Cl^-\text{-}Ot}(r)$ and b) $g_{Cl^-\text{-}Ht}(r)$

The water molecule positioned alongside the Cl^- ion illustrates how it might be orientated to result in the g(r) peaks in a) and b)

The first Cl^--H_t interaction peak is similar for each concentration (~2.2 Å), which is close to the value in the literature for isotopically enriched 2 molal $NiCl_2$ aqueous solutions studied by neutron scattering (2.28 Å [473]). The solvent-induced effects by the larger hydration shell in the 2 molal $NiCl_2$ solution could reduce the electrostatic attraction and result in Cl^--H_t atomic separation being greater. The second peak is at a slightly lower r for 6 M (3.50 Å) than for 2 & 4 M (3.55 Å)[48], which could be related to the steric effects of slightly different water molecule orientations around the Cl^- ion – a more linear hydrogen bond would reduce the second peak position. Alternatively, the second Cl^--H_t peak may include H_t in water molecules that are not in the first coordination shell (penetration effects [473]) thus

[48] Again slightly lower than the value in the literature (3.7 Å in the previously mentioned neutron scattering study [473])

distorting the peak position to a higher r-value for the 2 or 4 M solutions, where water:ion ratio is higher. The peak position of the third Cl^--H_t coordination shell is 4.7 Å for 2 and 4 M, and 4.6 Å for 6 M.

CoordN of water around Cl^- for each concentration is listed in Table 5.17. CoordN for the first shell of H_t around Cl^- reduces with concentration (Table 5.17), which reflects the decreasing water:GuHCl ratio with increasing concentration. CoordN in the literature is ~6 (for a low-concentration solution) [105,473]. Powell [473] found that the Cl-H_t (first peak) interaction should be considered as the primary CoordN of hydrated Cl^- ions, rather than the Cl-O_t CoordN, as penetration effects by water molecules outside the first solvation shell affect the Cl-O_t CoordN.

Table 5.17 CoordN of water around Cl^-

Sample conc. /M	Cl^--H_t 1st shell		Cl^--H_t 2nd shell		Cl^--O_t up to:	
	Range /Å	CoordN	Range /Å	CoordN	Range /Å	CoordN
2	< 2.8	4.24	2.8–4.1	9.76	< 4.0	6.23
4	< 2.8	3.68	2.8–4.1	8.42	< 4.1	5.96
6	< 2.8	3.20	2.8–4.2	7.93	< 4.2	5.61

The separation of the first and second Cl-H_t peaks in Figure 5.16 gives an indication of the linearity of the hydrogen bond O_t-H_t···Cl^-. The distribution of this interaction angle across the concentrations (Figure 5.17a) is extracted from the model using the Triangles auxiliary routine (Appendix E.1.3). The proportion of linear hydrogen bonds is slightly lower in the 2 M solution, where complete ion hydration shells may be possible if solvent molecules are shared by the ions (Table 5.18).

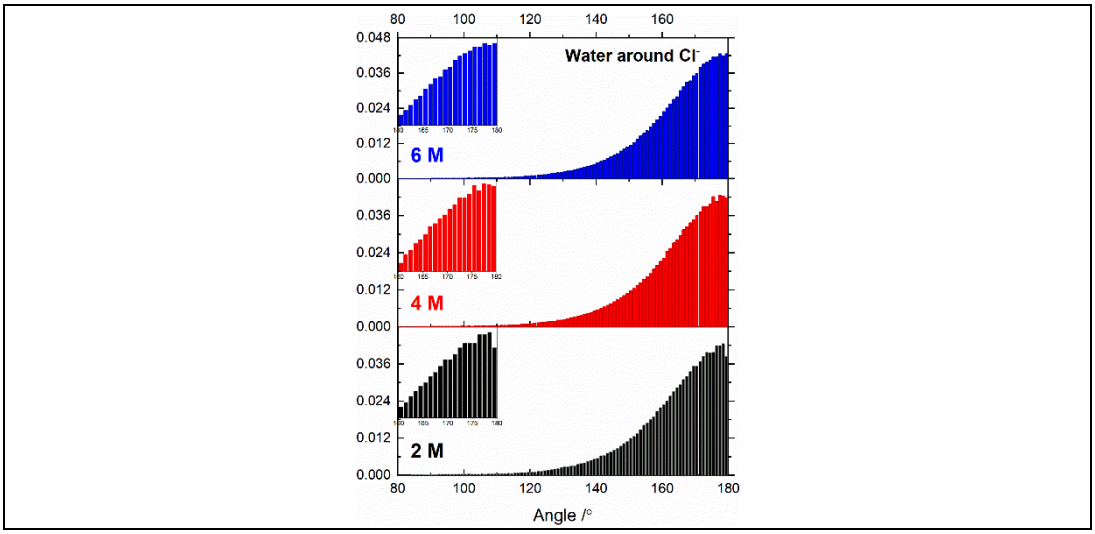


Figure 5.17 Angular distributions of coordination of water around Cl^- (O_t-H_t···Cl^-)

Table 5.18 Distribution of interaction angles of water around Cl^-

Sample conc. /M	% bonds	
	170-180°	175-180°
2	39.5	20.5
4	40.5	21.0
6	40.3	21.0

5.3.2.4 Water structure

There is extensive literature on the structure of pure water, and aqueous ionic solutions, in the literature, with refs. [103,105,107,232,235,323,375,494–497] being just a small sample. $g_{Ot\text{-}Ot}(r)$, $g_{Ot\text{-}Ht}(r)$ and $g_{Ht\text{-}Ht}(r)$ for water in the aqueous GuHCl solutions are shown in Figure 5.18a-c, with the relevant $g_{\alpha\beta}(r)$ for pure water from Soper (2013) [375] for comparison.

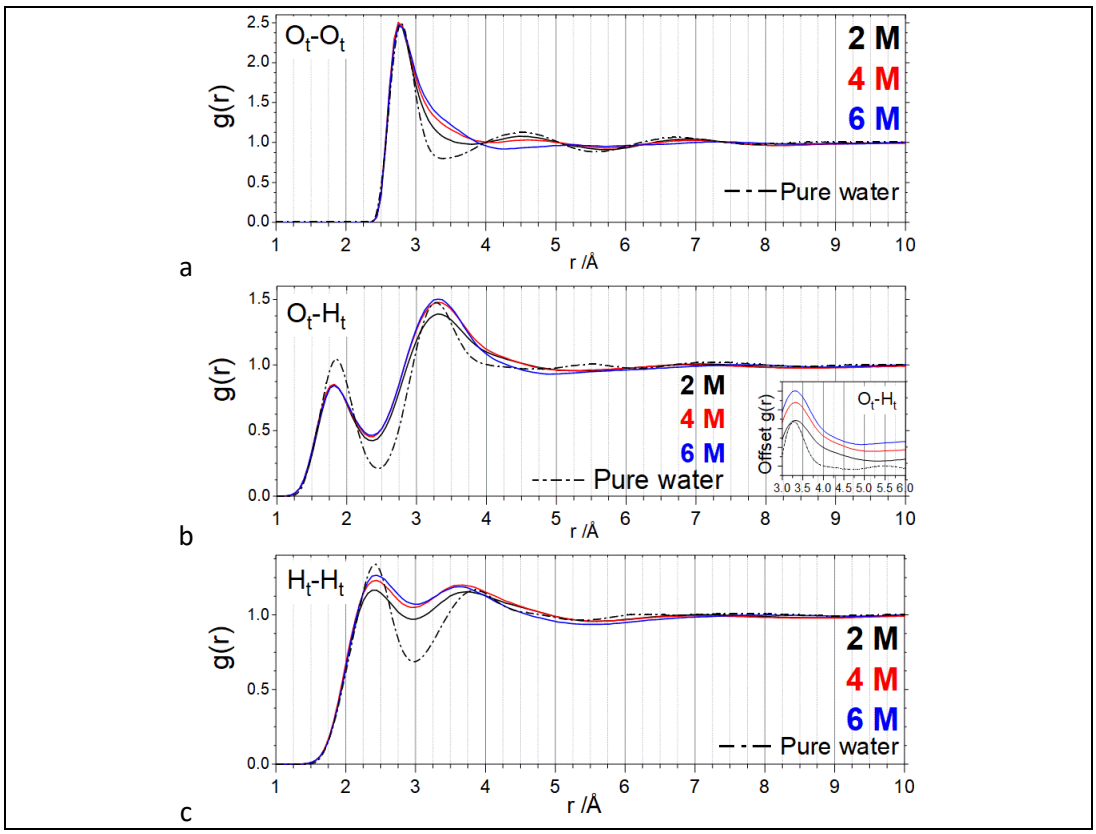


Figure 5.18 Intermolecular pair distribution functions for water: a) $g_{Ot\text{-}Ot}(r)$, b) $g_{Ot\text{-}Ht}(r)$ and c) $g_{Ht\text{-}Ht}(r)$. Pure water data (dashed lines) from ref. [375]

Although $g_{Ot\text{-}Ot}(r)$ has been shown to be a more sensitive measure of the distortion of medium and long-range water structure than $g_{Ot\text{-}Ht}(r)$ and $g_{Ht\text{-}Ht}(r)$ [105], the higher minima intensity in all water-water correlations from the GuHCl solutions when compared with pure water indicate disorder from the tetrahedral structure [235].

The first $g_{Ot\text{-}Ot}(r)$ peak, at 2.8 Å for all concentrations, broadens with increasing concentration. The second peak, a marker for tetrahedral bonding in water [105], disappears

with increasing concentration. The shoulder feature at 3.5 Å is also present in the *in situ* analysis of the 9 M GuHCl solution (Chapter 4), where it was proposed that it may be caused by the proximity of water molecules that are hydrating the same ion, or is the result of the electrostrictive effect of charged solutes on the water structure [98]. Electrostriction has been observed in aqueous solutions of NaOH [476], and NaCl and KCl (with water:ion ratios are similar to this study) [105], manifesting as a shift of the second O_t-O_t peak position to a lower r, becoming a shoulder or broadening of the first peak at high concentrations. Although Marcus [98] showed that Gdm^+ does not cause electrostriction, this effect could be due to the Cl^- counterion.

However, whilst a feature at 3.5 Å grows with increasing concentration, the second O_t-O_t peak position does not appear to merge with the first, as seen in NaCl and KCl solutions of increasing concentration [105]. Whilst the second O_t-O_t peak is reasonably well-defined for 2 M, and is consistent with pure water, the 4 M second peak is less defined at ~4.6 Å and the low-intensity 6 M second peak is at ~5.2 Å. This decreasing second shell structure with increasing concentration reflects decreasing water:Gdm^+ ratio, where bulk water is reduced thus there are insufficient water molecules to maintain structure in the higher concentration solutions. It would be interesting to acquire more data for lower concentration GuHCl solutions to monitor the effect of Gdm^+ and Cl^- ions on the O_t-O_t correlations.

The double peaks seen in the pure water O_t-H_t and H_t-H_t functions are characteristic of strongly directional hydrogen bonding [235]. The $g_{Ot\text{-}Ht}(r)$ first and second peak position shifts are negligible compared with pure water (~0.03 Å shorter for the first peak and ~0.02 Å longer for the second). However, the first peak is broader and less intense, which could be linked to hydrogen bond breaking [105]. Figure 5.19 show SPD illustrating water-water interaction sites.

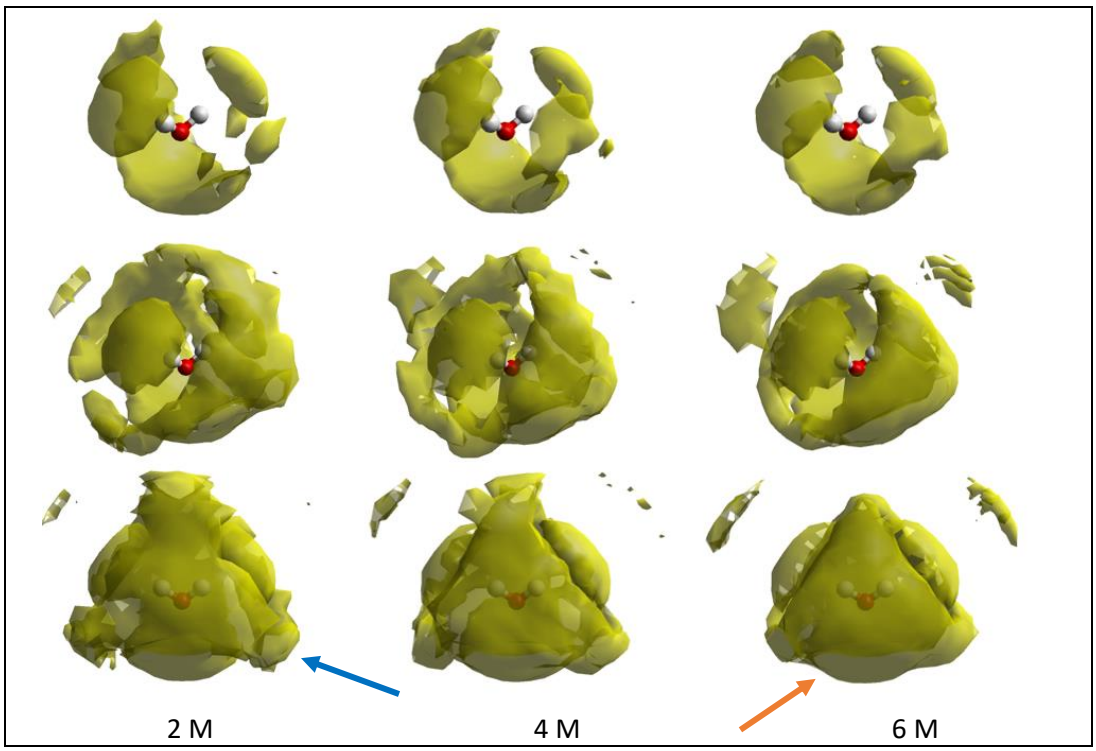


Figure 5.19 Spatial probability densities of water around water for a) 2, b) 4 and c) 6 M GuHCl solutions: top 1% (top row) and 2% (middle & bottom rows) most likely interaction sites

Based on data from original simulations only.

The decreasing water:Gdm^+ ratio with increasing concentration means that the second coordination shell will be increasingly disrupted. The 2 M solution has a defined second shell that appears in antiphase with the first shell (Figure 5.19, blue arrow), which is similar to, but less well formed than pure water at 280 K (Figure 4.45b). In the 6 M solution, there is a merging of the first and second shells (orange arrow).

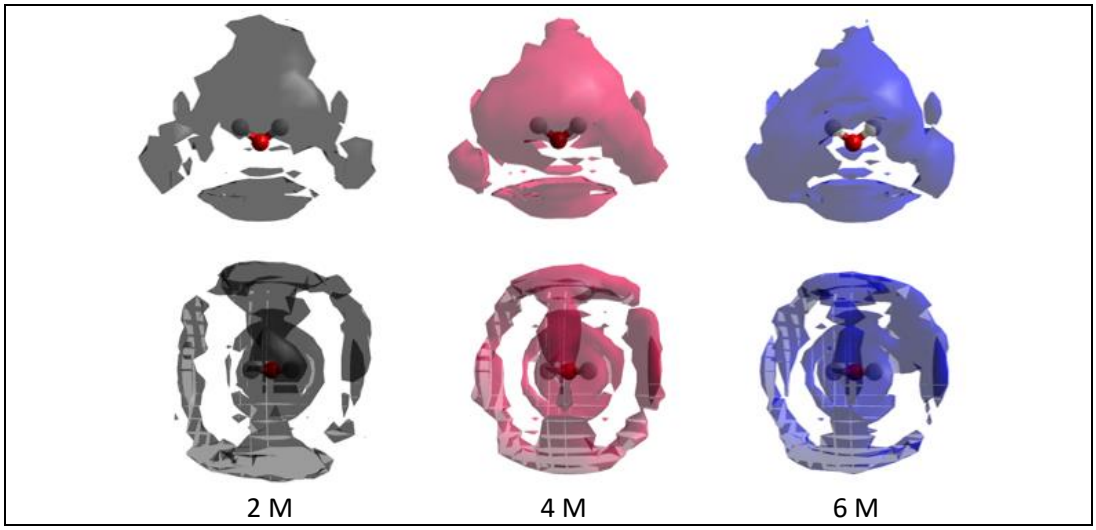


Figure 5.20 Spatial probability densities of water around water for a) 2, b) 4 and c) 6 M GuHCl solutions 3–5 Å: top 15% most likely interaction sites

Based on data from original simulations only.

The first shell shoulder in the $g_{Ot\text{-}Ot}(r)$ (Figure 5.18a) emerges between 3 Å and 5 Å. The 15% most likely interactions sites for r = 3–5 Å are shown in Figure 5.20. The density of the shells, and the merging of the first and second shells, increases from 2 M to 6 M.

Water internal bond angles: The bond angle and bond lengths of the water molecules are dynamic in the simulation and vary from the initial defined structure (104.5°, O-H length 0.96 Å). The distribution of the bond angles has been analysed using the Triangles auxiliary routine EPSR Appendix E.1.3). The analysis routine calculates the angle between three atom types (H_t-O_t-H_t) when they are within a specified distance (0.8–1.2 Å, based on the r_{min} and r_{max} of the intramolecular g(r)). However, there is no way to define whether or not the atom types should be bonded or non-bonded.

The distributions of the internal bond angles are quite similar when comparing the mean values, ranges and percentage of angles ±5° (99.5-109.5°) and ±25° (79.5-129.5°), to compare the distributions (Table 5.19). The mode values indicates that the water bond angles in the 2 M solution tend to be lower than in the higher concentrations.

Table 5.19 Water internal bond angle analysis (EPSR)

Sample conc. /M	Mean	Mode	% 99.5-109.5°	% 79.5-129.5°	Min	Max
2	108.9	98.5	0.5%	4.0%	51.5	178.5
4	109.3	103.5	0.5%	4.1%	54.5	177.5
6	109.5	104.5	0.4%	3.8%	51.5	177.5

The angle ranges of the three distributions are much greater than would be expected (Figure 5.21).

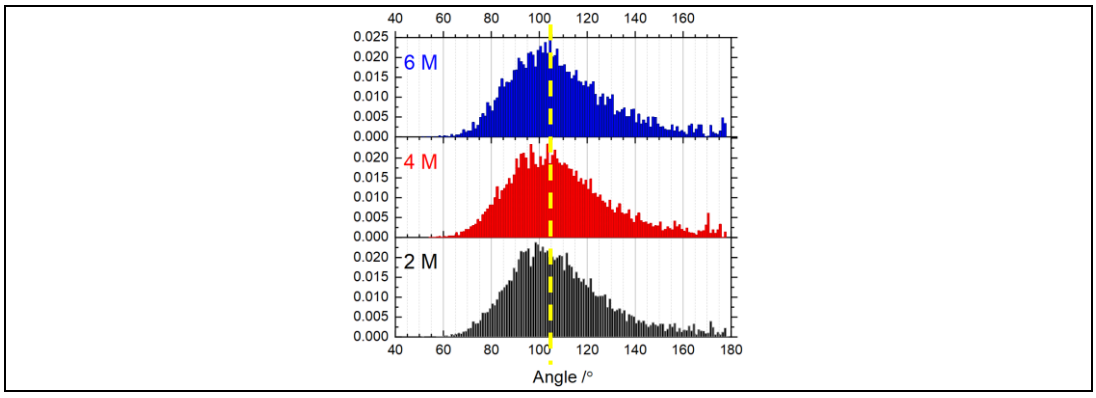


Figure 5.21 Angular distribution of water molecule internal bond angles

The yellow line represents the H_t-O_t-H_t bond angle of the starting geometry. Based on data from original simulations only.

As it was not possible to constrain the Triangles analysis to atoms within the same molecule, some of the data captured in the distribution—especially those with a very high or low

angle—may have been the result of more than one water molecule being close enough to be included in the analysis, i.e. the H_t from one molecule may be close enough to the O_t-H_t of another water molecule to be captured in the analysis. Alternatively, it could suggest that there is too much flexibility in the water molecule in the simulation, and that the geometries of the water molecules in the simulation are physically incorrect.

A second analysis of the geometry of the water molecules was undertaken with dlputils, as it is possible to be more specific about the atoms to be included in the study. The water molecule bond length distributions (Table 5.20) are similar in each concentration, as are the bond angles distributions (Table 5.21). The mean value is 2–3° smaller, as are the distributions ranges, compared with the EPSR Triangles analysis (Figure 5.21).

Table 5.20 Geometry of water molecules: bond lengths (dlputils)

Sample conc. /M	Bond length /Å											
	$H_t1\cdots O_t$				$O_t\cdots H_t2$				Distance between H_t			
	Min	Max	Ave	SD	Min	Max	Ave	SD	Min	Max	Ave	SD
2	0.57	1.33	0.97	*0.09*	0.62	1.27	0.97	*0.09*	0.96	2.04	1.53	*0.15*
4	0.65	1.33	0.97	*0.09*	0.65	1.29	0.97	*0.09*	0.91	2.04	1.52	*0.15*
6	0.64	1.30	0.97	*0.09*	0.65	1.32	0.97	*0.09*	0.96	1.98	1.52	*0.15*

Table 5.21 Geometry of water molecules: bond angles (dlputils)

Sample conc. /M	Angle /°			
	Min	Max	Mean	SD
2	52.40	176.64	106.36	*18.09*
4	54.53	176.06	106.13	*18.95*
6	54.52	177.74	106.33	*18.57*

Water-water hydrogen-bonding angles: an analysis of the angles of water-water hydrogen bonding ($H_t\cdots O_t$-H_t) was undertaken using *intertorsion* in dlputils to compare the structure of the water molecules (Table 5.22, up to 2.5 Å (left) and up to 2.8 Å (right) apart).

For both r_{max} analyses, the proportion of bonds in a perfect tetrahedral arrangement, 109.5° ±5° is similar, with slightly more in 2 M than 6 M. This data suggests that none of the solution water distributions is tetrahedral. The peak at around 30° in the plot of hydrogen bond angles for r_{max} = 2.8 Å would appear to be capturing H_t of another water molecule hydrogen-bonded to the second water molecule (as shown in Figure 5.22).

Table 5.22 Bond angles between water molecules

Sample conc. /M	r_{max} = 2.5 Å			r_{max} = 2.8 Å		
	Mean	Mode	% 104.5-114.5°	Mean	Mode	% 104.5-114.5°
2	111.4	106.6	15.7%	105.0	105.4	13.6%
4	111.4	108.8	15.5%	104.9	106.1	13.4%
6	112.0	108.9	15.5%	105.3	109.0	13.3%

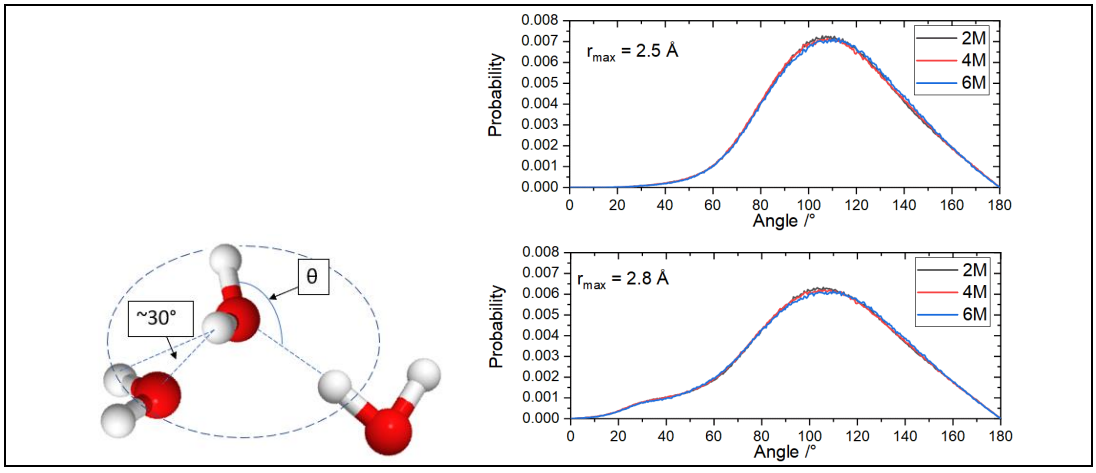


Figure 5.22 Angular distribution of water-water hydrogen bond angles (top r_{max} = 2.5 Å, bottom r_{max} = 2.8 Å), with possible cause of the peak at 30°

The combination of XPDF with computational modelling has the potential to be a helpful tool in understanding more about how ions, and in particular organic salts, in solution influence the solvent structure. However, further work is required to ensure that the geometries and flexibilities of the molecules in the model are physically real. The water internal bond angle analysis suggests that the water molecules could be too flexible, which could impact the modelled fit of g(r) in the low-r region (Figure 5.2). Although an imperfect fit may be expected at low-r (< 1 Å) as they are considered to be non-physical, analysis of the water internal angles suggests that the water molecules may be too flexible, which could affect the low-r g(r) fit. However, the g(r) for O_t-O_t, O_t-H_t and H_t-H_t (Figure 5.18) appear to be reasonable when compared with those in the literature for pure water and water in salt solutions.

5.4 Conclusions

X-ray total scattering data were collected for 2 M, 4 M and 6 M aqueous GuHCl solutions, and XPDF patterns were generated by Fourier transform of the reduced structure factors. The decreasing contribution to scattering of the water with increasing concentration was evident in the G(r), where the second O-O interaction peak was much less intense in the 6 M solution (Figure 5.18). In addition to the intramolecular non-bonded interactions becoming clearer from 2 M to 6 M, peaks that could relate to Gdm^+-Cl^- ion pairing (specifically N-Cl^- interactions) are identified.

The EPSR modelling provided reasonable F(Q) and g(r) fits of the experimental data, although the fit in the low-r region was not good. Although features below 1 Å are typically considered to be non-physical, and therefore an imperfect fit may be expected, the analysis

of the water internal angles suggests that the water molecules may be too flexible, which could affect the low-r g(r) fit. However, the g(r) for O_t-O_t, O_t-H_t and H_t-H_t appear to be reasonable when compared with those in the literature for pure water and water in salt solutions.

Despite the adjustments made in the EPSR parameterisation study, a feature between 4 and 5 $Å^{-1}$ (F(Q)) was slightly offset in all three concentrations, and another at ~4.5 Å was not picked up in the simulated g(r). For future projects, especially where the high-throughput sample holder is used to maximise sample measurement at the beamline, which results in high quantities of data being generated, a machine learning approach to force field parameterisation could be employed to optimise force field selection.

EPSR analysis provides further evidence that cation-anion ion pairing is present in the three concentrations analysed, with C-Cl^- CoordN for probable interactions in the plane of Gdm^+ ranging from 0.13 in 2 M to 0.34 in 6 M. There is consistent atomic separation, indicating single or double NH···Cl^- hydrogen bonds, as proposed by a previous NDIS analysis [97]. It is demonstrated that contact Gdm^+-Gdm^+ ion pair association is indicated in the three concentrations, with interaction distances much shorter (~4 Å) than the average distribution of the ions in solution (Table 5.3).

Based on this data and data from the literature [179,188], it is proposed that the first $g_{C\text{-}C}(r)$ peak position could increase with concentration, i.e. the like-charged interactions are closer at lower concentrations, although more rigorous sampling of the concentration space is needed to provide evidence for this hypothesis.

The main atomic pair correlations of Gdm^+-Gdm^+ (($g_{C\text{-}C}(r)$, $g_{C\text{-}N}(r)$ and $g_{N\text{-}N}(r)$) change with increasing concentration, unlike those found in some computational studies in the literature [182,481]. Some features in the 2 M concentration solution structure $g_{\alpha\beta}(r)$ appear to be significantly different to the higher concentration solutions, which could indicate that counterintuitively, there are interesting structural features due to ionic interactions in the low concentration GuHCl solution that are not present at higher concentrations. However, due to the relatively low number of ions in the simulation, the ion-ion interaction results from the 2 M simulation were noisier than those of the 4 and 6 M simulations, and could therefore have benefited from increased data accumulation.

Alternatively, the differences between the 2 m and higher concentration simulations could relate to data processing errors (from background subtraction or Fourier transform processes), or the geometry of the 2 mm capillary in the high-throughput sample holder.

Remeasurement of the solutions would be proposed, using higher-energy beamlines for higher-resolution data, and using either 1.5 mm capillaries or the flow cell with the 6 mm Kapton tube for more robust alignment and to minimise the potential for gravitational solution separation in the vertical static sample holder.

The analysis of the water structure revealed a breakdown in the characteristic tetrahedral structure, even at the lowest concentration. There is interest in the effect of ions on water structure, and it has been demonstrated here that the combined XPDF/EPSR technique could be used to examine solvent-solvent interactions, including the possible electrostrictive effect of ions on water structure for lower solution concentrations.

There were, however, problems with the water internal angles analysis, which ranged from ~50° to almost 180°. These extreme angles could have been as a result of the Triangle analysis including atoms from more than one molecule, as the proximity of atom types, rather than bonding of atoms, is specified in the auxiliary routine. Alternatively, the range of angles could be as a result of an incorrect characterisation of the water molecule. Further investigation is required to ensure that the molecule parameterisation has not resulted in overly flexible molecules with non-physical behaviour.

To fully understand the interatomic interactions, peak fitting of $g_{\alpha\beta}(r)$ using software such as Fityk, could be applied. However, allocating the peak to specific interactions in the local structure can be difficult, due to the complexity of the possible molecular arrangements in the solution.

Repeated EPSR simulations generated models with similar goodness of fit for the F(Q) and g(r), but with different $g_{\alpha\beta}(r)$ (Appendix G). Averages of the simulation outputs were used in this analysis. It was noted that although increasing the number of iterations that the model was run for reduces g(r) noise, the g(r) would tend towards same solution for that molecular model after relatively few iterations. However, further analysis of longer simulations at each stage, the use of larger simulation boxes to include more combinations of possible local structure arrangements, as well as repeated simulations to assess the average structures, should be explored. Dissolve [407], a software in development for the analysis of neutron and X-ray total scattering data, will incorporate a multiple-box approach, as it is acknowledged that it will be more representative of heterogeneous sample structures. It is proposed that more robust quantitative comparisons of the solution structures could be made with additional results. Access to computational resources were limited during the data analysis section of this work due to working from home as a result of Covid restrictions, but opportunities to improve the quality of the EPSR simulations have been identified.

This combined XPDF/EPSR technique has great potential to reveal the local intermolecular interactions of organic salts in solution, and this method development process has identified some areas that will need addressing in future studies to ensure robust data analysis.

Chapter 6 Concluding remarks and further work

6.1 Concluding remarks

The aim of this project was to develop experimental X-ray and computational calculation and modelling techniques to characterise the phase from which organic salts crystallise. This work has demonstrated the feasibility of core-level X-ray spectroscopic techniques, combined with computational analysis, to probe the electronic structure of organic salts in solution.

X-ray Raman scattering (XRS), combined with near-ambient pressure X-ray photoelectron spectroscopy (NAP-XPS) data and time-dependent density functional theory (TDDFT) calculations, generates quantitative descriptions of the molecular orbitals of Imid and its protonated analogue, $ImidH^+$, in aqueous solution by way of C and N K-edge fine-structure spectra. The pseudoequivalence of the $ImidH^+$ N moieties' single $1s\rightarrow1\pi^*$ electronic transition peak contrasts with the double $1s\rightarrow1\pi^*$ peaks observed in Imid. TDDFT calculations provide a qualitative indication of the core-level electronic transitions in the probed atoms up to the IP energy, revealing $1s\rightarrow3\pi^*$ resonances in the C K-edge spectrum of $ImidH^+$ that are not present in the Imid spectrum.

N and C 1s binding energy (E_B) of Imid and $ImidH^+$, acquired with NAP-XPS, were used to assign relative IP energies to the aqueous Imid and $ImidH^+$ XRS spectra. These E_B data had previously been identified using a synchrotron-based liquid-jet XPS [160]. In this work, the measurements were undertaken at 11 mbar, with the solute remaining in solution in the humid sample chamber, using water vapour to maintain the required pressure. NAP-XPS has the potential to be a solution characterisation technique that is more accessible than synchrotron-based XPS measurements, but equally as effective.

The potential of the combined X-ray Pair Distribution Function (XPDF)/Empirical Potential Structure Refinement (EPSR) modelling approach to provide insight into the local molecular structure of a crystallising solution has been explored. The XPDF analysis method has previously been used to understand the locally-disordered structures of crystals with long-range order, as well as the local structures of liquids and solutions. However, limitations on the time- and length-scales that could be accessed by available experimental techniques meant it has not been possible until now to probe intermolecular interactions at the

required resolution to determine the solution structural transformations during the cooling crystallisation of a pharmaceutical salt.

Local ordering in the solutions is present to ~10 Å throughout, with long-range order only appearing in the Transition frame (~20°C). Principal component analysis of the temperature-resolved XPDF patterns identified three temperature-related pre-crystallisation structural regions—one undersaturated solution stage, and two structurally-distinct supersaturated stages. Analysis of the X-ray total scattering data from the first cooling crystallisation run, processed using an empty Kapton background, showed that the transition from stage 1 (undersaturated solution) to stage 2 (supersaturated solution) occurred at ~51°C. Stage 3, a further phase with a changed molecular structure, emerged at ~36°C. Processing of the X-ray total scattering data from the second cooling crystallisation experiment is needed to confirm the transition temperatures between the stages[49].

EPSR analysis, where the molecular models are refined to fit the experimental structure factors collected during the *in situ* cooling crystallisation of GuHCl, has been used to drill down into the molecular interactions that could be present in the solution. The EPSR analysis serves as proof of concept for the examination of how the molecular interactions in a crystallising solution evolve, as it is cooled through the MSZ from an unsaturated solution and crystallisation, and demonstrates the potential power of this combined experimental and computational technique.

The ion-ion and ion-water atomic pair distribution functions $g_{\alpha\beta}$® derived from the molecular modelling suggest that the supersaturated solutions could be dominated by stacked dimer structures. Possible structures of molecular clusters in each stage of the cooling crystallisation have been proposed (Figure 4.28). The dimers in the stage 2 solution are stabilised with more water molecules (and/Cl^- Cl^-) than in stage 3. The water clustering analysis support the hypothesis that there is a partial desolvation event between stages 2 and 3.

[49] Runs 1 and 2 were both processed with a water in Kapton background. Analysis of both runs identified three stages with different molecular structures in the MSZ, although at slightly different temperatures to that seen with the empty Kapton background: stage1 → stage 2 at 55.5°C, and stage 2 → stage 3 at 32°C.

This combined XPDF/EPSR method was also used to compare aqueous GuHCl solutions of different concentrations, providing the opportunity to explore the effect of solution concentration on solution structure, specifically the clustering of Gdm^+ ions. Although like-charged ion pairing has been predicted by several computational modelling techniques in the literature, the approach presented here is the first example of the molecular model being refined to the experimental data through the perturbation of the reference potential. Continuing the trend of modelling data from the literature, the like-charged Gdm^+-Gdm^+ interactions are identified to be closer at lower concentrations, although more rigorous sampling of the concentration space is needed to provide further evidence for this hypothesis.

Cation-anion ion pairing is present in all concentrations analysed, and throughout the MSZ of the crystallising solution, with Cl^- C-Cl^- CoordN for probable interactions in the plane of Gdm^+ ranging from 0.13 in 2 M to 0.34 in 6 M. Inspection of the models suggest that in some cases, more than one Cl^- will be associated with one Gdm^+.

The limitations of the EPSR simulations are acknowledged. The molecular structures are highly dependent on the Lennard-Jones parameters and partial charges applied. The parameterisation studies (Appendix F) were designed to optimise parameter selection. EPSR is designed to model systems in equilibrium, but due to the nature of the *in situ* cooling crystallisation experiment, the temperature of the solution reduced during data collection by 0.5–0.7°C. However, with the new generation of higher-brilliance synchrotrons, faster data acquisition with higher counting statistics may in part overcome this issue.

There were limitations to laboratory access, which could have improved the understanding of the GuHCl MSZ and measurement of the solution density, and access to computational resources, which limited the simulation box sizes and number of iterations that were possible during the data accumulation stage of the simulations, as a result of the campus being closed during the Covid lockdown.

Core-level near-edge spectroscopy of the N and C K-edges clearly differentiates between the neutral and protonated species of a biological molecule in solution. A method for using XPDF pattern analysis to reveal distinct structural pre-nucleation phases in the crystallising solution through the MSZ is shown. Statistical analysis of the structural models refined to *in situ* X-ray total scattering data has been performed for the first time and provides an indication of the possible structural motifs that may give rise to the identified phases. These *in situ* and concentration study analyses indicate that both cation-anion and like-charged

Gdm^+-Gdm^+ ion pairs are present in solution. The techniques developed here could be extended in further studies.

6.2 Further work

The work described in this thesis demonstrates the potential for X-ray techniques to provide insights into the electronic and molecular structures of the phase from which organic salts can be crystallised. Further technique development, and carrying out further experiments, could improve our understanding of the systems studied in this thesis.

X-ray Raman scattering of a crystallising organic salt

This initial exploration of the molecular orbital structure of a salt in solution using XRS could be continued to study the electronic structural changes of an organic salt during cooling crystallisation, allowing changes in local bonding and short-range structure to be evaluated, as has been applied to the crystallisation of a neutral molecule (Imid [67]). The higher-energy, fourth generation X-ray sources now available, e.g. at ESRF, will reduce the data acquisition time required for each scan. Beam damage that could be caused by the greater brilliance of the beam is mitigated by the recirculation of the crystallising solution.

The TDDFT analysis presented in Chapter 3 indicates that the extent of equivalence of the $ImidH^+$ N K-edge pre-edge $1s \rightarrow 1\pi^*$ peaks could be indicative of the degree of cation-anion pairing in solution. Initial exploratory work to crystallise ImidHCl from aqueous solution in the laboratory was not successful, so an alternative organic salt, e.g. GuHCl, which has degenerate N moieties in the protonated Gdm^+, could be investigated to complement the XPDF/EPSR studies in this thesis. Linear combination analysis of the near-edge fine structure of the solid-phase and solution-phase GuHCl may be a feasible approach for estimating the yield of crystallised salt, which is required for the accurate molecular modelling of the crystallising solution.

Synthesis of ImidHCl

Related, but less significant to the main theme of the study, is the investigation into the solid-phase structure of ImidHCl, which could benefit from further study. A technique was developed for the synthesis of ImidHCl, the structure of which is not currently deposited in the CCDC. SCXRD revealed that the rapid evaporation technique resulted in only seven of the eight Imid in the unit cell of ImidHCl being protonated. However, all eight Imid in the unit cell of ImidHCl were protonated when water was evaporated under vacuum from the ImidHCl solution, although two of Imid rings showed disorder. In discussion with Dr. Chris

Pask[50] and Dr. Anu Pallipurath, the possibilities of ImidHCl having an incommensurate structure (indicated by the high symmetry unit cell [498]) or twinning were identified. The characterisation of imidazolium bromide by SCXRD (amongst other techniques) indicated highly-disordered $ImidH^+$ ions in the structure and a first-order phase transition at ~200 K [499]. Further investigation of the structure of ImidHCl would provide greater understanding of the system.

Combined X-ray total scattering and molecular modelling technique

There are opportunities to build on the combined experimental and modelling approach detailed in Chapters 4 and 5.

XPDF pattern analysis of run 2 of the cooling crystallisation, and the dissolution of GuHCl: The background subtraction method described in Section 4.3.2.1.2 for the analysis of the cooling crystallisation run 1 should also be applied to data collected in the second cooling crystallisation experiment, as well as to the data collected whilst the solution was reheated to dissolve the GuHCl that had been crystallised during the first experiment. PCA and K-means cluster analysis should be carried out on the data to check that the same pre-crystallisation stages are found in both sets of crystallisation data.

Improvements to the experimental and modelling techniques applied in this study: Sample mounting in the X-ray beam could be modified for the *in situ* and capillary data collection. The Kapton tube was mounted horizontally in the beam (Figure 2.18c), and on occasion, bubbles formed at the top of the tube, which could affect the data collection. It is proposed that mounting the Kapton tube vertically could reduce the possibility of bubbles forming in the tube, improving the quality of the data collection.

Better understanding of the MSZ of the crystallising GuHCl system in the flow cell apparatus used in the XPDF data collection would be beneficial[51]. In addition, the development of an FTIR flow cell could be used to analyse the hydration of Gdm^+ through the MSZ.

In the case of the concentration study, repeating the data collection with solutions mounted in the horizontal, spinning capillary holder in 1.5 mm capillaries would provide some quality assurance of the data collection. Although the comparison of X-ray total scattering from the

[50] Department of Chemistry, University of Leeds

[51] MSZ characterisation work had been planned as a Master's project to complement the work presented here, but could not be completed due to the lack of access to the laboratory as a result of Covid restrictions.

high-throughput and spinning capillary holders for water data (Section 2.5.4.1.4) indicates that the main features in the F(Q) and g(r) are consistent, this process should be repeated for the GuHCl solutions. This suggestion is put forward for two reasons. Firstly, the beam alignment was not carried out for all positions in the high-throughput capillary sample holder, meaning that misalignment, which could introduce errors into the data processing of the scattering data, was possible. Secondly, density gradients in the solution would be exacerbated in the static, vertically-orientated 2 mm capillary should the solute settle in the capillary. The acquisition of further data points at a range of concentrations up to 6 M (room temperature solubility limit) is required to investigate the atomic separation of Gdm^{+}-Gdm^{+} ion pairs in solution. It would also be interesting to extend analysis to other GuHCl-solvent combinations.

The accuracy of the XPDF/EPSR technique could be improved if purification of GuHCl is undertaken, as the data post-processing and modelling require the atomic composition of the solution. An alternative method to determine the solution density for future system analysis could improve the modelling accuracy. A high-temperature, high-pressure vibrating-tube densimeter technique was developed to measure the densities of dilute aqueous solutions of organic substances [500]. This technique has been applied to aqueous GuHCl in the literature [168], although the solutions tested were up to a concentration of 6 $mol.kg^{-1}$, which is lower than the concentration of solution used here (~7.3 $mol.kg^{-1}$). A densimeter could be available for use at ISIS Neutron and Muon Source at the Rutherford Appleton Laboratory of the Science and Technology Facilities Council, on the Harwell Science and Innovation Campus, where DLS is located, and this option should be explored for future work. Crystal16 analysis and laboratory flow cell experiments undertaken in this study have highlighted the sensitivity of crystallisation to solution flow dynamics and the surfaces contacted by the crystallising solution (Section 2.2.3.3), especially for high concentration solutions. Therefore, the suitability of this technique would need to be tested for the high-concentration solution used in the *in situ* crystallisation experiment.

As an alternative, the experimental flow cell could be modified to include a graduated glass capillary positioned vertically out of the vessel. The changing height (thus volume) of solution introduced into the capillary could be used to calculate the solution density change, whilst maintaining the crystalliser dimensions and fluid dynamic characteristics during the cooling crystallisation.

The analysis of an aqueous GuHCl solution using NDIS and EPSR modelling, could also inform the selection of Lennard-Jones parameters and partial charges to be applied to the analysis of the XPDF analysis.

Further analysis of the EPSR simulations could yield further information about the modelled solutions. It has previously been demonstrated that the combined X-ray total scattering/EPSR modelling technique used in this thesis provides insight into non-ideal binary solutions [242], so the X-ray techniques used in this thesis may be extended to explore the thermodynamic behaviour and properties of non-ideal ionic solutions.

Exploiting advances in the XPDF technique: Even more insightful characterisation of the structure of an organic salt in solution should become achievable in the future with fourth generation synchrotron sources and X-ray free electron laser (XFEL) sources. Increased X-ray energy and beam brilliance will improve spatial and temporal resolution.

One limitation of the PDF method is that it only generates an average representation of the sample being studied [452]. Spatial mapping of the structures forming in solution could be achieved by PDF with computational tomography (PDF-CT), as an X-ray PDF can be extracted from each voxel [452]. However, this method may be more appropriate for tracking the changing XPDF for the growth of crystals on a surface, rather than during crystal nucleation in a flowing solution.

Advances in molecular modelling: The EPSR single-box modelling method has been successfully applied to the analysis of liquids and solutions [223,228,235,269,501,502]. However, EPSR was not developed to handle heterogenous systems [407], as would be present in the transition and crystallisation stages of a crystallisation experiment. The length-scales of the assemblies in a crystallising solution cannot be represented in the EPSR models presented here, where the simulation box size was limited by both EPSR functionality and access to sufficient computational power[52]. Furthermore, EPSR does not handle the Ewald calculations required for the simulation of crystallites in solution.

The Dissolve software [407], which builds on the EPSR approach, is currently in the beta phase of testing and may provide a multiple-box approach to the modelling of crystallising systems. This approach could allow sample heterogeneity to be investigated and may overcome the challenge of representative modelling with a relatively small number of

[52] Computational power limitations were exacerbated by no or reduced access to resources on campus during the Covid lockdown during the data analysis phase of the PhD work plan.

components through the linear combination of models that are characteristic of the different phases in solution [407]. Analysis tools would also be integrated into the same software, rather than applying separate tools, e.g. applying dlputils analysis to EPSR models.

For future projects, a machine learning approach to force field parameterisation could be employed to optimise force field selection and expedite the analysis of X-ray total scattering data. Machine learning would be beneficial when the high-throughput sample holder is used to generate large quantities of data, and could increase the number of frames of data that can be analysed from a temperature-resolved *in situ* crystallisation study, especially as newer synchrotron technology allows the faster acquisition of time-resolved data. This technique will bring us closer to quantifying the evolution of the intermolecular interactions in a crystallising solution and reveal system-specific nucleation mechanisms.

References

[1] Stahl PH, Wermuth CG. Handbook of Pharmaceutical Salts: Properties, Selection, and Use. VHCA; 2008. https://doi.org/10.1515/ci.2002.24.3.20a.

[2] Morissette SL, Almarsson O, Peterson ML, Remenar JF, Read MJ, Lemmo A V, et al. High-throughput crystallization: polymorphs, salts, co-crystals and solvates of pharmaceutical solids. Adv Drug Deliv Rev 2004;56:275–300. https://doi.org/10.1016/j.addr.2003.10.020.

[3] Almarsson Ö, Zaworotko MJ. Crystal engineering of the composition of pharmaceutical phases. Do pharmaceutical co-crystals represent a new path to improved medicines? Chem Commun 2004:1889–96. https://doi.org/10.1039/b402150a.

[4] Cerreia Vioglio P, Chierotti MR, Gobetto R. Pharmaceutical aspects of salt and cocrystal forms of APIs and characterization challenges. vol. 117. Elsevier B.V.; 2017. https://doi.org/10.1016/j.addr.2017.07.001.

[5] Lipinski CA, Dominy BW, Feeney PJ, Lombardo F. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23:3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.

[6] Shan N, Zaworotko MJ. The role of cocrystals in pharmaceutical science. Drug Discov Today 2008;13:440–6. https://doi.org/10.1016/j.drudis.2008.03.004.

[7] Aitipamula S, Banerjee R, Bansal AKK, Biradha K, Cheney MLL, Roy Choudhury A, et al. Polymorphs, salts, and cocrystals: What's in a name? Cryst Growth Des 2012;12:2147–52. https://doi.org/10.1021/cg3002948.

[8] Wöhler F. Untersuchungen über das Chinon. Justus Liebigs Ann Chem 1844;51:145–63. https://doi.org/10.1002/jlac.18440510202.

[9] Desiraju GR. Pharmaceutical Salts and Co-crystals: Retrospect and Prospects. In: Wouters J, Quéré L, editors. Pharm. Salts Co-crystals, Royal Society of Chemistry; 2012. https://doi.org/10.1039/9781849733502-00001.

[10] Davey RJ, Allen K, Blagden N, Cross WI, Lieberman HF, Quayle MJ, et al. Crystal engineering - nucleation, the key step. CrystEngComm 2002;4:257–64. https://doi.org/10.1039/b203521a.

[11] Collier EA, Davey RJ, Roberts RJ, Black SN. A crystallisation / crystal engineering approach to aid salt selection - anions. Acta Cryst 2002;A58 (Suppl:159.

[12] Raijada D, Müllertz A, Cornett C, Munk T, Sonnergaard J, Rantanen J. Miniaturized approach for excipient selection during the development of oral solid dosage form. J Pharm Sci 2014;103:900–8. https://doi.org/10.1002/jps.23840.

[13] Savjani KT, Gajjar AK, Savjani JK. Drug solubility: importance and enhancement techniques. ISRN Pharm 2012;2012:195727. https://doi.org/10.5402/2012/195727.

[14] Biswas N. Solid Forms and Pharmacokinetics. In: Wouters J, Quéré L, editors. Pharm. Salts Co-crystals, Royal Society of Chemistry; 2012. https://doi.org/10.1039/9781849733502-00128.

[15] Takagi T, Ramachandran C, Bermejo M, Yamashita S, Yu LX, Amidon GL. A provisional biopharmaceutical classification of the top 200 oral drug products in the United States, Great Britain, Spain, and Japan. Mol Pharm 2006;3:631–43. https://doi.org/10.1021/mp0600182.

[16] Fahr A, Liu X. Drug delivery strategies for poorly water-soluble drugs. Expert Opin Drug Deliv 2007;4:403–16. https://doi.org/10.1517/17425247.4.4.403.

[17] Chen J, Sarma B, Evans JMB, Myerson AS. Pharmaceutical Crystallization. Cryst Growth Des 2011;11:887–95. https://doi.org/10.1021/cg101556s.

[18] Black SN, Collier EA, Davey RJ, Roberts RJ. Structure, solubility, screening, and synthesis of molecular salts. J Pharm Sci 2007;96:1053–68. https://doi.org/10.1002/jps.20927.

[19] Bastin RJ, Bowker MJ, Slater BJ. Salt selection and optimisation procedures for pharmaceutical new chemical entities. Org Process Res Dev 2000;4:427–35. https://doi.org/10.1021/op000018u.

[20] Sigfridsson K, Ahlqvist M, Lindsjö M, Paulsson S. Salt formation improved the properties of a candidate drug during early formulation development. Eur J Pharm Sci 2018;120:162–71. https://doi.org/10.1016/J.EJPS.2018.04.048.

[21] McNaught AD, Wilkinson A. IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Online 201. Oxford: Blackwell Scientific Publications; 1997.

[22] Blagden N, Coles SJ, Berry DJ. Pharmaceutical co-crystals – are we there yet? CrystEngComm 2014;16:5753–61. https://doi.org/10.1039/c4ce00127c.

[23] Grothe E, Meekes H, Vlieg E, ter Horst JH, de Gelder R. Solvates, Salts, and Cocrystals: A Proposal for a Feasible Classification System n.d. https://doi.org/10.1021/acs.cgd.6b00200.

[24] Childs SL, Stahly GP, Park A. The Salt-Cocrystal Continuum: The Influence of Crystal Structure on Ionization State. Mol Pharm 2007;4:323–38. https://doi.org/10.1021/mp0601345.

[25] Stevens JS, Coultas S, Jaye C, Fischer DA, Schroeder SLM. Core level spectroscopies locate hydrogen in the proton transfer pathway-identifying quasi-symmetrical hydrogen bonds in the solid state. Phys Chem Chem Phys 2020;22:4916–23. https://doi.org/10.1039/c9cp05677g.

[26] Cooke CL, Davey RJ. On the Solubility of Saccharinate Salts and Cocrystals. Cryst Growth Des 2008;8:3483–5. https://doi.org/10.1021/CG800621Q.

[27] Stilinović V, Kaitner B. Salts and Co-crystals of gentisic acid with pyridine derivatives: The effect of proton transfer on the crystal packing (and vice versa). Cryst Growth Des 2012;12:5763–72. https://doi.org/10.1021/cg301267h.

[28] Guthrie JP. Short strong hydrogen bonds: Can they explain enzymic catalysis? Chem Biol 1996;3:163–70. https://doi.org/10.1016/s1074-5521(96)90258-6.

[29] Acharya PC, Marwein S, Mishra B, Ghosh R, Vora A, Tekade RK. Role of Salt Selection in Drug Discovery and Development. Dos. Form Des. Considerations Vol. I, Elsevier; 2018, p. 435–72. https://doi.org/10.1016/B978-0-12-814423-7.00013-7.

[30] Gao Z, Rohani S, Gong J, Wang J. Recent Developments in the Crystallization Process: Toward the Pharmaceutical Industry. Engineering 2017;3:343–53. https://doi.org/10.1016/J.ENG.2017.03.022.

[31] Mersmann A. Crystallization Technology Handbook. CRC Press; 2001.

[32] Crystal16 - Crystallization Systems n.d. https://www.crystallizationsystems.com/crystal16 (accessed September 6, 2021).

[33] Price SL. Predicting crystal structures of organic compounds. Chem Soc Rev 2014;43:2098. https://doi.org/10.1039/c3cs60279f.

[34] Harris KDM. Explorations in the Dynamics of Crystalline Solids and the Evolution of Crystal Formation Processes. Isr J Chem 2017;57:154–70. https://doi.org/10.1002/ijch.201600088.

[35] Jones AG. Crystallization Process Systems. Elsevier Science; 2015.

[36] Hulliger J. Chemistry and Crystal Growth. Angew Chemie Int Ed English 1994;33:143–62. https://doi.org/10.1002/anie.199401431.

[37] Crafts P. The role of solubility modeling and crystallization in the design of active pharmaceutical ingredients. Comput. Aided Chem. Eng., vol. 23, Elsevier; 2007, p. 23–85. https://doi.org/10.1016/S1570-7946(07)80005-8.

[38] Fitzner M, Sosso GC, Pietrucci F, Pipolo S, Michaelides A. Pre-critical fluctuations and what they disclose about heterogeneous crystal nucleation. Nat Commun 2017;8:1–7. https://doi.org/10.1038/s41467-017-02300-x.

[39] Quayle MJ, Davey RJ, McDermott AJ, Tiddy GJT, Clarke DT, Jones GR. In situ monitoring of

rapid crystallisation processes using synchrotron X-ray diffraction and a stopped-flow cell. Phys Chem Chem Phys 2002;4:416–8. https://doi.org/10.1039/B108773H.

[40] Mullin JW. Crystallization. Elsevier Science; 2001.

[41] Smeets PJM, Finney AR, Habraken WJEM, Nudelman F, Friedrich H, Laven J, et al. A classical view on nonclassical nucleation. Proc Natl Acad Sci United States Am Am 2017;114:E7882–90. https://doi.org/10.1073/pnas.1700342114.

[42] Davey R, Garside J. From Molecules to Crystallizers. Oxford University Press; 2000.

[43] Davey RJ, Schroeder SLM, Ter Horst JH. Nucleation of organic crystals - A molecular perspective. Angew Chemie - Int Ed 2013;52:2166–79. https://doi.org/10.1002/anie.201204824.

[44] Dok AR, Legat T, de Coene Y, van der Veen MA, Verbiest T, Van Cleuvenbergen S. Nonlinear optical probes of nucleation and crystal growth: recent progress and future prospects. J Mater Chem C 2021;9:11553–68. https://doi.org/10.1039/d1tc02007b.

[45] Erdemir D, Lee AY, Myerson AS. Nucleation of Crystals from Solution: Classical and Two-Step Models. Acc Chem Res 2009;42:621–9. https://doi.org/10.1021/ar800217x.

[46] Pienack N, Bensch W. In-situ monitoring of the formation of crystalline solids. Angew Chemie - Int Ed 2011;50:2014–34. https://doi.org/10.1002/anie.201001180.

[47] Vekilov PG. The two-step mechanism of nucleation of crystals in solution. Nanoscale 2010;2:2346–57. https://doi.org/10.1039/c0nr00628a.

[48] Karthika S, Radhakrishnan TK, Kalaichelvi P. A Review of Classical and Nonclassical Nucleation Theories. Cryst Growth Des 2016;16:6663–81. https://doi.org/10.1021/acs.cgd.6b00794.

[49] Gebauer D, Raiteri P, Gale JD, Cölfen H. On classical and non-classical views on nucleation. Am J Sci 2018;318:969–88. https://doi.org/10.2475/09.2018.05.

[50] Van Vleet MJ, Weng T, Li X, Schmidt JR. In Situ, Time-Resolved, and Mechanistic Studies of Metal-Organic Framework Nucleation and Growth. Chem Rev 2018;118:3681–721. https://doi.org/10.1021/acs.chemrev.7b00582.

[51] Garside J, of Chemical Engineers (Great Britain) I. Separation Technology: The Next Ten Years. Institution of Chemical Engineers; 1994.

[52] Vekilov PG. Nucleation. Cryst Growth Des 2010;10:5007–19. https://doi.org/10.1021/cg1011633.

[53] Mullin JW, Söhnel O. Expressions of supersaturation in crystallization studies. Chem Eng Sci 1977;32:683–6. https://doi.org/10.1016/0009-2509(77)80114-0.

[54] Coquerel G. Crystallization of molecular systems from solution: phase diagrams, supersaturation and other basic concepts. Chem Soc Rev 2014;43:2286–300. https://doi.org/10.1039/C3CS60359H.

[55] Bhamidi V, Kenis PJA, Zukoski CF. Probability of Nucleation in a Metastable Zone: Induction Supersaturation and Implications. Cryst Growth Des 2017;17:1132–45. https://doi.org/10.1021/acs.cgd.6b01529.

[56] Vetter T, Iggland M, Ochsenbein DR, Hä FS, Mazzotti M, Hänseler FS, et al. Modeling Nucleation, Growth, and Ostwald Ripening in Crystallization Processes: A Comparison between Population Balance and Kinetic Rate Equation. Cryst Growth Des 2013;13:4890–905. https://doi.org/10.1021/cg4010714.

[57] Morris L, Simone E, Glover ZJ, Powell H, Marty-Terrade S, Francis M, et al. Dynamic monitoring of glycine crystallisation with low power ultrasound reflection spectroscopy. Chem Eng Res Des 2021;170:213–23. https://doi.org/10.1016/J.CHERD.2021.04.003.

[58] Brown CJ, Ni XW. Determination of metastable zone width, mean particle size and detectable number density using video imaging in an oscillatory baffled crystallizer. CrystEngComm 2012;14:2944–9. https://doi.org/10.1039/c2ce06628a.

[59] Camacho Corzo DM, Borissova A, Hammond RB, Kashchiev D, Roberts KJ, Lewtas K, et al. Nucleation mechanism and kinetics from the analysis of polythermal crystallisation data: methyl stearate from kerosene solutions 2014;16:941–1218. https://doi.org/10.1039/c3ce41098f.

[60] Dubrovskii V. Nucleation Theory and Growth of Nanostructures. In: Dubrovskii VG, editor., Berlin, Heidelberg: Springer Berlin Heidelberg; 2014, p. 1–73. https://doi.org/10.1007/978-3-642-39660-1_1.

[61] Kalikmanov VI. Classical Nucleation Theory, Springer, Dordrecht; 2013, p. 17–41. https://doi.org/10.1007/978-90-481-3643-8_3.

[62] Juelsholt M, Lindahl Christiansen T, Jensen KMØ. Mechanisms for Tungsten Oxide Nanoparticle Formation in Solvothermal Synthesis: From Polyoxometalates to Crystalline Materials. J Phys Chem C 2019;123:5110–9. https://doi.org/10.1021/acs.jpcc.8b12395.

[63] Greer HF. Non-classical crystal growth of inorganic and organic materials. Mater Sci Technol 2014;30:611–26. https://doi.org/10.1179/1743284713Y.0000000433.

[64] Gibbs JW. On the Equilibrium of Heterogeneous Substances. Trans Connect Acad Arts Sci 1876;3:108–248.

[65] Gibbs JW. On the Equilibrium of Heterogeneous Substances. Trans Connect Acad Arts Sci 1878;16:343–524.

[66] Zimmermann NER, Vorselaars B, Quigley D, Peters B. Nucleation of NaCl from Aqueous Solution: Critical Sizes, Ion-Attachment Kinetics, and Rates. J Am Chem Soc 2015;137:13352–61. https://doi.org/10.1021/jacs.5b08098.

[67] Al-Madhagi LH, Chang SY, Balasubramanian M, Kroner AB, Shotton EJ, Willneff EA, et al. X-ray Raman scattering: a new in situ probe of molecular structure during nucleation and crystallization from liquid solutions. CrystEngComm 2018;20:6871–84. https://doi.org/10.1039/c8ce00929e.

[68] Kowacz M, Putnis C V., Putnis A. The effect of cation:anion ratio in solution on the mechanism of barite growth at constant supersaturation: Role of the desolvation process on the growth kinetics. Geochim Cosmochim Acta 2007;71:5168–79. https://doi.org/10.1016/J.GCA.2007.09.008.

[69] Gaines E, Maisuria K, Di Tommaso D, Laio A, Bussi G, Ferguson DC, et al. The role of solvent in the self-assembly of m-aminobenzoic acid: a density functional theory and molecular dynamics study. CrystEngComm 2016;18:2937–48. https://doi.org/10.1039/C6CE00130K.

[70] Bøjesen ED, Iversen BB. The chemistry of nucleation. CrystEngComm 2016;18:8332–53. https://doi.org/10.1039/c6ce01489e.

[71] Blow KE, Quigley D, Sosso GC. The seven deadly sins: When computing crystal nucleation rates, the devil is in the details. J Chem Phys 2021;155:040901. https://doi.org/10.1063/5.0055248.

[72] Mirabello G, Ianiro A, Bomans PHH, Yoda T, Arakaki A, Friedrich H, et al. Crystallization by particle attachment is a colloidal assembly process. Nat Mater 2020;19:391–6. https://doi.org/10.1038/s41563-019-0511-4.

[73] Demichelis R, Raiteri P, Gale JD, Quigley D, Gebauer D. Stable prenucleation mineral clusters are liquid-like ionic polymers. Nat Commun 2011;2:1–8. https://doi.org/10.1038/ncomms1604.

[74] Chung S-Y, Kim Y-M, Kim J-G, Kim Y-J. Multiphase transformation and Ostwald's rule of stages during crystallization of a metal phosphate. Nat Phys 2009;5:68–73. https://doi.org/10.1038/NPHYS1148.

[75] ten Wolde PR, Frenkel D. Enhancement of protein crystal nucleation by critical density fluctuations. Science 1997;277:1975–8.

[76] Davey RJ, Back KR, Sullivan RA. Crystal nucleation from solutions – transition states, rate

determining steps and complexity. Faraday Discuss 2015;179:9–26. https://doi.org/10.1039/c5fd00037h.

[77] Chakraborty D, Patey GN. Evidence that crystal nucleation in aqueous NaCl solution Occurs by the two-step mechanism. Chem Phys Lett 2013;587:25–9. https://doi.org/10.1016/J.CPLETT.2013.09.054.

[78] Bruno M. A two-step nucleation model based on diffuse interface theory (DIT) to explain the non-classical view of calcium carbonate polymorph formation. CrystEngComm 2019;21:4918–24. https://doi.org/10.1039/c9ce00610a.

[79] Wang T, Lu H, Wang J, Xiao Y, Zhou Y, Bao Y, et al. Recent progress of continuous crystallization. J Ind Eng Chem 2017;54:14–29.

[80] Su Q, Nagy ZK, Rielly CD. Pharmaceutical crystallisation processes from batch to continuous operation using MSMPR stages: Modelling, design, and control. Chem Eng Process 2015;89:41–53. https://doi.org/10.1016/J.CEP.2015.01.001.

[81] McGinty J, Yazdanpanah N, Price C, Horst JH ter, Sefcik J. Nucleation and Crystal Growth in Continuous Crystallization. Handb. Contin. Cryst., Royal Society of Chemistry; 2020, p. 1–50. https://doi.org/10.1039/9781788013581-00001.

[82] Jones HP. Crystallisation of Polymorphic Organic Salts. University of Manchester, 2006.

[83] Steudel R. Theory of the Chemical Bond. In: Nachod FC, Zuckerman JJ, editors. Chem. non-metals with an Introd. to At. Struct. Chem. Bond., Berlin ; New York: De Gruyter; 1976, p. 59–150.

[84] Szwarc M. Ions and Ion Pairs. Their meaning and Significance in Organic Reactions. Pure Appl Chem 1976;48:247–50. https://doi.org/10.1351/pac197648030247.

[85] Qiu Y, Chen Y, Zhang GGZ, Mantri R V, Yu L. Developing Solid Oral Dosage Forms: Pharmaceutical Theory and Practice. Elsevier Science & Technology Books; 2016.

[86] Klimeš J, Bowler DR, Michaelides A. Understanding the role of ions and water molecules in the NaCl dissolution process. J Chem Phys 2013;139:234702. https://doi.org/10.1063/1.4840675.

[87] Bian H, Wen X, Li J, Chen H, Han S, Sun X, et al. Ion clustering in aqueous solutions probed with vibrational energy transfer. Proc Natl Acad Sci U S A 2011;108:4737–42. https://doi.org/10.1073/pnas.1019565108.

[88] Collins KD, Neilson GW, Enderby JE. Ions in water: Characterizing the forces that control chemical processes and biological structure. Biophys Chem 2007;128:95–104. https://doi.org/10.1016/j.bpc.2007.03.009.

[89] Roux B, Allen T, Bernèche S, Im W. Theoretical and computational models of biological ion channels. Q Rev Biophys 2004;37:15–103. https://doi.org/10.1017/S0033583504003968.

[90] Andreev M, De Pablo JJ, Chremos A, Douglas JF. Influence of Ion Solvation on the Properties of Electrolyte Solutions. J Phys Chem B 2018;122:25. https://doi.org/10.1021/acs.jpcb.8b00518.

[91] Vazdar M, Ri Vym J, Heyda J, Ri Vondr J, Jungwirth P. Like-Charge Guanidinium Pairing from Molecular Dynamics and Ab Initio Calculations. J Phys Chem A 2011;115:11193–201. https://doi.org/10.1021/jp203519p.

[92] Cooper RJ, Heiles S, Ditucci MJ, Williams ER. Hydration of Guanidinium: Second Shell Formation at Small Cluster Size. J Phys Chem A 2014;118:18. https://doi.org/10.1021/jp506429a.

[93] Vazdar M, Heyda J, Mason PE, Tesei G, Allolio C, Lund M, et al. Arginine “Magic”: Guanidinium Like-Charge Ion Pairing from Aqueous Salts to Cell Penetrating Peptides. Acc Chem Res 2018;51:1455–64. https://doi.org/10.1021/acs.accounts.8b00098.

[94] Hofmeister F. Zur Lehre von der Wirkung der Salze - Dritte Mittheilung. Arch Für Exp Pathol Und Pharmakologie 1888;25:1–30. https://doi.org/10.1007/BF01838161.

[95] Tilman D, Downing JA. Effect of high salt concentrations on water structure. Nature 1994;354:363–5.

[96] Scoppola E, Sodo A, McLain SE, Ricci MA, Bruni F. Water-peptide site-specific interactions: A structural study on the hydration of glutathione. Biophys J 2014;106:1701–9. https://doi.org/10.1016/j.bpj.2014.01.046.

[97] Mason PE, Neilson GW, Enderby JE, Saboungi M-L, Dempsey CE, Mackerell AD, et al. The Structure of Aqueous Guanidinium Chloride Solutions. J Am Chem Soc 2004;126:11462–70. https://doi.org/10.1021/ja040034x.

[98] Marcus Y. The guanidinium ion. J Chem Thermodyn 2012;48:70–4. https://doi.org/10.1016/j.jct.2011.11.031.

[99] Marcus Y. Electrostriction in electrolyte solutions. Chem Rev 2011;111:2761–83. https://doi.org/10.1021/cr100130d.

[100] McLain SE, Soper AK, Watts A. Water structure around dipeptides in aqueous solutions. Eur Biophys J 2008;37:647–55. https://doi.org/10.1007/s00249-008-0292-1.

[101] Busch S, Pardo LC, O'Dell WB, Bruce CD, Lorenz CD, Mclain SE. On the structure of water and chloride ion interactions with a peptide backbone in solution. PhysChem Chem Phys 2013;15. https://doi.org/10.1039/c3cp53831a.

[102] Collins KD. Ions from the Hofmeister series and osmolytes: Effects on proteins in solution and in the crystallization process. Methods 2004;34:300–11. https://doi.org/10.1016/j.ymeth.2004.03.021.

[103] Mancinelli R, Botti A, Bruni F, Ricci MA, Soper AK. Hydration of sodium, potassium, and chloride ions in solution and the concept of structure maker/breaker. J Phys Chem B 2007;111:13570–7. https://doi.org/10.1021/jp075913v.

[104] Marcus Y. Effect of Ions on the Structure of Water: Structure Making and Breaking 2007. https://doi.org/10.1021/cr8003828.

[105] Mancinelli R, Botti A, Bruni F, Ricci MA, Soper AK. Perturbation of water structure due to monovalent ions in solution. Phys Chem Chem Phys 2007;9:2959–67. https://doi.org/10.1039/b701855j.

[106] Jungwirth P. Hofmeister series of ions: A simple theory of a not so simple reality. J Phys Chem Lett 2013;4:4258–9. https://doi.org/10.1021/jz402369u.

[107] Juurinen I, Pylkkänen T, Ruotsalainen KO, Sahle CJ, Monaco G, Hämäläinen K, et al. Saturation Behavior in X-ray Raman Scattering Spectra of Aqueous LiCl. J Phys Chem B 2013;117:16506–11. https://doi.org/10.1021/jp409528r.

[108] Kang B, Tang H, Zhao Z, Song S. Hofmeister Series: Insights of Ion Specificity from Amphiphilic Assembly and Interface Property. ACS Omega 2020;5:6229–39. https://doi.org/10.1021/acsomega.0c00237.

[109] Kunz W. Specific ion effects in colloidal and biological systems. Curr Opin Colloid Interface Sci 2009;15:34–9. https://doi.org/10.1016/j.cocis.2009.11.008.

[110] Zoski CG. Handbook of Electrochemistry. 1st ed. Elsevier; 2007.

[111] Marcus Y, Hefter G. Ion pairing. Chem Rev 2006;106:4585–621. https://doi.org/10.1021/cr040087x.

[112] Fini A, Fazio G, Gonzalez-Rodriguez M, Cavallari C, Passerini N, Rodriguez L. Formation of ion-pairs in aqueous solutions of diclofenac salts. Int J Pharm 1999;187:163–73. https://doi.org/10.1016/S0378-5173(99)00180-5.

[113] A van der Vegt NF, Haldrup K, Roke S, Zheng J, Lund M, Bakker HJ. Water-Mediated Ion Pairing: Occurrence and Relevance. Chem Rev 2016;116:7626–41. https://doi.org/10.1021/acs.chemrev.5b00742.

[114] Gao J, Boudon S, Wipff G. Ab Initio and Crystal Structure Analysis of Like-Charged Ion Pairs. J Am Chem Soc 1991;113:3958.

[115] No KT, Nam K-Y, Scheraga HA. Stability of Like and Oppositely Charged Organic Ion Pairs in Aqueous Solution. J Am Chem Soc 1997;119:12917–22.

[116] Desiraju GR. A bond by any other name. Angew Chemie - Int Ed 2011. https://doi.org/10.1002/anie.201002960.

[117] Morokuma K. Why Do Molecules Interact? The Origin of Electron Donor-Acceptor Complexes, Hydrogen Bonding, and Proton Affinity. Acc Chem Res 1977;10:294–300. https://doi.org/10.1021/ar50116a004.

[118] Aakeröy CB. Crystal Engineering: Strategies and Architectures. Acta Crystallogr Sect B 1997;53:569–86. https://doi.org/10.1107/S0108768197008446.

[119] Hardacre C, Holbrey JD, Nieuwenhuyzen M, Youngs TGA. Structure and solvation in ionic liquids. Acc Chem Res 2007;40:1146–55. https://doi.org/10.1021/ar700068x.

[120] Jahn S, Kowalski PM. Theoretical approaches to structure and spectroscopy of Earth materials. Rev Mineral Geochemistry 2014;78:691–743. https://doi.org/10.2138/rmg.2014.78.17.

[121] van der Lubbe SCC, Fonseca Guerra C. The Nature of Hydrogen Bonds: A Delineation of the Role of Different Energy Components on Hydrogen Bond Strengths and Lengths. Chem - An Asian J 2019;14:2760–9. https://doi.org/10.1002/asia.201900717.

[122] Dong K, Zhang S, Wang D, Yao X. Hydrogen bonds in imidazolium ionic liquids. J Phys Chem A 2006;110:9775–82. https://doi.org/10.1021/jp054054c.

[123] Hunt PA. Quantum Chemical Modeling of Hydrogen Bonding in Ionic Liquids. Top Curr Chem 2017;375:59. https://doi.org/10.1007/s41061-017-0142-7.

[124] Pauling L. The nature of the chemical bond and the structure of molecules and crystals : an introduction to modern structural chemistry. Third edit. Ithaca, N.Y: Cornell University Press; 1960.

[125] Pace CN. Evaluating contribution of hydrogen bonding and hydrophobic bonding to protein folding. Methods Enzymol 1995;259:538–54. https://doi.org/10.1016/0076-6879(95)59060-9.

[126] Derewenda ZS, Lee L, Derewenda U. The occurrence of C - H ⋯ O hydrogen bonds in proteins. J Mol Biol 1995;252:248–62. https://doi.org/10.1006/jmbi.1995.0492.

[127] Neves MAC, Yeager M, Abagyan R. Unusual arginine formations in protein function and assembly: Rings, strings, and stacks. J Phys Chem B 2012;116:7006–13. https://doi.org/10.1021/jp3009699.

[128] Pethes I, Bakó I, Pusztai L. Chloride ions as integral parts of hydrogen bonded networks in aqueous salt solutions: The appearance of solvent separated anion pairs. Phys Chem Chem Phys 2020;22:11038–44. https://doi.org/10.1039/d0cp01806f.

[129] Horikawa Y, Arai H, Tokushima T, Shin S. Spectral fingerprint in X-ray absorption for hydrogen-bonded dimer formation of acetic acids in solution. Chem Phys Lett 2012;522:33–7. https://doi.org/10.1016/j.cplett.2011.11.061.

[130] Sarma B, Sreenivas Reddy L, Nangia A. The role of π-stacking in the composition of phloroglucinol and phenazine cocrystals. Cryst Growth Des 2008;8:4546–52. https://doi.org/10.1021/cg800585d.

[131] Hayes R, Warr GG, Atkin R. Structure and Nanostructure in Ionic Liquids. Chem Rev 2015;115:6357–426. https://doi.org/10.1021/cr500411q.

[132] Atkins P, de Paula J. Physical Chemistry. 8th Editio. New York: W. H. Freeman; 2006.

[133] Grimmett MR. 4.06 - Imidazoles and their Benzo Derivatives: (i) Structure. In: Katritzky AR, Rees CW, editors. Compr. Heterocycl. Chem. [as Ref. Ebel, K., Koehler, H., Gamer, A. O. Jäckh, R. Imidazole Deriv. Ullmann's Encycl. Ind. Chem. (Wiley-VCH Verlag GmbH Co. KGaA, 2000)], Oxford: Pergamon; 1984, p. 345–72. https://doi.org/https://doi.org/10.1016/B978-008096519-2.00074-6.

[134] McMurry J. Organic Chemistry. Sixth. London: Brooks/Cole; 2004.

[135] Verma A, Joshi S, Singh D. Imidazole: Having Versatile Biological Activities. J Chem 2013;2013:1–12. https://doi.org/10.1155/2013/329412.

[136] Duboué-Dijon E, Mason PE, Fischer HE, Jungwirth P. Changes in the hydration structure of imidazole upon protonation: Neutron scattering and molecular simulations. J Chem Phys 2017;146:185102. https://doi.org/10.1063/1.4982937.

[137] Li G-S, Martins-Costa MTC, Millot C, Ruiz-Lopez MF. AM1/TIP3P molecular dynamics simulation of imidazole proton-relay processes in aqueous solution. Chem Phys Lett 1998;297:38–44.

[138] Craven BM, McMullan RK, Bell JD, Freeman HC. The crystal structure of imidazole by neutron diffraction at 20°C and –150°C. Acta Crystallogr Sect B Struct Crystallogr Cryst Chem 1977;33:2585–9. https://doi.org/10.1107/s0567740877008954.

[139] Scheiner S, Yi M. Proton Transfer Properties of Imidazole. J Phys Chem 1996;100:9235–41. https://doi.org/10.1021/jp9600571.

[140] Hu J, Fu R, Nishimura K, Zhang L, Zhou HX, Busath DD, et al. Histidines, heart of the hydrogen ion channel from influenza A virus: Toward an understanding of conductance and proton selectivity. Proc Natl Acad Sci U S A 2006;103:6865–70. https://doi.org/10.1073/pnas.0601944103.

[141] Chen WC, Zhu ZL, Lee CS. Organic Light-Emitting Diodes Based on Imidazole Semiconductors. Adv Opt Mater 2018;6:1800258. https://doi.org/10.1002/adom.201800258.

[142] Ehlert C, Holzweber M, Lippitz A, Unger WES, Saalfrank P. A detailed assignment of NEXAFS resonances of imidazolium based ionic liquids. Phys Chem Chem Phys 2016;18:8654. https://doi.org/10.1039/c5cp07434g.

[143] Razmkhah M, Hamed Mosavian MT, Moosavi F. Structural analysis of an amino acid ionic liquid: Bulk and electrical double layer. J Mol Liq 2018;268:506–16. https://doi.org/10.1016/j.molliq.2018.07.047.

[144] Fukumoto K, Yoshizawa M, Ohno H. Room Temperature Ionic Liquids from 20 Natural Amino Acids. J AM CHEM SOC 2005;127:2398–9. https://doi.org/10.1021/ja043451i.

[145] Qiao B, Krekeler C, Berger R, Site LD, Holm C. Effect of anions on static orientational correlations, hydrogen bonds, and dynamics in ionic liquids: A simulational study. J Phys Chem B 2008;112:1743–51. https://doi.org/10.1021/jp0759067.

[146] Wang H, Kelley SP, Brantley JW, Chatel G, Shamshina J, Pereira JFB, et al. Ionic fluids containing both strongly and weakly interacting ions of the same charge have unique ionic and chemical environments as a function of ion concentration. ChemPhysChem 2015;16:993–1002. https://doi.org/10.1002/cphc.201402894.

[147] Hardacre C, McMath SEJ, Nieuwenhuyzen M, Bowron DT, Soper AK. Liquid structure of 1,3-dimethylimidazolium salts. J Phys Condens Matter 2003;15:S159–66. https://doi.org/10.1088/0953-8984/15/1/320.

[148] Wilkes JS. Properties of ionic liquid solvents for catalysis. J Mol Catal A Chem 2004;214:11–7. https://doi.org/10.1016/j.molcata.2003.11.029.

[149] Smith EF, Garcia IJV, Briggs D, Licence P. Ionic liquids in vacuo; solution-phase X-ray photoelectron spectroscopy. Chem Commun 2005;2:5633–5635. https://doi.org/10.1039/b512311a.

[150] Welton T. Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis. Chem Rev 1999;99:2071−2083. https://doi.org/10.1021/cr980032t.

[151] Singh JK, Sharma RK, Ghosh P, Kumar A, Khan ML. Imidazolium based ionic liquids: A promising green solvent for water hyacinth biomass deconstruction. Front Chem 2018;6. https://doi.org/10.3389/fchem.2018.00548.

[152] Rani S, Bagchi D, Pal U, Kumari M, Sharma M, Bera A, et al. The Role of Imidazolium-Based

Surface-Active Ionic Liquid to Restrain the Excited-State Intramolecular H-Atom Transfer Dynamics of Medicinal Pigment Curcumin: A Theoretical and Experimental Approach. ACS Omega 2020;5:25582–92. https://doi.org/10.1021/acsomega.0c02438.

[153] Hardacre C, Holbrey JD, McMath SEJ, Bowron DT, Soper AK. Structure of molten 1,3-dimethylimidazolium chloride using neutron diffraction. J Chem Phys 2003;118:273–8. https://doi.org/10.1063/1.1523917.

[154] Mrestani-Klaus C, Richardt A, Wespe C, Stark A, Humpfer E, Bordusa F. Structural studies on ionic liquid/water/peptide systems by HR-MAS NMR spectroscopy. ChemPhysChem 2012;13:1836–44. https://doi.org/10.1002/cphc.201100927.

[155] Lingscheid Y, Arenz S, Giernoth R. Heteronuclear NOE spectroscopy of ionic liquids. ChemPhysChem 2012;13:261–6. https://doi.org/10.1002/cphc.201100622.

[156] Stassen HK, Ludwig R, Wulf A, Dupont J. Imidazolium Salt Ion Pairs in Solution. Chem EurJ 2015;21:8324–35.

[157] Zanatta M, Girard A-L, Simon NM, Ebeling G, Stassen HK, Livotto PR, et al. The Formation of Imidazolium Salt Intimate (Contact) Ion Pairs in Solution. Angew Chemie Int Ed 2014;53:12817–12821. https://doi.org/10.1002/anie.201408151.

[158] Skarmoutsos I, Welton T, Hunt PA. The importance of timescale for hydrogen bonding in imidazolium chloride ionic liquids. Phys Chem Chem Phys 2014;16. https://doi.org/10.1039/c3cp54551b.

[159] Thomason MJ, Seabourne CR, Sattelle BM, Hembury GA, Stevens JS, Scott AJ, et al. Self-association of organic solutes in solution: a NEXAFS study of aqueous imidazole. Faraday Discuss 2015;179:269–89. https://doi.org/10.1039/C5FD00005J.

[160] Nolting D, Ottosson N, Faubel M, Hertel I V., Winter B. Pseudoequivalent nitrogen atoms in aqueous imidazole distinguished by chemical shifts in photoelectron spectroscopy. J Am Chem Soc 2008;130:8150–1. https://doi.org/10.1021/ja8022384.

[161] Meyer F, Blum M, Benkert A, Hauschild D, Jeyachandran YL, Wilks RG, et al. Site-specific electronic structure of imidazole and imidazolium in aqueous solutions. Phys Chem Chem Phys 2018;20:8302–10. https://doi.org/10.1039/c7cp07885d.

[162] Sapse AM, Massa LJ. Guanidinium Ion: SCF Calculations. J Org Chem 1980;45:719–21. https://doi.org/10.1021/jo01292a034.

[163] Gund P. Guanidine, Trimethylenemethane, and "Y-Delocalization". J Chem Educ 1972;49.

[164] Voloshenko Rossin A, Sladkevich S, Gasser G, Melman A, Lev O. Sensitive Analysis of Nitroguanidine in Aqueous and Soil Matrices by LC-MS. Anal Chem 2017;89:9990–6. https://doi.org/10.1021/acs.analchem.7b02364.

[165] Saczewski F, Balewski Ł. Biological activities of guanidine compounds, 2008-2012 update. Expert Opin Ther Pat 2013;23:965–95. https://doi.org/10.1517/13543776.2013.788645.

[166] Williams ML, Gready JE. Guanidinium-Type Resonance Stabilization and Its Biological Implications. I. The Guanidine and Extended-Guanidine Series. J Comput Chem 2006;10:186–202.

[167] Angyal SJ, Warburton WK. The basic strengths of methylated guanidines. J Chem Soc 1951:2492–4. https://doi.org/10.1039/JR9510002492.

[168] Polák J, Morávek P, Brkljača Z, Vazdar M, Cibulka I, Heyda J. Computation and volumetric insight into (p,T) effect on aqueous guanidinium chloride. J Chem Thermodyn 2021;158. https://doi.org/10.1016/j.jct.2021.106450.

[169] Boudon S, Wipff G, Maigret B. Monte Carlo simulations on the like-charged guanidinium-guanidinium ion pair in water. J Phys Chem 1990;94:6056–61. https://doi.org/10.1021/j100378a078.

[170] Haas DJ, Harris DR, Mills HH. The crystal structure of guanidinium chloride. Acta Crystallogr 1965;19:676–9. https://doi.org/10.1107/S0365110X65004085.

[171] Nelyubina Y V., Lyssenko KA. Probing Ionic Crystals by the Invariom Approach: An Electron Density Study of Guanidinium Chloride and Carbonate. Chem EurJ 2015;21:9733–41. https://doi.org/10.1002/chem.201500296.

[172] Mishiro K, Hu F, Paley DW, Min W, Lambert TH. Macrosteres: The Deltic Guanidinium Ion. European J Org Chem 2016;2016:1655–9. https://doi.org/10.1002/ejoc.201600137.

[173] Macrae CF, Sovago I, Cottrell SJ, Galek PTA, McCabe P, Pidcock E, et al. Mercury 4.0: from visualization to analysis, design and prediction. J Appl Crystallogr 2020;53:226–35. https://doi.org/10.1107/S1600576719014092.

[174] Meuzelaar H, Panman MR, Woutersen S. Guanidinium-Induced Denaturation by Breaking of Salt Bridges. Angew Chemie Int Ed 2015;54:15255–9. https://doi.org/10.1002/anie.201508601.

[175] Bandyopadhyay D, Bhanja K, Mohan S, Ghosh SK, Choudhury N. Effects of Concentration on Like-Charge Pairing of Guanidinium Ions and on the Structure of Water: An All-Atom Molecular Dynamics Simulation Study. J Phys Chem B 2015;119:11262–74. https://doi.org/10.1021/acs.jpcb.5b03064.

[176] Greene Jr RF, Pace CN. Urea and guanidine hydrochloride denaturation curves. J Biol Chem 1974;249:5388–93. https://doi.org/10.1385/0-89603-301-5:177.

[177] Graziano G. Contrasting the denaturing effect of guanidinium chloride with the stabilizing effect of guanidinium sulfate. Phys Chem Chem Phys 2011;13:12008–14. https://doi.org/10.1039/c1cp20843h.

[178] Mason PE, Dempsey CE, Neilson GW, Brady JW. Nanometer-scale ion aggregates in aqueous electrolyte solutions: Guanidinium sulfate and guanidinium thiocyanate. J Phys Chem B 2005;109:24185–96. https://doi.org/10.1021/jp052799c.

[179] Vazdar M, Uhlig F, Jungwirth P. Like-charge ion pairing in water: An ab initio molecular dynamics study of aqueous guanidinium cations. J Phys Chem Lett 2012;3:2021–4. https://doi.org/10.1021/jz3007657.

[180] Van Der Post ST, Tielrooij KJ, Hunger J, Backus EHG, Bakker HJ. Femtosecond study of the effects of ions and hydrophobes on the dynamics of water. Faraday Discuss 2013;160:171–89. https://doi.org/10.1039/c2fd20097j.

[181] Heiles S, Cooper RJ, DiTucci MJ, Williams ER. Hydration of guanidinium depends on its local environment. Chem Sci 2015;6:3420–9. https://doi.org/10.1039/c5sc00618j.

[182] Mandal M, Mukhopadhyay C. Concentration-dependent like-charge pairing of guanidinium ions and effect of guanidinium chloride on the structure and dynamics of water from all-atom molecular dynamics simulation. Phys Rev E - Stat Nonlinear, Soft Matter Phys 2013;88:52708. https://doi.org/10.1103/PhysRevE.88.052708.

[183] Allolio C, Baxova K, Vazdar M, Jungwirth P. Guanidinium Pairing Facilitates Membrane Translocation. J Phys Chem B 2016;120:143–53. https://doi.org/10.1021/acs.jpcb.5b10404.

[184] Kubíčková A, Křížek T, Coufal P, Wernersson E, Heyda J, Jungwirth P. Guanidinium cations pair with positively charged arginine side chains in water. J Phys Chem Lett 2011;2:1387–9. https://doi.org/10.1021/jz2005495.

[185] Scherlis DA, Fattebert J-L, Gygi F, Cococcioni M, Marzari N. A unified electrostatic and cavitation model for first-principles molecular dynamics in solution. J Chem Phys 2006;124:74103. https://doi.org/10.1063/1.2168456.

[186] Vondrášek J, Mason PE, Heyda J, Collins KD, Jungwirth P. The molecular origin of like-charge arginine - arginine pairing in water. J Phys Chem B 2009;113:9041–5. https://doi.org/10.1021/jp902377q.

[187] Hebert MJ, Russell DH. Hydration of Guanidinium Ions: An Experimental Search for Like-Charged Ion Pairs. J Phys Chem Lett 2019;10:1349–54. https://doi.org/10.1021/acs.jpclett.9b00268.

[188] Shih O, England AH, Dallinger GC, Smith JW, Duffey KC, Cohen RC, et al. Cation-cation contact pairing in water: Guanidinium. J Chem Phys 2013;139:35104. https://doi.org/10.1063/1.4813281.

[189] Davey R. Using SAXS and WAXD for in situ observation of crystallisation from liquid phases. Acta Cryst 2002;A58 (Suppl:22.

[190] Ge L, Bernasconi L, Hunt P. Linking electronic and molecular structure: Insight into aqueous chloride solvation. Phys Chem Chem Phys 2013;15:13169–83. https://doi.org/10.1039/c3cp50652e.

[191] Altemose Q, Raichle K, Schnable B, Schwarz C, Kang M, Pantano C, et al. In Situ X-Ray Diffraction Studies of Crystallization Growth Behavior in ZnO-Bi2O3-B2O3 Glass as a Route to Functional Optical Devices. MRS Adv 2018;3:563–7. https://doi.org/10.1557/ADV.2017.640.

[192] Huotari S. Spectroscopy in the frequency domain. Lect Notes Phys 2012;837:15–28. https://doi.org/10.1007/978-3-642-23518-4_2.

[193] Towey JJ. A Structural Approach to Reveal the Croprotective Action of Glycerol. University of Leeds, 2013.

[194] Socrates G. Infrared and Raman Characteristic Group Frequencies: Tables and Charts. Third. Chichester: Wiley; 2001.

[195] Harris DC, Bertolucci MD. Symmetry and spectroscopy : an introduction to vibrational and electronic spectroscopy. Dover Publications; 1989.

[196] Marques MAL, Maitra NT, Nogueira FMS, Gross EKU, Rubio A. Fundamentals of Time-Dependent Density Functional Theory. Springer Berlin Heidelberg; 2012.

[197] McPhee C, Reed J, Izaskun Z. Nuclear Magnetic Resonance (NMR). Core Anal. - A Best Pract. Guid., Elsevier; 2015.

[198] Spitaleri A, Hunter CA, McCabe JF, Packer MJ, Cockroft SL. A 1H NMR study of crystal nucleation in solution. CrystEngComm 2004;6:489–93. https://doi.org/10.1039/b407163h.

[199] Hughes CE, Williams PA, Keast VL, Charalampopoulos VG, Edwards-Gau GR, Harris KDM. New in situ solid-state NMR techniques for probing the evolution of crystallization processes: Pre-nucleation, nucleation and growth. Faraday Discuss 2015;179:115–40. https://doi.org/10.1039/c4fd00215f.

[200] Harris KDM, Hughes CE, Williams PA, Edwards-Gau GR. “NMR Crystallization”: In-situ NMR techniques for time-resolved monitoring of crystallization processes. Acta Crystallogr Sect C Struct Chem 2017;73:137–48. https://doi.org/10.1107/S2053229616019811.

[201] Petkov V. Pair Distribution Functions Analysis. In: Kaufmann EN, editor. Charact. Mater., John Wiley & Sons; 2012, p. 1361–72. https://doi.org/10.1002/0471266965.com159.

[202] Stevens JS, Byard SJ, Muryn CA, Schroeder SLM. Identification of Protonation State by XPS, Solid-State NMR, and DFT: Characterization of the Nature of a New Theophylline Complex by Experimental and Computational Methods. J Phys Chem B 2010;114:13961–9. https://doi.org/10.1021/jp106465u.

[203] Stevens JS, Gainar A, Suljoti E, Xiao J, Golnak R, Aziz EF, et al. Chemical Speciation and Bond Lengths of Organic Solutes by Core-Level Spectroscopy: pH and Solvent Influence on p-Aminobenzoic Acid. Chem - A Eur J 2015;21:7256–63. https://doi.org/10.1002/chem.201405635.

[204] Jagoda-Cwiklik B, Slavíček P, Nolting D, Winter B, Jungwirth P, Slavíček P, et al. Ionization of Aqueous Cations: Photoelectron Spectroscopy and ab Initio Calculations of Protonated Imidazole. J Phys Chem B 2008;112:7355–8. https://doi.org/10.1021/jp802454s.

[205] Sahle CJ, Schroer MA, Jeffries CM, Niskanen J. Hydration in aqueous solutions of ectoine and hydroxyectoine. PCCP 2018;20:27917–23. https://doi.org/10.1039/c8cp05308a.

[206] Messer BM, Cappa CD, Smith JD, Drisdell WS, Schwartz CP, Cohen RC, et al. Local Hydration Environments of Amino Acids and Dipeptides Studied by X-ray Spectroscopy of Liquid

Microjets. J Phys Chem B 2005:21640–6. https://doi.org/10.1021/jp053802v.

[207] Stöhr J. NEXAFS Spectroscopy. Springer Berlin Heidelberg; 1992. https://doi.org/10.1017/CBO9781107415324.004.

[208] Thompson AC, Attwood DT, Gullikson EM, Howells MR, Kortright JB, Robinson AL, et al. X-Ray Data Booklet. 2nd ed. Berkeley: Lawrence Berkeley National Laboratory; 2001.

[209] Huotari S, Sahle CJ, Henriquet C, Al-Zein A, Martel K, Simonelli L, et al. A large-solid-angle X-ray Raman scattering spectrometer at ID20 of the European Synchrotron Radiation Facility. J Synchrotron Radiat 2017;24:521–30. https://doi.org/10.1107/S1600577516020579.

[210] Schülke W. Electron Dynamics by Inelastic X-Ray Scattering. OUP Oxford; 2007.

[211] Sahle CJ, Mirone A, Niskanen J, Inkinen J, Krisch M, Huotari S. Planning, performing and analyzing X-ray Raman scattering experiments. J Synchrotron Radiat 2015;22:400–9. https://doi.org/10.1107/S1600577514027581.

[212] Wernet P, Nordlund D, Bergmann U, Cavalleri M, Odelius M, Ogasawara H, et al. The Structure of the First Coordination Shell in Liquid Water. Science (80-) 2004;304:995–9. https://doi.org/10.1126/science.1096205.

[213] Nilsson A, Nordlund D, Waluyo I, Huang N, Ogasawara H, Kaya S, et al. X-ray absorption spectroscopy and X-ray Raman scattering of water and ice; an experimental view. J Electron Spectros Relat Phenomena 2010;177:99–129. https://doi.org/10.1016/j.elspec.2010.02.005.

[214] Nilsson A, Schlesinger D, Pettersson L. X-ray and simulation studies of water. Riv Del Nuovo Cim Della Soc Ital Di Fis 2016;39.

[215] Naslund L-Å, Edwards DC, Wernet P, Bergmann U, Ogasawara H, Pettersson LGM, et al. X-ray Absorption Spectroscopy Study of the Hydrogen Bond Network in the Bulk Water of Aqueous Solutions. J Phys Chem A 2005;109:5995–6002. https://doi.org/10.1021/jp050413s.

[216] Huang N, Nordlund D, Huang C, Bergmann U, Weiss TM, Pettersson LGMM, et al. X-ray Raman scattering provides evidence for interfacial acetonitrile-water dipole interactions in aqueous solutions. J Chem Phys 2011;135:164509. https://doi.org/10.1063/1.3655468.

[217] Waluyo I, Huang C, Nordlund D, Bergmann U, Weiss TM, Pettersson LGM, et al. The structure of water in the hydration shell of cations from x-ray Raman and small angle x-ray scattering measurements. J Chem Phys 2011;134:064513. https://doi.org/10.1063/1.3533958.

[218] Niskanen J, Sahle CJ, Juurinen I, Koskelo J, Lehtola S, Verbeni R, et al. Protonation Dynamics and Hydrogen Bonding in Aqueous Sulfuric Acid. J Phys Chem B 2015;119:11732–9. https://doi.org/10.1021/acs.jpcb.5b04371.

[219] Gueriau P, Rueff J-P, Bernard S, Kaddissy JA, Goler S, Sahle CJ, et al. Noninvasive Synchrotron-Based X-ray Raman Scattering Discriminates Carbonaceous Compounds in Ancient and Historical Materials. Anal Chem 2017;89:10819–26. https://doi.org/10.1021/acs.analchem.7b02202.

[220] Inkinen J, Niskanen J, Talka T, Sahle CJ, Müller H, Khriachtchev L, et al. X-ray induced dimerization of cinnamic acid: Time-resolved inelastic X-ray scattering study. Sci Rep 2015;5:15851. https://doi.org/10.1038/srep15851.

[221] Sahle CJ, Henriquet C, Schroer MA, Juurinen I, Niskanen J, Krisch M. A miniature closed-circle flow cell for high photon flux X-ray scattering experiments. J Synchrotron Radiat 2015;22:1555–8. https://doi.org/10.1107/S1600577515016331.

[222] Sahle CJ, Schroer MA, Juurinen I, Niskanen J. Influence of TMAO and urea on the structure of water studied by inelastic X-ray scattering. Phys Chem Chem Phys 2016;18:16518–26. https://doi.org/10.1039/c6cp01922f.

[223] Al-Madhagi LH, Callear SK, Schroeder SLM. Hydrophilic and hydrophobic interactions in concentrated aqueous imidazole solutions: a neutron diffraction and total X-ray scattering study. Phys Chem Chem Phys 2020;22:5105–13. https://doi.org/10.1039/c9cp05993h.

[224] Al-Madhagi LHA. Core Level Spectroscopy of Organic Systems - Experimental Studies and

Development of Data Analysis Techniques. University of Leeds, 2019.

[225] Fischer HE, Barnes AC, Salmon PS. Neutron and x-ray diffraction studies of liquids and glasses. Reports Prog Phys 2006;69:233. https://doi.org/10.1088/0034-4885/69/1/R05.

[226] Francis RJ, O'Hare D. The kinetics and mechanisms of the crystallisation of microporous materials. J Chem Soc Dalt Trans 1998:3133–48. https://doi.org/10.1039/A802330A.

[227] Yu L, Hudak BM, Ullah A, Thomas MP, Porter CC, Thisera A, et al. Unveiling the Microscopic Origins of Phase Transformations: An in Situ TEM Perspective. Chem Mater 2020;32:639–50. https://doi.org/10.1021/acs.chemmater.9b03360.

[228] Towey JJ, Soper AK, Dougan L. The structure of glycerol in the liquid state: A neutron diffraction study. Phys Chem Chem Phys 2011;13:9397–406. https://doi.org/10.1039/c0cp02136a.

[229] Sillrén P, Swenson J, Mattsson J, Bowron D, Matic A. The temperature dependent structure of liquid 1-propanol as studied by neutron diffraction and EPSR simulations. J Chem Phys 2013;138:214501. https://doi.org/10.1063/1.4807863.

[230] Soper AK. Is water one liquid or two? J Chem Phys 2019;150:234503. https://doi.org/10.1063/1.5096460.

[231] Schefer J, Birk J, Holm SL, Mannix D, Iversen K, Kolevatov R, et al. HEIMDAL@ESS – fast neutron powder diffraction for materials science. Acta Crystallogr Sect A Found Adv 2018;74:e414–5. https://doi.org/10.1107/S2053273318088964.

[232] Soper AK. Water: Two Liquids Divided by a Common Hydrogen Bond. J Phys Chem B 2011;115:14014–22. https://doi.org/10.1021/jp2031219.

[233] Soper AK, Ricci MA. Structures of High-Density and Low-Density Water. Phys Rev Lett 2000;84:2881–4. https://doi.org/10.1103/PhysRevLett.84.2881.

[234] Soper AK. Recent water myths. Pure Appl. Chem., vol. 82, 2010, p. 1855–67. https://doi.org/10.1351/PAC-CON-09-12-16.

[235] Soper AK. Water and ice structure in the range 220 - 365K from radiation total scattering experiments. Vol. 187 Int. Sch. Phys. "Enrico Fermi" Proc. Int. Sch. Phys. "Enrico Fermi," 2015, p. 151–71.

[236] Bowron D. Empirical Potential Structure Refinement: European Synchrotron Radiation Facility Data Analysis Tutorial 2006.

[237] Fu L, Bienenstock A, Brennan S. X-ray study of the structure of liquid water. J Chem Phys 2009;131:234702. https://doi.org/10.1063/1.3273874.

[238] Hammond OS, Atri RS, Bowron DT, De Campo L, Diaz-Moreno S, Keenan LL, et al. Structural evolution of iron forming iron oxide in a deep eutectic-solvothermal reaction. Nanoscale 2021;13:1723–37. https://doi.org/10.1039/d0nr08372k.

[239] Thomä SLJ, Krauss SW, Eckardt M, Chater P, Zobel M. Atomic insight into hydration shells around facetted nanoparticles. Nat Commun 2019;10. https://doi.org/10.1038/s41467-019-09007-1.

[240] Galicia-Andrés E, Pusztai L, Temleitner L, Pizio O. Microscopic structure of methanol-water mixtures: Synchrotron X-ray diffraction experiments and molecular dynamics simulations over the entire composition range. J Mol Liq 2015;209:586–95. https://doi.org/10.1016/j.molliq.2015.06.045.

[241] Olsson C, Swenson J. Structural Comparison between Sucrose and Trehalose in Aqueous Solution. J Phys Chem B 2020;124:3074–82. https://doi.org/10.1021/acs.jpcb.9b09701.

[242] Pallipurath AR, Evans B, Pugejs A, Chater PA, Schroeder SLM. Molecular Origins of Non-Ideality in Binary Solvents: Synchrotron X-ray Pair Distribution Function (XPDF) Analysis of the Acetone/Water System. [Pre-Print] 2021.

[243] Marangos JP. The measurement of ultrafast electronic and structural dynamics with X-rays. Philos Trans R Soc A Math Phys Eng Sci 2019;377:20170481.

https://doi.org/10.1098/rsta.2017.0481.

[244] Huang S, Ding Y, Feng Y, Hemsing E, Huang Z, Krzywinski J, et al. Generating Single-Spike Hard X-Ray Pulses with Nonlinear Bunch Compression in Free-Electron Lasers. Phys Rev Lett 2017;119:1–6. https://doi.org/10.1103/PhysRevLett.119.154801.

[245] Billinge SJL. Nanometre-scale structure from powder diffraction: total scattering and atomic pair distribution function analysis. In: C. J. Gilmore, J. A. Kaduk, H. Schenk, editors. Int. Tables Crystallogr. Vol. H Powder Diffr., vol. H, 2019, p. 649–72. https://doi.org/10.1107/97809553602060000972.

[246] Yang J. Femtosecond Electron Diffraction Using Relativistic Electron Pulses, Novel Imaging and Spectroscopy. Nov. Imaging Spectrosc., IntechOpen; 2019. https://doi.org/10.5772/intechopen.88511.

[247] Yang J, Zhu X, Wolf TJA, Li Z, Pedro Nunes JF, Coffee R, et al. Imaging CF3I conical intersection and photodissociation dynamics with ultrafast electron diffraction. Science (80-) 2018;361:64–7. https://doi.org/10.1126/science.aat0049.

[248] Ma L, Yong H, Geiser JD, Moreno Carrascosa A, Goff N, Weber PM. Ultrafast x-ray and electron scattering of free molecules: A comparative evaluation. Struct Dyn 2020;7. https://doi.org/10.1063/4.0000010.

[249] Christensen M, Lefmann K, Holm SL, Bertelsen M, Schefer J, Norby P, et al. ESS Instrument Construction Proposal HEIMDAL . 2014.

[250] Aminpour M, Montemagno C, Tuszynski JA. An overview of molecular modeling for drug discovery with specific illustrative examples of applications. Molecules 2019;24. https://doi.org/10.3390/molecules24091693.

[251] Jensen JH. Molecular modeling basics. Taylor & Francis; 2010.

[252] Fang C, Durbeej B. Calculation of Free-Energy Barriers with TD-DFT: A Case Study on Excited-State Proton Transfer in Indigo. J Phys Chem A 2019. https://doi.org/10.1021/acs.jpca.9b05163.

[253] Casida ME. Time-dependent density-functional theory for molecules and molecular solids. vol. 914. 2009. https://doi.org/10.1016/j.theochem.2009.08.018.

[254] da Silva GCQ, Silva GM, Tavares FW, Fleming FP, Horta BAC. Are all-atom any better than united-atom force fields for the description of liquid properties of alkanes? J Mol Model 2020;26:1–17. https://doi.org/10.1007/s00894-020-04548-5.

[255] Yin D, Mackerell AD. Combined ab initio / empirical approach for optimization of Lennard-Jones parameters. J Comput Chem 1998;19:334–48. https://doi.org/10.1002/(SICI)1096-987X(199802)19:3<334::AID-JCC7>3.0.CO;2-U.

[256] Allen MP. Computer simulation of liquids. Oxford: Clarendon Press; 1989.

[257] Soper AK, Bowron D, Thompson H, Bruni F, Ricci MA, Mclain S, et al. EPSRshell User Manual Version 25. 2017.

[258] Jen Chen I, Yin D, MacKerell AD. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. J Comput Chem 2002;23:199–213. https://doi.org/10.1002/jcc.1166.

[259] Zangi R. Refinement of the OPLSAA Force-Field for Liquid Alcohols. ACS Omega 2018;3:18089–99. https://doi.org/10.1021/acsomega.8b03132.

[260] McDonald NA, Jorgensen WL. Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles. J Phys Chem B 1998;102:8049–59. https://doi.org/10.1021/jp981200o.

[261] Jorgensen WL, Tirado-Rives J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. J Am Chem Soc 1988;110:1657–66. https://doi.org/10.1021/ja00214a001.

[262] Jorgensen WL, Mcdonald NA. Development of an all-atom force field for heterocycles.

Properties of liquid pyridine and diazenes. vol. 424. 1998.

[263] Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. vol. 118. 1996. https://doi.org/10.1021/ja9621760.

[264] Grant GH, Richards WG. Computational chemistry. Oxford University Press; 1996. https://doi.org/10.1109/MCISE.2000.881703.

[265] Bruhn G, Davidson ER, Mayer I, Clark AE. Löwdin Population Analysis With and Without Rotational Invariance. Int J Quantum Chem 2006;106:2065–72. https://doi.org/10.1002/qua.20981.

[266] Canongia Lopes JN, Pádua AAH. Molecular force field for ionic liquids composed of triflate or bistriflylimide anions. J Phys Chem B 2004;108:16893–8. https://doi.org/10.1021/jp0476545.

[267] Canongia Lopes JN, Deschamps J, Padua AAH. Modeling Ionic Liquids Using a Systematic All-Atom Force Field. J Phys Chem B 2004;108. https://doi.org/10.1021/jp0362133.

[268] Doherty B, Zhong X, Gathiaka S, Li B, Acevedo O. Revisiting OPLS Force Field Parameters for Ionic Liquid Simulations. J Chem Theory Comput 2017;13:6131–45. https://doi.org/10.1021/acs.jctc.7b00520.

[269] Mccune JA, Turner AH, Coleman F, White CM, Callear SK, Youngs TGA, et al. Association and liquid structure of pyridine-acetic acid mixtures determined from neutron scattering using a “free proton” EPSR simulation model. Phys Chem Chem Phys 2015;17:6767–77. https://doi.org/10.1039/c4cp05746e.

[270] Leontyev I, Stuchebrukhov A. Accounting for electronic polarization in non-polarizable force fields. Phys Chem Chem Phys 2011;13:2613–26. https://doi.org/10.1039/C0CP01971B.

[271] Walker JM. Biomolecular simulations: methods and protocols. First edit. Humana Press; 2013.

[272] Gallo P, Amann-Winkel K, Angell CA, Anisimov MA, Caupin F, Chakravarty C, et al. Water: A Tale of Two Liquids. Chem Rev 2016;116:7463–500. https://doi.org/10.1021/acs.chemrev.5b00750.

[273] Clark GNI, Cappa CD, Smith JD, Saykally RJ, Head-Gordon T. The structure of ambient water. Mol Phys 2010;108:1415–33. https://doi.org/10.1080/00268971003762134.

[274] Fransson T, Harada Y, Kosugi N, Besley NA, Winter B, Rehr JJ, et al. X-ray and Electron Spectroscopy of Water. Chem Rev 2016;116:7551–69. https://doi.org/10.1021/acs.chemrev.5b00672.

[275] Huang C, Wikfeldt KT, Tokushima T, Nordlund D, Harada Y, Bergmann U, et al. The inhomogeneous structure of water at ambient conditions. Proc Natl Acad Sci U S A 2009;106:15214–8. https://doi.org/10.1073/pnas.0904743106.

[276] Ouyang JF, Bettens RPA. Modelling water: A lifetime enigma. Chimia (Aarau) 2015;69:104–11. https://doi.org/10.2533/chimia.2015.104.

[277] Guillot B. A reappraisal of what we have learnt during three decades of computer simulations on water. J. Mol. Liq., vol. 101, 2002, p. 219–60. https://doi.org/10.1016/S0167-7322(02)00094-6.

[278] Cieplak P, Dupradeau F-Y, Duan Y, Wang J. Polarization effects in molecular mechanical force fields. J Phys Condens Matter 2009;21:21. https://doi.org/10.1088/0953-8984/21/33/333102.

[279] Berendsen HJC, Grigera JR, Straatsma TP. The Missing Term in Effective Pair Potentials. J Phys Chem 1987;91:6269–71.

[280] Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A 2001;105:9954–60. https://doi.org/10.1021/jp003020w.

[281] Heyda J, Mason PE, Jungwirth P. Attractive interactions between side chains of histidine-histidine and histidine-arginine-based cationic dipeptides in water. J Phys Chem B 2010;114:8744–9. https://doi.org/10.1021/jp101031v.

[282] Soetens JC, Millot C, Chipot C, Jansen G, Ángyán JG, Maigret B. Effect of polarizability on the potential of mean force of two cations. The guanidimum-guanidinium ion pair in water. J Phys Chem B 1997;101:10910–7. https://doi.org/10.1021/jp972113j.

[283] Mason PE, Neilson GW, Kline SR, Dempsey CE, Brady JW. Nanometer-scale ion aggregates in aqueous electrolyte solutions: Guanidinium carbonate. J Phys Chem B 2006;110:13477–83. https://doi.org/10.1021/jp0572028.

[284] Fuentes-Azcatl R, Mendoza N, Alejandre J. Improved SPC force field of water based on the dielectric constant: SPC/ε. Phys A Stat Mech Its Appl 2015;420:116–23. https://doi.org/10.1016/j.physa.2014.10.072.

[285] Di Gioacchino M, Bruni F, Sodo A, Imberti S, Ricci MA. Ectoine hydration, aggregation and influence on water structure. Mol Phys 2019;117:3311–9. https://doi.org/10.1080/00268976.2019.1649484.

[286] Di Gioacchino M, Ricci MA, Imberti S, Holzmann N, Bruni F. Hydration and aggregation of a simple amino acid: The case of glycine. J Mol Liq 2020;301:112407. https://doi.org/10.1016/j.molliq.2019.112407.

[287] Ekholm V, Vazdar M, Mason PE, Bialik E, Walz M-M, Öhrwall G, et al. Anomalous surface behavior of hydrated guanidinium ions due to ion pairing. J Chem Phys 2018;148:144508. https://doi.org/10.1063/1.5024348.

[288] Röntgen WC. Ueber die Constitution des flüssigen Wassers. Ann Phys 1892;281:91–7. https://doi.org/10.1002/andp.18922810108.

[289] Soper AK, Teixeira J, Head-Gordon T. Is ambient water inhomogeneous on the nanometer-length scale? Proc Natl Acad Sci U S A 2010;107. https://doi.org/10.1073/pnas.0912158107.

[290] Shi R, Tanaka H. Direct Evidence in the Scattering Function for the Coexistence of Two Types of Local Structures in Liquid Water. J Am Chem Soc 2020;142:2868–75. https://doi.org/10.1021/jacs.9b11211.

[291] Egami T, Billinge SJL. Underneath the Bragg Peaks: Structural Analysis of Complex Materials. 1st ed. Elsevier Science; 2003.

[292] van Bokhoven JA, Lamberti C. X-Ray Absorption and X-Ray Emission Spectroscopy: Theory and Applications. Wiley; 2016.

[293] Thornton ST, Rex A. Modern Physics For Scientists and Engineers. 4th ed. Cengage Learning; 2013.

[294] Rubensson J-E. Synchrotron Radiation. Morgan & Claypool Publishers; 2016. https://doi.org/10.1088/978-1-6817-4115-4.

[295] Diamond Light Source Ltd. How Diamond Works - Diamond Light Source. Https//WwwDiamondAcUk/Public/How-Diamond-WorksHtml Date Accessed 2018-06-25 2017.

[296] ESRF. What is a synchrotron? n.d. http://www.esrf.eu/about/synchrotron-science/synchrotron (accessed June 25, 2018).

[297] Calvin S. XAFS for Everyone. Taylor & Francis; 2013.

[298] Attwood DT. X-rays and extreme ultraviolet radiation : principles and applications. Second edi. Cambridge: Cambridge University Press; 2016.

[299] Hulbert SL, Weber JM. Flux and brightness calculations for various synchrotron radiation sources. Nucl Instruments Methods Phys Res Sect A 1992;319:25–31. https://doi.org/10.1016/0168-9002(92)90526-A.

[300] Shpyrko OG. Experimental X-ray studies of liquid surfaces. Harvard University, 2004.

[301] Raimondi P. ESRF-EBS: The Extremely Brilliant Source Project. Synchrotron Radiat News 2016;29:8–15. https://doi.org/10.1080/08940886.2016.1244462.

[302] Pettersson LGM. A two-state picture of water and the funnel of life. Springer International

Publishing; 2019. https://doi.org/10.1007/978-3-030-21755-6.

[303] Hanna L, Lockard J V. From IR to x-rays: gaining molecular level insights on metal-organic frameworks through spectroscopy. J Phys Condens Matter 2019;31:483001. https://doi.org/10.1088/1361-648X/ab38da.

[304] Geloni G, Huang Z, Pellegrini C. The Physics and Status of X-ray Free-electron Lasers. In: Bergmann U, Yachandra V, Yano J, editors. X-Ray Free Electron Lasers Appl. Mater. Chem. Biol., The Royal Society of Chemistry; 2017, p. P001-044. https://doi.org/10.1039/9781782624097.

[305] Nilsson A. Applications of core level spectroscopy to adsorbates. J Electron Spectros Relat Phenomena 2002;126:3–42. https://doi.org/10.1016/S0368-2048(02)00141-X.

[306] Smith JW, Saykally RJ. Soft X-ray Absorption Spectroscopy of Liquids and Solutions. Chem Rev 2017;117:13909–34. https://doi.org/10.1021/acs.chemrev.7b00213.

[307] Bergmann U, Glatzel P, Cramer SP. Bulk-sensitive XAS characterization of light elements: From X-ray Raman scattering to X-ray Raman spectroscopy. Microchem J 2002;71:221–30. https://doi.org/10.1016/S0026-265X(02)00014-0.

[308] Wilson KR, Rude BS, Catalano T, Schaller RD, Tobin JG, Co DT, et al. X-ray Spectroscopy of Liquid Water Microjets. J Phys Chem B 2001;105:3346–9. https://doi.org/10.1021/jp010132u.

[309] Schwartz CP, Uejio JS, Duffin AM, Drisdell WS, Smith JD, Saykally RJ. Soft X-ray absorption spectra of aqueous salt solutions with highly charged cations in liquid microjets. Chem Phys Lett 2010;493:94–6. https://doi.org/10.1016/j.cplett.2010.05.037.

[310] Raman C V. A new radiation. Indian J Phys 1928;2:387–98.

[311] Sternemann C, Wilke M. Spectroscopy of low and intermediate Z elements at extreme conditions: in situ studies of Earth materials at pressure and temperature via X-ray Raman scattering. High Press Res 2016;36:275–92. https://doi.org/10.1080/08957959.2016.1198903.

[312] Sokaras D, Nordlund D, Weng TC, Mori RA, Velikov P, Wenger D, et al. A high resolution and large solid angle x-ray Raman spectroscopy end-station at the Stanford Synchrotron Radiation Lightsource. Rev Sci Instrum 2012;83:043112. https://doi.org/10.1063/1.4704458.

[313] Näslund L-Å, Lüning J, Ufuktepe Y, Ogasawara H, Wernet P, Bergmann U, et al. X-ray Absorption Spectroscopy Measurements of Liquid Water. J Phys Chem B 2005;109:13835–9. https://doi.org/10.1021/jp052046q.

[314] Suzuki T. X-ray raman scattering. experiment. I. J Phys Soc Japan 1967;22:1139–50. https://doi.org/10.1143/JPSJ.22.1139.

[315] Soininen JA, Rehr JJ, Mattila A, Galambosi S, Hämäläinen K. Recent developments in the analysis of X-ray Raman scattering. AIP Conf. Proc., vol. 882, AIP; 2007, p. 102–4. https://doi.org/10.1063/1.2644441.

[316] Soininen JA, Ankudinov AL, Rehr JJ. Inelastic scattering from core electrons: A multiple scattering approach. Phys Rev B 2005;72:045136. https://doi.org/10.1103/PhysRevB.72.045136.

[317] Soininen JA, Mattila A, Rehr JJ, Galambosi S, Hamalainen K. Experimental determination of the core-excited electron density of states. J Phys Condens Matter 2006;18:7327–36. https://doi.org/10.1088/0953-8984/18/31/025.

[318] Georgiou R, Gueriau P, Sahle CJ, Bernard S, Mirone A, Garrouste R, et al. Carbon speciation in organic fossils using 2D to 3D x-ray Raman multispectral imaging. Sci Adv 2019;5:eaaw5019. https://doi.org/10.1126/sciadv.aaw5019.

[319] von der Heyden BP, Hauser EJ, Mishra B, Martinez GA, Bowie AR, Tyliszczak T, et al. Ubiquitous Presence of Fe(II) in Aquatic Colloids and Its Association with Organic Carbon. Environ Sci Technol Lett 2014;1:387–92. https://doi.org/10.1021/ez500164v.

[320] Petitgirard S, Sahle CJ, Weis C, Gilmore K, Spiekermann G, Tse JS, et al. Magma properties at deep Earth's conditions from electronic structure of silica. Geochem Persp Let 2019;9:32–7. https://doi.org/10.7185/geochemlet.1902.

[321] Lee SK, Eng PJ, Mao HK, Meng Y, Newville M, Hu MY, et al. Probing of bonding changes in B2O3 glasses at high pressure with inelastic X-ray scattering. Nat Mater 2005;4:851–4. https://doi.org/10.1038/nmat1511.

[322] Lee SK, Eng PJ, Mao HK. Probing of pressure-induced bonding transitions in crystalline and amorphous earth materials: Insights from X-ray raman scattering at high pressure. Rev Mineral Geochemistry 2014;78:139–74. https://doi.org/10.2138/rmg.2014.78.4.

[323] Wernet P, Testemale D, Hazemann J-L, Argoud R, Glatzel P, Pettersson LGM, et al. Spectroscopic characterization of microscopic hydrogen-bonding disparities in supercritical water. J Chem Phys 2005;123:154503. https://doi.org/10.1063/1.2064867.

[324] Lehmkü F, Forov Y, Bü T, Sahle CJ, Steinke I, Julius K, et al. Intramolecular structure and energetics in supercooled water down to 255 K. Phys Chem Chem Phys 2016;18:6925. https://doi.org/10.1039/c5cp07721d.

[325] Lelong G, Radtke G, Cormier L, Bricha H, Rueff JP, Ablett JM, et al. Detecting non-bridging oxygens: Non-resonant inelastic X-ray scattering in crystalline lithium borates. Inorg Chem 2014;53:10903–8. https://doi.org/10.1021/ic501730q.

[326] Braun A, Nordlund D, Song SW, Huang TW, Sokaras D, Liu X, et al. Hard X-rays in-soft X-rays out: An operando piggyback view deep into a charging lithium ion battery with X-ray Raman spectroscopy. J Electron Spectros Relat Phenomena 2015;200:257–63. https://doi.org/10.1016/j.elspec.2015.03.005.

[327] Ketenoglu D, Spiekermann G, Harder M, Oz E, Koz C, Yagci MC, et al. X-ray Raman spectroscopy of lithium-ion battery electrolyte solutions in a flow cell. J Synchrotron Rad 2018;25:537–42. https://doi.org/10.1107/S1600577518001662.

[328] Fehse M, Sahle CJ, Hogan MP, Cavallari C, Kelder EM, Alfredsson M, et al. Bulk-Sensitive Soft X-ray Edge Probing for Elucidation of Charge Compensation in Battery Electrodes. J Phys Chem C 2019;123:24396–403. https://doi.org/10.1021/acs.jpcc.9b06552.

[329] Schülke W. Inelastic x-ray scattering. Nucl Inst Methods Phys Res A 1989;280:338–48. https://doi.org/10.1016/0168-9002(89)90930-3.

[330] Huotari S, Pylkkänen T, Verbeni R, Monaco G, Hämäläinen K. Direct tomography with chemical-bond contrast. Nat Mater 2011;10:489. https://doi.org/10.1038/NMAT3031.

[331] Cai YQ, Chow P, Chen CC, Ishii H, Tsang KL, Kao CC, et al. Optical design and performance of the Taiwan inelastic X-Ray scattering beamline (BL12XU) at SPring-8. AIP Conf. Proc., vol. 705, American Institute of Physics Inc.; 2004, p. 340–3. https://doi.org/10.1063/1.1757803.

[332] Fister TT, Seidler GT, Wharton L, Battle AR, Ellis TB, Cross JO, et al. Multielement spectrometer for efficient measurement of the momentum transfer dependence of inelastic x-ray scattering. Rev Sci Instrum 2006;77:063901. https://doi.org/10.1063/1.2204581.

[333] Kao CC, Hamalainen K, Krisch M, Siddons DP, Oversluizen T, Hastings JB. Optical design and performance of the inelastic scattering beamline at the National Synchrotron Light Source. Rev Sci Instrum 1995;66:1699–702. https://doi.org/10.1063/1.1146467.

[334] Verbeni R, Pylkkänen T, Huotari S, Simonelli L, Vankó G, Martel K, et al. Multiple-element spectrometer for non-resonant inelastic X-ray spectroscopy of electronic excitations. J Synchrotron Radiat 2009;16:469–76. https://doi.org/10.1107/S090904950901886X.

[335] Yu Q, Mishra B, Fein JB. Role of bacterial cell surface sulfhydryl sites in cadmium detoxification by Pseudomonas putida. J Hazard Mater 2020;391:122209. https://doi.org/10.1016/j.jhazmat.2020.122209.

[336] Bergmann U, Groenzin H, Mullins OC, Glatzel P, Fetzer J, Cramer SP. Carbon K-edge X-ray Raman spectroscopy supports simple, yet powerful description of aromatic hydrocarbons and asphaltenes. Chem Phys Lett 2003;369:184–91. https://doi.org/10.1016/S0009-

2614(02)02003-1.

[337] Petitgirard S, Jacobs J, Cerantola V, Collings IE, Tucoulou R, Dubrovinsky L, et al. A versatile diamond anvil cell for X-ray inelastic, diffraction and imaging studies at synchrotron facilities. Rev Sci Instrum 2019;90:095107. https://doi.org/10.1063/1.5119025.

[338] Sahle CJ, Sternemann C, Schmidt C, Lehtola S, Jahn S, Simonelli L, et al. Microscopic structure of water at elevated pressures and temperatures. Proc Natl Acad Sci U S A 2013;110:6301–6. https://doi.org/10.1073/pnas.1220301110.

[339] Sternemann C, Sahle CJ, Mende K, Schmidt C, Nyrow A, Simonelli L, et al. X-ray Raman scattering: An exciting tool for the study of matter at conditions of the Earth's interior. J Phys Conf Ser 2013;425:202011. https://doi.org/10.1088/1742-6596/425/20/202011.

[340] Henderson GS, De Groot FMF, Moulton BJA. X-ray Absorption Near-Edge Structure (XANES) Spectroscopy. Spectrosc Methods Mineral Geol 2014;78:75–138. https://doi.org/10.2138/rmg.2014.78.3.

[341] Jayaraman A. Diamond anvil cell and high-pressure physical investigations. Rev Mod Phys 1983;55:65–108. https://doi.org/10.1103/RevModPhys.55.65.

[342] Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ. Algorithms and applications for approximate nonnegative matrix factorization. Comput Stat Data Anal 2007;52:155–73. https://doi.org/10.1016/j.csda.2006.11.006.

[343] Kjærvik M, Schwibbert K, Dietrich P, Thissen A, Unger WES. Surface characterisation of Escherichia coli under various conditions by near-ambient pressure XPS. Surf Interface Anal 2018;50:996–1000. https://doi.org/10.1002/sia.6480.

[344] Cushman C V., Chatterjee S, Major GH, Smith NJ, Roberts A, Linford MR. Trends in Advanced XPS Instrumentation. I. Overview of the Technique, Automation, High Sensitivity, Imaging, Snapshot Spectroscopy, Gas Cluster Ion Beams, and Multiple Analytical Techniques on the Instrument. Vac Technol Coat 2016:2–9.

[345] Moulder JF, Chastain J. Handbook of X-ray Photoelectron Spectroscopy: A Reference Book of Standard Spectra for Identification and Interpretation of XPS Data. Physical Electronics Division, Perkin-Elmer Corporation; 1992.

[346] Greczynski G, Hultman L. X-ray photoelectron spectroscopy: Towards reliable binding energy referencing. Prog Mater Sci 2020;107:100591. https://doi.org/10.1016/j.pmatsci.2019.100591.

[347] Cushman C, Dahlquist C, Dietrich P, Bahr S, Thissen, Schaff O, et al. Trends in Advanced XPS Instrumentation. 5. Near-Ambient Pressure XPS. Vac Technol Coat 2017;18:23–31.

[348] Dietrich PM, Bahr S, Yamamoto T, Meyer M, Thissen A. Chemical surface analysis on materials and devices under functional conditions – Environmental photoelectron spectroscopy as non-destructive tool for routine characterization. J Electron Spectros Relat Phenomena 2019;231:118–26. https://doi.org/10.1016/j.elspec.2017.12.007.

[349] Proffen T, Page KL, McLain SE, Clausen B, Darling TW, TenCate JA, et al. Atomic pair distribution function analysis of materials containing crystalline and amorphous phases. Zeitschrift Fur Krist 2005;220:1002–8. https://doi.org/10.1524/zkri.2005.220.12_2005.1002.

[350] Prill D, Juhás P, Schmidt MU, Billinge SJL. Modelling pair distribution functions (PDFs) of organic compounds: describing both intra-and intermolecular correlation functions in calculated PDFs. J Appl Cryst 2015;48:171–8. https://doi.org/10.1107/S1600576714026454.

[351] Eberhart JP. Structural and Chemical Analysis of Materials: X-Ray, Electron and Neutron Diffraction; X-Ray, Electron and Ion Spectrometry; Electron Microscopy. Wiley; 1991.

[352] Randall JT. The determination of structure in liquids by x-ray methods. Trans Faraday Soc 1937;33:105–9. https://doi.org/10.1039/tf9373300105.

[353] Billinge SJL. The atomic pair distribution function: past and present. Z Krist 2004;219:117–21.

[354] Neder RB, Proffen T. Diffuse Scattering and Defect Structure Simulations: A Cook Book Using

the Program DISCUS. OUP Oxford; 2008.

[355] Proffen T, Billinge SJL, Egami T, Louca D. Structural analysis of complex materials using the atomic pair distribution function - A practical guide. Zeitschrift Fur Krist 2003;218:132–43. https://doi.org/10.1524/zkri.218.2.132.20664.

[356] Keen DA. Total scattering and the pair distribution function in crystallography. Crystallogr Rev 2020;26:141–99. https://doi.org/10.1080/0889311X.2020.1797708.

[357] Billinge S, Bowron Rutherford Appleton Laboratory Andy Dent D, Florence A, Goodwin A, Hriljac J, Keen Rutherford Appleton Laboratory Gopinathan Sankar D, et al. XPDF - A Dedicated X-ray PDF Beamline Phase III Beamline Proposal. 2010.

[358] Vineyard GH. Neutron Diffraction Study of Liquid Mercury. J Chem Phys 1954;22:5237. https://doi.org/10.1063/1.1739869.

[359] Neilson GW, Mason PE, Ramos S, Sullivan D. Neutron and X-ray scattering studies of hydration in aqueous solutions. Philos Trans R Soc A Math Phys Eng Sci 2001;359:1575–91. https://doi.org/10.1098/rsta.2001.0866.

[360] Kameda Y, Amo Y, Usuki T, Umebayashi Y, Ikeda K, Otomo T. Origin of the Difference in Ion-Water Distances Determined by X-ray and Neutron Diffraction Measurements for Aqueous NaCl and KCl Solutions. Bull Chem Soc Jpn 2019;92:754–67. https://doi.org/10.1246/bcsj.20180283.

[361] Hughes CE, Hamad S, Harris KDM, Catlow CRA, Griffiths PC. A multi-technique approach for probing the evolution of structural properties during crystallization of organic materials from solution. Faraday Discuss 2007;136:71–89. https://doi.org/10.1039/b616611c.

[362] Chupas PJ, Qiu X, Hanson JC, Lee PL, Grey CP, Billinge SJLL. Rapid-acquisition pair distribution function (RA-PDF) analysis. J Appl Crystallogr 2003;36:1342–7. https://doi.org/10.1107/S0021889803017564.

[363] Pitak M. Exploring the advantages of single-crystal x-ray diffraction in pharma. Chem World 2020.

[364] Sheldrick GM. SHELXT - Integrated space-group and crystal-structure determination. Acta Crystallogr Sect A Found Crystallogr 2015;71:3–8. https://doi.org/10.1107/S2053273314026370.

[365] Soper AK, Barney ER. Extracting the pair distribution function from white-beam X-ray total scattering data. J Appl Crystallogr 2011;44:714–26. https://doi.org/10.1107/S0021889811021455.

[366] Dreele RBV. Protein Crystal Structure Analysis from High-Resolution X-Ray Powder-Diffraction Data. Methods Enzymol 2003;368:254–67. https://doi.org/10.1016/S0076-6879(03)68014-6.

[367] Terban MW, Cheung EY, Krolikowski P, Billinge SJL. Recrystallization, Phase Composition, and Local Structure of Amorphous Lactose from the Total Scattering Pair Distribution Function. Cryst Growth Des 2016;16:210–20. https://doi.org/10.1021/acs.cgd.5b01100.

[368] Soper AK. Rutherford Appleton Laboratory Technical Report RAL-TR-2011-013 GudrunN and GudrunX: programs for correcting raw neutron and X-ray diffraction data to differential scattering cross section. 2011.

[369] Keen DA. A comparison of various commonly used correlation functions for describing total scattering. J Appl Cryst 2001;34:172–7.

[370] Peterson PF, Olds D, McDonnell MT, Page K. Illustrated formalisms for total scattering data: a guide for new practitioners. J Appl Crystallogr 2021;54:317. https://doi.org/10.1107/S1600576720015630.

[371] Egami T, Billinge SJL. Underneath the Bragg Peaks: Structural Analysis of Complex Materials. 2nd ed. Elsevier Science; 2012.

[372] Head-Gordon T, Johnson ME. Tetrahedral structure or chains for liquid water. Proc Natl Acad Sci U S A 2006;103:7973–7. https://doi.org/10.1073/pnas.0510593103.

[373] Sutter JP, Chater PA, Hillman MR, Keeble DS, Tucker MG, Wilhelm H. Three-energy focusing Laue monochromator for the diamond light source x-ray pair distribution function beamline I15-1. AIP Conf Proc 2016;1741:1–5. https://doi.org/10.1063/1.4952877.

[374] Toby BH, Egami T. Accuracy of Pair Distribution Function Analysis Applied to Crystalline and Non-Crystalline Materials. vol. 48. 1992.

[375] Soper AK. The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure? ISRN Phys Chem 2013;2013:1–67. https://doi.org/10.1155/2013/279463.

[376] Thomas S, Rouxel D, Ponnamma D. Spectroscopy of Polymer Nanocomposites. Elsevier Science; 2016.

[377] Zhou W, Apkarian R, Wang ZL, Joy D. Fundamentals of Scanning Electron Microscopy (SEM). In: Zhou W, Wang ZL, editors. Scanning Microsc. Nanotechnol. Tech. Appl., New York, NY: Springer New York; 2007, p. 1–40. https://doi.org/10.1007/978-0-387-39620-0_1.

[378] Newbury DE, Ritchie NWM. Is scanning electron microscopy/energy dispersive X-ray spectrometry (SEM/EDS) quantitative? Scanning 2013;35:141–68. https://doi.org/10.1002/sca.21041.

[379] Neese F. The ORCA program system. WIREs Comput Mol Sci 2012;2:73–8. https://doi.org/10.1002/wcms.81.

[380] Neese F. Software update: the ORCA program system, version 4.0. WIREs Comput Mol Sci 2018;8:e1327. https://doi.org/10.1002/wcms.1327.

[381] Kirschner KN, Reith D, Heiden W. The performance of Dunning, Jensen, and Karlsruhe basis sets on computing relative energies and geometries. Soft Mater 2020;18:200–14. https://doi.org/10.1080/1539445X.2020.1714656.

[382] Monticelli L, Salonen E. Biomolecular Simulations: Methods and Protocols. Humana Press; 2012.

[383] Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. vol. 98. UTC; 1994. https://doi.org/10.1021/j100096a001.

[384] Zheng J, Xu X, Truhlar DG. Minimally augmented Karlsruhe basis sets. Theor Chem Acc 2011;128:295–305. https://doi.org/10.1007/s00214-010-0846-z.

[385] Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Phys Chem Chem Phys 2005;7:3297–305. https://doi.org/10.1039/b508541a.

[386] DeBeer George S, Petrenko T, Neese F. Time-dependent density functional calculations of ligand K-edge X-ray absorption spectra. Inorganica Chim Acta 2008;361:965–72. https://doi.org/10.1016/J.ICA.2007.05.046.

[387] Wennmohs F, Aravena D, Atanasov M, Becker U, Bykov D, Chilkuri VG, et al. ORCA manual Version 4.0 2017:797.

[388] Neese F, Wennmohs F, Hansen A, Becker U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A “chain-of-spheres” algorithm for the Hartree-Fock exchange. Chem Phys 2009;356:98–109. https://doi.org/10.1016/j.chemphys.2008.10.036.

[389] Giannone G, Della Sala F. Minimal auxiliary basis set for time-dependent density functional theory and comparison with tight-binding approximations: Application to silver nanoparticles. J Chem Phys 2020;153:84110. https://doi.org/10.1063/5.0020545.

[390] Elliott P, Furche F, Burke K. Excited states from time-dependent density functional theory. Rev Comput Chem 2009;26:91–165. https://doi.org/10.1002/9780470399545.ch3.

[391] Pescitelli G, Bruhn T. Good Computational Practice in the Assignment of Absolute Configurations by TDDFT Calculations of ECD Spectra. Chirality 2016;28:466–74. https://doi.org/10.1002/CHIR.22600.

[392] Papajak E, Zheng J, Xu X, Leverentz HR, Truhlar DG. Perspectives on Basis Sets Beautiful: Seasonal Plantings of Diffuse Basis Functions. J Chem Theory Comput 2011;7:3027–34. https://doi.org/10.1021/ct200106a.

[393] Wagner JP, Schreiner PR. London Dispersion in Molecular Chemistry—Reconsidering Steric Effects. Angew Chemie Int Ed 2015;54:12274–96. https://doi.org/10.1002/ANIE.201503476.

[394] Cohen AJ, Mori-Sánchez P, Yang W. Challenges for Density Functional Theory. Chem Rev 2011;112:289–320. https://doi.org/10.1021/CR200107Z.

[395] Caldeweyher E, Ehlert S, Hansen A, Neugebauer H, Spicher S, Bannwarth C, et al. A generally applicable atomic-charge dependent London dispersion correction. J Chem Phys 2019;150. https://doi.org/10.1063/1.5090222.

[396] Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 2010;132:154104. https://doi.org/10.1063/1.3382344.

[397] Becke AD. Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 1993;98:5648–52. https://doi.org/10.1063/1.464913.

[398] Yang L, Juhás P, Terban MW, Tucker MG, Billinge SJL. Structure-mining: Screening structure models by automated fitting to the atomic pair distribution function over large numbers of models. Acta Crystallogr Sect A Found Adv 2020;76:395–409. https://doi.org/10.1107/S2053273320002028.

[399] Juhás P, Farrow CL, Yang X, Knox KR, Billinge SJLL. Complex modeling: a strategy and software program for combining multiple information sources to solve ill posed structure and nanostructure inverse problems. Acta Crystallogr Sect A Found Adv 2015;71:562–8. https://doi.org/10.1107/S2053273315014473.

[400] Shi C. xINTERPDF : a graphical user interface for analyzing intermolecular pair distribution functions of organic compounds from X-ray total scattering data. J Appl Cryst 2018;51:1498–9. https://doi.org/10.1107/S1600576718012359.

[401] Coelho AA. TOPAS and TOPAS-Academic : an optimization program integrating computer algebra and crystallographic objects written in C++. J Appl Crystallogr 2018;51:210–8. https://doi.org/10.1107/S1600576718000183.

[402] Farrow CL, Juhas P, Liu JW, Bryndin D, Božin ES, Bloch J, et al. PDFfit2 and PDFgui: computer programs for studying nanostructure in crystals. J Phys Condens Matter 2007;19:335219. https://doi.org/10.1088/0953-8984/19/33/335219.

[403] Playford HY, Owen LR, Levin I, Tucker MG. New insights into complex materials using reverse Monte Carlo modeling. Annu Rev Mater Res 2014;44:429–49. https://doi.org/10.1146/annurev-matsci-071312-121712.

[404] McGreevy RL, Pusztai L. Reverse Monte Carlo Simulation: A New Technique for the Determination of Disordered Structures. Mol Simul 1988;1:359–67. https://doi.org/10.1080/08927028808080958.

[405] Tucker MG, Keen DA, Dove MT, Goodwin AL, Hui Q, Qun Hui. RMCProfile: reverse Monte Carlo for polycrystalline materials. J Phys Chem Lett 2007;19:335218.

[406] Soper AK. Computer simulation as a tool for the interpretation of total scattering data from glasses and liquids. Mol Simul 2012;38:1171–85. https://doi.org/10.1080/08927022.2012.732222.

[407] Youngs T. Dissolve: next generation software for the interrogation of total scattering data by empirical potential generation. Mol Phys 2019;117:3464–77. https://doi.org/10.1080/00268976.2019.1651918.

[408] Youngs TGA. Aten - An application for the creation, editing, and visualization of coordinates for glasses, liquids, crystals, and molecules. J Comput Chem 2010;31:639–48. https://doi.org/10.1002/JCC.21359.

[409] Soper AK. Empirical potential Monte Carlo simulation of fluid structure. Chem Phys 1996;202:295–306. https://doi.org/10.1016/0301-0104(95)00357-6.

[410] White C, Page K. Advanced Simulation Techniques for Total Scattering Data. Neutron News 2013;24:13–4. https://doi.org/10.1080/10448632.2013.750213.

[411] Soper AK. Partial structure factors from disordered materials diffraction data: An approach using empirical potential structure refinement. Phys Rev B - Condens Matter Mater Phys 2005;72:104204. https://doi.org/10.1103/PhysRevB.72.104204.

[412] Soper AK, Castner EW, Luzar A. Impact of urea on water structure: A clue to its properties as a denaturant? Biophys Chem 2003;105:649–66. https://doi.org/10.1016/S0301-4622(03)00095-4.

[413] Towey JJ, Soper AK, Dougan L. What happens to the structure of water in cryoprotectant solutions? Faraday Discuss 2013;167:159. https://doi.org/10.1039/c3fd00084b.

[414] Finney JL, Bowron DT, Soper AK. The structure of aqueous solutions of tertiary butanol. J Phys Condens Matter 2000;12:A123–8. https://doi.org/10.1088/0953-8984/12/8A/313.

[415] Bowron DT, Moreno SD. The structure of a trimolecular liquid: Terf-butyl alcohol :Cyclohexene: Water. J Phys Chem B 2005;109:16210–8. https://doi.org/10.1021/jp052912z.

[416] Pagnotta SE, McLain SE, Soper AK, Bruni F, Ricci MA. Water and trehalose: How much do they interact with each other? J Phys Chem B 2010;114:4904–8. https://doi.org/10.1021/jp911940h.

[417] Marcus Y. Mutual Effects of Ions and Solvents. Ions Solut. their Solvation, Hoboken: John Wiley & Sons Inc.; 2015, p. 156–92. https://doi.org/10.1002/9781118892336.ch5.

[418] Panagiotopoulos AZ. Simulations of activities, solubilities, transport properties, and nucleation rates for aqueous electrolyte solutions. J Chem Phys 2020;153:010903. https://doi.org/10.1063/5.0012102.

[419] Sahle CJ, Niskanen J, Schmidt C, Stefanski J, Gilmore K, Forov Y, et al. Cation Hydration in Supercritical NaOH and HCl Aqueous Solutions. J Phys Chem B 2017:11383–9. https://doi.org/10.1021/acs.jpcb.7b09688.

[420] Egorova KS, Gordeev EG, Ananikov VP. Biological Activity of Ionic Liquids and Their Application in Pharmaceutics and Medicine. Chem Rev 2017;117:7132–89. https://doi.org/10.1021/acs.chemrev.6b00562.

[421] Fedotova M V., Kruchinin SE, Chuev GN. Local ion hydration structure in aqueous imidazolium-based ionic liquids: The effects of concentration and anion nature. J Mol Liq 2017;247:100–8. https://doi.org/10.1016/j.molliq.2017.09.087.

[422] Blum M, Weinhardt L, Fuchs O, Bär M, Zhang Y, Weigand M, et al. Solid and liquid spectroscopic analysis (SALSA)-a soft x-ray spectroscopy endstation with a novel flow-through liquid cell. Rev Sci Instrum 2009;80. https://doi.org/10.1063/1.3257926.

[423] Nolting D, Aziz EF, Ottosson N, Faubel M, Hertel I V, Winter B. pH-Induced Protonation of Lysine in Aqueous Solution Causes Chemical Shifts in X-ray Photoelectron Spectroscopy. J Am Chem Soc 2007;129:14068–73. https://doi.org/10.1021/ja072971l.

[424] Messer BM, Cappa CD, Smith JD, Wilson KR, Gilles MK, Cohen RC, et al. pH Dependence of the Electronic Structure of Glycine. J Phys Chem B 2005;109:5375–82. https://doi.org/10.1021/jp0457592.

[425] Brown MA, Vila F, Sterrer M, Thurmer S, Winter B, Ammann M, et al. Electronic structures of formic acid (HCOOH) and formate (HCOO-) in aqueous solutions. J Phys Chem Lett 2012;3:1754–9. https://doi.org/10.1021/jz300510r.

[426] Gainar A, Stevens JS, Jaye C, Fischer DA, Schroeder SLM. NEXAFS Sensitivity to Bond Lengths in Complex Molecular Materials: A Study of Crystalline Saccharides. J Phys Chem B 2015;119:14373–81. https://doi.org/10.1021/acs.jpcb.5b07159.

[427] Stevens JS, Gainar A, Jaye C, Fischer DA, Schroeder SLM. NEXAFS and XPS of p-Aminobenzoic Acid Polymorphs: The Influence of Local Environment. J Phys Conf Ser 2016;712:012133. https://doi.org/10.1088/1742-6596/712/1/012133.

[428] Leinweber P, Kruse J, Walley FL, Gillespie A, Eckhardt K-U, Blyth RIR, et al. Nitrogen K-edge XANES – an overview of reference compounds used to identify "unknown" organic nitrogen in environmental samples. J Synchrotron Rad 2007;14:500–11. https://doi.org/10.1107/S0909049507042513.

[429] Stevens JS, Gainar A, Suljoti E, Xiao J, Golnak R, Aziz EF, et al. NEXAFS Chemical State and Bond Lengths of p- Aminobenzoic Acid in Solution and Solid State. J. Phys. Conf. Ser., vol. 712, 2016, p. 012136. https://doi.org/10.1088/1742-6596/712/1/012136.

[430] Stevens JS, Newton LK, Jaye C, Muryn CA, Fischer DA, Schroeder SLM. Proton transfer, hydrogen bonding, and disorder: Nitrogen near-edge X-ray absorption fine structure and X-ray photoelectron spectroscopy of Bipyridine−acid salts and co-crystals. Cryst Growth Des 2015;15:1776–83. https://doi.org/10.1021/cg5018278.

[431] Stevens JS, Byard SJ, Zlotnikov E, Schroeder SLM. Detection of free base surface enrichment of a pharmaceutical salt by X-ray photoelectron spectroscopy (XPS). J Pharm Sci 2011;100:942–8. https://doi.org/10.1002/jps.22334.

[432] Stevens JS, Seabourne CR, Jaye C, Fischer DA, Scott AJ, Schroeder SLM. Incisive probing of intermolecular interactions in molecular crystals: core level spectroscopy combined with density functional theory. J Phys Chem B 2014;118:12121–9. https://doi.org/10.1021/jp506983s.

[433] Stevens JS, Byard SJ, Seaton CC, Sadiq G, Davey RJ, Schroeder SLM. Proton transfer and hydrogen bonding in the organic solid state: a combined XRD/XPS/ssNMR study of 17 organic acid–base complexes. Phys Chem Chem Phys 2014;16:1150–60. https://doi.org/10.1039/c3cp53907e.

[434] Stevens JS, Byard SJ, Seaton CC, Sadiq G, Davey RJ, Schroeder SLM. Crystallography aided by atomic core-level binding energies: Proton transfer versus hydrogen bonding in organic crystal structures. Angew Chemie - Int Ed 2011;50:9916–8. https://doi.org/10.1002/anie.201103981.

[435] SPECS Surface Nano Analysis GmbH. Application note #000394 XPS surface chemical analysis of aqueous solutions with EnviroESCA. SPECS Surface Nano Analysis GmbH; n.d.

[436] Patel DI, Roychowdhury T, Jain V, Shah D, Avval TG, Chatterjee S, et al. Introduction to near-ambient pressure x-ray photoelectron spectroscopy characterization of various materials. Surf Sci Spectra 2019;26. https://doi.org/10.1116/1.5109118.

[437] Walton J. Peak Fitting with CasaXPS: A Casa Pocket Book. Acolyte Science; 2010.

[438] Stevens JS, Byard SJ, Schroeder SLM. Characterization of Proton Transfer in Co-Crystals by X-ray Photoelectron Spectroscopy (XPS). Cryst Growth Des 2010;10:1435–42. https://doi.org/10.1021/cg901481q.

[439] Briggs D, Seah MP. Practical Surface Analysis, Auger and X-ray Photoelectron Spectroscopy. Wiley; 1990.

[440] Pathak A, Sarre PJ. Protonated PAHs as carriers of diffuse interstellar bands. Mon Not R Astron Soc Lett 2008;391:10–4. https://doi.org/10.1111/j.1745-3933.2008.00544.x.

[441] Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 2012;4:17. https://doi.org/10.1186/1758-2946-4-17.

[442] Jagoda-Cwiklik B, Slavíček P, Cwiklik L, Nolting D, Winter B, Jungwirth P. Ionization of imidazole in the gas phase, microhydrated environments, and in aqueous solution. J Phys Chem A 2008;112:3499–505. https://doi.org/10.1021/jp711476g.

[443] Liem SY, Shaik MS, Popelier PLA. Aqueous imidazole solutions: A structural perspective from simulations with high-rank electrostatic multipole moments. J Phys Chem B 2011;115:11389–

98. https://doi.org/10.1021/jp2053234.

[444] Apen E, Hitchcock AP, Gland JL. Experimental Studies of the Core Excitation of Imidazole, 4,5-Dicyanoimidazole, and s-Triazine. J Phys Chem 1993;97:6859–66. https://doi.org/10.1021/j100128a019.

[445] Villar-Garcia IJ, Smith EF, Taylor AW, Qiu F, Lovelock KRJ, Jones RG, et al. Charging of ionic liquid surfaces under X-ray irradiation: The measurement of absolute binding energies by XPS. Phys Chem Chem Phys 2011;13:2797–808. https://doi.org/10.1039/c0cp01587c.

[446] Stevens JS, Byard SJ, Schroeder SLM. Salt or Co-Crystal? Determination of Protonation State by X-Ray Photoelectron Spectroscopy (XPS). J Pharm Sci 2010;99:4453–7. https://doi.org/10.1002/jps.22164.

[447] Winter B, Aziz EF, Hergenhahn U, Faubel M, Hertel I V. Hydrogen bonds in liquid water studied by photoelectron spectroscopy. J Chem Phys 2007;126:124504. https://doi.org/10.1063/1.2710792.

[448] Shah D, Patel DI, Bahr S, Dietrich P, Meyer M, Thißen A, et al. Liquid water, by near-ambient pressure XPS. Surf Sci Spectra 2019;26:024003. https://doi.org/10.1116/1.5119259.

[449] Patel DI, Shah D, Bahr S, Dietrich P, Meyer M, Thißen A, et al. Water vapor, by near-ambient pressure XPS. Surf Sci Spectra 2019;26:014026. https://doi.org/10.1116/1.5111634.

[450] Yamamoto S, Kendelewicz T, Newberg JT, Ketteler G, Starr DE, Mysak ER, et al. Water Adsorption on α-Fe2O3 (0001) at near Ambient Conditions. J Phys Chem C 2010;114:2256–2266. https://doi.org/10.1021/jp909876t.

[451] Beale AM, Reilly LM, Sankar G. Watching the crystallisation of complex oxides by in situ X-ray techniques. Appl Catal A Gen 2007;325:290–5. https://doi.org/10.1016/J.APCATA.2007.02.046.

[452] Jacques SDM, Michiel M Di, Kimber SAJ, Yang X, Cernik RJ, Beale AM, et al. Pair distribution function computed tomography. Nat Commun 2013;4. https://doi.org/10.1038/ncomms3536.

[453] Cho K, Saheb DN, Choi J, Yang H. Real time in situ X-ray diffraction studies on the melting memory effect in the crystallization of β-isotactic polypropylene. Polymer (Guildf) 2002;43:1407–16. https://doi.org/10.1016/S0032-3861(01)00729-7.

[454] Fogg AM, Price SJ, Francis RJ, O'Brien S, O'Hare D. Determination of the kinetics of crystallisation of gibbsite using time resolved in situ energy dispersive powder X-ray diffraction. J Mater Chem 2000;10:2355–7. https://doi.org/10.1039/B005198P.

[455] Becker J, Bremholm M, Tyrsted C, Pauw B, Jensen KMO, Eltzholt J, et al. Experimental setup for in situ X-ray SAXS/WAXS/PDF studies of the formation and growth of nanoparticles in near-and supercritical fluids. J Appl Crystallogr 2010;43:729–36. https://doi.org/10.1107/S0021889810014688.

[456] Potter ME, Light ME, Irving DJM, Oakley AE, Chapman S, Chater P, et al. Exploring the origins of crystallisation kinetics in hierarchical materials using in situ X-ray diffraction and pair distribution function analysis. Phys Chem Chem Phys 2020;22:18860–7. https://doi.org/10.1039/D0CP00670J.

[457] Hill RL, Schwartz HC, Smith EL. The effect of urea and guanidine hydrochloride on activity and optical rotation of crystalline papain. J Biol Chem 1959;234:572–6. https://doi.org/10.1016/s0021-9258(18)70247-0.

[458] Wernersson E, Heyda J, Vazdar M, Lund M, Mason PE, Jungwirth P. Orientational Dependence of the Affinity of Guanidinium Ions to the Water Surface. J Phys Chem B 2011;115:32. https://doi.org/10.1021/jp207499s.

[459] Mason PE, Neilson GW, Dempsey CE, Barnes AC, Cruickshank JM. The hydration structure of guanidinium and thiocyanate ions: Implications for protein stability in aqueous solution. PNAS 2003;100:4557–61. https://doi.org/10.1073/PNAS.0735920100.

[460] Youngs TGA. dlputils ProjectAten. Dlputils Doc 2007. https://www.projectaten.com/dlputils (accessed October 11, 2021).

[461] Canongia Lopes JN, Pádua AAH. Molecular force field for ionic liquids III: Imidazolium, pyridinium, and phosphonium cations; chloride, bromide, and dicyanamide anions. J Phys Chem B 2006;110:19586–92. https://doi.org/10.1021/jp063901o.

[462] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35. https://doi.org/10.1063/1.445869.

[463] Izadi S, Onufriev A V. Accuracy limit of rigid 3-point water models. J Chem Phys 2016;145:74501. https://doi.org/10.1063/1.4960175.

[464] Marcus Y. Concentration Dependence of Ionic Hydration Numbers. J Phys Chem B 2014;229:10471−10476. https://doi.org/10.1021/jp5039255.

[465] Marcus Y. On water structure in concentrated salt solutions. J Solution Chem 2009;38:513–6. https://doi.org/10.1007/s10953-009-9398-z.

[466] Sellberg JA, Huang C, Mcqueen TA, Loh ND, Laksmono H, Schlesinger D, et al. Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature 2014. https://doi.org/10.1038/nature13266.

[467] Jollife IT, Cadima J. Principal component analysis: A review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 2016;374. https://doi.org/10.1098/rsta.2015.0202.

[468] Bosman M, Watanabe M, Alexander DTL, Keast VJ. Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images n.d. https://doi.org/10.1016/j.ultramic.2006.04.016.

[469] Ngo L. Principal component analysis explained simply. BioTuring 2018. https://blog.bioturing.com/2018/06/14/principal-component-analysis-explained-simply/.

[470] Horan B. Conversation with Bethan Evans 2022.

[471] Bowron DT. Conversation with Bethan Evans, Anuradha Pallipurath and Arturs Pugejs 2020.

[472] Masunov A, Lazaridis T. Potentials of mean force between ionizable amino acid side chains in water. J Am Chem Soc 2003;125:1722–30. https://doi.org/10.1021/ja025521w.

[473] Powell DH, Neilson GW, Enderby JE. The structure of Cl- in aqueous solution: An experimental determination of gClH(r) and gClO(r). J Phys Condens Matter 1993;5:5723–30. https://doi.org/10.1088/0953-8984/5/32/003.

[474] Enderby JE. Ion solvation via neutron scattering. Chem Soc Rev 1995;24:159–68. https://doi.org/10.1039/CS9952400159.

[475] Rozas I, Alkorta I, Elguero J. Hydrogen bonds and ionic interactions in Guanidine/Guanidinium complexes: A computational case study. Struct Chem 2008;19:923–33. https://doi.org/10.1007/s11224-008-9377-9.

[476] Botti A, Bruni F, Imberti S, Ricci MA, Soper AK. Ions in water: The microscopic structure of concentrated NaOH solutions. J Chem Phys 2004;120:10154–62. https://doi.org/10.1063/1.1705572.

[477] Bin G, Feng-Shou Z, Yu-Gai H, Xia F. Variations of water's local-structure induced by solvation of NaCl. Chinese Phys B 2010;19:36101.

[478] Balos V, Bonn M, Hunger J. Anionic and cationic Hofmeister effects are non-additive for guanidinium salts. Phys Chem Chem Phys 2017;19:9724–8. https://doi.org/10.1039/c7cp00790f.

[479] Scott JN, Nucci N V., Vanderkooi JM. Changes in water structure induced by the guanidinium cation and implications for protein denaturation. J Phys Chem A 2008;112:10939–48. https://doi.org/10.1021/jp8058239.

[480] England JL, Haran G. Role of Solvation Effects in Protein Denaturation: From Thermodynamics

to Single Molecules and Back. Annu Rev Phys Chem Is Online 2011;62:257–77. https://doi.org/10.1146/annurev-physchem-032210-103531.

[481] Weerasinghe S, Smith PE. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. J Chem Phys 2004;121:2180–6. https://doi.org/10.1063/1.1768938.

[482] Mandal M, Mukhopadhyay C. Microsecond molecular dynamics simulation of guanidinium chloride induced unfolding of ubiquitin. Phys Chem Chem Phys 2014;16:21706–16. https://doi.org/10.1039/C4CP01657B.

[483] Hunger J, Niedermayer S, Buchner R, Hefter G. Are nanoscale ion aggregates present in aqueous solutions of guanidinium salts? J Phys Chem B 2010;114:13617–27. https://doi.org/10.1021/jp101520h.

[484] Hunger J, Neueder R, Buchner R, Apelblat A. A conductance study of guanidinium chloride, thiocyanate, sulfate, and carbonate in dilute aqueous solutions: Ion-association and carbonate hydrolysis effects. J Phys Chem B 2013;117:615–22. https://doi.org/10.1021/jp311425v.

[485] Chitra R, Smith PE. Molecular Association in Solution: A Kirkwood-Buff Analysis of Sodium Chloride, Ammonium Sulfate, Guanidinium Chloride, Urea, and 2,2,2-Trifluoroethanol in Water. J Phys Chem B 2002;106:1491–500. https://doi.org/10.1021/jp011462h.

[486] Chitra R, Smith PE. Molecular dynamics simulations of the properties of cosolvent solutions. J Phys Chem B 2000;104:5854–64. https://doi.org/10.1021/jp000558t.

[487] Kaminski G, Jorgensen WL. Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids. vol. 100. American Chemical Society; 1996. https://doi.org/10.1021/jp9624257.

[488] Weerasinghe S, Smith PE. A Kirkwood–Buff derived force field for sodium chloride in water. J Chem Phys 2003;119:11342. https://doi.org/10.1063/1.1622372.

[489] Soper AK. On the uniqueness of structure extracted from diffraction experiments on liquids and glasses. J Phys Condens Matter 2007;19:415108. https://doi.org/10.1088/0953-8984/19/41/415108.

[490] Narten AH, Vaslow F, Levy HA. Diffraction pattern and structure of aqueous lithium chloride solutions. J Chem Phys 1973;58:5017. https://doi.org/10.1021/j100699a015.

[491] Fetisov EO, Mundy CJ, Schenter GK, Benmore CJ, Fulton JL, Kathmann SM. Nanometer-Scale Correlations in Aqueous Salt Solutions. J Phys Chem Lett 2020;11:2598–604. https://doi.org/10.1021/acs.jpclett.0c00322.

[492] Buckner JK, Jorgensen WL. Energetics and Hydration of the Constituent Ion Pairs of Tetramethylammonium Chloride. J Am Chem Soc 1989;111:2507–16.

[493] Wojdyr M. Fityk: a general-purpose peak fitting program. J Appl Crystallogr 2010;43:1126–8.

[494] Nagasaka M, Yuzawa H, Kosugi N. Interaction between Water and Alkali Metal Ions and Its Temperature Dependence Revealed by Oxygen K-Edge X-ray Absorption Spectroscopy 2017. https://doi.org/10.1021/acs.jpcb.7b09789.

[495] Soper AK. Orientational correlation function for molecular liquids: The case of liquid water. J Chem Phys 1994;101:6888–901. https://doi.org/10.1063/1.468318.

[496] Zhang Q, Li Y, Zhuang W. Ion effect on the structures and dynamics of aqueous solutions. Sci Sin Chim 2016;46:51–8. https://doi.org/10.1360/N032015-00151.

[497] Liu C-W, Wang F, Yang L, Li X-Z, Zheng W-J, Gao YQ. Stable Salt–Water Cluster Structures Reflect the Delicate Competition between Ion–Water and Water–Water Interactions. J Phys Chem B 2014;118:743–51. https://doi.org/10.1021/JP408439J.

[498] Steed KM, Steed JW. Packing problems: High Z' crystal structures and their relationship to cocrystals, inclusion compounds, and polymorphism. Chem Rev 2015;115:2895–933. https://doi.org/10.1021/cr500564z.

[499] Ggor A, Piecha A, Jakubas R, Miniewicz A. Crystal structure and characterization of a novel acentric imidazolium analog [C3N2H5+][Br-]. Chem Phys Lett 2011;503:134–8. https://doi.org/10.1016/j.cplett.2010.12.083.

[500] Hnědkovský L, Cibulka I. An Automated Vibrating-Tube Densimeter for Measurements of Small Density Differences in Dilute Aqueous Solutions. Int J Thermophys 2004;25:1135–42. https://doi.org/10.1023/B:IJOT.0000038505.88771.2e.

[501] Soper AK, Edler KJ. Coarse-grained empirical potential structure refinement: Application to a reverse aqueous micelle. Biochim Biophys Acta - Gen Subj 2017;1861:1652–60. https://doi.org/10.1016/j.bbagen.2017.02.028.

[502] Bowron DT, Finney JL, Soper AK. The structure of liquid tetrahydrofuran. J Am Chem Soc 2006;128:5119–26. https://doi.org/10.1021/ja0583057.

[503] Blanco D, Antikainen O, Räikkönen H, Yliruusi J, Juppo AM. Effect of colloidal silicon dioxide and moisture on powder flow properties: Predicting in-process performance using image-based analysis. Int J Pharm 2021;597:120344. https://doi.org/10.1016/J.IJPHARM.2021.120344.

[504] Thomason MJ. Soft X-ray Spectroscopy of Molecular Species in Solution: Studies of Imidazole and Imidazole/Water Systems. The University of Manchester, 2012.

[505] Stevens JS, De Luca AC, Pelendritis M, Terenghi G, Downes S, Schroeder SLM. Quantitative analysis of complex amino acids and RGD peptides by X-ray photoelectron spectroscopy (XPS). Surf Interface Anal 2013;45:1238–46. https://doi.org/10.1002/sia.5261.

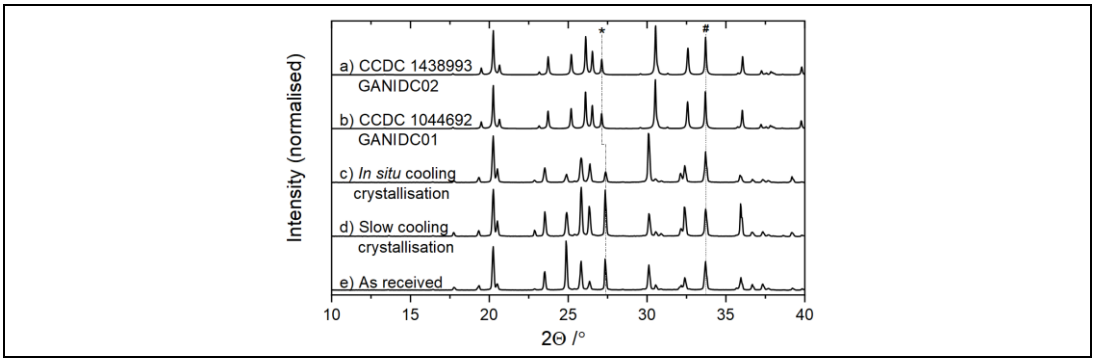
[506] de Groot F, Kotani A. Fundamental aspects of core level spectroscopies. Core Lev. Spectrosc. Solids, CRC Press; 2008, p. 11–38.

[507] Gordon ML, Tulumello D, Cooper G, Hitchcock AP, Glatzel P, Mullins OC, et al. Inner-Shell Excitation Spectroscopy of Fused-Ring Aromatic Molecules by Electron Energy Loss and X-ray Raman Techniques. J Phys Chem A 2003;107:8512–20. https://doi.org/10.1021/jp035607r.

Appendix A Sample characterisation

A.1 X-ray powder diffraction

Samples of the *as received* and recrystallised GuHCl were analysed by XRD to identify any differences in the polymorphic form. The XRD patterns for two GuHCl depositions in CCDC[53] are also shown in App. Figure 1. The crystallographic structures of GANIDC01 and GANIDC02 were collected at 100 K, compared with ambient temperature for the *as received* and recrystallised samples. All patterns have been aligned using the first main peak (20.3°).



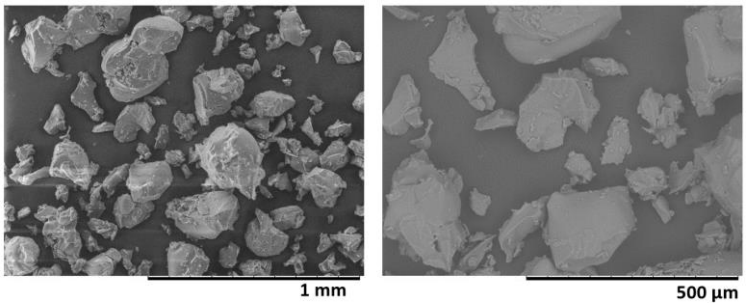
App. Figure 1 XRD analysis of GuHCl: a) *as received*, b&c) GuHCl from cooling crystallisation processes and d&e) from CCDC

There is a strong similarity in the peak distribution for all five structures, although the peak positions of a and b are shifted to higher 2θ values in nearly all cases. This difference could be due to the structures on which patterns a and b are based being measured at 100 K, so some shifts in the unit cell parameters would be expected at a lower temperature. The cell volume of sample (a) is 892.91 Å^3 compared to 898.28 Å^3 for sample (c). If the atomic spacing and thermal displacement of the atoms in the structure reduce with temperature, then it would be expected that the peaks would shift to higher 2θ values. The line indicated with (*) show that the peak in a and b are just below those in c-e, and the line (#) indicates where the peaks in a-e are aligned. Subtle differences in the pattern could be due to difference in morphology because of the synthesis or crystallisation methods, and collection temperature. The XRD patterns suggest that the *as received* and recrystallised samples are of the same GuHCl polymorph.

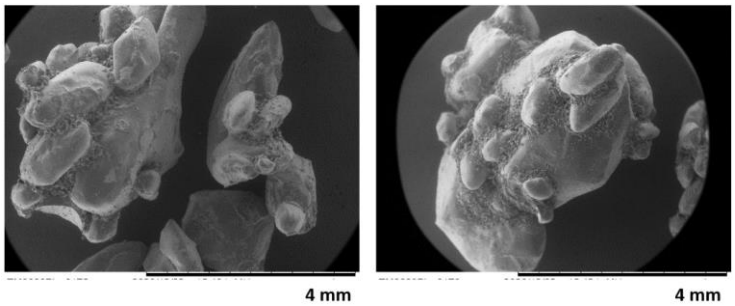
[53] Data acquired from powder pattern function in Mercury

A.2 SEM and EDX analysis

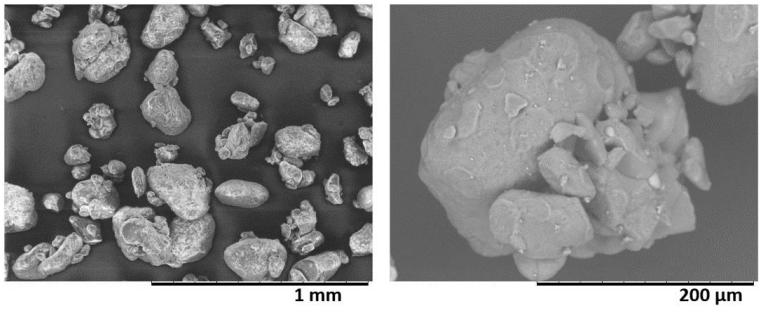
As expected, the crystals formed by cooling crystallisation (App. Figure 2a) are much smaller than those formed by slow cooling crystallisation from solution at room temperature over a longer period of time (App. Figure 2b). The size of the *as received* GuHCl crystals (App. Figure 2c) appear similar to those formed by cooling crystallisation, although the *as received* crystals appear to have fewer smaller-diameter particles than in the recrystallised material, which also have a slightly rougher appearance.



a) Controlled cooling crystallisation
BSE. Accelerating voltage = 15 keV, working distance = 8.60mm, Magnification = left: x100, right: x200



b) Slow cooling crystallisation
SE. Accelerating voltage = 5 keV, working distance = left: 9.50 mm, right: 6.70 mm, Magnification = left: x25, right: x25



c) As received
Left: SE. Accelerating voltage = 15 keV, working distance = 5.90mm, Magnification = x100.
Right: BSE. Accelerating voltage = 5 keV, working distance = 6.00mm, Magnification = x500.

App. Figure 2 SEM images of GuHCl, recrystallised and as received

(a) temperature-controlled cooling crystallisation, as carried out at XPDF beamline, (b) crystallisation from concentrated solution with longer induction time and (c) as received

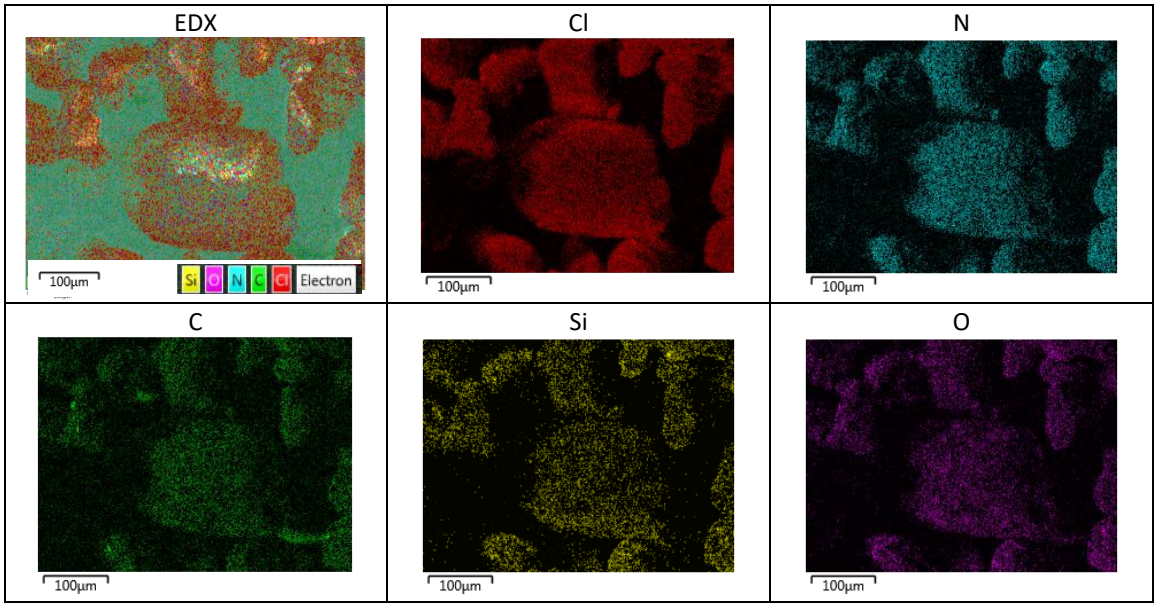
Due to the hazy appearance of GuHCl in solution under some conditions (App. Figure 3), elemental analysis by EDX was undertaken to look for the presence of silicon, as it was suspected it may have been incorporated into the material to aid flowability e.g. in the form of silicon dioxide [503].

App. Figure 3 Aqueous GuHCl solution at 60°C, and with a hazy appearance at 30°C

The images in App. Figure 4 show that whilst there is some evidence of Si in the material, it is a low proportion (0.68 atomic % - App. Table 1).

App. Table 1 Elemental analysis of GuHCl by EDX

Element	Weight %	Atomic %
N	54.69	61.21
Cl	17.23	7.62
O	14.03	13.75
Si	1.22	0.68
C	12.82	16.73
Total	100	100



App. Figure 4 EDX analysis of GuHCl, as received

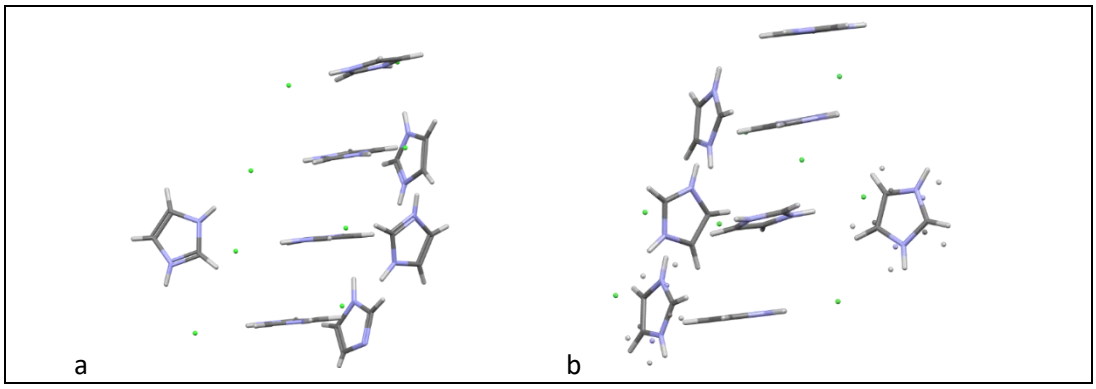
A.3 SCXRD including PDFgui refinement

SCXRD analysis was performed on ImidHCl and GuHCl. XPDF patterns were collected for each material, with the intention of refining the crystallographic structures derived by SCXRD to the XPDF patterns.

A.3.1 ImidHCl

In total, three SCXRD data collections of ImidHCl were undertaken[54]. The first analysis was carried out using ImidHCl that was synthesised by rapid evaporation (Section 2.2.1). The SCXRD analysis (ImidHCl SCXRD 1) identified that one of the Imid rings appeared to remain unprotonated. SCXRD was carried out on a crystal produced using a second method for ImidHCl synthesis (evaporation under vacuum, Section 2.2.1). This analysis (ImidHCl SCXRD 2) indicated that all Imid had been protonated, but that two of the Imid rings were disordered. As the data were only collected for structure solution (1/4 sphere), the measurements were repeated over a full sphere (ImidHCl SCXRD 3).

ImidHCl SCXRD 1: ImidHCl was solved in the $P2_1/n$, monoclinic space group, with eight formula units in the asymmetric unit (Z'=8) (App. Figure 5a). The data were obtained to 99% completeness and the structure model has a goodness-of-fit of 1.032%, with an R-factor of 4.9%. Hydrogen atom positions were refined from the Fourier Map (electron density difference map). Seven of the eight Imid molecules were protonated, although eight chloride ions were located in the structure.



App. Figure 5 Structure of ImidHCl from SCXRD

Synthesised by a) rapid evaporation (unprotonated $ImidH^+$ at bottom right) and b) evaporation under vacuum

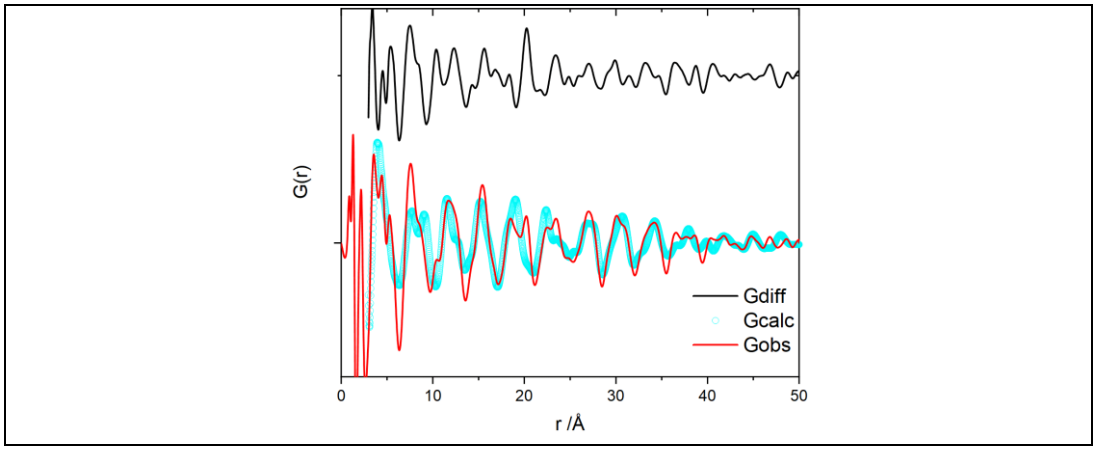
[54] Data collection by Dr. Chris Pask (Department of Chemistry, University of Leeds) and structure solutions were modelled by Dr. Pask and Dr. Anuradha Pallipurath (University of Leeds)

A high Z' is unusual and there is a possibility of finding a higher symmetry super-structure in this case [498]. An ADDSYM check using Platon software did not reveal a possible higher symmetry.

ImidHCl SCXRD 2: This ImidHCl sample also crystallised in the $P2_1/n$, monoclinic space group, with eight formula units in the asymmetric unit (Z'=8) (App. Figure 5b). All $ImidH^+$ were protonated. Non-hydrogen atoms were located in the Fourier Map and refined anisotropically, whilst the hydrogen atoms were placed in calculated positions and refined isotropically using a 'riding model'. Overall indexing is very high (95% of peaks are correctly indexed).

Four $ImidH^+$ ions form a slipped π-stack, while the other four units form T-shaped interactions with the stacked ions. Hydrogen bond lengths ($Cl^-\cdots$H-N) of 2.2-2.3 Å are observed. However, the motion of two the $ImidH^+$ cations overcomes the hydrogen bonding, as two $ImidH^+$ ions exhibit positional disorder. Each $ImidH^+$ ring was refined in two parts, each with 50% occupancy. Bond lengths in the rings were restrained to be the same using a SADI restraint.

X-ray total scattering data for solid ImidHCl was collected at beamline I15-1. The aim was to use the XPDF data to refine the crystal structure from the SCXRD analysis using PDFgui (Section 2.9.2.1) to acquire the molecular structure of the recrystallised ImidHCl (App. Figure 6). The structure did not refine to the XPDF data, possibly due to the disorder displayed by two $ImidH^+$ ions.



App. Figure 6 PDFgui refinement of ImidHCl

This poor refinement was discussed with the crystallographer who had collected the SCXRD data and a third data collection was carried out.

ImidHCl SCXRD 3: SCXRD data collection, as described in Section 2.5.2, was carried out to acquire a full sphere of data. Two datasets were collected using Cu K_α radiation (λ = 1.54184 Å) and Mo Kα radiation (0.7107 Å).

The diffraction layers were reconstructed, and the reciprocal lattices were reviewed. In both cases > 95% spots have been indexed to the unit cell. It was not clear what was going on with the structure. It was agreed that there may be some incommensurate scattering, but if it was present, it was very weak. Consequently, it is not trivial to treat. It was suggested that the high Z' (Z' = 8) may be a result of a possible incommensurate nature of the sample, crystal twinning, or the observed disordered behaviour.

Further study of the compound could be undertaken. The study of the crystal structure of imidazolium bromide [499] found that, like ImidHCl, it is unstable in air, as it is dissolved [499]. Imidazolium bromide was found to be highly disordered and exhibited first order phase transition at around 200 K. At 210 K (Phase 1), imidazolium bromide had a trigonal structure (R3 space group), and twinning was observed. In Phase 2 (120 K), the crystal structure was identified as monoclinic space group $P2_1/m$. Interestingly, and unlike ImidHCl, the $ImidH^+$ ions in imidazolium bromide appear to be stabilised at this temperature.

A.3.2 GuHCl

The recrystallised GuHCl sample was produced, as described in Section 2.2.2.

SCXRD analysis: GuHCl crystallises as a colourless compound in an orthorhombic cell. It was solved in the *Pbca* space group, as seen in the structures in CCDC, with one formula unit in the asymmetric unit. Non-hydrogen atoms were located in the Fourier Map and refined anisotropically, whilst hydrogen atoms were located in the Fourier Map and refined isotropically using a riding model.

Crystallographic Information File: The crystallographic information file (.cif) that was generated from the SCXRD analysis is presented is in App. Table 2.

A.3.2.1 Solid GuHCl structure refinement with PDFgui

X-ray total scattering data for solid GuHCl collected at beamline I15-1 (Section 2.5.4.1) was used to refine the .cif structure and acquire the molecular structure of the recrystallised GuHCl using PDFgui. The refinement is detailed in Section 2.9.2.1.

Crystal structure: The refinement of the GuHCl crystal structure (from SCXRD) to XPDF data was undertaken using PDFgui. Cell coordinates of the measured GuHCl compared with those of GuHCl structures deposited in the Cambridge Structural Database are in App. Table 3.

App. Table 2 crystallographic information file for GuHCl crystallised from *in situ* cooling crystallisation

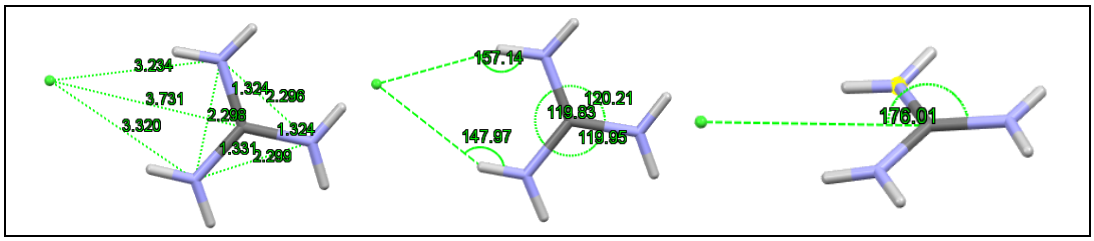
```
data_guhcl_1
_symmetry_cell_setting          orthorhombic
_symmetry_space_group_name_H-M   'P b c a'
_symmetry_Int_Tables_number      61
_space_group_name_Hall           '-P 2ac 2ab'
loop_
_symmetry_equiv_pos_site_id
_symmetry_equiv_pos_as_xyz
1 x,y,z
2 1/2-x,-y,1/2+z
3 -x,1/2+y,1/2-z
4 1/2+x,1/2-y,-z
5 -x,-y,-z
6 1/2+x,y,1/2-z
7 x,1/2-y,1/2+z
8 1/2-x,1/2+y,z
_cell_length_a          7.63070(10)
_cell_length_b          9.02640(10)
_cell_length_c          13.0417(2)
_cell_angle_alpha       90
_cell_angle_beta        90
_cell_angle_gamma       90
_cell_volume            898.283
loop_

_atom_site_label
_atom_site_type_symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
_atom_site_thermal_displace_type
N1 N 0.71024(16) 0.53316(13) 0.36904(8) 0.0270 Uani
H1A H 0.743(2) 0.5007(17) 0.3112(12) 0.0280 Uiso
H1B H 0.757(3) 0.4998(16) 0.4253(13) 0.0320 Uiso
C1 C 0.60918(15) 0.65312(13) 0.37410(8) 0.0205 Uani
N2 N 0.57014(14) 0.71169(12) 0.46434(8) 0.0256 Uani
H2A H 0.618(2) 0.6754(17) 0.5172(12) 0.0330 Uiso
H2B H 0.507(2) 0.7881(19) 0.4664(12) 0.0360 Uiso
N3 N 0.55074(14) 0.71518(12) 0.28866(8) 0.0249 Uani
H3A H 0.5726(19) 0.6741(16) 0.2329(12) 0.0290 Uiso
H3B H 0.478(2) 0.7844(18) 0.2950(12) 0.0340 Uiso
Cl1 Cl 0.68601(4) 0.48283(3) 0.11729(2) 0.0233 Uani
loop_
_atom_site_aniso_label
_atom_site_aniso_U_11
_atom_site_aniso_U_22
_atom_site_aniso_U_33
_atom_site_aniso_U_23
_atom_site_aniso_U_13
_atom_site_aniso_U_12
N1 0.0314(6) 0.0314(6) 0.0182(5) -0.0026(4) -0.0020(4) 0.0102(4)
C1 0.0190(6) 0.0231(6) 0.0195(5) -0.0008(4) -0.0003(4) -0.0029(4)
N2 0.0293(6) 0.0291(6) 0.0185(5) -0.0030(4) -0.0024(4) 0.0070(5)
N3 0.0305(6) 0.0262(5) 0.0181(5) -0.0002(4) 0.0015(4) 0.0069(5)
Cl1 0.0284(2) 0.02423(18) 0.01731(18) -0.00219(9) 0.00315(9) -0.00278(10)
#END
```

App. Table 3 Cell coordinates of known GuHCl structures

Structure name / *CCDC identifier	Measurement temp. /K	Edge length /Å			Internal angles / °		
		a	b	c	α	β	γ
GANIDC [170]*	Room Temp. (283-303)	9.184	13.039	7.765	90	90	90
GANIDC01 [171]*	100	7.61440	9.01530	13.0516	90	90	90
GANIDC02 [172]*	100	7.6043	9.0052	13.0393	90	90	90
Recrystallised GuHCl SCXRD	130	7.6307	9.0264	13.042	90	90	90
Recrystallised GuHCl ***structure refined with PDFgui***	290	7.7929	9.19815	13.0748	90	90	90

The atomic separations of the formula unit are shown in App. Figure 7 and Table 4.4.



App. Figure 7 Atomic separations and angles in GuHCl unit cell

Atomic pair interactions: The atomic separations in GuHCl up to 20 Å were identified with PDFgui. The data have been collated to atomic separations to ±0.1 Å, and the frequency of each interaction has been calculated (41,344 site-site interactions across 722 different separation distances). Interactions (< 6 Å) between C, N and Cl are included in App. Table 4.

App. Table 4 Atomic separations in solid GuHCl, up to 6.0 Å

Interaction		Å	Freq.
C	N	1.3	16
C	N	1.4	8
N	N	2.3	48
Cl	N	3.3	40
Cl	N	3.4	8
Cl	N	3.6	8
N	N	3.6	32
C	Cl	3.7	8
Cl	N	3.7	8
C	Cl	3.8	24
C	N	3.8	8
Cl	N	3.8	16
N	N	3.8	32
Cl	N	3.9	8
N	N	4.0	16
C	N	4.1	8
Cl	N	4.1	8
N	N	4.1	32
C	N	4.2	8
Cl	Cl	4.2	8

Interaction		Å	Freq.
N	N	4.2	8
C	N	4.3	8
N	N	4.3	32
C	N	4.5	8
C	N	4.6	8
N	N	4.6	16
C	C	4.7	8
C	Cl	4.7	8
C	N	4.7	8
Cl	Cl	4.7	16
C	Cl	4.8	16
N	N	4.8	24
C	N	4.9	32
N	N	5.0	48
C	C	5.1	32
Cl	N	5.1	16
Cl	Cl	5.2	16
Cl	N	5.2	16
N	N	5.2	48
C	N	5.3	8

Interaction		Å	Freq.
Cl	N	5.3	16
C	C	5.4	16
C	N	5.4	8
N	N	5.4	48
C	N	5.5	8
Cl	N	5.5	24
N	N	5.5	24
C	N	5.6	8
C	N	5.7	32
N	N	5.7	56
C	N	5.8	16
Cl	Cl	5.8	8
Cl	N	5.8	8
N	N	5.8	32
C	C	5.9	16
C	N	5.9	16
Cl	N	5.9	16
N	N	5.9	48
N	N	6.0	32

A.2 Helium pycnometer to measure density of GuHCl

As received GuHCl (1.3 g cm^{-3}) was recrystallised by cooling crystallisation in the jacketed vessel from a 9 M GuHCl aqueous solution. The sample was filtered and rinsed with chilled deionised water before being dried out in an oven (~60°C). A Thermo Pycnomatic ATC helium pycnometer (see Section 2.7) measured the density of a 1.6855 g sample of GuHCl. The sample density was 1.33848 g cm^{-3}, with a standard deviation of 0.00027 g cm^{-3} (0.01995 %).

A.3 Crystal16

Aqueous GuHCl solutions, prepared at 1.0, 2.0, 4.0 and 6.0 M (unit 1) and 7.0, 7.8, 9.3 and 9.8 M (unit 2), were heated and cooled between 5 and 75°C at cooling rates of 0.1, 0.2, 0.5, 0.75, 1, 2, 5 and 10 $K.min^{-1}$ (App. Figure 8). Each solution was stirred at 700 rpm throughout.

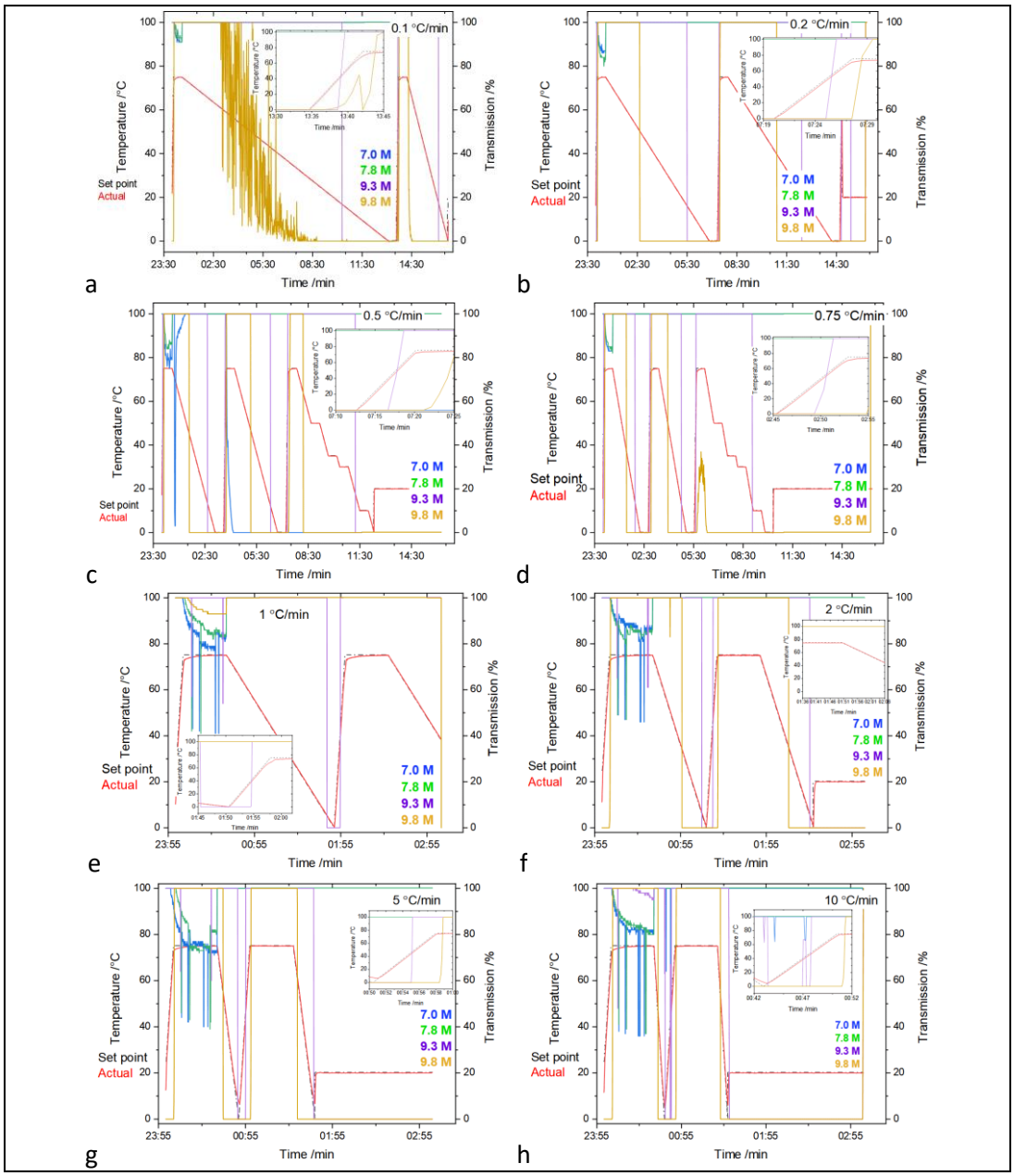
		Cooling rate /K.min^{-1}			
		0.1	0.2	0.5	0.75
Conc. /M	1.0	Unit 1 Run 1			
	2.0				
	4.0				
	6.0				

		Cooling rate /K.min^{-1}			
		0.1	0.2	0.5	0.75
Conc. /M	7.0	Unit 2 Run 1			
	7.8				
	9.3				
	9.8				

		Cooling rate /K.min^{-1}			
		1	2	5	10
Conc. /M	1.0	Unit 1 Run 2			
	2.0				
	4.0				
	6.0				

		Cooling rate /K.min^{-1}			
		1	2	5	10
Conc. /M	7.0	Unit 2 Run 2			
	7.8				
	9.3				
	9.8				

App. Figure 8 Configuration of Crystal16 runs



App. Figure 9 Crystal16 results

a-h: data for 7.0 (blue), 7.8 (green), 9.3 (purple) and 9.8 (orange) M solutions, with cooling rates of 0.1, 0.2, 0.5, 0.75, 1, 2, 5 and 10 Kmin^{-1}

Only the 9.3 and 9.8 M solutions were observed to consistently completely dissolve and crystallise (App. Figure 9). The 9.8 M solution dissolved at ~75°C, the maximum range for the temperature cycling in the Crystal16, and above the maximum temperature that could be safely achieved *in situ* using the water bath. The 9.3 M solution dissolved 46-59°C, depending on the cooling rate, and crystallised at 5-17°C, except at 10 $K.min^{-1}$, where crystallisation did not occur within the cooling cycle. As the actual desolvation and crystallisation temperatures will be different in the jacketed crystallising vessel with recirculation tubing, it was decided to run the recrystallisation with a 9 M solution.

Appendix B NAP-XPS analysis of Imid and $ImidH^+$

B.1 Overview

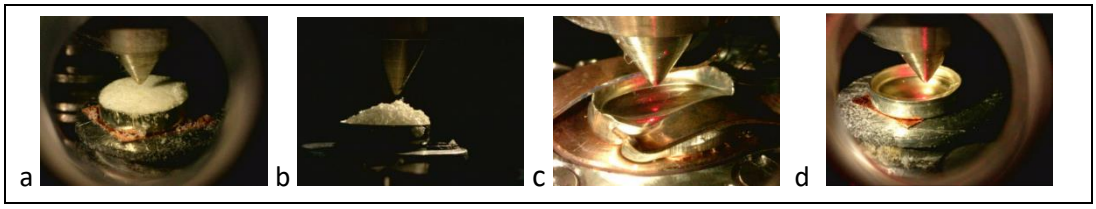
Aim: Calibration of the NAP-XP spectra of Imid and ImidHCl (dry and aqueous solution), required to inform fitting of the N K-edge and C K-edge XR spectra.

In addition to the collection of C, N and O 1s data for the solution samples, N and C 1s data were also collected for dry samples of each compound to assist with the peak fitting. The collection parameters for the XPS data collection are in App. Table 5.

App. Table 5 XPS collection parameters

SAMPLE HOLDER	Al DSC pans			
PRESSURE	11 mbar			
ENVIRONMENT	• Dry samples – 3 ml/min argon (Ar) • Solution samples – humid from large sample dish / deionised water in open glass bottle ($ImidH^+$)			
ENVIROESCA SETTINGS	Slit: 4:7x20 / C: grid, Mode: Fixed Analyzer Transmission, Excitation Energy: 1486.71 eV, Detector Voltage: 1600 V, Bias Voltage: 90.00 V			
SCAN SETTINGS	**Step /eV**	survey: 1 / elemental: 0.1 (0.15 aq. Imid)		
	Dwell time /s	survey: 0.1 / elemental: 0.25		
	Epass /eV	survey: 100 / elemental: 50		
NO. OF SCANS (SURVEY/ ELEMENTAL)				
Imid	**Aq. Imid**	**ImidHCl**	**Aq. ImidHCl**	
2 (start & end) / 16 (N, C), 4 (O, Cl2p, Ar)	2 (start & end) / 16 (N, C), O (9)	1 (start) / 4 (N, O, C1s Cl2s, Cl2p)	2 (start & end) / 16 (N, C), 9 (O), 4 (Cl2p, Ar)	

Images of the samples during measurement are shown in App. Figure 10. The aqueous ImidHCl sample has completely deliquesced to form a solution (unknown concentration). The structure of Imid and $ImidH^+$ are shown in Figure 1.8.

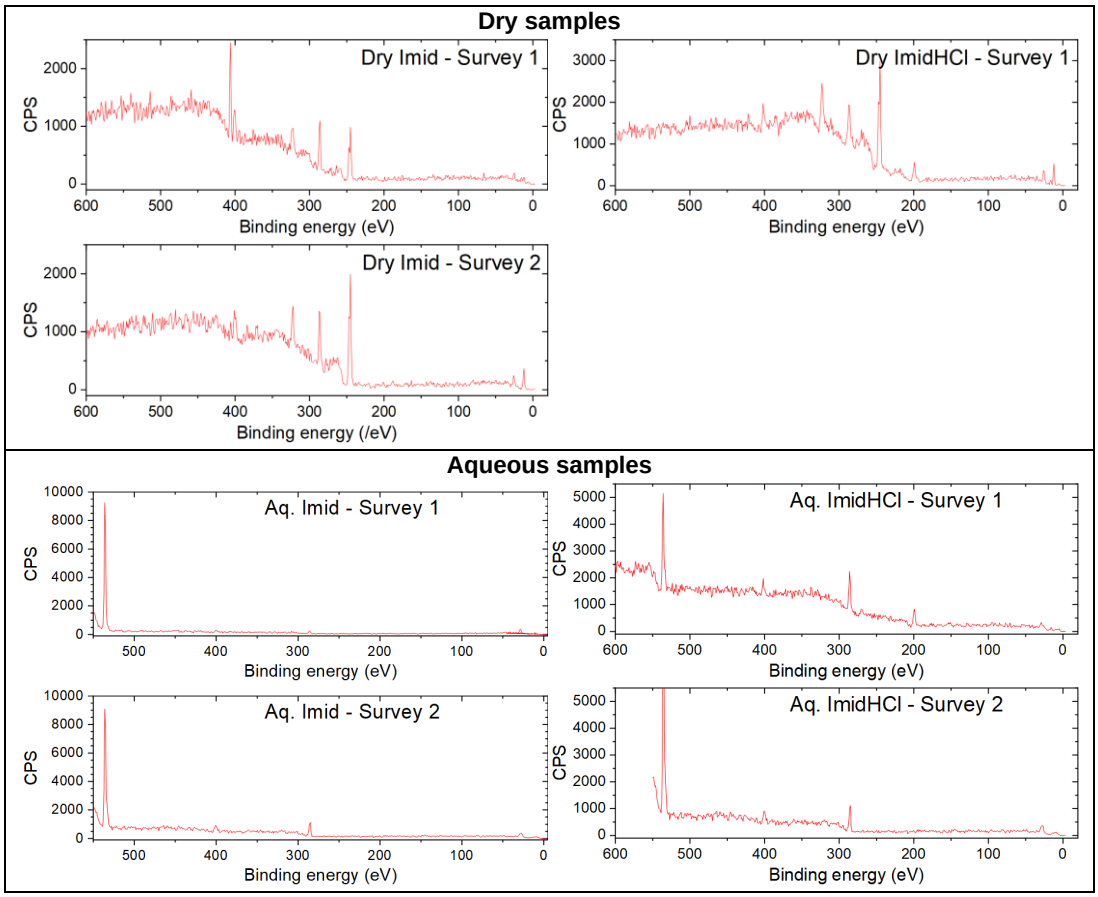


App. Figure 10 XPS samples: a) dry Imid b) dry ImidHCl c) aq. Imid d) aq. ImidHCl

(Aq. Imid from earlier measurement session.)

Charge compensation in the EnviroESCA is achieved by electron-ion pair formation in the surrounding gas that is present in the sample chamber, in this case argon (dry samples) and water vapour (solution samples).

To determine the quantity of adventitious carbon (C_{adv}) present in the NAP-XP spectra, quantitative elemental analysis of the survey scans (App. Figure 11) was undertaken to estimate the C present above the stoichiometric quantity expected based on the Imid or ImidHCl structures alone.



App. Figure 11 XPS survey scans for dry (top) and aqueous samples (bottom)

Top left: dry Imid, top right: dry ImidHCl: Bottom left: aqueous Imid, bottom right: aqueous ImidHCl: survey 1 prior to high-resolution scans and survey 2 after high-resolution scans

For imidazole (Imid), two N species (N1 and N3 in a 1:1 ratio), and two C species (C2, and C4&C5 in a 1:2 ratio) are expected. For imidazolium, one N species (N1&N3 are equivalent) and two C species (C2, and C4&C5 in a 1:2 ratio) are expected. This process assumes that no N contamination is present.

C_{adv} was fitted based on the excess carbon from the second survey scan, except for dry ImidHCl with only one survey scan (App. Figure 11). Initially this peak area was constrained, but once the other peaks were fitted, it was allowed to float, as the starting value was only an estimate.

App. Table 6 XPS survey scan C:N ratios

	Survey #	% carbon	% nitrogen	C:N ratio	C_{adv}:C_{imid}:N_{imid}		
Dry Imid	1	74.97	25.03	3.0	3.0	3	2
	2	77.50	22.50	3.4	3.9	3	2
Aq. Imid	1	60.51	39.49	1.5	0.1	3	2
	2	22.57	9.12	2.5	2.0	3	2
Dry ImidHCl	1	62.81	23.75	2.6	2.3	3	2
Aq. ImidHCl	1	82.61	17.39	4.8	6.5	3	2
	2	82.23	17.77	4.6271	6.255	3	2

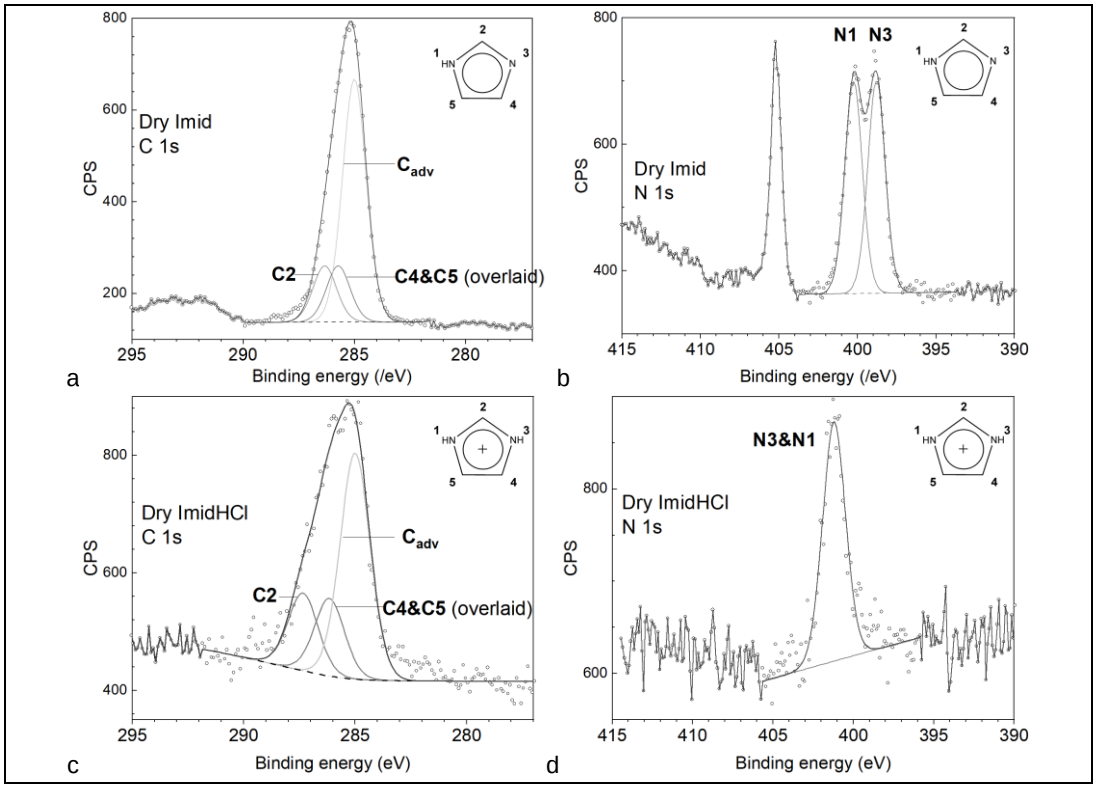
All spectra (N, C and O 1s) are fitted with GL(30) lineshapes (70% Gaussian, 30% Lorentzian using the Gaussian/Lorentzian product form), selected based on the resolution of the detector. The background is U 2 Tougaard, as it appeared to fit the background well in early data fitting, and so it was retained for consistency across all experimental spectra.

B.2 Results and discussion

The core-level C 1s and N 1s XP spectra and spectral features for dry Imid and ImidHCl are presented in App. Figure 12 and App. Table 7. The spectra and spectral features for aqueous Imid and ImidHCl are presented in Chapter 3.

B.2.1 Imidazole

The split between the N1 and N3 peaks in solid Imid (1.46 eV, App. Table 7) is similar to the IP separation in the previous NEXAFS and XRS studies of solid Imid, and slightly higher than the difference between the N1 & N3 E_B of solid Imid (1.35 eV (unpublished XPS data by authors of ref. [444]). There is a lower electron density around N3 than N1, as there is an increased electron at the N1 centre from the polarised N-H bond [504]. Therefore, the E_B of N3 is lower than that of N1 E_B.



App. Figure 12 XP spectra of a) C 1s dry Imid b) N 1s dry Imid c) C 1s dry ImidHCl and d) N 1s dry ImidHCl

Experimental data points (grey dots), background (dotted lines), fitted envelope (black line)

App. Table 7 XPS data for dry and aqueous Imid and ImidH$^+$

Compound	Transition	Peak energy /eV	ΔE /eV	
Dry Imid	N 1s (N3)	398.79	(N1-N3) =	***1.46***
	N 1s (N1)	400.25		
	C 1s (C4)	285.73	(C5-C4) =	0.00
	C 1s (C5)	285.73	(C2-C5) =	0.61
	C 1s (C2)	286.34	(C2-C4) =	***0.61***
	C 1s adventitious	*285.00*	*59.0% area*	
Dry ImidHCl	N1 & N3 1s	401.21	(N1-N3) =	***0.00***
	C 1s (C4)	286.15	(C5-C4) =	0.00
	C 1s (C5)	286.15	(C2-C5) =	1.18
	C 1s (C2)	287.33	(C2-C4) =	***1.18***
	C 1s adventitious	*285.00*	*55.6% area*	
Aqueous Imid	N 1s (N3)	399.30	(N1-N3) =	***1.65***
	N 1s (N1)	400.95		
	C 1s (C4)	285.44	(C5-C4) =	0.30
	C 1s (C5)	285.74	(C2-C5) =	0.87
	C 1s (C2)	286.61	(C2-C4) =	***1.17***
	C 1s adventitious	*285.00*	*45.6% area*	
Aqueous ImidHCl	N1 & N3 1s	401.57	(N1-N3) =	***0.00***
	C 1s (C4)	286.37	(C5-C4) =	0.00
	C 1s (C5)	286.37	(C2-C5) =	1.31
	C 1s (C2)	287.68	(C2-C4) =	***1.31***
	C 1s adventitious	*285.00*	*65.3% area*	

Previous NEXAFS studies of solid Imid have shown that that the energy separation of $\Delta E_{(N1-N3)}$ 1s→1π* transition peaks were found to be roughly equivalent to the separation of the IPs

and therefore E_B [159]. These values are shown to be 1.425 eV from XRS [67] and 1.5 eV from NEXAFS [159]).

Hydrogen bonds between the Imid molecules (chains) when in solid form or in solution, or between Imid and water when solvated, change the electron density distribution. As C4 & C5 and N3 participate in the aromatic ring, their electron density will be redistributed [67]. The ΔE_B of the C species in solid Imid was found to be lower in the NAP-XPS study than had been found in the XRS study.

App. Table 8 Dry and aqueous Imid E_B data (right-hand column). NEXAFS [159] **and XRS** [67] **IP positions, and liquid-jet XPS** [160] **and NAP-XPS E_B data**

Note: NEXAFS energy scale not calibrated

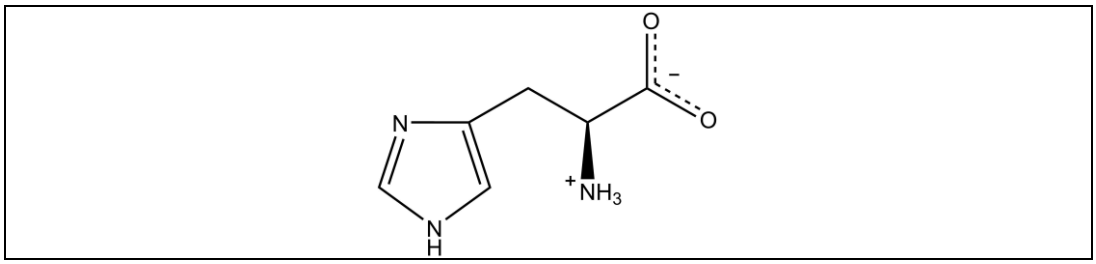
Nitrogen - Solid				
	NEXAFS IP /eV [159]	**XRS IP /eV** [67]		**NAP-XPS E_B /eV**
N1	404.7	404.216		400.25
N3	403.2	402.792		398.79
ΔE	*1.5*	*1.424*		*1.46*
Nitrogen - Solution				
		XRS IP /eV [67]	**Liquid-jet XPS E_B /eV** [160]	**NAP-XPS (aqueous) E_B /eV**
N1		404.186	405.6	400.95
N3		402.591	403.9	399.30
ΔE		*1.595*	*1.7*	*1.65*
Carbon - Solid				
		XRS IP /eV [67]		**NAP-XPS E_B /eV**
C2		290.887		286.34
C5		290.065		285.73
C4		289.424		285.73
ΔE (C2-C5)		*0.822*		*0.61*
ΔE (C5-C4)		*0.641*		*0.00*
ΔE (C2-C4)		*1.463*		*0.61*
Carbon - Solution				
		XRS IP /eV [67]	**Liquid-jet XPS E_B /eV** [160]	**NAP-XPS (aqueous) E_B /eV**
C2		290.697	291.7	286.61
C5		289.897	290.9	285.74
C4		289.497	290.5	285.44
ΔE (C2-C5)		*0.800*	*0.8*	*0.87*
ΔE (C5-C4)		*0.400*	*0.4*	*0.30*
ΔE (C2-C4)		*1.200*	*1.2*	*1.17*

The NEXAFS study reported that the intermolecular N-H···N bonding between Imid in the solid structure causes electron density redistribution between the N1 and N3 centres due to the weakening N1-H bond [159]. Therefore, it is reasonable to expect a reduction in the difference in the electronic structure of C4 and C5 as well, although not necessarily to the extent that the E_B of C4 and C5 would be equivalent.

In solution, water does not bind strongly to the C atoms (a recent NDIS/EPSR study demonstrated that the hydrogen bonding between water and the hydrophobic C moieties was highly improbable [223]) so changes to the electron density and therefore the E_B will be as a result of changes to the hydrogen bonding via the N centres. The ΔE_B of the C moieties are comparable to those found in the liquid-jet XPS study [160].

In the gas-phase molecule, it would be expected that the C4 and C5 are inequivalent with a small ΔE, as each are bonded to different N species in the heterocycle—C4 is bonded to the unprotonated N3, C5 to N1 (-NH) (Figure 1.8).

Imid forms the side-chain of histidine (App. Figure 13). In an XPS study of crystalline histidine [505], $\Delta E_{B(N1-N3)}$ = 1.6 eV (App. Table 9), which is comparable to the solid Imid data (1.46 eV, App. Table 7). The solid Imid $\Delta E_{B(N1-N3)}$ also compares well to that in the Imid ring found on theophylline (1.4 eV) [446].



App. Figure 13 Schematic of histidine molecule: zwitterionic form, as found in solid form and aqueous solution

E_B(N3) in solid histidine (398.8 eV) is equal to that of N3 in solid Imid. The hydrogen bonding via the N3 in the histidine solid crystal is with the -NH_3 functional group, and the hydrogen bonding via the N1 is with both C and O in the carboxylic group (-CO), meaning that the electron density will be different to the solid Imid, changing $\Delta E_{B(N3-N1)}$.

App. Table 9 XPS measurements of species found in the Imid side chain of histidine: ref. [505]

Species	Binding energy /eV	Comment	NAP-XPS E_B /eV
N=C-NH	398.8	Equivalent to N3	398.79
N=C-NH	400.4	Equivalent to N1	400.25
N=C-NH	286.7	Equivalent to C2	286.34
C=C-N	285.6	Equivalent to C 4 & C5 in imidazole, although in the side chain, C4 is linked to the aliphatic chain	285.73

The histidine $\Delta E_{B(C2-C5)}$ = 1.1 eV is greater than that found in dry Imid (0.61 eV). This difference could be due to the substitution effect of the aliphatic chain on histidine, or the different bonding in histidine (the Imid side-chains of histidine do not form the chains seen in solid Imid). The E_B for C atoms in theophylline were not determined.

B.2.2 Imidazole hydrochloride

As expected, the protonation of N3 in ImidHCl has resulted in the pseudoequivalence of the N moieties, resulting in a symmetrical ion. In the solid and aqueous ImidHCl, the E_B of N3 and N1 are equal (App. Table 10). The E_B is slightly higher in the aqueous sample (0.36 eV).

SCXRD analysis (Appendix A.3.1) showed that the two -NH in $ImidH^+$ interact via Cl^- ions in the solid structure. Cl^- appears to be positioned further from N1 and N3 [N-H···Cl^-] than the water in the explicit model of solvated $ImidH^+$ [N-H···O]. The hydrogen bonding of $ImidH^+$ with the water may reduce the electron density around the N more than the Cl^- does in the solid ImidHCl, resulting in a reduction in the N1&N3 electron density and hence a higher E_B. The presence of the Cl^- ions in solution in addition to the water molecules may also result in increased E_B.

App. Table 10 Dry and aqueous ImidHCl E_B data (right-hand column). Liquid-jet XPS [160] **and NAP-XPS E_B data**

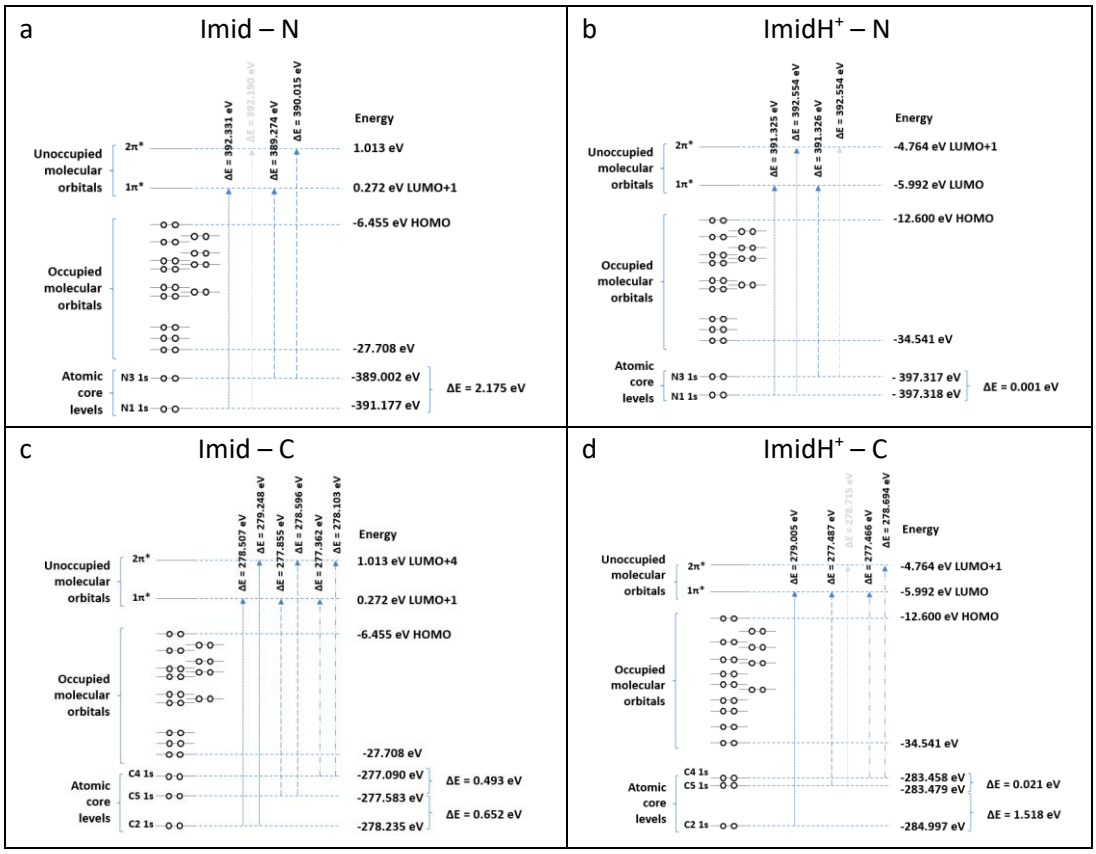
Nitrogen - Solid		
		NAP-XPS E_B /eV
N1 & N3		401.21
ΔE		*0.00*
Nitrogen - Solution		
	Liquid-jet XPS E_B /eV [160]	**NAP-XPS (aqueous) E_B /eV**
N1 & N3	406.6	401.57
ΔE	*0.0*	*0.00*
Carbon - Solid		
		NAP-XPS E_B /eV
C2		287.33
C5		286.15
C4		286.15
ΔE (C2-C5)		*1.18*
ΔE (C5-C4)		*0.00*
ΔE (C2-C4)		*1.18*
Carbon - Solution		
	Liquid-jet XPS E_B /eV [160]	**NAP-XPS (aqueous) E_B /eV**
C2	292.4	287.68
C5	291.2	286.37
C4	291.2	286.37
ΔE (C2-C5)	*1.2*	*1.31*
ΔE (C5-C4)	*0.0*	*0.00*
ΔE (C2-C4)	*1.2*	*1.31*

As expected, the N 1s E_B in ImidHCl will be higher than for the neutral Imid. A study of 3,5-diaminobenzoic acid showed that the protonation of N (-NH_2 to -NH_3^+) increases the E_B of the protonated N by 2.4 eV (399.5 eV to 401.9 eV) [434]. However, in the case of 3,5-diaminobenzoic acid, the protonated N is not part of the aromatic ring. As such, charge delocalisation does not occur, so the -$\underline{N}H_2$ E_B peak in 3,5-diaminobenzoic acid does not change position when protonation takes place. The E_B of the protonated N moiety (N3)

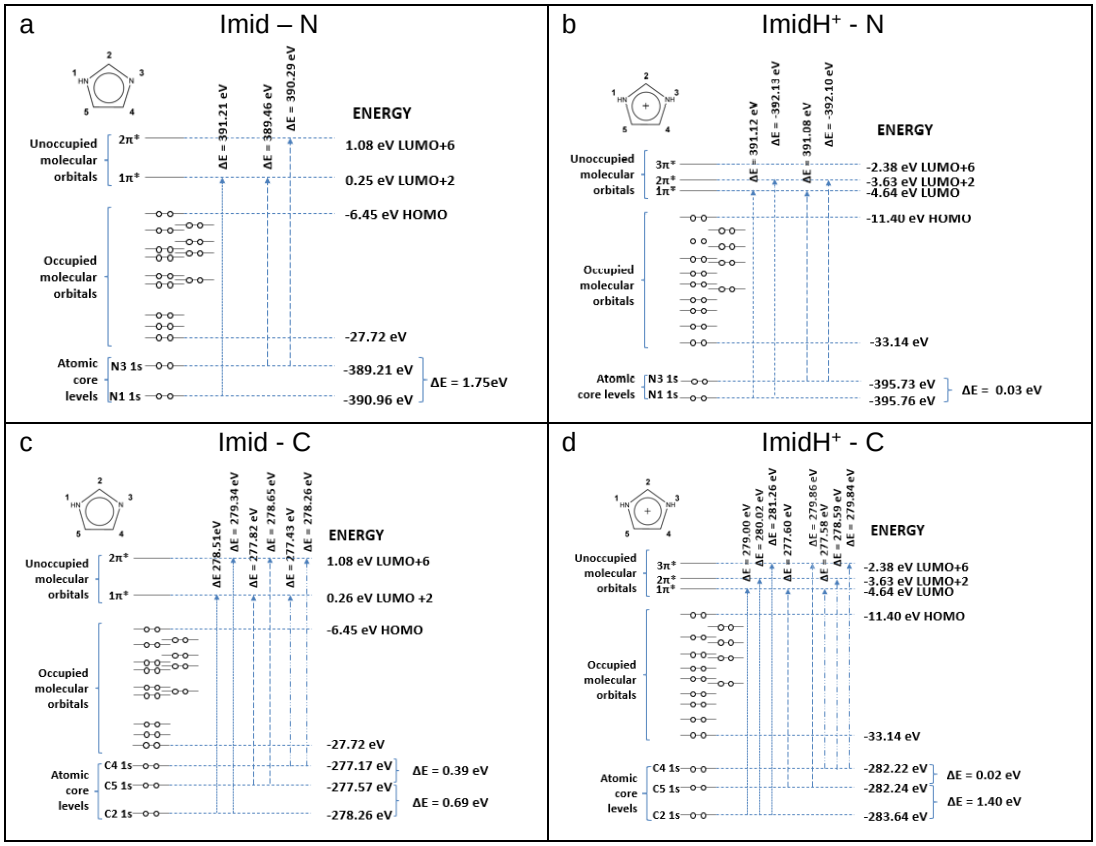
increases (+2.42 eV) in $ImidH^+$, and as the charge will delocalise, the E_B of the N1 moiety also increases (+0.96 eV). The three C moieties also feel the effect of protonation of the ring thus their E_B also increase in ImidHCl compared to Imid. As discussed elsewhere, it would be interesting to undertake further DFT calculations to study the effect of the anion in solution on the $ImidH^+$ ion, as it was not considered in the explicit or implicit solvation models.

Appendix C DFT and TDDFT of solid imidazole

C.1 Molecular orbitals: ground state Imid & $ImidH^+$



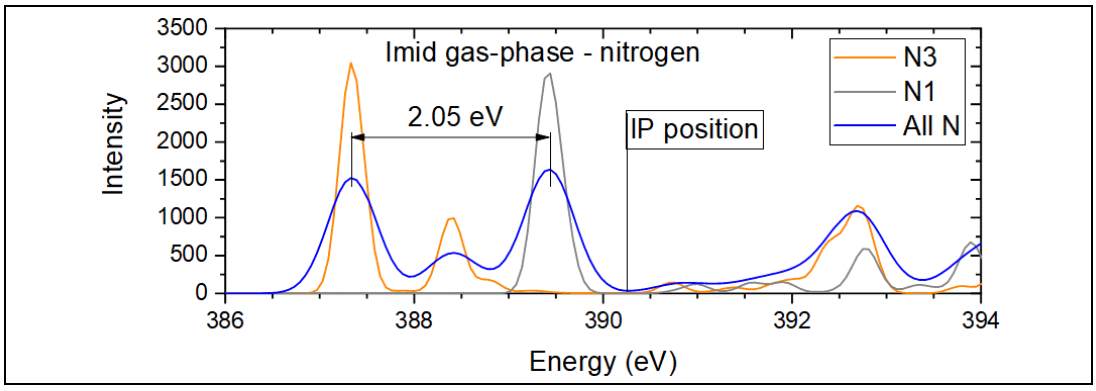
App. Figure 14 Gas-phase imidazole (left) and imidazolium (right) molecular orbital energies for nitrogen and carbon, with the electron densities of the two lowest unoccupied π* molecular orbitals. (Occupied MOs represent a schematic illustration only.)



App. Figure 15 Explicitly-solvated imidazole (left) and imidazolium (right) for nitrogen and carbon, with the electron densities of the lowest unoccupied π* molecular orbitals (two for Imid, three for ImidH⁺). (Occupied MOs represent a schematic illustration only.)

C.2 TDDFT calculations

Gas phase Imid and ImidH⁺

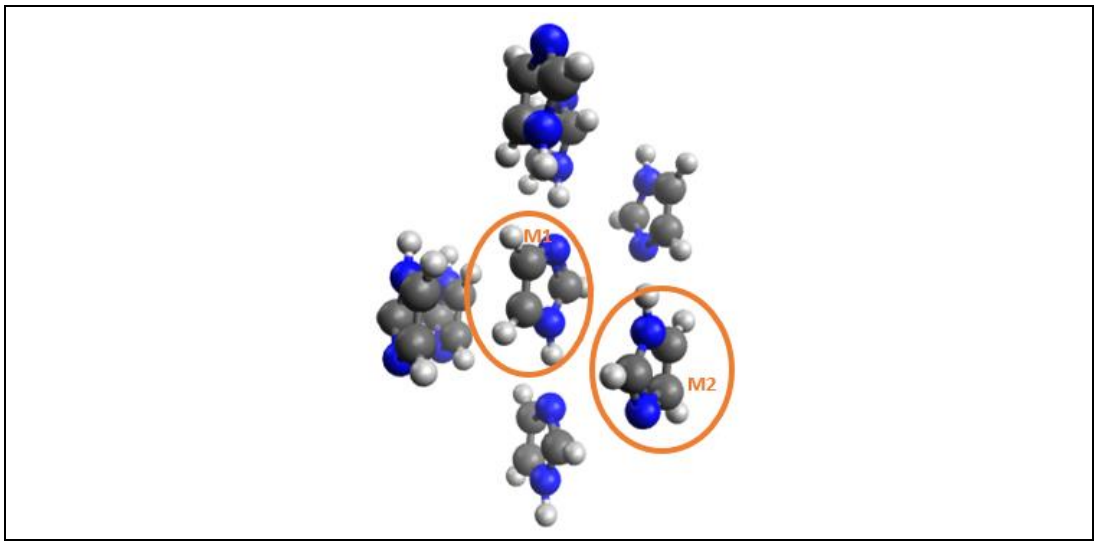


App. Figure 16 N1 and N3 π* transitions for gas-phase Imid

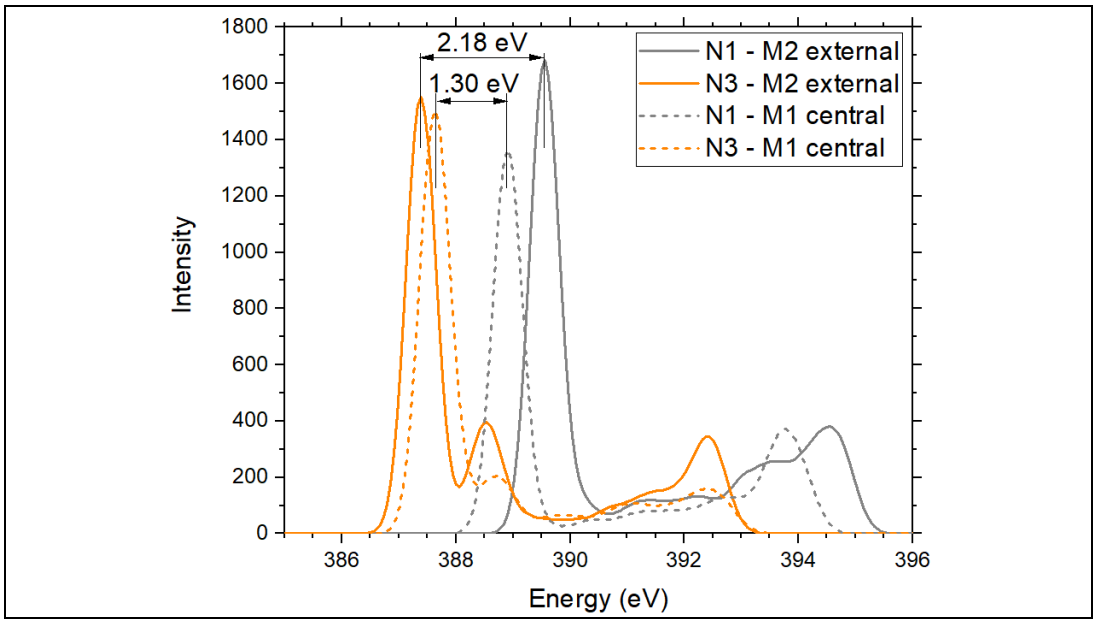
S**olid Imid cluster:** To illustrate the sensitivity of the electron transition energy of the probed atom to its environment, TDDFT calculations have been carried out for the N atoms of a central molecule (M1) and external molecule (M2) in Imid.

App. Table 11 TDDFT 1s→1π* calculated energies for N atoms in Imid

	Atom	1s→1 π* transition energy /eV	ΔE (N1-N3)/eV
Central molecule (M1)	N1	388.90	
	N3	387.60	1.30
External molecule (M2)	N1	389.56	
	N3	387.38	2.18



App. Figure 17 Solid cluster model of imidazole (IMAZOL04 from CCDC, ref. [138]**)**



App. Figure 18 N1 and N3 π* transitions for central (dotted line) and external (solid line) Imid molecules

Appendix D GuHCl cooling crystallisation XPDF pattern and EPSR analysis

The X-ray total scattering data used for the XPDF pattern analysis was initially processed by subtracting a background of water in Kapton. Due to the low ratio of water molecules to ions in the sample, it is likely that most water is used for solvating the ions, so the number of water-water interactions may be limited. Therefore the analysis was repeated for Run 1 of the cooling crystallisation experiment using an empty Kapton background (Section 4.3.2.1).

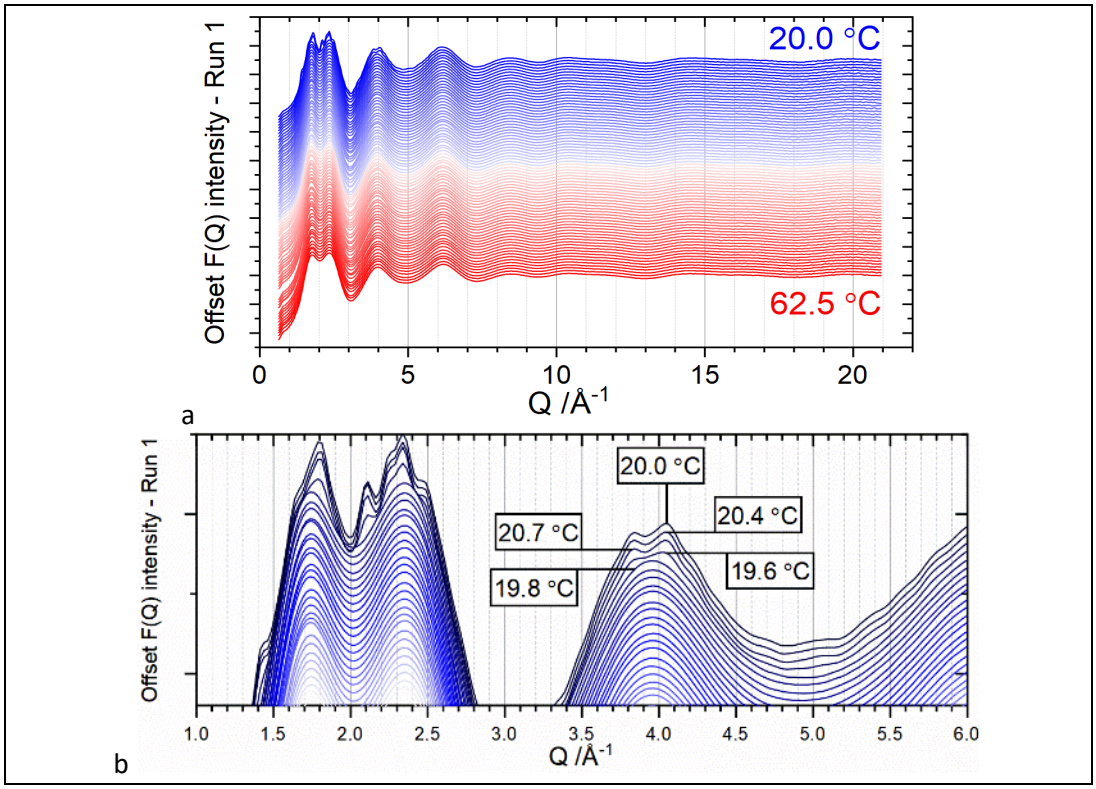
The original analysis showed that the clustering of the data with principal component analysis resulted in similar stage transitions in cooling crystallisation runs 1 and 2. This suggests that if the X-ray total scattering data from cooling crystallisation Run 2 were processed with the empty Kapton background, the clustering in Run 2 may be similar to that found for Run 1. It is the clustering from the PC analysis of these XPDF patterns that was used to group the EPSR simulations.

The original analyses of cooling crystallisation Runs 1 and 2 and the reheating solution showing GuHCl dissolution, where a water in Kapton background was used in the data processing, are presented here (D.1.1-D.1.3). Rerunning the data processing of the Run 2 of the cooling crystallisation and the dissolution process using an empty Kapton background is proposed in the Further Work section (6.2).

D.1 XPDF pattern analysis

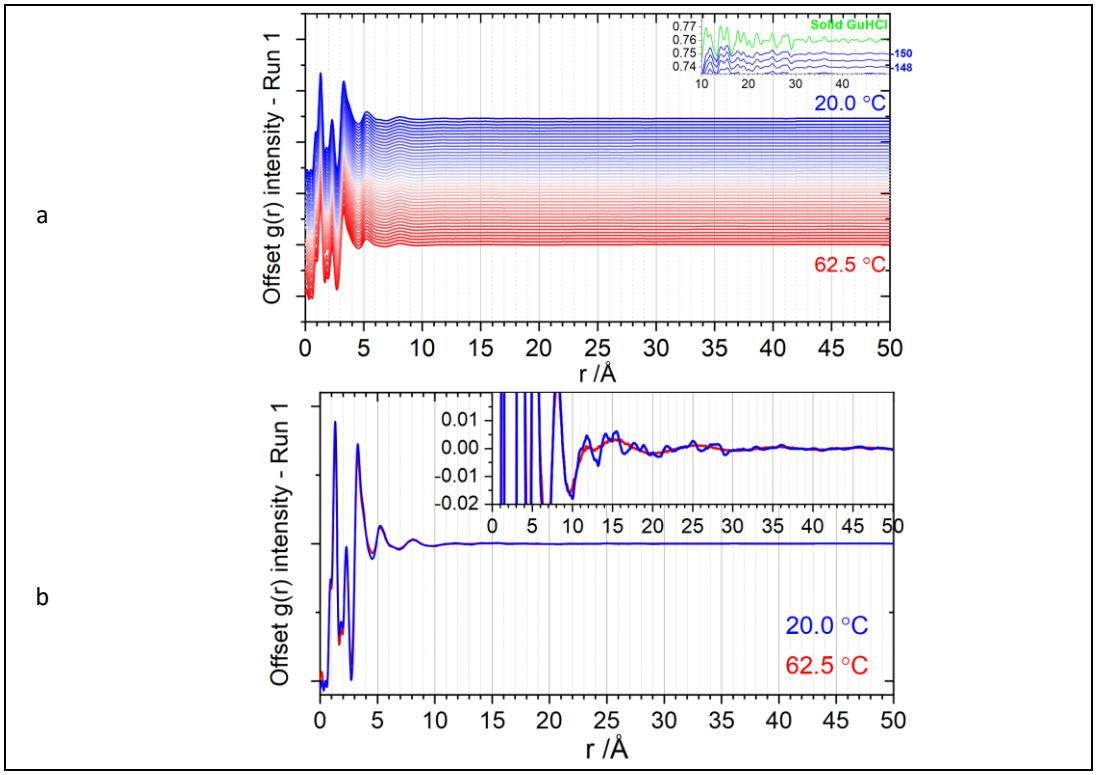
D.1.1 Cooling crystallisation: Run 1

The final four frames of the F(Q) indicate the presence of long-range structure (App. Figure 20a&b). The first principal peaks in the GuHCl solution are at 1.8 $Å^{-1}$ and 2.4 $Å^{-1}$, followed by peaks at 4.0 $Å^{-1}$, 6.2 $Å^{-1}$ and 8.4 $Å^{-1}$.



App. Figure 19 Run 1: a) offset alternate F(Q) and b) offset low-Q F(Q)

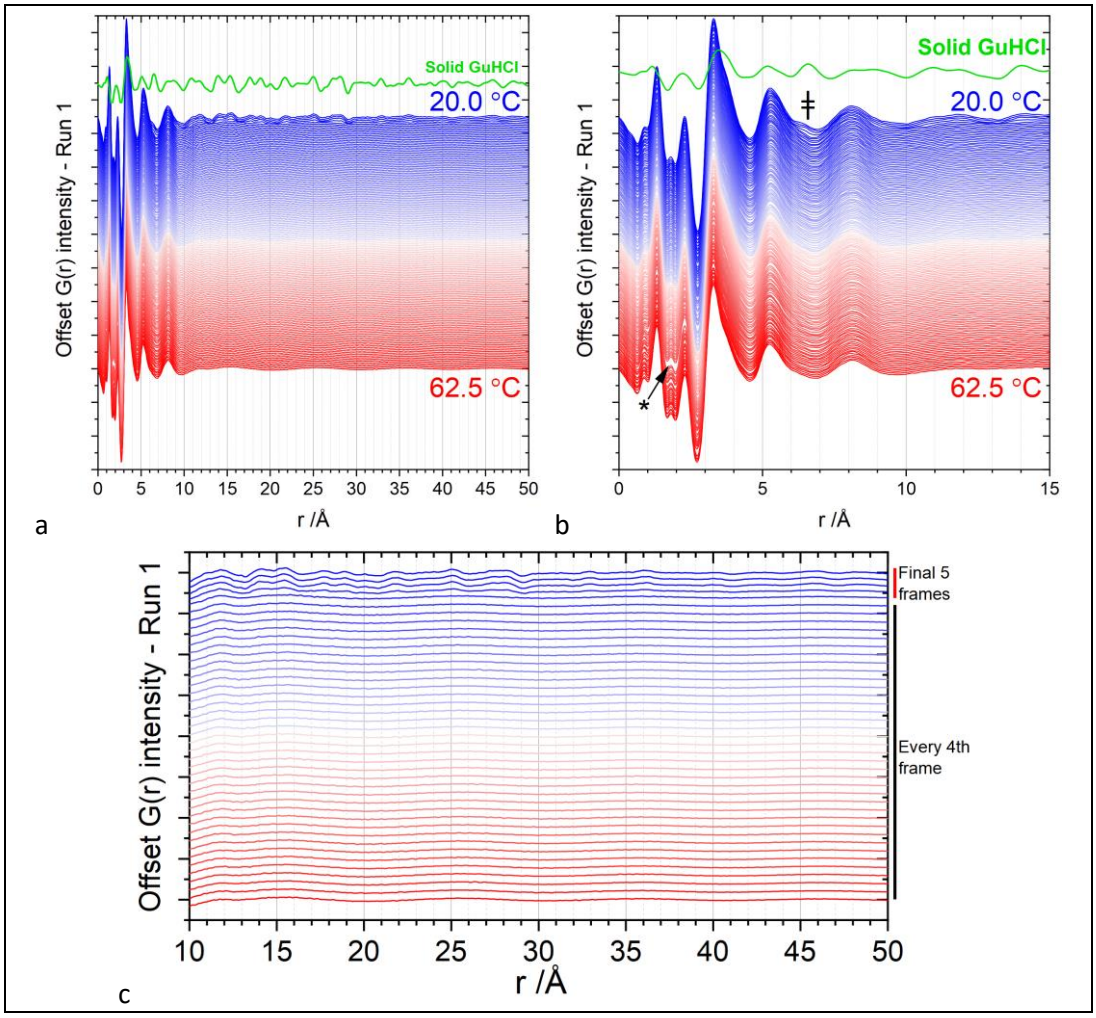
Long-range order can be seen emerging up to 50 Å in the in the final frames of the g(r) (App. Figure 20a). The difference between the first and last patterns is shown in App. Figure 20b.



App. Figure 20 Run 1: a) offset g(r) (0-50 Å) and b) g(r) comparison of the first and last XPDF patterns

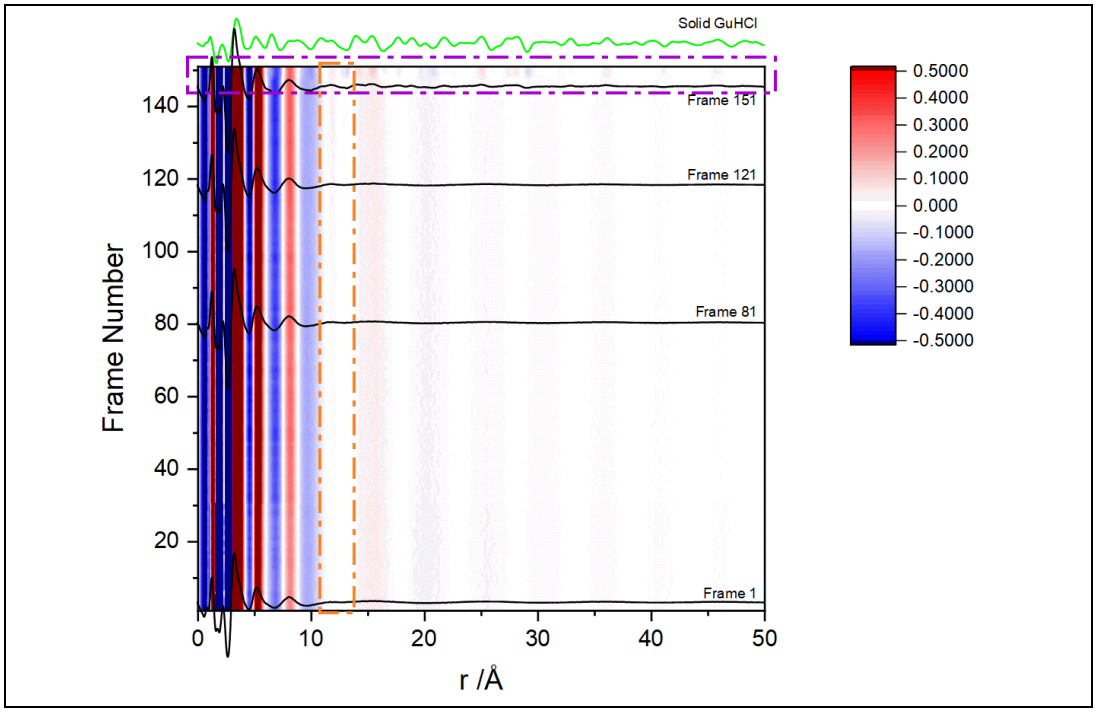
Some differences in the G(r) can be easily seen more easily by eye, for example long-range order can be seen in the top four G(r) XPDF patterns in App. Figure 21Figure 4.15a&b. There is a peak forming at 6.5 Å (App. Figure 21b, ‡) and a change is also observed at 1.8 Å (App. Figure 21b,*). Before the final four frames, no evidence of long-range interactions in the solution is observed.

App. Figure 21c shows the long-range order of the XPDF patterns more clearly, and illustrates that, aside from an increased intensity around 12 Å, there are no significant features indicating molecular interaction of GuHCl in this range until the final four frames of data.



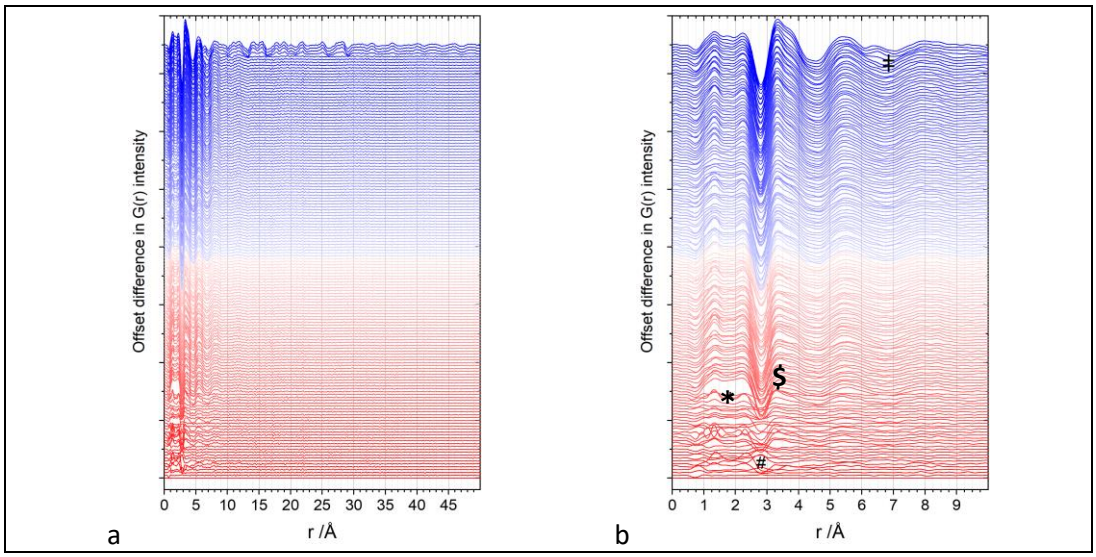
App. Figure 21 Run 1: a) offset G(r) (0-50 Å) b) offset G(r) data (0-15 Å) and c) every fourth G(r), 10-50 Å

The emergence of long-range structure of GuHCl, matching the solid GuHCl XPDF pattern in green, can be seen in the final few frames of the contour plot of the G(r) (highlighted with purple box, App. Figure 22). There in an increasing intensity at ~12 Å (based on the intensity parameters set for the contour plot, orange box).



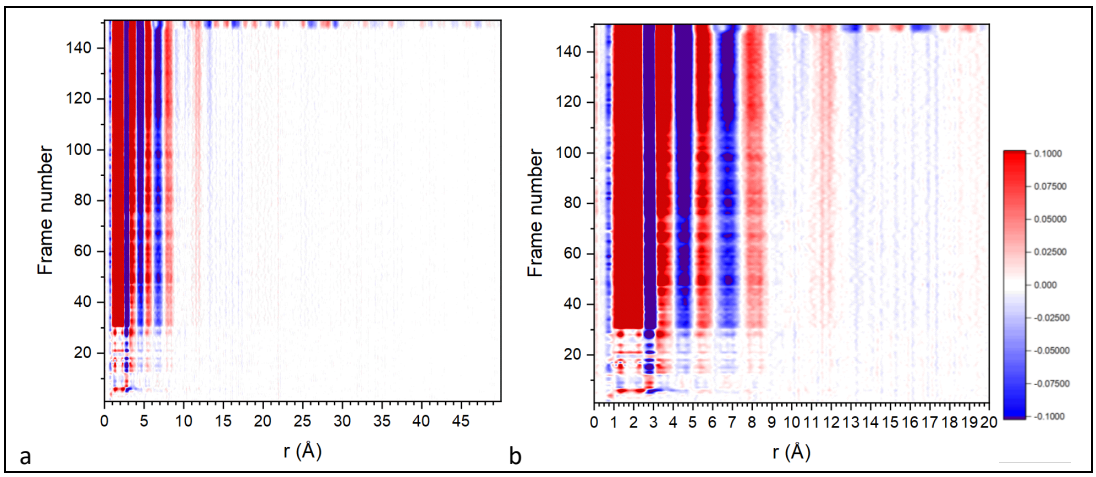
App. Figure 22 Run 1: G(r) contour plot 0-50 Å

The intensity changes are seen in the G(r) difference plots (G(r) - Frame 0 G(r) (highest temperature)) in App. Figure 23a&b. Aside from the emergence of long-range order in the final four frames, the obvious differences appear to be for short-range correlations, indicated by * (1.8 Å), # (2.8 Å) and $ (3.2 Å) in App. Figure 23b, where features become more defined after ~30 frames (55.5°C), and ‡ (~6.5 Å), where a feature forms upon crystallisation.



App. Figure 23 Run 1 - G(r) difference patterns a) 0-50 Å and b) 0-10 Å

Contour plots of the difference plots (App. Figure 24) further highlight that the appearance of long-range order occurs in the final four frames only. From frame 30 (~55.5°C), there are changes in the medium-range order, ~5-15 Å (App. Figure 24).

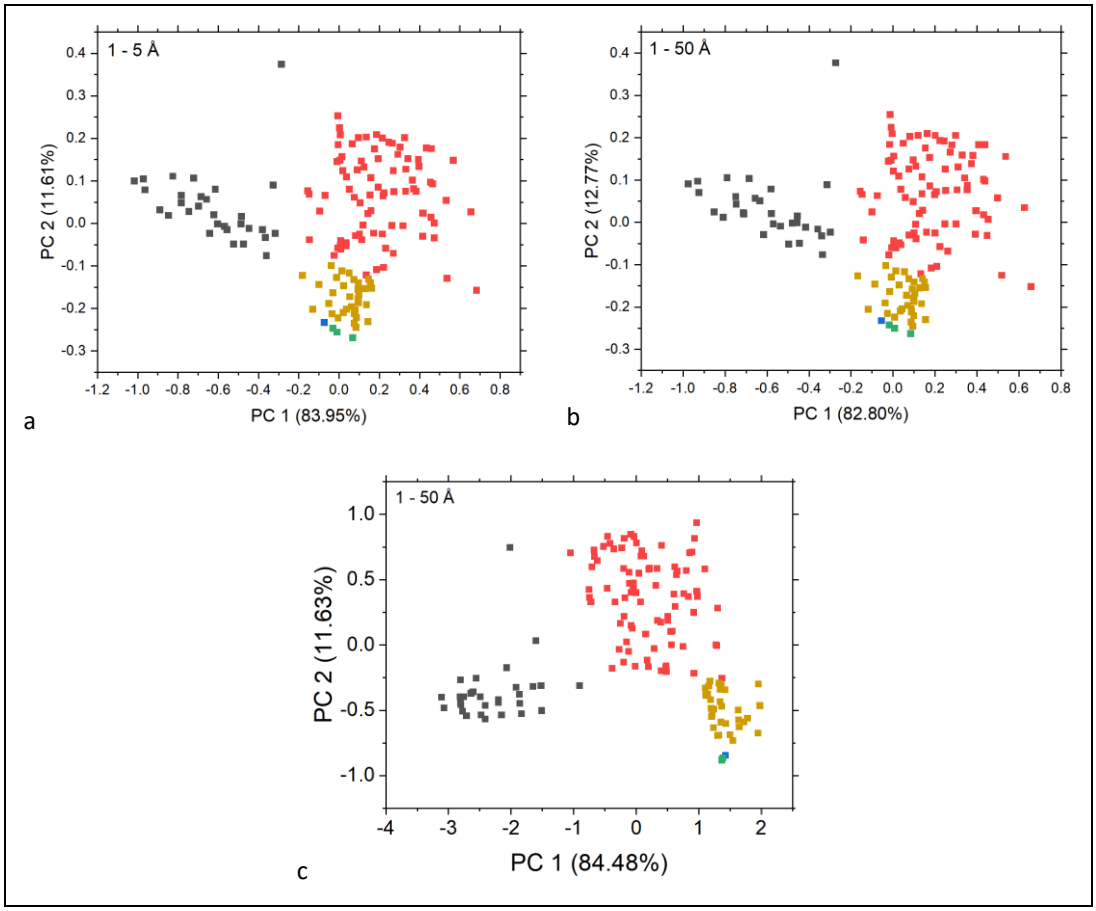


App. Figure 24 Run 1 – G(r) difference contour plot set between +0.1 (red) and -0.1 (blue): a) 0-50 Å and b) 0-20 Å

Principal component analysis (PCA): PCA was carried out as described in Section 4.3.2.1.1, using covariance matrix PCA in OriginLab 2019.

App. Figure 25a&b shows PCA scores plots for the XPDF patterns (g(r)). As features at low-r (< 1 Å) can be spurious and non-physical [291], PCA analysis from 1 A has been carried out. The distribution of points is similar in both, suggesting that the changes in structures at low-r are responsible for variation in the XPDF patterns. Used alongside the PCA of G(r) (App. Figure 25c), the frames have been grouped by eye, with three solution stages (black, red and gold) before transition to crystallisation (blue) and the visible presences of crystals in the solution (green).

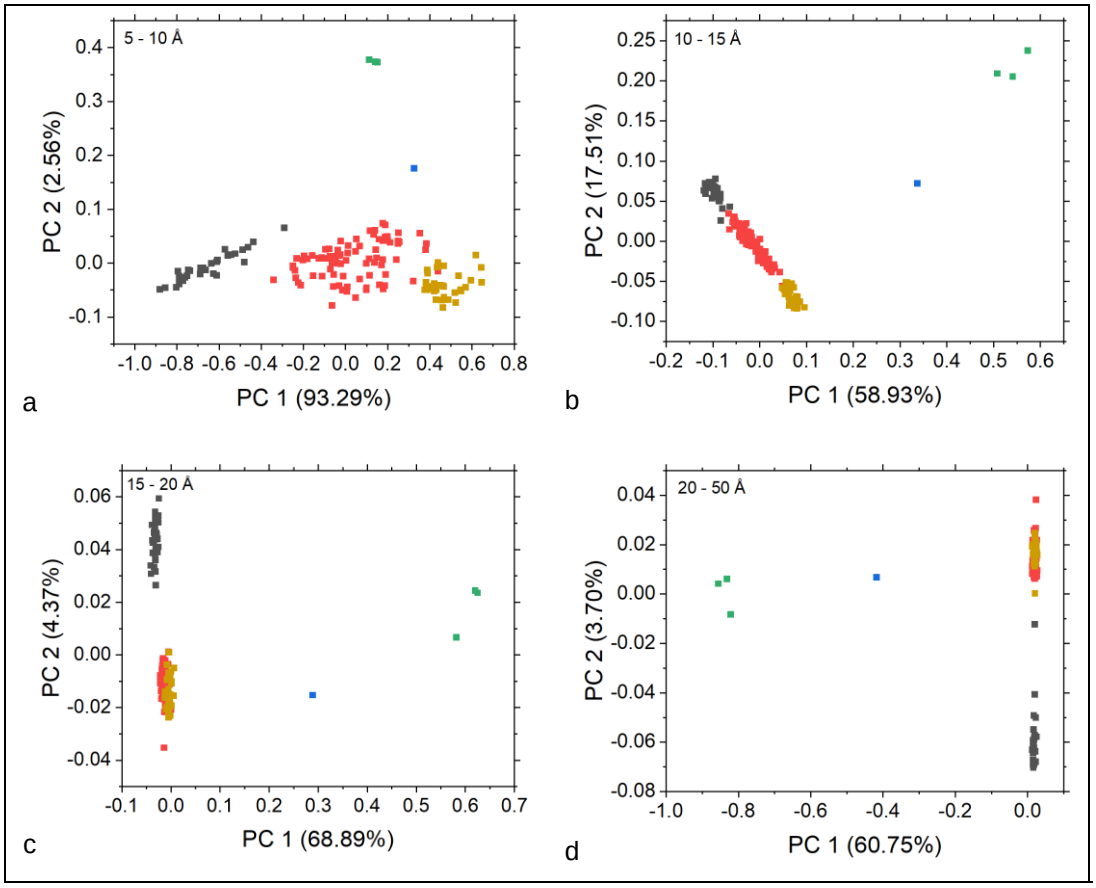
Across all analysed length ranges, the first 30 frames are grouped together, indicating that there is a difference in the patterns in the range 1–5 Å. As features at low-r (below 1 Å) can be spurious and non-physical [291], removing this range from the PCA further separate the frames. PCA of the G(r) reveals further clustering (Figure 4.20) and can be used to separate the frames into groups, which suggest that there are three stages (highlighted black, red and gold) before transition to crystallisation (blue) and the visible presences of crystals in the solution (green).



App. Figure 25 PCA run 1: XPDF score plots

PC1 v PC2 of g(r) for a) 1–5 Å and b) 1–50 Å and c) G(r) for 1–50 Å

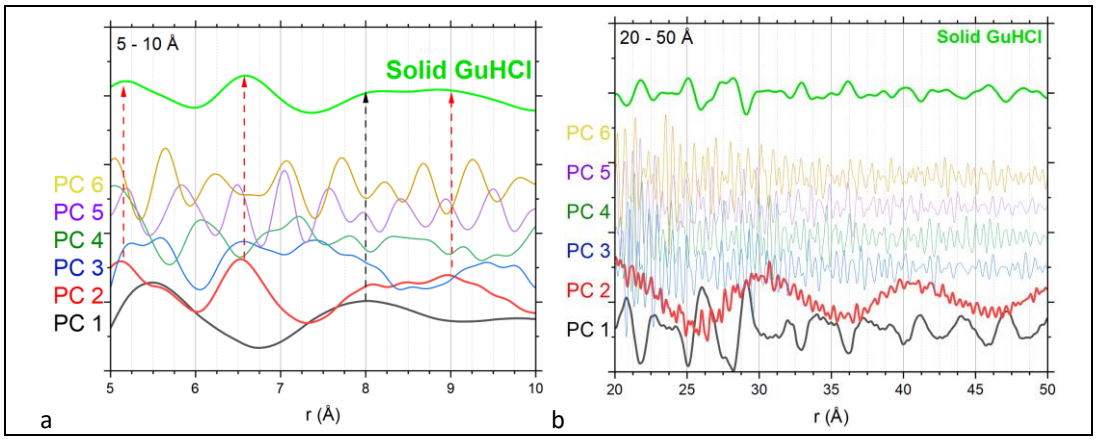
Further PCA score plots across other ranges confirm the clustering (App. Figure 26a–d) and show the separation of the scattering patterns of the solvated solute from the scattering patterns where crystalline materials are present (frames 147-150). Stage 1 (black) XPDF patterns appear to be differentiated from stages 2 and 3 in the range 5–10 Å, and stage 3 points are distinct from stages 1 and 2 in the range 10–15 Å.



App. Figure 26 PCA scores plots for run 1 G(r)

PC1 v PC2 for a) 5–10 Å b) 10–15 Å c) 15–20 Å and d) 20–50 Å

The weightings plots (App. Figure 27) show the extent to which each variable (r-values in this case) influences the PC. For the range 5–10 Å, PC 1 and 2 contain features from the solid structure of GuHCl. In the high-r region (20–50 Å), PC1 dominates when compared to the solid XPDF pattern of solid GuHCl.



App. Figure 27 Run 1 G(r) PCA loadings plots for a) 5-10 Å and b) 20-50 Å

The groupings (App. Table 12) have been applied to the G(r) difference patterns (App. Figure 28). The low-r difference in the G(r), peaking at 1.3 Å and 2.2 Å suggest that there may be a

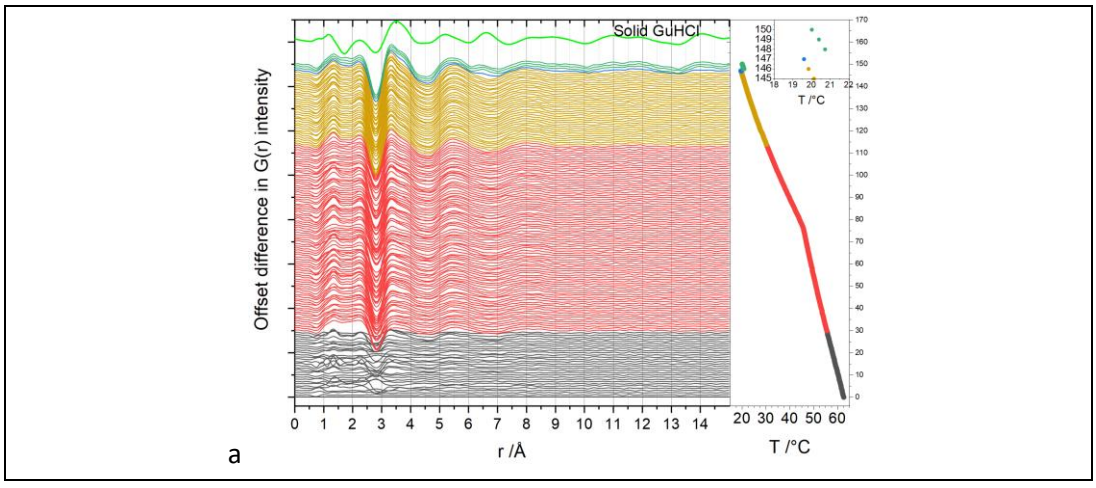
change in hydrogen bonding between stages 1 and 2. Interactions of these bond lengths may be associated with the hydrogen bonding, for example between Gdm^+ or Cl^- and water, or even between water molecules if the water structure in the solution is different to the pure water background.

There is a reduction in interaction intensity ~2.9 Å and an increase at 5.5 Å. The differences for the stage 2 to 3 transition are less clear. The increased intensity signalling an interaction at 6.5 Å is clear between stage 3 and the transition stage, and longer-range features are seen in the crystallisation stage. The crystallisation exotherm associated with nucleation is aligned with the transition stage.

App. Table 12 PCA-derived stages for Run 1

	Frame numbers	Temperature /°C	Average cooling rate /°C min^{-1}
Stage 1	0-29	62.5–55.7	0.47
Stage 2	30-113	55.5–30.5	0.60
Stage 3	114-146	30.1–19.8	0.64
Transition	147	19.6	-2.26
Crystallisation	148-150	20.7–20.0	0.47

Medium-range local structure changes in the G(r) difference plot are observed at 3.3, 3.8, 5.5 and 8 Å (increased intensity) and 4.6 and 6.8 Å (reduced intensity) at the onset of stage 2. Beyond 12 Å, there are no features indicating intermolecular interactions until frame 146 (transition), indicating that pre-nucleation clusters > 12 Å are not identified at higher temperatures in the MSZ. It is acknowledged that the time resolution of the data collection, with 30 s XPDF scattering frames, means that there could be a series of structural changes during the transition stage. However, this study also shows the potential to capture these changes with higher-brilliance X-rays.

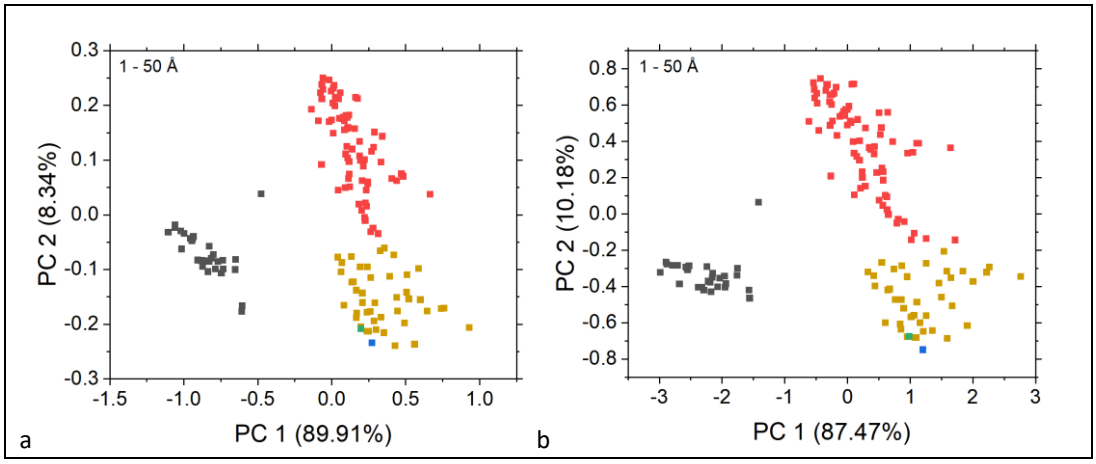


App. Figure 28 Run 1 – G(r) difference plots (offset) alongside solution temperature, with the crystallisation exotherm shown in the inset

The clustering of the XPDF patterns in Cooling Crystallisation Run 1 will be compared with the same analysis for Cooling Crystallisation Run 2 to establish if the stages are consistent, and the intensity changes across the stages can be compared with EPSR-simulated g(r) to evaluate the changes in molecular structure that may be occurring.

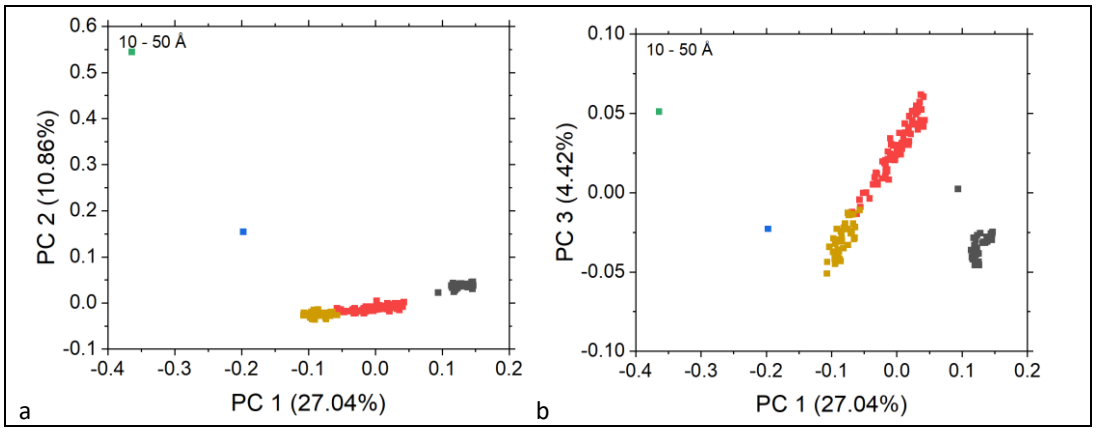
D.1.2 Cooling crystallisation: Run 2

App. Figure 29a&b compares the PCA score plots of g(r) and G(r) over the range 1–50 Å. As with run 2, removing r < 1 Å separates the points into three regions. Comparing the 1–50 Å score plots, the G(r) plot has a greater spread of PC values across PC 1 and 2 than the equivalent g(r) plot. Further analysis will be based on G(r) patterns.



App. Figure 29 Run 2 – PCA scores plot for a) g(r) and b) G(r) XPDF patterns

When the data for 10–50 Å are analysed, the transition and crystallisation stages can be seen, as well as the stage 1 points (App. Figure 30a&b). The stages are identified and compared with those from Run 1 in App. Table 13.

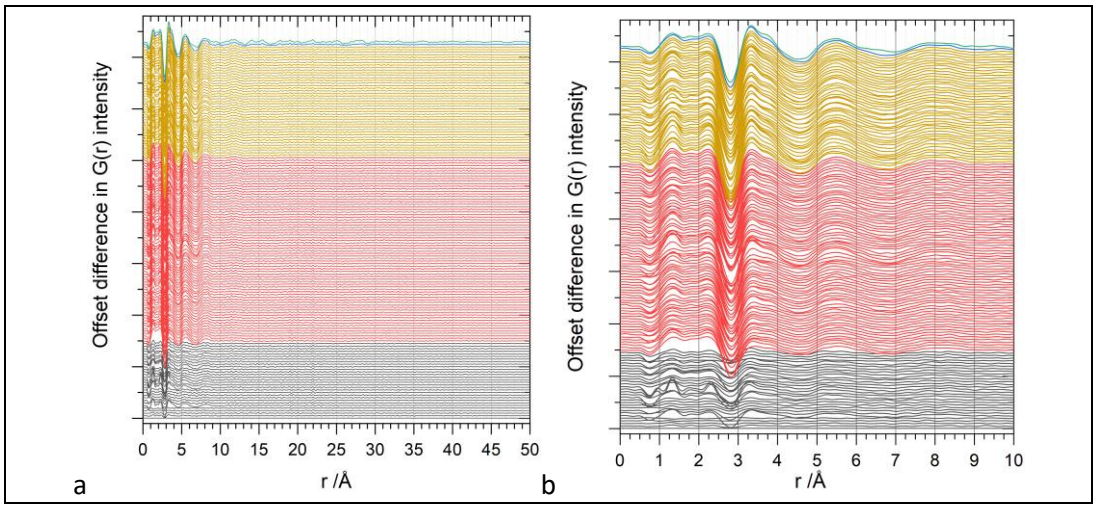


App. Figure 30 Run 2 – PCA scores plots of G(r) XPDF patterns: a) PC1 and PC2 & b) PC1 and PC3

App. Table 13 PCA-derived stages for Run 2 (with Run 1 data for comparison)

	Run 1		**Run 2**	
	Frame numbers	Temperature /°C	**Frame numbers**	**Temperature /°C**
Stage 1	0–29	62.5–55.7	**0–29**	**62.6–55.8**
Stage 2	30–113	55.5–30.5	**30–105**	**55.5–33.7**
Stage 3	114–146	30.1–19.8	**106–144**	**33.3–20.5**
Transition	147	19.6	**145**	**20.2**
Crystallisation	148–150	20.7–20.0	**146**	**20.7**

The G(r) difference plots identify the interactions that change during the *in situ* crystallisation process (App. Figure 31).



App. Figure 31 Run 2 – G(r) difference plots

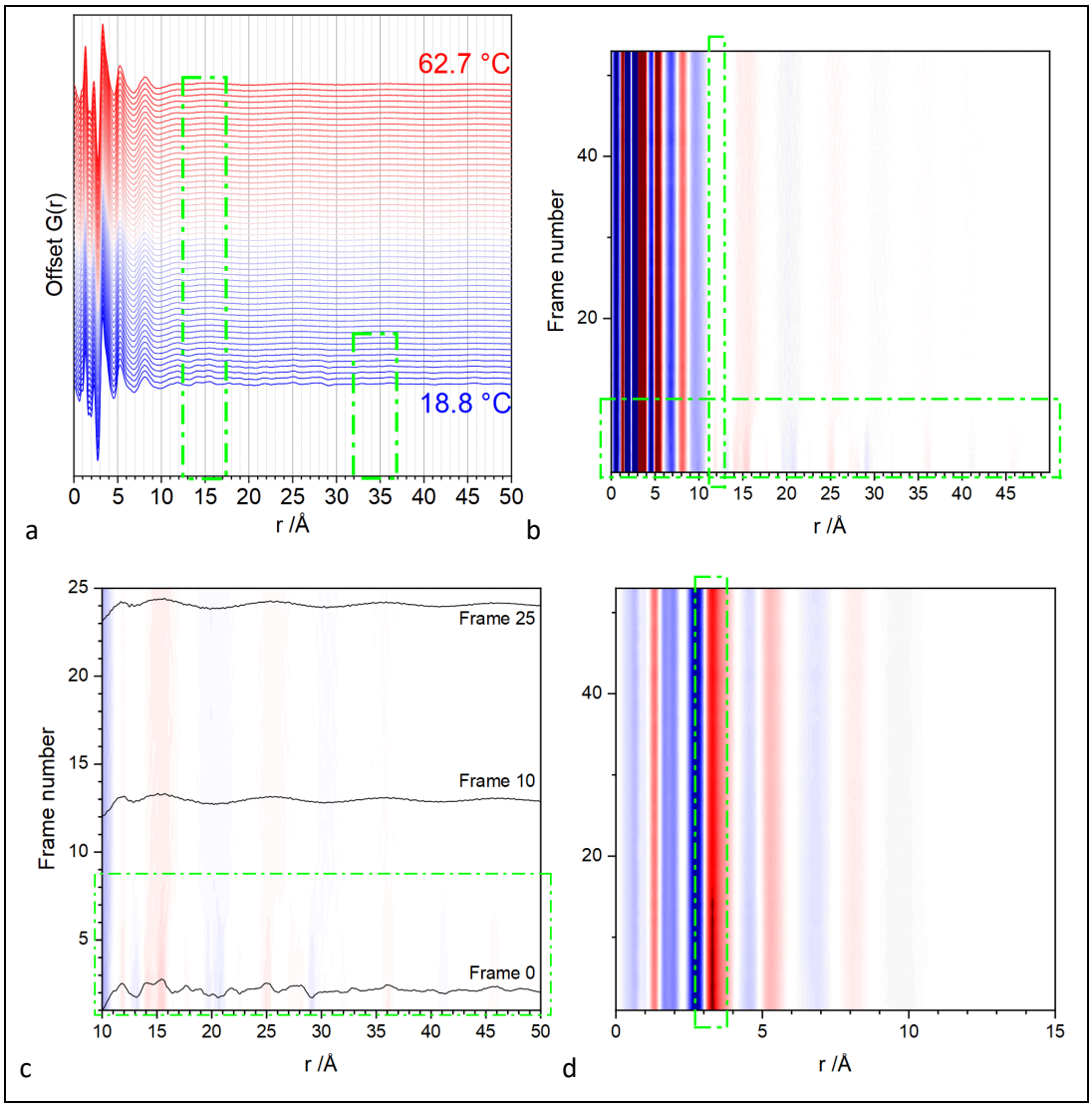
It appears that from stage 1 to stage 2, the principal structural changes occur between 1.0 and 2.3 Å and at 5.5 Å and 8.2 Å, where intensity increases, whereas a decrease in intensity is observed at 2.8 Å at the start of stage 2.

From stage 2 to stage 3, there are peak shifts from 8.2 Å to 8.0 Å, from 3.4 Å to 3.3 Å, which also increases in intensity. The peaks at 1.3 Å and 2.2 Å also increase in intensity. At larger r-values, there are no significant changes to the XPDF patterns until transition and crystallisation, which is in agreement with the data for Run 1, suggesting that the structural changes are reproducible.

D.1.3 Heating the solution: dissolution

Between Runs 1 and 2, X-ray total scattering data were collected whilst the solution was reheated (30 s frames, 18.8–62.7°C). 53 data frames were collected (the 54th has been omitted here as there was a lot of noise in the data, possibly due to an air bubble being present).

Examples of medium and long-range order are highlighted in the offset G(r) plot (App. Figure 32a). Significant long-range structure is seen up to 50 Å until frame 6 (in the horizontal green box, App. Figure 32b&c). At 13 Å, the trough reduces in intensity, starting at frame 4 (61.6°C), disappearing by frame 10 (60.3°C). At 12 Å, the peak loses intensity, but can be seen after other medium and long-range interactions disappear. In the range 0–15 Å (App. Figure 32d), the peak at 3.3 Å weakens in intensity by frame 15 (59.0°C).



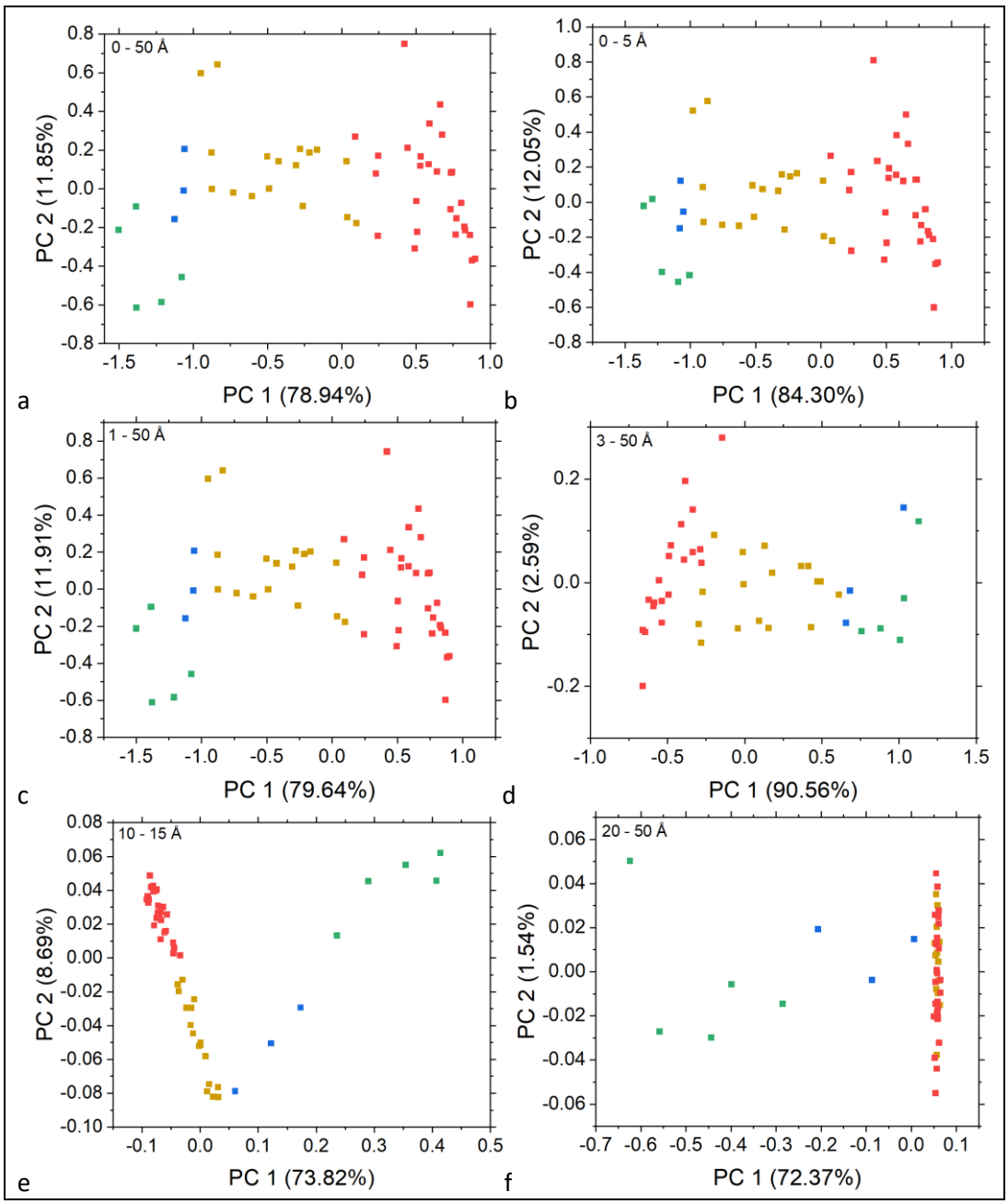
App. Figure 32 Dissolution of GuHCl in water: a) offset G(r) XPDF patterns, and contour plots of b) all G(r), c) long-range, and d) short- to medium-range interactions

The green boxes in a) pick out features at ~15 Å that remain present up to a higher temperature than those at ~30 Å. The vertical green box in b) indicates a feature at 12 Å that reduces in intensity over a greater temperature range than the longer-range interactions shown in the horizontal green box, which can be seen more clearly in the green box in c). d) indicates that an interaction at 3.2 Å is of a much greater intensity when there is solid present.

The contour plot parameters (range and increments) will affect how clearly features can be seen, so comparing across the plots does not necessarily indicate the relative importance of

the intensity changes, but this analysis provides an indication of the changing interactions during dissolution.

As with the crystallising solutions, PC analysis of the G(r) show that there is little difference when r < 1 Å are removed (App. Figure 33a&c). In the score plots of XPDF patterns over a range > 10 Å (App. Figure 33e&f), the frames where longer-range order (> 15 Å) is present are clearly separated from the solution stage scans and are coloured blue and green. The remaining frames are clustered in two main groups in App. Figure 33e (stage 1H, gold and stage 2H, red, detailed in App. Table 14). These data would benefit from K-means clustering analysis of XPDF patterns where empty Kapton is used for background subtraction.



App. Figure 33 PCA of dissolution G(r) XPDF patterns

App. Table 14 PCA-derived stages for the dissolution (reheating solution)

	Heating	
	Frame numbers	Temperature /°C
Crystal	0–4	18.8–21.3
Transition	5–7	22.1–24.1
Stage 1H	8–24	25.2–45.5
Stage 2H	25–52	46.9–62.7

D.1.4 Summary of the XPDF pattern analysis

The onset temperatures of the stages are similar in Runs 1 & 2, the biggest difference being between stages 2 and 3 (App. Table 15). It is likely that stage 1 is the undersaturated solution state,[55] as the starting temperature of the measurements was selected to be higher than the temperature at which the solute dissolved in Crystal16. The second stage may indicate the onset of supersaturation, which will be the driving force for crystallisation. The third stage would also therefore be in a supersaturated state, although PCA suggests there are structural differences between stages 2 and 3. These differences may be due to the different clustering of the ions in solution or changes in ion hydration.

Upon reheating, there is an extended transition phase, followed by two further stages according to this analysis.

App. Table 15 Temperatures for onset of PCA-derived stages for Runs 1 & 2, based on analysis of XPDF patterns: water in Kapton background

	Run 1	Run 2	Average
	Temperature /°C	Temperature /°C	Temperature /°C
Stage 1	62.5	62.6	62.6 ±0.05
Stage 2	55.5	55.5	55.5 ±0.0
Stage 3	30.1	33.3	31.7 ±1.6
Transition	19.6	20.2	19.9 ±0.3
Crystallisation	20.7	20.7	20.7 ±0.0

D.2 EPSR analysis: Coordination number (CoordN) data tables

CoordN data is presented as the average per stage rather than per simulation. Whilst there is some variation in CoordN for the simulations in each stage, taking the average of each stage indicates the general trend in the stage. The SD indicates the spread of data in the stage. The r-range considered for each stage is noted in the relevant CoordN table.

[55] The extent of undersaturation should be measured to fully characterise the system

D.2.1 Water atom-atom correlations

App. Table 16 O_t-O_t CoordN – 1st shell – peak and shoulder: averaged per stage

	O_t-O_t 1st shell (to shoulder)			O_t-O_t 1st shell shoulder		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-3.1	1.39	*0.02*	3.1-5.5	3.33	*0.10*
Stage 2	0-3.1	1.39	*0.07*	3.1-5.5	3.41	*0.03*
Stage 3	0-3.1	1.44	*0.02*	3.1-5.5	3.35	*0.07*

App. Table 17 Ot-Ht CoordN – 1st and 2nd shells: averaged per stage

	O_t-H_t 1st shell			O_t-H_t 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-2.3	0.62	*0.00*	2.3-4.7	10.16	*0.23*
Stage 2	0-2.3	0.60	*0.04*	2.3-4.7	10.38	*0.19*
Stage 3	0-2.3	0.65	*0.01*	2.3-4.7	10.29	*0.14*

App. Table 18 H_t-H_t CoordN – 1st and 2nd shells: averaged per stage

	H_t-H_t 1st shell			H_t-H_t 2nd shell		
	Ave. Range /Å	Ave. CoordN	SD	Ave. CoordN	Ave. Range /Å	SD
Stage 1	0-3.0	2.41	*0.03*	3.0-5.0	10.55	*0.27*
Stage 2	0-3.0	2.43	*0.07*	3.0-5.0	10.72	*0.13*
Stage 3	0-3.0	2.44	*0.02*	3.0-5.0	10.66	*0.17*

D.2.2 Gdm^+-Cl^- atom-atom correlations

App. Table 19 Cl^--N CoordN – 1st and 2nd shells: averaged per stage

	Cl^--N 1st shell			Cl^--N 2nd shell			Cl^--N 1st 2 shells Ave. CoordN
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*	
Stage 1	0-4.5	6.64	*0.04*	4.5-6.3	11.36	*0.08*	0.37
Stage 2	0-4.5	6.60	*0.07*	4.5-6.4	12.07	*0.14*	0.35
Stage 3	0-4.5	6.71	*0.07*	4.5-6.3	11.66	*0.12*	0.37

App. Table 20 Cl^--H CoordN – 1st and 2nd shells: averaged per stage

	Cl^--H 1st shell			Cl^--H 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-3.05	4.06	*0.03*	3.05-4.7	10.42	*0.08*
Stage 2	0-3.05	3.93	*0.04*	3.00-4.7	10.57	*0.04*
Stage 3	0-3.05	4.11	*0.10*	3.05-4.7	10.58	*0.08*

App. Table 21 H-Cl^- CoordN – 1st and 2nd shells: averaged per stage

	H-Cl- 1st shell			H-Cl- 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-3.05	0.68	*0.01*	3.05-4.7	1.74	*0.01*
Stage 2	0-3.05	0.66	*0.01*	3.00-4.7	1.76	*0.01*
Stage 3	0-3.05	0.69	*0.02*	3.05-4.7	1.76	*0.01*

App. Table 22 C-Cl⁻ CoordN – 1st and 2nd shells: averaged per stage

	C-Cl⁻ 1st shell			C-Cl⁻ 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-4.1	1.81	*0.05*	4.1-5.5	2.59	*0.08*
Stage 2	0-4.1	1.79	*0.01*	4.1-5.5	2.62	*0.06*
Stage 3	0-4.1	1.81	*0.01*	4.1-5.5	2.66	*0.06*

App. Table 23 N-Cl⁻ CoordN – 1st and 2nd shells: averaged per stage

	N-Cl⁻ 1st shell			N-Cl⁻ 2nd shell			N-Cl- 1st 2 shells Ave. CoordN	2nd: 1st shell ratio
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*		
Stage 1	0-4.5	2.21	*0.01*	4.5-6.3	3.79	*0.03*	6.00	1.71
Stage 2	0-4.5	2.20	*0.02*	4.5-6.4	4.07	*0.01*	6.27	1.85
Stage 3	0-4.5	2.24	*0.02*	4.5-6.3	3.89	*0.04*	6.12	1.74

D.2.3 Gdm^+-Gdm^+ and Gdm^+-water atom-atom correlations

App. Table 24 C-C CoordN – 1st and 2nd shells: averaged per stage

	C-C 1st shell			C-C 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-6.1	3.88	*0.04*	6.1-9.1	12.21	*0.08*
Stage 2	0-5.9	3.57	*0.08*	5.9-9.4	14.29	*0.24*
Stage 3	0-6.0	3.73	*0.08*	6.0-9.3	13.95	*0.08*

App. Table 25 C-C CoordN – 1st shell shoulder and peak: averaged per stage

	C-C shoulder			C-C 1st peak (after shoulder)			CoordN ratio (peak): (shoulder)
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*	
Stage 1	0-4.1	0.45	*0.03*	4.1-6.1	3.43	*0.02*	7.68
Stage 2	0-4.2	0.58	*0.04*	4.2-5.9	2.99	*0.05*	5.18
Stage 3	0-4.2	0.59	*0.08*	4.2-6.0	3.14	*0.13*	5.39

App. Table 26 C-N CoordN – 1st and 2nd shells: averaged per stage

	C-N 1st shell			C-N 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-4.5	3.21	*0.11*	4.5-5.3	3.69	*0.05*
Stage 2	0-4.5	3.35	*0.21*	4.5-5.3	3.79	*0.06*
Stage 3	0-4.5	3.32	*0.18*	4.5-5.3	3.72	*0.04*

App. Table 27 N-N CoordN – 1st and 2nd shells: averaged per stage

	N-N 1st shell			N-N 2nd shell			N-N 1st 2 shells 0-~7.2 Å
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*	Ave. CoordN
Stage 1	0-4.1	2.19	*0.06*	4.1-7.2	19.87	*0.14*	22.06
Stage 2	0-4.1	2.80	*0.13*	4.3-6.8	15.59	*0.14*	18.39
Stage 3	0-4.1	2.50	*0.16*	4.1-7.2	19.07	*0.14*	21.57

App. Table 28 H-O_t CoordN – 1st and 2nd shells: averaged per stage

	H-O_t 1st shell			H-O_t 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-2.5	0.40	*0.01*	2.5-4.4	3.39	*0.05*
Stage 2	0-2.5	0.41	*0.01*	2.5-4.3	3.05	*0.01*
Stage 3	0-2.5	0.41	*0.01*	2.5-4.4	3.49	*0.02*

App. Table 29 C-O_t CoordN – 1st shell (peak and shoulder): averaged per stage

	C-O_t 1st shell (to shoulder)			C-O_t 1st shell shoulder		
	Range /Å	Ave. CoordN	*SD*	Range /Å	Ave. CoordN	*SD*
Stage 1	0-4.2	3.57	*0.02*	4.2-5.4	4.27	*0.15*
Stage 2	0-4.2	3.59	*0.04*	4.2-5.4	4.17	*0.06*
Stage 3	0-4.2	3.66	*0.08*	4.2-5.4	4.36	*0.09*

App. Table 30 C-O_t CoordN – 1st and 2nd shells: averaged per stage

	C-O_t 1st shell (inc. shoulder)			C-O_t 2nd shell		
	Range /Å	Ave. CoordN	*SD*	Range /Å	Ave. CoordN	*SD*
Stage 1	0-5.4	7.84	*0.14*	5.4-8.2	22.56	*0.22*
Stage 2	0-5.4	7.76	*0.03*	5.4-8.2	22.94	*0.07*
Stage 3	0-5.4	8.02	*0.06*	5.4-8.0	18.13	*0.12*

App. Table 31 C-H_t CoordN – 1st and 2nd shells: averaged per stage

	C-H_t 1st shell			C-H_t 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-5.7	18.99	*1.31*	5.7-8.2	34.66	*1.53*
Stage 2	0-5.7	18.65	*0.10*	5.7-7.9	32.74	*13.87*
Stage 3	0-5.8	19.82	*0.57*	5.8-7.7	28.51	*1.13*

App. Table 32 N-O_t CoordN – 1st and 2nd shells: averaged per stage

	N-O_t 1st shell			N-O_t 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-3.2	0.96	*0.01*	4.3-5.8	2.48	*0.04*
Stage 2	0-3.2	0.97	*0.02*	4.3-5.8	2.44	*0.04*
Stage 3	0-3.2	0.98	*0.03*	4.3-5.8	2.55	*0.02*

App. Table 33 N-H_t CoordN – 1st and 2nd shells: averaged per stage

	N-H_t 1st shell			N-H_t 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-4.5	8.00	*0.11*	4.5-7.1	27.87	*0.27*
Stage 2	0-4.4	7.57	*0.28*	4.4-6.8	24.02	*0.14*
Stage 3	0-4.5	8.24	*0.08*	4.5-7.1	28.35	*0.07*

D.2.4 Cl^--Cl^- and Cl^--water atom-atom correlations

App. Table 34 Cl^--Cl^- CoordN – 1st shell, and 1st-3rd shells: averaged per stage

	Cl^--Cl^- 1st shell			Cl^- /10^3 $Å^3$	Cl^--Cl^- 1st 3 shells			Cl^- /10^3 $Å^3$
	Ave. Range /Å	Ave. CoordN	*SD*		Ave. Range /Å	Ave. CoordN	*SD*	
Stage 1	0-5.9	3.47	*0.05*	4.03	0-9.2	16.78	*0.11*	5.15
Stage 2	0-5.4	3.45	*0.02*	5.22	0-9.3	17.43	*0.16*	5.17
Stage 3	0-5.6	3.51	*0.06*	4.77	0-9.3	17.62	*0.05*	5.23

App. Table 35 Cl^--O_t CoordN – 1st and 2nd shells: averaged per stage

	Cl^--O_t 1st shell			Cl^--O_t 2nd shell			Cl^--O_t 1st 2 shells 0-~6.5 Å
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*	Ave. CoordN
Stage 1	0-4.5	4.54	*0.09*	4.5-6.5	8.60	*0.05*	13.13
Stage 2	0-4.5	4.58	*0.07*	4.5-6.4	7.91	*0.10*	12.49
Stage 3	0-4.5	4.64	*0.05*	4.5-6.5	8.04	*0.08*	12.68

App. Table 36 Cl^--H_t CoordN – 1st and 2nd shells: averaged per stage

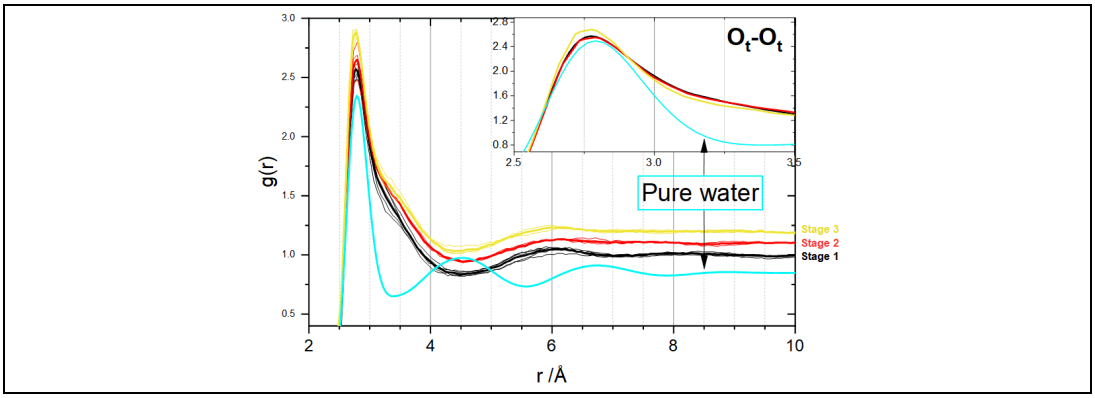
	Cl^--H_t 1st shell			Cl^--H_t 2nd shell		
	Ave. Range /Å	Ave. CoordN	*SD*	Ave. Range /Å	Ave. CoordN	*SD*
Stage 1	0-2.9	2.42	*0.04*	2.9-4.2	4.99	*0.03*
Stage 2	0-2.9	2.44	*0.08*	2.9-4.2	5.06	*0.07*
Stage 3	0-2.9	2.46	*0.03*	2.9-4.2	5.14	*0.06*

D.3 EPSR analysis: g(r) graphs

Average per stage with individual simulations g(r)

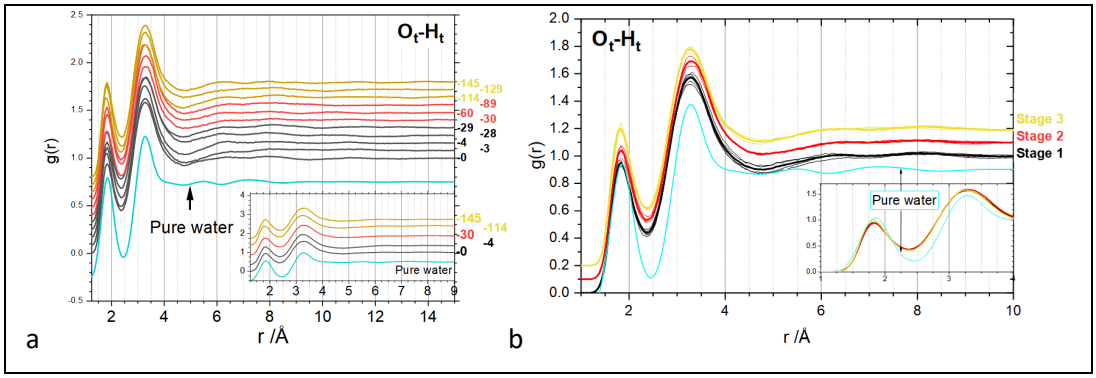
For each partial pair correlation for atoms α-β, the graphs show the $g_{\alpha\text{-}\beta}(r)$ for each individual simulation (thin lines, offset by stage) with average $g_{\alpha\text{-}\beta}(r)$ for the stage in bold (offset by stage), inset: average $g_{\alpha\text{-}\beta}(r)$.

D.3.1 Water-water atomic interactions



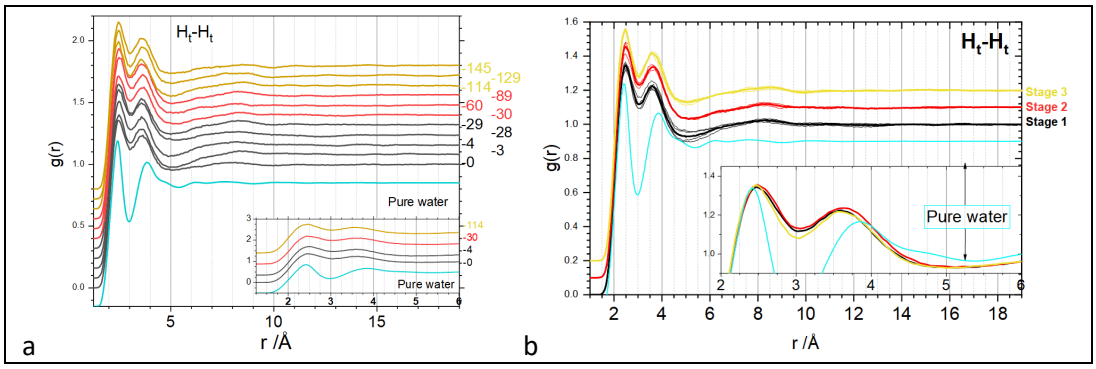
App. Figure 34 Average O_t-O_t partial pair distribution functions $g_{Ot\text{-}Ot}(r)$

g(r) of pure water (light blue) [375]



App. Figure 35 O_t-H_t partial pair distribution functions $g_{Ot\text{-}Ht}(r)$

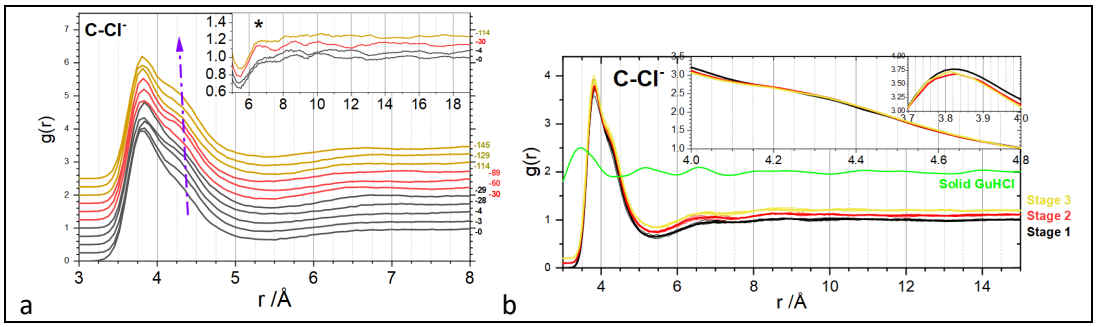
a) offset $g_{Ot\text{-}Ht}(r)$, and b) $g_{Ot\text{-}Ht}(r)$ averaged by stage, with g(r) of pure water (light blue) [375]



App. Figure 36 H_t-H_t partial pair distribution functions $g_{Ht\text{-}Ht}(r)$

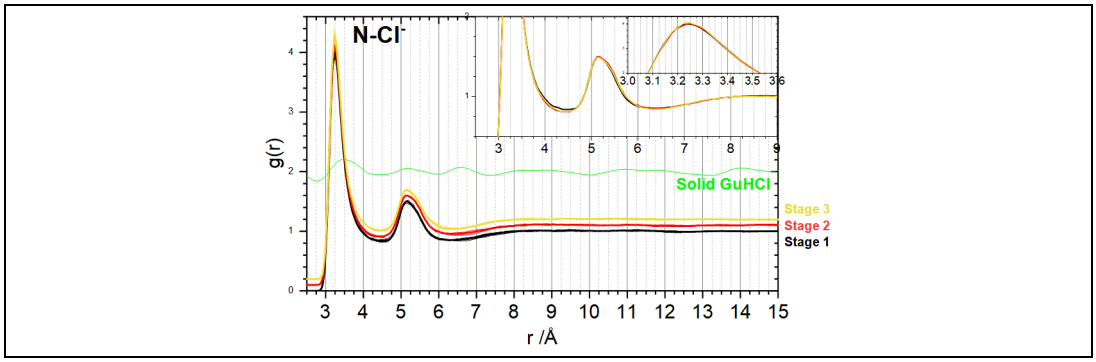
a) offset $g_{Ht\text{-}Ht}(r)$, and b) $g_{Ht\text{-}Ht}(r)$ averaged by stage, with g(r) of pure water (light blue) [375]

D.3.2 Gdm^+-Cl^- atomic interactions



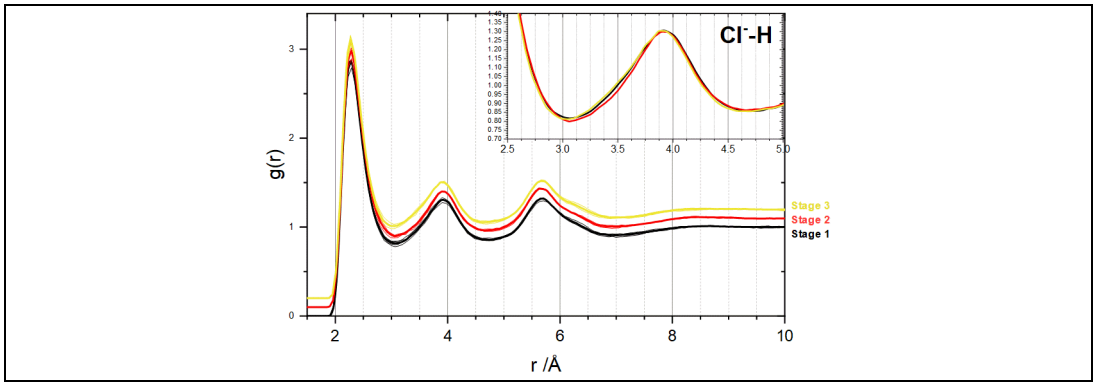
App. Figure 37 C-Cl^- partial pair distribution functions $g_{C\text{-}Cl^-}(r)$

a) Offset $g_{C\text{-}Cl^-}(r)$: purple arrow indicating the reducing r of the shoulder position and the (*) in the inset highlighting the increasing intensity of the feature at ~6.5 Å, and b) $g_{C\text{-}Cl^-}(r)$ averaged by stage, with g(r) of solid GuHCl (light green)



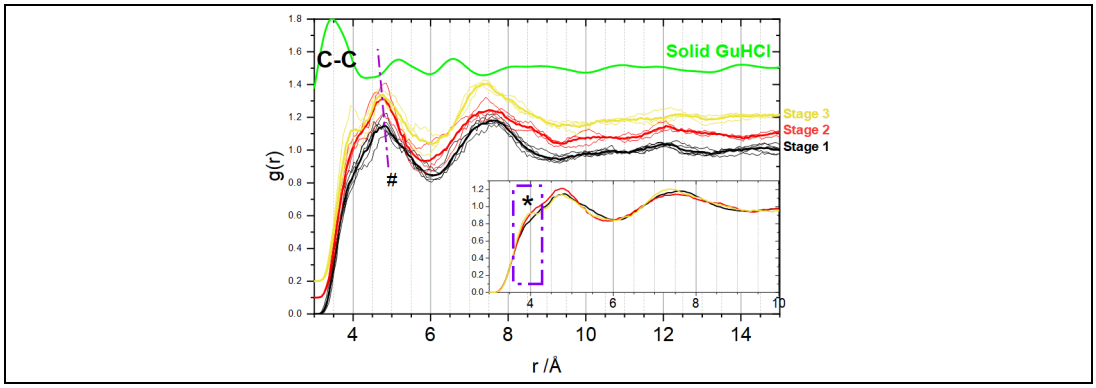
App. Figure 38 Average N-Cl$^-$ partial pair distribution functions $g_{N\text{-}Cl^-}(r)$

With g(r) of solid GuHCl (light green)



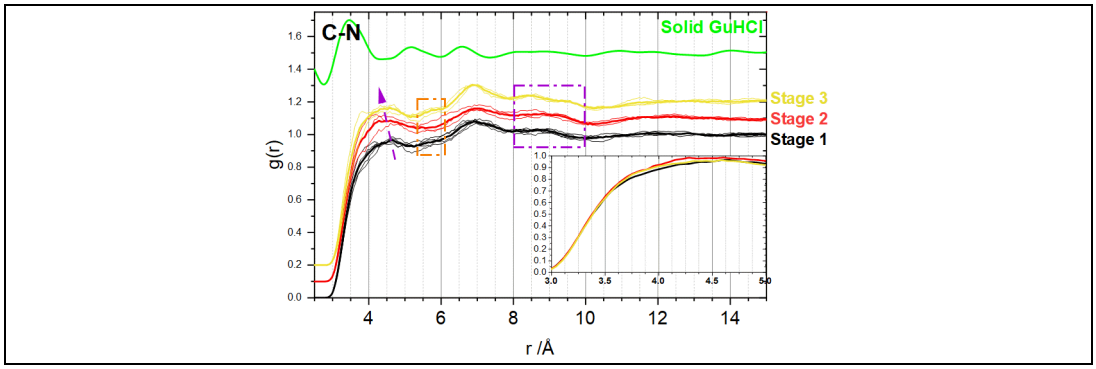
App. Figure 39 Average Cl$^-$-H partial pair distribution functions $g_{Cl^-\text{-}H}(r)$

D.3.3 Gdm$^+$-Gdm$^+$ and Gdm$^+$-water atomic interactions



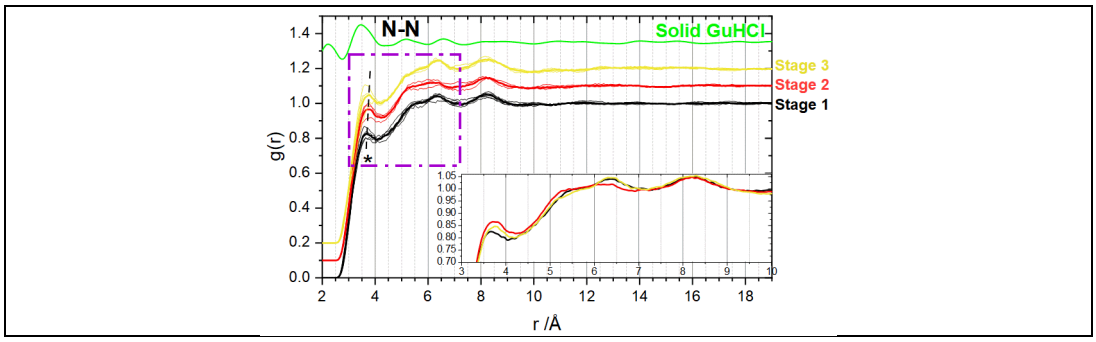
App. Figure 40 Average C-C partial pair distribution functions $g_{C\text{-}C}(r)$

Line indicated with # shows decreasing interatomic distance of the first peak, and * in the purple box (inset) showing the increasing shoulder position with increasing stage. g(r) of solid GuHCl (light green)



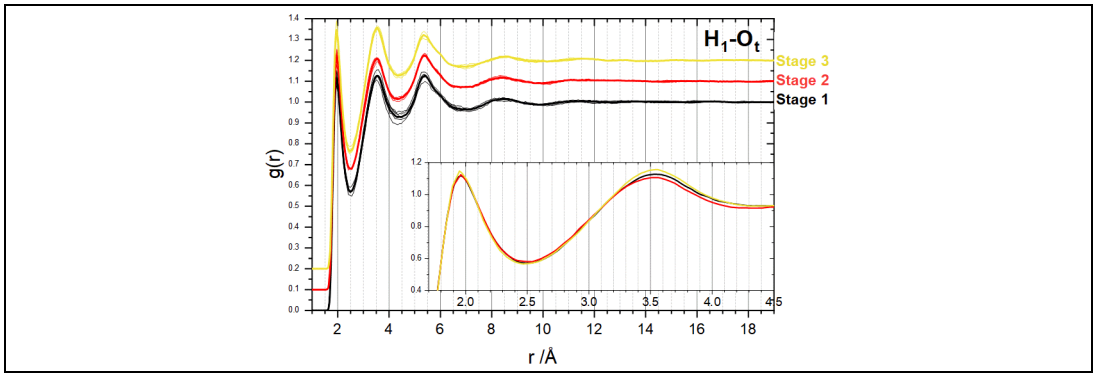
App. Figure 41 Average C-N partial pair distribution functions $g_{C\text{-}N}(r)$

Purple arrow indicates the reducing distance of the first interaction peak, orange box: highlights differing features between stages at ~5.75 Å, purple box: highlights differing features between stages at 8–10 Å, with g(r) of solid GuHCl (light green)

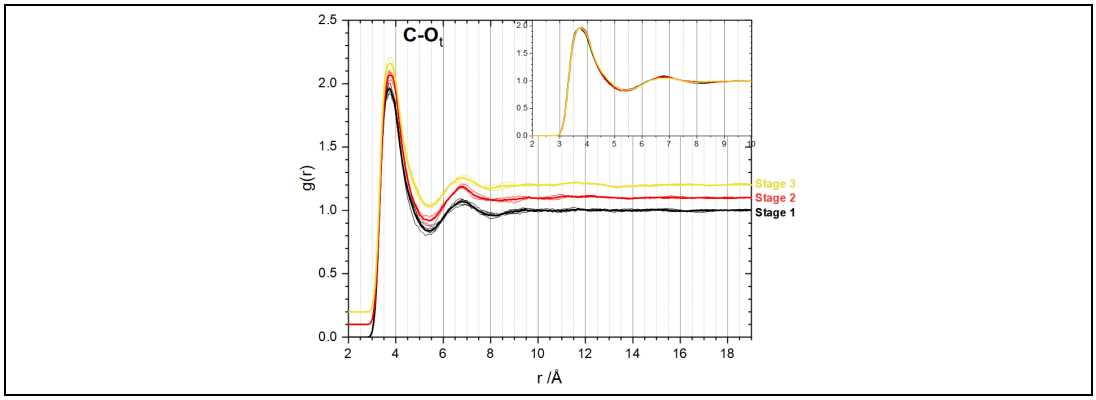


App. Figure 42 Average N-N partial pair distribution functions $g_{N\text{-}N}(r)$

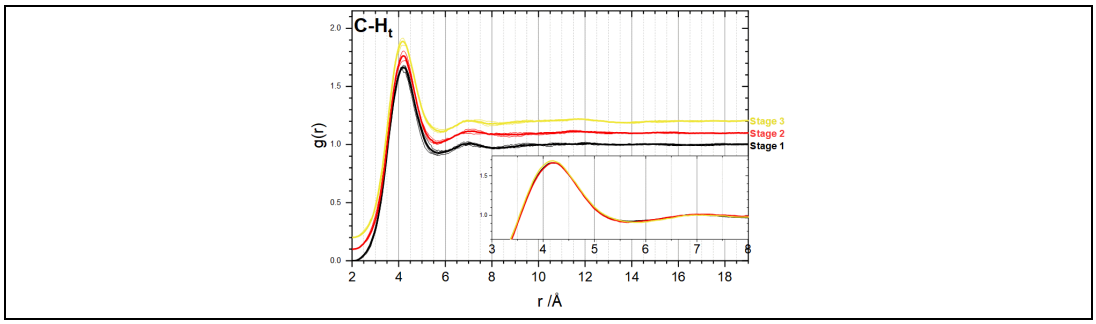
Short- to medium-range features highlighted by the purple box, and the line indicated with * showing the change in first peak position, with g(r) of solid GuHCl (light green)



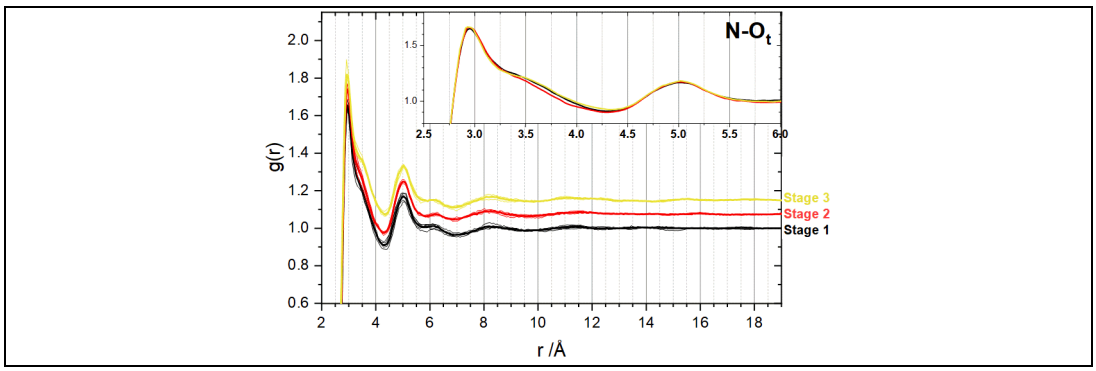
App. Figure 43 Average H-O_t partial pair distribution functions $g_{H\text{-}Ot}(r)$



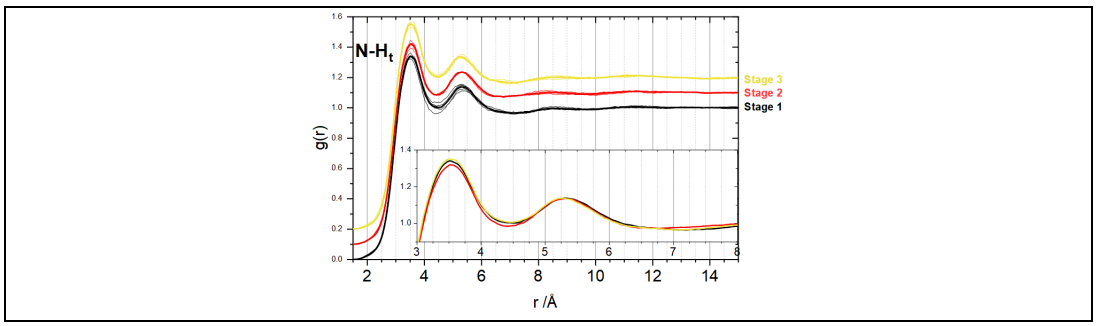
App. Figure 44 Average C-O_t partial pair distribution functions $g_{C\text{-}Ot}(r)$



App. Figure 45 Average C-H_t partial pair distribution functions $g_{C\text{-}Ht}(r)$

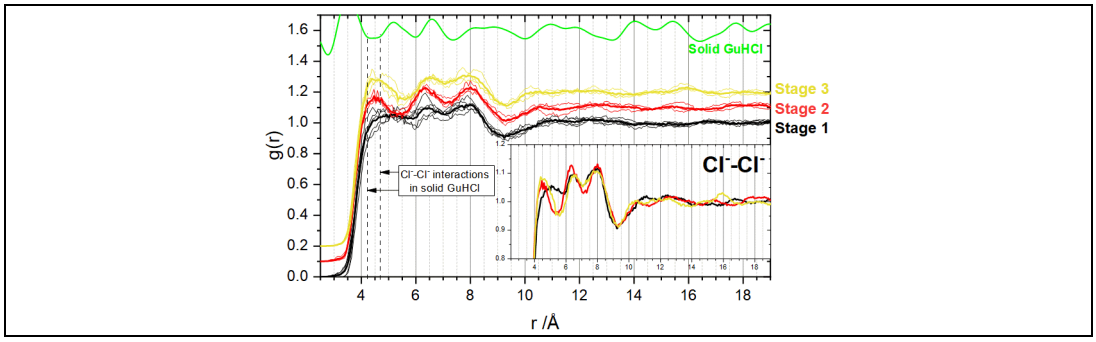


App. Figure 46 Average N-O_t partial pair distribution functions $g_{N\text{-}Ot}(r)$



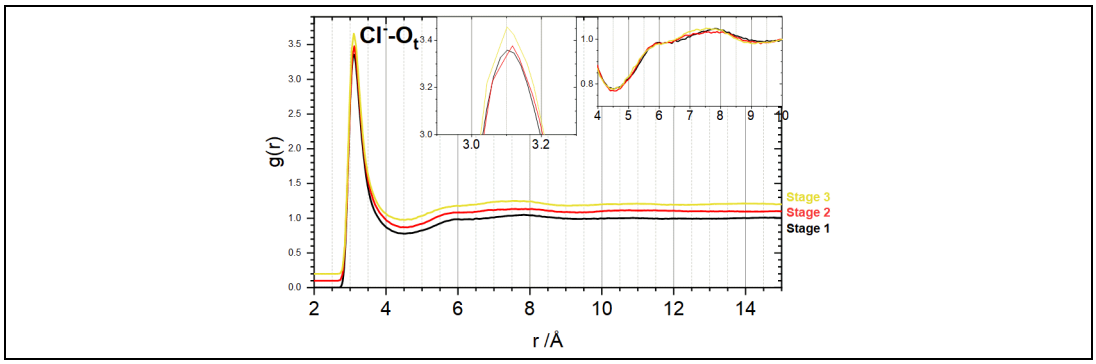
App. Figure 47 Average N-H_t partial pair distribution functions $g_{N\text{-}Ht}(r)$

D.3.4 Cl^--Cl^- and Cl^--water atomic interactions

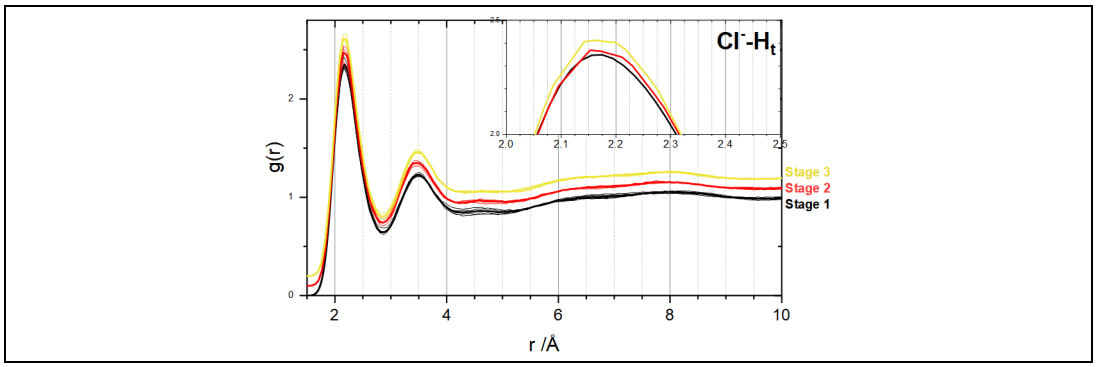


App. Figure 48 Average Cl^--Cl^- partial pair distribution functions $g_{Cl^-\text{-}Cl^-}(r)$

With Cl^--Cl^- interaction distances in solid GuHCl indicated by the dash black lines and the g(r) of solid GuHCl (light green)

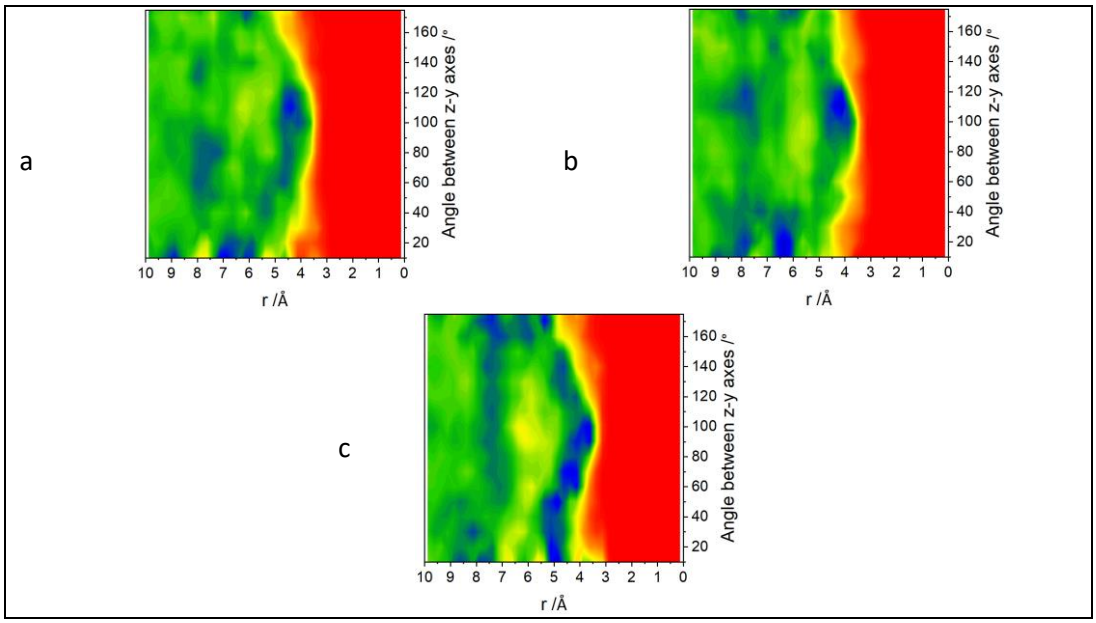


App. Figure 49 Average Cl^-O_t partial pair distribution functions $g_{Cl^-\text{-}Ot}(r)$



App. Figure 50 Average Cl^-H_t partial pair distribution functions $g_{Cl^-\text{-}Ht}(r)$

D.4 Angular radial distribution functions



App. Figure 51 Gdm^+-Gdm^+ angular radial distribution function between the z-axes of Gdm^+ (central molecule) and the y-axis of the surrounding Gdm^+

a) **Stage 1**: GuHCl_3, b) **Stage 2**: GuHCl_60, c) **Stage 3**: GuHCl_129

Appendix E Auxiliary routines for statistical analysis

Statistical analysis of the molecular model was undertaken using auxiliary routines in EPSR, with further analysis using dlputils. The following routines have been applied.

E.1 EPSR analysis

E.1.1 Partials

The Partials routine provides accumulated site-site radial distribution functions. The distributions are calculated to the largest spherical shell radius allowed by the box size.

E.1.2 Clusters

A molecule is assigned to a cluster if a specific atom pair is within a given separation and a cluster distribution is then determined.

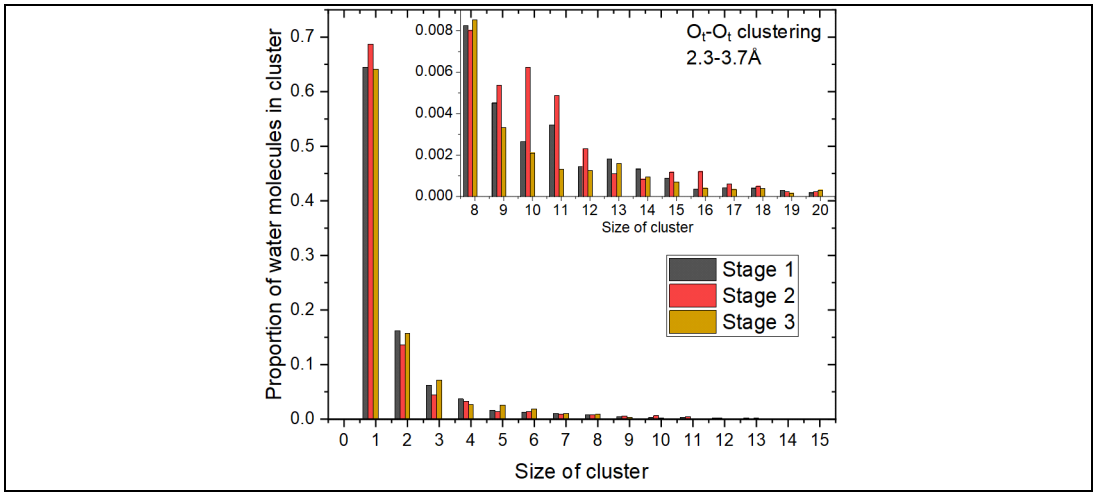
For water, two analyses were undertaken to evaluate O_t-O_t clustering with different selection parameters.

App. Table 37: Cluster analysis: r-ranges for first O_t-O_t peak

Analysis	r_{min} /Å	r_{max} /Å
1	2.3	2.9
2	2.3	3.7

r_{min} was based on the r-value at the start of the first O_t-O_t interaction peak. Max r for analysis 1 was selected to avoid including the O_t-O_t interactions that were related to the 'shoulder' seen in the $g_{Ot\text{-}Ot}(r)$ at ~3.5 Å, as it has been hypothesised that these may be indirect interactions of water molecules that are hydrating an ion. The second max r-value was selected to estimate the maximum separation of the O_t on hydrogen-bonded water molecules and considers the intramolecular O-H bond in water (max r of this peak = 1.2 Å), plus the maximum hydrogen bond length (2.5 Å) = 3.7 Å .

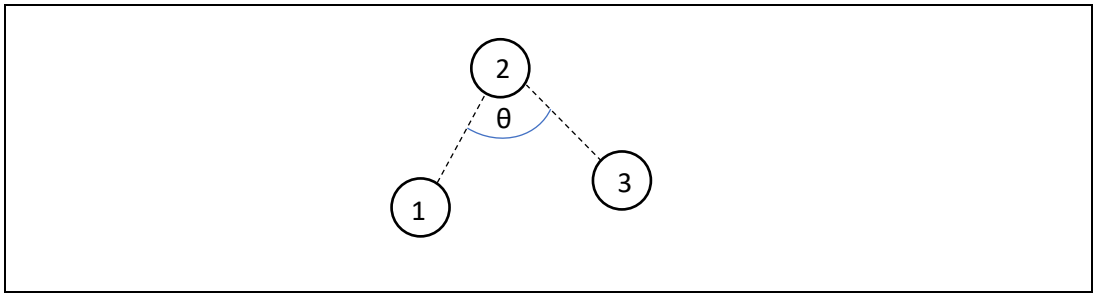
The plot for the Analysis 1 radii is in Chapter 4 (Figure 4.41). App. Figure 52 shows the distribution of water cluster sizes based on the radii considered in Analysis 2.



App. Figure 52 Water clustering: O_t-O_t interactions between 2.3 and 3.7 Å

E.1.3 Triangles

The Triangles routine allows the study of the angle (θ) distribution between three atoms (bonded or non-bonded), where the distance ranges for 1-2 and 2-3 are specified (App. Figure 53).



App. Figure 53 Inputs for the Triangles auxiliary routine in EPSR

E.2 dlputils analysis

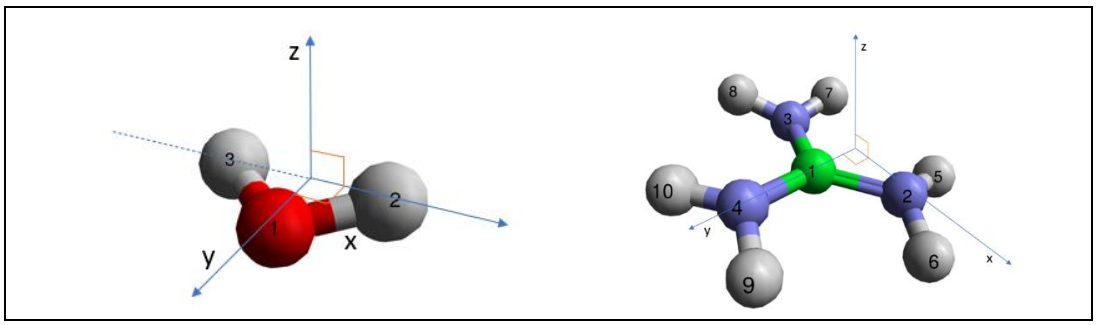
E.2.1 *ardf* for angular radial distribution functions

Angular radial distribution function (ARDF) calculates the radial distribution function of sites, binning the results by the angle defined by vectors on each target molecule [460]. A switch in the dlputils script was used to centre the axes origin on the central C, although there appeared to be no difference in the output irrespective of whether centring was applied.

In the *in situ* analysis (Chapter 4), the data are for four individual simulations and are not averages of the simulations for each stage. The trajectory data were collected over around 350 iterations of the structure per simulation. A greater number of data points may improve the data, but the data is indicative of the molecular structures of the simulations analysed.

E.2.2 *pdens* for spatial probability densities

pdens calculates the spatial probability densities of one species around another. *pdens* uses the defined axis origins (App. Figure 54) or species centre of mass to represent the positions of surrounding components around a central species.



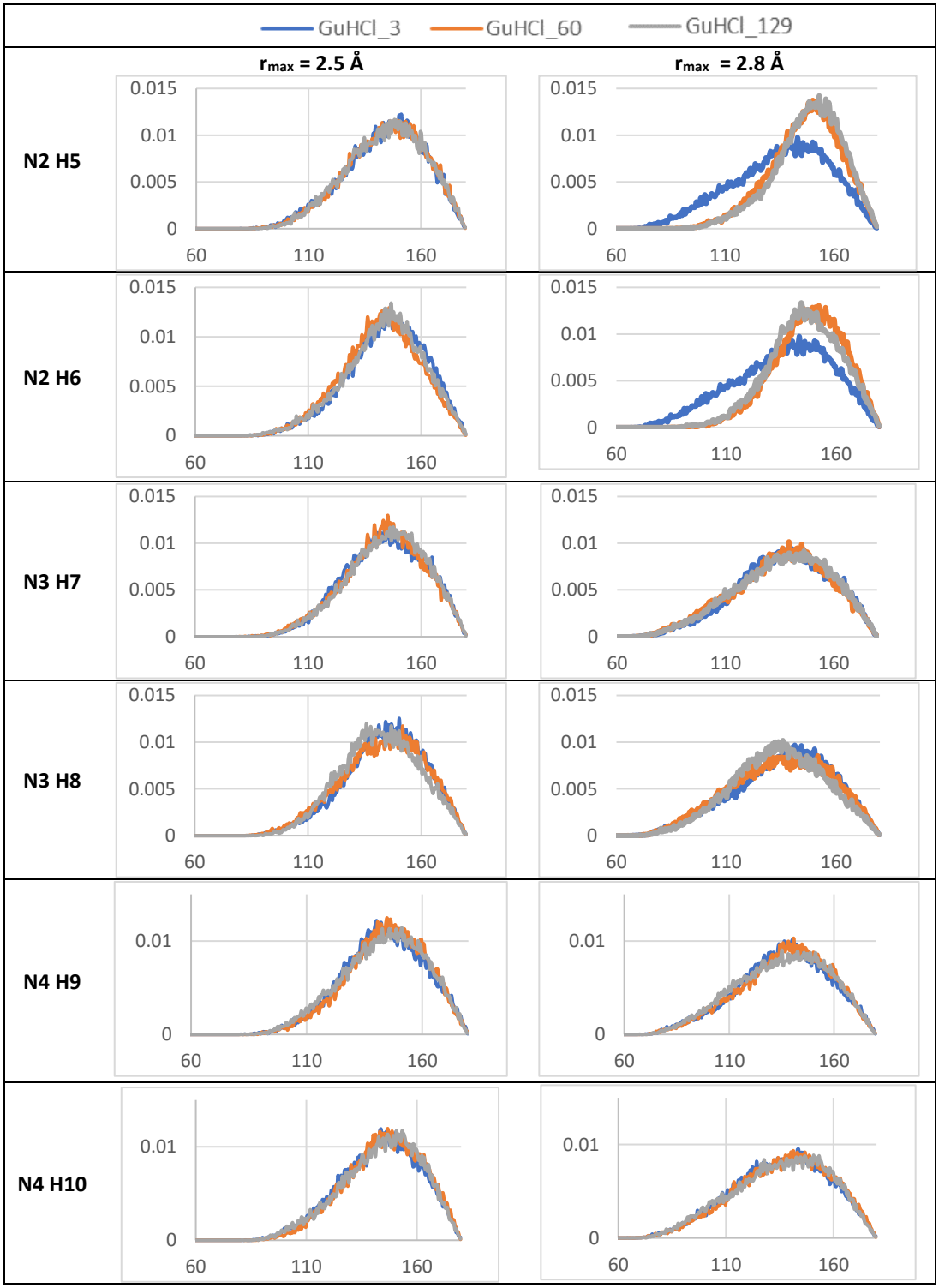
App. Figure 54 Axes applied to water and Gdm^+ for the spatial probability density analyses

The analysis used here does not consider the orientation of the surrounding species.

E.2.3 *intratorsion* and *intertorsion* for angles analysis

intratorsion analysis generates a histogram of a single torsion angle for a given species (here it was used for analysis of water molecules in Chapter 4). *intertorsion* analysis generates a histogram of torsion angles defined by two atoms on one species and two atoms on another. It requires the atom numbers, rather than atom types, to be specified. To analyse the hydrogen bonds between Gdm^+ and water (N-H···O), six analyses were carried out (for each -NH in Gdm^+ with one specified H_t). Averages of the six analyses were calculated. It was expected that the distributions of each N-H···O would be very similar, if not identical. The analysis was undertaken for N-H···O interactions < 2.5 Å and < 2.8 Å (App. Figure 55).

However, further investigation is needed, as the N2-H5···O_t and N2-H6···O_t distribution for GuHCl_3 (r_{max} = 2.8 Å) are a significantly different shape. It is not clear why the interactions from one amino group would be so different from the other two on the Gdm^+, as the Lennard-Jones potentials and partial charges are the same.



App. Figure 55 Comparison of hydrogen bond angle distributions N-H···O_t from each H atom on Gdm^+

Left-hand column: r_{max} = 2.5 Å and right-hand column: r_{max} = 2.8 Å

To describe the average hydrogen bonding between Gdm^+ and water via the H atoms, the average of the six analyses was calculated (r_{max} = 2.5 Å).

App. Table 38: Hydrogen bond angle analysis (N-H···O_t from each H atom on Gdm^+)

		Stage 1	Stage 2	Stage 3
N-H5···O_t	**Mean /°**	143.5	143.9	143.9
	SD /°	0.0039	0.0038	0.0038
	Mode /°	151.25	146.25	148.25
	% 150-180°	40.6%	40.6%	40.6%
	% 160-180°	19.2%	20.0%	20.3%
	% 170-180°	5.0%	5.4%	5.6%
		Stage 1	**Stage 2**	**Stage 3**
N-H6···O_t	**Mean /°**	144.2	141.7	143.1
	SD /°	0.0039	0.0040	0.0040
	Mode /°	146.25	145.75	146.75
	% 150-180°	40.7%	33.6%	37.7%
	% 160-180°	19.8%	15.1%	17.4%
	% 170-180°	5.2%	3.8%	4.7%
		Stage 1	**Stage 2**	**Stage 3**
N-H7···O_t	**Mean /°**	143.5	143.3	144.2
	SD /°	0.0038	0.0039	0.0039
	Mode /°	145.75	145.25	147.75
	% 150-180°	38.7%	38.2%	41.2%
	% 160-180°	20.3%	18.6%	20.4%
	% 170-180°	5.8%	5.4%	5.6%
		Stage 1	**Stage 2**	**Stage 3**
N-H8···O_t	**Mean /°**	143.5	142.6	141.3
	SD /°	0.0039	0.0037	0.0039
	Mode /°	150.25	151.75	135.75
	% 150-180°	39.3%	38.5%	33.3%
	% 160-180°	18.7%	19.2%	15.4%
	% 170-180°	5.0%	5.5%	4.2%
		Stage 1	**Stage 2**	**Stage 3**
N-H9···O_t	**Mean /°**	143.0	144.1	142.9
	SD /°	0.0039	0.0040	0.0038
	Mode /°	140.25	144.75	151.25
	% 150-180°	37.1%	40.1%	38.9%
	% 160-180°	18.5%	19.6%	19.1%
	% 170-180°	5.5%	5.0%	5.0%
		Stage 1	**Stage 2**	**Stage 3**
N-H10···O_t	**Mean /°**	142.9	144.7	143.8
	SD /°	0.0038	0.0039	0.0038
	Mode /°	143.75	146.75	150.75
	% 150-180°	38.5%	41.9%	41.7%
	% 160-180°	19.2%	21.8%	20.5%
	% 170-180°	5.4%	6.3%	5.3%
		Stage 1	**Stage 2**	**Stage 3**
Ave N-H···O_t	**Mean /°**	143.4	143.4	143.2
	SD /°	0.0039	0.0039	0.0039
	Mode /°	146.25	146.75	146.75
	% 150-180°	39.1%	38.8%	38.9%
	% 160-180°	19.3%	19.1%	18.8%
	% 170-180°	5.3%	5.2%	5.1%

Appendix F EPSR GuHCl parameterisation studies

Two parameterisation studies were undertaken to optimise the EPSR simulation of aqueous GuHCl solutions. The second parametrisation study assessed the effects of changing the molecule flexibility, the charge on the central carbon atom of Gdm^+ and the distance range applied to the potential truncation.

F.1 Parametrisation study 1

Lennard-Jones potentials and partial charges

The first study assessed the Lennard-Jones potentials, partial charges, simulation parameters, including the EP for model refinement, and water models to apply in the *in situ* crystallisation study (Chapter 4). In EPSR, the molecular model is first generated using the reference potentials (RP) alone, without refinement to the experimental data. The calculated F(Q) should be as close as possible to the experimental data before the empirical potential (EP) is applied to minimise the required perturbation to optimise the fit. Therefore, the selection of the RP is critical to the goodness of fit. The selected parameters were also applied to the GuHCl concentration study (Chapter 5).

A cubic simulation box of side length 38.44 Å was constructed, containing: 306 Gdm^+, 306 Cl^- and 694 water molecules (total components: 1306) and with an atomic number density of 0.095940 atoms/$Å^3$. The Gdm^+ structure was initially geometry-optimised using Gaussian, although neither this structure nor the water molecules remain rigid during the simulation.

Gdm^+ and Cl^- Lennard-Jones potentials and partial charges were identified from previous modelling studies in the literature (App. Table 39). Each iteration of the analysis was carried out with the SPC/E water model, the TIP3P water model and a 50:50 combination of both (App. Table 40).

The parameterisation study used F(Q) from the cooling crystallisation at 30 °C (Scan 29_114), as density data was available for this temperature.[56] The EP to optimise the fit was selected

[56] Due to the laboratories being unavailable during the Covid-19 lockdown, data required for EPSR could not be acquired. Solution density (30 °C) had been roughly measured prior to the *in situ* XPDF beam time (1.22 g/cm^3), so this value was used until the data could be checked experimentally.

using the EP search function in EPSR. The application of the parameter to fix the minimum allowed approach distance between atoms was also compared.

App. Table 39 Lennard-Jones potentials and partial charges for Gdm^+ and Cl^-

		q	σ /Å	ε /kJ.mol^{-1}	Reference
Gdm^+ (#1)	**C**	**0.7969**	**3.77**	**0.417**	**Ref.** [458] **with 20% charge reduction**
	N	**-0.7594**	**3.11**	**0.500**	
	H	**0.3802**	**1.58**	**0.088**	
Gdm^+ (#2)	C	0.9961	3.77	0.417	Ref. [458]
	N	-0.9493	3.11	0.5	
	H	0.4753	1.58	0.088	
Cl^- (#1)		**-0.8**	**3.77**	**0.619**	**Ref.** [268] **with 20% charge reduction**
Cl^- (#2)		-1	3.77	0.619	Ref. [268]
Cl^- (#3)		-1	3.65	0.83	Ref. [461]

App. Table 40 Lennard-Jones potentials for water models

		q	σ /Å	ε /kJ.mol^{-1}	Bond angle /°	Bond length /Å	Reference
SPC/E	O	-0.8476	3.166	0.65	109.5	0.1	Ref. [279][a]
	H	0.4238	0	0			
TIP3P	**O**	**-0.834**	**3.1506**	**0.6364**	**104.5**	**0.9572**	**Ref.** [462][a]
	H	**0.417**	**0**	**0**			

[a] Lennard-Jones parameter values detailed in [463]

A matrix of the EPSR simulations (App. Table 41) shows the simulations were evaluated using output parameters listed on the right of the table, along with visual inspection of the comparison of the experimental and calculated F(Q) and g(r). ***The reference potentials with reduced partial charges for Gdm^+ (#1) and Cl^- (#1) (App. Table 39) and TIP3P only, with no minimum approach distance, were selected for use in the simulations.*** Using a single water model type was preferred for some statistical analysis routines, which work based on atom types or atom numbers and would not easily accommodate two water types.

The methodologies for generating the experimental F(Q) (.int01 file) using GudrunX and running the EPSR simulation (including interpretation of the EP search function) were refined between the initial parametrisation EPSR simulations and the final simulations in Chapters 4 and 5. As a result, the EP in the parametrisation simulations are much higher than in the final simulations.

App. Table 41 Matrix of EPSR models for parameter selection

Model	Lennard-Jones & charges applied / other	Water model	EP	Chi-square	R-space coeff.	Quality factor	R-factor for data file	Noise in average g(r)	Inter-molecular E	Inter-molecular P	Average dipole moment
GuHCl_2	Gdm+ (#1) & Cl- (#1)	50/50	59	0.00128	0.00078	0.00113	0.00128	0.02662	27.274	58.219	1.945
GuHCl_2t	As GuHCl_2	TIP3P	63	0.00137	0.00046	0.00102	0.00137	0.01233	473.421	245.489	1.955
GuHCl_2s	As GuHCl_2	SPC/E	88	0.00148	0.00053	0.00115	0.00148	0.00950	1303.450	579.335	1.933
GuHCl_3	Gdm^{+} (#2) & Cl^{-} (#2)	50/50	49	0.00150	0.00097	0.00148	0.00150	0.00652	266.306	168.601	2.034
GuHCl_3t	As GuHCl_3	TIP3P	40	0.00146	0.00132	0.00171	0.00146	0.01697	370.736	213.636	2.068
GuHCl_3s	As GuHCl_3	SPC/E	60	0.00154	0.00093	0.00135	0.00154	0.01671	880.415	424.959	2.048
GuHCl_4	Gdm^{+} (#2) & Cl^{-} (#3)	50/50	52	0.00141	0.00087	0.00121	0.00141	0.00655	-156.762	-1.807	2.040
GuHCl_4t	As GuHCl_4	TIP3P	48	0.00158	0.00104	0.00156	0.00158	0.00353	321.362	200.133	2.044
GuHCl_4s	As GuHCl_4	SPC/E	71	0.00202	0.00071	0.00152	0.00202	0.01768	470.845	264.836	2.045
GuHCl_2v2	As GuHCl_2	50/50	28	0.00138	0.00100	0.00131	0.00138	0.00405	3.302	45.038	1.962
GuHCl_2tv2	As GuHCl_2	TIP3P	28	0.00112	0.00064	0.00097	0.00112	0.00373	55.967	67.912	1.962
GuHCl_2sv2	As GuHCl_2	SPC/E	28	0.00149	0.00093	0.00125	0.00149	0.00338	-120.586	-0.019	1.941
GuHCl_2v3	As GuHCl_2, min roverlap = 1.2	50/50	28	0.00136	0.00106	0.00147	0.00136	0.01182	116.019	91.774	1.984
GuHCl_2tv3	As GuHCl_2t, min roverlap = 1.2	TIP3P	28	0.00116	0.00080	0.00117	0.00116	0.00785	-62.965	20.154	1.957
GuHCl_2sv3	As GuHCl_2s, min roverlap = 1.2	SPC/E	28	0.00145	0.00101	0.00134	0.00145	0.00773	-72.401	18.709	1.925
GuHCl_2v4	As GuHCl_2v3 with updated AR	50/50	27	0.00162	0.00092	0.00128	0.00162	0.00893	171.955	113.987	1.979
GuHCl_2tv4	As GuHCl_2tv3 with updated AR	TIP3P	32	0.00134	0.00068	0.00132	0.00134	0.00913	67.206	74.866	1.972
GuHCl_2sv4	As GuHCl_2vs3 with updated AR	SPC/E		model unstable							
GuHCl_3v2	As GuHCl_3 but min roverlap = 1.2	50/50	48	0.00147	0.00104	0.00129	0.00147	51.01730	-107.844	20.515	2.036
GuHCl_3tv2	As GuHCl_3t but min roverlap = 1.2	TIP3P	32	0.00338	0.00177	0.00193	0.00338	0.04660	14.534	77.838	2.050
GuHCl_3sv2	As GuHCl_3s but min roverlap = 1.2	SPC/E	34	0.00190	0.00174	0.00216	0.00191	0.03246	242.939	163.115	2.041
GuHCl_4v2	As GuHCl_4 but min roverlap = 1.2	50/50	54	0.00216	0.00173	0.00221	0.00216	0.01388	561.717	289.225	2.036
GuHCl_4tv2	As GuHCl_4t but min roverlap = 1.2	TIP3P		model unstable							
GuHCl_4sv2	As GuHCl_4s but min roverlap = 1.2	SPC/E	36	0.00174	0.00170	0.00193	0.00174	0.01406	393.302	222.720	2.043
GuHCl_2v5	LJ - as 2v3, with RESP charges	50/50	29	0.00159	0.00098	0.00130	0.00159	0.00371	-40.826	25.292	1.973
GuHCl_2tv5	LJ - as 2v3, with RESP charges	TIP3P		holes in sim box - aborted							
GuHCl_2sv5	LJ - as 2v3, with RESP charges	SPC/E		holes in sim box - aborted							
GuHCl_2v6	As 2v5, C1 σ reduced (3.77 to 3.75)	50/50	20	0.00285	0.00188	0.00276	0.00285	0.16563	-190.599	-32.271	1.973

App. Table 42 Selected Lennard-Jones potentials and partial charges

		q	σ /Å	ε /kJ.mol^{-1}	Reference
Gdm^+ (#1)	C	0.7969	3.77	0.417	Ref. [458] with 20% charge reduction
	N	-0.7594	3.11	0.500	
	H	0.3802	1.58	0.088	
Cl^-	Cl^-	-0.8	3.97	0.619	Ref. [268] with 20% charge reduction and increased σ
TIP3P	O	-0.834	3.1506	0.6364	Ref. [462][a]
	H	0.417	0	0	

F.2 Parametrisation study 2

Molecule flexibility

The effect of changing EPSR parameters (App. Table 43) on the goodness of fit to the experimental data were considered for the concentration study. The quality factor was used to compare the fit of the simulations, as it is derived from the goodness of fits of g(r) and F(Q), although this value applies to the most recent iteration rather than an accumulated value and therefore could vary (increasing and decreasing) during the simulations. Some discretion was applied when comparing the fits.

App. Table 43 EPSR parameters in study

Molecule flexibility	Molecule flexibility has been studied by a) removing dihedral angles in the starting geometry, and b) adjusting the parameters *vibtemp*, *angtemp*, and *dihtemp,* which weight the constraints on the bond lengths, bond angles and dihedral angles, were set to (65/1/1) by default. Decreasing the default values creates a floppier molecule. The values (50/1/1), (45/0.7/0.7) and (30/0.5/0.5) were considered.
Lennard-Jones potentials for C	σ and ε values for C adjusted from σ: 3.77 Å and ε: 0.417 kJ/mol to - σ: 3.87 Å and ε: 0.412 kJ/mol - σ: 3.67 Å and ε: 0.422 kJ/mol
Range of potential truncation	This parameter affects the radius over which the non-Coulomb part of the reference potentials take effect. The potential truncation (r_{min} = 9 Å and r_{max} = 12 Å) were changed to r_{min} = 6 Å and r_{max} = 9 Å and r_{min} = 9 Å and r_{max} = 15 Å.

Summary of results: removing the dihedral angles resulted in a molecule that was too floppy and did not improve the simulations. A more flexible molecule by setting *vibtemp*, *angtemp*, and *dihtemp* to 45/0.7/0.7 optimised the F(Q) and g(r) fits. This approach keeps the rigidity of the Gdm^+ structure, whilst accounting for the fact that in solution, there may be some variation in the structure due to the changing environment that the ion will be experiencing. The original σ and ε values for C (σ: 3.77 Å, ε: 0.417 kJ/mol) provided the best fits, keeping them consistent with the *in situ* analysis. The default potential truncation range produced the best quality fits.

Appendix G Concentration study: repeat simulations

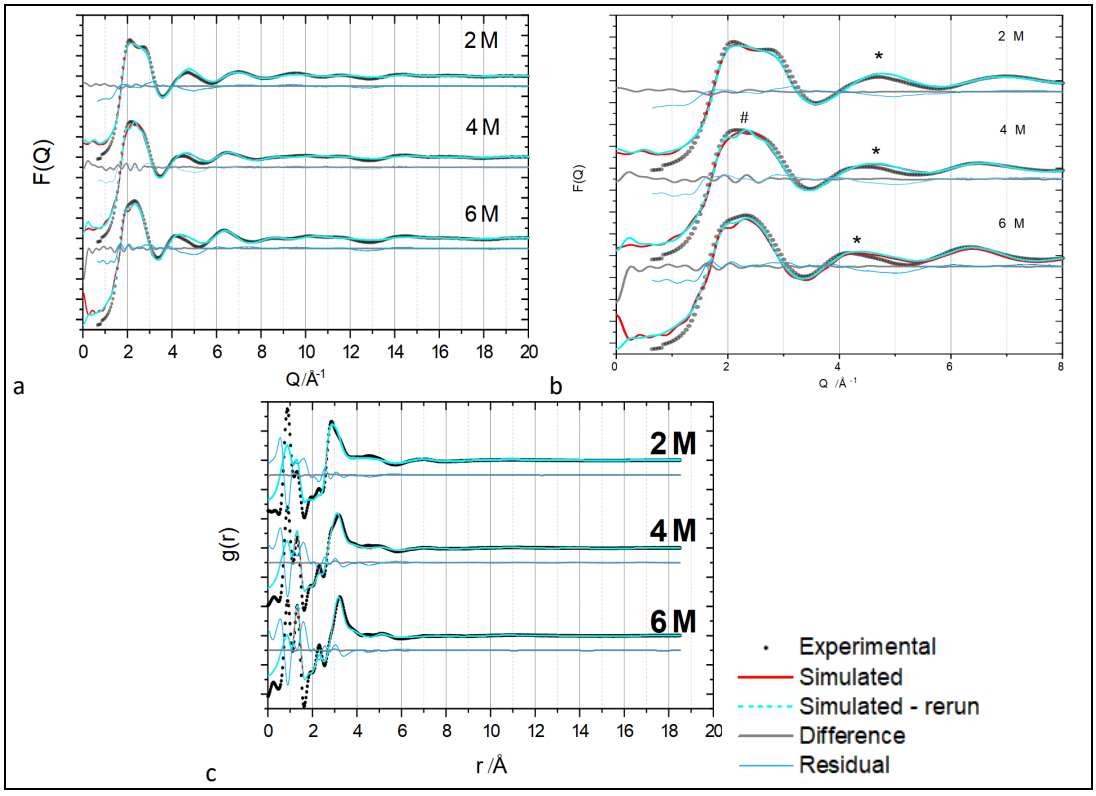
G.1 Repeat simulations: 2, 4 & 6 M GuHCl solutions

Due to the random nature of a Monte Carlo simulation, and the fact that an ensemble of molecular structures could be consistent with a single X-ray total scattering pattern [489], it is not unreasonable that there may be more than one set of correlations. This fact simply reflects the complexity of modelling a many-body system, where forces acting in the condensed phase are higher-order than simply pairwise-additive [489].

To examine this possibility, the EPSR simulations for the 2, 4 and 6 M GuHCl solutions were rerun. Each molecular model was re-generated with randomised position parameters, and restarted with the box expanded to allow the molecules to shift in case they had been stuck in a configuration that was a local energy minimum, i.e. following the same procedure as used the original models (Section 5.2.2).

The data from the original simulations have been included in the GuHCl concentration analysis in Chapter 5.

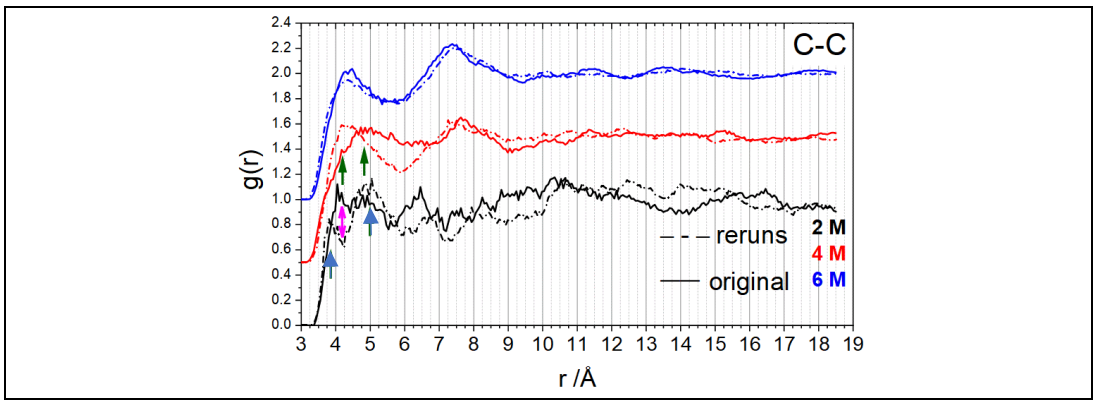
F(Q) and g(r): The simulated F(Q) and g(r) are compared with the experimental data in App. Figure 56a,b&c.



App. Figure 56 F(Q) and g(r) of 2, 4 and 6 M original and repeated simulations

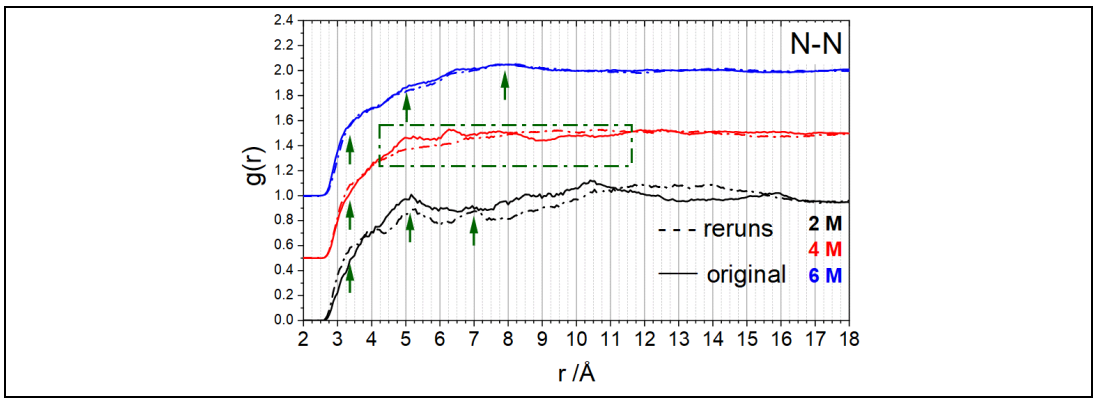
The F(Q) are similar, as the dashed cyan line for the rerun almost overlays the original simulated F(Q) in red. Both simulations have F(Q) peaks slightly to the right of the experimental data at ~4.5 $Å^{-1}$ (* on App. Figure 56b). The peaks on the feature at ~2.0 $Å^{-1}$ and 2.7 $Å^{-1}$ on the 4 M F(Q) are picked up in the original and rerun simulation, but there is an additional peak in the rerun 4 M F(Q) at ~2.3 $Å^{-1}$. The original and rerun simulated g(r) overlay each other across the r-range (App. Figure 56c).

The $g_{C\text{-}C}(r)$ of the original and repeated simulations in App. Figure 57 show that the two 6 M simulation $g_{C\text{-}C}(r)$ up to ~9 Å have similar peak positions and intensities. In the r < 5 Å region of the 4 M $g_{C\text{-}C}(r)$, the original simulation has a shoulder feature at ~4.2 Å and a peak at ~4.8 Å, whereas in the rerun, the positions are similar, but the peak are shoulder are reversed (green arrows, App. Figure 57). This data suggests that the interactions are occurring in similar positions, although the intensities vary. The peaks and minima at 7.5 Å and 9 Å in the 4 M and 6 M simulations are in similar positions, after which the $g_{C\text{-}C}(r)$ intensity is ~1. The two 2 M $g_{C\text{-}C}(r)$ show the least correlation. At ~3.8 Å, the shoulder feature in the original simulation is at the same point as the peak in the rerun simulation, and both have peaks at ~5 Å (blue arrows), but the feature at 4.2 Å is a peak in one and a minimum in the other (pink arrow). After 5.5 Å , there is little correlation.



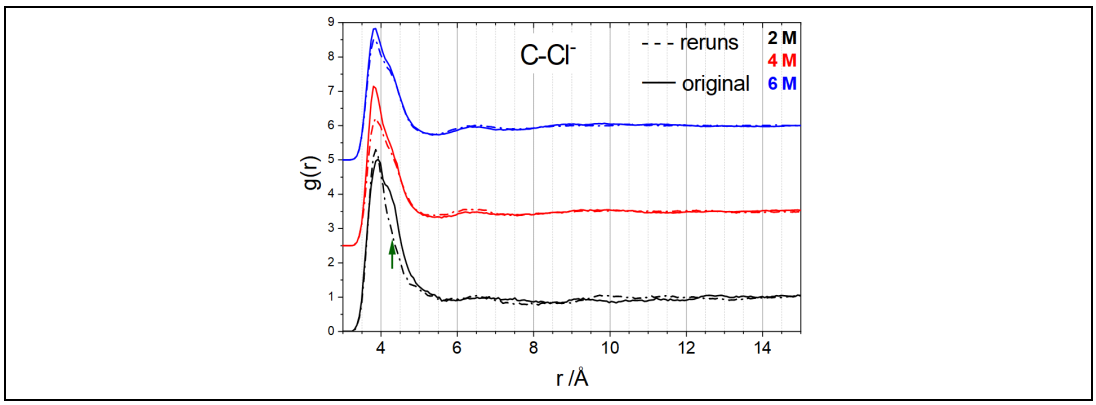
App. Figure 57 Repeat 2, 4 and 6 M simulations: C-C partial pair distribution functions $g_{C\text{-}C}(r)$

In the $g_{N\text{-}N}(r)$ (App. Figure 59), the shoulder features indicated with arrows at ~3.4 Å suggest there are interactions at this distance in all concentration solutions (less clear in the original 2 M solution). Whilst arrows indicate similar features in the original and rerun simulations for 2 and 6 M solutions, the rerun 4 M simulation $g_{N\text{-}N}(r)$ has few features up to 9 Å, after which the intensity is ~1 (green dotted box).



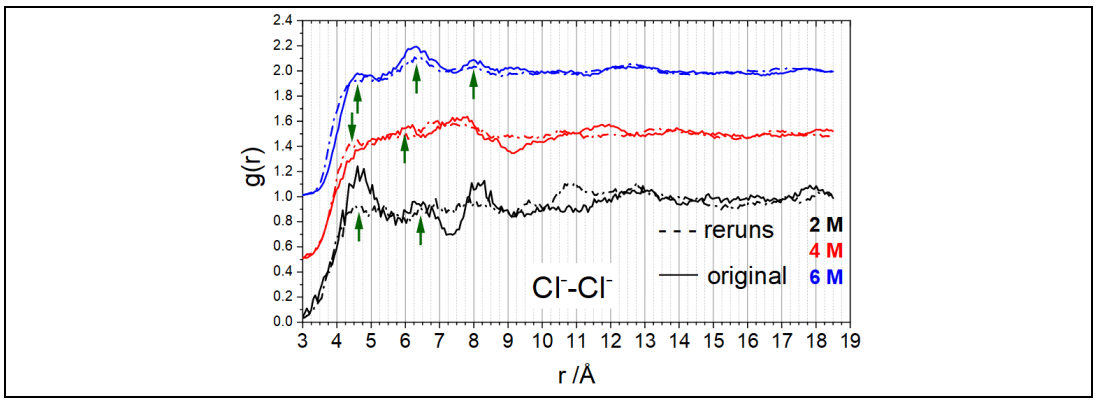
App. Figure 58 Repeat 2, 4 and 6 M simulations: N-N partial pair distribution functions $g_{N\text{-}N}(r)$

Ion-pairing is indicated in each of the concentrations in both the original and rerun simulations. The intensities of the 3.8 Å peaks in the 4 and 6 M $g_{C\text{-}Cl^-}(r)$ differ (lower in the reruns), but peak positions are the same (App. Figure 59). Small peaks at ~6.5 Å are captured in both simulations at 4 and 6 M. Again there is greater variation in the 2 M simulations, as the clear shoulder in the original simulation at 4.3 Å is less obvious in the rerun.



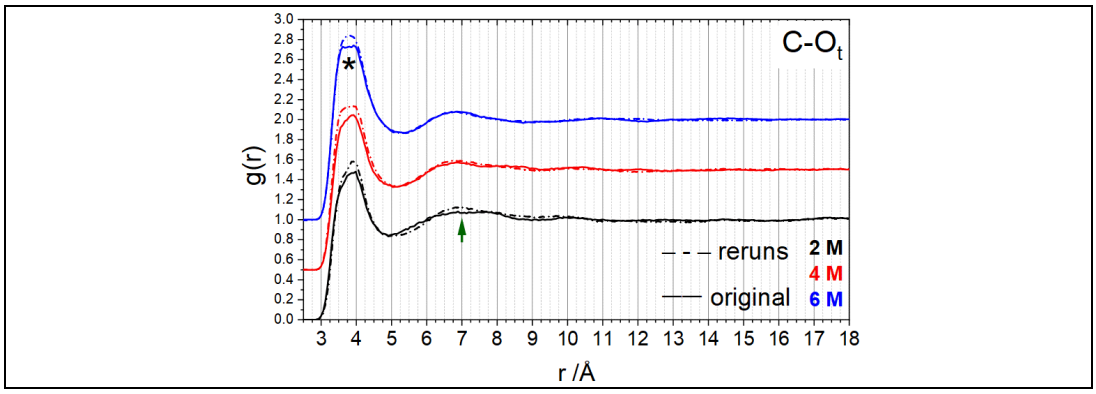
App. Figure 59 Repeat 2, 4 and 6 M simulations: C-Cl^- partial pair distribution functions $g_{C\text{-}Cl^-}(r)$

Similarities in the 6 M $g_{Cl^-\text{-}Cl^-}(r)$ extend over a greater r than in the 2 and 4 M solutions (App. Figure 60). In the 2 M solution, the peak positions up to ~7 Å are similar, but the relative intensities are not the same, as noted in $g_{C\text{-}C}(r)$ and $g_{C\text{-}Cl^-}(r)$. Although similar up to 6 Å, the 4 M $g_{Cl^-\text{-}Cl^-}(r)$ show less correlation beyond this point.

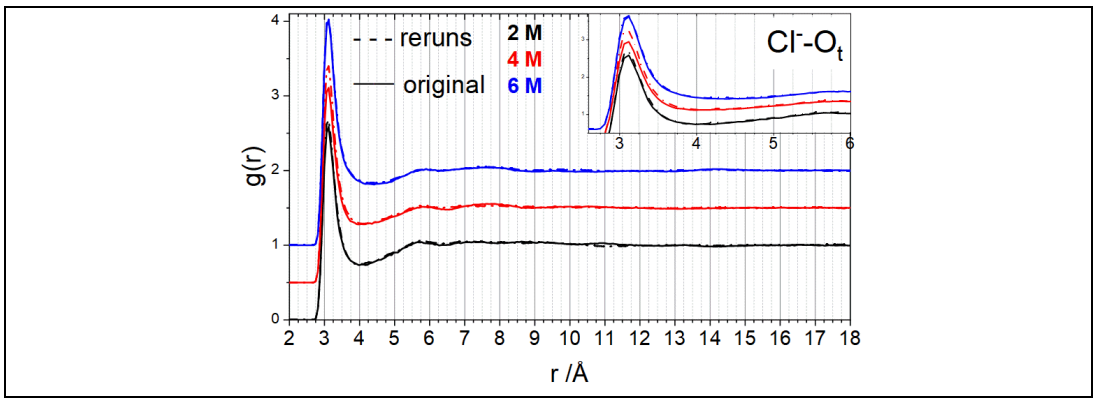


App. Figure 60 Repeat 2, 4 and 6 M simulations: Cl^--Cl^- partial pair distribution functions $g_{Cl^-\text{-}Cl^-}(r)$

The most significant difference in the $g_{C\text{-}O}(r)$ first peak shapes and intensities are seen in the 6 M first coordination shell, where the peak is practically flattened in the original 6 M simulation (App. Figure 61). From ~5 Å, the higher concentration simulations $g_{C\text{-}Ot}(r)$ are more comparable than the 2 M simulations. The irregularities around 7 Å may be due to there being fewer Gdm^+-water interactions, as there is a much higher water:ion ratio in the 2 M solution. The $g_{Cl^-\text{-}Ot}(r)$ again have intensity differences in the 4 M simulation in the first peak, and the rerun simulation $g_{Cl^-\text{-}Ot}(r)$ is fairly flat from 6 Å, although there are small features in the original simulation (6–9 Å, App. Figure 62).

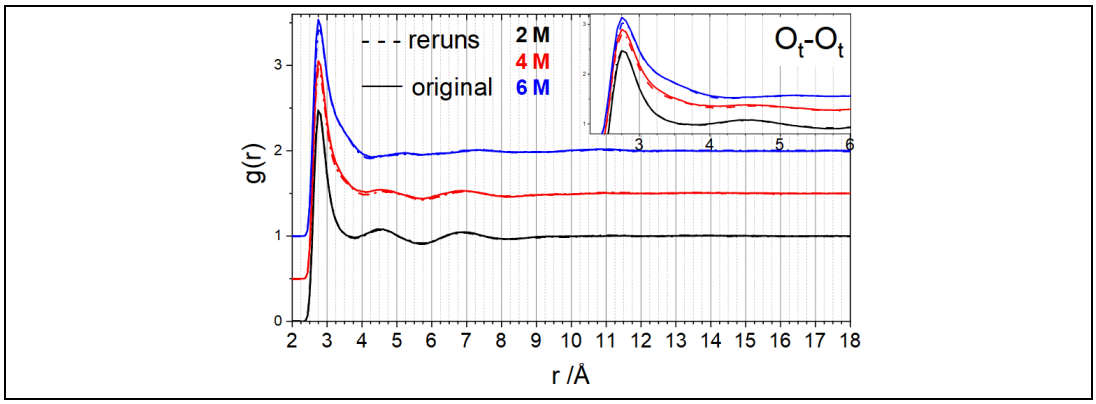


App. Figure 61 Repeat 2, 4 and 6 M simulations: C-O_t partial pair distribution functions $g_{C\text{-}Ot}(r)$



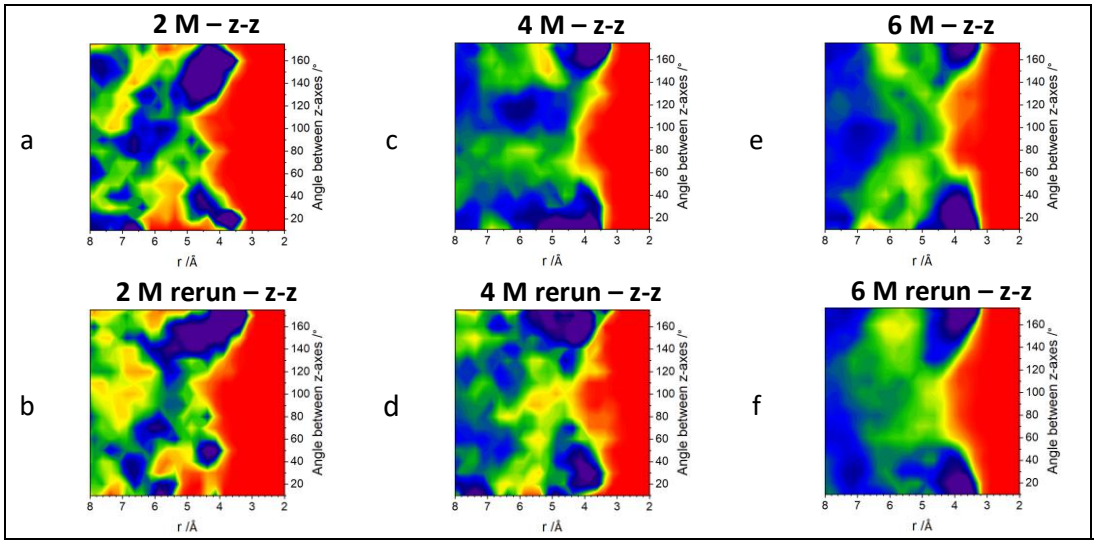
App. Figure 62 Repeat 2, 4 and 6 M simulations: Cl^--O_t partial pair distribution functions $g_{Cl^-\text{-}Ot}(r)$

With the greatest number of water molecules in the simulation box, the 2 M $g_{Ot\text{-}Ot}(r)$ practically overlay each other. There is a slight intensity difference in the first peak of the 4 M simulation $g_{Ot\text{-}Ot}(r)$, and the 6 M $g_{Ot\text{-}Ot}(r)$ are similar.

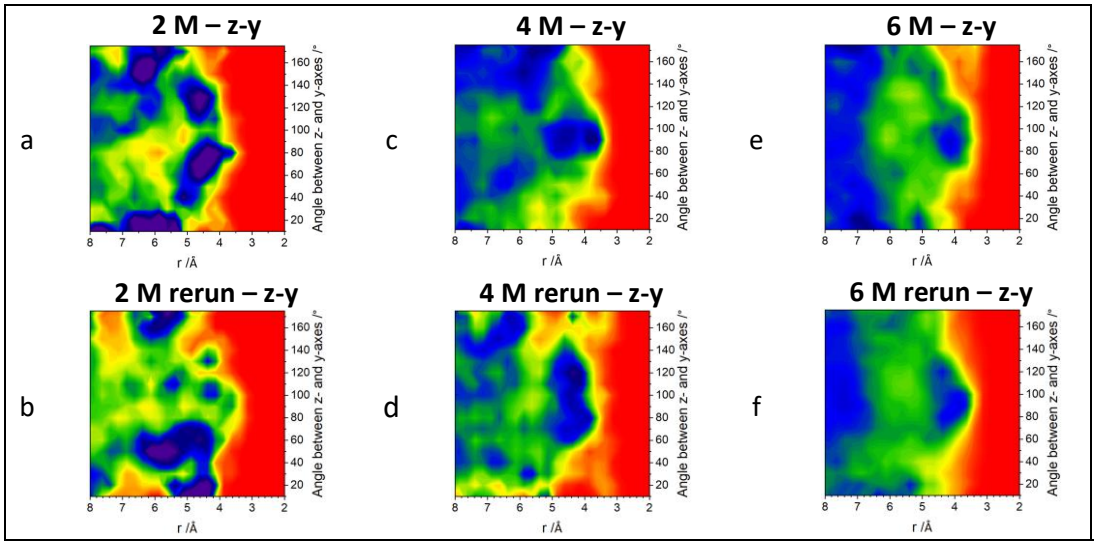


App. Figure 63 Repeat 2, 4 and 6 M simulations: O_t-O_t partial pair distribution functions $g_{Ot\text{-}Ot}(r)$

Angular radial distribution functions (ARDF) of Gdm^+-Gdm^+: ARDF analysis of the rerun simulations was also undertaken to compare with the original simulation ARDF. For ease of comparison, the ARDF from Figure 5.5 are repeated side-by-side with the ARDF from the rerun simulations (App. Figure 64 and App. Figure 65). There is a strong similarity in the 4 and 6 M simulation ARDF for the z-z and z-y analyses, probably because the number of Gdm^+-Gdm^+ interactions is greater at the higher concentration.



App. Figure 64 Gdm^+-Gdm^+ angular radial distribution function between the z-axes of Gdm^+ ions for the original simulations of a) 2, c) 4 and e) 6 M solutions, and the rerun simulations of b) 2, d) 4 and f) 6 M solutions



App. Figure 65 Gdm^+-Gdm^+ angular radial distribution function between the z-axis if the central Gdm^+ and the y-axes of the surrounding Gdm^+ ions for the original simulations of a) 2, c) 4 and e) 6 M solutions, and the rerun simulations of b) 2, d) 4 and f) 6 M solutions

Summary of rerun simulations:

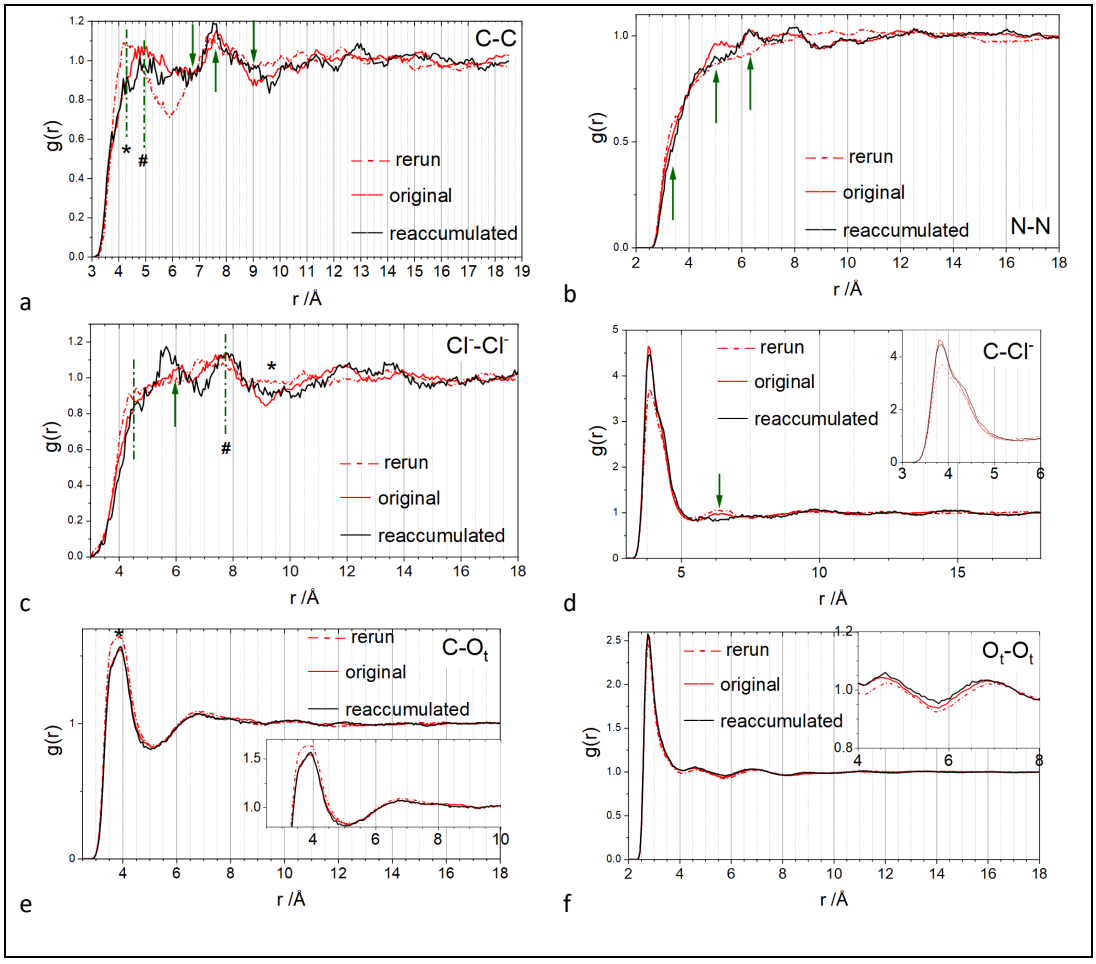
- F(Q) and g(r) for simulations refined to the same experimental data set are almost identical, except from some low-Q deviation for the 4 M solution (App. Figure 56).
- The greatest differences occur in the $g_{\alpha\text{-}\beta}(r)$ involving GuHCl interactions. It could be due to the large number of water molecules and therefore the relatively low number of interactions between Gdm^+ and Cl^- in the simulation box.
- Most partial g(r) show reasonable agreement between the same concentration models up to 5 Å, although the 2 M $g_{C\text{-}C}(r)$ did show deviation at ~4 Å, a key atomic separation when considering the potential for Gdm^+-Gdm^+ stacking.

- The 4 M showed greatest first peak intensity variation in the g(r) with relatively sharp first peaks ($g_{C\text{-}Ot}(r)$, $g_{Cl^-\text{-}Ot}(r)$ and $g_{Cl^-\text{-}C}(r)$). This higher apparent variability may be related to the small sample of Gdm^+ molecules at 2M.

G.2 Reaccumulated 4 M GuHCl solution simulation

Across the interactions reviewed above, the 4 M simulation data tend to show larger intensity differences compared to the 2 and 6 M solutions where there is a large first coordination peak (App. Figure 59, App. Figure 61 and App. Figure 62).

The 4 M EPSR simulation was rerun from the accumulation stage to assess changes to the $g_{\alpha\beta}(r)$. The resulting partial g(r) for the original simulation (Simulation 1), the rerun simulation (Simulation 2) and the reaccumulated simulation (Simulation 3) can be seen in App. Figure 66a-f.



App. Figure 66 Partial pair distribution functions $g_{\alpha\beta}(r)$, where α and β are a) C-C, b) N-N, c) Cl^--Cl^-, d) C-Cl^- e) C-O_t and f) O_t-O_t for 4 M GuHCl solution (original, reaccumulated and rerun) simulations

The positions of the features in $g_{Cl^--Cl^-}(r)$ up to 8 Å are similar, even though the intensities are different. At 9 Å, minima in Simulations 1 and 3 correspond to a small peak in Simulation 2. The same is seen in $g_{C\text{-}C}(r)$, where * and # indicate peak positions in all three simulations, but of varying intensities (App. Figure 66a). The large minimum in Simulation 3 ~6 Å is not as extreme in the other two simulations.

It was noted earlier that the $g_{N\text{-}N}(r)$ for Simulation 2 had fewer features than Simulation 1. Although Simulation 1-3 all have a shoulder feature (~3.5 Å: first arrow on App. Figure 66b), the next main feature at 5 Å (second arrow) is not present for Simulation 2 and barely there for Simulation 3, whereas at ~6.2 Å, the peak is only present for Simulations 1 and 3.

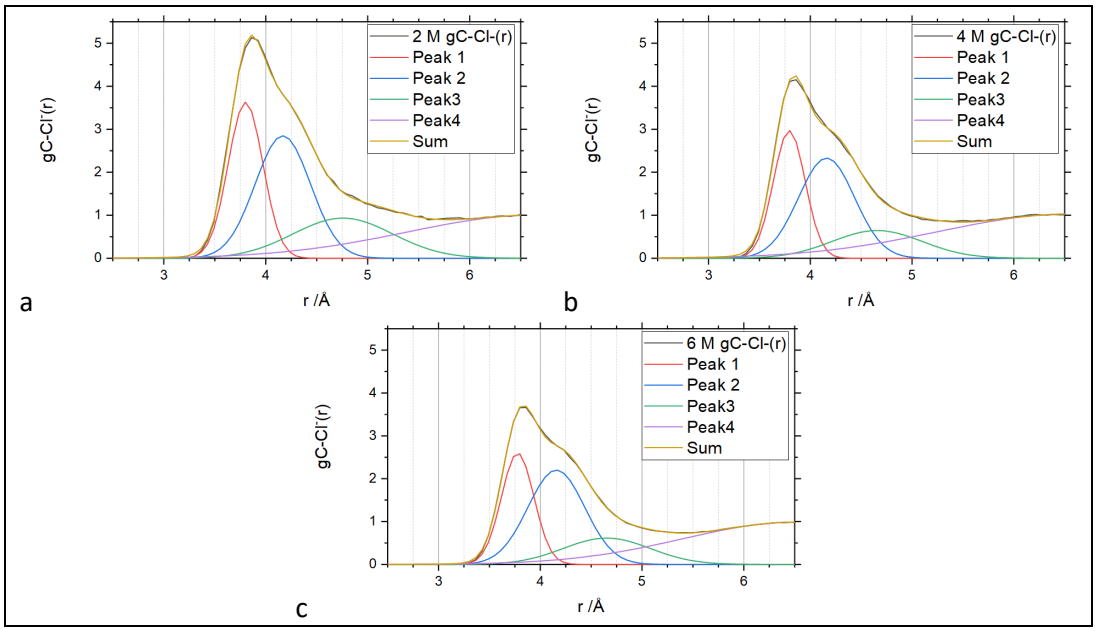
The $g_{\alpha\beta}(r)$ Sims 1 and 3 are similar for C-O_t and O_t-O_t in terms of intensity and features (App. Figure 66e&f). In App. Figure 66c, the $g_{C\text{-}Cl^-}(r)$ first coordination peaks of the Simulations 1 and 3 are similar in intensity, but the peak at 6.3 Å is present for the Simulations 1 and 2, but not Simulation 3.

Summary of 4 M original, rerun and reaccumulated simulations:

- There is variation between the partial g(r) from all three simulations. The data accumulated from the same starting EP molecular model are similar in some places and show variation in others.
- In most partial g(r), up to 5 Å the peak or shoulder positions are at quite similar atomic spacings. In most of the ion-ion partial g(r) > 5 Å, there is strong variation in the peak positions and intensities in some or all simulations.
- The $g_{C\text{-}C}(r)$ look dissimilar, even below 5 Å, although the positions of some features are similar even if the intensity varies, suggesting that similar interactions may be present in all three simulations but to different extents.

G.3 C-Cl$^-$ peak fitting

Peaks were fitted to the first ~6 Å of the $g_{C\text{-}Cl}$-(r) (App. Figure 67) using Fityk [493].



App. Figure 67 $g_{C\text{-}Cl^-}$ (r) for a) 2 M, b) 4 M and c) 6 M GuHCl solution simulations: peaks have been fitted with Fityk