Developing Structure Determination from Powder X-ray Diffraction Data for Applications in Molecular Organic Materials

Christopher J.H. Smalley

Thesis submitted for

Doctor of Philosophy

School of Chemistry

Cardiff University

December 22, 2022

Personal Acknowledgements

I would like to dedicate this thesis to my Mum, without her I would not be the person I am today and would not have had a chance of getting here. Thank you for your constant support both in my PhD and elsewhere, I try to make you proud every day - I love you. I also want to partly dedicate this to my late grandparents Peter and Margaret as well as my Aunty Barbara whom I always try to make proud in everything I do, we all miss you but will never forget you.

I would like to thank my family including Pete, Tasha, Mick, Peter, Philip, Gill and Dad as well as the rest of my extended family not mentioned here. Thanks for all the laughs, and for listening to all my complaining throughout my PhD, I swear I will shut up one day.

Thank you to our dogs on the farm Sam, Bugs and Milo whilst not forgetting Jack, Nelly and Ziggy - you always keep me smiling.

I would like to thank Bowland High School for helping mold me into the person I am today. Specifically, Danny Breen and Saranne Costigan from the Science Department. The confidence and belief you inspired me with will never be forgotten, without both of your support I would not have been able to pursue my love for science to reach where I am.

I want to take this chance to thank my friends in no particular order. Thank you to Emily Nash, Pippa Patterson, Lee Stanley, Joanne Stanley, Chloe Stanley and Aimee Stanley for always being there for a laugh and a chat. Charlie Forrest, Ben Mason, Dom Ward, Jamie Carway, Jordan Mackreal, Ryan Booth and Tommy Bennett cheers boys for the races and the dubs. Aly Ahmed, Connor Maguire, Lauren Bailey, Bridie Collins, Rachel Mathomes, George Dutton, Dill Chuyat, Josh Eckersley and Kira Heslop - thanks guys for being such good friends. I will never give you up, let you down, run around or desert you.

A particular thank you must go to Will Stockburn, who has always been there to provide advice and guidance in both my academic life and outside of it. It truly is appreciated, thank you for always listening.

I want to thank my colleagues at Bruker for their support in my final months of finishing my PhD. To Mick Carr, Cheryl Haidon and Archie Harris, thanks for putting your faith in taking me on as a member of the team, and for helping me in getting my PhD finished whilst working with you all, it's a pleasure to be part of the team. I must extend my thanks to the rest of my colleagues at Bruker UK and at Bruker AXS for making me feel welcome.

I would be remiss not to thank the following actors, and authors for providing me entertainment during my research project which kept me going throughout this time: Stewart Griffin, B. B. Dook, Theodore E. Mosby and Kenneth McCormick.

Professional Acknowledgements

I would like to give thanks to my supervisor Kenneth Harris for his supervision and constructive advice throughout the course of this PhD project. It has been very insightful working in his research group and has provided me with valuable experience for my career following the completion of this project. I give thanks to Mark Young for providing additional supervision in this project and always being willing to assist when needed. I would also like to thank Colan Hughes for his assistance in aspects of this project, particularly for assisting in collecting solid-state NMR data and helping with its analysis as well as helping me at the beginning of my PhD.

My thanks extend to the PGR staff at Cardiff University. George, Caru and Moira who supported me beyond belief throughout the challenging stages of my PhD, without them I would have not completed this and they are one of the greatest assets to Cardiff University. It is without a doubt that the PGR administrative staff provide an exceptional service to all PGR students in the School of Chemistry and their support cannot go unnoticed. Thank you all so much.

I would like to extend my thanks to Benson Kariuki for his generous support, going very much out of his way to help me throughout the project. Thank you for helping with all the stupid questions, and for providing extensive advice about research and life. I would like to thank the members of the research group whom I have worked with in my time at Cardiff University including Andrew Williams, Rhian Patterson, Okba Al-Rahal, Hatty Hoskyns, Poppy Hindle, Emma Green, Quentin Bernard and Ffion Cartwright, it was great working with you all and engaging with the research group spirit. Thanks also to the other members of the office, particularly Ashish, Alan and Adriana for being so welcoming in the office and always having a chat.

Thank you to Andy Logsdail for performing many DFT-D calculations on our behalf in this work, and for his very clear communication of these results to me. Thank you to Chris Pickard and Duncan Johnstone for performing structure prediction calculations on our behalf. Thank you to Duncan Johnstone, Tom Willhammar and Paul Midgley for your expertise in the 3D electron diffraction measurements and analysis. I would also like to give thanks to the UK 850 MHz Solid-State NMR Facility (funded by the EPSRC and BBSRC, as well as the University of Warwick including via part funding through Birmingham Science City Advanced Materials Projects 1 and 2 supported by Advantage West Midlands and European Regional Development Fund) for use of their 850 MHz solid-state NMR spectrometer and to Dinu Iuga for his assistance at the facility. I would also like to thank the ARCCA team who created and continue to support the HAWK supercomputer, this work could not have been conducted without their facility.

Abstract

The research presented in this PhD thesis is based on the theme of crystal structure determination from powder X-ray diffraction (XRD) data. More specifically, the work here advances existing methodologies primarily by implementing information from other complementary structural techniques such as solid-state NMR spectroscopy, density functional theory calculations, structure prediction calculations, and electron diffraction data. The materials studied here are a range of molecular organic solids, ranging from well-known pharmaceuticals to materials with interesting optoelectronic properties.

Chapter 1 provides information on the background of the materials of interest in this study. The phenomenon of polymorphism is discussed as well as further information regarding biominerals and amino acids, whilst concluding with the aims of the research presented in this work.

Chapter 2 details the experimental techniques used in the scope of this work. This includes powder X-ray diffraction, solid-state NMR spectroscopy, density functional theory calculations, electron diffraction, thermogravimetric analysis, and crystal structure prediction calculations.

Chapter 3 discusses the conventional approach of using powder XRD data for crystal structure determination. This details the entire process, from peak selection and indexing all the way through to structural refinement by the Rietveld method, with subsequent structural validation using a variety of experimental and theoretical methods.

Chapter 4 provides insight into the crystal structure of alloxazine and riboflavin. In the riboflavin section, the effort in determining the crystal structure is detailed, providing an alternative structure to that published by Guerain *et al*. with evidence from complementary techniques to discriminate between the two. In the alloxazine section, the crystal structure was determined with relative ease; however, determining the tautomeric form present within the crystal structure (alloxazine or isoalloxazine) required using other experimental and theoretical methods.

Chapter 5 details the discovery and subsequent crystal structure determination of a new polymorph of L-tyrosine using powder XRD data and electron diffraction data. The crystal structure was validated using solid-state NMR data and periodic DFT-D calculations. The use of crystal structure predictions calculations was explored, leading to the discovery of multiple potential polymorphs of L-tyrosine.

Chapter 6 details the elucidation of the crystal structure of racemic lysine. Several hydrates are also discovered, with their crystal structures determined. One of the hydrates is isostructural to the pure form and exhibits disorder between two arrangements; periodic DFT-D calculations were used to successfully model this disorder.

Chapter 7 describes the preparation of a new polymorph of the popular analgesic ibuprofen. The crystal structure of this polymorph is determined from synchrotron powder XRD data and subsequently validated using periodic DFT-D calculations, with subsequent comparison to the two known polymorphs. The crystal structure is noticeably higher in energy than both two known polymorphs, determined through periodic DFT-D study.

Finally, Chapter 8 describes two materials, DBTMA and PDICH, and the respective strategies used to determine their crystal structures, primarily using powder XRD data. They both show molecular similarity defined by rigid cores. Periodic DFT-D calculations were imperative in scrutinizing the molecular geometry of the DBTMA molecules, where powder XRD was inadequate.

In conclusion, this thesis has described how structure determination from powder XRD data can be augmented by complementary methods to extract direct or indirect structural information. By employing methods such as 3D-ED data, it has been shown how to avoid typical bottlenecks such as indexing or structure solution, which are often impossible to overcome with imperfect data. Theoretical techniques such as periodic DFT-D calculations have been employed to probe plausible structures, and to rigorously validate or correct them, which is extremely difficult to perform with powder XRD data solely.

Future work to follow on from this thesis would clearly have to focus on two aspects. One aspect being the application of the methodology presented in this thesis towards molecules of different natures, or more complex natures to those studied here. This can include inorganic or organic extended solids such as zeolites, metal-organic frameworks, or covalent organic frameworks. However, the studies of molecular organic solids with a significantly greater structural complexity e.g. peptides with ten or more amino acids, would be a huge challenge for this methodology. The other aspect would be to develop the methodology presented in this thesis even further. For example, optimizing the experimental set-up for 2D-PXRD to allow for more ideal data to be collected and used in the structure determination process. Another area for significant development would be the use of crystal structure prediction in the indexing stage, for example predicting a variety of crystal structures and using the information of their hypothetical unit cells and crystallographic symmetry to aid the indexing procedure, particularly in the cases of impure data or poorly crystalline samples.

Contents

Chapter 1
Introduction

Molecular organic solids are some of the most common materials in both nature and technology. They have applications in a wide variety of industries such as pharmaceuticals, agrochemicals, and cosmetics [1]. They often exist in a crystalline form and their physicochemical properties can be rationalized from knowledge of their crystal structure. Unfortunately, sometimes it is difficult or even impossible to prepare single crystals of these materials to allow the crystal structure to be determined through the conventional method of using single-crystal X-ray diffraction (XRD). When this is the case, other methods must be applied to determine the crystal structure; powder XRD is a particularly useful alternative in this regard [2]. In the last 25 years, since the advent of the direct-space strategy for structure solution from powder XRD data [3], the use of powder XRD for crystal structure determination of organic materials has become more common, although it is still usually only conducted by expert crystallographers; single-crystal XRD is more readily accessible to non-experts due to higher automation and "black box" procedures in the data collection and analysis stages. In recent years, the use of additional methods such as solid-state NMR [4, 5] and/or periodic DFT-D calculations [6] to supplement the process of crystal structure determination from powder XRD data has been initiated.

1.1 Polymorphism

Polymorphism is defined as the ability of a given molecule to exist in two or more different crystal structures, with different crystal packing, and in some cases, different molecular conformations [7]. These structures can be vastly different or closely related, but polymorphs have an identical chemical and molecular composition in all cases. As a consequence of having different crystal structures, polymorphs often exhibit different properties. These can include, but are not limited to, melting point, solubility, and crystal colour [8]. Transformations between polymorphs can occur, usually from a higher energy polymorph to a lower energy polymorph, although high energy polymorphs can still exist as metastable phases if no favourable kinetic pathway exists for the transformation to a polymorph of lower energy [8]. Polymorphism is common for molecular organic materials. For example, in the case of pharmaceutical materials, it has been shown that polymorphism is observed in approximately 50% of cases [9]. To illustrate the phenomenon of polymorphism, three examples are now considered: i) 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophene-carbonitrile (ROY), which exhibits differences in colour between polymorphs; ii) 1,3,5-trinitrobenzene (TNB), for which the polymorphs were prepared through additives within the

crystallization procedure; iii) Ritonavir, for which the polymorphs show vastly different solubilities. Currently, the material with the greatest number of structurally characterized polymorphs is known famously as ROY [10], the molecular structure of which is shown in Figure 1.1. This molecule has been proven to exist in 12 different polymorphic states [10], of which all 12 have crystal structures determined. Some properties of the first 10 identified polymorphs of ROY are shown in Figure 1.2, which include the crystal colour and morphology.

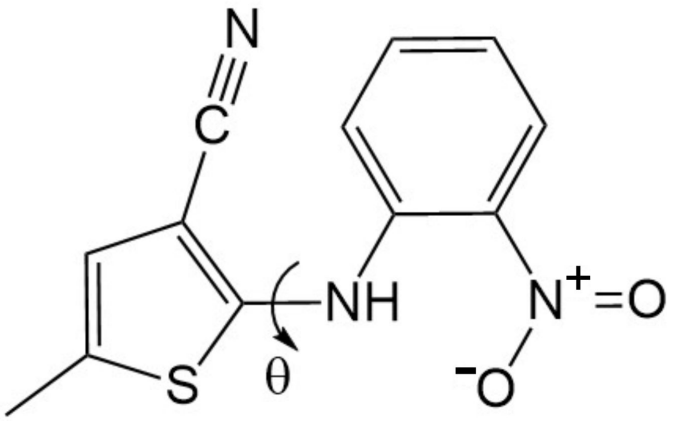

Figure 1.1: The molecular structure of 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophene-carbonitrile (ROY). The torsion angle, θ, between the thiophene ring and the rest of the system has a considerable influence on the colour of the different polymorphs.

As seen in Figure 1.3, crystalline TNB can exist as three different polymorphs [11]. Their crystal structures are similar, and all contain an axis close to 9 Å in length, with a center of inversion. For many years, the existence of polymorphs II and III was known but their crystal structures were undetermined. After many years they were determined, and with the use of lattice energy calculations it was established that the first known structure (form I) was the least energetically stable, with form II being intermediate in stability, and form III being the most stable [11]. Polymorphs II and III were prepared using an additive during crystallization, which can facilitate the preferential formation of a particular polymorph. The additive in this example was trisindane (TI), a phenyl ring fused with three cyclopropyl groups at the 1-2, 3-4 and 5-6 positions, which is structurally similar to the TNB molecule. But in this case, the highly polar nitro-groups are replaced by non-polar cycloalkane groups. It is theorized that the TI additives indiscriminately adsorb onto the faces of growing TNB crystals during crystallization, due to their similar three-fold molecular symmetries [11]. This example clearly demonstrates that the most stable polymorph is not necessarily the easiest to prepare experimentally.

It is important to note that polymorphism is not only an interesting chemical phenomenon but it has a significant impact on a variety of industrial processes, most notably in the pharmaceutical industry [12]. As previously mentioned, different polymorphs exhibit different physicochemical properties such as different solubility, which in the context of pharmaceuticals can lead to increased

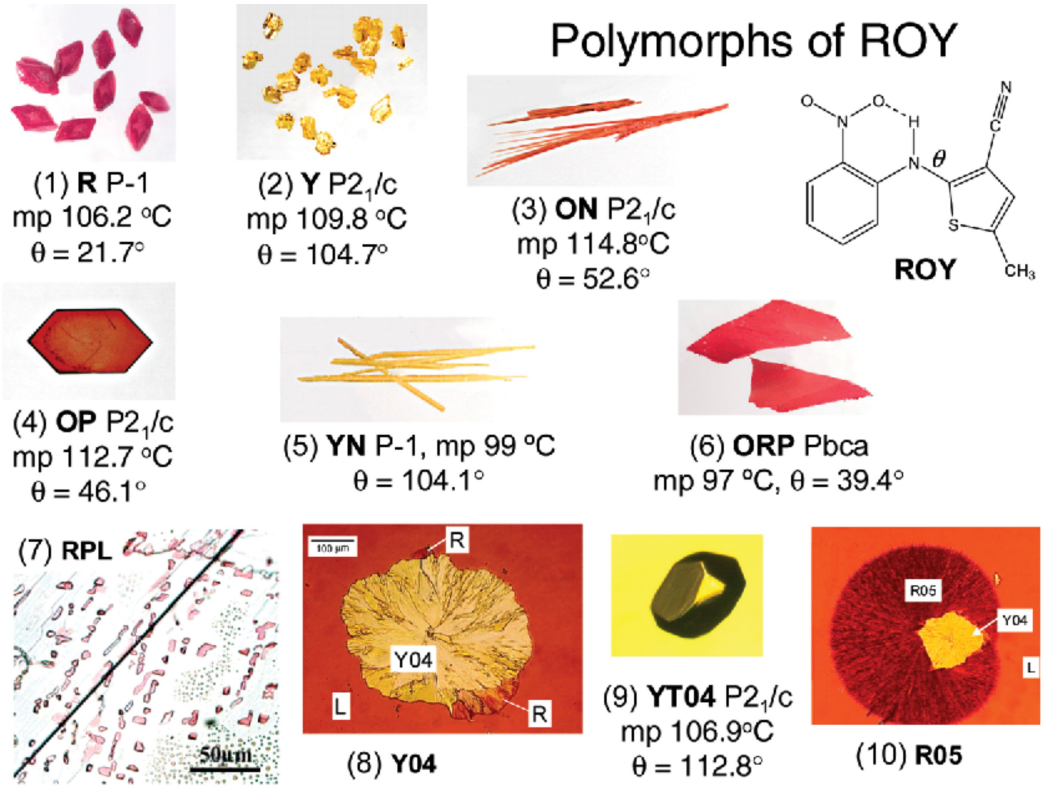

Figure 1.2: The polymorphs of ROY and their melting points, crystal structure space group and θ value in the crystal structure (where known). Reprinted with permission from L. Yu Acc. Chem. Res., 43, 1257-1266 (2010). Copyright 2010 American Chemical Society.

bioavailability of an active pharmaceutical ingredient (API) and increased potency. Ritonavir is the classic example of the importance of understanding polymorphism of pharmaceutical materials. The drug is a protease inhibitor for the treatment of HIV and is not bioavailable in the solid-state and so it was formulated as a liquid and a gel, dispersed in a solution of ethanol in water [13]. However, within two years following commercial production, the drug transformed from the known polymorph (later identified as metastable) into a more stable polymorph, previously unknown. This new, more stable polymorph was far less soluble than the commercial drug material. Consequently, when a small amount of this new polymorph was present in the production process, it facilitated a transition of the whole batch to the new polymorph, massively decreasing solubility. The drug material was rendered unviable and required the development of a new formulation procedure. Crystal structure prediction has also been explored for other pharmaceuticals to avoid this type of issue in the future, which concluded that, in 15% - 45% of cases, the most stable solid form of the drug has not yet been observed experimentally [14], and therefore the known experimental phases of many pharmaceuticals are likely metastable.

The preparation of new, sometimes metastable polymorphs can require more innovative preparation procedures. Rather than a typical recrystallization, in which material is dissolved in a small amount of hot solvent and then cooled to induce precipitation, less conventional approaches are sometimes

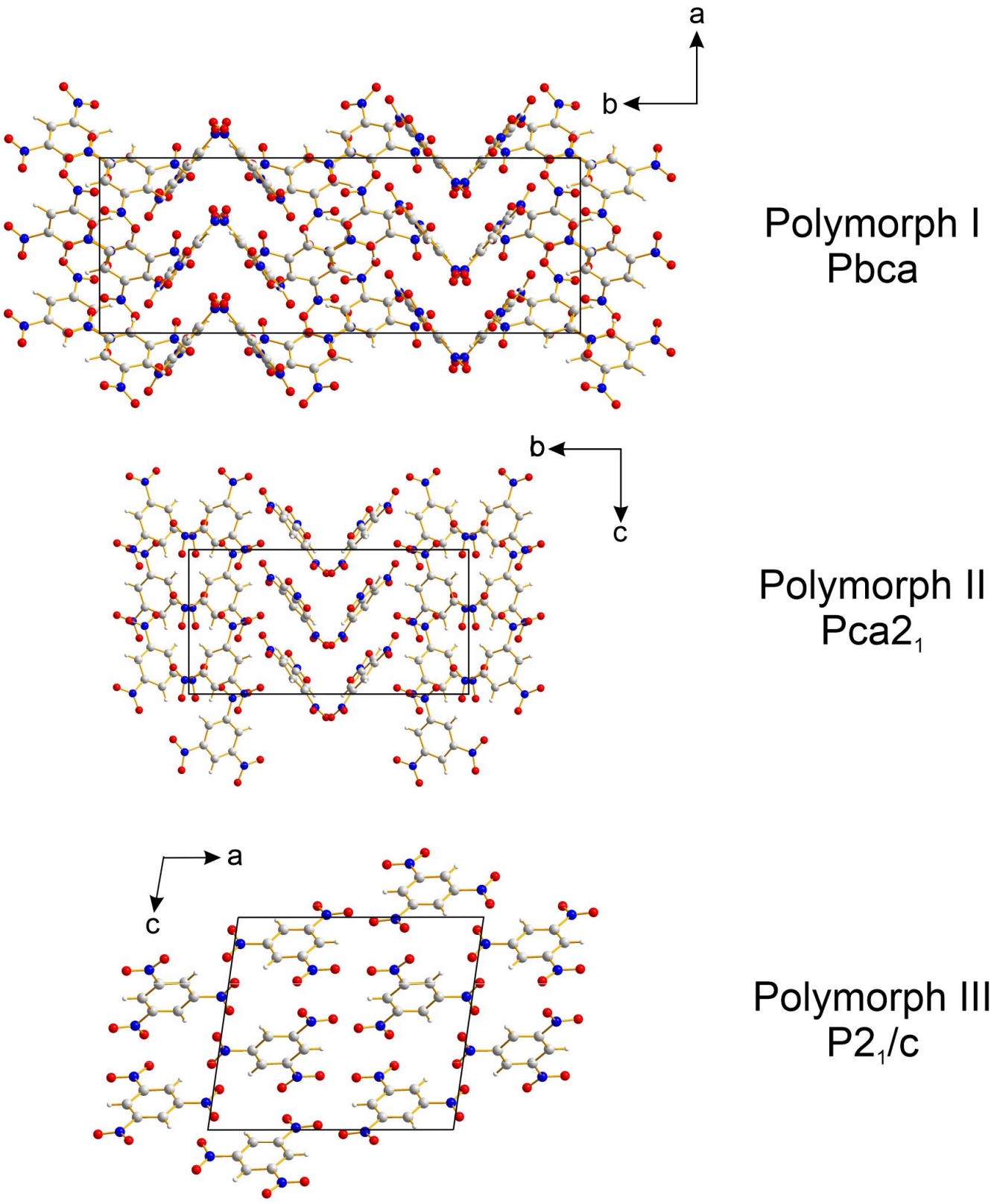

Figure 1.3: The crystal structures of the three polymorphs of 1,3,5-trinitrobenzene.

required to prepare more elusive polymorphs. These alternative methods include, but are not limited to, crystallization with the use of an additive (as with TNB discussed previously) [11], mechanochemical grinding (ball-milling) [15], and crystallization from the gas phase (sublimation)

[16]. Mechanochemical grinding and crystallization from the gas phase have proven to be useful in the preparation of new polymorphs of molecular organic materials, using the methodology shown in Figure 1.4. In the gas phase (Figure 1.4(a)), the molecules occupy sufficient space to be able to rearrange into a range of conformations (which may have low population in solution) and have an increased capability to crystallize as a new polymorph. In mechanochemical grinding by ball-milling (Figure 1.4(b)), the powder material is placed inside a sealed vessel with multiple solid balls, which is then shaken vigorously for a period of time. This is usually performed for *ca*. 5 minutes, however this can be substantially longer or shorter. This process inputs a huge degree of mechanical energy into the system which allows solid-state energetic barriers to be overcome and gives the materials flexibility for crystallizing in a unique form e.g. as a new polymorph. The process of ball-milling can be used in the preparation of co-crystals, for example in the preparation of ibuprofen-proline co-crystals [17]. The main factors that can be varied in a ball-milling experiment include: the amount of material placed inside the vessel, the vessel size, the size of the balls used in milling, the frequency of the milling, and also the time milling is allowed to occur for.

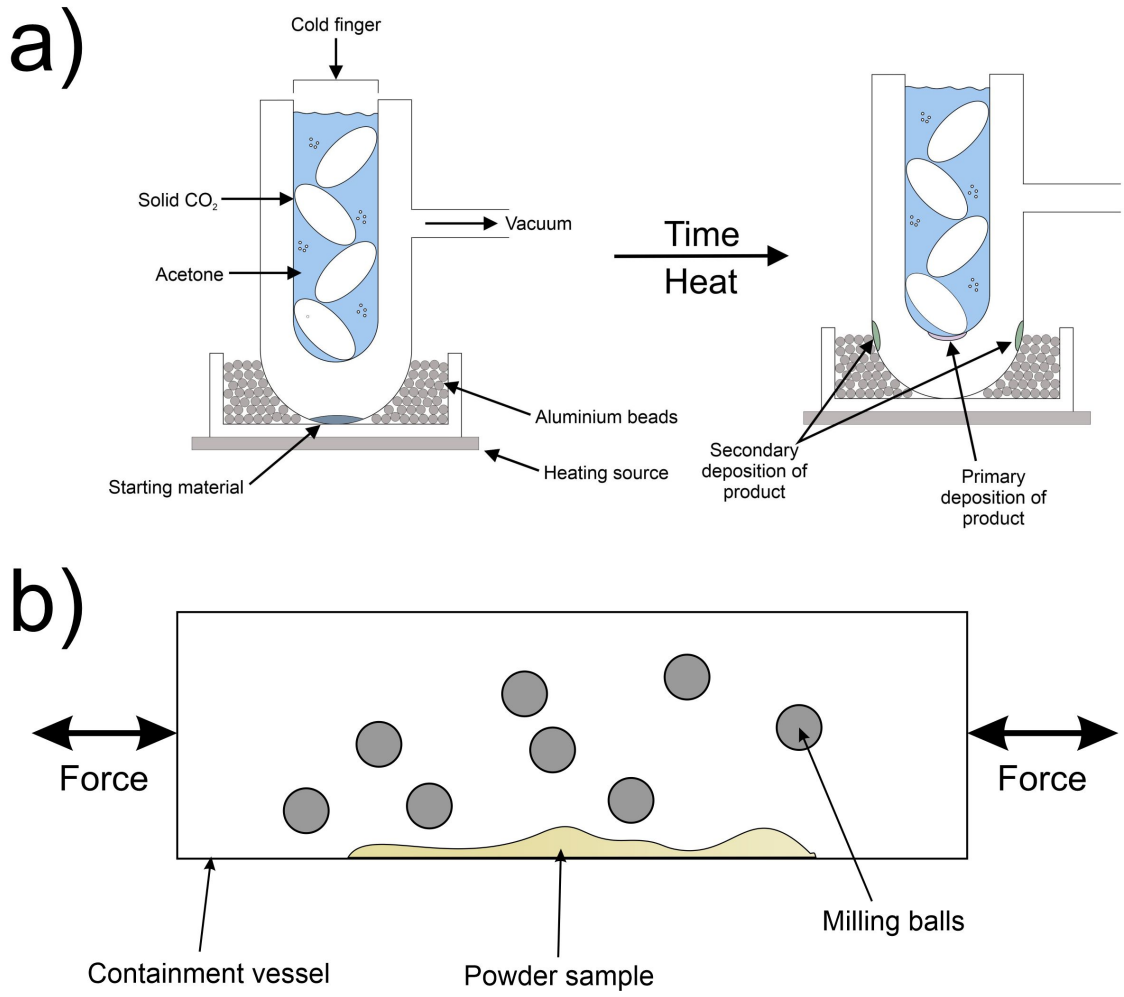

Figure 1.4: A diagram showing two innovative methods to prepare new polymorphs of materials: (a) crystallization from the gas phase following sublimation, and (b) ball-milling (mechanochemical grinding).

1.2 Biominerals

Biominerals are materials produced by living organisms for the purpose of improved strength (or other physical properties) of existing tissues or structures [18]. Many organisms are able to manipulate the structures of biominerals in their biological setting to bring about improved physical properties of the tissues or structures, such as increased tensile strength [19]. The composition of a biomineral is often based on a carbonate, phosphate or silica structure. Other inorganic components can exist in biominerals, as well as organic components (e.g. the presence of guanine crystals within scallop eyes), although these are less common [20]. Biominerals are found in a wide variety of different organisms. For example, carbonates can be found in bird eggshells, phosphates are present in vertebrate bones, and silica is present in marsh horsetail [21], which is a type of plant found in North America and Eurasia.

One of the more common groups of inorganic biominerals are the carbonates. Calcium carbonate is the most abundant type of biomineral found in biological organisms and exists as three crystalline polymorphs: calcite (the most stable), aragonite (metastable) and vaterite (metastable) [22].

The crystal structures of calcite and aragonite are seen in Figure 1.5. Vaterite is not shown as there is not a consensus among crystallographers as to the true structure, but most agree that the carbonate group is disordered. The hypothetical structures point towards a $\sqrt{3}$ x $\sqrt{3}$ superstructure with relative consensus, there are also several polytypes that have been theorized, and another suggestion is a structure with stacking faults in the *xy* plane [23]. Calcium carbonate is an important component to the shells of molluscs such as snails, clams and oysters by providing enhanced protection of the shell due to the greater strength a higher concentration of calcium carbonate provides. These shells are more than 95% (by weight) calcium carbonate, which is usually in the crystallographic form of calcite or aragonite, with the remaining component being organic material.

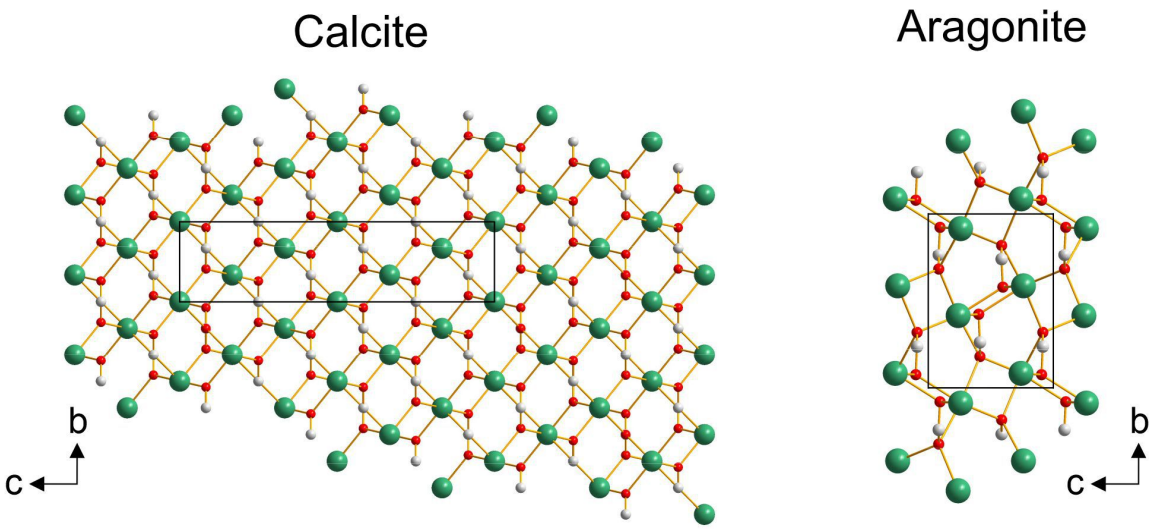

Figure 1.5: The crystal structures of calcite and aragonite viewed along the *a*-axis in both cases. Calcium atoms are shown in green, carbon atoms in grey, oxygen atoms in red.

The type of composite material found in the shells of molluscs is a combination of a crystalline calcium carbonate polymorph and an amorphous organic component, which can serve to promote a specific polymorph upon crystallization [24]. These composite materials exhibit exceptional resistance to fracturing in comparison to the pure crystalline polymorphs. This is because, although cracks can travel easily within the material, the 'crossed lamellar' structure (Figure 1.6) prevents them from travelling a significant distance [25]. In this type of 'crossed lamellar' structure, every so often the structure is rotated and thus a propagating crack meets a crystal orientation in which the crack is unable to travel. As shown in Figure 1.6, this type of structure is made up of individual lamellar components that form conglomerate sheets in a specific direction. A series of sheets are then stacked upon one another in different directions, preventing cracks from travelling easily throughout the structure.

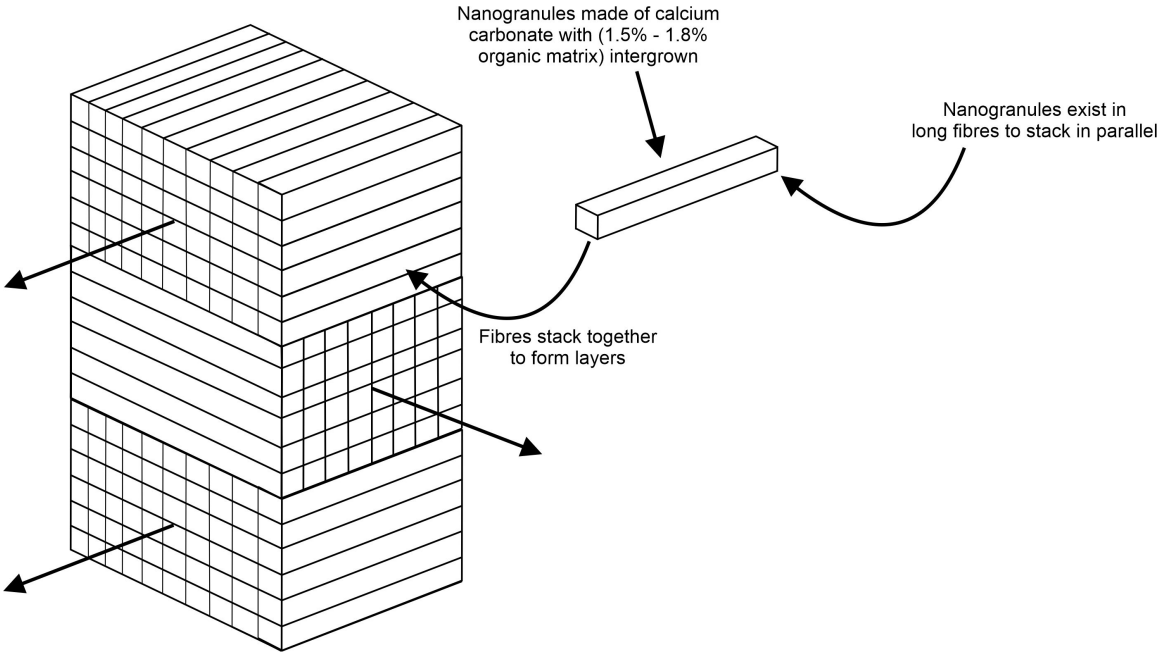

Figure 1.6: The crossed lamellar structure as observed in mollusc shells, showing both the individual lamellae as well as arrows to indicate the overall direction of the high order lamellar structure.

Although biominerals composed solely of organic material (organic biominerals) are less common than inorganic biominerals, they nevertheless represent an important class of biomineral. Organic biominerals are found prominently in the ocular tissues of organisms [26]. Their ability to form specific and highly regular crystal morphologies, and to form specific polymorphs, means they can provide an organism with enhanced physical properties. As a result they can give rise to physical phenomena, such as guanine crystals in silver spiders which are assembled to provide a 'silver mirror' effect [27].

Often in nature, materials exist as polycrystalline samples rather than single crystals, and thus powder XRD (and other powder methods) are important for fully understanding the structural

properties of these materials. Previous study has reported the structural properties of light-reflecting polymorphs of isoxanthopterin, using powder XRD to determine the crystal structure of the synthetic polymorph. This work was then supplemented by periodic DFT-D calculations to rationalize the high performance of the biogenic polymorph in its biological function [28]. Calculation of refractive indices from the known crystal structure showed that the biogenic polymorph has an exceptionally high refractive index (1.96), greater than that of the synthetic polymorph (1.83). This clearly rationalizes why the biogenic polymorph of isoxanthopterin found within decapod crustaceans operates so efficiently as a reflector in optical tissues [29]. Due to the poor crystallinity of the biogenic materials extracted from decapod crustaceans, powder XRD was imperative in this case, and is important with many other cases in the future.

To summarize, biominerals and other biogenic materials are some of the most interesting natural occurrences of crystalline materials, whether they are inorganic materials like calcium carbonate or organic materials such as isoxanthopterin. The crystal morphologies and crystal structures of these materials are inherently important for their biological roles and can inspire materials scientists in the manipulation of solid-state materials to influence their macroscale properties.

1.3 Amino Acids

Amino acids are a class of small organic molecules that contain a carboxylic acid group and an amine group, as well as a side chain (-R) group, as shown in Figure 1.7. These molecules are chiral, meaning they can exist as right-handed or left-handed isomers in a 3-dimensional arrangement [30]. Therefore they can exist in two different enantiomeric forms that are non-superimposable mirror images of one another, except in the case of glycine where the side chain is a hydrogen atom. D-amino acids are rare in biological systems, although they remain important in some roles [31]. Most proteins are composed of twenty standard amino acids that are directly encoded by DNA (known as directly encoded proteinogenic amino acids) and are the basic components of polypeptide chains in both oligopeptides and proteins. Amino acids are abbreviated either by a three letter code, or a one letter code, shown in Table 1.1, along with their side-chain characteristics.

Amino acids are not only essential building blocks in structural biology for the architecture of proteins, but they also serve a useful purpose within the pharmaceutical industry. For example, when an active pharmaceutical ingredient (API) shows efficacy in a biological application, its physicochemical properties (e.g. solubility) are rarely optimal for a specific application. Finding a similarly effective API with suitable physicochemical properties would be a long, costly and unlikely venture. However, a cheap and effective alternative is to prepare a co-crystal of the biologically active API with a biologically safe molecule, such as an amino acid [32]. The resultant co-crystal would still exhibit efficacy in its targeted role as well as showing superior physicochemical properties (such as improved solubility) because of the presence of the amino acid, recalling that

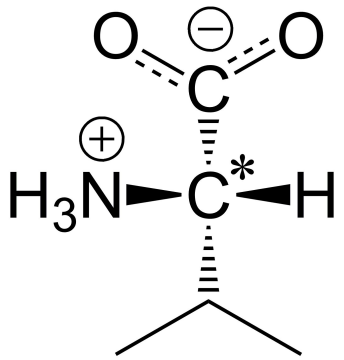

Figure 1.7: The structure of the amino acid valine in its zwitterionic form, with the α-carbon denoted with an *.

improved solubility promotes the bioavailability of the drug, thus improving its performance and overall marketability.

The first crystal structure of an amino acid (glycine) was determined in 1939 [33] using a combination of X-ray diffraction rotation and oscillation photographs. However, it was only in 2015 that finally all 20 directly encoded proteinogenic amino acids had at least one crystal structure determined; the final crystal structure to complete this set was the crystal structure of L-lysine, determined by powder XRD methods [34]. At present, L-phenylalanine has the highest number of known polymorphs among the 20 directly encoded proteinogenic amino acids with six reported polymorphs, four of which have known crystal structures, while the other two crystal structures are still unknown [34–36].

Many amino acids are hygroscopic and readily form hydrates, which clearly hinders the structural investigation of the anhydrous (pure) solid phases. To circumvent this problem, alternative processes to form the anhydrous phases are required, such as dehydration of the hydrate phase(s) or crystallization from the gas phase under vacuum. Although these methods are useful for preparing new crystalline phases, they often lead to microcrystalline samples which require the use of powder XRD for structural characterization. Powder XRD methods have been essential in the structure determination of many amino acids, their polymorphs and hydrates, such as L-arginine [35], L-lysine [34, 37] and L-tryptophan [36, 38].

It is likely due to the aforementioned methods for preparing pure phases of materials such as amino acids, the powder XRD samples will be poorly crystalline and therefore will likely lead to non-ideal XRD data. This can be in the case of broad peaks, which can hinder the indexing stage, or preferred orientation which can affect structure solution and/or Rietveld refinement. Clearly by employing other structural methods, these bottlenecks can be overcome and crystal structures of materials which are more challenging to prepare can be determined.

Amino Acid	3 Letter Code	1 Letter Code	Side-Chain Charge	Side-Chain Polarity
Alanine	Ala	A	Neutral	Non-polar
Arginine	Arg	R	Positive	Polar
Asparagine	Asn	N	Neutral	Polar
Aspartic acid	Asp	D	Negative	Polar
Cysteine	Cys	C	Neutral	Polar
Glutamic acid	Glu	E	Negative	Polar
Glutamine	Gln	Q	Neutral	Polar
Glycine*	Gly	G	Neutral	Non-polar
Histidine	His	H	Positive	Polar
Isoleucine	Ile	I	Neutral	Non-polar
Leucine	Leu	L	Neutral	Non-polar
Lysine	Lys	K	Positive	Polar
Methionine	Met	M	Neutral	Non-polar
Phenylalanine	Phe	F	Neutral	Non-polar
Proline*	Pro	P	Neutral	Non-polar
Serine	Ser	S	Neutral	Polar
Threonine	Thr	T	Neutral	Polar
Tryptophan	Trp	W	Neutral	Non-polar
Tyrosine	Tyr	Y	Neutral	Non-polar
Valine	Val	V	Neutral	Non-polar

Table 1.1: All the proteinogenic amino acids, along with their identification codes and side-chain properties. *These amino acids are unique. Proline is unique due to the lack of a free α-amino and free α-carboxyl group due to its cyclic structure. Glycine is unique due to a hydrogen atom being in place of a side chain completely.

1.4 Aims of the Project

Recent developments in methodology for structure determination from powder XRD data include the use of solid-state NMR data and periodic DFT-D calculations to augment specific aspects of the structure determination process. These also prove useful for final structure validation. The main aim of this project is to develop and optimize a multi-technique approach for structure determination from powder XRD data whereby supplementation of the approach by other structural techniques is more consolidated. The range of techniques used in this thesis are described in more detail in Chapter 2. A more synergistic approach will ensure that more complex crystal structures become easier to determine, as well as representing a more rigorous and robust process for the determination of other molecular organic crystal structures.

The materials chosen for this project mostly have biological relevance, in particular the structures

of riboflavin, L-tyrosine and ibuprofen. Riboflavin exists as a polycrystalline material in its biological application, yet there is no crystal structure attributed to it. L-tyrosine has been shown to exhibit polymorphism, without a crystal structure being known for the discovered polymorph, as for ibuprofen. Researchers do often not even consider the use of powder XRD data, even though it has the advantage of not requiring the preparation of single crystals of suitable size and quality for single-crystal XRD. The aim of the project from this aspect is to provide inspiration on how powder XRD data can be applied within the domain of structural biology, specifically to address structural problems involving small molecules. Although other materials studied in this project (alloxazine, PDICH and DBTMA) are not necessarily biologically important, their crystal structure determination has helped to push the boundaries of structure determination from powder XRD data.

This thesis focuses primarily on the optimization of a procedure for determining crystal structures from powder XRD data using a multi-technique approach in which the structure determination process is augmented by information from other structural techniques. The main challenges that this thesis aims to overcome are the typical bottlenecks in the crystal structure determination process by utilizing these other structural techniques. In particular, the inability to index a powder XRD pattern due to impurity from multiple phases, as well as difficulties in structure solution from issues such as preferred orientation. The powder XRD methodology is applied to various materials, which have been chosen based on their applications and importance and/or their structural complexity. The applications of the selected materials range from their biological role in the ocular tissues of mammals (riboflavin) [39] to their use as an over-the-counter analgesic (ibuprofen) [40]. The hypothesis of this thesis is that the usage of complementary techniques in the crystal structure determination from powder XRD procedure to be improved by allowing additional information to bypass some common bottlenecks in the process, as well as further scrutinising crystal structures following Rietveld refinement.

Chapter 2

Experimental Methods

2.1 X-ray Diffraction

X-rays are electromagnetic radiation, with a short wavelength compared to visible light, in the range of 1 Å (= 0.1 x 10^{-10} m) to 100 Å (= 1 x 10^{-8} m). They were discovered by Wilhelm Röntgen in 1895 [41], and less than two decades later it was discovered that crystals could be used as a diffraction grating for X-rays by Max Von Laue [41, 42]. Following this development by Von Laue, the field of X-ray diffraction took off rapidly with the publication of Bragg's law by the father and son team of William Henry Bragg and William Lawrence Bragg [43]. Bragg's Law relates the diffraction order (n), radiation wavelength (λ), d-spacing (d) and the diffraction angle (θ), the derivation of Bragg's Law is shown in Figure 2.1.

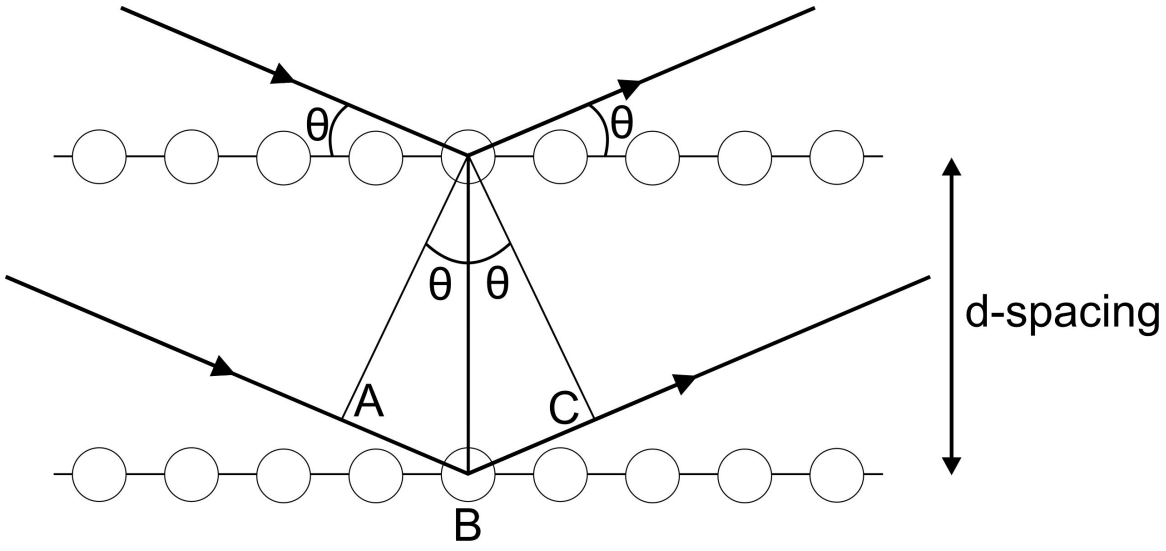

Figure 2.1: Derivation of Bragg's Law from the scattering of X-rays by parallel lattice planes making up a crystal.

As seen in Figure 2.1:

$$AB = BC = d \sin\theta \tag{2.1}$$

which, leads to the following representation of Bragg's Law:

$$n\lambda = 2d \sin\theta \tag{2.2}$$

X-ray radiation is scattered by the electron clouds within atoms. The electron cloud will scatter the X-ray in all directions when only an isolated atom, however when atoms are arranged in a periodic array as part of a crystal the X-rays are scattered in specific directions. Most of the scattered X-rays interfere destructively with one another. However, when scattered in certain specific directions they interfere constructively which results in diffraction maxima. For constructive interference to happen the X-rays must be 'in-phase' with each other, such that the difference in path between the waves must equal $n\lambda$ (where n = integer, corresponding to an integral number of wavelengths).

The intensity of a Bragg reflection, I_{hkl}, with given Miller indices *hkl*, is proportional to the square of the magnitude of the structure factor.

$$I_{\mathrm{hkl}} \propto |F_{\mathrm{hkl}}|^2 \tag{2.3}$$

Miller indices are a set of three integers which represent the orientation of a set of lattice planes within a crystal, explained in more detail in Section 3.1. The structure factor depends on the arrangement of atoms within the unit cell, and is given by the following equation:

$$F_{\mathrm{hkl}} = \sum_{j=1}^{n} f_{\mathrm{j}} N_{\mathrm{j}} e^{2\pi(hx_{\mathrm{j}}+ky_{\mathrm{j}}+lz_{\mathrm{j}})} exp^{(\frac{-B_{\mathrm{j}}\sin^2\theta}{\lambda^2})} \tag{2.4}$$

where the summation is over all atoms within the unit cell (j = *l,...,n*), f_j is the atomic scattering factor for atom j (which depends on the type of atom), N_j is the site occupancy for atom j. The atomic coordinates for atom j are (x_j, y_j, z_j) and B_j is the isotropic displacement factor (B_{iso}) for atom j. The value of B_{iso} describes the displacement of atom j from its average position (x_j, y_j, z_j) due to thermal vibration and/or disorder. Another common term used to describe B_{iso} is U_{iso}. It is usually used simply as a different notation for thermal displacement rather than having a different physical meaning, and is often implemented in programs due to historic use by the designer of the computer program. It is related to B_{iso} by:

$$U_{\mathrm{iso}} = \frac{B_{\mathrm{iso}}}{8\pi^2} \tag{2.5}$$

The atomic scattering factor f_j describes the scattering of the incident radiation by an individual atom (shown in Figure 2.2). The scattering factor depends on the Bragg angle, the wavelength of the radiation and the type of atom. Electrons that are closely held to the nucleus scatter coherently, otherwise known as Rayleigh scattering. The equation describing the atomic scattering (form) factor is given by:

$$f = \sum_{n=1}^{4} a_{\mathrm{n}} exp(\frac{-b_{\mathrm{n}}\sin^2\theta}{\lambda^2}) + c \tag{2.6}$$

where θ is the Bragg angle, λ is the radiation wavelength, and a_n, b_n, and c are coefficients determined from fitting to a simulated atomic scattering curve. These coefficients may be calculated theoretically using a variety of techniques, including Dirac-Slater wavefunctions and relativistic Hartree-Fock [44].

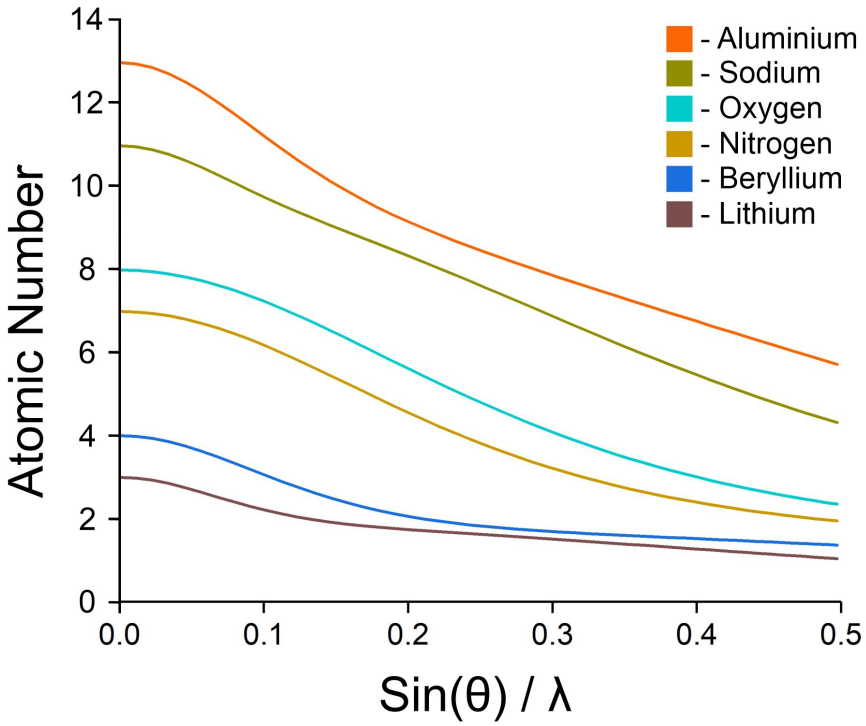

Figure 2.2: The atomic form factors for different atoms as a function of the scattering angle θ for a specific wavelength λ.

2.1.1 Powder X-ray Diffraction

A powder sample comprises many crystallites of small size (approximately one μm in diameter) and is referred to as a polycrystalline material. In an 'ideal' powder the crystallites have completely random orientations, in contrast to a single crystal which has a single orientation. When a powder sample is analysed using XRD, the diffraction data from three dimensions is compressed into a single dimension (expressed in 2θ). A disadvantage of powder XRD data compared to single crystal XRD data is the issue of overlapping peaks. A significant degree of peak overlap is intrinsic for materials with low-symmetry and/or large unit cells, even when individual peaks are narrow. For this reason severe overlap is often exhibited in the powder XRD patterns of molecular organic solids. Peak overlap is also problematic in cases for which the individual peaks are broad, for example due to small crystallite size.

Sometimes a sample can exhibit 'preferred orientation', which usually arises when the crystallites have a non-uniform shape e.g. flat-plates or needles, and they pack together in a way that causes

partial alignment crystallites within the sample as opposed to a complete random distribution of orientations, as displayed in Figure 2.3. Preferred orientation leads to the relative intensities of peaks in the powder XRD pattern differing from the diffraction intensities characteristic of the crystal structure, which can cause failure in the structure solution and refinement stages (discussed in Section 3.3.3). Examples of powder XRD and single-crystal XRD data collection in which the intensity distribution for the powder sample is idealized (unaffected by phenomena such as preferred orientation) are shown in Figure 2.5.

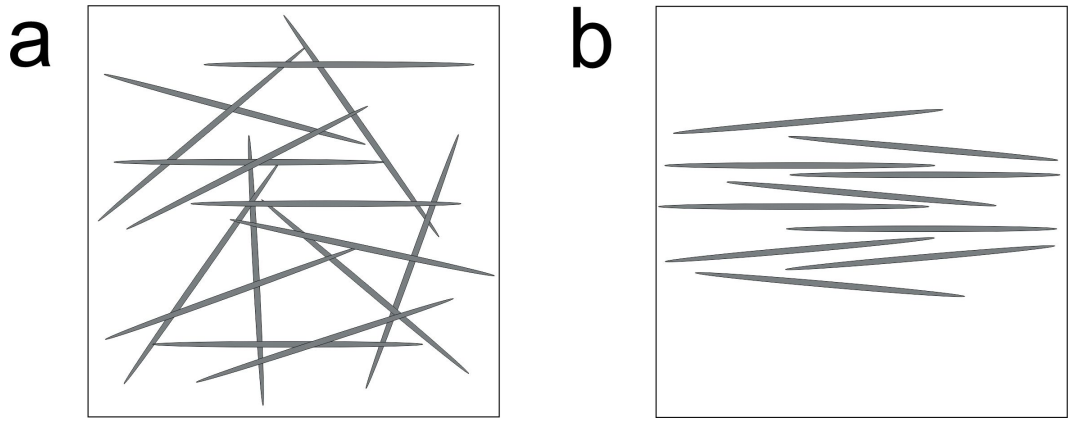

Figure 2.3: Illustration of a) random orientations of needle-like crystals and b) preferential orientation of needle-like crystals with all the long axes of the needles parallel. Note that in both these cases, the crystals are preferentially oriented in the plane of the paper.

There are two types of experimental set-up for a powder XRD experiment, specifically transmission and reflection. Reflection geometry (also referred to as Bragg-Brentano) is the more versatile set-up. In this design, the incoming divergent beam is reflected off the surface of the powder sample, and it converges on the detector. The advantage of reflection geometry is that there is a significantly smaller degree of X-ray absorption, which in turn leads to a much more intense powder XRD pattern, facilitating a shorter collection time to obtain data with a good signal-to-noise ratio. However, an important disadvantage is the packing of sample within the sample holder, which can induce preferential alignment of the crystallites and thus the intensities of the powder pattern peaks may be inaccurate. Both the basic reflection and transmission geometry set-ups are depicted in Figure 2.4.

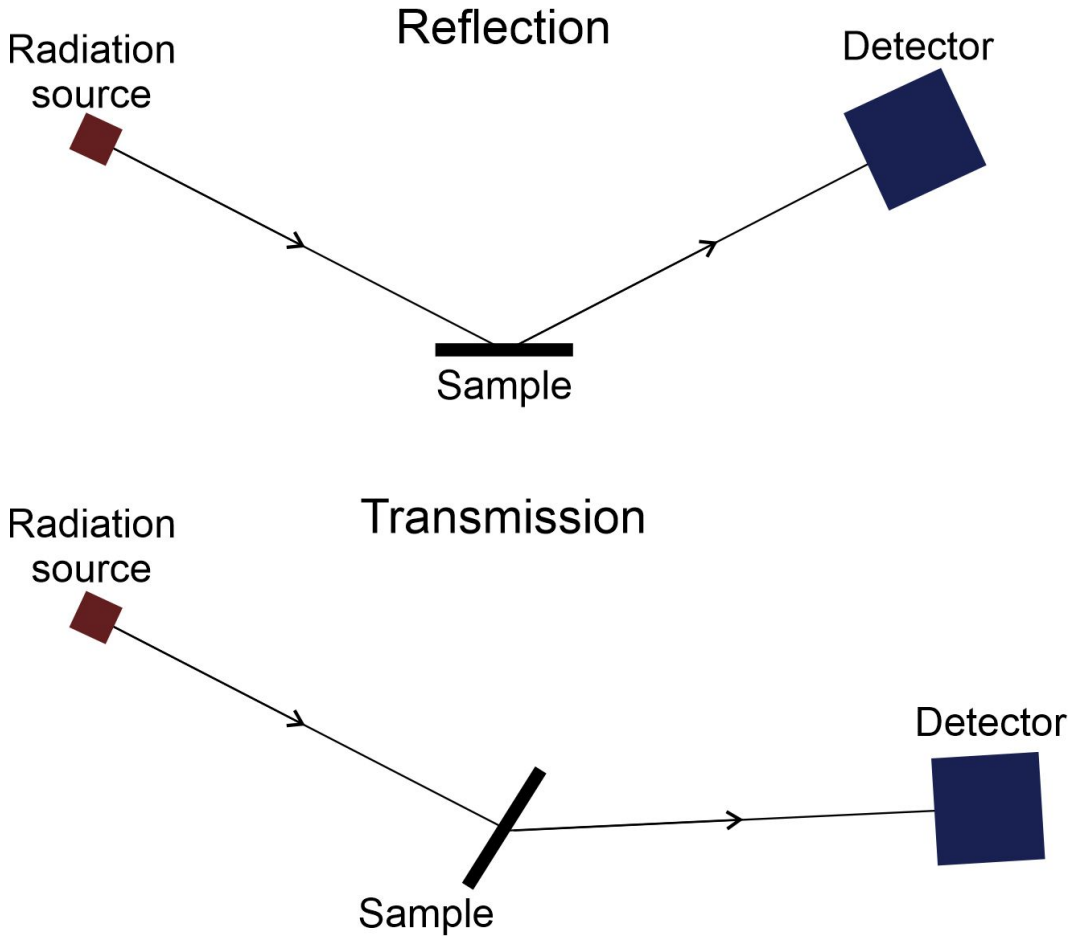

Figure 2.4: Experimental set-ups for powder XRD measurements in reflection mode (top) and transmission mode (bottom).

Transmission geometry (also referred to as Debye-Scherrer) is used to obtain data where high resolution is required, particularly when the aim is crystal structure determination. In this set-up, the X-ray beam passes through a monochromator (often Ge-110 for Cu-Kα_1 radiation), and then passes through the sample where it is reflected and then converges on the detector. The sample is usually spun in this geometry, allowing some averaging of crystallite orientations and partially reducing the problem of preferred orientation. The main benefit of transmission geometry is the significantly improved peak resolution, where diffraction peak widths can be 0.1° or better, and due to the sample spinning, the effects of preferred orientation are usually less severe. The sample can also be spun in reflection geometry.

2.1.2 Two-Dimensional Powder X-ray Diffraction

Instrumentally, the measurement of two-dimensional powder XRD (2D-powder XRD) data is simple. A standard powder diffraction experiment (in reflection or transmission geometry) is conducted, but instead of using a one-dimensional position-sensitive-detector (PSD) that scans as a function of 2θ, a two-dimensional area detector (perpendicular to the incident beam direction) is used, which

allows the whole Debye-Scherrer ring to be measured. As described in Section 2.1.1, preferred orientation in a powder sample distorts the measured intensities in the powder XRD pattern, which is qualitatively observed in 2D-powder XRD data as a non-uniform distribution of intensities around each Debye-Scherrer ring.

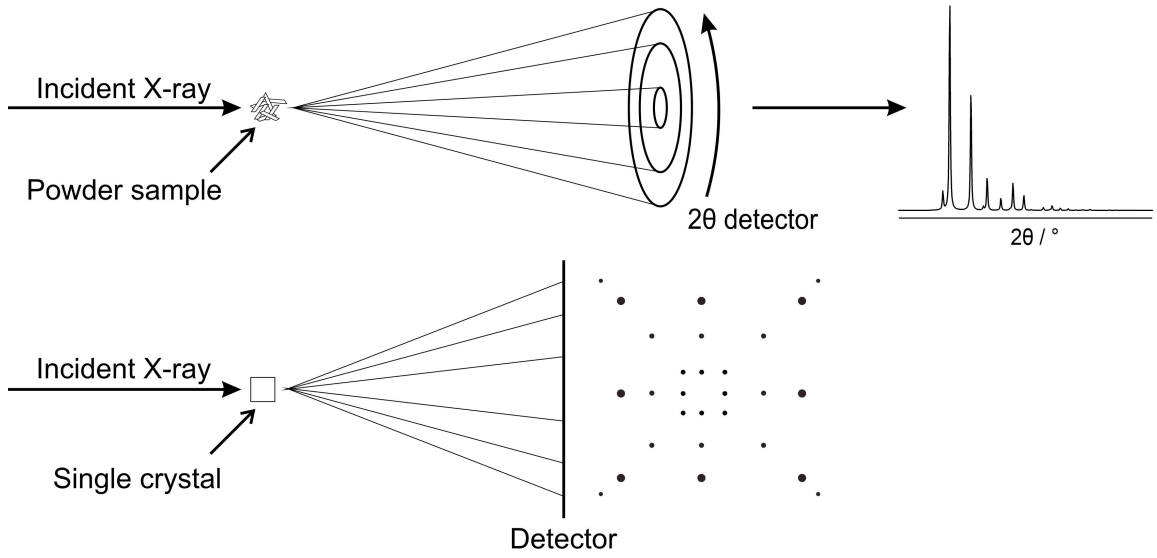

Figure 2.5: Intensity distributions for a powder that does not exhibit preferred orientation (top) and a single-crystal sample (bottom).

In this thesis, 2D-powder XRD measurements are used to qualitatively assess the extent of preferred orientation in a powder sample, and to provide justification to the use of corrections for preferred orientation in Rietveld refinement. The degree of preferred orientation can be assessed directly from the 2D powder XRD data, by assessing the uniformity of the rings. It can also be assessed upon integration of the two-dimensional powder XRD data into one-dimensional powder XRD data, which may then be compared to the powder XRD data recorded for the same sample using a one-dimensional PSD detector.

2.1.3 X-ray Sources

The X-rays required for diffraction experiments can be generated either through a high-voltage cathode ray tube, as used in most laboratory diffractometers, or using a synchrotron radiation source.

Laboratory X-ray sources generate X-rays by heating up a metal filament to produce electrons by thermionic emission, which are then accelerated towards a metallic target anode. Interaction of the electrons with the metal target cause excitation or ejection of core electrons from the K shell of the metal atoms. Electrons from the L or M shell then descend in energy to fill the core hole in the K-shell, resulting in X-ray emission. This is denoted as Kα or Kβ radiation depending on whether the electron descends from the L shell or the M shell respectively (Figure 2.6). α radiation has two components, denoted Kα_1 and α_2, arising from the small difference in energy between

different spin states for the 2p orbital. Usually only $K\alpha$ is used for diffraction experiments in a lab X-ray diffractometer, which can be achieved using of a nickel filter to remove Kβ radiation. For experiments requiring monochromatic radiation, a monochromator is placed at a specific Bragg angle. For example, when using Cu radiation, a Ge(111) monochromator crystal at a Bragg angle of 27.6° can be used to remove any Kα_2 radiation. It is worth noting that whenever a monochromator crystal is used, the intensity loss is substantial, which usually leads to longer measurement durations.

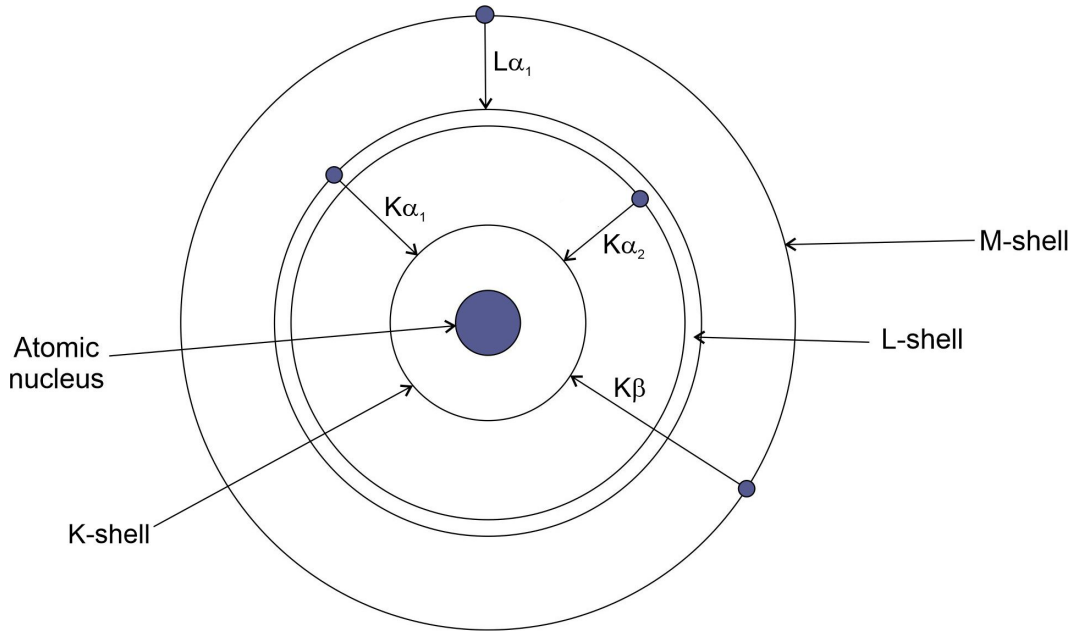

Figure 2.6: Schematic illustration of the emission of Kα_1, Kα_2 and Kβ radiation from an atom.

An alternative to laboratory X-ray radiation is the use of synchrotron radiation, for which the X-rays produced are highly polarized and extremely intense. Synchrotron radiation is also fully tunable *i.e.* all X-ray wavelengths are available, with the use of an appropriate monochromator. An electron beam is accelerated using a linear accelerator, followed by injection into a storage ring where electrons move in a circulator trajectory at constant speed, close to the speed of light. At several points of the storage ring, the electron beam is bent using bending magnets, which causes acceleration by changing direction; this is accompanied by emission of radiation across the whole electromagnetic spectrum. From this broad continuum, X-rays of a specific wavelength can be selected depending on the experiment of interest. Due to the high collimation of the radiation from the synchrotron, much narrower peaks are generally obtained in powder XRD data as a result of significantly better instrumental resolution, compared to a laboratory diffractometer. However, the fitting of line shapes from synchrotron radiation data can often prove difficult as sample dependent line-broadening effects often dominate.

2.2 Solid-State Nuclear Magnetic Resonance Spectroscopy

2.2.1 Introduction to NMR

Solid-state nuclear magnetic resonance (NMR) is a technique that is being used increasingly in conjunction with powder XRD [45], particularly as it can provide direct insights on specific aspects of the local structural properties of solids. It can also provide information on dynamic properties of solids. NMR exploits the fact that an atomic nucleus with a non-zero value of spin quantum number *I* has spin angular momentum of magnitude $\sqrt{I(I+1)}\hbar$ and correspondingly has a nuclear magnetic moment, μ. The constant of proportionality is the gyromagnetic ratio, γ, which is characteristic of a given type of nucleus. Thus, as each NMR active isotope has a unique gyromagnetic ratio, it has a unique NMR resonance frequency.

Nuclei with $I \geq 1$ are known as quadrupolar nuclei. When a nucleus with I > 0 is placed in an applied magnetic field, the component of spin angular momentum along the magnetic field axis is quantized according to the quantum number m_I ($m_I = I, I-1, ..., -I$). Each value of m_I corresponds to a different energy of interaction (Zeeman interaction) between the nuclear magnetic moment and the applied magnetic field. For a nucleus with spin quantum number I, there are $2I+1$ non-degenerate energy levels. When the nuclei are irradiated at the appropriate Larmor frequency, transitions between the nuclear spin energy levels will occur, these transitions give rise to the NMR spectrum [46].

There is a significant difference between how molecules in the liquid state and those in the solid state behave in regard to the NMR phenomenon. In the liquid state, the molecules undergo rapid isotropic tumbling, which averages out all the anisotropic NMR interactions. In the solid state, on the other hand, there is often no significant molecular motion, and anisotropic NMR interactions strongly influence the NMR spectrum, which leads to significant peak broadening [47]. The important anisotropic interactions that need to be considered in solid state NMR are shielding anisotropy, direct dipole-dipole interaction and the quadrupolar interaction. Each of the anisotropic NMR interactions depends on a term of the type $(3cos^2\theta - 1)$, where θ refers to the angle between the principal axis of the tensor that described the anisotropic NMR interaction and the axis of the applied magnetic field B_0. In the case of a polycrystalline powder sample, the observed NMR spectrum is the sum of the individual spectra for all crystal orientations (i.e. all values of θ) represented in the powder (for a randomly oriented powder, all orientations in space are equally populated). Consequently, NMR spectra for polycrystalline powder samples are typically very broad compared to those for liquid samples or single crystal samples. The details of the shape of the NMR "powder pattern" for a polycrystalline sample depend on details of the crystal structure and the properties of the specific NMR-active nucleus under investigation.

There are two general strategies for recording solid-state NMR spectra: i) broad-line spectra [48], and ii) high-resolution spectra [49]. The aim of broad-line spectra is to study the details of the

broadening due to anisotropic interactions. Broad-line spectra give detailed information about the solid, such as insights into structure, dynamics and molecular geometry, by giving information on the tensors that describe the anisotropic NMR interactions. However, broad-line spectra are often complicated, and interpretation may be difficult unless one specific anisotropic NMR interaction is dominant. The aim of high-resolution spectra is to record the spectrum in such a way that the anisotropic NMR interactions are removed, giving spectra that contain sharp lines reminiscent of liquid-state NMR spectra. As a result, high-resolution spectra are usually straightforward to interpret. However, they are essentially "isotropic" spectra as all the information about the anisotropic NMR interactions is removed by the measurement technique. Therefore, high-resolution solid-state NMR spectra contain much less detailed information about the material than broad-line spectra.

The solid-state NMR spectra recorded in this research are high-resolution solid-state ^{13}C NMR spectra. For most organic materials at natural isotopic abundance, the two main anisotropic NMR interactions that need to be eliminated in recording high-resolution solid-state ^{13}C NMR spectra are: i) chemical shift (shielding) anisotropy, and ii) heteronuclear ^{13}C$\cdots^{1}$H direct dipole-dipole interaction. It is noted that $I = \frac{1}{2}$ for ^{13}C, and therefore no quadrupolar interaction is present. The experimental strategy to record high-resolution solid-state ^{13}C NMR spectra is to use the following techniques: i) magic-angle sample spinning to reduce the chemical shift anisotropy to the isotropic value (i.e. the isotropic chemical shift) [50], and ii) high-power ^{1}H decoupling to eliminate the heteronuclear ^{13}C$\cdots^{1}$H direct dipole-dipole interaction [51]. In the technique of magic-angle sample spinning (MAS), the powder sample is contained in an NMR rotor. During the measurement of the NMR spectrum, the rotor is spun rapidly (at frequency ν_r) about an axis making an angle β with the direction of the applied magnetic field B_0, as shown in Figure 2.7.

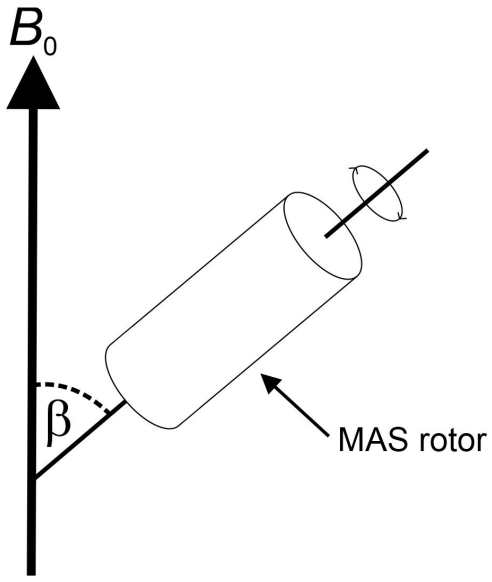

Figure 2.7: Schematic of the experimental set-up for recording solid-state NMR data using magic-angle spinning (MAS). The value of β is the magic-angle, β = 54.7°.

For one specific value of β, in particular $\beta = cos^{-1}(\frac{1}{\sqrt{3}})$ = 54.7° (the "magic-angle") it can be

shown that all the anisotropic NMR interactions are reduced to their isotropic value. This is only the case if the spinning frequency ν_r is sufficiently fast. Typically the maximum spinning frequency for most solid-state NMR instruments is about 20 kHz, although some modern instruments can achieve spinning frequencies up to about ν_r = 126 kHz [52].

When the MAS frequency ν_r is sufficiently fast, all the anisotropic NMR interactions are eliminated and the observed high-resolution solid-state NMR spectrum contains sharp lines at the isotropic chemical shifts, resembling a liquid-state NMR spectrum. However, if the frequency of magic angle spinning (denoted as ν_r) is not fast enough to eliminate a particular anisotropic interaction, the way in which MAS affects the solid-state NMR spectrum depends on the particular anisotropic NMR interaction. In the case of chemical shift anisotropy, the spectrum is split into a set of sharp isotropic peaks, corresponding to the isotropic chemical shift for each crystallographically distinct site in the structure. For each isotropic peak, there is a set of sharp "spinning sidebands" that are separated from the corresponding isotropic peak by multiples of the MAS frequency - i.e. separated from the isotropic peak by $\pm n\nu_r$, where n is a positive integer. For higher values of ν_r, the spinning sidebands are further apart from each other (and for sufficiently high ν_r, no spinning sidebands are observed). The positions of the isotropic peaks in the spectrum are independent of the MAS frequency ν_r used to record the spectrum. The effect of spinning speed on the position of spinning sidebands relative to the isotropic peak is shown in Figure 2.8.

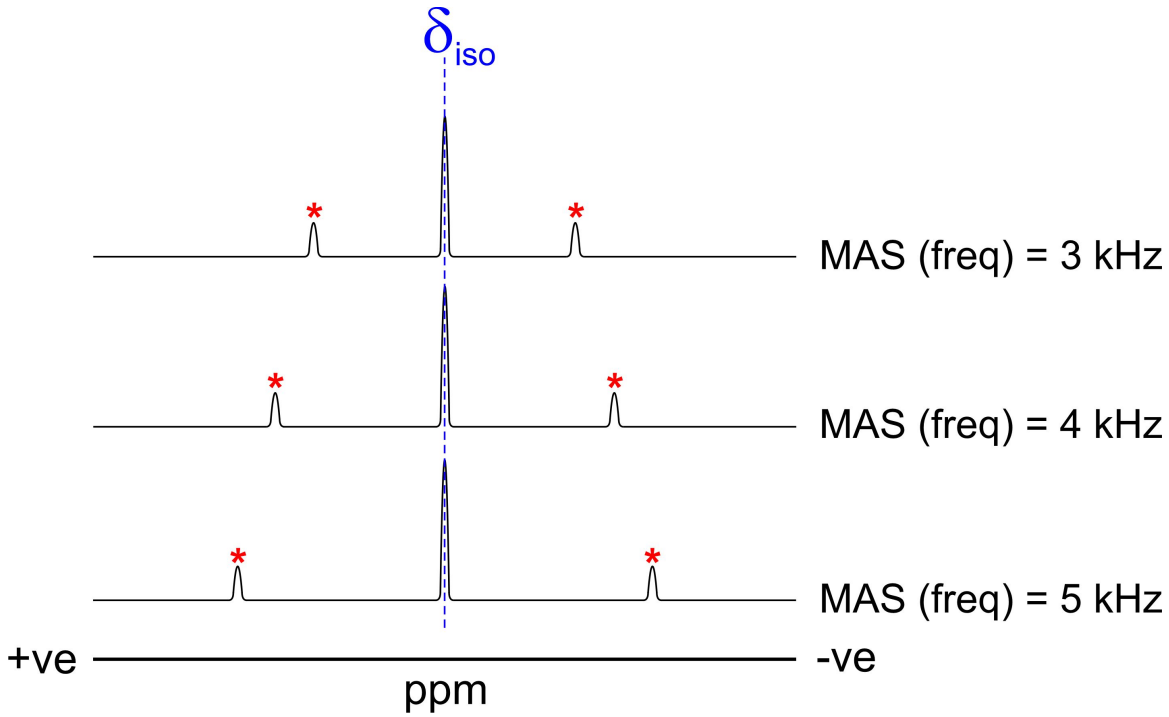

Figure 2.8: A schematic showing the relative distance between spinning sidebands (denoted with a red asterisk) and the isotropic peak (denoted as δ_{iso}) for a series of three different spinning speeds for an ideal, arbitrary NMR spectrum.

In general, the magnitude of the heteronuclear $^{13}C\cdots{}^{1}H$ direct dipole-dipole interaction in organic

solids is too large to be reduced by magic-angle sample spinning. Instead, the $^{13}C\cdots{}^{1}H$ direct dipole-dipole interaction can be eliminated by high-power ^{1}H decoupling. In this technique, the ^{13}C NMR spectrum is recorded while simultaneously irradiating the sample at the ^{1}H resonance frequency. As mentioned above, the combination of magic-angle sample spinning and high-power ^{1}H decoupling is a widely used strategy for recording high-resolution solid-state ^{13}C NMR spectra of organic solids. Another experimental method commonly used in recording high-resolution solid-state ^{13}C NMR spectra is $^{1}H\rightarrow{}^{13}C$ cross-polarization (CP), which is designed to enhance the signal intensity in recording solid-state NMR spectra for low-abundance nuclei, such as ^{13}C [53].

Cross-polarization involves the transfer of magnetization from high Larmor frequency nuclei (*e.g.* ^{1}H) to nuclei with a low Larmor frequency such as ^{13}C. This transfer of magnetization from ^{1}H to ^{13}C has two benefits in improving signal intensity compared to conventional measurements of ^{13}C NMR spectra by direct polarization. The first benefit arises from the polarization transfer process, which gives a maximum enhancement equal to the ratio of the Larmor frequencies of the ^{1}H and ^{13}C nuclei given by:

$$\frac{\omega_0(^{1}H)}{\omega_0(^{13}C)} = 3.98 \qquad (2.7)$$

This ratio is equal to the ratio of the gyromagnetic ratios γ^H/γ^C, where γ^H and γ^C are the gyromagnetic ratios of ^{1}H and ^{13}C respectively. Secondly, because the magnetization originates from the ^{1}H nuclei, the spin recovery before repetition of the pulse sequence depends on $T_1(^{1}H)$ rather than $T_1(^{13}C)$ (where T_1 denotes the spin-lattice relaxation time). As the T_1 values for ^{1}H nuclei are typically much shorter than for ^{13}C nuclei, the recycle delay is shorter in a $^{1}H\rightarrow{}^{13}C$ CP measurement than a direct excitation ^{13}C NMR measurement, and therefore the $^{1}H\rightarrow{}^{13}C$ CP pulse sequence needs to be repeated more often within a given total experiment time. A diagram of the pulse sequence for a CP-MAS NMR experiment is shown in Figure 2.9.

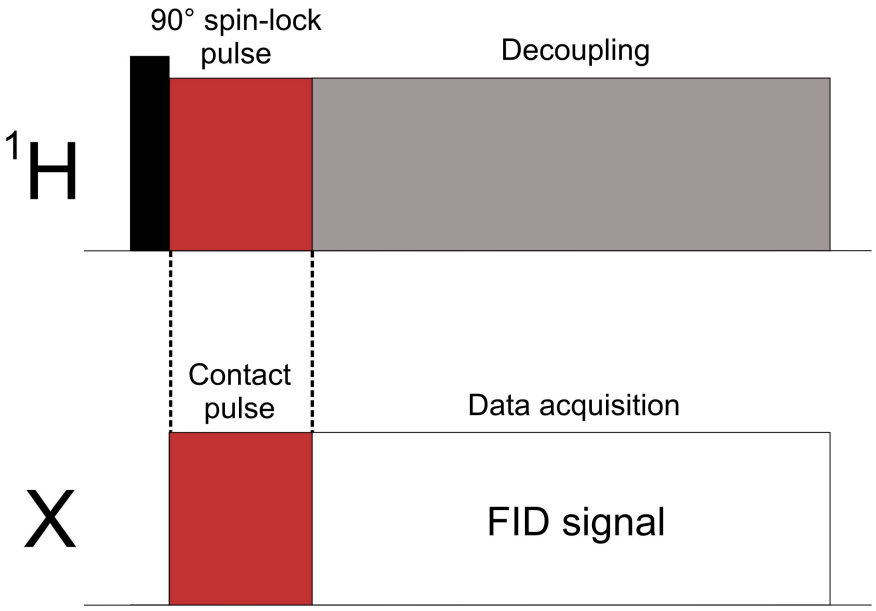

Figure 2.9: The pulse sequence for recording the solid-state NMR spectrum of nucleus X under conditions of cross-polarization from ^{1}H. Free induction decay (FID) is the observable NMR signal generated by non-equilibrium nuclear spin magnetization precessing about the magnetic field.

2.2.2 Applications of Solid-State NMR in Structure Determination

Solid-state NMR can be applied to enhance the process of structure determination from powder XRD data. This can be manifested as validation of the crystal structures determined following Rietveld refinement but can also provide specific structural insights that may assist various stages of the structure determination process. One such application is to provide information on the contents of the asymmetric unit of a crystal structure, and to establish the number of crystallographically independent molecules within the asymmetric unit (Z'), which can be crucial in space group assignment.

Another way that solid-state NMR can be utilized in relation to structure determination is to study specific internuclear couplings, which can yield information on specific internuclear distances or molecular conformation. This information can aid structure determination from powder XRD data by providing independent information to assess the correctness of trial structures obtained in structure solution, as well as facilitating additional restraints in Rietveld refinement. The approach has been used in the structure determination of a material with high degree of conformational flexibility, 3',5'-bis-O-decanoyl-2'-deoxyguanosine (dG(C_{10})$_2$), for which a variety of trial structures can give a reasonable fit to the powder XRD data. However, by considering ^{15}N$\cdots$N distances established from solid-state ^{15}N NMR studies, the number of potentially correct trial structures was reduced substantially [5]. A further approach that exploits the synergy between solid-state NMR and powder XRD involves the comparison of chemical shifts (such as ^{13}C and ^{15}N). Where the calculated chemical shifts were generated by periodic DFT-D calculations for a crystal structure

determined from powder XRD data and experimental chemical shifts measured from solid-state NMR data for the material. This approach works exceptionally well as a strategy for validation of the final structure obtained from Rietveld refinement. This approach was first applied in structure determination of the pharmaceutical co-crystal indomethacin nicotinamide [54], and also applied in the above work on $dG(C_{10})_2$ as well as in the structure determination of cimetidine hydrochloride [4].

2.3 Density Functional Theory Calculations

2.3.1 Introduction

Density functional theory (DFT) is a quantum mechanical modelling process utilized by physicists, chemists and materials scientists, which can provide reliable insights into chemical systems that are not accessible (for example, on account of their size) to traditional quantum mechanical calculation based on Hartree-Fock theory. The recent developments in density functional theory have moved the technique from relatively niche to a tool that is regularly used by many chemists from a range of disciplines. Hohenberg & Kohn showed, in the 1960s, that the ground state properties (for example, the energy E) of a molecule can be determined as a functional of the electron density, ρ [55]:

$$E = f[\rho] \tag{2.8}$$

In principle this approach can replace wave mechanics completely for molecules, although in practice the correct functional is unknown. Modern research is therefore focused on finding approximate functionals that allow the energy E to be estimated from the electron density ρ. DFT calculations are set up for a given molecule using a hybrid functional and a basis set. The hybrid functionals available include, but are not limited to, B3LYP [56], PBE [57] and PBE0 [58]. DFT calculations in the gas phase can be used for tasks, such as geometry optimization of an isolated molecule, calculating molecular energy, and predicting molecular spectra. In this thesis, the geometry optimization and molecular energy calculations in the gas phase were conducted using GAUSSIAN [59], accessed through the web interface WebMO [60] and processed using the HAWK supercomputer at Cardiff University.

2.3.2 Periodic DFT-D Calculations

Periodic DFT-D (D representing the use of dispersion corrections) calculations are the specific implementation of DFT calculations for periodic solids. Periodic DFT-D calculations involve molecules arranged within a unit cell, which is repeated in three-dimensions in a crystal structure. These calculations require additional physical information (such as space group symmetry and unit cell dimensions) compared to calculations on isolated molecules.

The application of periodic DFT-D calculations for optimizing the geometry of a crystal structure is now widely used in conjunction with structure determination from powder XRD data and serves

a variety of purposes within the structure determination process. For example, it can generate more reliable bond lengths and angles (including hydrogen bonding geometries) for the crystal structure than those generated in an initial Rietveld refinement. This information can then be used to define the restraints used in subsequent Rietveld refinements. Furthermore, as the geometry is optimized by minimizing the energy of a crystal structure, the periodic DFT-D calculations can determine whether a crystal structure lies within (or close to) an energy minimum. NMR parameters such as chemical shifts can also be calculated reliably for crystal structures using periodic DFT-D calculations. The calculated solid-state NMR data can be used for comparison against the experimental solid-state NMR data for the same material as an additional method for validating the final crystal structure obtained in Rietveld refinement.

Accurate calculations of the energies of crystal structures are extremely important in certain contexts, particularly in relation to the phenomenon of polymorphism, which requires accurate determination of the relative energies to determine an accurate energetic ranking. Periodic DFT-D methods have developed over recent years and a variety of different functionals, and different dispersion correction schemes are available. It has been shown [61] that the combination of the PBE0 functional and the many-body dispersion (MBD) correction scheme is the most reliable method in calculating the relative energies of polymorphs of organic molecular material.

2.4 Electron Diffraction

Electron diffraction arises from the scattering of an electron beam of appropriate energy by its interaction with a crystalline material, producing a characteristic diffraction pattern [62]. The concept was initially predicted by Louis de Broglie in 1924 before two groups demonstrated the phenomenon experimentally: i) Davisson and Germer in the range of low electron energies [63], and ii) Thomson and Reid in the range of high electron energies [64]. The wavelength of radiation used in an electron diffraction experiment is typically shorter than that used in X-ray diffraction experiments (*ca.* 0.5-2.5 Å for XRD and *ca.* 0.02-0.025 Å for electron diffraction). The wavelength of an electron beam depends on the accelerating voltage used to produce the electron beam. Unlike X-ray diffraction, a relativistic correction (shown below in Equation 2.9) must be made as the electron velocity approaches the speed of light. At low accelerating voltages (e.g. 10 kV), the value of the wavelength changes by around 0.5% due to relativistic effects, whereas at higher accelerating voltages (*e.g.* 200 kV), the value of wavelength changes by *ca.* 8.5%.

$$\lambda = \frac{h}{\sqrt{2m_{\mathrm{e}}E}} \frac{1}{\sqrt{1 + \frac{E}{2m_{\mathrm{e}}c^2}}} \tag{2.9}$$

where λ is the electron wavelength, h is Planck's constant, m_e is the mass of an electron, E is the energy of the electron beam, and c is the speed of light.

A common feature of electron diffraction and powder XRD diffraction is that they are both suitable techniques for determining the crystal structures of microcrystalline samples, unlike traditional single-crystal X-ray diffraction methods which require a larger crystallite size. Nevertheless, electron diffraction and powder XRD differ significantly in other aspects. Regarding data collection, electron diffraction requires a large, expensive electron microscope, which is not routinely available within most laboratories. Therefore the ability to source time for data collection may be limited. Conversely, collecting powder XRD data is straightforward and is cost-effective. Another aspect of data collection concerns radiation damage, which can be a significant problem for electron beams compared to X-rays [65]. For this reason electron diffraction measurements are often conducted at low temperature to mitigate the effects of radiation damage. For powder XRD data collection radiation damage this is only usually an issue at synchrotron radiation sources, and is usually mitigated using area detectors that allow for rapid data acquisition. Regarding the relative merits of powder XRD and electron diffraction in structure determination, powder XRD data can provide precise unit cell parameters, provided that the indexing calculations are successful. In contrast, unit cell determination from electron diffraction is usually more straightforward, but the unit cell parameters are much less accurate (although systematic absences are determined more easily). However, a major difference is that electron diffraction is significantly affected by multiple scattering, whereas X-ray diffraction is well described by the single scattering approximation (exhibited in Figure 2.10) [66]. Consequently, the kinematical theory of diffraction may be used in conducting structure solution and refinement from powder XRD data. However accurate structure solution and refinement from electron diffraction data require significantly more complicated calculations based on the dynamical theory of diffraction.

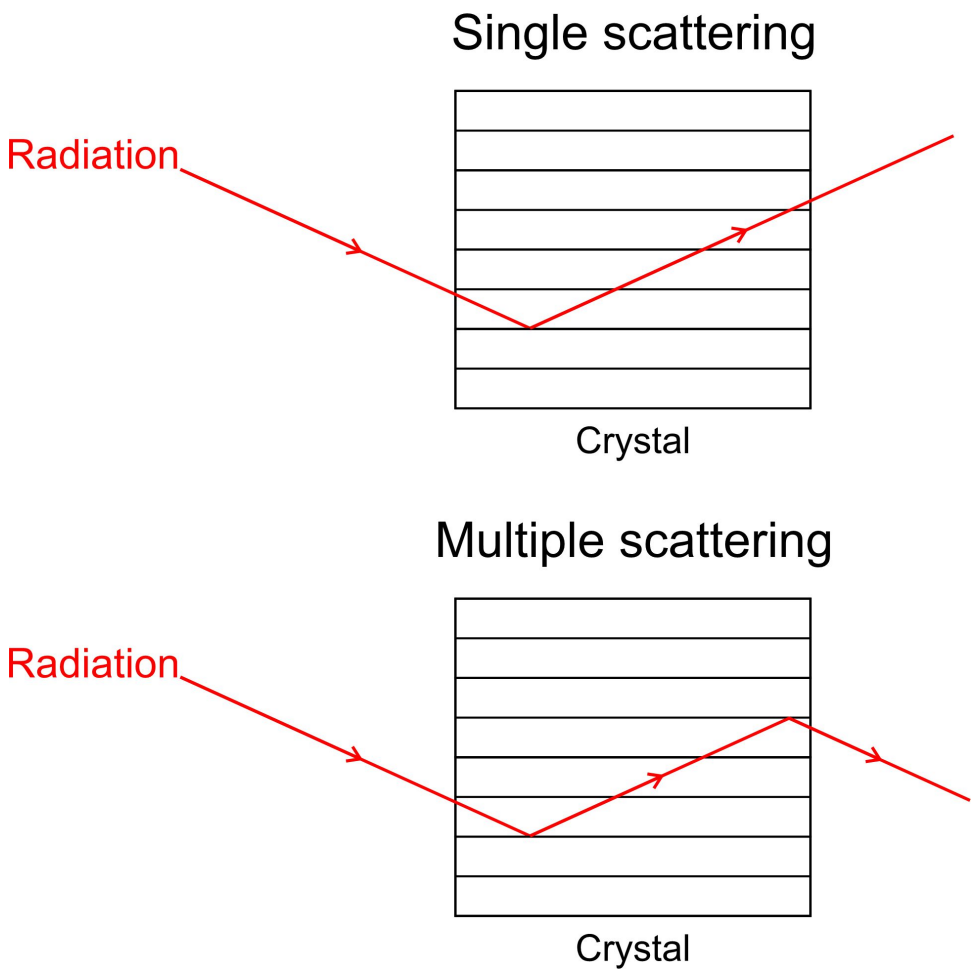

Figure 2.10: The difference in scattering kinematically (top) and dynamically (bottom).

Although in some respects electron diffraction data is inferior, there are some aspects that make it a powerful tool to be used in conjunction with powder XRD diffraction data. Unlike powder XRD data, electron diffraction data are unaffected by the effects of preferred orientation as the data are measured on a single micro-crystal, and we recall that preferred orientation can be one of the biggest limiting factors in the use of powder XRD data in structure determination. Another advantage of electron diffraction is that, because an individual crystallite is studied, the measurement is unaffected by the presence of an impurity phase in the bulk sample, whereas the presence of an impurity phase can significantly impede the process of structure determination from powder XRD methods (particularly regarding indexing the powder XRD data).

The main limitation when analysing electron diffraction data is the effect of dynamical scattering, which occurs when a Bragg reflection is scattered multiple times by the crystal before being recorded by the detector. This is a consequence of the high scattering power of electrons with solids. It is possible to obtain a valid structure solution using the kinematical theory of diffraction, even when the electron diffraction data are affected by dynamical scattering, although this may lead to high values of figure of merit and a low-quality structure. For structure refinement, programs utilizing dynamical theory of diffraction are available, but the calculations are complex and associated with

high computational cost [67]. Some of the main advances in developing electron crystallography instrumentation has been the development of techniques to reduce the effects of dynamical scattering, including known as 3D electron diffraction (3D-ED) technique [68]. One of the most widely used methods is precession electron diffraction (PED), which was developed by Vincent and Midgley in 1994 [69]. PED works by rotating the electron beam around the electron-optical axis, which allows only a few Bragg reflections to arise at any one time, which reduces the likelihood of dynamical scattering (shown in Figure 2.11). Another recently developed method is the automated diffraction tomography (ADT) method developed by Kolb in 2007 [70]. ADT utilizes off-zone electron diffraction patterns (non-oriented) which reduces the dynamical scattering effects due to the inclination of the electron beam from the zone axis. The electron diffraction data used as part of this work was obtained using the 3D-ED approach and was collected and processed by Dr Duncan Johnstone at the University of Cambridge Electron Microscopy group with Professor Paul Midgley.

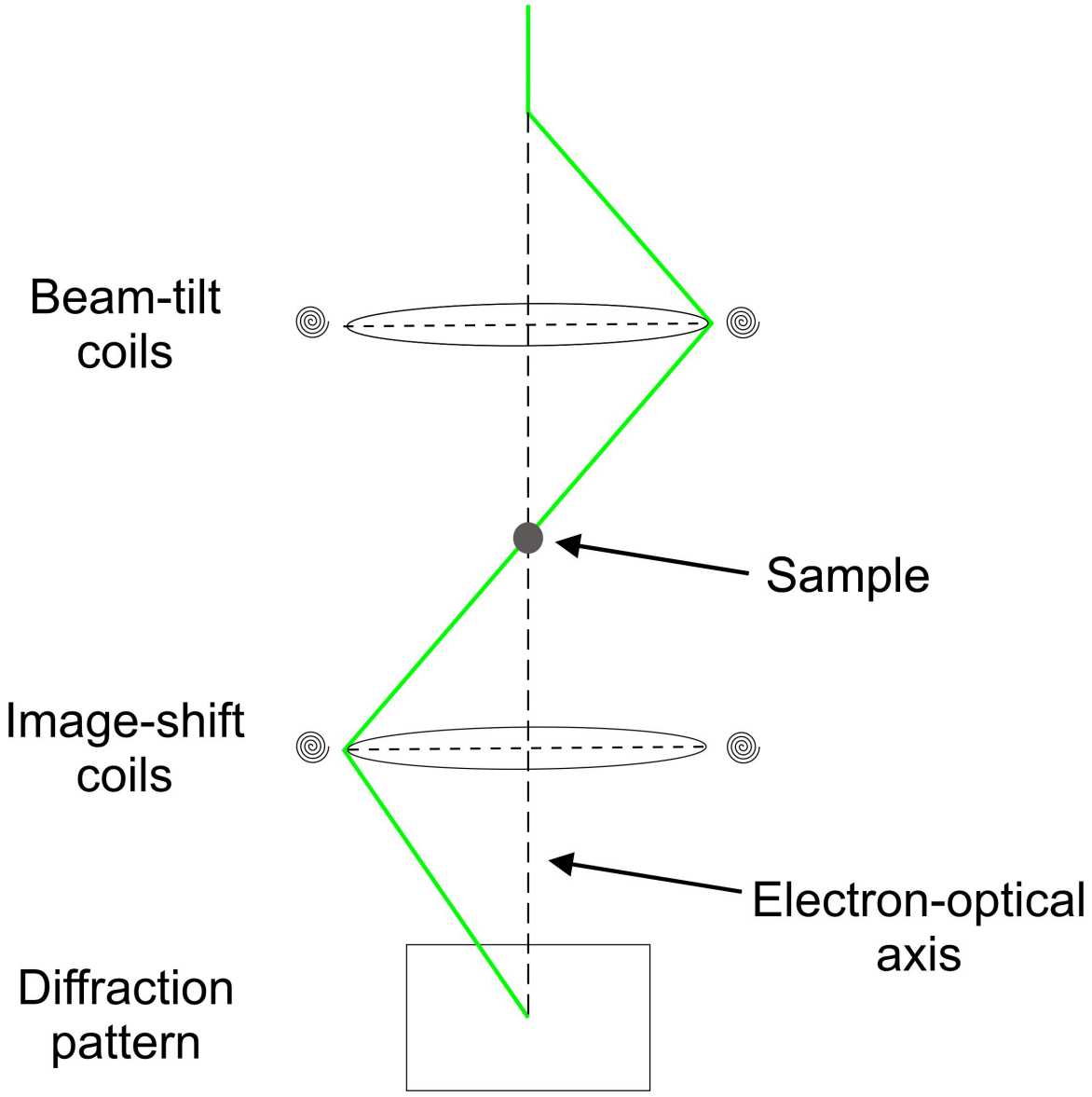

Figure 2.11: Schematic of a basic instrumental set-up for an electron diffraction experiment.

2.5 Thermogravimetric Analysis

Thermogravimetric analysis (TGA) is a technique that measures the mass of a substance as a function of temperature, while the substance is subject to a temperature control program. This technique is a simple, yet informative method of determining thermal characteristics of a material. The instrumental set-up involves placing a sample (typically ten milligrams) in a pan, which is then attached to a balance to determine changes in mass. In most cases, the furnace controlling the temperature is in direct contact with the sample, although external furnaces can be used for experiments at elevated temperature involving flammable gases such as hydrogen.

A variety of experimental parameters may be explored in a TGA measurement that will inherently affect the quality of the data collected, including (but not limited to) heating rate, furnace atmosphere, pressure and humidity. Heating rate is one of the most important experimental variables. For samples that exhibit multiple mass-loss steps upon heating, a slower heating rate may be essential to ensure that these steps do not overlap, which may be especially problematic when they occur within a narrow temperature range. By using a slower heating rate, the mass change is more clearly resolved in the data and the temperature at which each step occurs will be clearer. Also, if the heating rate is too fast, the primary step of a multi-step process may not be complete before the secondary step begins, which may lead to difficulties in data interpretation. A heating program known as 'sample controlled TGA' is now commonly provided in most TGA software and differs from a manually programmed heating schedule. It uses information on the rate of change in sample mass to automatically control the heating rate (e.g. when the sample mass changes rapidly, the heating rate is decreased). This approach is advantageous when overlapping processes are observed in TGA analysis. However care must be taken to prevent artificial steps in the data.

The processes observed as mass loss steps in TGA data are commonly attributed to either desolvation or a chemical reaction. Figure 2.12 shows a schematic set of TGA data for a single-step dehydration process, characterized by a sharp decrease in mass, followed by a plateau after completion of the dehydration process. This information can be used in the context of structure determination from powder XRD data, to understand whether a material is a solvate phase or a non-solvate phase. In the case of a solvate phase, this information assists in rationalizing the stoichiometry of the material (discussed in Sections 3.3, 3.4 and 3.5).

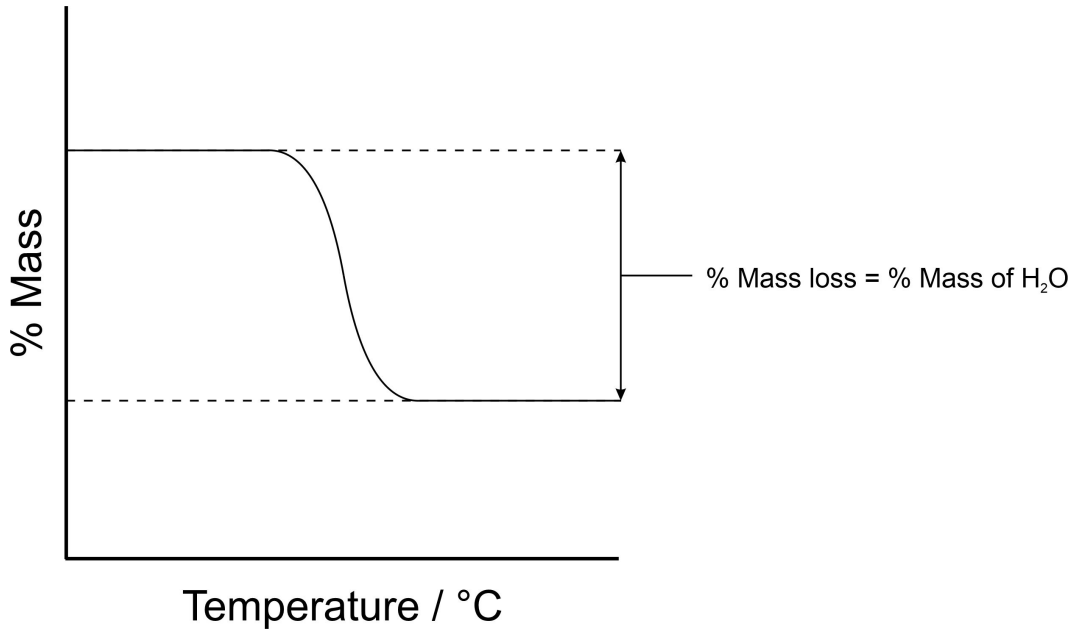

Figure 2.12: A model set of TGA data, showing the reduction in mass typical of a single-step dehydration.

2.6 Crystal Structure Prediction

Crystal structure prediction (CSP) refers to the prediction of all feasible crystal structures that may be formed for a given molecule based only on knowledge of its molecular structure [71]. Nowadays CSP is an accessible technique, although over three decades ago it was still far from a reality albeit new ground was slowly being broken. *"One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition"* [72] was a controversial quote from the former editor of *Nature* - John Maddox. This served to stimulate interest in tackling the challenge of crystal structure prediction in the 1980s. In the past three decades there has been significant advancement in this field, both from the theoretical perspective (including new methodologies for structure prediction) as well as developments in computer power, which have significantly expanded the scope of CSP calculations.

There are four conventional stages of CSP: i) molecular structure analysis, ii) initial generation of structures, iii) optimization of crystal structure energies, and iv) further optimization of crystal structures and property calculations [73].

i) Molecular structure analysis

In the first stage, several factors should be considered, including (but not limited to) the tautomeric identity of the molecules and consideration of the internal molecular geometry, both of which can significantly affect crystal structure energy. In the case of organic materials, tautomeric identity can strongly influence the possible hydrogen bonding arrangements and can significantly affect the crystal structure energy.

ii) Initial generation of structures

The second stage is the initial generation of crystal structure candidates. In the case of an achiral molecule or a racemic material, in principle any of the 230 space groups are possible, while in the case of a single enantiomer of a chiral molecule, only a subset of the 230 space groups is possible. In each case, this represents a huge search space, which must be significantly reduced for CSP to proceed efficiently. In this regard, it is recognized that certain space groups are more prevalent than others, in particular $P2_1/c$, which is the space group of over a third of all crystal structures in the Cambridge Structural Database [74]. Most CSP programs restrict the search for structures by considering only the most common space groups. The second consideration is the number (Z') of independent molecules in the calculation. In early work in the field, studies of structures with $Z' > 1$ were not feasible due to the increased computational complexity, although nowadays such structures can be predicted using CSP. The crystal structure generation can also be influenced using constraints. Minimum atom-atom intermolecular distances can be specified to eliminate structures containing unreasonable intermolecular interactions. The unit cell parameters can also be allowed to vary within a small range to allow the density to fluctuate towards the local energetic minima [75].

iii) Initial optimization of structures

At this stage, the structures generated from the structure prediction algorithms are filtered using relatively crude DFT calculations (in comparison to the level of calculations used in later stages), followed by more rigorous optimization. This is the case in both the AIRSS [76] program and the GRACE program [77] follow this approach, and both are arguably the two most widely used crystal structure prediction programs. In this stage, each structure is usually subjected to geometry optimization, with relaxed unit cell, at a higher level of accuracy than in the initial stage (in the AIRSS method, the PBE functional and TS dispersion-correction scheme are often used). Based

on the results, the structures are filtered, with high-energy structures discarded and only the most promising structures are allowed to proceed to the next and final stage [78].

iv) Final energy ranking and property calculations of predicted structures

This final stage begins with the final energy ranking. As discussed earlier, the PBE-TS methodology is fairly inexpensive from a computational aspect, although not as accurate as other approaches, such as PBE0-MBD. However, after filtering the predicted structures down to a relatively small number, it is feasible to apply a PBE0-MBD approach as a single-point energy calculation to reveal the most reliable energy ranking of the structures.

2.6.1 Analysis of Predicted Structures

Initially, the CSP calculations explore a huge search space, covering thousands or millions of trial structures, which are then filtered by the rejection of implausible structures and/or discarding duplicates. As the number of remaining structures is reduced, the level of computation can be increased proportionally. Often, the initial structure generation stage uses rather crude force fields for basic energy comparisons, but as the number of structures under consideration decreases, it becomes feasible to use periodic DFT-D calculations at the level of the PBE-TS methodology. As the search space is significantly reduced to a few trial structures, more comprehensive computational procedures can be used, effectively becoming more rigorous as the search space decreases, as shown in Figure 2.13.

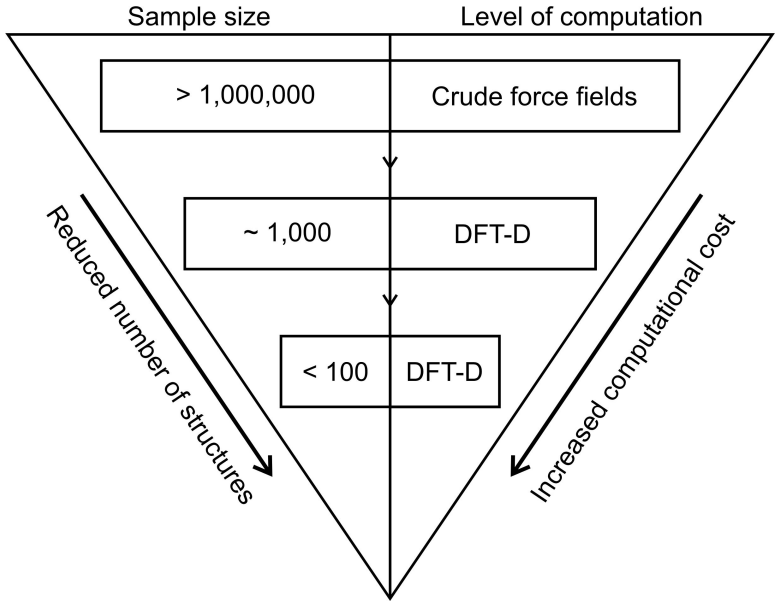

Figure 2.13: A diagram displaying how, as the number of structures under consideration in a structure prediction calculation decrease, the level of computation can increase.

The main approach for analysis of the energy landscape takes the form of a plot of crystal structure energy versus density, as seen in Figure 2.14. Such plots often show clusters of implausible

structures with high energy and low density, which become more spaced out as density increases and energy decreases. Figure 2.14 shows two common situations encountered within CSP [79]. In Figure 2.14 (a) the structure generated with the lowest energy (bottom right of the figure) corresponds to the known polymorph of a given material. In this case, the predicted structures of higher energy could be searched for experimentally using a range of crystallization techniques, especially if these structures have better properties for a desired application (e.g. increased solubility). In Figure 2.14 (b), the known polymorph is not the structure of lowest energy among those generated in the CSP calculation. In this scenario, further exploration would be required to obtain a more stable form of the material, and could shine light on whether samples of the known material could be contaminated with a more stable polymorph, as was the case with Ritonavir [80].

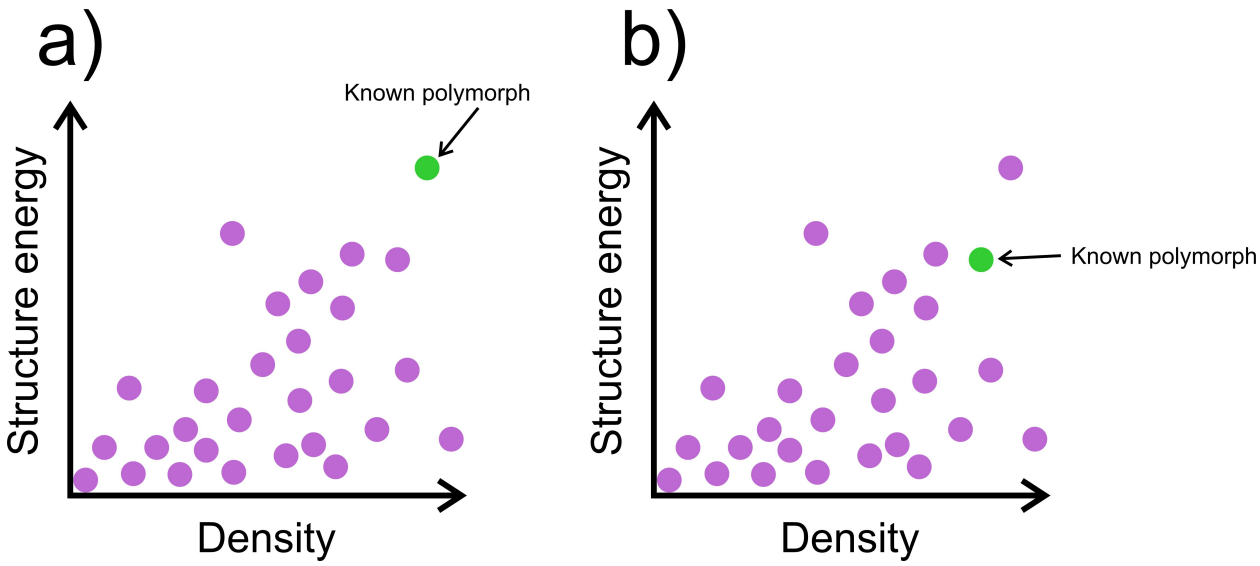

Figure 2.14: A visualization of the energy landscape in which structures generated from CSP are plotted according to their density and structure energy. Two different scenarios are represented in the case of a system in which only one polymorph is known experimentally: a) a case in which the most stable predicted structure is the known polymorph, and b) a case in which the most stable predicted structure is not the known polymorph.

In this work, all crystal structure prediction calculations were performed, and the data processed, by Prof Chris Pickard at the University of Cambridge.

Chapter 3
Crystal Structure Determination from Powder X-ray Diffraction Data

Crystal structures are usually determined through single-crystal X-ray diffraction. The procedure of determining a crystal structure by single-crystal XRD is nowadays typically routine, due to the highly automated steps involved in the process. Conversely, structure determination from powder XRD data is significantly more challenging and requires more expertise to avoid several pitfalls that can arise in the structure determination process.

Structure determination from single-crystal XRD data relies on measuring accurate values of the relative intensities of individual diffraction maxima, which are then used in structure solution, typically using direct methods or Patterson methods [81,82]. However, in the powder XRD diffraction data, severe peak overlap results in difficulties in extracting reliable values of the intensities of individual diffraction peaks. This can lead to failure of structure solution calculations using the types of "traditional" structure solution methods used for single-crystal XRD data. As discussed previously (in Section 2.1.1), the intensities of peaks in an XRD pattern are related to the corresponding crystal structure through the structure factor, F_{hkl}. The calculation of the diffraction pattern from a known crystal structure is an automatic calculation; however, determining a crystal structure from experimental diffraction pattern is inherently challenging. This is because only the amplitudes of the structure factor can be measured from experimental diffraction data, while the phases of the structure factors cannot be measured (the so-called "Phase Problem"). Direct methods or Patterson methods, as discussed earlier, provide estimated phase amplitudes. These estimated values of the phase amplitude can be used alongside the experimentally determined structure factor values to determine an approximate description of the electron density. It is from this approximate description of the electron density that the crystal structure can be derived; the quality of the description is subsequently improved through an iterative sequence that improves the quality of estimated amplitudes [83].

The process of determining a crystal structure from powder diffraction data begins with indexing (unit cell determination), followed by whole pattern profile fitting, structure solution and structure refinement. Following structure refinement, various approaches are typically used for "structure validation". This is depicted in Figure 3.1.

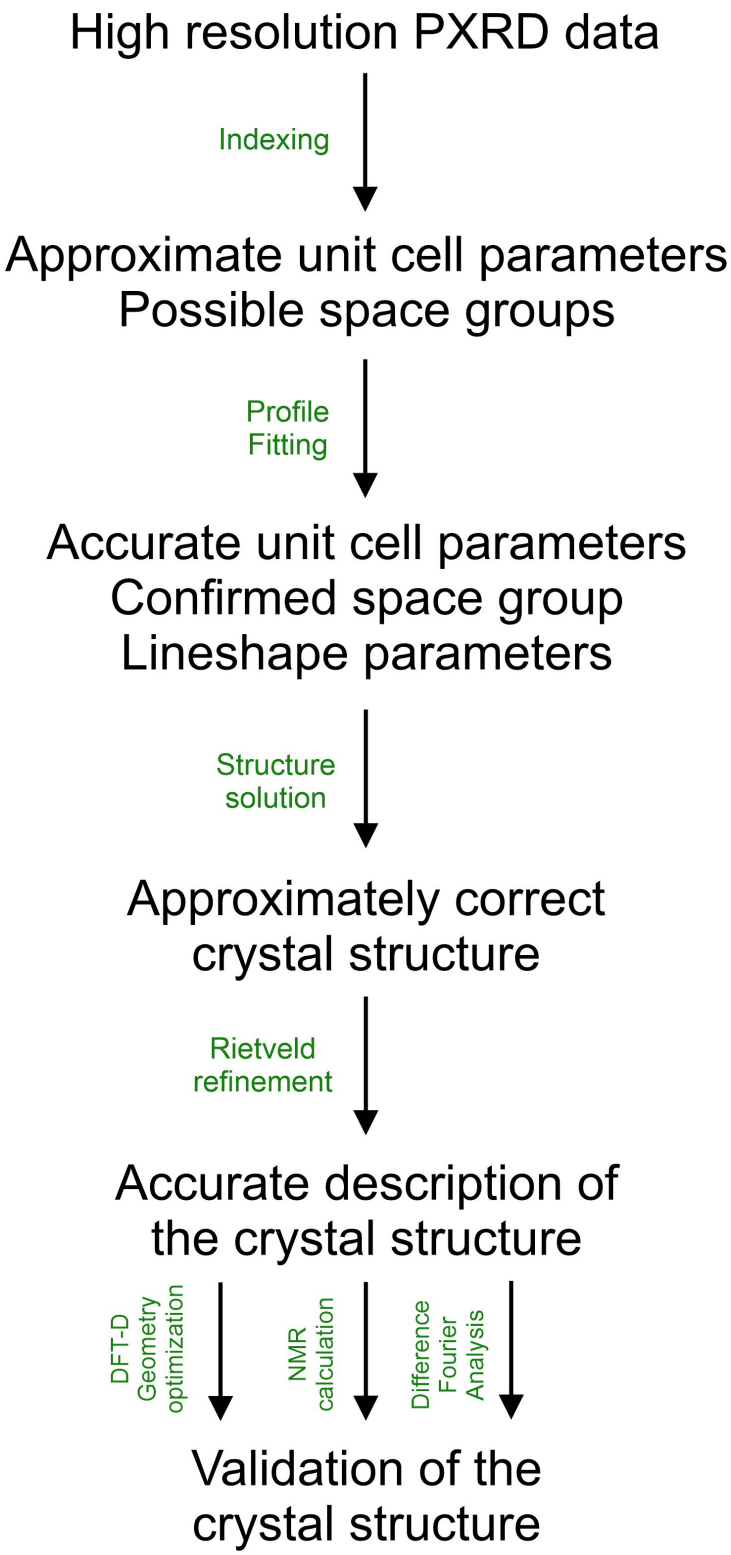

Figure 3.1: A flowchart depicting an ideal method of determining a crystal structure directly from powder XRD data, naming each step in green.

3.1 Indexing

The indexing procedure involves extracting peak positions from a powder diffraction pattern to determine the unit cell parameters (*a, b, c,* α*,* β*,* γ). Each peak is 'indexed' by having Miller indices (*h, k, l*) assigned to it. The space group of the structure can be speculated at this stage by considering which groups of Miller indices correspond to unobserved peaks, known as systematic absences.

To index a powder diffraction pattern, usually 20 or more peaks are identified in 'peak picking' software such as XFit [84] or TOPAS [85]. These programs assist in the peak picking process by applying a Pseudo-Voigt peak shape to fit each peak individually, to distinguish between overlapped and non-overlapped peaks.

The measured peak positions are then input into an indexing algorithm and, if the indexing calculation is successful, the correct unit cell parameters are obtained. There are many types of indexing algorithms, which can either process the information on peak positions exhaustively or deductively. An indexing algorithm that is exhaustive can be considered to systematically proceed through a variety of parameters, such as changing unit cell dimensions incrementally until a good match to the experimental positions has been obtained. A deductive method aims to make deductions to limit the size of the solutions field to increase the speed of the calculation. Some of the most common programs are ITO [86], TREOR [87] and DICVOL [88], which are deductive, semi-exhaustive and exhaustive respectively. These have been available for many years. These algorithms may be accessed using the CRYSFIRE interface [89], which applies a wide range of indexing algorithms to an experimental set of peak positions. In this work, all unit cells determined from powder XRD data used algorithms found in the CRYSFIRE interface, however as described later, there are a variety of different programs available which utilise a range of methodologies.

Some other, less converntional indexing strategies are also available to use, although not focused on in this work. This includes the use of a genetic algorithm, for example GAIN [90], or using artificial neural networking, implemented in PIANNO [91]. Other recent indexing algorithms include; N-TREOR [92], an evolved version of TREOR to make it fully exhaustive; McMaille [93], which indexes via a Monte-Carlo approach; a Singular Value Decomposition (SVD) method [94], which is implemented in the TOPAS software from Bruker AXS; X-CELL [95], which is implemented in Materials Studio from Accelrys, and performs an exhaustive search similar to the DICVOL approach.

3.2 Whole Pattern Profile Fitting

The process of profile fitting involves refining the unit cell obtained in the indexing stage against the experimental powder XRD data, as well as refining the peak shape parameters and instrumental

parameters. The factors affecting peak shape can include; equatorial and axial divergence, the emission profile of the radiation, the crystallite size, and crystallite microstrain. By careful inspection of the quality of fit obtained in profile fitting, using different unit cells and space groups suggested in the indexing stage, the correct unit cell and the correct space group is established from the powder XRD data. Thus, the unit cell and space group, along with various other parameters that accurately describe features of the powder XRD pattern, including the peak shape and peak width functions, can be extracted following profile fitting, and are then used in the structure solution process. There are two commonly used methods for carrying out profile fitting, developed by Le Bail [96] and Pawley [97] respectively. The main difference between the two methods is that in the Le Bail method the calculation of peak intensities (which are arbitrary) is performed by using a modified version of the Rietveld code, which requires a smaller search space for full optimization of all refined parameters. In the Pawley method, all the reflection intensities are subjected to non-linear least squares refinement, which is more comprehensive albeit more computationally expensive. However, it is worth mentioning that computational resources are sufficiently high powered that, the methods take the same length of time and give the same agreement factor in most refinement programs.

The peak shape can be fitted using a variety of mathematical functions, including the pseudo-Voight function described by Thompson, Cox and Hastings [98], which contains a correction for the peak asymmetry caused by axial divergence provided by Finger, Cox and Jephcoat [99]. Peak shape is dependent on several factors that are either instrumental or sample based. Sample effects include peak broadening caused by particle size, strain and stress, stacking faults (disorder) and sample roughness. The instrumental effects contributing to peak shape include the diffractometer geometry, the nature of the radiation (e.g. the type of radiation and how monochromatic it is), detector type, and axial divergence (which leads to peak asymmetry at low angles). The pseudo-Voight function can be split into two parts, the Gaussian part (Equation 3.1) and the Lorentzian part (Equation 3.2).

$$fwhm_{G} = \sqrt{U\tan^{2}\theta} + V\tan\theta + W + \frac{Z}{\cos^{2}\theta} \tag{3.1}$$

$$fwhm_{L} = X\tan\theta + \frac{Y}{\cos\theta} \tag{3.2}$$

where U and X are related to microstrain, Z and Y are related to domain size, and W is related to particle size.

The 'agreement' between the experimental and calculated powder diffraction patterns is described by the weighted (R_{wp}) or unweighted (R_{p}) powder profile R-factors, calculated using:

$$R_{wp} = 100 \times \sqrt{\frac{\sum_{i} w_{i}(y_{i} - y_{ci})^{2}}{\sum_{i} w_{i} y_{i}^{2}}} \tag{3.3}$$

$$R_p = 100 \times \sqrt{\frac{\sum_i |y_i - y_{ci}|}{\sum_i |y_i|}} \tag{3.4}$$

where y_i is the intensity of the i^{th} data point in the experimental data, y_{ci} is the intensity of the i^{th} data point in the calculated powder XRD data, and w_i is a weighting factor defined by:

$$w_i = \frac{1}{y_i} \tag{3.5}$$

The resulting R_{wp} and R_p values from a Le Bail or Pawley fit to an experimental powder XRD pattern using the correct unit cell and space group provide the 'best possible' fit that could be achieved in a subsequent Rietveld refinement [96, 97]. The closer the value of the R_{wp} value from a Rietveld refinement is to the Le Bail or Pawley fit R_{wp}, the more reliable the refinement. However, visual inspection and not simply quantitative analysis is critical in investigating the difference profile between the two.

3.3 Structure Solution

The process of solving crystal structures using powder diffraction data was originally conducted using the types of calculation (e.g. direct methods of Patterson methods) that had been developed to solve crystal structures from single-crystal diffraction data. The downside to this approach is the requirement for accurate extraction of individual intensities I_{hkl} from the power XRD data, which is often impossible due to severe peak overlap. The initiation of the direct-space strategy in 1994 brought forward a new method for solving structures from powder XRD data. This strategy (which follows a close analogy to "global optimization") involves the generation of trial structures using chemical knowledge of the given material, such as internal molecular geometry. This is then combined with the unit cell and space group determined from the indexing and profile-fitting stages [3]. The trial structures can then be optimized using global optimization methods. Initially Monte-Carlo (MC) methods [100] were used for this process although more sophisticated methods are used now, in particular the genetic algorithm (GA), which is used in this work [101].

There are, however, a wide variety of techniques that utilize the direct-space strategy for structure solution rather than just the GA approach used in this work, and these are detailed further in this section. The simplest method is a standard grid-search method, in which a grid is defined over the parameter search space, and it is systematically explored via every grid point. This is simple and easily implemented however is only suitable to search spaces with a small number of parameters [102]. Another set of methods that is more widely used in the powder diffraction field are the MC methods, which are stochastic search methods based on a random sampling of the parameter space. In the MC approach, a sequence of structures is generated and each structure

is defined by internal and external degrees of freedom. The structures along the sequence are different due to random variations of these degrees of freedom, which are known as MC moves. Simulated annealing (SA) [103] techniques are also performed and, whilst they are similar to MC methods, they differ in how they are scaled in the sampling algorithm. In MC, the scaling factor is fixed or varied empirically, whereas the scaling factor is slowly reduced in SA.

Structure solution calculations using the direct-space strategy involve the definition of molecules (fragments) that comprise the asymmetric unit of the crystal structure constructed using standard bond lengths and angles for the molecule of interest. As such, the only unknown information about the intramolecular geometry in the crystal structure are the torsion angles that define the molecular conformation. The variables involved in the direct-space search are then the position (*x, y, z*) of the molecule in the unit cell, the orientation (θ, ϕ, ψ) of the molecule relative to the unit cell axes, and the unknown torsion angles, τ_1, τ_2....τ_n, required to define the molecular conformation. These variables are depicted in Figure 3.2.

For each trial structure, a powder XRD pattern is simulated using the profile parameters determined in the profile fitting stage, and the R_{wp} value is determined to assess the quality of fit between the simulated and experimental powder XRD data.

The program EAGER (Evolutionary Algorithm Generalized for Energy and R-factor) uses evolutionary principles, by analogy to those developed by Charles Darwin, as depicted in Figure 3.5. Each trial structure in the calculation is defined by a set of structural variables, which define the "genetic code" of each trial structure, shown as:

$$(x, y, z, \theta, \phi, \psi, \tau_1, \tau_2 ... \tau_n) \tag{3.6}$$

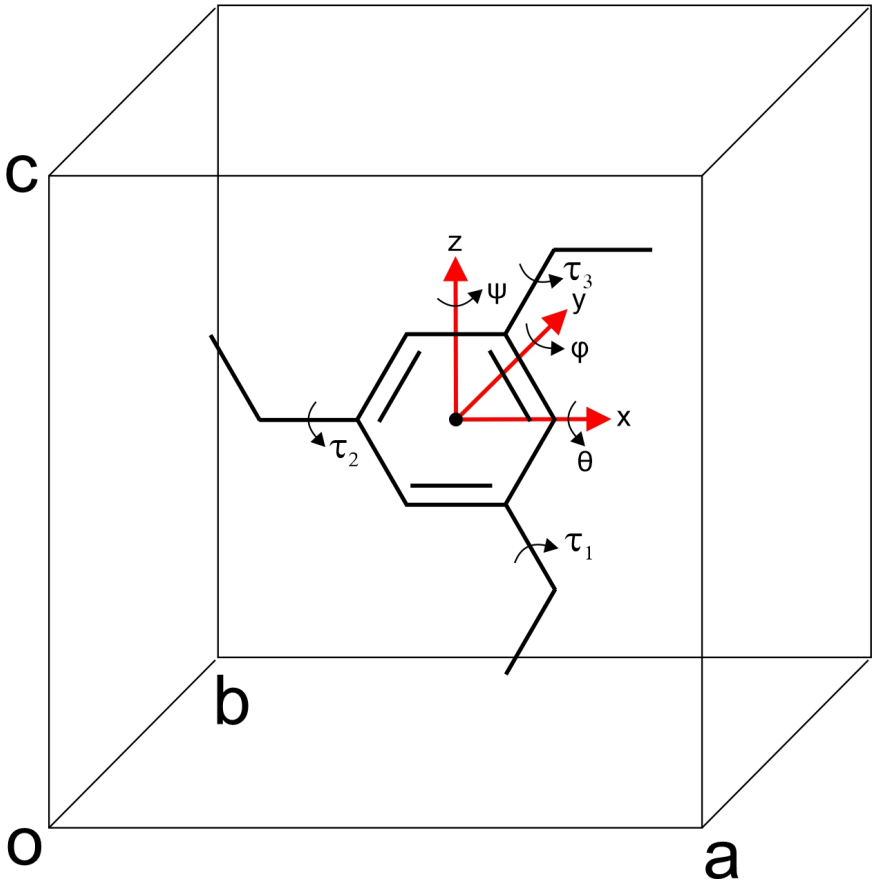

Figure 3.2: An example illustrating the definition of the structural variables in a direct-space structure solution calculation.

In the initial population, the genetic code defining each trial structure in the population is generated using random numbers for each variable. Within the GA, one of the evolutionary operations is "mating", in which the trial structures are allowed to mate with each other to form new structures (offspring). This is done by selecting a pair of structures (known as parents) from the population; the probability of selecting a trial structure for mating is proportional to its fitness. The fitness of a structure is determined by the quality of fit between the powder XRD data simulated for the trial structure and the experimental data. Superior fitness corresponds to a better fit to the experimental data (i.e. lower R_{wp}). In the mating procedure, groups of variables from the genetic codes of the two parent structures are swapped to generate two new structures known as offspring, shown in Figure 3.3.

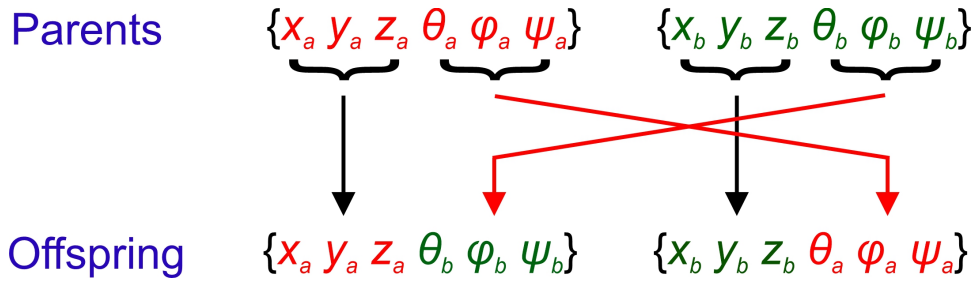

Figure 3.3: Definition of one method for conducting mating between two trial structures within the GA involving swapping of the positional and orientational variables between the two parent structures to form two offspring.

Following the creation of the offspring, they are then subjected to local minimization of R_{wp}. This is analogous to Lamarckian principles of evolution, defined by Jean-Baptiste Lamarck [104]. In this local minimization, the variables defining each trial structure are varied locally to find the nearest local minimum on the R_{wp} hypersurface. The offspring structures following local minimization are then added to an intermediate population of trial structures containing all the offspring together with the structures in the previous population from which the offspring were generated. At this stage, natural selection takes place allowing only the best trial structures from the intermediate population (assessed on the basis of their fitness) to pass to the next generation. Simultaneously, several mutant structures are generated from the structures in the intermediate population by assigning new random values (denoted ^{R}) to some of the structural variables or making random displacements to some of the variables (Figure 3.4). Mutant structures are then subjected to local minimization as described above for the offspring. This set of mutants, as well as the best trial structures from the intermediate population that passed through the natural selection process, are then used to make up the next population of trial structures. The whole GA process conducted by the EAGER program is depicted in Figure 3.5.

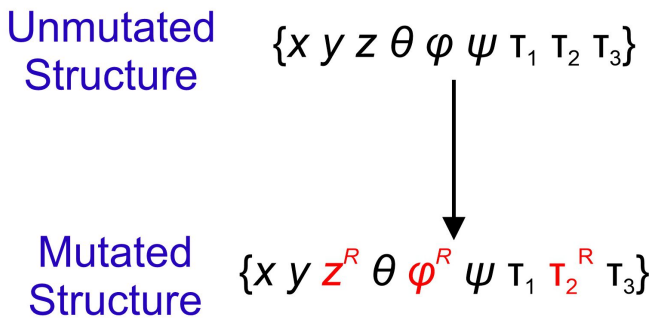

Figure 3.4: Definition of one method for generating a mutant structure within the GA. A mutant structure is one which contains one or more structural variables that have either been assigned new random values, or have been given a random displacement to change the value, both are denoted with a ^{R} notation.

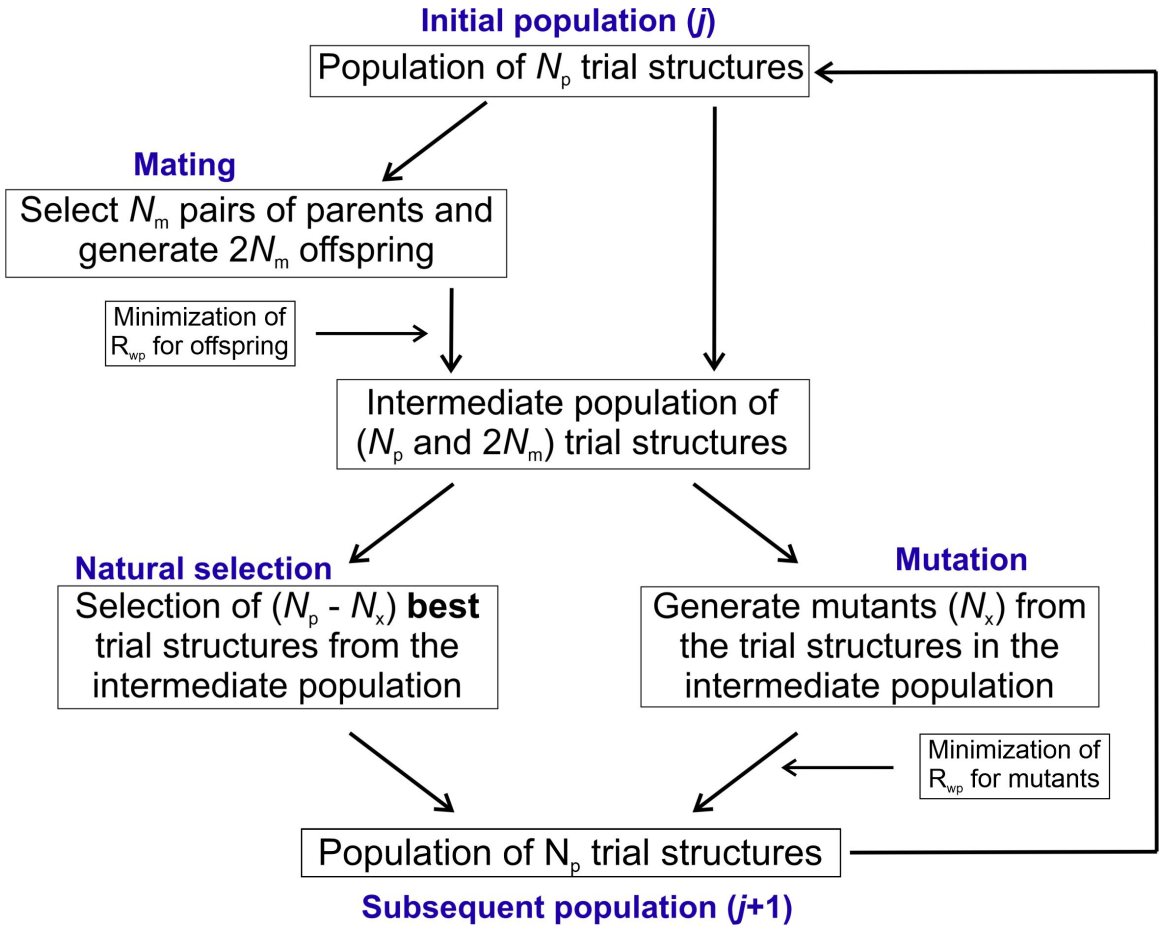

Figure 3.5: Flow chart showing the steps involves in the evolution from population *j* to population *j+1* in the GA used in the direct-space structure solution program EAGER.

The GA calculation can run indefinitely, only stopping at a practical number of generations which is manually defined by the user, to fit within the maximum calculation time allowed. However, at some point it is not pertinent to keep running the calculation depending on the progress of the GA. The calculation progress can be analysed by observing the change in R-factor as a function of the number of generations; the calculation is usually stopped when the R-factor stops reducing substantially, which can be subjective. The changes in R-factor may be minimal, however they could have a noticeable effect on the crystal structure that it represents, which is why a subjective evaluation as such as this one is not ideal. One has a greater degree of understand of the calculation when viewing whether the structure is changing significantly over a number of generations or whether the structure is already close to the global minimum. Another method to determine when a GA calculation can be concluded is to compare the R-factor of the lowest R-factor structures from the whole array of processors (as the calculation is run in parallel among multiple processors). When most, if not all, of these structures exhibit a similar R-factor the calculation can be concluded. These structures can also be assessed to determine how similar they are, often by calculating the RMSD of all other structures by using the lowest R-factor structure as an origin point. If they show very low RMSD in atom positions, then the array of processors may have converged

on a common structure solution. Another function of EAGER is the ability to perform a "diversity check". This diversity check identifies when such a convergence upon a single crystal structure has occurred and intends to re-invigorate the structure solution calculation with additional diversity. This is done by taking a small sample of structures in a population and varying the structural variables randomly to then mate with the existing, stagnated structures. The key to achieving a successful structure solution calculation using a GA is to strike the correct balance between elitism (which leads to a quick convergence) and diversity, which leads to a slow convergence but tends to explore a greater search space. The elitist approach is useful with ideal data and simple structures, whereas the diversity approach tends to be more exhaustive and is favourable for non-ideal data (e.g. affected strongly by preferred orientation) and more complex structures.

3.4 Structure Refinement by the Rietveld Method

The next step of structure determination from powder XRD data is Rietveld refinement, which takes the structure of best quality from the structure solution calculation and aims to derive a more accurate description of this structure by variation of the structural variables (atomic coordinates in the unit cell, atomic displacement parameters and unit cell parameters) and profile parameters. This is done to achieve the optimal fit between calculated and experimental powder XRD data. The Rietveld method was invented by Hugo Rietveld in 1969 and is still to this day the most widely used method for refining crystal structures using powder diffraction data [105]. In a Rietveld refinement, the variables affecting the intensities of peaks in the powder XRD data are allowed to refine, including atomic coordinates and isotropic atomic thermal displacement parameters. In some cases, site occupancies as well as parameters that define preferred orientation in the powder sample are also allowed to refine.

Preferred orientation can be modelled in the Rietveld calculation by one of two methods: i) the March-Dollase method [106], which is the most common and most simple correction or ii) the Spherical Harmonics method [107], which is more comprehensive although is often the cause of over-refinement. Firstly, the March-Dollase method involves the definition of one or two lattice planes (*hkl*) that are most affected by the preferred orientation, which are then corrected for either over- or under-representation in the powder XRD pattern relative to random distribution. For each direction defined there is an abundance ratio that is greater or lower than one; when equal to one then a total random distribution is present. The second method, the Spherical Harmonics approach involves defining a shape that defines the relative amounts of crystallites in each direction. This process involves a much greater number of refinable parameters in the Rietveld calculation and can often correct for preferred orientation where the March-Dollase approach cannot. The method can often lead to over-refinement so care must be taken to use this type of correction sparingly and only when supplemented by experimental evidence, such as two-dimension powder XRD data.

When a crystal structure has been refined against the powder diffraction data, an ideal result would be a perfect fit to the data. However, this is not practical as sample and instrumental factors, as well as the presence of random background noise in the powder XRD data, means no fit to experimental data can be perfect. The profile fitting result can be considered a "best case" fit for the subsequent Rietveld fit, as the peak intensities are free parameters rather than calculated from the crystal structures, meaning that there is less restraint towards this fitting. Therefore the aim in a Rietveld fit is to fit as close to the peak profile and intensity as possible, which can be achieved qualitatively by viewing the fit manually, or quantitatively by analysing the R_p and R_{wp} values. A poor fit would indicate that some aspects of the structure or the description of the powder diffraction profile may not be correct; for example, if the experimental and calculated peak intensities do not match completely, the possibility to correct for preferred orientation in the sample may be considered, by using either the March-Dollase model or by the Spherical Harmonics approach. Following Rietveld refinement, other methods to validate the structure can be performed, such as geometry optimization by periodic DFT-D calculations.

3.5 Validation

When determining crystal structures primarily from powder XRD diffraction data, each stage of the structure determination process requires careful scrutiny to ensure that the resultant crystal structure is correct and accurate. Without careful diligence, simple errors in any step of the process can lead to the wrong structure being obtained. Prior to the structure solution stage, the composition of the unit cell needs to be fully understood, as success in direct-space structure solution relies on significant prior chemical knowledge of the material (i.e. molecular identity), as well as the knowledge of presence of any solvent in the crystal structure. The number of molecules in the unit cell of a crystal structure can be deduced from density considerations, which typically falls within a narrow range for organic materials. The identity of the molecules present in the crystal structure can be understood through simple analytical techniques such as liquid ^{1}H NMR spectroscopy as well as solid-state ^{13}C NMR, supplemented by TGA measurements to identify the presence and number of solvent molecules within the crystal structure.

Following completion of an acceptable Rietveld refinement, giving a quality of fit that is close to the whole pattern profile fitting stage of the structure determination process (see above), other validation strategies should be applied to confirm the veracity of the crystal structure. Careful inspection of the structure is a classic method of determining whether there are any voids in the structure, rationalizing any short contacts and especially for organic materials, whether the hydrogen bonding interactions are chemically and structurally reasonable. Nowadays the CheckCIF facility provided by the International Union of Crystallography (IUCr) is a common tool used in the validation step, although relying solely on this is insufficient. Fortunately, the crystallographic community has access to various other computational techniques to allow for additional quantitative

assessment. The program PLATON can be used to determine whether any solvent accessible voids are present within the structure. To support this analysis, a Fourier difference map can be generated using the Rietveld refinement code (e.g. GSAS or TOPAS) and can be used to qualitatively and quantitatively analyze any residual electron density that is not accounted for by the refined structural model that (which can for example, be inferred by missing solvent). This tool can also be used to explore the possibility of disorder.

As discussed in Section 2.3.1, periodic DFT-D calculations provide additional scope for validation. The use of a geometry optimization calculation during minimization of the overall crystal structure energy is valuable for ensuring that the refined crystal structure corresponds to an energy minimum, on the energy landscape, for the molecule of interest. If this process leads to improved geometric features (e.g. hydrogen bonding) in the crystal structure, it would then be advisable to perform subsequent Rietveld refinement calculations in which the structural restraints used in the refinement are updated to reflect these improved geometric features. Another aspect of periodic DFT-D calculations is that they can be used to calculate NMR chemical shifts for a crystal structure, which can then be used to compare against experimental solid-state NMR data, adding another aspect of validation into the procedure.

Chapter 4
Insights on the Crystal Structures of Riboflavin and Alloxazine

4.1 Introduction

4.1.1 Riboflavin

While the work presented in this chapter was close to completion, a structure of riboflavin was published by Guerain *et al*. [108]. This structure is almost identical to the structure determined from the initial powder XRD study in the present work, but now known (as discussed in Section 4.3.2) to have an incorrect description of the intermolecular hydrogen-bonding at the end of the ribityl side-chain.

Riboflavin, shown in Figure 4.1, otherwise known as vitamin B_2, is a vitamin well known for its crucial role in human health. It acts as a precursor to both flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), which take part in a range of redox reactions in human metabolism [109]. Recent research suggests riboflavin may have a role in protecting against cancer and cardiovascular disease. Given the clear importance of riboflavin in human health, the endemic riboflavin deficiency among some populations is a major cause of concern [110].

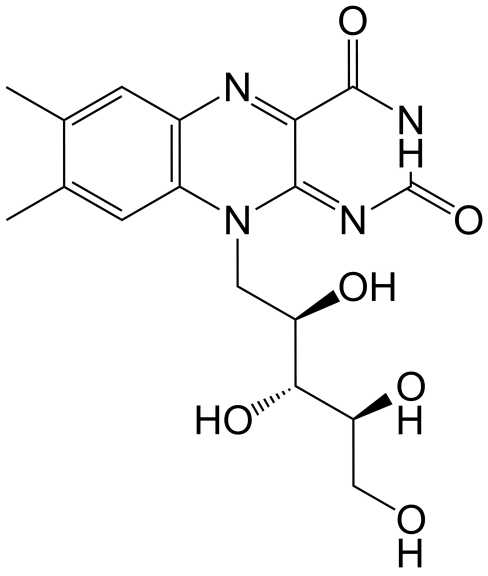

Figure 4.1: The molecular structure of riboflavin.

Riboflavin can be crystallized as orange crystals, which have been shown to be present within the tapetum lucidum in the eyes of lemurs and cats, leading to “eye-shine” phenomena in these animals (Figure 4.2) [26, 39]. However, the ability to understand how crystalline riboflavin achieves this biological function has been limited by the absence of knowledge of the structural properties of riboflavin in the crystalline state.

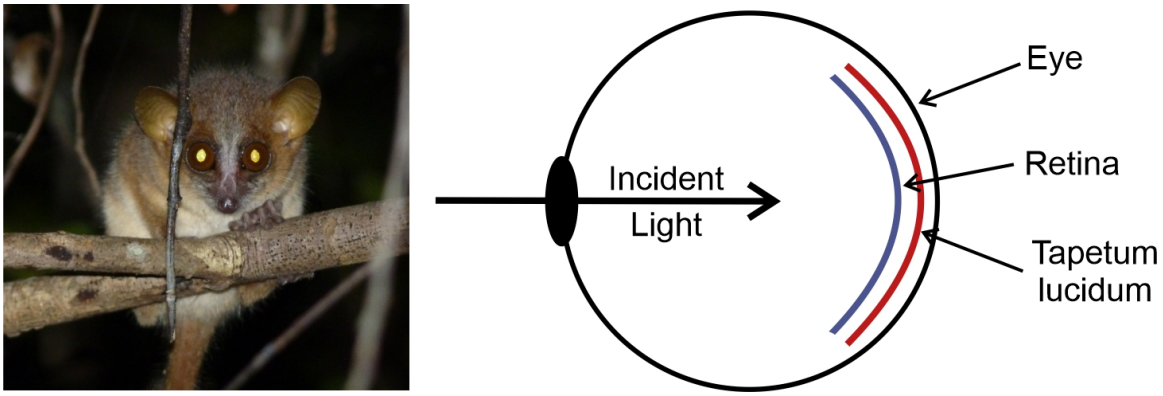

Figure 4.2: An image of a lemur exhibiting ’eyeshine’ [111] as a result of the tapetum lucidum (left) and a diagram showing the retroreflective mechanism of the tapetum lucidum (right).

The tapetum lucidum is a layer of tissue behind the retina of the eye, which acts as a secondary reflector of light. The light which is transmitted through the retina hits the tapetum lucidum and is reflected back onto the retina, which increases the light available to the photoreceptors [112]. This process facilitates superior night vision in nocturnal animals as well as deep sea organisms, and it also leads to the ’eyeshine’ phenomena seen by some animals in low light (example in Figure 4.2). The reflective material in the tapetum lucidum differs between different organisms. For example, riboflavin is present in the tapetum lucidum of lemurs and cats; guanine is present in numerous types of fish; and a zinc-cysteine complex is present in dogs and ferrets [26].

The tapetum lucidum of the lemur was studied extensively by Antoinette Pirie, who showed that the riboflavin crystals extracted from lemurs have a plate morphology [39]. It is reported [39] that powder X-ray diffraction photographs were recorded for biogenic samples of riboflavin and compared with that of a sample of riboflavin crystallized *in vitro*, showing that they were identical crystal forms. However, the powder XRD data are unavailable in the public domain, and the specific solid form (e.g. polymorph) of riboflavin present in the biological systems remains unknown. It is known [113] that crystals of riboflavin extracted from the lemur have a plate-like morphology, whereas those extracted from the cat have a rod-like morphology, suggesting (but not proving) that they may represent different crystal structures of riboflavin. The flat plate single crystal has two relatively long edges and a short edge, whereas the needle-like single crystals have two relatively short edges and a long edge. Work done by Mohnicke and Gloor has shown that riboflavin can exist in different anhydrous polymorphs and hydrate structures [114, 115], but the crystal structures were not reported by these authors. At this stage it is unknown which (if any) of

these structures correspond to the biogenic forms of riboflavin and further structural studies are required to understand the true crystallographic nature of the biogenic form. There were a variety of methods used to produce the forms discovered by both Mohnicke and Gloor. These methods included fermenting riboflavin in a reaction slurry produced by microorganisms, varying the relative humidity of the environment to dehydrate hydrated riboflavin forms, and by crystallizing riboflavin from a formaldehyde solution.

In the work presented herein, the crystal structure of riboflavin has been determined, initially by powder XRD methods (in conjunction with DFT-D calculations), followed by further investigation into the structure using micro-crystal XRD data. The subsequent investigation of the structure using micro-crystal XRD data was initiated by the opportunity to provide samples for analysis at Rigaku Japan on a microfocus rotating anode X-ray diffractometer to showcase developments in analysing small crystals. The riboflavin sample contained needle-like crystals approx. 1 micron in width, and approx. 10-15 microns in length, which provided a substantial challenge for a modern day X-ray diffractometer.

As discussed below, while the crystal structures obtained from these approaches are very similar, they differ on the details of the intermolecular hydrogen bonding at the end of the ribityl side-chain. Additional investigations using higher level DFT-D calculations and analysis of solid-state NMR data support the conclusion that the hydrogen bonding arrangement in the structure determined from micro-crystal XRD is the correct structure.

4.1.2 Alloxazine

The heterocyclic alloxazine ring system is found in many biologically important molecules. For example, it is present in all flavins, such as riboflavin, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Both FMN and FAD are crucial coenzymes for redox processes in a wide range of biological systems [116]. Alloxazine can exist in two tautomeric forms, referred to as alloxazine and isoalloxazine (shown in Figure 4.3), which only differ in the location of one hydrogen atom. Recently, alloxazine has been shown to have potential in battery technology as an energy storage material, with functionalization of alloxazine by electron-donating groups leading to improved performance in this application [117]. To date, the crystal structure of alloxazine has not yet been reported; clearly, knowledge of the crystal structure of alloxazine and the preferred modes of intermolecular interaction may yield important insights in understanding the structural chemistry of flavins in different environments.

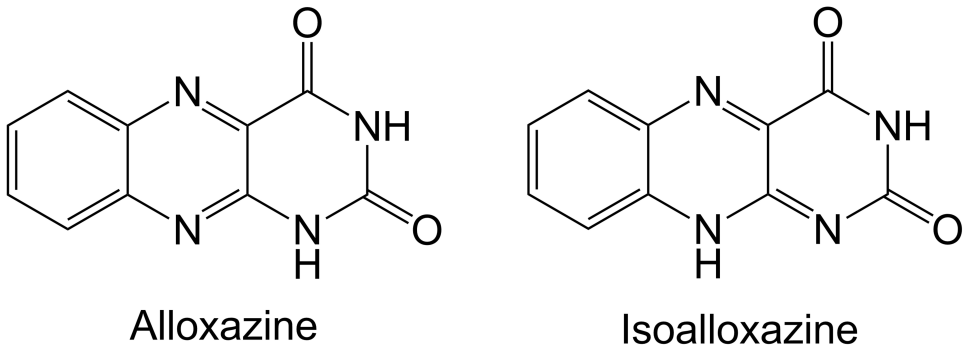

Figure 4.3: The structures of the alloxazine tautomer (left) and the isoalloxazine tautomer (right).

4.2 Riboflavin

4.2.1 Experimental Methods

Sample Preparation

(-)-Riboflavin, produced from Eremonthecium ashbyii ($\geq$ 98%) was purchased from Sigma-Aldrich. Crystallization of riboflavin was conducted by adding an amount of this sample (0.0315 g) to pre-heated deionized water (50 mL; 80 °C) and allowing dissolution to occur over a 20 minute period. The mixture was then filtered while hot to remove any excess solid riboflavin. The solution was then placed in an incubator at 80 °C, and the temperature decreased at a rate of 0.02 °C min^{-1}. When the solution had reached room temperature, it was transferred to a vacuum desiccator (containing dry silica as desiccant). Each day the desiccant was dehydrated, and the desiccator was re-evacuated. After 28 days, a sufficient amount of powder (with needle-like morphology) had precipitated and powder XRD data were recorded.

Powder X-ray Diffraction

High-quality powder XRD data were recorded at 21 °C on a Bruker D8 Advance Series II diffractometer operating in transmission mode using Ge-monochromated Cu-Kα_1 radiation (2θ range, 4° - 70°; step size, 0.016°; data collection time, 86 hr). The recrystallized sample of riboflavin was subjected to light manual grinding using a pestle and mortar and loaded into a 0.7 mm glass capillary, which was then flame sealed.

Micro-crystal X-ray Diffraction

Micro-crystal XRD data were collected at room temperature on a Rigaku XtaLAB SynergyCustom FR-X diffractometer with a microfocus rotating anode X-ray generator source (Cu-Kα radiation), using multi-later confocal optics as a monochromator. The crystal structure was solved by SHELXTL 2018/2 and refined using SHELXL2018/3 [118, 119]. Non-hydrogen atoms were refined with anisotropic displacement parameters. Ideal geometry and riding coordinates were used for hydrogen atoms with U_{iso} set to 1.2 times (for CH, CH_2 and all NH groups) or 1.5 times (for CH_3 and OH groups) the equivalent isotropic displacement parameter for the atom to which they are bonded.

Solid-state ^{13}C NMR Spectroscopy

High-resolution solid-state ^{13}C NMR data for riboflavin were recorded on a Bruker AVANCE III spectrometer (magnetic field, 9.4 T; ^{13}C Larmor frequency, 100.64 MHz), with the powder sample packed in a zirconia solid-state NMR rotor. The solid-state ^{13}C NMR spectrum was recorded at ambient temperature using ramped ^{1}H $\rightarrow$ ^{13}C cross-polarization (CP), magic-angle-spinning (MAS spinning frequency,12 kHz) and ^{1}H decoupling using the SPINAL-64 sequence [120]. The ^{13}C

NMR spectrum was referenced [121] against the α polymorph of glycine, with the carboxylate resonance set to 176.5 ppm, corresponding to tetramethylsilane (TMS) as the primary reference.

Periodic DFT-D Calculations

Periodic DFT-D calculations for geometry optimization were conducted (with fixed unit cell) using CASTEP [122] (Academic Release version 8.0). These calculations used ultrasoft pseudopotentials [123], PBE functional [124], semi-empirical dispersion corrections using the TS correction scheme [125], conserved space group symmetry, fixed unit cell, periodic boundary conditions, a basis set cut-off energy of 700 eV and a Monkhorst-Pack grid [126] of minimum sample spacing (0.05 x 2π) Å^{-1}. The convergence criteria for geometry optimization were 0.01 eV Å^{-1} for atomic forces, 0.00001 eV per atom in the total energy, and 0.001 Å for atomic displacements.

Solid-state NMR chemical shifts were calculated using CASTEP (Academic Release version 8.0) following DFT-D geometry optimization of the respective crystal structures investigated in this work. The Gauge Including Projector Augmented Wave (GIPAW) approach [127–131] was used with a cut-off energy of 700 eV and PBE functional. From the isotropic ^{13}C NMR shielding value (σ_{calc}) calculated for each ^{13}C environment in the crystal structure, the corresponding calculated value of isotropic ^{13}C NMR chemical shift (δ_{calc}) was determined [129] from the equation $\delta_{calc} = <\delta_{exp}> + <\sigma_{calc}> - \delta_{calc}$, where $<\delta_{exp}>$ denotes the mean of the isotropic ^{13}C NMR chemical shifts determined from the experimental high-resolution solid-state ^{13}C NMR spectrum ($<\delta_{exp}>$ = 107.08 ppm) and $<\sigma_{\text{calc}}>$ denotes the mean of the calculated isotropic ^{13}C NMR shielding values. For the three riboflavin structures (structure 1, structure 2, and the Guerain structure as mentioned previously) the corresponding $<\sigma_{calc}>$ values were 65.01 ppm, 65.64 ppm, and 65.01 ppm, respectively.

4.2.2 Crystal Structure Determination from Powder XRD data

The high-resolution powder XRD pattern for riboflavin was indexed using the TREOR90 algorithm in the CRYSFIRE package [87, 89], giving the following unit cell with orthorhombic metric symmetry: a = 20.12 Å, b = 15.15 Å, c = 5.35 Å (V = 1630.8 Å^3). The high-resolution solid-state ^{13}C NMR spectrum suggests that there is one molecule of riboflavin in the asymmetric unit (Z' = 1). From assessment of systematic absences, the space group $P2_12_12_1$ was assigned. With one molecule in the asymmetric unit, the approximate density is 1.53 g cm^{-3}, which is reasonable for a molecular organic material. Unit cell and profile refinement was conducted using the Le Bail method [96, 132] in the GSAS program, which gave a good quality of fit to the experimental powder XRD data (R_{wp} = 0.54%, R_p = 0.39%), shown in Figure 4.4.

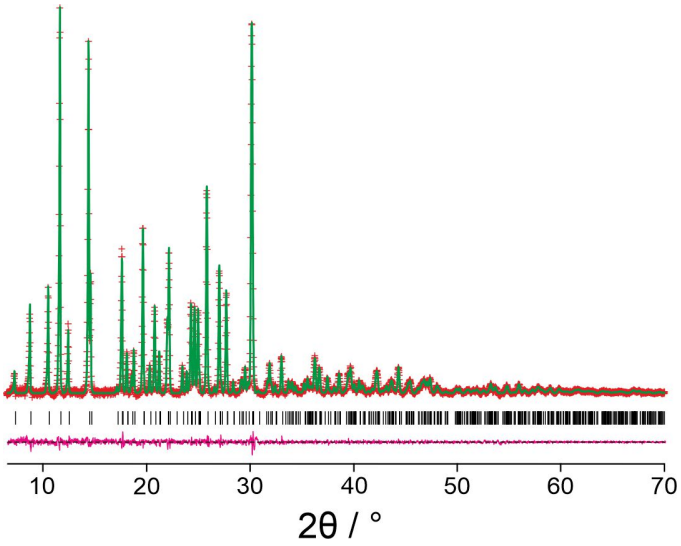

Figure 4.4: Profile fitting (using the Le Bail method) of the powder XRD data (background subtracted) recorded for riboflavin, showing the experimental powder XRD data (red crosses), calculated powder XRD data (green line), predicted peak positions (black ticks) and difference between the experimental and calculated powder XRD data (magenta line).

Structure solution was conducted using the direct-space genetic algorithm search procedure implemented in the EAGER program [133–135]. The molecular model for this calculation was generated in the WebMO interface and was optimized using the B3LYP hybrid functional and 3-21G basis set in the GAUSSIAN program [56, 59, 60]. The structure solution was defined by a total of 11 structural variables (three positional, three orientational and five torsional). In total, 40 independent GA calculations were conducted in parallel from different initial random populations. Each calculation involved a population size of 100 structures, with 10 mating operations and 50 mutation operations per generation. From the trial structures generated, the trial structure giving the best fit to the experimental powder XRD data (lowest R_{wp}) was selected as the initial structural model for Rietveld refinement.

In the Rietveld refinement, conducted using the GSAS program [105], standard restraints were applied to bond lengths and bond angles, with three overlapping planar restraints applied to the isoalloxazine ring system. A common isotropic displacement parameter was refined for all non-hydrogen atoms, with the value for hydrogen atoms taken as 1.2 times the value for the non-hydrogen atoms. Following initial refinement, the refined structure was subjected to geometry optimization using periodic DFT-D in CASTEP with a fixed unit cell, which generated a structure with a sensible hydrogen-bonding arrangement. This structure was then used as the starting structural model for a subsequent Rietveld refinement, with additional restraints applied to the hydrogen bonding geometries based on the result following DFT-D geometry optimization. A correction for preferred orientation was applied using spherical harmonics [107]. This refinement gave a high-quality fit to the powder XRD data (R_{wp} = 0.78%, R_p = 0.53 %), comparable to the level of

agreement obtained in Le Bail profile fitting, with the following refined unit cell: a = 20.0878(6) Å, b = 15.1260(5) Å, c = 5.33440(13) Å, (V = 1620.85(10) Å^3), as shown in Figure 4.5.

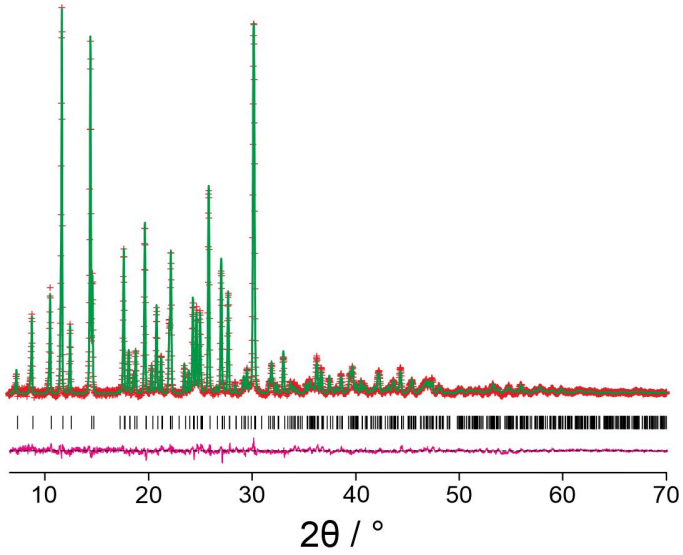

Figure 4.5: Final Rietveld refinement for riboflavin powder XRD data (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Following the final Rietveld refinement, a final DFT-D geometry optimization was conducted using CASTEP. The geometry optimization did not result in any significant change to the structure, as reflected in the low RMSD value (not including hydrogen atoms) of 0.12 Å. In this context, the structure is sufficiently close to an energy minimum to be considered acceptable.

4.2.3 Assessment of the Crystal Structures Determined from Powder XRD and from Micro-crystal XRD

The structure determined directly from powder XRD data (denoted structure 1), and the structure determined directly from micro-crystal XRD data (denoted structure 2) are plausible. Both structures contain hydrogen-bonding arrangements that are structurally and chemically sensible, although differing in the detail of the hydrogen-bonding connectivity. We now assess in more detail which of these structures is the "correct" description, or whether a disordered structural model involving both hydrogen-bonding arrangements may be plausible. However, to simplify this discussion, we conclude both structure 1 and the structure proposed by Guerain *et al* are essentially the same. In this scenario, we can compare the suitability of both the structure determined in this work and the structure proposed by Guerain *et al*. by assessing their fit to the powder XRD data, their relative energies from DFT-D calculations, and their completed solid-state ^{13}C NMR data for comparison to the experimental solid-state ^{13}C NMR data.

The values of DFT-D energy for the micro-crystal XRD structure and that of Guerain *et al*. following geometry optimization (Table 4.1) show clearly that the structure determined in this work has the

lowest energy by 16.5 kJ mol^{-1}. The energy difference is significant and provides strong support that the structure determined by micro-crystal XRD in this work is the correct structure. The significant difference in energy between both crystal structures suggest that no significant degree of disorder involving the hydrogen bonding arrangements in these structures will exist in practice. Indeed, analysis of difference Fourier maps in the structure determination from micro-crystal XRD data provides no evidence for disorder in the hydrogen bonding arrangement.

Structure	This work	Guerain *et al.*
Relative energy / kJ mol^{-1}	0.00	16.48

Table 4.1: DFT-D energy of both the structure from micro-crystal XRD and the structure of Guerain *et al*.

Next, we consider the quality of fit to the powder XRD data of riboflavin. Qualitatively, both structures fit the experimental powder XRD data very closely (Figure 4.6), and in a similar way. However, qualitative comparison using the figures-of-merit R_p and R_{wp}, shows that the micro-crystal XRD structure fits the powder XRD data slightly better (Guerain *et al.*: R_{wp} = 0.78%, R_p = 0.53%; micro-crystal XRD structure: R_{wp} = 0.71%, R_p = 0.50%). Although the structure proposed by Guerain *et al.* was determined directly from structure solution using the powder XRD data, it is clear that the structure determined directly from micro-crystal XRD data gives a better fit to the powder XRD data, which could be a consequence of preferred orientation in the powder sample used to record the data. Preferred orientation can be taken into consideration in Rietveld refinement (in this case through a spherical harmonics approach). This avoids the need for reducing a small degree of preferred orientation by time consuming sample preparation tasks, providing the correction is valid and not over-used.

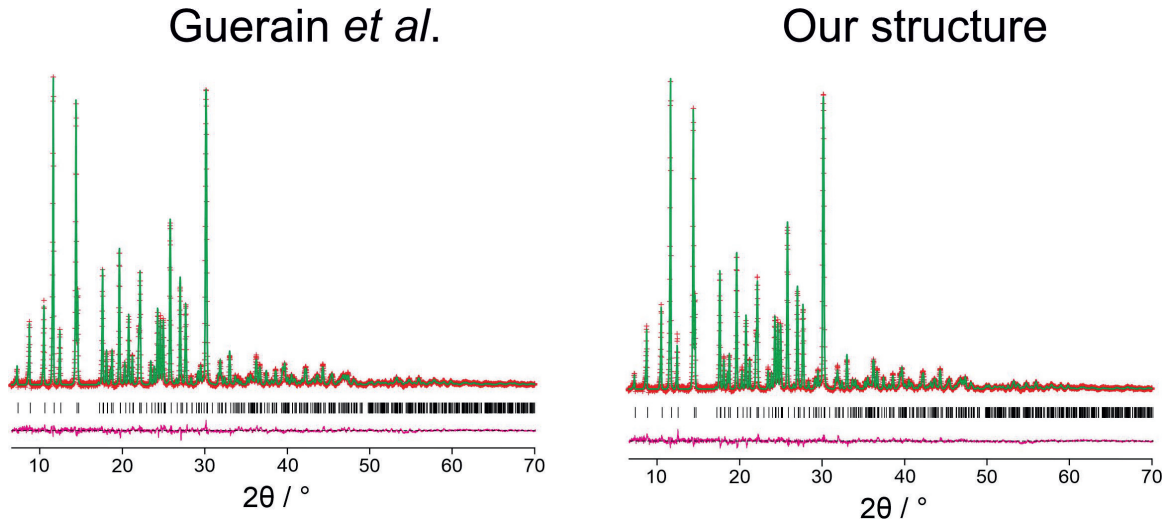

Figure 4.6: The Rietveld fits to the powder XRD data using the structure of Guerain *et al.* and the micro-crystal XRD structure.

Solid-state ^{13}C NMR data have been calculated for structure determined in this work and for the structure of Guerain *et al.* using the DFT-GIPAW methodology and compared to the experimental

high-resolution solid-state ^{13}C NMR spectrum recorded for riboflavin (Figure 4.7). In general, the calculated values of isotropic ^{13}C NMR chemical shifts are similar for the two structures, although the overall match between calculated and experimental ^{13}C NMR chemical shifts is better for the micro-crystal XRD structure (RMSD = 1.96 ppm) than the structure of Guerain *et al*. (RMSD = 2.54 ppm). More significantly, for the terminal $^{13}CH_2OH$ group of the side-chain, the calculated value of the ^{13}C NMR chemical shift (δ_{calc}) is closer to the experimental value (δ_{exp} = 65.20 ppm) for the micro-crystal XRD structure (δ_{calc} = 66.14 ppm; $\delta_{calc} - \delta_{exp}$ = 0.94 ppm) than for the structure of Guerain *et al*. (δ_{calc} = 68.82 ppm; $\delta_{calc} - \delta_{exp}$ = 3.62 ppm). The agreement between calculated and experimental ^{13}C NMR chemical shifts for the terminal $^{13}CH_2OH$ group is significantly better for the micro-crystal XRD structure than for the structure of Guerain *et al*. which lends further support that the structure determined in this work represents the correct hydrogen-bonding arrangement. The difference in the level of agreement between both the calculated and experimental ^{13}C chemical shifts between the two structures is likely a result of the difference in the hydrogen-bonding pattern involving the terminal $^{13}CH_2OH$ group.

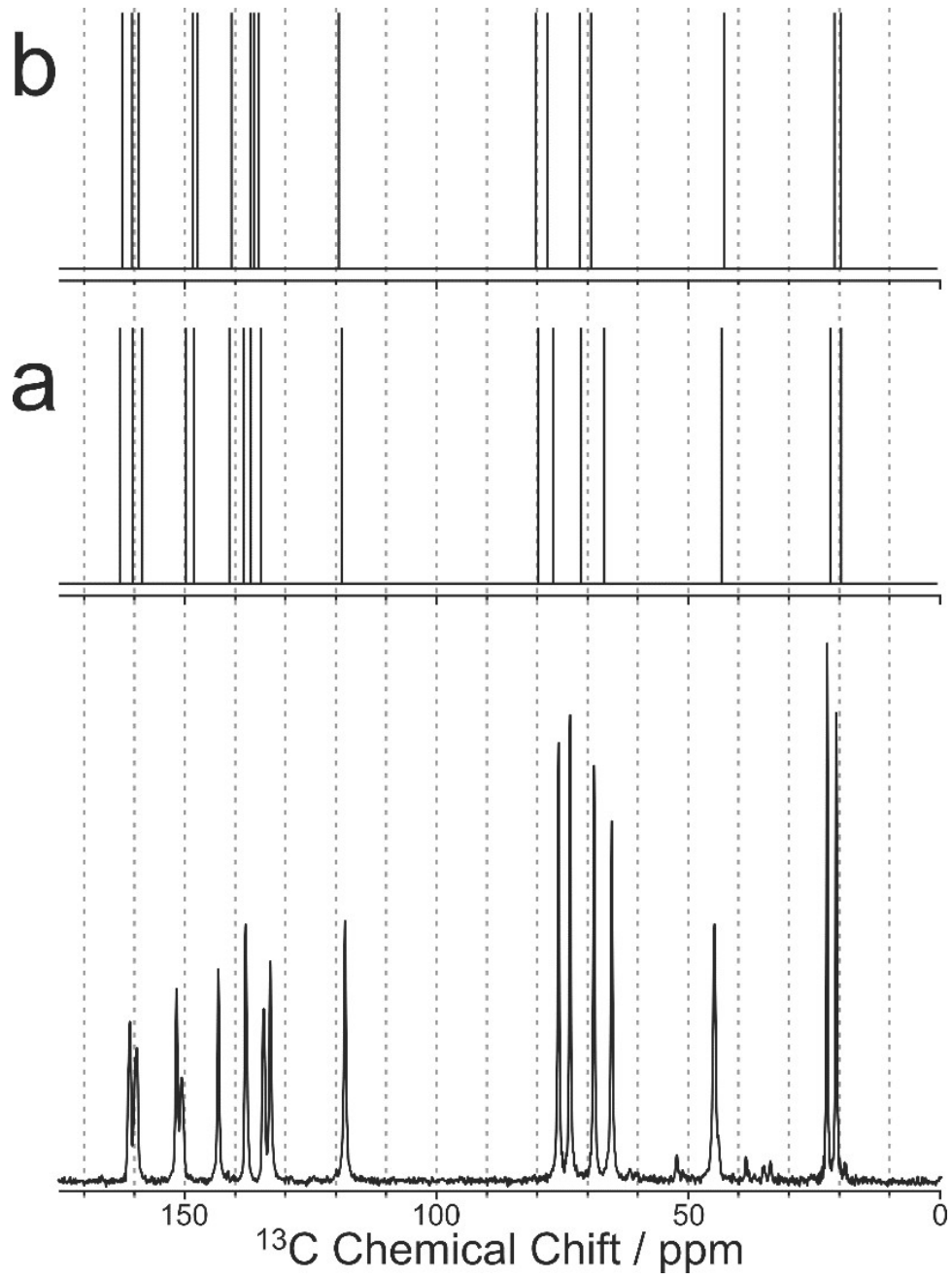

Figure 4.7: The experimental high-resolution solid-state ^{13}C NMR spectrum for riboflavin (bottom), and the values of isotropic ^{13}C NMR chemical shifts calculated using DFT-GIPAW methodology for: (a) the crystal structure from the micro-crystal XRD study at 293 K, and (b) the crystal structure reported by Guerain *et al*. at 293 K.

4.2.4 Structural Discussion

First, we emphasize that the chirality of the side-chain in the enantiomer of riboflavin found in biological systems is S,S,R, for the chiral centers C20, C22, and C24, respectively; these are defined in Figure 4.8. We note that Guerain *et al*. incorrectly define the chirality as {R, R, S}, both in the molecular structure shown on page 801 of their paper, and in specifying the molecule (on page 800 of their paper) as 7,8-dimethyl-10-[(2R,3R,4S)-2,3,4,5-tetrahydroxypentyl]benzo[g]pteridine-2,4(3H,10H)-dione. In spite of these errors, the crystal structure reported by Guerain *et al*. and deposited in the CSD actually contains the correct {S, S, R} chirality of the molecule.

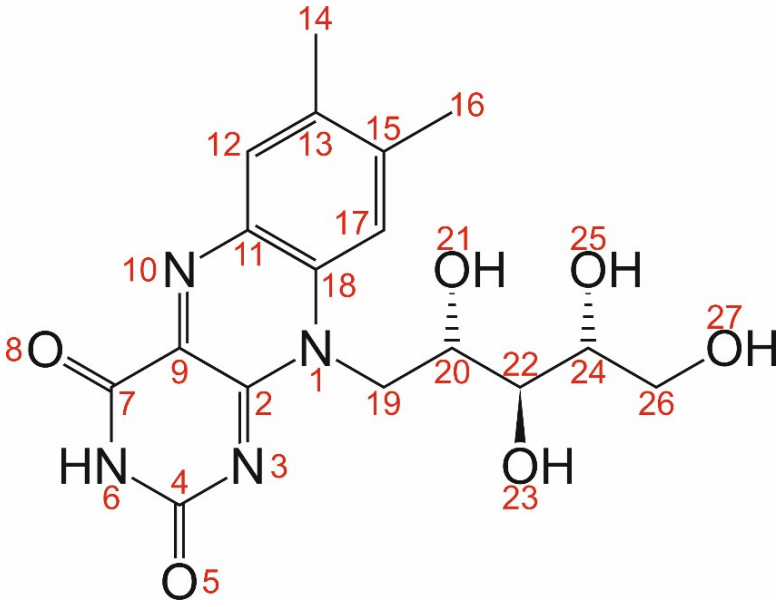

Figure 4.8: The riboflavin molecule with chirality {S, S, R} (for the chiral centers C20, C22, and C24, respectively), showing the atom numbering scheme used throughout this discussion.

To describe the similarities and differences between this structure (as determined from micro-crystal XRD) and the structure of Guerain *et al*., we focus on the results from the micro-crystal XRD studies which, as stated previously, represent the most definitive structure determination among the techniques used in this work. First, we note that the structures determined by micro-crystal XRD at 100 K and 293 K are identical, except for the effects of lattice contraction upon cooling, indicating no solid-phase transitions occur in this temperature range. The comparison to the structure of Guerain *et al*., which was determined at 293 K, is focused on the structure determined from micro-crystal XRD at the same temperature.

As shown in the overlay plot in Figure 4.9, most of the atomic positions are essentially identical in the two crystal structures, particularly concerning the aromatic ring system and the parts of the side-chain closest to the aromatic ring. However, the main difference between the two structures concerns the hydrogen-bonding terminal OH group of the side-chain (containing O27). In the micro-crystal XRD structure, the terminal OH group (O27) is the donor in an intermolecular O—H···O

hydrogen bond with the OH group (O25) of the side-chain as the acceptor. However, in the structure of Guerain *et al.*, the terminal OH group (O27) is the OH donor in an intermolecular O—H···O hydrogen bond with one of the C=O groups (O8) of the aromatic ring of a different neighbouring molecule as the O acceptor. To accommodate the formation of an intermolecular O—H···O hydrogen bond to a different neighbouring molecule, some minor differences arise in the conformation at the end of the side-chain in the two structures. These conformational differences are such that O27 is closer to the O25 neighbour than the O8 neighbour in the micro-crystal XRD structure, whereas O27 is closer to the O8 neighbour than the O25 neighbour in the structure of Guerain *et al.* These structural differences at the end of the side-chains in the two structures are highlighted in Figures 4.10, 4.11 and 4.12.

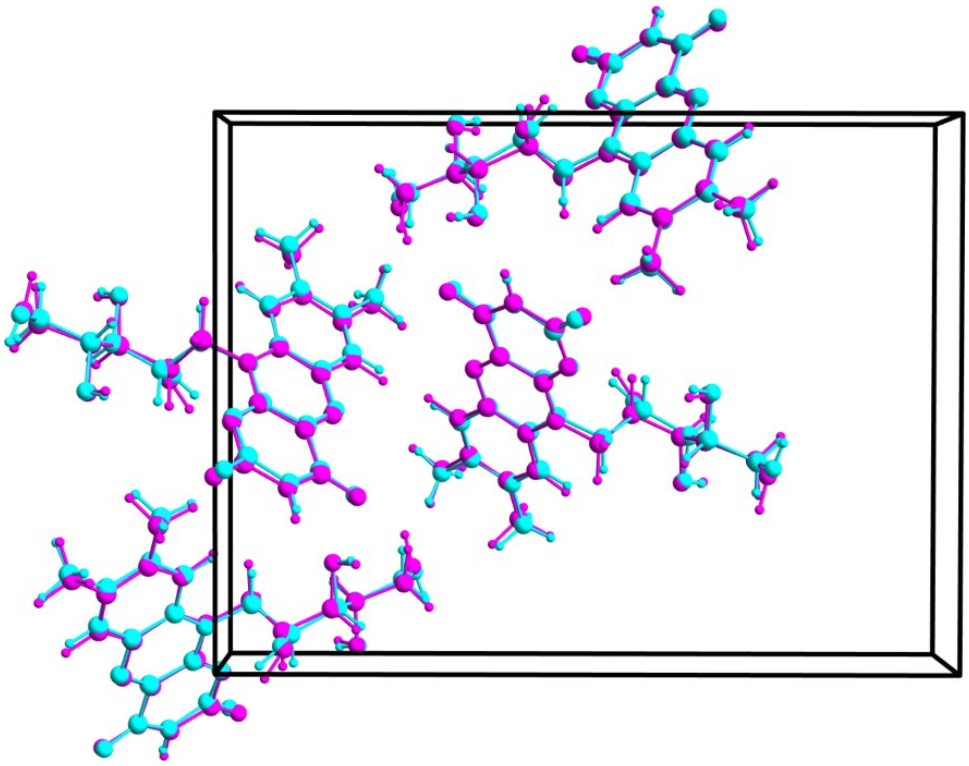

Figure 4.9: Overlay of the crystal structure from the micro-crystal XRD study (cyan) and the crystal structure reported by Guerain *et al.* (magenta), in both cases at 293 K.

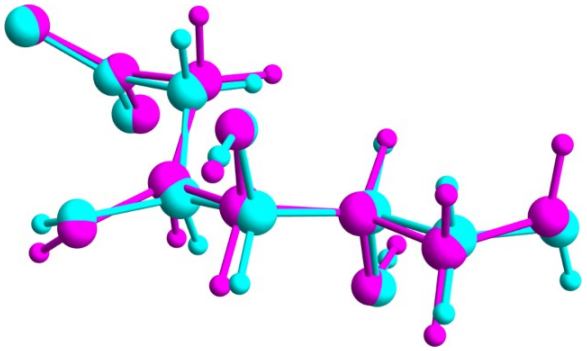

Figure 4.10: Overlay of the side-chain of the riboflavin molecule (also including the C2—N1—C18 portion of the aromatic ring system) in the crystal structure from the micro-crystal XRD study (cyan) and the crystal structure reported by Guerain *et al*. (magenta), in both cases at 293 K.

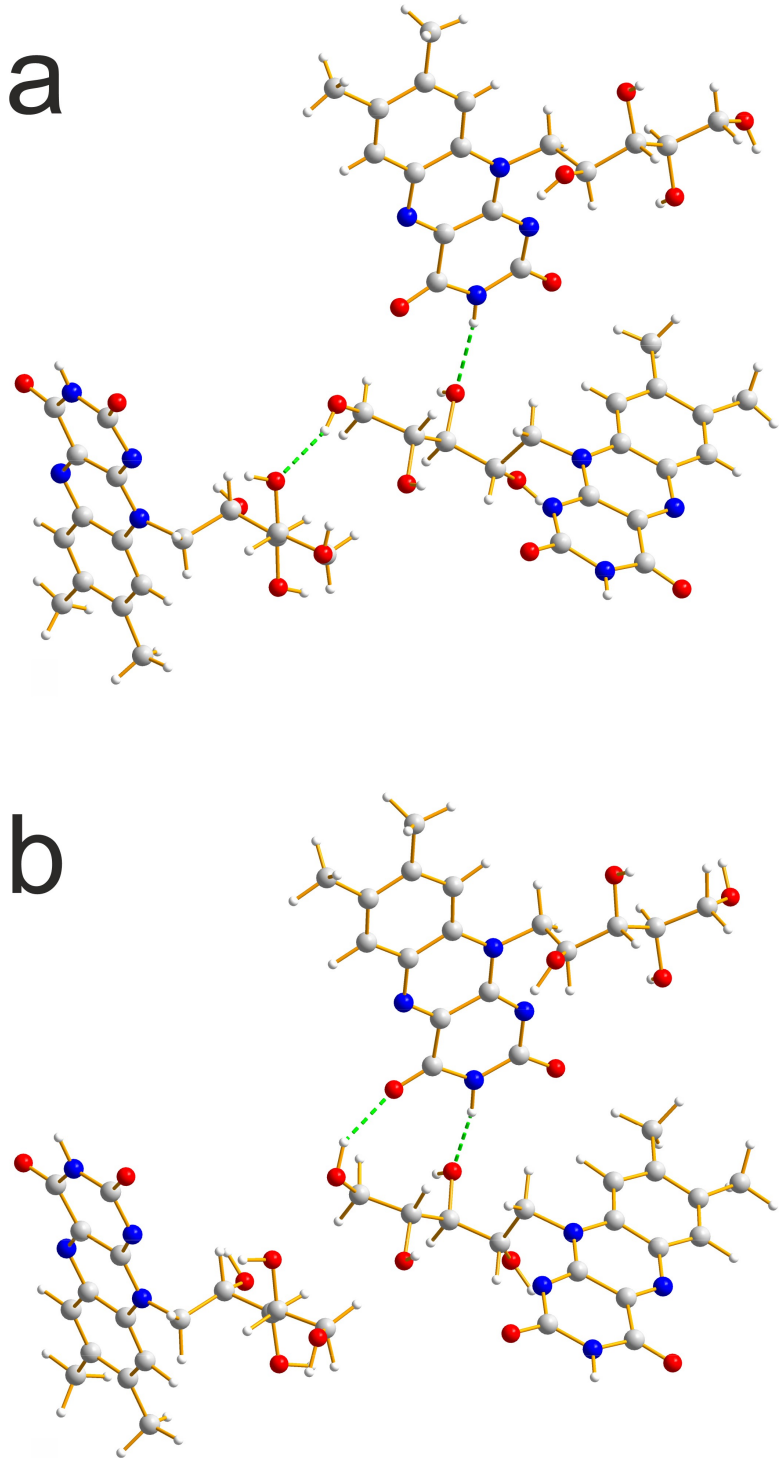

Figure 4.11: The hydrogen-bonding arrangement between a given molecule of riboflavin (the central molecule shown) and two neighbouring molecules in (a): the crystal structure from the micro-crystal XRD study at 293 K, and (b) the crystal structure reported by Guerain *et al.* at 293 K. The terminal OH group of the side-chain is the OH donor in an O—H···O hydrogen bond to a different neighbouring molecule in each structure.

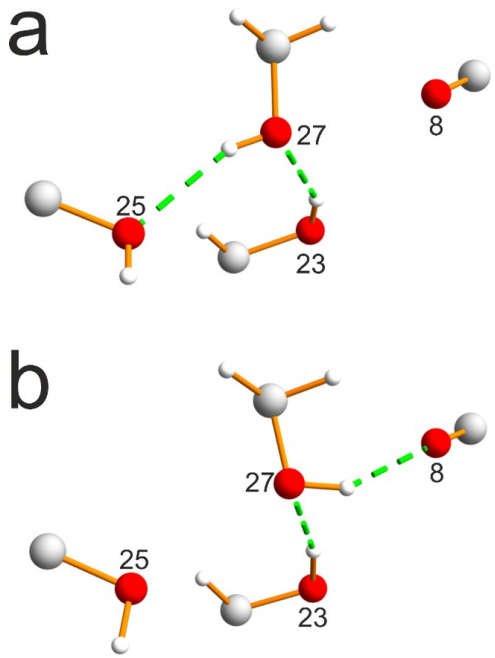

Figure 4.12: The local environment of the terminal OH group (containing O27) of the side-chain of the riboflavin molecule in: (a) the crystal structure from the micro-crystal XRD study at 293 K, and (b) the crystal structure reported by Guerain *et al.* at 293 K. The different hydrogen-bonding arrangement is also manifested by differences in O···O distances, as shown in Table 4.2.

Torsion angle	Structure from this work	Structure of Guerain *et al.*
C2—N1—C19—C20	80.7°	89.4°
N1—C19—C20—C22	171.4°	171.1°
C19—C20—C22—C24	69.4°	60.0°
C20—C22—C24—C26	175.3°	175.8°
C22—C24—C26—C27	175.4°	161.7°
C24—C26—C27—H	79.7°	267.6°
O—O Distance	Structure from this work / Å	Structure of Guerain *et al.* / Å
O27—O25	2.97	3.15
O27—O8	3.11	2.88
O27—O23	2.94	2.95

Table 4.2: Values of torsion angles for the side-chain of the riboflavin molecule and intermolecular O···O distances between the terminal OH group (containing O27) and OH groups of neighbouring molecules in the two crystal structures.

4.3 Conclusion

In conclusion, the crystal structure of riboflavin has been determined using a combination of techniques including powder XRD, solid-state NMR, periodic DFT-D calculations and micro-crystal XRD. Initially, structure 1 was considered as a plausible structure of riboflavin (subsequently also proposed by Guerain *et al*. [108]). However, analysis of micro-crystal XRD data led instead to structure 2. Comparison of structure 1 and structure 2 based on the fit to both powder XRD data and solid-state NMR data, as well as consideration of relative energies by DFT-D calculations, it was concluded that structure 2 is the correct structure of riboflavin. Structure 1 and the structure determined by Guerain *et al.* [108] represents a similar structure but differing in the hydrogen bonding of the terminal OH group of the ribityl side-chain. The reason that these two structure determined from powder XRD data both contain the incorrect hydrogen-bonding arrangement is suspected to be a consequence of the way in which the DFT-D geometry optimization was managed between structure solution and structure refinement stages.

In the structure solution, the hydrogen atom of the terminal OH group (O27) is directed more towards the O8 neighbour than the O25 neighbour, and the application of DFT-D geometry optimization on this structure effectively pulled O27 and O8 together to form an O27—H···O8 hydrogen bond, rather than forming the O27—H···O25 hydrogen bond that was inferred (see discussion above) as the "correct" hydrogen-bonding in the structure solution based on consideration of O···O distances. If the structure following DFT-D geometry optimization was then used as the starting structural model for Rietveld refinement, with sufficiently strong restraints applied to preserve the intermolecular hydrogen-bonding arrangement, then the Rietveld refinement would retain the O27—H···O8 hydrogen bond and would hence correspond to the structure of Guerain *et al*.

It is not known whether the scenario outlined arose in the work of Guerain *et al*. Nevertheless, analysis of this situation highlights a potential danger in applying DFT-D geometry optimization (now a frequent practice) as a bridging step after completing structure solution from powder XRD data and before commencing Rietveld refinement. Given that the location of hydrogen atoms in structure solution from powder XRD data is intrinsically unreliable, hydrogen atoms can be located in arbitrary positions in the structure solution, but it is clear from the above analysis that the locations of the hydrogen atoms in the structure solution can have a critical influence on the structural rearrangements that take place upon subsequent DFT-D geometry optimization of the structure solution.

The structure determination of this commonly prepared phase of riboflavin is promising in facilitating the determination of the structures of other polymorphs of riboflavin. Mohnicke *et al*. [136] prepared a variety of polymorphs and solvate phases of riboflavin with unknown purity, and characterized them using powder XRD, although the data was poorly resolved. If these polymorphs and solvates could be prepared as phase-pure materials, the structure determination procedure developed in the

present work could be exploited for determining the crystal structures of these other polymorphs and solvate phases of riboflavin. The final point regarding riboflavin is to emphasize that, to know the crystalline form of riboflavin in the biological context (in cat and lemur eyes), powder XRD data must be recorded for a biogenic sample, which would reveal whether the biogenic sample represents the same phase as the structure determined in this work a different crystalline solid form (or even an amorphous form) of riboflavin. Regardless, understanding the solid form that exists in biological systems will help to rationalize the role of riboflavin as a functional optical material, either experimentally or via theoretical calculations.

4.4 Alloxazine

Prior to submission of this thesis, the work presented on the crystal structure of alloxazine in this section was published in the journal *Crystal Growth& Design*. The information required to find this article can be found in Appendix D.

4.4.1 Experimental Methods

Sample Preparation

Alloxazine was purchased from Sigma-Aldrich as a fine, yellow powder. The material was recrystallized by heating a sample of alloxazine powder in DMSO at 130°C until fully dissolved, followed by uncontrolled cooling in a fume cupboard to room temperature. Alloxazine swiftly precipitated out from the DMSO solution at room temperature. Powder XRD patterns recorded for the commercial sample and the sample following recrystallization showed these phases were identical.

Powder X-ray Diffraction

High-quality powder XRD data suitable for structure determination were recorded for the recrystallized sample of alloxazine at 21°C on a Bruker D8 Advance Series II diffractometer operating in transmission mode using Ge-monochromated Cu-Kα_1 radiation (2θ range, 4° - 70°; step size, 0.016°; data collection time, 15.5 hr). The recrystallized material was subjected to light manual grinding using a pestle and mortar before loading into a 0.7 mm glass capillary which was then flame sealed.

Solid-state NMR Spectroscopy

The high-resolution solid-state ^{13}C NMR spectrum for the recrystallized sample of alloxazine was recorded on a Bruker AVANCE III spectrometer (magnetic field, 9.4 T; ^{13}C Larmor frequency, 100.64 MHz), with the powder sample packed in a zirconia solid-state NMR rotor. The solid-state ^{13}C NMR spectrum was recorded at ambient temperature using ramped ^{1}H $\rightarrow$ ^{13}C cross-polarization (CP) [120], magic-angle-spinning (MAS frequency, 10 kHz) and ^{1}H decoupling using the SPINAL-64 sequence [120]. The solid-state ^{13}C NMR spectrum was referenced [121] against the α polymorph of glycine, with the carboxylate resonance set to 176.5 ppm, corresponding to tetramethylsilane (TMS) as the primary reference.

The high-resolution solid-state NMR experiments performed at the U.K. High-Field (850 MHz) Solid-State NMR Facility were performed by Dr Dinu Iuga at the University of Warwick. The high-resolution solid-state ^{15}N spectrum of the recrystallized sample of alloxazine was recorded at 20.0 T on a Bruker AVANCE III spectrometer at the U.K. High-Field (850 MHz) Solid-State NMR Facility (^{15}N Larmor frequency, 86.16 MHz; ^{1}H Larmor frequency, 850.24 MHz) using a 4 mm HY MAS

probe. The solid-state ^{15}N NMR spectrum was recorded at ambient temperature using ramped ^{1}H $\rightarrow$ ^{15}N cross-polarization (CP contact time, 2 ms), magic-angle spinning (MAS frequency, 8 kHz) and high-power ^{1}H decoupling. The ^{15}N referencing was based on the method of Bertani *et al*. [137], using the isotropic ^{15}N NMR chemical shift of the δ-^{15}N site in histidine hydrochloride monohydrate at 189.5 ppm.

Periodic DFT-D Calculations

Periodic DFT-D calculations for geometry optimization were conducted (with fixed unit cell) using CASTEP [122] (Academic Release version 8.0). These calculations utilized ultrasoft pseudopotentials [123], PBE functional [124], semi-empirical dispersion corrections using the TS correction scheme [125], conserved space group symmetry, fixed unit cell, periodic boundary conditions, a basis set cut-off energy of 700 eV and a Monkhorst-Pack grid [126] of minimum sample spacing (0.05 x 2π) Å^{-1}. The convergence criteria for geometry optimization were 0.01 eV Å^{-1} for atomic forces, 0.00001 eV per atom in the total energy, and 0.001 Å for atomic displacements.

Solid-state NMR chemical shifts were calculated using CASTEP (Academic Release version 8.0) for the crystal structures containing the alloxazine and isoalloxazine tautomers determined directly from powder XRD data, followed by DFT-D geometry optimization (with fixed unit cell) as described above. The Gauge Including Projector Augmented Wave (GIPAW) approach [127–131] was used with a cut-off energy of 700 eV and PBE functional. From the isotropic ^{13}C NMR shielding value (σ_{calc}) calculated for each ^{13}C environment in the crystal structure, the corresponding calculated isotropic ^{13}C NMR chemical shift (δ_{calc}) was determined [129] from the equation δ_{calc} = $<\delta_{exp}>$ + $<\sigma_{calc}>$ $-$ σ_{calc}, where $<\delta_{exp}>$ denotes the mean of the isotropic ^{13}C NMR chemical shifts determined from the experimental high-resolution solid-state ^{13}C NMR spectrum ($<\delta_{exp}>$ = 140.45 ppm) and $<\sigma_{calc}>$ denotes the mean of the calculated isotropic ^{13}C NMR shielding values ($<\sigma_{calc}>$ = 30.18 ppm for alloxazine, and 30.71 ppm for isoalloxazine).

From the isotropic ^{15}NMR shielding value (σ_{calc}) calculated for each ^{15}N environment in the crystal structure, the corresponding calculated isotropic ^{15}N NMR chemical shift (δ_{calc}) was determined from the equation δ_{calc} = δ_{exp}(ref) + σ_{calc}(ref) $-$ σ_{calc}, where δ_{exp}(ref) denotes the experimental value of the isotropic ^{15}N NMR chemical shift [δ_{exp}(ref) = 189.5 ppm] for the reference material (the δ-^{15}N site in histidine hydrochloride monohydrate) and σ_{calc}(ref) denotes the isotropic ^{15}N NMR shielding value [σ_{calc}(ref) = 29.2 ppm] for the δ-^{15}N site in histidine hydrochloride monohydrate calculated for the known crystal structure of this material [138] using the same DFT-D methodology as that used to calculate the ^{15}N shielding values for the alloxazine and isoalloxazine crystal structures.

To determine the relatively energies of the crystal structures containing the alloxazine and isoalloxazine tautomers and to assess the energy barriers to convert between these crystal structures

high-level DFT-D calculations were conducted using FHI-aims [139] (date stamp: 191029) coupled with the Atomic Simulation Environment (ASE) [140] and Catlearn Python package [141]. These calculations were performed by Dr Andrew Logsdail at Cardiff University. Geometry optimization was conducted using the BFGS algorithm implemented in FHI-aims, with a convergence criterion of 0.01 eV $Å^{-1}$ for atomic forces. Transition states were determined using the machine-learning nudged-elastic band (MLNEB) algorithm provided by Catlearn [58], with a convergence criterion of 0.05 eV $Å^{-1}$ for atomic forces. In both types of calculation, the PBE version [124] of the general gradient approximation (GGA) was coupled with the Tkatchenko-Scheffler (TS) dispersion correction [125]. Following convergence of the geometry optimization procedures, single-point energy calculations were conducted using both GGA and hybrid-GGA exchange-correlation functionals, specifically PBE [124] and PBE0 [142], coupled with either the TS method [125] or the many-body dispersion method for dispersion correction [116] (MBD). Thus, the complete set of exchange-correlation functionals considered in this work were: PBE-TS, PBE-MBD, PBE0-TS and PBE0-MBD. The electronic structure calculations were conducted with an "intermediate" basis set and relativistic effects were included via the scaled zeroth order regular approximation [124]. A Γ-centered **k**-grid was used with a minimum sample spacing of (0.05 x 2π) $Å^{-1}$ (testing with a denser **k**-grid sampling of (0.04 x 2π) $Å^{-1}$ gave changes in relative energies of $<$1 meV). The electronic structure self-consistent field (SCF) cycle was considered to be converged when changes in the electron density, the total energy, and the sum of the eigenvalue energies were below 10^{-6} e $a_0{}^{-3}$, 10^{-6} eV, and 10^{-6} eV, respectively.

4.4.2 Crystal Structure Determination

The high-resolution powder XRD data for alloxazine were indexed using the program LZON in the CRYSFIRE package [89, 143], giving the following unit cell with triclinic metric symmetry: a = 5.86 Å, b = 7.61 Å, c = 10.16 Å, α = 69.9°, β = 83.5°, γ = 83.1° (V = 421.2 $Å^3$). Unit cell and profile refinement were conducted using the Le Bail method in the GSAS program [96, 132], giving a good quality of fit to the experimental powder XRD data (R_{wp} = 1.07%, R_p = 0.80%), shown in Figure 4.13. Based on the measured unit cell volume and consideration of density, it was deduced that there are two molecules of alloxazine in the unit cell. Thus the asymmetric unit would comprise one molecule (Z' = 1) for space group $P\bar{1}$ or two molecules (Z' = 2) for space group $P1$.

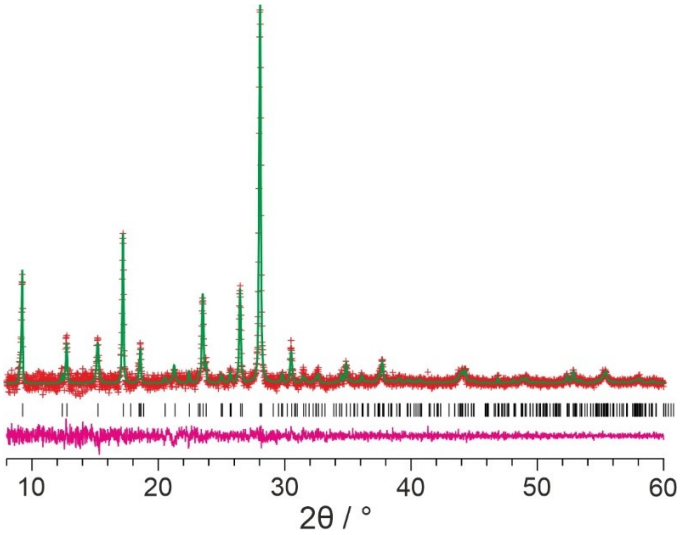

Figure 4.13: Profile fitting (using the Le Bail method) of the powder XRD data (background subtracted) recorded for alloxazine, showing the experimental powder XRD data (red crosses), calculated powder XRD data (green line), predicted peak positions (black ticks) and difference between the experimental and calculated powder XRD data (magenta line).

Structure solution was conducted using the direct-space genetic algorithm (GA) technique incorporated in the program EAGER [133–135]. Structure solution was conducted initially for space group $P\bar{1}$ ($Z' = 1$), using a molecular model for alloxazine constructed using standard bond lengths and bond angles from the Cambridge Structural Database using the MOGUL program [144], with bond lengths involving hydrogen atoms taken from Allen *et al*. [145] As there is one molecule in the asymmetric unit for the $P\bar{1}$ space group, each trial structure was defined by six structural variables (six positional and three orientational variables). In total, 40 independent GA structure-solution calculations were conducted, each starting from a different randomly generated initial population. Each GA calculation involved the evolution of a population of 100 trial structures for 100 generations, with 10 mating operations per generation and 50 mutation operations per generation. All 40 independent calculations produced the same trial structure giving the best fit the experimental powder XRD data (i.e. lowest R_{wp}). This structure was considered reasonable on the basis of structural and chemical criteria and was used as the starting model for Rietveld refinement [105].

Structure solution was also conducted for space group $P1$, with $Z' = 2$. In this case, each trial structure was defined by nine variables (three positional and six orientational variables; for $P1$, the position of one molecule in the unit cell can be arbitrarily fixed). The trial structure giving the best fit to the experimental powder XRD data was essentially identical to the structure obtained for space group $P\bar{1}$, confirming that the description of the structure in the space group $P\bar{1}$ is correct.

The best trial structure obtained in the GA structure solution calculations for space group $P\bar{1}$ was used as the initial structural model for Rietveld refinement, which was conducted using the GSAS program. In the Rietveld refinement, standard restraints were applied to bond lengths and bond

angles, with planar restraints applied to the tricyclic ring system. A common isotropic displacement parameter was refined for all non-hydrogen atoms, with the value for hydrogen atoms set to 1.2 times the refined value for the non-hydrogen atoms. The Rietveld refinement gave an excellent fit to the experimental powder XRD data (R_{wp} = 1.15%, R_p = 0.86%) (seen in Figure 4.14), comparable to the quality of fit obtained in the Le Bail profile fitting, with the following final refined unit cell parameters: a = 5.8638(6) Å, b = 7.6066(6) Å, c = 10.1644(9) Å, α = 69.972(7)°, β = 83.493(8)°, γ = 83.061(6)°, (V = 421.57(9) Å^3).

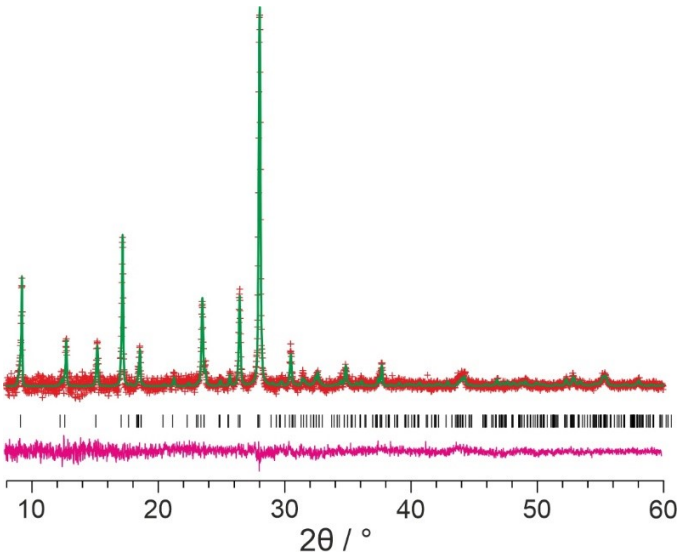

Figure 4.14: Rietveld fit of the powder XRD data (background subtracted) recorded for alloxazine, refined using the alloxazine crystal structure, showing the experimental powder XRD data (red crosses), calculated powder XRD data (green line), predicted peak positions (black ticks) and difference between the experimental and calculated powder XRD data (magenta line).

For structure validation, the final refined crystal structure was subjected to geometry optimization (with fixed unit cell) using periodic DFT-D calculations. The geometry optimized structure showed only minor differences in atomic positions compared to the structure obtained in the Rietveld refinement, as shown in Figure 4.15 (RMSD in the positions of non-hydrogen atoms = 0.082 Å), confirming that the final refined crystal structure is close to a minimum on the energy landscape.

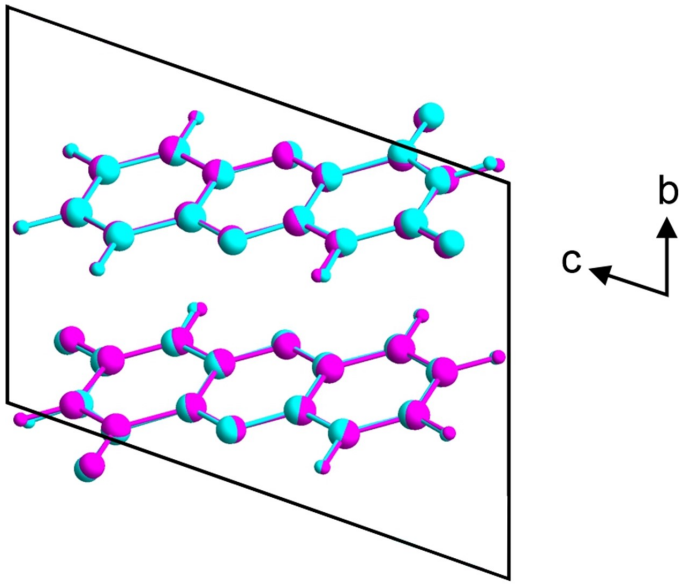

Figure 4.15: Overlays of the crystal structure of alloxazine from the final Rietveld refinement (magenta) and following DFT-D geometry optimization of this structure with fixed unit cell (cyan).

4.4.3 Structural Discussion

The crystal structure of alloxazine comprises layers of molecules connected by three different hydrogen-bonding motifs (Figure 4.16) between a given molecule and three neighbouring molecules (Figure 4.16 shows the intermolecular interactions between reference molecule A and three neighbouring molecules; B, C and D). The hydrogen bonding to one neighbour, molecule B, is a self-complementary motif involving two N—H···N interactions (N···N, 3.110(3) Å, N—H···N, 173.7(3)°) and two C—H···O interactions (C···O, 3.175(7) Å; C—H···O, 170.6(3)°). The two molecules involved in this interaction motif lie in the same plane and are related through an inversion center located in the same plane; therefore, the two N—H···N interactions are symmetry equivalent, and the two C—H···O interactions are also symmetry equivalent.

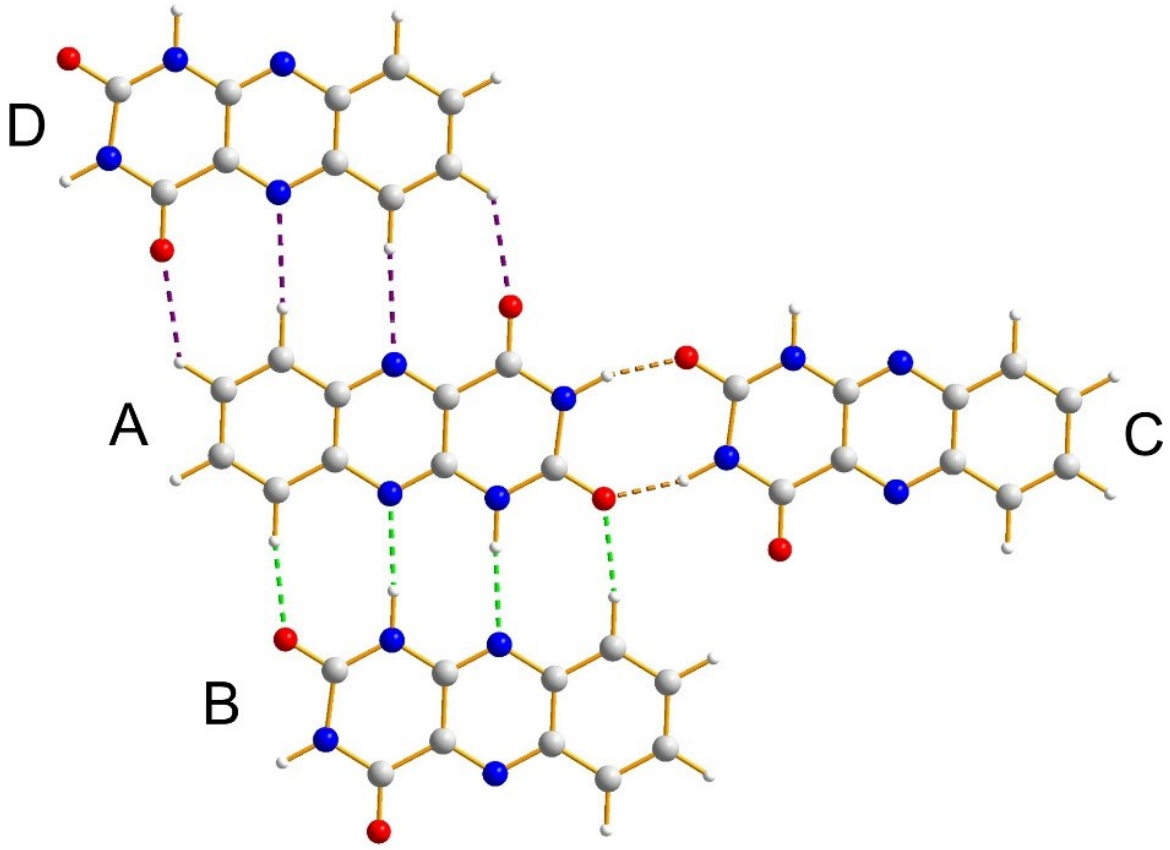

Figure 4.16: A single hydrogen-bonded layer in the crystal structure of alloxazine viewed perpendicular to the plane of the two-dimensional hydrogen-bonded sheet. Hydrogen bonds are indicated by dashed lines with the following colour code: A-B interaction motif, green; A-C interaction motif, gold; A-D interaction motif, purple.

The intermolecular interaction to another neighbouring molecule in the same layer (denoted D in Figure 4.16) comprises two C—H···O interactions and two C—H···N interactions (shown as purple dashed lines in Figure 4.16). The two molecules involved in this interaction motif are very close to co-planar and are related across an inversion center; therefore the two C—H···O interactions are symmetry equivalent and the two C—H···N interactions are symmetry equivalent. Based on geometric criteria, these C—H···O interactions (C···O, 3.235(8) Å, C—H···O, 131.0(7)°) and C—H···N interactions (C···N, 3.543(6) Å, C—H···N, 173.3(7)°) are on the borderline of being classified as C—H···O and C—H···N hydrogen bonds. For this reason, this pair of unique interactions are noticeably less significant than the A-B interaction motif described previously. The combination of the A-B and A-D interaction motifs form a one-dimensional ribbon of molecules along the $[2\bar{1}0]$ direction, with all molecules in the ribbon lying close to the same plane.

The hydrogen bonding to another neighbouring molecule (the A-C interaction motif, indicated by gold dashed lines) is a cyclic array, with an $R_2^2(8)$ graph set classification [146], constructed from two N—H···O hydrogen bonds (N···O, 2.857(5) Å, N—H···O, 147.6(6)°). This interaction motif

links adjacent ribbons within the hydrogen-bonded layer. The two molecules in this interaction motif are related across an inversion centre; however, while the planes of the two molecules are parallel, they are not co-planar, as the positions of the two molecules are slightly offset in the direction perpendicular to the molecular planes (therefore, the inversion center lies mid-way between the planes of the two molecules). As a consequence, the hydrogen-bonded layers are “step-like” as opposed to flat, with each step corresponding to the hydrogen-bonded “interface” (comprising A-C interactions) between adjacent ribbons. In Figure 4.16, the stacking of adjacent hydrogen-bonded layers is viewed parallel to the molecular plane, clearly showing the step-like character of each layer. The perpendicular distance between adjacent layers [taking the mean plane of the stepped layers as (122)] is 3.17 Å. The perpendicular distance between the mean planes of the molecular ribbons in adjacent layers is 3.26 Å, as shown in Figure 4.17. The crystal structure of alloxazine as determined in this work from powder XRD data can be accessed through the accession code: CCDC 2111430 (CSD Refcode: XARROB).

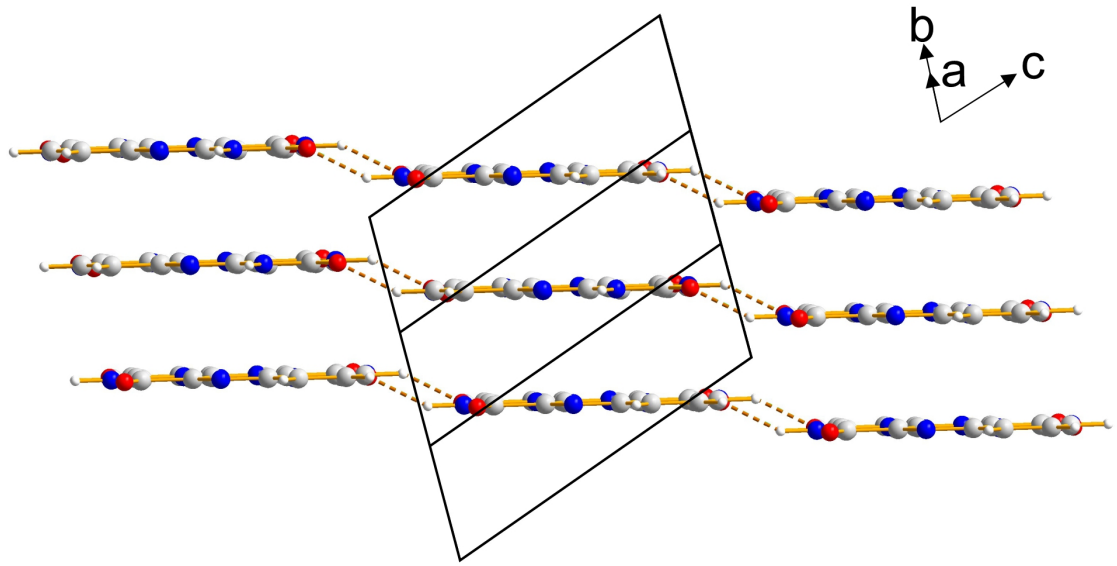

Figure 4.17: Crystal structure of alloxazine, viewed parallel to the molecular planes, showing three layers of molecules constructed by hydrogen-bonded ribbons. As shown, the molecules in the layers are not completely co-planar, as the planes of neighbouring molecules linked by two N—H···O hydrogen bonds (gold dashed lines) are offset slightly in the direction perpendicular to the molecular plane, giving rise to the stepped arrangement shown.

Assessment of Alloxazine/Isoalloxazine Tautomerism

As the alloxazine and isoalloxazine tautomers differ only in the position of one hydrogen atom (with only small differences in the remainder of the molecular geometry resulting from the different position of the hydrogen atom), the isoalloxazine tautomer should be able to form an essentially identical crystal structure to alloxazine, differing only in the positions of the hydrogen atoms in each intermolecular N—H···N hydrogen bond (N—H···N verses N···H—N) that links pairs of

neighbouring molecules in each 2D sheet. As the crystal structures containing alloxazine and isoalloxazine tautomers would differ only in the position of one hydrogen atom in the asymmetric unit, the powder XRD patterns of the two structures would be virtually indistinguishable.

To explore this issue in more detail, the molecule in the alloxazine crystal structure was converted to the isoalloxazine tautomer and then subjected to Rietveld refinement against the same experimental powder XRD dataset. The Rietveld refinement gave a good fit to the powder XRD data (R_{wp} = 1.18%, R_p = 0.88%), which is very similar in the quality of fit to the refinement using the alloxazine tautomer, although with marginally greater differences between the experimental and calculated powder XRD data (Figure 4.18), as reflected in the slightly higher R_{wp} and R_p values. The fixed unit cell parameters from this Rietveld refinement (a = 5.8633(7) Å, b = 7.6064(6) Å, c = 10.1637(10) Å, α = 69.976(7)°, β = 83.497(9)°, γ = 83.063(6)° (V = 421.51(9) Å^3) are essentially the same, within estimated standard deviations, as those obtained in the Rietveld refinement for alloxazine.

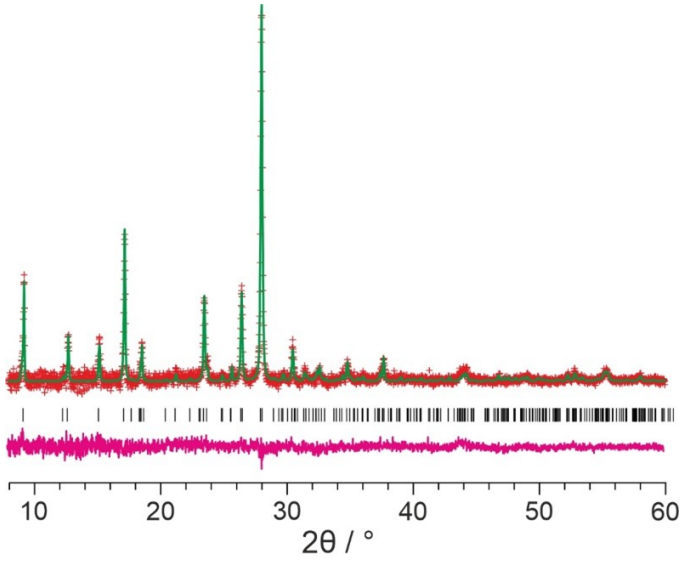

Figure 4.18: Rietveld fit of the powder XRD data (background subtracted) recorded for alloxazine, but refined using the isoalloxazine crystal structure, showing the experimental powder XRD data (red crosses), calculated powder XRD data (green line), predicted peak positions (black ticks) and difference between the experimental and calculated powder XRD data (magenta line).

The relative energetic stabilities of the crystal structures containing alloxazine and isoalloxazine were assessed using periodic DFT-D calculations. Initially the crystal structure for each tautomer obtained from the final Rietveld refinement was subjected to geometry optimization (with fixed unit cell) using the PBE-TS method. In the resulting geometry-optimized structures, the N—H···N hydrogen bond distances are significantly shorter for the alloxazine structure (N—N: alloxazine, 3.09 Å; isoalloxazine, 2.99 Å). For each geometry-optimized structure, single-point energy calculations were conducted using different functionals (PBE and PBE0), along with different dispersion correction methods (TS and MBD). For each calculation method, the calculated energy is higher for the isoalloxazine structure than the alloxazine structure (Table 4.3).

Calculation type	PBE-TS	PBE-MBD	PBE0-TS	PBE0-MBD
Relative energy / kJ mol^{-1}	17.8	17.2	23.8	23.3

Table 4.3: The relative energy of the isoalloxazine structure compared to the alloxazine structure, for the four types of DFT-D calculation.

The magnitude of these energy differences suggests that the crystal structure containing the alloxazine tautomer will be strongly favoured in the experimental situation. Calculation of the relative Boltzmann factor for the two tautomeric states indicates that, under conditions of thermal equilibrium at 298 K, more than 99.9% of the molecules in the crystal structure would exist as the alloxazine tautomer. The significant difference in energy between the alloxazine and isoalloxazine crystal structures is dominated by the difference in intramolecular energy between the alloxazine and isoalloxazine tautomers, which favours alloxazine by 54.9 kJ mol^{-1} (PBE0-MBD), rather than differences in the energies of intermolecular interactions within the crystal structures, which favours isoalloxazine by 30.7 kJ mol^{-1} (PBE0-MBD). Thus, the intermolecular cohesion is actually weaker in the crystal structure containing the alloxazine tautomer. To seek direct experimental verification of the tautomeric form present in the crystal structure, high-resolution solid-state ^{15}N NMR and ^{13}C NMR data were recorded for the same sample of alloxazine used in the powder XRD study.

First, we consider the solid-state ^{1}H→^{13}C CP-MAS NMR, with chemical shifts also calculated for the crystal structures of alloxazine and isoalloxazine determined in this work (Table A.1). The experimental solid-state ^{13}C NMR spectrum shows nine resolved isotropic peaks, which are compared to the isotropic ^{13}C NMR chemical shifts calculated by DFT-D calculations on the crystal structure, containing the alloxazine and isoalloxazine tautomers (Figure 4.19). Although the overall level of agreement between experimental and calculated data is better for the alloxazine structure than the isoalloxazine structure, the quality of agreement is not sufficiently high to merit a definitive assignment of tautomeric form. At this point, solid-state ^{1}H→^{15}N CP-MAS NMR data were recorded to further assess the assignment of the tautomeric form in the crystal structure.

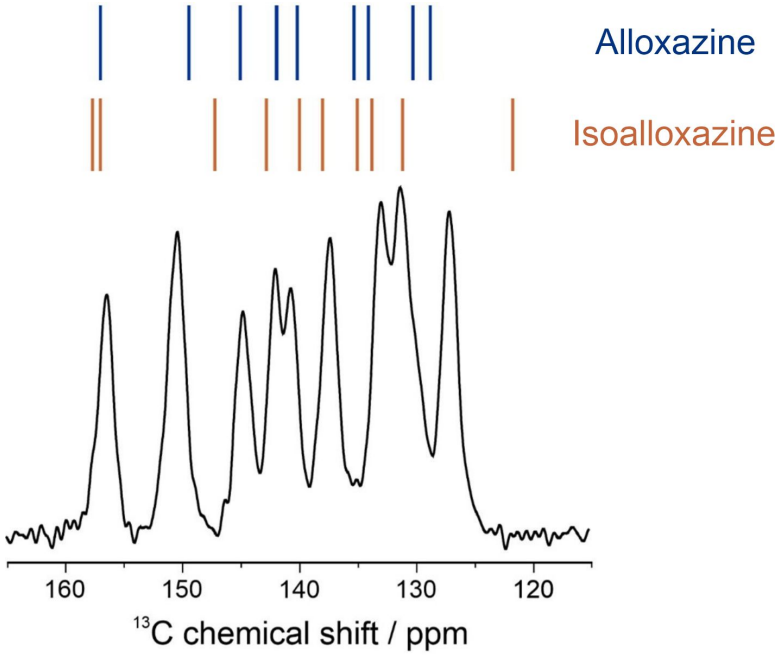

Figure 4.19: The experimental solid-state ^{13}C NMR spectrum recorded for the sample of alloxazine, together with stick plots which correspond to the calculated ^{13}C NMR chemical shifts for the crystal structures containing alloxazine (blue) and isoalloxazine (orange).

The high-resolution solid-state ^{1}H→^{15}N CP-MAS NMR spectrum contains isotropic peaks at 123.81 ppm and 156.56 ppm (chemical shift difference, 32.75 ppm), shown in Figure 4.20. Although there are four distinct nitrogen environments in the structure, with one molecule in the asymmetric unit, the conditions of the ^{1}H→^{15}N CP-MAS NMR measurement (^{1}H → ^{15}N cross-polarization with contact time τ_{cp} = 2 ms) are such that signals of significant intensity are only observed for ^{15}N nuclei directly bonded to ^{1}H nuclei. Interpretation of the experimental solid-state ^{15}N NMR spectrum is greatly facilitated by the capability to calculate the isotropic ^{15}N chemical shifts for the crystal structures containing alloxazine and isoalloxazine tautomers determined from the powder XRD data, giving resonances at 126.68 and 161.43 ppm (chemical shift difference, 34.75 ppm) for the two N—H environments in the alloxazine crystal structure, and at 149.94 and 165.64 ppm (chemical shift difference, 15.70 ppm) for the two N—H environments in the isoalloxazine crystal structure (seen in Table 4.4). The corresponding ^{15}N chemical environments are shown in Figure 4.21. The match between the calculated and experimental values of the isotropic ^{15}N chemical shifts is significantly better for the crystal structure containing alloxazine than the crystal structure containing isoalloxazine. Correspondingly, the chemical shift difference between the two isotropic peaks in the experimental solid-state NMR data (32.75 ppm) is in far better agreement with the calculated difference for alloxazine (34.75 ppm) than the calculated difference for isoalloxazine (15.70 ppm). These observations give strong support to the conclusion that the material studied in this work contains the alloxazine tautomer rather than the isoalloxazine tautomer.

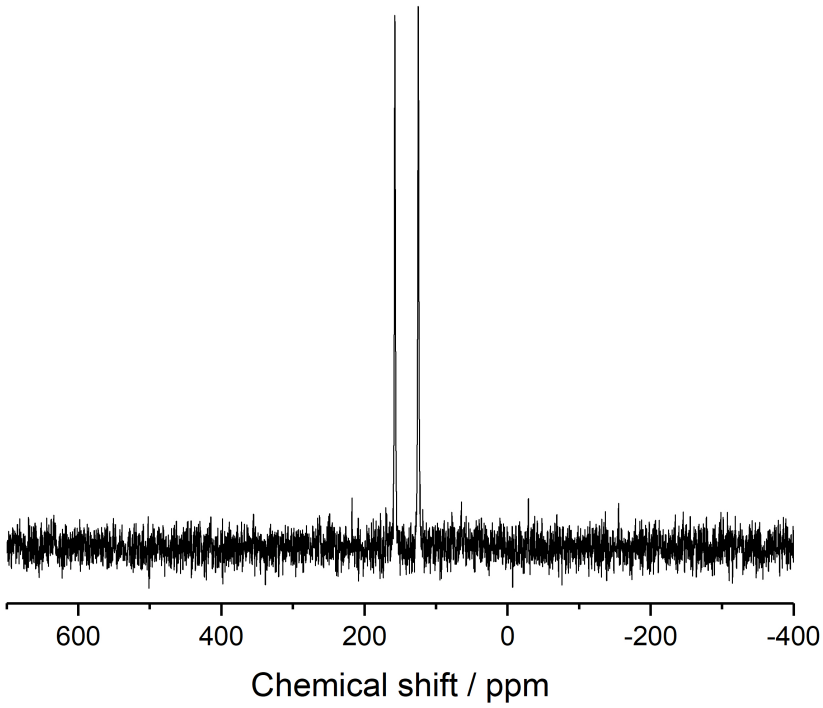

Figure 4.20: The high-resolution solid-state ^{15}N NMR spectrum recorded for alloxazine.

^{15}N site	δ_{calc} Alloxazine / ppm	δ_{exp} / ppm	δ_{calc} Isoalloxazine / ppm
1	255.37	-	**149.94 (NH)**
2	**126.68 (NH)**	123.81 (NH)	182.85
3	**161.43 (NH)**	156.56 (NH)	**165.64 (NH)**
4	337.01	-	349.36

Table 4.4: Isotropic ^{15}N NMR chemical shifts measured from the experimental solid-state ^{15}N NMR spectrum (δ_{exp}) and those determined from DFT-D calculations for the crystal structures containing the alloxazine tautomer (δ_{calc} Alloxazine) as well as the isoalloxazine tautomer (δ_{calc} Isoalloxazine). The isotropic ^{15}N NMR chemical shifts calculated for the NH environments in each crystal structure are highlighted in bold. The numbering of the ^{15}N sites in each tautomer is defined in Figure 4.21.

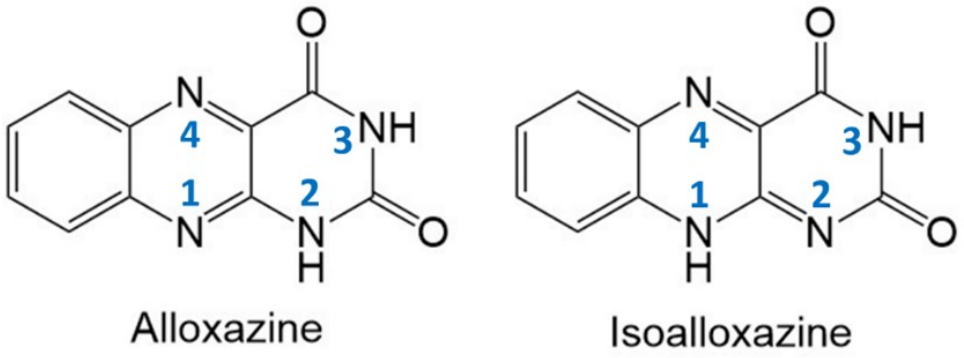

Figure 4.21: The numbering of the ^{15}N sites in both alloxazine and isoalloxazine.

Computational Study of Tautomeric Interconversion in the Crystal Structure of Alloxazine

The work in this section, and descriptions detailed in this section, were contributed by Dr Andrew Logsdail at Cardiff University as part of the publication of this work detailed in Appendix D.

Another intriguing issue relating to the crystal structure of alloxazine concerns the feasibility of interconversion between the alloxazine and isoalloxazine tautomers in the crystal structure, which is most likely to occur by proton transfer along the two N—H···N hydrogen bonds which link pairs of neighbouring molecules. To explore the energetic and mechanistic aspects of this tautomeric interconversion process, transition state calculations were conducted using a nudged elastic band (NEB) method, in which the reaction path between the initial and final structures is explored using a "chain" of intermediate structures. These calculations were performed, and the data processed, by Dr Andrew Logsdail at Cardiff University. In this work, the initial structure was the crystal structure containing the isoalloxazine tautomer, and the final structure was the crystal structure containing the alloxazine tautomer. Each structure within this "chain" of intermediates was subjected to periodic DFT-D geometry optimization, whilst maintaining equal "spacing" to the adjacent structures on the potential energy landscape, which is achieved using a coupling spring constant. A surrogate machine-learned atomistic landscape was used to accelerate convergence (MLNEB), with the efficiency of the transition state search favourably decoupled from the length of the chain representing the reaction path [147].

By analogy to proton dynamics in similar hydrogen bonded systems involving double proton transfer processes within a pair of hydrogen bonds (e.g. the carboxylic acid "dimer" [148–150]), the interconversion between the alloxazine and isoalloxazine tautomers may be envisaged to proceed via a concerted mechanism in which the protons in the two distinct N—H···N hydrogen bonds linking adjacent molecules simultaneously jump to the other molecule. This mechanism converts both molecules from one tautomer to the other tautomer via a transition state that contains two symmetrical N···H···N hydrogen bonds. This mechanism for tautomeric interconversion as the concerted mechanism, depicted in Figure 4.22.

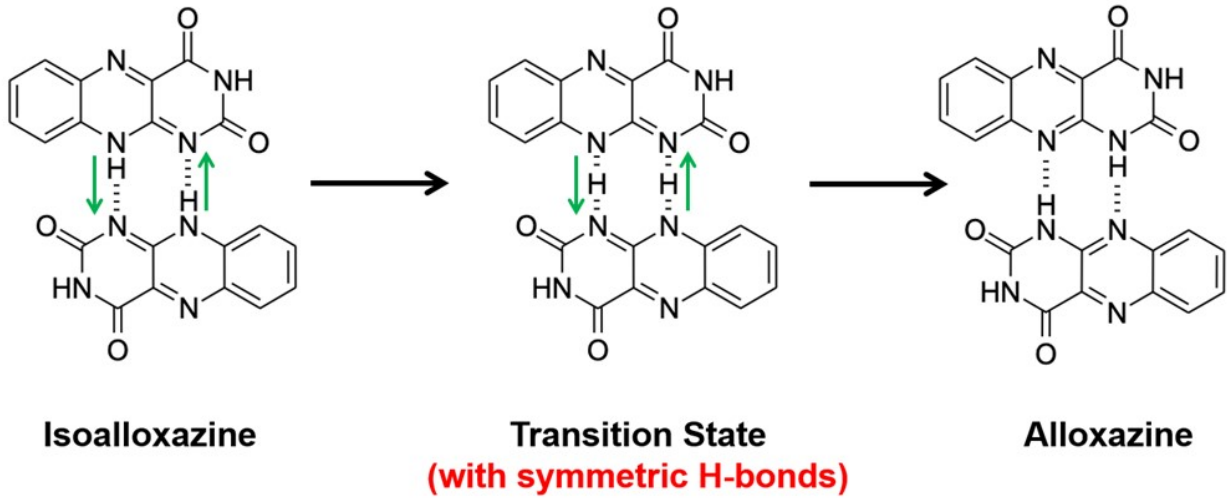

Figure 4.22: Schematic of the transformation from the isoalloxazine crystal structure to the alloxazine crystal structure via the concerted mechanism. Green arrows indicate the directions of proton transfer along N—H···N hydrogen bonds. This figure was prepared by Dr Andrew Logsdail and Prof Kenneth Harris at Cardiff University.

However, DFT-D studies on this system revealed another pathway an ionic intermediate. This mechanism is referred to as the sequential mechanism, shown in Figure 4.23, and involves two steps (described here in the context of conversion from isoalloxazine to alloxazine). The first step involves proton transfer along only one of the N—H···N hydrogen bonds between adjacent molecules, which leads to the formation of an ionic intermediate with a protonated cation and a deprotonated anion, linked by two N—H···N hydrogen bonds. The second step involves proton transfer along the other N—H···N hydrogen bond, which leads to the formation of two molecules of the alloxazine tautomer. First, we note that DFT-D geometry optimization conducted (using PBE-TS and a fixed unit cell) on the ionic intermediate confirms that it is a local energy minimum, representing a metastable structure on the transformation pathway. The energy of the ionic intermediate (calculated following PBE-TS geometry optimization) is higher than the energy of the crystal structure containing isoalloxazine by 9.2 kJ mol^{-1} (PBE-TS), 8.6 kJ mol^{-1} (PBE-MBD), 9.9 kJ mol^{-1} (PBE0-TS), 9.5 kJ mol^{-1} (PBE0-MBD). The two N—H···N hydrogen bonds linking the anion-cation pair in the ionic intermediate are shorter (N···N: 2.86, 2.91 Å) than those linking pairs of molecules in the crystal structures of alloxazine (N···N, 3.09 Å) and isoalloxazine (N···N, 2.99 Å). The inversion center relating the two N—H···N hydrogen bonds in the crystal structures containing the isoalloxazine and alloxazine tautomers (which have $P\bar{1}$ space group symmetry) is lost when forming the ionic intermediate, which has $P1$ space group symmetry.

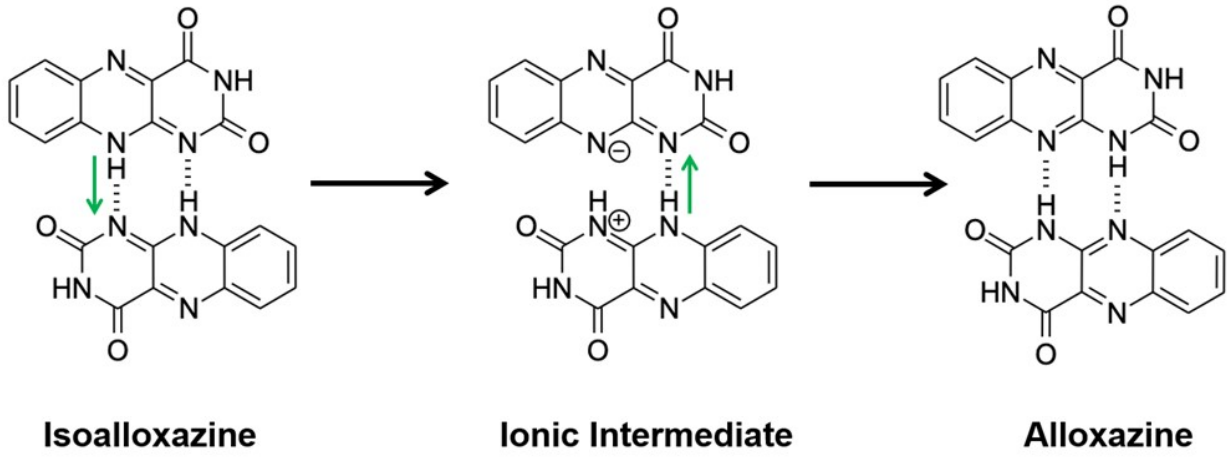

Figure 4.23: Schematic of the transformation from the isoalloxazine crystal structure to the alloxazine crystal structure via the sequential mechanism. Green arrows indicate the directions of proton transfer along the N—H···N hydrogen bonds. This figure was prepared by Dr Andrew Logsdail and Prof Kenneth Harris at Cardiff University.

For the calculations using the PBE-TS exchange-correlation functional, the energy barrier for conversion of the isoalloxazine crystal structure to the alloxazine crystal structure by the concerted mechanism is 21.6 kJ mol^{-1}, while the sequential mechanism has lower energy barriers of 17.0 kJ mol^{-1} to convert the isoalloxazine crystal structure to the ionic intermediate and 2.2 kJ mol^{-1} to convert the ionic intermediate to the alloxazine crystal structure. The corresponding calculations using the PBE-MBD and PBE0-TS methods also suggest that the highest energy barrier for the sequential mechanism (PBE-MBD, 15.8 kJ mol^{-1}; PBE0-TS, 21.5 kJ mol^{-1}), representing the barrier to convert the isoalloxazine crystal structure to the ionic intermediate, is lower than the energy barrier for the concerted mechanism (PBE-MBD, 20.0 kJ mol^{-1}; PBE0-TS, 28.4 kJ mol^{-1}). In stark contrast, however, the calculations using the PBE0-MBD exchange-correlation functional give a higher energy barrier for the sequential mechanism (20.5 kJ mol^{-1}) than the concerted mechanism (16.3 kJ mol^{-1}), indicating that the concerted mechanism is now the favoured pathway for the tautomeric transformation at this level of DFT-D computation. The energy profiles are shown in Figure 4.24.

Consideration of the individual contributions to the total energy for each calculation method shows that, for PBE0-MBD, the transition state in the concerted mechanism is particularly stabilized (by more than 20 kJ mol^{-1}) relative to the alloxazine and isoalloxazine crystal structures as a result of the many-body dispersion corrections. In the transition state for the concerted mechanism, the hydrogen atom in each N—H···N hydrogen bond is essentially equidistant between the two nitrogen atoms and these N—H···N hydrogen bonds are considerably shorter (N···N, 2.69 Å) than those in the crystal structures of alloxazine and isoalloxazine (N···N in the range 2.85 - 3.09 Å). These shorter N—H···N hydrogen bonds may facilitate the stabilization of the transition state when the many-body dispersion correction is considered and may explain the lower overall

energy barrier calculated using PBE0-MBD. These observations further highlight the importance of considering high-level dispersion corrections in future investigations of other proton transfer processes in hydrogen bonds. We note that, among these methods, PBE0-MBD is considered to give the most reliable assessment of the energetic properties of organic materials [61].

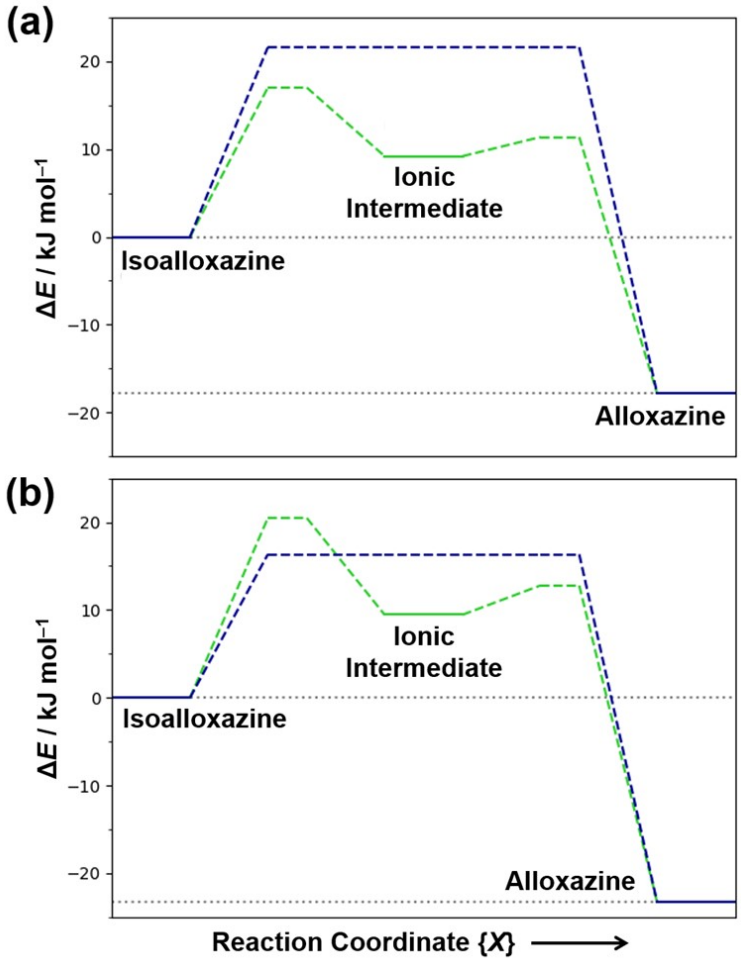

Figure 4.24: Energy profiles for tautomeric interconversion by the concerted mechanism (blue) and the sequential mechanism (green) calculated using (a) PBE-TS and (b) PBE0-MBD. In each case, the horizontal axis represents the general reaction coordinate, and the vertical axis represents the energy relative to the crystal structure containing the isoalloxazine tautomer. Solid lines indicated stable/metastable structures, dashed lines indicate transition processes, with dotted lines indicating energy levels for the crystal structures containing the alloxazine and isoalloxazine tautomers. For calculations using PBE-TS, the sequential mechanism is more favourable. For calculations using PBE0-MBD, the concerted mechanism is more favourable. This figure was prepared by Dr Andrew Logsdail and Prof Kenneth Harris at Cardiff University.

4.4.4 Comparison of the Crystal Structures of Alloxazine and its Derivatives

We now compare the crystal structure of alloxazine with those of its derivatives, focusing initially on non-solvate structures. The crystal structure properties are detailed in 4.5. The crystal structure of 10-methyl alloxazine (Figure 4.25) contains several similarities to that of alloxazine. Firstly, it contains the same strong hydrogen bonding ring described using graph-set notation as $R_2^2(8)$, which consists of two short, linear N···O hydrogen bond (2.941 Å, 174.99°). This structure also contains two C—H···N interactions (3.566 Å, 165.37°) that are marginally longer than in alloxazine and slightly less linear. In this structure, the inversion center exists at the equivalent position in both structures (in the middle of the $R_2^2(8)$ ring). The molecules of 10-methyl alloxazine also exhibit coplanarity as small hydrogen bonding sheets; however, these are made up of more ribbon-like periodicity rather than a full two-dimensional sheet as seen in alloxazine. It is likely that the additional methyl group blocks a strong hydrogen bond accepting nitrogen atom, which would accept a C—H hydrogen donation in the alloxazine structure; therefore, the full potential of the self-complementary hydrogen bonding cannot be achieved and the structure is less extensive. This is likely to be the reason that the density of the structure is 1.603 g cm^{-3} as opposed to 1.686 g cm^{-3} for alloxazine, as the packing of the molecules is not as efficient.

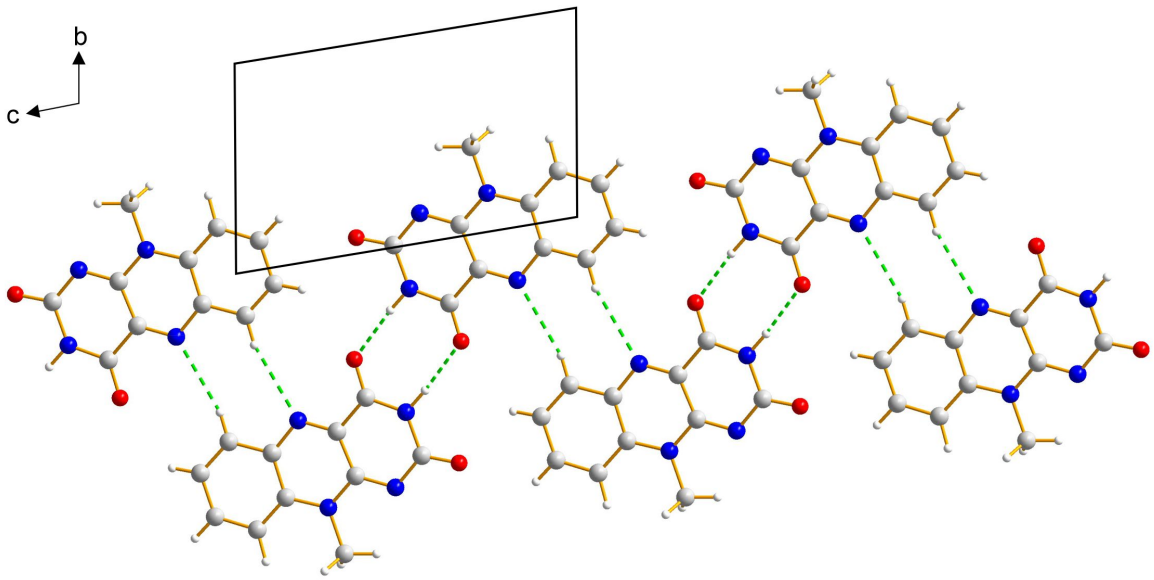

Figure 4.25: The crystal structure of 10-methyl alloxazine viewed along the *a*-axis, with hydrogen bonds indicated by green dashed lines.

The crystal structure of 7,10-dimethyl alloxazine (Figure 4.26) is far less similar to alloxazine than that of 10-methyl alloxazine, although it still contains a hydrogen bonded ring, denoted $R_2^2(8)$, containing two N—H···O=C interactions (2.885 Å, 176.67°). The space group symmetry is also different where Z = 4 is the case for 7,10-dimethyl alloxazine ($P2_1/n$) as opposed to Z = 2 in the other two structures ($P\bar{1}$ for both). Clearly the addition of a second methyl group significantly affects

the packing, with a density of 1.686 g cm^{-3} for alloxazine and 1.503 g cm^{-3} for 7,10-dimethyl alloxazine, representing less efficient packing in the latter case. The lower density may also be a consequence of the less extensive hydrogen bonding in this structure.

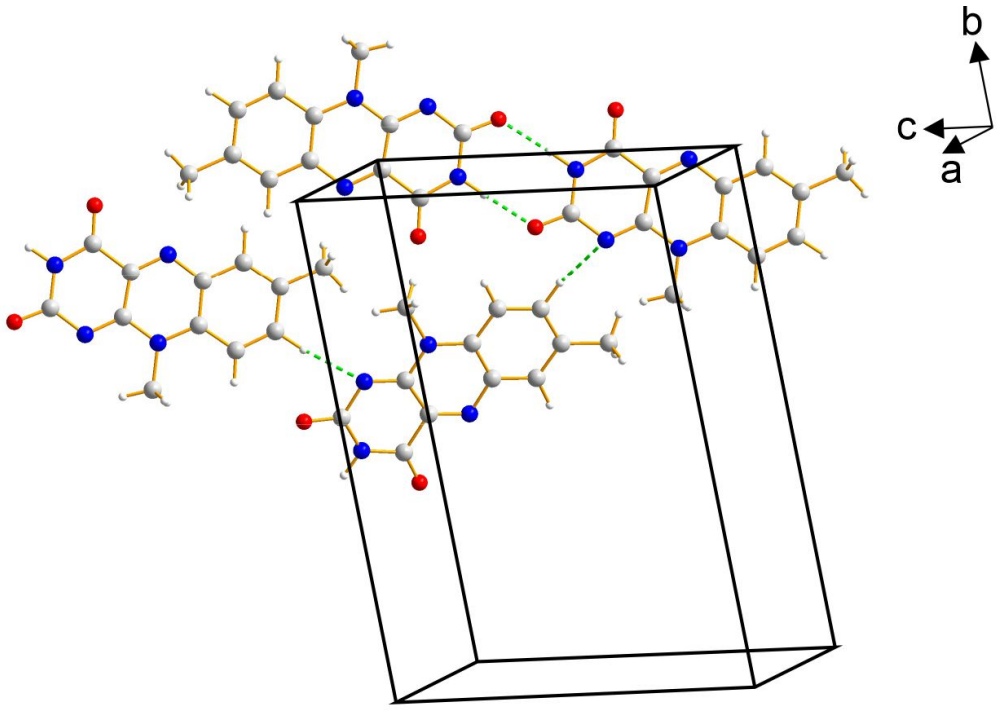

Figure 4.26: The crystal structure of 7,10-dimethyl alloxazine. Hydrogen bonds are indicated by green dashed lines.

We now consider crystal structures of other derivatives of alloxazine (CSD Refcodes: CODYEZ10, FUDWOQ, FISLID, AGEPEK [151–154]), that contain only functionalization of these C_6 ring so that the derivatives retain all the hydrogen-bond donor and acceptor groups that are present in alloxazine. However, we note that these structures are all solvates. Among these four structures, only the methanol solvate of 7,8-dimethylalloxazine contains hydrogen-bonding motifs that resemble those observed in the crystal structure of alloxazine. Specifically, a given molecule of 7,8-dimethyl alloxazine is linked to two adjacent molecules by the A-B and A-C interaction motifs, whilst the other N acceptor is engaged in an O—H···N hydrogen bond with the O—H bond of the methanol molecule (rather than forming the A-D interaction motif observed in the crystal structure alloxazine).

Structure	Alloxazine	10-methyl alloxazine	7,10-dimethyl alloxazine
a / Å	5.8638(6)	7.179(3)	4.6505(5)
b / Å	7.6066(6)	6.725(4)	16.3211(17)
c / Å	10.1644(9)	10.857(3)	14.1162(16)
α /°	69.972(7)	103.10(1)	90
β /°	83.493(8)	75.08(2)	92.109(2)
γ /°	83.061(6)	109.07(2)	90
Volume / Å^3	421.57(9)	472.829	1070.712
Space group	$P\bar{1}$	$P\bar{1}$	$P2_1$/n
Z	2	2	4
Z'	1	1	1
Density / g cm^{-3}	1.686	1.603	1.503
T / K	RT	RT	RT

Table 4.5: Table showing the values of all the crystal structure properties of the non-solvated derivatives of alloxazine: 10-methyl and 7,10-dimethyl alloxazine.

4.5 Conclusion

The solid-state structural properties of alloxazine, a tricyclic ring system found in many biologically important molecules, have been reported, with structure determination conducted directly from powder XRD data. As the crystal structures containing alloxazine and isoalloxazine tautomers both give an exceptionally good fit to the powder XRD data in Rietveld refinement, further techniques are required to determine the tautomeric form present in the crystal structure. In particular, high-resolution ^{15}N NMR data, supported by periodic DFT-D calculations for comparison of relative energy and calculation of solid-state NMR parameters, support the presence of the alloxazine tautomer in the crystal structure. The crystal structure containing the alloxazine tautomer shows a greater agreement with the experimental ^{15}N NMR data (via calculation of chemical shifts through periodic DFT-D) and have a significantly lower energy than the crystal structure containing the isoalloxazine tautomer. Furthermore, periodic DFT-D calculations at the PBE0-MBD level indicate that the crystal structures may interconvert between tautomeric forms via proton transfer along two intermolecular N—H···N hydrogen bonds. A concerted (rather than sequential) proton transfer along these two hydrogen bonds corresponds to a lower energy transition state. However, based on the relative energies of the crystal structures containing the isoalloxazine and alloxazine tautomers, it is estimated that more than 99.99% of the molecules in the crystal structure exist as the alloxazine tautomer at ambient temperature.

On the question of accurate determination of the positions of hydrogen atoms within the crystal structure, classically neutron diffraction is the go-to method, as the relative scattering of neutrons by hydrogen atoms are not negligible, unlike with X-rays. Therefore, future studies may include neutron powder diffraction measurements on alloxazine to verify more accurately the location of the hydrogen atoms in the structure. This would provide an experimental validation to the rigorous computational investigations as presented in this chapter. Performing these measurements *in situ* may also provide interesting insights on whether the proportions of alloxazine and isoalloxazine molecules change as a function of temperature or pressure, supplemented by further calculations to understand whether the amount of alloxazine reduces from 99.9% in response to changes in these conditions.

Chapter 5
Discovery and Structure Determination of a New Polymorph of L-Tyrosine using Powder XRD and 3D-ED Data

Prior to submission of this thesis, the work presented on the crystal structure of β L-tyrosine in this chapter was published in the journal *Chemical Science*. The information required to find this article can be found in Appendix D.

5.1 Introduction

Amino acids have important industrial applications, particularly in pharmaceuticals, as they can be used to alter the physicochemical properties of active pharmaceutical ingredients when incorporated in a co-crystal system. The understanding of the crystal structures of amino acids are highly important. This is because structure-property relationships can be inferred and these can be utilized in the design of materials with improved characteristics, such as solubility and/or melting point, among others [155]. As mentioned in Chapter 1, glycine was the first proteinogenic amino acid to have its crystal structure reported over 70 years ago [33]. Now all 20 directly encoded proteinogenic amino acids have had at least one crystal structure determined following recent progress [34]. Amino acids often exhibit polymorphism (which is described in Section 1.1), notable cases being glycine and phenylalanine, which have 5 and 6 known polymorphs, respectively (see Table 5.1). The crystal structures of the five known polymorphs of glycine (denoted the α, β, γ, δ and ϵ polymorphs) are shown in Figure 5.1. The α polymorph [156] and β polymorph [157] are primitive monoclinic structures and contain a 2_1 screw-axis, whereas the α polymorph has higher symmetry due to the presence of a glide-plane, which is not the case in the β polymorph. The γ polymorph [158] has the highest symmetry among the glycine polymorphs, defined by the $P3_2$ space group (see Figure 5.1). The structures of the δ polymorph [159] and ϵ polymorph [160] are very similar and are both prepared under high-pressure conditions; they are the only glycine structures not to contain any screw-axes, but do contain a glide-plane with Pn space group symmetry.

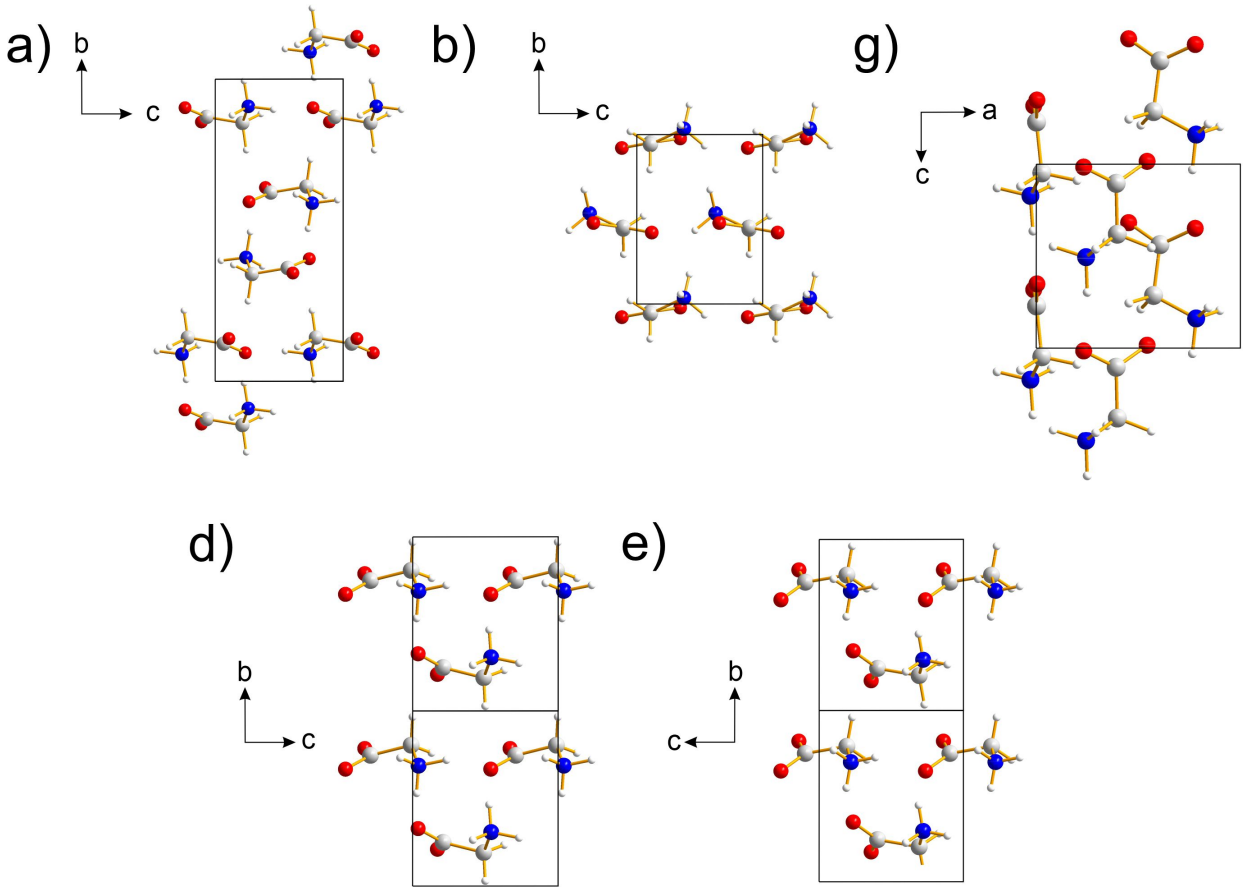

Figure 5.1: The five known crystal structures of glycine: (a) α polymorph, (b) β polymorph, (c) γ polymorph, (d) δ polymorph and (d) ϵ polymorph. The δ and ϵ polymorphs are only observed at high-pressure.

Amino Acid	Polymorphs	Amino Acid	Polymorphs	Amino Acid	Polymorphs
Alanine	1	Glycine	5	Phenylalanine	6
Arginine	1	Histidine	2	Serine	3
Asparagine	1	Isoleucine	1	Threonine	1
Aspartic acid	1	Leucine	2	Tryptophan	1
Cysteine	4	Lysine	1	Tyrosine	1
Glutamic acid	2	Methionine	1	Valine	1
Glutamine	1	Phenylalanine	4		

Table 5.1: Number of known polymorphs of the amino acids, prior to the work described in this chapter.

Due to the varied physicochemical properties exhibited by polymorphs of a material, it is of interest to materials scientists to obtain new polymorphs, which may provide improved physicochemical properties for specific applications. Particularly for amino acids, the most notable properties that can vary through a series of polymorphs would be solubility, melting point and energetic stability. However for some molecular crystals, polymorphism can also influence mechanical properties,

as in *p*-amino and *p*-nitro benzoic acid [161]. One method for preparing novel polymorphs is to crystallize the material directly from the gas phase following sublimation, and has been shown to lead to the formation of new polymorphs for several amino acids [162], as well as known metastable polymorphs. For example, this method led to the preparation of the metastable β polymorph of glycine via sublimation of either the α or γ polymorph. Further work by Liu *et al*. 2008 showed that, upon the sublimation of L-phenylalanine, L-tryptophan and L-tyrosine, new metastable polymorphs were prepared [163]. These preliminary studies inspired the work on L-tyrosine presented in this chapter, with the aim of using crystallization directly from the gas phase to prepare a new polymorph of L-tyrosine suitable for exploiting powder XRD analysis for accurate crystal structure determination of the new polymorph, recalling that the previous studies [163] did not report a structure determination of the new polymorphs prepared. The crystal structure of the known polymorph of L-tyrosine was published in 1972 [164] and is now referred to as the α polymorph of L-tyrosine.

As discussed above, crystallization from the gas phase previously [163] yielded a novel phase of L-tyrosine, among other amino acids. The aim of the work detailed in this chapter was to reproduce the preparation of this material and to obtain powder XRD data (as well as data from other experimental techniques) to allow for crystal structure determination. Alongside the development of a multi-technique approach, it is clear that other techniques (such as solid-state NMR and periodic DFT-D calculations) are useful in providing independent validation that the structure determined from powder XRD data is accurate. Another aspect of this work was to compare the relative energetics of the new polymorph, hereby referred to as the β polymorph, and the previously known α polymorph. By using a multi-technique approach, we have determined an accurate crystal structure for the new polymorph and have gained additional understanding of the energy landscape for this material.

5.2 Experimental Methods

5.2.1 Sample Preparation

The β polymorph of L-tyrosine was prepared by crystallization from the gas phase, following sublimation of a powder sample (*ca*. 12.5 mg) of the α polymorph. The sample was heated under vacuum inside a cylindrical curved-bottom vessel containing a cold finger. The cold finger was filled with acetone/dry ice at -78 °C (dry ice was added periodically to maintain the temperature at -78 °C). The temperature of the sample was increased from ambient temperature to a target temperature over a period of *ca*. 10 min and was then left at the target temperature until sublimation was complete. During this time, microcrystalline powder samples were observed to be deposited on the cold finger (primary deposition) and on the outer glass tube (secondary deposition). The powder samples collected from these various parts of the apparatus were analysed separately. In

the initial experiments involving sublimation at 300 °C, the sample collected from the outer glass tube was found (from PXRD analysis shown in Figure 5.2) to be a biphasic sample comprising primarily the β polymorph, but with small amounts of the α polymorph also present. In later experiments involving sublimation at 275 °C, the sample collected from the outer glass tube was found to be a monophasic sample of the β polymorph. In all experiments, the sample collected from the cold finger was a monophasic sample of the α polymorph.

5.2.2 Powder X-ray Diffraction

Powder XRD data for the original biphasic sample (Figure 5.2) were recorded at 298 K in transmission mode on a Bruker D8 Advance Series II diffractometer (Ge-monochromated Cu-Kα_1 radiation) between 4° and 70° in 2θ (step size; 0.016°). The total data collection period was 64.5 hours using a Vantec detector covering 3° in 2θ. Powder XRD data for the monophasic sample of the β polymorph were recorded using the same procedure, with a data collection time of 191 hrs. In each case, the polycrystalline sample was packed into a 0.7 mm glass capillary (flame sealed) to reduce the effects of preferred orientation and to prevent hydration and was placed on a foil type sample holder for data collection. Although the powder XRD data were measured in the 2θ range from 4° to 70°, no peaks are observed below 2θ = 8° and no peaks are predicted for the β polymorph below this angle. Therefore, the region below 2θ = 8° in the powder XRD data was excluded from Pawley fitting and Rietveld refinement.

To assess the degree of preferred orientation in the powder sample, two-dimensional powder XRD data were recorded at ambient temperature on an Agilent SuperNova diffractometer using Cu-K$_\alpha$ radiation and an Atlas detector. The instrument was configured with the two-dimensional detector perpendicular to the incident beam direction, with a sample-to-detector distance of 100 mm. The capillary containing the powder sample was oriented with its axis perpendicular to the incident X-ray beam. Data were collected over a period of 10 s, with the capillary rotated through an angular range of 10° about the capillary axis during data collection.

5.2.3 Solid-state ^{13}C NMR Spectroscopy

High-resolution solid-state ^{13}C NMR data were recorded at ambient temperature (20 °C) on a Bruker AVANCE III spectrometer at the U.K. High-Field (850 MHz) Solid-State NMR Facility (^{13}C Larmor frequency, 213.8 MHz; 4 mm HX probe; zirconia rotor; MAS frequency, 12 kHz) using ramped ^{1}H → ^{13}C cross-polarization with ^{1}H decoupling (using SPINAL-64 [120]) applied during acquisition. In total, 224 scans were acquired with a recycle delay of 60 s between scans. The solid-state ^{13}C NMR data were referenced using L-alanine, of which the ^{13}C resonance of the carboxylate was set to 177.9 ppm.

5.2.4 3D Electron Diffraction (3D-ED)

The dry powder of the β polymorph of L-tyrosine was dispersed directly onto a continuous carbon film supported by a 200-mesh copper grid, which was mounted onto a Gatan cryo-transfer tomography holder (No. 914). 3D-ED data were collected on this sample using a JEOL JEM-2100 transmission electron microscope (TEM). The accelerating voltage for data collection was 200 kV, and the material was cooled to 100 K. The continuous rotation electron diffraction (cRED) method [165] was used to collect data, involving continuous tilting of the goniometer with a speed of 0.454° s^{-1}. The crystal was tracked during tilting by sequential defocusing of the intermediate lens by using Instamatic software [166]. A data collection range of 98.41° was covered and the diffraction patterns were collected using an integration time of 0.5 s using a high-speed hybrid detection camera Timepix Quad (ASI). The resultant data were processed using X-ray Detector Software (XDS) to extract intensity information [167].

5.2.5 Periodic DFT-D Calculations

Energy calculations

Periodic DFT-D calculations conducted in conjunction with crystal structure determination of the β polymorph of L-tyrosine from powder XRD data were conducted using CASTEP [122] (Academic Release Version 8.0). The geometry optimization calculations used ultrasoft pseudopotentials [123], semi-empirical dispersion corrections [125] (TS correction scheme), PBE functional [124], fixed unit cell and preserved space group symmetry, periodic boundary conditions, a basis set cut-off energy of 700 eV and a Monkhorst-Pack grid [126] of minimum sample spacing 0.05 x 2π $Å^{-1}$. Convergence criteria for geometry optimization were 0.01 eV $Å^{-1}$ for forces, 0.00001 eV per atom for energy, and 0.001 Å for atomic displacements.

Following geometry optimization using the methodology described above, higher level calculations utilizing other functionals and dispersion correction schemes were conducted as a single-point energy calculation, performed using FHI-aims [139] (date stamp: 198013), recognizing that this software is highly efficient for periodic hybrid-DFT calculations. Generalized gradient approximation (GGA) and hybrid-GGA exchange correlation functionals, specifically PBE [124] and PBE0 [142], were considered in combination with the TS dispersion correction scheme [125], and many-body dispersion correction (MBD) [61]. Thus, the combination of exchange-correlation functionals and dispersion corrections considered in these calculations were PBE-TS, PBE-MBD, PBE0-TS, PBE0-MBD. The calculations were conducted using an 'intermediate' basis set with relativistic effects included through the scaled zeroth order regular approximation [124]. A Γ-centered **k**-grid, with a minimum sample spacing of (0.05 x 2π) $Å^{-1}$, was used; testing with a denser **k**-grid sampling of (0.04 x 2π) $Å^{-1}$ gave changes in relative energy less than 1 meV. The self-consistent field (SCF) cycle was considered converged at the point where the changes in the electron density, the

total energy and the sum of eigenvalue energies were under 10^{-6}e $a_0{}^{-3}$, 10^{-6} eV and 10^{-6} eV respectively.

NMR chemical shift calculations

Solid-state ^{13}C NMR chemical shifts were calculated using CASTEP (Academic Release version 8.0) for the crystal structure of the β polymorph of L-tyrosine obtained in the Rietveld refinement and for the published crystal structure of the α polymorph of L-tyrosine, in each case following DFT-D geometry optimization (with fixed unit cell). The Gauge Including Projector Augmented Wave (GIPAW) approach [127–131] was used with a cut-off energy of 700 eV and PBE functional. A set of isotropic ^{13}C NMR shielding values were generated from the CASTEP calculations. From the isotropic ^{13}C NMR shielding value (σ_{calc}) calculated for each ^{13}C environment in the crystal structure, the corresponding calculated isotropic ^{13}C NMR chemical shift (δ_{calc}) was determined from the equation $\delta_{calc} = <\delta_{exp}> + <\sigma_{calc}> - \sigma_{calc}$, where $<\delta_{exp}>$ denotes the mean of the isotropic ^{13}C NMR chemical shifts determined from the experimental high-resolution solid-state ^{13}C NMR spectrum ($<\delta_{exp}>$ = 116.23 ppm. β polymorph) and $<\sigma_{calc}>$ denotes the mean of the calculated isotropic ^{13}C NMR shielding values ($<\sigma_{calc}>$ = 54.21 ppm, β polymorph: $<\sigma_{calc}>$ = 53.33 ppm, α polymorph).

5.2.6 Crystal Structure Prediction

Crystal structure prediction for L-tyrosine was conducted using the Ab-Initio Random Structure Searching (AIRSS) methodology [76, 168]. This method involves two separate steps: 1) generating trial structures through 'intelligent' random searching, and 2) subsequent geometry optimization of the periodic structure followed by ranking of the trial structures based on crystal structure energy. The resultant structures from this procedure are subjected to more advanced DFT-D calculations (e.g. using the PBE0-MBD approach) for a more accurate energy ranking.

In the crystal structure prediction of L-tyrosine, the initial stage involved crystal structures for which the number of molecules per unit cell was: 1, 2, 3, 4 or 6. Structures were generated initially by randomly selection of unit cell parameters and random positioning of the molecule within the unit cell, with space group symmetry operators applied (for all space groups). The unit cell parameters were allowed to change within limits to give a volume per molecule within 10% of the initial estimate of molecular volume, which was based on the published crystal structure of the α polymorph of L-tyrosine (CSD REFCODE: LTYROS10), for which the unit cell volume is 850 Å^3 with four molecules in the unit cell. Minimum allowed intermolecular atomic distances were specified to eliminate structures with chemically unreasonable intermolecular interactions. Searches with smaller unit cells (with Z = 1 or 2) were conducted with a minimum separation of 2 Å globally. The minimum separations between pairs of atoms in different molecules in the resulting low-energy

structures were then used to define minimum separations in searches with larger unit cells (Z = 3, 4 or 6).

The second stage involved initial geometry optimization (including unit cell relaxation) of the structures generated in the first stage, and subsequent energetic ranking, conducted by DFT calculations using the program CASTEP [122]. These calculations used the PBEsol functional [169], a plane-wave cut-off energy of 800 eV and a Brillouin zone sampling size of (0.07 x 2π) Å^{-1}. The criteria for convergence of the total energy, maximum atomic force, and maximum atomic displacement, and maximum stress component were as follows: 0.00002 eV per atom, 0.05 eV Å^{-1}, and 0.001 Å, and 0.1 GPa, respectively. Following the initial geometry optimization, it is clear that reasonably low-energy structures were only obtained for crystal structures with $Z = 2$ and $Z = 4$. Low-energy structures in achiral space groups were removed to leave only chiral structures.

5.3 Crystal Structure Determination

Initial attempts to progress with structure determination of the β polymorph of L-tyrosine from powder XRD data were unsuccessful as indexing using algorithms in the CRYSFIRE [89] suite could not account for all peaks present in the powder XRD pattern. However, 3D-ED data recorded for the same sample were indexed successfully using REDp [165] and subsequently XDS [167] to give the following unit cell with monoclinic metric symmetry: a = 7.92 Å, b = 6.13 Å, c = 9.90 Å and β = 94.82°. The space group was assigned from the 3D-ED data as $P2_1$, based on systematic extinctions.

The unit cell determined from the 3D-ED data provided insights into the failure of the initial indexing from powder XRD data. From the powder XRD pattern, it was clear that the β polymorph of L-tyrosine was the dominant phase, but with some peaks also present that were attributed to the known (α) polymorph of L-tyrosine, as shown in Figure 5.2. Although the sample was biphasic, the proportion of the α polymorph of L-tyrosine was sufficiently low, with sufficiently few overlapping peaks, such that structure solution of the β polymorph using the powder XRD data of the biphasic sample was considered viable. Profile fitting, using the Pawley method in TOPAS [85, 97] was conducted using the data set for the biphasic sample. Pawley profile fitting (Figure 5.2) provided the following unit cell for the β polymorph with monoclinic metric symmetry: a = 7.6767(1) Å, b = 5.8825(6) Å, c = 9.6095(1) Å, β = 94.549(7)° (V = 432.6(1) Å^3). A good quality fit to the data was obtained (R_{wp} = 0.41%, R_p = 0.28%) with the only significant discrepancies arising due to the presence of impurity peaks from the α polymorph. It is clear that the presence of the two impurity peaks at low angle (as well as others at high angle) were the reason that successful indexing from powder XRD data alone was unsuccessful.

Structure solution was conducted independently from the 3D-ED data and powder XRD data using the direct-space strategy implemented in the program EAGER [133–135]. While EAGER

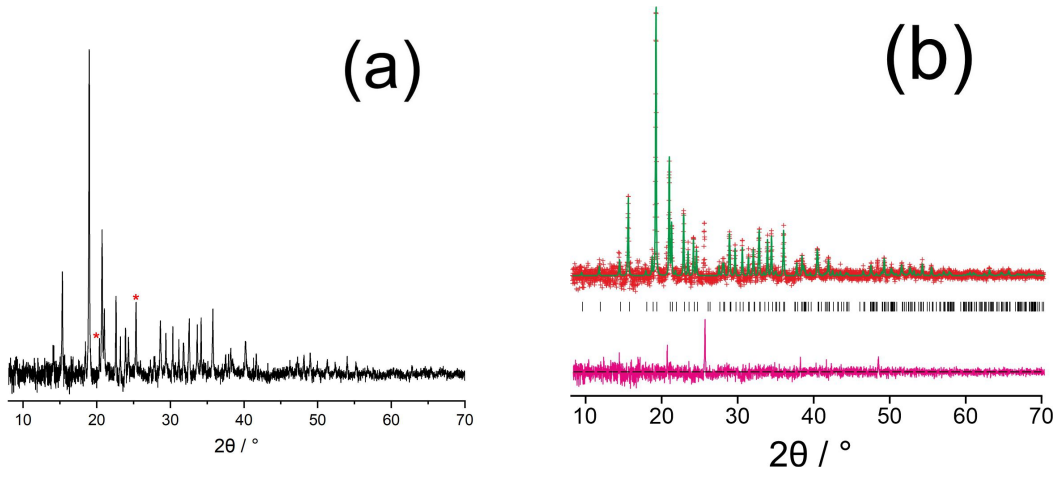

Figure 5.2: The powder XRD pattern for the biphasic sample of L-tyrosine (a) with red asterisks denoting the peaks attributed to the α polymorph using the unit cell for the β polymorph only in the profile fitting calculation. The Pawley fit for the powder XRD pattern of the biphasic sample is also shown (b).

was originally developed and applied for structure solution from powder XRD data, it has been extended recently to allow structure solution to be conducted from 3D-ED data within the kinematical scattering approximation [170]. In the direct-space GA structure solution calculations, the contents of the asymmetric unit comprised one molecule of L-tyrosine in the zwitterionic form, constructed using standard bond lengths and bond angles (based on geometric information from MOGUL [144] and, for bonds involving hydrogen, from Allen *et al*. [145]). Trial crystal structures were defined by a total of eight structural variables (two positional, three orientational and three torsional, shown in Figure 5.3). It is noted that, for space group $P2_1$, the position of the molecule along the *b*-axis can be fixed; thus only two positional variables are required in the direct-space search. The torsion angle between the α-carbon and the ammonium (NH_3^+ group) was not varied in this calculation. Due to the low scattering power of the hydrogen atoms, it is clear the fit to the powder XRD data and 3D-ED data would be insensitive to the positions of the H-atoms of the NH_3^+ group, and other techniques such as geometry optimization using periodic DFT-D calculations would provide a more reliable description of the orientation of this group in the crystal structure.

In the structure solution calculations using the 3D-ED data and the powder XRD data (for the original biphasic sample), a total of 40 independent GA structure solution calculations were conducted, in each case from a different random initial population. Each GA calculation involved the evolution of a population of 100 trial structures for 100 generations, with 10 mating operations and 50 mutation operations conducted per generation. The R-factors used to assess the quality of trial structures (based on the level of agreement between calculated and experimental diffraction data) in the structure solution calculations from the 3D-ED data (R_F) and powder XRD data (R_{wp}) are defined in Equations 5.1 and 3.3 respectively. In Figure 5.4 the evolution of the lowest R-factor in the population as a function of generation number, for all 40 independent calculations for both the powder XRD data and 3D-ED data, are shown. After 100 generations, the success rate in finding

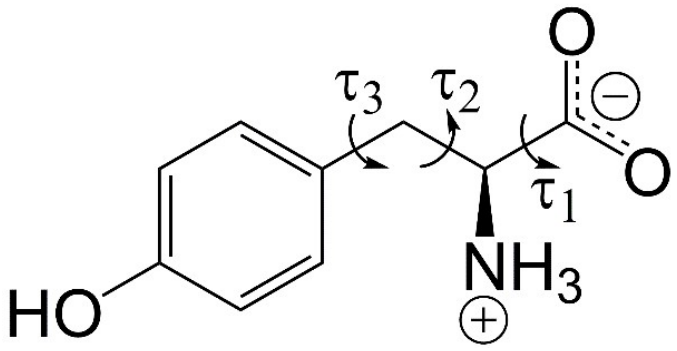

Figure 5.3: Molecular structure of L-tyrosine in the zwitterionic form. The variable torsion angles (τ_1, τ_2 and τ_3) in the direct-space GA structure solution calculations are indicated.

the correct structure solution is significantly higher for the powder XRD data than the 3D-ED data. For the 3D-ED data, 7 of 40 independent GA calculations generated essentially the same structure solution with low *R*-factor; these 7 structures are among those with R_F in the range 41.8% - 43.2% shown in Figure 5.4(a). For the powder XRD data, 38 of the 40 independent GA calculations generated essentially the same structure solution with the lowest *R*-factor, correspond to those with R_{wp} ca. 23.5% shown in Figure 5.4(b). The best structure solutions (with lowest R-factor) from the 3D-ED data and the powder XRD data represent essentially the same structure (see Figure 5.5), with root-mean-squared difference (RMSD) of only 0.33 Å between the positions of the non-hydrogen atoms in these structures.

$$R_F = 100 \times \left(\frac{\sum_i \left|\left|F_{o,i}\right| - \left|F_{c,i}\right|\right|}{\sum_i \left|F_{o,i}\right|}\right) \tag{5.1}$$

The structure factor amplitude (the square-root of the measured intensity) for the *i*th reflection in the experimental 3D-ED dataset is given as $|F_{o,i}|$, the structure factor amplitude for the corresponding reflection calculated for the structural model is given as $|F_{c,i}|$.

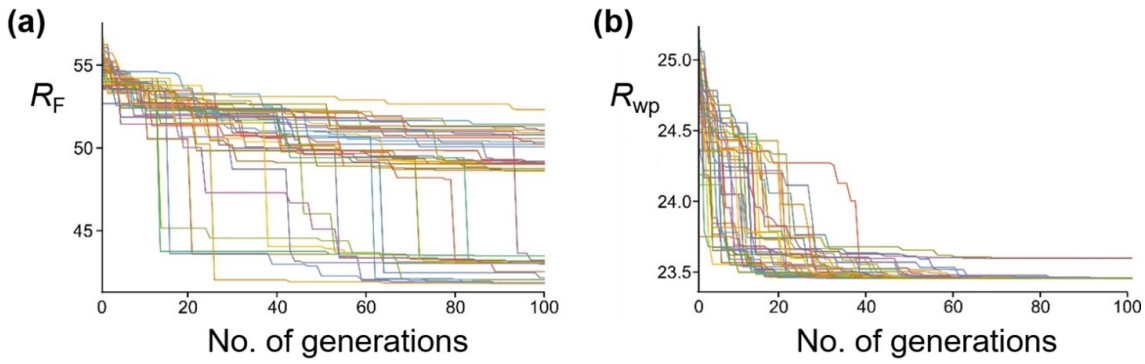

Figure 5.4: Evolutionary progress in the GA structure solution calculations using (a) the 3D-ED data and (b) the powder XRD data for the original biphasic sample. In each case, the evolution is shown for 40 independent GA calculations (each starting from a different random initial population of trial structure). Each continuous line represents the evolution of one of the 40 independent GA calculations and show the lowest *R*-factor among all 100 trial structures in the population as a function of generation number. The *R*-factors used in the analysis of the 3D-ED data (R_F) and the powder XRD data (R_{wp} are defined in Equation 5.1 and 3.3 respectively (note that the absolute values of R_F and R_{wp} cannot be compared directly).

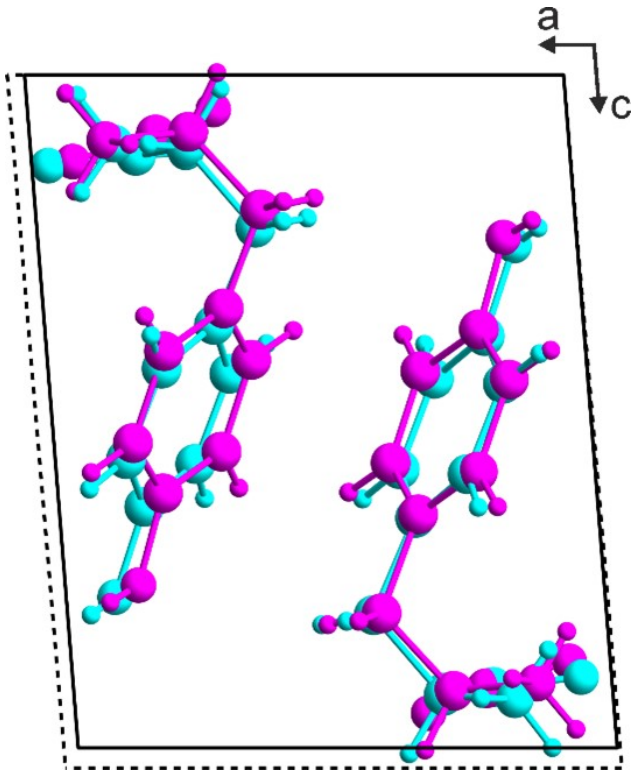

Figure 5.5: Overlay of the best structure solutions (with lowest R-factor) viewed along the *b*-axis for the β polymorph of L-tyrosine obtained from the direct-space GA structure solution calculations using 3D-ED data (cyan; unit cell shown by the dashed black lines) and powder XRD data (magenta; unit cell shown by the solid black lines).

Our initial Rietveld refinement [105] using TOPAS [85] used the powder XRD dataset for the original biphasic sample that was used in the structure solution calculations described above. However, at this stage, further attempts to prepare the β polymorph of L-tyrosine by crystallization from the gas phase under different experimental conditions were successful in producing a monophasic sample, confirmed by powder XRD. High-quality powder XRD data were recorded for this monophasic sample, which was used for the final Rietveld refinement calculations. Initially, profile fitting using the Pawley method [97] gave a good fit to this experimental powder XRD data (R_{wp} = 0.29%; R_p = 0.22%). The best structure solution obtained from the EAGER calculations on the powder XRD data for the biphasic sample discussed above was used as the initial structural model for the Rietveld refinement, with standard restraints applied to bond lengths and bond angles, and planar restraints applied to the phenyl ring and carboxylate group. A common isotropic displacement parameter was refined for all non-hydrogen atoms, and the isotropic displacement parameter for hydrogen atoms was set at 1.2 times this value. As powder XRD data recorded for the same powder sample using a two-dimensional detector (Figure 5.6) clearly showed non-uniform intensity of the Debye-Scherrer rings, it was concluded that the powder sample exhibits preferred orientation; thus a preferred orientation correction was applied in the Rietveld refinement using the March-Dollase method [106], with a refined March parameter of 0.67 for the (0 1 0) plane. After completing the Rietveld refinement, the refined structure was subjected to DFT-D geometry optimization with fixed unit cell, which led to only minor structural changes. In utilizing DFT-D geometry optimization in this way to facilitate improvements in the structural model, in tandem with Rietveld refinement, it is crucial that the unit cell is fixed in the DFT-D geometry optimization calculation in order that the unit cell matches the experimental powder XRD data. The geometry optimized structure was then used as the starting model for the final Rietveld refinement, which included additional restraints on the geometries of hydrogen bonds based on the structure obtained in the DFT-D geometry optimization. The final Rietveld refinement (Figure 5.7) gave a high-quality fit to the powder XRD data (R_{wp} = 0.30%, and R_p = 0.23%), comparable to the quality of fit obtained in Pawley fitting (Figure 5.7), with the following final refined parameters: a = 7.6764(6) Å, b = 5.8870(3) Å, c = 9.6143(6) Å and β = 94.575°, V = 433.10(5) Å^3.

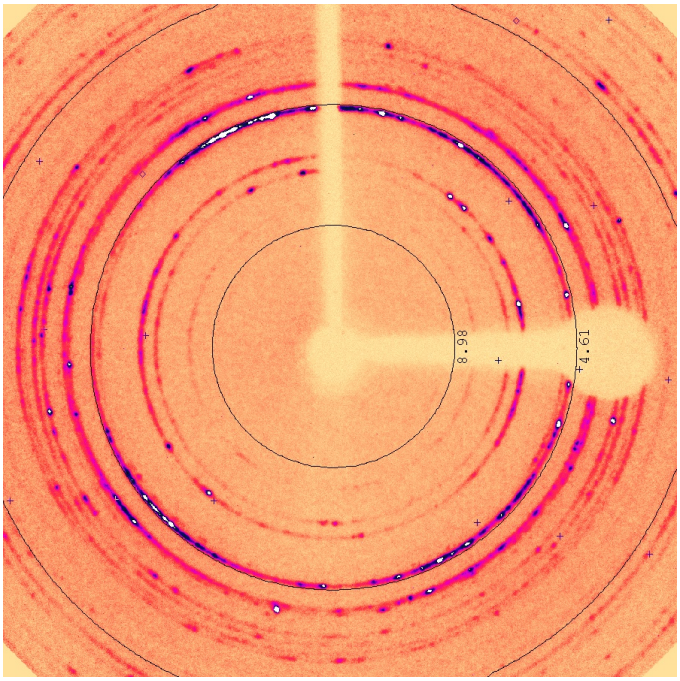

Figure 5.6: Two-dimensional powder XRD data for the monophasic sample of the β polymorph of L-tyrosine providing evidence for preferred orientation within the sample.

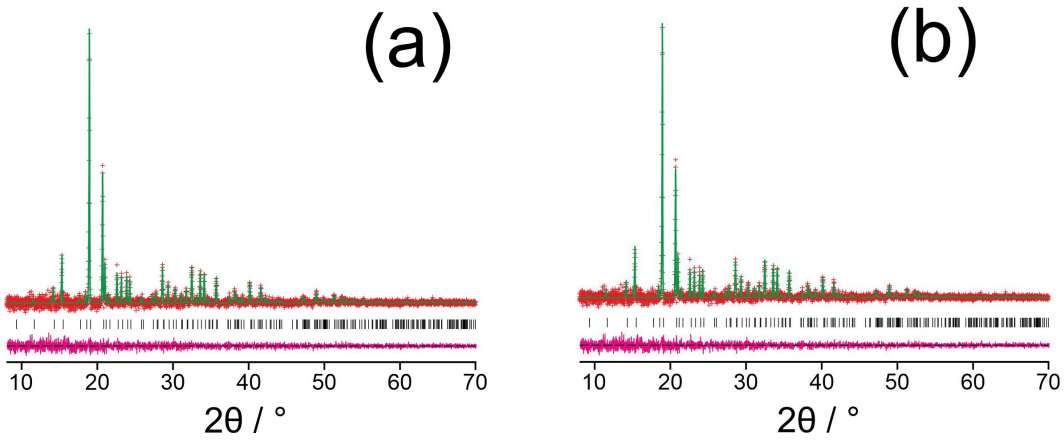

Figure 5.7: (a) Pawley fit and (b) the final Rietveld fit for the monophasic sample of the β polymorph of L-tyrosine.

Least-squares refinement was also conducted against the 3D-ED data using the kinematical treatment. The starting structural model for the refinement was the structure solution obtained from the 3D-ED data using EAGER, with unit cell parameters taken from the profile-fitting (in this case Pawley fitting) of the powder XRD data and with hydrogen atoms included based on geometric restraints. An initial refinement using isotropic atomic displacement parameters converged with R_1 = 25.1%. A final refinement using additional restraints on the C–C, C–O and C–N distances to improve the bonding geometries converged with R_1 = 25.1%. The final refined structures from the powder XRD data and 3D-ED data are in good agreement (RMSD for non-hydrogen atoms, 0.135 Å), shown as an overlay in Figure 5.8.

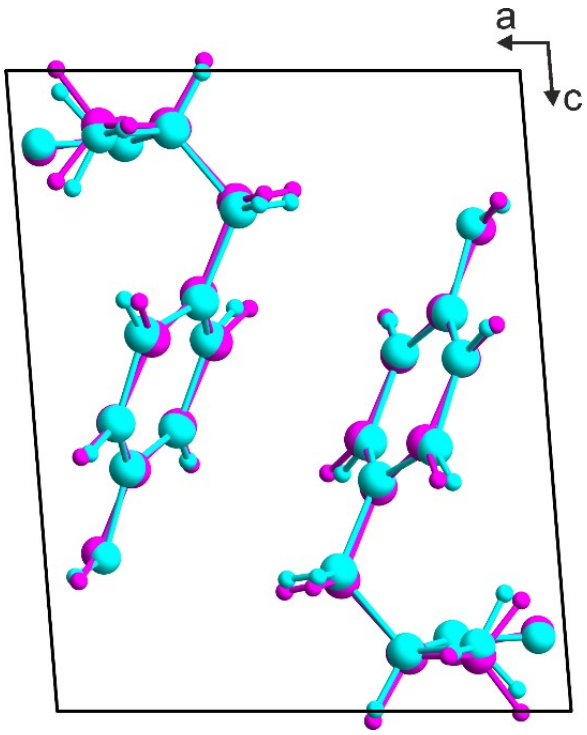

Figure 5.8: Overlay of the crystal structures (viewed along the *b*-axis) of the β polymorph of L-tyrosine obtained by Rietveld refinement from the powder XRD data (magenta) and by refinement from the 3D-ED data (cyan).

As further validation of the crystal structure of the β polymorph of L-tyrosine, DFT-D geometry optimization (using PBE-TS and with fixed unit cell) on the final refined structure from the Rietveld refinement led to only minor atomic displacements (RMSD for non-hydrogen atoms, 0.025 Å), confirming that the final refined crystal structure is very close to a minimum on the energy landscape, shown in an overlay in Figure 5.9. Furthermore, the isotropic solid-state ^{13}C NMR chemical shifts calculated using DFT-D/GIPAW [131] methodology for the crystal structure of the β polymorph are in close agreement with the isotropic chemical shifts in the experimental high-resolution solid-state ^{13}C NMR spectrum of this material. Thus, in addition to being in excellent agreement with the powder XRD data, the crystal structure of the β polymorph of L-tyrosine determined here is also confirmed to be in excellent agreement with the experimental solid-state ^{13}C NMR data.

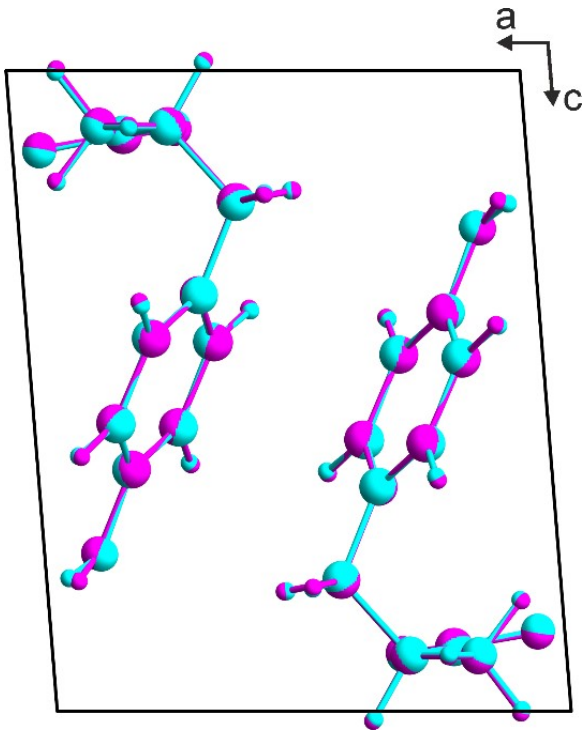

Figure 5.9: Overlay of the crystal structure of the β polymorph of L-tyrosine obtained in the final Rietveld refinement (magenta) and the crystal structure obtained after subjecting this structure to periodic DFT-D geometry optimization using PBE-TS with fixed unit cell (cyan).

It is noted that the high-resolution solid-state ^{13}C NMR spectrum of the β polymorph is clearly distinct from that of the α polymorph, and in each case there is good agreement between the experimental solid-state ^{13}C NMR chemical shifts and the isotropic ^{13}C NMR chemical shifts calculated from the crystal structure using DFT-D/GIPAW methodology as shown in Figure 5.10 (β polymorph) and Figure 5.11 (α polymorph). These observations highlight the utility of DFT-D/GIPAW [131] calculations to compute reliable solid-state NMR data for known crystal structures, both in the context of polymorph characterization and as a powerful strategy in conjunction with structure determination from powder XRD data. The absolute values of the calculated solid-state ^{13}C NMR chemical shifts for both the α and β polymorphs are given in Appendix A.3.

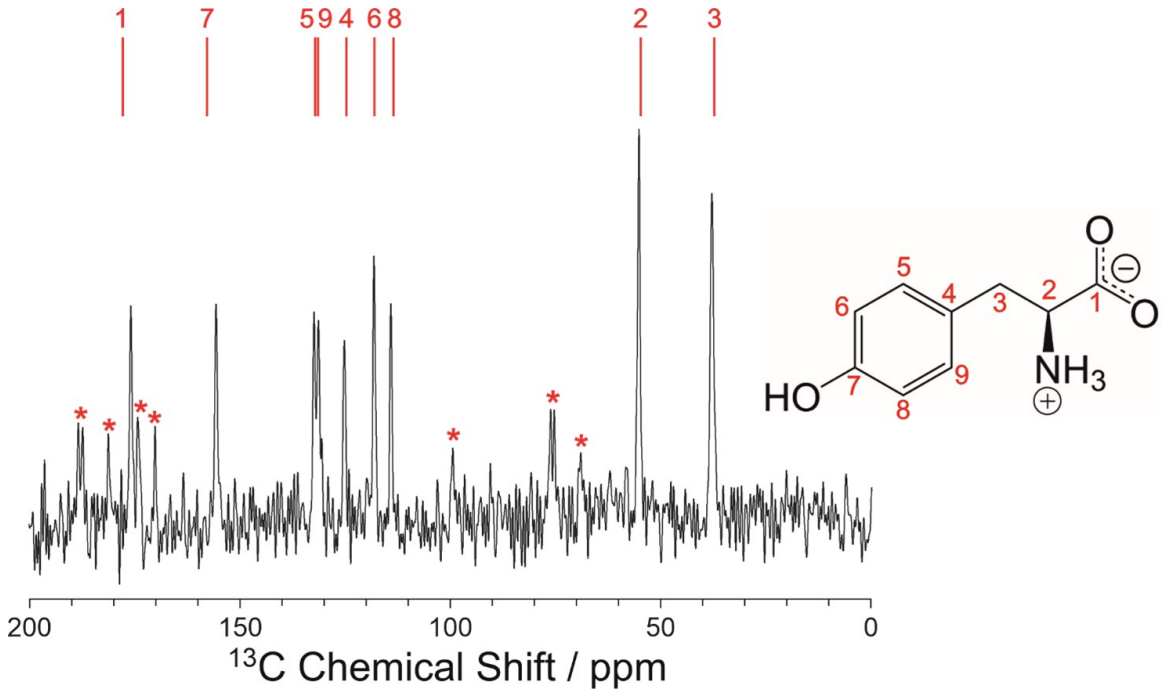

Figure 5.10: Experimental high-resolution solid-state ^{13}C NMR spectrum recorded for the β polymorph of L-tyrosine with the values of isotropic ^{13}C NMR chemical shifts calculated for the crystal structure of the β polymorph indicated by the red lines above the spectrum. The specific ^{13}C site corresponding to each calculated value is indicated. Spinning sidebands in the experimental spectrum are marked by red asterisks.

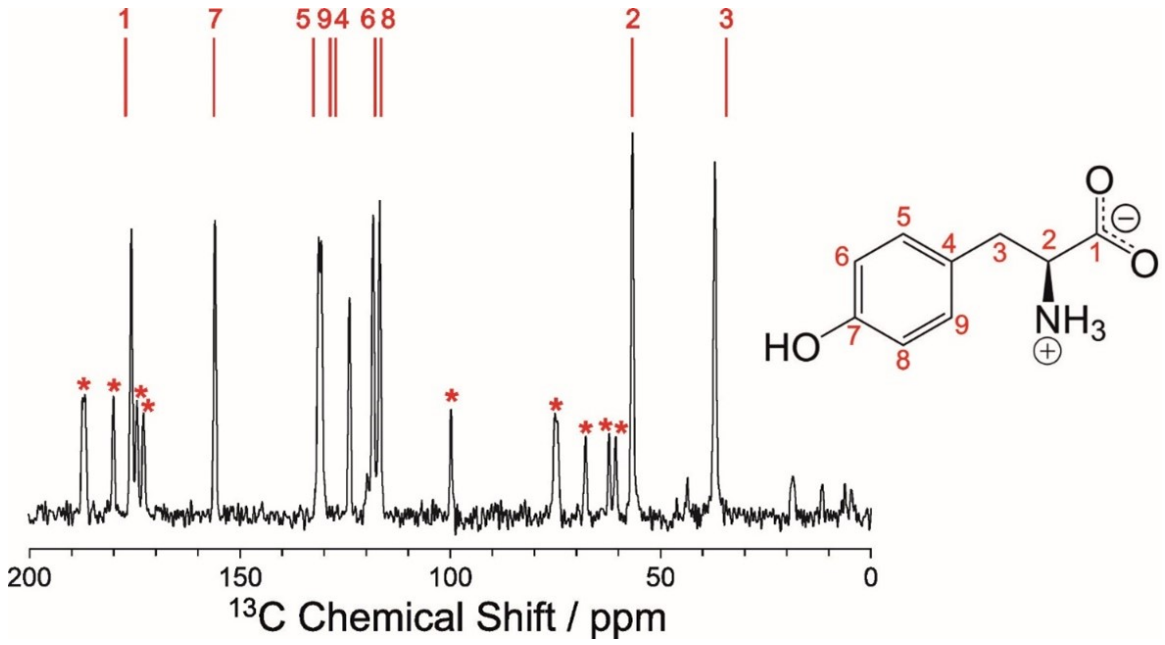

Figure 5.11: Experimental high-resolution solid-state ^{13}C NMR spectrum recorded for the α polymorph of L-tyrosine with the values of isotropic ^{13}C NMR chemical shifts calculated for the crystal structure of the α polymorph indicated by the red lines above the spectrum. The specific ^{13}C site corresponding to each calculated value is indicated. Spinning sidebands in the experimental spectrum are marked by red asterisks.

5.4 Energetic Comparison of the α and β Polymorphs of L-tyrosine

To assess the relative energetic properties of the crystal structure of the β polymorph of L-tyrosine reported here and the crystal structure of the α polymorph reported previously [171] (from single-crystal XRD), periodic DFT-D calculations were conducted using different functionals (PBE and PBE0) combined with different methods for dispersion correction (TS and MBD). These calculations were performed, and the data processed, by Dr Andrew Logsdail at Cardiff University. For each polymorph, the experimentally determined crystal structure at ambient temperature was taken as the initial structural model for DFT-D geometry optimization (with fixed unit cell) using PBE-TS. For the resulting geometry optimized structures, single-point energy calculations were conducted using each of the four combinations of functional and dispersion correction method discussed above. For each calculation method, the energy is higher for the β polymorph than the α polymorph, with the energy differences reported in Table 5.2. Among these methods, PBE0-MBD is considered to give the most reliable assessment of the relative energies of polymorphs of organic materials [61, 172]. The calculated energy difference using PBE0-MBD of 3.64 kJ mol^{-1} between the α and β polymorphs is fully consistent [173] with the β polymorph being an experimentally observable metastable polymorph of L-tyrosine. From the crystal structures determined at ambient

temperature, the densities of the α and β polymorphs are 1.414 g cm^{-3} and 1.389 g cm^{-3}, respectively. The higher density of the α polymorph may be a contributing factor to the lower energy of this polymorph established from the DFT-D calculations.

Calculation type	Relative energy / kJ mol^{-1}
PBE-TS	4.74
PBE0-TS	3.66
PBE-MBD	3.96
PBE0-MBD	3.64

Table 5.2: Relative energy (per molecule) of the β polymorph of L-tyrosine relative to the energy of the α polymorph, calculated using different functionals and methods of dispersion correction in DFT-D calculations.

5.5 Structural Properties of the β Polymorph of L-tyrosine

The β polymorph of L-tyrosine may be described as a bilayer structure, comprising alternate hydrophobic and hydrophilic layers parallel to the *ab*-plane (Figure 5.13). The hydrophilic region contains the amino acid head-groups and the OH groups of the side-chains, while the hydrophobic region contains the phenyl rings of the side-chains. The hydrogen-bonding involving the amino acid head-groups comprises a ribbon motif (Figure 5.12), propagating along the *b*-axis, and constructed from two strands of L-tyrosine molecules. Within a given strand, adjacent molecules are related by translation along the *b*-axis, while the two strands are related to each other by the 2_1 screw along the *b*-axis. In each molecule, the atoms of the N–C(α)–CO_2 unit are essentially co-planar, and this plane is parallel to the crystallographic *ab*-plane. Within a given hydrogen-bonded ribbon, the planes of the N–C(α)–CO_2 units of the molecules in each strand are parallel to each other, but displaced slightly along the *c*-axis, as shown in Figure 5.12. The hydrogen-bonded ribbon (Figure 5.12) is constructed from short and relatively linear N–H$\cdots$O hydrogen bonds, both between adjacent molecules in a given strand (N$\cdots$O, 2.70 Å; N–H$\cdots$O, 178.4°) and between molecules in the two strands (N$\cdots$O, 2.79 Å; N–H$\cdots$O, 159.1°), giving rise to a cyclic hydrogen-bonded array described as $R_3^3(11)$ in graph set notation [146]. Significantly, a given hydrogen-bonded ribbon is not engaged in hydrogen bonding with any other ribbon. However, the hydrogen-bonded ribbon participates in additional hydrogen bonding with the OH groups of the side-chains of the molecules that form the hydrogen-bonded ribbons in the layers above and below along the *c*-axis. As shown in Figure 5.14, each OH group serves as the donor in an O–H$\cdots$O hydrogen bond (O$\cdots$O, 2.62 Å; O–H$\cdots$O, 165. 2°) and as the acceptor in an N–H$\cdots$O hydrogen bond (N$\cdots$O, 2.86 Å; N–H$\cdots$O, 129.7°) with carboxylate and ammonium groups, respectively, in the hydrogen-bonded ribbon, giving a cyclic motif described as $R_3^3(11)$ in graph set notation [146]. The crystallographic data for both the β and α polymorphs are shown in Table 5.3.

Structure	α polymorph	β polymorph
a / Å	6.913(5)	7.6764(6)
b / Å	21.116(3)	5.8870(3)
c / Å	5.829(4)	9.6143(6)
β /°	90	94.575(5)
Volume / Å^3	850.88	433.10(5)
Space group	$P2_12_12_1$	$P2_1$
Z	4	2
Z'	1	1
Density / g cm^{-3}	1.414	1.389
Method	Single crystal XRD	Powder XRD
Experimental temperature / K	RT	RT
CSD Refcode	LTYROS10	2114085

Table 5.3: Structural properties of the α and β polymorphs of L-tyrosine.

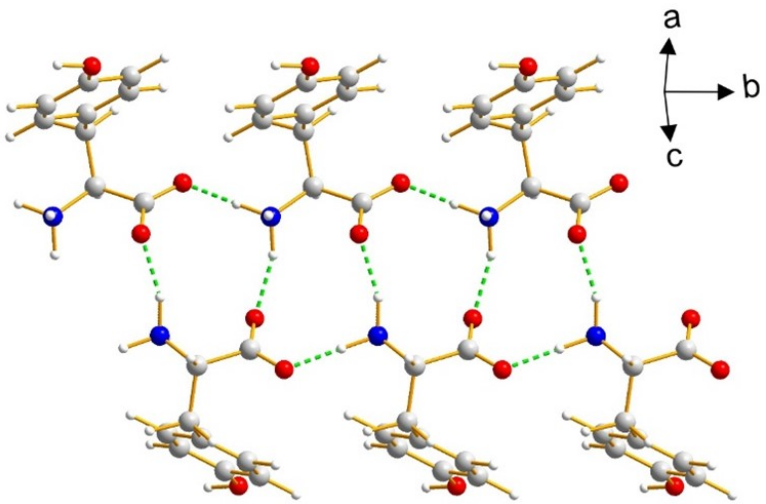

Figure 5.12: The hydrogen-bonded ribbon involving the amino acid head-groups in the β polymorph of L-tyrosine. Hydrogen bonds are indicated by green dashed lines.

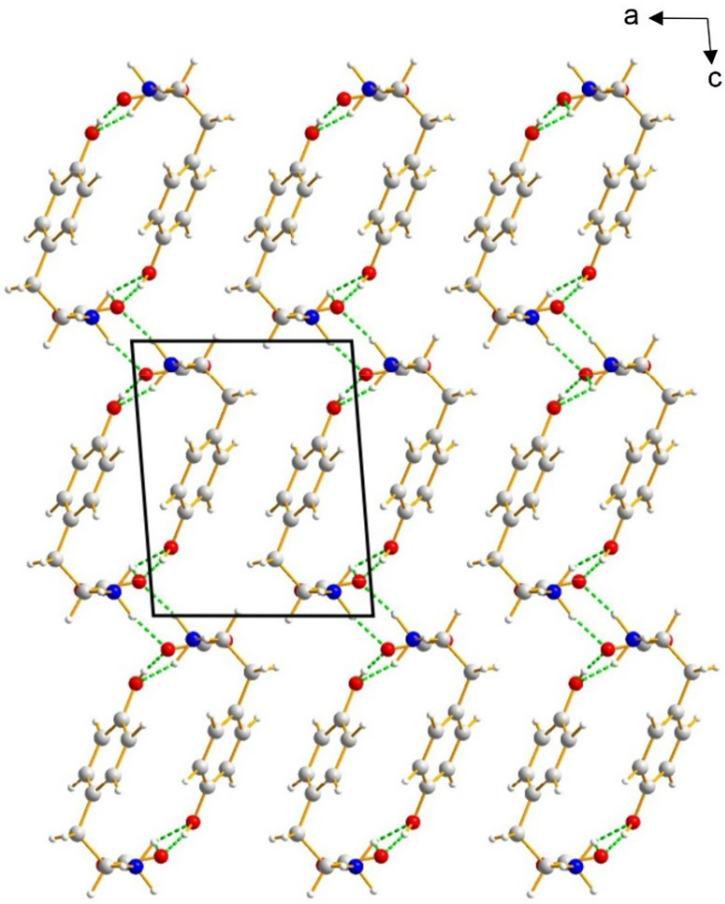

Figure 5.13: Crystal structure of the β polymorph of L-tyrosine viewed along the *b*-axis, showing the alternating hydrophilic and hydrophobic layers parallel to the ab-plane (horizontal). The hydrogen-bonded ribbons involving the amino-acid head-groups propagate along the direction of view. Hydrogen bonds are indicated by green dashed lines.

Comparing the structural properties of the α and β polymorphs of L-tyrosine, the molecular conformations (defined by torsion angles τ_1, τ_2 and τ_3; see Figure 5.16) are similar in each case: τ_1 (O–C–C–C) = –71.06° (α), –62.78° (β); τ_2 (C–C–C–C)= –53.08° (α), –53.99° (β); τ_3 (C–C–C–C) = 95.51° (α), 99.96° (β). However, in contrast to the one-dimensional hydrogen-bonded ribbons propagating along the *b*-axis in the β polymorph, the hydrophilic region of the α polymorph is a two-dimensional hydrogen-bonded array (parallel to the *ac*-plane; see Figures 5.15 and 5.16) involving the amino acid head-groups and the OH groups of the side-chains of L-tyrosine molecules. As a consequence, the α polymorph is a three-dimensionally connected hydrogen-bonded structure, whereas the contiguous hydrogen-bonded network in the β polymorph comprises corrugated slabs with a mean plane parallel to the *bc*-plane (Figure 5.12). This is constructed from the hydrogen-bonded ribbons parallel to the *b*-axis and hydrogen-bonded linkages (involving the OH groups) to the adjacent ribbons along the *c*-axis. Adjacent corrugated slabs in the β polymorph are related by translation along the *a*-axis and nestle into each other through van der Waals interactions, with no hydrogen-bonding interactions between adjacent corrugated slabs.

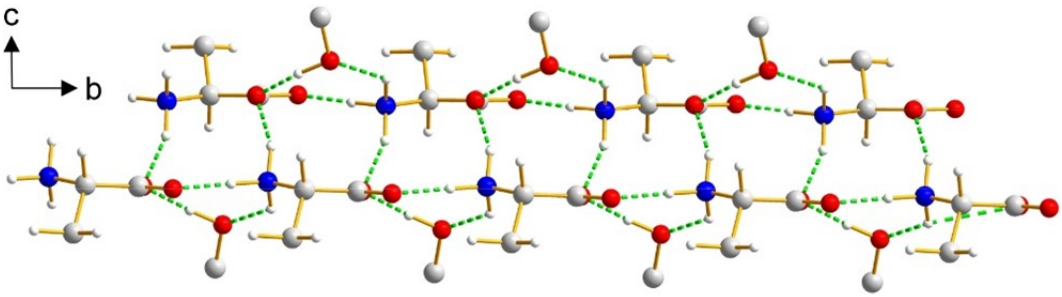

Figure 5.14: Hydrogen bonding between the ribbon containing the amino acid head-groups and the OH groups of the side-chains (from molecules above and below the ribbon shown) in the β polymorph of L-tyrosine. For clarity, only the $CH_2CH(NH_3^+)CO_2^-$ unit of each head-group and only the COH unit of each side-chain are shown. The two strands of the ribbon propagate along the *b*-axis (horizontal) and are offset slightly from each other along the *c*-axis (vertical). Hydrogen bonds are indicated by green dashed lines.

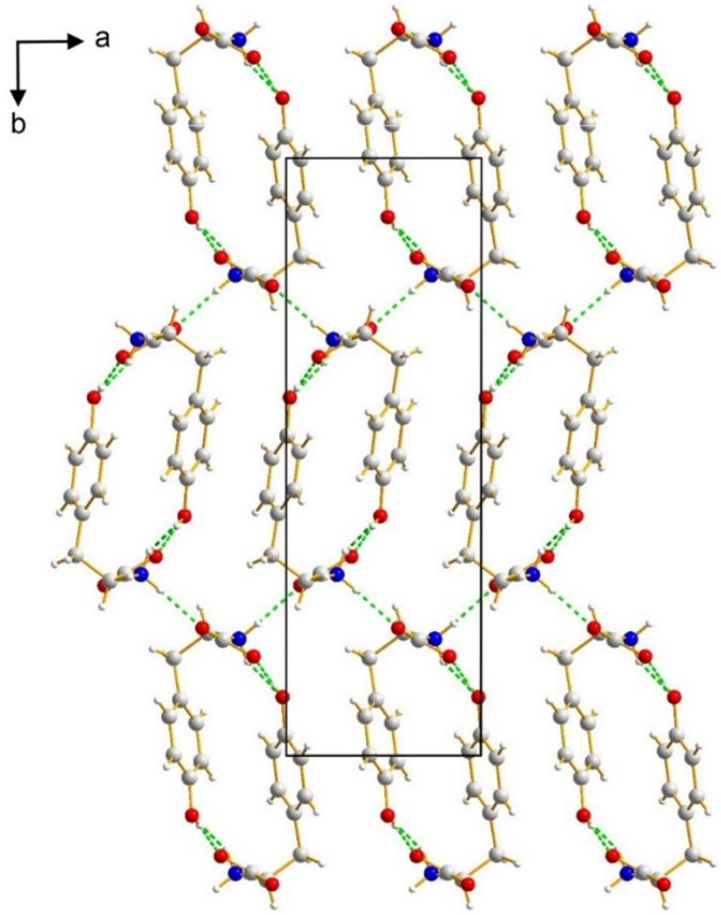

Figure 5.15: Crystal structure of the α polymorph of L-tyrosine (determined previously: A. Mostad, H. M. Nissen, C. Romming, *Acta Chem. Scand.*, **1972**, *26*, 3819-3833) viewed along the *c*-axis. The structure comprises alternating hydrophilic and hydrophobic layers parallel to the *ac*-plane. Hydrogen bonds are indicated by green dashed lines.

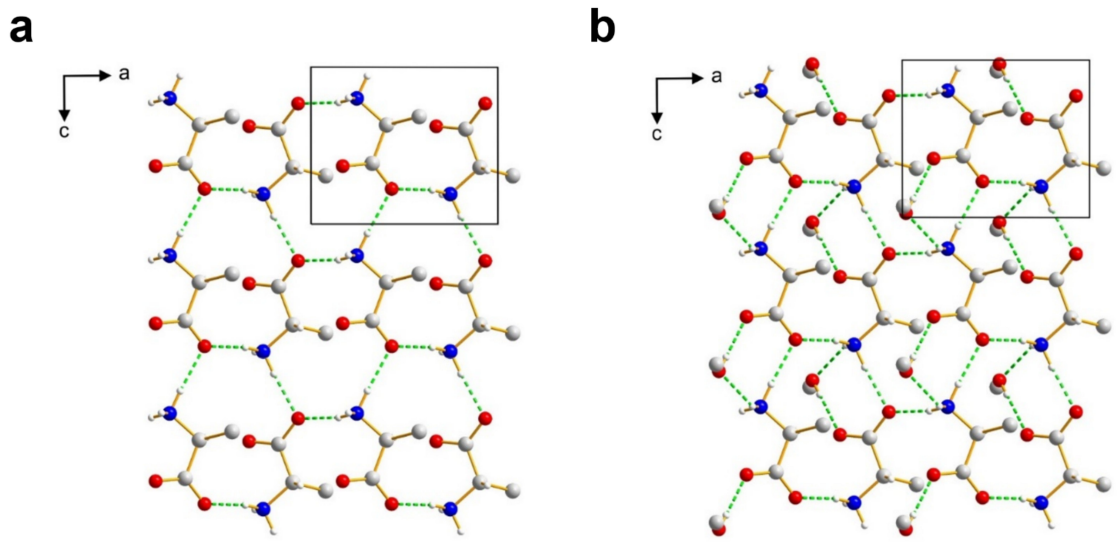

Figure 5.16: The two-dimensional hydrogen-bonding arrangement in the hydrophilic layer of the crystal structure of the α polymorph of L-tyrosine (determined previously: A. Mostad, H. M. Nissen, C. Romming, *Acta Chem. Scand.*, **1972**, *26*, 3819-3833) viewed along the *b*-axis, showing: (a) only the amino acid head-groups, and (b) both the amino acid head-groups and the OH groups of the side-chains. For clarity, only the $CCH(NH_3^+)CO_2^-$ unit of each head-group and the COH unit of each side-chain are shown. Hydrogen bonds are indicated by green dashed lines.

5.6 Crystal Structure Prediction of L-Tyrosine

To further explore the polymorphic landscape of L-tyrosine, structure prediction calculations were conducted using the AIRSS methodology [76]. Structure prediction calculations and some periodic DFT-D calculations were performed, and the data processed, by Prof Chris Pickard at the University of Cambridge. Dr Andrew Logsdail at Cardiff University performed calculations utilising the PBE0 methodology, shown in Table 5.5. The initial geometry optimization in AIRSS produced seven trial crystal structures for further consideration, with the energies of these structures given in Table 5.4. One pair of trial structures (ranked 1 and 3 in the initial energy ranking using PBE-TS) resembled the α polymorph and another pair of trial structures (ranked 2 and 6 in the initial energy ranking using PBE-TS) resembled the β polymorph.

The seven trial structures from AIRSS were then subjected to precise geometry optimization (using PBE-TS and including relaxation of the unit cell) in FHI-aims, and the results are summarized in Table 5.5. In the precise geometry optimization, trial structures 1 and 3 converged (with an energy difference of only 0.04 kJ mol^{-1}) on the same structure corresponding to the α polymorph, while trial structures 2 and 6 converged (with an energy difference of only 0.06 kJ mol^{-1}) on the same structure corresponding to the β polymorph. The good agreement between the crystal structure corresponding to the β polymorph from this calculation (specifically structure 6, which had slightly lower energy than structure 2) and the crystal structure of the β polymorph determined from

Structure from AIRSS	ΔE (PBE-TS) / kJ mol^{-1}
1	0.00
2	6.03
3	6.64
4	14.58
5	15.79
6	18.65
7	17.92

Table 5.4: Relative energy (ΔE, expressed per mole of L-tyrosine molecules) for the seven predicted crystal structures of L-tyrosine generated by AIRSS. For each crystal structure generated by AIRSS, the single-point PBE-TS energy was calculated using FHI-aims. The values of ΔE are given relative to crystal structure 1 (the structure of lowest energy generated by AIRSS).

Rietveld refinement of powder XRD is shown in Figure 5.17. Trial structures 4, 5 and 7 represent three distinct polymorphs that have not been reported previously in experimental studies and are hereafter denoted predicted structures A, B and C, respectively.

Structure (AIRSS)	Assignment	ΔE (PBE-TS) / kJ mol^{-1}	ΔE (PBE0-MBD) / kJ mol^{-1}
3	α polymorph	0.00	0.00
1	α polymorph	0.02	0.04
6	β polymorph	4.38	4.10
2	β polymorph	4.41	4.16
4	Predicted structure A	11.69	11.11
5	Predicted structure B	13.28	11.83
7	Predicted structure C	16.07	27.56

Table 5.5: Relative energy (ΔE, expressed per mole of L-tyrosine molecules) after subjecting the crystal structures of L-tyrosine generated by AIRSS to precise geometry optimization (including relaxation of unit cell parameters and nuclear coordinates) using PBE-TS in FHI-aims. After geometry optimization, structures 1 and 3 converge on an equivalent structure corresponding to the experimentally observed α polymorph, and structures 2 and 6 converge on an equivalent structure corresponding to the experimentally observed β polymorph. Results are presented for the more stable of these structures for each polymorph (i.e., structure 3 for the α polymorph and structure 6 for the β polymorph). Values of ΔE are expressed relative to the α polymorph (*i.e.*, structure 3). The single-point PBE0-MBD energy was also calculated using FHI-aims for each structure following the "precise" geometry optimization, giving the high-accuracy relative energies reported in the manuscript.

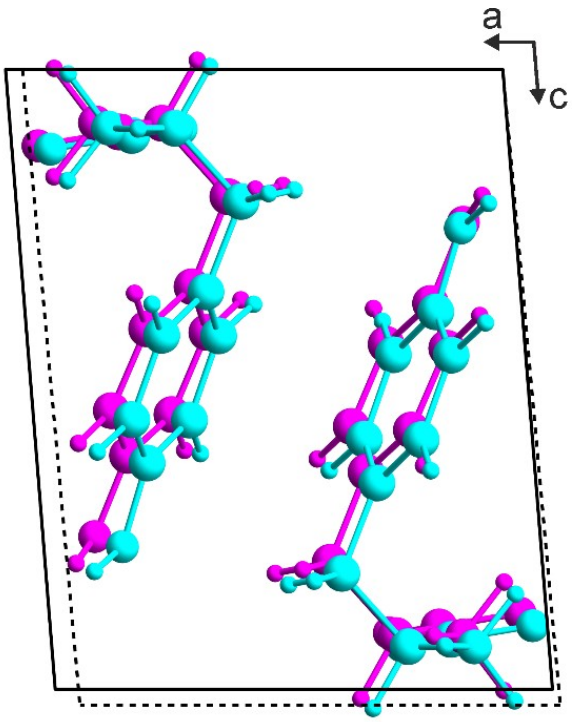

Figure 5.17: Overlay of the crystal structure of the β polymorph of L-tyrosine from the final Rietveld refinement (magenta; unit cell shown by the solid black lines) and the predicted crystal structure corresponding to the β polymorph generated by AIRSS followed by "precise" geometry optimization including unit cell relaxation (cyan; unit cell shown by the dashed black lines).

Thus, the seven initial trial structures from AIRSS were reduced to five distinct structures, representing five predicted polymorphs of L-tyrosine. Following the "precise" geometry optimization, the energy of each of the five polymorphs was then determined using a single-point PBE0-MBD calculation. The structure corresponding to the α polymorph is lowest in energy and the structure corresponding to the β polymorph is next lowest in energy. The calculated PBE0-MBD energies of each polymorph, relative to the energy of the α polymorph, are: 4.10 kJ mol^{-1} (β polymorph); 11.11 kJ mol^{-1} (predicted structure A); 11.79 kJ mol^{-1} (predicted structure B); 27.56 kJ mol^{-1} (predicted structure C). As predicted structures A, B and C are all significantly higher in energy than the α polymorph, it is unlikely that they represent experimentally accessible metastable polymorphs of L-tyrosine. However, the computational approach used here neglects entropic factors, which could have a significant influence on the relative energetic properties of the different structures. Crystallographic data for the five structures predicted by AIRSS (following "precise" geometry optimization) are summarized in Table 5.6.

Structure	Space group	Z	a / Å	b / Å	c / Å	α / °	β / °	γ / °	Volume / Å^3
α polymorph	$P2_12_12_1$	4	5.86	21.13	6.80	90	90	90	841.99
β polymorph	$P2_1$	2	7.41	5.90	9.87	90	95.15	90	429.76
Predicted structure A	$P2_12_12_1$	4	5.90	9.75	15.39	90	90	90	885.31
Predicted structure B	$P2_12_12_1$	4	5.92	12.08	11.88	90	90	90	849.58
Predicted structure C	$P2_12_12_1$	4	5.87	11.80	12.13	90	90	90	840.20

Table 5.6: Crystallographic data for the five distinct structures of L-tyrosine generated from the AIRSS structure prediction calculations, following "precise" geometry optimization (including relaxation of unit cell parameters). The structures labelled as the α polymorph and β polymorph correspond to the experimentally determined crystal structures of these polymorphs. The structures labelled as the predicted structures A, B and C have not been observed in experimental studies.

It is interesting to note that predicted structure A and predicted structure C (Figures 5.18 and 5.19 respectively) share several features in common with the β polymorph of L-tyrosine. In particular, these structure have a similar hydrogen-bonded ribbon motif comprising two strands constructed from amino acid head-groups, with N—H···O hydrogen bonding between adjacent molecules in a given strand and between molecules in the two strands, and the hydrogen-bonded ribbon is also engaged in hydrogen bonding with the OH groups of molecules involved in adjacent hydrogen-bonded ribbons. The contiguous hydrogen-bonded network in these structures comprises corrugated slabs of L-tyrosine molecules [with mean plane parallel to the *bc*-plane for the β polymorph and parallel to the *ab*-plane for predicted structures A and C], which interact with adjacent corrugated slabs through van det Waals interactions. The structures of the β polymorph and predicted structures A and C differ in the details of the topology and the relative arrangement of these corrugated slabs.

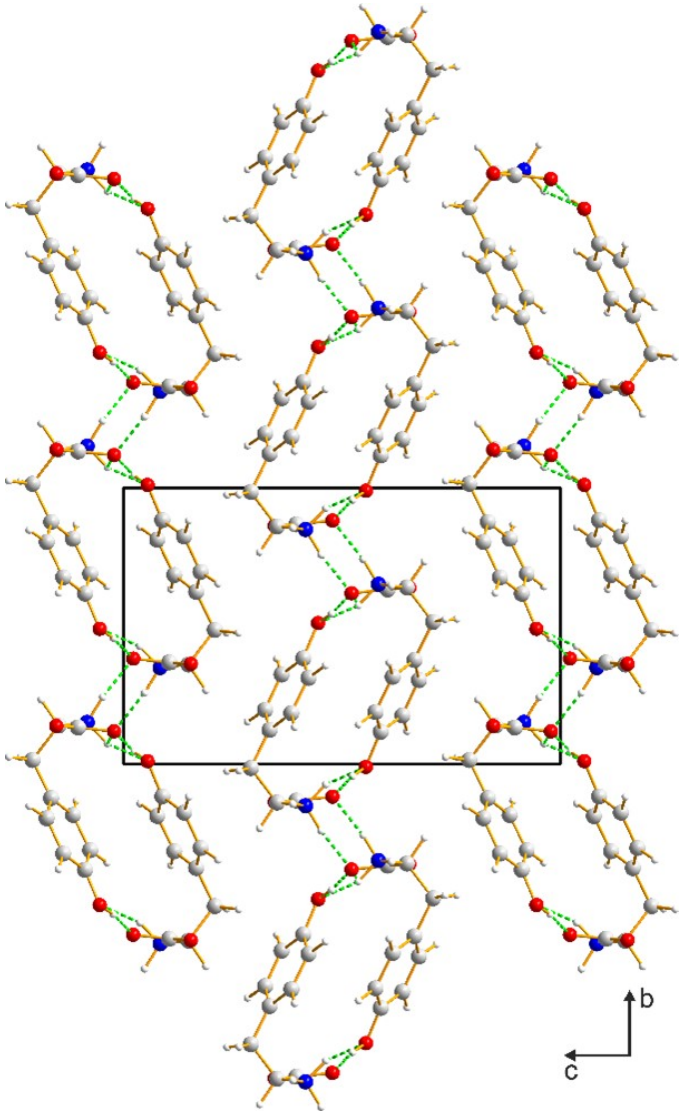

Figure 5.18: Predicted structure A of L-tyrosine (generated by AIRSS, followed by precise geometry optimization) viewed along the *a*-axis. Hydrogen bonds are indicated by green dashed lines.

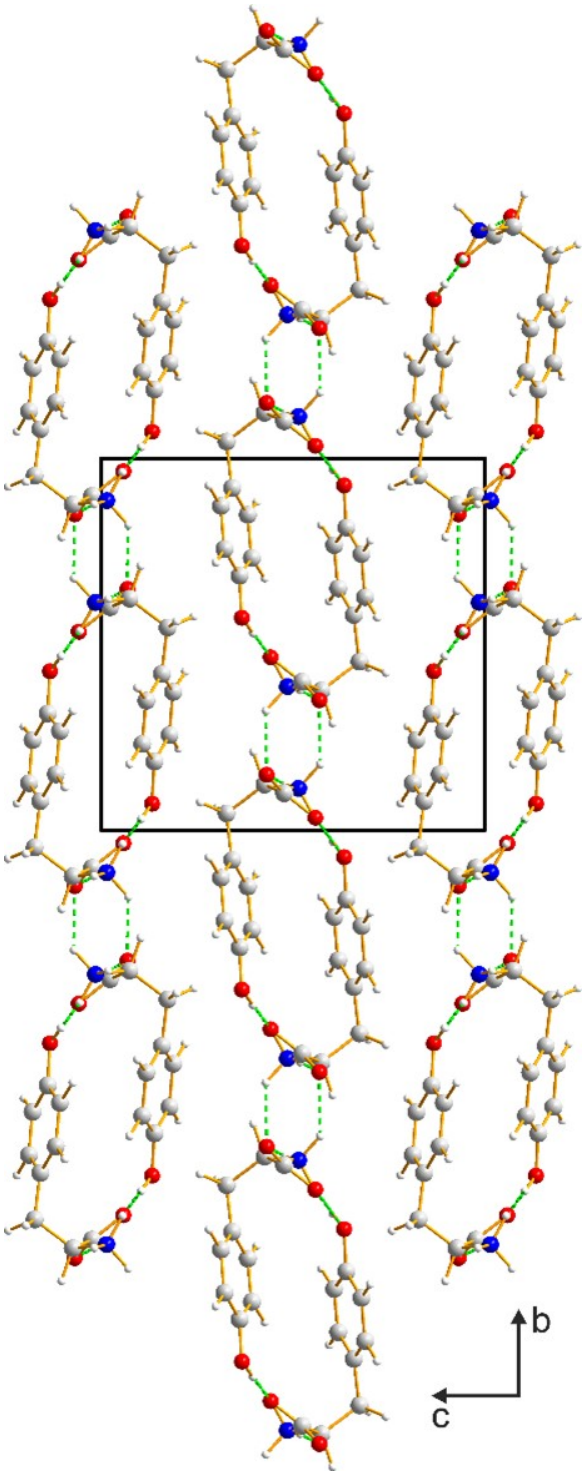

Figure 5.19: Predicted structure C of L-tyrosine (generated by AIRSS, followed by precise geometry optimization) viewed along the *a*-axis. Hydrogen bonds are indicated by green dashed lines.

In contrast, predicted structure B (shown in Figure 5.20) contains hydrogen bonded chains of L-tyrosine molecules linked by N—H···O interactions between amino acid head-groups (analogous to a single strand of the hydrogen-bonded ribbon motif observed in the β polymorph). Each hydrogen-bonded chain is also engaged in hydrogen bonding with the OH groups of molecules in an adjacent chain. In this structure, the contiguous hydrogen-bonded network comprises columns

of molecules (extending along the *a*-axis) that interact with adjacent columns through the van der Waals interactions.

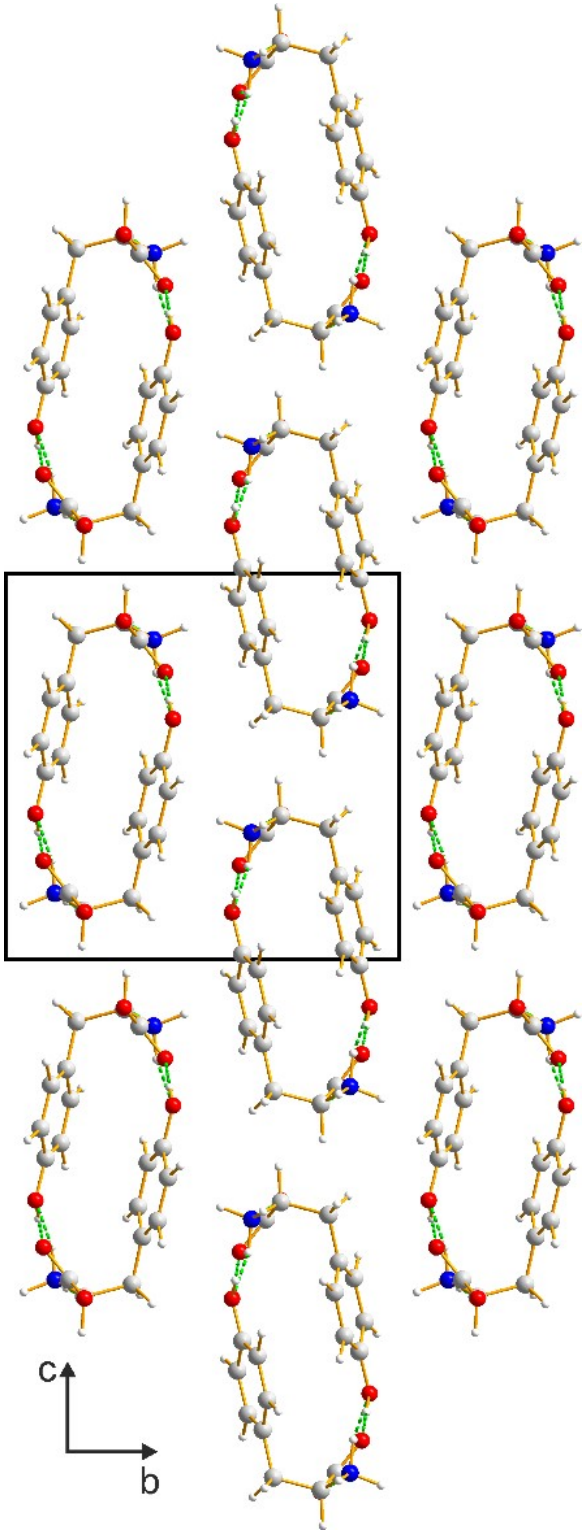

Figure 5.20: Predicted structure B of L-tyrosine (generated by AIRSS, followed by precise geometry optimization) viewed along the *a*-axis. Hydrogen bonds are indicated by green dashed lines.

5.7 Conclusions

From the work described in this Chapter, it is clear that the results obtained previously for crystallization of amino acids from the gas phase [163] are reproducible, at least in the case of L-tyrosine. The β polymorph was produced using crystallization directly from the gas phase via sublimation, and this is common with the study of the polymorphism of L-tryptophan [36]. Clearly there is scope to continue this work with a wide range of organic molecules, however it would be wise to begin with amino acids due to their biological importance.

The crystal structure of the β polymorph was determined directly from powder XRD data using the direct-space approach, however at several stages of the structure determination process other techniques proved invaluable. In the indexing stage, due to the presence of an impurity phase, difficulties were encountered, which can be a common issue when attempting to determine the unit cell from powder XRD data. The use of 3D-ED data to independently determine a good quality unit cell was key to the successful structure determination in this case. Subsequently it was shown that the crystal structure could also be solved from the 3D-ED data using the direct-space GA approach, which was also used for successful structure solution from the powder XRD data. The opportunity to use 3D-ED data in this way may be invaluable in the future when powder XRD data are of low quality, for example if the intensities are affected heavily by preferred orientation, such that successful structure solution cannot be achieved. The use of solid-state NMR data (both experimental and calculated from periodic DFT-D methods) to validate the final refined crystal structure of the β polymorph is an important aspect of the structure determination process. Finally, the use of periodic DFT-D calculations was invaluable in ensuring that the final refined crystal structure of the β polymorph corresponded closely to an energy minimum.

Following the crystal structure determination on the β polymorph of L-tyrosine, periodic DFT-D calculations at varying levels of theory were conducted to assess the relative energies of the β and α polymorphs of L-tyrosine. Following this energetic assessment, focusing in particular on the results from the PBE0-MBD calculations, it is clear that the β polymorph is higher in energy than the α polymorph. Crystal structure prediction calculations using the AIRSS methodology generated five unique, plausible polymorphic crystal structures for L-tyrosine, two of which correspond to the known α and β polymorphs. Among these five structures, the lowest energy structure is the α polymorph, followed by the β polymorph. The other three predicted structures are significantly higher in energy than the β polymorph. High-pressure studies could also be conducted on L-tyrosine in the quest to make these polymorphs experimentally, as in the case of the δ and ϵ polymorphs of glycine, discussed in Section 5.1.

Chapter 6
DL-Lysine: Determination of The Crystal Structures of Anhydrate, Hemihydrate and Monohydrate Forms

6.1 Introduction

Lysine is one of the 20 directly encoded proteinogenic amino acids with the molecular structure of its L-enantiomer shown in Figure 6.1. Since the publication of the crystal structure of L-lysine in 2015 [34], all 20 directly encoded proteinogenic amino acids now have at least one crystal structure reported, for the naturally occurring enantiomerically pure form. In addition to the structural properties of enantiopure amino acids in the solid-state, the crystal structures of their racemates can provide additional structural insights. Crystallization of a racemic sample of a chiral material may form either a racemic crystalline phase (containing both enantiomers), or a conglomerate, in which the two enantiomers crystallize separately in enantiopure crystalline phases of opposite chirality. Following the publication of the structure of DL-arginine in 2017 [35], there remained four amino acids for which a racemate crystal structure had yet to be reported: DL-asparagine, DL-phenylalanine, DL-threonine, and DL-lysine. However, in this work the crystal structure of DL-lysine has been determined, leaving DL-asparagine, DL-phenylalanine, and DL-threonine left for their crystal structures to be determined.

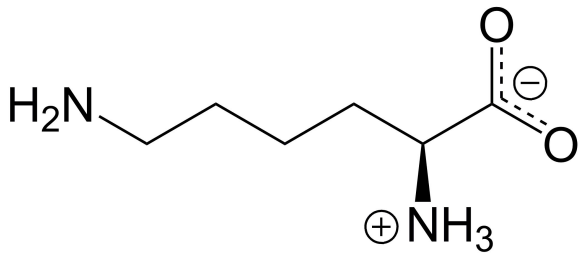

Figure 6.1: The molecular structure of L-lysine.

When analysing racemate structures, Wallach's rule is an interesting consideration [174]. In 1895, Otto Wallach proposed that a racemic crystal structure tends to have a higher density than the corresponding enantiopure crystal structure. While Wallach's rule is often obeyed and is supported by extensive comparison of racemic crystal structures and their enantiopure counterparts, it is more

of a general trend than a hard rule. In the case of the directly encoded proteinogenic amino acids, several exceptions to Wallach's rule exist. In fact, the crystal structures of several of the enantiopure amino acids have higher density than their racemic counterparts, as shown in Table 6.1.

Three crystal structures containing L-lysine in anhydrous or hydrate forms have been reported, specifically anhydrous (CSD refcode CUFFUG) [34], hemihydrate (CSD refcode UPUKUN) [37] and monohydrate (CSD refcode UPUKOH) [37] forms. The anhydrous and hemihydrate crystal structures (shown in Figure 6.2) share similar packing and hydrogen bonding schemes, both containing a set of strong hydrogen bonds between the ammonium and carboxylate groups. These structures differ in the hydrogen bonding, involving the terminal amino groups on the side-chains. In the hemihydrate structure, the water molecule engages in hydrogen bonding with the amino groups of the side-chains, whereas the anhydrate structure contains hydrogen bonded chains of N—H$\cdots$N between the amino groups of the side chains of the neighbouring molecules. The main feature that distinguishes the monohydrate structure from the anhydrate and hemihydrate structures is the formation of an extended hydrogen-bonded network in which the amino acid head-groups, the NH_2 groups of the side-chains, and the water molecules interact together, as shown in Figure 6.2. In the present work, the crystal structure of DL-lysine is determined, alongside both hemihydrate and monohydrate form.

Amino acid	Racemate ρ / g cm^{-3}	Enantiopure ρ / g cm^{-3}	Wallach's Law observed?
Alanine	1.389	1.367	Yes
Arginine	1.291	1.340	No
Aspartic acid	1.644	1.643	Yes
Cysteine	1.478	1.505(α)	No
Glutamic acid	1.593	1.532(α)	Yes
Glutamine	1.364	1.525	No
Histidine	1.522	1.441(α)	Yes
Isoleucine	1.237	1.203(α)	Yes
Leucine	1.191	1.166(α)	Yes
Methionine	1.341(β)	1.350	No
Proline	1.385	1.412(α)	No
Serine	1.512	1.544(α)	No
Tyrosine	1.436	1.414	Yes
Valine	1.331	1.262(α)	Yes

Table 6.1: A range of amino acids with their respective densities (for amino acids that have a crystal structure for both the racemate and enantiopure material), whether they obey Wallach's rule. All densities reported are for crystal structures determined using data collected at room temperature. Asparagine, phenylalanine and threonine have no racemate structure in the Cambridge Structural Database; up until this work, lysine had no racemate crystal structure available. Tryptophan is omitted as the only structure for the racemate contained in the Cambridge Structural Database is at low temperature, and therefore density comparisons are unreliable. The values in brackets next to density values represent the polymorph of that material to which the value corresponds to.

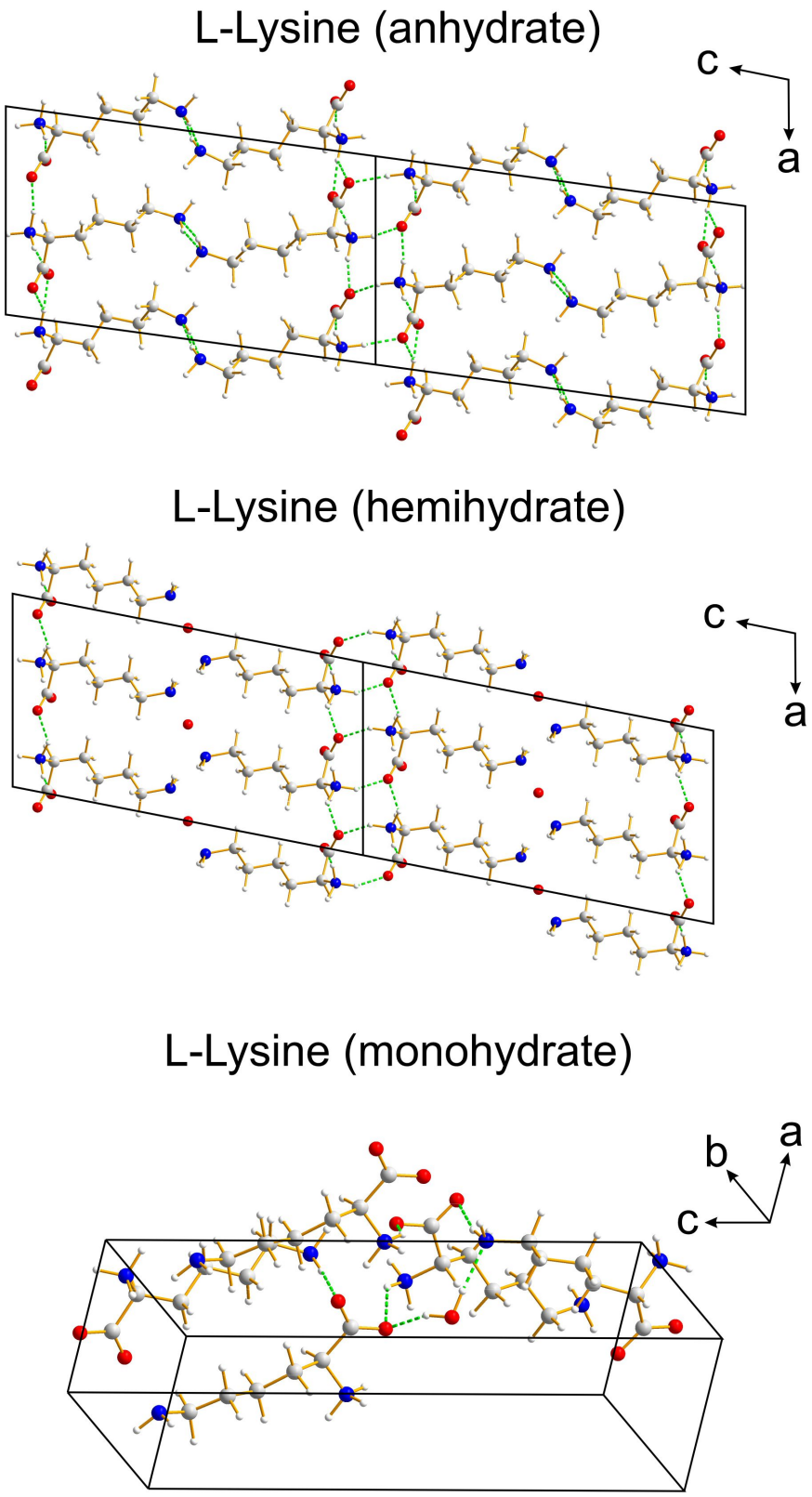

Figure 6.2: The structures of three known phases of L-lysine: anhydrate (top), hemihydrate (middle), monohydrate (bottom).

In this chapter, the crystal structures of the anhydrous phase of DL-lysine, as well as the hemihydrate and monohydrate phases of this material, have been determined primarily through the analysis of powder XRD data. These structures have been verified, and in some cases the possibility for disorder probed, through the use of periodic DFT-D calculations. This chapter explores how periodic DFT-D calculations can be a useful tool for exploring the feasibility of a disordered crystal structure, and for determining the most favourable hydrogen bonding arrangement specifically in the case of molecular organic materials.

6.2 Experimental Methods

6.2.1 Sample Preparation

A sample of DL-lysine was purchased from Alfa Aesar as an off-white microcrystalline powder, which was later determined to be the hemihydrate phase. To prepare the anhydrous phase, a powder sample of the hemihydrate phase was packed into a 0.7 mm glass capillary and placed in an oven at 80 °C for 16 hr, after which it was removed and immediately flame sealed to prevent rehydration. The monohydrate phase was prepared by subjecting a powder sample of the hemihydrate phase to a high humidity environment for 7 days (at room temperature) by placing a small amount of DL-lysine hemihydrate in a glass vial, within a larger sealed glass vial containing water.

6.2.2 Powder X-ray Diffraction

Powder XRD data were recorded in transmission mode at 293 K on a Bruker D8 Advance Series II diffractometer (Ge-monochromated Cu-Kα_1 radiation) between 3.5° and 70° in 2θ (step size; 0.017°) using a Vantec detector covering 3° in 2θ. The polycrystalline sample of the anhydrous phase of DL-lysine was packed into a 0.7 mm glass capillary (flame sealed) to prevent hydration before being placed on a foil type sample holder. The powder sample of the monohydrate phase was also packed into a 0.7 mm glass capillary after mixing with amorphous starch to reduce the effects of preferred orientation in the powder sample (a two-dimensional powder XRD pattern recorded for the same material (Figure 6.3) showed evidence for significant preferred orientation). The powder sample of the hemihydrate phase was loaded between two pieces of tape and then placed on a foil type sample holder. Data collection times for recording the powder XRD data were 17 hr for each sample.

To assess the degree of preferred orientation in the powder sample, two-dimensional powder XRD data were recorded at ambient temperature on an Agilent SuperNova diffractometer using Cu-K$_\alpha$ radiation and an Atlas detector. The instrument was configured with the two-dimensional detector perpendicular to the incident beam direction, with a sample-to-detector distance of 100 mm. The

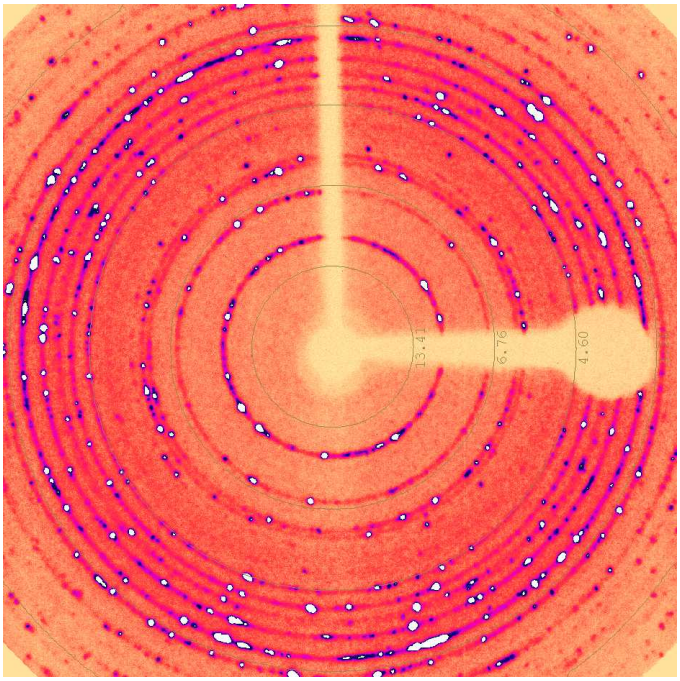

Figure 6.3: Two-dimensional powder XRD data recorded for the monohydrate phase of DL-lysine, providing evidence for preferred orientation within the sample.

capillary containing the powder sample was oriented with the capillary axis perpendicular to the incident X-ray beam. Data were collected for 10 s, with sample rotation through an angular range of 10° about the capillary axis during data collection.

6.2.3 Structure Solution

The crystal structures of the anhydrous, hemihydrate and monohydrate phases of DL-lysine were solved using the same methodology. Structure solution was conducted from powder XRD data using the EAGER program, which implements a genetic algorithm search procedure within the direct-space strategy [133–135]. In the structural models for calculations, the molecular geometry of lysine was based on that in the crystal structure of L-lysine (CSD Refcode: CUFFUG) reported previously [34]. For the space groups *I*2/a (for the anhydrate and hemihydrates) and $P2_1/n$ (for the monohydrates), the asymmetric unit comprised one molecule of L-lysine (with the molecule of D-enantiomer being generated by the space group symmetry). For the structure solution calculation of the monohydrate there was one molecule of water included in the calculation, whereas for the hemihydrate structure solution calculation there was one molecule of water in the calculation where the occupancy of all atoms in the water molecule were 0.5. The molecular structures of D-lysine and L-lysine are shown in Figure 6.4. In the GA structure solution calculations, the population size was 100, with 10 mating operations and 50 mutation operations per generation, and the evolution of the population was conducted for 100 generations. In the structural models for the hemihydrate and monohydrate phases, the water molecule (with hydrogen atoms included) was generated

using standard bond lengths and angles. The L-lysine molecule in the asymmetric unit was in the zwitterionic form (Figure 6.4), as all amino acids exist in this form in the solid-state under ambient conditions. The $-NH_2$ and $-NH_2^+$ groups were not allowed to rotate as the low scattering power of the hydrogen atoms in these groups has a negligible effect on the powder XRD pattern, while the introduction of the two additional torsion-angle variables would increase the computational cost. The number of structural variables for the structure solution calculation for the anhydrous phase of DL-lysine included 3 positional variables, 3 orientational variables and 5 torsional variables, as seen in Figure 6.4, leading to a total of 11 structural variables for optimization. For the hemihydrate and monohydrate phases, there were an additional 6 structural variables due to the addition of a water molecule in each calculation, which had 3 positional and 3 orientational variables in both calculations, leading to 17 structural variables for optimization in both the hydrated phases of DL-lysine.

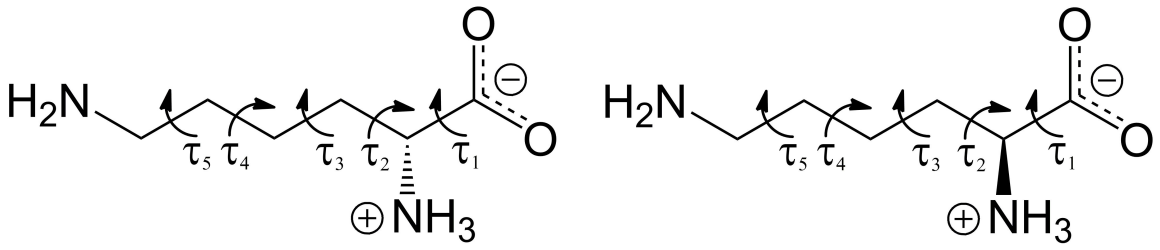

Figure 6.4: Molecular structures of D-lysine (left) and L-lysine (right) in the zwitterionic form. The variable torsion angles in the direct-space structure solution calculations from powder XRD data are indicated.

6.2.4 Periodic DFT-D Calculations

Periodic DFT-D calculations used in conjunction with the determination and validation of the experimental crystal structures of the anhydrous, hemi-hydrate and monohydrate phases of DL-lysine were conducted using CASTEP [122] (Academic Release Version 8.0). The geometry optimization calculations used ultrasoft pseudopotentials [123], semi-empirical dispersion corrections [125] (TS correction scheme), PBE functional [124], fixed unit cell and preserved space group symmetry, periodic boundary conditions, a basis set cut-off energy of 700 eV and a Monkhorst-Pack grid [126] of minimum sample spacing (0.05 x 2π) $Å^{-1}$. Convergence criteria for geometry optimization were 0.01 eV $Å^{-1}$ for forces, 0.00001 eV per atom for energy, and 0.001 Å for atomic displacements.

6.3 Crystal Structure Determination

6.3.1 Anhydrous Phase of DL-Lysine

The powder XRD pattern of the anhydrous phase of DL-lysine was indexed using the LZON algorithm within the CRYSFIRE package [89, 143], which gave the following unit cell with monoclinic

metric symmetry (space group *I*2/a used): a = 9.84 Å, b = 4.74 Å, c = 34.72 Å and β = 94.55° (V = 1614.5 Å^3). Unit cell refinement and profile fitting were conducted using the Le Bail method in the GSAS program [96, 132], which gave an excellent fit to the powder XRD data (R_{wp} = 3.77%, R_p = 2.83%), shown in Figure 6.5. Based on systematic absences, the space group was determined to be *I*2/a.

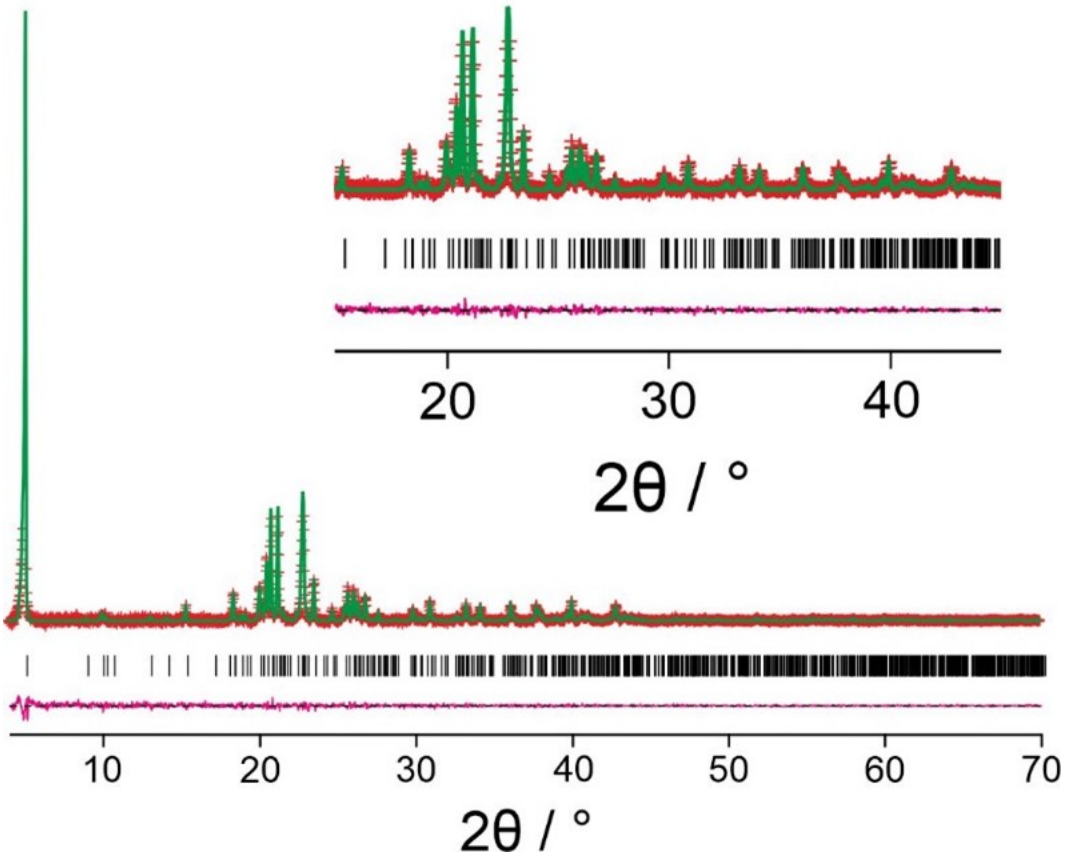

Figure 6.5: Profile fit using the Le Bail method for the anhydrate phase of DL-lysine (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Structure solution was conducted using the direct-space genetic algorithm technique, as described in Section 6.2.3. Following density considerations of the unit cell provided by Le Bail fitting, one molecule of L-lysine was incorporated in the asymmetric unit. The structure with lowest R-factor (R_{wp}) was structurally sensible and gave a good fit to the experimental powder XRD data, and was used as the initial structural model for Rietveld refinement, which was conducted using the GSAS program [132]. Soft restraints were applied to maintain sensible bond lengths and bond angles, and a planar restraint was applied to the carboxylate group and α-carbon atom. A common isotropic displacement parameter was refined for the non-hydrogen atoms, with the value for hydrogen atoms set to 1.2 times the refined value for the non-hydrogen atoms. The initial Rietveld refinement showed a reasonable fit to the powder XRD data (R_{wp} = 4.17%, R_p = 3.08%).

Inspection of the structure suggested the possibility of several favourable hydrogen bonding

schemes involving neighbouring NH_2 groups from the lysine side chains, which are related by the crystallographic two-fold rotation axis. DFT-D geometry optimization calculations were conducted in the lower symmetry Ia space group to investigate the potential for alternative hydrogen bonding arrangements to exist (for example, as ordered components within a disordered description of the structure in *I*2/a). As the —NH_2 groups of the side chains are related by the two-fold rotation axis, they cannot move independently of one another, therefore the symmetry must be lowered to Ia to allow them to move independently (e.g. to form local hydrogen-bonding arrangements that do not present the two-fold rotation axis). Three calculations were conducted, one in which the —NH_2 group of the side chain remained unchanged, and two others in which the NH_2 group is rotated by 120° and 240°, respectively as shown in Figure 6.6.

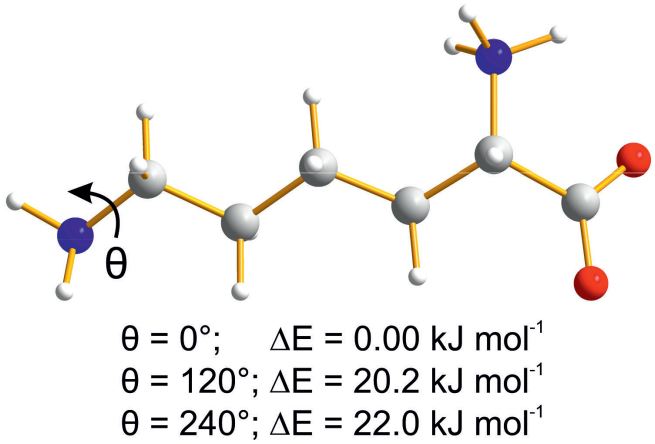

Figure 6.6: The local structure of the L-lysine molecule in the anhydrous crystal structure, with the rotation of the —NH_2 shown by the symbol θ with the relative energies for the three rotations given under the molecule.

The optimized structure with no rotation of the NH_2 group gave the lowest energy, while the two other structures gave energies that are higher by 20.2 kJ mol^{-1} (120°) and 22.0 kJ mol^{-1} (240°). As the two structures containing the rotated —NH_2 groups have significantly greater energy than the structure with the original conformation, it is concluded that the structure is not disordered. For the final Rietveld refinement, the ordered model of the DL-lysine structure in space group *I*2/a was used.

In the final Rietveld refinement, additional restraints were applied to preserve the hydrogen bonding geometries, based on those in the structure following DFT-D geometry optimization. A good fit to the powder XRD data (Figure 6.7) was obtained (R_{wp} = 4.49%, R_p = 3.30%), comparable to the quality of fit in the Le Bail profile fitting, with the final refined parameters: a = 9.8395(5) Å, b = 4.73838(17) Å, c = 34.7211(17) Å and β = 94.566° (V = 1613.67(12) $Å^3$). DFT-D geometry optimization of the final refined structure showed only minimal atomic displacements, with RMSD of 0.14 Å for non-hydrogen atoms.

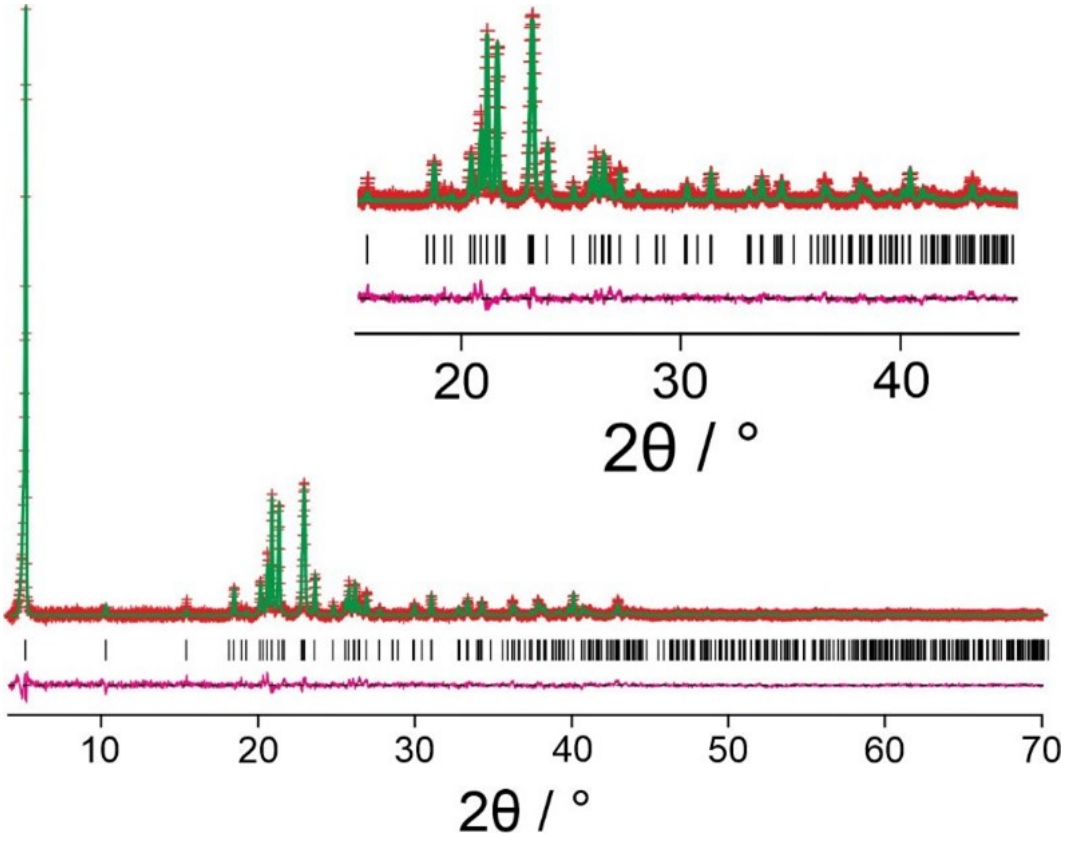

Figure 6.7: Final Rietveld refinement for the anhydrate phase of DL-lysine (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

6.3.2 Hemihydrate Phase of DL-Lysine

The powder XRD pattern of DL-lysine hemihydrate was indexed using the KOHL algorithm in the CRYSFIRE package [89, 175], giving the following unit cell with monoclinic metric symmetry: a = 9.85 Å, b = 4.73 Å, c = 36.32 Å and β = 95.38° (V = 1683.9 Å^3). Profile fitting and unit cell refinement were conducted using the Le Bail method in the GSAS program [96, 132], giving a good fit to the powder XRD data (R_{wp} = 3.77%, R_p = 2.83%), although with a small discrepancy in the peak shape at approximately 2θ = 10° (shown in Figure 6.8). Based on consideration of systematic absences, the space group was assigned as *I*2/a. Density considerations of this unit cell suggests that for one molecule of L-lysine in the asymmetric unit with the space group *I*2/a applied, the density is approximately 1.2 g cm^{-3} which is in the range anticipated for a typical molecular organic material.

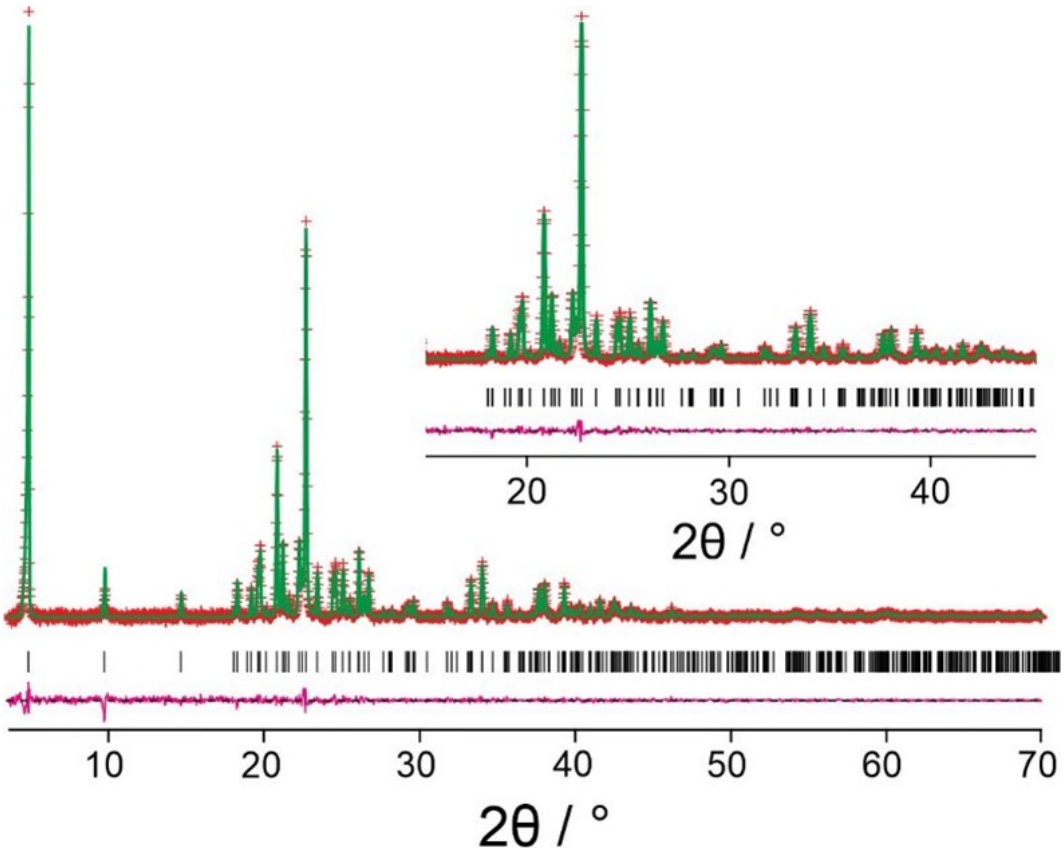

Figure 6.8: Profile fit using the Le Bail method for the hemihydrate phase of DL-lysine (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Three separate structure solution calculations were conducted using the direct-space genetic algorithm technique, corresponding to different lysine:water stoichiometries. Specifically, the calculations included zero, half and one water molecule, respectively, per molecule of lysine. The calculation with one water (with a site occupancy of 0.5 for every atom in the water molecule) per molecule of lysine gave the best fit (R-factor of approx. 15.1%, as opposed to values of 19% for no water, and 17% for one water) whilst remaining structurally reasonable. The asymmetric unit in this calculation comprised one molecule of L-lysine (with D-lysine generated by the space group symmetry) and one half molecule of water, with the oxygen atom located on the two-fold rotation axis.

The structure selected from the structure solution calculations (as described previously) was used as the starting model for a Rietveld refinement using the GSAS program [132]. Soft restraints were applied to maintain sensible bond lengths and angles during the refinement, and a planar restraint was applied to the carboxylate group and α-carbon of the lysine molecule. A common isotropic displacement parameter was refined for non-hydrogen atoms, and the value for hydrogen atoms was taken as 1.2 times the value of non-hydrogen atoms. The initial Rietveld refinement gave a reasonable fit to the powder XRD data (R_{wp} = 4.17%, R_p = 3.08%).

Inspection of the crystal structure of DL-lysine hemihydrate suggested that there may be potential

for multiple favourable hydrogen bonding schemes involving the —NH_2 groups of the side-chains and the water molecule. When viewing the crystal structure without consideration of the hydrogen atoms on the —NH_2 side-chain or the water molecule (as seen in Figure 6.9), it is clear to see that the two adjacent —NH_2 groups are related by the *a*-glide (translation by *a*/2 with reflection in the plane of the page. There are also —NH_2 groups generated by the symmetry of the inversion present in this space group symmetry. The oxygen atom for the water molecule is then present at the position of the two-fold axis of rotation, which is perpendicular to the page. It is from this view that various hydrogen-bonding schemes can be considered.

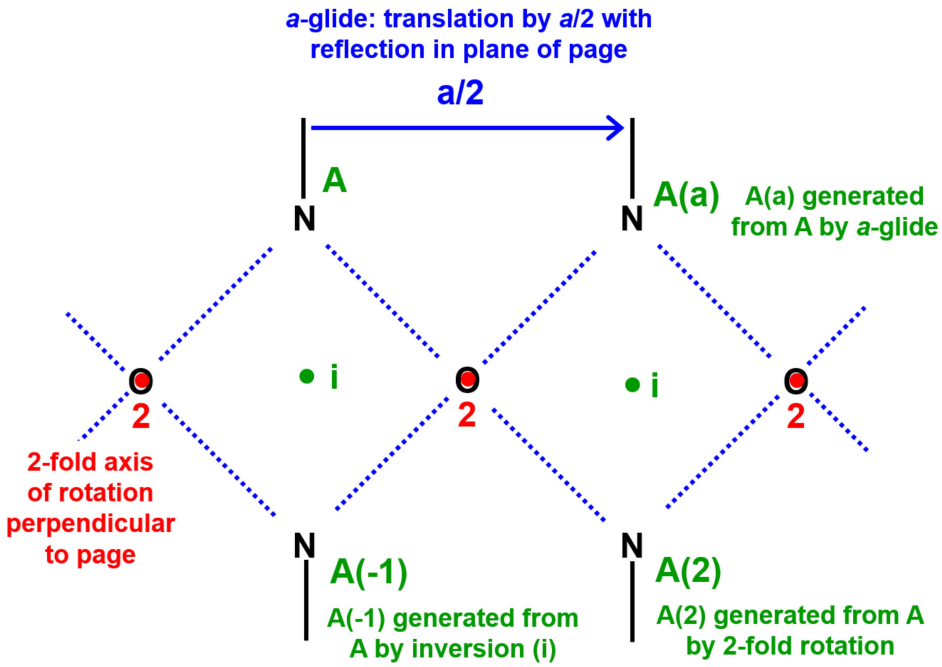

Figure 6.9: Diagram of the hemihydrate crystal structure from the structure solution calculation specifically on the region of the —NH_2 and H_2O interaction region (with hydrogens omitted), viewed along the *b*-axis.

The first arrangement (arrangement I) exists in the I2/a space group and can be observed in Figure 6.10. In this arrangement both O—H groups of the water molecule are not hydrogen bond donors, although both N—H of the —NH_2 groups of the side-chain are in fact hydrogen bond donors. The second arrangement (arrangement II) exists in the space group Ia, where the two-fold axis is not present. In this arrangement, both of the O—H groups of the water molecule are hydrogen bond donors to the —NH_2 groups of the side-chain. This —NH_2 group has one N—H group that is a hydrogen bond donor to the water molecule. The third arrangement (arrangement III) exist also in the Ia space group, similarly where there is no two-fold axis. In this arrangement, both of the O—H groups of the water molecule are hydrogen bond donors to the —NH_2 groups of the side-chain. This —NH_2 group has one N—H that group is a hydrogen bond donor to the water molecule.

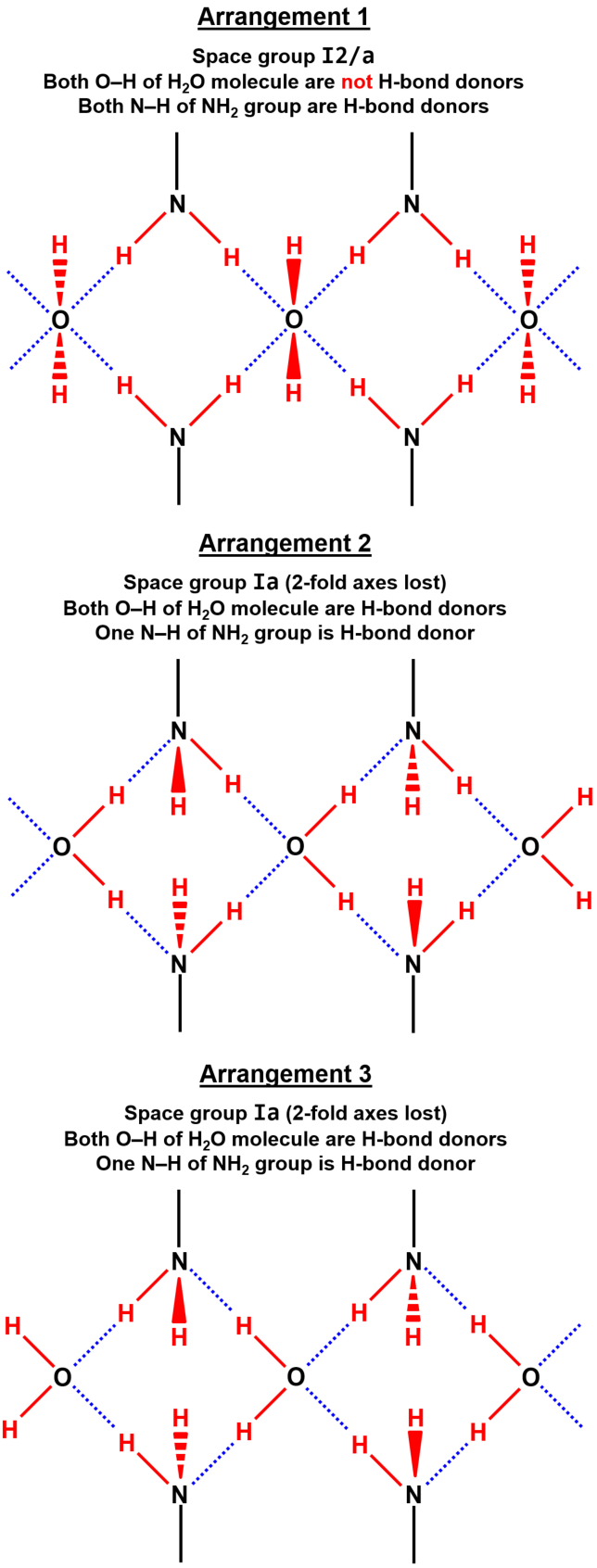

Figure 6.10: Diagram of the three arrangements described for the different hydrogen bonding arrangements possible for the hemihydrate of DL-lysine.

There are two possible combinations of these arrangements, both of which can be seen in Figure 6.11. One combination includes both arrangement II and arrangement III and, in this case, they are related to each other by rotation about the two-fold axis (existing in space group *I*2/a). Here, the average of the two arrangements must exist in the *I*2/a space group where the hydrogen atoms from the $-NH_2$ groups and water molecules have a site occupancy of 0.5 each. The other combination of multiple arrangements is the combination of arrangement I, II and III. In this combination, there is a fraction (1-*x*) of arrangement I, fraction *x*/2 of arrangement II and fraction *x*/2 of arrangement III, where the hydrogen atom occupancies in the averaged structure (in the *I*2/a space group) are shown in Figure 6.11.

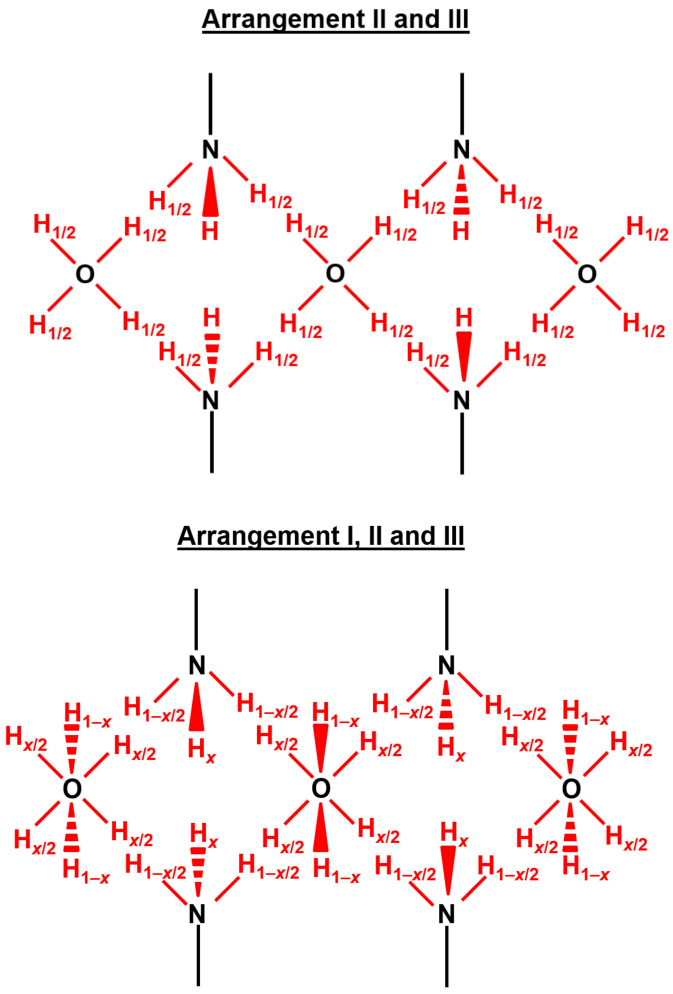

Figure 6.11: Diagram of the two combinations of the three different hydrogen bonding arrangements hypothesized for the hemihydrate of DL-lysine.

To compare between these series of arrangement and decide whether the structure is disordered, and if it is then which combination of arrangements are present, periodic DFT-D calculations were employed to independently assess the relative energies of these arrangements. From the relative energies reported in Table 6.2 it can be seen that arrangement I has a much higher relative energy than either arrangement II or arrangement III which would make the possibility of it being involved

in a disordered structure unlikely. The relative energies for arrangement II and III are so similar, that they are within the margin of error for PBE-TS methodology, it is reasonable to consider a structural model for the DL-lysine hemihydrate in which the two local hydrogen-bonding arrangements are present as a disordered structure, with 50% occupancy of each arrangement. The combination of the two ordered components, each with 50% occupancy in space group Ia, generates the resultant disordered structure in space group *I*2/a observed in Figure 6.11.

Arrangement	ΔE (PBE-TS) / kJ mol^{-1}
I	28.59
II	0.00
III	1.67

Table 6.2: The relative energies of each potential hydrogen bonding arrangement for the hemihydrate phase of DL-lysine. As discussed by Hoja *et al*. [176] the relative energies of polymorphs ranked by the PBE-TS methodology can vary by up to 10 kJ mol^{-1}, which validates the discarding of arrangement I as a plausible structure as it sits outside this range. It also allows both arrangements II and III to be considered in a disordered structure, as their energies are much closer than the typical margin of error for a PBE-TS calculation. Other periodic DFT-D calculations such as the PBE0-MBD type calculation are the most comprehensive and thus give a more reliable energetic ranking, however the most reliable energetic ranking of organic materials by periodic DFT-D calculations is considered the PBE0-MBD+F_{vib} where F_{vib} is used to take into account the vibrational contribution to the free energy

In the final Rietveld refinement, the two hydrogen-bonding arrangements were included with 50% occupancy in the *I*2/a space group. The NH_2 group of the lysine side chain has a different orientation in arrangements I and II. One N—H bond (containing H_{23}) is in virtually the same orientation in both arrangements, while the other N—H bond is in a different orientation in the two arrangements (containing H_{24} in arrangement I with 50% occupancy and containing H_{30} in arrangement II with 50% occupancy). The water molecule is in a different position and orientation in each arrangement, with occupancy of 50% in each arrangement, with the water molecules in the two arrangements related to each other by the two-fold rotation axis of space group *I*2/a. Standard restraints were applied as in the initial Rietveld refinement, with additional restraints applied to geometric features of the hydrogen bonding arrangements. The refinement gave a good quality of fit to the powder XRD data (R_{wp} = 4.49%, R_p = 3.30%) (Figure 6.12, comparable to the quality of fit obtained in the Le Bail profile fitting with the following final refined parameters: a = 9.8395(5) Å, b = 4.73838(17) Å, c = 34.7211(17) Å and β = 94.566° (V = 1684.12(11) Å^3). Due to the disordered nature of the structure, geometry optimization on the whole, averaged structure was not possible.

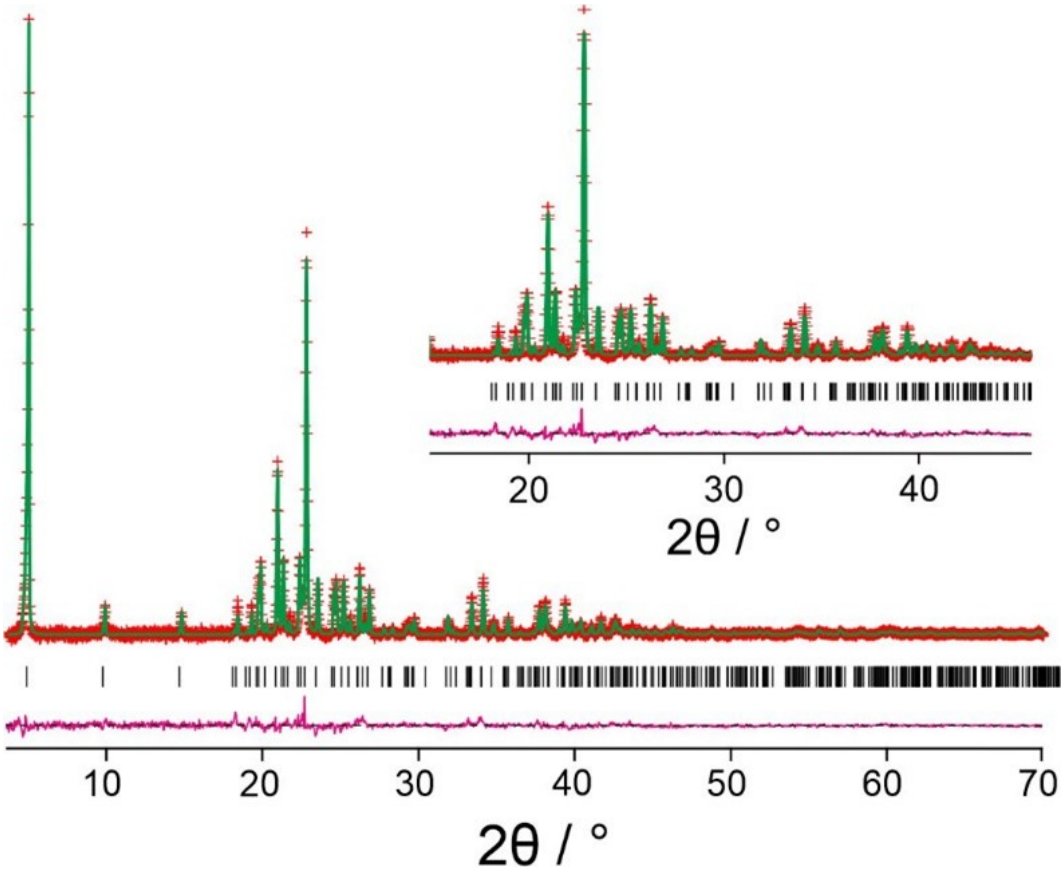

Figure 6.12: Final Rietveld refinement for the hemihydrate phase of DL-lysine (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

6.3.3 Monohydrate Phase of DL-Lysine

The powder XRD pattern of DL-lysine monohydrate was indexed using the ITO algorithm within the CRYSFIRE package [86, 89], giving the following unit cell with monoclinic metric symmetry: a = 5.88 Å, b = 7.51 Å, c = 19.86 Å and β = 95.31° (V = 872.8 Å^3). Profile fitting and unit cell refinement were conducted using the Le Bail method in the GSAS program [96, 132], resulting in a good fit to the powder XRD data (R_{wp} = 3.88%, R_p = 2.90%) with the exception of a peak shape issue for the peak at 2θ = 9°. The Le Bail fit is shown in Figure 6.14. The space group was determined to be $P2_1/n$ based on consideration of systematic absences. Density considerations for four molecules of lysine in the unit cell gave an approximate density of 1.11 g cm^{-1}, which is reasonable for a molecular organic material. When considering that the mass loss observed in the TGA experiment (seen in Figure 6.13) is 10.64%, it is clear that there should be 1:1 ratio of lysine molecules to water molecules in this structure as the theoretical mass loss for 1 molecule of water is 10.97%.

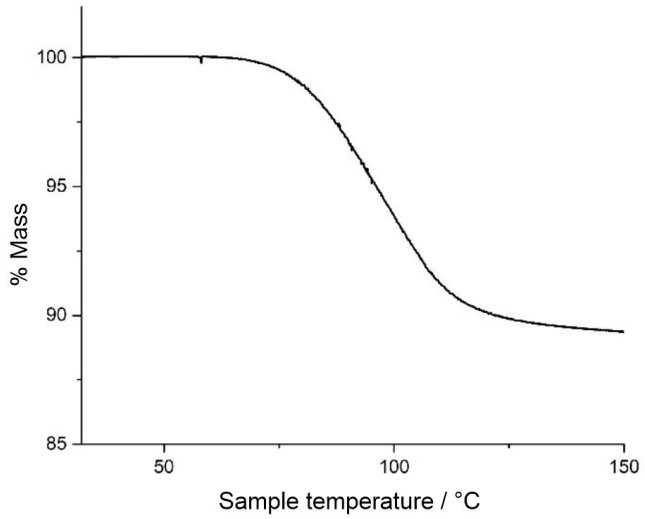

Figure 6.13: Thermogravimetric analysis data for the monohydrate phase of DL-lysine showing a mass loss of approx. 10% on heating from ambient temperature to 150 ℃. The mass loss shown is 10.64%.

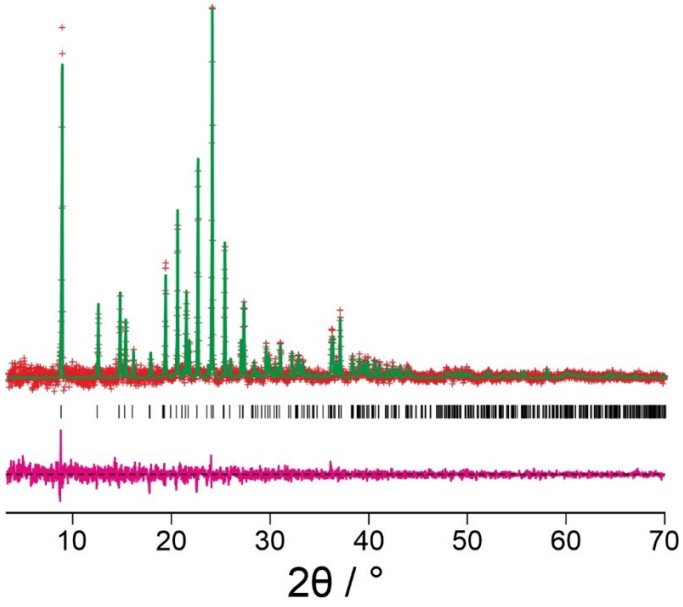

Figure 6.14: Profile fit using the Le Bail method for the monohydrate phase of DL-lysine (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Structure solution was conducted using the direct-space genetic algorithm technique, as described in Section 6.2.3, with one molecule of L-lysine and one molecule of water in the asymmetric unit. The best trial structure from the structure solution calculation was used as the starting model for Rietveld refinement using the program GSAS. Soft restraints were applied to the bond lengths and angles, and a planar restraint was applied to the carboxylate group and the α-carbon of the lysine molecule. A common isotropic displacement parameter was refined for non-hydrogen atoms, and the value for hydrogen atoms was taken as 1.2 times the value of non-hydrogen atoms. The initial Rietveld refinement shows a reasonable fit to the powder XRD data (R_{wp} = 4.28%, R_p = 3.22%).

The structure was then subjected to geometry optimization by periodic DFT-D methods. Although most atoms show only small displacements, the positions of the hydrogen atoms in the NH_2 group of the lysine side chain change significantly (corresponding to rotation of the NH_2 group), giving rise to a more favourable hydrogen bonding geometry. In the final cycles of Rietveld refinement, additional restraints, applied to preserve the hydrogen bonding geometry, resulting in a good fit to the powder XRD data (R_{wp} = 4.09%, R_p = 3.02%), shown in Figure 6.15, which was comparable to the quality of fit obtained in Le Bail profile fitting. The final refined parameters were: a = 5.87548(25) Å, b = 7.49624(30) Å, c = 19.8380(13) Å and β = 95.2648° (V = 870.06(10) Å^3). The final refined structure was subjected to further geometry optimization using periodic DFT-D methods, giving an RMSD for non-hydrogen atoms of only 0.086 Å, indicating that the final structure obtained in the Rietveld refinement is sufficiently close to an energy minimum.

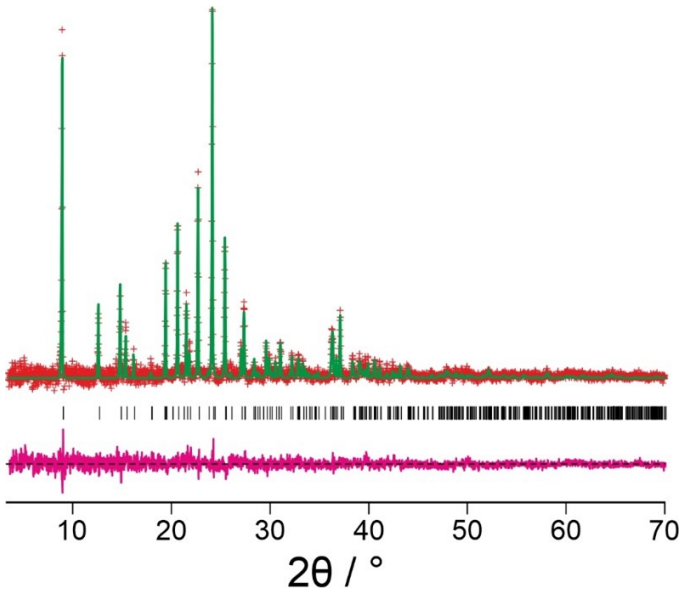

Figure 6.15: Final Rietveld refinement for the monohydrate phase of DL-lysine (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

6.4 Structural Discussion

The structures of the anhydrous and hemihydrate phases of DL-lysine are essentially isostructural (structural data are given in Table 6.3), except for the presence of water in the latter case. Upon dehydration of the hemihydrate phase, the *c*-axis contracts by *ca.* 1.6 Å, and the density of the crystal reduces slightly by 0.02 g cm^{-3}. The two structures have the same space group symmetry, *I*2/a, although the hemihydrate structure exhibits disorder in the hydrogen bonding involving the NH_2 groups and water molecules. The two individual ordered components are shown in Figure 6.16. In both the anhydrous and hemihydrate phases, the amino acid head-groups interact only with each other. The N—H bonds of the ammonium group are the donors in three short and very linear N—H···O hydrogen bonds to carboxylate groups of neighbouring molecules. The N—O hydrogen bond distances and N—H···O hydrogen bond angles for the anhydrate structure are: 2.852 Å, 161.4°; 2.772 Å, 175.7°; 2.758 Å, 157.8°. For the hemihydrate crystal structure, the values of the N—O hydrogen bond distances and N—H···O hydrogen bond angles are: 2.790 Å, 164.1°; 2.732 Å, 173.9°; 2.803 Å, 173.7°. Both structures can be observed in Figures 6.19 and 6.17, with specifically zoomed in regions to observe the hydrogen bonding between the amino acid head groups; from these figures the isostructural features are qualitatively observed.

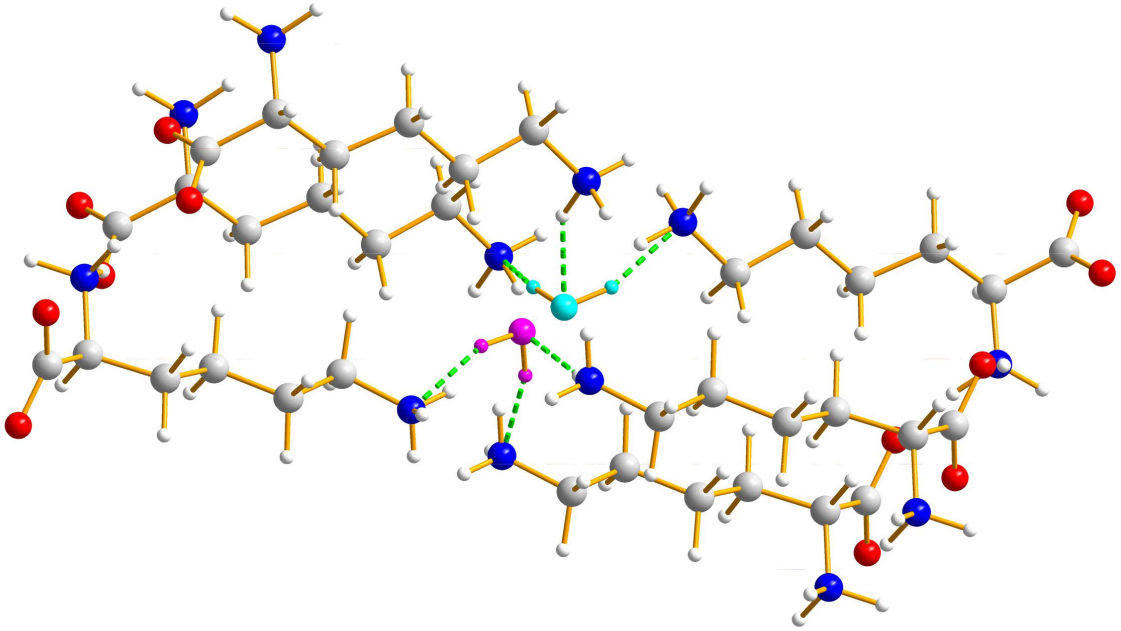

Figure 6.16: The two hydrogen-bonding arrangements involving the water molecules and the NH_2 groups of the lysine side-chains in the disordered structure of the hemihydrate phase of DL-lysine (following DFT-D geometry optimization). Each hydrogen-bonding arrangement has 50% occupancy and the two symmetrically opposed water molecules are differentiated by being coloured cyan and magenta respectively. The hydrogen bonds are represented by green dashed lines.

DL-Lysine			
Structure	Monohydrate	Hemihydrate	Anhydrate
a / Å	5.87548(25)	9.84809(34)	9.8402(5)
b / Å	7.49624(30)	4.72825(12)	4.73835(16)
c / Å	19.8380(13)	36.3276(16)	34.7195(17)
α / °	90	90	90
β / °	95.2648(30)	95.3772(28)	94.567(4)
γ / °	90	90	90
Volume / Å^3	870.06(10)	1684.12(11)	1613.71(12)
Space group	$P2_1$/n	*I*2/a	*I*2/a
Z	4	8	8
Z'	1	1	1
Density / g cm^{-3}	1.254	1.224	1.204
Temperature / K	293	293	293
Disorder	No	Yes	No
L-Lysine			
Structure	Monohydrate	Hemihydrate	Anhydrate
a / Å	5.93493(10)	9.53090(27)	9.5060(7)
b / Å	6.94133(11)	5.21728(17)	5.1289(4)
c / Å	20.5754(6)	17.6011(8)	16.9941(13)
α / °	90	90	90
β / °	90	101.078(4)	97.697(9)
γ / °	90	90	90
Volume / Å^3	847.63	858.913	821.09
Space group	$P2_12_12_1$	C2	$P2_1$
Z	4	4	4
Z'	1	2	2
Density / g cm^{-3}	1.287	1.247	1.187
Temperature / K	293	293	293
Disorder	No	No	No

Table 6.3: Crystal structural parameters for the three crystal structures containing DL-lysine in different hydrated forms, and the crystal structure parameters for three differently hydrated forms of L-lysine.

The only significant difference between the structures of the hemihydrate and anhydrate phases concerns the hydrogen bonding of the NH_2 groups of the side-chains. In the anhydrous form, the terminal $-NH_2$ groups engage in weak N—H$\cdots$N hydrogen bonding (3.188 Å, 120.4°) as shown in Figure 6.17. An overlay of the lysine molecules of the hemihydrate and anhydrate phases

(Figure 6.18) shows that they have very similar conformations, with only a slight difference at the end of the side-chain (torsion angle τ_5); the torsion angle values are given in Table 6.4. In the hemihydrate, the NH_2 groups of the side chains engage in two different hydrogen bonding arrangements with the water molecule (shown in Figure 6.19 and Figure 6.16). In hydrogen-bonding arrangement II, the water molecule is the acceptor in an N—H···O hydrogen bond (2.045 Å, 150.7°) and is the donor in two O—H···N hydrogen bonds to NH_2 groups of two different neighbouring lysine molecules (1.902 Å, 163.3°; 2.014 Å, 159.3°). Hydrogen-bonding arrangement III is essentially the same but the water is pointing in an opposite direction to interact with the terminal $-NH_2$ groups of the side-chains. As previously discussed in Section 6.3.2, the structure of the hemihydrate is an averaged structure of both arrangements II and III. Both arrangements are essentially the same and are equivalent by symmetry across the two-fold axis. In the anhydrous phase, although no water is present, there are still two plausible hydrogen-bonding arrangements for the terminal amino group. The terminal amino group interacts weakly via a chain of N—H···N hydrogen bonds (3.188 Å, 120.4°).

Torsion angle	Anhydrate	Hemihydrate
τ_1	88.23°	91.23°
τ_2	−174.43°	−177.04°
τ_3	−176.11°	−175.93°
τ_4	−166.80°	−167.37°
τ_5	167.42°	−175.35°
τ_6	86.09°	75.67°

Table 6.4: Values of torsion angles defining the conformation of the lysine molecule in the anhydrate and hemihydrate phases of DL-lysine. The values of the torsion angles are defined by the torsion angles shown in Figure 6.4, apart from τ_6 which was not optimized in the structure solution calculation; this torsion angle is the H—N—C—C torsion angle at the end of the lysine side chain.

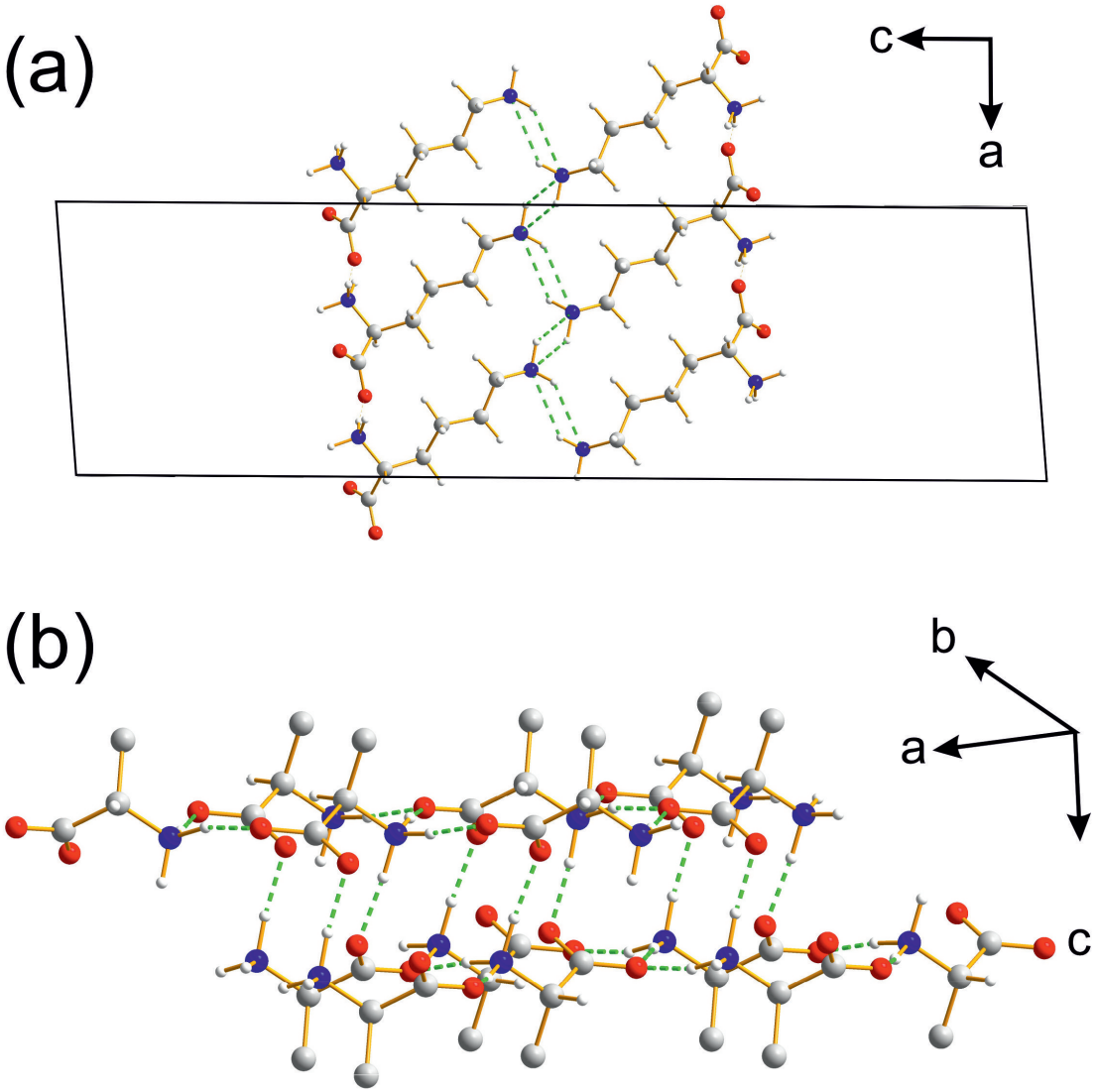

Figure 6.17: The crystal structure of the anhydrous phase of DL-lysine: (a) viewed along the *b*-axis, showing a bilayer type structure, and (b) showing the hydrogen-bonding arrangement between the amino acid head-groups of the lysine molecules. Hydrogen bonds are shown as green dashes in both cases.

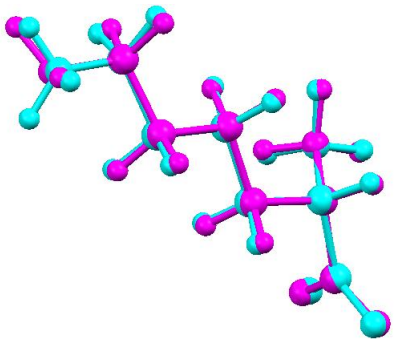

Figure 6.18: An overlay of the molecules of L-lysine in the asymmetric unit of the anhydrate phase (cyan) and the hemihydrate phase (magenta), illustrating the similar molecular conformation in these phases.

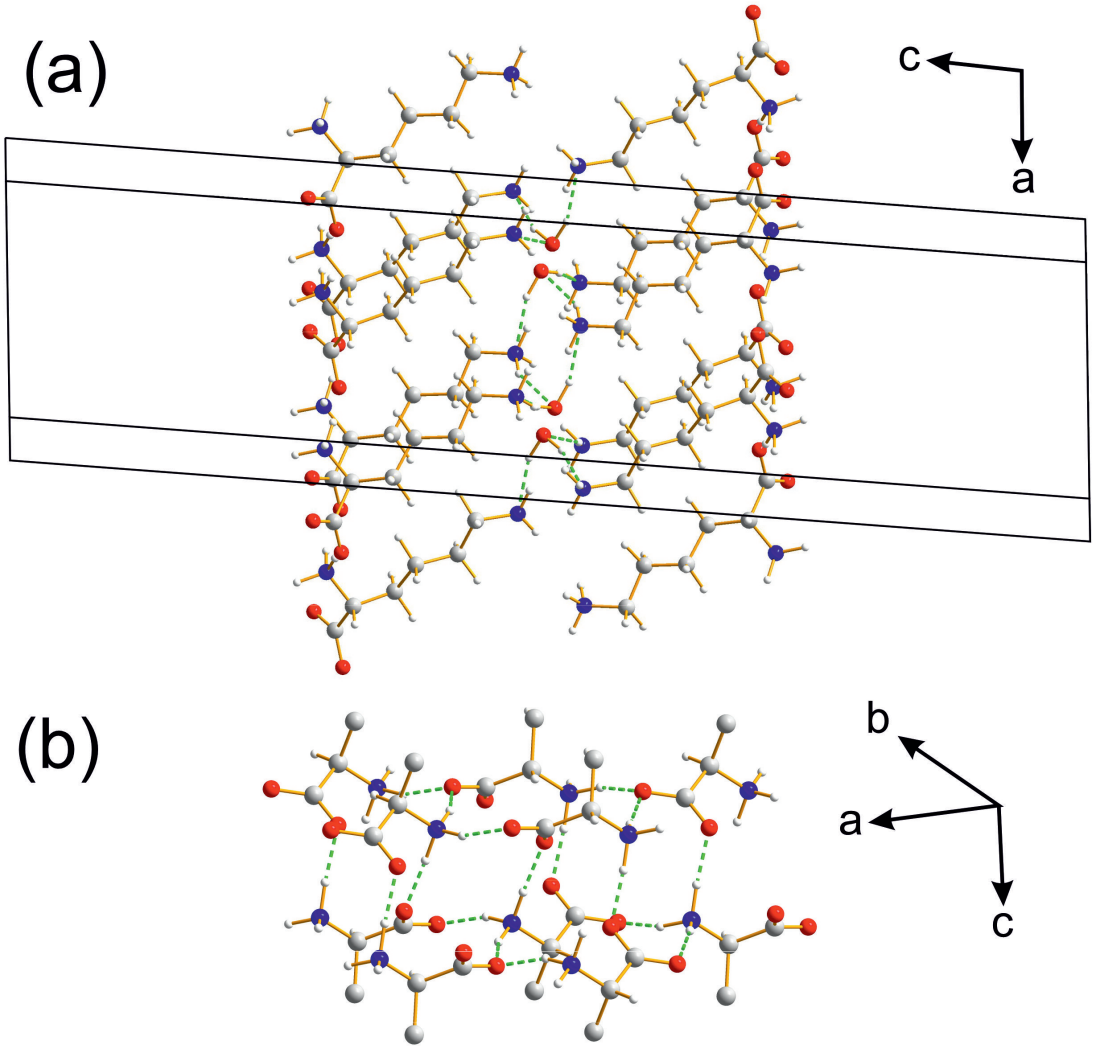

Figure 6.19: Crystal structure of the hemihydrate of DL-lysine: (a) viewed along the *b*-axis, showing a bilayer type structure, and (b) showing the hydrogen bonding interactions involving the amino acid head groups (the rest of the lysine molecules are omitted for clarity). Hydrogen bonds are shown as green dashes in both cases.

The anhydrate and hemihydrate structures both exhibit the same arrangement of L-lysine and D-lysine molecules, which are stacked in layers, when viewed along the *b*-axis. Within each bilayer, the L-lysine and D-lysine molecules alternate along the *a*-axis, similar to the arrangement in DL-tryptophan [177].

The crystal structure of the anhydrous phase of DL-lysine shows some similarity to the structure of L-lysine. Both structures are based on bilayers. The amino acid head groups form hydrophilic layers in both structures, with the alkyl chains are located in hydrophobic layers. The major difference between the DL-lysine and L-lysine structures is that the structure of L-lysine contains a chain of N—H···N hydrogen bonds involving the NH_2 groups of the side-chains repeated periodically along the *b*-axis, as shown in Figure 6.20. The interactions between the head groups in DL-lysine and the hemihydrate is common with the crystal structure of L-lysine. The unit cell dimensions of the DL-lysine and L-lysine crystal structures are remarkably similar; both the *a*-axis and *b*-axis are within 0.4 Å of each other, and the *c*-axis for DL-lysine is double that for L-lysine. It is noted that the density of DL-lysine is higher than that of L-lysine (ρ = 1.204 g cm^{-3} for the racemate and 1.184 g cm^{-3} for the enantiopure crystal structure), in agreement with Wallach's rule. It is worth noting that DL-lysine exists in the *I*2/a space group (with Z' = 1) whereas L-lysine exists in the $P2_1$ space group (with Z' = 2). The internal molecular conformations of both molecules of lysine in the asymmetric unit of L-lysine are particularly similar to the molecule in the asymmetric unit of DL-lysine.

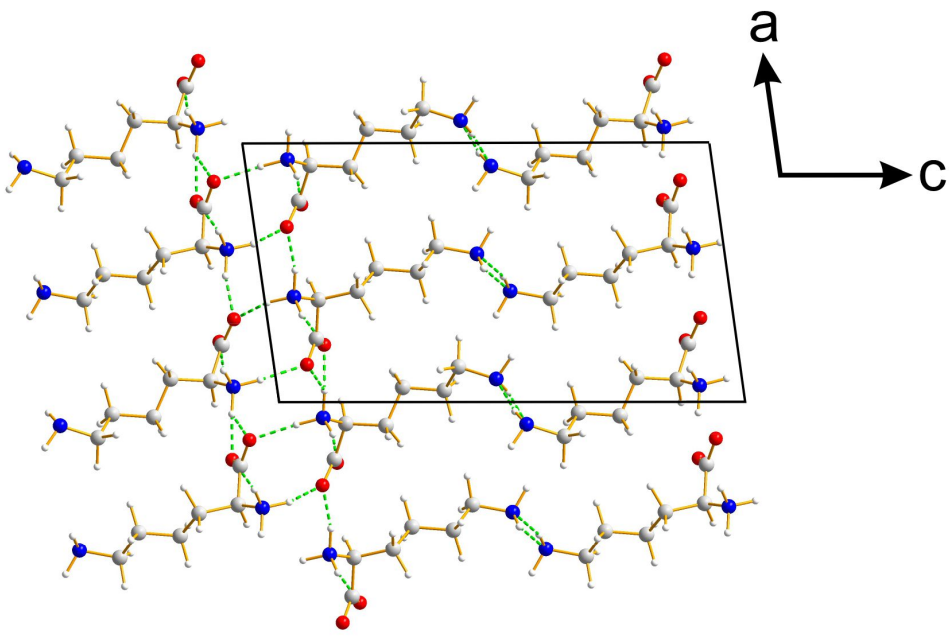

Figure 6.20: Crystal structure of L-lysine viewed along the *b*-axis, showing a bilayer type structure, with hydrogen bonds are shown as green dashes in both cases.

The monohydrate phase of DL-lysine does not share many structural features in common with the hemihydrate and anhydrate phases of DL-lysine, or with the structure of L-lysine. The monohydrate

structure comprises layers of D and L enantiomers parallel to the *ab*-plane, which alternate as L-D-L-D along the *c*-axis. The molecules in adjacent layers are offset from each other by b/4 along the *b*-axis, shown in Figure 6.21. This arrangement is reminiscent of the structure of DL-glutamic acid [178], which also has a side chain that contains hydrogen bonding functionality (CO_2H group) at the end of the side-chain. The water molecule is ordered and participates in a hydrogen bonded ring described in graph-set notation as $R_5^5(12)$ in an O—H···O hydrogen bond to the carboxylate group (2.750 Å, 160.4°) of a neighbouring molecule and the donor in an O—H···N hydrogen bond to the NH_2 group (2.692 Å, 165.9°) of another neighbouring molecule. The water molecule is also the acceptor in an N—H···O hydrogen bond with an N—H donor in the ammonium group (2.693 Å, 171.9°) of another neighbouring molecule.

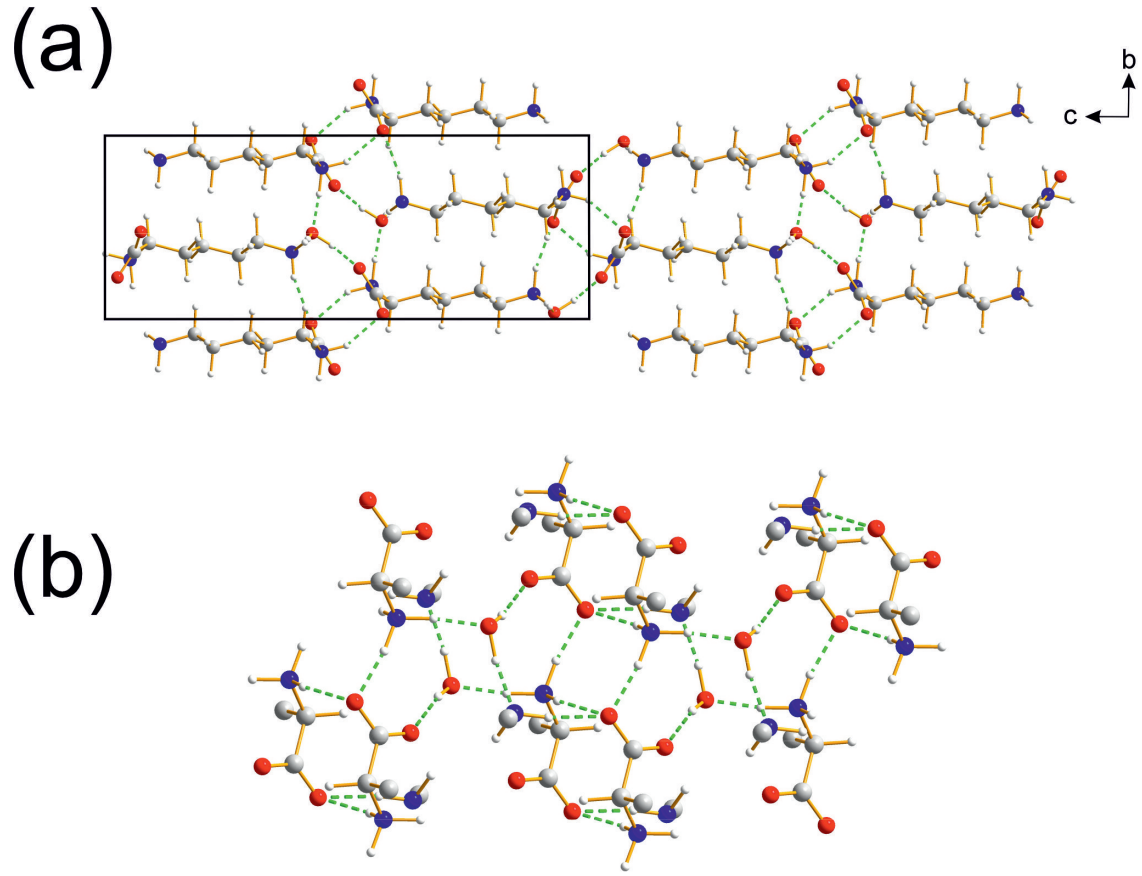

Figure 6.21: Crystal structure of the monohydrate phase of DL-lysine: (a) view of the structure along the *a*-axis, showing bilayers (parallel to the *ab*-plane), and (b) the hydrogen-bonding involving the water molecule, the $-NH_2$ groups from the lysine side chains, and the amino acid head groups.

6.5 Conclusions

In conclusion, the crystal structures of anhydrous, hemihydrate and monohydrate phases of DL-lysine have been determined directly from powder XRD data and periodic DFT-D calculations. All of these structures required the use of periodic DFT-D calculations to determine the most favourable hydrogen bonding networks present, and to resolve possible disorder. The methodology described in this chapter provides a good illustration of the identification of disorder by inspection of the average structure determined from diffraction data, together with periodic DFT-D calculations to assess the energetic feasibility of individual ordered local hydrogen bonding arrangements within the disordered structure.

Due to the difficulty in locating accurate information on the positions of hydrogen atoms in structure determination from powder XRD data, obtaining the correct hydrogen bonding scheme directly from structure solution and Rietveld refinement can be challenging. However, in such situations the application of periodic DFT-D calculations can facilitate a robust search for the most favourable hydrogen bond arrangement. This approach allows for conclusive assignment to hydrogen atom positions and provides the ability to establish (or rule out) the existence of disorder within the structure.

Chapter 7
The Discovery and Structure Determination of a New Metastable Polymorph of Ibuprofen

7.1 Introduction

Ibuprofen [Figure 7.1; (RS)-2-(4-(2-methylpropyl)phenyl)propanoic acid] was first derived from propionic acid in the 1960s by the Boots Group [40], and is mainly used as a non-steroidal anti-inflammatory drug (NSAID) [179]. It is often used to treat headaches and arthritis due to its analgesic (painkilling) properties. In its pure form, ibuprofen is poorly soluble in water. However, a salt co-crystal containing ibuprofen and lysine is more soluble in water, and ibuprofen thus has higher bioavailability in this form [180]. Ibuprofen is most commonly administered as the racemic mixture. The S-enantiomer is more biologically active, but the R-enantiomer is converted to the S-enantiomer in the body due to enzymatic inversion [181].

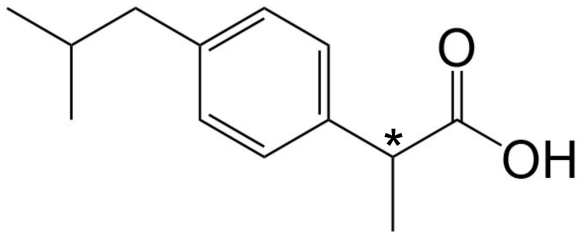

Figure 7.1: A figure showing the molecular structure of ibuprofen, an asterisk is used to define the chiral carbon of this molecule.

So far, two polymorphs of racemic ibuprofen have been reported. The crystal structure of the α polymorph was determined in 1974 by F.J. McConnell [182]. The crystal structure of the β polymorph was determined initially in 2010 by Derollez *et al*. [183] although several geometric aspects of this structure were far from optimal. Subsequently, van de Streek *et al*. optimized this crystal structure using DFT-D calculations [184]; the RMSD following geometry optimization was 0.38 Å, representing a significant difference between the published crystal structure and the geometry optimized structure. Van de Streek *et al*. also note that the isopropyl group is disordered, which would make energetic comparisons to other polymorphs less conclusive. The work of van

de Streek *et al*. does not further investigate the suspected disorder of the isopropyl group to determine whether it is static or dynamic, however the conformation of the group is noticeably different following geometry optimization. Both of the crystal structures of the α and β polymorphs of racemic ibuprofen can be seen in Figure 7.2.

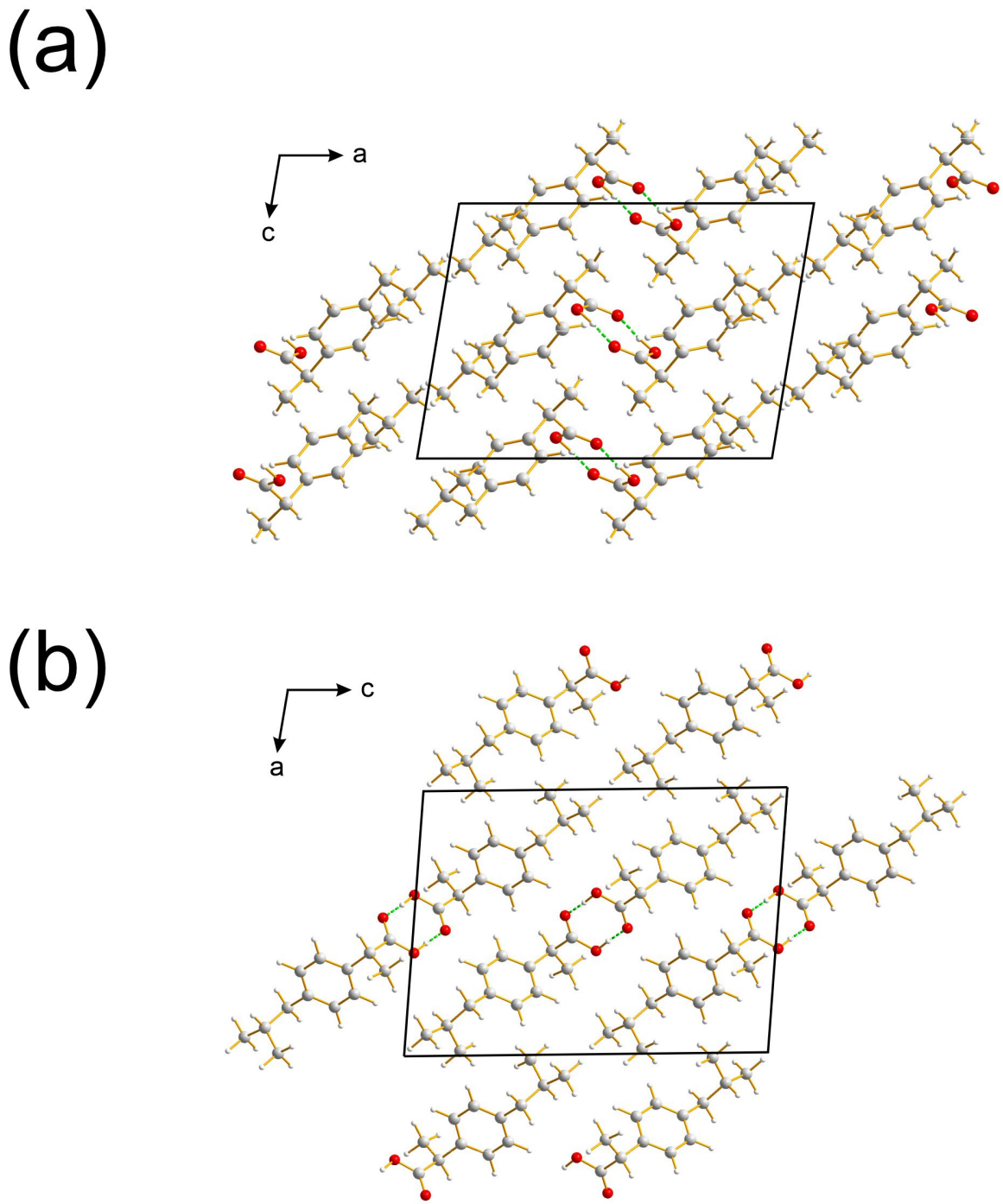

Figure 7.2: The crystal structures of the (a) α and (b) β polymorphs of racemic ibuprofen, both crystal structures are viewed along the *b*-axis. Hydrogen bonds are indicated with dashed green lines.

The enantiopure phase of ibuprofen has one crystal structure determined and is shown in Figure 7.3. It is observed from here that the enantiopure form of ibuprofen retains the dimer characterized by the $R_2^2(8)$ hydrogen bonding ring. These are made up of short, linear O—H···O hydrogen bonds in a similar arrangement to both the α and β polymorphs of the racemic form of ibuprofen. However, the enantiopure crystal structure of ibuprofen lacks an inversion center between these adjacent carboxylic acid groups due to the racemically pure nature of the material. Therefore S-enantiomers interact via hydrogen bonding to other S-enantiomers, unlike in the racemate.

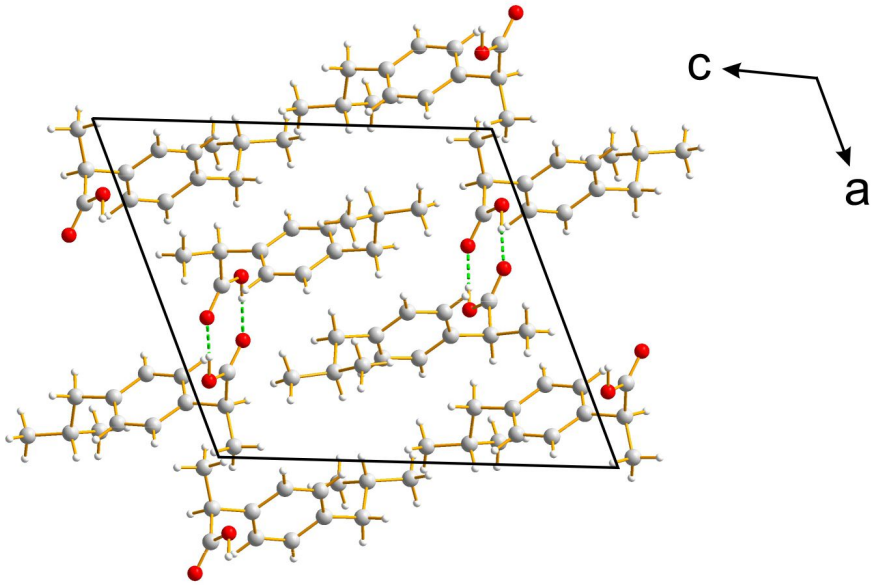

Figure 7.3: The crystal structure of the S-enantiomer of ibuprofen, viewed along the *b*-axis. Hydrogen bonds are indicated by dashed green lines.

7.2 Experimental Methods

7.2.1 Sample Preparation

A new polymorph of ibuprofen (named the γ polymorph) was prepared by crystallization from the melt. The starting material was the α polymorph of racemic ibuprofen, which was melted at 373 K (the melting point of the α polymorph of racemic ibuprofen is 349 K [185]). The molten sample was cooled to 258 K and held at this temperature until crystallization occurred, over several hours.

7.2.2 Synchrotron Powder X-ray Diffraction

Powder XRD data were recorded collected at 258 K on the I11 beamline at Diamond Light Source for the sample prepared by cooling from the molten phase, as described above. The powder XRD data were recorded using the multianalyzer-crystal (MAC) detector (wavelength, 0.826968 Å; 2θ range, 1° - 150°; step size, 0.004°; total time, 15 minutes).

7.2.3 Periodic DFT-D Calculations

Periodic DFT-D calculations for geometry optimization were conducted (with fixed unit cell) using CASTEP [122] (Academic Release version 19.1.1). These calculations utilized ultrasoft pseudopotentials [123], PBE functional [124], semi-empirical dispersion corrections using the TS correction scheme [125], conserved space group symmetry, fixed unit cell, periodic boundary conditions, a basis set cut-off energy of 700 eV and a Monkhorst-Pack grid [126] of minimum sample spacing (0.05 x 2π) Å^{-1}. The convergence criteria for geometry optimization were 0.01 eV Å^{-1} for atomic forces, 0.00001 eV per atom in the total energy, and 0.001 Å for atomic displacements.

7.3 Crystal Structure Determination

The high-resolution powder XRD data for the γ polymorph were indexed using the DICVOL algorithm within the CRYSFIRE package [88, 89], giving the following unit cell with monoclinic metric symmetry: a = 5.63 Å, b = 10.60 Å, c = 42.34 Å, β = 91.53° (V = 2323.3 Å^3). The indexing process omitted the two peaks observed at $2\theta \approx 3.3°$ and 3.8°, which correspond to the peaks of highest intensity for the α and β polymorphs of racemic ibuprofen. The powder XRD data are shown in Figure 7.4, with the two peaks due to the impurity phases highlighted.

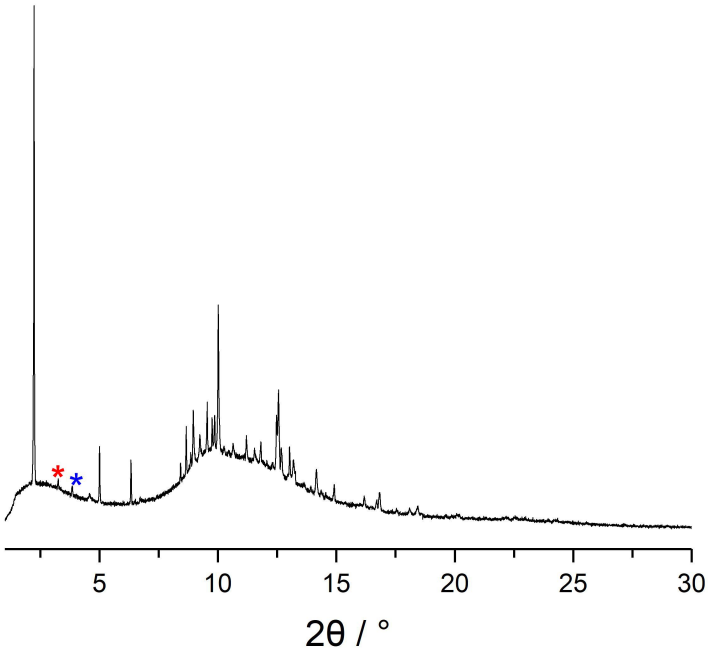

Figure 7.4: The powder XRD data for the γ polymorph of Ibuprofen. Peaks due to the impurities of the α and β polymorphs are indicated by the red and blue asterisks respectively.

Profile fitting using the Pawley method in the TOPAS program was conducted starting from the unit cell obtained in the indexing stage [85, 97], giving the fit shown in Figure 7.5 using the *P*2

space group. Although this unit cell gave a good quality of fit to the experimental data (R_{wp} = 3.10%, R_p = 2.21%), there were a large number of predicted peaks at low 2θ for which there is no corresponding peak in the experimental data. The possibility to fit all experimental peaks using smaller unit cells was then considered. Halving the *c*-axis (corresponding to *ca*. 42 Å) of the original unit cell yielded a good quality of Pawley fit (R_{wp} = 3.24%, R_p = 2.33%). The issue of having a large number of predicted peaks with no corresponding experimental peak at low 2θ was eliminated. This Pawley fit gave the following refined unit cell parameters: a = 5.6273(3) Å, b = 10.5809(6) Å, c = 21.191(2) Å, β = 91.530(6)° (V = 1261.31(6) Å^3). While the quality of fit with this smaller unit cell is still excellent (Figure 7.6), it is noted that there is some peak shape error for the dominant peak at 2θ = 2.2°. Assessment of different space groups using Pawley fitting concluded that the only space groups that are compatible with the powder XRD data are: *P*2, *P*2$_1$, Pm, *P*2/m and *P*2$_1$/m.

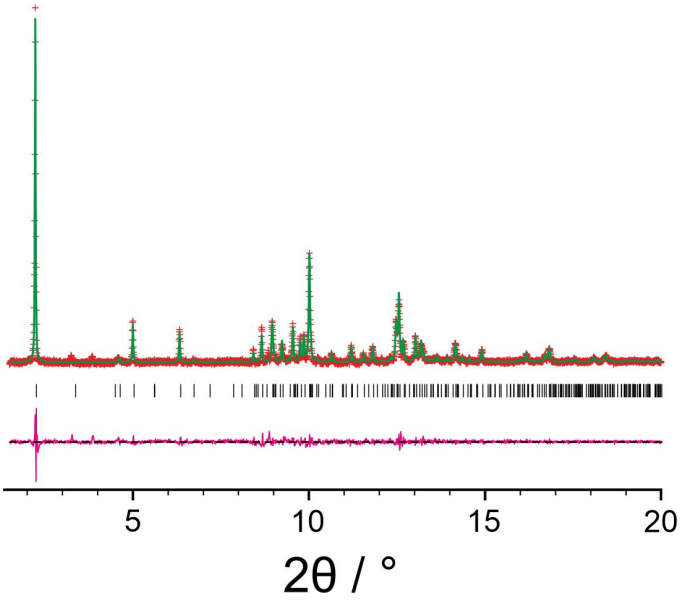

Figure 7.5: Profile fit using the Pawley method for the initial unit cell of the γ polymorph of ibuprofen (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

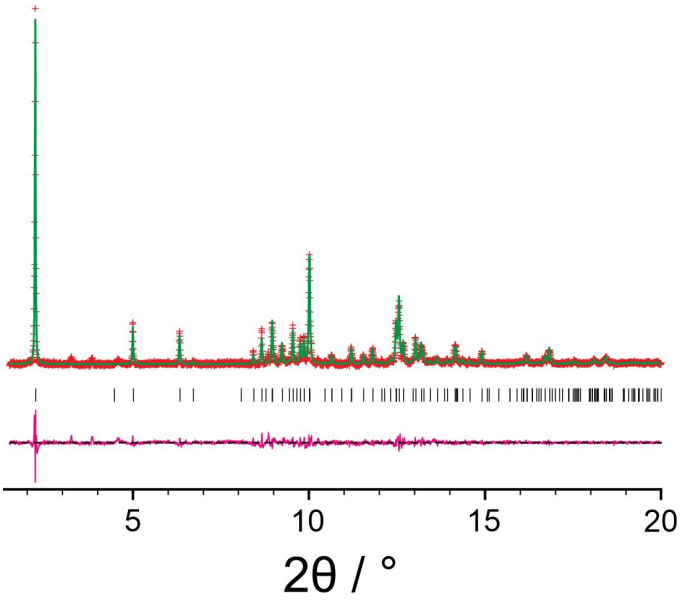

Figure 7.6: Profile fit using the Pawley method for the smaller unit cell of the γ polymorph of ibuprofen (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Structure solution was conducted using the direct-space genetic algorithm (GA) method implemented within the EAGER program [133–135]. The molecular model used for structure solution was generated using the molecular geometry from the reported crystal structure of the α polymorph (CSD Refcode: IBPRAC) [182], as shown in Figure 7.7. Structure solution calculations were conducted independently for space groups $P2$, $P2_1$, Pm, $P2/m$ and $P2_1/m$. For $P2$, $P2_1$

and Pm, density considerations indicate that there should be two molecules of ibuprofen in the asymmetric unit, whereas for $P2/m$ and $P2_1/m$ density considerations suggest that there should be one molecule in the asymmetric unit. With four molecules of ibuprofen in the unit cell, the approximate density was 1.09 g cm^{-3}. The densities of the α and β polymorphs of the racemate of ibuprofen are approximately 1.12 g cm^{-3} and 1.13 g cm^{-3} respectively. In total, 40 independent GA structure solution calculations were conducted, all of which started from a different randomly generated initial population. In the GA structure solution calculation, the population size was 100, the number of mating operations per generation was 10 and the number mutation operations per generation was 50. Although the material used in the sample preparation was the racemate, the material could potentially form a conglomerate (a mixture of enantiopure crystals containing the R-enantiomer, and enantiopure crystals containing the S-enantiomer).

Structure solution was conducted for both an enantiopure structure and for a racemate structure. In the case of the enantiopure structures, the S-enantiomer was used as the molecular model. Here only the $P2$ and $P2_1$ space groups are relevant, as the other mentioned space groups contain an inversion center and would cause the structure to be racemic. For the racemic structures, all space groups mentioned previously were considered, and both the R- and S-enantiomers were present in the asymmetric unit for the $P2$, $P2_1$ and Pm structure solution calculations, for the $P2/m$ and $P2_1/m$ structure solution calculations, there was only one molecule in the asymmetric unit which was the S-enantiomer. The structure solution calculations for the space groups $P2$ and $P2_1$ were defined by a total of 19 structural variables; 5 positional (the *b*-axis can be arbitrarily fixed for these space groups), 6 orientational and 8 torsional. The structure solution calculations for the space group Pm was defined by a total of 18 structural variables; 4 positional (the *a*- and *c*-axes can be arbitrarily fixed for these space groups), 6 orientational and 8 torsional. The structure solution calculations for the space groups $P2/m$ and $P2_1/m$ were defined by a total of 10 structural variables: 3 positional, 3 orientational and 4 torsional. The torsion angles allowed to vary in the structure solution calculations are shown in Figure 7.7. The structure giving the best fit to the powder XRD data in the case of enantiopure structures and in the case of racemate structures were used as the initial structural models for Rietveld refinement [105], which was conducted using TOPAS. For both the enantiopure structural model and the racemic structural model the space group that gave the lowest R_{wp} was $P2_1$.

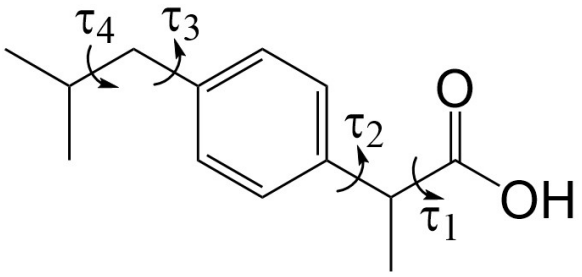

Figure 7.7: The molecular structure of ibuprofen with the four variable torsion angles in the GA structure solution calculation indicated by τ_n (n = 1, 2, 3, 4).

Prior to Rietveld refinement (directly from the structure solution calculation) the best fitting enantiopure structure contains carboxylic acid groups of neighbouring molecules that are in close proximity and potentially may be consistent with an $R_2^2(8)$ hydrogen bonding ring, although with a poor hydrogen bonding angle and low planarity between the two carboxylic acid groups. On the other hand, the best fitting racemate structure shows a much more linear hydrogen bond with most atoms in a similar plane. A rigid body model was used, with the same torsion angle variables as in the structure solution calculations, with additional bond distance and bond angle restraints applied to the hydrogen bonding between the two carboxylic acid groups in the structure. These hydrogen bonds were present between two independent molecules of ibuprofen, forming a hydrogen bonded dimer containing ibuprofen molecules of opposite chirality in the racemate structure, although it is clear they were not related by an inversion centre by performing a check for higher symmetry using the PLATON program [186]. A common isotropic displacement parameter was allowed to refine for all non-hydrogen atoms, with the value for hydrogen atoms taken as 1.2 times the refined value for non-hydrogen atoms.

Although an alternative racemate (as modelled in Figure 7.8) was not found in the structure solution stage, the possibility of an alternate racemate crystal structure was explored. If it does correspond to the correct structure, it may not have been generated by the structure solution process due to imperfect data i.e. powder XRD data affected by preferred orientation. The racemate used in the initial refinement was used as the starting model for this alternative structure; however, the chirality of each molecule in the asymmetric unit for the initial structure was inverted to form a second racemate. This was performed by swapping the positions of the methyl group and hydrogen atom on the α carbon atom in ibuprofen. At this point the “initial” racemate is known as racemate (a) and the structure with swapped chirality is known as racemate (b). This is shown in Figure 7.8, in which the two structures have the same positions of the molecules, but with the chirality inverted at each position. Both of these racemate models, as well as the structure representing the enantiopure model, were then subjected to geometry optimization in the program CASTEP, following the initial Rietveld refinements.

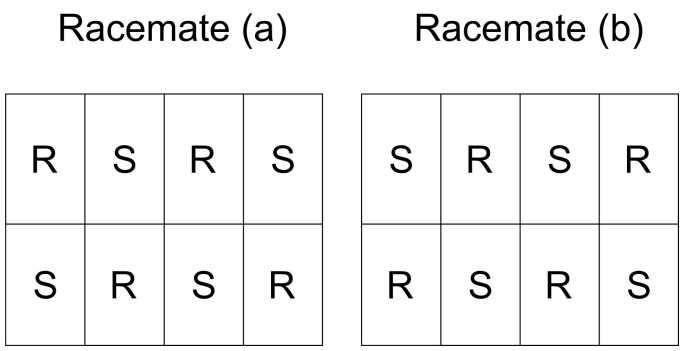

Figure 7.8: Schematic of two alternative racemate structures for ibuprofen, showing the molecules at the same positions in each structure, but with the chirality of the molecule inverted at each position.

Clearly the enantiopure structure is far higher in energy than either racemate structure (shown in Table 7.1). This is unsurprising given molten phase from which the γ polymorph was crystallized was a racemic liquid phase, which is a relatively "harsh" method of crystallization compared to standard processes.

	Enantiopure	Racemate (a)	Racemate (b)
Relative energy / kJ mol^{-1}	7.27	0.03	0.00

Table 7.1: Table showing the relative energies of the enantiopure structure and the two racemate structures for the γ polymorph of ibuprofen, following an initial Rietveld refinement.

At this point, both racemate (a) and (b) show very similar energies, well within the margin of error for the PBE-TS methodology and both were subjected to a final rigid body Rietveld refinement in TOPAS; however, the enantiopure structure was considered too high in energy to be a likely candidate for the crystal structure so was not taken to the next stage. The Rietveld refinement procedure was the same as described in the initial stages, with hydrogen bonding restrained again and a preferred orientation correction made using the spherical harmonics method [107]. Both refinements gave similar fits to the powder XRD data. Racemate (a): a = 5.6262(4) Å, b = 10.5831(5) Å, c = 21.1907(11) Å, β = 91.510(10)° (V = 1261.31(13) Å^3) (R_{wp} = 5.02%, R_p = 3.37%). Racemate (b): a = 5.6273(4) Å, b = 10.5838(4) Å, c = 21.1886(11) Å, β = 91.550(9)° (V = 1261.49(12) Å^3) (R_{wp} = 4.64%, R_p = 3.26%). Figure 7.9 shows the Rietveld fit of racemate (a) and Figure 7.10 shows the Rietveld fit of racemate (b).

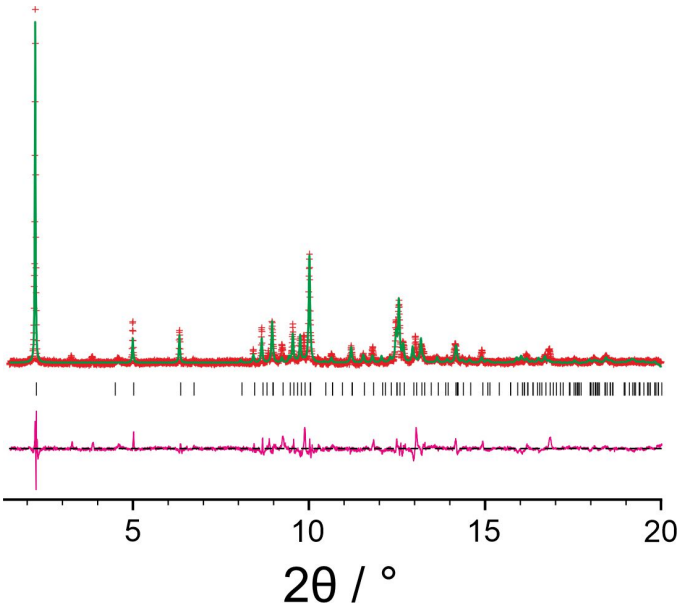

Figure 7.9: Rietveld refinement for the racemate (a) structure of the γ polymorph of ibuprofen (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

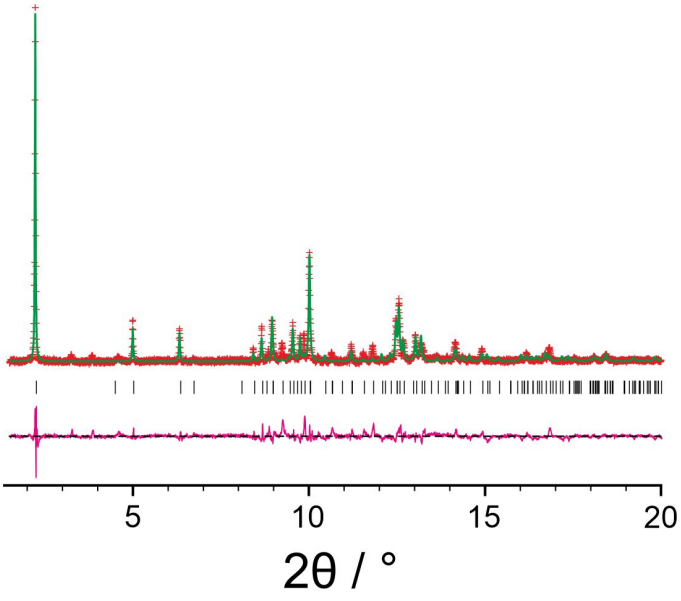

Figure 7.10: Rietveld refinement for the racemate (b) structure of the γ polymorph of ibuprofen (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Following the final refinements, both structures gave reasonably good fits to the data but, the fit for racemate (b) was better than racemate (a). At this point, a second geometry optimization was performed to determine how close both structures resided to an energy minimum. The overlays, shown in Figure 7.11, of the structures before and after optimization show that racemate (a) has a much lower displacement when the structure is optimized compared to racemate (b), with RMSD

values of 0.14 Å and 0.44 Å, respectively. From the values of the relative energies presented in Table 7.2, the structural model of racemate (b) is also marginally more stable upon geometry optimization. With both the fit to the powder XRD being more agreeable, and a lower relative energy, the proposed correct crystal structure is that of racemate (b).

	Racemate (a)	Racemate (b)
Relative energy / kJ mol^{-1}	0.31	0.00

Table 7.2: Table showing the relative energies of two racemate structures for the γ polymorph of ibuprofen, following the final Rietveld refinement.

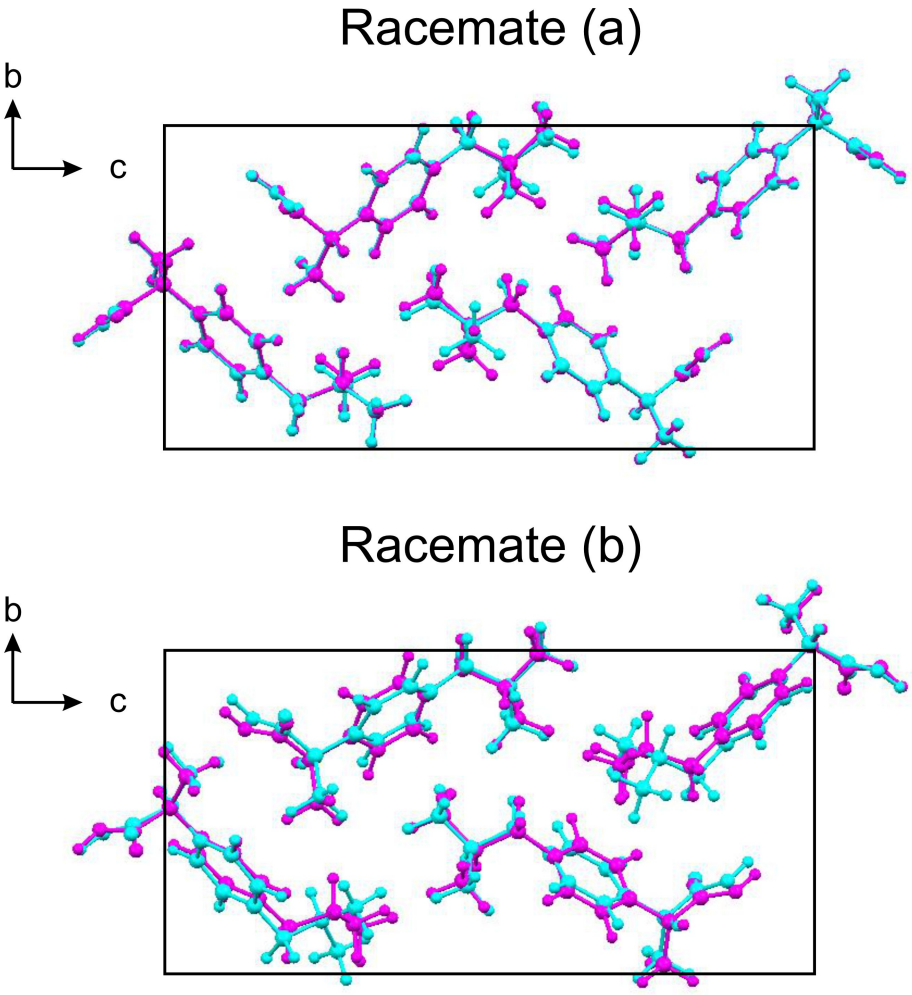

Figure 7.11: Overlay of the structures of racemate (a) and racemate (b) of the γ polymorph before (magenta) and after (cyan) DFT-D geometry optimization (with fixed unit cell). Racemate (a) is above, racemate (b) is below.

7.4 Structural Discussion

The crystal structure of the γ polymorph of racemic ibuprofen (shown in Figure 7.12) is dominated by a pair of O—H···O hydrogen bonds between the carboxylic acid groups of two ibuprofen molecules of opposite chirality. Although the α and β polymorphs contain an inversion center in the middle of a similar carboxylic acid hydrogen bonded "dimer", it is not the case in this new crystal structure because the two molecules of ibuprofen in the asymmetric unit have different conformations of the aromatic rings with respect to the carboxylic acid groups (τ_2: R-enantiomer, −52.97°; S-enantiomer, −9.14°). However, for further validation the structure was processed in the PLATON program [186] to evaluate for higher symmetry, which was not found.

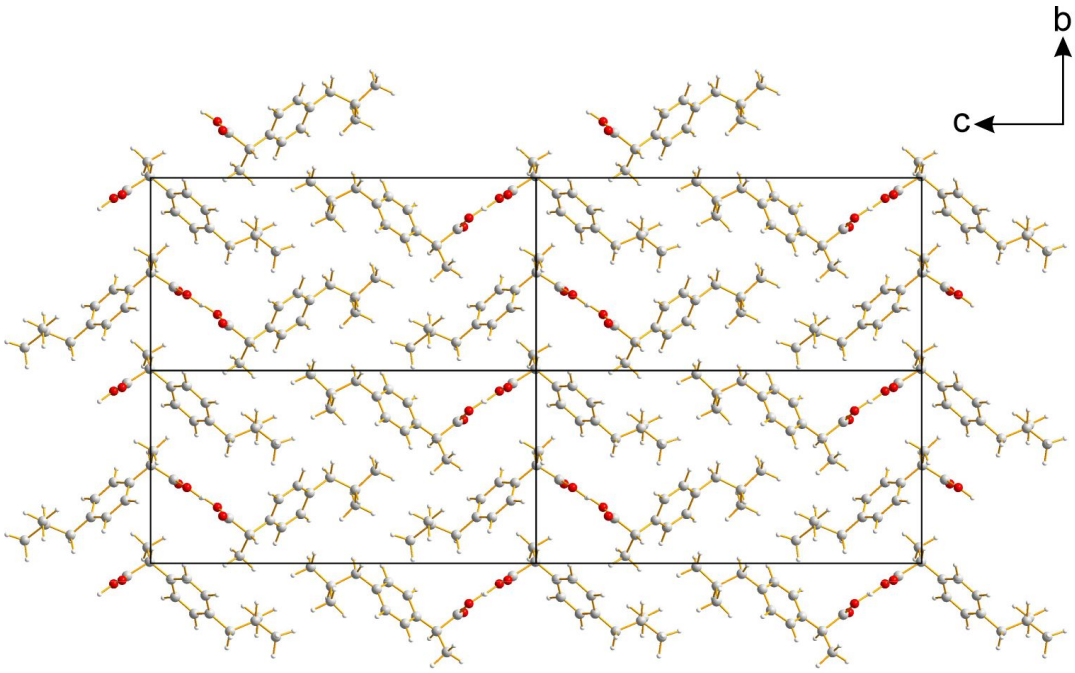

Figure 7.12: The crystal structure of the γ polymorph of ibuprofen viewed along the *a*-axis, following refinement by the Rietveld method.

The hydrogen-bonding arrangement, representing an $R_2^2(8)$ ring in graph-set notation, is shown in Figure 7.13. These hydrogen bonds are of similar geometry due to the geometry for the O—H···O hydrogen bond donated by the R-enantiomer and accepted by the S-enantiomer being 2.636 Å and 175.9° (O···O distance) whereas for the O—H···O hydrogen bond donated by the S-enantiomer and accepted by the R-enantiomer it is 2.668 Å, 176.0 ° (O···O distance). This hydrogen bonding geometry is similar to the α and β polymorphs, although in these structures, the hydrogen bonds are symmetry equivalent due to the presence of the inversion centre.

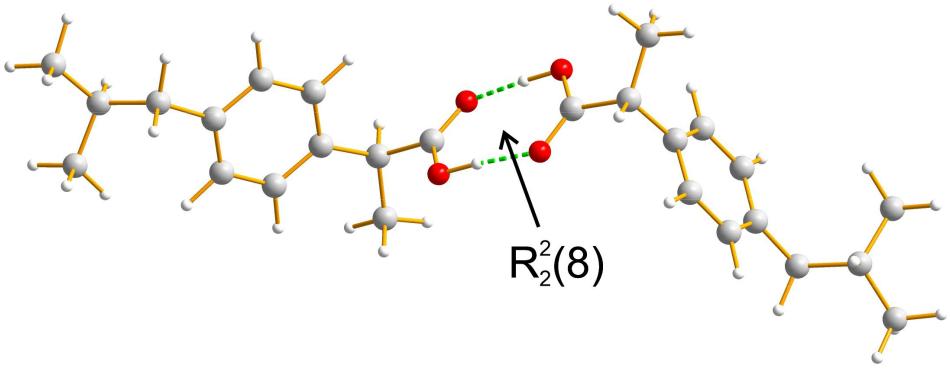

Figure 7.13: The $R_2^2(8)$ hydrogen bonded ring in the crystal structure of the γ polymorph of ibuprofen.

The crystal structures of all three polymorphs are similar (both the α and β polymorphs are shown in Figure 7.2 with a similar view provided for the γ polymorph in Figure 7.12). The structure of the β polymorphs has been shown to exhibit disorder in the isopropyl group. Disorder in the γ polymorph has not been investigated, although this issue could be investigated using molecular dynamics simulations to determine whether there is one clear preferential conformation of the isopropyl group or whether different conformations are plausible and of sufficiently close energy to be populated in a disordered situation. In contrast, there is no disorder in the α polymorph. Structural parameters for the α, β and γ polymorphs are given in Table 7.3.

Polymorph	γ	α	β
a / Å	5.6262(4)	14.667	12.569
b / Å	10.5831(5)	7.886	5.605
c / Å	21.1907(11)	10.73	17.209
β / °	91.510(10)	99.362	94.944
Volume / Å^3	1261.31(13)	1224.54	1207.91
Space group	$P2_1$	$P2_1$/c	$P2_1$/c
Z	4	4	4
Z'	2	1	1
Density / g cm^{-3}	1.086	1.119	1.134
Temperature / K	258	283-303	258
Method	PXRD & DFT-D	SC-XRD	PXRD & DFT-D
τ_1 / °	29.99 & -1.68	-35.85	36.89
τ_2 / °	-52.97 & -65.59	42.05	-7.70
τ_3 / °	-12.64 & 1.03	18.77	-39.84
τ_4 / °	51.80 & 80.12	-57.95	-177.65

Table 7.3: Table showing the structural parameters for the α, β and γ polymorphs of ibuprofen.

7.4.1 Assessment of the Relative Energies of the α, β and γ Polymorphs of Ibuprofen

The structure of the γ polymorph of ibuprofen has been determined as described above. We now compare the relative energies of the α, β and γ polymorphs using the PBE-TS methodology in CASTEP, shown in Table 7.4. The results indicate that the energy of the γ polymorph is much higher than the energies of the α and β polymorphs. Due to the rapid cooling of the molten phase of ibuprofen, from which the new polymorph was produced, it is clear to see this polymorph is the least stable by 22.50 kJ mol^{-1}, which is twice the relative energy difference between the β polymorph and the α polymorph. These results suggest that the γ polymorph is a metastable form of racemic ibuprofen. We note that the γ polymorph was observed as the initial crystallization product in a rapid cooling of the molten phase, consistent with this polymorph being a metastable kinetic product.

	α	β	γ
Relative energy / kJ mol^{-1}	0	13.72	22.50

Table 7.4: Table showing the relative energies of the α, β and γ polymorphs of ibuprofen.

7.5 Conclusions

The γ polymorph of ibuprofen has been prepared in this work through crystallization from the melt, with the crystal structure determined using synchrotron powder XRD data in conjunction with DFT-D calculations. As a variety of trial structures generated by the structure solution process were considered structurally reasonable, periodic DFT-D calculations were essential in determining the correct structure by discriminating these structures as a function of their energy. Periodic DFT-D calculations were also particularly useful in determining whether the structure was enantiopure or racemic. As the sample was metastable and was only produced on one occasion, further structural analysis using other techniques (*e.g.* solid-state NMR) was not feasible.

In comparison to the known α and β of racemic ibuprofen, the γ polymorph is the only structure that does not contain an inversion centre. The molecular conformations are quite different from those in the α and β polymorphs, as shown in Figure 7.14. The structure determination of the γ polymorph is a fantastic example of the utility of simple periodic DFT-D calculations when limited experimental data are available to ensure the structure is validated. Further work could entail crystal structure prediction calculations to determine whether the α, β and γ polymorphs are observed from these calculations, as well as to any other energetically feasible structures of racemic ibuprofen.

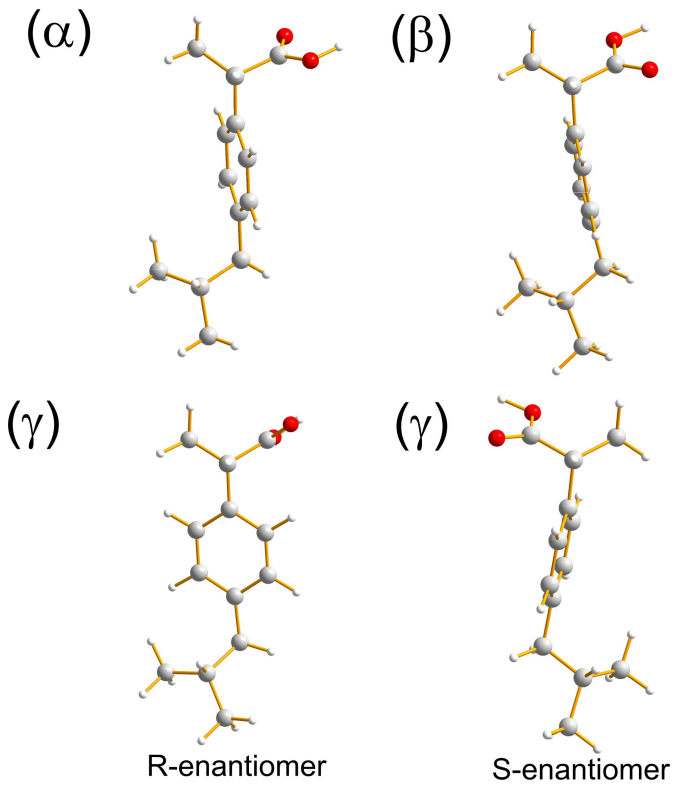

Figure 7.14: The molecules of ibuprofen that compose the asymmetric units in the crystal structures of the α, β and γ polymorphs, viewed along the α-C—H bond. The R- and S-enantiomers from the γ polymorph are shown separately.

Chapter 8

Structural Characterization of Symmetrical Molecules with Rigid Cores

8.1 Introduction

Studying symmetrical molecules with rigid cores has been conducted using powder XRD data for decades now, and in some instances, they have proved challenging. In work published in 2002, the study of substituted arenesulfonamides were published from powder XRD data [100], using a differential evolution technique for structure solution. This paper mentioned the determination of these types of crystal structures had previously been attempted yet were unsuccessful. The authors commented that, upon the development and implementation of a differential evolution (DE) approach for direct-space structure solution (analogous to the genetic algorithm search procedure), successful structure solution was finally achieved. This methodology works by creating a child for each member of a population of trial structures, based on summation of the weighted differences of randomly selected members of the population Then, the recombination (mating) and mutation of structures are conducted in a single step to produce a new population. In this process, the child and parent structures are compared (by assessing the quality of fit to the powder XRD data) and the superior structure is added to the new population. As in the GA search procedure, the DE procedure is repeated for a fixed number of generations, or until convergence is achieved. This approach was targeted specifically to arenesulfonamide materials, including 4-nitrobenzenesulfonamide (shown in Figure 8.1) which contains a σ_v plane through the center of the benzene ring, the nitro group and the sulfonamide group.

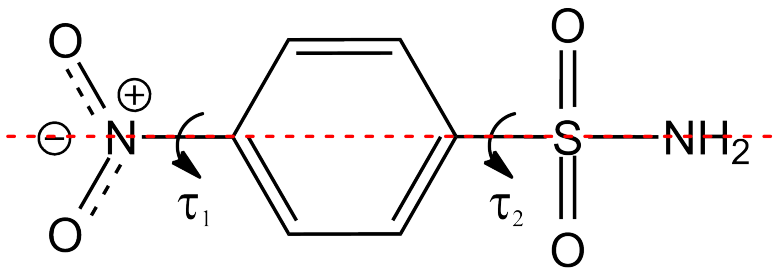

Figure 8.1: The molecular structure of 4-nitrobenzenesulfonamide, with the two variable torsion angles denoted as τ_1 and τ_2 respectively, and the σ_v mirror plane shown by the dashed red line.

This symmetry could be observed in an isolated molecule of 4-nitrobenzenesulfonamide if the τ_1 and τ_2 values are such that the nitro group is coplanar with the benzene ring (or perpendicular to the benzene ring) and with the bisector of the two S=O bonds lying in the σ_v mirror plane. This conformation could exist in a periodic crystal structure containing a mirror plane (as a space group symmetry element) and with the mirror plane of the molecule consistent with the crystallographic mirror plane, giving rise to a Z' value of 0.5. This is assuming there are no other molecules present in the asymmetric unit. Whilst this type of scenario is a trivial issue in crystal structure determination from single-crystal XRD data, which is a highly automated process nowadays, it is more challenging to manage this type of situation when determining the crystal structure using the direct-space approach from powder XRD data. In this scenario, knowledge of the composition of the asymmetric unit is essential prior to commencing the structure solution calculation. In the case of 4-nitrobenzenesulfonamide, the nitro group is in fact coplanar with the benzene ring, but the sulfonamide group is not in the conformation required to retain a σ_v mirror plane through the middle of the molecule (as shown in Figure 8.2). For a space group of $P2_1/n$, it is clear there is no mirror plane, and there is a single molecule of 4-nitrobenzenesulfonamide in the asymmetric unit. In the crystal structure two noticeably short and linear N—H$\cdots$O hydrogen bonds are formed between the amine group and oxygen atoms on the nitro and sulfonamide groups of neighbouring molecules, such that the sulfonamide group is not in a suitable conformation to retain an intramolecular mirror plane.

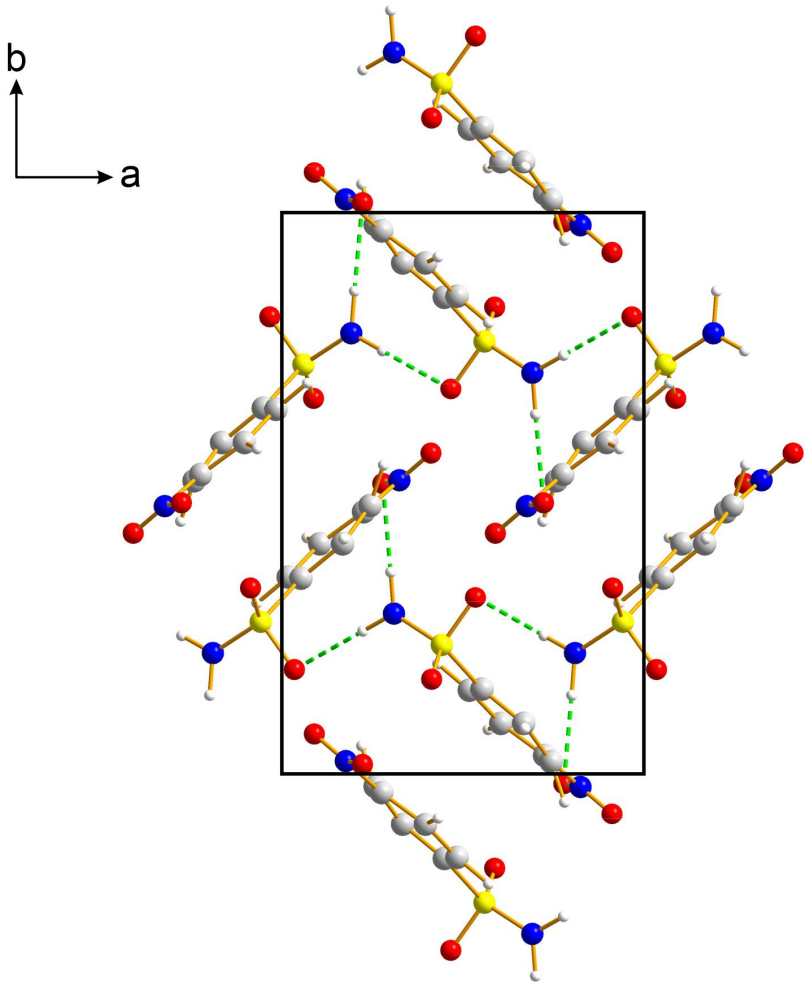

Figure 8.2: The crystal structure of 4-nitrobenzenesulfonamide, viewed along the *c*-axis, with hydrogen bonds shown as dashed green lines.

Molecules that possess an intramolecular mirror plane have the potential to retain this symmetry in a periodic crystal structure, for example a planar molecule residing on a crystallographic mirror plane in the Pm space group. Such possibilities must be considered when conducting crystal structure determination from powder XRD data.

This chapter describes determination of two materials containing molecules that contain several symmetry elements for the isolation molecule and contain rigid cores, and highlights the types of issues that arise in tackling structure determination in such cases.

8.1.1 Background to Trimesic Acid

Trimesic acid (TMA), also known in IUPAC nomenclature as benzene-1,3,5-tricarboxylic acid, is shown in Figure 8.3. TMA is a versatile compound in a wide range of materials applications [187], as a consequence of the hydrogen bonding networks that can be generated due to 1,3,5-tricarboxylic acid functionalization of the molecule. As a result, TMA can form supramolecular structures

with low density and high pore volume [188]. An example of a specific application of TMA is in spectrofluorimetry [189].

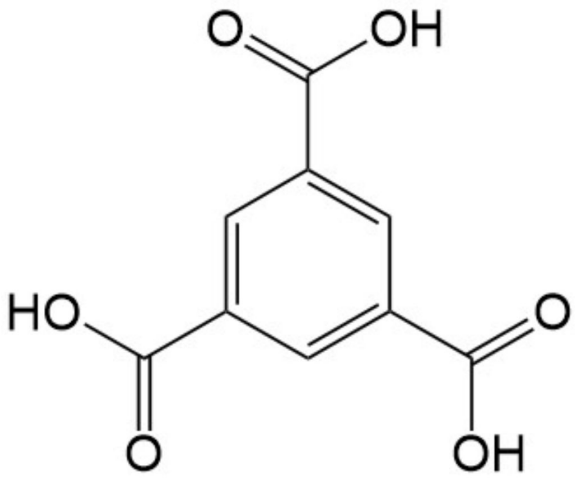

Figure 8.3: The molecular structure of trimesic acid.

Trimesic acid can self-assemble into an interesting honeycomb network containing cyclic hydrogen-bonded rings, constructed from the well-known double hydrogen-bonded carboxylic acid "dimer" motif, denoted by graph set notation as $R_2^2(8)$ and shown in Figure 8.4 [190]. The ability to form this type of open hexagonal network is a reason trimesic acid is commonly used as a linger in metal-organic frameworks. TMA can also form a wide range of solid inclusion compounds [191]. Upon substitution of further functional groups on the aromatic ring at the 2,4 and/or 6 positions, the structures that form show far lower symmetry, usually with far higher densities than the open network structures formed by unsubstituted TMA. Even the monohydrate and dihydrate crystal structures of trimesic acid are noticeably lower symmetry than the non-solvate phase of TMA, with the dihydrate showing orthorhombic symmetry and the monohydrate exhibiting triclinic symmetry [192, 193]. When a water molecule is accommodated within the crystal structure, the $R_2^2(8)$ hydrogen-bonding schemes are disrupted. Rather than forming the standard carboxylic acid hydrogen-bonding dimers with short and linear hydrogen bonds, the carboxylic acid groups form less ideal hydrogen bonds with a mixture of water molecules and carboxylic acid groups of adjacent molecules. Both have densities much higher than the non-solvate phase, about 50% higher in both cases (1.54 g cm^{-3} for the monohydrate and 1.58 g cm^{-3} for the dihydrate, 0.94 g cm^{-3} for the non-solvate). In this work, we study the dibromo derivative of TMA, in both its hydrate and anhydrous crystalline forms. It is referred to here as DBTMA.

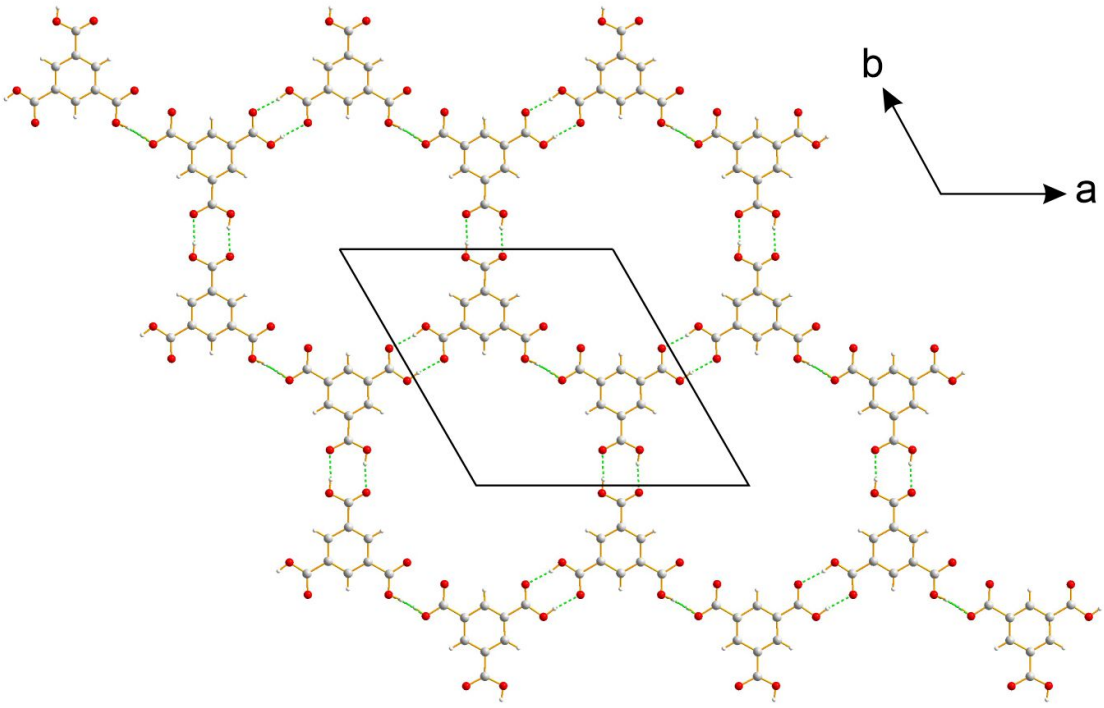

Figure 8.4: The crystal structure of one layer of TMA molecules in the crystal structure of guest-free trimesic acid, viewed along the *c*-axis.

8.1.2 Background to Perylene Diimide and its Derivatives

Perylene diimide, hereafter referred to as PDI, is a planar aromatic molecule shown in Figure 8.5. The two N atoms at each end of the ring system can be functionalized with a variety of groups, such as aliphatic rings or alkyl chains. PDI derivatives have a variety of applications, based in their luminescent, electronic and optical properties [194]. The most notable application is as industrial dyes, in which the N—H group is replaced by N-R groups such as: R = CH_3 (crimson), R = $C_6H_5CH_2CH_2$ (black), or R = 3,5-$(CH_3)_2C_6H_3$ [195] (medium red).

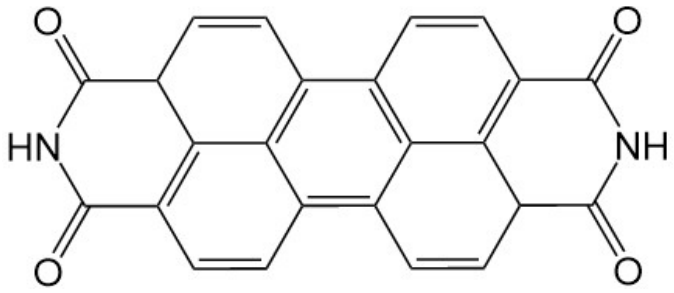

Figure 8.5: The molecular structure of perylene diimide.

A specific example of a PDI-functionalized material with useful applications is the cyclohexyl-functionalized PDI, referred to as PDICH throughout the thesis. This specific functionalization of the PDI ring facilitates improved optoelectronic properties due to the formation of 100 μm long

"nanobelts", referring to crystals that are long and flat [196]. The crystal structure of PDICH is reported in [196] using 3D-ED methods; however, as a consequence of problems due to dynamical scattering in the 3D-ED data, the geometry of the PDICH molecule in the reported structure was far from ideal, both in terms of the PDI ring and the cyclohexyl groups. In this chapter, we report the structure determination of PDICH from powder XRD data, using the direct-space strategy for structure solution, followed by Rietveld refinement and DFT-D validation, and leading to a structure of significantly higher quality than that reported previously [196].

8.2 Crystal Structure Determination of 2,4-Dibromo Trimesic Acid (DBTMA)

8.2.1 Experimental Methods

Synthesis of DBTMA

2,4-dibromomesitylene (10.16 g) was added to a solution of sodium hydroxide (2.59 g in 125.06 mL of deionized water) with stirring at ambient temperature, and potassium permanganate (42.77 g) was then added over a period of 2 hr. The mixture was then heated to 95 °C and stirred over a period of time (15 days) until the mixture turned from purple to brown, indicating full oxidation. This reaction is depicted in Figure 8.6. The reaction mixture was then cooled to ambient temperature and filtered, giving a clear filtrate. Hydrochloric acid (37%) was then added dropwise to the clear solution until pH = 1 was achieved, to ensure full protonation of all carboxylate groups. The product was then extracted using diethyl ether as a separating agent, leaving the product within the organic layer. The organic layer was left in a ventilated fume cupboard to allow the diethyl ether to evaporate. White crystals with long-needle morphology, identified subsequently (from powder XRD) as the dihydrate phase of DBTMA, were also determined in this work and therefore the sample prepared here is confirmed as being a pure sample of DBTMA.

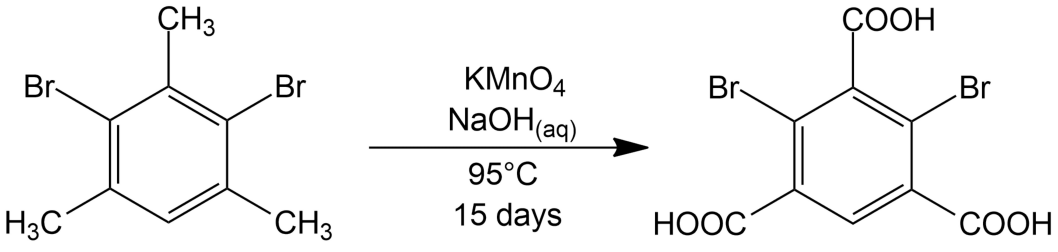

Figure 8.6: Reaction scheme for the oxidation of 2,4-dibromomesitylene to 2,4-dibromo trimesic acid.

Preparation of the Anhydrous Phase of DBTMA

A sample of the dihydrate phase of DBTMA (ground with a mortar and pestle) was packed into three glass capillaries (diameter, 0.7 mm) and placed in an oven at 80 °C. After 16 hr, the capillaries were removed from the oven and quickly flame sealed to prevent rehydration. The powder XRD data for the sealed capillaries showed no consistency with the powder XRD data for the dihydrate phase, indicating a full phase transition to a lesser hydrated phase.

Single-crystal X-ray Diffraction of the Dihydrate Phase of DBTMA

The harvesting of the single-crystal sample, collection and processing of single-crystal XRD data, and structure solution followed by structural refinement was performed by Dr Benson Kariuki at Cardiff University.

Single-crystal XRD data were collected for the dihydrate phase of DBTMA at room temperature on an Agilent SuperNova Dual Atlas diffractometer with mirror monochromator Cu radiation. The crystal structure was solved by SHELXS [197] and refined using SHELXL2018 [118].

Non-hydrogen atoms were refined with anisotropic displacement parameters whereas all hydrogen atoms were located in difference Fourier maps. For the final refinement, hydrogen atoms were inserted in idealized positions for the DBTMA molecule using a riding model. Geometric restraints were applied to water molecules. For the final refinement, hydrogen atoms inserted in idealized positions for DBTMA were set at 1.2 or 1.5 times the value of U_{eq}(C,O) for the atom to which they are bonded. Elongated ADP ellipsoids indicated disorder for one carboxylic acid group, which was thereafter modelled as two components, which refined with occupancies ca. 50%. Crystallographic and refinement parameters are shown in Appendix B.3.

Powder X-ray Diffraction of the Anhydrous Phase of DBTMA

High-quality powder XRD data for structure determination of the anhydrous phase of DBTMA were recorded at ambient temperature (21 °C) on a Bruker D8 Advance Series II diffractometer operating in transmission mode (Ge-monochromated Cu-Kα_1; Vantec detector covering 3° in 2θ; range, 4° to 70°; step size, 0.016°). The three flame-sealed glass capillaries containing the powder sample were placed on a foil type sample holder. The total data collection time was 46.5 hr.

Thermogravimetric Analysis of the Dihydrate Phase of DBTMA

Thermogravimetric analysis (TGA) was conducted using a Perkin Elmer 4000 ThermoGravimetric analyzer. The TGA measurement was conducted starting from a powder sample of the dihydrate phase of DBTMA (*ca.* 15 mg) in an open aluminium pan, with the following measurement parameters: initial temperature, 30 °C; final temperature, 120 °C; heating rate, 10 °C min^{-1}; nitrogen flow

rate, 30 mL min^{-1}. Following each temperature increase of 10 °C, the material was held at that temperature for 10 min.

The TGA experiment shown in Figure 8.7 indicates a percentage mass loss of approx. 10.1% from the dihydrate phase of DBTMA, which is consistent with the loss of two water molecules. Based on this result, it was concluded that an anhydrous phase of DBTMA can be prepared by dehydration of the dihydrate phase of DBTMA at 80°C.

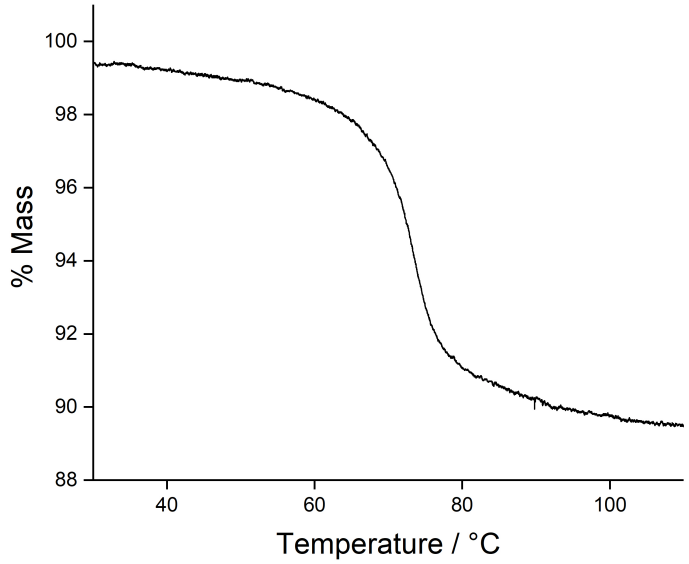

Figure 8.7: Thermogravimetric analysis data for the dihydrate phase of DBTMA.

Periodic DFT-D Calculations

Periodic DFT-D calculations were conducted using CASTEP (Academic release 19.1.1). The calculations used ultrasoft pseudopotentials, the PBE functional, semiempirical dispersion corrections using the TS correction scheme, fixed unit cell, preserved space group symmetry, periodic boundary conditions, a basis set cutoff energy of 700 eV, and a Monkhorst-Pack grid of minimum sample spacing (0.05 x 2π) $Å^{-1}$. The convergence criteria for geometry optimization were: 0.01 eV $Å^{-1}$ for forces, 0.00001 per atom for energy, and 0.001 Å for atomic displacements.

8.2.2 Crystal Structure Determination of the Anhydrous Phase of DBTMA from Powder XRD data

The powder XRD pattern of the anhydrous phase of DBTMA was indexed using the TAUP algorithm in the CRYSFIRE package [89,198], giving the following unit cell with orthorhombic metric symmetry: a = 7.21 Å, b = 13.25 Å, c = 16.44 Å (V = 1570 $Å^3$). Profile fitting and unit cell refinement were conducted using the Pawley method in the program TOPAS [85, 97]. Initially, space group $P222$

was used for the orthorhombic unit cell (R_{wp} = 1.02%, R_p = 0.76%; Figure 8.8. However, the β angle was subsequently allowed to relax in the Pawley fitting, this was to observe whether the fit would improve noticeably and/or whether the β angle would deviate noticeably from 90°. Using monoclinic space group *P*2, a slightly better fit to the powder XRD data was achieved (R_{wp} = 0.99%, R_p = 0.80%; Figure 8.9) with the following refined unit cell: *a* = 7.2085(4) Å, *b* = 13.2563(8) Å, *c* = 16.4421(17) Å, β = 90.086(7) ° (*V* = 1571.2(2) Å^3). The space groups C2, Pc, *P*2/m, *P*2/c, *P*2$_1$/m and *P*2$_1$/c were ruled out. However the space groups *P*2, *P*2$_1$ and Pm could not be ruled out by profile fitting, therefore independent structure solution calculations were conducted for these space groups.

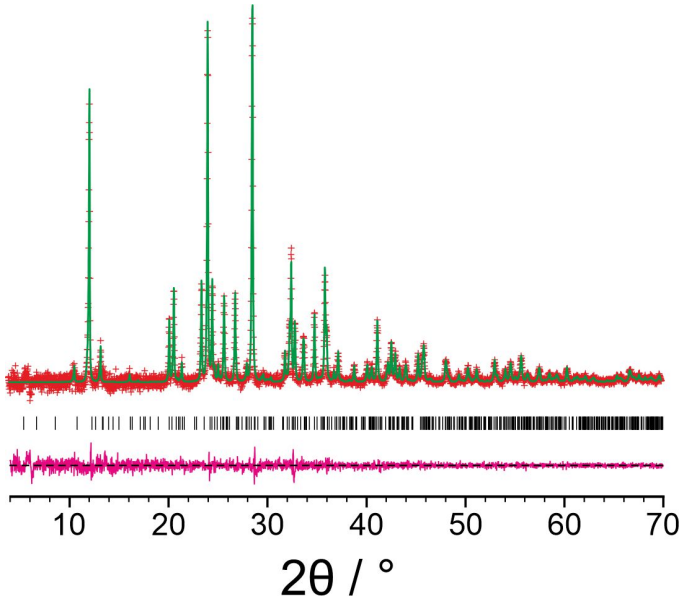

Figure 8.8: Profile fit using the Pawley method for the orthorhombic unit cell of the anhydrous phase of DBTMA, in the orthorhombic setting (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

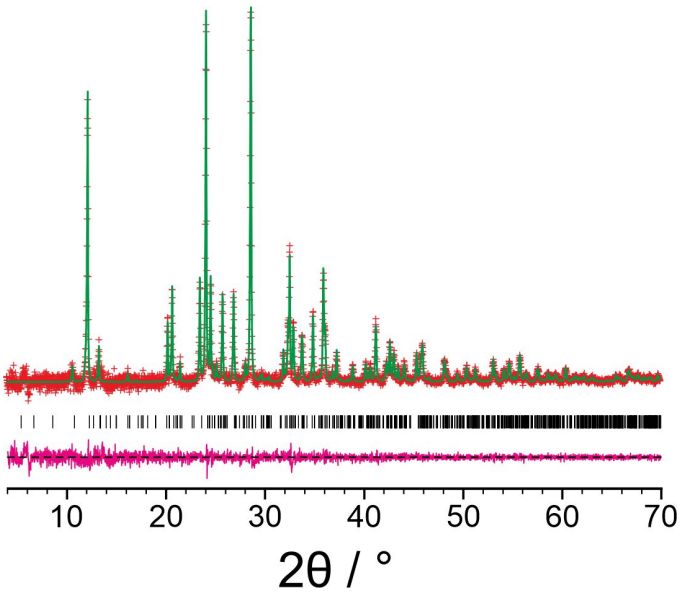

Figure 8.9: Profile fit using the Pawley method for the anhydrous phase of DBTMA allowing relaxation of the β angle for monoclinic space group $P2$ (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Structure solution was conducted using the direct-space genetic algorithm in the EAGER program [133–135]. One molecule of DBTMA was assumed to be present in the asymmetric unit as a result of density considerations. Three molecules of DBTMA were predicted to comprise the asymmetric unit, with a total of six molecules in the unit cell for each GA calculation for each space group, as every space group has two symmetry operators. Each trial structure was defined by twenty six structural variables for the $P2$ and $P2_1$ space groups: eight positional variables (the position along the *b*-axis may be fixed for both space groups for one molecule in the asymmetric unit), nine orientational variables and nine torsion-angle variables (defined in Figure 8.11). For space group Pm, the number of positional variables were seven, the position along the *b*-axis for one molecule in the asymmetric unit. The molecular model was generated and optimized within the GAUSSIAN program using the B3LYP methodology [56, 59]. The GA structure-solution calculation was conducted for a population of 100 trial structures and was run for 100 generations with 25 mating operations per generation and 10 mutation operations per generation. The number of mating and mutation operations was lowered for this structure. This was due to the intrinsic simplicity of the structural model leading to a very small number of variables, which allowed for a shorter calculation time due to a smaller search space. From visual inspection of the lowest R-factor (R_{wp}), structure for each calculation, the structures in space groups $P2$ and Pm were chemically unfeasible, whilst the structure for space group $P2_1$ was sensible and provided a good quality of fit to the powder XRD data. The three molecules in the asymmetric unit appeared to be related by translation along the *b*-axis, suggesting that the actual unit cell should correspond to one third of this unit cell, giving $Z' = 1$. Profile fitting for space group $P2$ using the Pawley method in TOPAS for the unit cell with the original *b*-axis divided by three gave a good fit to the powder

XRD data with the following refined unit cell: a = 7.2103(4) Å, b = 4.42010(10) Å, c = 16.4453(6) Å, β = 90.076(6)° (V = 1571.2(2) Å^3) (R_{wp} = 0.99%, R_p = 0.80%). The fit was reasonable except for a peak at 2θ = 28.1°, attributed to an unknown impurity phase; the fit is shown in Figure 8.10.

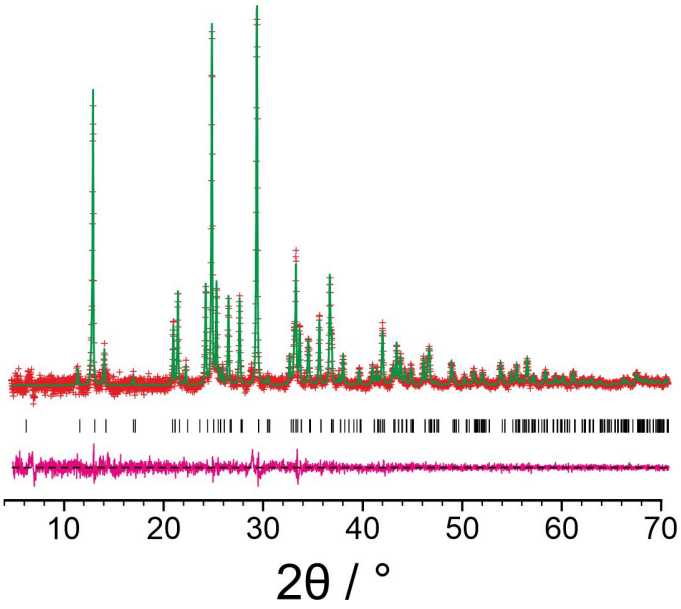

Figure 8.10: Profile fit using the Le Bail method for the smaller monoclinic unit cell of the anhydrous phase of DBTMA, in the monoclinic setting (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

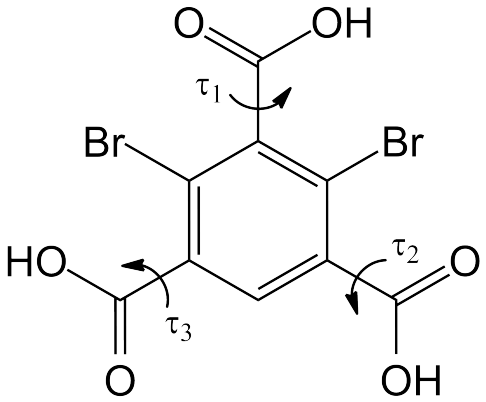

Figure 8.11: The molecular structure of DBTMA indicating the variable torsion angles τ_1, τ_2 and τ_3 in the structure solution calculation.

The EAGER calculation was then repeated on the smaller unit cell for the $P2_1$ space group. The structure was subjected to geometry optimization (with fixed unit cell) using periodic DFT-D calculations using the program CASTEP. Due to the low X-ray scattering power of hydrogen atoms, the positions of hydrogen atoms cannot be determined be accurately from powder XRD data, so they were fixed in the structure solution calculations. Thus, the assignment of C=O and C-OH groups within each carboxylic acid substituent was determined by inspecting the local geometry to

establish the most favourable hydrogen-bonding arrangement based on geometric criteria, together with periodic DFT-D calculations.

A series of structural models were generated based on the structure from geometry optimization using periodic DFT-D calculations. They were generated by moving the hydrogen atom from the oxygen atom of each carboxylic acid group and moving it to the other oxygen atom of the same group. The C-O—H bond angle was then rotated until it engaged in satisfactory hydrogen bonding before being optimized using a periodic DFT-D calculation. The energies of each structure were compared to determine the hydrogen-bonding arrangement of lowest energy. The numbering of the molecule is given in Figure 8.12.

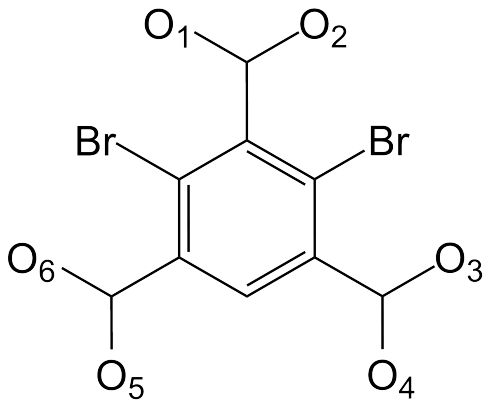

Figure 8.12: The numbering of the oxygen atoms within the carboxylic acid groups in DBTMA.

The structure of lowest energy, as detailed in Table 8.1, following DFT-D geometry optimization with fixed unit cell corresponds to hydrogen bonding configuration 4, which also corresponds to the most geometrically optimal hydrogen bonding arrangement. However, visual inspection of the structure at this stage suggested a possible mirror plane perpendicular to the aromatic ring and parallel to the plane of the carboxylic acid group between the two bromine atoms, as highlighted in Figure 8.13. The torsion angles defining the orientations of the carboxylic acid groups relative to the aromatic ring were -164.53 ° and 163.99 ° respectively. The carboxylic acid groups are involved in O—H···O hydrogen bonding to neighbouring molecules (1.650 Å, 175.66 °; 1.655 Å, 175.35 °).

Hydrogen bonding configuration	Assignment of O—H groups	Relative energy / kJ mol^{-1}
1	O_2-H, O_4-H, O_6-H	34.05
2	O_2-H, O_4-H, O_5-H	15.51
3	O_2-H, O_3-H, O_6-H	20.65
4	O_2-H, O_3-H, O_5-H	0.00

Table 8.1: Relative energies for the various configurations of hydrogen bonding for the carboxylic acid groups in anhydrous DBTMA.

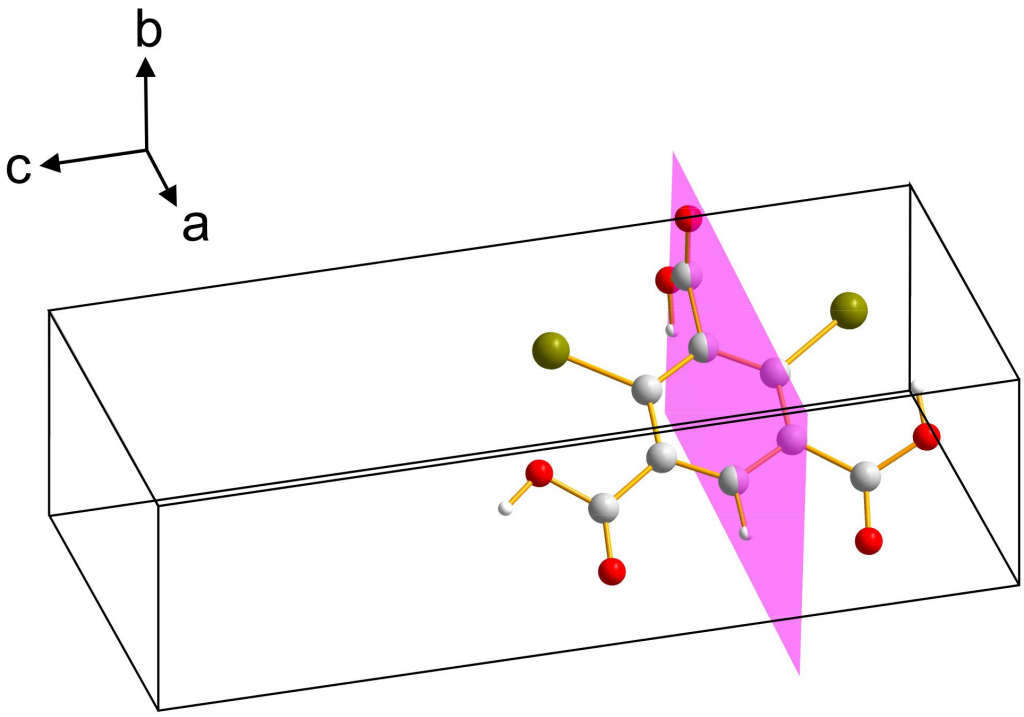

Figure 8.13: The crystal structure of DBTMA in the small unit cell with the $P2_1$ space group, with the subsequently added mirror plane identified with the solid magenta line.

A test for higher symmetry, conducted using the PLATON program [186], suggested that the structure may be compatible with the orthorhombic space group $Pmn2_1$. Accordingly, the structure was transformed from $P2_1$ to $Pmn2_1$, with the 2_1 screw axis remaining along the 4.4 Å axis (now the *c*-axis) and the plane of symmetry of the molecule coinciding with the mirror plane parallel to the bc-place in $Pmn2_1$. In this space group, the asymmetric unit comprises one half molecule (Z' = 1/2). Profile fitting was conducted using the Pawley method for space group $Pmn2_1$, giving a good fit to the powder XRD data (R_{wp} = 1.11%, R_p = 0.81%) with the following refined unit cell: a = 16.4441(4) Å, b = 7.2092(3) Å, c = 4.41979(13) Å (V = 523.96(3) Å^3). The fit is shown in Figure 8.14.

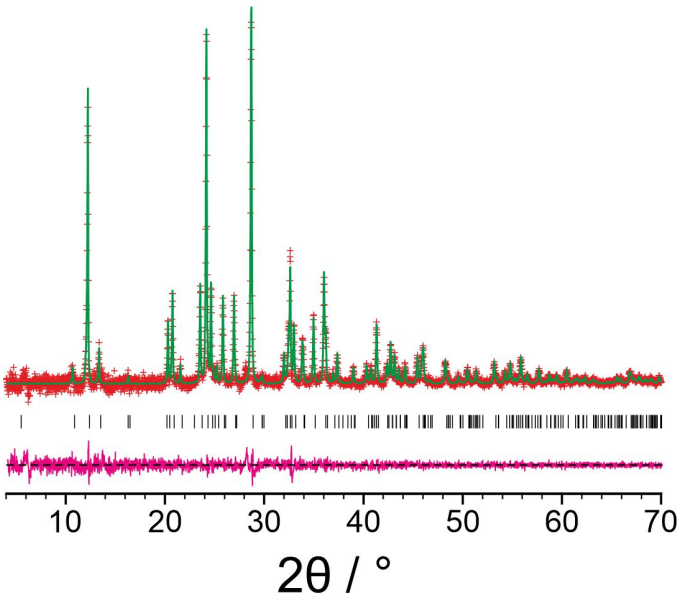

Figure 8.14: Profile fit using the Pawley method for the anhydrous phase of DBTMA based on the orthorhombic space group $Pmn2_1$ (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

This structure was used as the starting point for Rietveld refinement using TOPAS [105]. Restraints were applied to the bond lengths and angles based on the molecular geometry obtained by DFT-D geometry optimization of the crystal structure, and planar restraints were applied to each carboxylic acid group, to the aromatic ring, and to the atoms directly bonded to the aromatic ring (i.e. the bromine atoms and the carbon atoms of the carboxylic acid groups). A common isotropic displacement parameter was refined for the non-hydrogen atoms, and the value for the hydrogen atoms was set to 1.2 times this value. The *x*-coordinates of the atoms located on the mirror plane were fixed. Each pair of hydrogen bonds (both donor and acceptor groups from the carboxylic acid groups) for all three carboxylic acid groups had bond length and angle restraints applied. The Rietveld refinement gave a good quality of fit between the calculated and experimental data (R_{wp} = 1.20%, R_p = 0.87%) with the following refined unit cell: a = 16.4445(4) Å, b = 7.2093(2) Å, c = 4.41995(12) Å (V = 524.00(2) Å^3). The final Rietveld fit is shown in Figure 8.15.

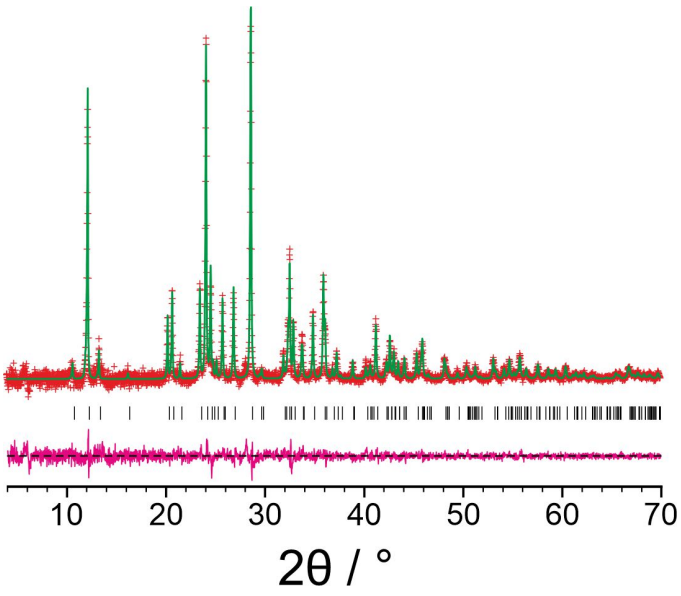

Figure 8.15: Final Rietveld fit for the anhydrous phase of DBTMA in the $Pmn2_1$ space group (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

The final refined structure was subjected to periodic DFT-D geometry optimization (with fixed unit cell) using CASTEP, which led to only very minor structural changes (RMSD for non-hydrogen atoms, 0.14 Å), confirming that the refined structure is close to a minimum on the energetic landscape. An overlay of the DBTMA structure before and after DFT-D geometry optimization is shown in Figure 8.16.

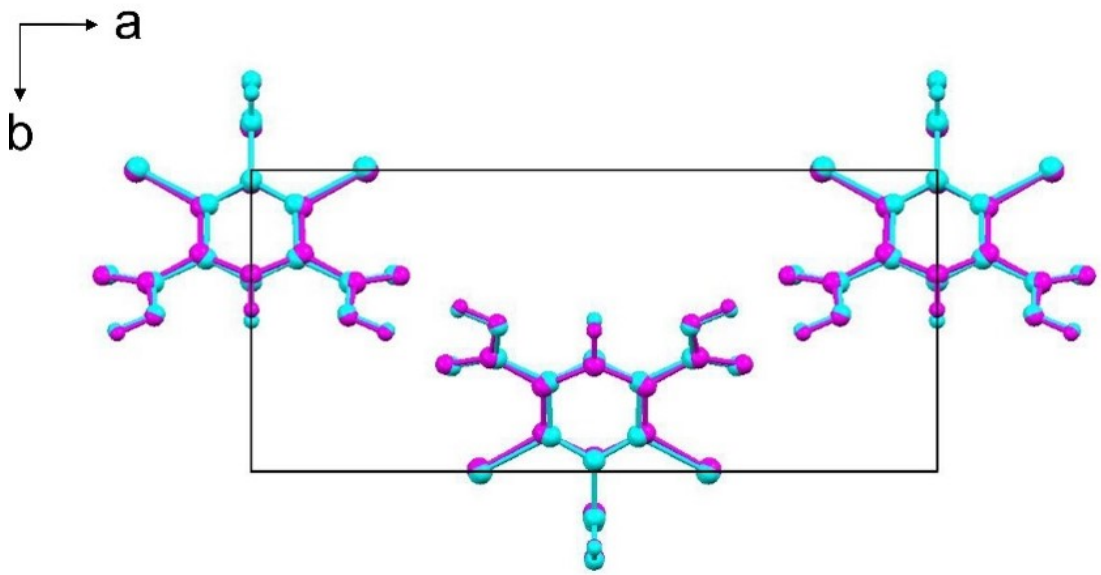

Figure 8.16: Overlay of the crystal structure of the anhydrous phase of DBTMA before (magenta) and after (cyan) DFT-D geometry optimization (with fixed unit cell).

8.2.3 Structural Discussion

Structural Properties of the Dihydrate Phase of DBTMA

The crystal structure of the dihydrate phase of DBTMA is shown in Figure 8.18 and structural data are given in Table 8.2. The structure comprises of stacks of DBTMA molecules along the *c*-axis, where the main interaction is the $\pi\cdots\pi$ stacking of the phenyl rings, which are almost completely in line with one another when viewed along the *c*-axis. Adjacent molecules along the slack are rotated close to 120° from one another when viewed along the *b*-axis, which is likely to accommodate the bromine atoms. One water molecule is an acceptor for an O—H···O hydrogen bond (2.558 Å, 169.48°) from the carboxylic acid group at position 3 (as assigned in Figure 8.17, which is short and linear. This water molecule is also the donor in a longer and less linear O—H···O hydrogen bond (2.822 Å, 149.09°) to the C=O of a carboxylic acid group. The second water molecule is the acceptor for an O—H···O hydrogen bond involving the carboxylic acid group at position 5 as the donor (2.714 Å, 172.41°), and this water is also the donor in a long and non-linear O—H···O interaction (2.869 Å, 128.96 °). The carboxylic acid group at position 1 is the donor in an O—H···O hydrogen bond to a water molecule (2.626 Å, 151.83 °), whereas those at positions 3 and 5 are donors in somewhat shorter and more linear O—H···O hydrogen bonds (2.558 Å, 169.48 °; 2.714 Å, 172.41°), as shown in Figure 8.19.

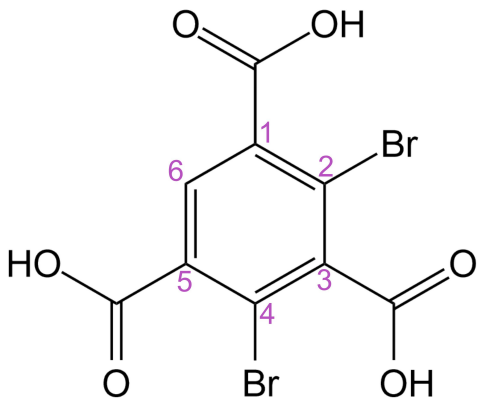

Figure 8.17: The structure of a molecule of DBTMA with the position numbering of all carbon atoms denoted in magenta.

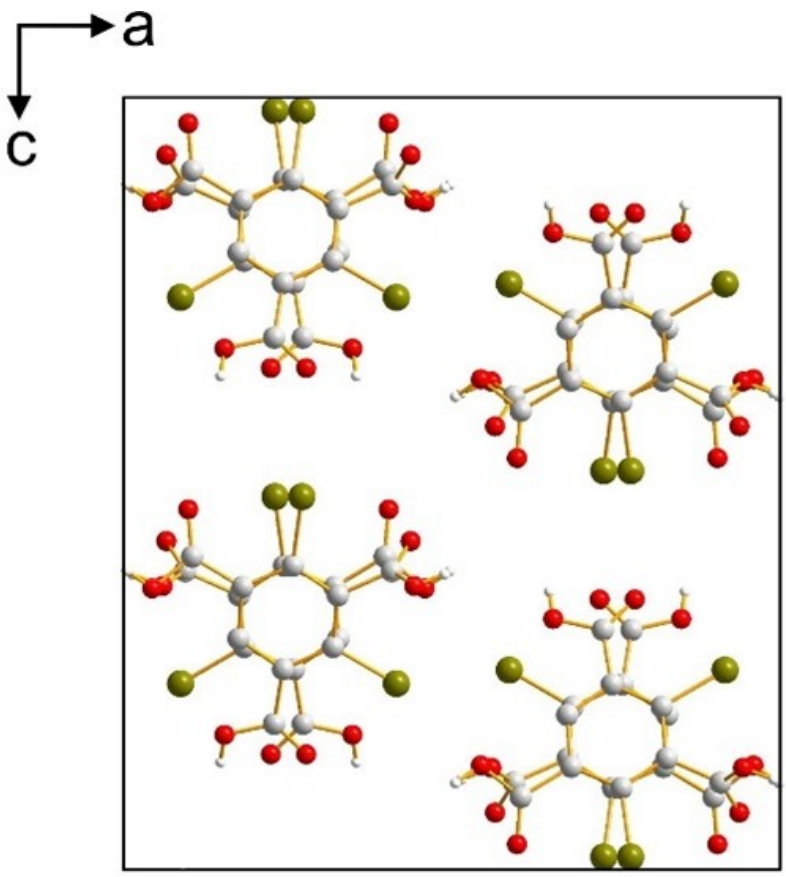

Figure 8.18: Crystal structure of the dihydrate phase of DBTMA viewed along the *b*-axis, with water molecules excluded.

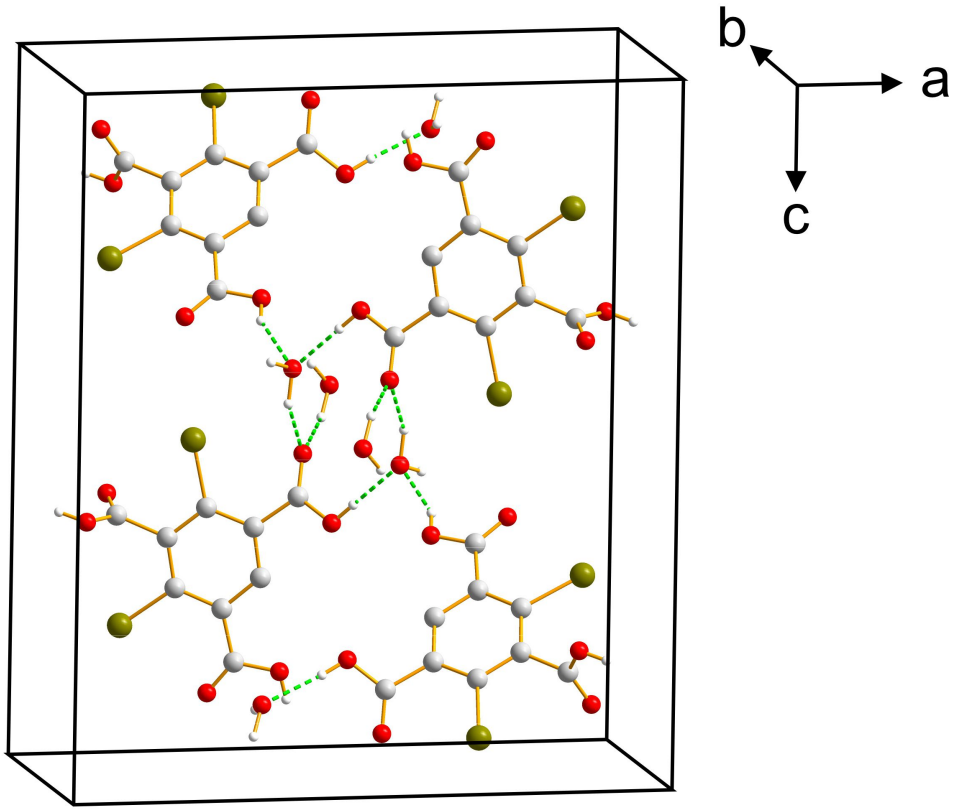

Figure 8.19: The hydrogen bonding between DBTMA molecules and the two independent water molecules in the crystal structure of the dihydrate phase of DBTMA, viewed along the *b*-axis.

Structural Properties of the Anhydrous Phase of DBTMA

The crystal structure of the anhydrate phase of DBTMA is shown in Figures 8.20 and 8.21, and structural data are given in Table 8.2. The asymmetric unit comprises one half molecule, with the DBTMA molecule located on a crystallographic mirror plane, specifically (100). The carboxylic acid group between the two bromine atoms lies in this mirror plane and is exactly perpendicular to the plane of the phenyl ring, clearly avoiding unfavourable steric interactions with the adjacent bromine atoms on the ring. The carboxylic acid groups at the 1 and 5 carbon positions on the phenyl ring are related to each other by the mirror plane and therefore have an identical geometric relation to the aromatic ring. Both of these carboxylic acid groups are close to parallel to the plane of the ring ($\tau_2 = \tau_3 = 22.40^\circ$). The structure contains two networks of hydrogen bonds. One network propagates along the *c*-axis and involves the carboxylic acid groups at the 3-position on the phenyl ring (between the bromine atoms), giving rise to a periodically repeating chain of C=O···H—O interactions (2.674 Å, 174.12°), along the *c*-axis. Each carboxylic acid group involved in this chain serves as both the donor and acceptor in interactions of this type to adjacent molecules in the chain

(Figure 8.20). The second network involves the other two carboxylic acid groups in each molecule and propagates as a helical chain where the chains are of opposite handedness, of C=O···H—O hydrogen bonds along the *c*-axis (2.607Å, 154.94°), as highlighted in Figure 8.21.

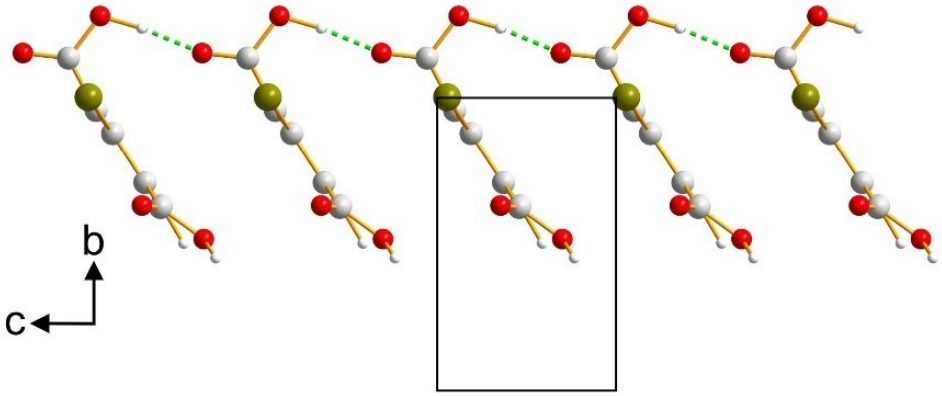

Figure 8.20: The hydrogen-bonded chain involving the carboxylic acid groups at the 3-position in the phenyl ring in the crystal structure of the anhydrous phase of DBTMA, viewed along the *a*-axis. Hydrogen bonds are indicated by green dashed lines.

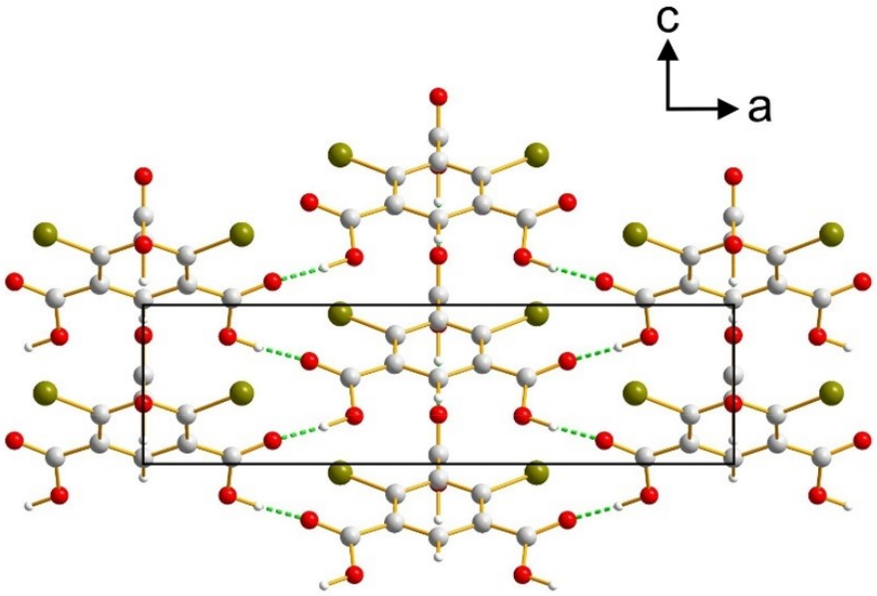

Figure 8.21: The hydrogen-bonded chain involving the carboxylic acid groups at the 1 and 5 positions in the phenyl ring in the crystal structure of the anhydrous phase of DBTMA, viewed along the *b*-axis. Hydrogen bonds are indicated by green dashed lines.

The intermolecular geometries of the DBTMA molecules in both the hydrate and anhydrous phases are similar, the carboxylic acid group between the two bromine substituents is close to perpendicular to the aromatic ring (τ_1 = 0° for the anhydrate structure, τ_2 = 85.57° for the dihydrate structure).

The other two carboxylic acid groups are close to coplanar with the aromatic ring (τ = 22.40° for the anhydrate structure, τ = -1.40° and τ = 26.08° for the dihydrate structure).

Structure	Dihydrate	Anhydrate
Molecular formula	$C_9H_4Br_2O_6 \cdot 2H_2O$	$C_9H_4Br_2O_6$
a / Å	16.6279(3)	16.4445(4)
b / Å	7.7983(2)	7.2093(2)
c / Å	19.5444(4)	4.41995(12)
Volume / Å^3	2534.31	524.00(2)
Space group	Pbca	$Pmn2_1$
Z	8	4
Z'	1	0.5
Density / g cm^{-3}	2.11	2.33
Temperature / K	293	293
Method	Single-crystal XRD	Powder XRD

Table 8.2: Structural parameters for the dihydrate and anhydrous phases of DBTMA.

8.3 Structure Determination of Cyclohexyl Functionalized Perylene Diimide (PDICH)

8.3.1 Experimental Methods

Sample Preparation

The PDICH 'nanobelts' were prepared through self-assembly of PDICH using a phase transfer. A solution of PDICH (0.5 mL, 0.03 mM) was prepared in chloroform and stored within a small glass vial. This solution was then injected into ethanol (11:3 ratio of ethanol to PDICH/chloroform solution, by volume), which was then allowed to sit in a dark room at room temperature for a 24 hour period. During this time, the nanobelts formed at the solvent interface. The solvents were then mixed, allowing the nanobelts to diffuse into the entire solution phase, after which they were pipetted onto glass coverslips. Solvent evaporation from the cover slips was conducted in a glove box (N_2 atmosphere) prior to extraction of the nanobelts [196].

Powder X-ray Diffraction

Powder XRD data were recorded in transmission mode at 293 K on a Bruker D8 Advance Series II diffractometer (Ge-monochromated Cu-Kα_1 radiation) between 4°-70° in 2θ (step size; 0.016°) using a Vantec detector covering 3° in 2θ. The polycrystalline sample of PDICH was loaded between two pieces of tape before being placed on a foil type sample holder. The data collection time was 15.5 hr.

Periodic DFT-D Calculations

Periodic DFT-D calculations to validate the structure of PDICH determined from powder XRD data were conducted using CASTEP [122] (Academic Release Version 19.1.1). The geometry optimization calculations used ultrasoft pseudopotentials [123], semi-empirical dispersion corrections [125] (TS correction scheme), PBE functional [124], fixed unit cell and preserved space group symmetry, periodic boundary conditions, a basis set cut-off energy of 700 eV and a Monkhorst-Pack grid [126] of minimum sample spacing 0.05 x 2π Å^{-1}. Convergence criteria for geometry optimization were 0.01 eV Å^{-1} for forces, 0.00001 eV per atom for energy, and 0.001 Å for atomic displacements.

8.3.2 Crystal Structure Determination

The powder XRD pattern recorded for PDICH was consistent with an orthorhombic unit cell (a = 22.9 Å, b = 37.6 Å, c = 7.2 Å, V = 6199.5 Å^3) and space group ($Pn2_1a$) determined previously from 3D-ED measurements [196]. Profile fitting was conducted using the Le Bail method in the GSAS program [96, 132], which led to the following refined unit cell: a = 21.1956(17) Å, b = 7.08567(15) Å,

c = 36.766(4) Å (V = 5521.7(9) Å^3), for the Pna2$_1$ setting of the same space group. A good quality of fit was obtained (R_{wp} = 0.80%, R_p = 0.57%), as seen in Figure 8.22. We note that, for the unit cell refined against powder XRD data, the 2$_1$ screw axis (*ca.* 37 Å axis) is the *c*-axis, rather than the *b*-axis in the space group setting derived from 3D-ED data, allowing structure solution from the powder XRD data to be conducted in the conventional setting of the space group.

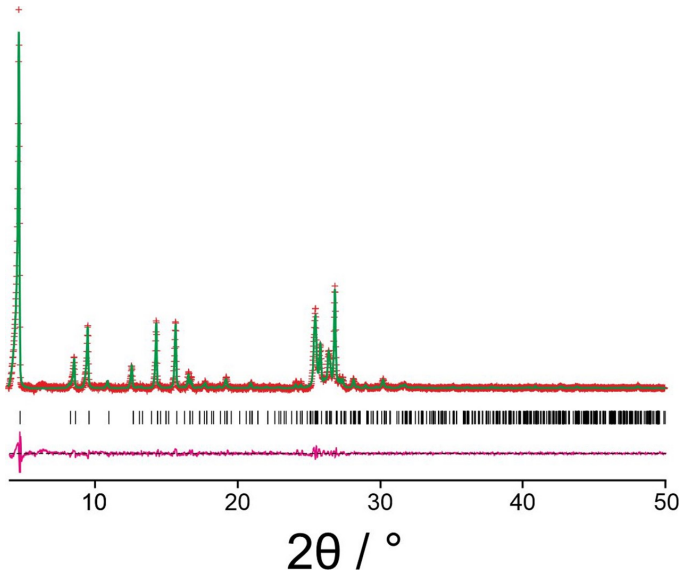

Figure 8.22: Profile fit using the Le Bail method for PDICH (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

Structure solution was conducted using the direct-space genetic algorithm in the EAGER program. The molecular model of PDICH was generated by optimizing the geometry of a model constructed in WebMO by using the GAUSSIAN program. A B3LYP/6-31G(d) setting was using and the resulting molecular model was used in the structure solution calculations (Figure 8.23). Based on density considerations, there are two molecules of PDICH in the asymmetric unit. In total, the direct-space structure solution calculation involved a total of 15 structural variables: five positional variables (the *z*-coordinate can be fixed for one molecule in the asymmetric unit for space group Pna2$_1$), six orientational variables, and four torsional variables.

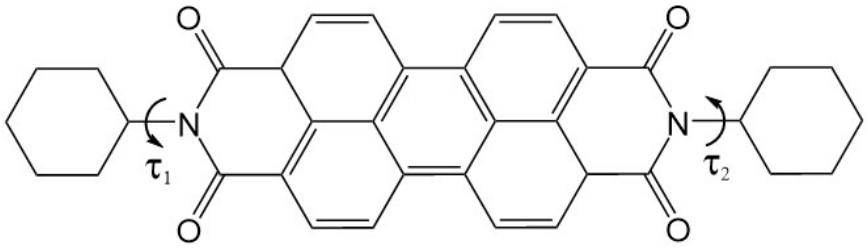

Figure 8.23: Molecular structure of PDICH, showing the torsion angles (τ_1 and τ_2) that were allowed to optimize in structure solution calculations.

The trial structure of lowest R-factor from the EAGER calculation was used as the starting point for Rietveld refinement using GSAS [105]. Restraints were applied to the bond lengths and angles, and

multiple overlapping planar restraints were applied to ensure the perylene ring system remained planar during refinement. A common isotropic displacement parameter was allowed to refine for the non-hydrogen atoms, with the value for hydrogen atoms taken as 1.2 times this value. The Rietveld refinement gave a quality of fit (R_{wp} = 1.08%, R_p = 0.77%) comparable to the quality of fit obtained in profile fitting, with the following refined unit cell: a = 21.178(4) Å, b = 7.08423(27) Å, c = 36.647(9) Å (V = 5498.2(19) Å^3). The final Rietveld fit is shown in Figure 8.24.

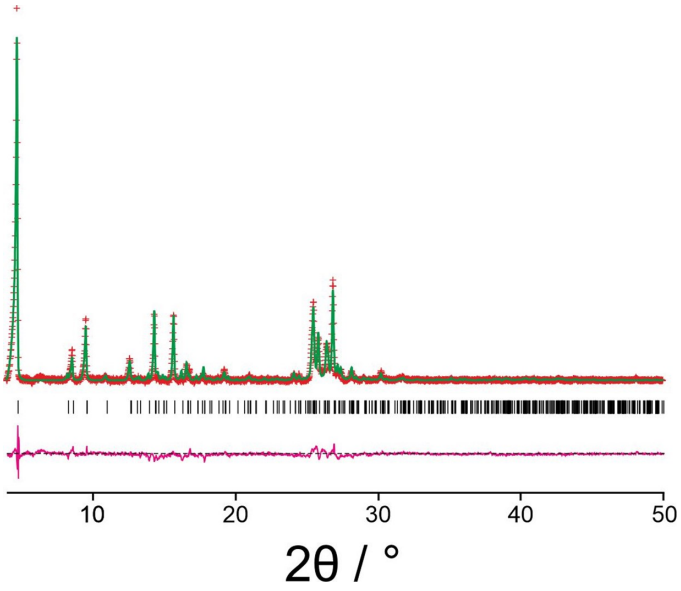

Figure 8.24: Final Rietveld fit for PDICH, in the $Pna2_1$ space group (predicted peaks – black ticks; difference plot – magenta; calculated data – green; experimental data – red).

The final refined structure was subjected to periodic DFT-D geometry optimization (with fixed unit cell) using CASTEP, which led to only very minor structural changes (RMSD for non-hydrogen atoms, 0.45 Å), confirming that the refined crystal structure is close to a minimum on the energy landscape. An overlay of the structure before and after geometry optimization is shown in Figure 8.25; it is clear to see from this overlay that the "core" perylene diimide ring system exhibits the smallest deviation in RMSD, whereas the cyclohexyl rings are the main contributors to the RMSD, which is relatively high compared to other examples presented in this thesis. It is conceivable that the cyclohexyl rings are in some way disordered compared to the core PDI ring system, even if not totally disordered. This would be difficult to model via Rietveld refinement containing a single model, and therefore the high deviation between pre and post-optimized structures is justified.

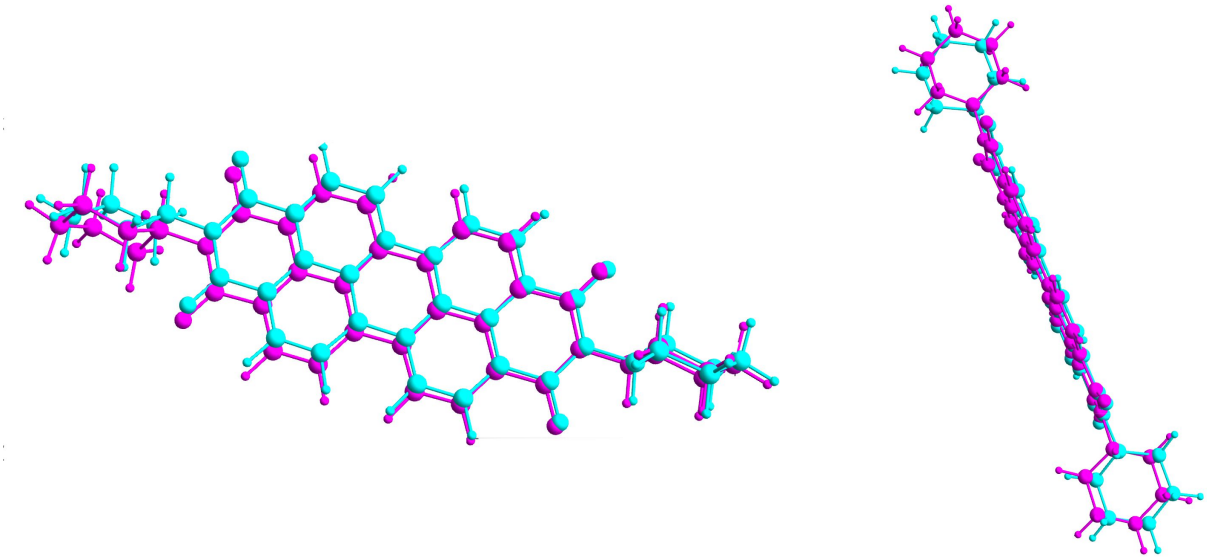

Figure 8.25: Overlay of the structure of PDICH before (magenta) and after (cyan) DFT-D geometry optimization (with fixed unit cell).

8.3.3 Structural Discussion

The crystal structure of PDICH (shown in Figure 8.26) has two independent molecules in the asymmetric unit. The molecules stack along the *c*-axis (which contains the 2_1 screw-axis symmetry operation) in a way that ensures the cyclohexyl groups do not clash and do not create an unnecessary number of empty voids. Along the *b*-axis the molecules stack alternately whilst the mean planes of the cyclohexyl groups are parallel to the direction of the *b*-axis. In one molecule, the two torsion angles defining the orientation of the cyclohexyl rings with respect to the PDI system are 3.07° and 3.94° respectively, which are extremely similar, indicating that the conformation of this molecule is close to centrosymmetric. For the other molecule, the torsion angles are 14.23° and -15.98° respectively. The torsion angles are measured by plane of the first carbon atom of the PDI ring connected to the nitrogen atom preceding the cyclohexyl group, and the plane of the first carbon of the cyclohexyl group and the hydrogen atom attached (not part of the cyclic chain of carbon atoms). As this hydrogen is perpendicular to the mean plane of the cyclohexyl group it is the most reliable indication of the orientation of the cyclohexyl ring compared to the PDI ring system. The closer this torsion angle is to 0°, the more perpendicular the cyclohexyl ring is to the PDI ring system. As these torsion angles are close to 0°, it is clear to see that all four unique cyclohexyl groups in the structure are almost perpendicular to their respective PDI ring systems which is likely to avoid steric clashes between other cyclohexyl groups in neighbouring molecules.

Another interesting aspect of the structure, seen in Figure 8.27, concerns the relative orientations of the PDI ring systems of adjacent molecules along the stacks. The orientations of adjacent molecules differ by an angle of *ca.* 30°, where the middle phenyl ring in the PDI ring systems are almost overlapping, as can be seen in Figure8.26. The alternation of orientations of adjacent

molecules along the stack, combined with the cyclohexyl substituents being close to perpendicular to the PDI rings, allows a reasonable degree of packing throughout the whole structure and prevents any steric conflict between the cyclohexyl rings. In general, the stacks can be considered to be arranged in "slabs" parallel to the *ab*-plane, with the cyclohexyl substituents located at the interface between adjacent slabs.

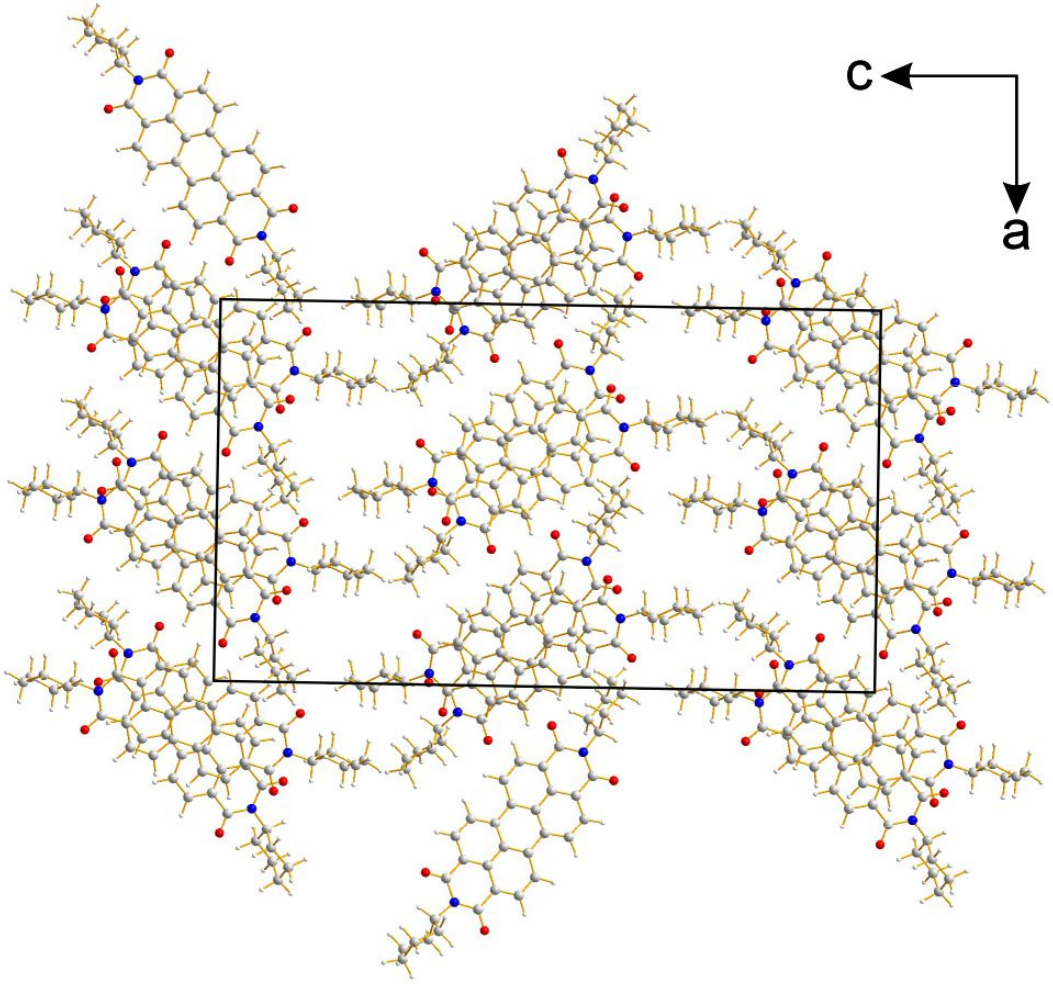

Figure 8.26: The crystal structure of PDICH viewed along the *b*-axis (the direction of the stacking axis).

8.4 Conclusions

This chapter has detailed a variety of methods that can be incorporated to determine crystal structures of symmetrical molecules with rigid cores. The study of DBTMA has shown that, due to the degradation of crystal quality in forming the anhydrous phase by dehydration of the dihydrate phase, powder XRD is the only reasonable method for determination of the crystal structure of the resulting phase.

A further development in the multi-technique approach was the determination of specific rotations

of carboxylic acid groups, which through powder XRD data alone is impossible due to the negligible scattering of the hydrogen atoms by X-ray radiation. Empirical evaluation of the possible hydrogen bond lengths and angles can be performed; however, in this case, where either rotation can provide a perfectly suitable crystal structure, more rigorous methods were applied. Each carboxylic acid group was rotated and a geometry optimization was performed using periodic DFT-D methods. It was clear that one hydrogen-bonding configuration was noticeably lower in energy than the others, which was not clear through visual inspection of the structure, highlighting the importance of vigorous validation of crystal structures following refinement.

The second example focused on the determination of the PDICH crystal structure from powder XRD data (and validation using periodic DFT-D calculations), following initial structure solution using 3D-ED data which yielded an inadequate quality representation of the structure. The unit cell and space group were independently determined from 3D-ED data, which circumvented one of the most challenging stages of crystal structure determination from powder XRD data. Thus, the starting values for the profile fitting using the Le Bail method were derived from the 3D-ED data, although the refined unit cell from profile fitting of the powder XRD data is significantly more accurate than that determined from 3D-ED data. The PDICH molecule could contain a σ_v plane of symmetry (for a specific conformation of the CH substituents); therefore, it would be conceivable for the internal mirror plane to be superimposed on the crystal structure mirror plane, as in DBTMA, leading to a Z' value of 0.5. However, in this case, due to the stacking of the PDI rings (offset from one another along the *b*-axis, as shown in Figure 8.27) and the conformation of cyclohexyl groups, this situation does not occur, and there are in fact two independent molecules of PDICH in the asymmetric unit. The structure is primarily dictated by forming $\pi\cdots\pi$ stacking between PDI ring systems, with the conformations of the cyclohexyl groups driven by the need to avoid steric conflicts.

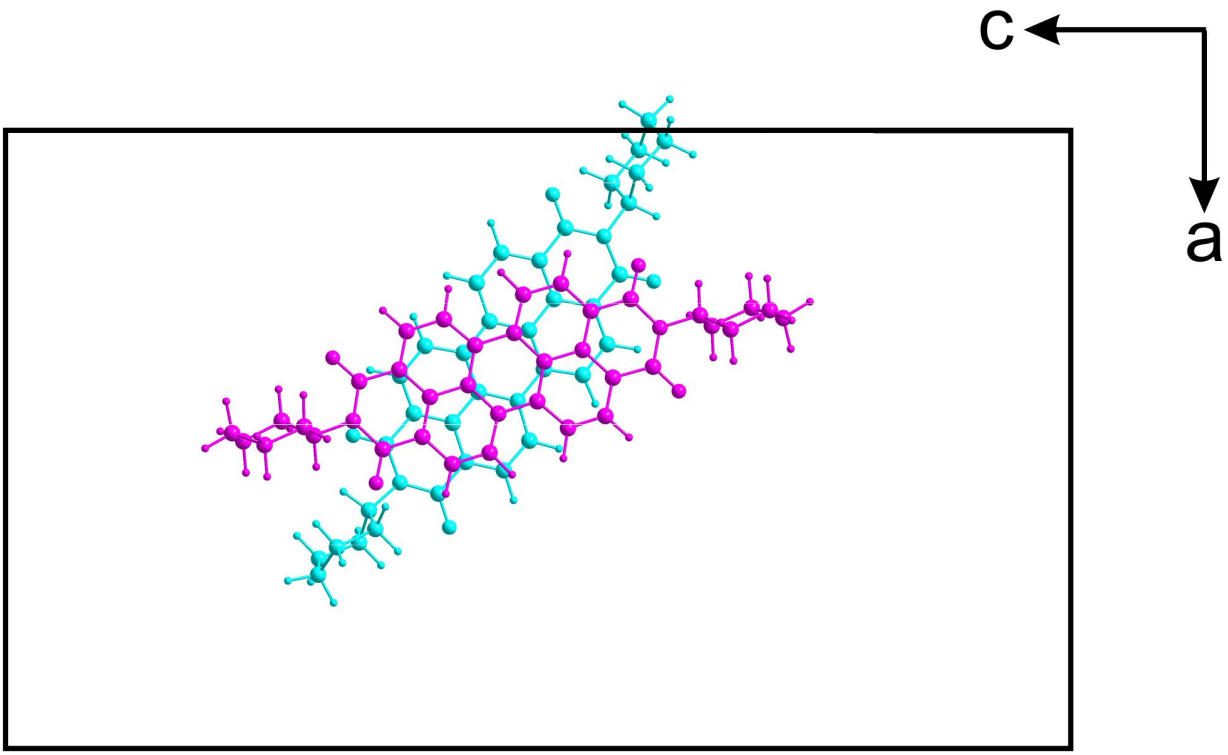

Figure 8.27: The crystal structure of PDICH viewed along the *b*-axis with the cyclohexyl rings excluded and the molecules toward the front of the page in magenta, with those toward the back in cyan.

The determination of the PDICH crystal structure has shone light on another approach for determining the crystal structures of symmetrical molecules with rigid cores; however, even more importantly it provides a methodology for determining a crystal structure from powder XRD data with inspiration from 3D-ED methods. For example, the unit cell determination and space group determination processes are circumvented in this stage, but powder XRD data is important for a straightforward and accurate crystal structure refinement (in contrast, accurate structure refinement from 3D-ED data requires dynamical scattering calculations). It is due to this use of the Rietveld method that the crystal structure determined by powder XRD data showed a significantly improved and more plausible molecular geometry. The poor molecular geometry in the published work was partly due to this aspect, but also because it was simply a structure solution by traditional methods. Direct-space structure solution provides a significant advantage as it already contains a reliable representation of the molecular geometry.

Chapter 9
Conclusions

To conclude, this thesis describes the latest advancement in structure determination from powder XRD data. This advancement has been performed by augmenting the existing methodologies for structure determination from powder XRD data with complementary structural methods, such as: solid-state NMR data (both experimental and theoretical), density functional theory, 3-dimensional electron diffraction data and crystal structure prediction. This work has shown how they can overlap with one another to form a cohesive, yet fluid approach to determine crystal structures directly from powder XRD data. As described in Chapter 3, there are numerous pitfalls that may be an issue in this process, most notably difficulty in indexing due to the presence of an impurity phase, as well as difficulty in crystal structure solution when severe preferred orientation has affected the distribution of peak intensities in powder XRD data. By employing these complementary methods, the aforementioned issues can be evaded, as well as other ambiguities in crystal structure determination, usually in the final stages of Rietveld refinement and subsequent structural validation. A wide variety of materials have been studied, some with biological relevance and others with applications in optoelectronic materials. This showcases the need for: i) performing powder XRD experiments on these types of materials to gain important structural insights, ii) advanced protocols to be in place for crystal structure determination using said powder XRD data, and iii) that every material showcases their own intrinsic structural challenges, whereby the importance of a multi-technique approach for crystal structure determination is paramount.

Chapter 10
Future Work

The work presented in this thesis represents a significant advance in the use of complementary techniques to augment structure determination from powder XRD data. However, several avenues of research arising from this work should be explored in the future and are discussed in this chapter.

In Chapter 4, the crystal structures of riboflavin and alloxazine were determined from powder XRD data. The section on riboflavin noted that other phases of riboflavin have been reported by Mohnicke *et al*. [136]. A clear direction for future work is to conduct crystallization to produce and isolate these phases of riboflavin, with the aim of determining the crystal structures subsequently determined using the protocols presented in this thesis. It was also mentioned that while riboflavin has been extracted previously from the ocular tissues in lemurs [39], the specific phase of riboflavin in these biogenic samples was not recorded. Further work to extract biogenic samples of riboflavin from the eyes of animals such as cats and lemurs would be an interesting direction for future research, with a view to structure determination from powder XRD data. This is so that the polymorph/s present in the biological context (the biogenic form) can be determined. There is a possibility that the rod-shaped crystals in felines and plate-shaped crystals in lemurs could be caused by different polymorphic forms.

With respect to alloxazine, it was concluded from the present work that the proportion of the alloxazine tautomer was more than 99.9% based on consideration of the relative energies of the crystal structure containing the alloxazine and isoalloxazine tautomers. Further investigation of alloxazine by neutron diffraction would provide more definitive experimental evidence of whether a detectable amount of the isoalloxazine tautomer is present in the crystal structure this, as neutron diffraction is far more sensitive to the presence of hydrogen atoms. Thus, neutron diffraction has the potential to provide direct experimental evidence on this issue as opposed to the indirect evidence inferred from periodic DFT-D modelling. Further experiments at higher temperature could be conducted to determine whether alloxazine does, or does not, exhibit interconversion to isoalloxazine at elevated temperature.

The work in Chapter 5 concerned the discovery and subsequent crystal structure determination from powder XRD data, in conjunction with other techniques, of the β polymorph of L-tyrosine. The work provided inspiration on how electron diffraction data can supplement the indexing and structure solution stages of the structure determination process from powder XRD data. The scope

for future work however concerns the crystal structure prediction calculations. Three hypothetical crystal structures were found in the crystal structure prediction calculations, which provides the inspiration and motivation for future endeavours to produce these materials and identify them experimentally. The periodic DFT-D calculations suggest the predicted structures A, B and C are metastable relative to the two known polymorphs (α and β polymorphs) of L-tyrosine. However, as structures B and C are much higher in energy than the α polymorph, it is likely that only structure A that has a chance to be observed experimentally. Therefore unconventional methods of crystallization, as described in this thesis, may provide the best prospects for producing these materials. These methods can include, but are not limited to, crystallization from the gas phase following sublimation, crash cooling from a molten phase, and mechanochemical grinding.

Chapter 6 concerned the crystal structure determination of DL-lysine and two hydrate phases of DL-lysine, with the existence of disorder in the crystal structures of the anhydrous and hemihydrate phases probed by periodic DFT-D calculations. A clear avenue for further exploration will be to prepare pure phases of the three remaining amino acids for which a racemate crystal structure has not yet been reported (DL-asparagine, DL-phenylalanine, and DL-threonine). This would ideally be followed with crystal structure determination following the protocols described in this thesis. The methodology for determining crystal structures presented in this thesis could be followed, including the application of other structural methods such as solid-state NMR. If disorder is inferred from inspection of the crystal structure, the approach used here for DL-lysine phases using periodic DFT-D calculations would provide valuable insights on the nature of the disorder.

In Chapter 7, the discovery and subsequent crystal structure determination from powder XRD data of a new polymorph of ibuprofen was presented. While the structure was validated by periodic DFT-D calculations, the high energy of this meta-stable polymorph suggests that preparing a pure phase that would survive for a significant period of time may be challenging. Suitable future work for this material would be to attempt to re-prepare the material and to attempt to conserve the new polymorph in a way that would allow complementary structural techniques to be applied; for example, solid-state NMR would be valuable to provide validation of the crystal structure presented here. Furthermore, *in-situ* powder XRD data could be recorded using a large area detector to collect data rapidly to provide insights on the transformational pathway of this meta-stable polymorph as it evolves into polymorphs of greater stability.

Finally, in Chapter 8, both the crystal structures of DBTMA and PDICH were determined from powder XRD data. The structure of DBTMA is a derivative of trimesic acid, and the crystal structure was validated and optimized using periodic DFT-D calculations to determine an accurate description of the hydrogen bonding arrangement. The same methodology could be applied in the structure determination of other derivatives of trimesic acid from powder XRD data, particularly those that have the internal molecular symmetry of a mirror plane located on a crystallographic mirror plane, as in the case of DBTMA. The PDICH material is a derivative of perylene diimide functionalized with

cyclohexyl groups, but it is known that other derivatives of perylene diimide also exhibit interesting optoelectronic properties. The same approach for structure determination from powder XRD data and 3D-ED data could be applied in the case of other derivatives, and other structural methods such as solid-state NMR can be utilized. Evidently, obtaining reliable information on the crystal structures of these materials is critical to understand their optoelectronic properties.

Chapter 11
Bibliography

[1] W. Jones. Taylor & Francis Group, 1st edition, 1997.

[2] G. K. Lim, K. Fujii, K. D. M. Harris, and D. C. Apperley. *Crystal Growth & Design*, 11(12):5192–5199, 2011.

[3] K. D. M. Harris, M. Tremayne, P. Lightfoot, and P. G. Bruce. *Journal of the American Chemical Society*, 116(8):3543–3547, 1994.

[4] A. E. Watts, K. Maruyoshi, C. E. Hughes, S. P. Brown, and K. D. M. Harris. *Crystal Growth & Design*, 16(4):1798–1804, 2016.

[5] C. E. Hughes, G. N. M. Reddy, S. Masiero, S. P. Brown, P. A. Williams, and K. D. M. Harris. *Chemical Science*, 8(5):3971–3979, 2017.

[6] C. E. Hughes, I. Boughdiri, C. Bouakkaz, P. A. Williams, and K. D. M. Harris. *Crystal Growth & Design*, 18(1):42–46, 2018.

[7] Joel Bernstein. volume 14. 2002.

[8] E. H. Lee. *Asian Journal of Pharmaceutical Sciences*, 9(4):163–175, 2014.

[9] K. Kersten, R. Kaur, and A. Matzger. *IUCrJ*, 5(2):124–129, 2018.

[10] G. J. O. Beran, I. J. Sugden, C. Greenwell, D. H. Bowskill, C. C. Pantelides, and C. S. Adjiman. *Chemical Science*, 13(5):1288–1297, 2022.

[11] P. K. Thallapally, R. K. R. Jetti, A. K. Katz, H. L. Carrell, K. Singh, K. Lahiri, S. Kotha, R. Boese, and G. R. Desiraju. *Angewandte Chemie International Edition*, 43(9):1149–1155, 2004.

[12] A. Bhatia, S. Chopra, K. Nagpal, P. K. Deb, M. Tekade, and R. K. Tekade. book section 2, pages 31–65. Academic Press, 2018.

[13] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris. *Pharmaceutical Research*, 18(6):859–866, 2001.

[14] M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber, and O. Grassmann. *Nature Communications*, 6(1):7793, 2015.

[15] N. Kamali, K. Gniado, P. McArdle, and A. Erxleben. *Organic Process Research & Development*, 22(7):796–802, 2018.

[16] M. A. Solomos, C. Capacci-Daniel, J. F. Rubinson, and J. A. Swift. *Crystal Growth & Design*, 18(11):6965–6972, 2018.

[17] O. Al Rahal, P. A. Williams, C. E. Hughes, B. M. Kariuki, and K. D. M. Harris. *Crystal Growth & Design*, 21(4):2498–2507, 2021.

[18] V. Michel, D. Yoshiharu, H. Karl-Heinz, H. Michael, H. Philip, K. Przemyslaw, R. Marguerite, and S. François. *Pure & Applied Chemistry*, 84(2):377–410, 2012.

[19] M. Cusack and A. Freer. *Chemical Reviews*, 108(11):4433–4454, 2008.

[20] B. A. Palmer, G. J. Taylor, V. Brumfeld, D. Gur, M. Shemesh, N. Elad, A. Osherov, D. Oron, S. Weiner, and L. Addadi. *Science*, 358(6367):1172–1175, 2017.

[21] M. S. Reddy. *Frontiers in Microbiology*, 4:1–13, 2013.

[22] A-W. Xu, Y. Ma, and H. Cölfen. *Journal of Materials Chemistry*, 17(5):415–449, 2007.

[23] A. G. Christy. *Crystal Growth & Design*, 17(6):3567–3578, 2017.

[24] R. Liu, S. Huang, X. Zhang, Y. Song, G. He, Z. Wang, and B. Lian. *RSC Advances*, 11(24):14415–14425, 2021.

[25] J. D. Currey. *Journal of Experimental Biology*, 202(23):3285–3294, 1999.

[26] F. J. Ollivier, D. A. Samuelson, D. E. Brooks, P. A. Lewis, M. E. Kallberg, and A. M. Komáromy. *Veterinary Ophthalmology*, 7(1):11–22, 2004.

[27] M. S. Reddy. *Frontiers in Microbiology*, 4:1–13, 2013.

[28] A. Hirsch, B. A. Palmer, A. Ramasubramaniam, P. A. Williams, K. D. M. Harris, B. Pokroy, S. Weiner, L. Addadi, L. Leiserowitz, and L. Kronik. *Chemistry of Materials*, 31(12):4479–4489, 2019.

[29] B. A. Palmer, A. Hirsch, V. Brumfeld, E. D. Aflalo, I. Pinkas, A. Sagi, S. Rosenne, D. Oron, L. Leiserowitz, L. Kronik, S. Weiner, and L. Addadi. *Proceedings of the National Academy of Sciences*, 115(10):2299–2304, 2018.

[30] S. M. Barlow and R. Raval. *Surface Science Reports*, 50(6):201–341, 2003.

[31] A. Aliashkevich, L. Alvarez, and F. Cava. *Frontiers in Microbiology*, 9:683–694, 2018.

[32] I. Nugrahani and M. A. Jessica. *Molecules*, 26(11):3279, 2021.

[33] G. Albrecht and R. B. Corey. *Journal of the American Chemical Society*, 61:1087–1103, 1939.

[34] P. A. Williams, C. E. Hughes, and K. D. M. Harris. *Angewandte Chemie - International Edition*, 54(13):3973–3977, 2015.

[35] E. Courvoisier, P. A. Williams, G. K. Lim, C. E. Hughes, and K. D. M. Harris. *Chemical Communications*, 48(22):2761–2763, 2012.

[36] O. Al Rahal, G. E. Hughes, P. A. Williams, A. J. Logsdail, Y. Diskin-Posner, and K. D. M. Harris. *Angewandte Chemie - International Edition*, 58(52):18788–18792, 2019.

[37] P. A. Williams, C. E. Hughes, J. Martin, E. Courvoisier, A. B. M. Buanz, S. Gaisford, and K. D. M. Harris. *The Journal of Physical Chemistry C*, 120(17):9385–9392, 2016.

[38] C. H. Görbitz, K. W. Tornroos, and G. M. Day. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 68:549–557, 2012.

[39] A. Pirie. *Nature*, 183(4666):985–986, 1959.

[40] G. M. Halford, M. Lordkipanidzé, and S. P. Watson. *Platelets*, 23(6):415–422, 2012.

[41] W. C. Röntgen. *Science*, 3(59):227–231, 1896.

[42] M. von Laue. *Physikalische Zeitschrift*, 14:1040–1041, 1913.

[43] W. H. Bragg and W. L. Bragg. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 88(605):428–438, 1913.

[44] R. E. Watson and A. J. Freeman. *Acta Crystallographica*, 14(1):27–37, 1961.

[45] K. D. M. Harris. *Crystals*, 12(9):1277, 2022.

[46] P. J. Hore. Oxford Chemistry Primers. Oxford University Press, 2nd edition, 2015.

[47] D. C. Apperley, R. K. Harris, and B. Hodgkinson. Momentum Press, LLC, 2012.

[48] P. Batamack, C. Doremieux-Morin, R. Vincent, and J. Fraissard. *The Journal of Physical Chemistry*, 97(38):9779–9783, 1993.

[49] J. E. Pacilio, J. T. Tokarski, R. Quiñones, and R. J. Iuliucci. *Journal of Chemical Education*, 91(8):1236–1239, 2014.

[50] T. Polenova, R. Gupta, and A. Goldbourt. *Analytical Chemistry*, 87(11):5458–5469, 2015.

[51] M. Hanaya and R. K. Harris. *Solid State Nuclear Magnetic Resonance*, 8(3):147–151, 1997.

[52] S. Penzel, A. Oss, M. L. Org, A. Samoson, A. Böckmann, M. Ernst, and B. H. Meier. *Journal of Biomolecular NMR*, 73(1):19–29, 2019.

[53] W. Kolodziejski and J. Klinowski. *Chemical Reviews*, 102(3):613–628, 2002.

[54] K. Maruyoshi, D. Iuga, O. N. Antzutkin, A. Alhalaweh, S. P. Velaga, and S. P. Brown. *Chemical Communications*, 48(88):10844–10846, 2012.

[55] P. Hohenberg and W. Kohn. *Physical Review B*, 136(3):864, 1964.

[56] J. Tirado-Rives and W. L. Jorgensen. *Journal of Chemical Theory & Computation*, 4(2):297–306, 2008.

[57] J. P. Perdew, K. Burke, and M. Ernzerhof. *Physical Review Letters*, 77(18):3865–3868, 1996.

[58] C. Adamo and V. Barone. *Journal of Chemical Physics*, 110(13):6158–6170, 1999.

[59] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, Staroverov V.N., T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, and D.J. Fox. *GAUSSIAN09*, 2016.

[60] J.R. Schmidt and W.F. Polik. *WebMO Enterprise 19.0.009e*, 2020.

[61] J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio, and A. Tkatchenko. *Science Advances*, 5(1):eaau3338, 2019.

[62] R. Glauber and V. Schomaker. *Physical Review*, 89(4):667–671, 1953.

[63] F. Weinert. pages 150–152. Springer Berlin Heidelberg, 2009.

[64] G. P. Thomson and A. Reid. *Nature*, 119:890–890, 1927.

[65] W. H. Massover. *Journal of Synchrotron Radiation*, 14:116–127, 2007.

[66] G. Radi. *Zeitschrift Für Physik*, 212(2):146, 1968.

[67] T. B. Blum, D. Housset, M. T. B. Clabbers, E. van Genderen, M. Bacia-Verloop, U. Zander, A. A. McCarthy, G. Schoehn, W. L. Ling, and J. P. Abrahams. *Acta Crystallographica Section D - Structural Biology*, 77:75–85, 2021.

[68] T. Gruene, J. T. C. Wennmacher, C. Zaubitzer, J. J. Holstein, J. Heidler, A. Fecteau-Lefebvre, S. De Carlo, E. Muller, K. N. Goldie, I. Regeni, T. Li, G. Santiso-Quinones, G. Steinfeld, S. Handschin, E. van Genderen, J. A. van Bokhoven, G. H. Clever, and R. Pantelic. *Angewandte Chemie-International Edition*, 57(50):16313–16317, 2018.

[69] R. Vincent and P. A. Midgley. *Ultramicroscopy*, 53(3):271–282, 1994.

[70] U. Kolb, Y. Krysiak, and S. Plana-Ruiz. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 75(4):463–474, 2019.

[71] G. M. Day and C. H. Görbitz. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 72(4):435–436, 2016.

[72] J. Maddox. *Nature*, 335(6187):201–201, 1988.

[73] S. L. Price and J. G. Brandenburg. book section 11, pages 333–363. Elsevier, 2017.

[74] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 72(2):171–179, 2016.

[75] M. Zilka, D. V. Dudenko, C. E. Hughes, P. A. Williams, S. Sturniolo, W. T. Franks, C. J. Pickard, J. R. Yates, K. D. M. Harris, and S. P. Brown. *Physical Chemistry Chemical Physics*, 19(38):25949–25960, 2017.

[76] C. J. Pickard and R. J. Needs. *Journal of Physics: Condensed Matter*, 23(5):053201, 2011.

[77] A. Oganov (editor). Wiley-VCH, 2011.

[78] A. R. Oganov. *Faraday Discussions*, 211:643–660, 2018.

[79] S. L. Price. *Chemical Society Reviews*, 43(7):2098–2111, 2014.

[80] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris. *Pharmaceutical Research*, 18(6):859–866, 2001.

[81] A. L. Patterson. *Physical Review*, 46(5):372–376, 1934.

[82] Q. Hao, Y. W. Liu, and H. F. Fan. *Acta Crystallographica Section A*, 43:820–824, 1987.

[83] K. D. M. Harris and P. A. Williams. *Structure from Diffraction Methods (Editors: D. W. Bruce, D. O. O'Hare, R. I. Walton)*, pages 1–81, 2014.

[84] A. Coelho and R.W. Cheary. *XFit*, 1998.

[85] A. Coelho. *Topas Academic v5*, 2012.

[86] J. W. Visser. *Journal of Applied Crystallography*, 2:89–95, 1969.

[87] P. E. Werner, L. Eriksson, and M. Westdahl. *Journal of Applied Crystallography*, 18:367–370, 1985.

[88] A. Boultif and D. Louër. *Journal of Applied Crystallography*, 24:987–993, 1991.

[89] R. Shirley. *CRYSFIRE*, 1999.

[90] B. M. Kariuki, S. A. Belmonte, M. I. McMahon, R. L. Johnston, K. D. M. Harris, and R. J. Nelmes. *Journal of Synchrotron Radiation*, 6:87–92, 1999.

[91] S. Habershon, E. Y. Cheung, K. D. M. Harris, and R. L. Johnston. *Journal of Physical Chemistry A*, 108(5):711–716, 2004.

[92] A. Altomare, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, R. Rizzi, and P-E. Werner. *Journal of Applied Crystallography*, 33(4):1180–1186, 2000.

[93] A. Le Bail. *Powder Diffraction*, 19(3):249–254, 2004.

[94] A. Coelho. *Journal of Applied Crystallography*, 36(1):86–95, 2003.

[95] M. A. Neumann. *Journal of Applied Crystallography*, 36(2):356–365, 2003.

[96] A. Le Bail, H. Duroy, and J. L. Fourquet. *Materials Research Bulletin*, 23(3):447–452, 1988.

[97] G. S. Pawley. *Journal of Applied Crystallography*, 14:357–361, 1981.

[98] P. Thompson, D. E. Cox, and J. B. Hastings. *Journal of Applied Crystallography*, 20(2):79–83, 1987.

[99] L. W. Finger, D. E. Cox, and A. P. Jephcoat. *Journal of Applied Crystallography*, 27(6):892–900, 1994.

[100] M. Tremayne, C. Seaton, and C. Glidewell. *Acta Crystallographica. Section B - Structural Science*, 58:823–834, 2002.

[101] B. M. Kariuki, H. Serrano-Gonzalez, R. L. Johnston, and K. D. M. Harris. *Chemical Physics Letters*, 280(3-4):189–195, 1997.

[102] K. Goubitz, E. J. Sonneveld, V. V. Chernyshev, A. V. Yatsenko, S. G. Zhukov, C. A. Reiss, and H. Schenk. *Zeitschrift für Kristallographie*, 214:469–475, 1999.

[103] W. I. F. David, K. Shankland, and N. Shankland. *Chemical Communications*, (8):931–932, 1998.

[104] G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, and B. M. Kariuki. *Chemical Physics Letters*, 321(3-4):183–190, 2000.

[105] H. M. Rietveld. *Journal of Applied Crystallography*, 2:65–71, 1969.

[106] A. March. *Zeitschrift für Kristallographie*, 81:285–297, 1932.

[107] Von Dreele R. B. *Journal of Applied Crystallography*, 30(4):517–525, 1997.

[108] M. Guerain, F. Affouard, C. Henaff, C. Dejoie, F. Danède, J. Siepmann, F. Siepmann, and J.-F. Willart. *Acta Crystallographica Section C - Structural Chemistry*, 77(12):800–806, 2021.

[109] D. Edmondson and S. Ghisla. 131:157–179, 1999.

[110] H. J. Powers. pages 628–632. Academic Press, 2016.

[111] Ute Radespiel and TiHo Hannover, 2021.

[112] A. Pirie. In *Proceedings of the Third Annual Symposium of the British Small Animal Veterinary Association of Aspects of Comparative Ophthalmology*.

[113] H. Sakate. 1955.

[114] A. Gloor. *Process for the purification of riboflavin*, US7670800B2, 2004.

[115] M. Mohnicke. *Berichte aus der Verfahrenstechnik*, 2007.

[116] C. Walsh. *Accounts of Chemical Research*, 13(5):148–155, 1980.

[117] K. Lin, R. Gómez-Bombarelli, E. S. Beh, L. Tong, Q. Chen, A. Valle, A. Aspuru-Guzik, M. J. Aziz, and R. G. Gordon. *Nature Energy*, 1(9):16102, 2016.

[118] G. M. Sheldrick. *Acta Crystallographica A - Foundation & Advances*, 71:3–8, 2015.

[119] G. M. Sheldrick. *SHELXT 2018/2*, 2018.

[120] B. M. Fung, A. K. Khitrin, and K. Ermolaev. *Journal of Magnetic Resonance*, 142(1):97–101, 2000.

[121] R. E. Taylor. *Concepts in Magnetic Resonance Part A*, 22A(2):79–89, 2004.

[122] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M. C. Payne. *Zeitschrift für Kristallographie*, 220(5-6):567–570, 2005.

[123] D. Vanderbilt. *Physical Review B*, 41(11):7892–7895, 1990.

[124] J. P. Perdew, K. Burke, and M. Ernzerhof. *Physical Review Letters*, 77(18):3865–3868, 1996.

[125] A. Tkatchenko and M. Scheffler. *Physical Review Letters*, 102(7):4, 2009.

[126] D. J. Chadi. *Physical Review B*, 16(4):1746–1747, 1977.

[127] C. J. Pickard and F. Mauri. *Physical Review B*, 63(24):13, 2001.

[128] J. R. Yates, C. J. Pickard, and F. Mauri. *Physical Review B*, 76(2):11, 2007.

[129] R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates, and V. Zorin. *Magnetic Resonance in Chemistry*, 45:174–186, 2007.

[130] C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azais, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri, and C. J. Pickard. *Chemical Reviews*, 112(11):5733–5779, 2012.

[131] T. Charpentier. *Solid State Nuclear Magnetic Resonance*, 40(1):1–20, 2011.

[132] A.C. Larson and R. B. Von Dreele. *Los Alamos National Laboratory Report LAUR*, pages 86–748, 2000.

[133] K. D. M. Harris, R. L. Johnston, C. E. Hughes, Z. Zhou, S. Habershon, G. W. Turner, P. A. Williams, B. M. Kariuki, and E. Y. Cheung. *EAGER - A Computer Program for Direct-Space Structure Solution from Diffraction Data*.

[134] K. D. M. Harris, R. L. Johnston, and S. Habershon. *Applications of Evolutionary Computation in Chemistry*, 110:55–94, 2004.

[135] K. D. M. Harris, S. Habershon, E. Y. Cheung, and R. L. Johnston. *Zeitschrift für Kristallographie - Crystalline Materials*, 219(12):838–846, 2004.

[136] M. Mohnicke. *Berichte aus der Verfahrenstechnik*, 2007.

[137] P. Bertani, J. Raya, and B. Bechinger. *Solid State Nuclear Magnetic Resonance*, 61-62:15–18, 2014.

[138] S. E. Ashbrook and D. McKay. *Chemical Communications*, 52(45):7186–7204, 2016.

[139] H. Fuess, D. Hohlwein, and S. A. Mason. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 33(3):654–659, 1977.

[140] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. G. Ren, K. Reuter, and M. Scheffler. *Computer Physics Communications*, 180(11):2175–2196, 2009.

[141] M. H. Hansen, J. A.Garrido Torres, P. C. Jennings, Z. Wang, J. R. Boes, O. G. Mamun, and T. Bligaard. *ArXiv*, abs/1904.00904, 2019.

[142] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko. *Journal of Chemical Physics*, 140(18):14, 2014.

[143] R. Shirley and D. Louër. *Acta Crystallographica*, 1978.

[144] I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. E. Taylor, R. I. Cooper, S. E. Harris, and A. G. Orpen. *Journal of Chemical Information & Computer Sciences*, 44(6):2133–2144, 2004.

[145] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, and R. Taylor. *Journal of the Chemical Society - Perkin Transactions 2*, pages S1–S19, 1987.

[146] M. C. Etter, J. C. Macdonald, and J. Bernstein. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 46:256–262, 1990.

[147] J. A. Garrido Torres, P. C. Jennings, M. H. Hansen, J. R. Boes, and T. Bligaard. *Physical Review Letters*, 122(15):156001, 2019.

[148] B. H. Meier, F. Graf, and R. R. Ernst. *The Journal of Chemical Physics*, 76(2):767–774, 1982.

[149] S. Nagaoka, T. Terao, F. Imashiro, A. Saika, N. Hirota, and S. Hayashi. *The Journal of Chemical Physics*, 79(10):4694–4703, 1983.

[150] A. E. Aliev and K. D. M. Harris. *Supramolecular Assembly Via Hydrogen Bonds I*, 108:1–53, 2004.

[151] V. Kumar, K. A. Woode, R. F. Bryan, and B. A. Averill. *Journal of the American Chemical Society*, 108(3):490–496, 1986.

[152] I. Csoregh, P. Kierkegaard, J. Koziol, and F. Muller. *Acta Chemica Scandinavica*, 41B:383–390, 1987.

[153] E. Sikorska, I. V. Khmelinskii, M. Kubicki, W. Prukała, G. Nowacka, A. Siemiarczuk, J. Koput, L. F. V. Ferreira, and M. Sikorski. *The Journal of Physical Chemistry A*, 109(9):1785–1794, 2005.

[154] M. Sikorski and M. Kubicki. *CSD Communication (Private Communication); REFCODE: AGEPEK*, 2018.

[155] L. Pogliani. *Amino Acids*, 6(2):141–153, 1994.

[156] P.-G. Jonsson and A. Kvick. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 28(6):1827–1833, 1972.

[157] N. A. Tumanov, E. V. Boldyreva, and H. Ahsbahs. *Powder Diffraction*, 23(4):307–316, 2008.

[158] A. Kvick, W. M. Canning, T. F. Koetzle, and G. J. B. Williams. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 36(1):115–120, 1980.

[159] E. V. Boldyreva, S. N. Ivashevskaya, H. Sowa, H. Ahsbahs, and H.-P. Weber. *Zeitschrift für Kristallographie - Crystalline Materials*, 220(1):50–57, 2005.

[160] A. Dawson, D. R. Allan, S.A. Belmonte, S. J. Clark, W. I. F. David, P. A. McGregor, S. Parsons, C. R. Pulham, and L. Sawyer. *Crystal Growth & Design*, 5(4):1415–1427, 2005.

[161] B. P. A. Gabriele, C. J. Williams, M. E. Lauer, B. Derby, and A. J. Cruz-Cabeza. *CrystEngComm*, 23(10):2027–2033, 2021.

[162] Z. Liu, L. Zhong, P. Ying, Z. Feng, and C. Li. *Biophyiscal Chemistry*, 132:18–22, 2008.

[163] Z. Liu and C. Li. *Biophysical Chemistry*, 138(3):115–119, 2008.

[164] A. Mostad, H. M. Nissen, and C. Romming. *Acta Chemica Scandinavica*, 26(10):3819–3833, 1972.

[165] W. Wan, J. Sun, J. Su, S. Hovmoller, and X. Zou. *Journal of Applied Crystallography*, 46(6):1863–1873, 2013.

[166] M. O. Cichocka, J. Angstrom, B. Wang, X. Zou, and S. Smeets. *Journal of Applied Crystallography*, 51(6):1652–1661, 2018.

[167] W. Kabsch. *Acta Crystallographica Section D - Structural Biology*, 66(2):133–144, 2010.

[168] C. J. Pickard and R. J. Needs. *Physical Review Letters*, 97(4):045504, 2006.

[169] Ł Szeleszczuk, D. M. Pisklak, and M. Zielińska-Pisklak. *Journal of Computational Chemistry*, 39(19):1300–1306, 2018.

[170] T. Sun, C. E. Hughes, L. Guo, L. Wei, K. D. M. Harris, Y.-B. Zhang, and Y. Ma. *Angewandte Chemie - International Edition*, 59(50):22638–22644, 2020.

[171] A. Mostad, H. M. Nissen, and C. Romming. *Acta Chemica Scandinavica*, 26(10):3819–3833, 1972.

[172] A. G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkatchenko, A. R. Oganov, A. L. Rohl, M. E. Tuckerman, and B. Kahr. *Chemical Science*, 8(7):4926–4940, 2017.

[173] J. Nyman and G. M. Day. *CrystEngComm*, 17(28):5154–5165, 2015.

[174] C. P. Brock, W. B. Schweizer, and J. D. Dunitz. *Journal of the American Chemical Society*, 113(26):9811–9820, 1991.

[175] F. Kohlbeck and E. M. Horl. *Journal of Applied Crystallography*, 9(1):28–33, 1976.

[176] J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio, and A. Tkatchenko. *Science Advances*, 5(1):eaau3338, 2019.

[177] O. Bakke and A. Mostad. *Acta Chemica Scandinavica*, 34(8):559–570, 1980.

[178] J. D. Dunitz and W. B. Schweizer. *Acta Crystallographica Section C*, 51(7):1377–1379, 1995.

[179] L. L. Mazaleuskaya, K. N. Theken, L. Gong, C. F. Thorn, G. A. FitzGerald, R. B. Altman, and T. E. Klein. *Pharmacogenetics and Genomics*, 25(2):96–106, 2015.

[180] C. Saal and A. Becker. *European Journal of Pharmaceutical Sciences*, 49(4):614–623, 2013.

[181] M. Villanueva, R. Heckenberger, H. Strobach, M. Palmér, and K. Schrör. *Br J Clin Pharmacol*, 35(3):235–242, 1993.

[182] J. F. McConnell. *Crystal Structure Communications*, 3(73), 1974.

[183] P. Derollez, E. Dudognon, F. Affouard, F. Danède, N. T. Correia, and M. Descamps. *Acta Crystallogr B*, 66(Pt 1):76–80, 2010.

[184] J. van de Streek and M. A. Neumann. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 70(6):1020–1032, 2014.

[185] P. A. Williams, C. E. Hughes, and K. D. M. Harris. *Crystal Growth & Design*, 12(12):5839–5845, 2012.

[186] A. L. Spek. *Journal of Applied Crystallography*, 36:7–13, 2003.

[187] K. D. M. Harris and P. A. Williams. *Springer Japan*, pages 141–166, 2015.

[188] S. V. Kolotuchin, E. E. Fenlon, S. R. Wilson, C. J. Loweth, and S. C. Zimmerman. *Angewandte Chemie - International Edition*, 34(23-24):2654–2657, 1996.

[189] R. Nadrowitz, W. Schmidt, and K. Wulff. *Strahlenther Onkol*, 167(6):366–371, 1991.

[190] D. J. Duchamp and R. E. Marsh. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 25(1):5–19, 1969.

[191] C. B. Aakeröy, J. Desper, and J. F. Urbina. *CrystEngComm*, 7(31):193–201, 2005.

[192] F. Zhen-Zhong, L. Xin-Hua, and W. Guo-Ping. *Acta Crystallographica Section E - Crystallographic Communications*, 61(6):o1607–o1608, 2005.

[193] F. H. Herbstein and R. E. Marsh. *Acta Crystallographica Section B - Structural Science Crystal Engineering & Materials*, 33(8):2358–2367, 1977.

[194] F. Zhang, Y. Ma, Y. Chi, H. Yu, Y. Li, T. Jiang, X. Wei, and J. Shi. *Scientific Reports*, 8(1):8208, 2018.

[195] K. Hunger and W. Herbst. Ullmann's Encyclopedia of Industrial Chemistry. Wiley VCH.

[196] R. Pandya, R. Y. S. Chen, Q. F. Gu, J. Sung, C. Schnedermann, O. S. Ojambati, R. Chikkaraddy, J. Gorman, G. Jacucci, O. D. Onelli, T. Willhammar, D. N. Johnstone, S. M. Collins, P. A. Midgley, F. Auras, T. Baikie, R. Jayaprakash, F. Mathevet, R. Soucek, M. Du, A. M. Alvertis, A. Ashoka, S. Vignolini, D. G. Lidzey, J. J. Baumberg, R. H. Friend, T. Barisien, L. Legrand, A. W. Chin, J. Yuen-Zhou, S. K. Saikin, P. Kukura, A. J. Musser, and A. Rao. *Nature Communications*, 12(1):1–11, 2021.

[197] G. M. Sheldrick. *Acta Crystallographica Section A - Foundations & Advances*, 64(1):112–122, 2008.

[198] D. Taupin. *Journal of Applied Crystallography*, 6(5):380–385, 1973.

Appendix A

NMR Parameters for Structure Validation

A.1 Riboflavin

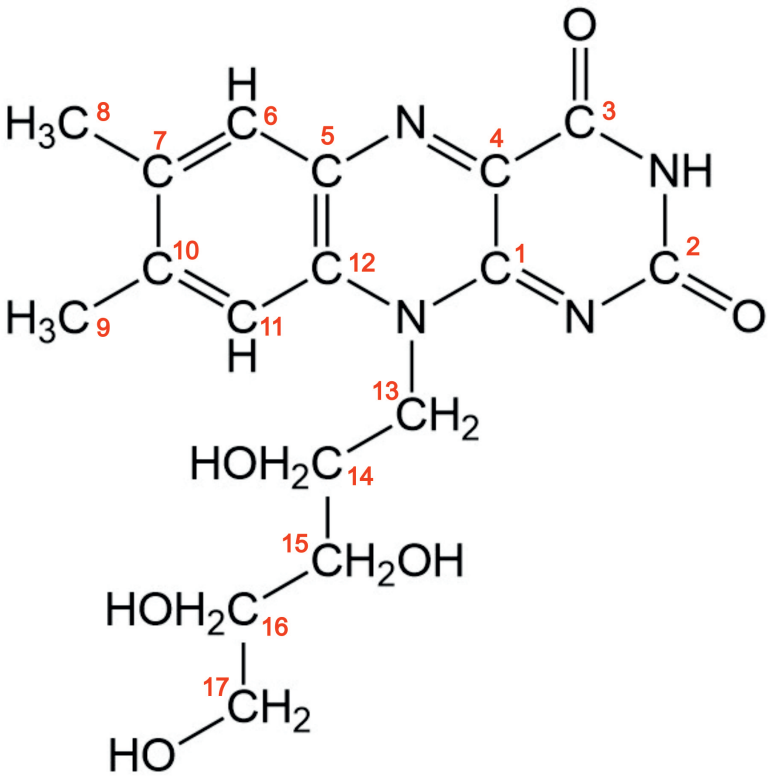

Figure A.1: The numbering of the ^{13}C sites in structure 1, structure 2 and the structure by Guerain *et al.*, of riboflavin.

^{13}C site	δ_{calc} Structure 1 / ppm	δ_{calc} Structure 2 / ppm	δ_{calc} Guerain Structure / ppm
1	146.93	147.56	146.90
2	158.62	160.13	158.71
3	162.17	162.56	162.04
4	136.54	137.66	136.46
5	140.20	140.68	140.23
6	135.71	136.20	135.66
7	147.96	149.38	147.87
8	18.65	19.28	19.08
9	20.37	21.34	20.39
10	159.79	158.12	159.99
11	118.88	118.29	118.87
12	134.76	134.37	134.86
13	42.28	42.81	42.31
14	71.12	70.67	70.95
15	77.43	79.07	77.37
16	79.96	76.24	79.76
17	68.95	65.94	68.81

Table A.1: Isotropic ^{13}3 NMR chemical shifts determined from DFT-D calculations for the crystal structures of the structure 1, structure 2 and the structure by Guerain *et al.*, of riboflavin. These values correspond to the peak positions shown in Figures 4.6 and 5.8.

A.2 Alloxazine

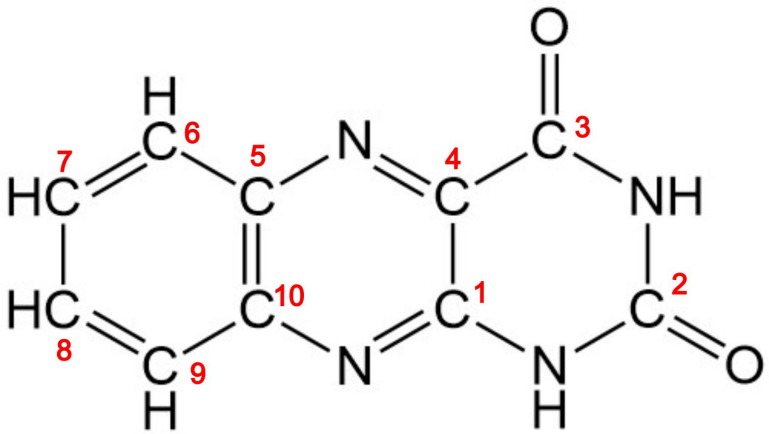

Figure A.2: The numbering of the ^{13}C sites in both alloxazine and isoalloxazine.

^{13}C site	δ_{calc} Alloxazine / ppm	δ_{calc} Isoalloxazine / ppm
1	145.097	138.012
2	149.477	157.002
3	157.027	157.682
4	130.347	147.232
5	141.967	131.172
6	134.137	121.772
7	135.387	142.822
8	140.227	135.062
9	142.007	133.792
10	142.007	139.992

Table A.2: Isotropic ^{13}C NMR chemical shifts determined from DFT-D calculations for the crystal structures containing the alloxazine tautomer (δ_{calc} Alloxazine) and the isoalloxazine tautomer (δ_{calc} Isoalloxazine). These values correspond to the peak positions shown in Figure A.2.

A.3 L-Tyrosine

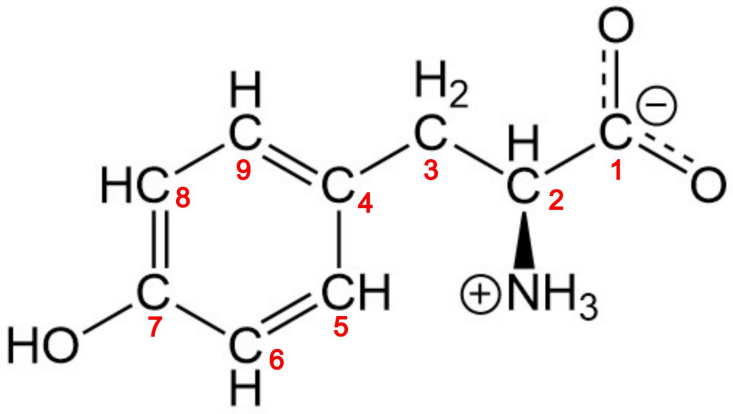

Figure A.3: The numbering of the ^{13}C sites in both the α and β polymorphs of L-tyrosine..

^{13}C site	δ_{calc} α / ppm	δ_{calc} β / ppm
1	179.47	177.93
2	55.99	54.45
3	38.49	36.95
4	126.19	124.65
5	133.59	132.05
6	119.45	117.91
7	159.43	157.89
8	114.95	113.41
9	132.97	131.43

Table A.3: Isotropic ^{13}3 NMR chemical shifts determined from DFT-D calculations for the crystal structures of the α and β polymorphs of L-tyrosine. These values correspond to the peak positions shown in Figures 5.7 and 5.8.

Appendix B
Single Crystal Data Information

B.1 Riboflavin

Parameter	Information
Wavelength / Å	1.54184
Absorption coefficient / mm^{-1}	1.074
F(000)	682
Crystal size / mm^{-1}	0.15 x 0.02 x 0.01
Theta range for data collection / °	3.646 to 74.224
Reflections collected	5848
Independent reflections	2549 [R(int) = 0.0281]
Completeness to theta = 66.97°	99.89%
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	2530 / 0 / 250
Goodness-of-fit on F^2	1.047
Final R indices [I>2sigma(I)]	R1 =0.0345
R indices (all data)	R1 = 0.0435, wR2 = 0.0833
Largest diff. peak and hole / e $Å^{-3}$	0.170 and -0.150

Table B.1: Table showing the data collection information and structure refinement data for riboflavin.

B.2 L-Tyrosine

Parameter	Information
Wavelength / Å	0.0251
Exposure time per frame / s	0.5
Tilt speed / ° s^{-1}	0.2321
Completeness / %	56.8
Resolution / Å	0.90
R_{int}	0.159
Independent reflections	801

Table B.2: Table showing the data collection information and structure refinement data for the β polymorph of L-tyrosine.

B.3 DBTMA - Dihydrate

Parameter	Information
Wavelength / Å	1.54184
Absorption coefficient / mm^{-1}	8.475
F(000)	1568
Crystal size / mm^{-1}	0.224 x 0.044 x 0.030
Theta range for data collection / °	4.525 to 76.845
Reflections collected	11538
Independent reflections	2260 [R(int) = 0.0250]
Completeness to theta = 67.684°	100.0%
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	2260 / 123 / 207
Goodness-of-fit on F^2	1.038
Final R indices [I>2sigma(I)]	R1 = 0.0278, wR2 = 0.0707
R indices (all data)	R1 = 0.0327, wR2 = 0.0
Largest diff. peak and hole / e $Å^{-3}$	0.433 and -0.913

Table B.3: Table showing the data collection information and structure refinement data for DBTMA dihydrate.

Appendix C
Atomic Parameters for Crystal Structures Determined

C.1 Riboflavin

Atom	x	y	z
O001	0.32124(9)	0.49731(11)	0.1775(4)
H001	0.350266	0.504668	0.072277
O002	0.35267(9)	0.32490(11)	0.6396(3)
H002	0.325155	0.317763	0.75058
O003	0.51482(10)	0.63250(13)	-0.1441(4)
O004	0.24073(10)	0.36013(13)	-0.0221(4)
H004	0.226604	0.407482	-0.07326
O005	0.42243(9)	0.47435(14)	0.8356(4)
H005	0.452721	0.509567	0.857937
N006	0.53514(10)	0.44895(13)	0.5275(4)
N007	0.65304(10)	0.53591(13)	0.6547(4)
N008	0.52027(10)	0.54457(13)	0.1958(4)
N009	0.59764(11)	0.66107(14)	0.1275(4)
H009	0.608012	0.706499	0.03985
O00A	0.68323(13)	0.68776(16)	0.3925(6)
C00B	0.55571(12)	0.51975(15)	0.3897(5)
C00C	0.34697(12)	0.41239(15)	0.5418(4)
H00C	0.312991	0.443608	0.638252
C00D	0.57368(12)	0.41773(16)	0.7242(5)
C00E	0.63282(12)	0.46282(15)	0.7831(5)
C00F	0.61628(12)	0.56186(16)	0.4696(5)
C00G	0.54263(12)	0.61252(16)	0.0519(5)
C00H	0.65739(12)	0.35856(16)	1.1154(5)
C00I	0.41297(12)	0.46125(17)	0.5750(5)
H00I	0.410192	0.518694	0.492006

C00J	0.59850(14)	0.31179(17)	1.0501(5)
C00K	0.47050(11)	0.40882(16)	0.4627(5)
H00A	0.469094	0.348792	0.524728
H00B	0.465772	0.406878	0.282541
C00L	0.32460(12)	0.40913(15)	0.2706(5)
H00L	0.357168	0.375847	0.17281
C00M	0.25642(12)	0.36746(17)	0.2374(5)
H00D	0.255999	0.309455	0.313408
H00E	0.223236	0.403337	0.320233
C00N	0.67319(13)	0.43245(17)	0.9803(5)
H00N	0.71162	0.463389	1.019839
C00O	0.55759(13)	0.34194(16)	0.8622(5)
H00O	0.518651	0.311576	0.82621
C00P	0.63672(14)	0.64237(17)	0.3307(5)
C00Q	0.70197(15)	0.32721(19)	1.3215(6)
H00F	0.676702	0.320852	1.472401
H00G	0.736831	0.369346	1.347476
H00H	0.720897	0.271376	1.276515
C00R	0.58049(16)	0.22791(19)	1.1836(6)
H00J	0.615756	0.185915	1.164776
H00K	0.540431	0.204249	1.113463
H00M	0.573802	0.240047	1.357734

Table C.1: Atomic coordinates for the crystal structure of riboflavin.

C.2 Alloxazine

Atom	x	y	z
N1	0.16303(29)	0.1303(5)	0.5537(5)
C2	-0.0222(4)	0.14933(18)	0.63446(35)
C3	-0.0390(8)	0.03011(18)	0.7854(4)
O4	0.1210(10)	-0.08619(15)	0.8341(6)
N5	-0.2341(9)	0.0502(5)	0.87051(24)
H6	-0.2458(12)	-0.0330(7)	0.97586(28)
C7	-0.4143(6)	0.1962(6)	0.79786(7)
O8	-0.5896(7)	0.2142(9)	0.87432(20)
N9	-0.39881(23)	0.30578(30)	0.65909(11)
H10	-0.53540(12)	0.4051(4)	0.61750(28)
C11	-0.21081(15)	0.28643(9)	0.57705(11)
N12	-0.19618(25)	0.39047(26)	0.44537(9)
C13	-0.00111(34)	0.3704(6)	0.36026(23)
C14	0.0146(7)	0.4824(8)	0.21845(19)
H15	-0.1272(9)	0.5855(7)	0.17530(15)
C16	0.2097(8)	0.4624(11)	0.13331(33)
H17	0.2219(11)	0.5496(13)	0.02290(29)
C18	0.3931(5)	0.3291(12)	0.1891(6)
H19	0.5472(6)	0.3132(15)	0.1219(7)
C20	0.37761(22)	0.2186(10)	0.3289(6)
H21	0.51942(16)	0.1155(11)	0.3721(8)
C22	0.17831(15)	0.2391(7)	0.4159(4)

Table C.2: Atomic coordinates for the crystal structure of alloxazine.

C.3 β L-Tyrosine

Atom	x	y	z
C1	0.340(2)	0.867(2)	0.6496(14)
C2	0.4266(14)	0.8360(17)	0.7950(11)
H3	0.539(3)	0.956(3)	0.813(2)
H4	0.483(3)	0.662(4)	0.807(2)
C5	0.3059(12)	0.8764(16)	0.9133(11)
H6	0.387(3)	0.867(3)	1.015(3)
N7	0.1722(11)	0.6942(14)	0.9141(10)
H8	0.098(4)	0.700(3)	1.002(3)
H9	0.080(2)	0.702(3)	0.827(2)
H10	0.230(2)	0.532(4)	0.914(2)
C11	0.2180(19)	1.110(2)	0.8993(19)
O12	0.319(2)	1.277(2)	0.909(2)
O13	0.053(3)	1.114(2)	0.877(2)
C14	0.362360(12)	1.068680(11)	0.575840(4)
H15	0.43611(6)	1.211880(11)	0.628430(5)
C16	0.279(2)	1.101(2)	0.4434(14)
H17	0.29863(7)	1.2631(2)	0.385590(11)
C18	0.183(2)	0.925(2)	0.3774(14)
O19	0.104(2)	0.9644(17)	0.2465(16)
H20	0.062(7)	0.817(6)	0.203(2)
C21	0.157760(11)	0.723320(11)	0.449670(4)
H22	0.08612(7)	0.578820(11)	0.396870(5)
C23	0.242(2)	0.691(2)	0.5817(15)
H24	0.22334(6)	0.53017(18)	0.638990(11)

Table C.3: Atomic coordinates for the crystal structure of L-tyrosine.

C.4 DL-Lysine Anhydrate

Atom	x	y	z
C1	0.96976(23)	0.18696(23)	0.182173(24)
H2	1.01136(28)	-0.02754(28)	0.18133(6)
N3	0.85255(10)	0.1747(4)	0.20672(5)
H4	0.78901(12)	0.0039(4)	0.19855(16)
H5	0.88829(15)	0.1394(13)	0.23577(5)
H6	0.79648(22)	0.3649(4)	0.20506(8)
C7	1.07926(19)	0.38506(19)	0.200963(24)
O8	1.03858(24)	0.59479(18)	0.219216(25)
O9	1.20293(19)	0.32368(35)	0.19678(4)
C10	0.9262(4)	0.2968(6)	0.14143(4)
H11	0.8977(7)	0.5207(5)	0.14370(9)
H12	1.0167(4)	0.2903(13)	0.12473(6)
C13	0.8074(4)	0.1390(9)	0.11996(7)
H14	0.8281(10)	-0.0895(8)	0.11980(21)
H15	0.7151(5)	0.1644(22)	0.13561(10)
C16	0.7768(4)	0.2480(17)	0.07876(8)
H17	0.8603(6)	0.1876(31)	0.06093(10)
H18	0.7747(9)	0.4807(17)	0.07876(22)
C19	0.6404(5)	0.1446(16)	0.05977(12)
H20	0.5601(5)	0.1905(29)	0.07923(12)
H21	0.6427(9)	-0.0879(15)	0.05621(24)
N22	0.60569(16)	0.2970(4)	0.02358(9)
H23	0.64890(10)	0.2030(4)	0.00104(11)
H24	0.50319(13)	0.2976(4)	0.01598(13)

Table C.4: Atomic coordinates for the crystal structure of DL-lysine anhydrate.

C.5 DL-Lysine Hemihydrate

Atom	x	y	z
C1	0.5431(7)	0.8155(12)	0.31601(11)
H2	0.5118(11)	0.0380(10)	0.31723(21)
N3	0.6558(9)	0.8012(27)	0.29159(23)
H4	0.7305(10)	0.9524(27)	0.2997(5)
H5	0.6187(13)	0.853(6)	0.26425(23)
H6	0.6996(11)	0.5976(28)	0.2917(5)
C7	0.4240(7)	0.6351(13)	0.29896(12)
O8	0.3049(7)	0.7106(21)	0.30562(21)
O9	0.4532(12)	0.4264(13)	0.27968(13)
C10	0.5893(8)	0.6992(16)	0.35459(12)
H11	0.6131(15)	0.4725(13)	0.35218(23)
H12	0.5011(9)	0.7128(33)	0.37103(17)
C13	0.7126(8)	0.8479(16)	0.37452(15)
H14	0.6951(15)	0.0777(16)	0.37514(31)
H15	0.8020(9)	0.8168(32)	0.35885(20)
C16	0.7457(8)	0.7343(25)	0.41361(15)
H17	0.6668(8)	0.802(4)	0.43141(19)
H18	0.7435(16)	0.5007(24)	0.41344(26)
C19	0.8849(8)	0.8322(25)	0.43060(19)
H20	0.9646(9)	0.732(4)	0.41545(23)
H21	0.908(9)	0.057(6)	0.432(4)
N22	0.90513(24)	0.7333(21)	0.46970(16)
H23	-0.00551(21)	0.7576(25)	0.48091(21)
H24	0.8647(7)	0.5360(22)	0.47271(13)
O25	0.79977(22)	0.2209(25)	0.50539(5)
H26	0.8254(16)	0.063(5)	0.4886(11)
H27	0.8906(10)	0.257(8)	0.5192(14)
H30	0.1407(7)	0.1706(31)	0.51398(18)

Table C.5: Atomic coordinates for the crystal structure of DL-lysine hemihydrate.

C.6 DL-Lysine Monohydrate

Atom	x	y	z
C1	0.8313(8)	0.0876(6)	0.40875(18)
H2	0.8078(7)	-0.0572(5)	0.41392(25)
N3	0.6607(7)	0.1786(5)	0.44822(19)
H4	0.4926(7)	0.1325(8)	0.4344(4)
H5	0.6966(11)	0.1459(9)	0.49872(19)
H6	0.6680(8)	0.3214(5)	0.44241(34)
C7	1.0743(7)	0.1357(6)	0.43931(18)
O8	1.2319(8)	0.0266(6)	0.42600(20)
O9	1.0999(7)	0.2746(6)	0.47414(18)
C10	0.7995(10)	0.1405(9)	0.33372(17)
H11	0.8419(17)	0.2829(11)	0.32963(21)
H12	0.9272(11)	0.0658(15)	0.30774(23)
C13	0.5610(10)	0.1038(9)	0.29872(17)
H14	0.5196(16)	-0.0388(10)	0.30437(30)
H15	0.4312(11)	0.1769(15)	0.32437(33)
C16	0.5403(12)	0.1537(12)	0.22372(21)
H17	0.6888(12)	0.1040(22)	0.19908(19)
H18	0.5425(20)	0.3003(13)	0.2186(4)
C19	0.3214(12)	0.0825(10)	0.18518(21)
H20	0.1722(12)	0.1264(19)	0.2109(4)
H21	0.3198(20)	-0.0646(10)	0.1871(4)
N22	0.2851(14)	0.1368(10)	0.11391(24)
H23	0.4098(18)	0.0838(16)	0.08710(32)
H24	0.2955(17)	0.2733(10)	0.1099(5)
O25	0.1434(20)	0.9662(6)	0.9281(8)
H26	-0.0256(20)	0.9345(21)	0.9191(18)
H27	0.205(6)	0.8994(21)	0.9686(14)

Table C.6: Atomic coordinates for the crystal structure of DL-lysine monohydrate.

C.7 γ Ibuprofen

Atom	x	y	z
C1	0.5438(19)	0.41822(19)	0.05400(7)
C2	0.548(3)	0.5097(2)	-0.00156(8)
H3	0.402(4)	0.5769(13)	0.0047(6)
C4	0.781(4)	0.583(2)	-0.0072(8)
H5	0.820(7)	0.634(3)	0.0370(9)
H6	0.764(5)	0.654(2)	-0.0452(8)
H7	0.932(3)	0.522(4)	-0.0176(14)
C8	0.476(6)	0.4350(7)	-0.0605(2)
O9	0.263(7)	0.413(3)	-0.0731(10)
O10	0.653(9)	0.3941(10)	-0.0948(5)
H11	0.588(12)	0.3406(4)	-0.1317(3)
C12	0.352(2)	0.4215(5)	0.09472(7)
H13	0.213(3)	0.4921(9)	0.08753(8)
C14	0.3345(10)	0.3342(6)	0.14343(7)
H15	0.1804(11)	0.3369(10)	0.17350(7)
C16	0.5079(3)	0.24088(16)	0.15371(7)
C17	0.4793(10)	0.14805(18)	0.20693(7)
H18	0.2925(16)	0.1173(16)	0.2074(6)
H19	0.586(3)	0.0625(9)	0.1988(5)
C20	0.551(5)	0.2027(8)	0.27255(16)
H21	0.467(10)	0.2968(17)	0.2762(8)
C22	0.820(6)	0.219(6)	0.2787(14)
H23	0.910(6)	0.127(8)	0.276(2)
H24	0.889(9)	0.279(8)	0.2409(16)
H25	0.872(9)	0.263(6)	0.3240(15)
C26	0.458(10)	0.120(3)	0.32580(11)
H27	0.533(12)	0.150(3)	0.3718(2)
H28	0.504(15)	0.0203(19)	0.3191(7)
H29	0.264(10)	0.125(7)	0.3284(11)
C30	0.7027(3)	0.2397(6)	0.11384(7)
H31	0.8404(8)	0.1679(10)	0.12028(7)
C32	0.7201(10)	0.3266(7)	0.06452(7)
H33	0.8717(9)	0.3215(11)	0.03354(8)
C34	0.9319(3)	0.72478(17)	1.31036(7)

C35	0.8090(4)	0.6494(2)	1.25749(8)
H36	0.6434(18)	0.6107(16)	1.2762(3)
C37	0.9555(19)	0.5400(15)	1.2314(9)
H38	1.1234(5)	0.574(3)	1.2123(12)
H39	0.8543(18)	0.4912(15)	1.1936(8)
H40	0.995(5)	0.4705(12)	1.2686(12)
C41	0.727(4)	0.7347(4)	1.2031(6)
O42	0.514(5)	0.7459(10)	1.1877(12)
O43	0.901(7)	0.792(2)	1.1739(3)
H44	0.835(10)	0.842(2)	1.1361(7)
C45	1.1460(3)	0.68794(17)	1.33998(8)
H46	1.2391(4)	0.60397(18)	1.32403(8)
C47	1.2450(3)	0.75701(17)	1.39036(7)
H48	1.4125(3)	0.72452(19)	1.41165(8)
C49	1.1360(3)	0.86657(16)	1.41294(7)
C50	1.2364(3)	0.94674(16)	1.46648(7)
H50	1.548(17)	0.759(5)	1.578(3)
H51	1.403(3)	0.992(2)	1.4515(5)
H52	1.112(3)	1.0255(18)	1.4731(10)
C53	1.286(7)	0.8822(9)	1.5313(6)
H54	1.136(12)	0.818(5)	1.5407(11)
C55	1.298(14)	0.982(2)	1.5838(4)
H56	1.434(17)	1.054(3)	1.5744(15)
H57	1.128(17)	1.032(6)	1.5881(11)
H58	1.343(17)	0.939(3)	1.6299(7)
C59	1.516(12)	0.804(6)	1.532(2)
H61	1.669(7)	0.866(10)	1.523(3)
H62	1.516(19)	0.730(7)	1.496(3)
C63	0.9227(3)	0.90342(17)	1.38252(7)
H64	0.8325(3)	0.98834(19)	1.39858(7)
C65	0.8227(3)	0.83422(18)	1.33275(7)
H66	0.6545(3)	0.8650(2)	1.31110(7)

Table C.7: Atomic coordinates for the crystal structure of γ ibuprofen.

C.8 2,4-Dibromo Trimesic Acid (DBTMA) - Dihydrate

Atom	x	y	z
C1	0.32510(15)	0.7414(3)	0.64444(13)
C2	0.25641(16)	0.7737(3)	0.60627(12)
C3	0.17989(16)	0.7479(4)	0.63362(13)
C4	0.17404(16)	0.6882(4)	0.70061(14)
C5	0.24175(17)	0.6508(3)	0.74016(13)
C6	0.31768(16)	0.6792(4)	0.71121(13)
C7	0.40730(15)	0.7670(4)	0.61310(13)
C8	0.10339(17)	0.7761(4)	0.59448(14)
C9	0.22816(18)	0.5821(4)	0.81090(14)
Br1	0.27049(2)	0.86260(4)	0.51676(2)
Br2	0.41414(2)	0.63389(6)	0.75875(2)
O1	0.43580(14)	0.6637(3)	0.57426(12)
O2	0.44136(14)	0.9095(3)	0.63236(13)
H2	0.487325	0.913444	0.617243
O3	0.09759(13)	0.7668(4)	0.53321(11)
O4	0.04268(12)	0.8099(3)	0.63536(11)
H4	0.00384	0.840518	0.612579
O5	0.27686(17)	0.5410(7)	0.85024(16)
O6	0.15267(17)	0.5704(7)	0.82500(16)
H6	0.146811	0.515542	0.860436

Table C.8: Atomic coordinates for the crystal structure of DBTMA dihydrate.

C.9 2,4-Dibromo Trimesic Acid (DBTMA) - Anhydrate

Atom	x	y	z
C1	0	0.04352(4)	0.40095
C2	0	-0.13506(6)	0.55510(6)
O3	0	-0.1524(16)	0.809(4)
O4	0	-0.2785(14)	0.378(4)
H5	0	-0.237(3)	0.147(6)
C6	-0.0727(3)	0.12242(3)	0.30915(2)
Br7	-0.16649(14)	-0.0018(6)	0.4430(17)
C8	-0.0744(2)	0.29076(3)	0.12861(3)
C9	-0.1494700(2)	0.3759800(11)	0.0412400(10)
O10	-0.2187200(2)	0.369210(3)	0.1451400(18)
O11	-0.1425900(3)	0.4837400(7)	-0.1996800(10)
H12	-0.1940500(3)	0.547900(2)	-0.266620(9)
C13	0	0.36957(4)	0.03698(4)
H14	0	0.49803(5)	-0.08762(6)

Table C.9: Atomic coordinates for the crystal structure of DBTMA anhydrate.

C.10 Cyclohexyl-Functionalized Perylene Diimide (PDICH)

Atom	x	y	z
N1	0.6832	0.71385	-0.11384
C2	0.70356	0.77008	-0.15141
H3	0.75494	0.76045	-0.15007
C4	0.68074	0.62781	-0.1803
H5	0.62938	0.63658	-0.18312
H6	0.69228	0.48406	-0.17139
C7	0.71244	0.67048	-0.21705
H8	0.69582	0.56949	-0.23762
H9	0.76372	0.64905	-0.21457
C10	0.69956	0.87439	-0.22971
H11	0.64925	0.88797	-0.23682
H12	0.72627	0.90421	-0.25473
C13	0.71598	1.0219	-0.20012
H14	0.70022	1.16262	-0.20873
H15	0.76765	1.03004	-0.19727
C16	0.68643	0.97412	-0.16264
H17	0.63506	0.9893	-0.16407
H18	0.70365	1.07268	-0.14178
C19	0.61969	0.6683	-0.10747
O20	0.57999	0.67326	-0.13182
C21	0.72977	0.7091	-0.08632
O22	0.78446	0.75009	-0.09343
C23	0.6016	0.61296	-0.06982
C24	0.71021	0.65301	-0.04879
C25	0.64655	0.60593	-0.04124
C26	0.53916	0.56726	-0.06281
H27	0.50541	0.57348	-0.08506
C28	0.75436	0.64597	-0.02066
H29	0.80321	0.68244	-0.02672
C30	0.62764	0.55149	-0.00526
C31	0.56278	0.50538	0.0012
C32	0.52023	0.51401	-0.0277
H33	0.4709	0.47951	-0.02346
C34	0.67399	0.54566	0.02315

C35	0.73633	0.59289	0.01472
H36	0.77241	0.58969	0.03566
C37	0.58873	0.44583	0.06631
C38	0.54247	0.45171	0.03787
C39	0.4799	0.40562	0.04589
H40	0.44423	0.4089	0.02472
C41	0.65374	0.49153	0.05998
C42	0.6958	0.48276	0.08922
H43	0.74528	0.51635	0.08524
C44	0.56929	0.39149	0.10218
C45	0.50535	0.3458	0.10915
C46	0.46132	0.35294	0.08106
H47	0.41256	0.31711	0.08697
C48	0.61362	0.38339	0.13108
C49	0.6763	0.42869	0.12427
H50	0.7098	0.42197	0.14665
C51	0.48497	0.29069	0.14632
O52	0.42996	0.2517	0.15256
C53	0.59426	0.32794	0.16865
O54	0.63262	0.321	0.19366
N55	0.53044	0.28442	0.17422
C56	0.50705	0.22779	0.21113
H57	0.45598	0.24407	0.20877
C58	0.52975	0.36211	0.24141
H59	0.5187	0.50818	0.23382
H60	0.58103	0.3514	0.24444
C61	0.49727	0.31057	0.2775
H62	0.44603	0.3321	0.2749
H63	0.51341	0.40593	0.29916
C64	0.51046	0.10333	0.28793
H65	0.48553	0.06775	0.31328
H66	0.56119	0.08687	0.29363
C67	0.49142	-0.03581	0.25743
H68	0.43963	-0.03793	0.25497
H69	0.50574	-0.17987	0.26469
C70	0.52072	0.01819	0.2203
H71	0.50133	-0.07189	0.19865
H72	0.57179	-0.00527	0.22108

N73	0.43052	0.76999	0.13008
C74	0.3854	0.70955	0.15918
H75	0.41567	0.69887	0.18319
C76	0.33311	0.85048	0.16838
H77	0.29602	0.84632	0.14762
H78	0.35239	0.99401	0.16901
C79	0.30626	0.79516	0.2059
H80	0.26888	0.89506	0.21355
H81	0.34354	0.80946	0.22669
C82	0.2805	0.58983	0.20669
H83	0.23494	0.58636	0.19207
H84	0.27036	0.54869	0.23506
C85	0.32532	0.44258	0.18912
H86	0.29972	0.31046	0.18393
H87	0.36314	0.40799	0.20866
C88	0.3551	0.51303	0.1531
H89	0.31885	0.52191	0.13187
H90	0.39132	0.41356	0.14364
C91	0.41311	0.81393	0.09393
O92	0.35816	0.80573	0.08354
C93	0.4929	0.77774	0.14304
O94	0.5072	0.73941	0.17441
C95	0.46472	0.87158	0.06848
C96	0.54245	0.83487	0.1173
C97	0.52817	0.88116	0.08065
C98	0.45135	0.91735	0.03246
H99	0.40274	0.90986	0.0232
C100	0.60365	0.8425	0.13018
H101	0.61281	0.80586	0.15845
C102	0.57789	0.93724	0.05659
C103	0.56299	0.98396	0.01956
C104	0.49974	0.97255	0.00854
H105	0.48671	1.00691	-0.01929
C106	0.64117	0.94433	0.07053
C107	0.65244	0.89698	0.10691
H108	0.69991	0.9016	0.11788
C109	0.67845	1.04752	0.0093
C110	0.61514	1.04133	-0.00485

C111	0.60593	1.08963	-0.04156
H112	0.55895	1.08655	-0.05324
C113	0.69288	1.00041	0.04637
C114	0.75526	1.00966	0.05859
H115	0.76679	0.97442	0.0866
C116	0.72932	1.10214	-0.01413
C117	0.7175	1.15058	-0.05109
C118	0.65606	1.14438	-0.06436
H119	0.64769	1.1817	-0.09267
C120	0.79202	1.10801	-0.00052
C121	0.8045	1.06146	0.03546
H122	0.85285	1.06733	0.04521
C123	0.76972	1.2072	-0.07593
O124	0.75857	1.24556	-0.10787
C125	0.84528	1.16329	-0.02455
O126	0.89952	1.1669	-0.0127
N127	0.83146	1.20988	-0.06095
C128	0.88628	1.26625	-0.08418
H129	0.86449	1.30429	-0.11015
C130	0.93482	1.10972	-0.09159
H131	0.91094	0.98677	-0.10346
H132	0.95653	1.06466	-0.06594
C133	0.98583	1.18401	-0.11782
H134	0.96401	1.21851	-0.14424
H135	1.02042	1.0721	-0.12319
C136	1.01906	1.36076	-0.10274
H137	1.05388	1.41297	-0.12248
H138	1.04544	1.32265	-0.07794
C139	0.97127	1.517	-0.0935
H140	0.95055	1.5714	-0.11895
H141	0.99535	1.63655	-0.0805
C142	0.91804	1.44533	-0.06831
H143	0.88219	1.55577	-0.06479
H144	0.93744	1.41358	-0.0413

Table C.10: Atomic coordinates for the crystal structure of PDICH.

Appendix D
Work Published

C. J. H. Smalley, A. J. Logsdail, C. E. Hughes, D. Iuga, M. T. Young, K. D. M. Harris; Solid-state structural properties of alloxazine determined from powder XRD data in conjunction with DFT-D calculations and solid-state NMR spectroscopy: unraveling the tautomeric identity and pathways for tautomeric interconversion; *Cryst. Growth Des.*, 2022, **22**, 524-534.

N. Pinsk, A. Wagner, L. Cohen, C. J. H. Smalley, C. E. Hughes, G. Zhang, M. J. Pavan, N. Casati, A. Jantschke, G. Gooves, K. D. M. Harris, B. A. Palmer; Biogenic guanine crystals are solid solutions of guanine and other purine metabolites; *J. Am. Chem. Soc.*, 2022, **144**, 5180-5189.

C. J. H. Smalley, H. E. Hoskyns, C. E. Hughes, D. N. Johnstone, T. Wilhammar, M. T. Young, C. J. Pickard, A. J. Logsdail, P. A. Midgley, K. D. M. Harris; A structure determination protocol based on combined analysis of 3D-ED data, powder XRD data, solid-state NMR data and DFT-D calculations reveals the structure of a new polymorph of L-tyrosine; *Chem. Sci.*, 2022 **13**, 5277-5288.