SDPD-SX: a methodology for powder diffraction data collection for structure solution

Marta Joan Ross

Submitted for the Degree of
Doctor of Philosophy

University of Reading

School of Pharmacy
PO Box 226
Whiteknights
Reading
Berkshire
RG6 6AP

May 2023

Declaration

I confirm that this is my own work and that all other material from various sources are properly and fully acknowledged.

Marta Joan Ross

Acknowledgements

The warmest of thanks to my PhD supervisors from the University of Reading; Dr. Elena Kabova and Prof. Kenneth Shankland as well as the support from my industrial supervisor, Dr. Charles Blundell from C4X Discovery, Manchester.

I am grateful for the financial support from UKRI and C4X Discovery that ensured the creation of this thesis.

For technical support given throughout the PhD I would like to acknowledge the help of Nick Spencer from the University of Reading's Chemical Analysis Facility. In addition, I would like to thank Dr. George Whitehead and Dr. Inigo Vitorica from the University of Manchester for support in data collection and analysis.

I would like to acknowledge the many hours of support provided by Laura Bassenger to help to create this thesis. This thesis would not have been completed without this consistent support.

For support during long data collection days, I would like to thank Connie and Sam for your thoughts and warm welcome each time I arrived and stayed with you.

I would like to thank my Auntie, Sally Gibson, for the many phone calls and words of encouragement.

Finally, I thank my partner, Jon, for support in teaching me the Python needed for this thesis. Additionally, thanks are due the unwavering support and encouragement both during experimental and writing of this work, particularly during times of ill health.

Abstract

Determination of structural information from diffraction patterns of crystalline materials can enable and enhance many ‘high impact’ areas of chemistry including the development of materials for photovoltaics, hydrogen storage cells and battery materials. In this thesis, drug discovery and drug development is explored with respect to their use of structural information. Given that these processes utilise conformational and packing information to enhance all stages of their development; from lead identification to formulation. The earlier conformational information is obtained within these processes, the greater potential impact on the direction of a project within these areas. Furthermore, in many cases, it is not feasible to grow single crystals, large enough what they are suitable for single crystal X-ray diffraction (SCXRD). Powder X-ray diffraction is the best alternative but unfortunately, a standard powder diffraction set up requires at least 10 mg of sample, which is simply not available at the early stages of drug development programmes. In this thesis, a trial of 81 small organic compounds found that, after multiple recrystallisations, 25 compounds produced one or more samples that were crystalline but were unsuitable for SCXRD.

This thesis presents a method for the routine powder diffraction data collection using an in-house single crystal diffractometer with < 0.1 mg of polycrystalline sample. The process of optimising of a single crystal diffractometer for optimal powder data collection is described. This method combines the use of powder diffraction data collected on a standard, lab-based single crystal diffractometer (SDPD-SX) using DASH, an open-source global optimisation structure solution approach. The applicability of the method is also demonstrated through a wide range of molecular and crystallographic complexity, with particular focus on compounds of pharmaceutical interest. SDPD-SX enabled the crystal structure determination from limited samples of a range of known and novel crystal structures. Novel structures were validated using DFT-D energy minimisation.

Table of contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Equations - Recreate

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Frequently used Abbreviations

APPE	Automatic Powder Pattern Evaluation
API	Active Pharmaceutical Ingredient
CSD	Cambridge Structural Database
DFT-D	Dispersion-corrected Density Functional Theory
DoF	Degrees of Freedom
ED	Electron Diffraction
ESD	Estimated Standard Deviation
FP	Flat Plate
FWHM	Full Width at Half Maximum
GO	Global Optimisation
PO	Preferred Orientation
PXRD	Powder X-ray Diffraction
RFP	Reflection Flat Plate
RMSD	Root Mean Squared Deviation
SA	Simulated Annealing
SCXRD	Single Crystal X-ray Diffraction
SDPD	Structure Determination from Powder Diffraction
SDPD-SX	Structure Determination from Powder Diffraction using a Single crystal X-ray diffractometer
TFP	Transmission Flat Plate
UoM	University of Manchester
UoR	University of Reading

Chapter 1

Introduction

This thesis explores a novel method for determining molecular crystal structures from powder diffraction data collection, using data collected from a dedicated single-crystal instrument. The following introduction gives a summary of relevant topics and current research developments. The chapter includes a discussion of the basics of crystallography, how the crystalline state is engineered to produce desirable properties with particular importance for how this information can be used within the pharmaceutical industry. Following this is a discussion on the basics of diffraction, the instrumentation used for data collection and the most-used methods for molecular crystal structure determination.

1.1 Crystals, polycrystalline materials, and their importance

Crystals, as a form of solid material, are found everywhere in nature and the modern world. Knowledge of this organised state has enabled the understanding of intermolecular interactions, material properties, solid-state behaviour, and much more.

1.1.1 Crystals

A crystal is a solid that possesses long-range order, meaning that it has a repeating pattern of atoms or molecules lying within a representative 'unit cell', this unit cell is repeated by translation in three-dimensions to produce the macroscopic crystal (Figure 1.1). The asymmetric unit is the smallest fraction of the unit cell that can be rotated and translated by symmetry operations to create the whole crystal. Although each crystal structure is unique with a unique set of lattice parameters, there are common properties that researchers exploit for structural analysis.[1]

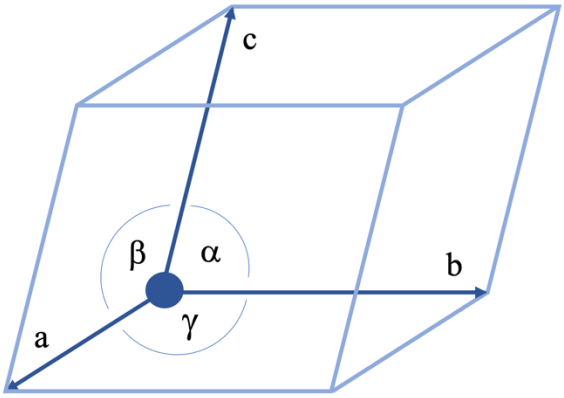


Figure 1.1. The unit cell is the simplest repeating unit that describes the entire lattice by translation, described by lengths a, b and c and angles α, β and γ.

The focus of this study is molecular crystals, crystallography is the study of the properties and structures of crystal structures and their resulting properties. Crystals form because the energy

gained from the formation of inter and intra molecular interactions outweighs the loss of entropy from the formation of a crystal from its components, giving an overall drop in Gibbs free energy. Crystals form in 1 of the 7 crystal systems (Bravais lattices), 1 of 32 point groups, and 1 of 230 space groups. The number of molecules or 'formula units' in the unit cell is referred to as Z, with the number of molecular units in the asymmetric unit given as Z′. For example, a crystal structure with 2 molecules in space group $P\bar{1}$ will have Z = 2 and Z′ = 1, as illustrated below in Figure 1.2.

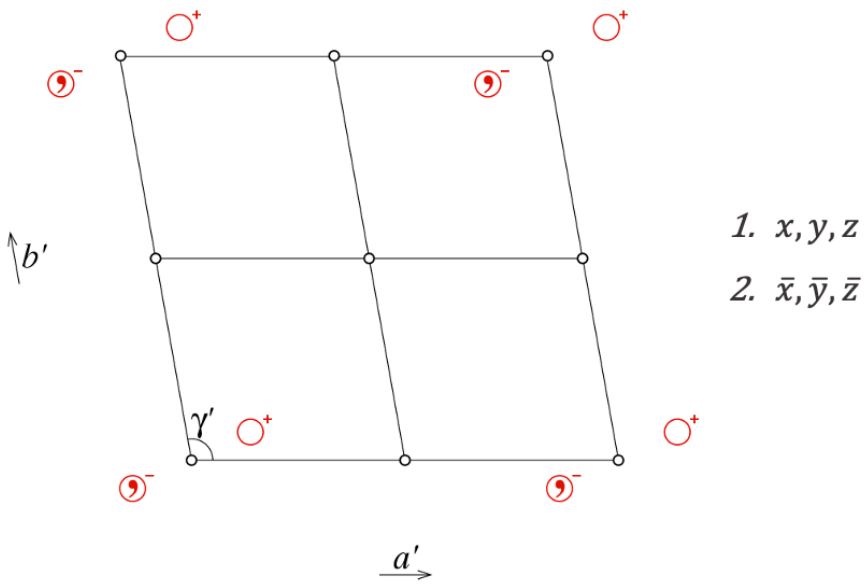


Figure 1.2. Space group diagram for $P\bar{1}$ (space group no 2), shown in projection down the c axis (symbol ' showing plains do not lie in plane of diagram). $P\bar{1}$ has only two symmetry operators: x, y, z and -x, -y, -z with a 2:1 ratio of Z and Z′.[2]

1.1.2 Cambridge structural database

Established in 1965, the Cambridge Structural Database (CSD) is "the comprehensive repository of validated and curated small-molecule organic and metal-organic structures". For the molecular materials that are the subject of this thesis, the main crystal systems and space groups of relevance are shown in Table 1.1. As of 2017, These 10 space groups account for 87.8% of crystal structures deposited in the CSD.[3]

Table 1.1. Most populated space groups in the Cambridge Structural Database (CSD), reported in 2017.

Space group	% of CSD
$P2_1/c$	34.5
$P\bar{1}$	24.7
$C2/c$	8.4
$P2_12_12_1$	7.2
$P2_1$	5.2
$Pbca$	3.3
$Pna2_1$	1.4
$Pnma$	1.1
Cc	1.0
$P1$	1.0

1.1.2 Analysis of polycrystalline materials

For many researchers, crystallisation is a process used to purify products; for others, the crystal structures themselves (the stereochemistry, composition, conformation, *etc*.) are the research focus. Both have in common the need to grow good-quality crystals of the material of interest; in general, a good-quality single crystal can be thought of as a crystal that is large enough to mount upon a single-crystal X-ray diffractometer, and which diffracts strongly enough to permit the collection of diffraction data. A calculated single-crystal diffraction pattern for ibuprofen (CSD refcode IBPRAC01) is shown in Figure 1.3 demonstrating how symmetry related reflections are separated in space making indexing and intensity integration straightforward. Typically, the chosen sample has dimensions no less than 50 μm, though much smaller crystals can be dealt with routinely, particularly at synchrotron sources.

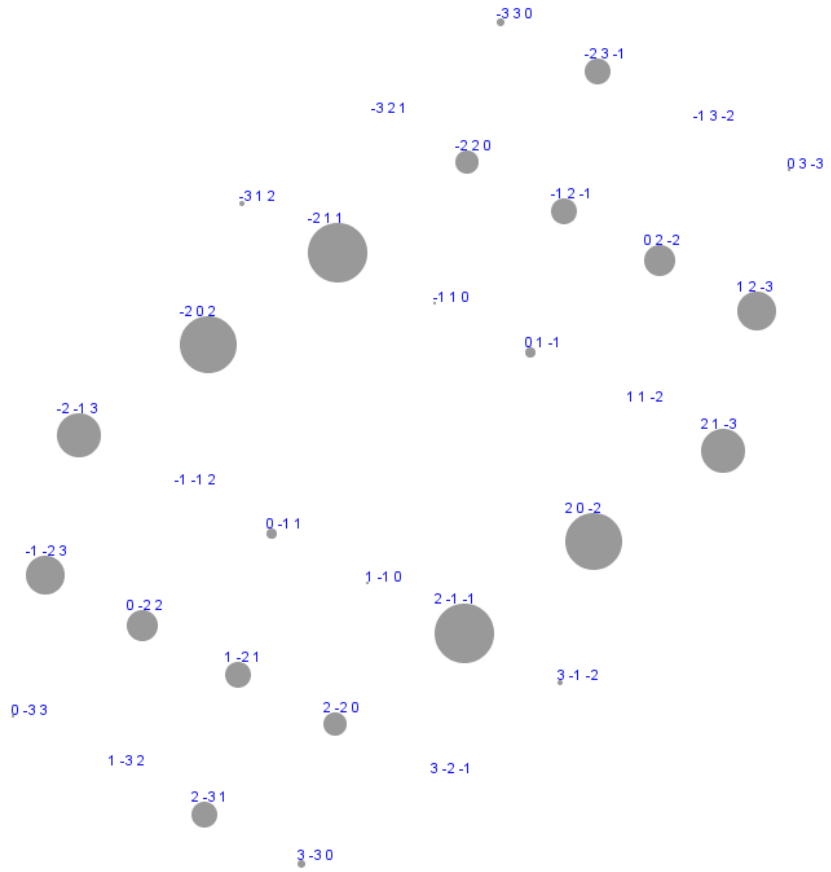


Figure 1.3. Simulated single-crystal diffraction pattern (111 projection, layer index 0) of ibuprofen (CSD refcode IBPRAC01), highlighting the separation of symmetry related space. Simulation created in ReciprOgraph, showing HKL <= |3|

It is often the case that crystals cannot easily be grown as suitable single crystals (often with barriers such as size, poor morphology, twinning, and multiple forms) and instead form a polycrystalline material, termed a 'powder'. Knowledge of the crystal structure is often still required but as long as the material in question is crystalline, a diffraction pattern can be collected. Powder X-ray diffraction (PXRD) is the result of diffraction from a combination of many very small single crystals to form a 1- or 2-dimensional diffraction pattern that can be used to deduce information the about crystal structure and the bulk properties. Unlike SCXRD, powder diffraction structure determination greatly benefits form knowledge of the unit cell contents *i.e.,* the molecular conformation, potential solvent molecules, stereochemistry *etc*. This is crucial for solving organic compounds, and so naturally affects the structure determination of pharmaceuticals. A further description of this is given is section 1.4.2. The calculated 1D powder pattern of ibuprofen (CSD refcode IBPRAC01) is shown below in Figure 1.4.[4]

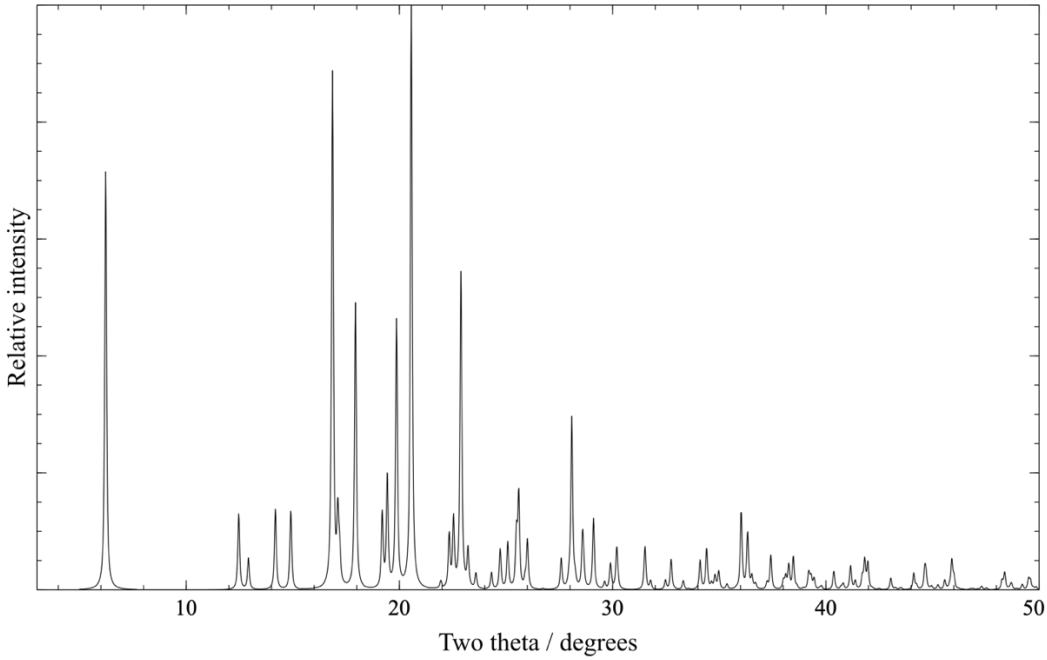


Figure 1.4. Simulated 1 dimensional powder diffraction pattern (Cu radiation, λ = 1.54056) of ibuprofen (CSD refcode IBPRAC01).

A few examples of polycrystalline materials and their importance are listed below.

1.1.2.1 Metal organic frameworks (MOFs)

Metal organic frameworks (MOFs) are highly tuneable crystalline materials known for their large surface areas, typically 3000-6000 m^2g^{-1}, and high gas storage capacities as a consequence of their porous structures. Formed by the linkage of metals as the nodes and organic ligands as the struts, MOFs are simple to functionalise, and their pore sizes can be adjusted to the desired guest-host interaction. One example of MOF application is shown in Figure 1.5 (MOF-74-Zn, created with 2,5-dihydroxyterephthalate and Zn^{2+}) and other similar structures have been shown to be effective for controlled gas storage and release. One application for this property is absorbing and releasing 1-methylcyclopropene to inhibit ethylene-triggered events in plants, preventing degreening and fruit ripening, for enhanced fruit preservation.[5, 6]

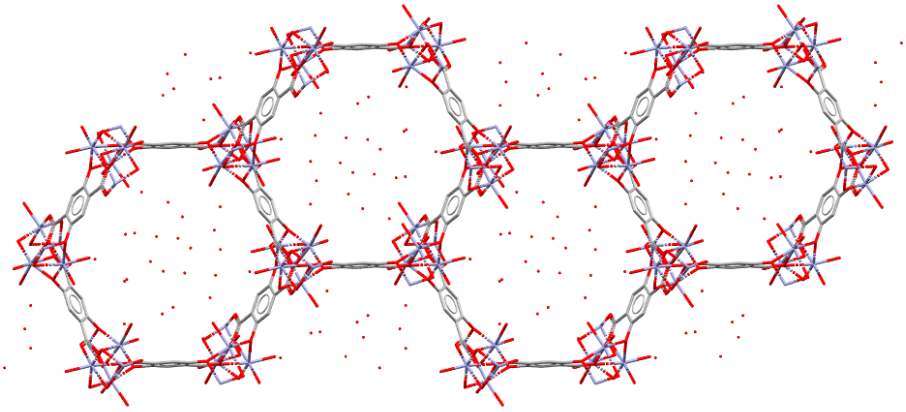


Figure 1.5. MOF-74-Zn (CSD refcode ORIVOC) with H_2O molecules in the channels, used for gas absorption and release applications for fruit preservation.[5, 6]

MOFs are, however, difficult to crystallise in forms other than polycrystalline solids, and so their crystallographic characterisation is primarily achieved through the use of PXRD.

1.1.2.2 Dyestuffs and pigments

PXRD plays a significant role in structural characterisation within the art world, and as such has become an invaluable tool for historians and art conservators. The analysis of paints is used to determine counterfeits, and the study of corrosion materials that form on the surface of works of art aids in the preservation of irreplaceable objects.[7]

Many dyestuffs and pigments have very limited solubility in a wide range of solvents and so are difficult to crystallise as anything other than as powders. The crystal structures of these materials largely determine their physical properties, which in order to engineer requires an understanding of the inter and intramolecular forces. A prime example of this is the investigation of pigment yellow (PY 138), a material whose properties depend on the tautomeric state (Figure 1.6). All attempts to grow single crystals have been unsuccessful, except for a toluene solvate. However, the combination of PXRD and NMR techniques was able to deduce that the dominant tautomer present is the NH form. [8]

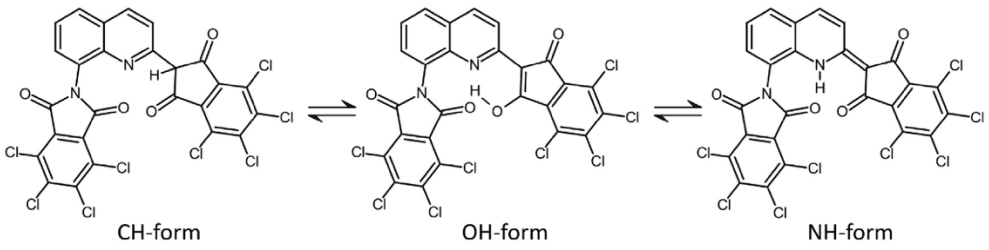


Figure 1.6. Pigment yellow 138 tautomeric forms, from which the correct form was deduced using powder diffraction, NMR, and DFT-D calculations.[8]

1.1.2.3 Food production

The control of crystal form and crystallite size is used throughout food research. For example, the form and size of fat crystals in chocolate affect properties such as texture, melting point, speed of melting, brittleness, and gloss, all of which affect the look and taste. These are all carefully controlled to create an ideal product for consumption. Products can be analysed by PXRD to ensure consistency as well as to shed light on properties such as the relative amounts of crystalline components in foodstuffs.[9, 10, 11]

1.1.2.4 *In-situ* studies and phase transformations

Often when a single crystal undergoes a phase transformation the crystal shatters, precluding investigation of the transformed structure. Crystal structures such as polymorphs of pyrene have been able to be studied with single crystal X-ray diffraction (SCXRD). After a phase transition upon lowering the temperature to less than 93 K, a crystal of the room temperature form of pyrene shattered; fortunately, a single-crystal piece of the low temperature form remained attached to the goniometer head fibre and was large enough to continue data collection. This is often not the case and so alternative structural characterisation methods are needed.[12]

Powder diffraction can be used to characterise the materials on either side of the transformation and so it is a powerful tool for studying processes *in-situ*, such as changes as a function of temperature or pressure over time. The speed of 1-dimentional data collection allows for many scans, precisely showing changes of a system. By way of example, zopiclone dihydrate, a cyclopyrrolone hypnotic drug, undergoes a transformation to a monoclinic anhydrous form at 75 °C and then an irreversible transformation to an orthorhombic anhydrous form at 112 °C. Crystal structures of these forms were obtained using PXRD.[13]

In addition to phase changes, materials that change as a function of mechanical grinding, or those that react with other materials when combined, can be investigated *in-situ* (particularly at synchrotron sources). Using specially designed ball mills, the material can be released at a steady rate into the path of the X-ray beam.[14]

1.1.2.5 Pharmaceuticals

The majority of small molecule active pharmaceutical ingredients (APIs) are processed in powder form, ensuring reliability and reproducibility, with well-defined physical characteristics. This, therefore, means that PXRD plays a considerable role in active API development,

characterisation, and analysis. Such materials are the focus of this thesis and are discussed in further detail below.

1.2 Crystals and pharmaceuticals

85% of pharmaceuticals in the UK and US are delivered orally; a large portion of these medications will be administered in a polycrystalline form, held together by excipients into a solid form (most commonly a tablet). Crystal forms and subsequent analysis, therefore, play an important role throughout the pharmaceutical development process. Furthermore, crystal structures can act as the basis for structure-based drug design, as well as the basis for patent applications (both chemical- and physical-form based).[15]

1.2.1 Drug discovery process

The development of novel pharmaceutical candidates is lengthy and expensive. A new pharmaceutical typically arrives to market 10-12 years after programme inception, and after examining around 10,000 candidate molecules. A high number of candidate molecules are necessary to find one that has all the desired activity and physical properties.

Each large company describes the steps of development within drug discovery differently, and there are overlapping terms that vary slightly in definition. Crucially, however, compounds have to meet strict criteria before moving from one stage to more complex assays, *i.e.*, safety and efficacy. The series of compounds at each stage are subject to standards for stability and safety that are more stringent as the programme progresses. An overall description of drug discovery and development (up to clinical research) is shown in Table 1.2, with more detailed descriptions, relevant to the work of this thesis below.[16]

Table 1.2. Early-stage drug discovery

Stage	Sub-stage	Goal	Investigative techniques	Assay
Discovery	Target validation	To find and validate a protein, enzyme or other target receptor involved in a disease process.	Genetic analysis, analysis of phenotypic markers.	
Discovery	Lead identification: HitID	To find a compound that binds to the target (mM potency).	Understanding of structure-activity relationships, pharmacophore. Screening of natural products and library of compounds.	Fluorescence, UV-Vis Isothermal titration calorimetry NMR fragment screening X-ray fragment screening
Discovery	Lead identification: Hit-to-lead	To find a compound that binds to the target (nM potency).	Fragment based design, Ligand based design, Structure based design	Cell based assay, capable of showing biological result of compound binding.
Development	Lead optimisation	Maintaining potency (nM) while improving other properties (ADMET)	Ligand based design	In order of complexity Cell based Organ based Mice / rat studies Dog studies
Development	Pre-clinical			

1.2.1.1 Lead identification (Hit ID and Hit to Lead)

Lead identification is the stage of initially finding a series (a core skeleton and functional groups) with desired pharmacological activity for a specific target within the body. This can be effective in high concentrations and show properties and undesirable side effects, as this is the start of drug design and a developmental process.

This is performed in numerous ways, for example, by screening natural products such as: plants, microbes, marine and animal sources, venoms / toxins, or by screening synthetic compound libraries. Sometimes an existing drug has multiple uses or produces a side effect that can be used as a starting point for a new programme. In addition to the study of physical tests, a study of the target protein can be performed with X-ray crystallography and solution NMR, where a lead compound can be designed to fit the binding site.

1.2.1.2 Hit ID

Fragment based drug design

Fragment based drug design (FBDD) utilises many of the techniques of both structure-based design and ligand-based design. However, this process uses 'fragments' (smaller compounds than the end API) of known active compounds to assess a protein's binding site. A typical example of this method in practice is to soak protein crystals in solutions containing different fragments and then perform single crystal X-ray diffraction to determine where fragments have crystalised within the binding site.

The aim is to design linkers between fragments that are found to bind in different parts of the binding pocket. It is hoped that if multiple fragments bind to the protein active site, the information can be combined to create a more active species (pharmacophore analysis).

1.2.1.3 Hit to Lead

Ligand based drug design

It is possible to carry out virtual screening and drug design without protein crystal structures. This is where an established API's conformation is analysed, with pharmacophore analysis to search a library of virtual compounds to identify similar shape molecules.

1.2.1.4 Structure based drug design

The requirement for structure-based design is experimentally verified structural information about the target *i.e.*, a protein crystal structure. Development in crystallisation, X-ray sources, and understanding of analysis has meant a large number of proteins thus far have been crystallised, and their crystal structures are stored in the Protein Data Bank (PDB).[17]

Virtual screening is a technique used within structure-based design where libraries (conformational data composed of small molecule single crystal data and/or computationally generated compounds) are modelled with the protein crystal structure binding site to find 'hits'. A scoring function is then used to evaluate the molecule (*e.g.*, analysis of bond strain) and the compatibility with the binding site, based on energy calculations (see Figure 1.7). The best scores indicate potential hits, which are then synthesised and studied in real-world assays for further analysis of binding affinities and structural information.

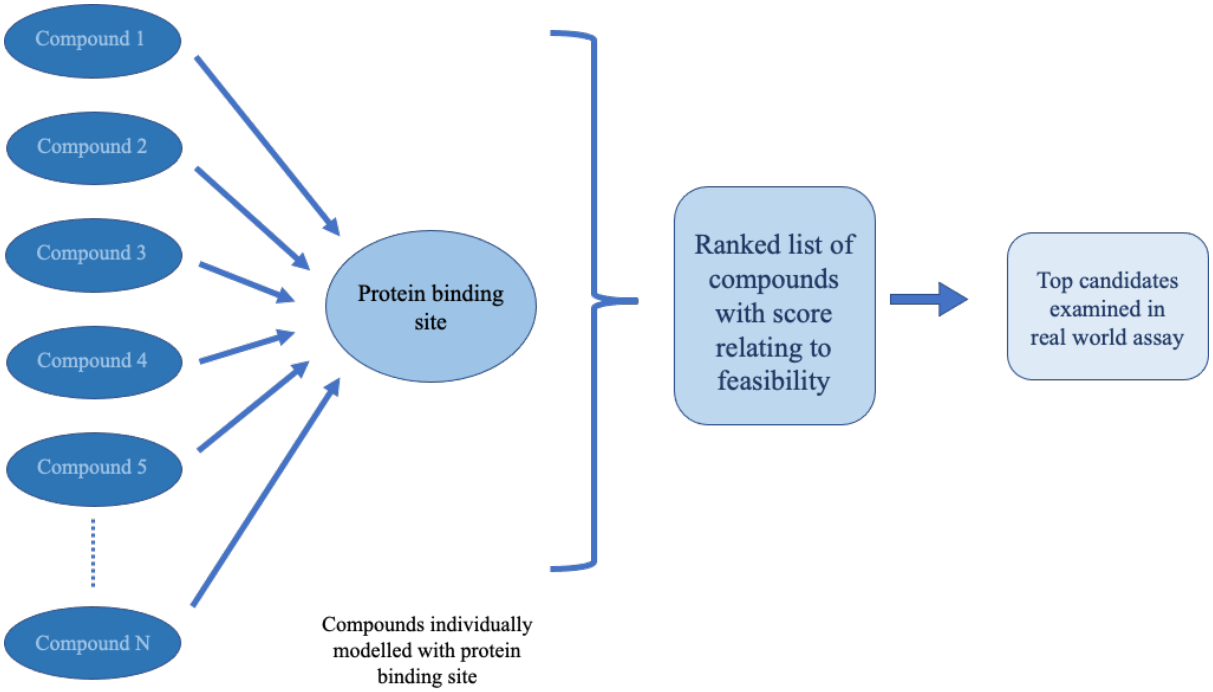


Figure 1.7. Simplified virtual screening roadmap using crystal structures and generated compound libraries to determine compounds with potentially good binding for real-world development.

Compounds that perform well in the development stage then move to the next. After animal studies are performed, to ensure safety, the compounds are used in clinical trials to study the side effects (common and rare), as well as to establish efficacy.

Each stage of drug discovery is enhanced with additional structural information, including knowledge of the binding site, the lead compounds, the ligands, *etc*. It is important to note that during the early stages of drug development, producing these novel compounds is a major cost to the study. These small quantities of synthesised compounds are consumed in assays, leaving very little material available for structural analysis.

Utilising in-house powder diffraction has the potential to reduce cost, time and materials for this analysis. This thesis aims to explore this and ultimately enable greater solid-state

structural information to be accessible from smaller amounts of material at earlier stages of drug programmes.

1.3 Modifying physical properties (crystal engineering)

The physical properties of a crystal form are dependent on the atoms and molecules' local environment within the crystal structure and the inter- and intra- molecular bonds formed. The concept of crystal engineering was first introduced by Pepinski (1955) and is defined by Desiraju as "the understanding of molecular interactions in the context of crystal packing and the utilization of such understanding in the design of new solids with desired physical and chemical properties.[18, 19]

For APIs that have undesirable physical properties (the most common being poor aqueous solubility), the options are (a) to develop an alternative compound, which is a lengthy and expensive process and not guaranteed to yield a better result, or (b) to attempt to change the crystalline state of the existing compound. This can be performed by finding a new polymorph or by changing the constituents of the crystal structure by *e.g.,* creating a salt, solvate or co-crystal.[20]

Crystal engineering is performed by manipulating the crystal growing conditions of a compound (or multiple compounds), such as changing the temperature, solvent, pH, seeding conditions, and more. Figure 1.8 shows the possible outcomes, guided by chance or design.[20]

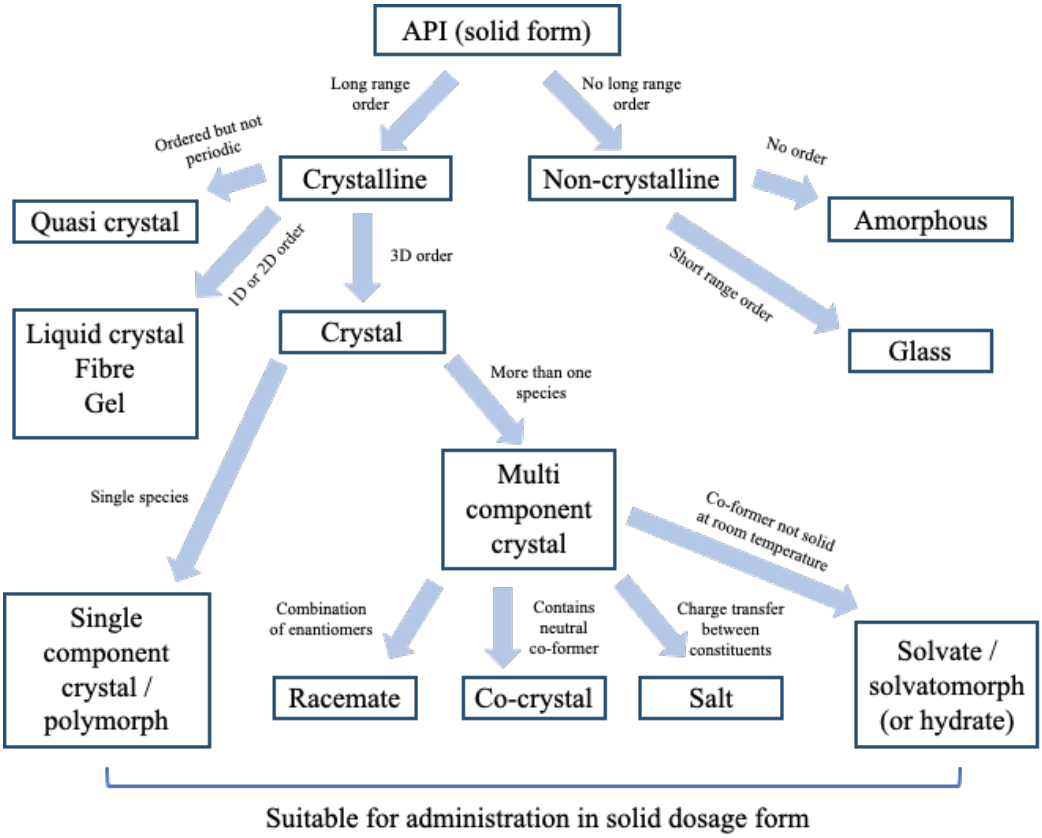


Figure 1.8. Flowchart for possible crystal engineering outputs, adapted from the role of co-crystals in pharmaceutical design.[21]

1.3.1 Polymorph screening

Polymorphs are different crystal structures (different unit cells and atomic coordinates) of the same chemical. The relative stability of polymorphs is temperature dependant; the polymorph that has the closest packing has the lowest Gibbs free energy at 0 K, but this changes as the temperature increases. The rate of change of Gibbs free energy is different for each crystal form, therefore on a phase diagram they generally intersect each other (Figure 1.9). If the point of intersection is above the melt temperature, then the cross-over is often not seen.

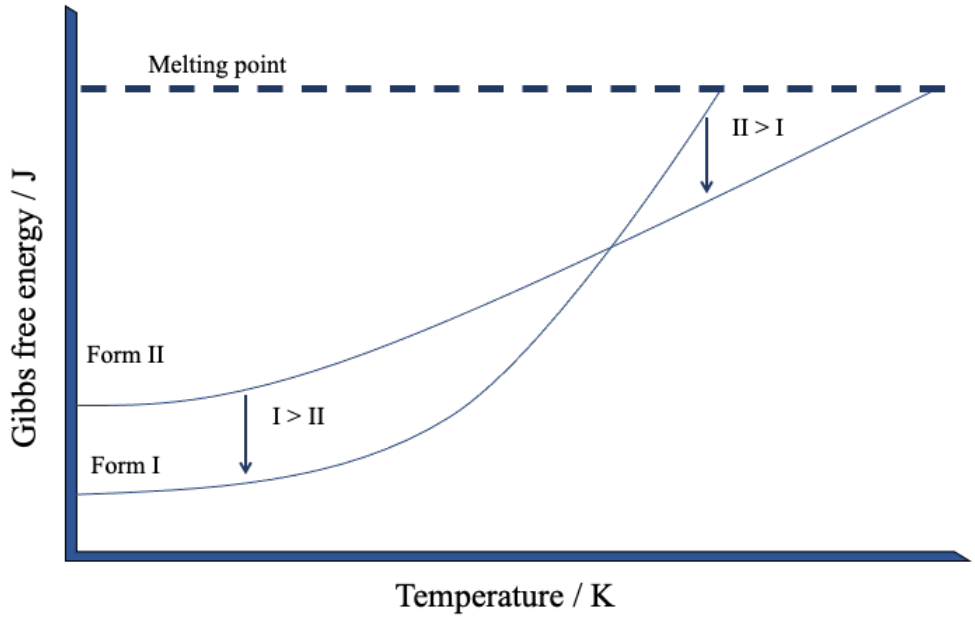


Figure 1.9. Phase diagram for enantiotropically related polymorphs, highlighting the different melting points of the different polymorphs.

The most stable crystal form of an API (at an ambient temperature range) is preferred for development, as this is often what is desired for the final solid dosage form. Other crystal forms found at this temperature are described as metastable polymorphs. The dissolution rates are invariably higher for these forms due to the metastable forms being more labile. The lowest energy crystalline form usually has the highest packing density and activation energy barriers to dissolution. Whilst meta-stable forms may have desirable properties, such as improved aqueous solubility or dissolution rate, there is a risk that they may transform over time to a more stable form (during production / storage) with consequences for production and clinical efficacy.

Ritonavir, an API developed to treat HIV-1, is arguably the most notorious example of polymorphism affecting the useability of a formulation. In 1998, two years after the introduction of ritonavir to the market, the appearance of a more stable polymorph (form II) was found in the product. This form's much-reduced aqueous solubility severely compromised the oral bioavailability of the compound and its consequent effectiveness as a treatment. Since then, 5 more forms of ritonavir have been found as a result of extensive laboratory screening.[22, 23]

Regulatory bodies such as the FDA recommend polymorph screening to establish the reliability of the solid form and stability under stress. It is not possible to say that the most stable form has been located, even following even an extensive polymorph screening process. However, this lack of certainty is typically mitigated with extensive laboratory polymorph screening and Crystal Structure Prediction (CSP) campaigns.

During pharmaceutical production, the chosen polymorph's physical properties (for example the compressibility and particle shape) affect the manufacturing process. This includes the excipients needed, coatings, glidants and disintegrants, *etc*., as well as the final solid form chosen for the API.

One of the best ways to search for new polymorphs is to create large numbers of recrystallisation experiments with a crystallisation robot. Initially designed for growing protein crystals, these robots enable the creation of many different chemical environments to best explore polymorph formation. As mentioned earlier, powder diffraction data obtained from crystals can also be compared with calculated powder diffraction data from hypothetical crystal structures returned by a Crystal Structure Prediction (CSP) campaign; this combined approach is increasingly gaining favour.[24]

Increased computing power, advancements in technology, and time spent on crystallisation experiments have led to the discovery of an increasing number of crystal structures with large values of Z'. It has been suggested by Desiraju that the most energetically stable form is the form that has a Z' = 1 and that high Z' structures are kinetic and meta stable products, whereas Anderson and Steed make the case that this may be an oversimplification. This will no doubt become an increasingly important field of study, with applications as yet unknown.[25, 26, 27, 28]

1.3.2 Multicomponent crystals

When polymorph screening of an API does not improve properties or produces an unstable form (that is therefore an unusable alternative), the API can be combined with co-formers or made into a salt form to improve the resulting material. A brief description of the most common ways to change the physical properties of an organic compound is given below.

Salts

Salt formation is the most common methodology for producing a series of crystal structures of the same compound but with varying physical properties, as most APIs contain basic nitrogen or acidic functional groups (such as carboxylic acids and tetrazoles). The salts are, of course, a physical form distinct from that of the original single component material, and these materials are most commonly used to improve aqueous solubility and dissolution rates. Salt formation can be with inorganic or organic co-formers. As long as the API has an easily ionisable functional group (*e.g.* -COOH, -NH_2), then there exists the possibility of producing a salt form. An important criterion in counterion selection is to use co-formers that have been previously used in approved drugs and are thereby Generally Recognized As Safe (GRAS).[19]

1.3.2.1 Solvates (and Hydrates)

A solvate is defined as any solid form that has a solvent molecule within its structure. Technically, solvates (of which hydrates are included) do not sit in the polymorphic landscape of an API as the crystal components are different. 'True' solvates or solvatomorphs (Figure 1.10) are those that have solvent molecules that "form an integral part of the unit cell of the crystal, are in stoichiometric ratio to the principal substrate, and are bound into the crystal lattice" by hydrogen bonds or other specific interactions.[19]

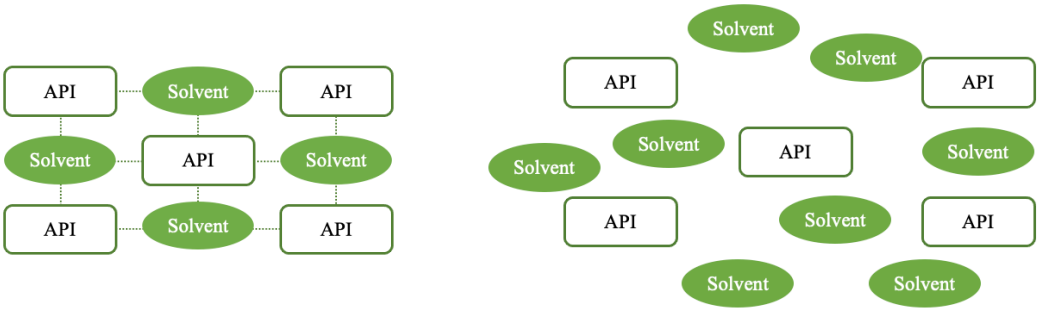


Figure 1.10. Solvent molecules as part of the crystal structure, solvate (Left). Solvent molecules trapped in void spaces / channels, in non-stoichiometric ratios, solvatomorph (Right).

Due to the toxicity and indigestibility of the majority of solvents, solvates are not favoured for pharmaceutical development, except for hydrates and occasionally ethanoates. Often however, hydrates are less water soluble than their anhydrous counterparts, which may be advantageous, depending on the desired outcome.[29]

1.3.2.2 Co-crystals

Co-crystals are often formed with the same methodologies as salts but they differ in being multicomponent solid forms where there is no charge transfer. An accepted definition of a co-crystal is that "co-crystals are solids that are crystalline, single-phase materials composed of two or more different molecular and/ or ionic compounds generally in a stoichiometric ratio which are neither solvates nor simple salts". Co-crystal formation can be used to improve physical properties by crystalising the API with another biocompatible organic compound. A Venn diagram showing the differences between different solid forms is shown in Figure 1.11. Although a wide variety of small organic compounds could be used as a co-former to manipulate the physical properties of an API, typically only compounds on the EAFUS (Everything Added to Food in the United States) or GRAS lists are studied.

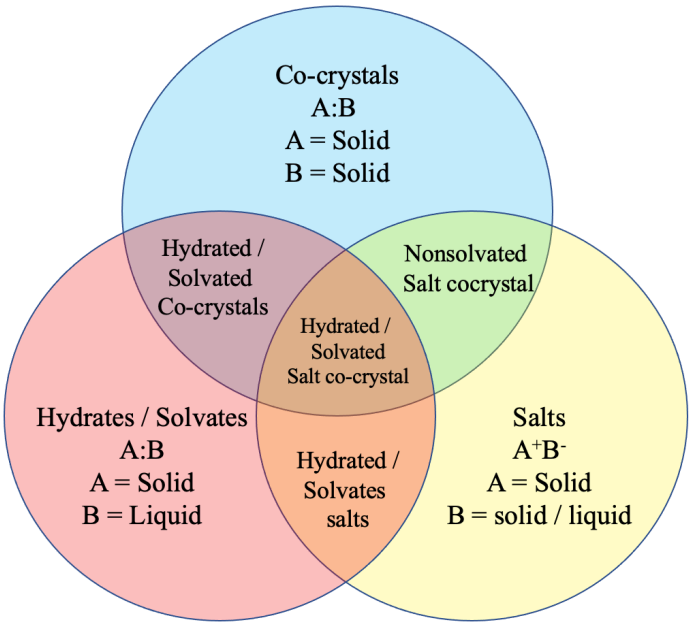


Figure 1.11. Venn diagram of terms given to multi-component crystal systems where A = API and B = co-former(s).[30]

In recent years, co-crystals have begun to move beyond the field of academia with multiple pharmaceutical co-crystals being given FDA approval. One example is Entresto, a co-crystal of valsartan (API) and sacubitril (used here as a co-former however itself is an active species) used as a treatment for cardiovascular disease (Figure 1.12). Entresto was approved by the FDA in July 2015 and subsequently has been approved for use in treating renal disease.[31]

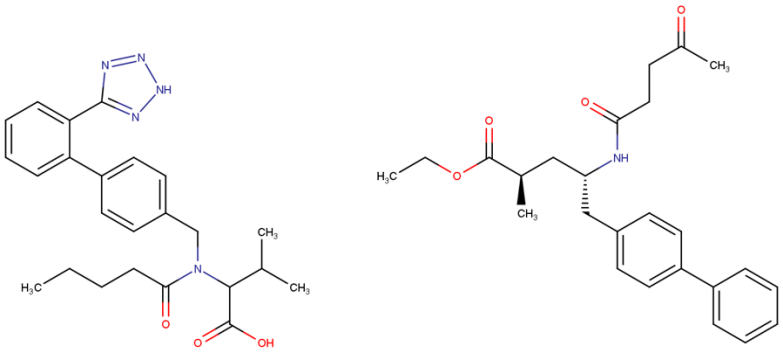


Figure 1.12. Entresto, a co-crystal of valsartan (API, left) and, sacubitril (co-former, right).

Unlike inorganic salts, organic co-crystals are known to be far more 'tuneable', as the co-crystal physical properties typically lie between those of the API and the conformer. For example, an API with a low melting point can be combined with a compound with a high melting point to produce a crystal with a melting point between the two, forming a eutectic.[32, 33, 34]

Predicting whether a co-crystal or salt will be formed is difficult. The general rule is that when Δ*pKa* [where $\Delta pK_a = pK_a(base) - pK_a(acid)$ *i*s < 0, there is a higher chance of co-crystal formation; for Δ*pKa* > 3, salt formation (complete charge transfer) is more likely; and for 0 < Δ*pKa* < 3, co-crystal or salt formation could take place.[35]

Typically, modern co-crystal development has focused on the manipulation of hydrogen bond formation, matching co-formers with APIs with complementary functional groups. Examples of these combinations are shown below in Figure 1.13. Co-crystals that form with hydrogen bonding from the same functional group are called homosynthons and different functional groups are called heterosynthons.

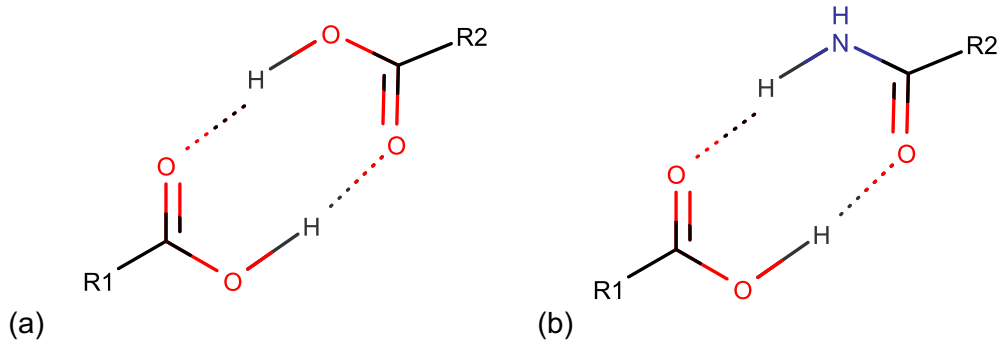


Figure 1.13. Examples of co-crystal formed of homosynthons (a) and heterosynthons (b).

It can be difficult for researchers to distinguish between the crystal structures of salts and co-crystals because they often only differ in the location of a single hydrogen atom. In practice, salt and co-crystals made of the same constituents can be identified by multiple analytical techniques. The most common is the use of high-quality single crystal X-ray diffraction to locate the position of the relevant H atom from a difference Fourier map.

If no single crystal can be obtained (*e.g.*, in the case of a sample produced by mechanochemical means), there are a variety of techniques used alone and in combination to analyse the hydrogen bond lengths. The most available of which is Fourier transform infrared (FTIR), used to analyse the change in intermolecular interactions, showing where hydrogen bonding has changed. When available, techniques such as solid-state NMR are highly effective for identifying charged states, the substantial change in chemical shift being caused by complete charge transfer. Lastly, neutron diffraction (available at pulsed sources) has been used to accurately determine hydrogen atom positions, to probe shifts in atom distances, and to study the changes from salts to co-crystals.[36]

For co-crystals, a good rule of thumb used by researchers is that the melting point of the co-crystal is between the melting points of two starting materials, whereas for protonated salts, ionic interactions play a dominant role, resulting in higher melting points.

1.3.2.3 inclusion complexes

Inclusion complexes such as those produced with cyclodextrins can also be used to help improve the poor aqueous solubility of an API. The API is sequestered inside the hydrophobic interior of the co-former, with the co-former's hydrophilic exterior providing enhanced aqueous solubility.[37, 38, 39, 40]

1.4 Diffraction

Diffraction is defined as the bending of waves around an edge and by extension the spreading out of waves that encounter an aperture of a similar magnitude to the wavelength. This is illustrated in Figure 1.14, showing maximum diffraction when the aperture is equal to the wavelength, λ.

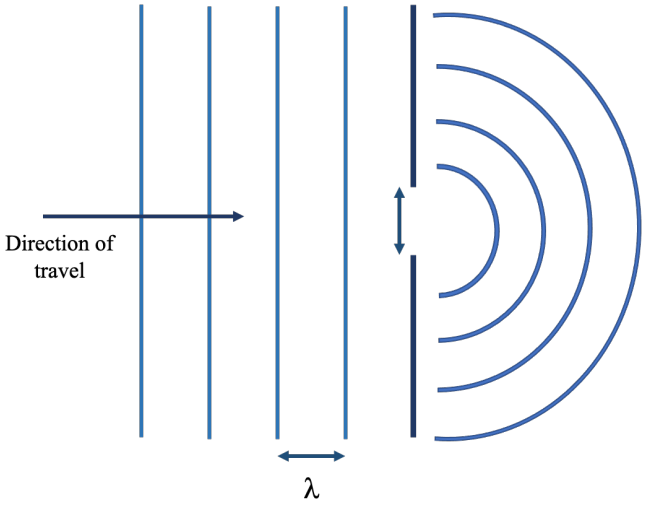


Figure 1.14. Diffraction of an obstructed wave, with aperture of similar magnitude to wavelength, λ.

When a barrier has multiple apertures (or slits), this creates a diffraction grating and a wave that passes through creates an interference patten, as the result of constructive and destructive interference. The resulting diffraction pattern depends on the parameters set out in the Bragg equation (Equation 1.1).[41]

$$n\lambda = 2d\sin\theta$$

Equation 1.1. The Bragg equation, where θ is the angle of incidence to the barrier, n is the order of diffraction, λ is the wavelength of light, and d is the slit separation.

1.4.1 Single crystal diffraction

Crystal structures act as 3-dimensional gratings for X-rays due to the periodic nature of the arrangement of atoms within them. Although X-rays span a range of wavelengths, the vast majority of structural analysis work carried out on crystals uses radiation in the range 0.1 – 5 Å, these being well suited to the distances being probed within the crystals. When a single crystal is exposed to an X-ray beam, it produces a series of constructive interference diffraction, which when intercepted by a 2D detector creates a pattern of spots. This pattern, which exists in so-called reciprocal space (Figure 1.15) can be interpreted in terms of reflections from sets of hypothetical planes (each set having its own d-spacing value) inside the crystal.

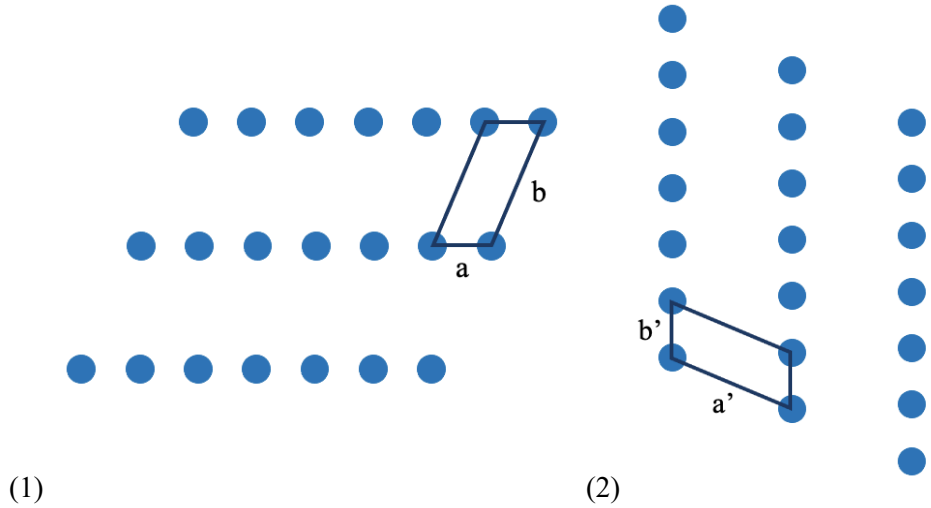


Figure 1.15. Array in real space with repeating unit, lengths [a, b] (1) and the corresponding lattice in reciprocal space with lengths [a', b'] (2).

The observed Bragg reflections can be characterised by their hkl indices, and their structure factors, F_{hkl}. The structure factor itself can be expressed in terms of the structure factor magnitude, $|F_{hkl}|$ and the phase angle ϕ_{hkl} as shown in Equation 1.2.

$$F_{hkl} = |F_{hkl}|\exp(i\phi_{hkl})$$

Equation 1.2 Structure factor equation, with structure factor F_{hkl}, structure factor amplitude, $|F_{hkl}|$, and phase angle ϕ_{hkl}.

If the intensity and phase (ϕ_{hkl}) values are known, the crystal structure (specifically the electron density in the unit cell) can be calculated directly by Fourier synthesis. The phase angles of diffracted beams cannot (except in exceptional circumstances) be determined experimentally; this loss of phase information during a diffraction experiment is known as the **phase problem** in crystallography.

Equation 1.3 is used to model the electron density map, and so determine the atom positions. This is performed by modelling the position and number of electrons at different coordinates within the unit cell iteratively until the calculated and observed are in agreement.

$$\rho_{(x,y,z)} = \frac{1}{V}\sum_{\mathrm{hkl}} \mathrm{F_{hkl}} \exp(-2\pi \mathrm{i}(hx + ky + lz)$$

Equation 1.3. Inverse Fourier equation, ρ is the electron density at point x, y, z within the unit cell, F_{hkl} is the structure factor, h, k, and l are Miller index values, V is the volume of the unit cell.

Several methods have been developed to address the phase problem; Patterson's approach, based on structure factor amplitudes alone, was very successful but was superseded by the so-called "direct methods" that became the mainstay of crystal structure determination. These are described in more detail in section 1.6.1.

1.4.2 Powder structure solution and inherent difficulties

As previously mentioned, powders or polycrystalline materials are the combination of many crystallites that are randomly orientated with respect to each other. Powder diffraction presents a different set of challenges to crystallographers since when the sample is hit with X-rays, the collection of small crystallites each diffracts and combines to give a pattern of rings rather than a 2-dimensional pattern of spots. What is seen by the researcher is the compression of the 3-dimensional array of spots from a single crystal into a 1-dimensional diffraction pattern, shown below in Figure 1.16. When a radial slice is taken of these rings, a 1-dimensional dataset consisting of the 2θ angle and the intensity is produced.

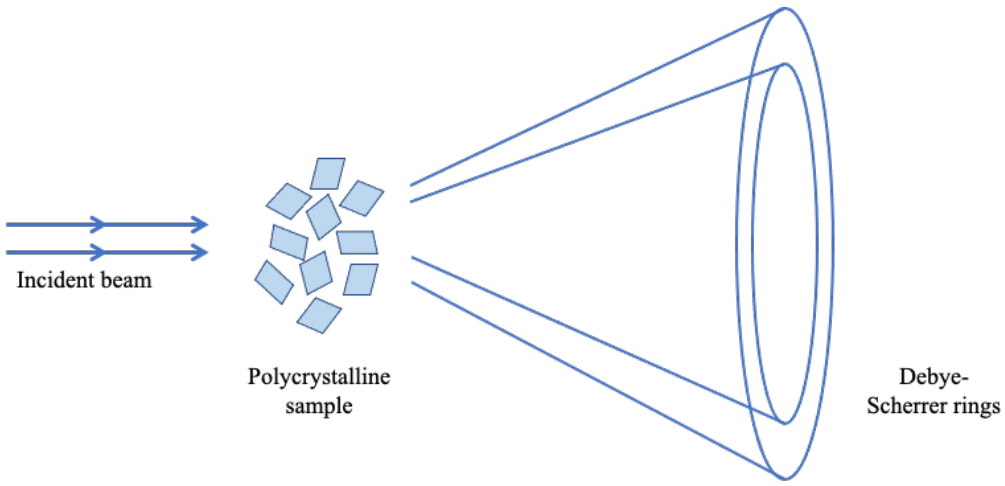


Figure 1.16. Diffraction cones produced from irradiation of polycrystalline material with X-rays.

As with single crystal diffraction, the diffracted intensity drops as a function of increasing 2θ angle due to scattering form, temperature, and Lorentz polarisation factors.[42]

Crystal structures with lower symmetry (organic molecules tend to crystallise in lower symmetry space groups, with over a quarter of CSD entries in P1 and P-1) have a greater chance of exhibiting accidental peak overlap. The greater the extent of such peak overlap, the more difficult it is to identify individual reflection positions, making indexing more difficult. Despite this, modern X-ray sources and detectors have meant instrumental resolution is sufficiently good that the majority of single-phase PXRD datasets can be indexed without too much difficulty.

During the collection of powder diffraction data, it is assumed that within the sample all orientations are represented equally. Unfortunately, often where the crystallites pack together some reflections are overrepresented (for example, needles packing in the same direction), known as preferred orientation (PO). Figure 1.17 shows the difference seen in a sample of ibuprofen (CSD refcode IBPRAC01) with no PO compared to PO along one axis. Other factors than can change the diffraction pattern include crystal strain, crystallite size, impurities, and amorphous component.

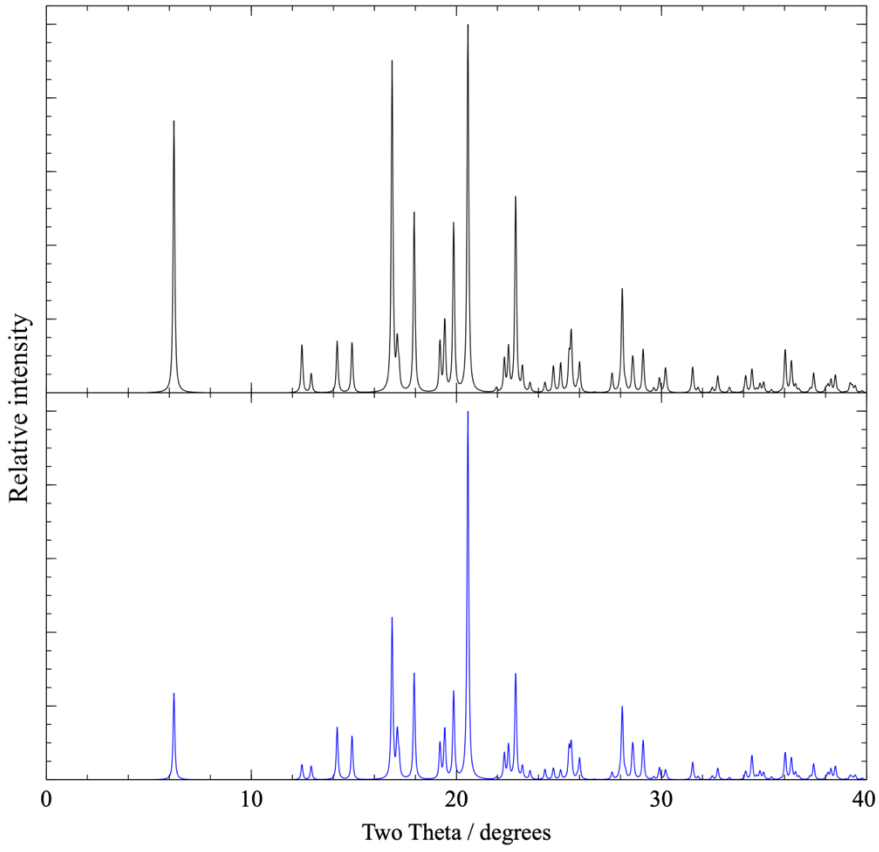


Figure 1.17. Ibuprofen generated pattern (CSD refcode IBPRAC01) with no preferred orientation (above) and with preferred orientation in the 100 plane with March-Dallase parameter set to 1.3 (below).

Structure solution from powder diffraction data is discussed in section 1.6.

1.5 Diffraction data collection techniques

1.5.1 In-house lab equipment

A variety of configurations have been designed to best collect diffraction data from different sample types. Outlined below is a description of what is available to most researchers both in industry and academic institutions *i.e.*, in-house equipment.

1.5.1.1 Single crystal diffractometers

Modern single crystal diffractometers are able to collect diffraction data suitable for structure solution from very small single crystals – typically of minimum single dimension of 50 μm. These diffractometers are widely available and practical for high-quality routine measurements. Single crystal goniometers are also typically capable of mounting a Lindeman capillary, enabling the collection of powder diffraction data (for phase identification and indexing).

The standard for in-house diffractometers is to use either Mo Kα (0.78 Å) or Cu Kα (1.54 Å) radiation from a stationary or rotating anode source (this latter producing higher intensity of emitted radiation). Other radiation sources are also becoming more common for in-house diffractometers, such as: Ag for Pair Distribution Function (PDF) analysis of amorphous materials; and liquid Ga (a brighter radiation source than Cu or Mo).[43, 44]

1.5.1.2 Powder diffractometers

Powder diffractometers are available with a range of configurations, depending on the desired data quality and purpose for data collection. For example, a smaller bench top diffractometer may be well suited to fast, routine measurements used for phase identification, whereas it is recommended that data collection intended for structure solution is collected with a more specialist configuration. Powder diffraction can be collected in two geometries, by reflection or by transmission, as illustrated in Figure 1.18. Different geometries require different methods of sample preparation to accommodate collection modes and to achieve the highest quality dataset.

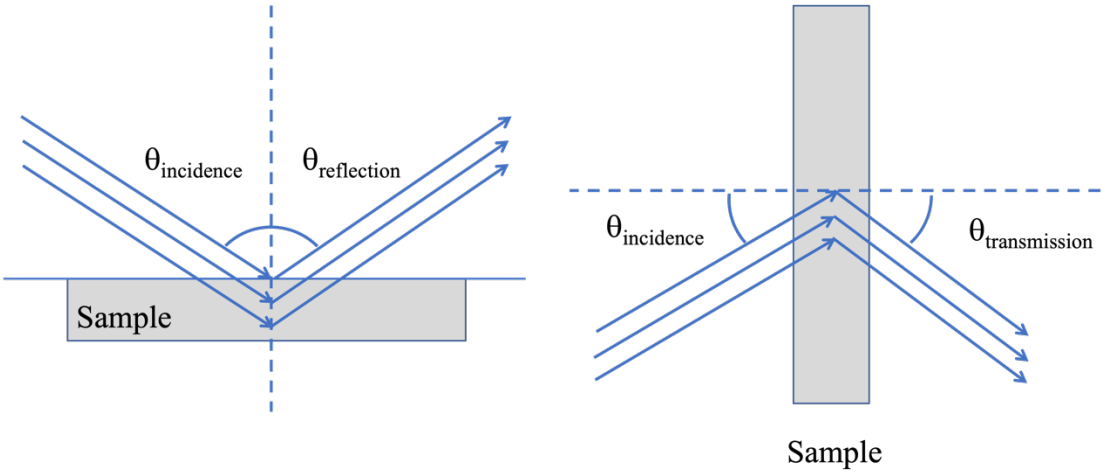


Figure 1.18. Simplified diffraction geometries, reflection (left), transmission (right). The incoming beam is shown as being parallel for simplicity."

These geometries are found in 3 main setups: reflection plate, transmission plate, and transmission capillary. Below (Table 1.3) is a summary of the advantages and limitations of each configuration.

Table 1.3. Methods for in-house diffraction data collection

Geometry setup	Sample amount (*approx.*)	Advantages	Limitations
Flat plate reflection	1 mg to several hundred milligrams	Flat plate reflection setups provide quick data collection, taking as little as 10 – 45 mins. The addition of low background sample holders means 1 - 2 mg of material can be used. These setups are useful for variable temperature / humidity experiments. Setups with sample-changing robots enable data collection automation, allowing for large numbers of samples to be collected in one session.	Reflection flat plate data collection is not used routinely for molecular structure determination because the data are often affected by preferred orientation, skewing peak height ratios. Setups are also susceptible to systematic errors from height displacement. Additionally, standard in-house setups are not equipped to collect data for air sensitive compounds.
Flat plate transmission	5 mg to several hundred milligrams	A high throughput methodology that uses a specialist sample holder, the diffraction patterns show reduced preferred orientation compared to equivalent reflection flat plate data collections.	Transmission flat plates require a specific configuration for data collection. Not ideal for single samples as trays designed for multiple samples to be collected in a single session. Therefore, transmission flat plate is not a benchtop technique.
Transmission capillary	5- 10 mg	Highest quality output for in-house data collection, preferred orientation is less prevalent in transmission capillary data collection. Capillaries are ideal for studying air sensitive compounds and can be mounted on powder diffractometers as well as a variety of SX diffractometer setups.	A high-quality dataset for Rietveld refinement takes a several hours to collect, so is not a high throughput technique. Capillaries require 5 – 10 mg of sample (organic compounds) to perform data collection, not accounting for damaged capillaries and materials that are difficult to load into capillaries *e.g.,* ‘sticky’ samples.

1.5.2 Synchrotron data collection

A synchrotron is a cyclic particle accelerator where the electrons are accelerated close to the speed of light (a schematic is shown below in Figure 1.19). At first the particles are generated and accelerated by an oscillating electronic potential in the linear accelerator and then steered by magnets in the storage ring. As the particles change direction, they emit X-rays (at a variety of wavelengths) at tangents to the storage ring.

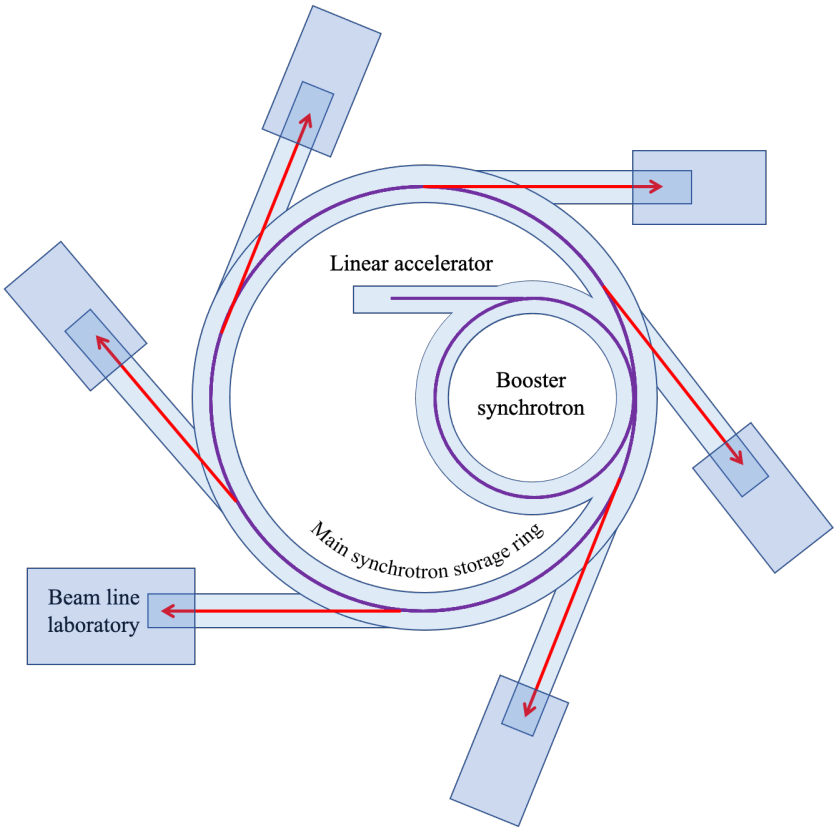


Figure 1.19. Simplified synchrotron schematic showing the line of the electrons (purple) and the radiation produced at tangents (red).

The wavelength of radiation can be altered using 'wigglers', producing a continuous spectrum of wavelengths or 'white' X-ray radiation. Individual wavelengths are chosen and produced with a monochromator. The resulting X-rays are typically 100-10,000 more intense than laboratory Kα X-rays. [45, 41]

As described in section 1.1.2, X-ray diffraction is an ideal methodology for studying changes in material states and chemical reactions. The beam lines offer superior setups for these studies as well as being superior for protein crystal analysis because the powerful source compensates for their typical small crystal size and large unit cells. This analysis is often destructive for organic samples and so typically multiple crystals are prepared for analysis.

1.5.2.1 XFELs

X-ray free electron lasers (XFELs) were first created at the Stanford Linear Accelerator Centre (SLAC) in 2009, like other lasers are produced by High Harmonic Generation (HHG). These short pulses, lasting between 10^{-12} and 10^{-15} seconds are used in various methods of analysis for probing systems to study reactions by absorption, diffraction, and imaging. Most relevant to this thesis is their use in structural analysis of micro sized proteins in a method called serial femtosecond crystallography (SFX). Often the high intensity radiation rapidly destroys the single crystal and so to compensate a flow of micro crystals is irradiated and the datasets are combined to give a whole single crystal dataset.[46, 47, 48]

1.5.3 Electron diffraction

Incident electrons are scattered by Coulombic interactions with the atoms, whereas X-rays are scattered by the electron cloud (the electron density). This ultimately means that electron scattering is greater, by a factor of 10^3 or 10^4 than is seen with X-rays, and electrons therefore interact with crystal samples much more strongly. As a consequence, electron diffraction can be used to analyse samples with a diameter between 1 and 10 μm. De Broglie found that all particles with a mass and a velocity have a wavelength, and since electrons have a mass their wavelength can be calculated from the Equation 1.4 below. [49, 50]

$$\lambda = \frac{h}{mv}$$

Equation 1.4. de Broglie equation, λ = wavelength (m) h = the plank constant = 6.63×10^{-34} Js, m = mass (Kg), v = velocity (ms^{-1})

This ultimately means the exact wavelength of the electrons can be controlled by changing their velocity.

Although electrons scatter more strongly than X-rays, electrons attenuate faster. Importantly, this means thinner samples are required for effective electron diffraction. Electron diffraction (ED) is able to analyse nano crystals, meaning samples used for data collection are a fraction of the size required for traditional in-house X-ray diffractometers.

One example of this is the study of ABTPA (Figure 1.20), a hydrogen bonded framework (HOF), where structure determination was performed using electron diffraction in combination with powder diffraction analysis.[51]

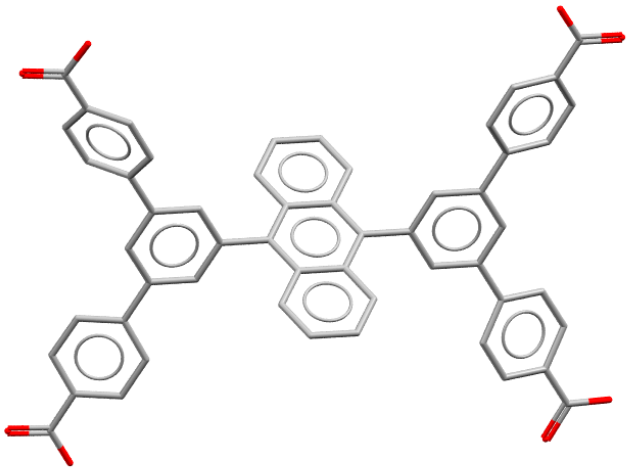


Figure 1.20. ABTPA, (CSD refcode KUPPUJ) a hydrogen bonded organic framework, solved using electron and powder diffraction.[51]

An inherent disadvantage of electron diffraction data collection is the damage to materials from the electron beam. Diffraction relies on elastic scattering, where the energy of the particle is the same before and after interaction with the structure. For high energy electrons, there is a proportion of particles that scatter inelastically, where the sample absorbs energy causing chemical bonds to break and altering the structure of the sample. This is problematic for biological materials *i.e.,* protein crystals, that are particularly sensitive to beam damage. [52]

1.5.3.1 Kinematic vs dynamic scattering

In X-ray diffraction analysis, it can be assumed for practical purposes that the X-rays scatter kinematically, meaning each photon is reflected from the sample only once. Dynamical scattering, (shown below in Figure 1.21) occurs when the reflected beam propagates though the crystal at the Bragg angle (θ), then the re-reflected beam then travels parallel to the incident beam within the crystal, causing destructive interference with the incident beam.

Dynamical scattering in X-ray diffraction is only seen with near-perfect crystals, as typical samples consist of a slightly mis-orientated mosaic. This is not the case for electron diffraction, where dynamical scattering plays a major role.

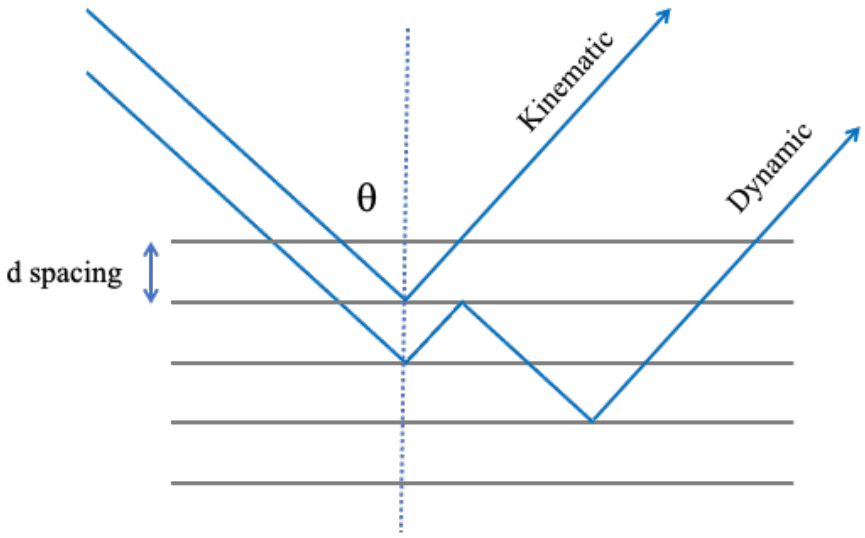


Figure 1.21. Kinematic (single reflection) vs Dynamical (multiple) scattering within a crystal lattice.

Electron diffraction, like SCXRD does not study bulk materials, therefore is generally not suitable for uses such as quality control. Currently, electron diffraction is not widely available due to the cost of analysis and limited expertise. Manufacturers, such as Rigaku, are developing laboratory electron diffractometers (a demanding task as ED is performed in a vacuum), meaning electron diffraction will become available as an in-house technique in the near future. Dynamical scattering effects require a modified approach to structure analysis than current X-ray diffraction software, which current crystallography software is not yet fully adapted for. This presents a challenge that needs to be overcome in order for analysis to become routine.

Despite these initial difficulties, ED will likely become a significant and routine structure determination technique for microcrystal analysis in the future.

1.6 Structure solution

Several methods, and subsequently computer programs, have been developed and refined to solve the phase problem (see section 1.4) and are summarised in Table 1.4. A brief description of the key approaches to structure determination, particularly for single crystal diffraction data, is given below.

1.6.1 Patterson methods

In 1935, Patterson introduced a function (see Equation 1.5) for calculating electron density without knowledge of the structure factor phases. The intensity of a reflection (proportional to the square of the magnitude of its structure factor) can be reliably and accurately measured and can be used in this method to determine interatomic vectors, from which atomic positions can be deduced. Known also as the heavy element method, the determination of one or more heavy elements enables a successive approximation of the unit cell, as peak heights are proportional to the number of electrons so act as anchors for structure solution.[53]

$$P_{(u,v,w)} = \frac{1}{V}\sum_{\mathrm{hkl}}|F_{hkl}|^2\cos\{2\pi(hu+kv+lw)\}$$

Equation 1.5. Patterson function $P(u,v,w)$, $|F_{hkl}|^2$ is the structure factor amplitude, h, k and l are Miller index values, V is the unit cell volume.

The Patterson method is less effective with structures containing many atoms, especially when none of the atoms are particularly strongly scattering. The result is a large number of interatomic vectors leading to a very crowded, and therefore difficult to interpret map.

1.6.2 Direct methods

'Direct methods' is a term for a collection of methods, based on the original work of Karle and Hauptman (1950). This approach exploits statistical correlations between reflection phases and the knowledge that the scattering density everywhere is positive, in order to calculate likely values for phases. Direct methods were transformational in allowing crystallographers to solve crystal structures routinely. Before direct methods the computational power needed to 'brute force' the problem was simply not available. The legacy of direct methods is that organic materials can be accurately and rapidly studied by X-ray crystallography and have become the basis for modern single crystal structure determination. Initial software packages such as SHELXS have been largely superseded by dual-space methods (*e.g.*, as implemented in SHEXLT) and charge flipping (*e.g.*, as implemented in SuperFlip, OLEX2).[54, 55, 56, 57]

1.6.3 Summary of current methods

Table 1.4. Methods for structure determination

Structure determination method	Developed for	Summary	Program(s) / publications that employ methodology
Patterson method for powder	Simple inorganic powder data and amorphous material	This method reveals to the user the inter-atomic distances (more detail is given below). Originally, 2D slices of the structure are solved and stacked for single crystal. See section 1.6.1.	EXPO2014[58]
Direct methods	Single crystal data – modified for powder	Direct methods use relationships between reflections to determine the coordinates of electron density. Modified for powder data by incorporating structural information in the intensity-extraction stage. This method does not require knowledge of the structure's connectivity or coordination of the atoms. The quality of the density map is dependent on the number and accuracy of the reflections.	EXPO packages COVMAP (part of EXPO) Altomare *et al*., (2012, 2013)[59,60]
Direct methods sum function	Single crystal data	Involves the maximisation of the sum function (S), such that the value of S is at maximum for the correct set of structure factor phases. This method utilises different maximisation approaches, *e.g.*, fast Fourier Transform and modified tangent, however it requires atomic resolution data.	XLENS part of fullprof suite[61, 62]
Anomalous scattering (in combination with maximum entropy)	SX protein crystallography	This technique uses Friedel pairs, h and –h (which overlap exactly in powder data). The method relies on differences in average intensity as a function of more than one wavelength. This approach has been shown to work well for single crystal proteins, however it requires a tuneable x-ray wavelength source, which requires synchrotron availability.	Burger *et al* (1998)[63]
Isomorphous replacement	SX Protein crystallography	The methodology is to introduce heavy atoms into the structure and observe the differences. The change in scattering will be due only to the introduced atom. The result	Basso *et al* (2010)[64]

(related to Patterson's)		is to correctly determine structure factors for the heavy atom which can be used to help uncover the rest of the structure. Isomorphous replacement is a good option in the crystallographers' tool kit as a way to begin structure determination with the help of the anchor atom(s). Unfortunately, this requires multiple crystallisations with different heavy atoms, and there is no guarantee the introduction of a heavy atom will crystallise the same way if at all.	
Maximum entropy	Powder data	An iterative method whereby the Fourier map is given as a basis set (p_i is pixel density) and is used to produce an inverse Fourier transform constrained by the fact that it must return the original structure amplitudes and phases, as well as new amplitudes and phases that are reprocessed as part of a new basis set. Continued until convergence. Maximum entropy methods have been shown to calculate solutions in good time (given enough computational power) and can be used in combination with other approaches such as Rietveld refinement. Importantly, the crystallographer does not require knowledge of chemical structure. However, because of the high computational demand, the method was not widely adopted and has since fallen from use.	MICE Gilmore *et al.,* (1991)[65] Nishibori *et al*., (2008) – used when other method failed.[66]
Charge flipping	Single crystal, modified for powder	The majority of charge density is located in a small volume of the whole cell. The density map is created by assigning random phase angles, unobserved magnitudes are assigned a magnitude of 0. The FT map is altered by changing the sign of the structure factors that lie below a certain threshold, δ. This is done iteratively until the map converges. The R factor determined comparison to the observed amplitudes. Charge flipping has been shown to work well with atomic resolution powder data however, it has not been shown to perform well against other methods because powder data often doesn't provide atomic resolution. Improvements made by combination with histogram matching.	SUPERFLIP Palatinus & Chapuis (2007) Sisak *et al.,* (2012) [67,56]
Molecular envelopes	Powder data	This is the idea of limiting the volume in which the atoms must fit. This is done by correctly phasing a few low index strong reflections. This is used to generate surfaces, highlighting the regions the rest of the atoms must lie. With low angle reflections, charge flipping has	McCusker *et al.,* (2001)[68]

		shown to be effective (as these reflections are less likely to suffer from peak overlap). The method is limited as it is not simple to carry out the initial phasing.	
Molecular replacement (Part of model building)	Powder data	A crystal structure of a related molecule to the molecule of interest is used as a model for structure solution. This assumes a very similar structure will achieve the actual structure without being too computationally demanding. Crucially, this method does not require atomic resolution data for a successful structure solution. Determining which related compound can be difficult and there is no guarantee the molecule chosen will be similar enough to be effective.	MOLREP Vagin & Teplyakov (2010)[69]
Global optimisation (Includes model building)	Powder data	This is a broad description of methods that use a model of the molecule of interest and uses programs to optimise the model to the observed powder pattern. This is discussed further on in the two most frequently used methodologies, genetic algorithms and simulated annealing. See section 1.6.4.1.	DASH[70]
Maximum likelihood methods (A part of global optimisation)	Powder data	Omitted parts of a molecular model, hydrogen atoms, counter ions or solvent are inputted into the model and treated as a 'blur'. The corresponding electron density is distributed across the unit cell and increases uncertainty in the calculated structure. This allows multi-fragment systems to be solved with global optimisation methods. The method is computationally expensive; generally superseded by modified DM and global optimisation methods	Markvardsen *et al.,* (2002)[71] Favre-Nicolin & Cerny (2004a)[72]
Local minimisation	Powder data	A large number of runs are created in random locations in the solution space; each is set to find the local minimum (no uphill moves). The law of averages predicts that 1 or more of the runs will be randomly placed near enough to the global minimum. Local minimisation is a fast and simple approach for each individual run and has shown to be competitive with simulated annealing (a global optimisation method). It should be noted that each run has a low probability of finding the global minimum. There is no structure on solution with this method and it is limited by the number of runs computationally viable.	TALP Vallcorba *et al.,* (2012)[73] Shankland *et al.,* (2010) in combination with simulated annealing[74]

1.6.4 Structure Determination from Powder Diffraction Data (SDPD)

The number of crystal structures that have been derived from powder diffraction has expanded considerably within the last four decades (Figure 1.22). The increasing ease of structure determination has been due to advances in powder diffraction methodologies, reduction of cost, and availability of powerful computers. In addition, there is a greater willingness to now try to solve crystal structures from powder data, in light of demonstrated successes.

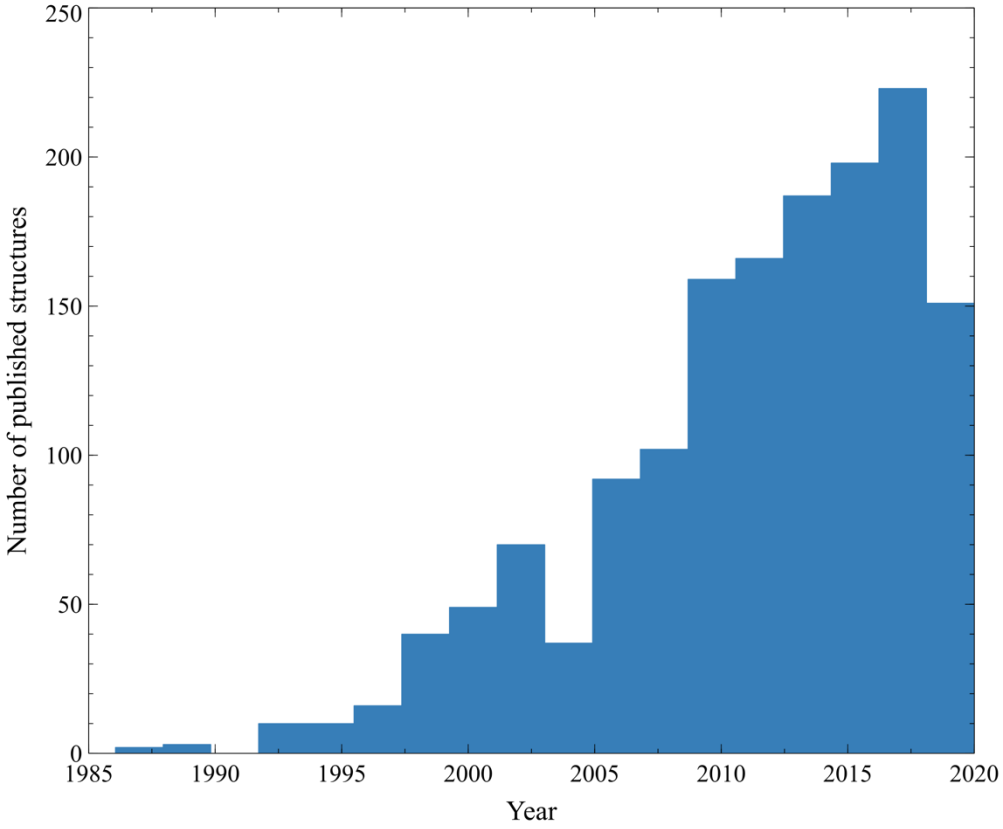


Figure 1.22. The number of unique, organic SDPD-derived structures deposited in the CSD between 1985 and mid 2020.[75]

There is one main class of SDPD methods that has transformed the ability of scientists to tackle the molecular materials of interest in this work – the so-called 'direct space' approaches.

1.6.4.1 Direct space methods

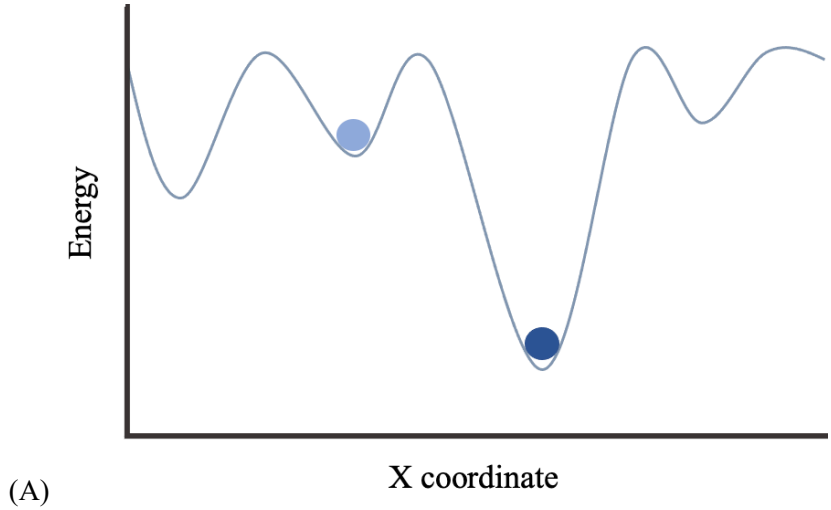
Direct space methods exploit *a priori* knowledge of the molecular connectivity within the crystal structure. The user builds a molecular model based on the known connectivity and well-characterised bond lengths and angles, and this model is inserted at a random position, orientation, and conformation (a function of torsion angles in the molecule that are free to rotate) within the previously determined unit cell. The structure solution is then performed by adjusting the model, comparing the calculated and observed diffraction pattern after each adjustment, and repeating until a good match is found. This adjustment can be performed

systematically, and this approach is used to locate the best fit to the observed data, provided sufficiently small steps are used to adjust the position, orientation, and rotatable torsion angles.

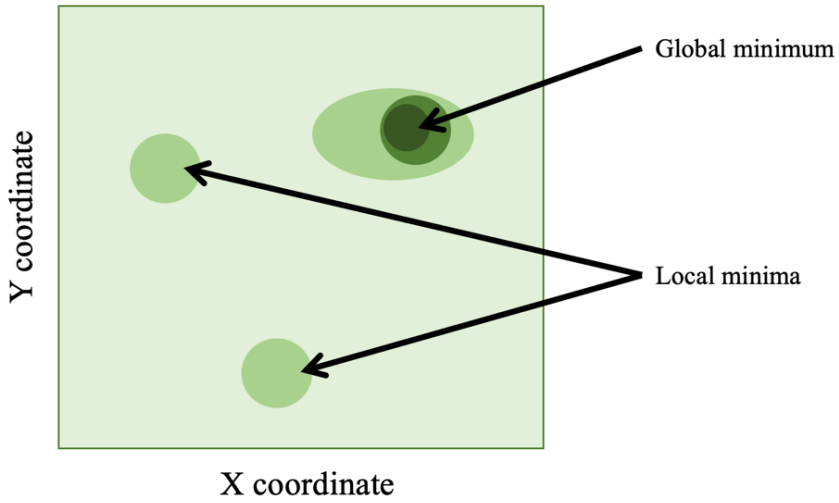
Unfortunately, systematic searches can become prohibitively computationally expensive and could take months (or even years) to run. As such, global optimisation methods are used to negotiate the space that describes the agreement between observed and calculated diffraction data, as a function of the variable parameters that describe the system. Two such methods have found particular usage in SDPD – genetic algorithms (GA) and simulated annealing (SA).

1.6.4.2 Global and local minima

Minima in a function hypersurface are most conveniently discussed in terms of the energy of a system that depends upon the configuration of the atoms within the system. A local energy minimum is simply a minimum on a hypersurface, whereas the global minimum is the lowest point, describing the lowest 'energy' configuration. This can be represented visually in many ways; below in Figure 1.23 are simple 1D and 2D representations. Let us assume for the moment that we wish to locate the global minimum on the hypersurface. It is clear that if we start a search from some random point (an x coordinate, in the case of Figure 1.23A) and use a local minimiser to find a lower energy point, we will end up trapped in a local minimum as the local minimiser will only proceed in downhill steps. In Figure 1.23A, the light blue circle is trapped within a local minima because the minimiser cannot make uphill (increasing energy) moves, and remains unaware that a much lower energy point (indicated by the dark blue circle *i.e.*, the global minimum) exists nearby. Figure 1.23B shows a similar situation for a surface that depends upon two variables, the X and Y coordinates.



(A)



(B)

Figure 1.23. Energy maps shown as 1-dimensional surface (A) and as a 2D-hypersurface (B). In the latter, the darker the shade, the lower the energy.

In the case of SDPD, we are not looking to minimise energy; rather, we are looking for the global minimum on the hypersurface that describes the level of agreement between observed and calculated diffraction data as a function of the unknown parameters (so-called degrees of freedom, DoF) that describe the crystal structure. The global minimum of this surface corresponds to the solved crystal structure. Typically, the level of agreement is expressed either as the familiar Rietveld R_{wp} value, or (as is the case in the widely used DASH program) a χ^2 value. The global minimum of this agreement hypersurface corresponds to the solved crystal structure, and also corresponds with an energetic minimum, a fact that can be exploited for crystal structure validation using Density Function Theory (DFT) calculations.

1.4.2.3 Model building

Direct space methods rely on known chemical information for structure solution. Software packages such as MarvinSketch or CHEMdraw produce optimised geometries and bond lengths for inputted chemical formulas. Model building is a 'best guess' and so is not a replacement for structure determination.[76, 77]

Mercury, part of the CCDC, utilises the MOGUL database to determine the likelihood of bond lengths and angles. Values that do not match the ranges (*i.e.* a bond is uncharacteristically long) within the database are flagged to the user for further optimisation and/or refinement.[78]

1.4.2.4 Monte Carlo methods

Monte Carlo methods are a group of algorithms that rely on randomness to aid with non-deterministic problems (*i.e.*, those that have a specific answer and that could be found simply if all of the information is available). The most powerful aspect of using randomness (known as a stochastic approach) in the solution-finding process is the idea that although the user may not have the time or computing power to visit every possible solution, they have the ability to visit any point. Essentially, samples are drawn randomly and used to approximate the desired quantity. [79]

To demonstrate this visually, Figure 1.24 shows an attempt to estimate the value of π. Given that π is an irrational number, but still has a specific value, it cannot be determined by an equation. However, the use of randomness can greatly simplify the solution. By randomly placing points onto a square (with side length 1) the ratio of points with a distance ≤ 1 from one corner to the total number can be used to derive p as the quadrant has a ratio of $\pi/4$. The more points, the more accurately the value of π can be estimated. Within 60 points the value of π is found, accurate to 2 d.p.[79]

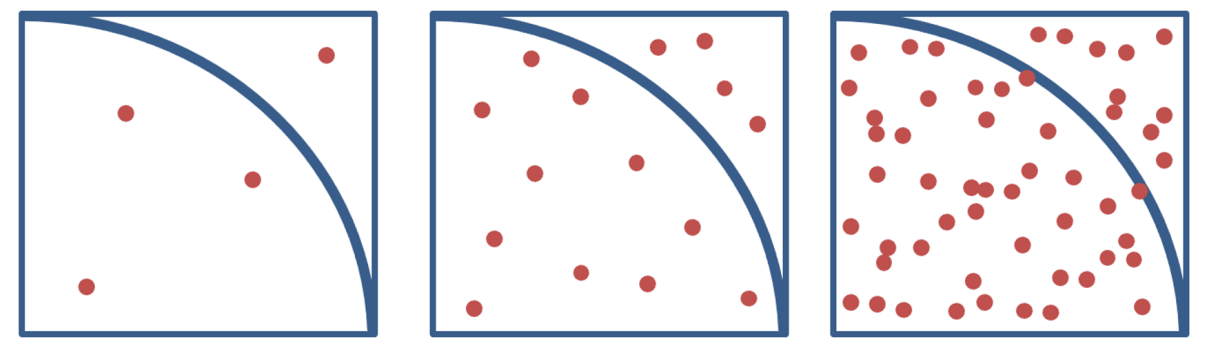


Figure 1.24. Random points are added to a square with unit side length to determine the value of π by the ratio of points total and within the circle quadrant. Left to right gives increasing accuracy with an increasing number of points.

1.6.4.5 Genetic algorithms

Genetic algorithms are a global optimisation method and are named as such due to the designed parallels to natural selection. Figure 1.25 shows a Genetic Algorithm (GA) general procedure which is performed by generating several random structures that are assessed in terms of their agreement with the observed data (*i.e.,* their so-called 'fitness'), and the best fitting solutions become 'parents'. The 'children' are new structures, resulting from a combination of features from the parents, for example, this could be the coordinates of one parent and the torsion angles of the other parent or some other arrangement. The children also have a possibility of mutation, meaning that some parameters are subject to random variation, in order to prevent the process from always terminating in the same local minima. The fitness of the children is assessed, and the 'successful' ones then go on to be parents of the next generation. These types of algorithms have been shown to be very successful when they have been fine tuned.

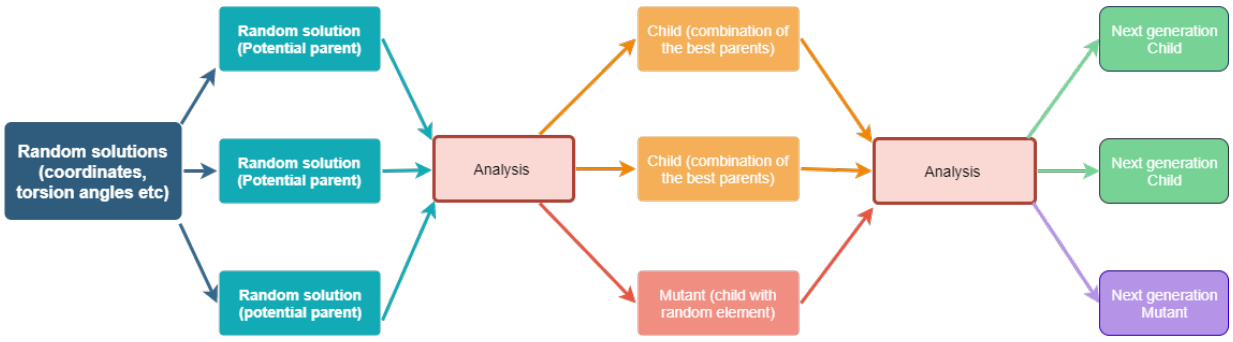


Figure 1.25. A simplified protocol for a typical genetic algorithm.

The first examples of genetic algorithms being used for molecules were shown by Kariuki *et al* (1997) and Shankland *et al* (1997), with the latter describing with the structure solution of the flexible molecule ibuprofen using an genetic algorithm approach, demonstrating the possibility of this approach in solving small organic crystal structures.[80, 81, 82]

The disadvantage of genetic algorithms is that there are no simple rules for setting the parameters, such as how many original random solutions are generated, the optimal crossover rate, and the optimal mutation rate *etc*. With so many parameters, the problem of fine tuning the best algorithm can become a bigger problem than the actual structure solution.[79, 82, 83]

1.6.4.6 DASH

DASH is a simulated annealing based global optimisation (GO) software package for structure solution from powder diffraction data. The main advantages of DASH are the speed of the methodology (working on extracted intensities from a Pawley fit of the observed data) and the

simplicity of the program to the user. The core concepts of the program were first published in 1998, and in 2001 DASH became part of the CSD software suite. Since 2022, DASH is an open-source program available for download from GitHub.

Since its release, DASH has been used to solve an increasing number of structures each year, ranging and increasing in molecular complexity, unit cell size, Z', and DoF. DASH has been widely used to solve the structures of many organic drug molecules, most recently the structure solution of carmustine, an anti-cancer drug, and riboflavin, vitamin B2 in 2021. DASH has also found application in other areas *e.g.*, coordination polymers research (with the study of one-dimensional spin chains of Cu^{II} and 2,5-Bis(pyrazol-1-yl)-1,4-dihydroxybenzene) and molecular magnets (with the study of -(2,3,4-trifluorophenyl)-1,2,3,5-dithiadiazolyl).[85, 86, 87,88]

One of the most complex structure to be solved from powder diffraction data is that of tenapanor hydrochloride (Figure 1.26), with 45 DoF, was solved using DASH, published 2018.

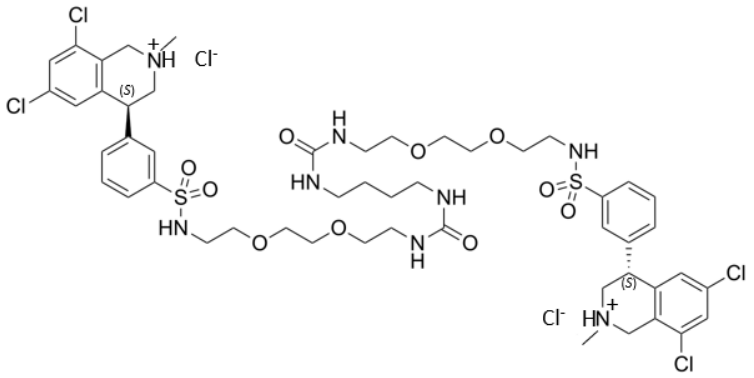


Figure 1.26. Tenapanor hydrochloride, dashed line highlighting pseudo centre of symmetry, 33 optimisable torsion angles, and 2 HCl, totalling 45 DoF.[84]

1.6.4.6.1 Using DASH

To operate DASH the user is guided through these steps with the help of the DASH wizard, shown below in Figure 1.27.

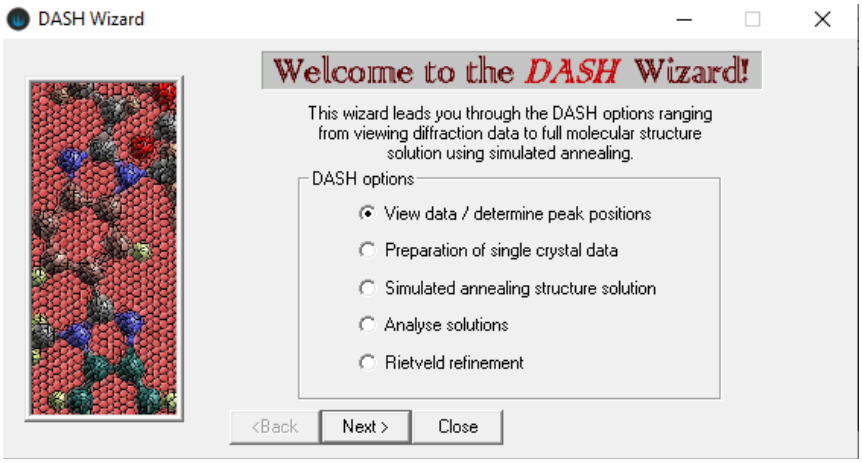


Figure 1.27. DASH wizard window.

The methodology for a user of DASH is described by the flow chart given in Figure 1.28, the main stages of which are (A) Data input, (B) indexing, (C) space group determination, (D) Pawley fitting, and (E) simulated annealing. Which are briefly described below:

A. **Data input:** A powder diffraction file is read as a series of data points (in the form of a Bruker .RAW or generic .XYE file) and the user then defines the 2θ range they wish to work with (typically 5-45° 2θ). The user then models the background removal carefully so as to not over or under subtract from the data.
B. **Indexing:** The first 20 peaks positions are selected by the user and from this, the pattern is indexed by DICVOL. Peak selection can be difficult when there is significant peak overlap or unknown impurities.
C. **Space group determination:** Once a unit cell is chosen (one that explains the observed peaks and makes chemical sense with respect to V_{cell} and V_{mol}) the user performs a Pawley fit (see step D) to the data using a space group that has no systematic absences (*e.g.,* P2 for monoclinic cells; P222 for orthorhombic cells) to confirm the cell. DASH then outputs a list of possible space groups ranked in order of probability.
D. **Pawley fitting:** This is the process of modelling the data as a function of background, peak position, peak shape and peak intensity. This procedure extracts reflection intensities whilst preserving the relationship between intensities of overlapping reflections. At the end of a successful Pawley fit in the determined space group, extracted intensities are written to file along with a covariance matrix that allows subsequent crystal structure models to be quickly evaluated against the extracted intensities. Properly carried out, this process is effectively equivalent to, but much faster to perform, a full-pattern based evaluation.

E. **Simulated annealing:** The SA runs are performed with a model of the model under investigation in ZMATRIX format and rely upon the extracted intensities as described in 'D' above. The simulated annealing is described in more detail in section 1.5.4.1.

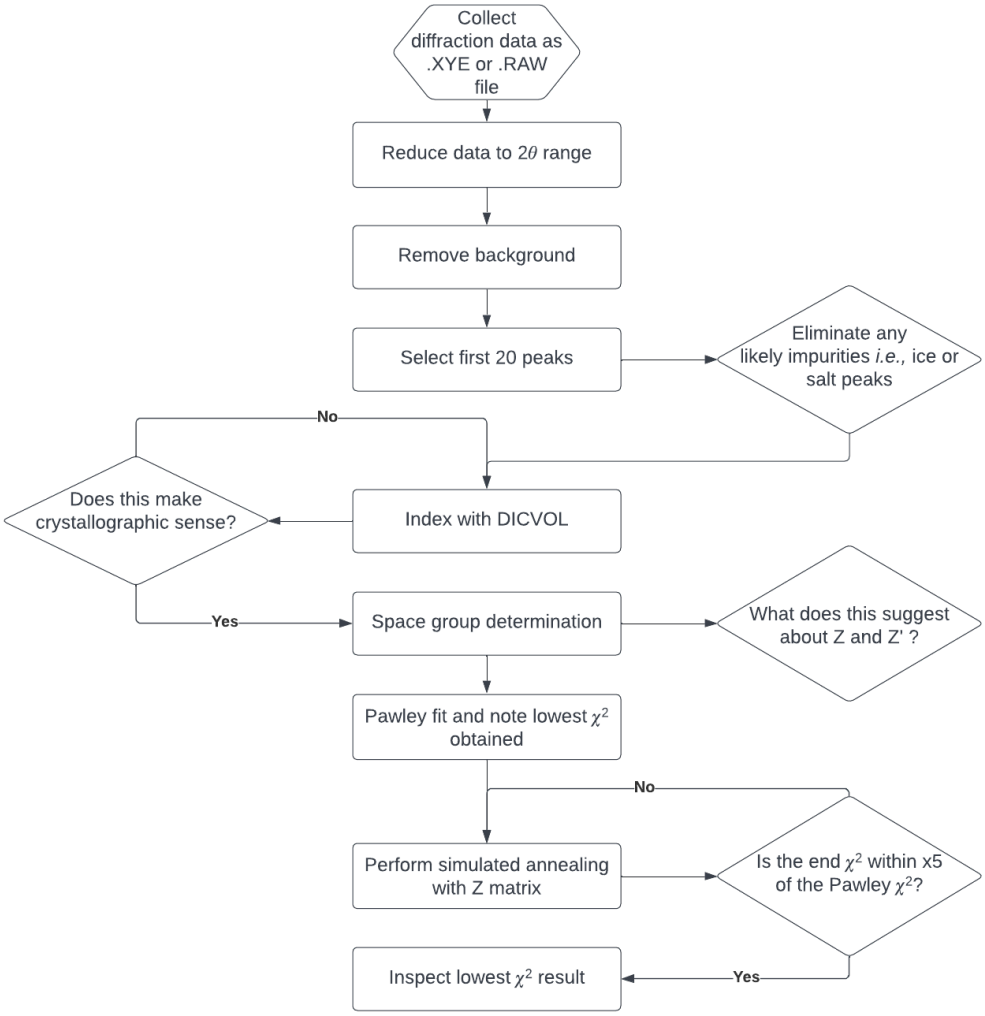


Figure 1.28. DASH protocol flow chart for structure solution from powder diffraction data.

1.6.4.6.2 Simulated annealing

The algorithm underlying DASH's structure solution is a global optimisation approach termed simulated annealing (SA). The concept stems from the practice of annealing metals where heating and controlled cooling leads to a low energy system in the latter, and so by analogy a good fit to observed diffraction data in the SDPD problem. Briefly, the system is heated such that any change of the variables (position, orientation, conformation) in the system is allowed, permitting the structure to span large areas of real space and escape local minima in the process.

The temperature is, of course, a notional concept and relates to the fact that there is a higher chance for the system to make uphill moves at a 'higher temperature'. As the temperature of

the system is reduced, the algorithm spends more time exploring lower-lying areas of the hypersurface, with the hope of locating the global minimum. SA performs a similar task to genetic algorithms but is considerably simpler to implement computationally and has fewer control parameters (start and end temperatures, cooling rate, and the number of steps taken before the cooling rate is applied), making it easier for general use. A single sufficiently long (potentially infinite) SA run is guaranteed to find the global minimum; a single normal finite SA run is not. The hope is that with multiple finite runs and enough time spent on each run, the global minimum will be found. This methodology has been found to be highly effective in DASH and has found widespread adoption with multiple other packages developed *e.g.*, PSSP, PowderSolve, FOX, *etc*. [26, 70, 72, 79]

Simulated annealing is utilised within DASH and the algorithm pseudo code is given below in Figure 1.29.

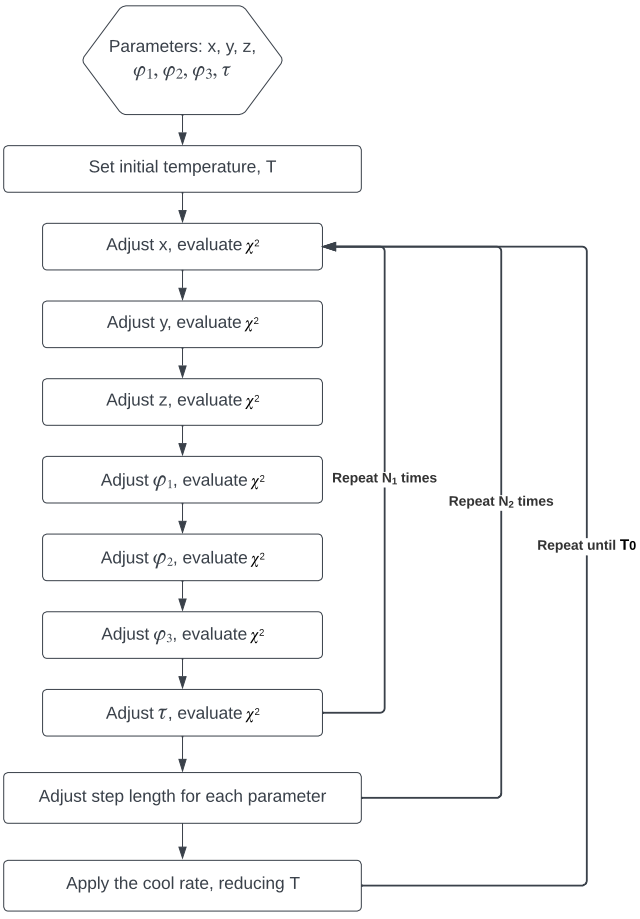


Figure 1.29. Simulated annealing algorithm pseudo code, iterating the model in three positional values (x, y, z,), three orientations (φ_1, φ_2, φ_3), and torsion angle (τ).

Simulated annealing has been shown to be improved with the use of torsion angle information, detailed below.

1.6.4.6.3 Model parameter ranges from MOGUL

The MOGUL database contains information from the Cambridge Structural Database (CSD) on bond lengths, angles and torsion angles (as shown in Figure 1.30 as a Newman projection). Bond angles are set to known values, but variable torsions are free to rotate as degrees of freedom. One use for this information is to understand preferred torsion angle distribution for the purposes of improving the structure solution within DASH. This application results in the explored structure solution space being reduced to the preferred space (*i.e.*, the most likely regions). This results in more time spent exploring 'good space' where good solutions are likely to lie *e.g.*, a secondary amide bond can be set to be planar with good confidence.

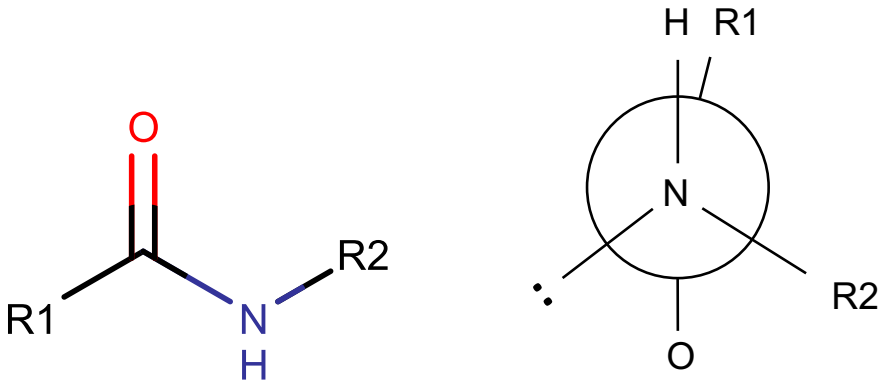


Figure 1.30. The torsion, or dihedral, angle for any given four atoms can be exploited to limit the search space, a secondary amide shown as a 2-dimensional structure (left) and a corresponding Newman projection (right).

It was shown by Kabova *et al* (2017). that adding torsion information into the structure solution dramatically increases the chances of solving a crystal structure. Evidence of this is given in Figure 1.31, the purple and blue (the best-case scenario) lines represent no conformational information with varying simulated annealing parameters, and the green and red lines represent the addition of MOGUL torsion constraints again with varying SA parameters. The number of degrees of freedom is still a limiting factor, but structures that previously had not been solved after weeks of calculations are now solvable within days. [70, 89, 90, 91]

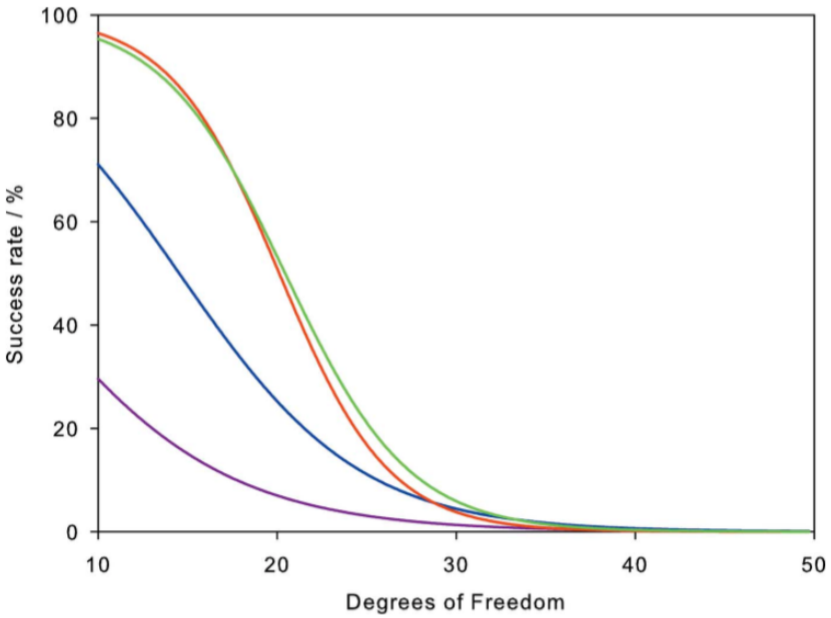


Figure 1.31. The impact of adding torsion restrains based on statistical data is given, by Kabova *et al*. The previous default DASH SA settings with no conformational information is given in purple, and the same settings with incorporation of MOGUL information is given in red. The use of optimised SA parameters with no conformational information is shown in blue and the addition of MOGUL with these improved settings in shown in green.[90]

1.6.4.6.4 Enhanced versions of DASH

A natural progression of simulated annealing is the structure solution of larger, more complex crystal structures, which requires increasingly larger amounts of computational power and time. One solution to this is to spread the computational load across cores within a computer. A crystal structure that requires 500 SA runs to reach a solution in DASH would potentially take several days to finish (where runs are performed consecutively) can be performed using MDASH (2008) in a fraction of the time. This is because the same 500 runs can be spread over N cores of a CPU, each performing 500/N runs on each core, achieving the same result ca. N-times faster. Published at the same time, GDASH allows the user to distribute the computational demands of DASH calculations over multiple computers on the same network.[92, 93]

In recent years there has been growing use of cloud computing as a way to perform very demanding calculations in a way that doesn't require a user's desktop power. Cloud computing is becoming more available year-on-year with companies such as Amazon and Google offering increasingly affordable cloud computing services. In 2015, CDASH, a cloud-enabled version of DASH was released, allowing crystallographers to take advantage of the low-cost use of cloud computing, enabling the structure solution of structures that were previously limited by computational effort.[94]

1.6.4.7 GALLOP

With advances in cloud computing, powder structure determination strategies, such as simulated annealing within DASH, can be enhanced, altered, and optimised. GALLOP (2021) uses a local optimisation and particle swarm approach, utilising a ‘multi-start’ approach. The method utilises cloud computing, which allows access to readily available graphics processing units (GPUs), which can be accessed via Google Colab. This approach has been shown to improve the speed and success of complex molecular crystal structure determination relative to DASH *e.g.,* by an order of magnitude in the cases of selexipag and chlorothiazide DMF_2 solvate (38 and 42 DoF respectively) [95]. GALLOP is open source and written in Python for other crystallographers to utilise and improve upon.

1.6 Recent developments in data collection using < 0.5 mg of material

In recent years there have been significant developments in the crystal structure analysis of small volumes. Below outlines a few examples of these developments.

1.6.1 The Gandolfi stage

In 2018, ThermoFisher developed an attachment for their powder diffractometer to enable the collection of accurate powder patterns from single crystals. The Gandolfi stage allows researchers to adapt what is at their disposal, increasing their analysis range. This 'make do and mend' approach is worthy of note, especially as they have shown that micron-sized crystals can produce high-quality powder patterns. Unfortunately, this is not a practical option for most researchers, as the apparatus requires the user to have the specific equipment before adding the additional Gandolfi stage.[96, 97]

1.6.2 In-house micro powder 3D orientation

In recent years, the Tsuboi group [98] developed a method to use a limited amount of powder (using ~5 μm crystals) to create a magnetically orientated microcrystal array (MOMA), in essence creating a pseudo-single crystal. So far, this technique has been shown to be effective for the analysis of inorganic, organic and protein crystal samples.

This method requires that that sample is biaxial (specifically, crystals whose magnetic susceptibility tensor χ has three different principal values) and the sample is prepared by suspending the powder in oil within a rotating capillary. The MOMA is then created by aligning the microcrystals with the aid of large magnets around a single crystal diffractometer. The user then collects a pseudo single crystal data set for structure solution. This is a highly innovative way to overcome the lack of 3D data found with powder patterns. However, this approach is unlikely to become commonplace in university or commercial lab spaces any time soon, because the initial report from this group used *ca.* 10 T magnets to generate the MOMAs. Much like synchrotron equipment, it is not routine for most research groups, but does have a place as a highly specialised technique, requiring a highly trained individual to run.

1.6.3 Crystal sponges

MOFs have recently been found to be useful for capturing structural information of organic compounds in a method referred to as a crystal sponge. By absorbing a compound and orientating each molecule into the pores a pseudo crystal is formed, making previously unstudied organic compounds observable by X-ray diffraction. This method requires very little material and does not require the compound to be crystalised, removing the need for multiple crystallisation attempts. However, this is not currently a widely used technique, due to pore sizes, limiting the size of the molecule that can fit into the crystal sponge.[99]

1.6.4 Structure Determination from Powder Diffraction using a Single Crystal Diffractometer: SDPD-SX

Another technique, which provided the stimulus for this Ph.D. was proposed in 2018 and published in 2022. The research conducted from this provided the background data and context to this thesis. SDPD-SX was pioneered as a collaboration between C4X Discovery, the University of Manchester, and the University of Reading, investigating the collection of high-quality powder diffraction data for SDPD on a single crystal diffractometer. Recently published, this pilot study describes the structural analysis of a range of polycrystalline materials using < 1 mg amounts.[100]

This study used a variety of known crystalline forms of APIs to test the accuracy of the structure solutions and vary the complexity of the structure. Whilst the instrument used for this study is "standard" (*i.e.*, purchasable off-the-shelf), it represents what is available at the high end of diffractometers: Rigaku FR-X, 2.97kW dual port rotating anode microfocus source providing Cu Kα wavelength radiation. For the 15 compounds in the study, complexity ranged from 9 - 31 DoF and produced excellent results. By way of example, a comparison of the known crystal structure Ritonavir (CSD refcode YIGPIO01, 29 DoF) with the structure solution from SDPD-SX is shown below in Figure 1.32 .

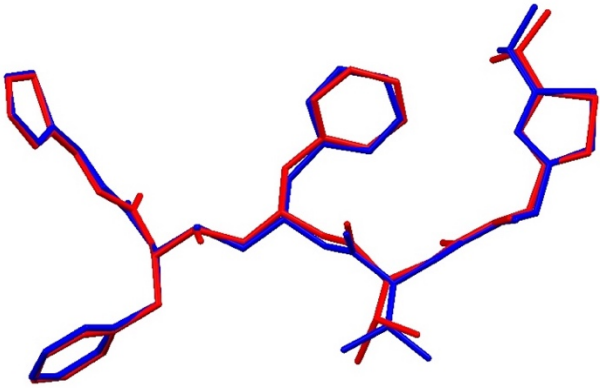


Figure 1.32. The asymmetric unit of ritonavir (form II), with the known SX structure (CSD refcode YIGPIO01) shown in blue and the structure as determined by SDPD-SX overlaid in red.

These data were collected without significant machine optimisation, and in principle, there is scope for further improvement by optimisation of the parameters which have the potential to affect data quality, including radiation source, beam divergence, detector distance, and mode of collection. To understand SDPD-SX fully, a series of experiments is required to optimise relevant machine parameters. Furthermore, this analysis should ideally be performed on a more typical lab-based setup to reflect what is generally accessible for the majority of researchers.

1.7 Aims

The primary aim of this research is to develop a methodology (so-called SDPD-SX) for the collection of high-quality PXRD data using a laboratory single-crystal diffractometer and < 0.5 mg of material. One application explored throughout is the potential to improve the availability of structural information of new compounds in early-stage drug discovery programs. The research sets out to explore the real-world possibilities for new crystal structures that cannot be accessed with exiting readily available techniques. Importantly, this thesis aims to define the capabilities and limitations of this methodology and highlight where there are areas for further investigation.

Objectives

The key objectives for this research are to:

1. Establish effective, efficient, and reproducible tools to aid optimal instrumental parameter determination for powder data collection using an in-house single crystal diffractometer (Chapter 2).
2. Using the tools created, to determine the optimal instrumental parameters for collecting powder data from small volumes of polycrystalline material using a single crystal diffractometer (Chapter 3).
3. To determine the optimal instrumental parameters of other in-house single crystal diffractometers, with different optimisable instrumental parameters (Chapter 4).
4. To establish the viability and limitations of powder data collected for structure solution purposes using a single-crystal diffractometer, and to compare the data quality with that of other in-house powder data collection methodologies (Chapter 5).
5. To attempt SDPD-SX on novel crystalline materials; in particular, materials that are not easily able to be studied by other available means (Chapter 6).

1.8 References

1 W. Clegg, *Oxford University Press Xray Crystallography*, 2013.

2 J. K. Cockcroft, High-resolution Space Groups and Tables, http://img.chem.ucl.ac.uk/sgp/large/sgp.htm

3 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2016, **72**, 171–179.

4 M. B. Hursthouse, L. S. Huth and T. L. Threlfall, *Org. Process Res. Dev.*, 2009, **13**, 1231–1240.

5 W. Wong-Ng, J. A. Kaduk, H. Wu and M. Suchomel, *Powder Diffr.*, 2012, **27**, 256–262.

6 L.-L. Huang, L. Yu, B. Li, B. Li, H. Wang and J. Li, *ACS Mater. Lett.*, 2022, **4**, 1053–1057.

7 R. E. Dinnebier, A. Fischer, G. Eggert, T. Runčevski and N. Wahlberg, *J. Vis. Exp.*, , DOI:10.3791/54109.

8 S. D. Gumbert, M. Körbitzer, E. Alig, M. U. Schmidt, M. R. Chierotti, R. Gobetto, X. Li and J. van de Streek, *Dye. Pigment.*, 2016, **131**, 364–372.

9 T. Koyano, I. Hachiya and K. Sato, *Food Struct.*

10 D. Nicholls, C. Elleman, N. Shankland and K. Shankland, *Acta Crystallogr. Sect. C Struct. Chem.*, 2019, **75**, 904–909.

11 D. Nicholls, K. Shankland, M. Spillman and C. Elleman, *Food Anal. Methods*, 2018, **11**, 2673–2681.

12 C. . Frampton, K. . Knight, N. Shankland and K. Shankland, *J. Mol. Struct.*, 2000, **520**, 29–32.

13 N. Shankland, W. I. F. David, K. Shankland, A. R. Kennedy, C. S. Frampton and A. J. Florence, *Chem. Commun.*, , DOI:10.1039/b107075d.

14 T. Runčevski, R. E. Dinnebier and D. Freyer, *Zeitschrift für Anorg. und Allg. Chemie*, 2014, **640**, 100–105.

15 K. T. Savjani, A. K. Gajjar and J. K. Savjani, *ISRN Pharm.*, 2012, **2012**, 1–10.

16 G. L. Patrick, *An introduction medicinal chemictry*, 3rd edn., 2005.

17 H. M. Berman, *Nucleic Acids Res.*, 2000, **28**, 235–242.

18 G. R. Desiraju, *Crystal engineering: The design of organic solids*, Elsevier, Amsterdam, 1989.

19 H. G. Brittain, *Polymorphism in Pharmaceutical Solids: Second edition*, 2016.

20 P. D. Beer, P. A. Gale and D. K. Smith, *Supramolecular Chemistry*, Oxford university Press, 1999.

21 J. W. Steed, *Trends Pharmacol. Sci.*, 2013, **34**, 185–193.
22 D. D. Ho, A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard and M. Markowitz, *Nature*, 1995, **373**, 123–126.
23 S. L. Morissette, S. Soukasene, D. Levinson, M. J. Cima and Ö. Almarsson, *Proc. Natl. Acad. Sci.*, 2003, **100**, 2180–2184.
24 N. E. Chayen, *Prog. Biophys. Mol. Biol.*, 2005, **88**, 329–337.
25 E. H. Lee, *Asian J. Pharm. Sci.*, 2014, **9**, 163–175.
26 T. Threlfall, *Org. Process Res. Dev.*, 2003, **7**, 1017–1027.
27 G. R. Desiraju, *CrystEngComm*, 2007, **9**, 91–92.
28 K. M. Anderson and J. W. Steed, *CrystEngComm*, 2007, **9**, 328–330.
29 P. Vishweshwar, J. A. McMahon, J. A. Bis and M. J. Zaworotko, *J. Pharm. Sci.*, 2006, **95**, 499–516.
30 S. Aitipamula, R. Banerjee, A. K. Bansal and E. Al, *Cryst. Growth Des.*, 2012, **12**, 2147–2152.
31 D. Soergel, *Cardiovascular and Renal Drugs Advisory Committee Meeting*, 2015.
32 S. A. Ross, D. A. Lamprou and D. Douroumis, *Chem. Commun.*, 2016, **52**, 8772–8786.
33 A. D. Bond, *CrystEngComm*, 2007, **9**, 833.
34 R. Davey and J. Garside, in *From molecules to crystallizers*, 2000, pp. 1–52.
35 U. Garg, Y. Azim, A. Kar and C. P. Pradeep, *CrystEngComm*, 2020, **22**, 2978–2989.
36 Y. Yano, T. Ono, T. Ohhara and Y. Hisaeda, *Chem. – A Eur. J.*, 2021, **27**, 17802–17807.
37 H. Dodziuk, in *Cyclodextrins and Their Complexes*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, 2006, pp. 1–30.
38 J. Szejtli, *Cyclodextrin Technology*, Springer Netherlands, Dordrecht, 1988, vol. 1.
39 M. R. Caira, *Rev. Roum. Chim.*, 2001, **4**, 371–386.
40 J. Frank, J. F. Holzwarth and W. Saenger, *Langmuir*, 2002, **18**, 5974–5976.
41 W. Clegg, A. J. Blake, J. M. Cole, J. S. O. Evans, P. Main, S. Parsons and D. J. Watkin, *Crystal Structure Analysis*, Oxford University Press, 2009, vol. 9780199219.
42 W. I. F. David, K. Shankland, L. B. McCusker and C. Baerlocher, *Structure Determination from Powder Diffraction Data*, Oxford University Press, 2006.
43 D. J. M. Irving, D. A. Keen and M. E. Light, *Rev. Sci. Instrum.*, 2021, **92**, 043107.
44 M. Otendal, T. Tuohimaa, U. Vogt and H. M. Hertz, *Rev. Sci. Instrum.*, 2008, **79**, 016102.
45 W. Clegg, *J. Chem. Soc. Dalt. Trans.*, 2000, 3223–3232.
46 P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J.

Frisch, S. Gilevich, J. Hastings, G. Hays, P. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H.-D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky and J. Galayda, *Nat. Photonics*, 2010, **4**, 641–647.

47 S. Boutet, L. Lomb, G. J. Williams, T. R. M. Barends, A. Aquila, R. B. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman, M. Messerschmidt, A. Barty, T. A. White, S. Kassemeyer, R. A. Kirian, M. M. Seibert, P. A. Montanez, C. Kenney, R. Herbst, P. Hart, J. Pines, G. Haller, S. M. Gruner, H. T. Philipp, M. W. Tate, M. Hromalik, L. J. Koerner, N. van Bakel, J. Morse, W. Ghonsalves, D. Arnlund, M. J. Bogan, C. Caleman, R. Fromme, C. Y. Hampton, M. S. Hunter, L. C. Johansson, G. Katona, C. Kupitz, M. Liang, A. V. Martin, K. Nass, L. Redecke, F. Stellato, N. Timneanu, D. Wang, N. A. Zatsepin, D. Schafer, J. Defever, R. Neutze, P. Fromme, J. C. H. Spence, H. N. Chapman and I. Schlichting, *Science (80-.).*, 2012, **337**, 362–364.

48 K. Asakura, K. J. Gaffney, C. Milne and M. Yabashi, *Phys. Chem. Chem. Phys.*, 2020, **22**, 2612–2614.

49 C. Hammond, *Meas. Sci. Technol.*, 2002, **13**, 232–232.

50 U. Kolb, K. Shankland, L. Meshi, A. Avilov and W. David, *Uniting Electron Crystallography and Powder Diffraction*, Springer Netherlands, Dordrecht, 2012.

51 P. Cui, E. Svensson Grape, P. R. Spackman, Y. Wu, R. Clowes, G. M. Day, A. K. Inge, M. A. Little and A. I. Cooper, *J. Am. Chem. Soc.*, 2020, **142**, 12743–12750.

52 M. S'ari, J. Cattle, N. Hondow, H. Blade, S. Cosgrove, R. M. Brydson and A. P. Brown, *J. Phys. Conf. Ser.*, 2015, **644**, 012038.

53 A. L. Patterson, 1934, **537**, 372–376.

54 H. Hauptman and J. Karle, *Phys. Rev.*, 1950, **80**, 244–248.

55 G. M. Sheldrick, *Acta Crystallogr. Sect. A Found. Adv.*, 2015, **71**, 3–8.

56 L. Palatinus and G. Chapuis, *J. Appl. Crystallogr.*, 2007, **40**, 786–790.

57 O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl. Crystallogr.*, 2009, **42**, 339–341.

58 A. Altomare, F. Ciriaco, C. Cuocci, A. Falcicchio and F. Fanelli, *Powder Diffr.*, 2017, **32**, S123–S128.

59 A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi, *J. Appl. Crystallogr.*, 2012, **45**, 789–797.

60 A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio, *J. Appl. Crystallogr.*, 2013, **46**, 1231–1235.

61 J. Rius, *Powder Diffr.*, 1999, **14**, 267–273.

62 J. Rodríguez-Carvajal, *Lab. Léon Brillouin (CEA-CNRS), CEA/Saclay, 91191 Gif sur*

Yvette Cedex,FRANCE.

63 K. Burger, D. Cox, R. Papoular and W. Prandl, *J. Appl. Crystallogr.*, 1998, **31**, 789–797.

64 S. Basso, C. Besnard, J. P. Wright, I. Margiolaki, A. Fitch, P. Pattison and M. Schiltz, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 756–761.

65 C. J. Gilmore, K. Henderson and G. Bricogne, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 1991, **47**, 830–841.

66 E. Nishibori, T. Ogura, S. Aoyagi and M. Sakata, *J. Appl. Crystallogr.*, 2008, **41**, 292–301.

67 D. Šišak, C. Baerlocher, L. B. McCusker and C. J. Gilmore, *J. Appl. Crystallogr.*, 2012, **45**, 1125–1135.

68 L. B. McCusker, C. Baerlocher, R. Grosse-Kunstleve, S. Brenner and T. Wessels, *Chimia (Aarau).*

69 A. Vagin and A. Teplyakov, *J. Appl. Crystallogr.*, 1997, **30**, 1022–1025.

70 W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell and J. C. Cole, *J. Appl. Crystallogr.*, 2006, **39**, 910–915.

71 A. J. Markvardsen, W. I. F. David and K. Shankland, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 2002, **58**, 316–326.

72 V. Favre-Nicolin and R. Černý, *Zeitschrift für Krist. - Cryst. Mater.*, 2004, **219**, 847–856.

73 O. Vallcorba, J. Rius, C. Frontera and C. Miravitlles, *J. Appl. Crystallogr.*, 2012, **45**, 1270–1277.

74 K. Shankland, A. J. Markvardsen, C. Rowlatt, N. Shankland and W. I. F. David, *J. Appl. Crystallogr.*, 2010, **43**, 401–406.

75 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2016, **72**, 171–179.

76 A. Volford, *Chem Axon MarvinSketch*.

77 *Chem. Eng. News Arch.*, 2000, **78**, obc.

78 C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler and J. van de Streek, *J. Appl. Crystallogr.*, 2006, **39**, 453–457.

79 S. S. Skiena, *The Algorithm Design Manual*, Springer London, London, 2008.

80 K. D. M. Harris and M. Tremayne, *Chem. Mater.*, 1996, **8**, 2554–2570.

81 K. Shankland, W. I. F. David and T. Csoka, *Zeitschrift für Krist. - Cryst. Mater.*, 1997, **212**, 550–552.

82 B. M. Kariuki, H. Serrano-González, R. L. Johnston and K. D. M. Harris, *Chem. Phys. Lett.*, 1997, **280**, 189–195.

83 W. Paszkowicz, *Mater. Manuf. Process.*, 2009, **24**, 174–197.

84 S. O. Nilsson Lill, C. M. Widdifield, A. Pettersen, A. Svensk Ankarberg, M. Lindkvist, P. Aldred, S. Gracin, N. Shankland, K. Shankland, S. Schantz and L. Emsley, *Mol. Pharm.*, 2018, **15**, 1476–1487.

85 C. Schlesinger, E. Alig and M. U. Schmidt, *Powder Diffr.*, 2021, **36**, 148–150.

86 M. Guerain, F. Affouard, C. Henaff, C. Dejoie, F. Danède, J. Siepmann, F. Siepmann and J.-F. Willart, *Acta Crystallogr. Sect. C Struct. Chem.*, 2021, **77**, 800–806.

87 R. Dinnebier, H.-W. Lerner, L. Ding, K. Shankland, W. I. F. David, P. W. Stephens and M. Wagner, *Zeitschrift für Anorg. und Allg. Chemie*, 2002, **628**, 310–314.

88 A. M. T. Bell, J. Nicholas B. Smith, J. Paul Attfield, J. M. Rawson, K. Shankland and W. I. F. David, *New J. Chem.*, 1999, **23**, 565–567.

89 I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris and A. G. Orpen, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 2133–2144.

90 E. A. Kabova, J. C. Cole, O. Korb, A. C. Williams and K. Shankland, *J. Appl. Crystallogr.*, 2017, **50**, 1421–1427.

91 E. A. Kabova, C. D. Blundell and K. Shankland, *J. Pharm. Sci.*, 2018, **107**, 2042–2047.

92 T. A. N. Griffin, K. Shankland, J. van de Streek and J. Cole, *J. Appl. Crystallogr.*, 2009, **42**, 360–361.

93 T. A. N. Griffin, K. Shankland, J. van de Streek and J. Cole, *J. Appl. Crystallogr.*, 2009, **42**, 356–359.

94 M. J. Spillman, K. Shankland, A. C. Williams and J. C. Cole, *J. Appl. Crystallogr.*, 2015, **48**, 2033–2039.

95 M. J. Spillman and K. Shankland, *CrystEngComm*, 2021, **23**, 6481–6485.

96 G. Schmidt, *The Gandolfi Stage: a novel approach for the analysis of single crystals and small volume samples with ARL EQUINOX Series*, 2018.

97 G. Gandolfi, *Miner. Petrogr. Acta*, 1967, **13**, 67–74.

98 C. Tsuboi, K. Aburaya, F. Kimura, M. Maeyama and T. Kimura, *CrystEngComm*, 2016, **18**, 2404–2407.

99 M. Hoshino, A. Khutia, H. Xing, Y. Inokuma and M. Fujita, *IUCrJ*, 2016, **3**, 139–151.

100 E. A. Kabova, C. D. Blundell, C. A. Muryn, G. F. S. Whitehead, I. J. Vitorica-Yrezabal, M. J. Ross and K. Shankland, *CrystEngComm*, 2022, **24**, 4337–4340.

Chapter 2

Creation and validation of Automated Powder Pattern Evaluation: APPE

2.1 Introduction

Before attempting to optimise the physical parameters of a single crystal diffractometer, for the purpose of powder diffraction data collection, it is important to create the tools necessary for this task. At present, the inbuilt software for Rigaku single crystal diffractometers (used in this thesis), CrysAlisPro, does not have a function for assessing pattern quality.

The focus of this chapter is the creation and validation of a program to partially automate the assessment of powder pattern quality. For this pattern assessment, the full width at half max (FWHM) is the parameter used to assess the peak quality. Analysis of the quality of the powder patterns can be performed 'manually', with powder diffraction programs such as DASH, but due to the quantity of data that was predicted to be created throughout this thesis, a program to reduce the time taken by this process was essential.

In this chapter, the term 'resolution' relates to instrumental resolution, *i.e.*, how the instrument geometry and distribution of wavelengths contribute to the observed peak widths. For samples whose diffraction is 'infinitely' sharp, all increases in peak width are due to the instrument setup. With maximum instrumental resolution, there is minimal reflection overlap, producing the highest quality diffraction pattern for a material under study.[1]

2.1.1 Automation of data analysis

As described above, due to the large number of data files produced in this research, it is advantageous to have a method whereby each data file can be processed quickly, efficiently, accurately, and consistently with minimal user intervention or subjective judgement.

Automation of data processing has several advantages, such as increasing throughput and eliminating 'human errors' due to inconsistent user interaction. Systematic errors are still an issue, but with careful program design, these errors can be caught when compared to a reference. Including automation within an analysis helps to gather and reduce data to the essential information, creating an abstraction and reducing the complexity of ideas and relationships.

Python is a user-friendly programming language that is easily accessible to individuals without a strong background in programming. It is widely used for scientific data analysis, mainly because of the various data science packages that are freely available and are well maintained by the Python community. Examples of open-source packages include the libraries SciPy (which includes, amongst others, modules for optimization, linear algebra, integration,

interpolation, special functions, signal and image processing) and NumPy (which adds support for large multidimensional arrays and matrices, and associated mathematical operations on these arrays). Furthermore, there is a growing global community of Python users that provides education and support in using the language within various settings. For use in diffraction, Boulle and Kieffer have noted that "in particular it is the various Python libraries that can enhance all aspects of crystallography".[2, 3, 4]

In this chapter, the creation of the Python program APPE (Automatic Powder Pattern Evaluation) is reported. APPE was created using object orientated programming (OOP) techniques. This widely-used approach relies upon writing code that adheres to the four principles of object orientation: encapsulation, inheritance, polymorphism, and abstraction. These principles are implemented via the use of classes which are the blueprints for objects. Objects are an example of persistence within a computer program, as they reside upon the 'heap' within the computer's memory. Each object can be given individual attributes and functions with the former relating to storage of data and the latter relating to specific actions an object can perform.[5]

This approach to programming was chosen for three reasons. Firstly, it allows the reuse of functions and variables in line with the paradigm of 'DRY' (Do not Repeat Yourself) *i.e.*, the principle of not repeating code within different sections of the program. Secondly, to allow for rapid prototyping, and thirdly to allow for modularisation.

APPE leverages the power of existing tried-and-tested algorithms (within Python libraries) to ultimately semi-automate the quality analysis of the powder patterns collected. This is achieved though the return of FWHM values for the diffraction peaks by APPE and subsequent human evaluation of the lowest value and trends. Although all the peaks within a powder pattern are assessed and reported in the APPE output, a simplified approach is used, where the analysis of the FWHM average is only performed on the first three peaks of a pattern. This is with the aim of avoiding the inclusion of any peak overlap and creating a convention for any material to be used in this analysis.

2.2 Aims

The overarching aim of this work is to be able to perform structure determination from powder data collected using a single-crystal diffractometer. To achieve this aim, the optimal instrumental parameters of the single-crystal diffractometer have to be determined; this results in a large volume of data which is both slow to process and processes that are prone to human error. The aim of this chapter is to create a tool (referred to throughout as APPE, Automatic Powder Pattern Evaluation) to aid efficient and accurate optimal instrumental parameter determination. Specifically, a tool that is able to read multiple powder data (.XYE) files, determine the breadth (FWHM) of chosen peaks and report the outcome in an easily accessible format *i.e* Excel. To validate the APPE program, the output of generated and experimental data needs to be compared to that of existing programs such as DASH, to ensure the production of accurate results.

2.3 Methodology

2.3.1 Software

Table 2.1 lists the software used in the creation, development, and testing of APPE.

Table 2.1. Software details

Software	Version	Use / description
Spyder [5]	4.1.5	Integrated development environment (IDE), used for advanced editing, interactive testing, and debugging of program creation.
Anaconda [6]	2.2.1	Hosts and manages Conda applications in a protected environment containing libraries for data science.
Topas[7]	3.8.1	PXRD analysis program used in this chapter for the creation of calculated powder patterns.
DASH[8]	3.4.9	PXRD crystal structure solution program, used in this chapter for peak width analysis,
SciPy and NumPy[4,3]	1.9.0, 1.21.0	Libraries, Scientific Python and Numerical Python, containing functions and various operations for data analysis.
CrysAlisPro[9]	1.171.41.93a	Software used on Rigaku single crystal diffractometers. Used for data collection of powder diffraction and exporting data to .XYE file format.

2.3.2 Program flow and operation

For each of the experimental datasets, the background was removed manually within DASH before all datasets were assessed by APPE.

2.3.2.1 APPE flow design

Firstly, APPE determines the location of the data file; in this chapter this is done by placing the files in a folder named ‘test-data’. The program then processes this folder and once each

dataset (.XYE file) is found; the program identifies the peaks present, determining the peak position and the FWHM (Full Width at Half Maximum).

The FWHM is used as a measure of data quality: a higher resolution pattern will mean lower FWHM values. The better resolution will allow peaks with similar 2θ values to be distinguished and so improving indexing of the diffraction pattern and improving the overall structure solution. This information is then stored within a bespoke data structure. Figure 2.1 shows a flow chart of the operation of APPE.

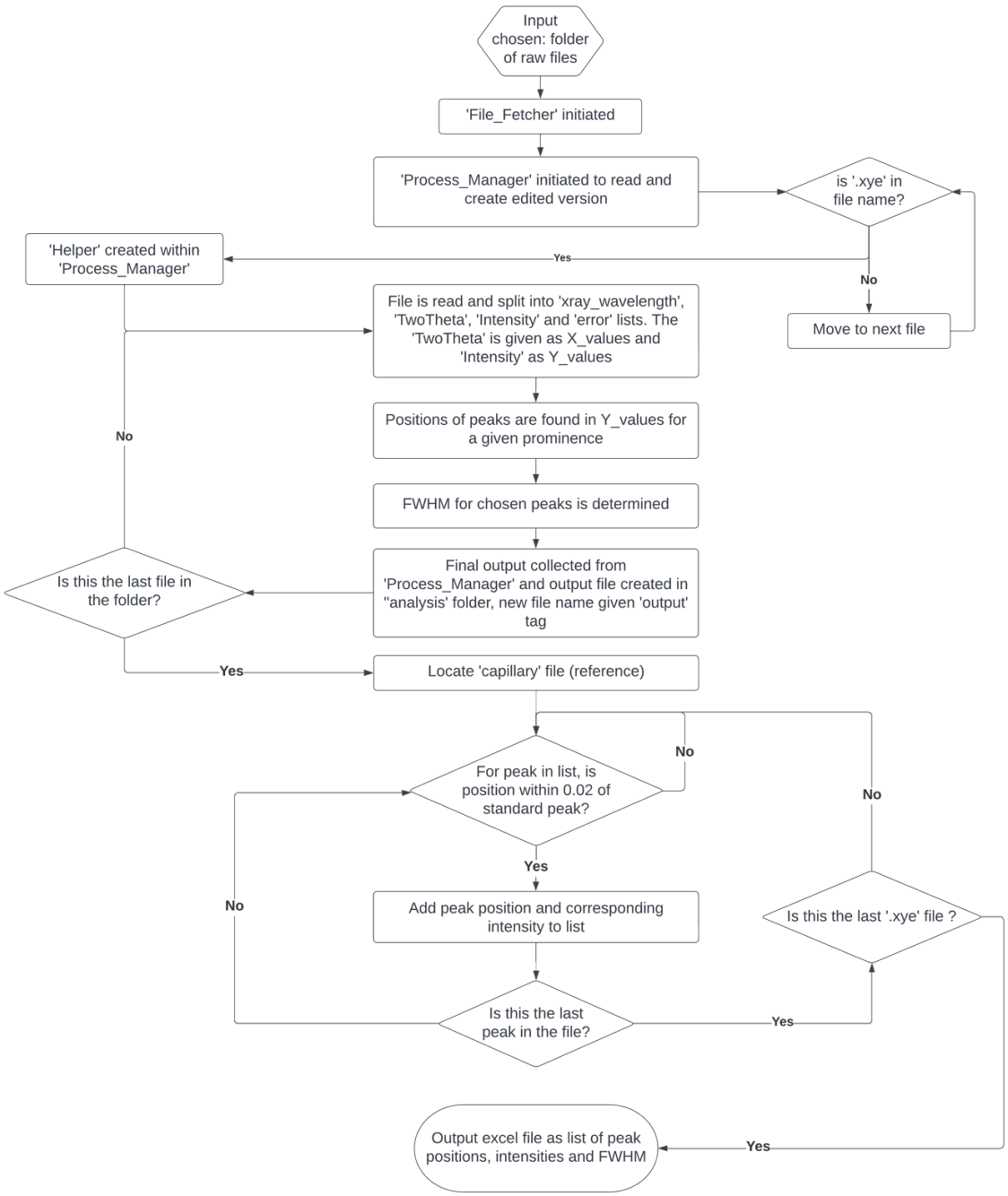


Figure 2.1. A flowchart showing the processes and actions implemented within APPE.

2.3.2.2 FWHM evaluation

The FWHM of peaks is affected by the physical parameters of the machine setup (such as beam divergence, detector distance and exposure time). Therefore, determining the differences found between datasets of the same sample, but with different setups, should lead to an optimal parameter set.

Inevitably, there is a need for a programme which is able to determine peak positions and associated FWHMs rapidly, in order to evaluate multiple data files efficiently, facilitating the parameter optimisation of the single-crystal diffractometer within a reasonable time scale.

To avoid the analysis of false peaks within a dataset, a reference is used with known peak positions to filter the peak list produced for each dataset. For this chapter, a calculated powder pattern of an in-house 'chosen standard material', L-glutamic acid, was used as the reference.

2.3.2.3 Result collation and output

Once the peak analysis and comparison as outlined above have been carried out, a separate spreadsheet is created with the averaged FWHM of specific peaks (chosen from the reference by the user as being low angle and a single reflection). Below this, the filenames with corresponding peak widths are ordered smallest to largest, allowing the user to quickly identify the range of values as well as the most resolved dataset.[4, 5]

2.3.3 Validation methodology

2.3.3.1 Calculated powder patterns

A series of calculated powder diffraction patterns of L-glutamic acid, with well-characterised FWHM were created using TOPAS, in order to measure the accuracy of APPE's output. The calculated peak width was controlled with the use of TOPAS's 'crystallite size' option; by increasing the crystallite size, the FWHM drops. Having such reference data available makes identifying errors in the program simple, the peak widths being determined are known and therefore any differences are due to program operation.[7]

As a control, the generated patterns were also analysed in DASH, to compare to APPE's FWHM values. Results from this comparison and the comparisons between APPE and TOPAS's generated data are presented in Table 2.2.

2.3.3.2 Experimentally collected powder patterns

For analysis of experimentally collected data, a series of diffraction patterns was collected using a sample of L-glutamic acid at varying detector distances. Collections were performed with a Rigaku Synergy-S single crystal diffractometer (HyPix-6000HE, Cu Kα). Standard phi scan data collection mode and a beam divergence of 9.5 mrad (*i.e.*, maximum incident X-ray intensity) were used. Background subtraction was performed using DASH and the peak widths subsequently analysed using the two separate methodologies. The average of the first three peak widths FWHM, as a function of detector distance is plotted in Figure 2.2

2.4 Results

The analysis of the synthetic PXRD data, generated using TOPAS, is shown below in Table 2.2.

Table 2.2. FWHM of calculated patterns, assessed using the Python program, APPE, and DASH.

		FWHM / °	
CS_G (crystallite size gaussian) crystallite size (TOPAS) /μm	Stated‡	APPE*	DASH*
150	0.07	0.06	0.06
120	0.08	0.08	0.08
100	0.09	0.09	0.09
80	0.11	0.11	0.11
70	0.13	0.13	0.13
60	0.15	0.15	0.15
50	0.18	0.18	0.18
45	0.20	0.20	0.20
40	0.22	0.22	0.22
35	0.25	0.26	0.25
32.5	0.27	0.27	0.27
30	0.30	0.30	0.30

‡ "Shankland, K: personal communication"

* The values reported are the average of the first 3 peaks in the powder diffraction patterns.

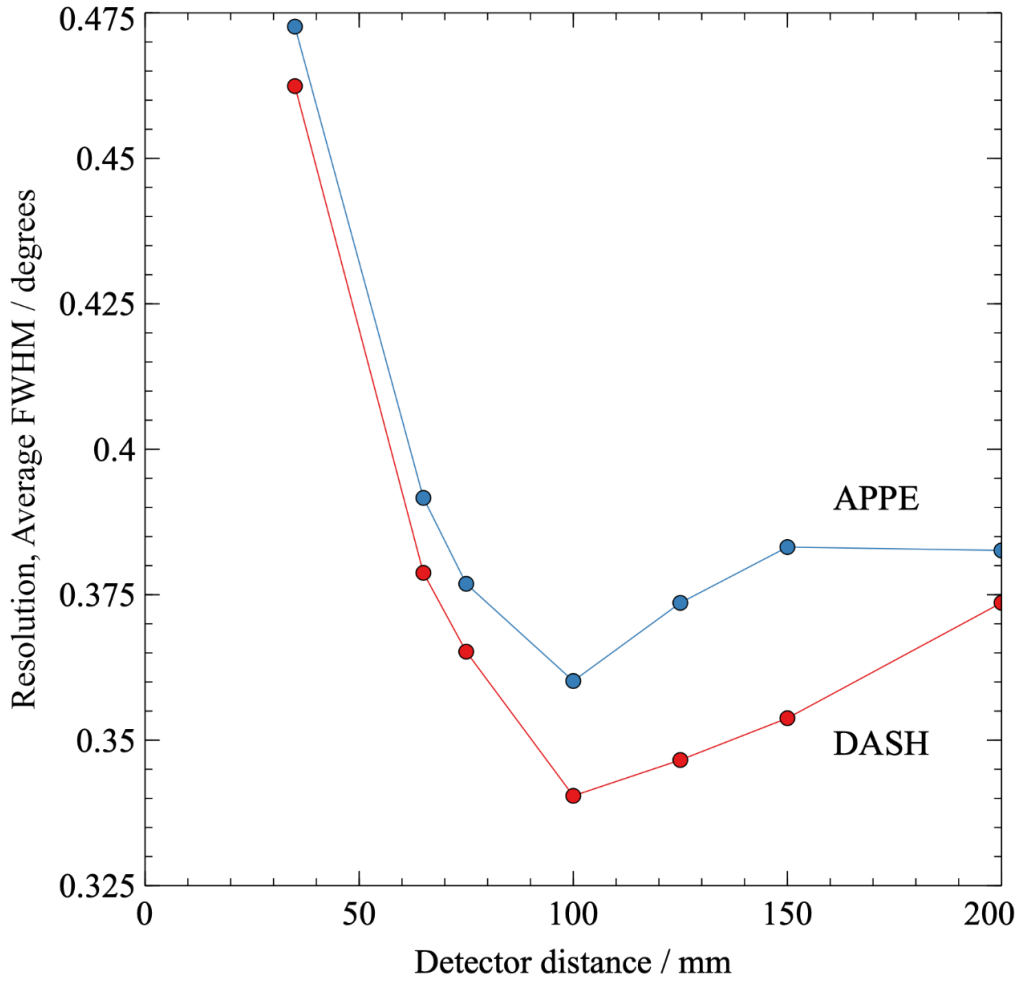


Figure 2.2. The variation in FWHM (the average of the first 3 peaks of L-glutamic acid) as a function of detector distance, measured by DASH (red) and APPE (blue).

2.5 Discussion

2.5.1 Creation and development of APPE

When working with Python programs, it is advisable to use a dedicated Integrated Development Environment (IDE) in which to write and test code. Furthermore, many IDEs provide a collection of additional features to aid the identification of coding errors and debugging. For the creation of APPE, the Spyder environment (part of Anaconda) was chosen as it is easily accessible and easy to use.[10]

Due to the varying quality of experimental data, the peak picking software can falsely identify peaks; this is an issue that can be partially removed by altering the sensitivity of the peak picking function. When analysing multiple datasets, altering the peak picking sensitivity is not a viable solution as the sensitivity needed for each dataset is different. The solution within APPE was to use a reference with sharp, well-defined peaks; for example from simulated powder patterns. These chosen powder pattern references are also relatively high symmetry and so little chance of accidental overlap at lower angles. The program then filters out false peaks by comparing peak positions to the reference and selecting those within a small error range of the reference peak list. The result is an output peak list with associated FWHM that is simple to interpret, given the number of peaks is equal to the reference.

It is important to note that the background removal is not a process that can easily be automated for this research. Removing a background signal manually is recommended since it is specific to each data set and any over-subtraction of the background could significantly affect the end results.

2.5.2 Validation of APPE

2.5.2.1 Calculated pattern validation

The validation of APPE using calculated powder patterns is shown in Table 2.2, and highlights that APPE was able to match the calculated FWHM to 2 decimal places for 10 of the 12 datasets. The DASH analysis of the same datasets shows a similar level of agreement with a single data point differing from the simulated data. For both methodologies, the values that did not match are all within 0.01° of the stated FWHM (in the case of the 0.06° stated value, outputted 0.07°). This therefore shows APPE is comparable to DASH, in the context of peak analysis within this methodology.

The use of an independent program to validate the results obtained by APPE is an important step in investigating its performance, particularly as it uses a fairly simple method for

estimating FWHM (see below). DASH was chosen as a control as it has a well-established methodology for powder pattern fitting using a full-Voigt peak shape function.

2.5.2.2 Experimental validation

Whilst generated patterns are good for validation against idealised scenarios, when working with real world data the references have to be experimentally collected. Different functions measure the FWHM of experimental peaks differently, so it is crucial for APPE to be tested using experimentally collected data against an established method, in this case, DASH. To examine if APPE can be used with experimental data, the detector distance was varied with all other parameters remaining constant.

Analysis of a range of data points collected at varying detector distances were performed to establish the detector distance that produces the minimum FWHM. The output of APPE and DASH vary from each other, producing different FHWM values varying by 2.2 – 7.6 %, as shown in Figure 2.2. Given the different peak fitting methods employed, this is not surprising, but importantly the trend in resolution variation as a function of distance is the same for both programs. Both DASH and APPE found the minimum FWHM to be 100 mm. It is important to note however that the optimal detector distance found in this chapter does not reflect the actual optimal detector distance with respect to other optimised variables. These have deliberately not been explored here to simplify the validation of APPE. Therefore, the detector distance of 100 mm represents a local minimum in the optimal parameter landscape.

2.5.3 Analysis of peak widths within DASH vs APPE

The primary function of DASH is structure solution, not individual peak analysis. DASH has an auto peak picking function as part of its interface however it was found that when extracting the FWHM values of the chosen peaks, the majority are not reported in the structural information window. When multiple overlapping peaks are selected at once, only the largest peak is reported in the list of FWHM.

Analysis of peak FWHM in DASH was therefore performed by individually selecting peaks to avoid errors. Unfortunately, this process is time-consuming and this methodology is prone to inconsistency, especially with the high number of datasets required for a systematic analysis.[8]

Two significant advantages of using APPE are the automatic peak picking ability and the speed of processing. Where DASH requires each peak to be selected individually, APPE is

able to significantly reduce the time needed to process data before analysis. This allows researchers to interpret the output spreadsheet to make experimental decisions in real-time.

2.5.3.1 Peak width analysis

Figure 2.2 highlights that although DASH and APPE arrive at the same conclusion for the ideal detector distance, the individual curves do not coincide due to the differences in the peak shape functions used within the two programs. DASH fits the selected peaks with a full Voigt function (a convolution of a Lorentzian and Gaussian function) and from this derives the FWHM value. The 'peak_width' function in the SciPy package, measures the FWHM differently, taking a much more simplistic approach. The function defines a prominence, identifying peak maximum height, identifying, and measuring a base line, and from this information the peak width is measured. Figure 2.3 (taken from the SciPy online manual) shows the use of the SciPy functions with a random dataset (blue) to identify peaks (orange crosses) in the data, the corresponding peak bases (red), and widths (green).

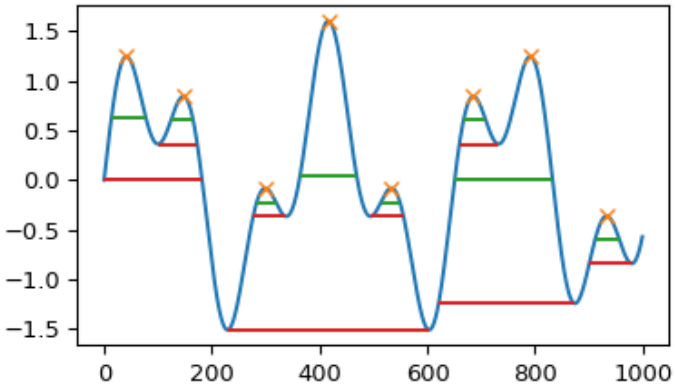


Figure 2.3. Peak picking, and peak width determination within SciPy. Peaks (orange crosses), peak bases (red), and widths (green), taken from the SciPy online manual.[4]

Unfortunately, automated analysis for peak picking can output incorrect peak widths if the baseline function sensitivity (selected by the user) is incorrect. If the function is too sensitive, then too many peaks are chosen; if it is not sensitive enough then the peaks of interest are not analysed. This is particularly challenging for overlapping peaks where the baseline is given where the peaks split (demonstrated with overlapping peaks above). It is therefore crucial that for the use of APPE, materials are chosen with intense single low-angle reflections to avoid errors. The sensitivity is related to the peak heights and so the sensitivity for APPE peak picking function for the reference (with peak heights of 10,000 -100,000) is different than for the diffraction data typically being output from the SX diffractometer (with peak heights of 10 – 500).

2.6 Conclusions and next steps

The optimisation of any hardware setup results in a large number of data sets that require analysis. APPE is able to measure the FWHM from a series of simulated and experimental PXRD patterns and to rank the measurements with appropriately labelled filenames corresponding to the lowest FWHM *i.e.,* the optimal instrumental parameters. The creation of APPE has the potential to greatly reduce the time needed for parameter optimisation as well as creating a more accessible (in both time and experience) path for researchers to explore the effect of parameter changes on data quality. With the tools created in this chapter. the next step is to determine the optimal parameters for the University of Reading's Rigaku Synergy single-crystal diffractometer.

2.7 References

1 W. I. F. David, K. Shankland, L. B. McCusker and C. Baerlocher, Structure Determination from Powder Diffraction Data, Oxford University Press, 2006.

2 A. Boulle and J. Kieffer, J. Appl. Crystallogr., 2019, **52**, 882–897.

3 S. van der Walt, S. C. Colbert and G. Varoquaux, Comput. Sci. Eng., 2011, **13**, 22–30.

4 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko and Y. Vázquez-Baeza, Nat. Methods, 2020, **17**, 261–272.

5 E.-E. Doberkat, Python 3, De Gruyter, 2018.

6 D. Rolon-Mérette, M. Ross, T. Rolon-Mérette and K. Church, Quant. Methods Psychol., 2020, **16**, S3–S11.

7 A. A. Coelho, J. Appl. Crystallogr., 2018, **51**, 210–218.

8 W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell and J. C. Cole, J. Appl. Crystallogr., 2006, **39**, 910–915.

9 Oxford Diffraction /Agilent Technologies UK Ltd, CrysAlisPro.

10 Anaconda Documentation, Anaconda Navigator.

Chapter 3

Optimisation of Rigaku Synergy diffractometer for accurate powder diffraction

3.1 Introduction

The focus of this chapter is the optimisation of the configuration of a single crystal diffractometer, with the aim of collecting accurate powder diffraction data that is ultimately suitable for structure solution purposes. A brief introduction is provided of the different instrument configurations available for single crystal X-ray diffraction (SCXRD) and powder X-ray diffraction (PXRD), to highlight the aspects of both techniques that are most pertinent to this work.

3.1.1 Single crystal data collection

Single crystal diffraction is a widely accessible technique that provides a rapid, routine route to obtaining high-quality structural information. As the technology for single crystal diffraction has become more widely available and less costly (through technical advances and reduction in the cost of machine parts), this has allowed for the technique to become relatively inexpensive. So much so that it is typical for most science research institutions to have either an in-house single crystal diffractometer or have a close relationship with another institution that does.

Figure 3.1 shows a typical modern in-house single crystal diffractometer (Rigaku Synergy). Of particular note are the high-intensity microsource X-ray tubes and the single-photon-counting detector, which together allow rapid data collection even from very small (50 μm length) single crystals. Data collection begins with a pre-experiment on a selected crystal taken from a sample, during which the diffractometer software attempts to work out the unit cell and space group of the crystal. The software then devises a strategy for the collection of a complete dataset that will permit structure solution and refinement. This process is generally fully automated, requiring little or no input from the user. Crucially, the pre-experiment only takes around 5-10 minutes and so allows the researcher to quickly assess the quality of the crystal. If the crystal is of poor quality (for example, if less than 70% of the observed reflections can be attributed to a calculated unit cell) a user can select another crystal from the sample.

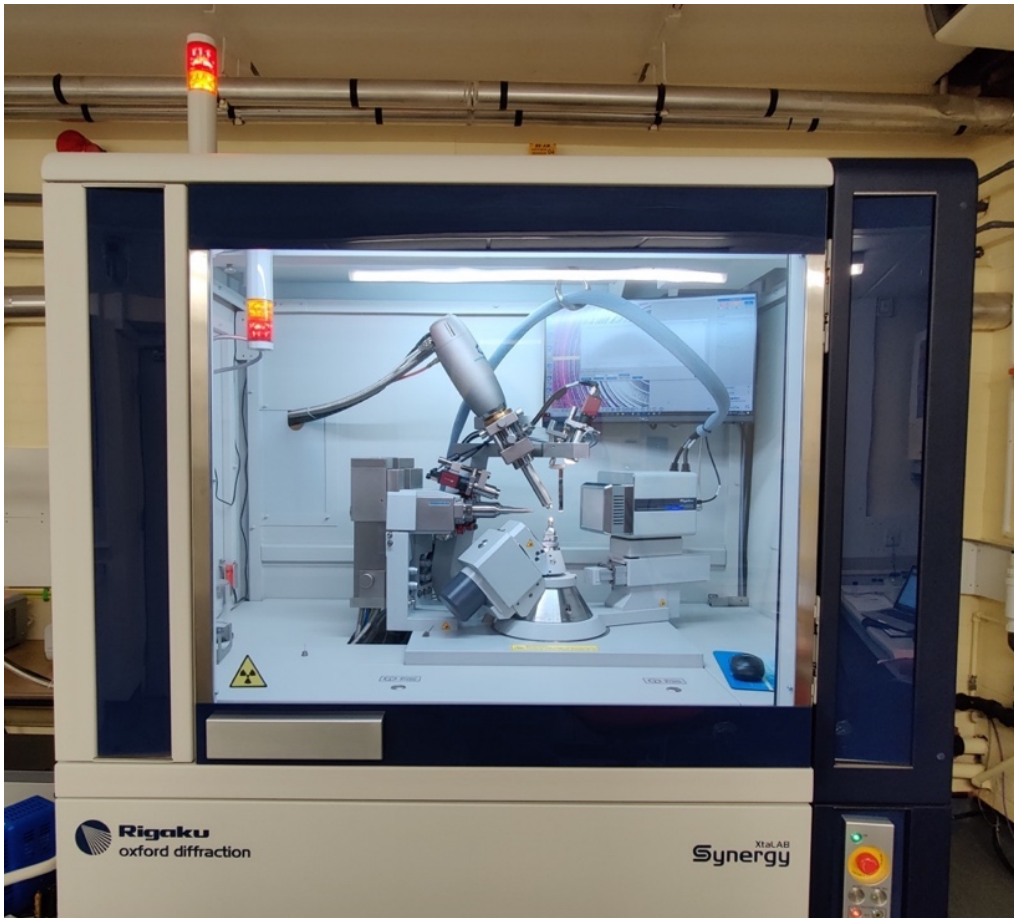


Figure 3.1 Rigaku Synergy single crystal diffractometer, with dual X-ray microsources and a HyPix 6000HE detector. (University of Reading, 2020)

3.1.2 Powder diffraction

When single crystals of a material are not available, or when the material's bulk properties are of interest, powder diffraction can be used for phase identification and / or testing crystallinity. The powder diffraction pattern obtained can be thought of as a "fingerprint" of the crystalline phases present in the sample under study and used both for structure identification and quantitative phase analysis. If both accurate peak positions and intensities are available, then structure solution and subsequent refinement can be attempted. Many research facilities have a flat-plate powder X-ray diffractometer setup operating in reflection mode, as this is well suited to high-throughput measurements. However, such an instrument is not well-suited to structure solution and refinement of molecular organic materials, mainly due to the problem of preferred orientation, but also to factors such as sample height displacement. In contrast, a capillary setup operating in transmission mode is very well suited to the structure solution and refinement of molecular organic materials. However, transmission capillary setups are not so widely used, largely due to the overhead of material needed for filling glass capillaries with

powder samples and a lack (in a laboratory environment, at least) of an easy route to high-throughput measurements.

Powder X-ray instrumentation has not changed as significantly in the last decade has its single-crystal equivalent; sealed-tube sources remain the norm, though detectors now collect data more efficiently than their decade-old equivalents, and encouragingly there is a trend towards instruments that can be easily swapped from reflection to transmission modes.

3.1.2.1 Instrument geometry and data collection

Both transmission and reflection instruments have the potential to operate in high-resolution modes, with the X-ray beam focussing on the detector after interacting with the sample. Aside from the differences in basic geometry, the data collection parameters are largely the same: 2θ step size, 2θ range, and acquisition time per step. Some instruments operate in transmission mode with samples placed not in capillaries, but in kapton-based multi-well plates. It has been shown that the collection of structure solution quality data is possible from such setups, but preferred orientation remains an issue in this technique.[1]

Data collection times in reflection mode are generally shorter than those in transmission, because of the area of sample exposed to the beam. Despite this, collection times are not prohibitive (at least in an academic environment) for capillary samples in transmission. For example, the initial 'pre-experiment' to get indexing-quality data might be 4° – 35° 2θ with a collection time of 2 hours, whilst a refinement-quality dataset would typically span 4° – 70° 2θ and run for 12 to 16 hours.

3.1.3 Powder diffraction data collection on a single crystal diffractometer

On Rigaku single crystal diffractometers, the control software (CrysAlisPro) has an inbuilt powder data collection function that drives the diffractometer in such a way as to permit the collection of diffraction data from polycrystalline samples. The operating window is shown in Figure 3.2. CrysAlisPro gives a variety of options to collect the powder diffraction pattern, but it is not necessarily obvious to users what settings should be used to yield the highest quality powder diffraction pattern. Also, it is not clear what, if any, relationship exists between the chosen values and the sample. Whilst default settings might be acceptable for the collection of PXRD data for phase identification purposes, they are not necessarily optimal for indexing, SDPD and refinement purposes and thus the instrument configuration needs to be optimised.

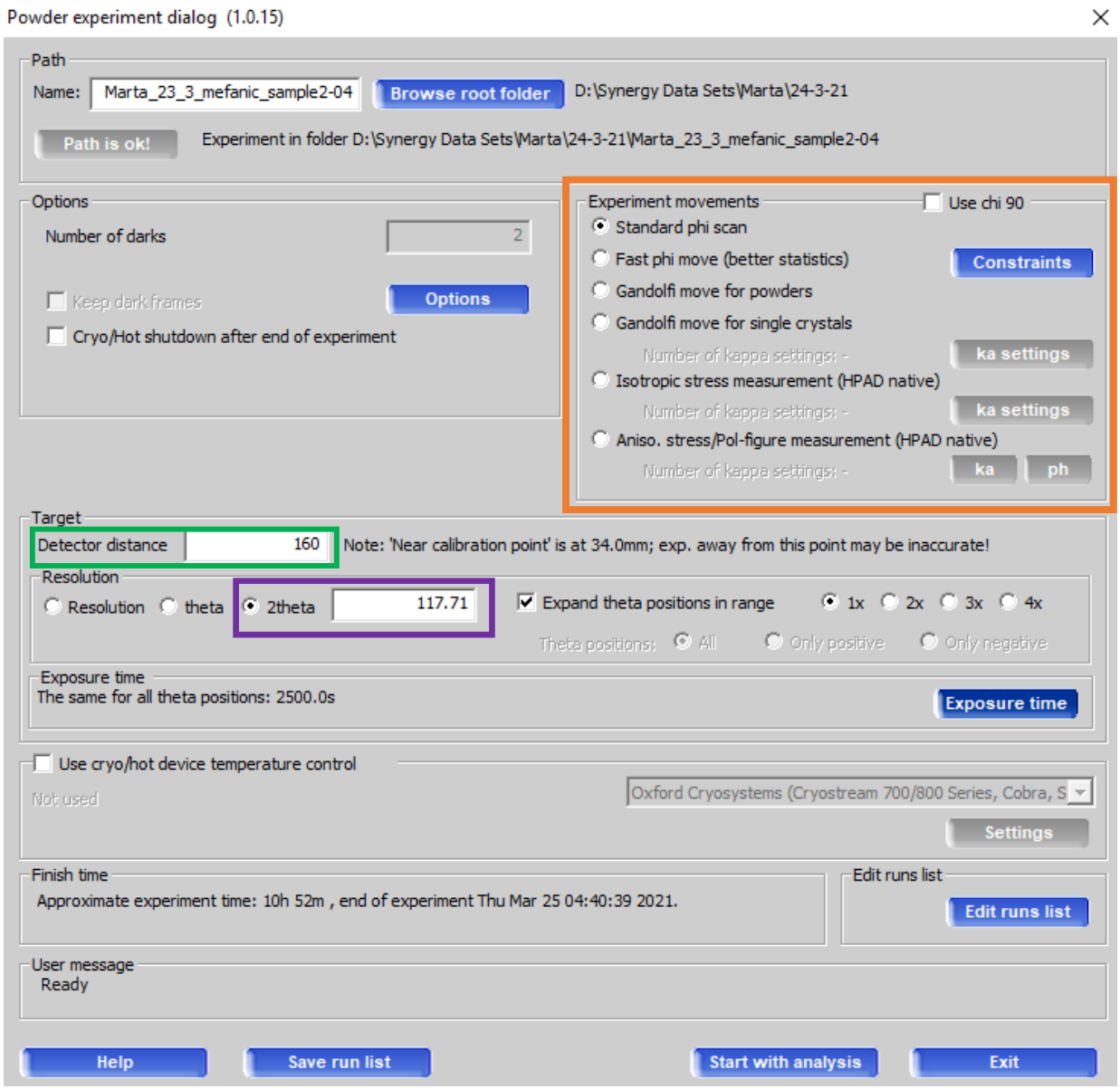


Figure 3.2. The powder diffraction experiment strategy control window in CrysAlisPro. The collection mode is highlighted in orange, detector distance in green and 2θ resolution in purple.

The physical parameters that can be altered for a powder diffraction experiment on a single crystal diffractometer are summarised in Figure 3.3.

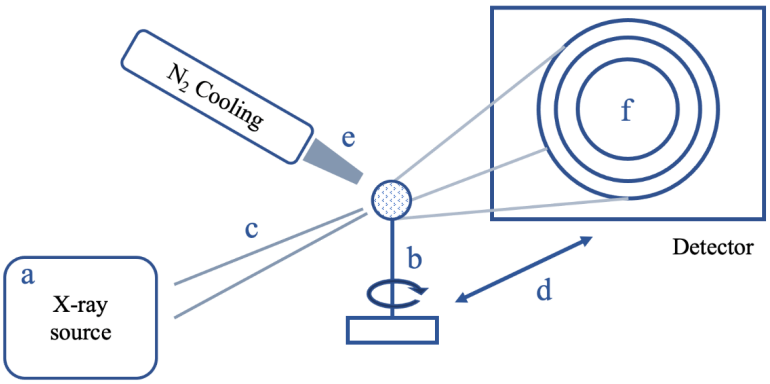


Figure 3.3. Basic schematic of single crystal X-ray diffractometer, with respect to powder diffraction data collection. Parameters, (**a**) radiation source, (**b**) mode of collection, (**c**) beam divergence, (**d**) detector distance and (**e**) temperature and (**f**) ring centring.

A further description of how each parameter (labelled as per Figure 3.3) affect the end powder data quality is shown in Table 3.1 and further details of each parameter are given below. It should be noted that the factors provided in Table 3.1 are specific to the Rigaku Synergy and the details will vary between different single crystal diffractometers.

Table 3.1 Factors affecting powder diffraction data quality on a single crystal diffractometer.

Factor	Choice / range	Effect on data
a. Radiation source	Cu / Mo	Different radiation wavelengths affect the scattering of the sample.
b. Collection mode	'Standard Phi', 'Fast Phi', 'Gandolfi for Powders', 'Gandolfi for SX'	Collection modes are used to externally average the sample, reducing preferred orientation seen in the diffraction pattern.
c. Beam divergence / %	2 – 100 (*ca.* 0.1 - 9.5mrad)	Reducing beam divergence reduces beam intensity as well as reducing diffraction peak width.
d. Detector distance / mm	34 – 200	Changing the detector distance changes how diffraction cones are intersected and diffraction data is recorded.
e. Temperature / K	100 - 300	Reducing temperature increases scattering from the sample by reducing thermal vibration of atoms.
f. Ring centring	N/A	Incorrect ring centring causes peak broadening (eventually peak splitting) where opposite sides of rings have different distances from the centre.

a. Radiation source

Typical lab-based X-ray diffractometers generate X-rays by hitting a metal target with high-energy electrons; excited atoms in the target then lose energy, emitting a characteristic spectrum of X-ray photons in the process. The Rigaku Synergy single-crystal diffractometer is dual source and is equipped to use either a molybdenum (*ca.* 0.71 Å) or copper target (*ca.* 1.54 Å). The chosen wavelength affects the peak separation, according to Bragg's law ($n\lambda = 2d\sin\theta$); it is obvious that for any fixed set of distances (d) in the crystal, the diffraction angle θ will scale with the wavelength, as illustrated in Figure 3.4.

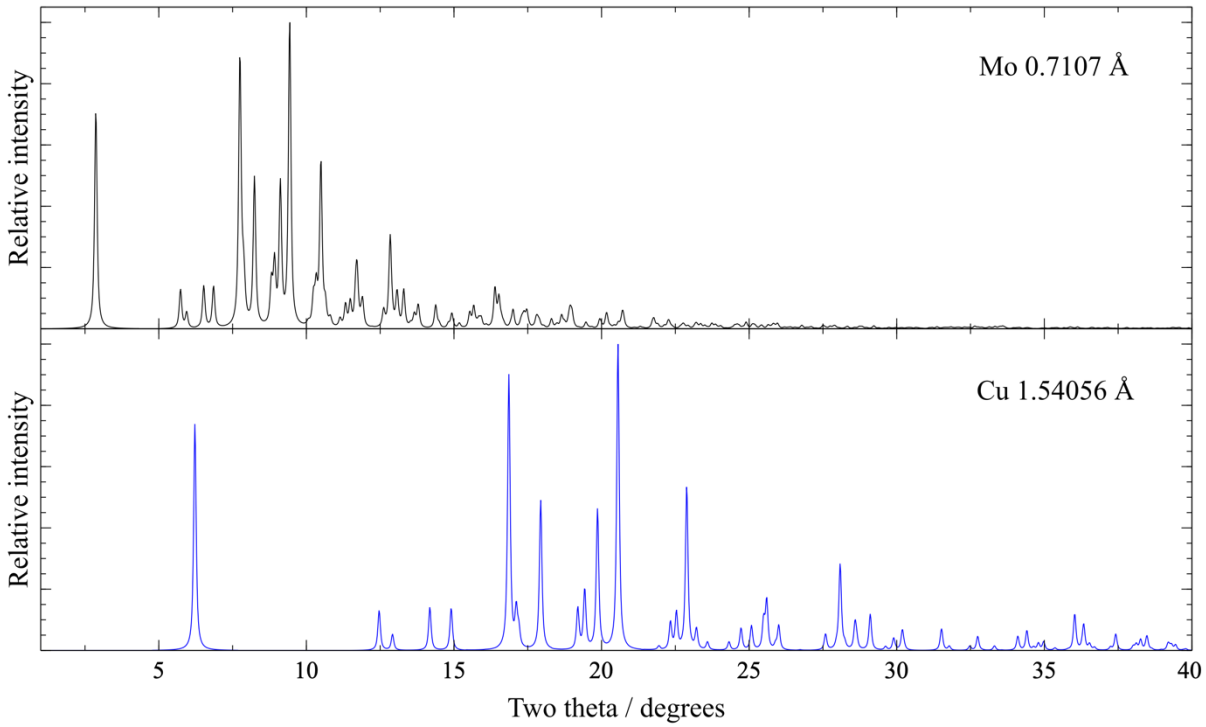


Figure 3.4. Simulated PXRD patterns for ibuprofen (form I, CSD refcode IBPRAC01) using Mo and Cu incident radiation. Simulations were performed in Mercury.[2]

Although using a different X-ray wavelength does not inherently change the diffraction, instrumental resolution factors (such as the ability of the detector to discriminate between two closely spaced reflections in 2θ) means that the observed diffraction can differ significantly. Scattering is proportional to λ^3 meaning that, in general, Cu radiation is preferred for low-symmetry molecular materials because of the increased range over which reflections are spread and increased scattered intensity.

It is also worth noting that SX diffractometers do not employ incident beam monochromators for the main emitted radiation; for example, incident Cu Kα radiation consists of Cu Kα_1 and Cu Kα_2 wavelengths. In contrast, most PXRD instruments designed for structural work use

monochromatic incident beams *e.g.*, Cu Kα_1. The presence of two incident wavelengths results in broader observed peaks at higher angle.

b. Mode of collection

The 'standard phi' mode of collection, where the sample rotates around the phi axis (shown in Figure 3.3) is the simplest mode and entirely suitable when the sample being studied exhibits a good powder average. In addition to 'Standard phi', the available collection modes (available on the Rigaku Synergy) are: Fast phi'', 'Gandolfi for powder' and 'Gandolfi for single crystal'. Each of these collection modes are best suited for different samples with the aim of reducing the effects of preferred orientation (PO). These modes and their role in optimal collection are not considered in this chapter but are explored further in Chapter 5.

c. Detector distance

The closer the detector is to the sample, the smaller is the diameter of each diffraction ring captured by the detector, resulting in a larger number of rings captured per scan. As a result, fewer frames are needed to capture a specific 2θ range when the detector is positioned closer to the sample. However, closer detector distances also result in greater peak overlap, diffracted beams with very similar diffraction angles arrive at the detector very close together. On the other hand, when the detector is too far from the sample, the diffracted intensity falls significantly due to air scatter. It is also worth noting that the diffracted beam is divergent, and so will broaden as the X-rays travel further away from the sample. Therefore, it is likely that there is a particular detector distance value at which these two competing effects give rise an optimal diffraction pattern.

d. Beam divergence

Microsource X-ray beams within SX instrumentation focus on the sample for maximum intensity, and the diffracted beams diverge en-route to the detector, as shown in Figure 3.5.

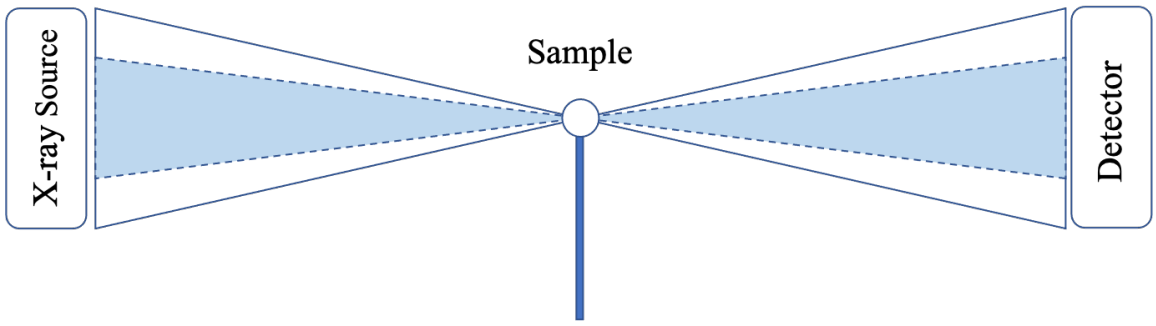


Figure 3.5. Beam geometry of single crystal diffraction setup. Maximum divergence, full beam, shown by solid line. Reduced beam divergence represented by dashed inner line.

The diffractometer used in this study permits the divergence of the Cu radiation beam to be changed by the user from the maximum (100% intensity) to a minimum divergence of *ca.* 2% intensity. Single crystal experiments typically use the beam at maximum intensity. For the purposes of powder data collection, where the diffraction pattern is 1-dimensional (from 2-dimensional rings), the effect of varying the incident beam divergence is likely to be a key element of optimising the diffracted data quality.

The beam divergence of the Rigaku Synergy can be altered using the inbuilt CrysAlisPro function, shown below in Figure 3.6.

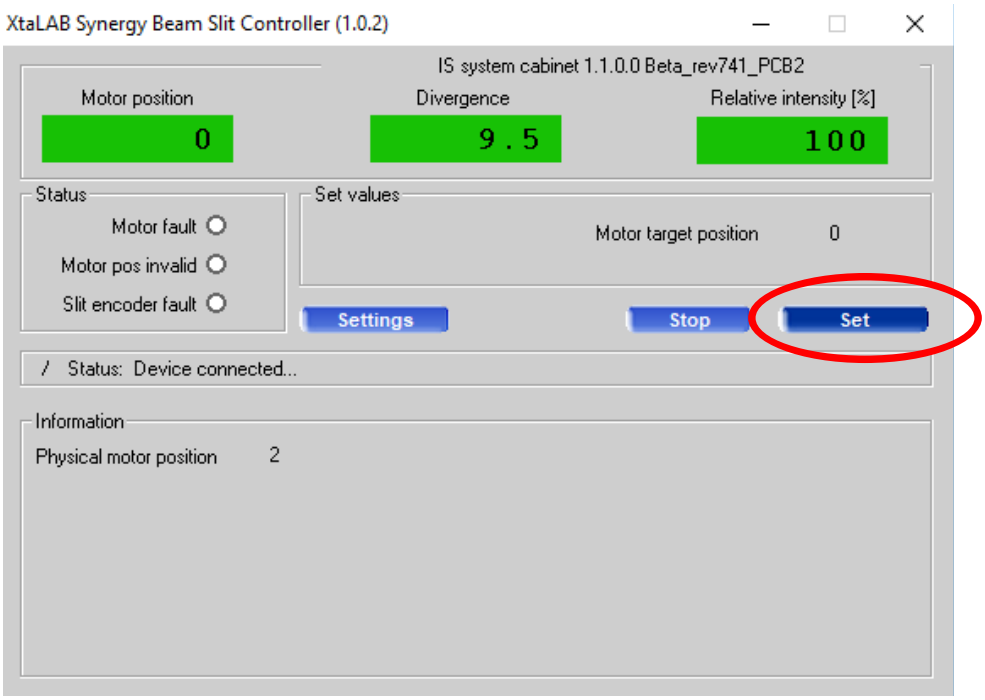


Figure 3.6. CrysAlisPro beam divergence control screen, with circle highlighting where the user can set the beam divergence between 2 – 100%.

e. Temperature

Single crystal experiments are typically carried out at low temperature in order to reduce thermal vibrations of atoms in the structure of interest. Similarly, cooling of powder samples improves the diffraction pattern via the same effect. Cooling of a sample, typically achieved with a low-T N_2 gas stream, can also be used to hold a sample in place during an experiment, by freezing the oil used to hold the crystal to the mounting fibre.

f. Ring centring

The use of an area detector with a single crystal diffractometer allows the rapid visualisation of the quality of the sample. In particular, the presence of individual larger crystallites and

preferred orientation in the sample can be seen from the uniformity or otherwise of the diffraction rings; rings with bright spots indicate the presence of large single crystals in the sample, whilst rings that are not uniformly intense around their circumference are indicative of preferred orientation. This is a significant advantage over point or linear PSD detectors more typically used in PXRD instrumentation. However, the use of an area detector also means that there is room for systematic errors in centring when processing the diffraction rings post experiment. The diffractometer used in this study is typically calibrated at a detector distance of 34 mm, and away from this distance, a centre coordinate is extrapolated, meaning that its accuracy is not guaranteed. The use of an incorrect centre point during ring integration (the process where software integrates diffracted intensity around the ring in order to create the familiar 1D PXRD diffraction profile) results in peak broadening, and in some cases peak splitting. Figure 3.7, shows how this arises.

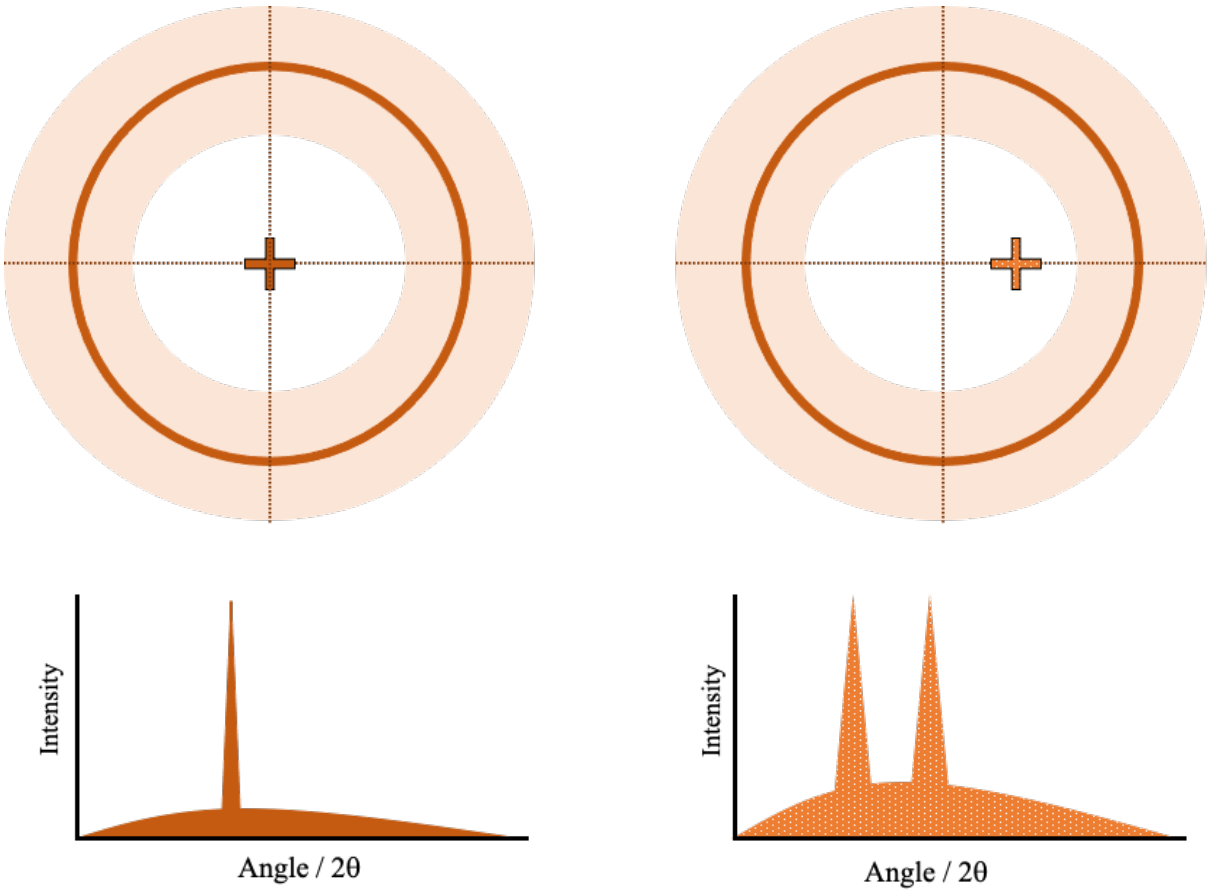


Figure 3.7. The outcome of correct ring centring (left) and incorrect centring (right), with the correct centre highlighted with dotted lines and the centre used by the peak extraction algorithm indicated by the cross. The background is shown in light orange around the diffraction ring.

3.2 Aims

To collect diffraction patterns that give the greatest opportunity for structure solution, the optimal instrumental parameters for data collection must be used. As single-crystal diffractometers are not expressly designed for powder data collection, these parameters must be determined. This chapter aims to determine the optimal instrumental parameters for the collection of high-quality powder diffraction data using a Rigaku Synergy single crystal diffractometer. Furthermore, it aims to establish a systematic approach to parameter optimisation that is transferrable to other single crystal diffractometers.

3.3 Methodology

All experiments were performed using the University of Reading's (UoR) Rigaku Synergy diffractometer with dual microfocus PhotonJet-S sources (Mo, Cu), a HyPix-6000HE detector and an Oxford Cryosystems Cryostream 800. Data visualisation and reduction were performed using CrysAlisPro (version 1.171.41.93a) and peak width analysis was performed using APPE (see chapter 2).

3.3.1 Approach to optimising diffractometer setup for high-quality PXRD data

An overall flowchart of the parameter optimisation is given below in Figure 3.8.

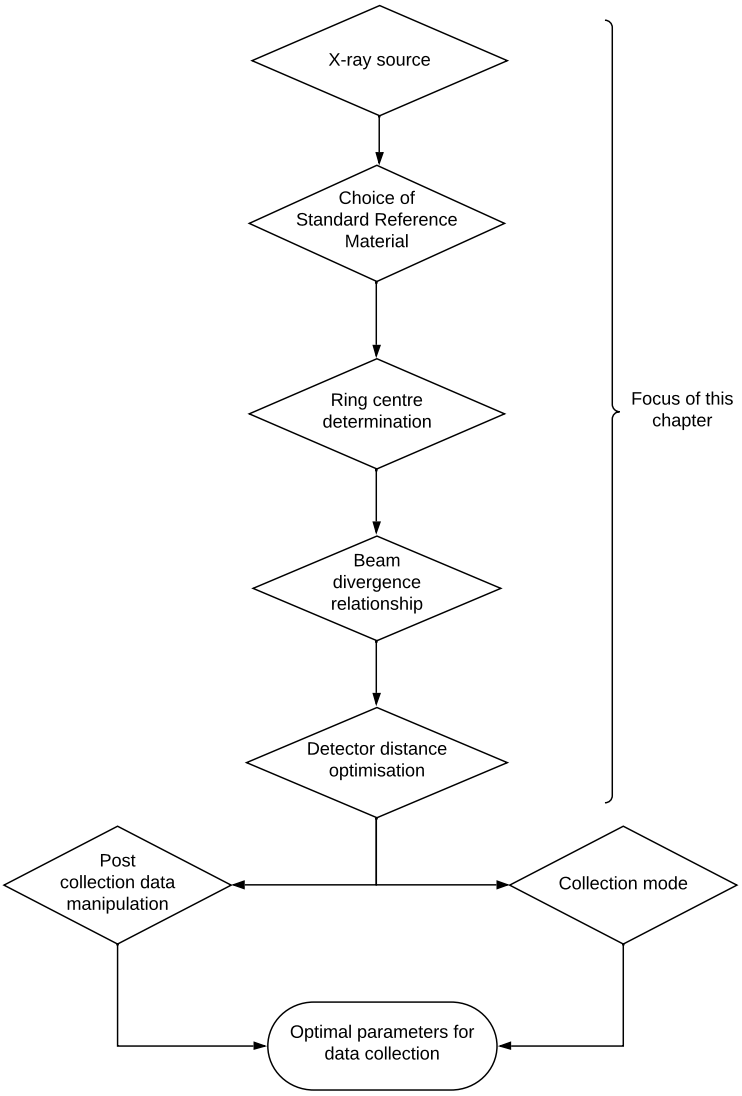


Figure 3.8. Flow chart of optimal parameter determination used for powder diffraction data collection on a single crystal diffractometer.

3.3.2 Standard materials, preparation, and mounting

All compounds used in this chapter were mounted by first mixing the powder with a very small amount of Paratone oil, in order to create a 'paste' from which a roughly spherical sample was used. This sphere was held (by the Paratone oil) to a sample holder, composed of a piece of fibre optic cable (*approx*. 3 cm length) 225 microns in diameter, anchored to a magnetic base. Further description and development of this mounting procedure is discussed in chapter 5.

Initial test scans were performed using the molecular organic material L-glutamic acid (form II, CSD refcode LGLUAC01). For the subsequent parameter optimisation and ring centring experiments, inorganic samples LaB_6 (NIST SRM 660c) and Si (NIST SRM 640c) were considered for use as SRMs, and LaB_6 was chosen used as standards (as shown below in Figure 3.9).

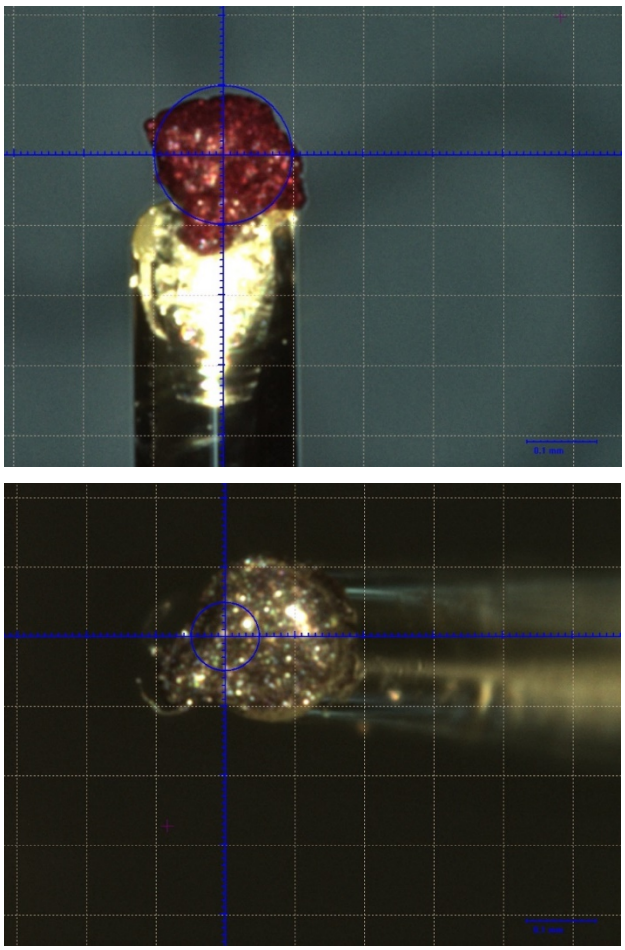


Figure 3.9. Standard samples used for parameter optimisation, above: LaB_6 sample; below: Si sample. Both samples were held with Paratone oil. Grid length is 0.1 mm.

The minimum amount of oil was used in order to reduce its background contribution to scattering. Due to the generally isotropic nature of the standards, the 'standard phi' mode of

collection was found to be perfectly acceptable and was used for all parameter optimisation scans. Initial attempts to cool the sample were frustrated by the N_2 gas flow blowing the powder off the sample holder; this could only be mitigated by moving the Cryostream further back from the sample, with consequent loss of temperature control. As such, all parameter optimisation scans were carried out at ambient temperature.

3.3.3 Parameter optimisation experiments

3.3.3.1 Radiation source

Test scans were performed on the sample L-glutamic acid using Mo and Cu radiation with a detector distance of 100 mm. Results are shown in Figure 3.10 and Figure 3.11. All further parameter optimisation experiments were performed using only Cu radiation.

3.3.3.2 Choice of SRM

To select the ideal standard material, diffraction patterns were collected from both Si and LaB_6 at 2% beam divergence (approximately 0.1 mrad) at 100 mm detector distance. Results are shown in Figure 3.12 and Figure 3.13. LaB_6 was selected for all subsequent optimisations.

3.3.3.3 Optimisation of data processing - Beam centring

Correct centring points were found by a grid search method of manually adjusting the centre coordinate of the diffraction image and reprocessing (where the rings are re-integrated from the new centre coordinates) within CrysAlisPro. Peak widths were evaluated by APPE, using background-subtracted data, with the background subtraction being performed within DASH. Only the first diffraction peak was used in order to minimise the influence of Cu $K\alpha_2$ broadening.

For the purpose of beam centring, diffraction patterns of LaB_6 were collected in standard phi mode at 2% beam divergence (*ca.* 0.1 mrad), at detector distances of 100, 160, and 190 mm. A linear relationship between centre coordinates and detector distance was assumed. The same procedure was performed for diffraction patterns at 100 mm detector distance using minimum and maximum beam divergence used for future beam divergence analysis.

Once determined, the correct centre coordinates remain correct until the diffractometer is recalibrated as part of its routine maintenance. Results for beam centring at varying detector distances are shown in Figures 3.14 – 3.16.

3.3.3.4 Optimisation of beam divergence

Diffraction patterns of LaB_6 were collected in standard phi mode at 100 mm detector distance, at approximately 8% beam intensity intervals spanning the available range of 2 – 100% of beam intensity (*ca.* 0.1 – 9.5 mrad). Results are shown in Figure 3.17 with results before and after improved centring. A further comparison of the 1st peak of LaB_6 at *ca.* 2 and 100% beam divergence is shown in Figure 3.18.

3.3.3.5 Optimisation of detector distance

Diffraction patterns of LaB_6 were collected in standard phi mode at room temperature. Datasets were collected at 5 mm detector distance intervals in the range 80 – 200 mm to determine the optimal detector distance. This optimisation was performed using Cu Kα radiation at minimum beam divergence. Results are shown in Figure 3.19 with a further comparison of the 1st peak of LaB_6 at 80 and 160 mm detector distances shown in Figure 3.20.

3.3.3.6 Impact of optimisation on FWHM and peak intensities

Diffraction patterns collected of LaB_6 using optimal settings (*ca.* 2% beam divergence, 160 mm detector distance) were overlayed with data collection of the sample using the 'least optimal' settings (single crystal operating settings at 100% beam divergence and 34 mm detector distance). This overlay was used to assess the overall improvement in FWHM from optimisation as well as the impact on peak intensity from moving from initial parameters. Results showing comparison of absolute intensities are shown in Figure 3.21 and comparison of normalised are shown in Figure 3.22.

3.4 Results

Presented below are the results of parameter optimisation in the form of data screenshots taken from CrysAlisPro, 1D diffraction patterns or plots derived from peak widths.

3.4.1 Choice of radiation

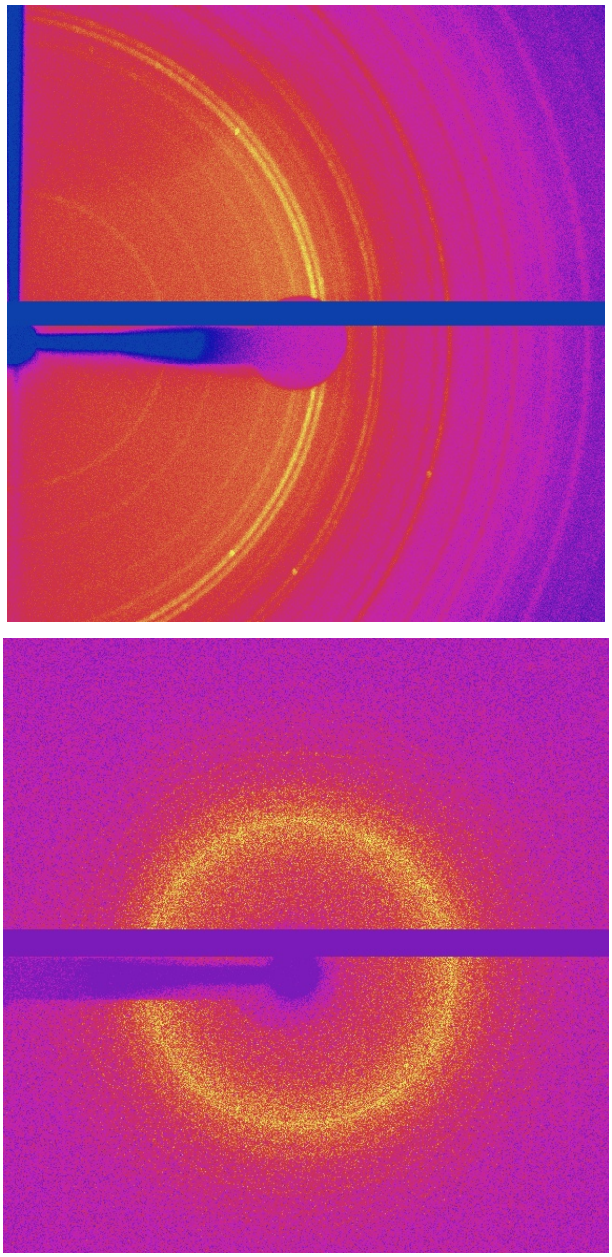


Figure 3.10. 2D images of diffraction from L-glutamic acid using Cu radiation (upper) and Mo radiation (lower). In both cases, detector distance = 100 mm and beam divergence at 100% (*ca.* 9.5 mrad).

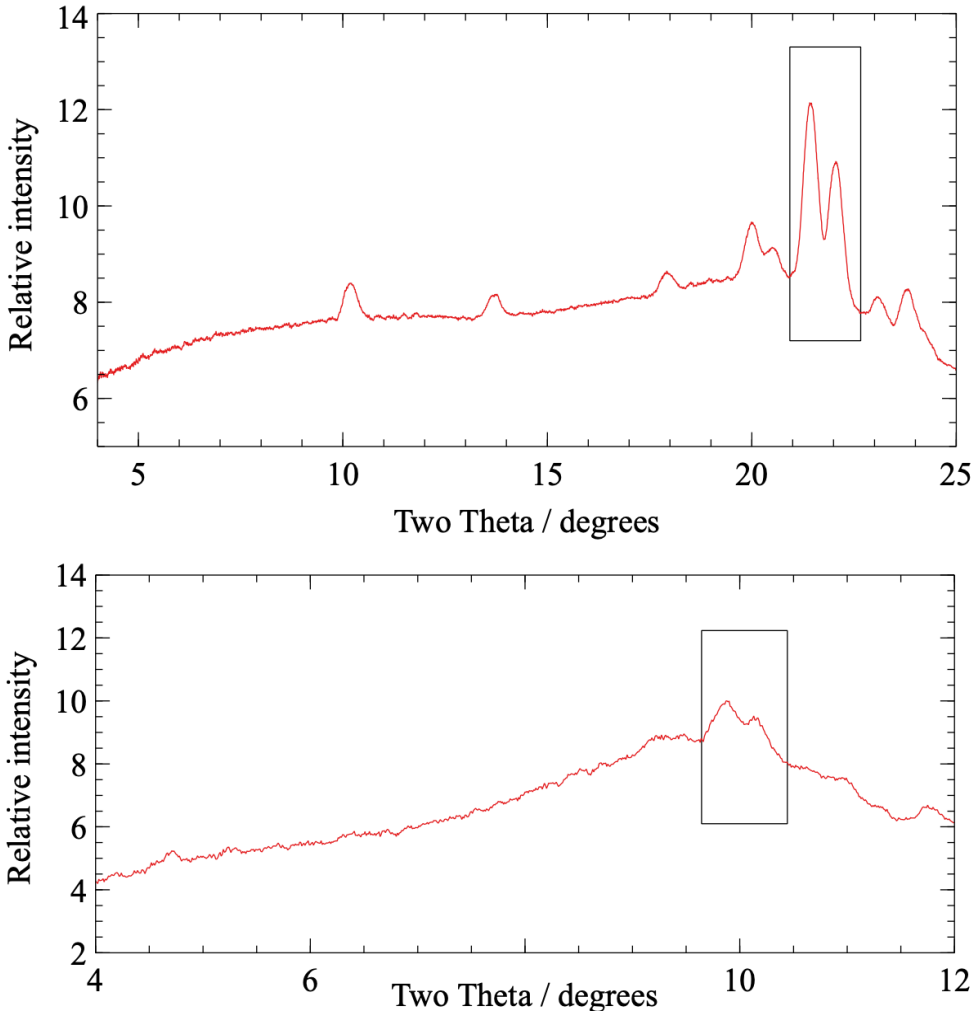


Figure 3.11. Integrated PXRD data of L-glutamic acid using Cu radiation (upper) and Mo radiation (lower). In both cases, detector distance is 100 mm and beam divergence is 100% (*ca.* 9.5 mrad). Equivalent reflections are highlighted by the boxes.

It is evident from the plots above that the diffraction pattern obtained using the Cu Ka source is significantly better in terms of signal to background ratio, and therefore all further optimisations were performed using only the Cu Kα source.

4.4.2 Choice of SRM for parameter optimisation

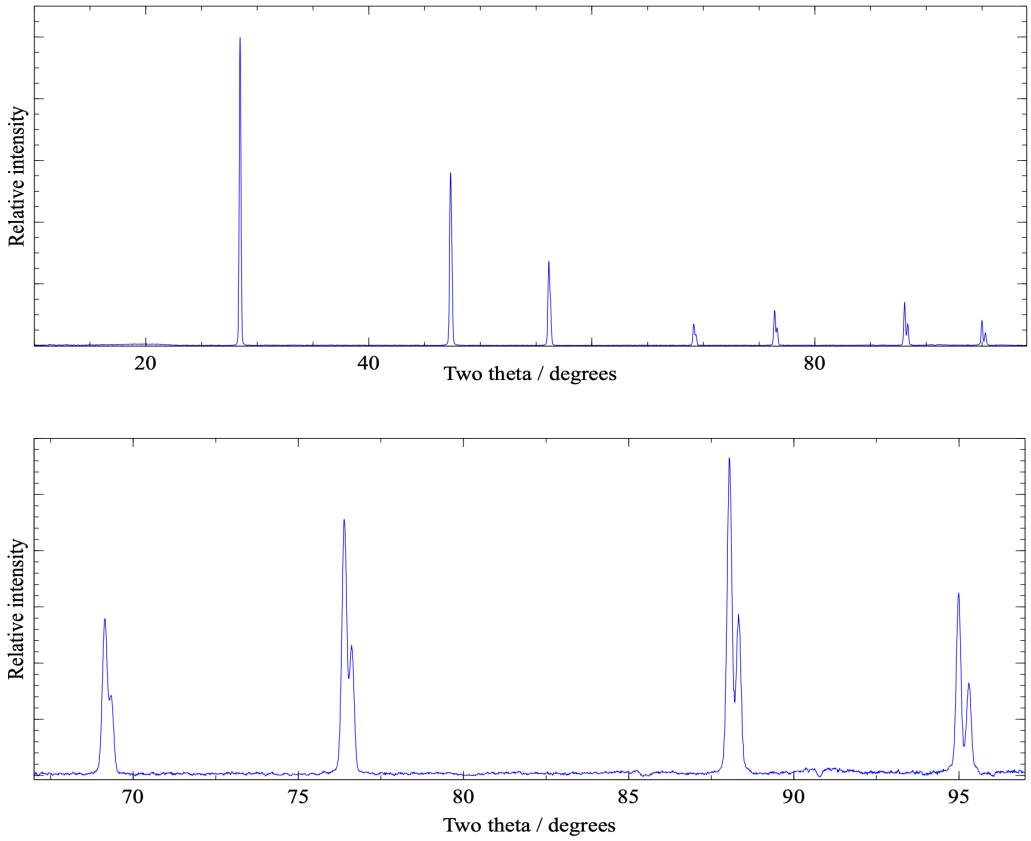


Figure 3.12 Integrated PXRD data for NIST 640c SRM silicon collected at detector distance: 100 mm, X-ray source: Cu, beam divergence: 2% (approximately 0.1 mrad). The first diffraction peak is at approximately 28° 2θ.

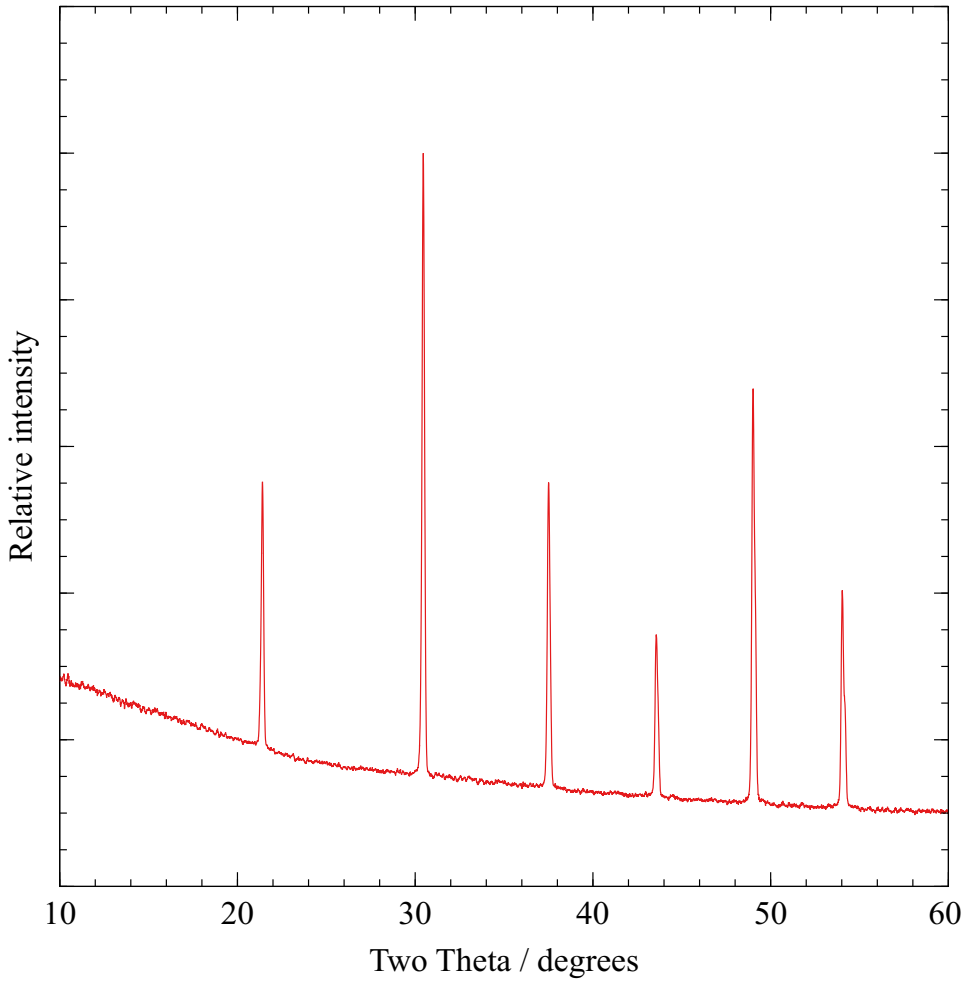
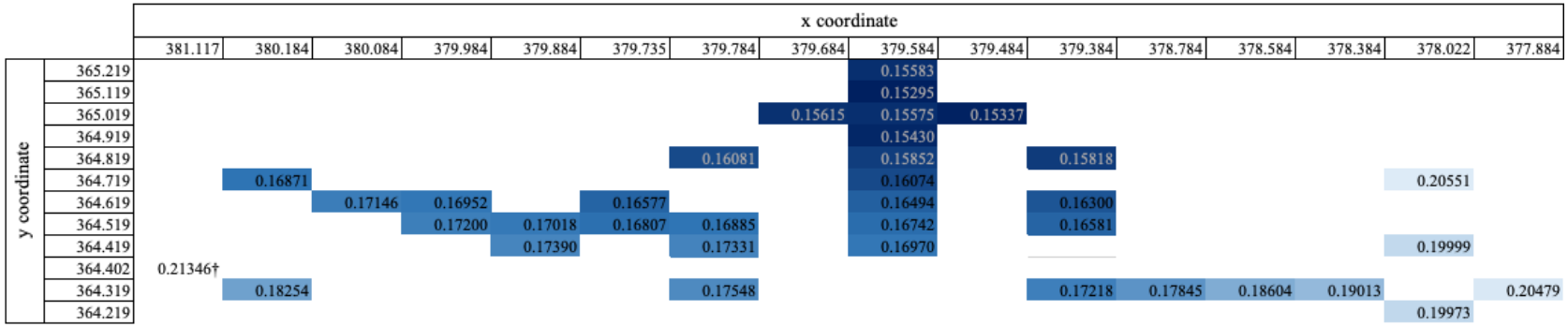


Figure 3.13. Integrated PXRD data for NIST 660c SRM LaB_6 collected at detector distance: 100 mm, X-ray source: Cu Ka, beam divergence *ca.* 0.2 mrad. The first diffraction peak is at approximately 22° 2θ.

LaB_6 was found to be the superior material for the purposes of parameter optimisation, due to the lower 2θ value of the 1st peak (*approx*. 22° vs 28° 2θ) and was therefore chosen as the SRM for further analysis.

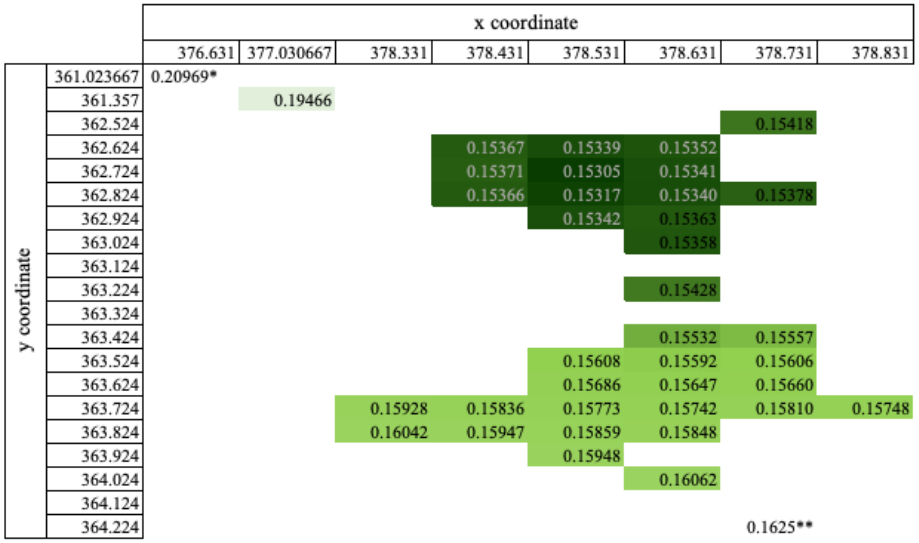
3.4.3 Ring centring analysis

Results for optimal centre coordinates are presented in Figures 3.14 – 3.16 corresponding to three detector distances (100, 160 and 190 mm). For all of the detector distances, diffraction data was collected at minimum beam divergence (*ca.* 0.2 mrad) using LaB_6. The FWHM values reported are those determined from the first peak of LaB_6, using the same sample for all three data collections. Optimal positions found iteratively with 2 – 4 data points per iteration.



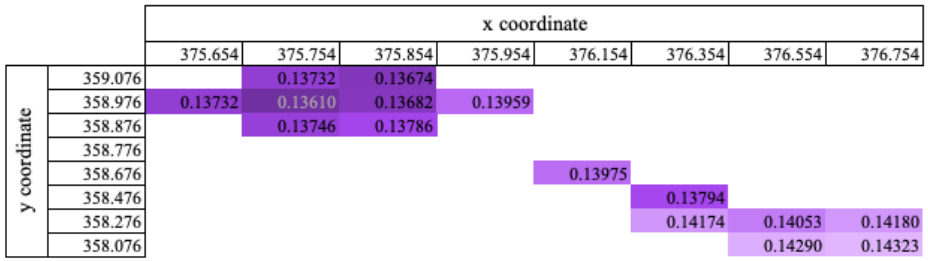
y coordinate \ x coordinate	381.117	380.184	380.084	379.984	379.884	379.735	379.784	379.684	379.584	379.484	379.384	378.784	378.584	378.384	378.022	377.884
365.219									0.15583							
365.119									0.15295							
365.019								0.15615	0.15575	0.15337						
364.919									0.15430							
364.819							0.16081		0.15852		0.15818					
364.719		0.16871							0.16074						0.20551	
364.619			0.17146	0.16952		0.16577			0.16494		0.16300					
364.519				0.17200	0.17018	0.16807	0.16885		0.16742		0.16581					
364.419					0.17390		0.17331		0.16970						0.19999	
364.402	0.21346†															
364.319		0.18254					0.17548				0.17218	0.17845	0.18604	0.19013		0.20479
364.219															0.19973	

Figure 3.14. A pseudo-contour map of the FWHM of the 1st Peak of LaB_6 at 100 mm detector distance against x and y ring centre coordinates. The darkest shade of blue indicates the lowest FWHM, and thus the new centre coordinates. † indicates starting point from 'auto centre' function.



y coordinate \ x coordinate	376.631	377.030667	378.331	378.431	378.531	378.631	378.731	378.831
361.023667	0.20969*							
361.357		0.19466						
362.524							0.15418	
362.624				0.15367	0.15339	0.15352		
362.724				0.15371	0.15305	0.15341		
362.824				0.15366	0.15317	0.15340	0.15378	
362.924					0.15342	0.15363		
363.024						0.15358		
363.124								
363.224						0.15428		
363.324								
363.424						0.15532	0.15557	
363.524					0.15608	0.15592	0.15606	
363.624					0.15686	0.15647	0.15660	
363.724			0.15928	0.15836	0.15773	0.15742	0.15810	0.15748
363.824			0.16042	0.15947	0.15859	0.15848		
363.924					0.15948			
364.024						0.16062		
364.124								
364.224							0.1625**	

Figure 3.15. A pseudo-contour map of the FWHM of the 1st Peak of LaB_6 at 160 mm detector distance against x and y ring centre coordinates. The darkest shade of green indicated the lowest FWHM, and thus the new centre coordinates. * indicates the machine model coordinate for 160 mm and ** indicates 'by eye' starting point.



y coordinate \ x coordinate	375.654	375.754	375.854	375.954	376.154	376.354	376.554	376.754
359.076		0.13732	0.13674					
358.976	0.13732	0.13610	0.13682	0.13959				
358.876		0.13746	0.13786					
358.776								
358.676					0.13975			
358.476						0.13794		
358.276						0.14174	0.14053	0.14180
358.076							0.14290	0.14323

Figure 3.16. A pseudo-contour map of the FWHM 1st Peak of the 1st LaB_6 at 190 mm detector distance against x and y ring centre coordinates. The darkest shade of purple indicated the lowest FWHM, and thus the new centre coordinates.

3.4.4 Optimal beam divergence value

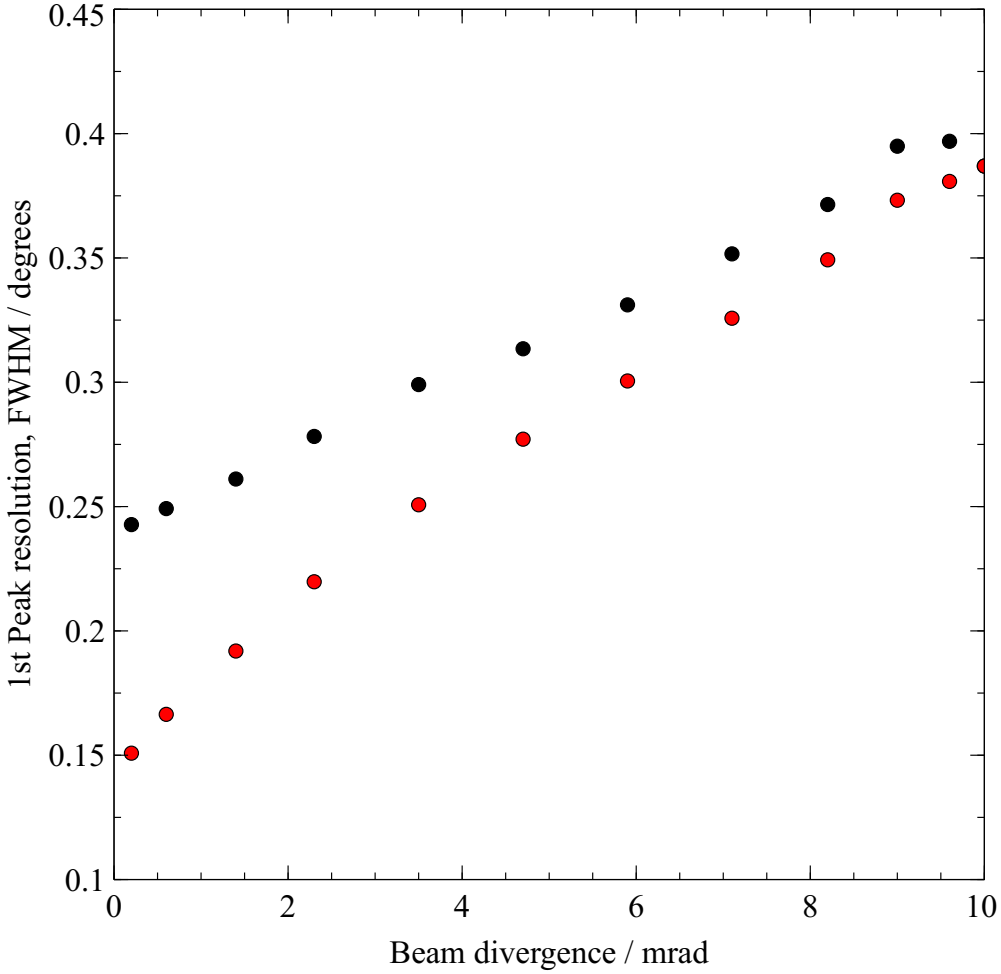


Figure 3.17. FWHM as a function of beam divergence, Cu radiation at 100 mm detector distance. Machine model centring coordinates (black) and improved centring coordinates (red).

A comparison of the diffraction patterns collected at full beam and at 2% beam divergence, both pre- and post-centring, is given below, showing the effect on peak shape when beam divergence is reduced without centring adjustment.

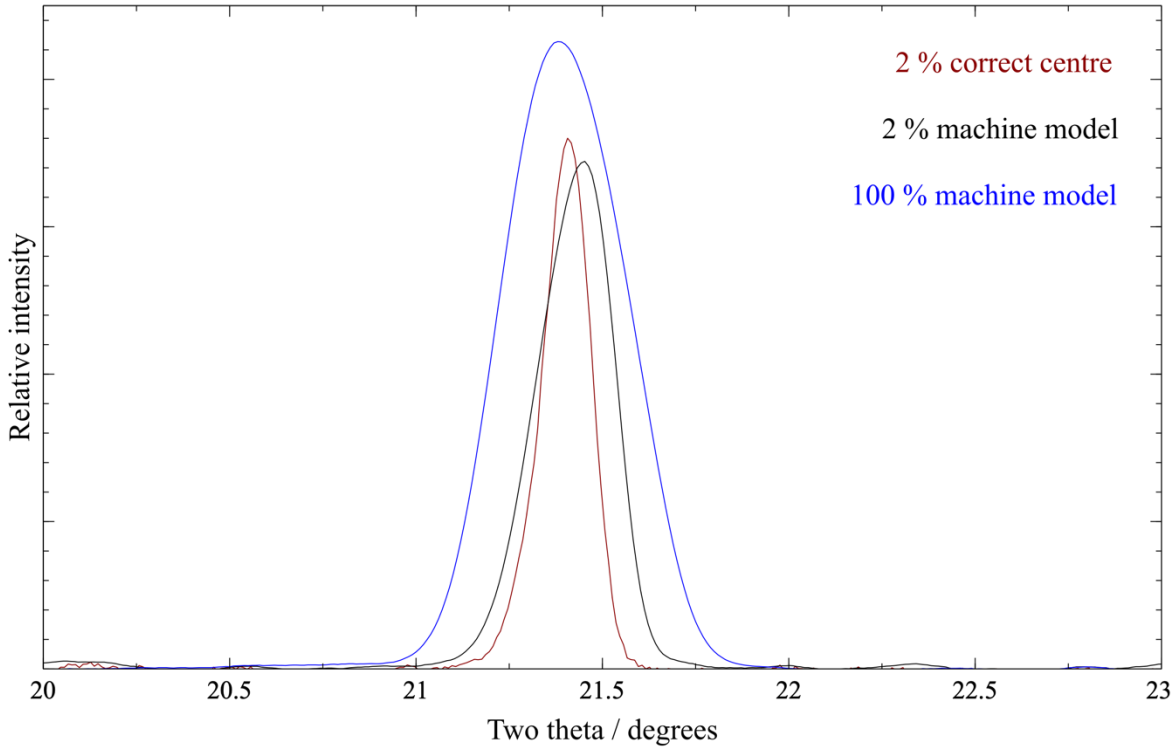


Figure 3.18. A Comparison of the shape of the 1st peak of LaB_6 at 100% (blue), 2% beam divergence with unaltered centring (black), and 2% beam divergence with improved centring (dark red). Background scattering has been removed for clarity.

Diffraction patterns collected at minimum beam divergence (*ca.* 0.2 mrad) resulted in the highest resolution datasets, and so this value was used for subsequent analysis and optimisation.

3.4.5 Optimal detector distance value

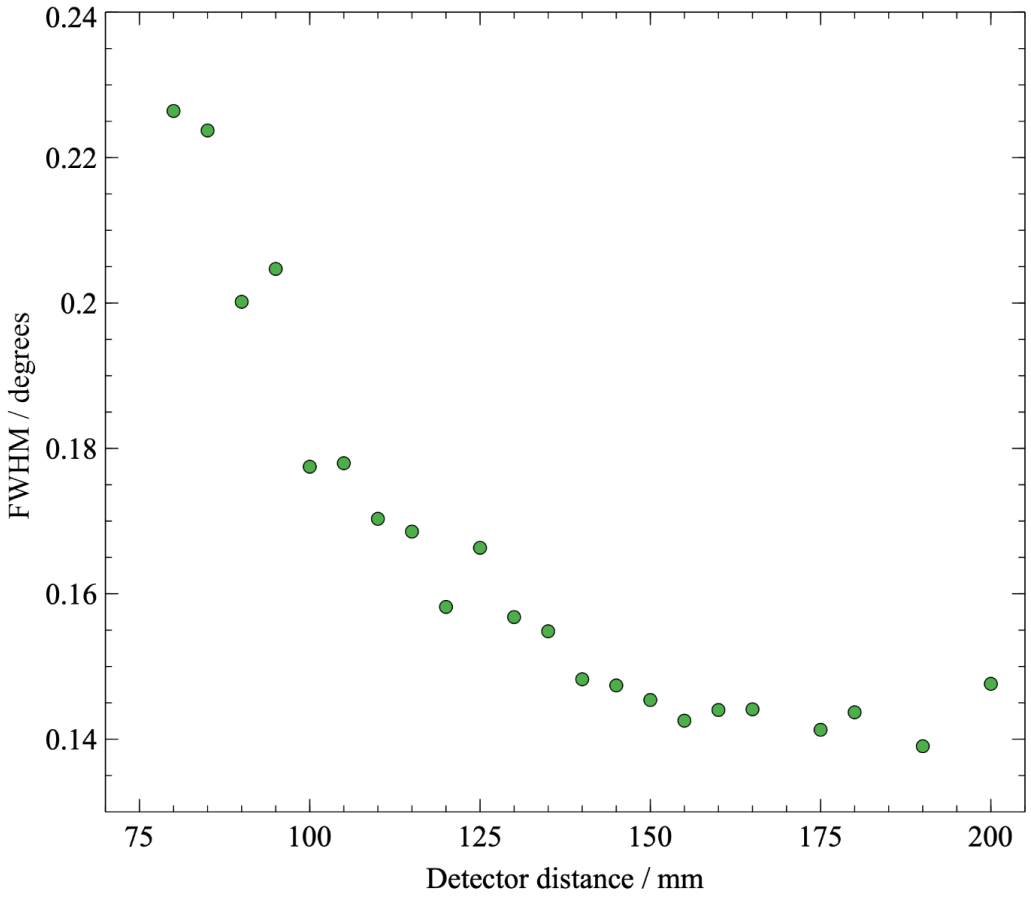


Figure 3.19. The FWHM of 1st Peak of LaB_6 as a function of detector distance, Cu Ka at minimal beam divergence (*ca.* 0.1 mrad).

The optimal detector distance was found to be approximately 160 mm and so this distance was selected for subsequent future data collections. Beyond this distance, there is no significant improvement in FWHM, but there is further loss of peak intensity.

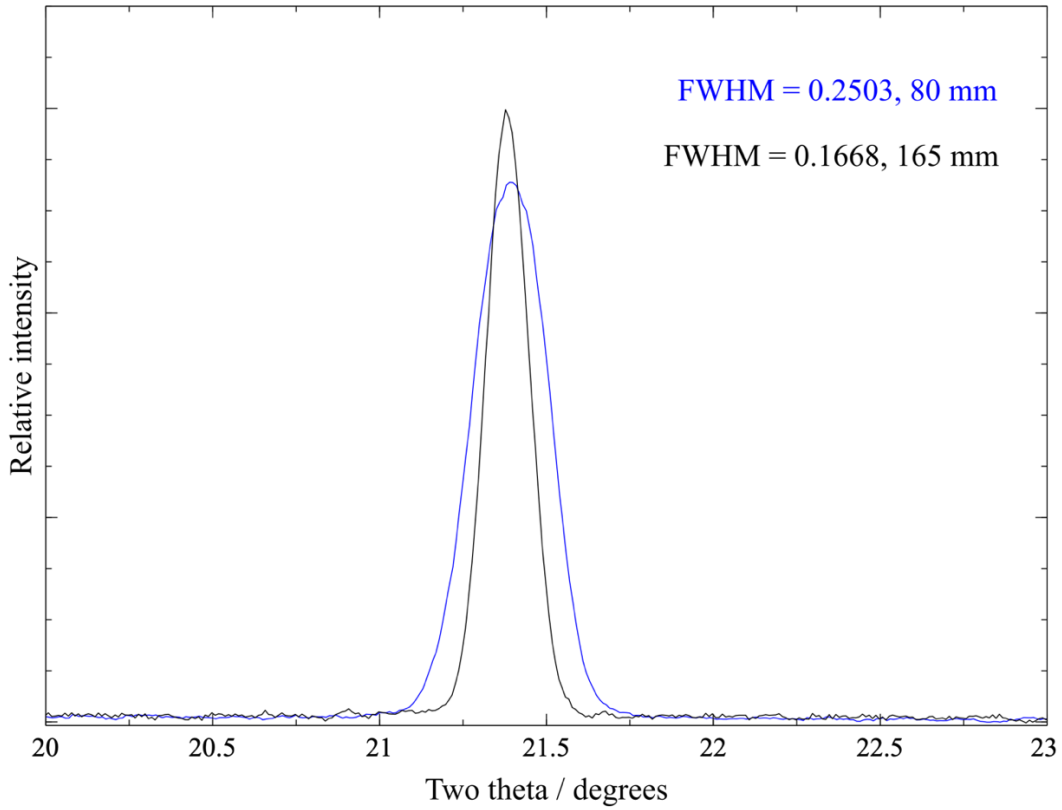


Figure 3.20. A Comparison of 1st peak of LaB_6 width at 2% beam divergence, at 80 and 165 mm detector distances, with improved centring and background removed. Both datasets were scaled to the same maximum peak (not visible within the regions of the axes shown).

3.4.6 Pre and post optimisation diffraction patterns

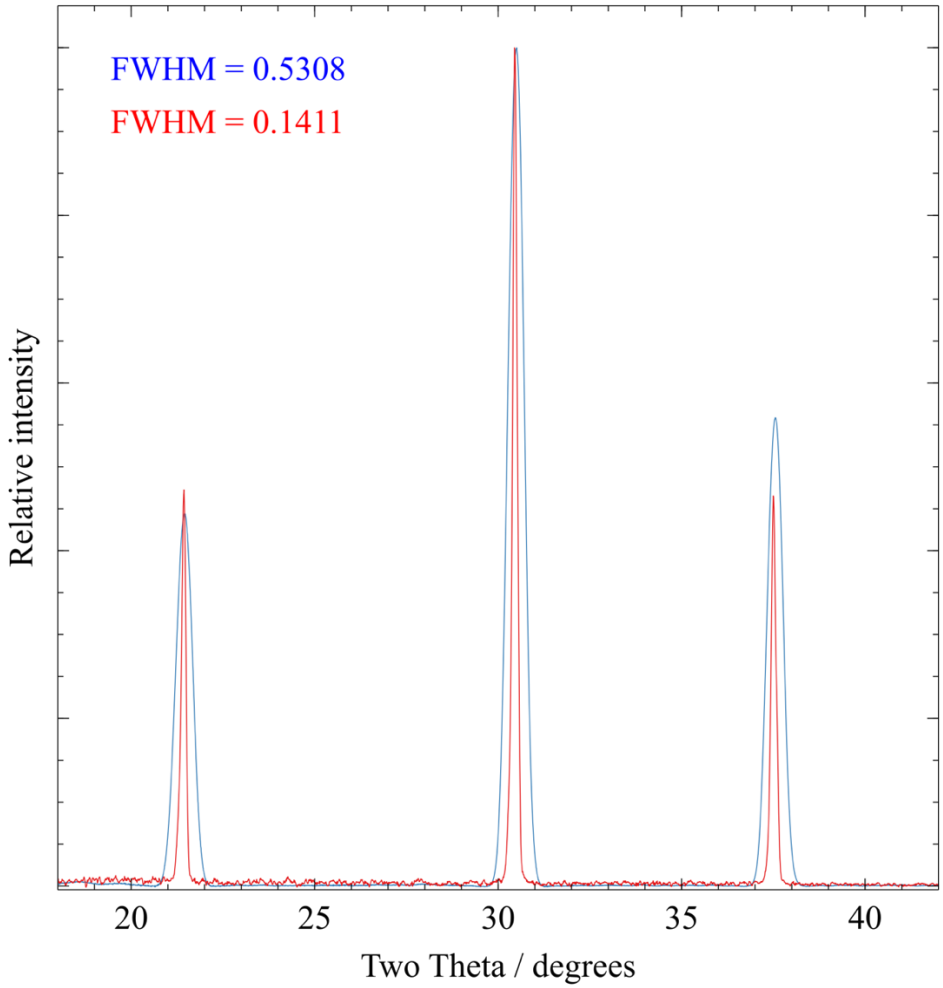


Figure 3.21. A Comparison of integrated PXRD data for NIST 640c SRM LaB_6 with (a) optimal centring, minimum beam divergence and 160 mm detector distance (red); and (b) default centring, maximum beam divergence and 34 mm detector distance (blue). Both data collections were performed in standard phi mode and scaled to the same maximum height after background removal.

The same diffraction data are represented, post-background removal and without normalisation, in Figure 3.22.

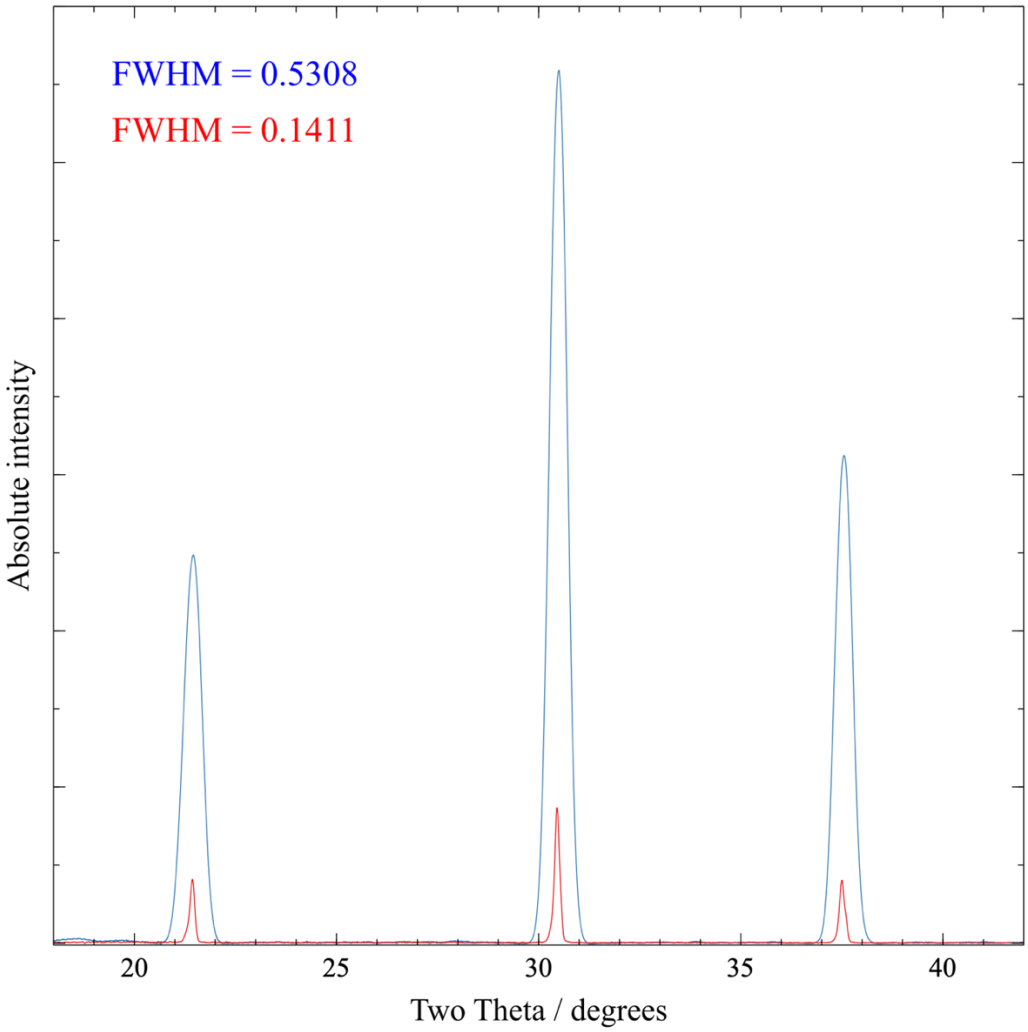


Figure 3.22. A Comparison of absolute peak intensities of integrated PXRD for NIST 640c SRM LaB_6 data collected using optimal (red) and least optimal, initial parameters (blue). Datasets are compared post background removal.

3.5 Discussion

3.5.1 Choice of X-ray source

The Rigaku Synergy is a dual source setup, with the option for Mo or Cu radiation. Equivalent test scans L-glutamic acid were taken at a detector distance of 100 mm, full beam (*ca.* 9.5 mrad) using Mo and Cu radiation, highlighting how a real-world "small molecule" crystal, such as L-glutamic acid, would diffract in this method without any optimisation. It was found from the initial scans (Figure 3.10 and Figure 3.11) that the diffraction patterns collected using the Mo source were significantly worse than equivalent data using Cu radiation, and so Cu was used for all further analysis.

Furthermore, the Mo source on the instrument does not include the ability to reduce the beam divergence, a parameter that was later found to have a significant effect upon diffraction quality, as exemplified for the Cu source. The poor profile quality of the Mo diffraction manifests both as a poor signal-to-background and less well resolved peaks. The Cu diffraction pattern, although is not particularly well resolved with all other settings at their default values, has a significantly better signal-to-background ratio, due to the effect of the λ^3 scattering power.

One of the limitations of using a lab-based single crystal diffractometer for PXRD data collection is the non-monochromatic nature of the beam, which whilst providing excellent incident intensity, contains both Cu $K\alpha_1$ and Cu $K\alpha_2$. With the copper source, (Figure 3.12) the presence of Cu $K\alpha_2$ introduces unwanted additional diffraction features that are particularly evident at higher values of 2θ. Although these can be removed with careful processing of the data but as demonstrated in Chapters 5 and 6, they are not particularly limiting for the range of materials studied in this work.

3.5.2 Choice of standard reference material

Both LaB_6 (NIST 660c) and Si (NIST 640c) were used as possible standard samples, both are highly crystalline, inorganic compounds that are regularly used as reference samples for PXRD. For the optimisation work, it was found that having the first peak at a relatively low value of 2θ was very important for the purposes of investigating the effect of changing parameters upon FWHM, at higher angles the reflections become broadened by the $K\alpha_1$ and $K\alpha_2$ splitting. Having a sharp low angle peak also means that is it possible to capture more of the ring at a greater detector distance, meaning the diffraction ring can be observed with one frame instead of multiple. The Si diffraction pattern (Figure 3.12) has the first peak at *approx*.

28° 2θ whereas LaB_6 (Figure 3.13) has the first reflection at *approx*. 22.5° 2θ, making LaB_6 the better choice for optimisation work.

It should be noted that for the use of APPE, (as outlined in chapter 3) the unit cell of the sample studied does not need to be known; rather only a single, low-angle peak needs to be present. Whilst a NIST material is not essential for the optimisation, it does have the significant benefit of being an established standard specifically used because of its sharp diffraction features.

3.5.3 Ring centring

Correct ring centring is crucial to the accurate optimisation of powder data collection using a single crystal diffractometer. Small variations in centre coordinates lead to large variations in peak widths, causing incorrect conclusions regarding optimal parameters. For this reason, correct ring centring was explored and determined before optimal beam divergence and detector distance values.

The standard calibration of the Rigaku Synergy for single crystal data collection is performed at 34 mm at 100% beam divergence; centre points are then calculated within CrysAlisPro for values away from this, referred to throughout this thesis as the 'machine model'. When data are collected beyond the calibration point, there is no adjustment procedure given to the user to ensure the correct centring. Finding the coordinates 'by-eye' is inefficient, unreliable and can lead to large reductions in data quality output. Each time a global calibration (for collection of single crystal diffraction) is performed as part of instrument maintenance, the coordinates for the powder data collection parameters must be redetermined.

The correct centre point model (at minimal beam divergence) as a function of detector distance was determined by a grid search method (Figures 3.14 – 3.16), measuring the FWHM of the 1st peak of LaB_6 at 100, 160 and 190 mm. The results show that for the best outcome, researchers are advised to take the time to manually determine the actual beam centre coordinates for at least two detector distances to create a reliable model. This was aided by the use of APPE (see Chapter 2), to obtain FWHM values.

To establish an accurate centre coordinate model as a function of beam divergence (at fixed detector distance) the centre coordinates were determined at maximum and minimum beam divergence (*ca.* 9.5 and 0.2 mrad) at a detector distance of 100 mm. This model is used to show the additional improvement of reduced beam divergence and correct centring (Figure 3.17).

3.5.3.1 Inbuilt CrysAlisPro recentring

Within CrysAlisPro, there are two options for semi-automated centre coordinates identification.

a. Utilising the peak hunting feature

Within the 'adjust model' function of the powder graph window in CrysAlisPro, the 'peak hunting' feature (Figure 3.23), can be used to create a list of peaks / d spacings from which centre coordinates are found. Depending on the sensitivity of this feature, dotted rings appear on the identified rings so the user can see visually how the software plots the observed rings. The user can manually adjust the centre point coordinates, with the help of the dotted rings, to find the best centre point or to use the 'auto-adjust' function to alter the centre point. It was found during the study that the in-built peak centring functions were inconsistent, with centre point coordinates changing between datasets of the same parameters. Furthermore, the ability for the software to 'peak hunt' varied with the sensitively level chosen, giving different results each time. These effects were compounded with diffraction patterns collected at further detector distances and with diffraction patterns collected at reduced beam divergence.

Figure 3.23. Powder adjustment tool (above), part of the CrysAlisPro powder diffraction function and the 2D Diffraction rings collected and visualised within CrysAlisPro software post peak hunting.

b. Utilising lattice parameter information

The alternative method explored within this chapter for determining centre coordinates is to use the in-built lattice information function (Figure 3.24) from the powder graph window. It is recommended within CrysAlisPro that this function is performed using a highly crystalline standard, such as LaB_6, and by determining where the diffraction rings match of the observed data. Crystallographic information on commonly used standards is included within the CrysAlisPro software, or a different compound can be used as a standard if known lattice information is provided by the user. The window for accessing this function is shown below in the data reduction tool, and options for editing the lattice information are highlighted in orange.

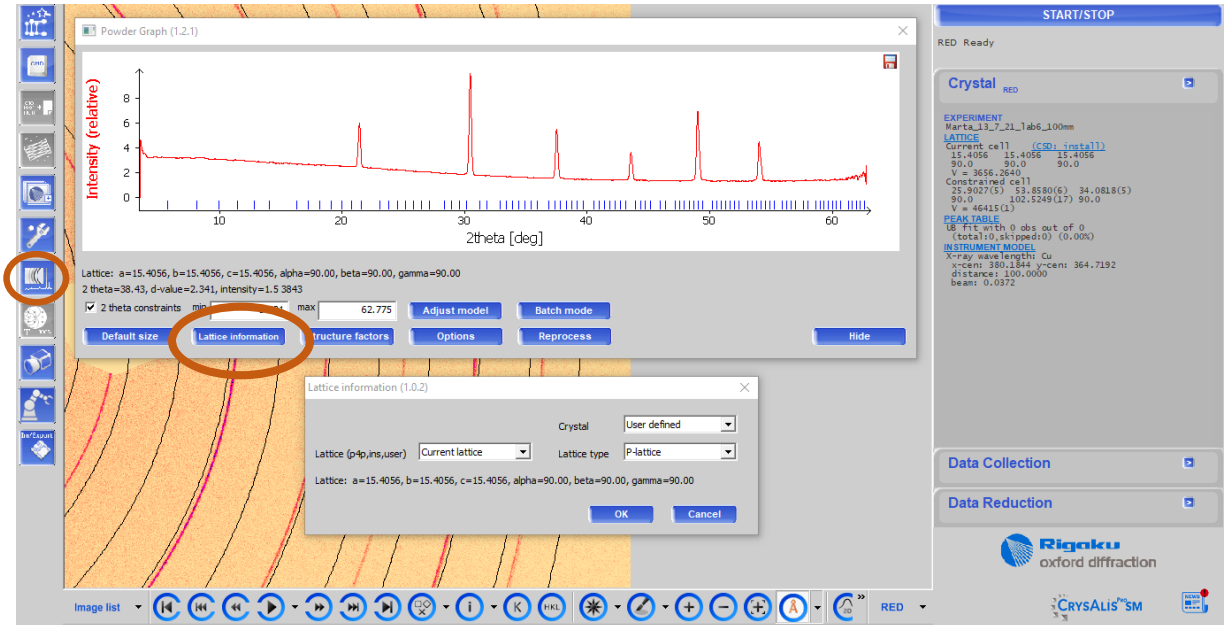


Figure 3.24. CrysAlisPro in-built centring, highlighting the use of standard materials to fit rings to chosen known lattice.

The inbuilt centring was not always found to be a reliable for during this investigation as the outcomes often did not align with the grid search output. A likely cause being from a weaker signal output greater detector distances/ reduced beam divergence. Additionally, this method function requires the use of standard materials that are often expensive and not commonly available to all research areas.

It is therefore recommended that researchers perform a manual grid search (either with the use of APPE or other software of choice) in addition to the use of the in-built centring functions to ensure the ideal starting point for parameter optimisation and data collection. As mentioned previously, the use of APPE does not require knowledge of lattice parameters or even an inorganic standard sample, so any highly crystalline material with a low angle peak can be used for this process.

3.5.4 Optimal beam divergence

The Rigaku Synergy has a beam divergence that can be manually controlled from within CrysAlisPro, ranging in the intensity of the Cu Kα source from 100% to 2%, corresponding to a beam divergence range of approximately 9.5 – 0.2 mrad. By reducing the beam divergence, the FWHM (of the 1st peak of LaB_6) reduces significantly, as shown in Figure 3.17. It is important to recognise that the beam reduction is asymmetric, and this is not pre-modelled within CrysAlisPro; in order to achieve the highest data quality correct centring is essential. With the correct centring, the FWHM of the peak is reduced by approximately 40%. Therefore, minimum beam divergence is optimal and was used for further optimisation of the Rigaku Synergy.

From this setup optimisation, the initial conclusion is that a minimal beam divergence will always produce the optimal data quality. This is an oversimplification as there is a balance between beam divergence, detector distance and detector sensitivity which will vary between setups.

It can be seen from Figure 3.18 that an indicator of incorrect centre coordinates is the presence of asymmetry within the peaks. This is caused by the measurement of the diffraction ring under-representing one portion of the diffraction ring and over-representing the rest, this effect is not even and so causing a hump for each peak. Because peak asymmetry is seen for peaks that have inaccurate ring centres, a possible future addition to APPE would be a skewness measurement to simplify the detection of incorrect centring coordinates. It may be that the peak asymmetry seen in this data is specific to the Rigaku Synergy or possibly specific to the diffractometer within this study. Therefore, for development of this feature, further investigation would be required of other single crystal diffractometer setups.

3.5.5 Optimal detector distance

The Rigaku Synergy has a detector distance range of 34 – 200 mm; most small molecule single-crystal experiments are performed (and so calibrated at) a distance of 34 mm. Although in theory, it is ideal to collect the data further out, to better separate the diffraction rings, in practice the intensity of the peaks drops significantly, (due to increased air scatter) lowering the data quality.

The optimal detector distance for the Synergy was determined using Figure 1.19 as this showed a minima around 160 mm. This plot does however show noise around the optimal region, due in part to the use of minimal beam divergence, causing significant intensity

reduction at greater distances. In addition, the individual data collections were short, using scans of only 5 mins, this allowed the total collection time to be less than 2 hours but is likely the most significant source of noise. Increased noise leads to uncertainty to the exact optimal detector distance, instead the optimal is chosen where the increase in distance does not improve peak resolution. If this method produced significantly more noise for other setups longer scans and repeats of measurements and longer scan times would improve optimal detector distance determination.

Data collection only detector distance optimsation shows an improvement in the peak width with a reduction in FWHM of 33 %. This can be seen in Figure 3.20 with data collection at 80 to 165 mm at minimal beam divergence.

It is important to note that this optimal detector distance is not applicable to all single-crystal diffractometer setups, each detector will vary in pixel density and therefore ability to extract resolved peaks at further distances. It is therefore important for other instruments to be tested with the same methodology as in this study, to ensure the optimal settings are achieved every time. For more intense beams and more sensitive detectors, the optimal detector distance may well be further than 160 mm and limited by the physical maximum detector distance.

3.5.6 Pre and post optimisation

A comparison of the pre and post optimisation parameters (initial SX parameters and parameter determined from this chapter) is given in Figure 3.21 with scaled maximum peak heights, with the same comparison with no scaling given in Figure 3.22. These results demonstrate that by altering the parameters (described throughout this chapter) for the Rigaku Synergy, the FWHM is significantly reduced. In this work, the FWHM of the first peak of a sample of LaB_6 dropped from 0.5308 to 0.1411 (4 d.p.), corresponding to a reduction of approximately 75% relative to SX default settings.

3.5.7 Peak shape

It is worth noting that powder diffraction peaks collected from a single crystal diffractometer and a transmission capillary powder diffractometer differ in shape. Area detectors (most commonly used in single crystal setups) produce symmetric peaks whereas strip detectors (used in dedicated powder diffractometers) tend to produce asymmetric peaks at low angles. This asymmetry (technically referred to as axial divergence) is caused by the finite width of the detector as it crosses a diffraction circle. Where the area detector collects 2D data, this effect is not evident.

3.6 Conclusions and next steps

This chapter has used the tools outlined in the previous chapter to help determine the optimal instrumental parameters for the collection of powder diffraction data using an in-house single crystal diffractometer; specifically, a Rigaku Synergy. Optimisation of radiation source, beam divergence, detector distance and accurate centre determination yielded a large improvement in data quality, as evidenced by an approx. 75 % reduction in the FWHM of peaks from a silicon standard sample. The work in this chapter enables the exploration of other in-house setups (see chapter 4), and establishes a confidence that the diffraction patterns will have the best possible instrumental resolution for structural work.

3.8 References

1 E. A. Kabova, C. D. Blundell and K. Shankland, J. Pharm. Sci., 2018, **107**, 2042–2047.

2 C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. Van De Streek and P. A. Wood, J. Appl. Crystallogr., 2008, **41**, 466–470.

Chapter 4

Optimisation of Rigaku FR-X for SDPD-SX

4.1 Introduction

The focus of this chapter is to apply the understanding gained from the optimisation of the Rigaku Synergy (Chapter 3) to a different configuration, a Rigaku FR-X based at the University of Manchester (UoM). The optimisation of the Synergy single crystal diffractometer (UoR) demonstrated that the powder diffraction data quality can be improved dramatically by systematically determining the optimal configuration parameters. A previous publication (Kabova *et al* 2022) has shown that structure determination of small organic molecules is possible using PXRD data collected on a FR-X single crystal diffractometer, a method termed SDPD-SX (structure determination from powder diffraction using a single crystal diffractometer). It is important to note that in that work, the FR-X setup used altered settings from those of a typical single crystal data collection, but was not optimised for powder diffraction data collection. Instead, the settings of 1 mrad divergence, 150 mm detector distance, and ring centring "by eye" were used as predicted optimal settings. The FR-X is a significantly different arrangement to the Synergy (see Table 4.1), with different parameter ranges and with a significantly more powerful rotating anode source, and so likely alternative considerations for optimisation.[1]

Table 4.1. Comparison of parameter ranges for the Synergy (UoR) and FR-X (UoM)

Feature	**FR-X**	**Synergy**
X-ray source	Rotating anode (2.97 kW)	PhotonJet-S
Radiation wavelength	Cu Kα	Cu Kα / Mo Kα
Approximate beam divergence range / mrad	0.2 – 4.6	0.1 – 9.5
Range of movement for the detector	5° anti-clockwise, full range clockwise	Full range of motion clockwise and anti-clockwise
Modes available	Standard Phi	Standard Phi, Fast Phi, Gandolfi for Powder Gandolfi for SX
Minimum time per scan / s	300	2
Detector	HyPix6000e	HyPix6000e
Detector distance range / mm	34 – 300	34 - 200

4.2 Aims

The work reported in this chapter aims to find the optimal instrumental settings for A Rigaku FR-X single-crystal diffractometer for high-quality powder diffraction data collection using limited amounts of material (< 0.5 mg). The optimisation is performed using the methods and understanding acquired in Chapter 3, with a view to better understanding how powder data quality differs between optimised single-crystal diffractometers.

4.3 Methodology

The experiments were performed using a Rigaku FR-X (UoM), 2.97 kW dual port rotating anode microfocus source (Cu Kα). The diffractometer has a VariMax microfocus variable divergence slit optic, AFC-11 RINC kappa with a HyPix-6000HE detector. Unlike the Synergy, the FR-X is only capable of Standard Phi collection mode. Data visualisation and reduction was performed using CrysAlisPro (version 1.171.41.93a) and peak width analysis was performed using APPE (see chapter 2).

4.3.1 Standard compounds, preparation and mounting

All compounds used in this chapter were mounted by first mixing the powder with a very small amount of Paratone oil, in order to create a 'paste' from which a roughly spherical sample was used. This sphere was then adhered (by the Paratone oil) to a sample holder, composed of a piece of standard '9/125 μm' fibre optic cable (length *approx.* 3 cm), 125 microns in diameter, anchored to a magnetic base. This is consistent with the sample holder used in Kabova *et al* (2022) and slimmer than the fibre described in chapter 3. The sample of mefenamic acid used in this chapter is shown below in Figure 4.1. Further description and development of this procedure is discussed in Chapter 5.

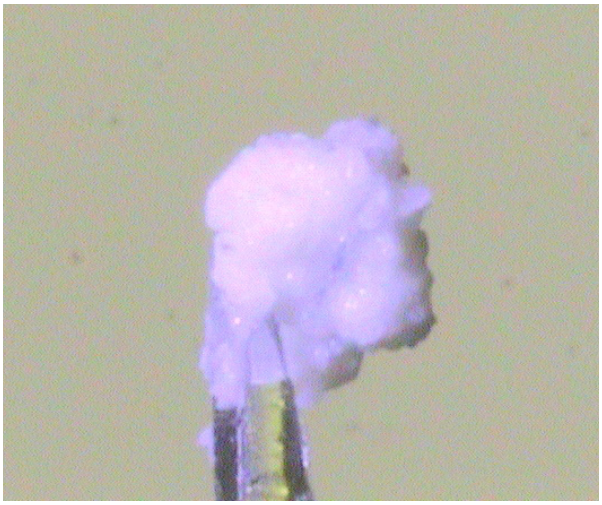


Figure 4.1. Mefenamic acid sample used for parameter optimisation, adhered to optical fibre. Fibre mount used has a 125 μm diameter.

Initial attempts to cool the sample were frustrated by the N_2 gas flow blowing the powder off the sample holder; this could only be mitigated by moving the Cryostream further back from the sample, making the temperature inconsistent. As such, all parameter optimisation scans were carried out at *approx.* ambient temperature.

4.3.2 Parameter optimisation experiments

4.3.2.1 Radiation source

All data collections were performed using Cu radiation.

4.3.2.2 Mode of collection

Due to the physical limitations of the setup as well as to replicate the procedure in Chapter 3, all diffraction patterns were collected in standard phi mode.

4.3.2.3 Choice of standard reference material

To select the ideal standard material, diffraction patterns were collected of LaB_6 (NIST 660c) and mefenamic acid (form I, CSD refcode XYANAC) at 150 mm detector distance, beam divergences of approximately 1 mrad and 0.2 mrad, respectively. Results are shown in Figure 4.2 and Figure 4.3.

4.3.2.4 Beam centring data processing

A ring centring model for varying detector distance was collected at minimum beam divergence (*ca.* 0.2 mrad) at 100, 200 and 298 mm detector distances. Results for the ring centring are shown in Figure 4.4 and Figure 4.5. This procedure was repeated to determine the correct centre for varying beam divergences, and performed using diffraction patterns of mefenamic acid at 1.0 and 4.6 mrad with a detector distance of 200 mm. These results are shown in Figure 4.6 and Figure 4.7.

The correct centre point was found by a grid search method of manually adjusting the centre coordinates, reprocessing within CrysAlisPro and assessing with APPE. Further details of this process are described in Chapter 3. Once determined the ring centre can be applied until the diffractometer is recalibrated.

4.3.2.5 Optimal beam divergence

Diffraction patterns of mefenamic acid were collected in standard phi mode at 200 mm detector distance, at increasing beam divergences in the range of approximately 0.2 – 4.6 mrad at a detector distance of 200 mm. Results are shown in Figure 4.8 and further analysis of the effect of reduced beam divergence are shown in Figure 4.9 and Figure 4.10.

4.3.2.6 Optimal detector distance

Diffraction patterns of mefenamic acid were collected at minimum beam divergence (*approx.* 0.2 mrad) at detector distances of 100 mm and at 10 mm intervals between 150 and 298 mm. Results are shown in Figure 4.11.

4.3.2.7 Overall improvement

The diffraction pattern of mefenamic acid with optimal machine settings (detector distance of 299 mm and *approx.* 0.2 mrad) is shown in Figure 4.12. A comparison of the optimal parameters dataset was compared to the mefenamic acid dataset (previously published by Kabova *et al* 2018), to demonstrate the improvement on this prior work, shown in Figure 4.13.

4.4 Results

This section details the parameter optimisation results, shown as screenshots from CrysAlisPro, 1D diffraction patterns, and/or plots derived from peak widths.

4.4.1 Choice of standard reference material

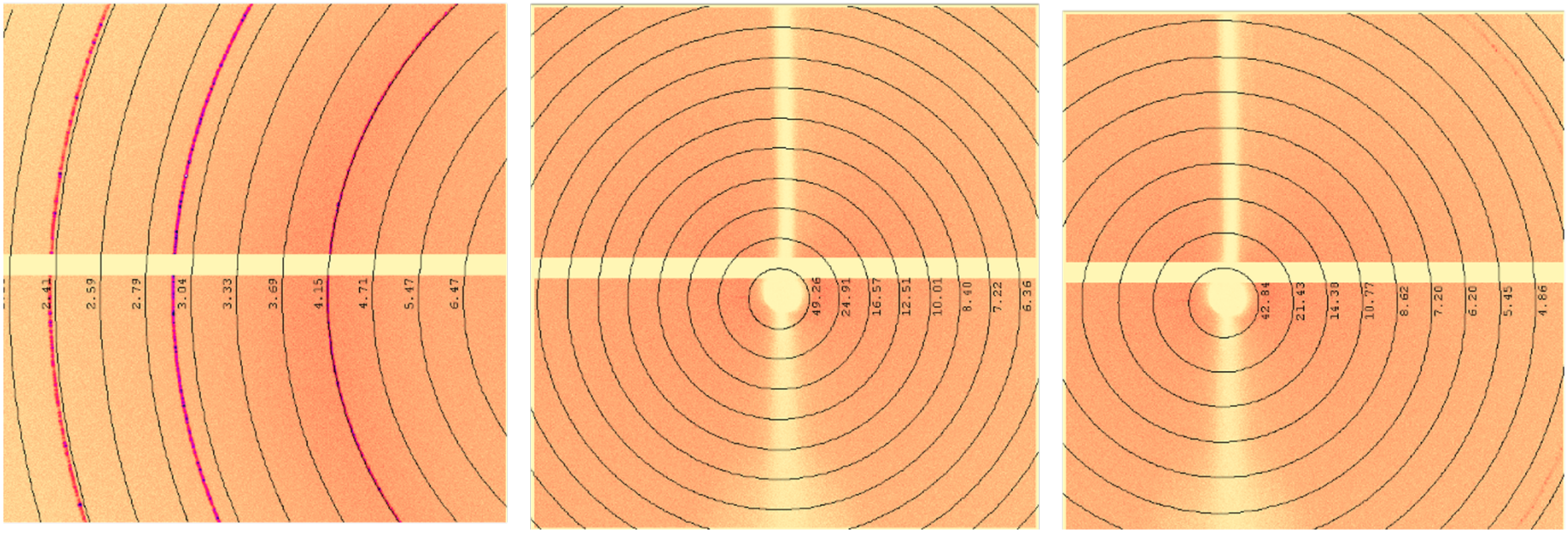


Figure 4.2. Diffraction images of LaB_6 using a detector distance of 150 mm, beam divergence approx. 1 mrad set to scan in the positive and negative directions for the purposes of ring centering. The scans highlight the assymetric range for the setup with a maximum angle of 5° and so do not show the LaB_6 rings clearly on the right image.

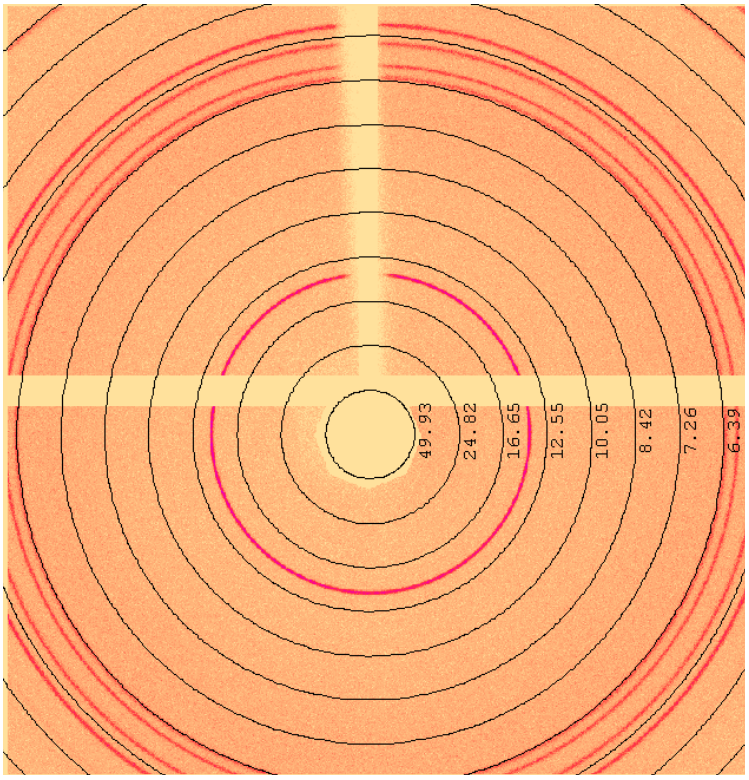


Figure 4.3. 2D image of diffraction from mefenamic acid (form I) with a detector distance 150 mm and beam divergence of *approx.* 0.2 mrad.

Diffraction patterns for LaB_6 were unsuitable for ring centring as it was found that a whole diffraction ring cannot be observed, making ring centring unreliable. Therefore LaB_6 was unsuitable as an optimisation material for the FR-X. In contrast, the first peak of mefenamic acid (form I) can be entirely seen, making this a suitable material for this configuration. All further optimisation was therefore performed using mefenamic acid.

4.4.2 Ring centring analysis

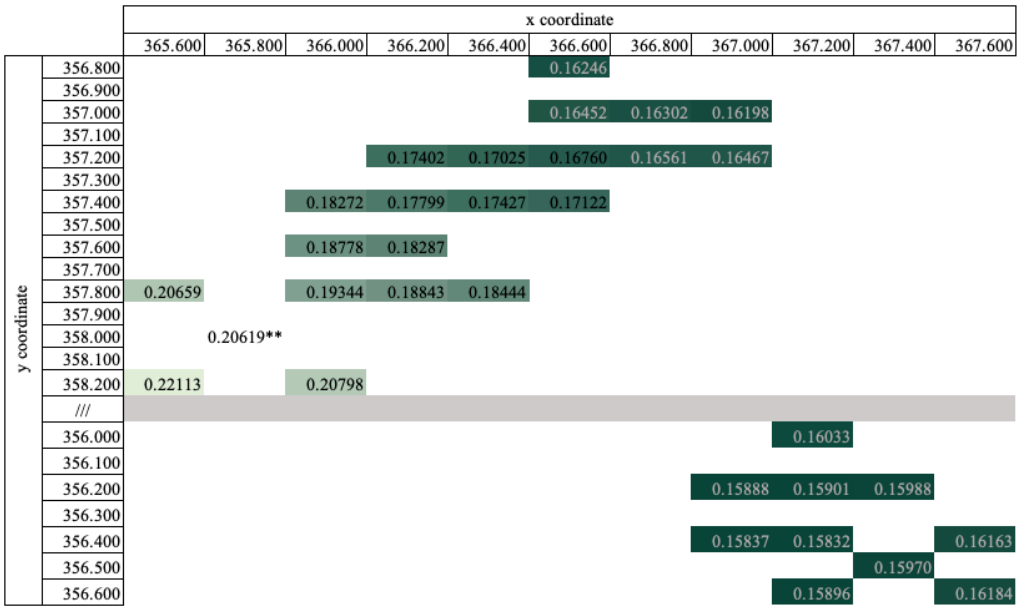
The optimal centre points were determined by measuring the FWHM of the 1st peak of mefenamic acid and performing an iterative grid search. 3-5 coordinates were examined at each iteration until the FWHM no longer improved (*i.e.*, the value no longer decreases).

The data collection for the ring centre models were performed separately, before and after a global recalibration of the diffractometer (as part of routine maintenance). This meant that the centre coordinates for equivalent parameters when separately investigated as a function of detector distance and beam divergence are inconsistent. Once the optimal parameters are determined for a setup, the centre point will need to be redetermined iteratively after each global calibration.

4.4.2.1 Ring centring model for variable beam divergence

		x coordinate			
		364.400	364.600	364.800	365.000
y coordinate	355.200	0.27159	0.27103	0.27161	
	355.300				
	355.400		0.27027**		
	355.500				
	355.600	0.27014	0.26970	0.26973	0.27068
	355.700				
	355.800		0.26986	0.26960	0.27046
	355.900				
	356.000			0.27009	
	356.100				
	356.200			0.27085	0.27191
	356.300				
	356.400			0.27281	

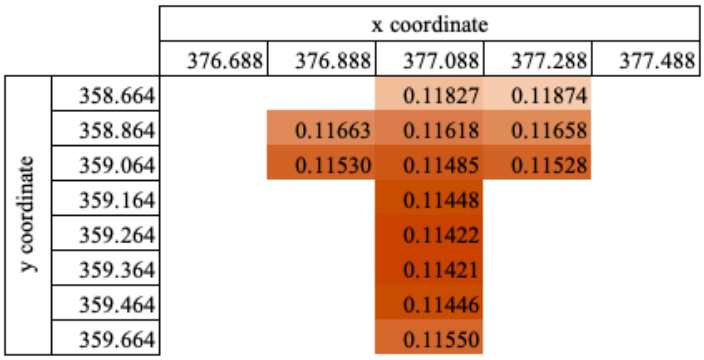
Figure 4.4. A pseudo-contour map of the FWHM of the 1st peak of mefenamic acid at beam divergence of *approx.* 4.6 mrad at 200 mm detector distance, against x and y ring centre coordinates. The darkest shade of blue indicates the lowest FWHM, and thus the new centre coordinates. ** indicates 'by eye' starting point.



y coordinate \ x coordinate	365.600	365.800	366.000	366.200	366.400	366.600	366.800	367.000	367.200	367.400	367.600
356.800						0.16246					
356.900											
357.000						0.16452	0.16302	0.16198			
357.100											
357.200				0.17402	0.17025	0.16760	0.16561	0.16467			
357.300											
357.400			0.18272	0.17799	0.17427	0.17122					
357.500											
357.600			0.18778	0.18287							
357.700											
357.800	0.20659		0.19344	0.18843	0.18444						
357.900											
358.000		0.20619**									
358.100											
358.200	0.22113		0.20798								
///											
356.000									0.16033		
356.100											
356.200								0.15888	0.15901	0.15988	
356.300											
356.400								0.15837	0.15832		0.16163
356.500										0.15970	
356.600									0.15896		0.16184

Figure 4.5. A pseudo-contour map of FWHM of the 1st peak of mefenamic acid at beam divergence of *approx.* 1 mrad at 200 mm detector distance, against x and y ring centre coordinates. The darkest shade of green indicates the lowest FWHM, and thus the new centre coordinates. ** indicates 'by eye' starting point.

4.4.2.2 Ring centring model for variable detector distance



y coordinate \ x coordinate	376.688	376.888	377.088	377.288	377.488
358.664			0.11827	0.11874	
358.864		0.11663	0.11618	0.11658	
359.064		0.11530	0.11485	0.11528	
359.164			0.11448		
359.264			0.11422		
359.364			0.11421		
359.464			0.11446		
359.664			0.11550		

Figure 4.6. A pseudo-contour map of FWHM of the 1st peak of Mefenamic acid at minimum beam divergence of *approx.* 0.2 mrad at 200 mm detector distance, against x and y ring centre coordinates. The darkest shade of orange indicates the lowest FWHM, and thus the new centre coordinates.

		x coordinate				
		380.676	380.776	380.876	380.976	381.076
y coordinate	358.942			0.1986		
	359.142			0.1951		
	359.242			0.1944		
	359.342	0.1940	0.1937	0.1938	0.1941	0.1953
	359.442			0.1935		
	359.542	0.1953		0.1938		0.1955
	359.742			0.1957		

Figure 4.7. A pseudo-contour map of FWHM of the 1st peak of Mefenamic acid at 100 mm detector distance at minimum beam divergence (*ca.* 0.2 mrad) against x and y ring centre coordinates. The darkest shade of grey indicates the lowest FWHM, and thus the optimal centre coordinates.

4.4.3 Optimal beam divergence value

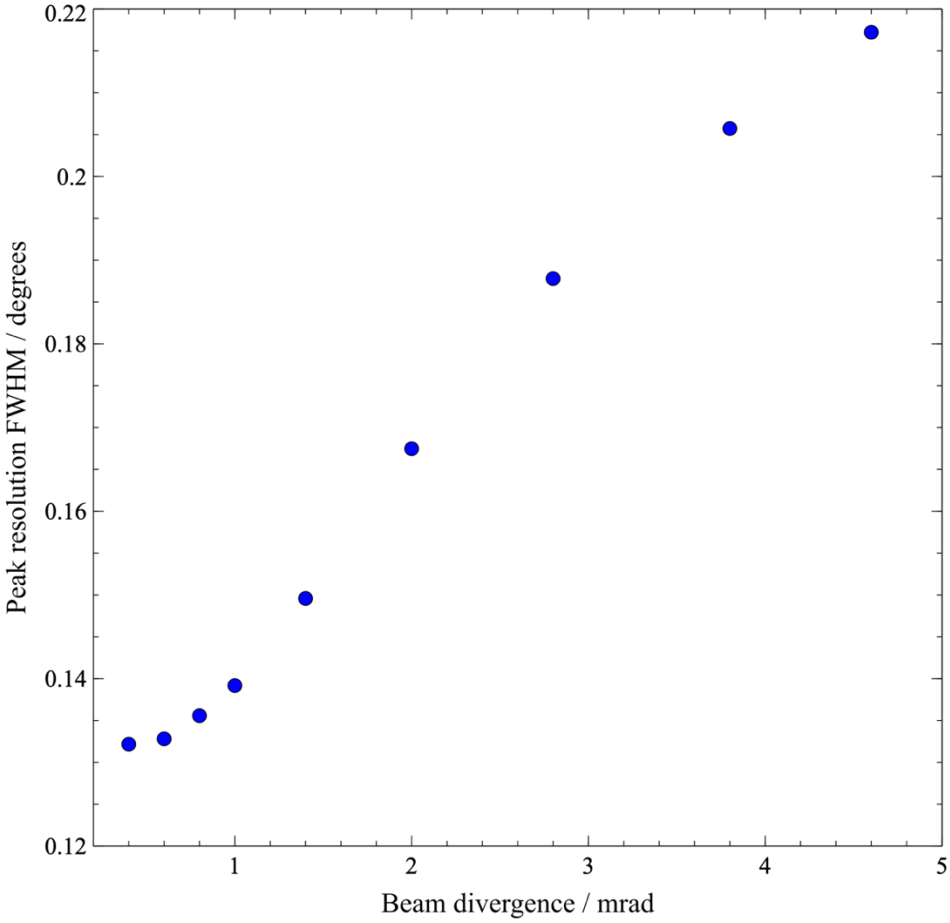


Figure 4.8. FWHM of the 1st peak of mefenamic acid as a function of beam divergence, Cu Kα radiation, detector distance at 200 mm, values are reported using improved centring coordinates.

A comparison (post background removal) of diffraction patterns collected at maximum (*ca.* 4.6 mrad) and minimum (*ca.* 0.2 mrad) beam divergence is given below.

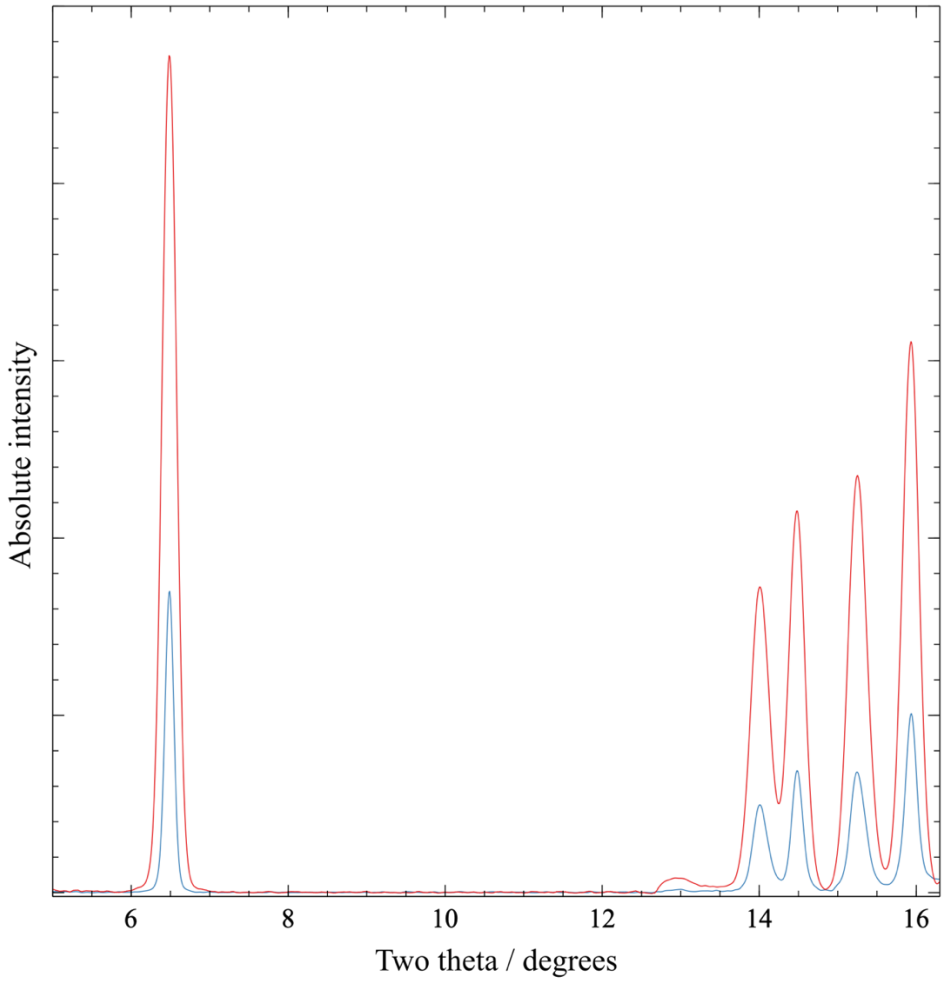


Figure 4.9. A comparison of the integrated PXRD data (unscaled) of mefenamic acid, 0.2 (blue) and 4.6 mrad (red) at 200 mm detector distance, showing the loss in intensity caused by reduced beam divergence.

Reducing the beam divergence significantly affects the absolute intensity of the diffraction peaks measured. A comparison of the same datasets (above) post normalisation is shown below (Figure 4.10) to demonstrate the effect on the FWHM, despite the reduction in intensity at further detector distances.

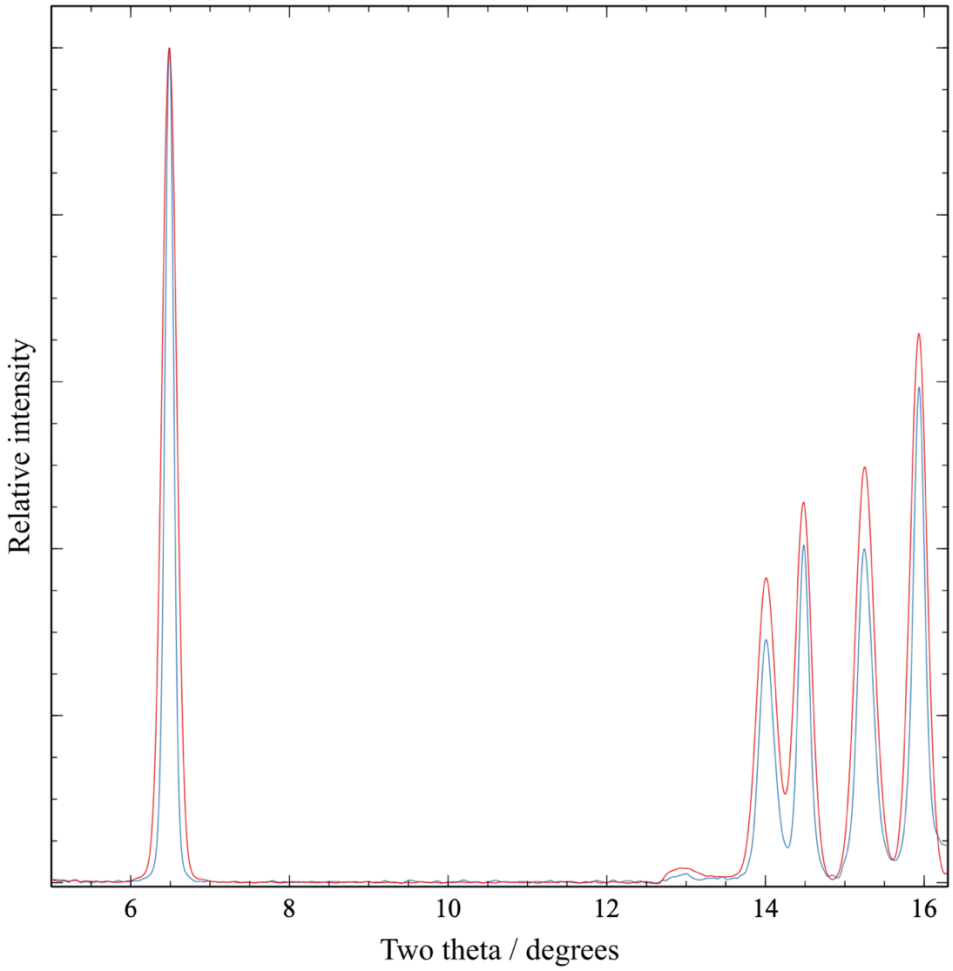


Figure 4.10. A comparison of the integrated PXRD data (normalised to maximum peak height) of mefenamic acid (form I), 0.2 (blue) and 4.6 mrad (red) at 200 mm detector distance.

The minimum beam divergence of *approx.* 0.2 mrad was used for further optimisation.

4.4.4 Optimal detector distance value

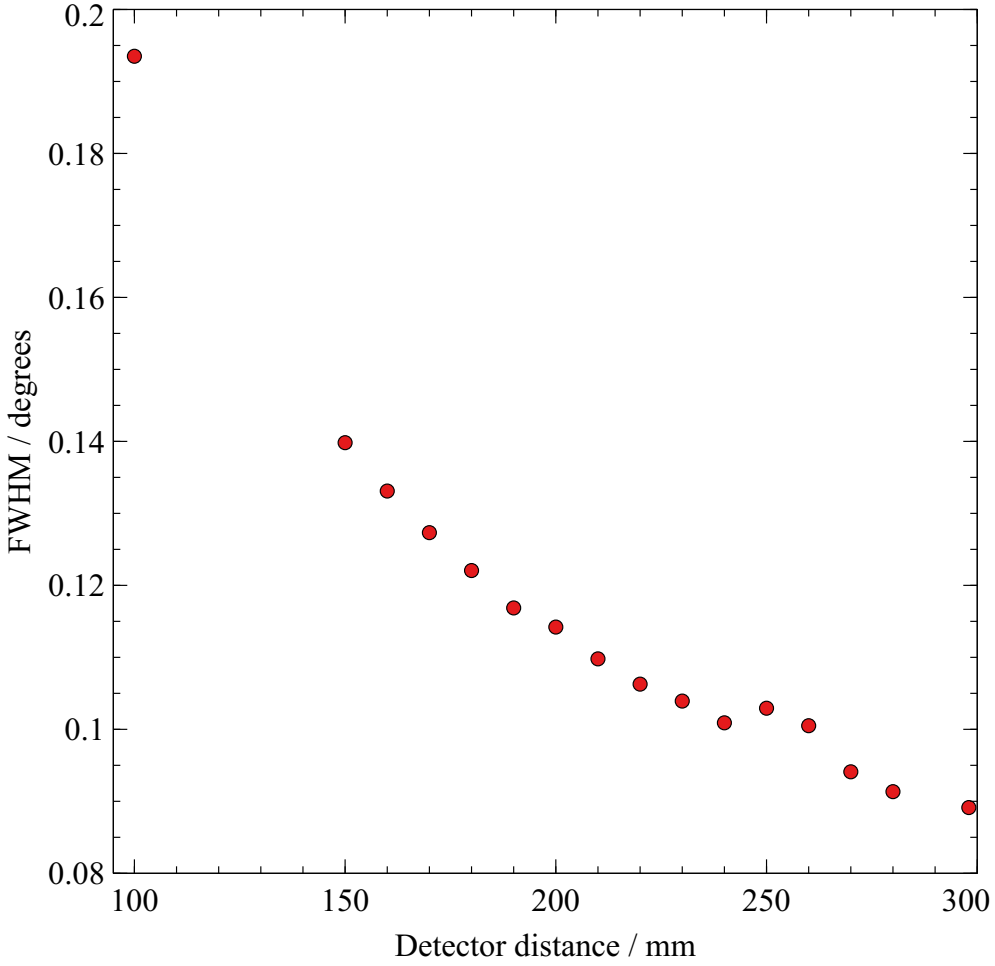


Figure 4.11. The FWHM of the 1st peak of mefenamic acid as a function of detector distance, at minimal beam divergence (*ca.* 0.2 mrad).

The further the detector distance, the lower the FWHM, making the optimal detector distance 299 mm (*i.e.*, at a physical maximum for the configuration). This effect tapers off at around 250 mm, where the difference of 50 mm only improves the FWHM by 0.05°.

The diffraction pattern of mefenamic acid using optimal parameters (correct centring, 0.2 mrad beam divergence and detector distance of 299 mm) is shown below in Figure 4.12.

4.4.5 Pre and post optimisation diffraction patterns

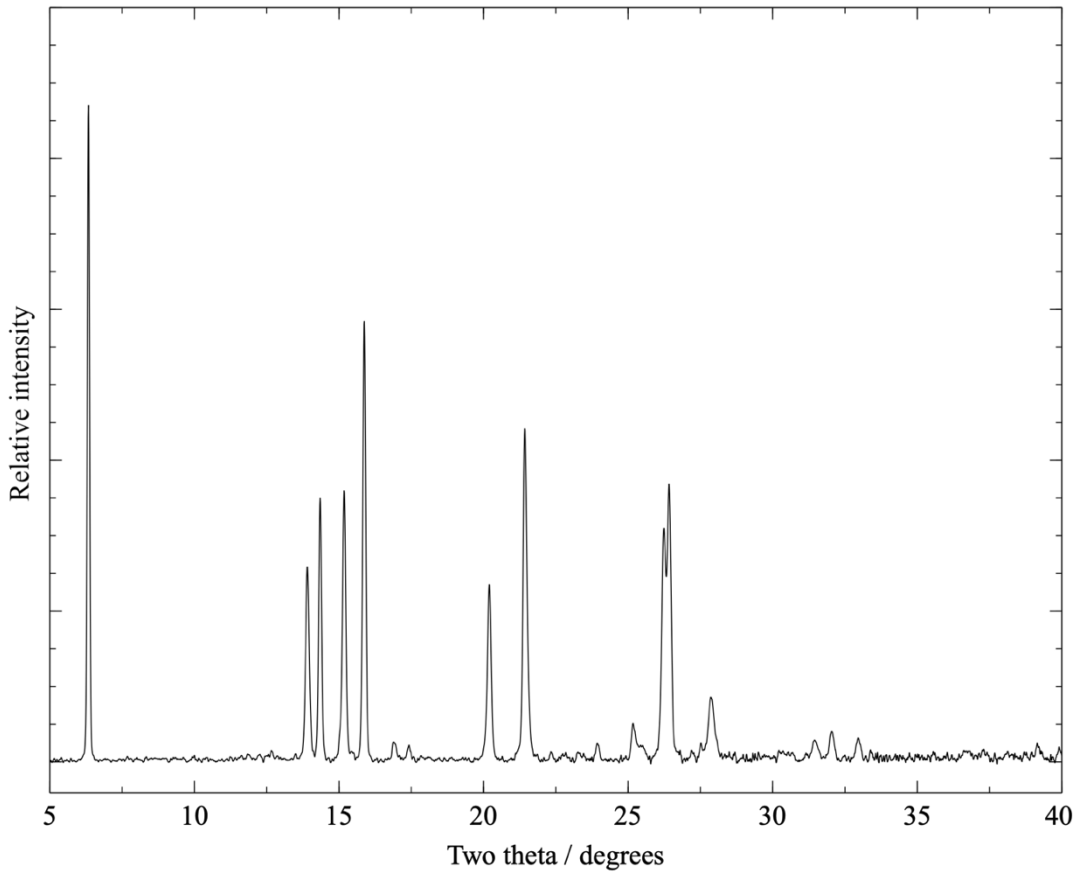


Figure 4.12. Mefenamic acid (form I) powder diffraction data collected using optimal parameters, detector distance 299 mm and minimal beam divergence (*approx.* 0.2 mrad). The data collection was performed using a limited 2θ range, causing data quality reduction beyond 30°.

An overlay of the optimal parameters and the previously used machine settings (from previous publication) for SDPD-SX using the FR-X is shown below in Figure 4.13.

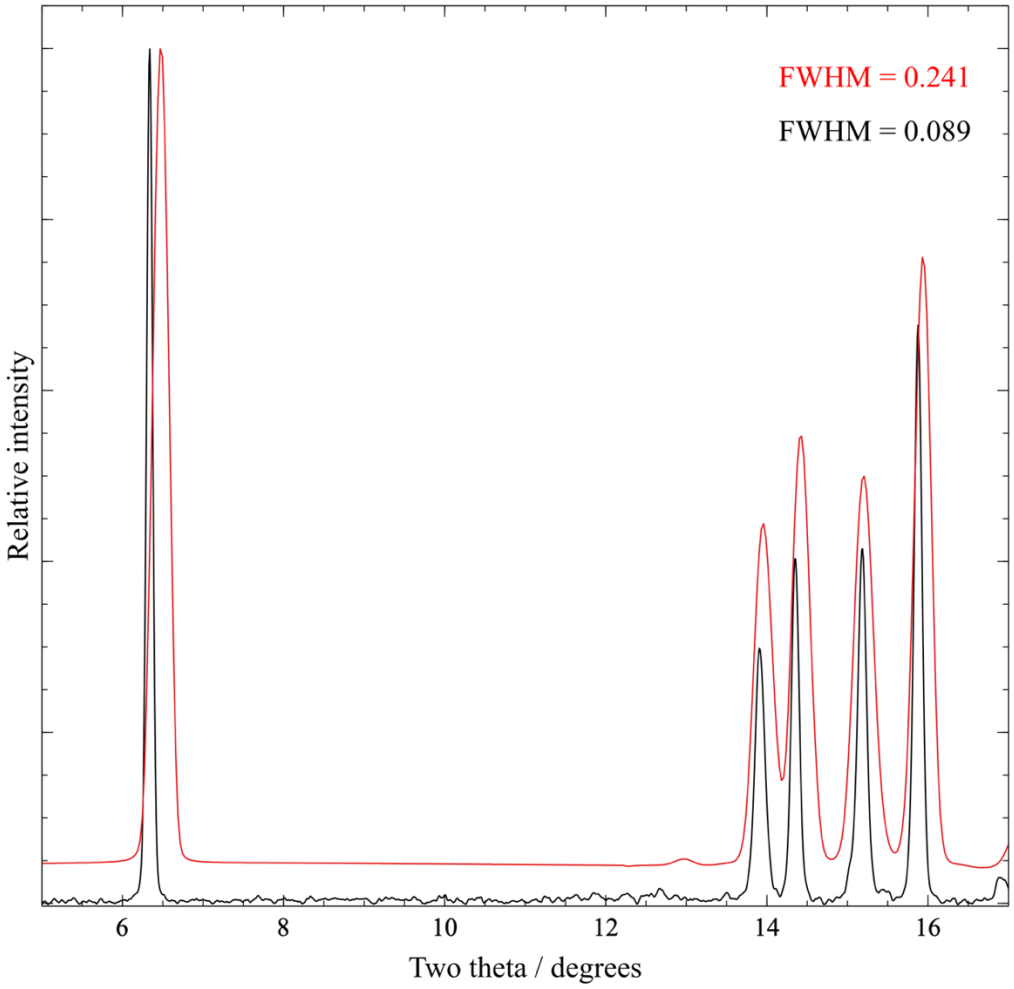


Figure 4.13. A comparison of mefenamic acid data collections on the FR-X. Data from the unoptimized setup (1 mrad divergence, detector distance = 150 mm and centering 'by eye'[1]) are shown in red, whilst data from the optimised setup (0.2 mrad divergence, detector distance = 250 mm and centering by grid search) are shown in black. The offset between the two patterns is attributable to thermal expansion of the structure due to the different data collection temperatures used.

4.5 Discussion

4.5.1 Choice of standard reference material

It was previously recommended in Chapter 3 that LaB_6 (NIST 660c) was an ideal material for ring centring due to its highly crystalline nature as well as its use as a standard material for calibration in other research areas. Due to the mechanical constraints of the FR-X, the detector is unable to scan further than 5 degrees in the anti-clockwise direction. This limits the range of frames that can be collected for powder X-ray diffraction. The outcome of this is a limit on the type of material that can be used as the standard reference material (SRM).

To perform the optimisation of parameters a whole ring should be visible to ensure correct centring. For the FR-X setup, it can be seen in Figure 4.2 that even at 150 mm (the midpoint of the detector distance range) the diffraction rings from LaB_6 are unsuitable as the 1st diffraction ring cannot be seen fully, making centring unreliable and therefore unsuitable for further machine configuration optimisation.

To compensate for the limited scanning range, a new standard was needed that was 1) highly crystalline and 2) had a single reflection at low angle. Mefenamic acid (form I) was chosen (Figure 4.3) because it is low cost, has a peak at approximately 6.3° 2θ, and shows minimal preferred orientation after light grinding.

4.5.2 Ring centring analysis

A detailed discussion on ring centring, particularly within the CrysAlisPro software package is given in Chapter 3. As previously demonstrated, accurate ring centring is vital for correct optimal parameters determination and ultimately optimal diffraction pattern output. CrysAlisPro creates the 1D powder pattern by integrating the rings according to a given centre point. Figure 4.14 demonstrates how drastically the data quality is affected when an inaccurate centre point is used. In this case the centre point is derived from the machine model, created from a calibration at single crystal settings (full beam divergence at a detector distance of 34 mm) but is then applied to settings far away from the calibration (*approx.* 0.2 mrad and 200 mm detector distance). When the centre model is inaccurate, this causes the output peaks to be broadened or even split.

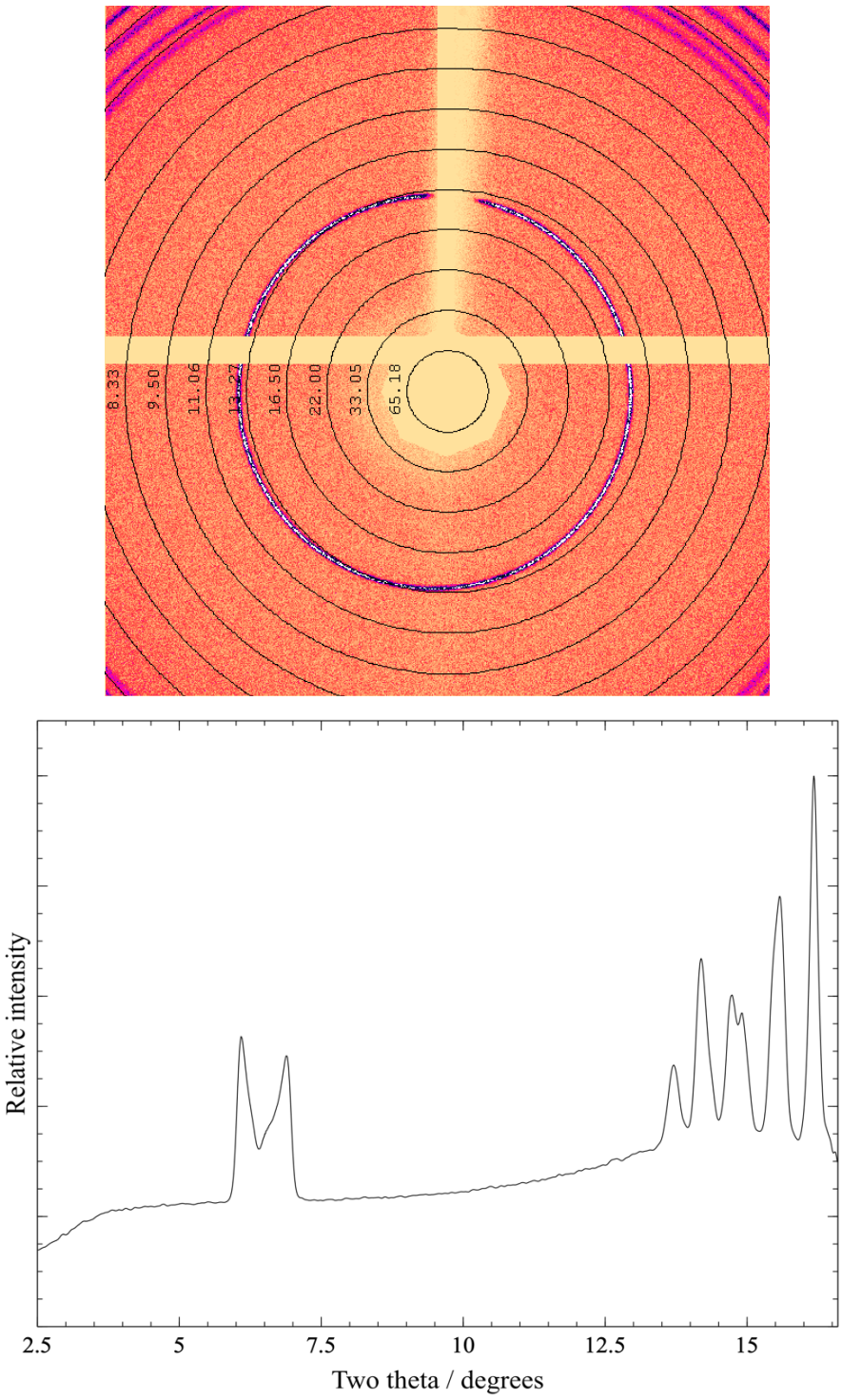


Figure 4.14. Mefenamic acid (form I) at detector distance 200 mm, beam divergence approximately 0.2 mrad with machine centre coordinates (from single crystal calibration). CrysAlisPro window (above) showing one low-angle ring and corresponding integrated 1D diffraction pattern (below).

As previously discussed, the correct centre point model (at 200 mm detector distance) as a function of beam divergence was determined by a grid search method (Figure 4.4 and Figure 4.5), measuring the FWHM of the 1st peak of LaB_6 at *approx.* 4.6 and 1 mrad. The FWHM is determined using the APPE program outlined in Chapter 2. These models were applied while optimising the detector distance and the beam divergence to ensure correct conclusions can

be drawn. This methodology was then repeated to find the correct centre model as a function of detector distance, at minimal beam divergence (*approx.* 0.2 mrad); measuring the 1st peak of LaB_6 at 100 and 200 mm (Figure 4.6 and Figure 4.7).

4.5.2.1 Centring 'by eye'

For different materials and machine settings the ability to obtain the correct centre 'by eye' will vary in success. Below (Figure 4.15) shows what is seen by the user when attempting to centre the rings. Image A shows the displacement of the centre after the single crystal calibration settings are used and image B shows the rings after being centered by eye. Lastly, image C shows the diffraction rings with respect to the CrysAlisPro dotted rings when the correct centre coordinates are used. Figure 4.15 is shown to demonstrate that although subtle, centring by eye is suboptimal and subjective.

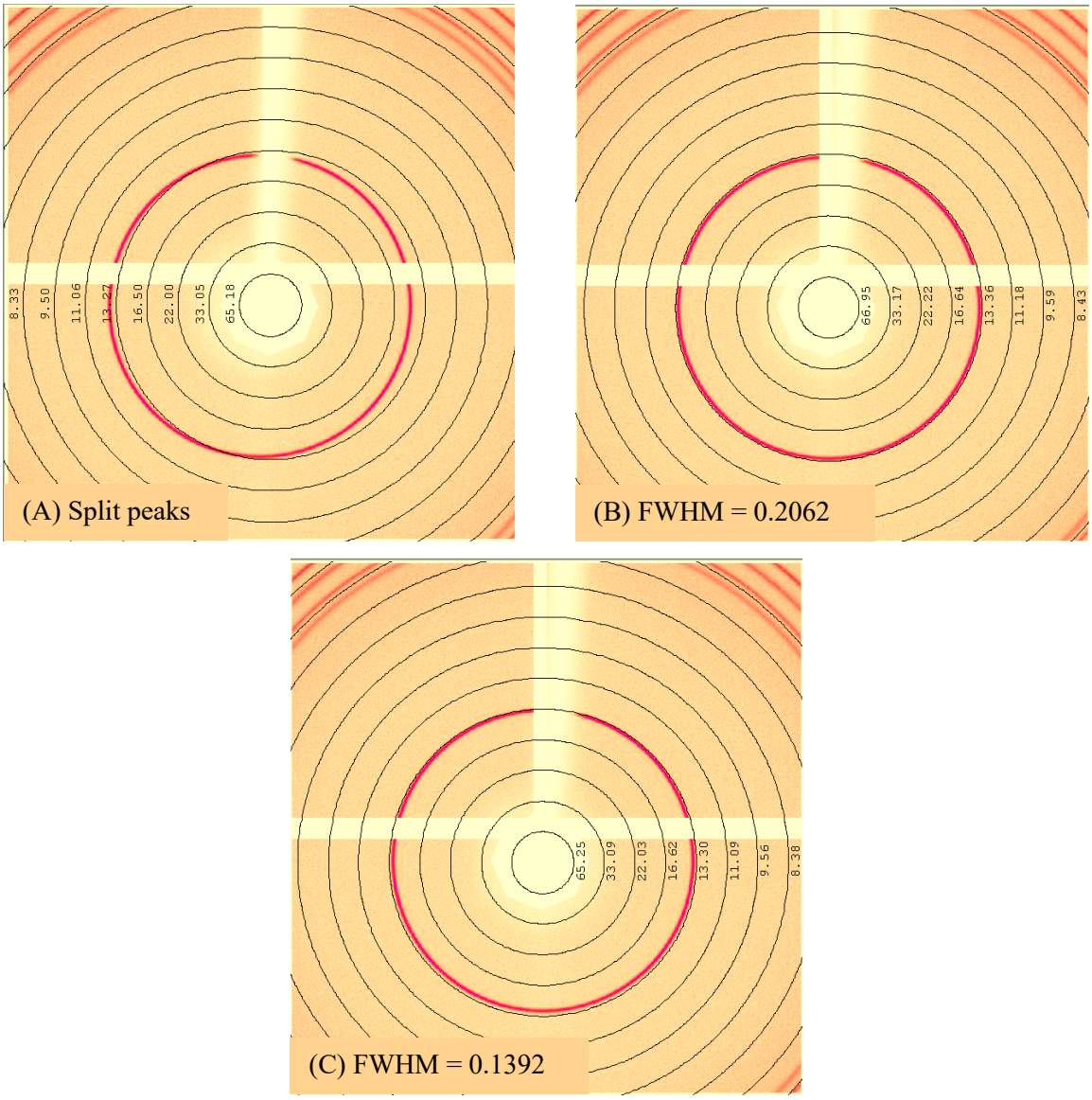


Figure 4.15. CrysAlisPro visual centring for mefenamic acid, beam divergence *approx.* 1 mrad at 200 mm detector distance. Machine model (A), correcting by eye (B) and grid search center (C).

4.5.2.2 Semi-automated ring centring

As discussed in Chapter 3, CrysAlisPro software allows the user to collect a diffraction pattern of a known (or unknown) standard and to select or input a unit cell to find the peaks according to the calculated ring positions. Since mefenamic acid is triclinic, this approach is not best suited for this material. Furthermore, ring centring also assumes sharp, intense peaks, which are expected when using known standards (such as LaB_6 and Si) but the same cannot be guaranteed when using other materials. As mentioned in the previous chapter, the APPE program does not require additional information about the crystal structure, and as such, any material with a single low angle sharp peak can be used as a standard.

It was found that by attempting to refine the instrument model, CrysAlisPro altered the physical coordinates of the detector (shown below in Figure 4.16) changing where the software places the detector in relation to the beam source. The result of this is the creation of separate islands within CrysAlisPro (version 171.41.93a), where different coordinates produced the sharpest peaks.

The purpose of the centring is to model the correct coordinates with respect to detector distance such that they can be inputted after each data collection. It is therefore not advised to attempt to improve the ring centring by adjusting the instrument model.

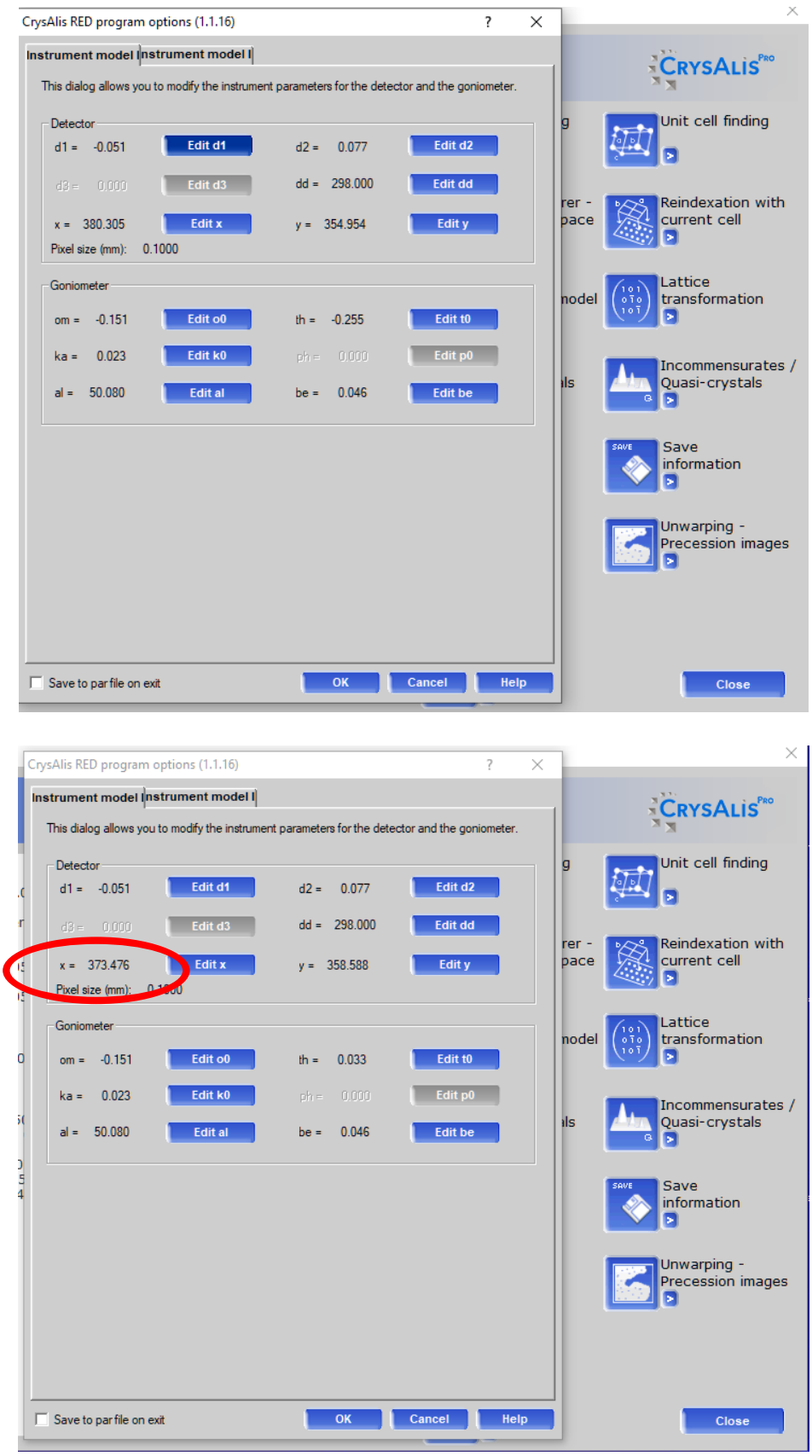


Figure 4.16. The instrument model window, found within the 'RED' program options, unrefined model (above) and refined model (below).

In addition, it was found that the FR-X single crystal calibration model did not change when the beam divergence was reduced. The effect of asymmetric beam reduction is far greater than that found with the Synergy, and so a comparison of machine model / correct centring could not be calculated for subsequent analysis.

4.5.3 Optimal beam divergence

The reduction of the beam, as with the Synergy, is asymmetric, causing both setups to alter in x, y coordinates when the beam divergence is reduced. The beam position was found (by removing the beam stop and using a fluorescent sample to measure the X-ray beam) that this effect is significant enough to move the beam centre out of the crosshairs seen on the camera used to align the sample. Ultimately this means using a larger sphere of compound for data collection to obtain a high-quality diffraction pattern to ensure the sample is being irradiated.

Figure 4.8 demonstrates that the lower the beam divergence, the lower the FWHM of the peaks is obtained, producing a higher resolution dataset, however this effect begins to diminish below *approx.* 1 mrad. It is therefore recommended that experiments are performed at minimal beam divergence where possible, with higher beam divergence for weaker diffracting samples. Figure 4.9 shows the comparison of using a reduced beam divergence, with the impact on the peak intensity being significant. When these outcomes are overlayed, however, (see Figure 4.10) the impact on data quality can be seen more clearly.

The beam divergence for the Rigaku FR-X is controlled externally from CrysAlisPro and so when the beam divergence is reduced, CrysAlisPro is not promped to change the machine model coordinated to compensate for this. The result is a large change from the correct centre coordinates, shown in Figure 4.14 as significant peak splitting. The FWHM of peaks at minimum beam divergence, unlike the Synergy, cannot be compared before and after the correct coordinates are determined. This is important to consider as many single crystal diffractometers are not purchased with in-built variable divergence and so may use external beam reduction.

4.5.4 Optimal detector distance

The FR-X has a detector distance range of 34 – 299 mm, a much greater range than the Synergy (34 – 200 mm). The optimal detector distance was found to be at 299 mm (Figure 4.11), the furthest point available, meaning that in principle, further resolution gains might be had beyond this limit. For practical purposes however, it may be more prudent to collect PXRD data on the FR-X with the detector closer in (for example at 250 mm) in order to gain diffracted

intensity at the cost of a small amount of resolution. This contrasts with the data collections on the Synergy which showed a defined minimum at approximately 160 mm (see Chapter 3). The cause of the difference in optimal detector distance is not obvious, but likey to be a combination of differences in beam intensity and beam geometry, due to the significantly different X-ray sources.

4.5.5 Pre and post parameter optimisation

Mefenamic acid (form I) was one of the 13 compounds used in the previous publication, chosen as a small organic API with 12 DoF in the unit cell. The final diffraction output after parameter optimisation, collected to 40° 2θ is shown in Figure 4.12.

Kabova *et al* (2022) demonstrated the use of SDPD-SX for use in the structure determination of small volumes of polycrystalline material, having successfully solved a variety of compounds from data collections on the FR-X. It should be noted that during the previous publication, the data were collected at a beam divergence of 1 mrad and a detector distance of 150 mm, resulting in a FWHM of the 1st peak of mefenamic acid being 0.241°. Figure 4.13 shows a comparison of the publication data and the collected post instrument optimisation, visually highlighting the significant improvement in data quality achieved from careful optimisation and centring. After parameter optimisation there is a 63% reduction in the FWHM, demonstrating the importance of optimisating a single crystal setup, even in the case of advanced configurations such as the FR-X. [1]

Crucially, with optimised parameters, the mefenamic acid diffraction pattern (with a FWHM of 0.089° for the first peak) has a resolution comparable to that of data from a dedicated transmission capillary diffraction instrument. The importance of this result cannot be understated, especially as SDPD-SX requires 1-5 % of the material.

4.5.6 Effect of machine differences between Synergy and FR-X

There are many differences between the FR-X and the Synergy, several of which are explored above, with the range of motion available and the power of the beam source being the most significant. These differences affect how samples are prepared and how the instruments are used in practice for optimal results.

4.5.6.1 Speed of collection and sample preparation

The Synergy is able to collect much shorter scans than the FR-X becuase although the sample rotates at the same rate, the FR-X requires the goniometer to make at least one full revolution for each scan. The impact of this is shown in Table 4.2.

Table 4.2. Comparison of time required for pre and full experiments

	Synergy	**FR-X**
Pre experiment scan time / s	50 – 150 (short times available)	300
Time to complete a pre-experiment / mins	5 - 10	30
Full experiment scan time / s	1200 - 3000	300 - 500
Time for a full experiment (*ca.* 60-70° span of 2θ) / hrs	2 – 6	1-2

The 'pre-experiment' performed with the Synergy saves considerable time compared to that of the FR-X, where a small range can be tested within a short time frame (*i.e.*, 50 s/ scan). The motor within the FR-X is limited to 300 s / scan so the time saved between a 'pre-experiment' and a full data collection would be minimal.

This changes the approach to data collection for these two configurations; more time is spent making the ideal sample when using the Synergy and multiple 5 min scans can prevent the repetition of long collections.

For a full experiment, shorter scans can be performed to collect high quality diffraction patterns with the FR-X. Moreover, the FR-X can effectively only scan in the 'positive' direction, reducing the time of a full collection further. It is possible to scan only in the 'positive' direction with the Synergy; however this is best avoided to improve ring averaging, reducing effects of preffered orientation.

4.5.6.2 Modes of collection

Due to the limited range and speed of motion, the FR-X is incapable of performing scanning modes other than the standard phi. For samples with large amounts of preferred orientation the only options are to attempt structure solution with the existing data or to grind the sample (at the risk of reducing the crystallinity) and remount.

4.6 Conclusions and next steps

The optimisation methodology outlined in Chapter 3 has been applied in this chapter to determine the optimal parameters of a Rigaku FR-X (University of Manchester) for powder diffraction data collection (< 0.5 mg). Despite the Synergy and FR-X both being in-house single crystal diffractometers, there are many practical differences that optimisation, as well as the range of data that can be collected. The work in this chapter determined the optimal parameters to be: Cu radiation, minimal beam divergence (specifically between 1 and 0.2 mrad), 250 - 300 mm detector distance and accurate diffraction circle centring prior to data integration. The previous use of NIST standard materials for parameter optimisation was not possible (due to instrument constraints) and so optimisation was performed using a sharply diffracting sample of mefenamic acid.

Kabova *et al* (2022) have previously demonstrated that the FR-X is able to successfully collect powder diffraction data capable of use for structure solution, even with minimal parameter optimisation. This likely reflects the very “high-end” specification of this particular instrument e.g. rotating anode source. For the more general adoption of SDPD-SX, the effectiveness of using of a more widely accessible instrument configuration is important. This is addressed in Chapter 5.

4.7 References

1 E. A. Kabova, C. D. Blundell, C. A. Muryn, G. F. S. Whitehead, I. J. Vitorica-Yrezabal, M. J. Ross and K. Shankland, CrystEngComm, 2022, **24**, 4337–4340.

Chapter 5

Application of SDPD-SX to known crystal structures using optimised Rigaku Synergy

5.1 Introduction

Previous chapters have shown how SX instrumentation can be optimised to collect high-quality PXRD data. This chapter focusses on using such data, collected on the Synergy diffractometer, to attempt the solution of several known crystal structures, with a view to establishing the efficacy of the approach.

5.1.1 In-house powder diffraction data collection methodologies

Previously, Kabova *et al* (2018) investigated the difference between laboratory powder data collection methodologies: reflection flat plate, transmission flat plate, and transmission capillary. This was performed to establish the difference in data quality based on the methods available for in-house data collection, and subsequently the effect that this quality had on the structure solution.[1]

The study was performed by the data collection and structure solution of known APIs and explored different measures of diffraction data quality: peak widths, quality of structure solution relative to known structure, rate of success of simulated annealing solutions, and the 2θ range (the real-space resolution) that was used for the simulated annealing runs within DASH.

The general limitations of SDPD have been outlined in Chapter 1. In terms of SDPD, the major issues are: the 1D nature of the PXRD pattern with consequent accidental reflection overlap, and the fall off in scattered intensity with increasing 2θ. Some instrument-specific issues are also relevant *e.g.*, for transmission capillary work, a few milligrams of material are needed, and it is not always straightforward to load material into a narrow (*e.g.*, 0.5 mm) capillary if the powder is “sticky”. The following work focusses on situations where only very small amounts (< 1 mg) of material are available.

5.1.2 Powder diffraction data collection on single crystal diffractometer

For a transmission capillary data collection, the capillary rotates along its axis orthogonal to the direction of the beam in order to minimise the effects of preferred orientation and improve sample averaging. On a Rigaku single-crystal diffractometer, this arrangement is termed throughout as ‘Standard phi’ mode *i.e.*, a capillary can be mounted on a goniometer head and rotated about the phi axis. Equally, when dealing with very small amounts of powder, a small “ball” of sample can be mounted on the end of a fibre optic that is rotated around the phi axis (see chapter 3 for details). Care must be taken however to “shape” the sample such that it remains in the narrow beam path as it rotates; the alignment microscope of the SX

diffractometer is essential in this regard. Sample preparation is discussed further in section 5.3.

5.1.2.1 Preferred orientation

During powder data collection, the assumption is made that every orientation is evenly represented and that ideally any large crystallites are removed / ground before the point of data collection. This is not always the case, and the extent of any preferred orientation (PO) is unknown when a 0D or 1D detector is used. Figure 5.1 illustrates how PO affects peak height ratios and peak width when a section of the Debye-Scherrer rings is integrated to create a 1D diffraction pattern. Only the outcome is observed with the use of a point detector, typically found with in-house powder diffractometers.

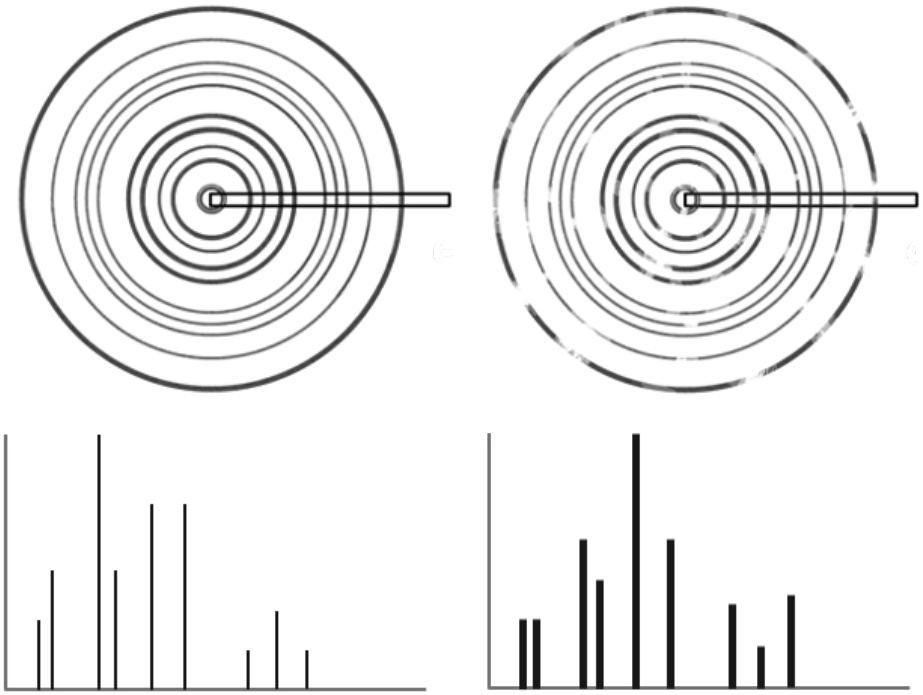


Figure 5.1. Diffraction rings and corresponding 1D pattern output from the view of a point detector. Left: ideal diffraction with uniform rings with correct peak height ratios; right: a typical example of preferred orientation affecting peak height ratios and peak widths.

The presence of PO is generally detrimental to analysis, especially when its extent is not known. The use of 2D (area) detectors allows for the detection of PO on PXRD instruments; for modern single crystal diffractometers, area detectors are the norm and can capture entire diffraction rings, allowing for the early detection of PO and/or large crystallites in the sample under study.

In 1967, Gandolfi optimised the data collection strategy for 2D Debye-Scherrer averaged rings. This was shown to be optimal when the sample centre of gravity is centred with respect to

axis A and B, shown in Figure 5.2, Axis B is at 45° to axis A. This allows for improved ring averaging, provided the sample is correctly aligned.[2]

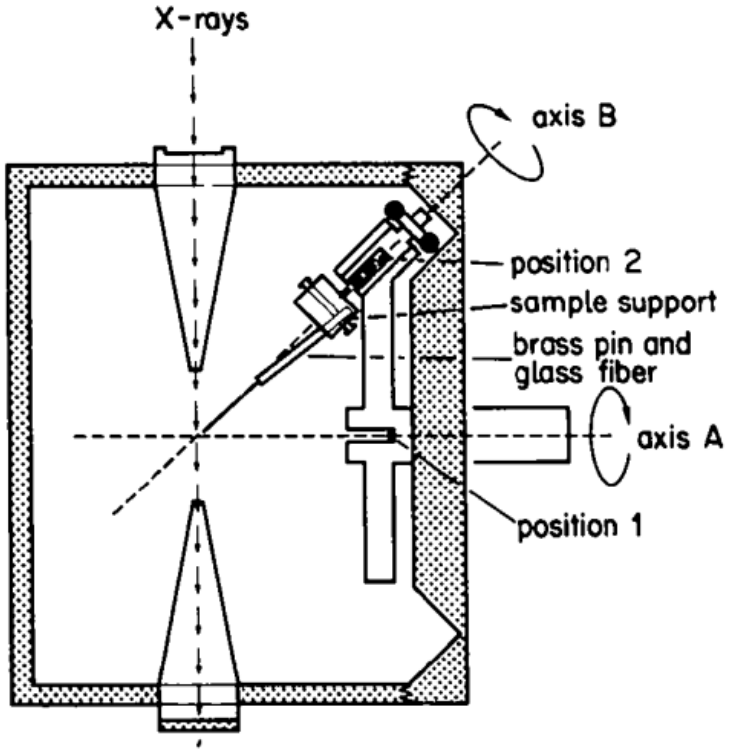


Figure 5.2. A schematic of Gandolfi movement, a transmission geometry.[3]

For the Synergy instrument, the available pre-programmed modes for PXRD data collection are ‘Standard phi’, ‘Fast phi’, ‘Gandolfi for powders’, and ‘Gandolfi for single crystals’.

5.2 Aims

Using the optimal parameters determined from chapter 3, this chapter aims to examine the usability of PXRD data collected on a Rigaku Synergy single crystal diffractometer for the purposes of structure solution (SDPD-SX) of compounds of pharmaceutical interest.

This broad aim can be subdivided into the following objectives:

- To establish a standard sample preparation and mounting procedure.
- To solve a set of pharmaceutical crystal structures from PXRD data collected on the Synergy using an amount of < 0.5 mg for each sample.
- To compare data quality and structure solution outcomes of SDPD-SX with those obtained by other PXRD data collection techniques (from previous publications).
- To compare data quality and structure solution outcomes obtained from SDPD-SX using the Rigaku Synergy to those obtained on a non-optimised Rigaku FR-X

5.3 Methodology

The experiments for this chapter were performed at the University of Reading (UoR). The diffractometer is a dual microfocus source (Mo, Cu) Synergy with PhotonJet-S X-ray source, equipped with a HyPix-6000HE detector.

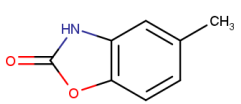
5.3.1 Sample preparation and pre-experiment

A test set of 18 compounds with known crystal structures (Table 5.1 and Table 5.2) spanning a range of molecular and crystallographic complexities were used. Details of the optimisation of the Synergy for PXRD purposes can be found in Chapter 3.

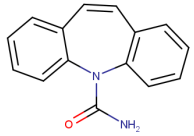
Table 5.1. Known compounds with corresponding CSD ref codes used throughout this chapter. Citations for the original CSD structures are given in Kabova *et al* (2018).[1]

Compound name	**Form**	**DoF**	**CSD refcode**
Chlorzoxazone	I	6	NEWKOP
Carbamazepine	III	7	CBMZPN01
Mefenamic acid	I	9	XYANAC
Ibuprofen	I	10	IBPRAC01
L-glutamic acid	II	10	LGLUAC01
Furosemide	I	11	FURSEM02
Salbutamol	I	11	BHHPHE
Sertraline hydrochloride	I	11	CAVVUQ
Sucrose	I	11	UCROS01
Indomethacin	III	11	INDMET01
Lansoprazole	I	12	XEGTIM
Chloramphenicol	I	13	CLMPCL01
Cefadroxil monohydrate	I	14	JOSWAP01
Mandelic acid	I	16	FEGHAA
Carvedilol	II	16	GIVJUQ01
Ritonavir	II	28	YIGPIO01
Sildenafil citrate monohydrate	I	30	FEDTEO
Paroxetine hydrochloride hemihydrate	I	31	GODVAW01

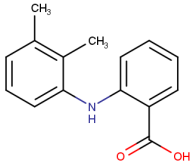
Table 5.2. 2D Molecular structures of the compounds used (see Table 5.1)



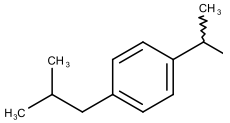
Chlorzoxazone



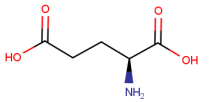
Carbamazepine



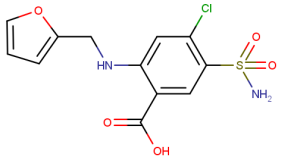
Mefenamic acid



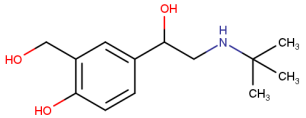
Ibuprofen



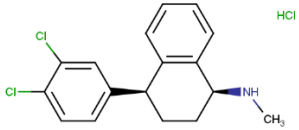
L-glutamic acid



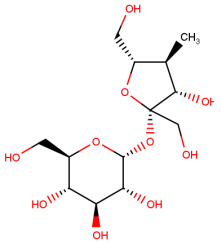
Furosemide



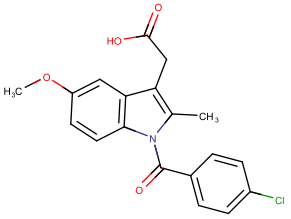
Salbutamol



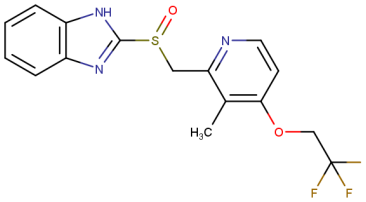
Sertraline hydrochloride



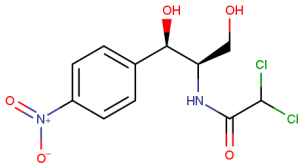
Sucrose



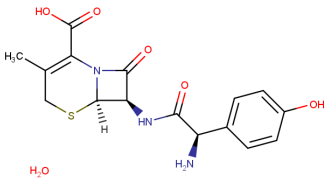
Indomethacin



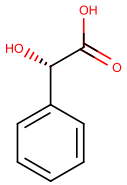
Lansoprazole



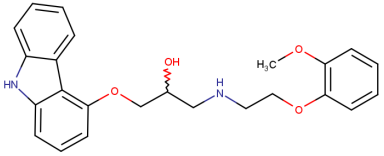
Chloramphenicol



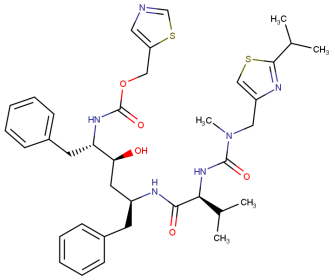
Cefadroxil monohydrate



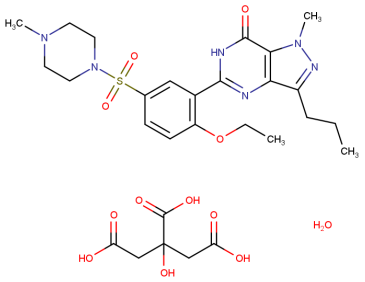
(S)-Mandelic acid



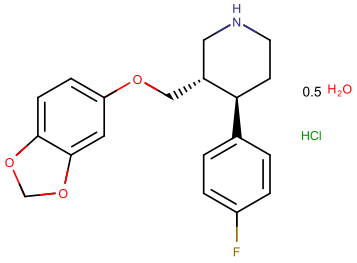
Carvedilol



Ritonavir



Sildenafil citrate monohydrate



Paroxetine hydrochloride hemihydrate

A small volume (< 1 mg) of each material was initially ground and transferred to a glass slide, and a small amount of Paratone® oil was placed next to the sample. A pointed metal tool (a clay modelling tool) was dipped into the oil and then mixed into the powder until the tip was coated. To produce the sample the tool was then pressed against the slide and rotated, forming a sphere of the sample with minimal oil content. The sample was then mounted onto the tip of a fibre optic cable mount (shown in chapter 3). Further description of the sample creation procedure is given in section 5.5.1.

5.3.2 Data collection parameters

The default mode for data collection was 'Standard phi' and was chosen because it is the simplest and most time-effective mode. A "pre-experiment" [1] (*approx*. 10 mins) allowed the extent of preferred orientation (PO), powder homogeneity, and signal-to-background ratio to be assessed before a full dataset was collected (details given below in Table 5.3). Although only 150 s is used per scan, multiple scans are needed at a further distance to assess the diffraction pattern out to 40 2θ, resulting in a total scan time of *approx*. 10 mins.

Table 5.3. Pre-experiment parameters

Parameter	**Value**
Temperature / K	275 - 298
Collection mode	Standard phi
2θ (max) / °	40
Detector distance / mm	160
Beam divergence / mrad	0.1
Time per scan / s	150

Based on the pre-experimental assessment, samples were then subject to the collection conditions of a 'full experiment' (Table 5.4). For samples showing strong PO, or large crystallites, the full data collection was performed in the 'Gandolfi for powders' mode.

[1] Not to be confused with the pre-experiment functionality of CrysAlisPro for single crystal work

Table 5.4. Full data collection parameters

Parameter	Value
Temperature / K	275 - 298
Collection mode	Standard phi / Gandolfi for powder
2θ (max) / °	60 - 70
Detector distance / mm	160
Beam divergence / mrad	0.1
Time per scan / s	1200 - 3000

5.3.3 Post-collection data handling

Following the full data collection, all powder diffraction patterns were centred with optimal coordinates, as discussed in Chapter 3. Datasets were then scaled using a program created within Python which determined the largest y coordinate value (peak intensity) then scaled the whole pattern to a chosen maximum height. For this work the maximum height was chosen as 50,000 or 100,000 and corresponding estimated standard deviation (ESD) values (square root of intensity) were calculated and outputted as a new .XYE file. Background removal was performed within DASH as part of structure solution.

5.3.4 Structure solution methodology

The DASH-based structure solution methodology has been outlined extensively elsewhere. EXPO2014 was used to help with indexing when DASH indexing proved ineffective or provided ambiguous results. Molecular models for the compounds under study were obtained from the CSD refcodes listed in Table 5.1; models were exported from Mercury in MOL2 format for use in DASH. Note that at the start of each DASH SA run, any variable torsion angles are assigned random values and as such no bias is introduced by using models derived from the CSD. All DASH SA runs were performed using its optimised parameters as reported by Kabova *et al* (2017), shown below in Figure 5.3.[4,5,6]

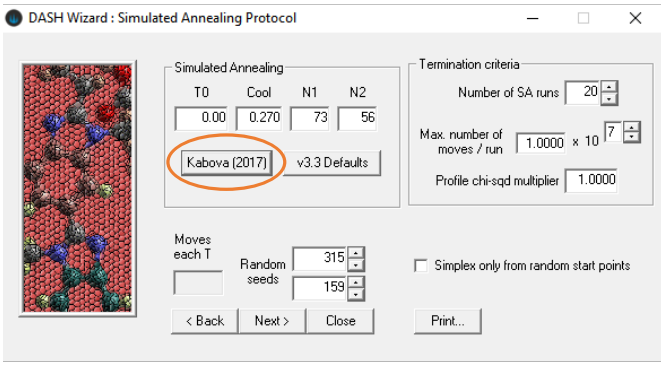


Figure 5.3. The use of optimal simulated annealing parameters (Kabova *et al* 2017) was selected within the DASH wizard.

The number of SA runs / moves required for each structure was determined by the number of degrees of freedom (DoF) of the system (details below in Table 5.5).

Table 5.5. Number of SA runs and SA moves chosen for each dataset, dependent on the DoF of the structure. For any structure ≤ 10 DoF that did not solve, the runs were repeated with 1 x 10^7 moves

DoF	Runs	SA moves
≤ 10	20	1 x 10^6
10 < x ≤ 28	20	1 x 10^7
>28	500	1 x 10^7

Successful crystal structure solution was confirmed by comparison of the best DASH solution and the equivalent CSD reference using the 'Crystal Packing Similarity' functionality of Mercury. The results are shown in Table 5.8.

5.3.5 Comparison to other PXRD approaches

Powder diffraction data for 8 compounds (Table 5.6), collected in three different instrumental geometries (Table 5.7), were obtained from Kabova *et al* (2018). The different data collection methodologies were evaluated for their structure solution capabilities, as well as comparing the FWHM of peaks as a means of facilitating data quality comparison.[1]

Table 5.6. Known structures used for data collection methodology comparisons

Compound	**DoF**
Mefenamic acid	9
Indomethacin	11
Salbutamol	11
Sertraline hydrochloride	11
Sucrose	11
Cefadroxil monohydrate	14
Carvedilol	16
Furosemide	22

Table 5.7. Instrument details as reported by Kabova *et al* (2018) for data quality comparisons.

Settings	**Transmission capillary**	**Transmission flat plate**	**Reflection flat plate**
System	D8 Advance	D8 Discover	D8 Advance
Generator	40 kV, 40 mA	40 kV, 40 mA	40 kV, 40 mA
Radiation	Cu Kα_1	Cu Kα (Göbel mirror)	Cu Kα_1
Geometry	Transmission	Transmission	Reflection
Detector	LynxEye	LynxEye	LynxEye
Sample holder	Borosilicate glass capillary (d = 0.7/0.5 mm)	Multi-well holder (Kapton foil wells)	Zero background silicon plate
Step size (°2θ)	0.017	0.016	0.017

The outcomes of the SDPD-SX performed on the Synergy are compared to the published results as shown in Table 5.9 – 5.13.

5.3.6 Comparison of FR-X to optimised Synergy

The structure solution of powder diffraction data collection of small organic molecules from single crystal diffractometer data (using < 1 mg of material) was first attempted and performed at the University of Manchester, using a Rigaku FR-X. As discussed previously (see Chapter 4), the Rigaku FR-X is a far more powerful setup, however the results were collected before any parameter optimisation was performed.

The results of the comparison of data quality from SDPD-SX using an optimised Synergy and non-optimised FR-X are shown in Table 5.13.

5.4 Results

Results are presented here as a series of tables.

Table 5.8 show the results of DASH structure solution runs for known structures using SDPD-SX on the Synergy.

Table 5.9 compares the FWHM of carvedilol peaks on the Synergy to FWHM values obtained on other powder diffraction instruments.

Table 5.10 shows a comparison of structure solution quality (using 50 SA runs at 5 x 10^6 SA moves) from the Synergy with that of other powder diffraction instruments, as assessed by the Crystal Packing Similarity functionality of Mercury.

Table 5.11 shows the maximum 2θ values and corresponding real-space resolution used in Pawley fitting for DASH structure determination for each of the different powder diffraction instruments.

Table 5.12 compares success rates for DASH structure solution with different powder diffraction instruments using 50 SA runs of 5 x 10^6 SA moves. A visual representation of this is given in Figure 5.4.

Table 5.13 compares real-space resolution, RMSD and success rate between an optimised Rigaku Synergy and non optimised Rigaku FR-X.

Table 5.8. Results of DASH structure solution runs for known structures using SDPD-SX on the Synergy

Compound	CSD refcode	DoF	SpGrp	Max 2θ / °	Res / Å	χ^2_{Paw}	N_{ref}	N_{moves}	$\%_{success}$	χ^2_{Prof}	RMSD / Å
Chlorzoxazone	NEWKOP	6	$P\bar{1}$	49.0	1.750	6.18	106	1 x 10^6	100	42.6	0.078
Carbamazepine	CBMZPN01	7	$P2_1/a$	35.5	2.527	5.77	76	1 x 10^6	10	98.64	0.033
Mefenamic acid	XYANAC	9	$P\bar{1}$	37.0	2.428	3.25	90	1 x 10^7	80	19.13	0.098
Ibuprofen	IBPRAC01	10	$P2_1/c$	37.5	2.396	4.21	92	1 x 10^6	80	12.55	0.139
L-glutamic acid	LGLUAC01	10	$P2_12_12_1$	59.0	1.564	1.50	120	1 x 10^6	100	9.61	0.031
Furosemide	FURSEM02	11	$P\bar{1}$	44.0	2.056	6.11	164	1 x 10^7	50	103.65	0.296
Salbutamol	BHHPHE	11	$Pbca$	43.7	2.070	4.47	163	1 x 10^7	100	14.80	0.040
Sertraline HCl	CAVVUQ	11	$P2_12_12_1$	45.5	1.992	6.77	151	1 x 10^7	100	27.25	0.034
Sucrose	SUCROS01	11	$P2_1$	44.9	2.017	13.70	107	1 x 10^7	100	36.90	0.040
Indomethacin	INDOMET01	11	$P\bar{1}$	37.0	2.428	3.99	122	1 x 10^7	100	22.51	0.101
Lansoprazole	XEGTIM	12	$P2_1/c$	41.2	2.189	10.57	163	1 x 10^7	20	34.24	0.067
Chloramphenicol	CLMPCL01	13	$A2_122$	49.6	1.836	0.99	154	1 x 10^7	55	24.05	0.056
Cefadroxil·H_2O	JOSWAP01	14	$P2_12_12_1$	44.5	2.034	3.55	146	1 x 10^7	20	32.81	0.080
Mandelic acid	FEGHAA	16	$P2_1$	44.3	2.030	1.1	113	1 x 10^7	85	9.46	0.209
Carvedilol	GIVJUQ01	16	$P2_1/c$	33.0	2.712	7.76	111	1 x 10^7	20	22.85	0.131
Ritonavir	YIGPIO01	28	$P2_12_12_1$	34.5	2.598	4.42	150	1 x 10^7	0.2	13.23	0.189
Sildenafil citrate H_2O	FEDTEO	30	$Pbca$	31.0	2.882	1.2	129	5 x 10^7	0.2	29.58	0.205†
Paroxetine·HCl·0.5H_2O	GODVAW01	31	$P2_1$	35.0	2.562	4.82	134	1 x 10^7	3.2	12.17	0.091

SpGrp = space group, Res= real-space resolution, χ^2_{Paw} = Best Pawley fit value, N_{ref} = Number of reflections, N_{moves}= Number of SA moves used, %success = % of runs achieving structure solution, RMSD = root mean square deviation as measured by Mercury's Crystal Packing Similarity functionality.

† The RMSD measurement was calculated excluding the water molecule

5.4.1 Comparison of Synergy and other diffraction methodologies

Table 5.9. Peak width, FWHM, as a measure of data quality using carvedilol‡

	FWHM / degrees			
Peak position / °	Flat plate	Transmission plate	SDPD-SX	Transmission capillary
5.843	0.089	0.089	0.1149	0.067
11.428	0.074	0.097	0.1250	0.069
11.919	0.086	0.096	0.1259	0.077
13.700	0.085	0.101	0.1282	0.067
15.252	0.070	0.100	0.1332	0.067
15.620	0.070	0.101	0.1310	0.068
Average FWHM	**0.079**	**0.097**	**0.1264**	**0.069**

‡ FWHM as reported by DASH

Table 5.10. Comparison of collection geometry structure-solution quality (from 50 runs at 5 x 10^6 moves), determined by comparison to CSD single crystal entry and reported RMSD using Mercury.

			RMSD / °				
Compound	**CSD code**	**DoF**	**FP**	**TP**	**TC**	**SDPD-SX**	**SDPD-SX$^{\perp}$**
Mefenamic acid	XYANAC	9	0.228	0.079	0.052	0.113	0.097
Indomethacin	INDMET01	11	0.212	0.248	0.158	0.176	0.101
Salbutamol	BHHPHE	11	/	0.097	0.072	0.248	0.04
Sertraline HCl	CAVVUQ	11	/	0.068	0.036	0.426	0.034
Sucrose	SUCROS01	11	0.095	0.071	0.038	0.04	0.04
Cefadroxil·H_2O	JOSWAP	14	0.208	0.100	0.055	0.065	0.08
Carvedilol	GIVJUQ01	16	0.452	0.118	0.115	0.12	0.131
Furosemide[§]	FURSEM01	22	/	0.987	0.888 (0.132)	N/A	N/A
Average of structures able to be solved	/	/	**0.239**	**0.221**	**0.082**	**0.144**	**0.074**

DoF = degrees of freedom, FP = flat plate, TP = transmission plate, TC = transmission capillary, ⊥ with 20 runs of 1 x 10^7 moves

[§] Furosemide studied using SDPD-SX is a different form to the form reported in *Kabova et al* (2022)

Table 5.11. Maximum 2θ used in Pawley fit, for structure determination in DASH for each of the collection methodologies and corresponding special resolution.

Compound	DoF	Reflection flat plate		Transmission flat plate		SDPD-SX (Synergy)		Transmission capillary	
		Max 2θ / °	Res / Å	Max 2θ / °	Res / Å	Max 2θ / °	Res / Å	Max 2θ / °	Res / Å
Mefenamic acid	9	45	2.01	40	2.25	37	2.42	45	2.01
Indomethacin	11	45	2.01	42	2.15	37	2.42	45	2.01
Salbutamol	11	45	2.01	45	2.01	43.7	2.07	45	2.01
Sertraline HCl	11	45	2.01	42.8	2.11	45.5	1.99	45	2.01
Sucrose	11	45	2.01	45	2.01	44.9	2.017	45	2.01
Cefadroxil·H_2O	14	45	2.01	45	2.01	44.5	2.03	45	2.01
Carvedilol	16	45	2.01	43	2.12	33	2.71	52	1.75
Furosemide[**]	22	40	2.25	33	2.73	N/A	N/A	45	2.01
Average	/	**44**	**2.04**	**42**	**2.17**	**40.8**	**2.24**	**46**	**1.98**

DoF = Degrees of Freedom, Res = real-space resolution

[**] Furosemide studied using SDPD-SX is a different form to the form reported in Kabova *et al* (2022)

Table 5.12. Comparison of success rates for structure solution with different powder diffraction data collection methodologies, for 50 runs at 5 x 10^6 moves.

		Success rate / %					
Compound	DoF	Reflection flat plate	Reflection flat plate (with PO adjustments)	Transmission plate	SDPD-SX	SDPD-SX (20 runs, 1 x 10^7)	Transmission capillary
Mefenamic acid	9	98	90	98	38	80	98
Indomethacin	11	2	42	50	100	100	100
Salbutamol	11	0	0	100	100	100	100
Sertraline HCl	11	0	0	100	100	100	100
Sucrose	11	48	50	100	100	100	100
Cefadroxil·H_2O	14	18	8	22	24	20	58
Carvedilol	16	0	10	32	36	20	22
Furosemide[††]	22	0	0	34	N/A	N/A	34

DoF = Degrees of Freedom

[††] Furosemide studied using SDPD-SX is a different form to the form reported in Kabova *et al* (2022)

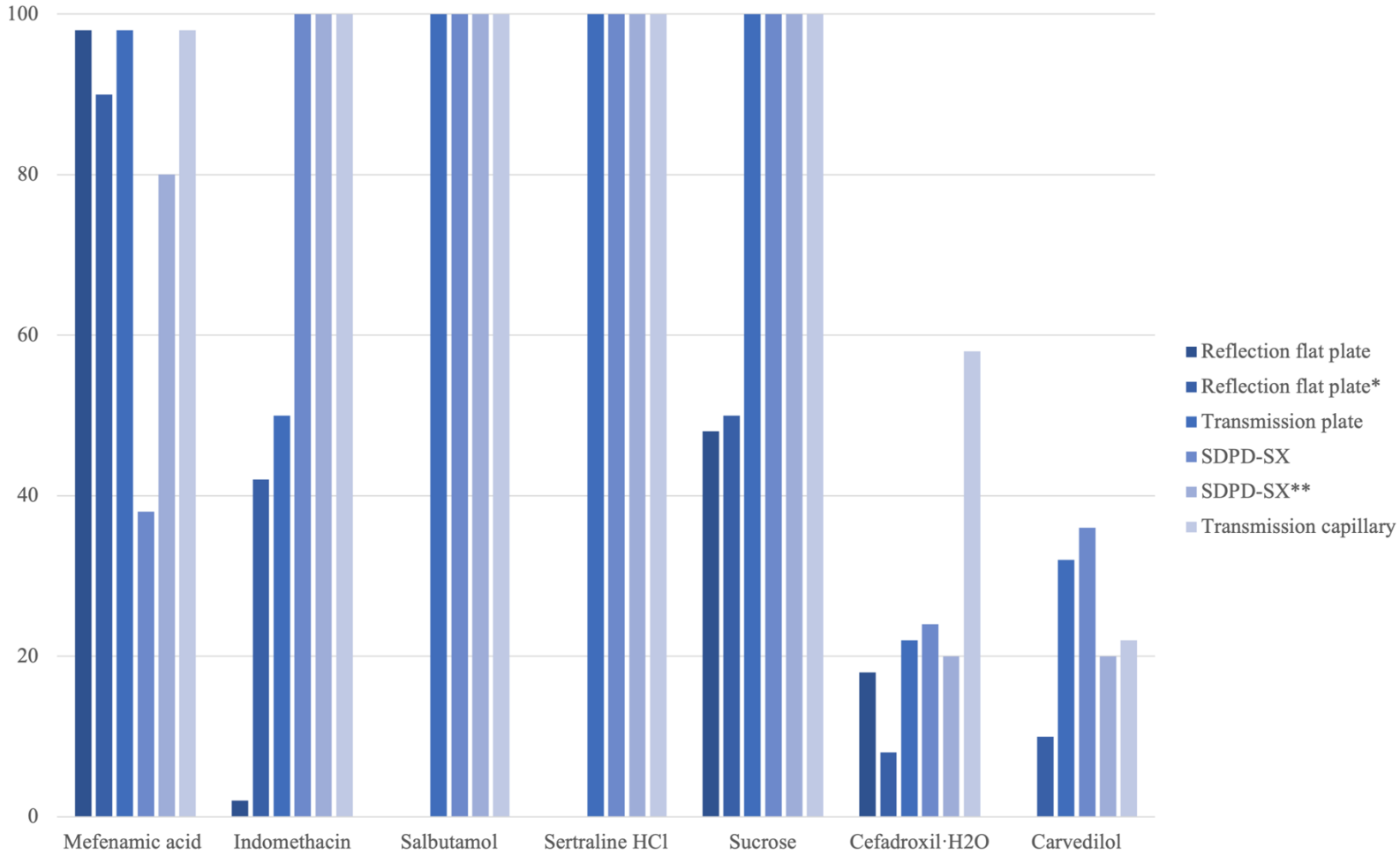


Figure 5.4. A visual representation of Table 5.12 comparing the success rate of different modes of data collection. Furosemide is excluded as this crystal structure was not available to be studied by SDPD-SX. * PO adjustments, ** 20 moves 1 x 10^7 moves.

5.4.2 Comparison of Synergy and FR-X (non-optimised settings).

Table 5.13. Comparison of data and structure solution (pre DFT minimisation), from an optimised Synergy and non-optimised FR-X.

Compound			**Resolution / Å**		**RMSD**		**Success rate / %**	
	DoF	Space group	Rigaku FR-X	Rigaku Synergy	Rigaku FR-X	Rigaku Synergy	Rigaku FR-X	Rigaku Synergy
Mefenamic acid	9	$P\bar{1}$	2.36	2.43	0.136	0.098	33	80
Ibuprofen	10	$P2_1/c$	2.14	2.40	0.099	0.139	100	80
L-glutamic acid	10	$P2_12_12_1$	1.75	1.56	0.027	0.031	33	100
Sertraline·HCl	11	$P2_12_12_1$	1.95	1.99	0.027	0.034	10	100
Indomethacin	11	$P\bar{1}$	2.23	2.43	0.118	0.101	20	100
Lansoprazole	12	$P2_1/c$	2.18	2.19	0.085	0.067	30	20
Chloramphenicol	13	$A2_122$	2.08	1.84	0.130	0.056	40	55
Cefadroxil·H_2O	14	$P2_12_12_1$	2.25	2.03	0.101	0.08	46	20
Carvedilol	16	$P2_1/c$	2.69	2.71	0.241	0.131	5	20
Ritonavir	28	$P2_12_12_1$	2.38	2.60	0.376	0.033	0.4	0.2
Sildenafil citrate·H_2O	30	$Pbca$	2.77	2.88	0.377	0.329	0.4	0.9[‡‡]
Paroxetine·HCl·0.5H_2O	31	$P2_1$	2.37	2.56	0.082	0.091	3.2	3.2
Average	/	/	**2.26**	**2.30**	**0.150**	**0.093**	**26.8**	**48.3**

DoF = Degrees of Freedom

[‡‡] The RMSD measurement was calculated excluding the water molecule

5.5 Discussion

The processes of sample preparation, sample mounting, data collection, and structure solution of the known materials is discussed to highlight best practice.

5.5.1 Sample creation

Sample preparation has a significant impact on the PXRD data quality and therefore the 'usability' of the resulting dataset. When selecting a single crystal, a small portion of the sample is submerged in oil to better separate the individual crystals. In contrast, established powder diffraction data collection methods (transmission capillary, transmission flat plate and reflection flat plate) avoid the use of oil altogether, and so the material is recoverable. The creation of suitable samples for SDPD-SX is challenging and somewhat counter intuitive to both the single crystal and powder sample preparation. As such, sample preparation and data collection for SDPD-SX requires a protocol to ensure consistency.

In the work of Kabova *et al* (2022), where ample lightly-ground sample was available, a sphere of each sample was produced by adding powder sample directly to a drop of oil. The resultant mixture was then pulled across the glass slide and more sample added until the mixture was of the correct consistency, a mouldable paste. Whilst reliable, using this approach the user runs the risk of running out of dry material (if it is in relatively short supply) before the correct consistency is reached, thus risking high-quality data collection.[7]

There is a variety of specialist equipment available for single crystal diffraction that can be used for sample creation; however, for this study, the most straightforward, durable, and inexpensive option was found to be a clay modelling tool. The tool used (shown in Figure 5.5) was chosen for being slim (diameter *approx*. 0.3 – 0.5 mm), rigid, sharp, and easy to handle. Multiple tools were explored and found to be inadequate; in particular, more expensive options (such as specialised tools designed for single crystal diffraction experiments) were found to bend against the glass slide or break at the tip if dropped, making them no longer usable.

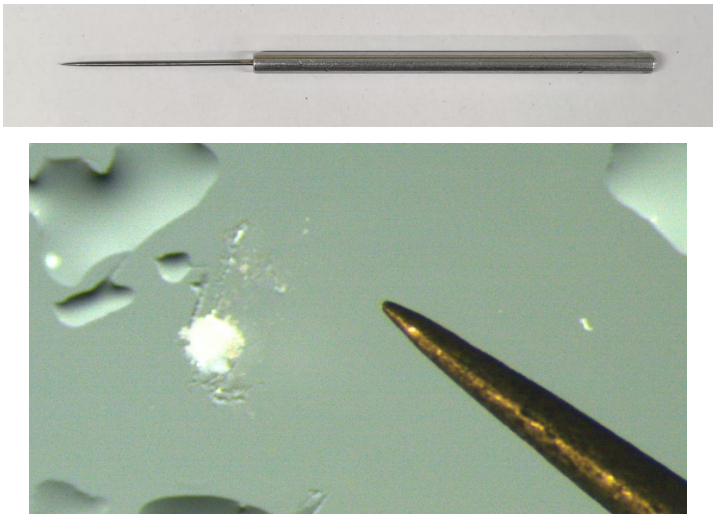


Figure 5.5. A clay modelling tool (above) shown with respect to a sphere of sucrose, used for data collection (Below). The tip is approximately 0.3 mm wide.

The most reliable method for sample preparation was found to be the addition of dry material and oil separately to a glass slide. To create the sphere of material for mounting, the tip of the clay modelling tool was dipped into the oil and then coated in the dry material. The rod was then pressed against the glass slide and rotated to form a paste which, when the material/oil ratio is correct, should form a stand-alone sphere that can be easily manipulated. This technique is shown below in Figure 5.6, any material mixed with an excess of oil can only be made into a useable sample by adding more dry material, so it is best to keep some dry material separate where possible.

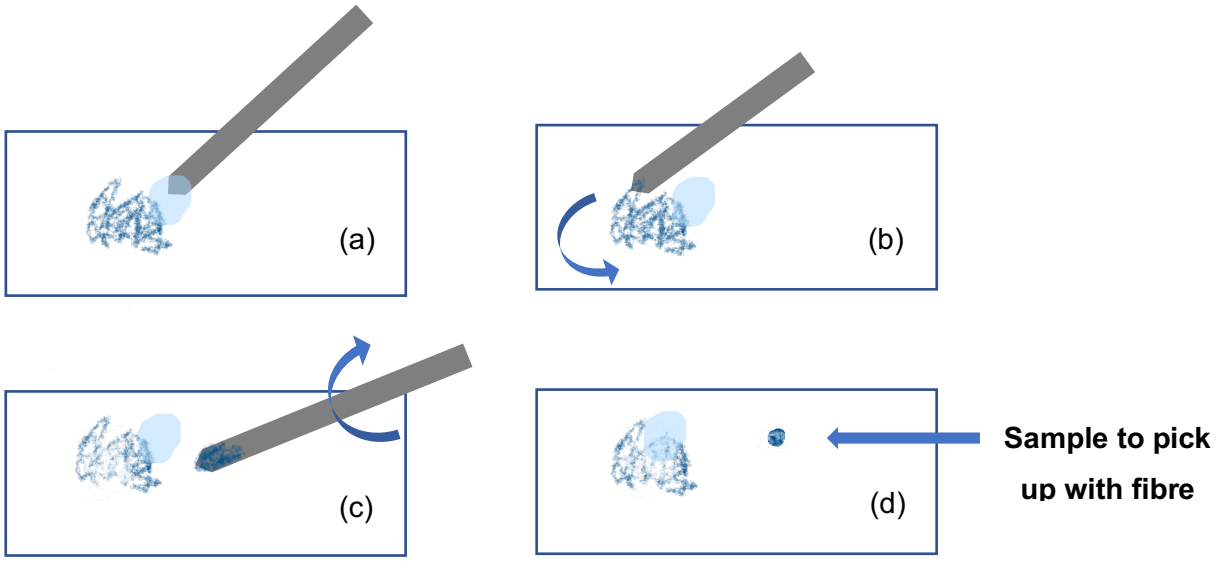


Figure 5.6. Simplified schematic of producing a good sample sphere (with minimal oil). Firstly, a small patch of oil is added to the side of the crystalline powder and dipping the tip of the tool in the oil (a). Then by rolling the tip of the tool in the dry powder (b) the tool can be held at an angle against the glass, rotating to create a sphere (c). Lastly, the sphere is lifted onto the fibre optic cable pip (d).

For material with large crystallites that are in limited quantities (too small for a pestle and mortar), the crystallites can be lightly crushed using the back of a small spatula. The process of sphere formation can then be performed.

5.5.1.1 Oil for sample preparation

Different oils (specifically oils used for single crystal diffraction experiments, *i.e.*, paraffin oil) were used at the beginning of the investigation to determine the best option for holding and adhering to the sample. It was found that Paratone® N oil was the most effective and oil that was less viscous than this was detrimental to sample formation. Some of the materials studied formed spheres without the need to mix with oil and in these instances, only the tip of the sample mount was dipped in Paratone® oil, in order to mount the sphere.

5.5.1.2 Sample mount

There is a large variety of sample mounts available to purchase for SX work, each with a design to accommodate a different sample size and shape. The sample mount when in the path of the beam, contributes to the background and therefore the signal-to-background ratio. For this reason, mounts and loops are made of X-ray transparent materials such as Kapton® or nylon, examples of which are shown below in Figure 5.7.

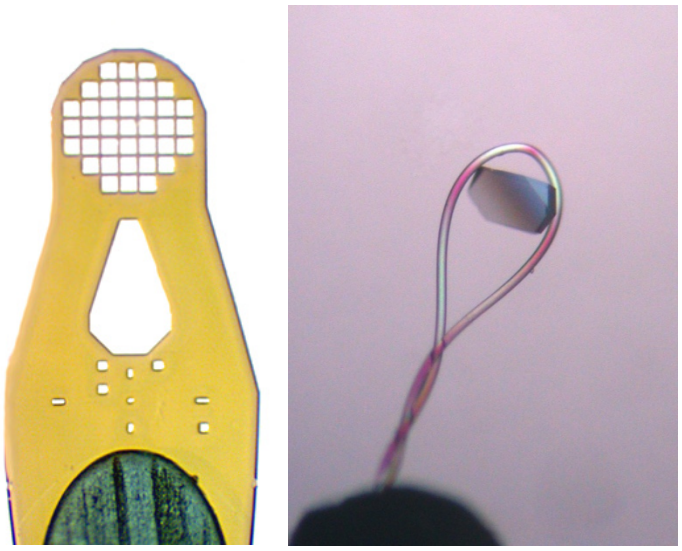


Figure 5.7. Examples of sample mount meshes available from MiTeGen, 400 μm head with 25 μm openings (left) and nylon loops, available from Hampton research 20 μm (right).

It was found during the initial tests that nylon loops were not a good choice as sample holders. The loop shape meant that the material was difficult to concentrate in one place and the addition of oil caused the sample to spread across the loop further. As a result, it was not easy

to centre the sample on the diffractometer, and the thinly spread sample was not concentrated in the beam.

The fibre optic cable sample mount (9/125 μm core) used by Kabova *et al* (2022) was found to be ideal, as the centre of mass of the sample is above the tip of the fibre, reducing the contribution of the fibre to the background. The sample holder was cut to a maximum length (with goniometer at minimum extension) and was shortened for subsequent samples, compensated for by increasing the sample height adjuster on the goniometer head. Furthermore, the majority of the material in the fibre optic cable is made from plastic, therefore making this superior to alternative fibres made from glass that would have increased x-ray absorption (from the heavier Si in the material). [7]

It was found during this study that due to the asymmetric reduction of the beam; a larger sphere (2-3 times greater than the width of the fibre) of material was beneficial for data collection. The optical fibre core used in this chapter is the same used by Kabova *et al* (2022) at 125 μm, but the cable used in this work had an extra coating, resulting in a diameter of 225 μm, allowing for a larger sphere to be mounted more easily. The greater fibre thickness was also found to be more durable than in the previous study and would break more cleanly, creating a flat surface for the sample to adhere.

5.5.1.3 Considerations regarding sample grinding

For most diffraction experiments the sample preparation begins with grinding the sample to reduce preferred orientation without fear of reducing data quality. In spite of the inferior instrumental resolution of SDPD-SX, it was found that some samples could be over-ground, leading to visible peak broadening (see Figure 5.8). This peak broadening constitutes an information loss for SDPD-SX purposes and can lead to problems with data fitting. It is not possible to reverse overgrinding (without recrystallisation) and therefore it is recommended perform data collection with the sample ‘as received’ when there is limited material available.

Subsequently, if PO is found to be severe and the effect upon the diffraction rings is not reduced by the ‘Gandolfi for Powders’ collection, the sample can be removed, lightly crushed on a glass slide, and remounted for further analysis. Significant PO is expected to be problematic for structure solution; however, this is not *always* a barrier, therefore it is pragmatic to have a full dataset with PO as a backup for structure solution.

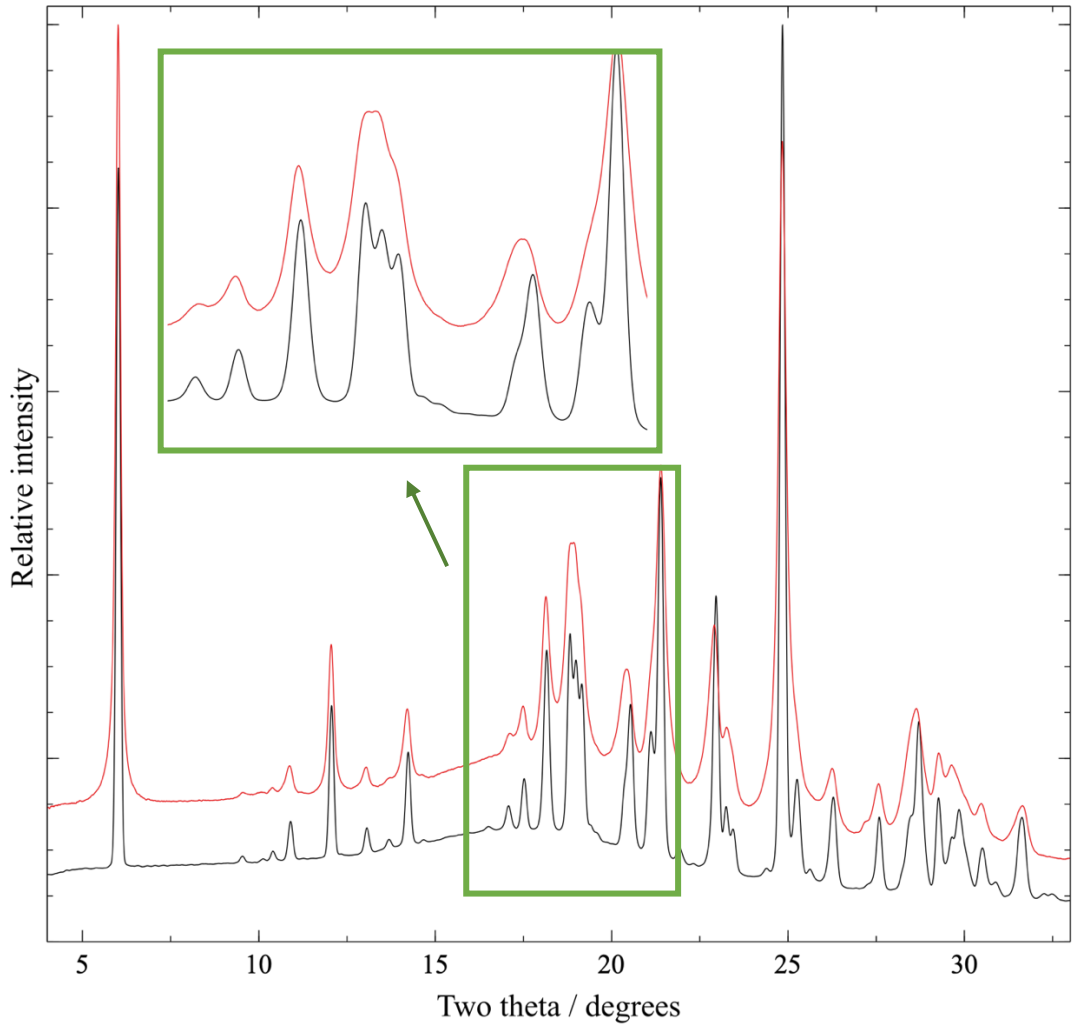


Figure 5.8. Effect of overgrinding demonstrated by a comparison of furosemide lightly ground using a pestle and mortar (red) and unground (black). Samples were collected in standard phi mode at minimal beam divergence, (*approx*. 0.1 mrad) at 160 mm detector distance. Significant peak broadening is highlighted in green.

5.5.2 Mode of data collection

External averaging) is a useful tool that can be used to compensate for preferred orientation or large crystallites within the sample. During transmission capillary data collection, this is done by rotating the sample in the phi axis (referred throughout at standard phi mode). Using a single crystal diffractometer, a greater range of external averaging modes can be used to compensate for a smaller volume. Figure 5.9 shows the effect that the 'Gandolfi for Powders' mode has on a sample with suspected large crystallites during a test scan of sucrose.

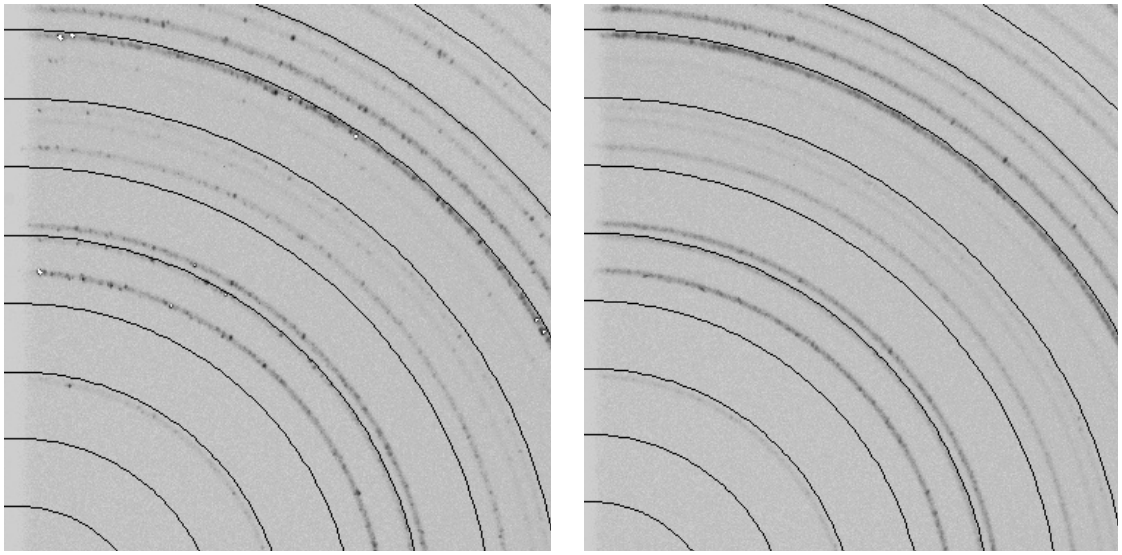


Figure 5.9. Sucrose with suspected large crystallites. Lightly ground sample using ‘Standard phi’ mode (left) and ‘Gandolfi for Powders’ (right) collected at minimal beam divergence with detector distance of 120 mm, for 150 s / scan.

It is important to note however that any mode of collection mode that moves the sample (particularly when moving in more than one axis) creates a risk for sample movement with respect to the beam. When the sample is not correctly aligned there is the potential for the centre of mass to move during collection, resulting in peak broadening. Thus, it is important to ensure optimal sample centring prior to full data collection.

Utilising different sample movements for powder diffraction data collection is a unique feature of SDPD-SX and enhances data collection due to the sample shape and mount. It should be noted that other data collection modes (for example, material within a transmission capillary) would not benefit from extra external averaging. This is because the minimal cross section of the glass covering is intersected by the incident beam when orthogonal to the capillary *i.e.*, a standard phi collection.

Although modes of collection other than ‘Standard Phi’ enable a greater range of data collection, many single crystal diffractometers such as the FR-X have physical constraints, preventing the use of other modes (see chapter 4).

One further refinement that could be used to improve data quality would be to determine the exact centre of the beam on the Synergy, when using minimum beam divergence, with the use of a fluorescent material. Exact centring of the sample on this point would maximise potential data quality.

5.5.3 Structure solution of known APIs

The results of SDPD-SX, presented in Table 5.8, demonstrate that powder diffraction performed on an optimised 'standard' single crystal diffractometer produces data of sufficient quality to achieve structure solution from crystals possessing a range of molecular and crystallographic complexity. It is worth noting, however, that none of the materials used were phase mixtures, or have particularly large unit cells, or high Z' values, all of which would make the collected PXRD data considerably more complex in terms of the number of observed peaks per degree two-theta. It is with such more complex data that instrumental resolution becomes a more important factor.

Of the 18 compounds studied, structure solution was achieved for all using readily available software and without the use of any of the available prior knowledge of the crystal structure.

Furthermore, the results from Table 5.8 show that the structure solution was not limited by high complexity; for example, structure solution of paroxetine.HCl.0.5H_2O with 31 DoF was straightforward, returning a structure with an RMSD (root mean square deviation using the Crystal Packing Similarity feature of Mercury) of 0.091 Å relative to the known SX structure. The success rates (*i.e.,* the number of SA runs achieving a solution as a % of the number of SA runs used) for the 18 compounds range from 0.2% to 100%, with lower success rates generally seen for those structures with a larger number of DoF, as expected. The accuracy of the structure solutions varied, with RMSD values of 0.033 – 0.296 Å, noting that 14 of the 18 structures had a RMSD value of < 0.15 Å. These are excellent results given the relatively low resolution (*ca*. 1.56– 2.71 Å) of the datasets collected, which in many cases is lower than that typically accessible by conventional transmission capillary experiments.

The relatively low resolution of structure solutions, like those presented in this work, is not an impediment for their inclusion in the early stages of drug discovery. This information can be used to aid the design of APIs because the molecular conformation and packing are still well determined in these structures.

5.5.3.1 Data scaling

When using a silicon strip detector such as those utilised in a variety of powder diffraction setups, the measure of intensity is given in counts. The range for these for a transmission capillary data collection is typically 10,000 – 100,000 counts (depending on exposure time). The maximum intensity is generally dependent on the amount of material, which for a fixed diameter of Lindemann capillary gives a consistent output. The exposure time does not

improve the overall signal to background but does however improve the signal-to-noise ratio. For each data point, an estimated standard deviation (ESD) of the intensity value can be quickly calculated as the square root of the number of counts. This means for a data point with an intensity of 100,000 counts, the associated ESD would be approximately 316 counts or 0.316 %. For data collected using SDPD-SX on the Synergy, with optimal parameters, the peaks have a maximum relative intensity of around 10 'counts', resulting in a maximum ESD of 3.16, or 31.6 %. Therefore, the low intensity values inevitably lead to relatively very high ESDs, a feature that becomes problematic during indexing and subsequent Pawley refinement; the large relative ESD value gives DASH a larger 'uncertainty range' for the program to create a fitting pattern, producing too many options for the program to find the correct output.

To counteract the low intensity values for this chapter and subsequent research, all powder datasets were scaled before transferring into DASH for indexing and structure solution steps. This procedure changes the magnitude of the peak intensities but crucially does not affect the peak ratios. As such, this does not affect the underlying information content of the powder pattern.

The option of scaling the data within CrysAlisPro was explored but found that the user could not choose the height of the peaks, and the maximum peak height produced was approximately 2000, which proved not to be suitable for analysis. Due to the number of datasets produced in this study, and with future work in mind, a program was created in Python to scale ".xye" files, outputting a new ".xye" file for structural analysis. This program, available from GitHub, allows the maximum intensity value to be chosen by the user. A suitable maximum peak height for structural analysis work is *approx*. 50,000 (to mirror the values produced by a transmission capillary dataset). Where Pawley fitting was ineffective a maximum peak of 100,000 was attempted.[8]

5.5.4 Comparison to other methods of powder data collection

For a typical crystal structure determination, a Rietveld refinement step or DFT calculation is performed to optimise the solved structure. These steps have not been performed here, as the SX structures are already known.

5.5.4.1 Data quality

Unlike dedicated powder diffractometers, the Synergy Cu beam source is non monochromatic i.e., the beam contains both $K\alpha_1$ and $K\alpha_2$ wavelengths. The result is peak broadening/ splitting

with increasing 2θ. This is reflected in Table 5.9, showing that diffraction peaks collected by the Synergy are significantly broader than when obtained with the other methodologies. In particular, the peak widths of the SDPD-SX data are approximately double those of the transmission capillary diffraction data reported for carvedilol.

5.5.4.2 Structure solution output

The quality of the structure solutions, when compared to other methodologies for collection is given in Table 5.10. The resulting RMSD values for the structure solutions, when compared to the CSD reference structures, show that the output is closest to the results from the transmission plate and superior to the flat plate data. When the length of individual SA runs is extended (from 5 x 10^6 to 1 x 10^7 moves), the RMSD values improve significantly, and the results are closer to those of the transmission capillary instrument.

However, the real-space resolution (as set by the max 2θ) of the data used for SDPD-SX, shown in Table 5.11, is on average significantly lower than the other data collection methods-where structure solution was possible. Table 5.12 shows that the structure solution obtained by SDPD-SX have higher success rates than reflection and transmission flat plate and were most similar to the success rates produced by the transmission capillary data collections.

5.5.4.3 SDPD-SX compared to transmission capillary

One of the aims of the work in this chapter was to create a comprehensive comparison of different in-house powder diffraction data collection methods. The most important of these comparisons is the comparison of SDPD-SX to the “gold standard” of a transmission capillary instrument.

Transmission capillary data collected from a typical molecular organic sample has a peak resolution of approximately 0.07° at low angle on a Bruker D8 Advance instrument with monochromatic incident radiation. A simulated $K\alpha_1$ dataset for ritonavir (CSD refcode YIGPI01) is shown below (Figure 5.10) to highlight the difference in data quality between the two approaches.

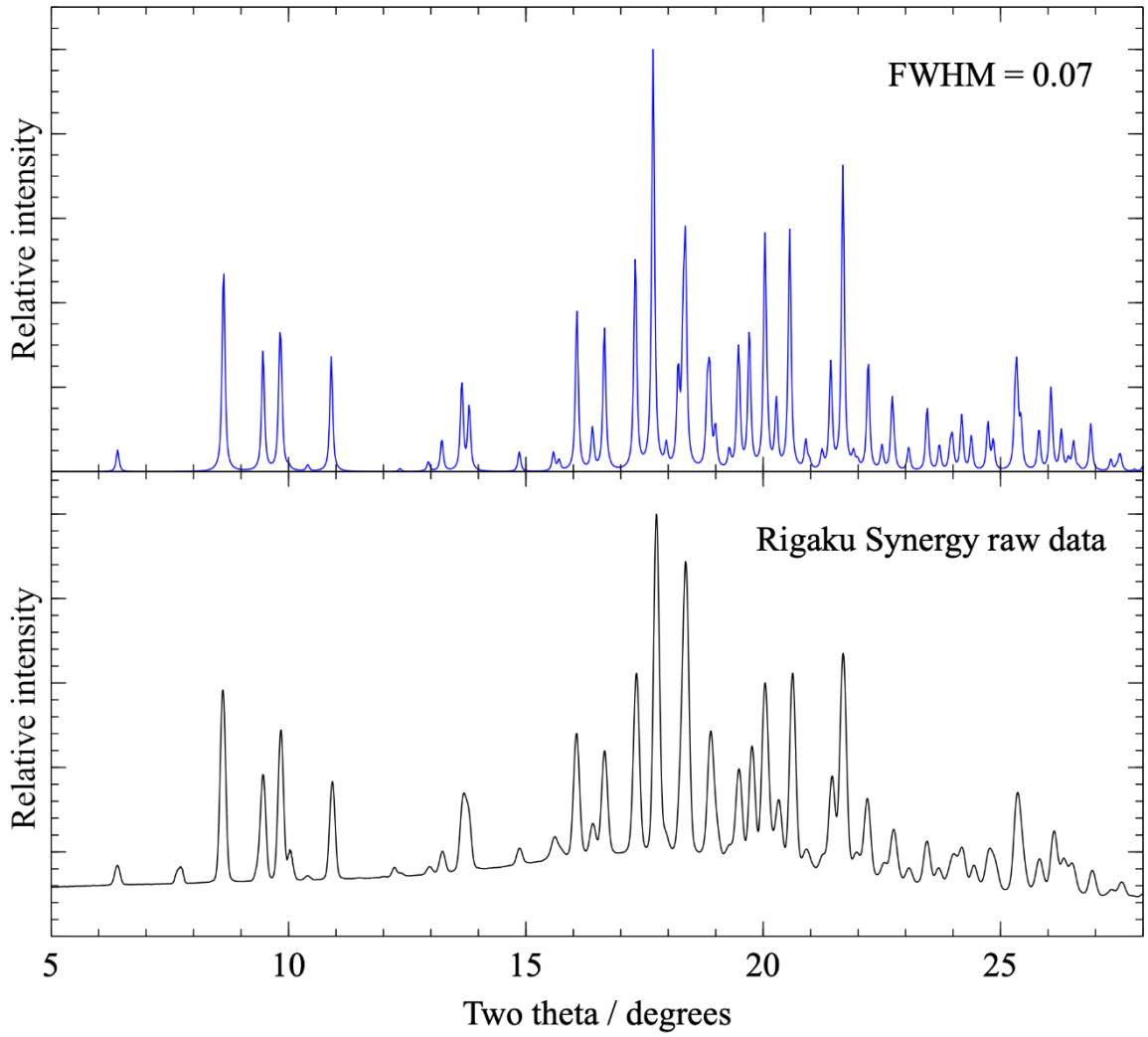


Figure 5.10. Powder diffraction patterns of ritonavir (CSD refcode YIGPI01), showing (upper) a calculated pattern with resolution typical of transmission capillary instrument and (lower) an experimentally collected powder pattern from the optimised Synergy single crystal diffractometer.

Transmission capillary data are obviously superior and benefit from several factors *e.g.,* consistent sample volume exposed to the beam, and focussing geometry (*i.e.,* the beam focal point is on the detector itself) leading to good instrumental resolution. That said, a capillary data collection requires approximately 20 to 100 times more material and a longer data collection time, as it uses a conventional X-ray tube as opposed to a microfocus source. The work in this chapter, as well as the work by Kabova *et al* (2022), has shown SDPD-SX is more than capable of structure solution for moderately complex molecular materials. As such, SDPD-SX can be attempted with a fair degree of confidence for single-phase samples possessing fewer than 30 DoF.

5.5.4.4 SDPD-SX compared to transmission flat plate

Transmission flat plate (TFP) is an in-house powder diffraction technique, designed to reduce the negative effects of preferred orientation whilst maintaining the benefit of automation seen with reflection flat plate setups. A typical setup involves the addition of different samples to a 96-well plate, sealed with Kapton® film. This is mounted on a dedicated diffractometer (such as the Bruker D8 Discover that is capable of transmission and reflection flat plate) where samples can be screened in series.

TFP data has been shown to be useable for structure solution and compared to SDPD-SX has superior instrumental resolution, which is unsurprising, given the use of focussing geometry. For structures solved previously using TFP, the real-space resolution is better for 5 out of 6 compounds (see Table 5.11). SDPD-SX, however, produces higher quality structure solutions with higher success rates. Although a big advantage of TFP is the high-throughput screening ability, for the purposes of structure solution it is recommended that SDPD-SX will yield more favourable results. This comparison suggests that the diffraction data collected using the single crystal diffractometer is of higher quality in the low angle range but drops in quality at higher angle (significantly beyond 40 2θ). As stated previously, the drop in data quality is due to a non-monochromatic source from the Synergy, which causes significant peak broadening. These differences in setup account for the improved success rates and lower RMSD values of the SDPD-SX structures compared to TFP but poorer spatial resolution (max 2θ).

5.5.4.5 SDPD-SX compared to reflection flat plate

Reflection flat plate instruments come in various configurations and flat “zero-background” sample holders offer the highest quality data output and very little sample is needed to use them. The sample is generally retrievable with this method, as only a thin film of grease is needed to cause the powder to adhere to the plate. Reflection flat plate data collection is simple to automate and is used widely in many research areas. As shown in Table 5.9 -5.12, reflection flat plate diffraction data nevertheless suffers considerably from preferred orientation, as crystallites tend to lie along the plane of the sample holder. Although Pawley fitting is demonstrated to be possible at higher 2θ angles compared to SDPD-SX, the solved structure is of higher RMSD and lower success rate (where structure solution is possible).

Kabova *et al* (2018) demonstrated that with adaptations to structure solution, the presence of PO can be somewhat accounted for. Unfortunately, this requires the knowledge of which

planes are affected. In general, the use of reflection flat plate geometry for SDPD is not recommended, other than if this is the only geometry available to the user.

5.5.5 Comparison to of data collection using the FR-X and Synergy

There are many similarities between the FR-X and the Synergy, including detector type and overall setup. The main differences, explored in this chapter, are the use of an optimal setup and the intensity of the beam, with the FR-X having a much more intense rotating anode source. The datasets collected and used for structure solution by Kabova *et al* (2022) were collected using a beam divergence of 1 mrad and a detector distance of 150 mm. Further investigation (see Chapter 4) found that the optimal parameters for this configuration were minimal beam divergence (*approx*. 0.2 mrad) and a detector distance between 250 and 300 mm. Table 5.13 shows that the FR-X outputs diffraction patterns with a superior real-space resolution for 8 of the 12 datasets; however the corresponding Synergy structure solutions (before DFT-D optimisation) have a lower RMSD output for 9 of the 12 compounds and have an average of double the success rate. This highlights the importance of optimising the instrumental setup, regardless of the hardware used.

5.6 Conclusions and next steps

The work in this chapter used the optimal parameters determined in chapter 3 to collect diffraction patterns of a variety of known pharmaceutically relevant compounds with < 0.5 mg of material. Each of the diffraction patterns collected was of suitable quality for structure determination and therefore has demonstrated the ability of SDPD-SX to produce good quality PXRD patterns from microgram quantities, using standard in-house equipment. Although the Rigaku Synergy is by no means a “basic” diffractometer (in that it has dual sources, with control over beam divergence for the Cu source, and a very efficient detector), equivalent diffractometers will increasingly become the laboratory standard in the near future.

The range of compounds studied, in terms of solution complexity, shows that SDPD-SX compares very favourably with other data collection methodologies. Compared to transmission capillary data collection, SDPD-SX is a time and sample efficient alternative , requiring only 1-5 % of the material needed for a transmission capillary experiment. Crucially, the quality of the resulting crystal structures is good and perfectly adequate for many purposes. The RMSD and success rate of the structures of SDPD-SX compared to both transmission flat plate and reflection flat plate shows that SDPD-SX offers a better path if the aim is to attempt structure solution.

Although the real-space resolution of the diffraction patterns collected is relatively low, (*ca.* 2 Å), SDPD-SX has the potential to enhance processes such as structure-based and ligand-based design, where the molecular geometry can be used to inform the early stages of drug development, despite only very small amounts of sample being available. The use of DFT-D for structure verification is also increasingly important.

Studying known compounds allows the determination of the capabilities of a methodology, but to determine if SDPD-SX is usable for “real-world” problems, then its application to materials whose crystal structures are not already know is essential (Chapter 6).

5.7 References

1 E. A. Kabova, C. D. Blundell and K. Shankland, *J. Pharm. Sci.*, 2018, **107**, 2042–2047.

2 G. Gandolfi, *Miner. Petrogr. Acta*, 1967, **13**, 67–74.

3 A. Gruttner, K. Yvon and M. Delaloye, *J. Appl. Crystallogr.*, 1978, **11**, 716.

4 W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell and J. C. Cole, *J. Appl. Crystallogr.*, 2006, **39**, 910–915.

5 A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni, R. Rizzi, N. Corriero and A. Falcicchio, *J. Appl. Crystallogr.*, 2013, **46**, 1231–1235.

6 E. A. Kabova, J. C. Cole, O. Korb, M. López-Ibáñez, A. C. Williams and K. Shankland, *J. Appl. Crystallogr.*, 2017, **50**, 1411–1420.

7 E. A. Kabova, C. D. Blundell, C. A. Muryn, G. F. S. Whitehead, I. J. Vitorica-Yrezabal, M. J. Ross and K. Shankland, *CrystEngComm*, 2022, **24**, 4337–4340.

8 M. Ross, 2023.

Chapter 6

Structure determination of novel materials using SDPD-SX

6.1 Introduction

Chapter 5 demonstrated that SDPD-SX can be used to solve a range of organic crystal structures with the use of an in-house SX diffractometer and non-specialist computing power. Although the quality of the data output of this method is lower, both in terms of spatial resolution (max 2θ) and greater FWHM (causing greater chance of peak overlap), compared to that of transmission capillary data collection, the complexity of structures solved is comparable. Data from SDPD-SX has been shown to be capable of structure solution of up to 31 DoF, demonstrated by the structure solution of 18 previously reported compounds with RMSD values ranging from 0.031 to 0.296 Å. Crucially all structure solutions were performed without the need for specialist computing power. Therefore, for samples that are not in quantities suitable for traditional in-house X-ray diffraction methods, SDPD-SX offers a viable solution, where previously there was none.

6.1.1 Modern drug discovery programmes

Modern drug discovery involves the synthesis and analysis of new drug candidates, both physically and computationally (see Chapter 1). When testing a new drug series, this largely involves screening the compounds with various assays to monitor the effect and assess the potency and effectiveness of the compound. An example of an assay would be screening compounds with the target protein and monitoring any changes as a result. In recent years, a growing proportion of drug design is performed computationally due to advancing computational methods and readily available conformational information. Examining a target protein single crystal structure with a library of small organic molecules (obtained from crystal structure determination or generated by virtual means) the potential hits can be identified and tailored to suit the active site. This approach has the potential to create more selective drug candidates which would potentially have fewer side effects.

Due to the large number of compounds that are considered during hit identification combined with the cost of synthesising novel compounds, the starting amount synthesised is less than 5-10 mg. After the chosen assays are performed, the amount remaining for each compound is negligible (< 0.5 mg), such that for structural analysis it is not routinely performed. As previously described, growing good quality single crystals can be challenging, often resulting in crystalline materials that are unsuitable for SCXRD, such as those shown in Figure 6.1. The quantities available also means transmission capillary data collections is inaccessible for the majority of the samples studied. Traditionally a crystallisation campaign typically consists of at least 10-20 mg of material to allow many conditions to be tested.

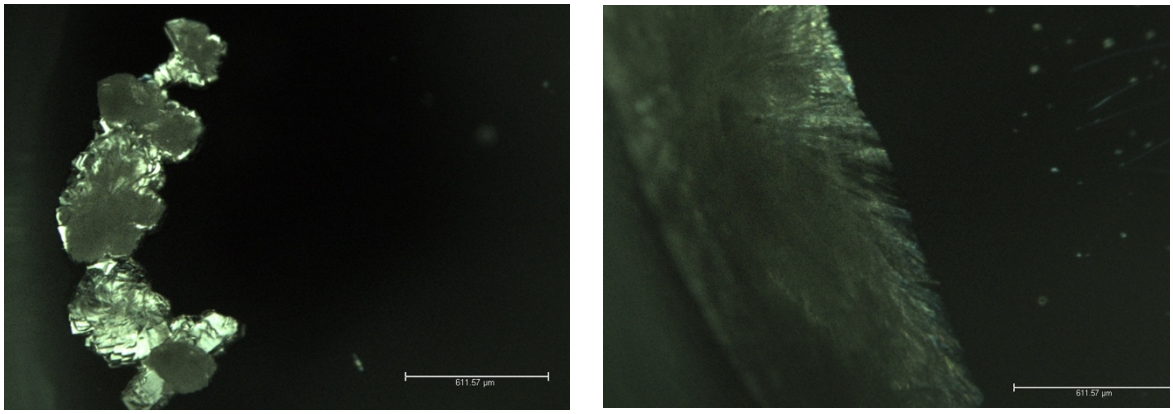


Figure 6.1. Examples of non-single crystal formation, carazolol (left) shows a ‘rose-like’ formation and FFA (right) giving an example of a ‘seashell-like’ formation.

6.1.1 Focus of this chapter

In order to assess how the application of SDPD-SX could affect real-world drug discovery, this work attempts to replicate an early-stage drug discovery program. To achieve this, this chapter uses an ensemble of drug-like compounds, composed of previously reported (available on the CSD) and novel structures. Each compound was only available in limited quantities (< 0.5 mg) and were chosen to represent compounds that are relevant to current drug discovery.

Ideally, for each compound, a vast number of solvents and crystal-growing conditions would be tested, but in practice (unless by use of automated methods) this is not viable in terms of time and expense. Therefore, as part of this work a strategy for solvent selection is explored to maximise crystal output. It is important to note that solubility data is generally not available for solubilities of novel compounds so typical recrystallisation solvent design would not be applicable.

The use of novel materials allowed the exploration of hurdles which were purposefully avoided in Chapter 5 (*i.e.,* phase mixtures, poor crystallinity, impurities). The use of serial recrystallisation was used with the aim of attempting a range of conditions without contaminating or reducing the compound between experiments. Only the study of apparently crystalline (by examining under cross polarised light) but unsuitable for SCXRD samples were investigated as part of this work.

6.2 Aims

The aim of this chapter is to apply the knowledge gained from the previous chapters to use SDPD-SX in real-world analysis-limiting conditions. Specifically, this work aims to attempt to solve novel structures using (< 0.5 mg) samples with non-ideal morphology (unsuitable for SCXRD). In addition, this chapter aims to explore the limitations of SDPD-SX as well as the differences in approach required for successful structure solution.

This chapter uses a large number of samples such that a recrystallisation protocol was used to maximise the number of crystalline samples. The secondary aim therefore is to explore the proportion of crystalline materials usable for SCXRD and SDPD-SX produced from this simple recrystallisation protocol.

6.3 Methodology

The methodology for this chapter is split into 3 sections:

1. Producing crystalline samples for SCXRD and SDPD-SX with a goal of obtaining the maximum number of single crystal / crystalline samples from a large collection of small amounts of material.
2. The collection of powder diffraction data for crystalline samples unsuitable for SCXRD using a SX diffractometer.
3. Attempting structure solution by SDPD-SX on the resulting powder diffraction data.
4. Determining the proportion of compounds that were able to be crystallised (both as SX and polycrystalline) to measure the potential impact of SDPD-SX.

Below is a description of each section in further detail.

6.3.1 Sample creation (recrystallisation protocol)

A collection of 81 drug-like organic compounds (maximum of 0.5 mg per sample) were used as a sample set for recrystallisation. These are composed of compounds available to purchase and those created 'in-house' by C4X Discovery LTD (C4XD). Prior to the recrystallisation experiments, all samples were anonymised by C4XD. Please note that whilst the recrystallisation experiments reported in this work were performed on all 81 compounds, not all compounds are listed here. Since the focus of the chapter is on (scoping) the application of SDPD-SX, only the 36 compounds found to be relevant are further discussed and reported. The compounds studied as part of this chapter are reported in Tables 6.2 and 6.3. Samples which crystallised into a suitable size single crystals were further analysed by C4XD and are not reported here.

Each starting compound was divided into two separate glass vials with septum lids (0.1 – 0.25 mg in each vial), group 1 was labelled 'blue' and group 2 was labelled 'red'. Where there was not enough material for two vials the available material was placed in a blue vial. Crucially, no vial contained more than 0.5 mg of material. Figure 6.2 illustrates the general recrystallisation steps undertaken for each of the vials with Table 6.1 outlining the solvent order used.

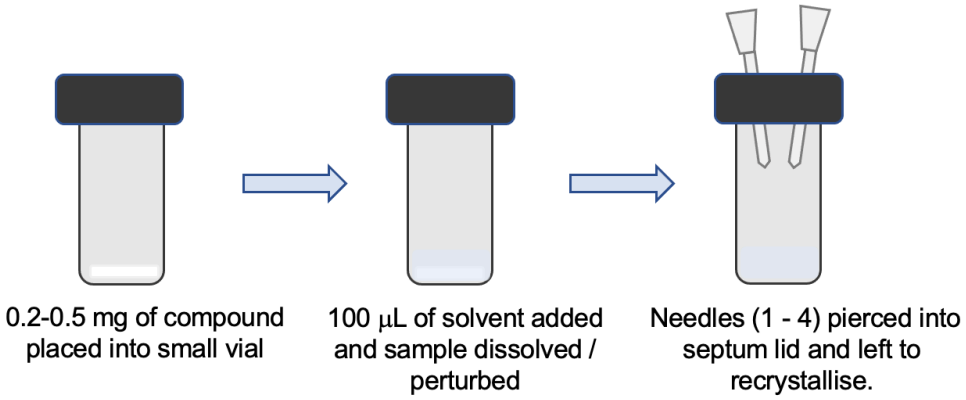


Figure 6.2. Schematic of the sample holders used for crystallisation experiments. Number of needles added depending on solvent volatility.

Each group of vials were attempted to be recrystalised with decreasingly volatile solvents (Table 6.1). After each attempt the vials were examined under cross-polarised light to assess the crystallinity / quality of each sample. Vials that contained suspected single crystals were set aside to be studied by SCXRD and were not included in the next recrystallisation iteration. As part of the 81 compounds examined, 10 biomolecules and derivatives (comprising many H-bond donors and acceptors) samples were included and were attempted to be recrystalised using methanol / ethanol.

Table 6.1. Solvent selection for recrystallisation

Group	**Solvent order***		
	1st	2nd	3rd
Red	Ethyl acetate	Diethyl ether	Tert-butyl methyl ether
Blue	Dichloromethane	Acetone	
Red (biomolecules and derivations)	Ethanol		
Blue (biomolecules and derivations)	Methanol		

Once all recrystallisation trials were completed, only potentially crystalline compounds with no single-crystal identified were investigated by SDPD-SX. Single crystal samples and SDPD-SX samples were separated after each round of recrystallisation. A general workflow of the methodology is given in Figure 6.3.

* Further solvents were not tested due to Covid-19 restrictions.

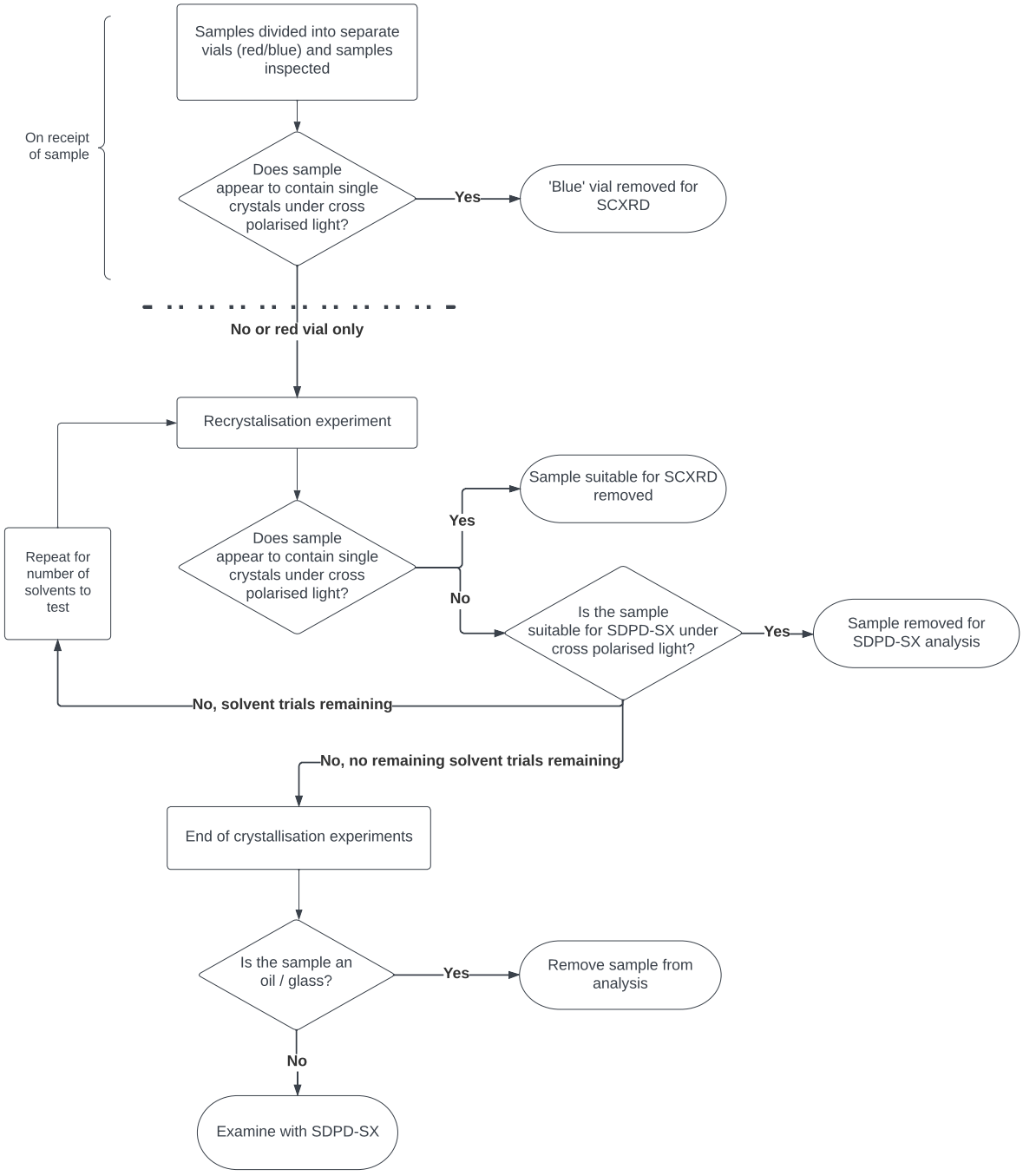


Figure 6.3. General flowchart of the recrystallisation protocol.

The compounds that were successfully crystallised with the described protocol are discussed in section 6.5.1.

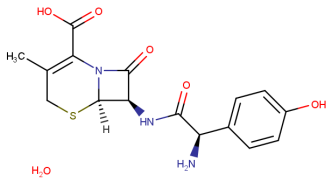
6.3.2 Data collection and processing

Of the 81 compounds (155 vials) that were recrystalised, 36 compounds (47 vials) were tested using SDPD-SX (see Table 6.2- 6.3).

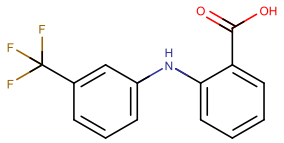
Table 6.2. Compounds analysed by SDPD-SX

Thesis	Compound name	Matching CSD reference
B1	Cefadroxil.H_2O	JOSWAP01
B2	Flufenamic acid	FPAMCA
B3	Mefenamic acid	XYANAC
B4	Sunitinib malate	PICVUU
B5	Suvorexant	RUKQOF
B6	Pemetrexed disodium heptahydrate	NOJXIW
C1	GSK1059865	-
C2	GSK-64-W0/2017/060854	-
C3	HC 030031	-
C4	-	-
C5	-	-
C6	-	-
C7	-	-
C8	-	-
C9	-	-
C10	3-(4-phenylbutanoylamino)pyridine-4-carboxylic acid	-
C11	61-10	-
C12	PDL1-BMS-W02015160641	-
C13	SU14813	-
C14	-	-
C15	-	-
C16	-	-
C17	-	-
C18	-	-
C19	-	-
C20	-	-
C21	Atorvastatin hemicalcium salt	-
C22	-	-
C23	-	-
C24	PHA665	-
C25	Rotigotine hydrochloride	-
C26	S-adenosyl-L-homocysteine (SAH)	-
C27	SB408124	-
C28	Streptomycin sulphate salt	-
C29	TCS1102	-
C30	TCS OX 29	-

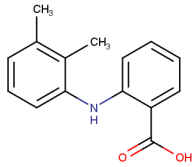
Table 6.3. Reported (available on the CSD) and previously unreported compound investigated by SDPD-SX



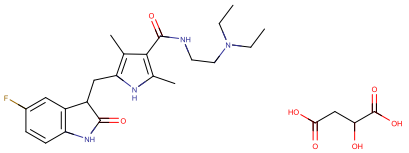
B1



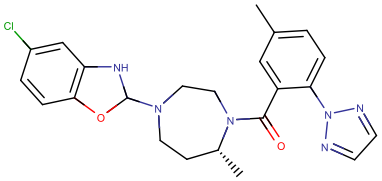
B2



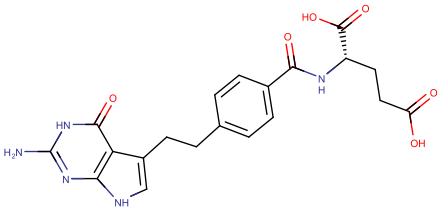
B3



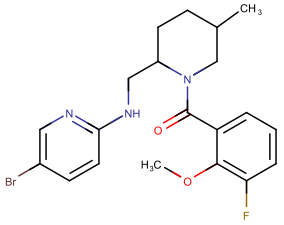
B4



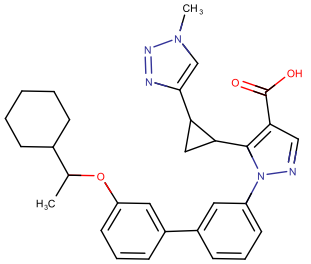
B5



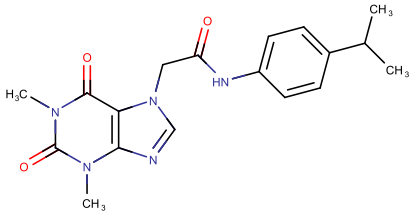
B6



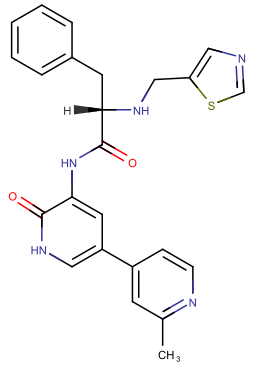
C1



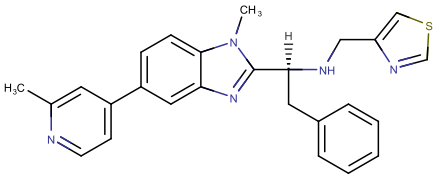
C2



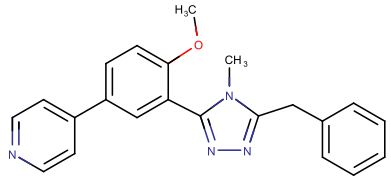
C3



C4

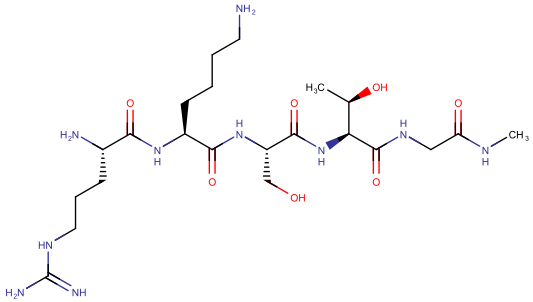


C5



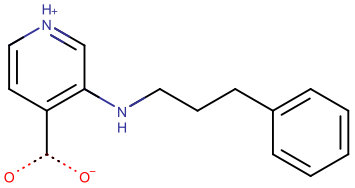
C6

Confidential

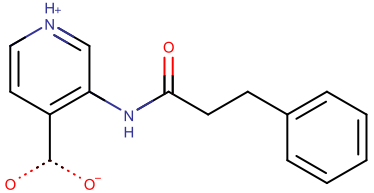


C7

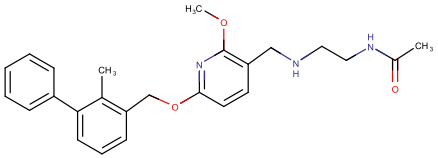
C8



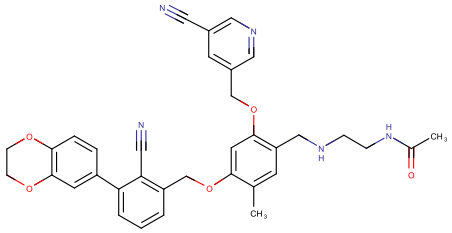
C9



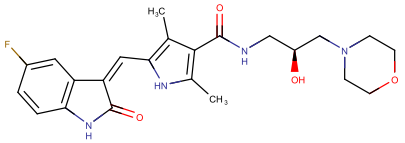
C10



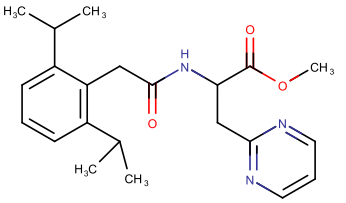
C11



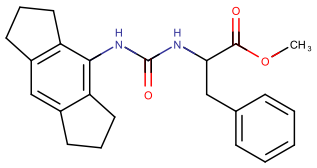
C12



C13



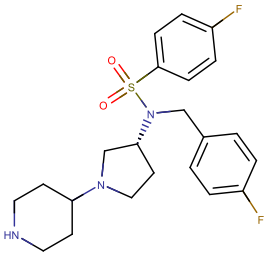
C14



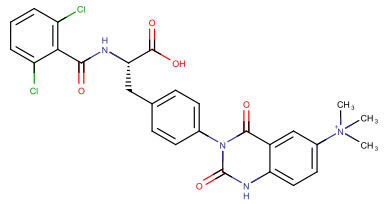
C15

Confidential

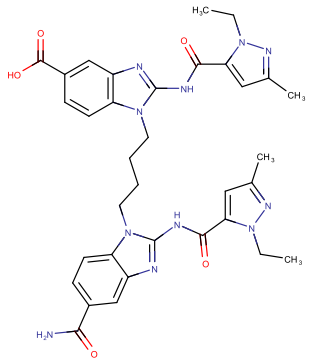
C16



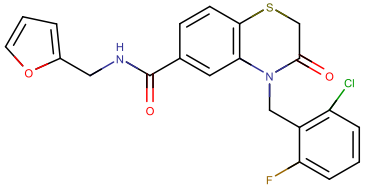
C17



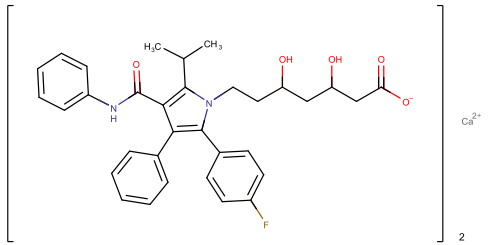
C18



C19



C20



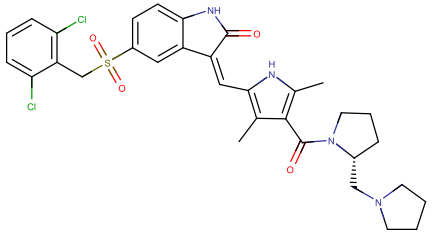
C21

Confidential

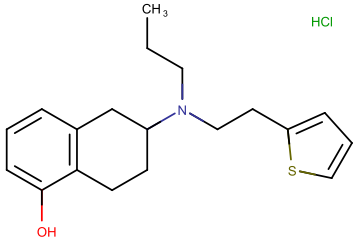
C22

Confidential

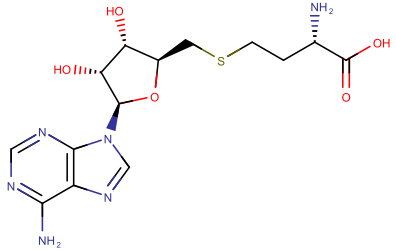
C23



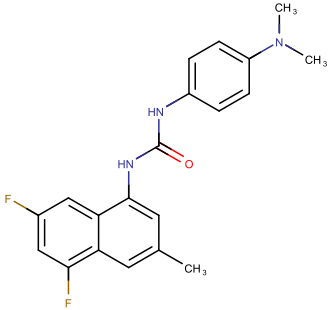
C24



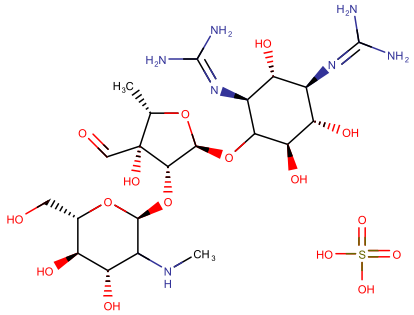
C25



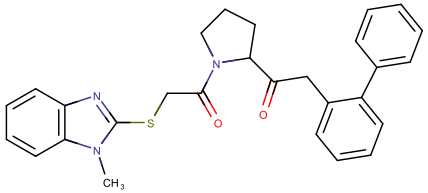
C26



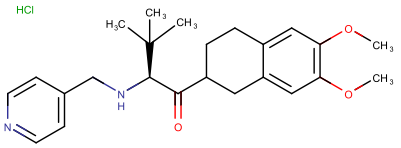
C27



C28



C29



C30

The SDPD-SX data collections for this chapter were performed at the University of Reading (UoR). The diffractometer is a dual microfocus source (Mo, Cu) Rigaku Synergy with PhotonJet-S X-ray source, equipped with a HyPix-6000HE detector.

A small sample was taken from the vial, combined with Paratone oil® and mounted to the tip of a fibre optic cable 'pip' (an in-depth description of this is given in chapter 5 section 5.5.1). A pre-experiment was performed for each sample to assess signal to background and to determine the optimal mode of collection (Table 6.4). Further details of data collection are discussed in Chapters 4 and 5.

Table 6.4. Pre-experiment parameters

Parameter	
Temperature / K	275 - 298
Collection mode	Standard phi
2θ (max) / °	40
Detector distance / mm	160
Beam divergence / mrad	0.1
Time per scan / s	150

For samples that were crystalline and with suitable signal to background, a full-experiment was performed (Table 6.5).

Table 6.5. Full experiment parameters

Parameter	
Temperature / K	275 - 298
Collection mode	Standard phi / Gandolfi for powder
2θ (max) / °	60 - 70
Detector distance / mm	160
Beam divergence / mrad	0.1
Time per scan / s	1200 - 3000

The diffraction rings were recentred from previously determined coordinates (see chapter 4) and .XYE file extracted. The datasets were scaled to a maximum peak height of 100,000 using the scaling python script (see Chapter 5).

6.3.3 Structure solution and assessment

6.3.3.1 SDPD-SX

After background removal, indexing was performed using DASH and EXPO2014. Indexable patterns were Pawley fitted in DASH. The unit cell was searched within the CSD, using ConQuest. For previously reported crystal structures the single crystal entry was used as a starting model for the simulated annealing algorithm within DASH. For novel structures, a 2D model of the structure was constructed in MarvinSketch and then optimised using Mercury's 'conformer generation feature'. The molecular model was then used as the input model for the simulated annealing structure solution step. Simulated annealing was performed using optimised parameters (Kabova *et al.,* 2017) using 20 - 250 runs, with 1 x 10^7 steps.[1,2,3,4,5]

The structure solution of previously reported structures returned by DASH were validated by comparing them to the available single crystal data using Mercury's 'crystal packing similarity' feature to obtain an RMSD value.

For previously unreported structures, the DASH solution with the lowest χ^2 value was energy minimised by DFT-D using Quantum Espresso (periodic density functional theory with van der Waals dispersion corrections, DFT-D3). Initial geometry optimizations were carried out with lattice parameters fixed at their crystallographic values, with subsequent variable cell geometry optimizations starting from the endpoint of the fixed-cell calculations. All calculations were carried out on the UK's National Tier 2 High Performance Computing Hub in Materials and Molecular Modelling, "Young".[6]

The end result of the DFT-D minimisation was compared to the input structure (the best DASH solution) using Mercury's 'crystal packing similarity' feature to obtain an RMSD value. For structures that differed significantly or had an end result that did not make chemical sense the structure solution was reattempted or considered unsolved. DFT minimised structures that were able to produce an RMSD value below *approx.* 0.4 with the best DASH structure (the input for DFT) were used as confirmation of the structure solution. Results for this are given in section 6.4.2.

6.3.3.2 Electron diffraction

3 samples which showed highly crystalline diffraction patterns but were unable to be indexed / Pawley fitted using SDPD-SX. These were analysed on my behalf by Rigaku, Japan using their Synergy-ED electron diffractometer. Results for this are given in section 6.4.3. The overall outcome of the compounds studied is given in Table 6.7 -6.10.

6.4 Results

6.4.1 Crystallisation protocol

The blind test started with 81 compounds (including 9 biomolecules and derivatives), 74 were divided into 2 vials, 7 were placed into a single vial, totalling of 155 vials before the initial recrystallisation experiments with various organic solvents.

Prior to first recrystallisation experiment, 6 compounds were identified under cross-polarised light as being suitable for single crystal analysis; for the samples with 2 vials, the blue vial was removed and the other was used for recrystallisation experiments. After the first recrystallisation attempt, 12 compounds (16 vials) were found to be suitable for SCXRD and were removed, whilst 5 compounds (5 vials) were removed for SDPD-SX.

After the second recrystallisation attempt, 10 compounds (11 vials) were found to be suitable for SCXRD and were removed, whilst 4 compounds (4 vials) were removed for SDPD-SX. After the third recrystallisation experiment, 13 compounds (13 vials) were found to be suitable for SCXRD and were removed.

In total, 31 unique compounds (46 vials) were removed as being suitable for SCXRD analysis. The remaining 109 vials were either candidates for SDPD-SX or were amorphous (for example in the form of an oil). The rest of the samples were examined under cross polarised light and apparent non-amorphous samples were examined by SDPD-SX

6.4.1.1 Pre-experiment

Samples that appeared to potentially be crystalline under cross- polarised light were examined by SDPD-SX. 30 unreported compounds (41 vials), and 6 previously reported compounds (6 vials), those with structures available on the CSD, were examined by a pre-experiment procedure described in chapter 5 (a total of 36 compounds in 47 vials). Compounds identified as amorphous or too poorly crystalline for further analysis are shown in Table 6.8.

Table 6.6. Samples found to be amorphous or poorly crystalline from pre-experiment

Reported compounds	**Previously unreported compounds**
Cefadroxil.H_2O (blue)	C4 (blue)
Suvorexant (red)	C7 (blue)
Pemetrexed disodium heptahydrate (blue)	C8 (red)
	C11 (blue)
	C17 (red)
	C29 (blue)
	Streptomycin (red)

6.4.1.2 Full experiment

For samples that were shown to be crystalline from the pre-experiment, a full-experiment procedure was performed. From the initial samples, 25 unreported compounds (33 vials) were collected in a long collection, 3 reported compounds (3 vials) were collected in a long collection. A total of 36 vials were used for a full data collection.

6.4.2 Structure solution attempts

Of the 6 APIs with a known crystal structure (with structures deposited on the CSD), 3 compounds were solved by SDPD-SX (Table 6.7).

Of the 30 previously unreported compounds tested by SDPD-SX, structure solution by SDPD-SX was obtained for 6 compounds (9 vials) and 1 compound was indexed but was unable to be solved a summary of the outcomes is given in Table 6.8.

A summary of the structure solutions/ indexed structure is given below in Table 6.9, demonstrating the use of SDPD-SX with previously unreported structures. The remaining 14 compounds were unsuitable for indexing due to a suspected large unit cell, impurities, and/or suspected mixtures. A summary and description of the outcomes for unindexed compounds is given below in Table 6.10.

Table 6.7. Previously reported compounds outcome post full-data collection

Compound name	**REFCODE**	**DoF**	**SpGrp**	**Res / Å**	**χ^2_{Paw}**	**N_{ref}**	**%success**	**χ^2_{prof}**	**RMSD / Å**
Flufenamic acid (red)	FPAMCA	10	*C*2/c	2.056	5.4	149	45	80.06	0.042
Sunitinib malate (blue)	PICVUU	24	$P2_1$	2.313	0.87	131	4	3.71	0.103
Mefenamic acid (red)	XYANAC01	9	$P\bar{1}$	2.274	3.56	111	50	17.04	0.055

DoF = Degrees of freedom, SpGrp = space group, Res= real-space resolution, χ^2_{Paw} = best Pawley fit value, N_{ref} = Number of reflections, %success = % of runs achieving structure solution, χ^2_{prof} = best DASH SA solution, RMSD with FC = root mean square deviation as measured by Mercury's Crystal Packing Similarity functionality between best DASH χ^2_{prof} with fixed cell and CSD refcode.

Table 6.8. Unreported compound outcome post full-data collection

Solved SDPD-SX	Solved by ED	Indexed	Unable to index
C1 (blue)	C13 (blue)	C14 (blue)	C2 (red/blue)
C3 (red/blue)	C20 (blue)		C5 (red)
C10 (red/blue)			C6 (blue)
C15 (red/blue)			C9 (red)
C25 (red/blue)			C11 (red)
C26 (red)			C12 (red/blue)
			C16 (red)
			C18 (blue)
			C19 (red/blue)
			C21 (red)
			C22 (blue)
			C23 (red)
			C24 (red)
			C27 (red/blue)
			C30 (red)

Table 6.9. Unreported compound indexing and structure solution summary

Compound	Mol. formula	DoF	V_{mol} / Å^3	V_{cell} / Å^3	SpGrp	χ^2_{Paw}	N_{ref}	Res / Å	%succ.	χ^2_{prof}	RMSD (FC) / Å
C1 (blue)	$C_{20}H_{23}N_3O_2BrF$	12	496.29	2027.29	$P2_12_12_1$	14.65	143	2.1741	85	117.27	0.269
C3 (blue)	$C_{18}H_{21}O_3N_5$	11	454.34	3634.80	$Pcab$	1.84	104	2.6049	47	9.15	0.163
C10 (blue)	$C_{15}H_{13}O_3N_2$	12	331.11	624.75	$P2_1$	22.66	137	1.7893	85	77.04	0.208
C10 (red)	$C_{15}H_{13}O_3N_2$	12	331.11	624.75	$P2_1$	7.96	135	1.8062	85	70.54	0.205
C14 (blue)	$C_{21}H_{29}O_3N_3$	32	508.16	2079.30	$P2_1$	/	/	/	/	/	/
C15 (red/blue)	$C_{26}H_{26}O_3N_2$	14	550.47	1951.20	$P2_12_12_1$	4.17	130	2.3190	95	24.82	0.176
C13 (blue)	$C_{22}H_{26}O_4N_3F$	14	529.35	1116.50	$P\bar{1}$	/	/	/	/	/	/
C20 (blue)	$C_{20}H_{16}SO_3N_2FCl$	24	478.62	4090.00	$Pca2_1$	/	/	/	/	/	/
C25 (red)	$C_{19}H_{24}ONS.HCl$	15	464.72	1895.76	$P2_12_12_1$	5.78	142	2.1158	95	85.41	0.224 (SX)
C25 (blue)	$C_{19}H_{24}ONS.HCl$	15	464.72	1897.98	$P2_12_12_1$	1.37	176	1.9513	95	23.77	0.307 (SX)
C26 (red)	$C_{14}H_{20}O_5N_6S$	13	448.73	1826.62	$P2_12_12_1$	13.32	121	2.4996	10	152.56	0.200

DoF = Degrees of freedom, SpGrp = space group, χ^2_{Paw} = best Pawley fit value, N_{ref} = Number of reflections, Res = real space resolution %succ. = % of runs achieving structure solution, χ^2_{prof} = best DASH SA solution, RMSD (FC) = root mean square deviation as measured by Mercury's Crystal Packing Similarity functionality between best DASH χ^2_{prof} with fixed cell and DFT-D minimised structure with fixed unit cell values.

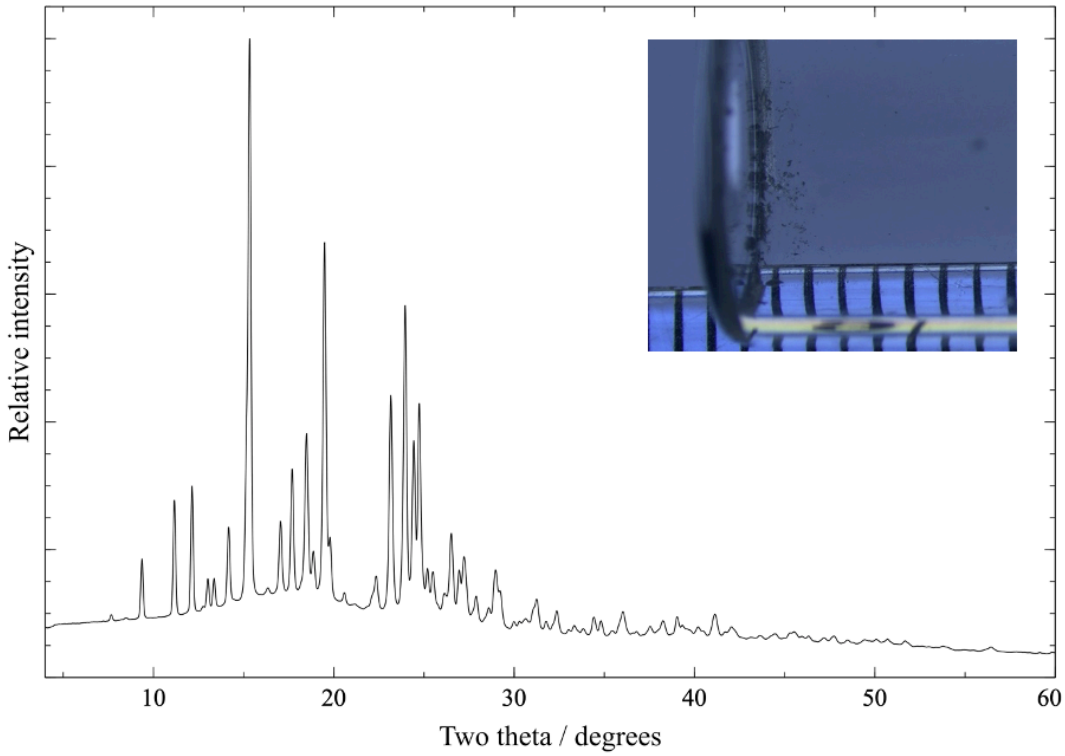
Table 6.10. Unreported crystalline compounds post full-data collection that could not be solved with SDPD-SX.

Compound	Outcome	Reasoning
C2 (red/blue)	Not indexed	Suggested large unit cell
C5 (red)	Not indexed	Both DASH and EPXO2014 suggest very large monoclinic cells; possible impurity peaks at 12.5° and 13.8° 2θ
C6 (red)	Not indexed	Possible mixture with different C6 compared to C6 blue SX form. Partially melted during collection
C9 (red)	Different index outcomes	Potential phase mixture
C11 (red)	Not confidently indexed	Suggested large unit cell, first peak below 3 two theta, potential phase mixture
C12 (red/blue)	Not indexed	Salt impurities
C14 (blue)	Indexed	Unable to solve due to disorder within unit cell
C16 (red)	Not indexed	Different form from C16 SX (blue)
C18 (blue)	Not indexed	Large background, ED was unsuccessful due to sample amount
C19 (red/blue)	Not indexed	Salt impurities
C21 (blue)	Not indexed	Poorly crystalline / large amorphous component
C22 (blue)	Not indexed	Suggested large unit cell
C23 (red)	Not indexed	Potential impurity
C24 (blue)	Not indexed	Poorly crystalline / large amorphous component
C27 (red/blue)	Not indexed	Multiple index outputs
C30 (red)	Not indexed	Poorly crystalline / large amorphous component
C13 (red/blue)	Solved with ED	Very large unit cell, solved with ED but not Pawley fitted, likely due to poor crystallinity
C20 (blue)	Solved with ED	Large unit cell

6.4.3 Previously unreported structure solution results by SDPD-SX

The following figures summarise the sample morphology, integrated diffraction data and solved structures for previously unreported structures of samples unsuitable for SCXRD.

6.4.2.1 Structure solution of C1



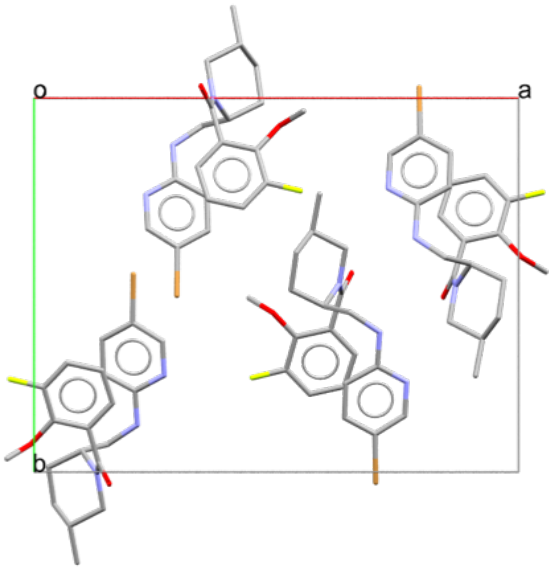


Figure 6.4. PXRD data collected from sample C1 (blue), above. The inset picture shows the polycrystalline sample of C1 in sample vial, with a 1 mm scale superimposed. The solved crystal structure of C1 is shown below. Periodic DFT optimisation of the solved structure gave an optimised structure that yielded an RMSD of 0.269 Å in a 15-molecule crystal packing similarity overlay in Mercury.

6.4.2.2 Structure solution of C3

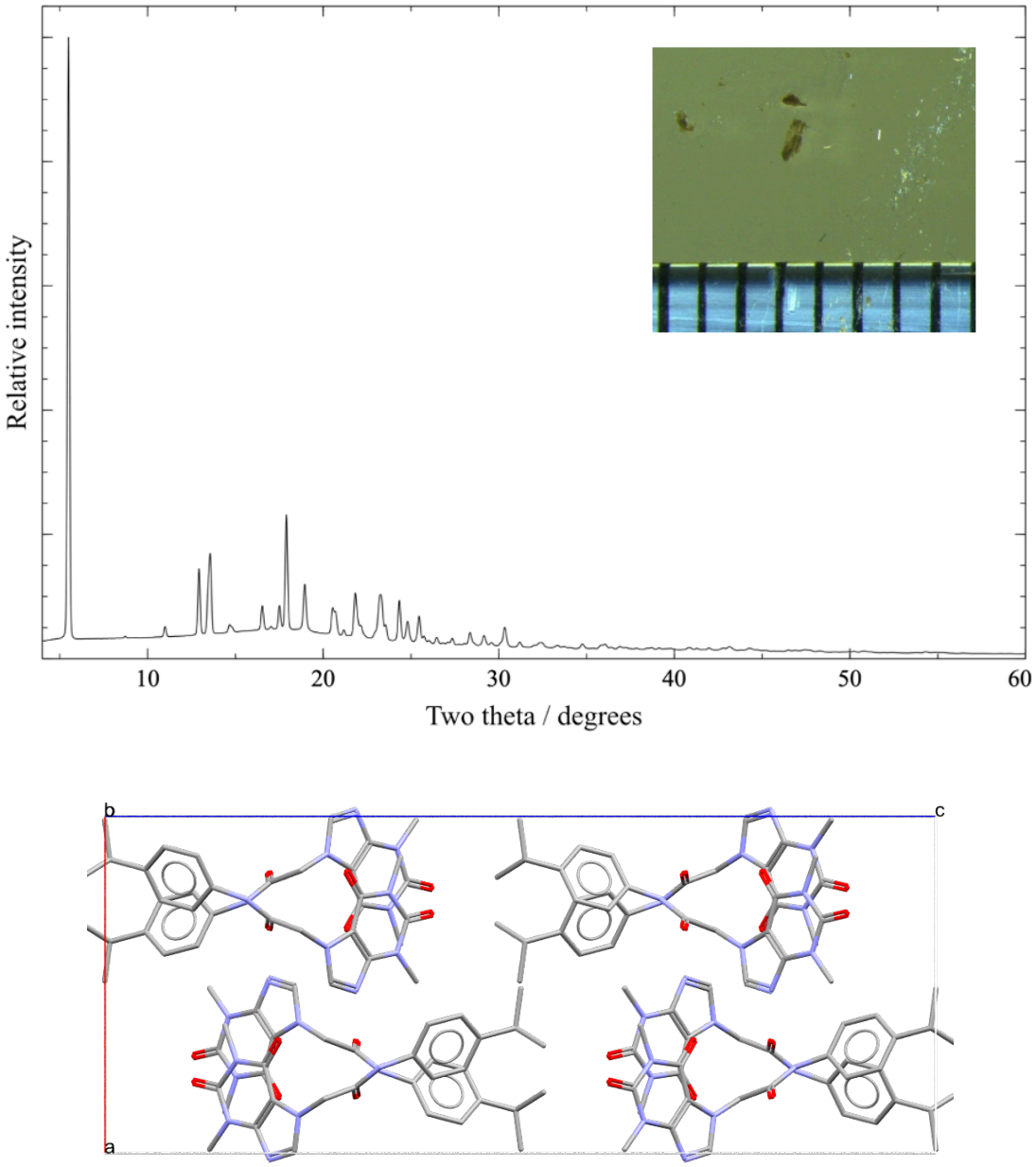


Figure 6.5. PXRD data collected from sample C3 (blue), above. The inset picture shows the polycrystalline sample of C3, with a 1 mm scale superimposed. The solved crystal structure of C3 is shown below. Periodic DFT optimisation of the solved structure gave an optimised structure that yielded an RMSD of 0.163 Å in a 15-molecule crystal packing similarity overlay in Mercury.

6.4.2.3 Structure solution of C10

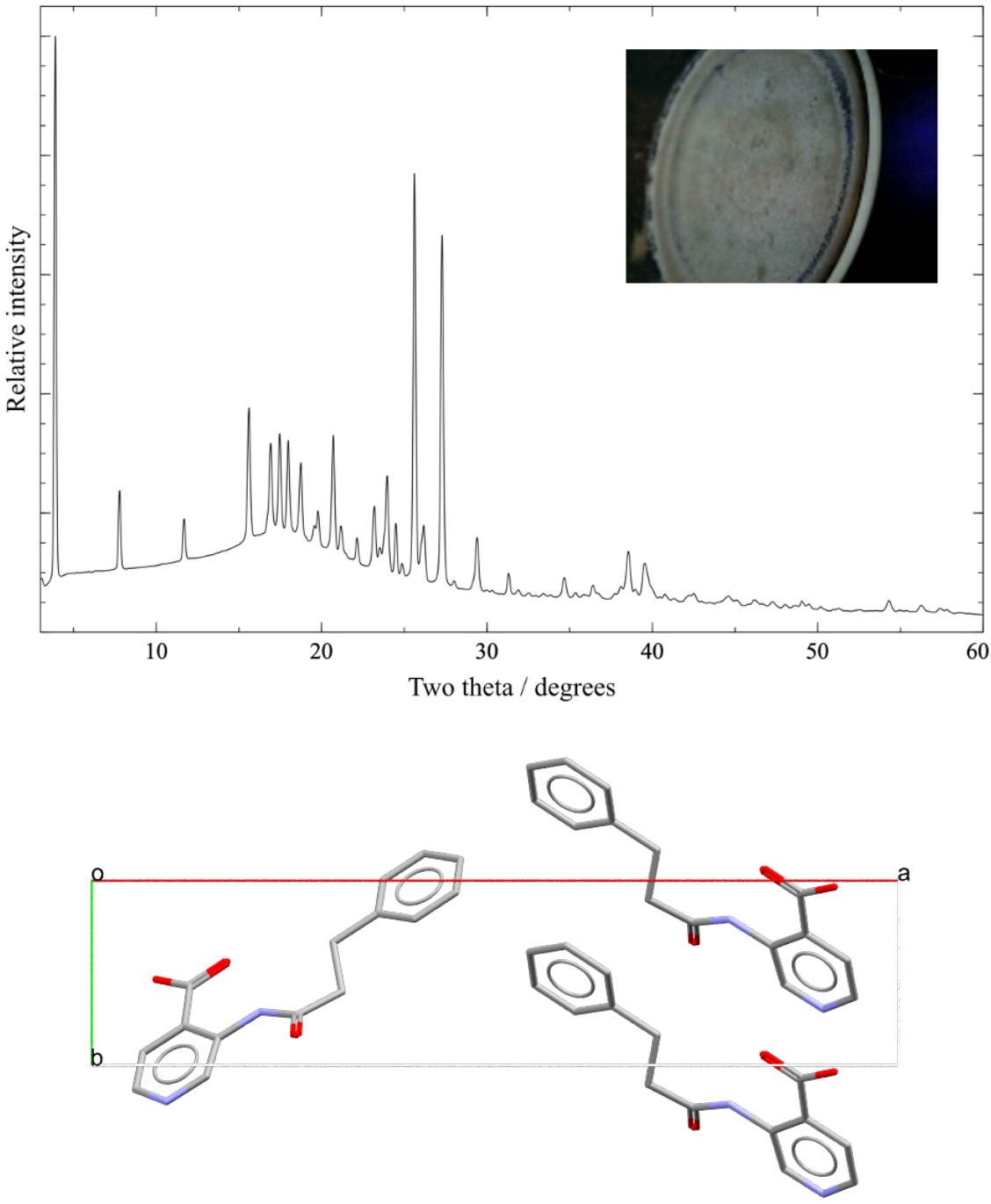
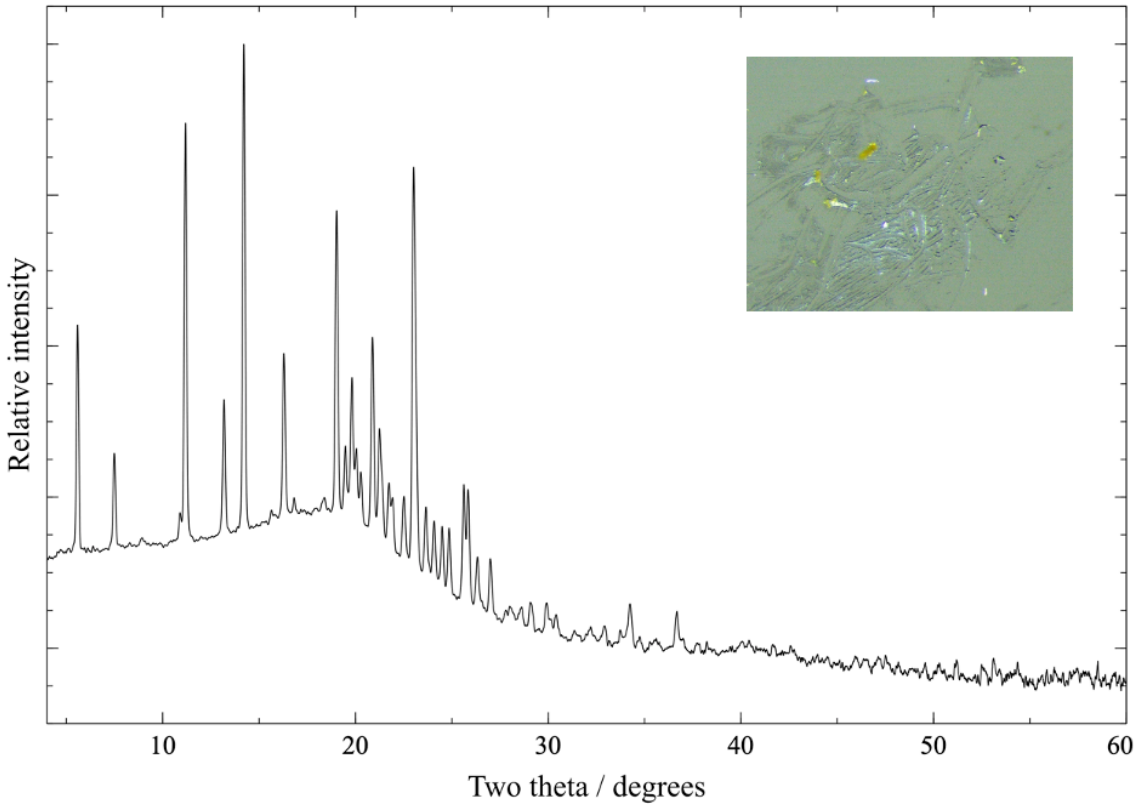


Figure 6.6. PXRD data collected from sample C10 (red), above. The inset picture shows the polycrystalline sample of C10 in sample vial. The solved crystal structure of C10 is shown below. Periodic DFT optimisation of the solved structure gave an optimised structure that yielded an RMSD of 0.205 Å in a 15-molecule crystal packing similarity overlay in Mercury.

6.4.2.4 Structure solution of C15



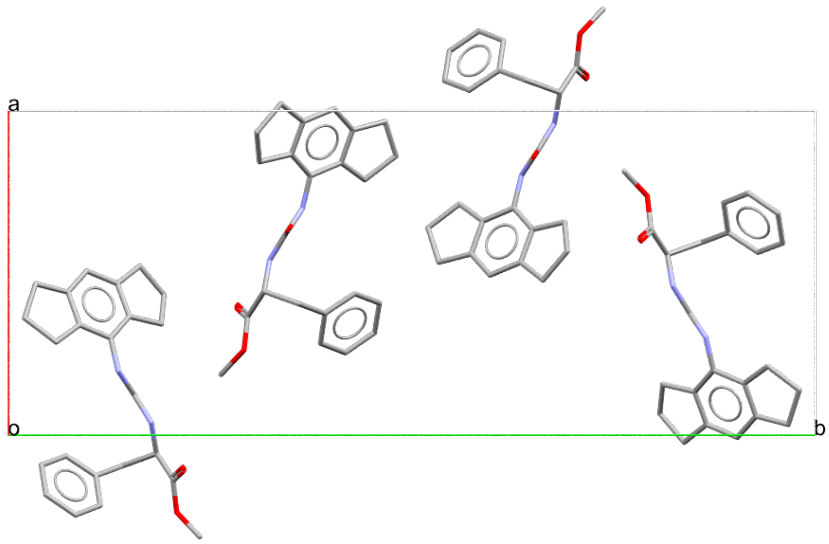


Figure 6.7. PXRD data collected from sample C15 (blue), above. The inset picture shows the polycrystalline sample of C15 on glass slide. The solved crystal structure of C15 is shown below. Periodic DFT optimisation of the solved structure gave an optimised structure that yielded an RMSD of 0.176 Å in a 15-molecule crystal packing similarity overlay in Mercury.

6.4.2.5. Structure solution of C25, rotigotine hydrochloride

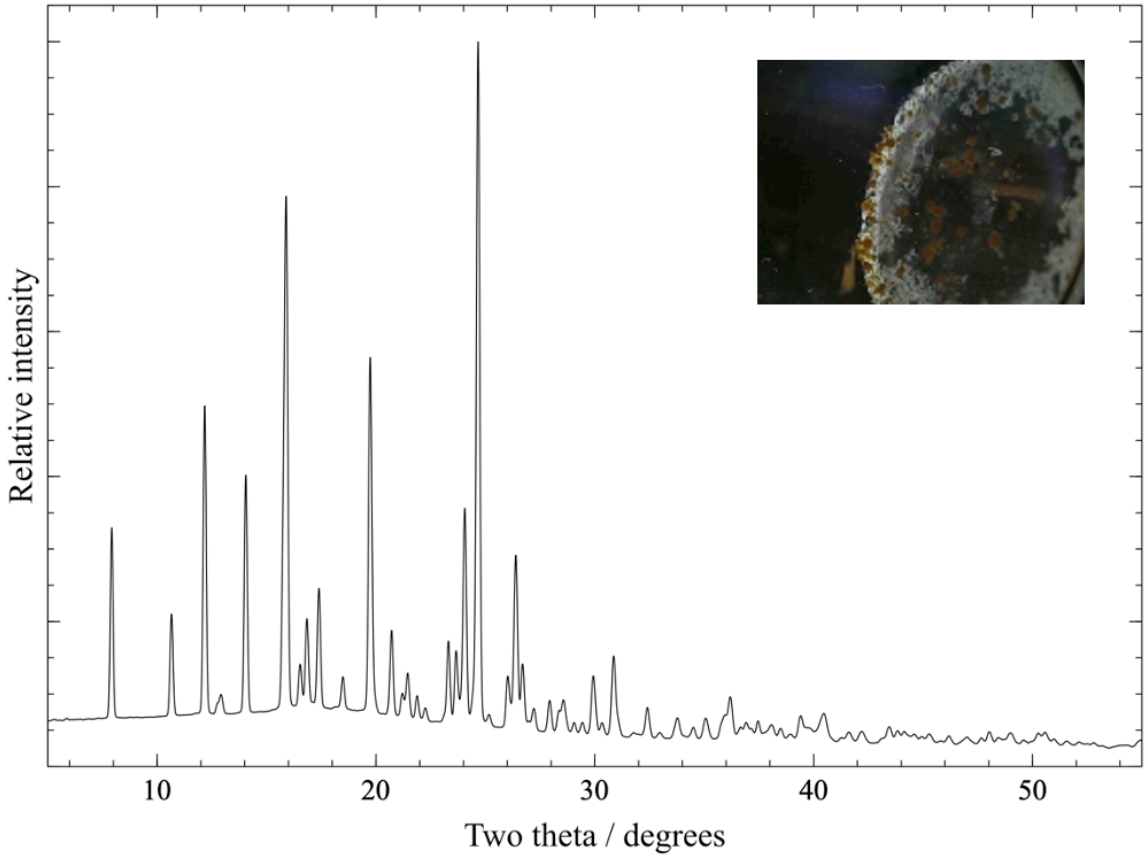


Figure 6.8. PXRD data collected from sample C25, rotigotine hydrochloride (red), above. The inset picture shows the polycrystalline sample of rotigotine hydrochloride in sample vial.

PXRD data were collected using samples from both the red and blue vials, both datasets indexed and Pawley-fitted to high accuracy. The 'red' and 'blue' solved structures differed only by a 180° rotation in the S-containing ring; a single crystal structure obtained subsequently by C4X showed this ring to be disordered. This is discussed in more detail in section 6.5.3.3.

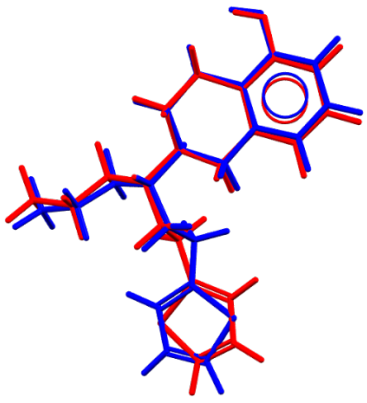


Figure 6.9. DASH's simulated annealing outcomes for C25, rotigotine hydrochloride (red and blue) overlayed using Mercury's molecule overlay feature.

6.4.2.6 Structure solution of C26, *S*-adenosyl-L-homocysteine (SAH) red

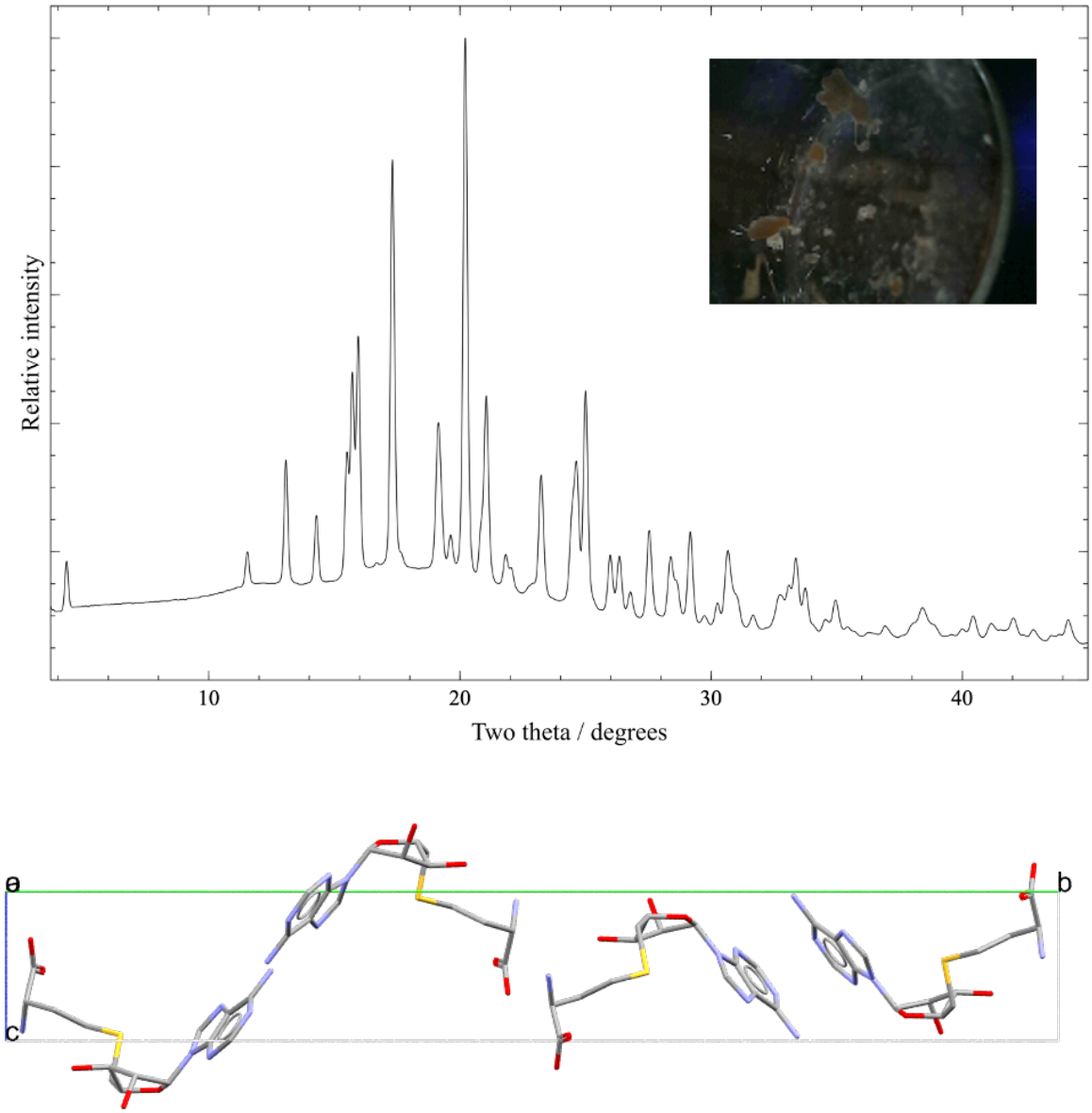


Figure 6.10. PXRD data collected from C26, *S*-adenosyl-L-homocysteine, SAH (red), above. The inset picture shows the polycrystalline sample of SAH in sample vial. The solved crystal structure of C10 is shown below. Periodic DFT optimisation of the solved structure gave an optimised structure that yielded an RMSD of 0.200 Å in a 15-molecule crystal packing similarity overlay in Mercury.

6.4.3 Structures obtained using Electron diffraction

Of the samples that were unable to be indexed, 3 compounds (C13 blue, C18 blue, C20 blue) were deemed sufficiently crystalline to merit sending for electron diffraction; of these, structures for C13 (blue) and C20 (blue) were obtained whilst there proved to be insufficient C18 (blue) sample to allow preparation for ED.

6.4.3.1 Structure solution of C13

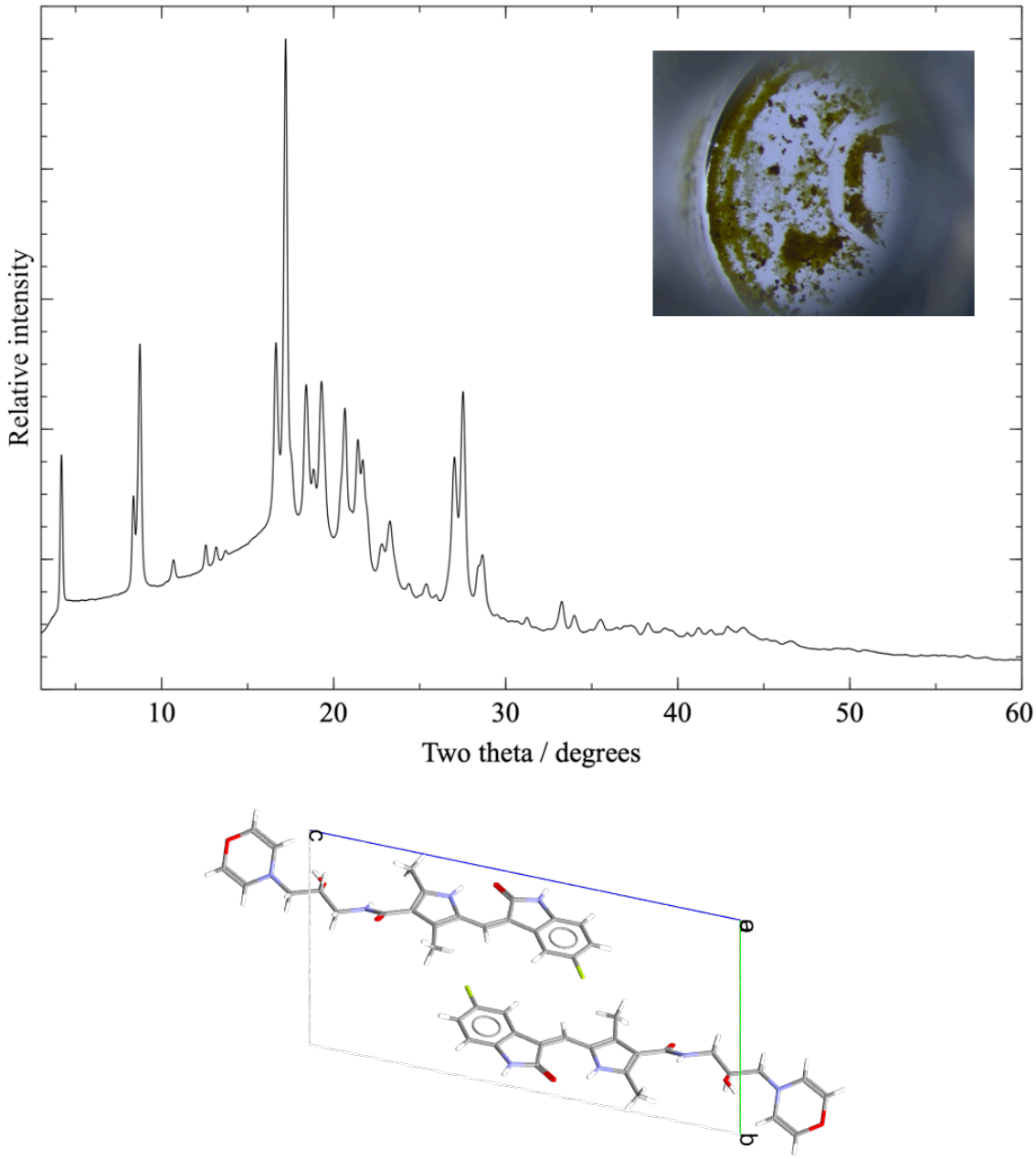


Figure 6.11. PXRD data collected from sample C13 (blue), above. The inset picture shows the polycrystalline sample of C13 in sample vial. The crystal structure of C13, determined by Rigaku, Japan using electron diffraction, below.

6.4.3.2 Structure solution of C20

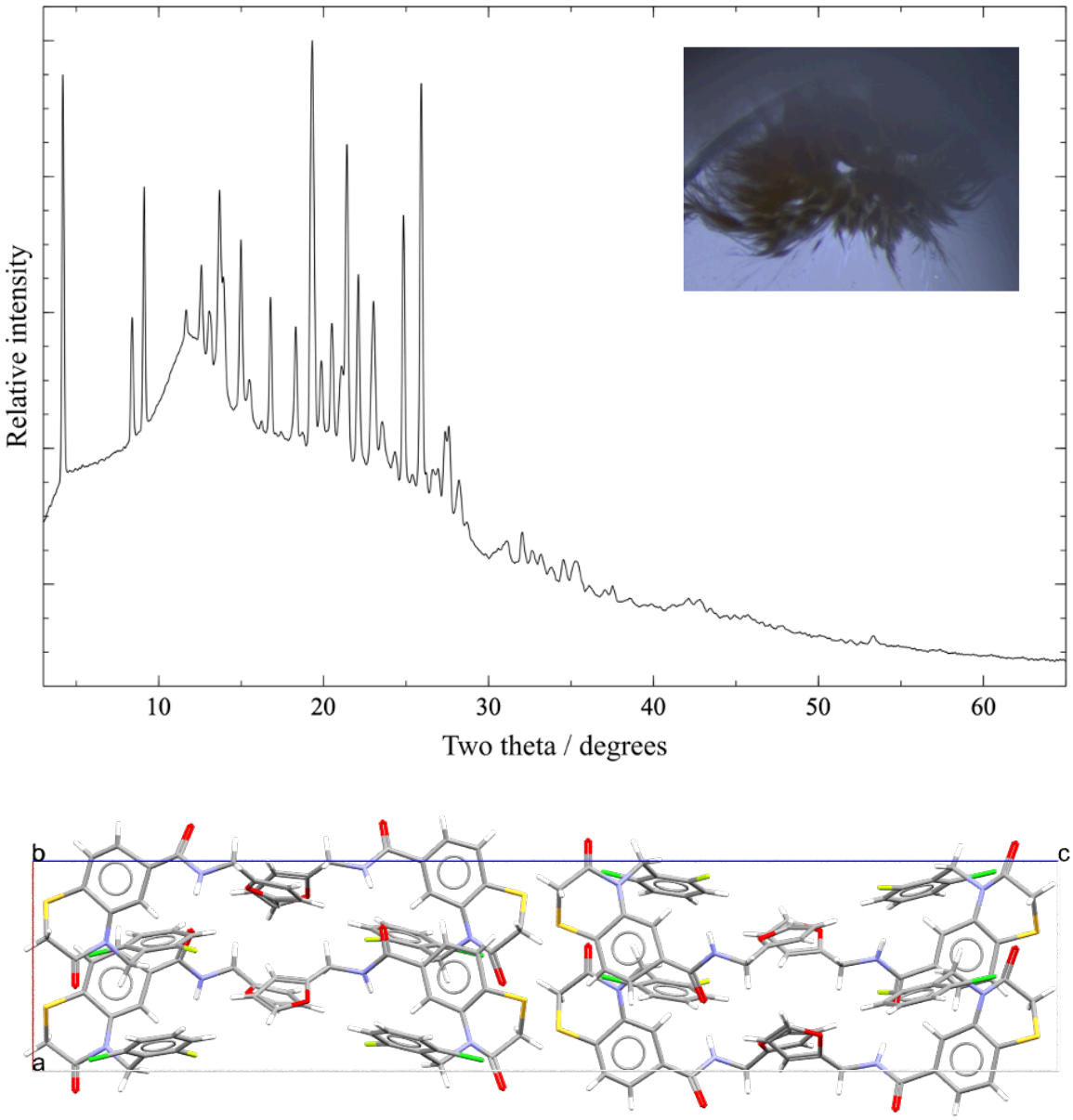


Figure 6.12. PXRD data collected from sample C20 (blue), above. The inset picture shows the polycrystalline sample of C20 in sample vial. The crystal structure of C20, determined by Rigaku, Japan using electron diffraction, below.

Analysis revealed that C20 (blue) was not able to be analysed by SDPD-SX because the powder pattern contained a phase mixture of two similar triclinic cells.

6.5 Discussion

6.5.1. Recrystallisation protocol

The recrystallisation protocol was applied to 81 organic compounds, composed of 72 drug-like organic compounds (137 vials) and 9 biomolecules and derivatives (18 vials). A summary of the number of crystalline samples in given below in Figure 6.13. 50 compounds were found to have at least one vial with a diffracting sample, corresponding to 11% of the biomolecules and derivatives and 80% of the drug-like molecules.

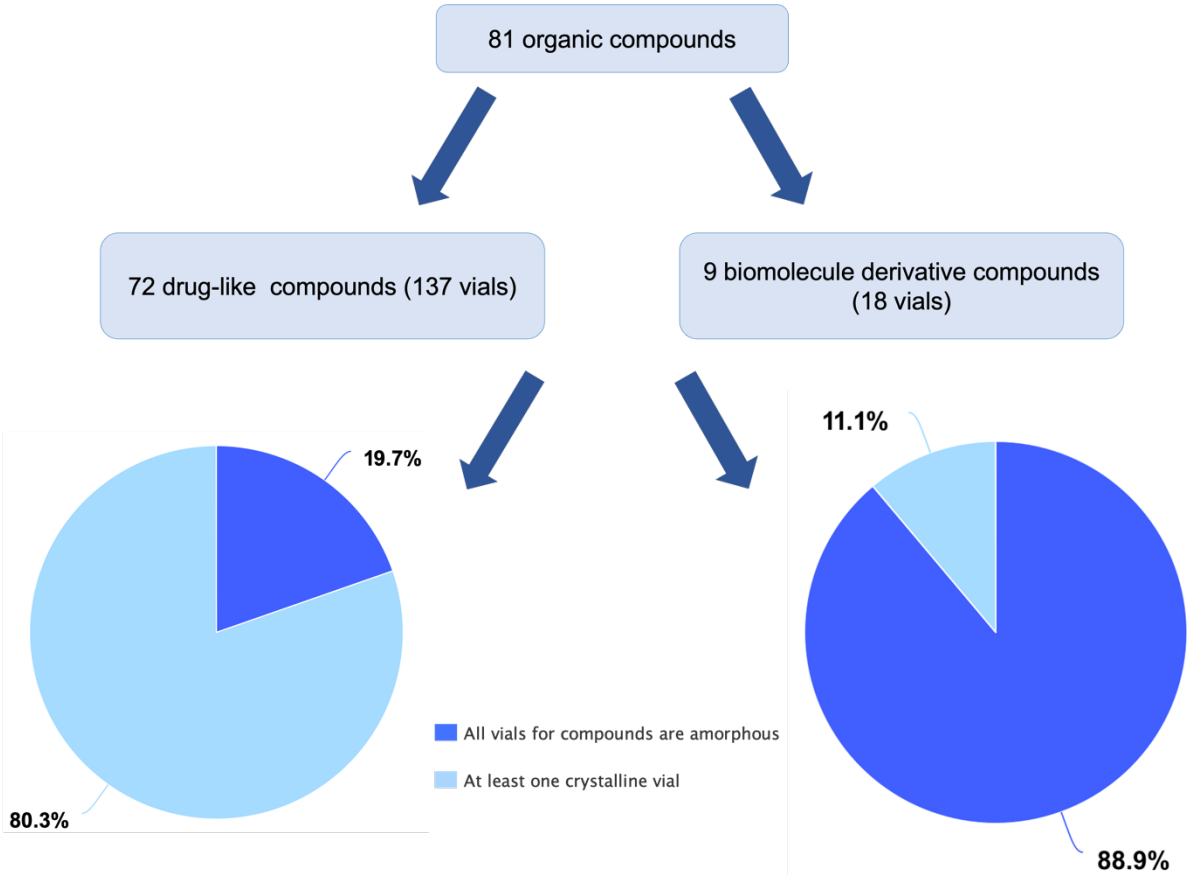


Figure 6.13. Outcome of the recrystallisation protocol, separating the different compound groups identifying the number of compounds that were crystalline in at least one vial.

From the vials containing a crystalline sample, a total of 33 compounds (44 vials) were separated as being suitable for SCXRD and were subsequently analysed by C4X Discovery. This resulted in 31 single crystal structures (38%) being determined from SCXRD. From the remaining samples (those not separated for SCXRD), 27 compounds (36 vials) were found, by SDPD-SX, to be crystalline. Of the 27 compounds used for a full SDPD-SX data collection,

9 were solved (11% of compounds). It should be noted that the only structure solution from the set of biomolecules and derivatives was C26, SAH, solved by SDPD-SX.

6.5.1.1 Protocol considerations

Crystallographers generally use a variety of recrystallisation methods (such as slow cooling, seeding, anti-solvent systems *etc*) to maximise the chances of obtaining a suitable single crystal. However, these different routes are more difficult to perform than slow evaporation to perform and risk losing/contaminating the limited sample, particularly when performed 'by hand' with > 50 samples. The focus of the protocol was to achieve the maximum crystalline output with the least input of time and expense in the most efficient way. Hence samples were recrystalised by slow evaporation at a constant temperature, with the evaporation controlled by the number of needles piercing the vial lid. Once each solvent evaporated, the samples were inspected using a cross polarising light microscope, and samples that appeared amorphous were redissolved with the next solvent. The initial protocol design was to perform the serial re-crystallisation with a greater number of solvents, however, due to the COVID-19 restrictions on work within laboratories, that was not possible; resulting in a total of 5 solvents being tested. Nevertheless, the protocol that is outlined here can easily be expanded to include a greater number of vials and or solvents.

6.5.1.2 Application to drug discovery

As mentioned throughout this thesis (see section 6.1), during the very early stages of drug discovery, relatively few attempts are made to crystalise individual samples. This is due to the focus of early research being key results from relevant assays to demonstrate the desired activity. The remaining material (<1 mg) often does not form single crystals and is not enough material for in-house powder diffraction, with the use of specialist facilities being too costly for the information gained. As a result, structural information is in many ways underutilised, particularly as advancements in computational chemistry allow structural information to play a crucial role in future drug discovery programs.

One of the aims for this chapter was to maximise the crystal structure output using a protocol with minimal time and materials spent recrystallising samples. It is important to note that novel materials are often initially amorphous because often the end step for a synthetic route for new materials is drying step via rotary evaporation. This means that for structural analysis (re)crystallisation is a necessary first step.

6.5.1.3 Solvent choice

Solvents were chosen initially on the basis of their volatility and ease-of-access; the more volatile the collection of solvents, the greater number of recrystallisations that can be performed within a chosen time frame. According to a recent study on solvent properties, certain parameters describing empirically solute/ solvent interactions can be used to help with solvent choice for a desired application. These parameters are dispersion and induction (DI), electrostatic interactions between permanent multipoles (ES), Lewis base / solvent Lewis acid interactions (α_1) and solute hydrogen bond donor / solvent hydrogen bond acceptor interactions (β_1). The seven solvents used in this work provide a reasonable span of this property space.[7]

6.5.2 Sample morphology

Ideal single crystals for SCXRD are simple to identify under a cross polarised light microscope. Such crystals are uniform, single, and change transmission under cross-polarised light. Practiced crystallographers often do not spend a significant amount of time identifying an 'ideal' crystal and experiments are performed without much delay. The same cannot be stated for non-ideal samples for SCXRD as they may be non-uniform, stacked or have no clear boundaries from each other (for example, in the case of dendrimers, see section 6.1).

However, a 'good powder' for powder diffraction on the scale of SDPD-SX is a term that had not yet been explored. It was found during this work that polycrystalline samples that formed 'frosted spheres' or similar were ideal for sample handling and resulted in good quality diffraction patterns collected. Figure 6.14 shows C18 (blue), C25, rotigotine hydrochloride (red) and mefenamic acid with a morphology that is easy to lift from the slide, with minimal oil, and does not need to be ground or crushed to form uniform diffraction rings. It is therefore recommended that samples with this morphology should be explored with SDPD-SX before further recrystallisation.

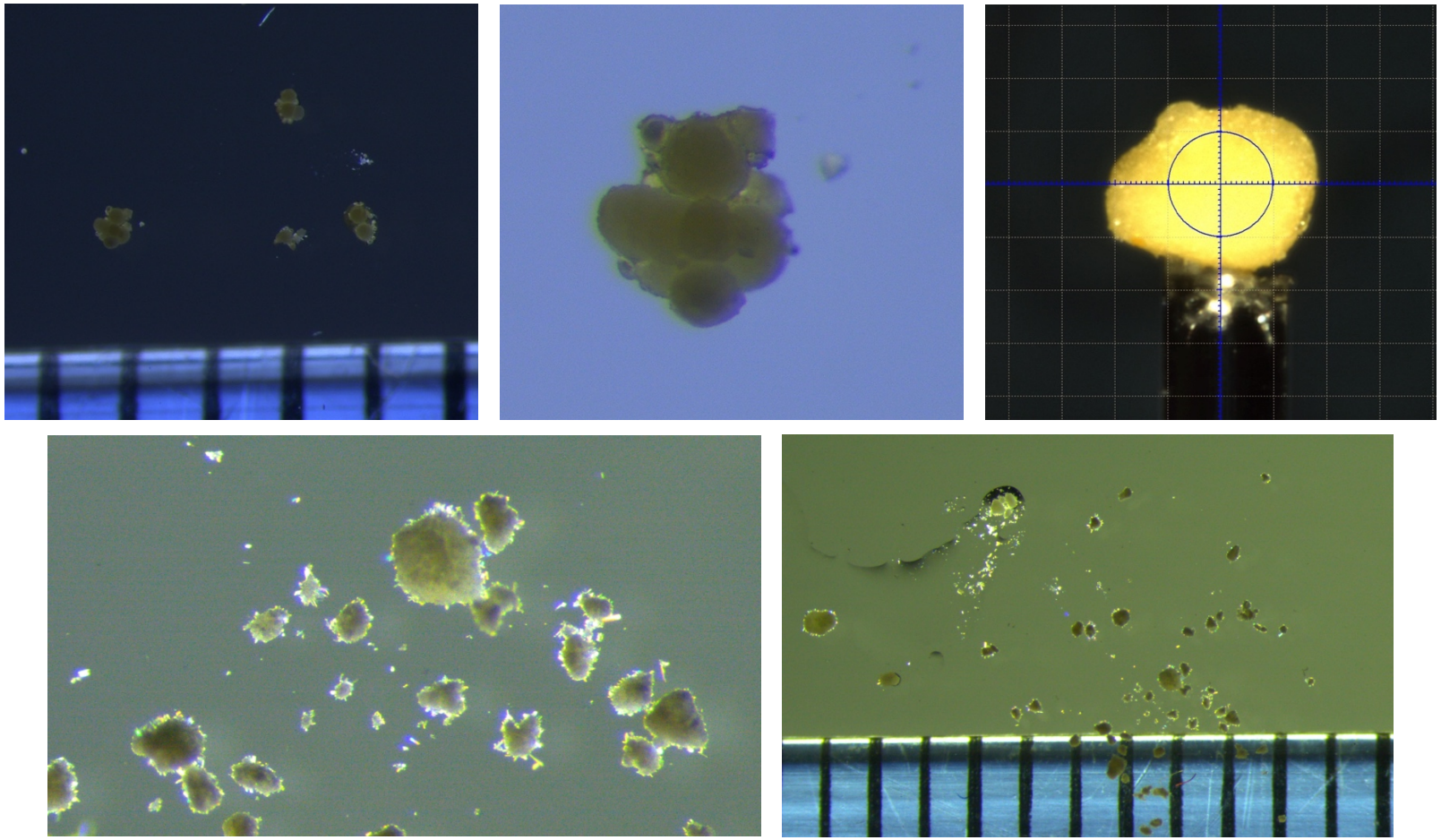


Figure 6.14. ‘Frosted sphere’ morphology seen in C18 (blue, above middle) with reference to 1 mm spacing (above left). C25, rotigotine hydrochloride (red) is shown with a similar morphology mounted on an optical fibre with 0.1 mm grid (above right) and mefenamic acid is shown (below left) and with respect to 1 mm spacing (below right).

Other morphologies that were opaque and with sharp edges indicated the presence of microcrystals under cross polarised light, such as those in Figure 6.15. These samples, once crushed and mounted with oil were found to yield a positive result with SDPD-SX.

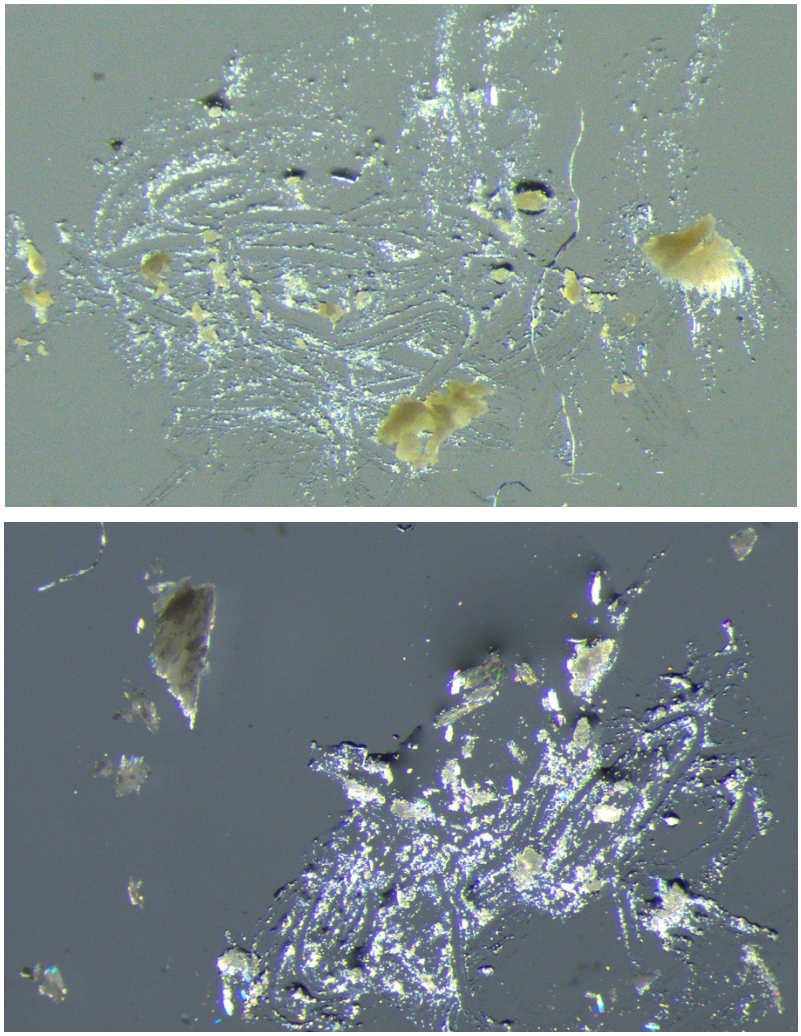


Figure 6.15. C2 (red) above and C16 (red) below, with morphologies indicating the presence of microcrystals SDPD-SX.

Transparent samples with non-uniform shapes and smooth edges were suspected to be amorphous and upon examination by SDPD-SX were found to be amorphous or too poorly crystalline for further analysis. An example of this is shown below in Figure 6.16.

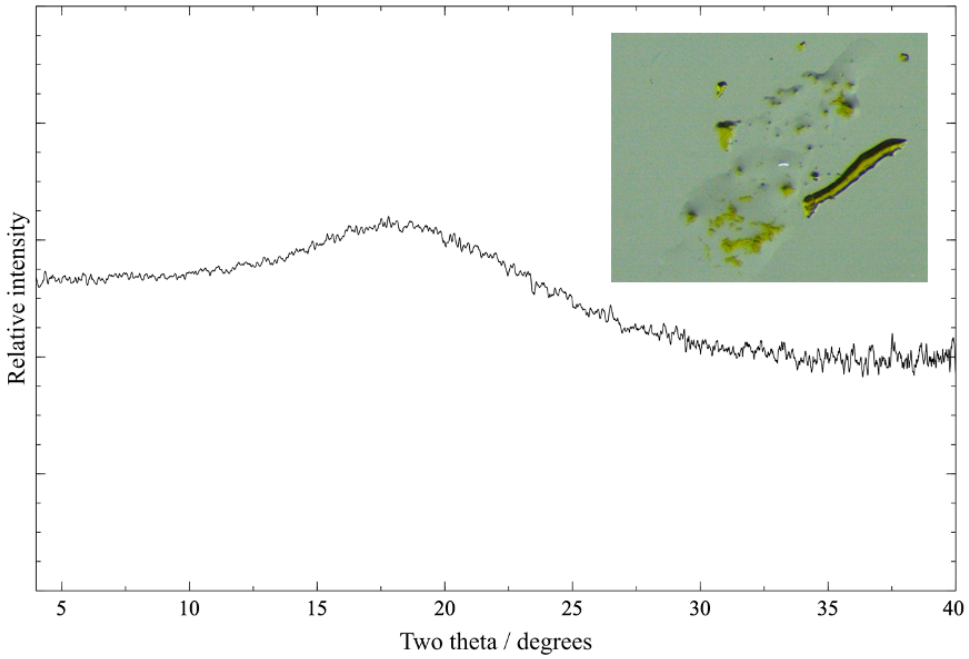


Figure 6.16. A sample of C7 (blue) inset, and its corresponding diffraction pattern.

There is no definite way to know how crystalline a powder is using a cross-polarised light microscope; some samples may appear amorphous but are actually crystalline and vice versa. Figure 6.17 shows a sample of streptomycin that on initial inspection appeared similar in appearance to other polycrystalline samples, but which gave an amorphous pattern.

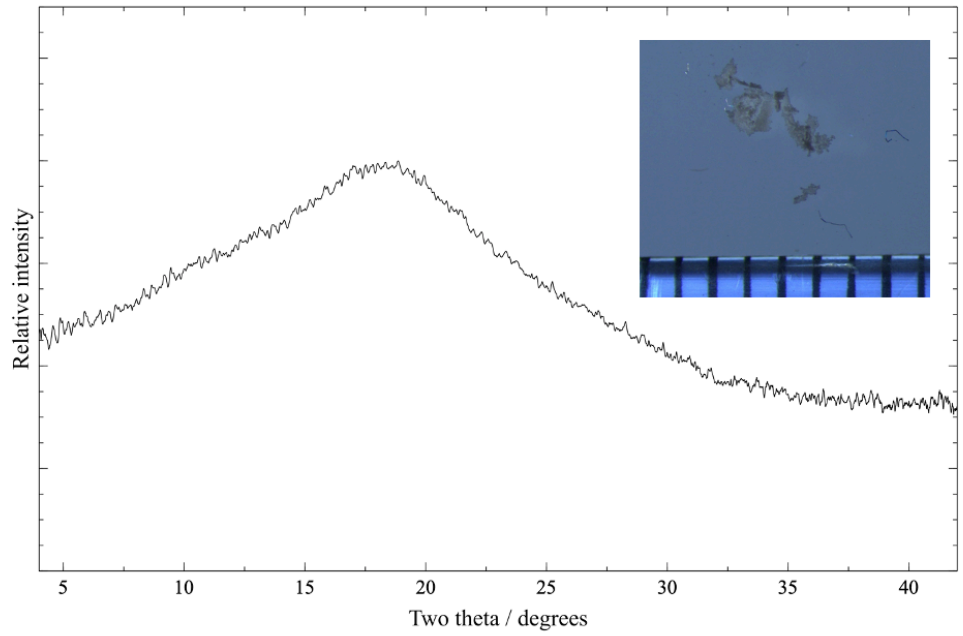


Figure 6.17. Streptomycin (red) sample and corresponding integrated diffraction pattern.

6.5.3 Structure solution with SDPD-SX

Structure solution was achieved for 9 compounds, ranging in complexity from 9 to 24 DoF, using SDPD-SX. All structure solutions were performed with DASH using standard desktop PCs. The typical PXRD data collection time for each sample was approximately 2-4 hours and used 20 to 256 SA runs with 1 x 10^7 SA moves per run to solve the structure. The time taken from initial data collection to final verified crystal structure varied considerably (a few hours to a few days), depending upon how straightforward indexing, Pawley fitting, model construction, SA, and DFT-D optimisation were. The whole process cannot be described as being as routine as a typical SCXRD experiment; however, it represents a valuable addition to the crystallographer's armoury, given that it uses "standard", widely-available laboratory instrumentation and software. It should be noted that the Rigaku Synergy single crystal diffractometer used in this work is by no means the most powerful lab-based diffractometer available, but it has shown that control of beam divergence is a prerequisite for high-quality PXRD data collection. The ability of SDPD-SX to work with tiny quantities of sample has the potential to aid drug design by providing conformational information early on; this kind of insight is important for finding novel and more selective chemical series earlier.

Figure 1.31, taken from Kabova et al (2017), shows that the average success rate of structure solution using transmission capillary data, for compounds in the range 10 to 25 DoF, spans from 100% to around 40%. By way of comparison, SDPD-SX of the structures examined in this chapter (10 – 24 DoF) exhibits a much greater spread of success rates across a similar complexity range, varying from 100% to 4 %. This range is somewhat elongated by the most complex structure having only a 4% success rate; the remaining structures (with a maximum 15 DoF) have an average success rate of 64%. Whilst significantly lower than the transmission capillary equivalent (ca. 80%, based on Figure 1.31) it is nevertheless very good. With less instrumental resolution than a good quality transmission capillary instrument, the use SDPD-SX data leads to significantly different agreement hypersurfaces, so it is not particularly surprising that success rates are impacted in the way summarised above.

6.5.3.1 Previously reported compounds solved using SDPD-SX

A few samples with known crystal structures were included in the recrystallisation protocol as they have the potential to produce new unreported phases, unstudied potentially due to 'unfavourable' morphology. All 3 of the previously reported structures (see Table 6.7) solved using SDPD-SX used a DASH input model derived from the CSD, to mirror the work in chapter 5. They ranged from 9 to 24 DoF and had a real space resolution spanning 2.313 to 2.056 Å. The accuracy of solved structures was validated using the 'crystal packing similarity'

functionality of Mercury; low RMSD values ranging from 0.042 Å to 0.103 Å, highlights the capability of SDPD-SX. These recrystallised samples could not have been analysed by in-house SCXRD, but were ideal for SDPD-SX, despite the small amount of material available.

6.5.3.2 Structure determination and validation of previously unreported compounds using SDPD-SX

Of the 24 previously unreported compounds that were crystalline, 6 were solved using DASH. The structures were validated using DFT-D based energy minimisation, following the approach of van de Streek.[8]

As previously described in Chapter 5, the real-space resolution of PXRD or SDPD-SX data is (in general) significantly lower than that of SCXRD. In this work the real-space resolution ranged from 1.806 to 2.605 Å whereas a typical single crystal structure has a resolution of approximately 0.8 Å. Despite the structures being of lower resolution they are nevertheless useful for many applications. Primarily, simply knowing whether a material is crystalline or amorphous is of value and if crystalline and solvable, the structure provides a wealth of information on conformation, H-bonding, packing etc.

6.5.3.3 Structure validation of previously unreported compounds

For a typical DASH structure solution (using transmission capillary data collection), a profile χ^2 for the solved structure is ca. 3-5 times the value of the Pawley χ^2 suggesting that the global minimum (*i.e.,* the correct crystal structure) has been found. Subsequent Rietveld refinement against the full PXRD data range can be used to refine and confirm the structure. However, it has been found during the current work that this χ^2 multiplier 'rule' is not always a good guide for an SDPD-SX solution and so further validation of novel materials solved by this method is required. Periodic Dispersion-corrected Density Functional Theory (DFT-D) based energy minimisation has been shown in recent years to be a valuable tool within crystallography to validate molecular crystal structures. In essence, optimising the initial solved crystal structure in its fixed unit cell, using an accurate energy minimisation technique (such as DFT-D), should yield an optimised structure that differs very little from the initial structure; the initial structure should be very close to the global energy minimum if it is correct. This difference can be quantified by an RMSD between the atoms in the initial and optimised structures. Assuming that there is no major difference (RMSD typical 0.2 Å or less) then one can be reasonably confident in the solved structure and further increase this confidence using a variable-cell DFT-D optimisation, starting from the endpoint of the fixed-cell optimisation. Of course, the fixed-cell optimised structure should also fit the PXRD data better than the initial structure, assuming

that all factors (such as preferred orientation) are correctly modelled and that the PXRD data are not the subject of some unknown artefacts.

DFT-D optimisation of structures of the complexity presented in this work *can* be performed on high-powered workstations (*e.g.,* 48 cores with 128 Gb or more of RAM) but takes many hours even with highly optimised codes. Ideally, however, access to a compute cluster is required in order to make the use of such optimisations routine; in this work, the EPSRC-funded "Young", a National Tier 2 High Performance Computing Hub in Materials and Molecular Modelling, was used.

6.5.3.4 Modelling disorder: structure solution of C25, rotigotine hydrochloride

Both C25, rotigotine hydrochloride red and blue vials were unsuitable for SCXRD but usable for SDPD-SX. Simulated annealing structure solution was performed on both patterns using a model derived from NMR experimental data (provided by C4X Discovery). Both structures yielded a "structure χ^2 / Pawley χ^2" ratio of 11 and were in excellent agreement with each other, with the exception of the thiophenyl ring which had a 180° ring flip between the minima (as shown in Figure 6.9, see section 6.4.2.4). In a separate investigation by C4X Discovery, a single-crystal structure was obtained that showed the thiophenyl ring to be disordered, with the two disordered conformations agreeing well individually with the 'red' and 'blue' conformations obtained by SDPD-SX. This is a case where small differences in sample presentation have resulted in slightly different PXRD patterns for 'red' and 'blue' that have favoured one diffraction minimum in one case and another in the other case. In the absence of any prior evidence of disorder, it would be difficult to reconcile this outcome without much additional work *e.g.,* periodic DFT-D of a supercell containing both ring orientations.

6.5.3.5 Identifying and modelling preferred orientation: structure solution of C3 (blue)

The use of an area detector for PXRD allows crystallographers to see the signs of preferred orientation. For C3 (blue), indexing and Pawley fitting of the integrated data was simple; however, structure solution proved more difficult. DFT-D energy minimisation of the best DASH structure showed significant movement of atoms, indicating an incorrect solution. Despite multiple efforts made to improve the structure solution, for example, using different starting models, the outcome did not improve. Upon review of the raw data in CrysAlisPro, shown below in Figure 6.18, significant PO can be clearly seen in the first peak, distorting the final integrated 1D diffraction pattern.

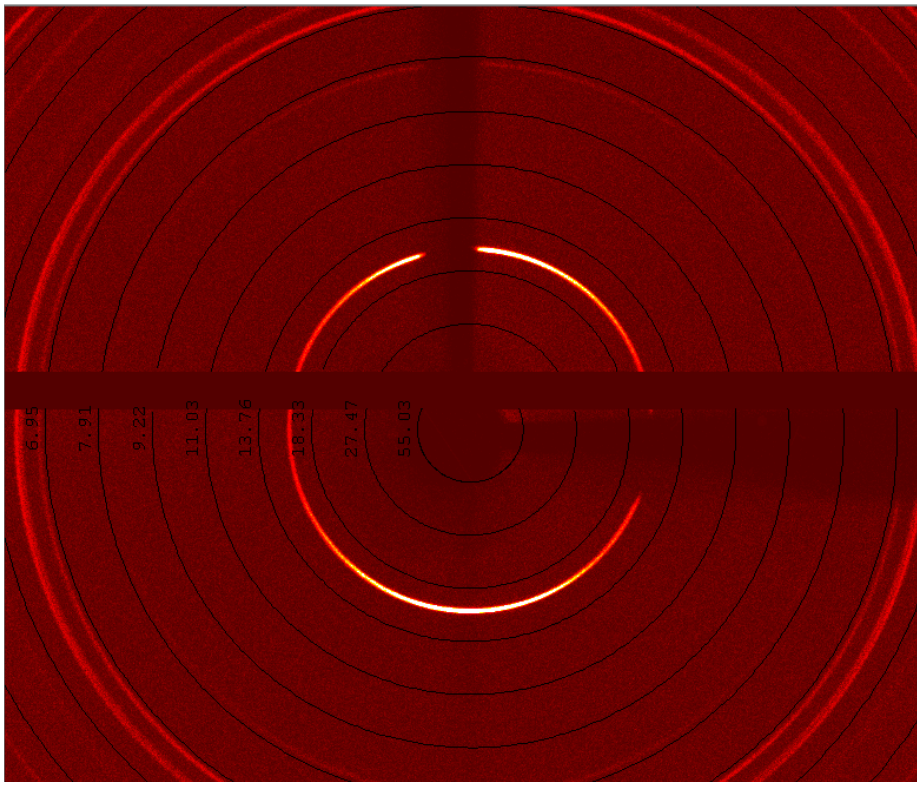


Figure 6.18. CrysAlisPro scan of C3 (blue) collected in standard phi mode. Significant PO is evidenced by the variable intensity around the first diffraction ring.

Armed with this knowledge, DASH structure solution was re-attempted with a 0 0 1 axis PO correction of variable extent included in the SA. The correct structure solution was obtained with a "structure χ^2 / Pawley χ^2" ratio of 4.97 and in comparison to the DFT optimisation end point, the structure gave an RMSD of 0.163. In addition, the structure solution gave a high success rate of 47%, indicating the relative ease of structure solution.

6.5.3.6 DASH input models: structure solution of *S*-adenosyl-L -homocysteine

Ideally a starting model for simulated annealing is created by editing a similar single crystal structure and then performing an isolated molecule DFT optimisation. This approach, which requires the use of non-periodic DFT codes was not utilised, due to the requirement for yet more specialist computational knowledge. Instead, this work focused on what is possible with access to open-source software and the CSD. Using a model generated from the CSD conformer generation tool, a low-quality *S*-adenosyl-L-homocysteine, SAH, (red) structure solution was obtained. In order to achieve a better structure solution, a C-O bond in the 5-member ring was broken to allow ring flexibility. The DASH structure solution using this model had a χ^2 ratio of 9.89, and a fixed-cell DFT-D energy minimisation yielded a high RMSD of 0.613 Å. Despite this large RMSD value, the overall shape of the structure remained relatively consistent, as shown below in Figure 6.19.

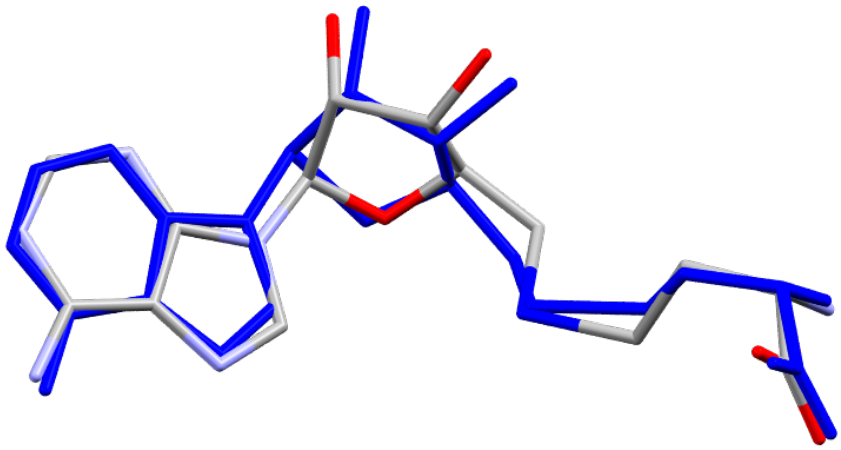


Figure 6.19. Overlay of C26, *S*-adenosyl-L-homocysteine, SAH, (red) best DASH output overlayed with the DFT-D minimisation output (in blue) with hydrogens removed, RMSD of 0.613 Å.

It was noted that the first DFT-D minimisation step was associated with a large drop in energy, attributable to the movement of a hydrogen atom, changing the molecule to the zwitterion form. Therefore, the DFT-D output structure was used as an improved starting model for a second SA structure solution attempt. This end result of this DASH run was much improved, with DFT-D minimisation resulting in an RMSD of just 0.2 Å, shown in section 6.4.2.6.

6.5.4 Structure determination using electron diffraction

Electron diffraction was performed by the Rigaku Corporation using their (at that point unreleased) XtalLab Synergy-ED instrument, but is included in this discussion, since it demonstrates what is possible for samples beyond the limitations of SDPD-SX.

SDPD-SX analysis is limited by the same constraints of all powder structure analysis, in particular the size of the unit cell and the number of structural DoF. Electron diffraction offers a solution to this limitation as it collects 3D diffraction data and uses direct methods of structure determination to solve the structure. Because ED requires very thin samples, a low sample amount available is not a limitation. ED as a technique has it's own challenges, in particular the presence of dynamic scattering which is discussed in Chapter 1. Three materials, analysed by SDPD-SX and found to be unindexable with any degree of confidence, were sent to Rigaku. These samples were selected because the SDPD-SX data showed that they were all highly crystalline and likely had large unit cells, possible impurities or were phase mixtures. From this ED data, crystal structures of C20 (blue) and C13 (blue) were obtained, confirming large unit cell of C20 (blue), shown in section 6.4.3. The third material, C18 (blue), was not able to be analysed as the limited amount of sample could not be retrieved from the sample vial.[9]

This work highlights a relationship between the two data collection methods that can be utilised beyond this thesis: materials first screened using SDPD-SX that are crystalline but are unable to be solved can be forwarded for ED. SDPD-SX is a much more cost-effective method to screen samples for ED, as single crystal diffractometers are widely accessible (and data collection and analysis is relatively simple. Screening can be used to confirm crystallinity and also detect likely impurities and/or multiple phases *e.g.,* as with C20 (blue).

6.5.5 Unsolved crystal structures

6.5.5.1 Indexed structures unable to be solved

6.5.5.1.1 C13

C13 (red/blue) datasets were able to be indexed by DASH but gave a low FOMs and were not able to be Pawley fitted with any degree of confidence. However, it is clear from Figure 6.20 that the PXRD data are good, as adjudged by the agreement between them and the calculated data based on the ED-derived structure.

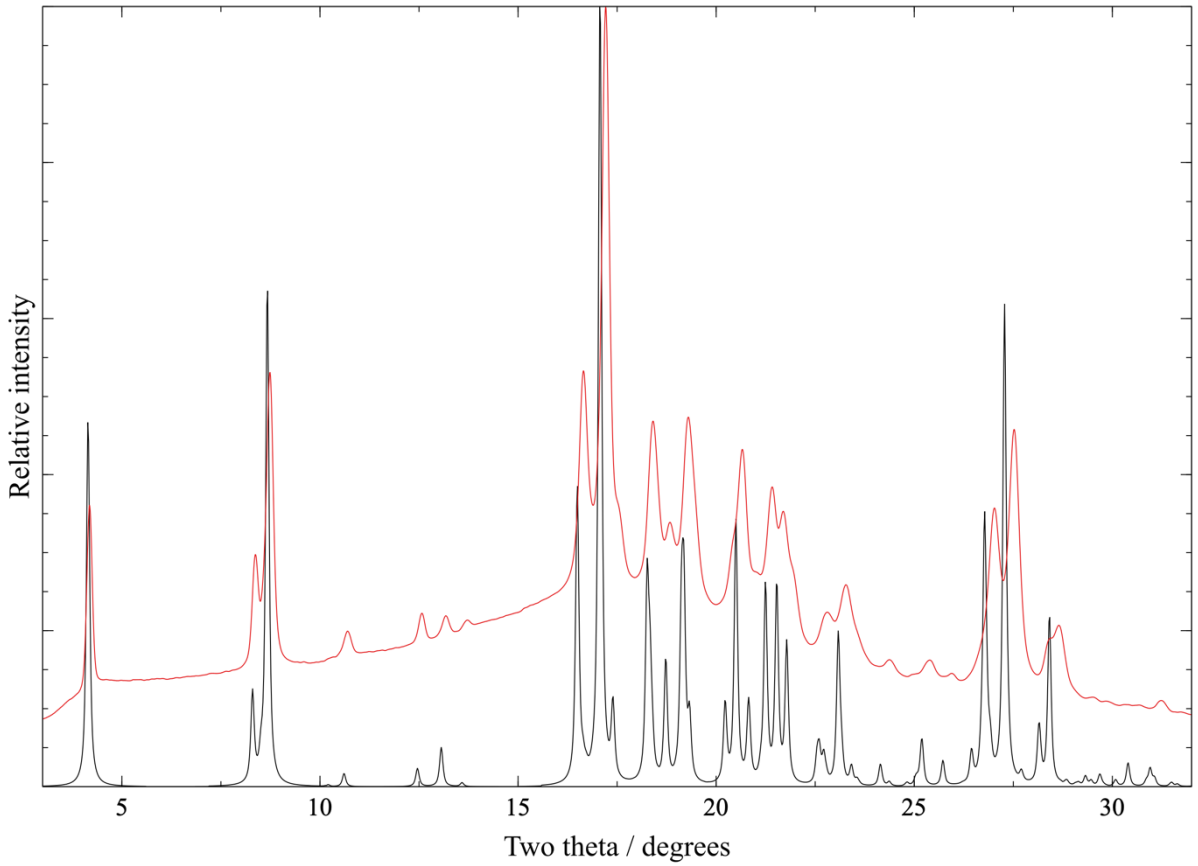


Figure 6.20. C13 (blue) generated diffraction pattern from EM structure (in black) *vs* SDPD-SX raw data (in red).

6.5.5.1.2 C14

C14 (blue) was indexed with confidence and was able to be Pawley fitted to 36.7° 2θ (162 reflections) but was unable to be solved. A single crystal obtained subsequently by C4X Discovery confirmed the indexing (Table 6.11) and showed the effect of differential thermal expansion upon the cell.

Table 6.11. Cell dimensions for C14 at varying collection temperatures.

Collection method	**T / K**	**Sp. Grp.**	**a / Å**	**b / Å**	**c / Å**	**β / °**	**V_{cell} / Å^3**
SX	100	$P2_1$	4.8880	18.6070	22.8810	92.357	2079.3
SDPD-SX	RT	$P2_1$	4.9575	18.7215	23.3921	91.787	2170.0

The single crystal structure is shown in Figure 6.21; with Z' = 2, and one molecule having substantial disorder, it is unsurprising that structure solution by SDPD-SX failed.

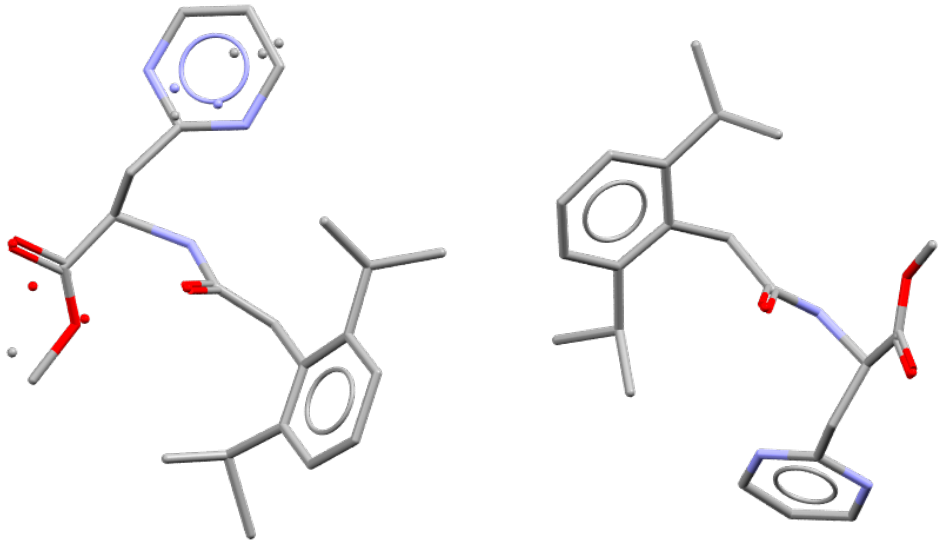


Figure 6.21. C14 red asymmetric unit with 0.5:0.5 occupancy of disordered molecule, determined by SCXRD at 100 K.

6.5.5.2 Patterns unable to be indexed with confidence

6.5.5.2.1 Amorphous and poorly crystalline materials

10 of the compounds that were examined by SDPD-SX were amorphous or too poorly crystalline to be analysed further. The boundary between poorly crystalline and amorphous is not clearly defined, as all of the samples appeared to contain some amorphous component. A large amorphous component in a crystalline sample can result in the powder diffraction

pattern being difficult to analyse. An example of this is atorvastatin (Figure 6.22) whose PXRD pattern could not be indexed by either DASH or EXPO2014. This sample does show some reasonable diffraction but ultimately was unusable for SDPD-SX. It is possible that with additional data collections and careful background subtraction, indexing might be possible, but there was not sufficient time to explore this option.

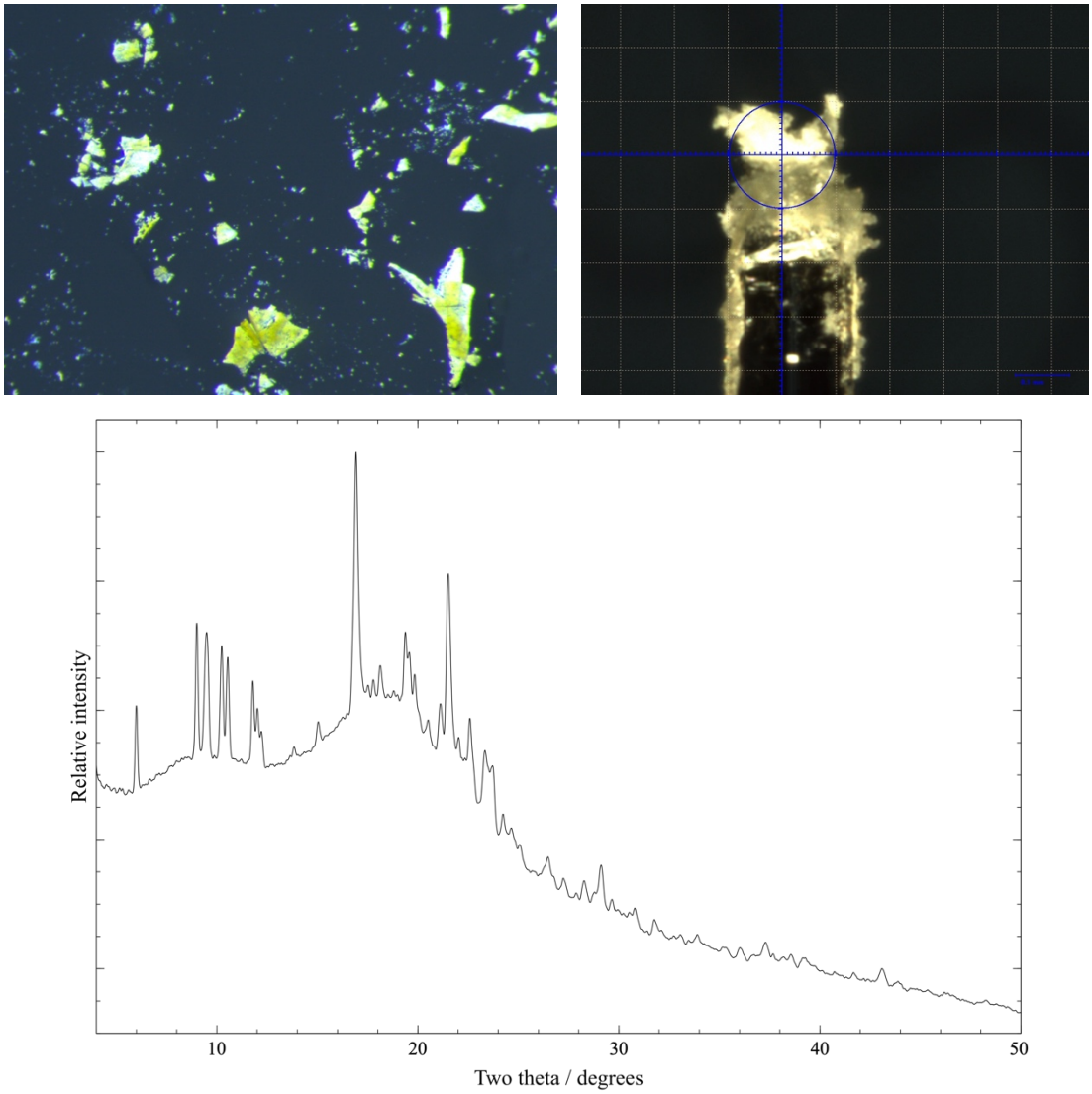


Figure 6.22. C21, atorvastatin hemicalcium salt (red) sample (above left), on the sample holder with minimal oil (above right) and resulting powder diffraction pattern (below).

6.5.5.2.2 Phase mixtures and/or anisotropic thermal expansion effect

It is often difficult to ascertain if a powder diffraction pattern is composed of contributions from more than one crystalline phase. The SDPD-SX powder diffraction pattern for C20 (blue) was not able to be indexed and was sent for ED analysis. The calculated PXRD pattern from the ED structure matches the observed PXRD well, apart from in a few locations (see Figure 6.23)

where additional peaks appear in the observed data. These are most likely attributable to the effect of anisotropic thermal expansion; the presence of a very small number of relatively strong peaks is not strongly indicative of another phase of C20 being present, given the size of the molecule and the likelihood of a low-symmetry phase.

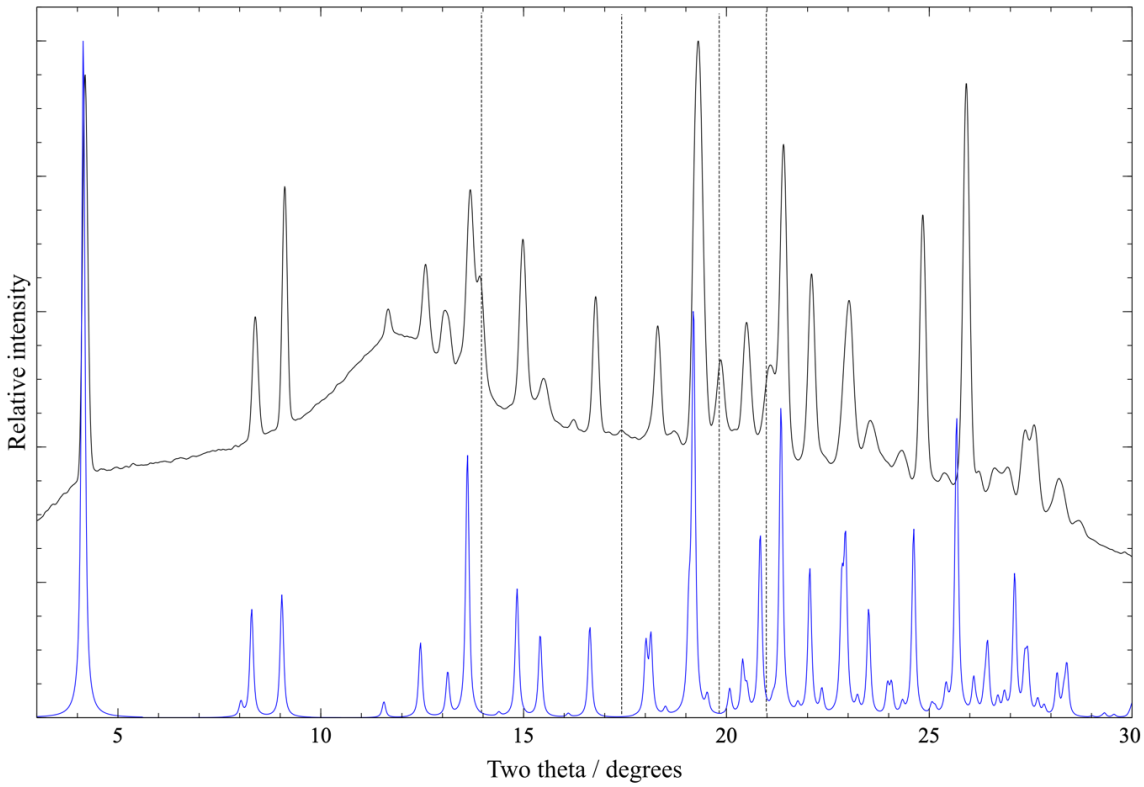


Figure 6.23. Comparison of C20' (blue) integrated diffraction pattern from collected by SDPD-SX (in black) and a generated powder patten of the single crystal structure found by ED (in blue). Anisotropic unit cell changes with temperature highlighted by dotted lines.

6.5.5.2.3 Inorganic salt impurities

Simple inorganic salts (*i.e.,* NaBr, $MgSO_4$) are commonly used throughout the synthesis of organic compounds, for example in the drying or separation steps. To remove impurities, the organic material is recrystalised, and once crystals have formed, the sample is filtered to remove material still in solution. Any impurities within materials in this study were not able to be filtered out, due to the significant loss of sample that would occur. For SCXRD, removing impurities is not always necessary; a single crystal can be removed and analysed without impurities affecting data collection. However, because powder diffraction records the average of the crystallites in the material, inorganic impurities (if present in sufficient quantity) are often easily identified by the presence of a few diffraction peaks (generally at higher values of 2θ) that are noticeably sharper than other nearby peaks. Inorganic salt impurities were observed

with multiple data collections in this study. The salt impurity within the C18 pattern was identified as NaCl, as shown in Figure 6.24.

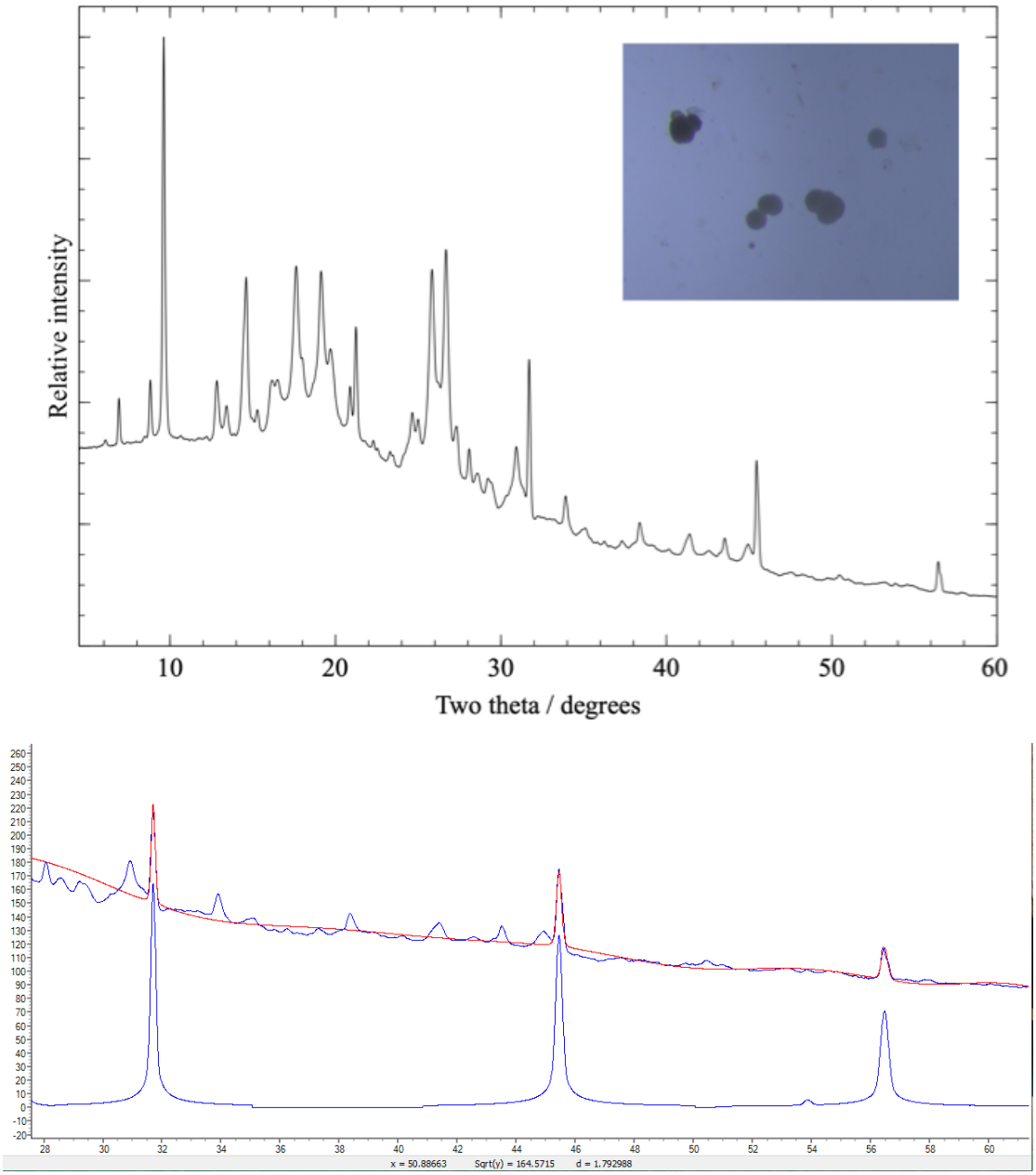


Figure 6.24. The integrated diffraction pattern of C18 blue, above from SDPD-SX showing inorganic salt impurities. NaCl modelled diffraction peaks modelled with the C18 (blue) diffraction pattern, performed in Topas, below.[10]

For the C19 (blue) diffraction pattern the salt impurity was found to be isostructural to KCl, shown below in Figure 6.25. In order to be able to perform structural analysis, either the peaks have to be removed from the diffraction data or the impurity removed from the original sample and data recollected. This analysis was beyond the scope of this chapter.

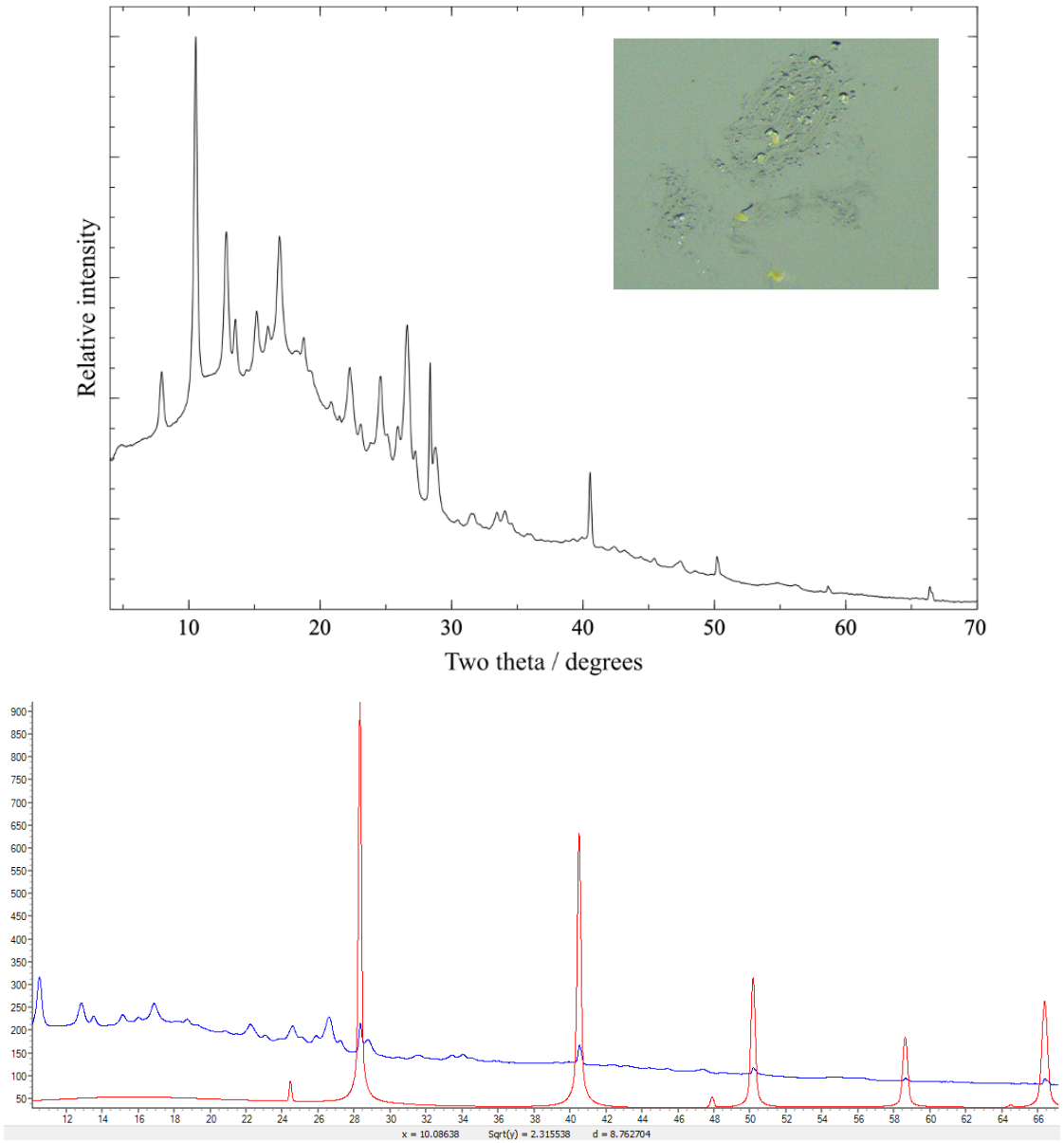


Figure 6.25. Integrated diffraction pattern of C19 (blue), above from SDPD-SX showing inorganic salt impurities. KCl modelled diffraction peaks overlayed with the C19 diffraction data, performed in Topas (below).

6.5.5.2.4 Possibility of hydrates / solvates

For unknown crystal structures it is difficult to know without prior analysis if the structure contains solvent or water molecules. For larger sample volumes, thermal analysis (hot stage microscopy, thermogravimetric analysis and differential scanning calorimetry) can be used to identify the presence of solvent molecules and the ratio of components. The sample amounts used in this work were limited (<0.1 mg), making this additional information unobtainable. Instead, the presence of solvent and water molecules can be deduced from a significant difference between V_{cell} and $Z \times V_{mol}$. Another tool for indicating the presence of possible

solvent inclusion is the presence of voids (calculated in Mercury) in the unit cell from a non-solvated structure solution attempt, although failure to model solvent in the structure solution often leads to incorrect and misleading sub-structures. Unfortunately, for structures that have solvent molecules not in a 1:1 or other similar ratio, then the structure solution becomes more complex. Multiple combinations can be attempted that make sense for the volume available, but finding the correct answer can be challenging and time consuming.

6.5.5.2.5 Sample degradation

The C6 (blue) sample was initially examined by a short (10 minute) SDPD-SX data collection, appearing to be crystalline with limited amorphous content. A longer data collection was, however, disappointing in comparison, with a much-reduced signal-to-noise ratio (Figure 6.26). It is likely that the sample was either losing solvent or suffering radiation damage, or possibly partially melting (unfortunately, the MPt of the sample is not available). Regardless, this shows the importance of looking for any time-dependent changes in the sample being studied, as opposed to only looking at the endpoint of a data collection.

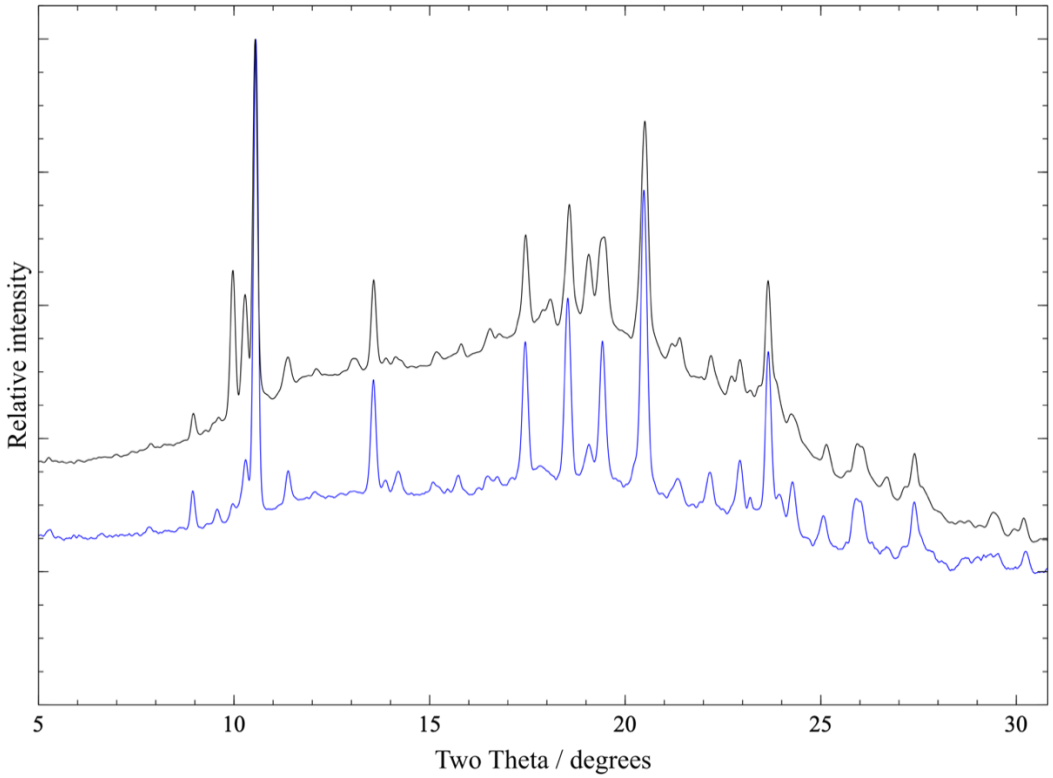


Figure 6.26. Integrated diffraction patterns for C6 (blue) test data set (in blue), overlayed with full data collection (black). No background removal, scaled to same max height.

6.5.5.2.6 Large and/or low-symmetry unit cells

Indexing powder patterns from low-symmetry molecular crystals becomes more challenging with increasing unit cell size; the chance of accidental peak overlap is increased, and it can be hard to obtain accurate positions for peaks close to the minimum value of 2θ achievable on the instrument. Figure 6.27 shows diffraction data for C2 (blue); despite being of good quality, it has resisted indexing, with low angle peaks suggesting a relatively large unit cell.

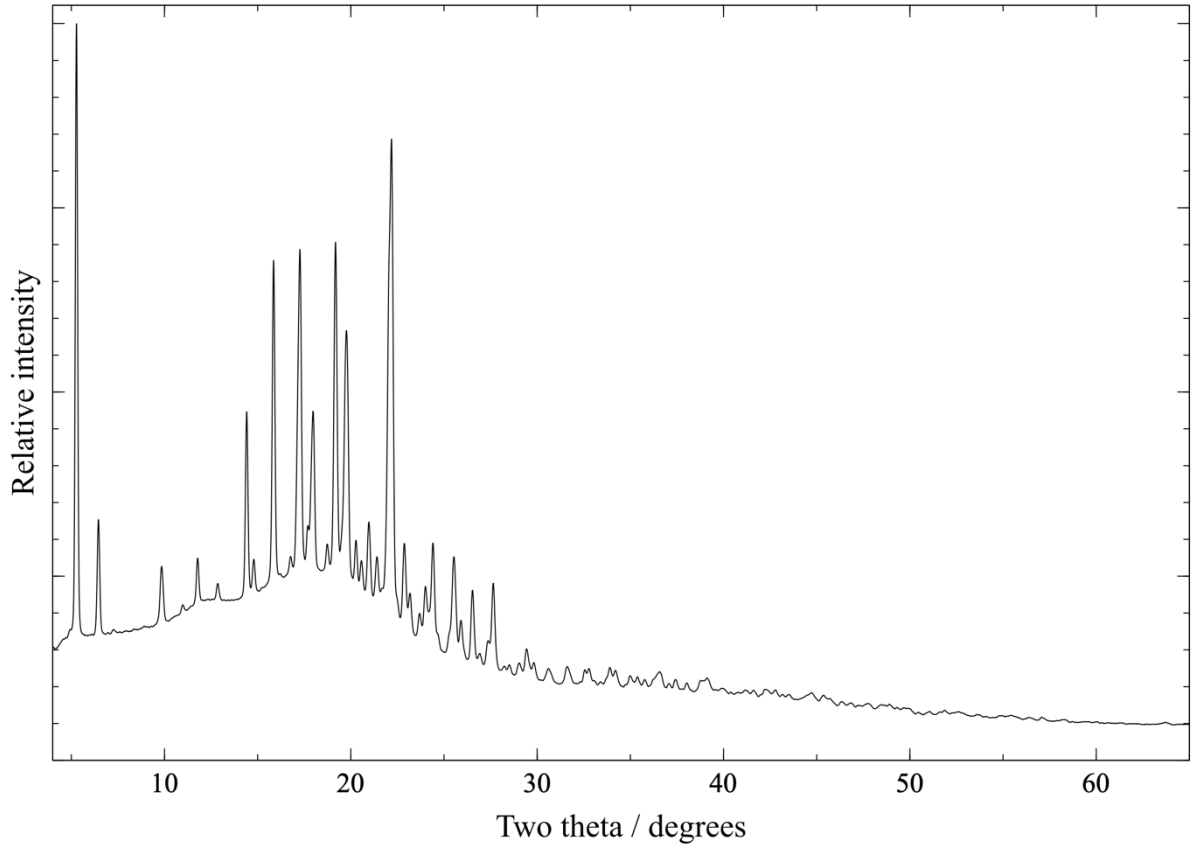


Figure 6.27. Integrated diffraction pattern of C2 (blue)

C9 (red) has a deceptively simple diffraction pattern (Figure 6.28) similar to that of C10 (red/blue) which had been solved easily. Indexing suggested three similar unit cells as shown in Table 6.12.

Table 6.12. Potential unit cells identified for C9 (red).

Data	**Sp. Grp**	**a / Å**	**b / Å**	**c / Å**	**β / °**	**V_{cell} / Å^3**
SDPD-SX	$P2_12_12_1$	5.348	5.141	49.073	90	1349.2
SDPD-SX	$P2_1$/n	5.365	49.058	4.822	93.914	1266.2
SDPD-SX	$P2_12_12_1$	10.719	5.138	49.145	90	2706.6
SCXRD	$P2_12_12_1$	5.0193	5.4498	47.402	90	1296.6

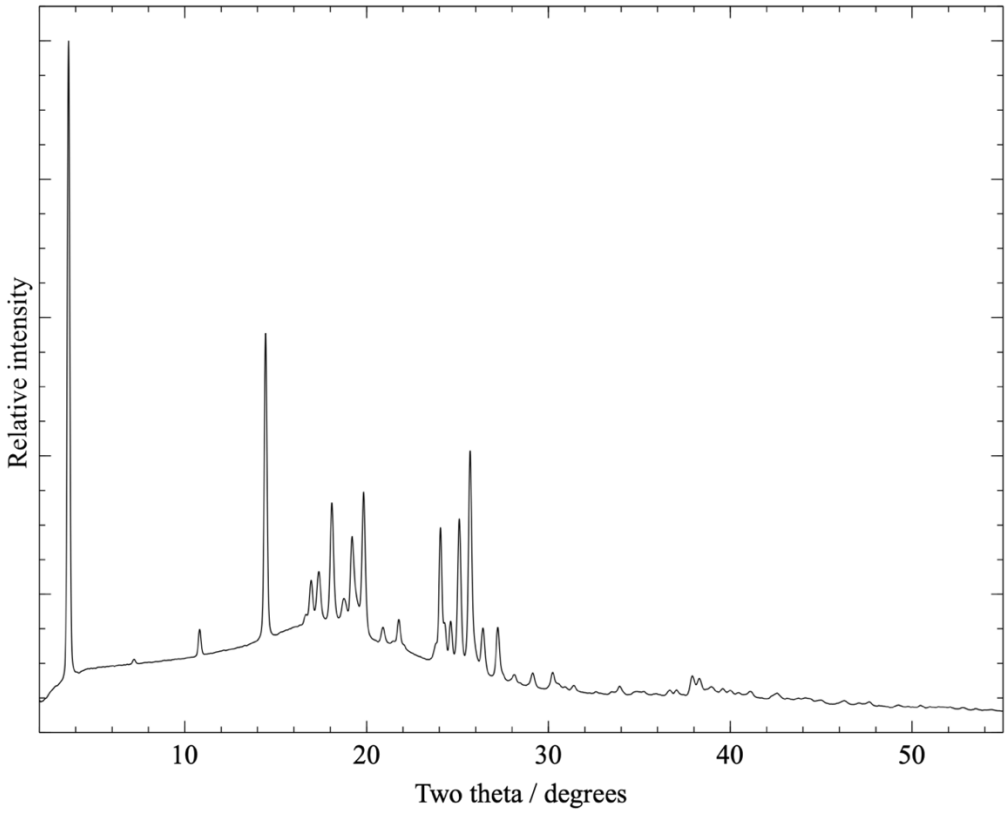


Figure 6.28. Integrated diffraction pattern of C9 (red).

Each of the putative unit cells gave a reasonable Pawley fit and so each was used as the basis of DASH runs using only the active molecule as input; the unit cell volumes, when compared to V_{mol}, suggested there was no solvent present. All DASH solutions obtained were subjected to DFT-D energy minimisation and the structures either moved significantly or the calculations failed to converge, both indicative of incorrect structures. Examination of the raw data showed the presence of significant preferred orientation (Figure 6.29), but inclusion of this information in the DASH runs did not yield any viable structure solutions.

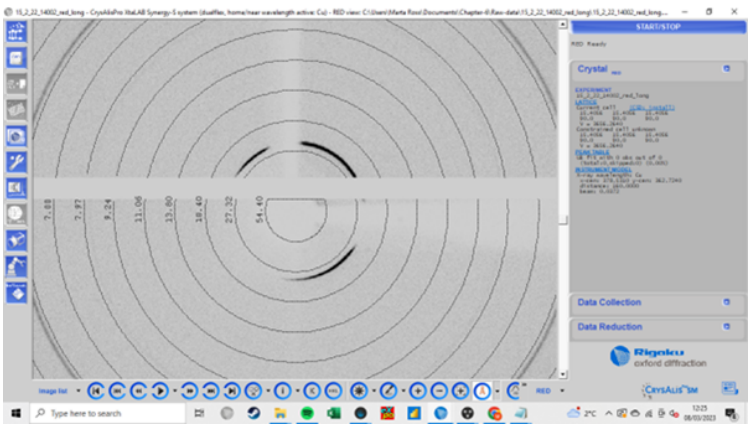


Figure 6.29. C9 (red) data collected in standard phi mode, showing significant preferred orientation in the first diffraction ring.

The unit cell of the orthorhombic crystal structure later obtained by C4X from a single crystal from C9 (blue) agreed well with one of the SDPD-SX cells, allowing for thermal expansion. However, a unit cell and scale-only Rietveld refinement of the single-crystal structure shows a clear problem with the data obtained by SDPD-SX; there is a large unaccounted-for feature at ca. 19° which we assume is an impurity.

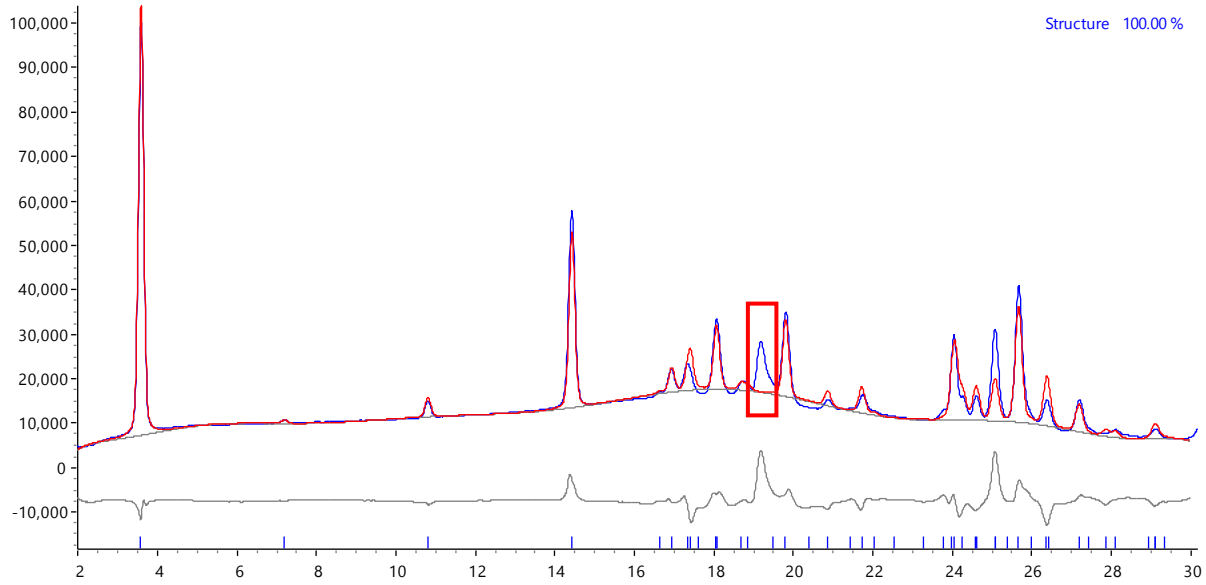


Figure 6.30. A unit cell and scale-only Rietveld refinement of the single-crystal structure of C9 (blue) against the SDPD-SX data with the anomalous peak highlighted in red.

The case of C9 (red) indicates several of the issues often encountered when dealing with PXRD data. Any uncertainty in the likely unit cell results in multiple structure solution pathways, that are not only time consuming but also may involve variations in the input structures for DASH, and as already discussed, structure validation is computationally intensive. Finally, even with a correct cell and space group, and a knowledge of the correct unit cell contents, disorder in the structure cannot be directly visualised and if suspected, has to be investigated by trial-and-error.

6.5.6 Considerations for data reduction within CrysAlisPro

When using a modern single crystal diffractometer, an area detector is used to capture several diffraction spots / rings at once. The beam stop is used to prevent the detector from radiation damage that would occur from direct exposure. Within CrysAlisPro the size of the beam stop is estimated (and larger than the physical beamstop shadow) and light within this region is hidden when the powder pattern is extracted. By default, the beam stop is set to 2.1 mm, which for the powder data extraction corresponds to approximately 4° 2θ.

When low angle peaks are observed (either as a partial low angle peak or low angle rings visible in CrysAlisPro) the size of the beam stop can be reduced to extend the useful 2θ range. This can be performed by selecting the 'Options RED' button from the CrysAlisPro command shell and reducing the size of the beam diameter, as shown in Figure 6.31. This adjustment should only be performed in the offline mode, so as not to affect future data collections.

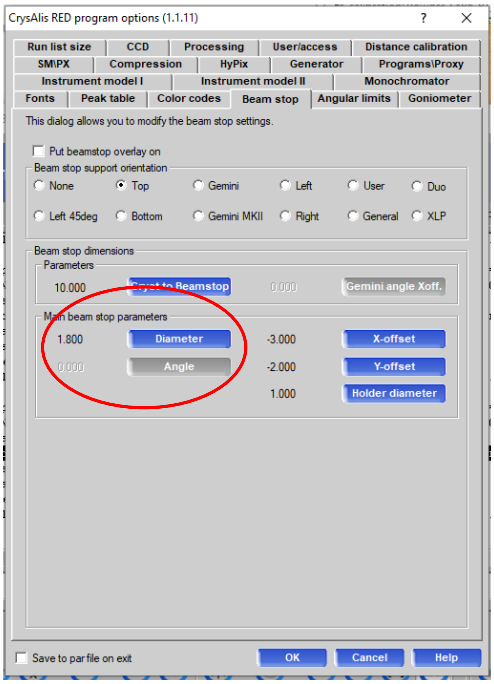


Figure 6.31. CrysAlisPro options for reducing beam stop size to increase 2θ range highlighted in red.

Figure 6.32 shows that for C14 (blue) the first diffraction peak begins below CrysAlisPro's default setting; the full first peak becomes visible once the size of the beam stop is reduced.

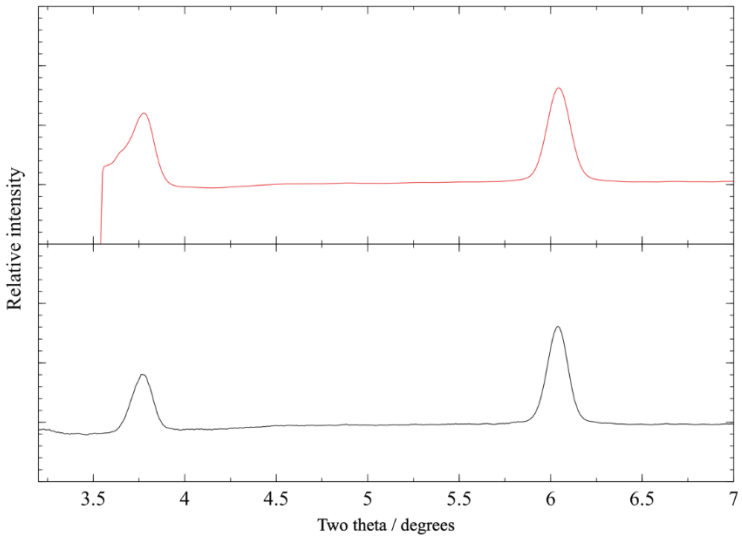


Figure 6.32. Diffraction pattern of C14 (blue) using standard CrysAlisPro beamstop setting (upper) and adjusted setting to extend useful data range.

6.6 Summary and conclusions

This work in this chapter focused on the (re)crystallisation of 81 small organic compounds. Of these, 36 were examined by SDPD-SX, leading to a full data collection of 27 compounds. A total of 9 structures were solved by SDPD-SX, of which 6 were previously unreported. Overall, 11% of the total compounds (corresponding to 32% of the samples examined by full SDPD-SX data collection) were solved by SDPD-SX. It is important to that the structure determinations were performed using standard PC computing power and open-source software.

Of the 81 compounds, 50 compounds (62%) were found to be crystalline from at least 1 vial, of which 31 compounds were able to be used for SCXRD. Of the initial 81 compounds, 9 were biomolecules and derivates (18 vials), of which only 1 vial was crystalline and unsuitable for SCXRD. This compound, C26, S-adenosyl-L-homocysteine (red), was able to be solved with the use of SDPD-SX and has not been reported elsewhere.

This chapter highlights the potential of SDPD-SX to fill an important gap in powder diffraction data collection where minimal (< 0.5 mg) sample is available. At the time of writing, SDPD-SX has already been used for structure determination for current drug discovery projects using the information gained from this thesis.

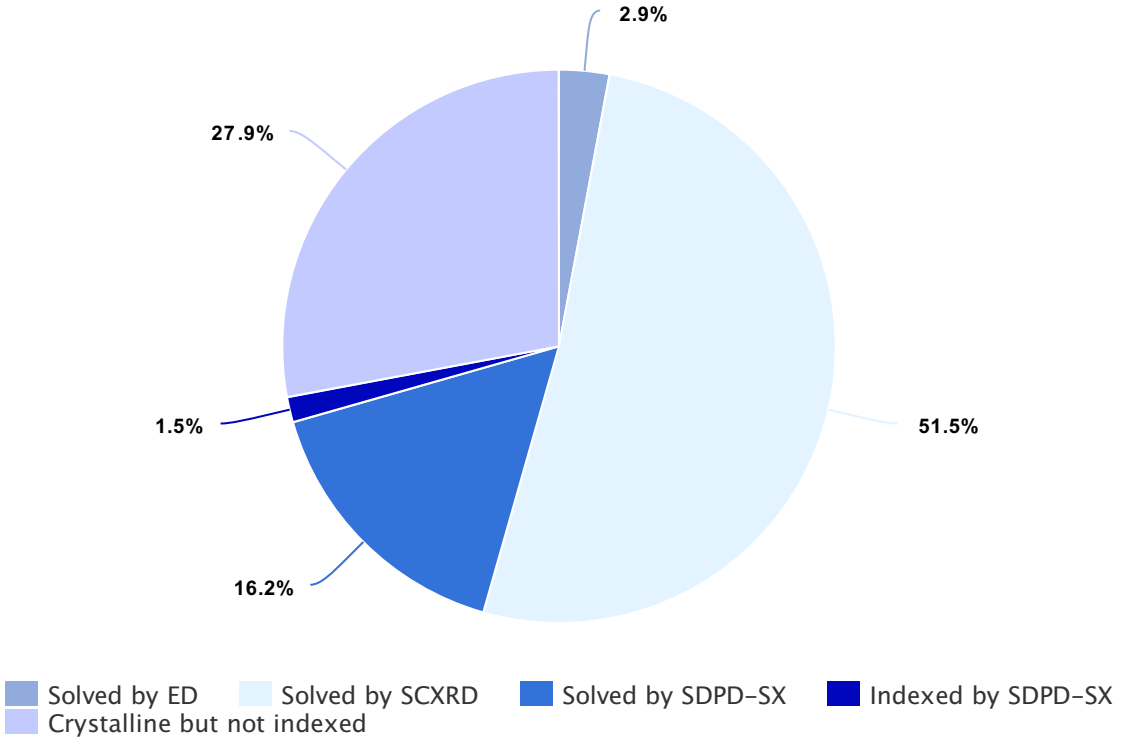


Figure 6.33 Summary of 68 crystalline vials (from 47 compounds) used for the recrystallisation protocol.

Structure solution by SDPD-SX was obtained for 32% of the crystalline unreported samples, with a further 4% indexed but not solved. A summary of structure solution found within this study is shown below in Figure 6.34.

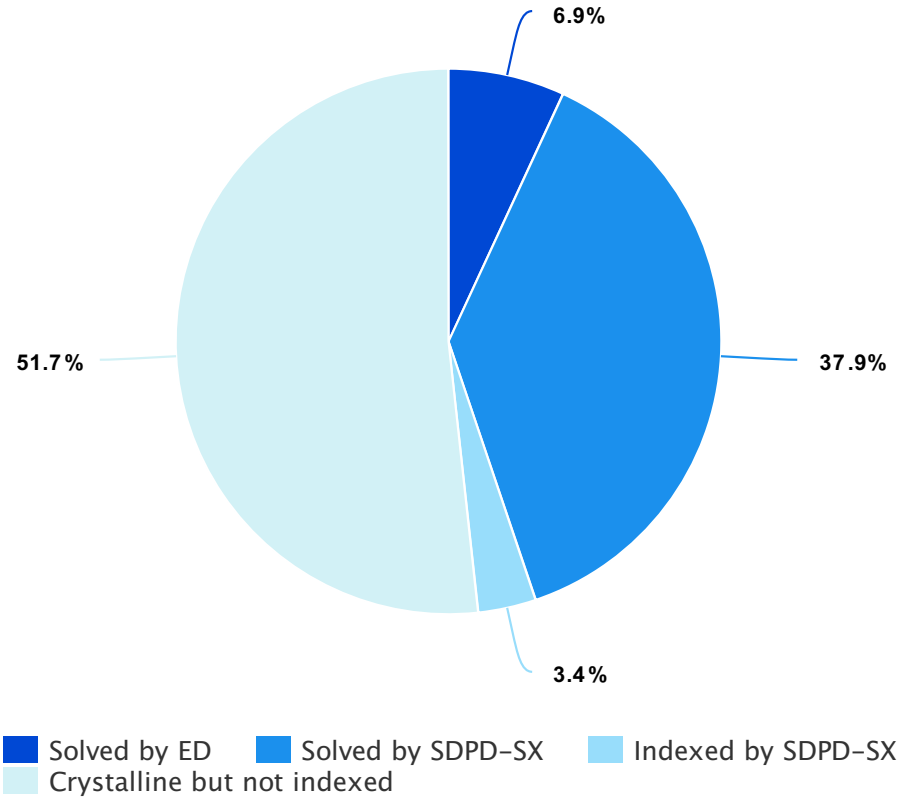


Figure 6.34. Summary of the 27 crystalline compounds that were examined with a full SDPD-SX data collection.

Results for the samples that were examined by SDPD-SX and the reasons preventing further structural analysis are given below in Figure 6.35. It is important to highlight the presence of salt impurities that contribute to 20% of the samples that were unable to be analysed by SDPD-SX. Samples identified as being unusable for SDPD-SX due to salt impurities or phase mixture are ideal candidates for ED, as these factors do not affect the structural analysis of micro single crystals. ED is currently a specialist technique, nevertheless it is an expanding area of research.

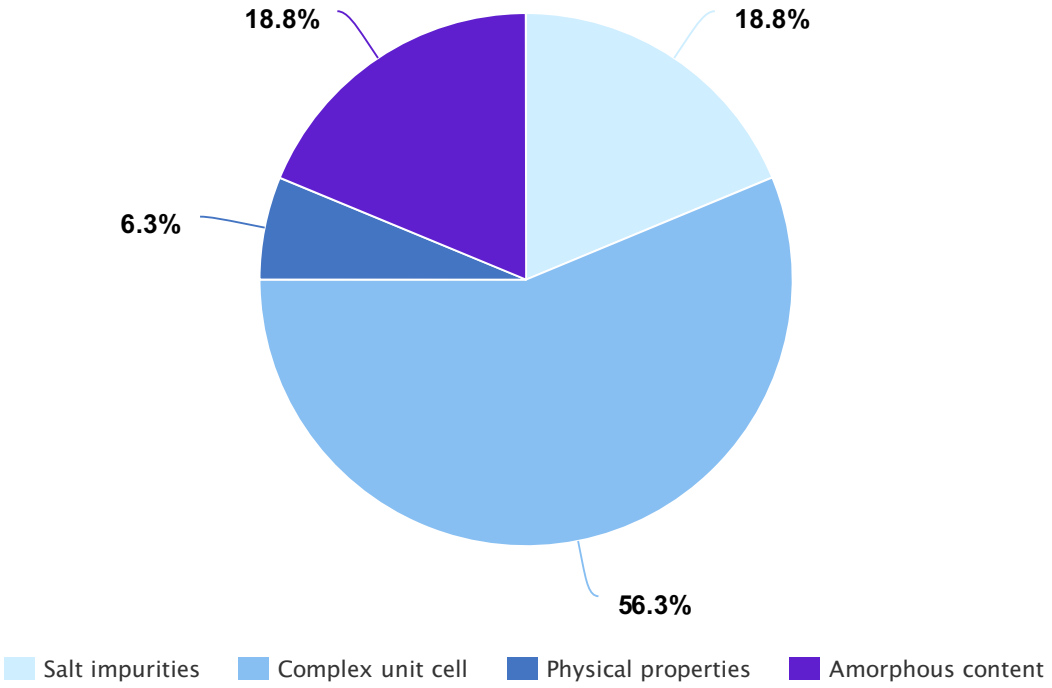


Figure 6.35. Summary of the 15 compounds that were unable to be indexed using SDPD-SX during this chapter. It is important to note that the 'Complex unit cell' may also be due to a phase mixture within the diffraction pattern.

6.7 References

1 A. Višnjevac and B. Kojić-Prodić, *Kem. u Ind. Chem. Chem. Eng.*

2 W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell and J. C. Cole, *J. Appl. Crystallogr.*, 2006, **39**, 910–915.

3 E. A. Kabova, J. C. Cole, O. Korb, M. López-Ibáñez, A. C. Williams and K. Shankland, *J. Appl. Crystallogr.*, 2017, **50**, 1411–1420.

4 E. A. Kabova, C. D. Blundell and K. Shankland, *J. Pharm. Sci.*, 2018, **107**, 2042–2047.

5 A. Volford, *Chem Axon MarvinSketch*.

6 P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru and S. Baroni, *J. Chem. Phys.*, 2020, **152**, 154105.

7 C. Laurence, J. Legros, A. Chantzis, A. Planchat and D. Jacquemin, *J. Phys. Chem. B*, 2015, **119**, 3174–3184.

8 J. van de Streek and M. A. Neumann, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2014, **70**, 1020–1032.

9 U. Kolb, K. Shankland, L. Meshi, A. Avilov and W. David, *Uniting Electron Crystallography and Powder Diffraction*, Springer Netherlands, Dordrecht, 2012.

10 A. A. Coelho, *J. Appl. Crystallogr.*, 2018, **51**, 210–218.

6.8 Appendix

Below are additional structures, with accompanying sample pictures, of crystalline materials that were unable to be analysed by SDPD-SX because of salt impurities, phase mixtures or poor data quality. Results are presented in numerical order.

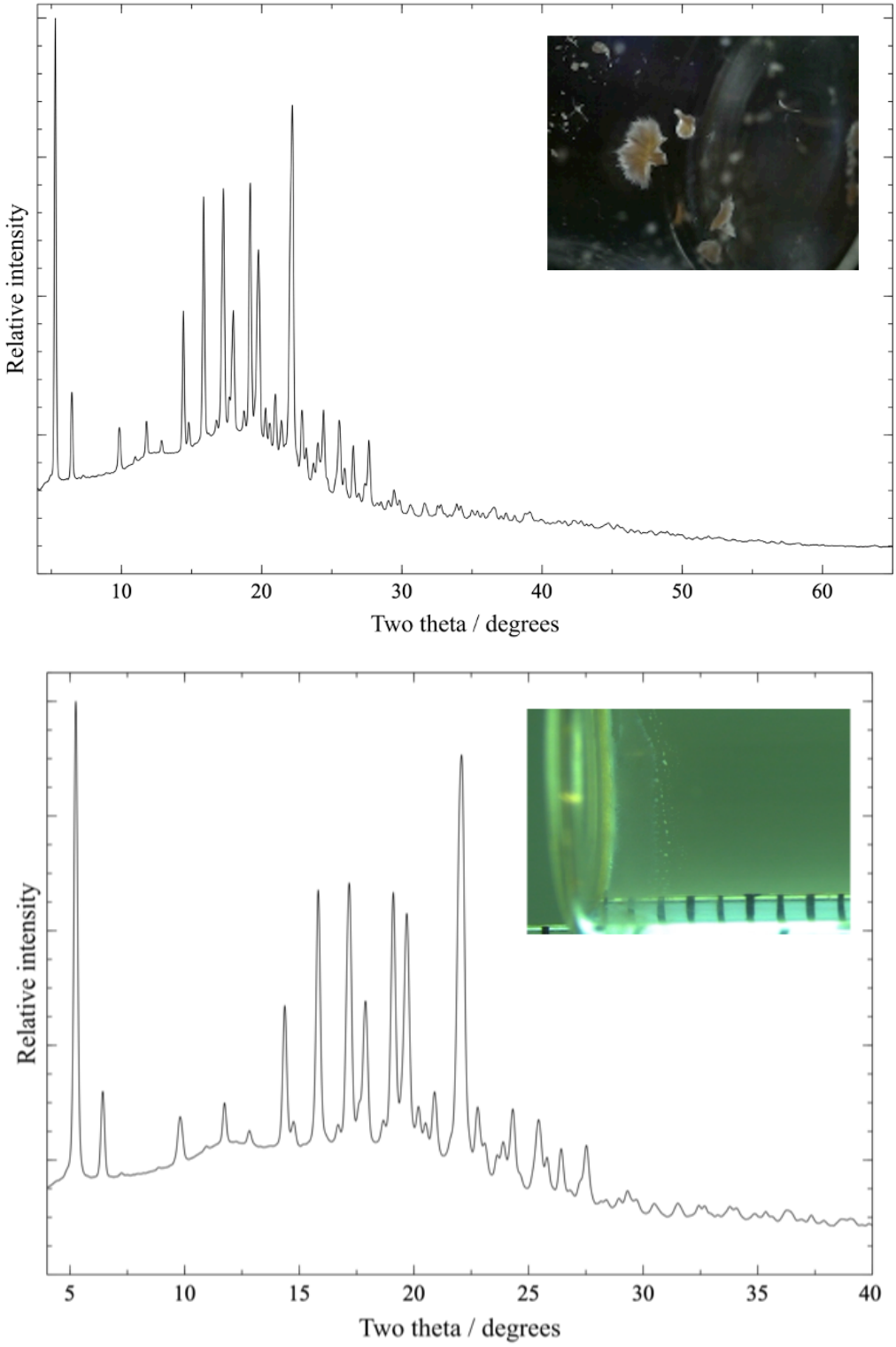


Figure 6.36. PXRD data of C2 (red) with inset picture of C2 (red) in vial, above. PXRD data of C2 (blue) with inset picture of C2 (blue) sample with 1'mm' reference, below.

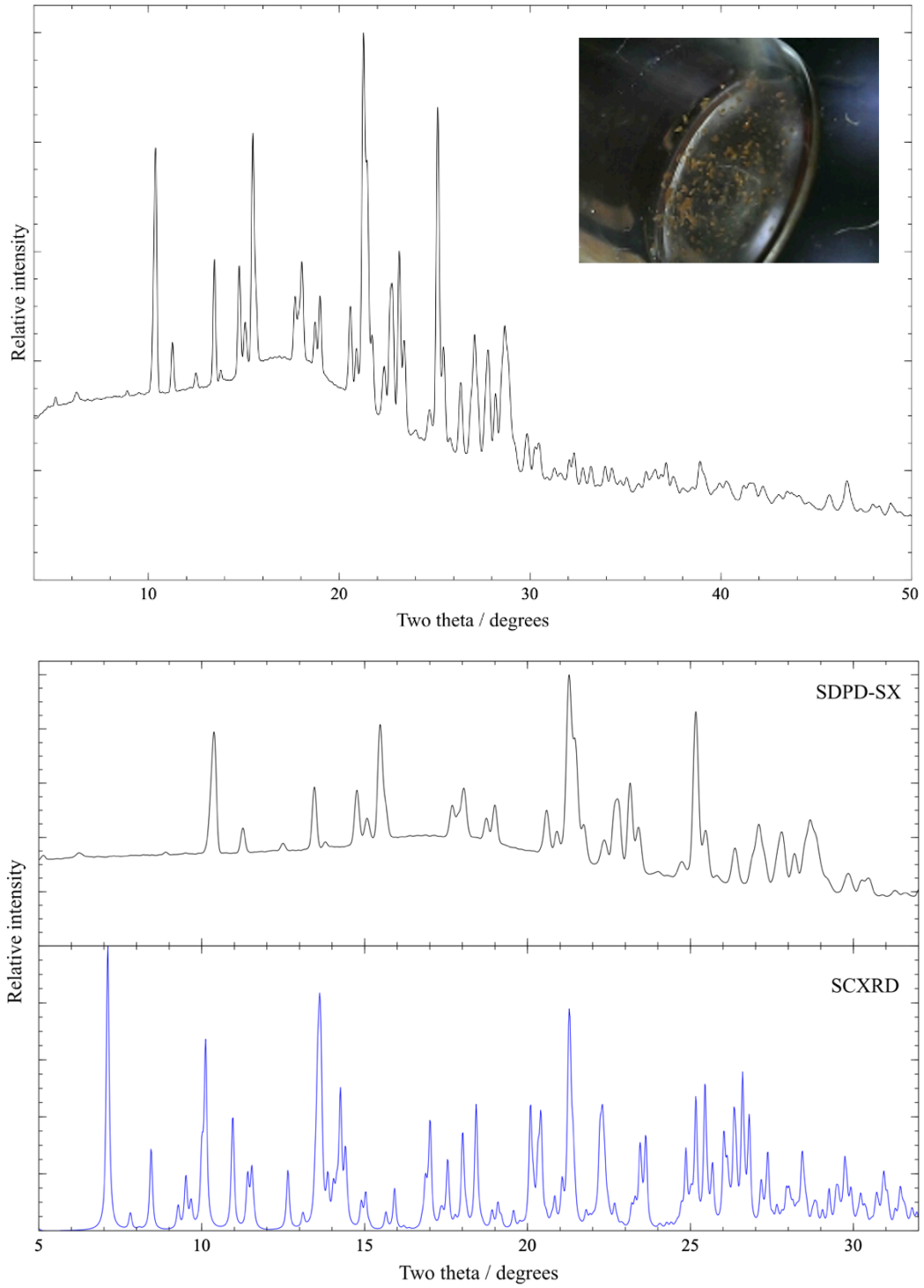


Figure 6.37. PXRD data of C5 (red), inset picture of C5 (red) sample in vial, above. Comparison to C5 (red) PXRD data compared to C5 (blue) generated PXRD pattern, below, highlighting phase difference.

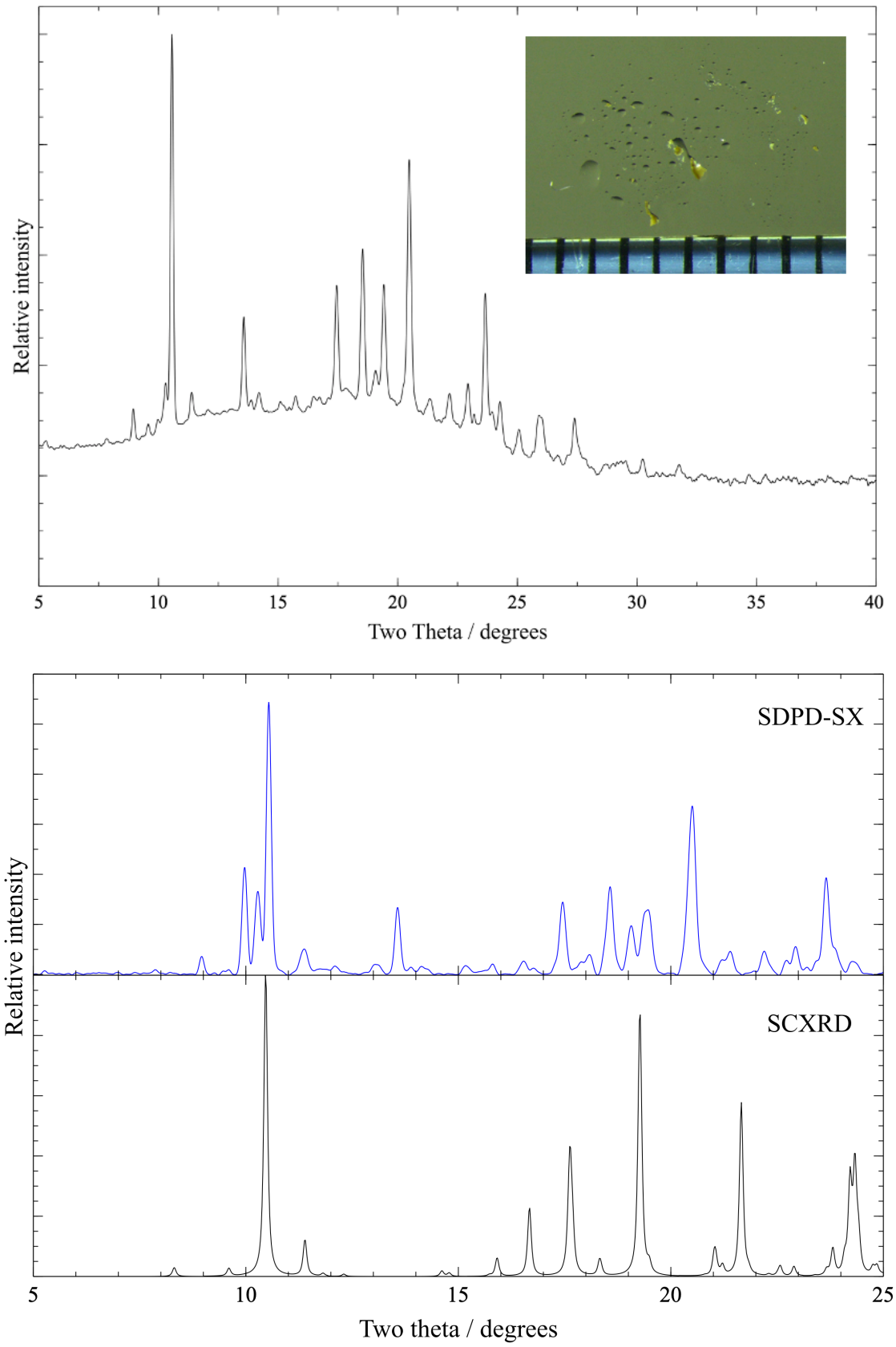


Figure 6.38. PXRD data of C6 (blue), inset picture of C6 (blue) sample with 1 mm reference above. Comparison to C6 (blue) PXRD data compared to C6 1:1 hydrate (red) generated PXRD pattern, below, highlighting potential mixture.

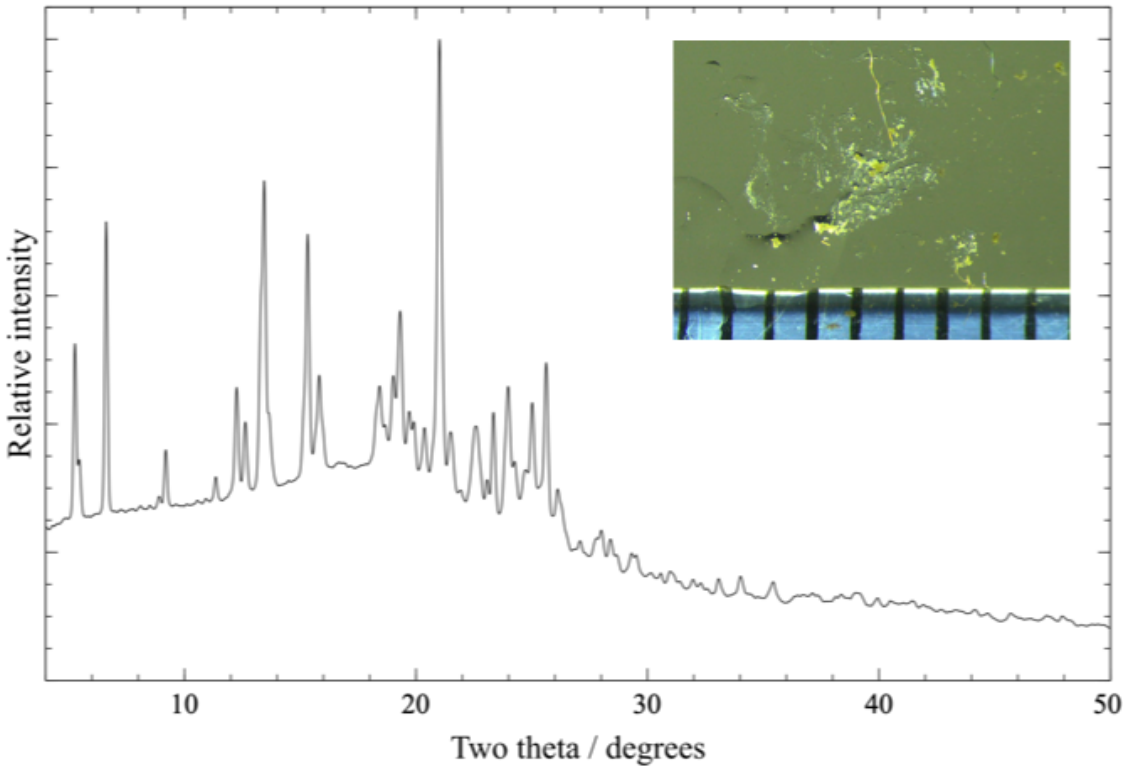


Figure 6.39. PXRD data of C11 (red), inset picture of C11 sample with 1 mm reference.

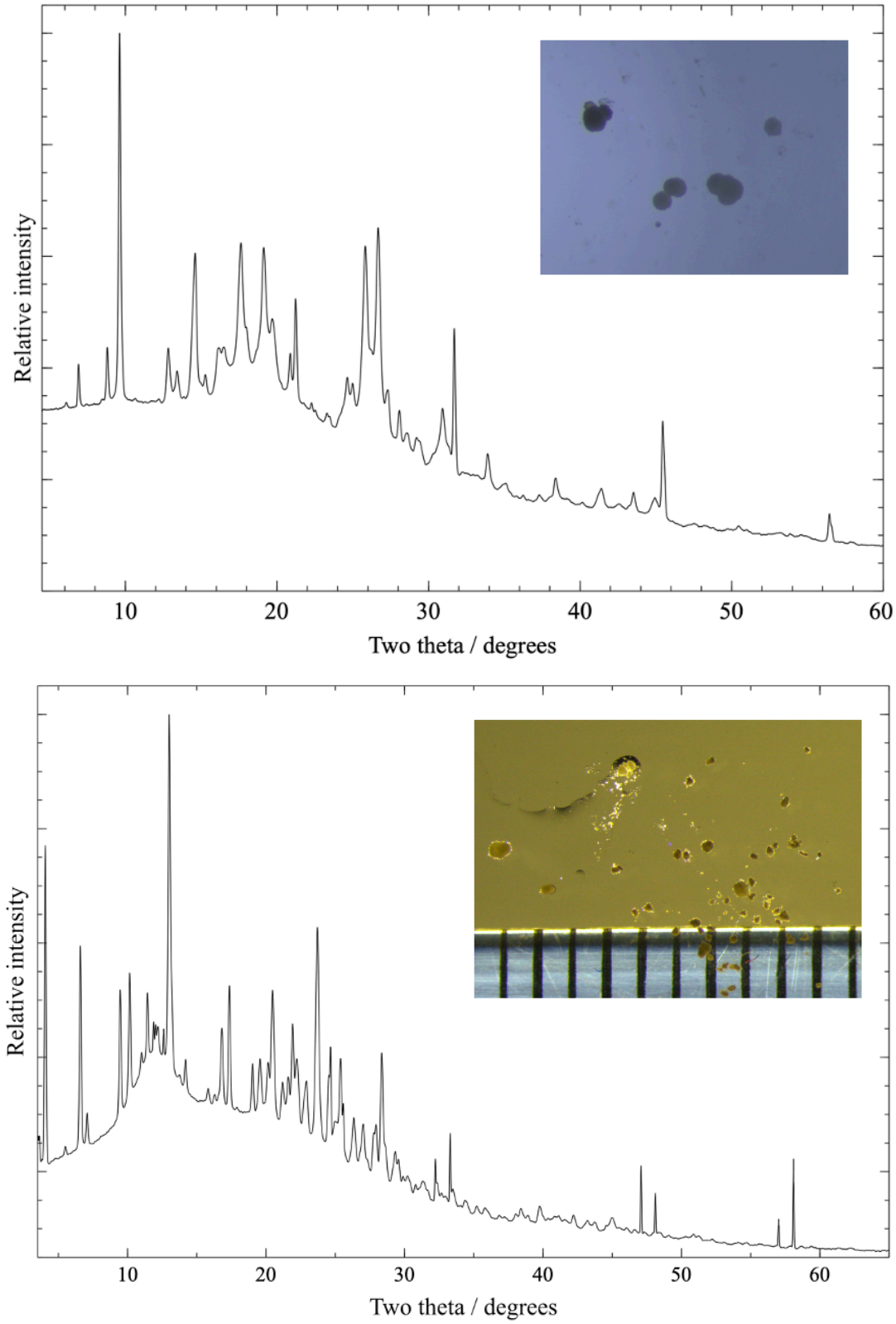


Figure 6.40. Data collections showing salt impurities that prevented structural analysis. PXRD Integrated diffraction pattern of C12 (blue) with inset picture of C12 sample, above. PXRD data collected from C18 (blue) with inset picture of C18 sample, below. C18 (blue) was sent for ED structure solution but was not able to be solved.

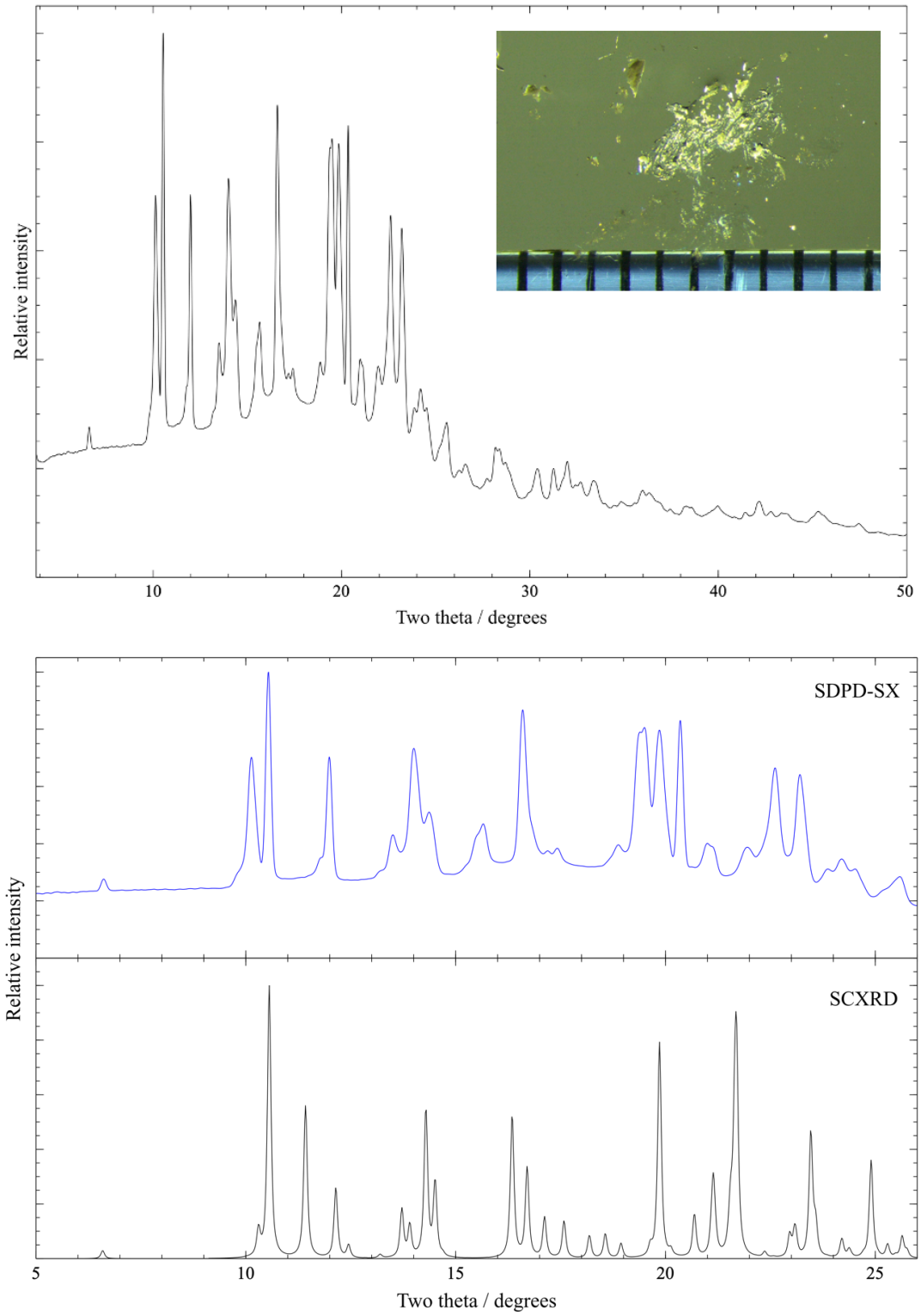


Figure 6.41. PXRD data of C16 (red), inset picture of C16 (red) sample with 1 mm reference above. Comparison to C16 (red) PXRD data compared to C16 (blue) generated PXRD pattern, below, highlighting phase difference.

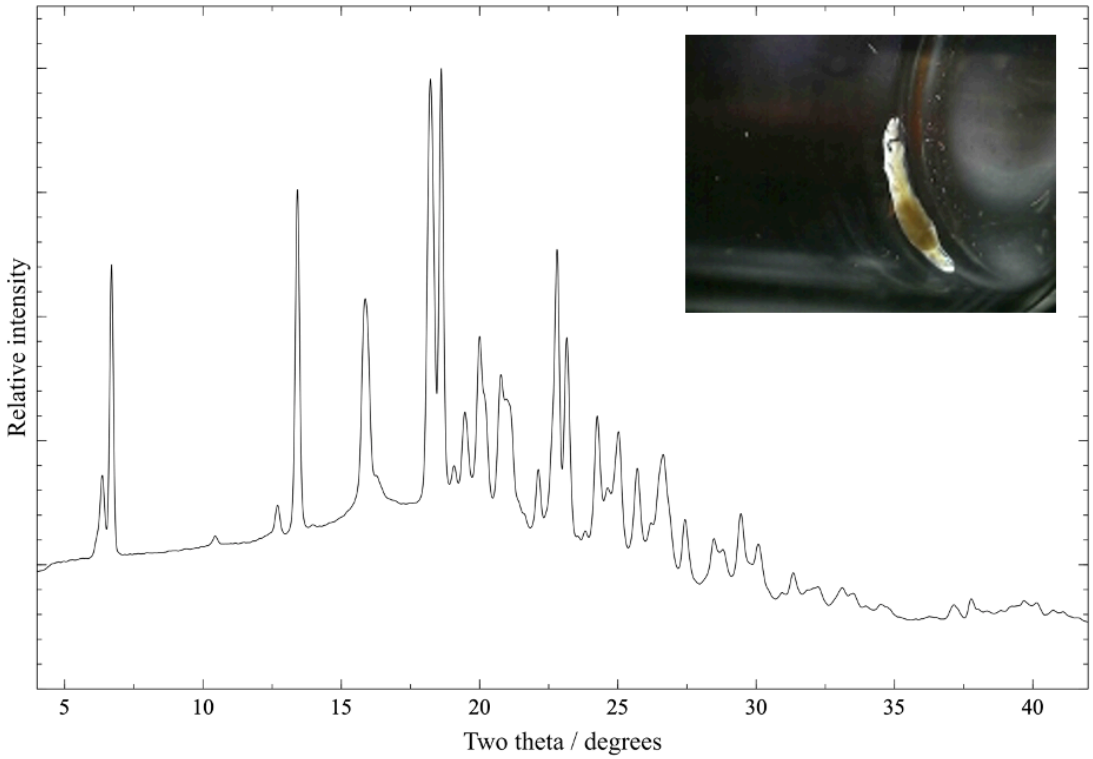


Figure 6.42. PXRD data of C22 (blue) with inset picture of C22 sample in vial.

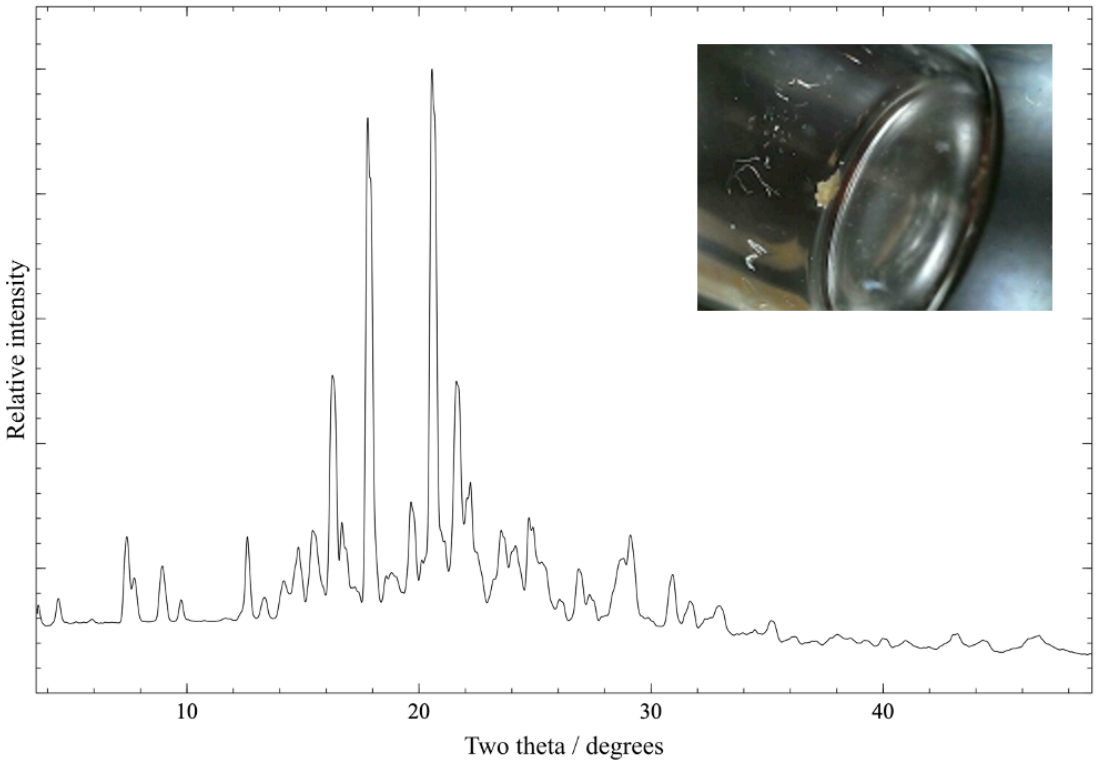


Figure 6.43. PXRD data of C23 (blue) with inset picture of C23 sample in vial.

Chapter 7

Summary and Conclusions

7.1 Key findings of the thesis

Crystal structures are widely used in drug discovery and development. In particular, a knowledge of a molecule's conformation is important for modelling, and thus predicting, how it will interact with biological systems. This information can be used to change the direction of a drug discovery programme; for example, to favour compounds that have conformations complementary to a particular protein active site. Often when an organic compound is crystallised, it does not form crystals suitable for SCXRD. Powder diffraction is a suitable alternative only when enough material (*approx.* 5-10 mg) is available. For samples that are unsuitable for SCXRD, and unsuitable for lab PXRD due to insufficient material being available, the next logical step for data collection is to use a large facility such as a synchrotron. However, the intense radiation and more accurate instrumentation available at such facilities come at a cost; beamtime applications are competitive, expensive and samples can often degrade in the intense incident beam. The work reported in this thesis aims to bridge this data collection gap by examining the viability of performing PXRD using lab-based single-crystal diffractometry on samples of polycrystalline material < 0.1 mg. Originally focussed on screening for sample crystallinity, it developed into a complete approach to structure determination; we refer to this approach as SDPD-SX.

7.1.1 Aims and outcomes

The overarching aim was to create a methodology for the collection of high quality PXRD data using SX instrumentation. This work has shown that it is possible to optimise data collection parameters for laboratory SX instrumentation in order to obtain PXRD data that permit not only simple crystallinity assessment but also full crystal structure determination. Exploring practical considerations of this method to create the best sample every time. The work has encompassed the following key steps:

1. Creating software to automate the evaluation of PXRD data generated during instrument optimisation (Chapter 2).
2. Establishing and optimising the key parameters that dictate the quality of the PXRD data collected on a lab instrument (Chapter 3).
3. Extending the principles determined in Chapter 3 to other instrumentation (Chapter 4).
4. Demonstrating the application of SDPD-SX with known pharmaceutical crystal structures and comparing SDPD-SX performance with that of standard lab-based capillary powder diffraction (Chapter 5).
5. Demonstrating the application of SDPD-SX with unknown pharmaceutical crystal structures and validating the resulting structures using DFT-D (Chapter 6).

It has been demonstrated convincingly that with careful sample preparation and relatively little machine optimisation, that a standard laboratory SX X-ray diffractometer can produce PXRD data that are sufficiently good for the crystal structure determination of moderately complex molecular materials. Some of the more complex SDPD-SX structures may not have the spatial resolution of conventional PXRD derived structures, but still provide (especially post-DFT optimisation) key structural information, such as molecular conformation, packing motifs, *etc.* Overall, SDPD-SX is certainly competitive with good quality lab-based PXRD, as evidenced by the range (unit cell size, conformational flexibility, Z') of structures solved and presented here. When compared to other known PXRD data collection techniques (transmission flat plate, reflection flat plate and transmission capillary), the use of a SX diffractometer requires *significantly* less material, and a generally shorter data collection times due to the use of an area detector. Of particular note are the previously undetermined crystal structures of 6 compounds provided by C4X Discovery Limited and solved by SDPD-SX; the sample quantities available in these cases meant that these structures could not have been solved by any other in-house method.

7.2 Main study contributions to research area

In chapter 6 over half of the materials studied that were not suitable for SCXRD were found to be crystalline, despite appearances. Often these crystalline materials, if not able to be grown into single crystals, are not studied at all. The work in this thesis demonstrates that these polycrystalline materials should be approached differently, particularly as the ideal sample morphology for SCXRD and SDPD-SX have an inverse relationship. The main contribution of this work is to change the view of these samples as something to be examined before undergoing recrystallisation.

During this work novel and known materials were studied in real-word conditions to produce materials ranging from highly crystalline to amorphous. From the full diffraction patterns collected, 32% were successfully used for structure determination. Not all the structure solution attempts were straightforward, and the work outlines various elements of structure solution that are not applicable to SCXRD or standard SDPD.

The work reported in this thesis advances the wider field of structural chemistry specifically in the following ways:

1. The process outlined allows widely-available SX instrumentation to be used to collect data of a quality that would otherwise only be easily available to those with dedicated in-house transmission capillary PXRD instrumentation.
2. It allows for the routine use of sub-milligram amounts of powder for PXRD experiments.
3. The work demonstrates crystal structure validation and troubleshooting for difficult structure solution.
4. As a consequence of (2), it expands the utility of PXRD to early-stage drug discovery programmes (as one area of research that this thesis focuses on), thus enabling crucial structural information to be obtained at a critical stage in development.
5. A large percentage of materials studied become usable, reducing the number of 'failed' crystallisation experiments and can be used to triage further analysis for example by ED.

Those who are in the best position to take advantage of SDPD-SX are those who have routine access to a single crystal diffractometer with a microfocus copper radiation source and the ability to control beam divergence; fortunately, such instrumentation is becoming increasingly common, specifically for Rigaku instruments. For those that are unfamiliar with SDPD, there is of course a learning curve associated with collecting and analysing PXRD data, but it is by no means insurmountable. On the other hand, regular PXRD users who are unfamiliar with SX instrumentation should find that collection of PXRD data on a SX instrument is relatively straightforward. The creation of open-source code and associated programs and materials (see chapter 3) aids the optimisation of equipment for all users and removes the bulk of the setup / calibration effort.

7.3 SDPD-SX as part of the wider research area

Below is a description of how SDPD-SX can be applied within the wider crystallography research landscape, in particular, fast growing methodologies and those already in use for small sample amounts.

7.3.1 Electron diffraction

Electron diffraction is described throughout this thesis in part because of its use in chapter 6 for samples that were unable to be studied by SDPD-SX. The other reason is its rapid expansion in recent years and capacity to solve crystal structures from nanocrystals. Much of the research in this area is still focused on the data collection and processing and analysis of

micro crystals but that is not to say that ED is not at present a very powerful technique. There are many examples of the use of ED to solve long standing problems; in 2019 for example, ED was used to answer a century-old question within the pharmaceutical industry, the crystal structure of orthocetamol (a regioisomer of one of the most common medicines, paracetamol). The structure had been inaccessible to X-ray diffraction due to the significant twinning, but this has not proven to be an obstacle for ED. In 2021/2, Andrusenko *et al* used ED to correct a 47-year misunderstanding of two polymorphs of indomethacin. The two structures were able to be analysed by ED using microdroplet melt crystallisation, something that was inaccessible to X-ray methods. There is no doubt that the future of ED is promising and that a variety of methodologies will be developed around it.[1,2,3,]

Regardless of which configuration that is used for ED, SDPD-SX can be used as a complimentary method. SDPD-SX can be used to screen materials for crystallinity before submitting them for (currently very expensive) electron diffraction experiments, saving researchers time and expense. Furthermore, SDPD-SX is able to capture diffraction from the bulk material, to potentially identify the presence of multiple phases, something that is not guaranteed by electron diffraction. This complimentary relationship will allow synergistic development of both techniques.

7.3.2 Micro single crystal using synchrotron radiation

This thesis focuses on the use of SDPD-SX in a lab-based setting, as this is what the majority of researchers have access to. Synchrotron serial crystallography is a very powerful tool for the study of protein crystals and continues to be the focus of many publications. Two examples of studies published in 2023 are the structure determination of 2 membrane proteins in different states using microcrystals at room temperature by Birch *et al* and the structure determination of human phosphatase PTP1B by Sharma *et al*. The information gathered from protein active sites from these studies has the potential to form the basis of drug discovery research.[4,5]

A natural progression from this thesis is to implement SDPD-SX in a synchrotron setting, where many single crystal diffractometers are installed. The strength of the incident radiation is likely to quickly destroy organic samples; however, starting from 0.5 mg of available material, a large number of individual samples could be measured, and thus data collection should not be a limitation.

7.3.3 CSP campaigns

Discussed in section 1.3.1., crystal structure prediction (CSP) is a collection of computational methods used for exploring possible crystal structures from a known compound. CSP is an ever-growing field of research, benefiting exponentially from increased understanding and supercomputer availability. In 1999 the CCDC hosted the first crystal structure prediction challenge (CSP), and since then has hosted another 6 (with the 7th being written-up for publication as the time of submission of this thesis) of increasing molecular and crystallographic complexity. The 6th CSP was reported in 2016 with 5 organic target systems to examine different computational approaches. The five examples were composed of a small rigid molecule, a co-crystal, a hydrate, an API polymorph and a bulky, flexible molecule. The starting point for this challenge was simply the chemical structure and growing conditions yet, despite the difficulty of the challenge, several participants were able to correctly predict the experimentally determined crystal structures using a variety of different techniques. The results of this blind challenge has provided new best practices and workflows with methods including DFT, molecular dynamic simulation and random search approaches.[6]

CSP is likely to have an increasing role within the pharmaceutical industry, particularly in early-stage drug discovery where the greatest impacts can be made. The most recent CSP (the 7th) also involved an example where additional information, in the form of a low-quality powder diffraction patten, was available to simulate a scenario frequently encountered when dealing with patent literature. It is clear that CSP can be used to complement the data collected by SDPD-SX, with the potential to aid structure solution where programs such as DASH are limited by peak overlap present.

7.3.3.1 Machine learning

Machine learning (ML) is a growing area of research and has shown to be beneficial for CSP. For example, the Oganov group is aiming to use machine learning to ultimately replace the DFT calculations that underpin the CSP of its USPEX code. Application of this approach in 2019 was able to demonstrate the algorithm learning interatomic potentials, leading to calculations that are orders of magnitude faster than those of traditional DFT. The application of machine learning can, and likely will, enhance many different aspects of crystallography and crystal structure determination. Machine learning opens up creative and non-traditional problem solving; for example, Kaufmann *et al* (2019) has shown how machine learning can be used in combination with electron back-scattered diffraction (EBSD) to index a crystal structure without prior knowledge of the sample. This approach was chosen because EBSD is faster and easier to perform than other traditional diffraction techniques and represents one of the future uses of machine learning in crystallography.[7,8]

In addition, Lui *et al* has demonstrated the use of machine learning with the atomic pair distribution function to aid with indexing of inorganic crystal structures. The learning model was trained using > 100,000 powder pattens, obtained from the inorganic structure crystal database (ISCD), weighted to the 45 most populated space groups. The model was able to identify the correct space group for 12 out of 15 within the top 6 predictions. It is clear this methodology is in its infancy but with the large number of datasets available to train ML models there is potential to aid more complicated, lower symmetry systems with indexing and beyond.[9] Many of the computational tools are part of a larger, open-source community and so it is likely that the use of available data collection will complement the use of powerful and low-cost structure determination methods. It is not clear what the full role of machine learning will have on powder diffraction or crystallography as a whole, but it is certain that the greatest benefits are still to come.

7.4 Limitations of the study

Sample handling for SDPD-SX requires some practice, and the main difficulty that can occur is the addition of an excess of oil to a sample. The tiny amounts of material used are easily dislodged from the sample mount when exposed to the N_2 flow of a Cryostream-type device and thus far, our experiments have been limited to data collections at room temperature. Of course, the additional requirements of SDPD still apply *i.e.,* a reasonably crystalline sample exhibiting sharp diffraction, and a knowledge of the 2D structure of the molecule under investigation.

7.5 Recommendations for future work

It is hoped that SDPD-SX will be adopted by other academic and industry-based researchers to explore the capabilities of this methodology. Work beyond this thesis has already been successful in using SDPD-SX for solving previously unreported inorganic structures, as well as small organic compounds for drug discovery projects.

Because SDPD-SX uses oil to coat the sample, this method has the potential to enable powder diffraction of air sensitive compounds that are not suitable for most reflection flat plate setups. As just outlined in section 7.4, the use of Cryostream cooling was limited in this work but given the almost universal nature of Cryostream-type devices on SX instruments, this offers one potential avenue of development. Ideally the sample preparation/ sample holder needs to produce a firmer sample so that is not likely to be lost during data collection. This would

improve data quality, particularly at higher angle and would enable highly air-sensitive compounds to be studied.

APPE (described in chapter 2) was created as a tool for this work and is open-source, but given its simple construction it could easily be implemented into existing single crystal diffraction softwares, such as CrysAlisPro. This would enable fully automated parameter optimisation, specifically accurate ring determination. In addition, many dedicated powder diffractometers incorporate automation to reduce the amount of time needed for a researcher to be present to change samples. A natural extension of this is for the manufacturers of single crystal diffractometers to explore how automation can be applied to SDPD-SX samples.

7.6 References

1 M. Lightowler, S. Li, X. Ou, X. Zou, M. Lu and H. Xu, *Angew. Chemie Int. Ed.*, , DOI:10.1002/anie.202114985.

2 I. Andrusenko, V. Hamilton, E. Mugnaioli, A. Lanza, C. Hall, J. Potticary, S. R. Hall and M. Gemmi, *Angew. Chemie Int. Ed.*, 2019, **58**, 10919–10922.

3 I. Andrusenko, V. Hamilton, A. E. Lanza, C. L. Hall, E. Mugnaioli, J. Potticary, A. Buanz, S. Gaisford, A. M. Piras, Y. Zambito, S. R. Hall and M. Gemmi, *Int. J. Pharm.*, 2021, **608**, 121067.

4 J. Birch, T. O. C. Kwan, P. J. Judge, D. Axford, P. Aller, A. Butryn, R. I. Reis, J. F. Bada Juarez, J. Vinals, R. L. Owen, E. Nango, R. Tanaka, K. Tono, Y. Joti, T. Tanaka, S. Owada, M. Sugahara, S. Iwata, A. M. Orville, A. Watts and I. Moraes, *J. Appl. Crystallogr.*, , DOI:10.1107/S1600576723006428.

5 S. Sharma, A. Ebrahim and D. A. Keedy, *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, 2023, **79**, 23–30.

6 A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio Jr, A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu and C. R. Groom, *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.*, 2016, **72**, 439–459.

7 E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev and A. R. Oganov, *Phys. Rev. B*, 2019, **99**, 064114.

8 K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, T. Harrington, E. Marin and K. Vecchio, *Microsc. Microanal.*, 2019, **25**, 2258–2259.

9 C.-H. Liu, Y. Tao, D. Hsu, Q. Du and S. J. L. Billinge, *Acta Crystallogr. Sect. A Found. Adv.*, 2019, **75**, 633–643.